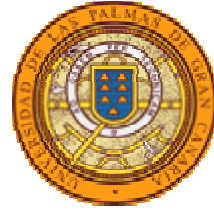




Fac. CC. Económicas y Empresariales
Universidad de La Laguna



Fac. CC. Económicas y Empresariales
Univ. de Las Palmas de Gran Canaria

Representación en el espacio de los estados y filtro de Kalman en el contexto de las series temporales económicas

Gloria Martín Rodríguez *

DOCUMENTO DE TRABAJO 2002-05

* Universidad de La Laguna. Departamento de Economía de las Instituciones,
Estadística Económica y Econometría.

REPRESENTACIÓN EN EL ESPACIO DE LOS ESTADOS Y FILTRO DE KALMAN EN EL CONTEXTO DE LAS SERIES TEMPORALES ECONÓMICAS

Gloria Martín Rodríguez

Departamento de Economía de las Instituciones, Estadística Económica y Econometría

Facultad de CC. EE. y Empresariales

Universidad de La Laguna

Camino de la Hornera, s/n

Campus de Guajara 38202

Teléfono: 922 317 034

Fax: 922 317 042

gmartinr@ull.es

Agradecimientos: Este trabajo no se hubiera llevado a cabo sin la ayuda y el apoyo de los profesores D. José Juan Cáceres Hernández y D. Julio Afonso Rodríguez, profesores de la UDI de Estadística y Econometría.

REPRESENTACIÓN EN EL ESPACIO DE LOS ESTADOS Y FILTRO DE KALMAN EN EL CONTEXTO DE LAS SERIES TEMPORALES ECONÓMICAS

Resumen

El tratamiento de los modelos estructurales de series temporales se encuadra en el marco más general de modelos en el espacio de los estados. La representación en el espacio de los estados es importante para el análisis estadístico de series temporales debido a que ofrece la posibilidad de recurrir al filtro de Kalman, que permite estimar cada uno de los componentes de la serie y evaluar la función de verosimilitud de una forma simple y directa. Desde el punto de vista práctico, otra ventaja de la representación en el espacio de los estados es su generalidad, ya que son muchos los modelos de series temporales gaussianos que pueden ser representados en el espacio de los estados.

El objetivo de este trabajo es describir la representación en el espacio de los estados, el filtro de Kalman y los algoritmos de suavizados, así como la estimación máximo verosímil en el dominio del tiempo. También se examina el papel que juegan las condiciones iniciales y las observaciones anómalas en el funcionamiento del filtro de Kalman. Finalmente, se expone la aplicación de esta metodología en el contexto de los modelos estructurales.

Clasificación JEL: C22.

Palabras claves: modelos en el espacio de los estados, filtro de Kalman, modelos estructurales.

Abstract

The key to handling structural time series models is the state space form. The importance of the state space model for the statistical analysis of time series is primarily due to the Kalman filter and its relation to the prediction error decomposition which allows the likelihood function of state space models to be evaluated in a simple and straightforward way. Another feature of the state space framework is its generality: most practical gaussian time series models can be formulated as a state space model.

The goal of this paper is to describe the state space form, the Kalman filter and smoothing algorithms and the maximum likelihood estimation in the time domain. Also, the role of initial conditions and outliers for the Kalman filter is explored. Finally, the procedure is applied to the most important structural models.

JEL classification: C22.

Key words: state space models, filtro de Kalman, structural models.

Representación en el espacio de los estados y filtro de Kalman

El tratamiento estadístico de los modelos estructurales de series temporales está basado en su representación en el espacio de los estados o *SSF*¹, una herramienta de enorme potencial que permite el tratamiento de un amplio rango de modelos de series temporales². Los modelos en el espacio de los estados surgieron en el campo de la ingeniería. El artículo de Kalman (1960) significó un importante salto cualitativo, ya que mostró que una extensa clase de problemas podía representarse mediante un modelo lineal simple en el espacio de los estados, cuya naturaleza markoviana permite que los cálculos necesarios para su aplicación práctica se efectúen de forma recursiva. Estas ideas se desarrollaron posteriormente en el ámbito de investigación de los ingenieros³; mientras que las contribuciones de estadísticos y econométricos fueron esporádicas hasta principios de los 80, aunque en los últimos años el interés mostrado por estos últimos ha crecido notablemente⁴.

El estado o situación de un sistema en un momento dado del tiempo se describe a través de un conjunto de variables que forman el llamado vector de estado. El espacio de los estados es, por tanto, el espacio donde los sucesivos vectores de estado describen la evolución del sistema como función del tiempo. En un modelo estructural de series temporales los elementos del vector de estado son los componentes no observables de la serie.

Una vez que el modelo ha sido formulado en el espacio de los estados, el filtro de Kalman proporciona el medio de estimar el estado inobservable a partir de alguna magnitud observable relacionada con éste, de forma que la estimación se actualice cada vez que se dispone de nueva información. Se pueden emplear varios algoritmos de alisado, que constituyen extensiones de la formulación original diseñada por Kalman para obtener una estimación óptima del estado en cualquier punto de la muestra.

Supóngase un sistema descrito por un vector de estado m -dimensional \mathbf{a}_t , en los instantes del tiempo $t = 0, 1, 2, \dots, T$, como por ejemplo una determinada economía nacional a lo largo de

¹ *State space form*.

² Concretamente, el espacio de los estados es una representación útil en el caso de magnitudes que puedan expresarse como suma de procesos estocásticos o que son observadas con errores de medida (véase Hamilton, 1994:376).

³ Cuando se publica el famoso artículo de Kalman, la metodología que en él se describe se estaba utilizando en la industria aeroespacial para estimar la posición de los planetas (véase Jones, 1993:77).

⁴ Véase Durbin y Koopman (2001:52).

diferentes años. Las variables de estado que describen la economía nacional no se observan en un estado puro sino que están contaminadas por un componente ruidoso⁵, de modo que es posible que lo realmente observado sea y_t , siendo

$$y_t = \mathbf{a}_t + \mathbf{e}_t,$$

donde \mathbf{e}_t es el componente ruidoso, o, de forma más general, se observa un vector $N \times 1$ de variables, y_t , relacionadas con el vector de estado a través de la expresión

$$y_t = \mathbf{Z}_t \mathbf{a}_t + \mathbf{e}_t,$$

donde \mathbf{Z}_t es alguna transformación lineal. Pues bien, si se desea estimar el valor del estado, magnitud que evoluciona en el tiempo y que no se observa directamente sino contaminada por dicho componente ruidoso, se puede recurrir a los denominados filtros adaptativos, que poseen características de autoajuste, es decir, permiten corregir la estimación cuando se recibe nueva información. Uno de estos filtros fue desarrollado por Kalman (1960) para el caso de sistemas dinámicos en tiempo discreto⁶.

El propósito del análisis en el espacio de los estados es inferir las propiedades relevantes de los estados $\mathbf{a}_1, \dots, \mathbf{a}_T$ a partir de los valores observados y_1, \dots, y_T de la magnitud relacionada con el estado y del conocimiento de las propiedades estadísticas del ruido. Para ello, es preciso establecer, además, una relación entre los estados en los momentos t y $t-1$, es decir, entre \mathbf{a}_t y \mathbf{a}_{t-1} . En la propuesta de Kalman se asume que

$$\mathbf{a}_t = \mathbf{S}_t \mathbf{a}_{t-1} + \mathbf{u}_t,$$

donde \mathbf{S}_t es alguna transformación lineal que describe la evolución en el tiempo del sistema, \mathbf{u}_t es un término de ruido y se asume que \mathbf{e}_t y \mathbf{u}_t son vectores de variables aleatorias gaussianas, de media nula y serialmente incorreladas. Kalman concluye que la mejor estimación para el estado \mathbf{a}_t puede ser obtenida recursivamente a partir de la mejor estimación de \mathbf{a}_{t-1} y su matriz de covarianzas.

A continuación, se hace una exposición más detallada de los aspectos fundamentales de la representación en el espacio de los estados y del filtro de Kalman.

⁵ "... los sistemas dinámicos están dirigidos no sólo por inputs sometidos a nuestro propio control, sino también por perturbaciones que no podemos controlar ni modelar determinísticamente" (Maybeck, 1979:2).

⁶ Kalman y Bucy (1961) desarrollaron el filtro para el caso de sistemas dinámicos en tiempo continuo. El filtro de Kalman se puede aplicar tanto para magnitudes estacionarias como para no estacionarias. Hasta su formulación por Kalman y Bucy en 1960-1961, se aplicaba otro filtro adaptativo, desarrollado por Wiener en los años 40, que sólo es válido para magnitudes estacionarias.

2.1. Representación en el espacio de los estados

Sea un sistema descrito por un vector de estado $m \times 1$, \mathbf{a}_t , en los instantes del tiempo $t = 1, \dots, T$ ⁷ y sea \mathbf{y}_t un vector $N \times 1$ de variables observables. La formulación básica del modelo en el espacio de los estados consta de dos ecuaciones. Bajo el supuesto de que el estado sigue un proceso de Markov de primer orden, la *ecuación de transición* establece una relación entre el estado actual y el estado en el instante del tiempo anterior, es decir,

$$\mathbf{a}_t = \mathbf{S}_t \mathbf{a}_{t-1} + \mathbf{c}_t + \mathbf{R}_t \mathbf{u}_t, \quad t = 1, \dots, T,$$

donde \mathbf{S}_t es una matriz $m \times m$, \mathbf{c}_t es un vector $m \times 1$, \mathbf{R}_t es una matriz $r \times m$ ($r \leq m$) y \mathbf{u}_t es un vector $r \times 1$ de perturbaciones no correlacionadas serialmente⁸ con media cero y matriz de covarianzas \mathbf{Q}_t . Dado que el estado es inobservable, es preciso establecer una relación entre el vector $N \times 1$ de variables observables y el vector de estado $m \times 1$ que se define a través de la *ecuación de medida*

$$\mathbf{y}_t = \mathbf{Z}_t \mathbf{a}_t + \mathbf{d}_t + \mathbf{e}_t, \quad t = 1, \dots, T,$$

donde \mathbf{Z}_t es una matriz $N \times m$, \mathbf{d}_t es un vector $N \times 1$ y \mathbf{e}_t es un vector $N \times 1$ de perturbaciones no correlacionadas serialmente⁹ con media nula y matriz de covarianzas \mathbf{H}_t ¹⁰. La especificación de la formulación en el espacio de los estados requiere dos supuestos adicionales:

⁷ Aunque hay muchas formas equivalentes de expresar un sistema en la representación *SSF*, cada una de ellas proporciona los mismos valores de la función de verosimilitud y, por tanto, las mismas predicciones del sistema. Así, resulta deseable definir \mathbf{a}_t de forma que contenga toda la información relevante del sistema en el momento t con el menor número de elementos. De modo que, si las demás condiciones se mantienen, se prefiere aquella especificación que posea una realización mínima, es decir, aquella que minimice la longitud del vector de estado. Téngase en cuenta que cualquier problema que se pretenda formular no tiene necesariamente una representación única.

⁸ En muchas aplicaciones $\mathbf{R}_t = \mathbf{I}_m$, con lo que la ecuación de transición puede escribirse como $\mathbf{a}_t = \mathbf{S}_t \mathbf{a}_{t-1} + \mathbf{c}_t + \mathbf{u}_t$, donde \mathbf{u}_t es un vector de dimensión $m \times 1$. Pero en otras conviene definir \mathbf{R}_t como una matriz $r \times m$, con $r < m$. Por ejemplo, Durbin y Koopman (2001:38) asumen que \mathbf{R}_t es un subconjunto de columnas de \mathbf{I}_m , de modo que se trata de una matriz que selecciona aquellas filas del conjunto de ecuaciones de estado con términos de perturbación distintos de cero. En ocasiones, la ecuación de transición se expresa como $\mathbf{a}_{t+1} = \mathbf{S}_t \mathbf{a}_t + \mathbf{c}_t + \mathbf{u}_t$, $t = 1, \dots, T$, (Jazwinski, 1970, y Anderson y Moore, 1979) y, aunque desde el punto de vista práctico existen pocas diferencias, es necesario tener cuidado en ciertos casos como, por ejemplo, aquéllos en los que existe correlación entre las perturbaciones de la ecuación de medida y de transición.

Nótese también que si $m = 1$, \mathbf{a}_t es una variable de estado unidimensional, \mathbf{S}_t es un coeficiente para cada instante del tiempo y \mathbf{u}_t es un término de perturbación con elementos serialmente incorrelados con media cero y varianza q_t .

⁹ Nótese que la incorrelación serial de los elementos del vector de perturbaciones implica la ausencia de correlación serial para cada uno de los elementos del vector así como para dos elementos cualesquiera del vector.

¹⁰ Si $N = 1$, \mathbf{y}_t es una variable unidimensional, \mathbf{Z}_t es un vector $1 \times m$ de coeficientes para cada instante del tiempo y \mathbf{e}_t es un término de perturbación con media cero y varianza h_t y, además, los elementos de este proceso estocástico están serialmente incorrelados.

1) Respecto a las condiciones iniciales, el vector de estado inicial, \mathbf{a}_0 , tiene media \mathbf{a}_0 y matriz de covarianzas \mathbf{P}_0 , es decir,

$$E[\mathbf{a}_0] = \mathbf{a}_0$$

y

$$Var[\mathbf{a}_0] = \mathbf{P}_0.$$

2) Los términos de perturbación de la ecuación de medida y de la ecuación de transición están incorrelados entre sí y con el vector de estado inicial en cualquier momento del tiempo, es decir,

$$E[\mathbf{e}_t \mathbf{u}'_s] = \mathbf{0}, \forall s, t = 1, \dots, T,$$

$$E[\mathbf{e}_t \mathbf{a}'_0] = \mathbf{0}, t = 1, \dots, T,$$

y

$$E[\mathbf{u}_t \mathbf{a}'_0] = \mathbf{0}, t = 1, \dots, T \quad ^{11}.$$

Las matrices \mathbf{Z}_t , \mathbf{d}_t , \mathbf{H}_t , \mathbf{S}_t , \mathbf{c}_t , \mathbf{R}_t y \mathbf{Q}_t se denominan matrices de sistema y, a menos que se especifique lo contrario, se consideran no estocásticas, aunque podrían cambiar a lo largo del tiempo de forma predeterminada. De esta manera, el sistema es lineal y para cualquier momento del tiempo, el vector de variables observables, \mathbf{y}_t , se puede expresar como una combinación lineal de valores presentes y pasados de perturbaciones, \mathbf{u}_t y \mathbf{e}_t , y del vector de estado inicial, \mathbf{a}_0 . Si las matrices del sistema no cambian con el tiempo, el modelo se denomina homogéneo o invariante respecto al tiempo¹² y los modelos no estacionarios constituyen casos particulares que, siendo invariantes respecto al tiempo, se transforman en estacionarios al aplicarles el operador diferencia.

Las matrices del sistema \mathbf{Z}_t , \mathbf{H}_t , \mathbf{S}_t , \mathbf{R}_t y \mathbf{Q}_t pueden depender de un conjunto de parámetros desconocidos que determinan las propiedades estocásticas del modelo. Se denominan hiperparámetros para distinguirlos de los que intervienen en los vectores \mathbf{d}_t y \mathbf{c}_t , que sólo afectarán al valor esperado del estado y las observaciones de forma determinística¹³.

¹¹ Este supuesto se puede relajar dando lugar a modificaciones en varios de los algoritmos.

¹² En un modelo homogéneo respecto al tiempo la ecuación de transición es un proceso VAR(1), pero este hecho no constituye un supuesto restrictivo debido a que procesos de órdenes superiores pueden ser tomados en consideración mediante una representación de Markov. Por otro lado, un modelo no tiene que ser autorregresivo para que pueda ser representado en el espacio de los estados sino que, por el contrario, cualquier proceso estocástico admite dicha representación.

¹³ Esta conclusión no se sostiene si \mathbf{d}_t es función de valores retardados de \mathbf{y}_t . Por otra parte, como señalan Durbin y Koopman (2001:30), los parámetros de modelos en el espacio de los estados se suelen denominar hiperparámetros

Si se adopta una representación simplificada en el espacio de los estados de la forma:

$$\mathbf{a}_t = \mathbf{S}\mathbf{a}_{t-1} + \mathbf{u}_t,$$

$$\mathbf{y}_t = \mathbf{Z}_t\mathbf{a}_t + \mathbf{e}_t,$$

donde \mathbf{a}_t es un vector de estado $m \times 1$, \mathbf{S} es una matriz de transición $m \times m$, \mathbf{u}_t es el vector $m \times 1$ de perturbaciones de la ecuación de transición tal que $\mathbf{u}_t \sim i.i.d. N_m(\mathbf{0}, \mathbf{Q})$, \mathbf{y}_t es un vector $N \times 1$ de variables observables, \mathbf{Z}_t es una matriz de medida $N \times m$ y \mathbf{e}_t es el vector $N \times 1$ de perturbaciones de la ecuación de medida tal que $\mathbf{e}_t \sim i.i.d. N_N(\mathbf{0}, \mathbf{H})$, entonces, por sustituciones recursivas, se tiene que

$$\mathbf{a}_t = \mathbf{S}^t \mathbf{a}_0 + \sum_{i=0}^{t-1} \mathbf{S}^i \mathbf{u}_{t-i}$$

y, por tanto,

$$\mathbf{y}_t = \mathbf{Z}_t \left[\mathbf{S}^t \mathbf{a}_0 + \sum_{i=0}^{t-1} \mathbf{S}^i \mathbf{u}_{t-i} \right] + \mathbf{e}_t.$$

A partir de estas igualdades, las propiedades estadísticas del modelo son¹⁴:

- 1) $E[\mathbf{a}_t] = \mathbf{S}^t E[\mathbf{a}_0] = \mathbf{S}^t \mathbf{a}_0$
- 2) $E[\mathbf{y}_t] = \mathbf{Z}_t \mathbf{S}^t E[\mathbf{a}_0] = \mathbf{Z}_t \mathbf{S}^t \mathbf{a}_0$
- 3) $E\left[(\mathbf{a}_t - E[\mathbf{a}_t])(\mathbf{a}_{t-k} - E[\mathbf{a}_{t-k}])'\right] = \left[\mathbf{S}^t \mathbf{P}_0 (\mathbf{S}^{t-k})' + \sum_{i=0}^{t-k-1} \mathbf{S}^{k+i} \mathbf{Q} (\mathbf{S}^i)'\right] = \mathbf{P}_{t,t-k},$

con $\mathbf{P}_{t,t} \equiv \mathbf{P}_t$.

$$4) E\left[(\mathbf{y}_t - E[\mathbf{y}_t])(\mathbf{y}_{t-k} - E[\mathbf{y}_{t-k}])'\right] = \left[\mathbf{Z}_t \mathbf{S}^t \mathbf{P}_0 (\mathbf{Z}_{t-k} \mathbf{S}^{t-k})' + \mathbf{Z}_t \sum_{i=0}^{t-k-1} \mathbf{S}^{k+i} \mathbf{Q} (\mathbf{Z}_{t-k} \mathbf{S}^i)'\right],$$

si $k \neq 0$

$$5) E\left[(\mathbf{y}_t - E[\mathbf{y}_t])(\mathbf{y}_t - E[\mathbf{y}_t])'\right] = \left[\mathbf{Z}_t \mathbf{S}^t \mathbf{P}_0 (\mathbf{Z}_t \mathbf{S}^t)'\right] + \mathbf{Z}_t \sum_{i=0}^{t-1} \mathbf{S}^i \mathbf{Q} (\mathbf{Z}_t \mathbf{S}^i)'\right] + \mathbf{H}$$

Nótese que la estacionariedad en media del estado se produce si $\mathbf{a}_0 = 0$, o bien, si \mathbf{S} posee algún valor propio unitario tal que $\mathbf{S}^t \mathbf{a}_0$ es constante. Por otra parte, se puede demostrar que los momentos de segundo orden son invariantes respecto a t si y sólo si todos los autovalores de \mathbf{S}

con objeto de distinguirlos de los elementos de los vectores de estado, que pueden ser considerados como parámetros aleatorios.

caen dentro del círculo de radio unidad (Steyn, 1996:10).

