1 Índice FEDEA

1.1 El Índice (Teoría)

1. Supongamos que queremos estimar el estado de la economía española (el Índice (x_t)). Sin embargo x_t resulta inobservable.

2. Pero disponemos de algunos datos observables que creemos que se ven influidos por el estado de la economía española (x_t) y que a su vez influyen en el estado de la economía española como pueden ser el PIB (y_t^1), el sentimiento de la economía (y_t^2), las matriculaciones de turismos (y_t^3), el consumo de energía eléctrica (y_t^4), el número de afiliados a la seguridad social (y_t^5), las ventas minoristas (y_t^6) o el índice de producción industrial (y_t^7).

3. En concreto, supongamos que el estado de la economía española (el Índice x_t) está relacionado con valores pasados del mismo índice de la siguiente forma:

$$x_t = \rho x_{t-1} + (1 - \rho^2)^{0.5} e_t \tag{1.1}$$

   donde $e_t \sim N(0, 1)$.

4. Mientras que los datos observables (y_t^i) se ven influidos por el estado de la economía (x_t) tal como por su propio pasado (y_{t-j}^i) del siguiente modo:

$$y_t^i = c_i + \beta^i x_t + \gamma^i y_{t-D_i}^i + u_t^i \tag{1.2}$$

5. Hay dos dimensiones importantes en los cuales los datos observables se distinguen:

   (a) Variables de flujo vs. variables de stock

   (b) Frecuencia con que se observan los datos (diario, semanal, mensual,..)

PIB (y_t^1)	flujo	trimestral
Sentimiento económico (y_t^2)	flujo	mensual
Matriculaciones turismos (y_t^3)	stock	mensual
Consumo energético (y_t^4)	flujo	mensual
Seguridad social (y_t^5)	stock	mensual
Ventas minoristas (y_t^6)	flujo	mensual
Producción industrial (y_t^7)	flujo	mensual

6. El sub-índice D_i en ecuaciòn 1.2 hace hincapié a las distintas frecuencias, donde $D_i = 1$ para datos mensuales y $D_i = 3$ para datos trimestrales.

7. El hecho de que nuestros datos se observan con distintas frecuencias implica que en algunos instantes, por ejemplo en el mes de marzo, tengamos nuevos datos para las matriculaciones de turismos pero no para el PIB, ya que el último solo se observa cada tres meses. Dicho de otra forma, para algunos datos vamos a tener variables ausentes en algunos momentos. Por eso, vamos que definir una nueva variable

$$\tilde{y}_t^i = \begin{cases} \sum_{j=0}^{Di-1} y_{t-j}^i & \text{para febrero, mayo, agosto, noviembre} \\ NA & \text{en cualquier otro mes} \end{cases} \tag{1.3}$$

donde $\sum_{j=0}^{D_i-1} y_{t-j}^i$ con $D_i = 3$ (en nuestro caso es la suma del producto interior bruto mensual). Usando ecuación 1.2 la nueva variable $\tilde{y}_t^i$ se puede escribir como

$$\tilde{y^i}_t = \sum_{j=0}^{D_i-1} c_i + \beta^i \sum_{j=0}^{D_i-1} x_{t-j} + \gamma^{i1} \sum_{j=0}^{D_i-1} y_{t-D_i-j}^i .... + \gamma^{iN} \sum_{j=0}^{D_i-1} y_{t-nD_i-j}^i + u_t^{i*} \tag{1.4}$$

donde $\gamma^{i1} \sum_{j=0}^{D_i-1} y_{t-D_i-j}^i = \tilde{y^i}_{t-1}$ es la nueva variable en $t-1$ y u_t^{i*} es la suma de los errores u_t^i de la ecuación 1.2.

