El filtro de Kalman y su aplicación como método de imputación

Miranda Belmomte Hairo Ulises

Presentación: Estadística Inferencial

7 de Diciembre del 2018

Estructura

Introducción

- El filtro de Kalman introducido por Rudolf Kalman en 1960 en :"New Approach to Linear Filtering and Prediction Problems".

Introducción

- El filtro estima estados no observables en base a series de mediciones observables con ruido de forma recursiva
- Funciona como filtro al tener componente *ruidoso* o algún otro inconveniente en los datos.
- Inclusive con eso, produce estimaciones óptimas.

Introducción

- Christensen, R.(1991), algoritmo que predice el vector de estados en base a los datos. Proceso recursivo de predicción y corrección de β_t basado en las variaciones de β_{t-1}.

Mejor predictor lineal

- Christesen (1991), Se utiliza para predicciones en series de tiempo, procesos estocásticos, modelos mixtos, análisis de datos espaciales, el filtro de Kalman, entre otros.
- El mejor predictor , esperanza condicional $E(y|x)$, necesario el conocimiento de la distribución conjunta de las variables aleatorias
- El predictor lineal $\hat{E}(y|x)$, solo requiere existencia media y la varianza.
- Se le conoce somo esperanza lineal.

¿Qué es?

- No es la esperanza condicional, ni tampoco es su estimación.
- Función que minimiza el error cuadratico medio a la hora de ser el predictor lineal de y.
- Distribución de las variables es normal multivariada, mejor predictor, $E(y|x)$, se le considera como el mejor predictor lineal,$\hat{E}(y|x)$, esperanza lineal

Definición

El mejor predictor lineal es aquella función $f(x)$ que minimice el error cuadrático medio.

$$E[(y - f(x))^{'}(y - f(x))]$$

A su vez, se le conoce somo esperanza lineal.

$$\hat{E}(y|x) \equiv \mu_y + \beta^{'}(x - \mu_x)$$

Donde β es la solución de $V_{xx}\beta = V_{xy}$.

Definición

- Se trabajo bajo del supuesto que β es conocida.
- En los casos en el cual no conocemos β, la teoría del mejor predictor lineal desarrolla el mejor predictor lineal insesgado.
- Toma importancia en modelos con efectos aleatorios (i.e., β se asume no conocido y aleatorio).

Notación

$$V_{xy} = Cov(x, y) \quad V_{yy} = Cov(y)$$
$$V_{xx} = Cov(x) \qquad Ey = \mu_y \qquad Ex = \mu_x$$

La esperanza lineal es una función de x.

$$E(y) = E[\hat{E}(y|x)]$$

Error de predicción de la matriz de varianza-covarianza

$$Cov(y, x) = V_{yy} - V_{yx} V_{xx}^{-1} V_{xy}$$

Propiedades

Proposición 3.1.1. La esperanza lineal es un operador lineal. Sea A una matriz mxn y a un vector $mx1$:

$$\hat{E}(Ay + a|x) = A\hat{E}(y|x) + a$$

Proposición 3.1.2. Sea y una variable aleatoria que se pretende predecir; entonces, si se predice con un conjunto de variables aleatorias el cual se encuentra contenida, la predicción seré ella misma. Sea y un vector:

$$\hat{E}(y_i|y) = y_i$$

Propiedades

Proposición 3.1.3. Transformaciones lineales nosingular del predictor, no cambiaran la esperanza lineal. Sea A una matriz mxn y a un vector $mx1$.

$$\hat{E}(y|Ax + a) = \hat{E}(y|x)$$

Proposición 3.1.4. Las variables predictoras no correlacionados con la predecida, dan como resultado su valor medio.

$$Cov(y, x) = 0; \quad \hat{E}(y|x) = \mu_y$$

Propiedades Adicionales

Dos vectores como predictores:

$$y\prime = \begin{bmatrix} y_1' & y_2' \end{bmatrix}$$

$$Cov(y) = Cov \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix}$$

$$Cov(y_i, x) = V_{ix}$$

$$E(y_i) = \mu_i$$

Proposición 3.1.5. Si $Cov(y|x) = 0$, entonces:

$$\hat{E}(y_2|y_1, x) = \hat{E}(y_2|x) + \hat{E}(y_2|y_1) - \mu_2$$

Propiedades Adicionales

Proposición 3.1.6. Sea $e(y_1|x) = y_1 - \hat{E}(y_1|x)$ el error de predicción de y_1 dado x.