Finalmente, la función de verosimilitud de los T vectores de observaciones se puede obtener del modo siguiente. Sean los vectores de vectores $\mathbf{y} = (\mathbf{y}_1, \dots, \mathbf{y}_T)'$, $\mathbf{a} = (\mathbf{a}_1, \dots, \mathbf{a}_T)'$ y $\mathbf{e} = (\mathbf{e}_1, \dots, \mathbf{e}_T)'$, y sea la matriz $\mathbf{Z} = \text{diag}(\mathbf{Z}_1, \dots, \mathbf{Z}_T)$, entonces,

$$\mathbf{y} = \mathbf{Z}\mathbf{a} + \mathbf{e},$$

donde

$$\mathbf{e} \sim N_{NT}(\mathbf{0}, (\mathbf{I}_T \otimes \mathbf{H})).$$

Por otro lado, se tiene que

$$\mathbf{a} = \mathbf{S}_1(\mathbf{i}_T \otimes \mathbf{a}_0) + \mathbf{S}_2\mathbf{u},$$

donde $\mathbf{S}_1 = \text{diag}(\mathbf{S}, \mathbf{S}^2, \dots, \mathbf{S}^T)$, \mathbf{i}_T es un vector unitario $T \times 1$,

$$\mathbf{S}_2 = \begin{bmatrix} \mathbf{I} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{S} & \mathbf{I} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{S}^{T-1} & \mathbf{S}^{T-2} & \dots & \mathbf{I} \end{bmatrix}$$

y

$$\mathbf{u} \sim N_{mT}(\mathbf{0}, (\mathbf{I}_T \otimes \mathbf{Q})).$$

Por tanto,

$$E[\mathbf{y}] = \mathbf{Z} \left[(\mathbf{S}\mathbf{a}_0)', \dots, (\mathbf{S}^T\mathbf{a}_0)' \right]' \equiv \mathbf{m},$$

mientras que, dado que,

$$\mathbf{y} = \mathbf{Z}[\mathbf{S}_1(\mathbf{i}_T \otimes \mathbf{a}_0) + \mathbf{S}_2\mathbf{u}] + \mathbf{e},$$

se obtiene que

$$E[(\mathbf{y} - \mathbf{m})(\mathbf{y} - \mathbf{m})'] = [\mathbf{Z}\mathbf{S}_1(\mathbf{i}_T\mathbf{i}_T' \otimes \mathbf{P}_0)\mathbf{S}_1'\mathbf{Z}' + \mathbf{Z}\mathbf{S}_2(\mathbf{I}_T \otimes \mathbf{Q})\mathbf{S}_2'\mathbf{Z}' + (\mathbf{I}_T \otimes \mathbf{H})] \equiv \mathbf{S}.$$

Como \mathbf{y} es normal multivariante de dimensión NT , con media \mathbf{m} y varianza \mathbf{S} , la función de verosimilitud y el logaritmo natural de ésta son

$$L(\mathbf{y}) = \frac{\sqrt{|\mathbf{S}|^{-1}}}{(2\pi)^{NT/2}} \exp\left\{-\frac{1}{2}(\mathbf{y} - \mathbf{m})\mathbf{S}^{-1}(\mathbf{y} - \mathbf{m})'\right\}$$

¹⁴ En el anexo I se pueden encontrar las demostraciones de estas propiedades.

y

$$\ln L(\mathbf{y}) = -\frac{1}{2} \ln |\mathbf{S}| - \frac{NT}{2} \ln(2\pi) - \frac{1}{2} (\mathbf{y} - \mathbf{m}) \mathbf{S}^{-1} (\mathbf{y} - \mathbf{m})',$$

respectivamente.

Para ilustrar los desarrollos del apartado siguiente, se tomará como punto de partida esta última representación simplificada del modelo en el espacio de los estados.

2.2. Filtro de Kalman y algoritmos de alisado

Una vez que un modelo ha sido formulado en el espacio de los estados, su aplicación a series temporales reales exige calcular cantidades tales como las medias y varianzas de las variables $\mathbf{a}_t | \mathbf{y}_1, \dots, \mathbf{y}_{t-1}$, $\mathbf{a}_t | \mathbf{y}_1, \dots, \mathbf{y}_t$ y $\mathbf{a}_t | \mathbf{y}_1, \dots, \mathbf{y}_T$, y ajustar el modelo a los datos obteniendo estimaciones máximo verosímiles de los parámetros desconocidos, tales como las varianzas de los términos de perturbación de las diferentes ecuaciones. En principio, como señalan Durbin y Koopman (2001:10-11), esta tarea puede ser abordada usando los métodos de inferencia clásica con el apoyo de los resultados de la teoría de la probabilidad respecto a la normal multivariante. Sin embargo, como consecuencia de la correlación serial entre las observaciones, la computación se complica notablemente cuando aumenta el número de éstas. En este sentido, las técnicas de filtrado y alisado constituyen algoritmos de computación que proporcionan los mismos resultados de forma más eficiente.

En concreto, el algoritmo conocido como filtro de Kalman permite obtener información sobre el estado \mathbf{a}_t a partir de la magnitud observada y_t relacionada con éste desde dos perspectivas diferentes, en función del nivel de información que se pretende alcanzar. La primera tiene por objeto encontrar un estimador insesgado del estado que minimice el error cuadrático medio (Harvey, 1989:104 y ss.). En la segunda, el objetivo se formula en términos más amplios, ya que se pretende obtener la distribución del estado \mathbf{a}_t condicionada al conjunto de información disponible $\mathbf{I}_t : \{\mathbf{y}_1, \dots, \mathbf{y}_t\}$, es decir, $f(\mathbf{a}_t | \mathbf{I}_t)$ (Steyn, 1996:14 y ss.). Los resultados de ambas aproximaciones no son esencialmente diferentes, dado que el estimador insesgado de error cuadrático medio mínimo viene dado, precisamente, por la media de la distribución condicionada

obtenida bajo el supuesto de normalidad (Steyn, 1996:23)¹⁵.

Para obtener la distribución del estado \mathbf{a}_t condicionada al conjunto de observaciones $I_t : \{y_1, \dots, y_t\}$, la representación simplificada del modelo en el espacio de los estados, dada al final del apartado anterior, se puede expresar como:

$$\text{Ecuación de transición: } f(\mathbf{a}_t | \mathbf{a}_{t-1}) \sim N_m(S\mathbf{a}_{t-1}, \mathbf{Q})$$

$$\text{Ecuación de medida: } f(y_t | \mathbf{a}_t) \sim N_N(\mathbf{Z}_t \mathbf{a}_t, \mathbf{H})$$

Dado que los términos de perturbación de las ecuaciones de medida y de transición están serialmente incorrelados e incorrelados entre sí, es decir,

$$E[\mathbf{e}_t \mathbf{e}_s'] = \mathbf{0}, t \neq s;$$

$$E[\mathbf{u}_t \mathbf{u}_s'] = \mathbf{0}, t \neq s;$$

$$E[\mathbf{e}_t \mathbf{u}_s'] = \mathbf{0}, \forall s, t = 1, \dots, T,$$

entonces,

$$f(\mathbf{a}_t | \mathbf{a}_{t-1}, y_1, \dots, y_{t-1}) = f(\mathbf{a}_t | \mathbf{a}_{t-1}),$$

es decir, conocido \mathbf{a}_{t-1} , la historia pasada de y_t , $\{y_1, \dots, y_{t-1}\}$, no aporta información adicional sobre \mathbf{a}_t , y

$$f(y_t | \mathbf{a}_t, y_1, \dots, y_{t-1}) = f(y_t | \mathbf{a}_t),$$

es decir, conocido \mathbf{a}_t , la historia pasada de y_t , $\{y_1, \dots, y_{t-1}\}$, no aporta información adicional sobre y_t .

El objetivo del filtro es obtener $f(\mathbf{a}_t | I_t)$, pero, para ello, conviene obtener primero $f(\mathbf{a}_t | y_t)$, lo que exige conocer $f(\mathbf{a}_{t-1} | \mathbf{a}_t)$, que se puede expresar como

$$f(\mathbf{a}_{t-1} | \mathbf{a}_t) = \frac{f(\mathbf{a}_{t-1}, \mathbf{a}_t)}{f(\mathbf{a}_t)} = \frac{f(\mathbf{a}_t | \mathbf{a}_{t-1}) f(\mathbf{a}_{t-1})}{f(\mathbf{a}_t)}.$$

Sabiendo que

$$f(\mathbf{a}_t | \mathbf{a}_{t-1}) \sim N_m(S\mathbf{a}_{t-1}, \mathbf{Q})$$

¹⁵ Téngase en cuenta que en el segundo planteamiento formulado se está adoptando un enfoque bayesiano. Y, una vez obtenida la distribución de \mathbf{a}_t condicionada a I_t , la estimación óptima podría ser tanto la media, la moda como la mediana, pero, bajo el supuesto de normalidad, las tres posibilidades coinciden.

$$f(\mathbf{a}_{t-1}) \sim N_m(\mathbf{S}^{t-1}\mathbf{a}_0, \mathbf{P}_{t-1})$$

y

$$f(\mathbf{a}_t) \sim N_m(\mathbf{S}^t\mathbf{a}_0, \mathbf{P}_t),$$

sustituyendo y reagrupando términos, se obtiene que

$$f(\mathbf{a}_{t-1}|\mathbf{a}_t) \sim N_m\left([\mathbf{P}_{t-1}^{-1} + \mathbf{S}'\mathbf{Q}^{-1}\mathbf{S}]^{-1}[\mathbf{S}'\mathbf{Q}^{-1}\mathbf{a}_t + \mathbf{P}_{t-1}^{-1}\mathbf{S}^{t-1}\mathbf{a}_0], [\mathbf{P}_{t-1}^{-1} + \mathbf{S}'\mathbf{Q}^{-1}\mathbf{S}]^{-1}\right)$$

(Steyn, 1996:15)¹⁶.

Además se sabe que

$$\begin{bmatrix} \mathbf{a}_t \\ \mathbf{y}_t \end{bmatrix} \sim N\left(\begin{bmatrix} \mathbf{S}^t\mathbf{a}_0 \\ \mathbf{Z}_t\mathbf{S}^t\mathbf{a}_0 \end{bmatrix}, \begin{bmatrix} \mathbf{P}_t & \mathbf{P}_t\mathbf{Z}_t' \\ \mathbf{Z}_t\mathbf{P}_t & \mathbf{Z}_t\mathbf{P}_t\mathbf{Z}_t' + \mathbf{H} \end{bmatrix}\right).$$

Nótese que

$$\begin{aligned} Cov[\mathbf{a}_t, \mathbf{y}_t] &= E\left[(\mathbf{a}_t - E[\mathbf{a}_t])(\mathbf{y}_t - E[\mathbf{y}_t])'\right] \\ &= E\left[(\mathbf{a}_t - E[\mathbf{a}_t])(\mathbf{Z}_t(\mathbf{a}_t - E[\mathbf{a}_t]) + \mathbf{e}_t)'\right] \\ &= E\left[(\mathbf{a}_t - E[\mathbf{a}_t])(\mathbf{a}_t - E[\mathbf{a}_t])' \mathbf{Z}_t'\right] + E[(\mathbf{a}_t - E[\mathbf{a}_t])\mathbf{e}_t'] = \mathbf{P}_t\mathbf{Z}_t' \end{aligned}$$

teniendo en cuenta que $E[(\mathbf{a}_t - E[\mathbf{a}_t])\mathbf{e}_t'] = \mathbf{0}$, dado que $\mathbf{a}_t - E[\mathbf{a}_t]$ es, por sustitución recursiva, función de \mathbf{a}_0 y retardos de \mathbf{u}_t .

Entonces, y dado que tanto las perturbaciones como el vector de estado inicial están distribuidos normalmente, la distribución de $\mathbf{a}_t|\mathbf{y}_t$ es también normal, de forma que queda totalmente caracterizada a partir del vector de medias y la matriz de covarianzas. En concreto, y dadas las propiedades de una distribución normal multivariante¹⁷, $f(\mathbf{a}_t|\mathbf{y}_t)$ sigue una distribución normal m -variante con media

$$E[\mathbf{a}_t|\mathbf{y}_t] = \mathbf{S}^t\mathbf{a}_0 + \mathbf{P}_t\mathbf{Z}_t'(\mathbf{Z}_t\mathbf{P}_t\mathbf{Z}_t' + \mathbf{H})^{-1}(\mathbf{y}_t - \mathbf{Z}_t\mathbf{S}^t\mathbf{a}_0)$$

y varianza

$$Var[\mathbf{a}_t|\mathbf{y}_t] = [\mathbf{P}_t^{-1} + \mathbf{Z}_t'\mathbf{H}^{-1}\mathbf{Z}_t]^{-1}.$$

¹⁶ Si \mathbf{a}_t no es estacionario, se puede asumir que los elementos de \mathbf{P}_t son suficientemente grandes como para despreciar \mathbf{P}_t^{-1} y, entonces, $f(\mathbf{a}_{t-1}|\mathbf{a}_t) \sim N_m\left([\mathbf{S}'\mathbf{Q}^{-1}\mathbf{S}]^{-1}[\mathbf{S}'\mathbf{Q}^{-1}\mathbf{a}_t], [\mathbf{S}'\mathbf{Q}^{-1}\mathbf{S}]^{-1}\right)$ y, si además, \mathbf{S} es invertible, $f(\mathbf{a}_{t-1}|\mathbf{a}_t) \sim N_m\left(\mathbf{S}^{-1}[\mathbf{S}'\mathbf{Q}^{-1}]^{-1}[\mathbf{S}'\mathbf{Q}^{-1}]\mathbf{a}_t, \mathbf{S}^{-1}\mathbf{Q}[\mathbf{S}^{-1}]'\right)$, es decir, $f(\mathbf{a}_{t-1}|\mathbf{a}_t) \sim N_m\left(\mathbf{S}^{-1}\mathbf{a}_t, \mathbf{S}^{-1}\mathbf{Q}[\mathbf{S}^{-1}]'\right)$

De forma más general, el objetivo del filtro de Kalman consiste en encontrar $f(\mathbf{a}_t | \mathbf{I}_k)$, donde $\mathbf{I}_k : \{\mathbf{y}_1, \dots, \mathbf{y}_k\}$. Dependiendo del conjunto de información, \mathbf{I}_k , empleado en la obtención de la distribución condicionada del estado se pueden distinguir tres casos. Si $k < t$, se trata de un problema de predicción, si $k = t$, se habla de un problema de filtrado, y si $k = T$ (o, en general, si $k > t$), se habla de un problema de alisado. En otras palabras, la solución al problema de predicción ofrecerá estimaciones que son extrapolaciones hacia el futuro de los componentes del vector de estado; por otro lado, el filtrado permitirá actualizar las estimaciones del estado no observable a medida que se disponga de una nueva observación de la magnitud observable relacionada con éste; y, finalmente, el alisado ofrece la mejor estimación del estado en cualquier instante del tiempo con toda la información de la muestra. A continuación se obtienen las distribuciones del estado \mathbf{a}_t condicionadas a \mathbf{I}_k en los diferentes casos.

- Predicción un período hacia delante y filtrado ¹⁸

La predicción un período hacia delante consiste en obtener $f(\mathbf{a}_t | \mathbf{I}_{t-1})$. Dicha función de densidad se puede expresar como

$$\begin{aligned} f(\mathbf{a}_t | \mathbf{I}_{t-1}) &= \int_{-\infty}^{\infty} f(\mathbf{a}_t, \mathbf{a}_{t-1} | \mathbf{I}_{t-1}) d\mathbf{a}_{t-1} \\ &= \int_{-\infty}^{\infty} f(\mathbf{a}_t | \mathbf{a}_{t-1}, \mathbf{I}_{t-1}) f(\mathbf{a}_{t-1} | \mathbf{I}_{t-1}) d\mathbf{a}_{t-1} \\ &= \int_{-\infty}^{\infty} f(\mathbf{a}_t | \mathbf{a}_{t-1}) f(\mathbf{a}_{t-1} | \mathbf{I}_{t-1}) d\mathbf{a}_{t-1} \end{aligned}$$

Por tanto, la función de densidad del problema de predicción un período hacia delante puede obtenerse si se conoce $f(\mathbf{a}_t | \mathbf{a}_{t-1})$, que viene dada por la ecuación de transición del modelo, y $f(\mathbf{a}_{t-1} | \mathbf{I}_{t-1})$, que es la solución al problema del filtro en el período anterior.