8. En caso de que se trata de una variable de stock a la que se observa con una frecuencia más baja, la nueva variable $\tilde{y^i}_t$ se define trivialmente como

$$\tilde{y}_t^i = \begin{cases} y_{t-j}^i & \text{para febrero, mayo, agosto, noviembre} \\ NA & \text{en cualquier otro mes} \end{cases} \tag{1.5}$$

9. Para poder estimar el Índice (x_t^i) usando los datos que consideramos oportunos para predecir el comportamiento de la economía española necesitamos estimar los parámetros desconocidos de las ecuaciones 1.1 y 1.2 - dicho de otra forma necesitamos estimar el impacto que cambios en los datos tendrán sobre el comportamiento de la economía española (sobre el Índice) y vice versa.

10. Para facilitar la estimación conviene representar las ecuaciones 1.1 y 1.2 - en "state-space form".

$$Y_t = Z_t X_t + \epsilon_t \tag{1.6}$$
$$X_{t+1} = T X_t + R\eta_t$$

donde $\epsilon_t \sim (0, H_t)$, $\eta_t \sim (0, Q)$ y $t = 1, ...T$ con $Y_t = \{\tilde{y}_t^1, \tilde{y}_t^2, ...\} \equiv \{\text{PIB, Sentimiento, etc.}\}$ Índice $\equiv X_t = X_t = \{x_{t-2}, x_{t-1}, x_t\}$ y $T \equiv$ última observación

11. Hay que tomar en cuenta que

(a) Dado que el PIB solo se observa cada tres meses, usamos la variable $\tilde{y}_t^i$. Lo que implica que en muchas estimaciones mensuales aparecerán valores NA en el vector Y_t

(b) Dado que el PIB es una variable de flujo, $\tilde{y}_t^i$ viene definida en 1.3. Lo que implica que el vector X_t tiene que incluir valores pasados de x_t. En el caso $D_i = 3$, eso implica que la dimensión del vector X_t será igual a 3 (si incluimos también datos diarios $D_i = 90$, y la dimensión del vector aumentaría muchísimo más).

(c) Para nuestro caso $Z_t = Z$, y $H_t = H$. En el caso de que usemos datos diarios hay que tomar e cuenta que cada mes cambia el número de días por mes lo que requiere un ajuste de estas matrices.

12. Para estimar los parámetros del sistema 1.6 recurrimos al método del filtro Kalmann. Para ello definimos

    (a) $YY_t = Y_1, Y_2, ..., Y_t$

    (b) $X_{t|t} = E(X_t|YY_t)$: esperanza condicional del Índice dado todos los valores de los datos (estimador a posteriori)

    (c) $Err1 = X_t - X_{t|t}$

    (d) $P(t|t) = var(X_t|YY_t) = E(Err1Err1')$: varianza condicional del Índice dado valores de los datos hasta $t-1$ (estimador a priori)

    (e) $\hat{X}t = E(X_t|YY_{t-1})$:

    (f) $Err2 = X_t - \hat{X}_t$

    (g) $P_t = var(X_t|YY_{t-1}) = E(Err2Err2')$

13. El método Kalman tiene como objetivo encontrar una ecuación que calcule un estimador a posteriori ($X_{t|t}$) como combinación linear del estimador a priori ($\hat{X}_t$) y la diferencia ponderada (con el peso de K) entre los datos observado (Y_t) y la predicción ($Z_t X_{t|t}$), tal que

$$X_{t|t} = \hat{X}_t + K(Y_t - Z_t X_{t|t}) \tag{1.7}$$

    donde K se elija tal que la varianza a posteriori (d)) se minimice. (se obtiene el resultado sustituyendo ecuación 1.7 en $Err1$ (c)) y posteriormente en (d)) para luego minimizar la expresión obtenida con respecto a K - quiere decir sacar las esperanza, y derivar la "trace" del resultado con respecto a K, igualarlo a cero y hallar K). Hay varias formas de escribir el resultado óptimo. Una de las más populares es la siguiente:

$$K = P_t Z_t'(Z_t P_t Z_t' + H_t)^{-1} \tag{1.8}$$

14. Cabe destacar que cuanto más pequeño la varianza del error (ϵ_t), $H_t \to 0$ $K \to Z_t^- 1$, i.e. el residual gana importancia, se confía más en el dato mismo Y_t que en la estimación. Mientras que cuanto más pequeño el estimador a posteriori de la varianza de X_t, $P_t \to 0$, $K \to 0$, i.e. el residual pierde importancia, se confía más en la estimación ($Z_t X_t$) que en el dato mismo.