$$Cov(e(y_1|x), x) = 0$$

Proposición 3.1.7.

$$\hat{E}(y_2|y_1, x) = \hat{E}(y_2|x) + Cov(y_2, e(y_1|x))[Cov(e(y_1|x))]^{-1}e(y_1|x)$$

Propiedades Adicionales

Proposición 3.1.8.

$$Cov(y_2 - \hat{E}(y_2|y_1, x)) = Cov(y_2 - \hat{E}(y_2|x)) - Cov(y_2, e(y_1|x)) \\ [Cov(e(y_1|x))]^{-1} Cov(e(y_1|x), y_2)$$

Proposición 3.1.9.

$$Cov(y, y - \hat{E}(y|x)) = Cov(y - \hat{E}(y|x))$$

Modelos Estado-Espacio

- Es una generalización para modelos multivariados en serie de tiempos.
- Pretenden medir variable no observables mediante aquellas que son medibles

Ecuación de medición

$$Y_t = X_t \beta_t + e_t \tag{1}$$

Donde:

Y_t	es un vector $(kx1)$ de series observables o endogenas
β_t	es el vector de estados $(nx1)$
X_t	Es una matriz (kxn) de parámetros de la ecuación (1)
e_t	Es un vectorde los errores de la ecuación

Ecuación de estado

- Captura estructura de dependencia en el tiempo.
- Genera dinámica en la ecuación de medición y el sistema.
- Es modelada con autoregresivo de orden uno (i.e. con un rezago)

$$\beta_t = \Phi \beta_{t-1} + \epsilon_{t-1} \tag{2}$$

Donde:

β_t	vector $(px1)$ de los estados
Φ	Matriz (pxp) de parámetros de la ecuación de estado
ϵ_t	Es un vector $(px1)$ de los errores de la ecuación

Supuestos

- Las matrices de parámetros se asumen conocida en el tiempo t.
- Pueden cambiar con el tiempo y por ende deben ser estimados.
- Los errores de las ecuaciones son ruido blanco, i.e., no se encuentra correlacionados.

$$E(e_t) = 0$$
$$E(\epsilon_t) = 0$$
$$Cov(e_i, e_j) = 0; i \neq j$$
$$Cov(\epsilon_i, \epsilon_j) = 0; i \neq j$$
$$Cov(e_i, \epsilon_i) = 0; \forall i, j$$
$$Cov(e_i) = V_t$$
$$Cov(\epsilon_i) = \sum_t$$

Supuestos Valores iniciales

$$E(\beta_0) = \bar{\beta}_0$$
$$Cov(\beta_0) = P_0$$
$$Cov(\beta_0, e_i) = 0$$
$$Cov(\beta_0, \epsilon_i) = 0$$

Propiedades

Hamlton, J. (1994)

- Conveniente para modelar sumas de procesos estocasticos
- Cualquier $AR(p)$ puede ser exprezado como un $AR(1)$, con el fin de obtener un mejor resumen de la dinámica del sistema.

Aoki, M.(1987)

- Modelos son estables. Sistema dinámico asintótico y el efecto de valores iniciales tiende a desapareces en el tiempo.
- Representación en estado espacio genera modelos parsimoniosos. Dimensiones se reducen por parte del vector de estados.

Ejemplo 1

Considere que se desea estudiar el comportamiento de la tasa de interés real ex ante (tasa de interés i_t menos la inflación esperada π_t^e). Sin embargo, esta variable es no observable debido a que no se tiene datos confiables de la inflación esperada. Entonces, se se asume que la tasa de interés real puede ser modelada como un proceso $AR(1)$.
De esta manera el vector de estado se encuentra representado por la siguiente ecuación:

$$\beta_t = i_t - \pi_t^e - \mu \tag{3}$$

$$\beta_{t+1} = \Phi \beta_t + \epsilon_{t+1}$$

con μ como el promedio ex ante de la tasa de interés real.