El problema de filtrado consiste en obtener $f(\mathbf{a}_t | \mathbf{I}_t)$. Si dicha función de densidad se expresa como

$$f(\mathbf{a}_t | \mathbf{I}_t) = f(\mathbf{a}_t | \mathbf{y}_t, \mathbf{I}_{t-1}) = \frac{f(\mathbf{a}_t, \mathbf{y}_t | \mathbf{I}_{t-1})}{f(\mathbf{y}_t | \mathbf{I}_{t-1})} = \frac{f(\mathbf{y}_t | \mathbf{a}_t, \mathbf{I}_{t-1}) f(\mathbf{a}_t | \mathbf{I}_{t-1})}{f(\mathbf{y}_t | \mathbf{I}_{t-1})} = \frac{f(\mathbf{y}_t | \mathbf{a}_t) f(\mathbf{a}_t | \mathbf{I}_{t-1})}{f(\mathbf{y}_t | \mathbf{I}_{t-1})},$$

la solución al problema requiere conocer $f(\mathbf{y}_t | \mathbf{a}_t)$, que viene dada por la ecuación de medida del

¹⁷ Véase anexo II.

¹⁸ La predicción y el filtrado se consideran simultáneamente porque los procesos recursivos de ambas derivaciones están relacionados.

modelo, $f(\mathbf{a}_t | \mathbf{I}_{t-1})$, que se obtiene en la fase de predicción, y $f(\mathbf{y}_t | \mathbf{I}_{t-1})$, que es un factor de normalización que se obtiene como se indica más adelante.

Para derivar las distribuciones del problema de predicción y de filtrado, se definen

$$E[\mathbf{a}_t | \mathbf{I}_k] = \mathbf{a}_{t/k},$$

con $\mathbf{a}_{t/t} \equiv \mathbf{a}_t$, y

$$E[(\mathbf{a}_t - \mathbf{a}_{t/k})(\mathbf{a}_t - \mathbf{a}_{t/k})' | \mathbf{I}_k] = \mathbf{P}_{t/k},$$

con $\mathbf{P}_{t/t} \equiv \mathbf{P}_t$, de forma que, conocidos \mathbf{a}_{t-1} y \mathbf{P}_{t-1} , se tiene que

$$f(\mathbf{a}_{t-1} | \mathbf{I}_{t-1}) \sim N_m(\mathbf{a}_{t-1}, \mathbf{P}_{t-1}).$$

A partir de la ecuación de transición del modelo, se sabe que

$$f(\mathbf{a}_t | \mathbf{a}_{t-1}) \sim N_m(\mathbf{S}\mathbf{a}_{t-1}, \mathbf{Q}).$$

Además,

$$f(\mathbf{a}_t | \mathbf{I}_{t-1}) = \int_{-\infty}^{\infty} f(\mathbf{a}_t | \mathbf{a}_{t-1}) f(\mathbf{a}_{t-1} | \mathbf{I}_{t-1}) d\mathbf{a}_{t-1} \sim N_m(\mathbf{S}\mathbf{a}_{t-1}, \mathbf{S}\mathbf{P}_{t-1}\mathbf{S}' + \mathbf{Q}),$$

es decir,

$$f(\mathbf{a}_t | \mathbf{I}_{t-1}) \sim N_m(\mathbf{S}\mathbf{a}_{t-1}, \mathbf{S}\mathbf{P}_{t-1}\mathbf{S}' + \mathbf{Q})^{19}.$$

Con la notación antes señalada, y dada la normalidad de la distribución, este resultado se puede expresar indicando que la media es $\mathbf{a}_{t/t-1} = \mathbf{S}\mathbf{a}_{t-1}$ y la varianza $\mathbf{P}_{t/t-1} = \mathbf{S}\mathbf{P}_{t-1}\mathbf{S}' + \mathbf{Q}$. Nótese que, dada la ecuación de transición, se tiene que

$$E[\mathbf{a}_t | \mathbf{I}_{t-1}] = \mathbf{S} E[\mathbf{a}_{t-1} | \mathbf{I}_{t-1}] = \mathbf{S}\mathbf{a}_{t-1},$$

y, dado que $Cov[\mathbf{a}_{t-1}, \mathbf{u}_t] = \mathbf{0}$, resulta que

$$Var[\mathbf{a}_t | \mathbf{I}_{t-1}] = Var[\mathbf{S}\mathbf{a}_{t-1} | \mathbf{I}_{t-1}] + Var[\mathbf{u}_t] = \mathbf{S}\mathbf{P}_{t-1}\mathbf{S}' + \mathbf{Q}.$$

Respecto al problema de filtrado, se necesita obtener la función de densidad condicionada $f(\mathbf{y}_t | \mathbf{I}_{t-1})$. Teniendo en cuenta que

$$f(\mathbf{y}_t | \mathbf{I}_{t-1}) = \int_{-\infty}^{\infty} f(\mathbf{y}_t | \mathbf{a}_t, \mathbf{I}_{t-1}) f(\mathbf{a}_t | \mathbf{I}_{t-1}) d\mathbf{a}_t = \int_{-\infty}^{\infty} f(\mathbf{y}_t | \mathbf{a}_t) f(\mathbf{a}_t | \mathbf{I}_{t-1}) d\mathbf{a}_t$$

se obtiene que

¹⁹ Si $f(\mathbf{x}) \sim N(\mathbf{m}, \mathbf{S})$, $f(\mathbf{y}|\mathbf{x}) \sim N(\mathbf{A}\mathbf{x}, \mathbf{W})$ y \mathbf{A} es invertible, entonces $\int_{-\infty}^{\infty} f(\mathbf{y}|\mathbf{x}) f(\mathbf{x}) d\mathbf{x} \sim N(\mathbf{A}\mathbf{m}, \mathbf{A}\mathbf{S}\mathbf{A}' + \mathbf{W})$.

$$f(\mathbf{y}_t | \mathbf{I}_{t-1}) \sim N_N(\mathbf{Z}_t \mathbf{S} \mathbf{a}_{t-1}, \mathbf{Z}_t \mathbf{P}_{t/t-1} \mathbf{Z}' + \mathbf{H}),$$

es decir,

$$f(\mathbf{y}_t | \mathbf{I}_{t-1}) \sim N_N(\mathbf{Z}_t \mathbf{S} \mathbf{a}_{t-1}, \mathbf{F}_t),$$

donde $\mathbf{F}_t = \mathbf{Z}_t \mathbf{P}_{t/t-1} \mathbf{Z}' + \mathbf{H}$ ²⁰.

Antes de obtener la solución al problema del filtrado, conviene expresar esta función de densidad a partir de los denominados errores de predicción, definidos como $\mathbf{e}_t = \mathbf{y}_t - E[\mathbf{y}_t | \mathbf{I}_{t-1}] = \mathbf{y}_t - \mathbf{Z}_t \mathbf{a}_{t/t-1}$, cuya media es nula y cuya varianza es, precisamente, la varianza de $\mathbf{y}_t | \mathbf{I}_{t-1}$, es decir,

$$E\left[(\mathbf{y}_t - E[\mathbf{y}_t | \mathbf{I}_{t-1}])(\mathbf{y}_t - E[\mathbf{y}_t | \mathbf{I}_{t-1}])' | \mathbf{I}_{t-1}\right] = \mathbf{F}_t.$$

Finalmente, se puede obtener la función de densidad del problema del filtro

$$f(\mathbf{a}_t | \mathbf{I}_t) = \frac{f(\mathbf{y}_t | \mathbf{a}_t) f(\mathbf{a}_t | \mathbf{I}_{t-1})}{f(\mathbf{y}_t | \mathbf{I}_{t-1})}$$

y, de hecho, se demuestra que

$$f(\mathbf{a}_t | \mathbf{I}_t) \sim N_m(\mathbf{a}_{t/t-1} + \mathbf{K}_t \mathbf{e}_t, \mathbf{P}_{t/t-1} - \mathbf{K}_t \mathbf{F}_t \mathbf{K}_t'),$$

donde $\mathbf{K}_t = \mathbf{P}_{t/t-1} \mathbf{Z}'_t \mathbf{F}_t^{-1}$ es la llamada ganancia de Kalman.

Con la notación antes señalada, y dada la normalidad de la distribución, este resultado se puede expresar indicando que la media es $\mathbf{a}_t = \mathbf{a}_{t/t-1} + \mathbf{K}_t \mathbf{e}_t$ y la varianza $\mathbf{P}_t = \mathbf{P}_{t/t-1} - \mathbf{K}_t \mathbf{F}_t \mathbf{K}_t'$. Nótese que la distribución de $\mathbf{a}_t | \mathbf{I}_t$ es también la distribución de $\mathbf{a}_t | \mathbf{I}_{t-1}, \mathbf{e}_t$, puesto que, dados \mathbf{I}_{t-1} y \mathbf{e}_t , \mathbf{I}_t queda determinado. Por tanto, dadas las propiedades de la normal multivariante²¹, resulta que

$$E[\mathbf{a}_t | \mathbf{I}_t] = E[\mathbf{a}_t | \mathbf{I}_{t-1}, \mathbf{e}_t] = E[\mathbf{a}_t | \mathbf{I}_{t-1}] + \text{Cov}[\mathbf{a}_t, \mathbf{e}_t | \mathbf{I}_{t-1}] [\text{Var}[\mathbf{e}_t]]^{-1} \mathbf{e}_t$$

y

$$\text{Var}[\mathbf{a}_t | \mathbf{I}_t] = \text{Var}[\mathbf{a}_t | \mathbf{I}_{t-1}, \mathbf{e}_t] = \text{Var}[\mathbf{a}_t | \mathbf{I}_{t-1}] + \text{Cov}[\mathbf{a}_t, \mathbf{e}_t | \mathbf{I}_{t-1}] [\text{Var}[\mathbf{e}_t]]^{-1} \text{Cov}[\mathbf{a}_t, \mathbf{e}_t | \mathbf{I}_{t-1}]',$$

y dado que

$$\text{Cov}[\mathbf{a}_{t-1}, \mathbf{e}_t | \mathbf{I}_{t-1}] = \text{Cov}[\mathbf{a}_{t-1}, \mathbf{y}_t | \mathbf{I}_{t-1}] = \mathbf{P}_{t/t-1} \mathbf{Z}'_t,$$

²⁰ Véase la nota anterior.

²¹ Véase Durbin y Koopman (2001:37).

se tiene que

$$E[\mathbf{a}_t | \mathbf{I}_t] = \mathbf{a}_{t/t-1} + \mathbf{P}_{t/t-1} \mathbf{Z}'_t \mathbf{F}_t^{-1} \mathbf{Z}_t \mathbf{e}_t$$

y

$$\text{Var}[\mathbf{a}_t | \mathbf{I}_t] = \mathbf{P}_{t/t-1} - \mathbf{P}_{t/t-1} \mathbf{Z}'_t \mathbf{F}_t^{-1} \mathbf{Z}_t \mathbf{P}_{t/t-1}.$$

Resumiendo, el problema de predicción un período hacia delante²² y el problema de filtrado se resuelven recursivamente, bajo el supuesto de normalidad, a través de las ecuaciones de predicción del estado, que permiten proyectar hacia el futuro el estado actual y estimar la varianza de la predicción, y las ecuaciones de actualización, que permiten incorporar una nueva observación a la estimación del estado y obtener una estimación mejorada del estado²³.

Ecuaciones de predicción	Predicción del estado	$\mathbf{a}_{t/t-1} = \mathbf{S} \mathbf{a}_{t-1}$
	Varianza de la predicción del estado	$\mathbf{P}_{t/t-1} = \mathbf{S} \mathbf{P}_{t-1} \mathbf{S}' + \mathbf{Q}$
Ecuaciones de actualización	Estimación filtrada del estado	$\mathbf{a}_t = \mathbf{a}_{t/t-1} + \mathbf{K}_t \mathbf{e}_t$
	Varianza de la estimación filtrada del estado	$\mathbf{P}_t = \mathbf{P}_{t/t-1} - \mathbf{K}_t \mathbf{F}_t \mathbf{K}'_t$

Es decir, el algoritmo se ejecuta en dos fases. En una primera fase se hace una predicción del estado en t a partir de la estimación filtrada del estado en $t-1$ mediante la ecuación $\mathbf{a}_{t/t-1} = \mathbf{S} \mathbf{a}_{t-1}$ y la varianza del error de predicción viene dada por $\mathbf{P}_{t/t-1} = \mathbf{S} \mathbf{P}_{t-1} \mathbf{S}' + \mathbf{Q}$. En la segunda fase, se actualiza la predicción incorporando la observación y_t y se obtiene la estimación filtrada del estado, es decir, $\mathbf{a}_t = \mathbf{a}_{t/t-1} + \mathbf{K}_t \mathbf{e}_t$, donde $\mathbf{e}_t = y_t - \mathbf{Z}_t \mathbf{a}_{t/t-1}$, que es una combinación lineal de la predicción del estado y de la diferencia observada entre la medida registrada en el instante t , y_t , y la predicción realizada para dicha medida con la estimación $\mathbf{a}_{t/t-1}$. El factor que pondera dicha diferencia, \mathbf{K}_t , es la ya definida ganancia de Kalman.

- Alisado.

²² La predicción m períodos hacia delante es relativamente trivial una vez solucionado el problema del filtro (Steyn, 1996:19).

²³ Como señalan Welch y Bishop (2001:4-5), las primeras pueden interpretarse como ecuaciones de predicción y las segundas de corrección. Cuando las secuencias $\{\mathbf{P}_{t/t-1}\}_{t=1, \dots, T}$ y $\{\mathbf{K}_t\}_{t=1, \dots, T}$ convergen a matrices fijas, el sistema alcanza un estado estacionario (*steady state*).

El problema de alisado de intervalo fijo²⁴ consiste en obtener la función de densidad condicionada $f(\mathbf{a}_t | \mathbf{I}_T)$, siendo T el número total de observaciones. Esta función de densidad se obtiene también recursivamente. Para ello, es preciso relacionar $f(\mathbf{a}_t | \mathbf{I}_T)$ con $f(\mathbf{a}_{t+1} | \mathbf{I}_T)$ y, en este sentido, la función de densidad conjunta viene dada por

$$f(\mathbf{a}_t, \mathbf{a}_{t+1} | \mathbf{I}_T) = f(\mathbf{a}_{t+1} | \mathbf{I}_T) f(\mathbf{a}_t | \mathbf{a}_{t+1}, \mathbf{I}_T) = f(\mathbf{a}_{t+1} | \mathbf{I}_T) f(\mathbf{a}_t | \mathbf{a}_{t+1}, \mathbf{I}_t).$$

Nótese que, conocido \mathbf{a}_{t+1} , la información $\{\mathbf{y}_{t+1}, \dots, \mathbf{y}_T\}$ es irrelevante para la distribución de \mathbf{a}_t . Además, a partir de la expresión de $f(\mathbf{a}_t, \mathbf{a}_{t+1} | \mathbf{I}_T)$ en $T = t$, se tiene que

$$f(\mathbf{a}_t, \mathbf{a}_{t+1} | \mathbf{I}_t) = f(\mathbf{a}_{t+1} | \mathbf{I}_t) f(\mathbf{a}_t | \mathbf{a}_{t+1}, \mathbf{I}_t)$$

y, entonces,

$$f(\mathbf{a}_t | \mathbf{a}_{t+1}, \mathbf{I}_t) = \frac{f(\mathbf{a}_t, \mathbf{a}_{t+1} | \mathbf{I}_t)}{f(\mathbf{a}_{t+1} | \mathbf{I}_t)}.$$

Por tanto, $f(\mathbf{a}_t, \mathbf{a}_{t+1} | \mathbf{I}_T)$ se puede expresar como

$$\begin{aligned} f(\mathbf{a}_t, \mathbf{a}_{t+1} | \mathbf{I}_T) &= f(\mathbf{a}_{t+1} | \mathbf{I}_T) \frac{f(\mathbf{a}_t, \mathbf{a}_{t+1} | \mathbf{I}_t)}{f(\mathbf{a}_{t+1} | \mathbf{I}_t)} = f(\mathbf{a}_{t+1} | \mathbf{I}_T) \frac{f(\mathbf{a}_{t+1} | \mathbf{a}_t, \mathbf{I}_t) f(\mathbf{a}_t | \mathbf{I}_t)}{f(\mathbf{a}_{t+1} | \mathbf{I}_t)} \\ &= f(\mathbf{a}_{t+1} | \mathbf{I}_T) \frac{f(\mathbf{a}_{t+1} | \mathbf{a}_t) f(\mathbf{a}_t | \mathbf{I}_t)}{f(\mathbf{a}_{t+1} | \mathbf{I}_t)}. \end{aligned}$$

Integrando la función de densidad conjunta con respecto a \mathbf{a}_{t+1} se obtiene la función de densidad deseada. Es decir,

$$f(\mathbf{a}_t | \mathbf{I}_T) = \int_{-\infty}^{\infty} f(\mathbf{a}_t, \mathbf{a}_{t+1} | \mathbf{I}_T) d\mathbf{a}_{t+1} = f(\mathbf{a}_t | \mathbf{I}_t) \int_{-\infty}^{\infty} f(\mathbf{a}_{t+1} | \mathbf{I}_T) \frac{f(\mathbf{a}_{t+1} | \mathbf{a}_t)}{f(\mathbf{a}_{t+1} | \mathbf{I}_t)} d\mathbf{a}_{t+1}.$$

A partir de esta expresión, se obtiene $f(\mathbf{a}_t | \mathbf{I}_T)$, una vez determinado $f(\mathbf{a}_{t+1} | \mathbf{I}_T)$. Dado que $f(\mathbf{a}_T | \mathbf{I}_T)$ es la última función de densidad obtenida en el problema del filtro, la función de densidad $f(\mathbf{a}_{T-1} | \mathbf{I}_T)$ se puede obtener a partir de la anterior. De este modo, se obtienen

²⁴ Para obtener las distribuciones alisadas se puede recurrir a diferentes algoritmos (véase Anderson y Moore, 1979:cap.7 y Harvey, 1989:149-155): a) alisado de punto fijo, que obtiene $\mathbf{a}_{\tau/T}$ para determinados instantes del tiempo τ ; b) alisado de retardo fijo, que obtiene $\mathbf{a}_{t-j/t}$, $j = 1, \dots, M$, siendo M el máximo retardo especificado (véase Anderson y Moore 1979:cap.7); c) alisado de intervalo fijo, que obtiene $\mathbf{a}_{t/T}$, $t = 1, \dots, T$. Este último algoritmo es el más utilizado (veáanse Anderson y Moore, 1979:cap.7, Ansley y Kohn, 1982, Kohn y Ansley, 1983, y Koopman, 1996).

recursivamente las funciones de densidad alisadas $f(\mathbf{a}_t | \mathbf{I}_T)$ para cada uno de los instantes del tiempo $t = 1, \dots, T$.