15. Nuestras ecuaciones de Kalman para la predicción son:

$$X_{t+1} = TX(t|t) \tag{1.9}$$

$$P_{t+1} = TP_{t+1}T' + RQR'. \tag{1.10}$$

Las ecuaciones para actualizar el filtro de Kalman son:

$$X(t|t) = \hat{X}_t + K(Y_t - Z_t X_{t|t}) \tag{1.11}$$

$$P(t|t) = P_t - KZ_t P_t'. \tag{1.12}$$

16. ¿Y cómo estimamos K, T, Z, H, Q?

17. Usando un "Gaussian pseudo log likelihood function"

   - Recuerda una regresión: $Y_t = \beta X_t + \epsilon_t$
   - Bajo el supuesto de que únicamente las observaciones de Y_t son relevantes para hacer interferencias sobre los parámetros desconocidos se puede usar el criterio de maximum liklihood
   - El método de maximum liklihood elige los parámetros tal que se maximice la probabilidad de generar la muestra observada.
   - Se supone que el vector de las variables aleatorias ϵ_t o Y_t están distribuidas de forma normal (mutivariate) y que $E(\epsilon_t \epsilon_t') = \sigma_t^2 I_t$, que los errores no están correlacionados.
   - Bajo estos supuestos se puede maximizar la siguiente función para obtener estimadores consistentes de β y σ_t
   - $$l(\beta, \sigma_t^2|y) = (2\pi\sigma_t^2)^(- T/2)exp(-(y - X\beta)'(y - X\beta)/2\sigma_t^2)$$

18. Para nuestro sistema 1.6 se puede hacer algo similar.

19. Dado una muestra de vectores (7x1) $Y_1, Y_2, ..Y_T$ distribuidas de forma normal se puede escribir el logaritmo de la función del log likilihood de la siguiente manera:

20.

$$lnl = -7T/2ln2\pi -1/2ln|\sigma(Y_t)| - 1/2(Y_t - \mu)' \sigma(Y)(-1)(Y_t - \mu)$$

21. donde μ es la esperanza de Y_t y σ (Y) la matriz de co-varianzas de Y_t. Se supone que tanto μ como σ (Y) son desconocidos y hay que estimarlas.

22. Dado nuestros supuestos recuerda que la inversa de la matriz Hessiana es igual a la matriz de las co-varianzas.

1 Índice Fedea

State space representation:

$$
\begin{aligned}
X_t &= A_t X_{t-1} + Be_t \qquad (1)\\
Y_t &= CX_t + \Gamma w_t + u_t^* \qquad (2)
\end{aligned}
$$

So we have 22 parameters that we need to estimate.

- Rho: ρ.
- Betas: β_1,β_2, β_3, β_4,β_5, β_6 and β_7.
- Gammas: γ_1,γ_2, γ_3, γ_4,γ_5, γ_6 and γ_7.
- Sigmas: σ_1,σ_2, σ_3, σ_4,σ_5, σ_6 and σ_7.