Ejemplo 1

En cambio, la tasa de interés real ex post, puede ser observada y tener datos al respecto (tasa de interés nominal i_t menos la inflación actual π_t).

$$i_t - \pi_t = (i_t - \pi_t^e) + (\pi_t^e - \pi_t) = \mu + \beta_t + e_t \qquad (4)$$

Siendo así, la ecuación de medición se encuentra dada por (3) , y la ecuación de estado (lo no observable) por (4). En el cual se pretende pronosticar la tasa de interés ex ante con la ex post.

Ejemplo 2

Considere un proceso $MA(1)$

$$y_t = \mu + \epsilon_t + \theta\epsilon_{t-1}$$

En forma de estado-espacio se tiene:
Vector de estados:

$$\begin{bmatrix} \epsilon_{t+1} \\ \epsilon_t \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \epsilon_t \\ \epsilon_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{t+1} \\ 0 \end{bmatrix}$$

La ecuación de medición:

$$y_t = \mu + \begin{bmatrix} 1 & \theta \end{bmatrix} \begin{bmatrix} \epsilon_t \\ \epsilon_{t-1} \end{bmatrix}$$

Filtro de Kalman

- Algoritmo que predice el vector de estados en base a los datos. Proceso recursivo de predicción de β_t basado en las variaciones de β_{t-1}.

Lütkepohl, H. (2005)

- Relevancia del filtro de Kalman como una herramienta en el análisis de los modelos estado espacio.
- El filtro de Kalman es de gran importancia a la hora de estimar los parámetros.

Supuestos

- Matrices de coeficientes conocidas
- Bajo el supuesto de normalidad, el proceso de estimación de los estados por medio del filtro de Kalman se puede ver como la esperanza condicional.
- Se desarrolla el filtro bajo el contexto de la esperanza lineal.

Supuestos Adicionales

$$Cov(\beta_t, e_t) = 0$$
$$Cov(\beta_{t-1}, \epsilon_t) = 0$$
$$Cov(\epsilon_{t-1}, x) = 0$$
$$Cov(\epsilon_t, \beta_t - \hat{E}(\beta_{t-1}|x)) = 0$$

Desarrollo

Desea predecir $\hat{E}(\beta_t|Y_t,x)$, y de acuerdo a la definición de esperanza lineal:

$$\hat{E}(\beta_t|Y_t,x) = \hat{E}(\beta_t|x) + Cov(\beta, Y_t - \hat{E}(Y_t|x)) \tag{5}$$

$$[Cov(Y_t - \hat{E}(Y_t|x))]^{-1}[Y_t - \hat{E}(Y_t|x)]$$

donde:

$$\hat{E}(\beta_t|x) = \Phi\hat{E}(\beta_t|x)) \tag{6}$$

$$\hat{E}(\epsilon_t|x) = 0$$

$$Y_t - \hat{E}(Yt|x) = Y_t - X_t\Phi\hat{E}(\beta_t|x) \tag{7}$$

Desarrollo

Utilizando la ecuación de estado y la ecuación de medición.

$$Cov(Y_t - \hat{E}(Y_t|x)) = X_t[\Phi P_t \Phi^{'} + \Sigma_t]X_t^{'} + V_t \tag{8}$$

$$P_{t-1} = Cov[\beta_t - \hat{E}(\beta_t|x))] \tag{9}$$

$$Cov[\beta_t, Y_t - \hat{E}(Y_t|x))] = (\Phi P_{t-1}\Phi^{'})X_t^{'} \tag{10}$$

Desarrollo

La forma estándar del filtro de Kalman sustituir en (5) las expresiones (6), (7), (8) y (10):

$$\hat{E}(\beta_t|Y_t,x) = \Phi\hat{E}(\beta_{t-1}|x) + R_t X_t^{'}[X_t R_t X_t^{'}]^{-1} \tag{11}$$

$$[Y_t - X_t\Phi\hat{E}(\beta_{t-1}|x)]$$

donde:

$$R_t = \Phi P_{t-1}\Phi^{'} + \Sigma_t \tag{12}$$

Algoritmo

Comenzar recursión.
Dado valores iniciales pronostico del vector de estados.