Si se asumen funciones de densidad normales, $\mathbf{a}_{t/T} = E[\mathbf{a}_t | \mathbf{I}_T]$ es también la moda de la distribución de $\mathbf{a}_t | \mathbf{I}_T$, de modo que $\mathbf{a}_{1/T}, \dots, \mathbf{a}_{T/T}$ pueden obtenerse como los valores de $\mathbf{a}_1, \dots, \mathbf{a}_T$ que maximizan $\log f(\mathbf{a}_1, \dots, \mathbf{a}_T | \mathbf{I}_T)$, es decir, $\log f(\mathbf{a}_1, \dots, \mathbf{a}_T | \mathbf{y}_1, \dots, \mathbf{y}_T)$ o, en forma compacta, $\log f(\mathbf{a} | \mathbf{y})$. Teniendo en cuenta que

$$\frac{\partial \log f(\mathbf{a} | \mathbf{y})}{\partial \mathbf{a}} = \frac{\partial [\log f(\mathbf{a}, \mathbf{y}) - \log f(\mathbf{y})]}{\partial \mathbf{a}} = \frac{\partial \log f(\mathbf{a}, \mathbf{y})}{\partial \mathbf{a}},$$

las estimaciones alisadas del estado pueden encontrarse como solución a las ecuaciones

$$\frac{\partial \log f(\mathbf{a}_1, \dots, \mathbf{a}_T, \mathbf{y}_1, \dots, \mathbf{y}_T)}{\partial \mathbf{a}_t} = \mathbf{0}, \quad t = 1, \dots, T \quad ^{25}.$$

Nótese que, dadas las propiedades de la normal multivariante²⁶, y teniendo en cuenta que \mathbf{I}_T está determinado si se conoce $\mathbf{I}_t, \mathbf{e}_{t+1}, \dots, \mathbf{e}_T$, así como el hecho de que $\mathbf{e}_{t+1}, \dots, \mathbf{e}_T$ son independientes de \mathbf{I}_t y, a su vez, independientes entre sí con medias nulas, resulta que

$$\mathbf{a}_{t/T} \equiv E[\mathbf{a}_t | \mathbf{I}_T] = E[\mathbf{a}_t | \mathbf{I}_t, \mathbf{e}_{t+1}, \dots, \mathbf{e}_T] = \mathbf{a}_t + \sum_{j=t+1}^T \text{Cov}[\mathbf{a}_t, \mathbf{e}_j | \mathbf{I}_t] \mathbf{F}_j^{-1} \mathbf{e}_j$$

y

$$\mathbf{P}_{t/T} \equiv \text{Var}[\mathbf{a}_t | \mathbf{I}_T] = \text{Var}[\mathbf{a}_t | \mathbf{I}_t, \mathbf{e}_{t+1}, \dots, \mathbf{e}_T] = \mathbf{P}_t - \sum_{j=t+1}^T \text{Cov}[\mathbf{a}_t, \mathbf{e}_j | \mathbf{I}_t] \mathbf{F}_j^{-1} \text{Cov}[\mathbf{a}_t, \mathbf{e}_j | \mathbf{I}_t]'$$

Estas expresiones pueden calcularse recursivamente a través de diferentes algoritmos²⁷. Nótese que $\mathbf{P}_{t/T} \leq \mathbf{P}_t$, dado que se está utilizando más información, y la diferencia entre \mathbf{P}_t y $\mathbf{P}_{t/T}$ mide la ganancia obtenida con el alisado.

Anderson y Moore (1979) propusieron su ya clásico algoritmo en el que la media y la varianza de $\mathbf{a}_t | \mathbf{I}_T$ se derivan recursivamente a través de las ecuaciones

$$\mathbf{a}_{t/T} = \mathbf{a}_t + \mathbf{P}_t^* [\mathbf{a}_{t+1/T} - \mathbf{S} \mathbf{a}_t]$$

y

²⁵ Véase Harvey y Shephard (1993).

²⁶ Véase Durbin y Koopman (2001:37).

²⁷ Véase Durbin y Koopman (2001:70-73).

$$\mathbf{P}_{t/T} = \mathbf{P}_t + \mathbf{P}_t^* (\mathbf{P}_{t+1/T} - \mathbf{P}_{t+1/t}) \mathbf{P}_t^{*'},$$

donde

$$\mathbf{P}_t^* = \mathbf{P}_t \mathbf{S} \mathbf{P}_{t+1/t}^{-1},$$

para $t = T-1, \dots, 1$, respectivamente, (Steyn, 1996:20)²⁸. Estas expresiones pueden obtenerse si se calcula primero $E[\mathbf{a}_t | \mathbf{a}_{t+1}, \mathbf{I}_t]$. Pues bien, se tiene que

$$E[\mathbf{a}_t | \mathbf{a}_{t+1}, \mathbf{I}_t] = E[\mathbf{a}_t | \mathbf{I}_t] + \text{Cov}[\mathbf{a}_t, \mathbf{a}_{t+1} | \mathbf{I}_t] \text{Var}[\mathbf{a}_{t+1} | \mathbf{I}_t]^{-1} [\mathbf{a}_{t+1} - E[\mathbf{a}_{t+1} | \mathbf{I}_t]].$$

Y, dado que,

$$\begin{aligned} \text{Cov}[\mathbf{a}_t, \mathbf{a}_{t+1} | \mathbf{I}_t] &= E\left[(\mathbf{a}_t - E[\mathbf{a}_t | \mathbf{I}_t]) (\mathbf{a}_{t+1} - E[\mathbf{a}_{t+1} | \mathbf{I}_t])' \right] \\ &= E\left[(\mathbf{a}_t - \mathbf{a}_t) (\mathbf{a}_t - \mathbf{a}_t)' \mathbf{S}' | \mathbf{I}_t \right] = \mathbf{P}_t \mathbf{S}' \end{aligned}$$

resulta que

$$\begin{aligned} E[\mathbf{a}_t | \mathbf{a}_{t+1}, \mathbf{I}_t] &= \mathbf{a}_t + \mathbf{P}_t \mathbf{S}' \mathbf{P}_{t+1/t}^{-1} [\mathbf{a}_{t+1} - \mathbf{a}_{t+1/t}] \\ &= \mathbf{a}_t + \mathbf{P}_t \mathbf{S}' \mathbf{P}_{t+1/t}^{-1} [\mathbf{a}_{t+1} - \mathbf{S} \mathbf{a}_t] = \mathbf{a}_t + \mathbf{P}_t^* [\mathbf{a}_{t+1} - \mathbf{S} \mathbf{a}_t] \end{aligned}$$

Si ahora se advierte que $E[\mathbf{a}_t | \mathbf{a}_{t+1}, \mathbf{I}_T] = E[\mathbf{a}_t | \mathbf{a}_{t+1}, \mathbf{I}_t]$ ²⁹, entonces, por la ley de las proyecciones iteradas³⁰, se tiene que

$$E[\mathbf{a}_t | \mathbf{I}_T] = E[E[\mathbf{a}_t | \mathbf{a}_{t+1}, \mathbf{I}_T] | \mathbf{I}_T] = \mathbf{a}_t + \mathbf{P}_t^* [E[\mathbf{a}_{t+1} | \mathbf{I}_T] - \mathbf{S} \mathbf{a}_t] = \mathbf{a}_t + \mathbf{P}_t^* [\mathbf{a}_{t+1/T} - \mathbf{S} \mathbf{a}_t].$$

Para obtener

$$\mathbf{P}_{t/T} \equiv E\left[(\mathbf{a}_t - \mathbf{a}_{t/T}) (\mathbf{a}_t - \mathbf{a}_{t/T})' | \mathbf{I}_T \right],$$

basta con considerar que

$$\mathbf{a}_t - \mathbf{a}_{t/T} = \mathbf{a}_t - \mathbf{a}_t - \mathbf{P}_t^* [\mathbf{a}_{t+1/T} - \mathbf{a}_{t+1/t}].$$

Entonces,

$$\mathbf{a}_t - \mathbf{a}_{t/T} = \mathbf{a}_t - \mathbf{a}_t - \mathbf{P}_t^* \mathbf{a}_{t+1/T} + \mathbf{P}_t^* \mathbf{a}_{t+1/t},$$

o bien,

$$\mathbf{a}_t - \mathbf{a}_{t/T} + \mathbf{P}_t^* \mathbf{a}_{t+1/T} = \mathbf{a}_t - \mathbf{a}_t + \mathbf{P}_t^* \mathbf{a}_{t+1/t},$$

²⁸ El algoritmo propuesto por Anderson y Moore (1979) es poco eficiente en términos de tiempo de cálculo y memoria. Un procedimiento más eficiente ha sido propuesto por Koopman (1993).

²⁹ Nótese que, conocido α_{t+1} , la información $\{y_{t+1}, \dots, y_T\}$ es irrelevante para la distribución de α_t .

³⁰ Véase Hamilton (1994:396).

de modo que

$$(\mathbf{a}_t - \mathbf{a}_{t/T} + \mathbf{P}_t^* \mathbf{a}_{t+1/T})(\mathbf{a}_t - \mathbf{a}_{t/T} + \mathbf{P}_t^* \mathbf{a}_{t+1/T})' = (\mathbf{a}_t - \mathbf{a}_t + \mathbf{P}_t^* \mathbf{a}_{t+1/t})(\mathbf{a}_t - \mathbf{a}_t + \mathbf{P}_t^* \mathbf{a}_{t+1/t})',$$

y tomando esperanzas, resulta que

$$\begin{aligned} & E\left[(\mathbf{a}_t - \mathbf{a}_{t/T} + \mathbf{P}_t^* \mathbf{a}_{t+1/T})(\mathbf{a}_t - \mathbf{a}_{t/T} + \mathbf{P}_t^* \mathbf{a}_{t+1/T})'\right] \\ &= E\left[(\mathbf{a}_t - \mathbf{a}_{t/T})(\mathbf{a}_t - \mathbf{a}_{t/T})'\right] + \mathbf{P}_t^* E[\mathbf{a}_{t+1/T} \mathbf{a}'_{t+1/T}] \mathbf{P}_t^{*'} \end{aligned}$$

y

$$\begin{aligned} & E\left[(\mathbf{a}_t - \mathbf{a}_t + \mathbf{P}_t^* \mathbf{a}_{t+1/t})(\mathbf{a}_t - \mathbf{a}_t + \mathbf{P}_t^* \mathbf{a}_{t+1/t})'\right] \quad 31. \\ &= E\left[(\mathbf{a}_t - \mathbf{a}_t)(\mathbf{a}_t - \mathbf{a}_t)'\right] + \mathbf{P}_t^* E[\mathbf{a}_{t+1/t} \mathbf{a}'_{t+1/t}] \mathbf{P}_t^{*'} \end{aligned}$$

Se obtiene que

$$\mathbf{P}_{t/T} + \mathbf{P}_t^* E[\mathbf{a}_{t+1/T} \mathbf{a}'_{t+1/T}] \mathbf{P}_t^{*'} = \mathbf{P}_t + \mathbf{P}_t^* E[\mathbf{a}_{t+1/t} \mathbf{a}'_{t+1/t}] \mathbf{P}_t^{*'},$$

de modo que

$$\begin{aligned} \mathbf{P}_{t/T} &= \mathbf{P}_t - \mathbf{P}_t^* E[\mathbf{a}_{t+1/T} \mathbf{a}'_{t+1/T}] \mathbf{P}_t^{*'} + \mathbf{P}_t^* E[\mathbf{a}_{t+1/t} \mathbf{a}'_{t+1/t}] \mathbf{P}_t^{*'} \\ &= \mathbf{P}_t + \mathbf{P}_t^* [-E[\mathbf{a}_{t+1/T} \mathbf{a}'_{t+1/T}] + E[\mathbf{a}_{t+1/t} \mathbf{a}'_{t+1/t}]] \mathbf{P}_t^{*'} \end{aligned}$$

y sumando y restando $E[\mathbf{a}_{t+1} \mathbf{a}'_{t+1}]$, se obtiene que

$$\mathbf{P}_{t/T} = \mathbf{P}_t + \mathbf{P}_t^* [\mathbf{P}_{t+1/T} - \mathbf{P}_{t+1/t}] \mathbf{P}_t^{*'}.$$

Bajo el supuesto de gaussianidad, se puede demostrar que las estimaciones alisadas del estado coinciden con los valores de una función *spline*³². Así, si deseamos aproximar una función alisada μ_t a una serie y_1, \dots, y_T de valores observados en instantes equidistantes en el tiempo, una aproximación estándar consiste en elegir μ_t tal que minimice la expresión

$$\sum_{t=1}^T (y_t - \mu_t)^2 + \lambda \sum_{t=1}^T (\Delta^2 \mu_t)^2,$$

³¹ Nótese que $E[(\mathbf{a}_t - \mathbf{a}_{t/T}) \mathbf{a}'_{t+1/T}] = E[\mathbf{a}_t \mathbf{a}'_{t+1/T}] - E[\mathbf{a}_{t/T} \mathbf{a}'_{t+1/T}] = \mathbf{0}$ y, del mismo modo, $E[(\mathbf{a}_t - \mathbf{a}_t) \mathbf{a}'_{t+1/t}] = E[\mathbf{a}_t \mathbf{a}'_{t+1/t}] - E[\mathbf{a}_t \mathbf{a}'_{t+1/t}] = \mathbf{0}$.

³² Véase Durbin y Koopman (2001:61-62).

para un valor dado de λ ($\lambda > 0$)³³. La función μ_t se conoce como función *spline*. Este problema puede formularse en el marco de los modelos en el espacio de los estados mediante las ecuaciones

$$y_t = \mu_t + \varepsilon_t$$

y

$$\Delta^2 \mu_t = \xi_t,$$

donde $Var[\varepsilon_t] = \sigma^2$ y $Var[\xi_t] = \lambda^{-1} \sigma^2$, $\lambda > 0$. Bajo el supuesto de normalidad, el logaritmo natural de la función de densidad de $\mathbf{a}_1, \dots, \mathbf{a}_T$, y_1, \dots, y_T es, al margen de las constantes irrelevantes, igual a

$$-\frac{\lambda}{2\sigma^2} \sum_{t=1}^T (\Delta^2 \mu_t)^2 - \frac{1}{2\sigma^2} \sum_{t=1}^T (y_t - \mu_t)^2,$$

y maximizar esta función es equivalente a minimizar la función señalada anteriormente, ocurriendo además que, como se comentó en los párrafos precedentes, los valores $\mathbf{a}_1, \dots, \mathbf{a}_T$ que minimizan la función de verosimilitud de $\mathbf{a}_1, \dots, \mathbf{a}_T$, y_1, \dots, y_T son, precisamente, las estimaciones alisadas del estado.

Una vez obtenidas las medias condicionales $E[\mathbf{a}_t | \mathbf{I}_k]$, $k = t-1, t, \dots, T$, resulta que dichas medias son sumas ponderadas de observaciones pasadas (predicción), observaciones pasadas y presentes (filtrado) o de todas las observaciones (alisado). El estudio de tales ponderaciones ayuda a comprender mejor las propiedades de los diferentes estimadores. Por ejemplo, la estimación alisada del componente de nivel $m_{t/T}$ en torno a $t = T/2$, $m_{(T/2)/T}$, debería ser una media ponderada de todas las observaciones,

$$m_{(T/2)/T} = \sum_{t=1}^T w_t y_t,$$

tal que las ponderaciones, w_t , decrezcan simétrica y exponencialmente a ambos lados de $t = T/2$. Aquellos modelos en los que los patrones de pesos difieran de lo que parece apropiado deberían ser investigados, como sugieren Durbin y Koopman (2001:81).

³³ Cuanto más pequeño sea el valor de λ , más cercanos estarán los valores de μ_t a los de y_t , pero menos alisada será μ_t .

2.3. Inicialización y estimación máximo verosímil

El carácter recursivo del filtro de Kalman exige obtener una distribución del estado inicial \mathbf{a}_0 o, cuando menos, determinar su media \mathbf{a}_0 y su varianza \mathbf{P}_0 para iniciar dicho proceso recursivo³⁴. La situación ideal es aquella en la que tanto el tipo de distribución como los parámetros \mathbf{a}_0 y \mathbf{P}_0 son conocidos. Pero, en general, no se conoce nada sobre la distribución de \mathbf{a}_0 . En este último caso, existen diferentes procedimientos de inicialización del filtro.

Una posibilidad consiste en derivar la distribución inicial a partir de información externa al modelo (Steyn, 1996:25-26), bien, estableciendo una determinada distribución *a priori*, o bien, considerando la estacionariedad del estado. Si se imponen restricciones de estacionariedad sobre el estado, puede concluirse que $\mathbf{a}_0 = \mathbf{0}$, mientras que la varianza \mathbf{P}_0 se determina a partir de la ecuación de transición resolviendo iterativamente la ecuación $\mathbf{P}_i = \mathbf{S}\mathbf{P}_{i-1}\mathbf{S}' + \mathbf{Q}$ hasta que se alcanza la convergencia³⁵. Puede demostrarse que

$$\text{vec}(\mathbf{P}_0) = [\mathbf{I}_{m^2} - \mathbf{S} \otimes \mathbf{S}]^{-1} \text{vec}(\mathbf{Q}),$$

donde $\text{vec}(\mathbf{P}_0)$ y $\text{vec}(\mathbf{Q})$ son vectores columna formados por las columnas de \mathbf{P}_0 y \mathbf{Q} ³⁶.