Where do they come from? From the transition equation (1), we have that (when considering only quarterly and monthly data):

$$
A_t = \begin{pmatrix} \rho & 0 \\ \rho & \xi_t^Q \end{pmatrix} \qquad (3)
$$

Where, ξ_t^Q takes the value 0 for the first day of the new quarter (when we know the GDP) and 1 otherwise. From the measurement equation (2), we know that

$$
C = \begin{pmatrix} 0 & \beta_1 \\ \beta_2 & 0 \\ \beta_3 & 0 \\ \beta_4 & 0 \\ \beta_5 & 0 \\ \beta_6 & 0 \\ \beta_7 & 0 \end{pmatrix} \qquad (4)
$$

Th subindexes follow this order:

1. GDP.
2. ESI (Economic Sentiment Indicator)
3. Matriculaciones de turismos.

4. Consumo de Energía Eléctrica.
5. Afiliados a la Seguridad Social.
6. Ventas Minoristas.
7. Índice de Producción Industrial.

From (2), we also know that:

$$\Gamma = \begin{pmatrix} 0 & ... & 0 & \gamma_1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & ... & 0 & 0 & \gamma_2 & 0 & 0 & 0 & 0 & 0 \\ 0 & ... & 0 & 0 & 0 & \gamma_3 & 0 & 0 & 0 & 0 \\ 0 & ... & 0 & 0 & 0 & 0 & \gamma_4 & 0 & 0 & 0 \\ 0 & ... & 0 & 0 & 0 & 0 & 0 & \gamma_5 & 0 & 0 \\ 0 & ... & 0 & 0 & 0 & 0 & 0 & 0 & \gamma_6 & 0 \\ 0 & ... & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \gamma_7 \end{pmatrix} \tag{5}$$

Finally, from (2), we have that the variance-covariance matrix of errors in measurement equation is:

$$\Sigma = \begin{pmatrix} \sigma_1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma_2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_4 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_5 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma_6 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sigma_7 \end{pmatrix} \tag{6}$$

How do we estimate these parameters?

1. We set initial parameter values (see below).
2. When starting the M-H, we present a proposal for the parameters: the current (initial) parameters plus an innovation (drawn from a normal distribution $N(0, 0.3) * D$ (we reduce this innovation for the parameter ρ, multiplying it by 0.01.
3. Before evaluating the likelihood at the proposed parameters, we set a prior: $betapdf(\rho^*, 18, 2) * Prob(\gamma_1^* > 0) * Prob(\gamma_1^* < 1) * ... * Prob(\sigma_1^* > 0) * Prob(\sigma_1^* < 1) * ... * Prob(\sigma_7^* > 0) * Prob(\sigma_7^* < 1)$ (stars * indicate that is the proposed paramenter, i.e. after adding the innovation).

4. Was that prior greater than 0? if yes, then evaluate the likelihood and continues with M-H.

θ	Initial parameter values	Posterior mean	s.d.	median
ρ	0.9600	0.9659	0.0001	0.9659
β_1	0.0051	0.0046	0.0003	0.0046
β_2	0.0009	-0.0033	0.0033	-0.0034
β_3	0.0734	0.0519	0.0172	0.0511
β_4	0.0232	0.0230	0.0033	0.0228
β_5	0.0042	0.0055	0.0007	0.0055
β_6	0.0257	0.0300	0.0044	0.0299
β_7	0.0546	0.0433	0.0081	0.0428
γ_1	0.7468	0.7198	0.0432	0.7196
γ_2	0.9046	0.9731	0.0228	0.9788
γ_3	0.3752	0.6349	0.0711	0.6353
γ_4	0.0680	0.2769	0.0747	0.2772
γ_5	0.8508	0.8590	0.0229	0.8598
γ_6	0.0507	0.2566	0.0774	0.2559
γ_7	0.0779	0.4082	0.0879	0.4124
σ_1	0.00001	0.0000	0.0000	0.0000
σ_2	0.0004	0.0009	0.0001	0.0009
σ_3	0.0053	0.0110	0.0012	0.0109
σ_4	0.0001	0.0003	0.0000	0.0003
σ_5	0.000004	0.0000	0.0000	0.0000
σ_6	0.0002	0.0005	0.0001	0.0005
σ_7	0.0001	0.0004	0.0001	0.0004

Table 1: Initial parameter values and Posterior statistics for the parameters