Si no se dispone de información externa, se suele asumir que $\mathbf{a}_0 \sim N(\mathbf{a}_0, \kappa \mathbf{I})$, siendo κ un número suficientemente grande (Steyn, 1996:26). En esta misma línea, puede fijarse un valor arbitrario \mathbf{a}_0 y considerar que $\kappa \rightarrow \infty$, de modo que $\mathbf{P}_0 \rightarrow \infty$. Se dice entonces que \mathbf{a}_0 tiene una distribución *a priori* difusa. En este último caso, el proceso recursivo inherente al filtro de Kalman se inicia calculando los límites cuando $\mathbf{P}_0 \rightarrow \infty$ de aquellas expresiones en las que aparece este término. Este proceso se denomina inicialización difusa y el filtro resultante se denomina filtro de Kalman difuso (Durbin y Koopman, 2001:100-101). Algunos filtros difusos fueron propuestos por Ansley y Kohn (1985) y De Jong (1991)³⁷. Koopman (1997) y Durbin y

³⁴ La cuestión de la inicialización es relevante, sobre todo, para la computación (Steyn, 1996:30).

³⁵ Nótese que, bajo el supuesto de que $\text{Var}[\mathbf{a}_t] = \text{Var}[\mathbf{a}_{t-1}] = \mathbf{P}_0$ y dado que $\mathbf{a}_t = \mathbf{S}\mathbf{a}_{t-1} + \mathbf{u}_t$, se obtiene que $\mathbf{P}_0 = \mathbf{S}\mathbf{P}_0\mathbf{S}' + \mathbf{Q}$.

³⁶ Es decir, $\text{vec}(\mathbf{P}_0)$ es un vector columna $m^2 \times 1$ cuyos primeros m elementos son los de la primera columna de \mathbf{P}_0 , los siguientes m elementos son los de la segunda columna y, así, sucesivamente. Dado que $\mathbf{P}_0 = \mathbf{S}\mathbf{P}_0\mathbf{S}' + \mathbf{Q}$, entonces

$$\text{vec}(\mathbf{P}_0) = (\mathbf{S} \otimes \mathbf{S}) \text{vec}(\mathbf{P}_0) + \text{vec}(\mathbf{Q}). \text{ Por tanto, } \text{vec}(\mathbf{P}_0) = (\mathbf{I}_{m^2} - \mathbf{S} \otimes \mathbf{S})^{-1} \text{vec}(\mathbf{Q}).$$

Véase Hamilton (1994:264-265), Steyn (1996:25) y Durbin y Koopman (2001:112).

³⁷ La propuesta de De Jong (1991) implica aumentar la dimensión del vector de estado, de ahí que se conozca como filtrado y alisado aumentado.

Koopman (2001:101-120), basándose en las ideas de Ansley y Kohn (1985), efectúan un tratamiento más transparente del filtrado y alisado difuso y proponen los métodos denominados filtrado y alisado inicial exacto.

La inicialización difusa es el mecanismo más frecuente cuando no se conoce nada sobre la distribución de \mathbf{a}_0 . Sin embargo, algunos autores opinan que asignar una varianza infinita al estado inicial es incongruente con el hecho de que los valores de la magnitud observada sean finitos. Una aproximación alternativa consiste en asumir que $\mathbf{P}_0 = \mathbf{0}$ y, por tanto, \mathbf{a}_0 es un parámetro desconocido que puede ser estimado por máxima verosimilitud (Rosenberg, 1973)³⁸.

Por último, se pueden sacrificar una o más observaciones y usar, por ejemplo, $f(\mathbf{a}_1|\mathbf{I}_1)$ como distribución inicial (Steyn, 1996:30). Por ejemplo, en el modelo de nivel local, y dado que $y_1|\alpha_1 \sim N(\mu_1, \sigma_\epsilon^2)$, se pueden utilizar las estimaciones $\hat{\mu}_1 = y_1$ y $Var[\hat{\mu}_1] = \sigma_\epsilon^2$. Obtenida la distribución de α_1 se puede continuar el proceso recursivo³⁹.

Como se acaba de ver, el filtro de Kalman permite obtener recursivamente estimaciones del vector de estado; pero éstas serán, en última instancia, función no sólo de los valores iniciales sino también de otros parámetros desconocidos tales como las varianzas de los términos de perturbación de las ecuaciones de transición y de medida. Pues bien, estos parámetros pueden ser estimados por máxima verosimilitud.

La función de verosimilitud de las observaciones, es decir, del vector de vectores $\mathbf{y} = (y_1, \dots, y_T)'$, se puede expresar en términos de distribuciones condicionadas de la forma

$$l(\mathbf{y}) = f(y_1, \dots, y_T) = f(y_1) \prod_{t=2}^T f(y_t | \mathbf{I}_{t-1})$$

y sabiendo que

$$f(y_t | \mathbf{I}_{t-1}) \sim N_N(\mathbf{Z}_t \mathbf{S} \mathbf{a}_{t-1}, \mathbf{F}_t),$$

donde

$$\mathbf{F}_t = \mathbf{Z}_t \mathbf{P}_{t/t-1} \mathbf{Z}_t' + \mathbf{H},$$

y, conocidos \mathbf{a}_0 y \mathbf{P}_0 ,

$$f(y_1) \sim N_N(\mathbf{Z}_1 \mathbf{S} \mathbf{a}_0, [\mathbf{Z}_1 \mathbf{P}_{1/0} \mathbf{Z}_1' + \mathbf{H}]).$$

³⁸ El método propuesto por Rosenberg (1973) implica también aumentar la dimensión del vector de estado y, por esta razón, se encuadra en el conjunto de métodos de filtrado aumentado.

La anterior función de verosimilitud se puede escribir también en términos de $\mathbf{e} = (e_1, \dots, e_T)'$. Dada la definición de los errores de predicción, $e_t = y_t - E[y_t | \mathbf{I}_{t-1}]$, el hecho de que el jacobiano de la transformación de $\mathbf{e} = (e_1, \dots, e_T)'$ en $\mathbf{y} = (y_1, \dots, y_T)'$ es unitario, y teniendo en cuenta que e_t es independiente de \mathbf{I}_{t-1} , se obtiene la función de densidad transformada

$$l(\mathbf{e}) = f(e_1, \dots, e_T) = \prod_{t=1}^T f(e_t).$$

Es decir, los errores de predicción son independientes⁴⁰, de forma que

$$l(\mathbf{e}) = f(e_1, \dots, e_T) = \prod_{t=1}^T f(e_t) = \frac{\prod_{t=1}^T |\mathbf{F}_t|^{-1/2}}{(2\pi)^{NT/2}} \prod_{t=1}^T \exp\left(-\frac{1}{2} \mathbf{e}_t' \mathbf{F}_t^{-1} \mathbf{e}_t\right).$$

Entonces

$$-2 \ln l = NT \cdot \ln(2\pi) + \sum_{t=1}^T \ln |\mathbf{F}_t| + \sum_{t=1}^T \mathbf{e}_t' \mathbf{F}_t^{-1} \mathbf{e}_t$$

y, a partir de esta expresión, se pueden estimar los parámetros del modelo.

Si se supone que \mathbf{a}_0 no es estocástico, sino un parámetro fijo desconocido, puede incorporarse como un parámetro más que debe ser estimado por máxima verosimilitud. Entonces conviene expresar la estimación máximo verosímil de \mathbf{a}_0 como función del resto de parámetros desconocidos y obtener la función de verosimilitud concentrada, en la que \mathbf{a}_0 se sustituye por su estimación máximo verosímil⁴¹.

En la práctica, la optimización de esta función se realiza a través de algoritmos numéricos de búsqueda basados, en muchos casos, en el método de Newton, que resuelve las ecuaciones $\partial l(\mathbf{y}/\mathbf{y})/\partial \mathbf{y} = \mathbf{0}$, siendo \mathbf{y} el vector de parámetros que se desea estimar usando desarrollos en serie de Taylor de primer orden⁴².

2.4. Observaciones anómalas y filtro de Kalman

³⁹ Otra opción consiste en asumir que en el punto $t = -m$, \mathbf{a}_{-m} tiene una varianza dada por \mathbf{P}_{-m} y, entonces, bajo ciertos supuestos, derivar el valor de \mathbf{P}_0 (Stein, 1996:26-29).

⁴⁰ Los errores de predicción son importantes desde el punto de vista estadístico porque son serialmente independientes cuando el modelo está especificado correctamente.

⁴¹ En el caso de que se recurra a la inicialización difusa, el logaritmo de la función de verosimilitud no converge cuando $\kappa \rightarrow \infty$, por lo que es preciso redefinir la función de verosimilitud para poder estimar los parámetros (véase De Jong, 1991, y Durbin y Koopman, 2001:139-141).

El funcionamiento del filtro de Kalman se puede ver afectado por la presencia de observaciones anómalas, también conocidas como *outliers*, que pueden tener efectos transitorios o, por el contrario, afectar a las observaciones posteriores.

En los modelos en el espacio de los estados, una observación anómala pasajera tiene lugar cuando la perturbación de la ecuación de medida toma un valor extraño en relación con la distribución asumida, pero en el período siguiente la perturbación vuelve a comportarse de acuerdo con lo esperado. Este tipo de observación anómala se conoce como *outlier* aditivo (AO). Sin embargo, un valor extraño en la perturbación de la ecuación de transición no sólo altera el estado actual \mathbf{a}_t , sino que, dada la naturaleza recursiva del modelo, afecta también a los valores siguientes del vector de estado. Este tipo de observación anómala se denomina *outlier* innovativo (IO).

Un valor inesperado de \mathbf{y}_t , es decir, un valor no concordante con el supuesto de que

$$f(\mathbf{y}_t | \mathbf{I}_{t-1}) \sim N_N(\mathbf{Z}_t \mathbf{S} \mathbf{a}_{t-1}, \mathbf{F}_t),$$

puede ser atribuido a la presencia de un *outlier* AO o IO⁴³. Dado que el error de predicción \mathbf{e}_t es tal que $\mathbf{e}_t \sim N_N(\mathbf{0}, \mathbf{F}_t)$, se podría considerar que existen *outliers* si $|\mathbf{e}_t| > \lambda \sqrt{\mathbf{F}_t}$, siendo λ una constante prefijada⁴⁴.

La presencia de observaciones anómalas tiene graves efectos sobre la estimación de filtro del estado, ya que $\mathbf{a}_t = \mathbf{a}_{t/t-1} + \mathbf{K}_t \mathbf{e}_t$. Pues bien, los procedimientos diseñados para minimizar, limitar o incluso eliminar estos efectos, se denominan procedimientos de robustecimiento⁴⁵. Algunos de estos métodos asumen distribuciones no normales que permiten recoger algunas de estas observaciones anómalas. Ahora bien, el supuesto de distribuciones invariantes en el tiempo choca con el problema de que la distribución que recoge las anomalías no se adapta a las observaciones que no son anómalas. Otro grupo de métodos emplea ecuaciones de actualización

⁴² Algunos de los algoritmos más utilizados, tales como el método de Broyden, Fletcher, Goldfarb y Shannon (BFGS) o el algoritmo EM, son descritos por Durbin y Koopman (2001:142-148).

⁴³ El mejor indicador de la presencia de observaciones anómalas es el error de predicción. Sin embargo, bajo el enfoque estructural se puede recurrir a otro conjunto de residuos, denominados residuos auxiliares, que son las estimaciones estandarizadas de los términos de perturbación asociados a los componentes inobservables. Dichos residuos auxiliares, a pesar de estar serialmente correlacionados aunque el modelo esté correctamente especificado, informan de la presencia de observaciones anómalas y, además, proporcionan una indicación de cuál es la fuente del problema (Harvey y Koopman, 1992)

⁴⁴ Una revisión de los métodos de diagnóstico para modelos en el espacio de los estados puede encontrarse en Harvey y Koopman (1992). Véase también Koopman y otros (2000) y Durbin y Koopman (2001:33-35, 152-154).

modificadas *ad hoc*, como por ejemplo $\mathbf{a}_t = \mathbf{a}_{t/t-1} + \mathbf{K}_t \min\{e_t, \lambda\sqrt{\mathbf{F}_t}\}$, que pretenden atenuar la influencia del error de predicción en la actualización de la estimación.

En cualquier caso, es interesante detectar si existen *outliers* AO o IO. Una forma de proceder si se detecta un valor sospechoso, y_t , consiste en predecir el siguiente valor, y_{t+1} , y juzgar si se ajusta a lo esperado teniendo en cuenta que

$$f(y_{t+1} | \mathbf{I}_{t-1}) \sim N_N(\mathbf{Z}_{t+1} \mathbf{S} \mathbf{a}_{t/t-1}, [\mathbf{Z}_{t+1} [\mathbf{S} \mathbf{P}_{t/t-1} \mathbf{S}' + \mathbf{Q}] \mathbf{Z}_{t+1}' + \mathbf{H}]).$$

Si y_{t+1} es concordante con lo esperado, y_t es *outlier* AO; si, por el contrario, y_{t+1} no es concordante con lo esperado, y_t es *outlier* IO.

Si el outlier es AO, el valor de y_t no es informativo y por tanto, la expresión

$$f(\mathbf{a}_t | \mathbf{I}_t) = \frac{f(y_t | \mathbf{a}_t) f(\mathbf{a}_t | \mathbf{I}_{t-1})}{f(y_t | \mathbf{I}_{t-1})}$$

se debe transformar en

$$f(\mathbf{a}_t | \mathbf{I}_t) = f(\mathbf{a}_t | \mathbf{I}_{t-1}),$$

es decir, no se debe tener en cuenta y_t . Si el outlier es IO, la función de densidad $f(\mathbf{a}_t | \mathbf{a}_{t-1})$ no es informativa y, por tanto,

$$f(\mathbf{a}_t | \mathbf{I}_{t-1}) = \int_{-\infty}^{\infty} f(\mathbf{a}_t | \mathbf{a}_{t-1}) f(\mathbf{a}_{t-1} | \mathbf{I}_{t-1}) d\mathbf{a}_{t-1} = \int_{-\infty}^{\infty} C f(\mathbf{a}_{t-1} | \mathbf{I}_{t-1}) d\mathbf{a}_{t-1} = C,$$

siendo C una constante; es decir, la función de densidad de la predicción no es informativa, ya que la información hasta $t-1$ no informa sobre el estado en t . En este caso, es preciso comenzar otra vez la aplicación del filtro de Kalman a partir de la observación anómala.

Por otro lado, se debe tener en cuenta que este tipo de observaciones pueden generar dudas sobre los supuestos distribucionales subyacentes en el modelo. Cuando no se trabaja bajo los supuestos de normalidad no hay garantías de que la estimación proporcionada por el filtro de Kalman coincida con la media condicional del vector de estado pero, la incorrelación entre las perturbaciones es suficiente para que el filtro de Kalman proporcione estimaciones óptimas lineales.

3. Modelos estructurales y filtro de Kalman

⁴⁵ Véase Meinhold y Singpurwalla (1989).

Como ya se ha comentado, los modelos estructurales de series temporales constituyen un caso particular de los modelos en el espacio de los estados donde los componentes típicos de la serie temporal son las variables de estado inobservables y la propia serie es la magnitud observable relacionada con el vector de variables de estado.

Supóngase una serie temporal unidimensional $y_t = \{y_1, \dots, y_T\}$ cuyas características más sobresalientes se pueden representar mediante un modelo estructural básico. Si se supone que s es par y se adopta la representación trigonométrica de la estacionalidad, el modelo anterior queda especificado por las ecuaciones siguientes

$$y_t = \mu_t + \gamma_t + \varepsilon_t, \quad \varepsilon_t \sim NID(0, \sigma_\varepsilon^2)$$

$$\mu_t = \mu_{t-1} + \beta_{t-1} + \eta_t, \quad \eta_t \sim NID(0, \sigma_\eta^2)$$

$$\beta_t = \beta_{t-1} + \zeta_t, \quad \zeta_t \sim NID(0, \sigma_\zeta^2)$$

$$\gamma_t = \sum_{j=1}^{\lfloor s/2 \rfloor} \gamma_{jt},$$

$$\gamma_{jt} = \cos\left(\frac{2\pi j}{s}\right) \gamma_{j,t-1} + \text{sen}\left(\frac{2\pi j}{s}\right) \gamma_{j,t-1}^* + \omega_{jt},$$

$$\gamma_{j,t}^* = -\text{sen}\left(\frac{2\pi j}{s}\right) \gamma_{j,t-1} + \cos\left(\frac{2\pi j}{s}\right) \gamma_{j,t-1}^* + \omega_{jt}^*,$$

para $j = 1, \dots, (s/2)-1$, y

$$\mathbf{g}_{\frac{s}{2},t} = -\mathbf{g}_{\frac{s}{2},t-1} + \mathbf{w}_{\frac{s}{2},t}$$

donde

$$\omega_{jt} \sim NID(0, \sigma_\omega^2),$$

$$\omega_{jt}^* \sim NID(0, \sigma_\omega^2),$$

y

$$\text{Cov}\left[\omega_{jt}, \omega_{jt}^*\right] = 0.$$

Por tanto, los elementos que aportan información sobre el sistema son los $s + 1$ términos del vector de estado

$$\mathbf{a}_t = \left(\mu_t, \beta_t, \gamma_{1,t}, \gamma_{1,t}^*, \dots, \gamma_{\frac{s}{2}-1,t}, \gamma_{\frac{s}{2}-1,t}^*, \gamma_{\frac{s}{2},t} \right)'$$

La ecuación de transición viene dada por

$$\mathbf{a}_t = \mathbf{S}\mathbf{a}_{t-1} + \mathbf{u}_t,$$

donde \mathbf{u}_t es un vector $(s+1) \times 1$ definido como

$$\mathbf{u}_t = \left(\eta_t, \zeta_t, \omega_{1,t}, \omega_{1,t}^*, \dots, \omega_{\frac{s}{2}-1,t}, \omega_{\frac{s}{2}-1,t}^*, \omega_{\frac{s}{2},t} \right)',$$

tal que $\mathbf{u}_t \sim NID(\mathbf{0}, \mathbf{Q})$, siendo $\mathbf{Q} = \text{diag}(\sigma_\eta^2, \sigma_\zeta^2, \sigma_\omega^2 \mathbf{I}_{s-1})$. La matriz \mathbf{S} está definida como,

$$\mathbf{S} = \begin{pmatrix} 1 & 1 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & 0 & \cos(2\pi/s) & \text{sen}(2\pi/s) & \cdots & 0 & 0 & 0 & 0 \\ 0 & 0 & -\text{sen}(2\pi/s) & \cos(2\pi/s) & \cdots & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & \cos\left(2\pi\left(\frac{s}{2}-1\right)/s\right) & \text{sen}\left(2\pi\left(\frac{s}{2}-1\right)/s\right) & 0 & 0 \\ 0 & 0 & 0 & 0 & \cdots & -\text{sen}\left(2\pi\left(\frac{s}{2}-1\right)/s\right) & \cos\left(2\pi\left(\frac{s}{2}-1\right)/s\right) & 0 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \end{pmatrix}$$

y la ecuación de medida adopta la expresión

$$\mathbf{y}_t = \mathbf{Z}_t \mathbf{a}_t + \mathbf{e}_t,$$

donde $\mathbf{Z}_t = (1, 0, 1, 0, \dots, 1, 0, 1)^{46}$.

Si se incorporan al modelo variables explicativas, la ecuación de medida quedaría especificada como

$$y_t = \mu_t + \gamma_t + \sum_{j=1}^k \delta_j x_{j,t} + \varepsilon_t.$$

A efectos ilustrativos, si se elimina el componente estacional, entonces el modelo definido como,

$$\begin{aligned} y_t &= \mu_t + \sum_{j=1}^k \delta_j x_{j,t} + \varepsilon_t, & \varepsilon_t &\sim NID(0, \sigma_\varepsilon^2) \\ \mu_t &= \mu_{t-1} + \beta_{t-1} + \eta_t, & \eta_t &\sim NID(0, \sigma_\eta^2) \end{aligned}$$

⁴⁶ Por ejemplo, en el caso de datos trimestrales ($s=4$), $\mathbf{a}_t = (\mu_t, \beta_t, \gamma_{1,t}, \gamma_{1,t}^*, \gamma_{2,t})'$, $\mathbf{u}_t = (\eta_t, \zeta_t, \omega_{1,t}, \omega_{1,t}^*, \omega_{2,t})'$, $\mathbf{Q} = \text{diag}(\sigma_\eta^2, \sigma_\zeta^2, \sigma_\omega^2 \mathbf{I}_3)$ y $\mathbf{Z}_t = (1, 0, 1, 0, 1)$.

$$\begin{aligned}\beta_t &= \beta_{t-1} + \zeta_t, & \zeta_t &\sim NID(0, \sigma_\zeta^2) \\ \delta_{j,t} &= \delta_{j,t-1}, & j &= 1, \dots, k\end{aligned}$$

podría formularse como un modelo en el espacio de los estados en el que

$$\mathbf{a}_t = (\mu_t, \beta_t, \delta_{1,t}, \dots, \delta_{k,t})'$$

La ecuación de transición conviene escribirla como

$$\mathbf{a}_t = \mathbf{S}\mathbf{a}_{t-1} + \mathbf{R}\mathbf{u}_t,$$

donde \mathbf{u}_t es un vector 2×1 definido como $\mathbf{u}_t = (\eta_t, \zeta_t)'$, tal que $\mathbf{u}_t \sim NID(\mathbf{0}, \mathbf{Q})$, siendo \mathbf{Q} una matriz definida como $\mathbf{Q} = \text{diag}(\sigma_\eta^2, \sigma_\zeta^2)$, y \mathbf{R} una matriz de dimensión $(k+2) \times 2$ tal que

$$\mathbf{R}' = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \end{pmatrix},$$

es decir, la matriz \mathbf{R} contiene dos vectores columna que seleccionan los términos de perturbación correspondientes a los elementos estocásticos del vector de estado. La matriz \mathbf{S} es una matriz de dimensión $(k+2) \times (k+2)$ definida como:

$$\mathbf{S} = \begin{pmatrix} 1 & 1 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{pmatrix}$$

y la ecuación de medida adopta la expresión

$$\mathbf{y}_t = \mathbf{Z}_t \mathbf{a}_t + \mathbf{e}_t,$$

donde $\mathbf{Z}_t = (1, 0, x_{1,t}, \dots, x_{k,t})$.

Para el modelo univariante de nivel local, la especificación anterior se transforma en la ecuación de medida

$$y_t = \alpha_t + \varepsilon_t,$$

con $\alpha_t = \mu_t$, donde ε_t es un término de perturbación tal que

$$E[\varepsilon_t] = 0,$$

$$\text{Var}[\varepsilon_t] = h$$

y

$$E[\varepsilon_t \varepsilon_s] = 0, \quad \forall s \neq t, \quad s, t = 1, \dots, T,$$

y la ecuación de transición

$$\alpha_t = \alpha_{t-1} + u_t,$$

con $u_t = \eta_t$, donde η_t es un término de perturbación tal que

$$E[\eta_t] = 0,$$

$$\text{Var}[\eta_t] = q$$

y

$$E[\eta_t \eta_s] = 0, \quad \forall s \neq t, \quad s, t = 1, \dots, T.$$

Además, se supone que

$$E[\mu_0] = m_0,$$

$$\text{Var}[\mu_0] = P_0,$$

$$E[\varepsilon_t \mu_0] = 0$$

$$E[\eta_t \mu_0] = 0$$

y

$$E[\varepsilon_t \eta_s] = 0, \quad \forall s \neq t, \quad s, t = 1, \dots, T.$$

En este último caso, la estimación del componente de nivel a través del filtro de Kalman y la estimación máximo verosímil de los hiperparámetros se efectúa como se indica a continuación.

Sea $m_{t/t-1}$, la estimación óptima del estado en el instante t con la información disponible hasta $t-1$ y sea m_t , la estimación óptima del estado en el instante t con la información disponible hasta t , es decir, una vez observada la medida y_t .

Las ecuaciones de predicción son

$$m_{t/t-1} = m_{t-1}$$

$$P_{t/t-1} = P_{t-1} + q$$

y las ecuaciones de actualización son

$$m_t = m_{t/t-1} + \frac{P_{t/t-1}}{P_{t/t-1} + h} [y_t - m_{t/t-1}]$$

$$P_t = \left(1 - \frac{P_{t/t-1}}{P_{t/t-1} + h} \right) P_{t/t-1}.$$

Dados m_0 y P_0 , se puede obtener el estimador óptimo de μ_t del siguiente modo:

$$\left. \begin{array}{l} m_{t/t-1} = m_{t-1} \\ P_{t/t-1} = P_{t-1} + q \end{array} \right\} \Rightarrow \begin{cases} m_t = m_{t/t-1} + \frac{P_{t/t-1}}{P_{t/t-1} + h} [y_t - m_{t/t-1}] \\ P_t = \left(1 - \frac{P_{t/t-1}}{P_{t/t-1} + h}\right) P_{t/t-1} \end{cases}$$

Por ejemplo, la predicción y la estimación filtrada del estado en el instante $t = 1$ vienen dadas por

$$\left. \begin{array}{l} m_{1/0} = m_0 \\ P_{1/0} = P_0 + q \end{array} \right\} \Rightarrow \begin{cases} m_1 = m_{1/0} + \frac{P_{1/0}}{P_{1/0} + h} [y_1 - m_{1/0}] = m_0 + \frac{P_0 + q}{P_0 + q + h} [y_1 - m_0] \\ P_1 = \left(1 - \frac{P_0 + q}{P_0 + q + h}\right) (P_0 + q) \end{cases}$$

y en el momento $t = 2$ por

$$\left. \begin{array}{l} m_{2/1} = m_1 = m_0 + \frac{P_0 + q}{P_0 + q + h} [y_1 - m_0] \\ P_{2/1} = P_1 + q = \left(1 - \frac{P_0 + q}{P_0 + q + h}\right) (P_0 + q) + q \end{array} \right\} \Rightarrow \begin{cases} m_2 = m_{2/1} + \frac{P_{2/1}}{P_{2/1} + h} [y_2 - m_{2/1}] \\ P_2 = \left(1 - \frac{P_{2/1}}{P_{2/1} + h}\right) P_{2/1} \end{cases}$$

Las ecuaciones de la estimación alisada son

$$m_{t/T} = m_t + P_t^* [m_{t+1/T} - m_t]$$

y

$$P_{t/T} = P_t + P_t^* (P_{t+1/T} - P_{t+1/t}) P_t^*,$$

donde

$$P_t^* = P_t P_{t+1/t}^{-1} = \frac{P_t}{P_t + q}, \quad t = T-1, \dots, 1.$$

Por tanto,

$$m_{t/T} = m_t + \frac{P_t}{P_t + q} [m_{t+1/T} - m_t] = \left(1 - \frac{P_t}{P_t + q}\right) m_t + \frac{P_t}{P_t + q} m_{t+1/T}.$$

Es decir, la estimación alisada del estado en el momento t es, simplemente, una media ponderada entre la estimación obtenida por el filtro de Kalman para el estado en el momento t y la estimación alisada del estado en el momento $t + 1$.

Si el modelo es gaussiano, es decir, si ε_t , η_t y μ_0 se distribuyen como variables aleatorias normales, entonces las estimaciones filtradas m_t y P_t proporcionadas por el filtro de Kalman coinciden con la media y la matriz de covarianzas de la distribución de μ_t condicionada a la información en t .

Como se puede observar, las estimaciones del nivel son función de m_0 , P_0 , q y h . Si se supone que $P_0 = 0$, m_0 se puede considerar como un parámetro más que debe ser estimado junto con los hiperparámetros q y h . Si se asume que el modelo es gaussiano, estos parámetros desconocidos, \mathbf{y} , se pueden estimar por máxima verosimilitud. La función de densidad conjunta del proceso estocástico observable $\{y_t\}$, $t = 1, 2, \dots, T$, se puede expresar en términos de las innovaciones o residuos e_t , que se distribuyen como normales independientes con $E[e_t] = 0$ y $Var[e_t] = F_t$. Es decir,

$$L(e_1, \dots, e_T, \mathbf{y}) = \prod_{t=1}^T \left\{ \frac{1}{\sqrt{2\pi}\sqrt{F_t}} \exp\left(-\frac{1}{2} \frac{e_t^2}{F_t}\right) \right\} = \frac{1}{(2\pi)^{T/2} \sqrt{F_1 \dots F_T}} \exp\left(-\frac{1}{2} \sum_{t=1}^T \frac{e_t^2}{F_t}\right)$$

$$\ln L(e_1, \dots, e_T, \mathbf{y}) = -\frac{T}{2} \ln(2\pi) - \frac{1}{2} \sum_{t=1}^T \ln(F_t) - \frac{1}{2} \sum_{t=1}^T \frac{e_t^2}{F_t}$$

donde $e_t = y_t - m_{t/t-1}$ y $F_t = P_{t/t-1} + h = P_{t-1/t-1} + q + h$.

Asignados unos valores iniciales se pueden obtener recursivamente las expresiones de $m_{t/t-1}$ y P_{t-1} como funciones de q y h . Una vez expresada la función de verosimilitud en términos de estos dos parámetros se obtienen directamente sus estimaciones máximo verosímiles. La función de verosimilitud se puede reparametrizar haciendo que todas las varianzas sean proporcionales a h , es decir,

$$Var[\eta_t] = hq,$$

$$Var[\alpha_0] = hP_0$$

y

$$Var[e_t] = hf_t.$$

Entonces,

$$\begin{aligned} \ln L(e_1, \dots, e_T, \mathbf{y}^*, h) &= -\frac{T}{2} \ln(2\mathbf{p}) - \frac{1}{2} \sum_{t=1}^T \ln(hf_t) - \frac{1}{2} \sum_{t=1}^T \frac{e_t^2}{hf_t} \\ &= -\frac{T}{2} \ln(2\mathbf{p}) - \frac{T}{2} \ln(h) - \frac{1}{2} \sum_{t=1}^T \ln(f_t) - \frac{1}{2h} \sum_{t=1}^T \frac{e_t^2}{f_t} \end{aligned}$$

y se tiene que

$$\frac{\partial \ln L(e_1, \dots, e_T, \mathbf{y}^*, h)}{\partial h} = -\frac{T}{2h} + \frac{1}{2h^2} \sum_{t=1}^T \frac{e_t^2}{f_t},$$

de modo que, igualando a cero la derivada anterior, resulta que

$$\hat{h}(\mathbf{y}^*) = \frac{1}{T} \sum_{t=1}^T \frac{e_t^2}{f_t}$$

Finalmente, se obtiene la función de verosimilitud concentrada

$$\ln L(e_1, \dots, e_T, \mathbf{y}^*, h) = -\frac{T}{2} \ln(2\mathbf{p}) - \frac{T}{2} \ln \left(\frac{1}{T} \sum_{t=1}^T \frac{e_t^2}{f_t} \right) - \frac{1}{2} \sum_{t=1}^T \ln(f_t) - \frac{T}{2},$$

que debe ser maximizada respecto a los elementos del vector \mathbf{y}^* .

En modelos más generales, la función de verosimilitud suele ser reparametrizada con objeto de reducir la dimensionalidad de la búsqueda numérica en el proceso de optimización. Por ejemplo, se puede optimizar respecto a σ_ε^2 y las varianzas de los elementos del vector de estado, o bien, utilizar σ_ε^2 y los ratios señal ruido⁴⁷ como parámetros que deben ser estimados. Entonces, puede obtenerse la estimación máximo verosímil de σ_ε^2 como función del resto de parámetros y sustituir esta estimación en la función de verosimilitud. Se obtiene de este modo la función de verosimilitud concentrada, que sólo dependerá de los ratios señal ruido o de las varianzas de los elementos del vector de estado.

En el caso multivariante, la estimación del modelo formulado en el espacio de los estados puede llevarse a cabo también a través del filtro de Kalman⁴⁸. Bajo el supuesto de homogeneidad, que implica que las distintas series tienen las mismas propiedades estocásticas, el modelo puede ser manejado aplicando el filtro de Kalman univariante a cada serie⁴⁹; sin embargo, si los hiperparámetros son desconocidos, usar un filtro de Kalman multivariante permite obtener estimaciones alisadas más eficientes de los componentes⁵⁰.

En general, dado un modelo estructural multivariante para la serie \mathbf{y}_t de dimensión $N \times 1$, puede ser conveniente, en términos computacionales, tratar el vector N -dimensional de T

⁴⁷ Los ratios señal ruido se definen como el cociente entre la varianza de cada uno de los componentes y la varianza del componente irregular.

⁴⁸ Véase Harvey y Shephard (1993:20-21).

⁴⁹ Para contrastar la homogeneidad, se puede recurrir al test de multiplicadores de Lagrange diseñado por Fernández y Harvey (1990).

⁵⁰ Véase Harvey (1989:438-442).

observaciones como una serie univariante de $T \times N$ observaciones⁵¹. Por ejemplo, sea el modelo definido por las ecuaciones

$$\begin{aligned} \mathbf{y}_t &= \mathbf{Z}_t \mathbf{a}_t + \mathbf{e}_t, \\ \mathbf{a}_t &= \mathbf{S}_t \mathbf{a}_{t-1} + \mathbf{R}_t \mathbf{u}_t, \end{aligned}$$

donde

$$\begin{aligned} \mathbf{e}_t &\sim N(\mathbf{0}, \mathbf{H}_t), \\ \mathbf{u}_t &\sim N(\mathbf{0}, \mathbf{Q}_t), \end{aligned}$$

siendo

$$\mathbf{y}_t = (y_{1,t}, \dots, y_{N,t})', \quad \mathbf{a}_t = (\alpha_{1,t}, \dots, \alpha_{m,t})', \quad \mathbf{e}_t = (\varepsilon_{1,t}, \dots, \varepsilon_{N,t})'.$$

Además, $\mathbf{Z}_t = (Z_{1,t}, \dots, Z_{N,t})'$ es un vector columna $N \times 1$ cuyos elementos son vectores fila de dimensión $1 \times m$, y, por último, la matriz \mathbf{H}_t , de dimensión $N \times N$, se define como

$$\mathbf{H}_t = \begin{pmatrix} \sigma_{1,t}^2 & 0 & \dots & 0 \\ 0 & \sigma_{2,t}^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_{N,t}^2 \end{pmatrix}.$$

Entonces, las T observaciones del vector $\mathbf{y}_t = (y_{1,t}, \dots, y_{N,t})'$, $t = 1, \dots, T$, se pueden escribir como una serie univariante $\{y_{i,t}\}_{i=1, \dots, N, t=1, \dots, T}$, definida como el conjunto de observaciones $\{y_{1,1}, \dots, y_{N,1}, y_{1,2}, \dots, y_{N,2}, \dots, y_{1,T}, \dots, y_{N,T}\}$. La ecuación de medida para esta representación univariante puede formularse ahora como

$$y_{i,t} = Z_{i,t} \alpha_t + \varepsilon_{i,t}, \quad i = 1, \dots, N, \quad t = 1, \dots, T,$$

y sobre esta nueva serie se pueden aplicar los algoritmos de filtrado y alisado⁵².

Los modelos estructurales descritos en los párrafos precedentes se caracterizan por asumir relaciones lineales e invariantes en el tiempo y se supone, además, que las observaciones se han generado a partir de distribuciones gaussianas. Un modelo lineal para el sistema se justifica, entre otras razones, por su manejo más fácil y porque algunas no linealidades pueden linealizarse⁵³. Por

⁵¹ Véase Durbin y Koopman (2001:128-134).

⁵² Si \mathbf{H}_t no es diagonal, puede introducirse \mathbf{e}_t como elemento del vector de estado, o bien, transformar las observaciones de modo que se diagonalice la matriz \mathbf{H}_t (véase Durbin y Koopman, 2001:131-132).

⁵³ Si las ecuaciones de transición o de medida fueran no lineales, es preciso diseñar un filtro que consiga linealizar las relaciones. Este tipo de filtros se denomina filtros de Kalman extendidos (véase Welch y Bishop, 2001).

otro lado, Maybeck (1982:7-9) justifica también el supuesto de gaussianidad y comportamiento ruido blanco para los términos de perturbación⁵⁴.

Además de la mayor facilidad para su tratamiento, los modelos estructurales de esta clase son los que con menor dificultad se relacionan con otros enfoques. Y lo cierto es que el modelo estructural lineal gaussiano, tras adecuadas transformaciones de las observaciones, permite analizar muchas de las series temporales que se encuentran en la práctica. Sin embargo, hay otras situaciones en las que una distribución no normal puede proporcionar un mejor modelo⁵⁵. Por ejemplo, si las observaciones registradas son datos de recuento puede resultar más apropiada una distribución de Poisson y algunas distribuciones con colas pesadas suelen ser convenientes para tratar observaciones anómalas.

También es importante permitir no linealidades con objeto, por ejemplo, de recoger situaciones en las que la tendencia y la estacionalidad se combinan de modo multiplicativo, mientras que el componente irregular se incorpora al modelo de forma aditiva. Estas no linealidades pueden introducirse en la ecuación de medida o en la ecuación de transición.

Entonces, la representación en el espacio de los estados viene dada, por ejemplo, por

$$y_t = Z_t(\alpha_t) + \varepsilon_t$$
$$\alpha_t = S_t(\alpha_{t-1}) + u_t,$$

⁵⁴ El supuesto de gaussianidad se puede justificar físicamente por el hecho de que cualquier ruido o error de sistema o medida se debe, generalmente, a un número de fuentes pequeñas y puede ser mostrado matemáticamente que cuando un número de variables aleatorias independientes se expresan de forma agregada, el efecto final puede ser descrito con bastante proximidad por una función de densidad de probabilidad gaussiana, a pesar de la forma de las densidades individuales. Al igual que antes, la gaussianidad hace que las operaciones matemáticas sean tratables; pero, además, normalmente un ingeniero conocerá, como mucho, los estadísticos de primer y segundo orden de un proceso ruidoso. En ausencia de estadísticos de cualquier orden mayor, no hay solución más conveniente que asumir gaussianidad. Este supuesto de índole práctica es también aplicable a la hipótesis de comportamiento ruido blanco. Maybeck recurre a sistemas físicos para mostrar que los comportamientos ruidosos suelen recogerse adecuadamente a través de procesos ruido blanco o de filtros basados en procesos ruido blanco. Pero, además de este hecho, el autor reconoce que otros supuestos no dan lugar a formulaciones tratables del problema de estimación y control. Afortunadamente, la clase de modelos en los que las operaciones matemáticas resultan tratables también proporcionan representaciones adecuadas para muchas aplicaciones de interés.

⁵⁵ Incluso en el caso de que los términos de perturbación del modelo sean no gaussianos, el filtro de Kalman puede ser empleado para obtener una proyección lineal de la serie a partir de sus valores pasados. El procedimiento de estimación *quasi*-máximo verosímil proporciona estimadores consistentes y asintóticamente normales de los elementos de las matrices del sistema (Hamilton, 1994:389).

donde Z_t y S_t están definidas de modo que y_t depende de α_t de forma no lineal y lo mismo ocurre con la relación entre α_t y α_{t-1} ⁵⁶.

ANEXO I. Propiedades estadísticas de un modelo en el espacio de los estados

Sea un sistema descrito por un vector $m \times 1$ de estados \mathbf{a}_t , en los instantes del tiempo $t = 1, 2, \dots, T$ y sea el vector $N \times 1$ de variables observables y_t . El modelo básico de la representación en el espacio de los estados consta de la ecuación de transición,

$$\mathbf{a}_t = S\mathbf{a}_{t-1} + \mathbf{u}_t, \quad t = 1, \dots, T,$$

y de la ecuación de medida

$$y_t = Z_t \mathbf{a}_t + \mathbf{e}_t, \quad t = 1, \dots, T,$$

donde S es una matriz de transición $m \times m$, \mathbf{u}_t es el vector $m \times 1$ de perturbaciones de la ecuación de transición tal que $\mathbf{u}_t \sim i.i.d. N_m(\mathbf{0}, \mathbf{Q})$, Z_t es una matriz de medida $N \times m$, y \mathbf{e}_t es el vector $N \times 1$ de perturbaciones de la ecuación de medida tal que $\mathbf{e}_t \sim i.i.d. N_N(\mathbf{0}, \mathbf{H})$.

Se asume que $E[\mathbf{a}_0] = \mathbf{a}_0$; $Var[\mathbf{a}_0] = \mathbf{P}_0$; $E[\mathbf{e}_t \mathbf{u}_s'] = \mathbf{0}$, $\forall t, s = 1, \dots, T$; $E[\mathbf{e}_t \mathbf{a}_0'] = \mathbf{0}$, $t = 1, \dots, T$ y $E[\mathbf{u}_t \mathbf{a}_0'] = \mathbf{0}$, $t = 1, \dots, T$.

Por sustitución recursiva, la ecuación de transición se puede expresar como

$$\mathbf{a}_t = S\mathbf{a}_{t-1} + \mathbf{u}_t = S(S\mathbf{a}_{t-2} + \mathbf{u}_{t-1}) + \mathbf{u}_t = S^2\mathbf{a}_{t-2} + S\mathbf{u}_{t-1} + \mathbf{u}_t = \dots = S^t\mathbf{a}_0 + \sum_{i=0}^{t-1} S^i \mathbf{u}_{t-i}$$

y la ecuación de medida, entonces, como

$$y_t = Z_t \left[S^t \mathbf{a}_0 + \sum_{i=0}^{t-1} S^i \mathbf{u}_{t-i} \right] + \mathbf{e}_t$$

A) Momentos de primer y segundo orden de las variables de estado \mathbf{a}_t y de las variables observables y_t .

- $E[\mathbf{a}_t] = S^t E[\mathbf{a}_0] = S^t \mathbf{a}_0$
- $E[y_t] = Z_t S^t E[\mathbf{a}_0] = Z_t S^t \mathbf{a}_0$

⁵⁶ Durbin y Koopman (2001:177-239) hacen un extenso tratamiento de los modelos estructurales no lineales y no gaussianos.

- Para calcular $Cov[\mathbf{a}_t, \mathbf{a}_{t-k}] = E[(\mathbf{a}_t - E[\mathbf{a}_t])(\mathbf{a}_{t-k} - E[\mathbf{a}_{t-k}])']$, se puede tener en cuenta que

$$\mathbf{a}_t = \mathbf{S}^k \mathbf{a}_{t-k} + \sum_{i=0}^{k-1} \mathbf{S}^i \mathbf{u}_{t-i}$$

y, entonces,

$$\begin{aligned} E[(\mathbf{a}_t - E[\mathbf{a}_t])(\mathbf{a}_{t-k} - E[\mathbf{a}_{t-k}])'] &= E\left[\left(\mathbf{S}^k \mathbf{a}_{t-k} + \sum_{i=0}^{k-1} \mathbf{S}^i \mathbf{u}_{t-i} - \mathbf{S}^k E[\mathbf{a}_{t-k}]\right)(\mathbf{a}_{t-k} - E[\mathbf{a}_{t-k}])'\right] \\ &= E\left[\left(\mathbf{S}^k (\mathbf{a}_{t-k} - E[\mathbf{a}_{t-k}]) + \sum_{i=0}^{k-1} \mathbf{S}^i \mathbf{u}_{t-i}\right)(\mathbf{a}_{t-k} - E[\mathbf{a}_{t-k}])'\right] \\ &= E\left[\left(\mathbf{S}^k (\mathbf{a}_{t-k} - E[\mathbf{a}_{t-k}])\right)(\mathbf{a}_{t-k} - E[\mathbf{a}_{t-k}])'\right] \\ &\quad + E\left[\left(\sum_{i=0}^{k-1} \mathbf{S}^i \mathbf{u}_{t-i}\right)(\mathbf{a}_{t-k} - E[\mathbf{a}_{t-k}])'\right] \end{aligned}$$

Considerando entonces que

$$\mathbf{a}_{t-k} - E[\mathbf{a}_{t-k}] = \mathbf{S}^{t-k} \mathbf{a}_0 + \sum_{i=0}^{t-k-1} \mathbf{S}^i \mathbf{u}_{t-k-i} - \mathbf{S}^{t-k} \mathbf{a}_0 = \mathbf{S}^{t-k} (\mathbf{a}_0 - \mathbf{a}_0) + \sum_{i=0}^{t-k-1} \mathbf{S}^i \mathbf{u}_{t-k-i}$$

resulta que

$$\begin{aligned}
& E\left[(\mathbf{a}_t - E[\mathbf{a}_t])(\mathbf{a}_{t-k} - E[\mathbf{a}_{t-k}])'\right] \\
&= E\left[\left(\mathbf{s}^k \left(\mathbf{s}^{t-k}(\mathbf{a}_0 - \mathbf{a}_0) + \sum_{i=0}^{t-k-1} \mathbf{s}^i u_{t-k-i}\right)\right) \left(\mathbf{s}^{t-k}(\mathbf{a}_0 - \mathbf{a}_0) + \sum_{i=0}^{t-k-1} \mathbf{s}^i u_{t-k-i}\right)'\right] \\
&+ E\left[\left(\sum_{i=0}^{k-1} \mathbf{s}^i u_{t-i}\right) \left(\mathbf{s}^{t-k}(\mathbf{a}_0 - \mathbf{a}_0) + \sum_{i=0}^{t-k-1} \mathbf{s}^i u_{t-k-i}\right)'\right] \\
&= \mathbf{s}' E\left[(\mathbf{a}_0 - \mathbf{a}_0)(\mathbf{a}_0 - \mathbf{a}_0)'\right] (\mathbf{s}^{t-k})' + \mathbf{s}^k \sum_{i=0}^{t-k-1} \mathbf{s}^i E\left[u_{t-k-i}(\mathbf{a}_0 - \mathbf{a}_0)'\right] (\mathbf{s}^{t-k})' \\
&+ \mathbf{s}' \sum_{i=0}^{t-k-1} E[(\mathbf{a}_0 - \mathbf{a}_0) u'_{t-k-i}] (\mathbf{s}^i)' \\
&+ \sum_{i=0}^{t-k-1} \mathbf{s}^{k+i} E[u_{t-k-i} u'_{t-k-i}] (\mathbf{s}^i)' + \sum_{\substack{i,j=0 \\ i \neq j}}^{t-k-1} \mathbf{s}^{k+i} E[u_{t-k-i} u'_{t-k-j}] (\mathbf{s}^j)' \\
&+ \sum_{i=0}^{k-1} \mathbf{s}^i E\left[u_{t-i}(\mathbf{a}_0 - \mathbf{a}_0)'\right] (\mathbf{s}^{t-k})' + \sum_{i=0}^{k-1} \sum_{j=0}^{t-k-1} \mathbf{s}^i E[u_{t-i} u'_{t-k-j}] (\mathbf{s}^j)' \\
&= \mathbf{s}' \text{Var}[\mathbf{a}_0] (\mathbf{s}^{t-k})' + \sum_{i=0}^{t-k-1} \mathbf{s}^{k+i} \text{Var}[u_{t-k-i}] (\mathbf{s}^i)' = \mathbf{s}' \mathbf{P}_0 (\mathbf{s}^{t-k})' + \sum_{i=0}^{t-k-1} \mathbf{s}^{k+i} \mathbf{Q} (\mathbf{s}^i)' \\
&= \left[\mathbf{s}' \mathbf{P}_0 (\mathbf{s}^{t-k})' + \sum_{i=0}^{t-k-1} \mathbf{s}^{k+i} \mathbf{Q} (\mathbf{s}^i)' \right] = \mathbf{P}_{t,t-k} \\
&- E\left[(\mathbf{y}_t - E[\mathbf{y}_t])(\mathbf{y}_{t-k} - E[\mathbf{y}_{t-k}])'\right]
\end{aligned}$$

$$\begin{aligned}
&= E\left[(\mathbf{Z}_t \mathbf{a}_t + \mathbf{e}_t - \mathbf{Z}_t E[\mathbf{a}_t])(\mathbf{Z}_{t-k} \mathbf{a}_{t-k} + \mathbf{e}_{t-k} - \mathbf{Z}_{t-k} E[\mathbf{a}_{t-k}])'\right] \\
&= E\left[(\mathbf{Z}_t (\mathbf{a}_t - E[\mathbf{a}_t]) + \mathbf{e}_t)(\mathbf{Z}_{t-k} (\mathbf{a}_{t-k} - E[\mathbf{a}_{t-k}]) + \mathbf{e}_{t-k})'\right] \\
&= E\left[\mathbf{Z}_t (\mathbf{a}_t - E[\mathbf{a}_t])(\mathbf{a}_{t-k} - E[\mathbf{a}_{t-k}])' \mathbf{Z}'_{t-k}\right] + E\left[\mathbf{e}_t (\mathbf{a}_{t-k} - E[\mathbf{a}_{t-k}])' \mathbf{Z}'_{t-k}\right] \\
&\quad + E\left[\mathbf{Z}_t (\mathbf{a}_t - E[\mathbf{a}_t]) \mathbf{e}'_{t-k}\right] + E\left[\mathbf{e}_t \mathbf{e}'_{t-k}\right] \\
&= \mathbf{Z}_t E\left[(\mathbf{a}_t - E[\mathbf{a}_t])(\mathbf{a}_{t-k} - E[\mathbf{a}_{t-k}])'\right] \mathbf{Z}'_{t-k} + E\left[\mathbf{e}_t (\mathbf{a}_{t-k} - E[\mathbf{a}_{t-k}])'\right] \mathbf{Z}'_{t-k} \\
&\quad + \mathbf{Z}_t E\left[(\mathbf{a}_t - E[\mathbf{a}_t]) \mathbf{e}'_{t-k}\right] + E\left[\mathbf{e}_t \mathbf{e}'_{t-k}\right] \\
&= \mathbf{Z}_t \mathbf{P}_{t,t-k} \mathbf{Z}'_{t-k}, \text{ si } k \neq 0 \\
- & E\left[(\mathbf{y}_t - E[\mathbf{y}_t])(\mathbf{y}_t - E[\mathbf{y}_t])'\right] \\
&= E\left[(\mathbf{Z}_t \mathbf{a}_t + \mathbf{e}_t - \mathbf{Z}_t E[\mathbf{a}_t])(\mathbf{Z}_t \mathbf{a}_t + \mathbf{e}_t - \mathbf{Z}_t E[\mathbf{a}_t])'\right] \\
&= E\left[(\mathbf{Z}_t (\mathbf{a}_t - E[\mathbf{a}_t]) + \mathbf{e}_t)(\mathbf{Z}_t (\mathbf{a}_t - E[\mathbf{a}_t]) + \mathbf{e}_t)'\right] \\
&= E\left[\mathbf{Z}_t (\mathbf{a}_t - E[\mathbf{a}_t])(\mathbf{a}_t - E[\mathbf{a}_t])' \mathbf{Z}'_t\right] + E\left[\mathbf{e}_t (\mathbf{a}_t - E[\mathbf{a}_t])' \mathbf{Z}'_t\right] \\
&\quad + E\left[\mathbf{Z}_t (\mathbf{a}_t - E[\mathbf{a}_t]) \mathbf{e}'_t\right] + E\left[\mathbf{e}_t \mathbf{e}'_t\right] \\
&= \mathbf{Z}_t E\left[(\mathbf{a}_t - E[\mathbf{a}_t])(\mathbf{a}_t - E[\mathbf{a}_t])'\right] \mathbf{Z}'_t + E\left[\mathbf{e}_t (\mathbf{a}_t - E[\mathbf{a}_t])'\right] \mathbf{Z}'_t \\
&\quad + \mathbf{Z}_t E\left[(\mathbf{a}_t - E[\mathbf{a}_t]) \mathbf{e}'_t\right] + E\left[\mathbf{e}_t \mathbf{e}'_t\right] \\
&= \mathbf{Z}_t \mathbf{P}_t \mathbf{Z}'_t + \mathbf{H}
\end{aligned}$$

Sustituyendo $\mathbf{P}_{t,t-k}$ y \mathbf{P}_t por sus valores respectivos, se tiene que

$$\begin{aligned}
&E\left[(\mathbf{y}_t - E[\mathbf{y}_t])(\mathbf{y}_{t-k} - E[\mathbf{y}_{t-k}])'\right] \\
&= \mathbf{Z}_t \left[\mathbf{S}'^t \mathbf{P}_0 (\mathbf{S}^{t-k})' + \sum_{i=0}^{t-k-1} \mathbf{S}^{k+i} \mathbf{Q} (\mathbf{S}^i)' \right] (\mathbf{Z}'_{t-k}) \\
&= \left[\mathbf{Z}_t \mathbf{S}'^t \mathbf{P}_0 (\mathbf{Z}_{t-k} \mathbf{S}^{t-k})' + \mathbf{Z}_t \sum_{i=0}^{t-k-1} \mathbf{S}^{k+i} \mathbf{Q} (\mathbf{Z}_{t-k} \mathbf{S}^i)' \right], \text{ si } k \neq 0
\end{aligned}$$

y que

$$\begin{aligned}
& E\left[(\mathbf{y}_t - E[\mathbf{y}_t])(\mathbf{y}_t - E[\mathbf{y}_t])'\right] \\
&= \left[\mathbf{Z}_t \left[\mathbf{S}' \mathbf{P}_0 (\mathbf{S}')' + \sum_{i=0}^{t-1} \mathbf{S}^i \mathbf{Q} (\mathbf{S}^i)' \right] \mathbf{Z}_t' + \mathbf{H} \right] \\
&= \left[\mathbf{Z}_t \mathbf{S}' \mathbf{P}_0 (\mathbf{Z}_t \mathbf{S}')' + \mathbf{Z}_t \sum_{i=0}^{t-1} \mathbf{S}^i \mathbf{Q} (\mathbf{Z}_t \mathbf{S}^i)' + \mathbf{H} \right]
\end{aligned}$$

B) Función de verosimilitud.

Sean los vectores de vectores aleatorios $\mathbf{y} = (\mathbf{y}_1, \dots, \mathbf{y}_T)'$ $N \cdot T \times 1$, $\mathbf{a} = (\mathbf{a}_1, \dots, \mathbf{a}_T)'$ $m \cdot T \times 1$ y $\mathbf{e} = (\mathbf{e}_1, \dots, \mathbf{e}_T)'$ $N \cdot T \times 1$, y la matriz de matrices $\mathbf{Z} = \text{diag}(\mathbf{Z}_1, \dots, \mathbf{Z}_T)$ $NT \times mT$. Entonces $\mathbf{y} = \mathbf{Z}\mathbf{a} + \mathbf{e}$, donde $\mathbf{e} \sim N_{NT}(\mathbf{0}, (\mathbf{I}_T \otimes \mathbf{H}))$, con $\text{Var}[\mathbf{e}]$ una matriz $T \times T$ de matrices de orden $N \times N$ definida como

$$\text{Var}[\mathbf{e}] = \begin{bmatrix} \mathbf{H} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{H} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{H} \end{bmatrix}$$

Por otro lado, se tiene que $\mathbf{a} = \mathbf{S}_1(\mathbf{i}_T \otimes \mathbf{a}_0) + \mathbf{S}_2 \mathbf{u}$, donde $\mathbf{S}_1 = \text{diag}(\mathbf{S}, \mathbf{S}^2, \dots, \mathbf{S}^T)$, \mathbf{i}_T es el vector unitario $T \times 1$,

$$\mathbf{S}_2 = \begin{bmatrix} \mathbf{I} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{S} & \mathbf{I} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{S}^{T-1} & \mathbf{S}^{T-2} & \dots & \mathbf{I} \end{bmatrix}$$

$\mathbf{y}\mathbf{u} \sim N_{mT}(\mathbf{0}, (\mathbf{I}_T \otimes \mathbf{Q}))$.

Nótese que

$$\begin{aligned}
\mathbf{a}_1 &= \mathbf{S}\mathbf{a}_0 + \mathbf{u}_1, \\
\mathbf{a}_2 &= \mathbf{S}^2\mathbf{a}_0 + \mathbf{u}_2 + \mathbf{S}\mathbf{u}_1, \dots \\
\mathbf{a}_T &= \mathbf{S}^T\mathbf{a}_0 + \mathbf{u}_T + \mathbf{S}\mathbf{u}_{T-1} + \dots + \mathbf{S}^{T-1}\mathbf{u}_1.
\end{aligned}$$

Por tanto,

$$E[\mathbf{y}] = \mathbf{Z}E[\mathbf{a}] = \mathbf{Z}\mathbf{S}_1(\mathbf{i}_T \otimes \mathbf{a}_0) = \mathbf{Z}[E[\mathbf{a}'_1], \dots, E[\mathbf{a}'_T]]' = \mathbf{Z}\left[(\mathbf{S}\mathbf{a}_0)', \dots, (\mathbf{S}^T\mathbf{a}_0)'\right]' \equiv \mathbf{m}$$

Para calcular la varianza del vector de vectores \mathbf{y} conviene expresarlo de la forma

$$\mathbf{y} = \mathbf{Z}[\mathbf{S}_1(\mathbf{i}_T \otimes \mathbf{a}_0) + \mathbf{S}_2\mathbf{u}] + \mathbf{e}$$

Entonces,

$$\begin{aligned} & E\left[(\mathbf{y} - \mathbf{m})(\mathbf{y} - \mathbf{m})'\right] \\ &= E\left[(\mathbf{Z}\mathbf{S}_1(\mathbf{i}_T \otimes \mathbf{a}_0) - \mathbf{Z}\mathbf{S}_1(\mathbf{i}_T \otimes \mathbf{a}_0) + \mathbf{Z}\mathbf{S}_2\mathbf{u} + \mathbf{e})(\mathbf{Z}\mathbf{S}_1(\mathbf{i}_T \otimes \mathbf{a}_0) - \mathbf{Z}\mathbf{S}_1(\mathbf{i}_T \otimes \mathbf{a}_0) + \mathbf{Z}\mathbf{S}_2\mathbf{u} + \mathbf{e})'\right] \\ &= E\left[\mathbf{Z}\mathbf{S}_1\left(\mathbf{i}_T\mathbf{i}_T' \otimes (\mathbf{a}_0 - \mathbf{a}_0)(\mathbf{a}_0 - \mathbf{a}_0)'\right)\mathbf{S}_1'\mathbf{Z}'\right] + E[\mathbf{Z}\mathbf{S}_2\mathbf{u}\mathbf{u}'\mathbf{S}_2'\mathbf{Z}'] + E[\mathbf{e}\mathbf{e}'] \\ &= \mathbf{Z}\mathbf{S}_1\left(\mathbf{i}_T\mathbf{i}_T' \otimes E\left[(\mathbf{a}_0 - \mathbf{a}_0)(\mathbf{a}_0 - \mathbf{a}_0)'\right]\right)\mathbf{S}_1'\mathbf{Z}' + \mathbf{Z}\mathbf{S}_2E[\mathbf{u}\mathbf{u}']\mathbf{S}_2'\mathbf{Z}' + E[\mathbf{e}\mathbf{e}'] \\ &= \mathbf{Z}\mathbf{S}_1(\mathbf{i}_T\mathbf{i}_T' \otimes \mathbf{P}_0)\mathbf{S}_1'\mathbf{Z}' + \mathbf{Z}\mathbf{S}_2(\mathbf{I}_T \otimes \mathbf{Q})\mathbf{S}_2'\mathbf{Z}' + (\mathbf{I}_T \otimes \mathbf{H}) \equiv \mathbf{S} \end{aligned}$$

Como \mathbf{y} es normal multivariante de dimensión NT , con media \mathbf{m} y matriz de covarianzas \mathbf{S} , la función de verosimilitud es

$$L(\mathbf{y}) = \frac{\sqrt{|\mathbf{S}|^{-1}}}{(2\pi)^{NT/2}} \exp\left\{-\frac{1}{2}(\mathbf{y} - \mathbf{m})\mathbf{S}^{-1}(\mathbf{y} - \mathbf{m})'\right\}$$

y su logaritmo neperiano

$$\ln L(\mathbf{y}) = -\frac{1}{2}\ln|\mathbf{S}| - \frac{NT}{2}\ln 2\pi - \frac{1}{2}(\mathbf{y} - \mathbf{m})\mathbf{S}^{-1}(\mathbf{y} - \mathbf{m})'$$

A partir de la función de verosimilitud pueden derivarse estimaciones máximo verosímiles de los parámetros del modelo, aunque pueden surgir problemas de identificación. Si no se imponen restricciones en las matrices del sistema, los parámetros de la representación en el espacio de los estados pueden quedar sin identificar; es decir, puede ocurrir que más de un conjunto de valores de los parámetros de lugar a idéntico valor de la función de verosimilitud y , por tanto, los datos no ofrecen una guía para elegir tales valores (Hamilton, 1994:387-388)⁵⁷.

⁵⁷ Se dice que un vector de parámetros θ , del cual depende la función de verosimilitud $f(y; \theta)$ está identificado cuando $f(y; \theta_1) = f(y; \theta_2) \Rightarrow \theta_1 = \theta_2$. Pues bien, se puede demostrar que, bajo ciertas condiciones iniciales, cualquier transformación lineal del vector de estado no cambia las covarianzas ni la media de \mathbf{y}_t y, por tanto, no modifica la función de verosimilitud. Esto quiere decir que los parámetros del sistema no pueden ser identificados por la función de verosimilitud (Steyn, 1996:12). Otter y Mittnik (1989) derivan las condiciones bajo las cuales la introducción de ciertas restricciones conduce a la identificación (Steyn, 1996:12-14).

ANEXO II. Distribución del estado condicionada a las observaciones, $f(\mathbf{a}_t/y_t)$

Si

$$\begin{bmatrix} \mathbf{a}_t \\ \mathbf{y}_t \end{bmatrix} \sim N \left(\begin{bmatrix} S' \mathbf{a}_0 \\ \mathbf{Z}_t S' \mathbf{a}_0 \end{bmatrix}, \begin{bmatrix} \mathbf{P}_t & \mathbf{P}_t \mathbf{Z}_t' \\ \mathbf{Z}_t \mathbf{P}_t & \mathbf{Z}_t \mathbf{P}_t \mathbf{Z}_t' + \mathbf{H} \end{bmatrix} \right),$$

entonces $f(\mathbf{a}_t/y_t)$ sigue una distribución normal m -variante con vector de medias dado por

$$E[\mathbf{a}_t/y_t] = E[\mathbf{a}_t] + Cov[\mathbf{a}_t, \mathbf{y}_t](Var[\mathbf{y}_t])^{-1}(\mathbf{y}_t - E[\mathbf{y}_t]),$$

es decir,

$$E[\mathbf{a}_t/y_t] = S' \mathbf{a}_0 + \mathbf{P}_t \mathbf{Z}_t' (\mathbf{Z}_t \mathbf{P}_t \mathbf{Z}_t' + \mathbf{H})^{-1} (\mathbf{y}_t - \mathbf{Z}_t S' \mathbf{a}_0).$$

Para obtener la matriz de varianzas-covarianzas, conviene tener en cuenta que si se define la matriz

$$\mathbf{M}_t = \begin{bmatrix} \mathbf{P}_t & \mathbf{P}_t \mathbf{Z}_t' \\ \mathbf{Z}_t \mathbf{P}_t & \mathbf{Z}_t \mathbf{P}_t \mathbf{Z}_t' + \mathbf{H} \end{bmatrix}$$

y

$$\mathbf{W}_t = \mathbf{M}_t^{-1} = \begin{bmatrix} \mathbf{W}_{1,t} & \mathbf{W}_{3,t} \\ \mathbf{W}_{3,t}' & \mathbf{W}_{2,t} \end{bmatrix},$$

entonces⁵⁸ $Var[\mathbf{a}_t/y_t] = \mathbf{W}_{1,t}^{-1}$, donde⁵⁹

$$\begin{aligned} \mathbf{W}_{1,t} &= Var[\mathbf{a}_t]^{-1} \\ &+ Var[\mathbf{a}_t]^{-1} Cov[\mathbf{a}_t, \mathbf{y}_t] (Var[\mathbf{y}_t] - Cov[\mathbf{y}_t, \mathbf{a}_t] Var[\mathbf{a}_t]^{-1} Cov[\mathbf{a}_t, \mathbf{y}_t])^{-1} Cov[\mathbf{y}_t, \mathbf{a}_t] (Var[\mathbf{a}_t])^{-1} \\ &= [\mathbf{P}_t^{-1} + \mathbf{P}_t^{-1} \mathbf{P}_t \mathbf{Z}_t' (\mathbf{Z}_t \mathbf{P}_t \mathbf{Z}_t' + \mathbf{H} - \mathbf{Z}_t \mathbf{P}_t \mathbf{P}_t^{-1} \mathbf{P}_t \mathbf{Z}_t')^{-1} \mathbf{Z}_t \mathbf{P}_t \mathbf{P}_t^{-1}] \\ &= [\mathbf{P}_t^{-1} + \mathbf{Z}_t' \mathbf{H}^{-1} \mathbf{Z}_t] \end{aligned}$$

Por tanto, $f(\mathbf{a}_t/y_t)$ sigue una distribución normal m -variante con media

$$E[\mathbf{a}_t/y_t] = S' \mathbf{a}_0 + \mathbf{P}_t \mathbf{Z}_t' (\mathbf{Z}_t \mathbf{P}_t \mathbf{Z}_t' + \mathbf{H})^{-1} (\mathbf{y}_t - \mathbf{Z}_t S' \mathbf{a}_0)$$

y matriz de covarianzas

⁵⁸ Véase Zoroa y Zoroa (1991:277-279).

⁵⁹ Véase Johnston (1987:160).

$$\text{Var}[\mathbf{a}_t/y_t] = [\mathbf{P}_t^{-1} + \mathbf{Z}'_t \mathbf{H}^{-1} \mathbf{Z}_t]^{-1}.$$

Si \mathbf{a}_t es no estacionario y \mathbf{Z}_t es de rango pleno por columnas, se puede demostrar que $f(\mathbf{a}_t/y_t)$ sigue una distribución normal m -variante con media

$$E[\mathbf{a}_t/y_t] = (\mathbf{Z}'_t \mathbf{Z}_t)^{-1} \mathbf{Z}'_t \mathbf{y}_t$$

y matriz de covarianzas

$$\text{Var}[\mathbf{a}_t/y_t] = (\mathbf{Z}'_t \mathbf{Z}_t)^{-1} \mathbf{Z}'_t \mathbf{H} \mathbf{Z}_t (\mathbf{Z}'_t \mathbf{Z}_t)^{-1}.$$

Nótese que

$$\mathbf{Z}_t \mathbf{a}_t = \mathbf{y}_t - \mathbf{e}_t$$

Por tanto,

$$\mathbf{Z}'_t \mathbf{Z}_t \mathbf{a}_t = \mathbf{Z}'_t \mathbf{y}_t - \mathbf{Z}'_t \mathbf{e}_t$$

$$(\mathbf{Z}'_t \mathbf{Z}_t)^{-1} \mathbf{Z}'_t \mathbf{Z}_t \mathbf{a}_t = (\mathbf{Z}'_t \mathbf{Z}_t)^{-1} \mathbf{Z}'_t \mathbf{y}_t - (\mathbf{Z}'_t \mathbf{Z}_t)^{-1} \mathbf{Z}'_t \mathbf{e}_t$$

$$\mathbf{a}_t = (\mathbf{Z}'_t \mathbf{Z}_t)^{-1} \mathbf{Z}'_t \mathbf{y}_t - (\mathbf{Z}'_t \mathbf{Z}_t)^{-1} \mathbf{Z}'_t \mathbf{e}_t$$

de modo que:

$$E[\mathbf{a}_t/y_t] = (\mathbf{Z}'_t \mathbf{Z}_t)^{-1} \mathbf{Z}'_t \mathbf{y}_t$$

y

$$\begin{aligned} \text{Var}[\mathbf{a}_t/y_t] &= E\left[\left((\mathbf{Z}'_t \mathbf{Z}_t)^{-1} \mathbf{Z}'_t \mathbf{e}_t \right) \left(\mathbf{e}'_t \mathbf{Z}_t (\mathbf{Z}'_t \mathbf{Z}_t)^{-1} \right) \right] \\ &= (\mathbf{Z}'_t \mathbf{Z}_t)^{-1} \mathbf{Z}'_t E[\mathbf{e}_t \mathbf{e}'_t] \mathbf{Z}_t (\mathbf{Z}'_t \mathbf{Z}_t)^{-1} \\ &= (\mathbf{Z}'_t \mathbf{Z}_t)^{-1} \mathbf{Z}'_t \mathbf{H} \mathbf{Z}_t (\mathbf{Z}'_t \mathbf{Z}_t)^{-1} \end{aligned}$$

Referencias bibliográficas

Anderson, B.D.O. y J.B. Moore (1979) *Optimal Filtering*, Prentice-Hall.

Ansley, C.F. y R. Kohn (1982) A geometrical derivation of the fixed interval smoothing algorithm, *Biometrika* 69:486-7.

- Ansley, C.F. y R. Kohn (1985) Estimation, filtering and smoothing in state space models with incompletely specified initial conditions, *Annals of Statistics* 13: 1286-1316.
- De Jong, P. (1991) The diffuse Kalman filter, *Annals of Statistics* 19:1073-1083.
- Durbin, J. y S.J. Koopman (2001) *Time Series Analysis by State Space Models*, Oxford University Press.
- Fernández, F.J. y A.C. Harvey (1990) Seemingly unrelated time series equations and a test for homogeneity, *Journal of Business and Economic Statistics* 8:71-82.
- Hamilton, D. (1994) *Time Series Analysis*, Princeton University Press.
- Harvey, A.C. (1989) *Forecasting, Structural Time Series and the Kalman Filter*, Cambridge University Press.
- Harvey, A.C. y S.J. Koopman (1992) Diagnostic checking of unobserved components time series models, *Journal of Business and Economic Statistics* 10:377-389.
- Harvey y Shephard (1993) Structural time series models, en G.S. Maddala, C.R. Rao y H.D. Vinod (eds.) *Handbook of Statistics*, vol. 11, Elsevier Science Publishers.
- Jazwinski, A.H. (1970) *Stochastic Processes and Filtering Theory*, Academic Press.
- Johnston, J. (1987) *Métodos de Econometría*, Vicens Universidad.
- Jones, R.H. (1993) *Longitudinal Data with Serial Correlation: A state-space approach*, Chapman & Hall.
- Kalman, R.E. (1960) A new approach to linear filtering and prediction problems, *Transactions ASME, Series D, Journal of Basic Engineering* 82:35-45.

- Kalman, R.E. y R.S. Bucy (1961) New results in linear filtering and prediction theory, *Transactions ASME, Series D, Journal of Basic Engineering* 83:95-108.
- Kohn, R. y C.F. Ansley (1983) Fixed interval estimation in state space models when some of the data are missing or aggregated, *Biometrika* 70:683-8.
- Koopman, S.J. (1993) Disturbance smoother for state space models, *Biometrika* 80: 117-126.
- Koopman, S.J. (1996) Kalman filtering and smoothing. Mimeo.
- Koopman, S.J. (1997) Exact initial Kalman filtering and smoothing for non-stationary time series models, *Journal of the American Statistical Association* 92:1630-8.
- Koopman, S.J., A.C. Harvey, J.A. Doornik y N. Shephard (2000) *Stamp 6.0: Structural Time Series Analyser, Modeller and Predictor*, Timberlake Consultants.
- Maybeck, P. (1982) *Stochastic Models, Estimation and Control*, Academic Press.
- Meinhold, R. y N. Singpurwalla (1989) Robustification of Kalman filter models, *Journal of the American Statistical Association* 84:479-486.
- Otter, P. y S. Mittnik (1989) *Parameter identifiability in Kalman filtering models*, Technical report, University of Groningen.
- Rosenberg, B. (1973) Random coefficient models: the analysis of a cross section of time series by stochastically convergent parameter regression, *Annals of Economic and Social Measurement* 2:399-428.
- Steyn, I.J. (1996) *State Space Models in Econometrics: a field guide*, Vrije Universiteit.

Welch, G. y G. Bishop (2001) «An introduction to the Kalman Filter»,
http://www.cs.unc.edu/~welch/kalman/kalman_filter/kalman.html.

Zoroa, P. y N. Zoroa (1991) *Introducción a la probabilidad y la medida*, PPU.