$$\hat{E}(\beta_0) = E(\beta_0) = \bar{\beta}_0$$

con el error cuadrático medio asociado.

$$Cov(\beta_0 - \hat{E}(\beta_0)) = Cov(\beta_0) = P_0$$

Algoritmo

Paso 1. Predicción $(1 < t < T)$
Generar pronóstico del estado hacia adelante

$$\hat{E}(\beta_t|x) = \Phi\hat{E}(\beta_{t-1}|x)$$

matriz del error cuadrático medio que asocia a estimación.

$$R_t = \Phi P_{t-1}\Phi^{'} + \Sigma_t$$

pronosticar valores de Y_t

$$\hat{E}(Y_t|x) = X_t\Phi\hat{E}(\beta_{t-1}|x)$$

covarianza del error de predicción (ECM)

$$Cov[Y_t - \hat{E}(Y_t|x)] = X_t[\Phi P_{t-1}\Phi^{'} + \Sigma_t]X_t^{'} + V_t$$

Algoritmo

Paso 2. Corrección o Actualización Generar un actualización y mejora en la estimación del estado. Calcular de la ganancia de Kalman. Factor de ponderación que se selecciona con el fin de minimizar la covarianza del error de la nueva estimación del estado.

Ganancia de Kalman

$$K_t = R_t X_t^{'}[X_t R_t X_t^{'} + V_t]^{-1}$$

Actualización o Predicción de β_t

$$\hat{E}(\beta_t|Y_t, x) = \hat{E}(\beta_{t-1}|x) + K_t[Y_t - X_t \Phi \hat{E}(\beta_t - 1|x)]$$

Calcular la covarianza del error para la actualización de la estimación

$$P_t = R_t - K_t X_t R_t$$

Algoritmo

Paso 3. Pronostico $(t > T)$ Una vez pronosticado para $t = 1$, entonces se puede realizar para $t > T$, los cuales se realizan de la siguiente manera:

$$\hat{E}(\beta_{t+1}|Y_t, x) = \Phi\hat{E}(\beta_t|Y_t, x)$$

Con su respectiva covarianza del error de predicción (MSE):

$$Cov[\beta_{t+1} - \hat{E}(\beta_{t+1}|Y_t, x)] = R_{t+1}$$

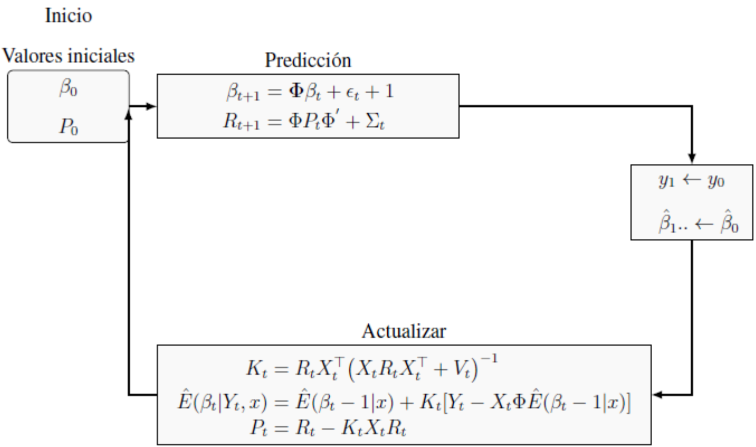
De forma similar, para la ecuación de medición:

$$\hat{E}(Y_{t+1}|Y_t, x) = X_{t+1}\Phi\hat{E}(\beta_t|Y_t, x)$$

Con la covarianza del errror correspondiente (MSE)

$$Cov[Y_{t+1} - \hat{E}(T_{t+1}|Y_t, x)] = X_{t+1} + R_{t+1} + X'_{t+1} + V_t$$

Algoritmo


Inicio
Valores iniciales
β_0
P_0
Predicción
$\beta_{t+1} = \Phi\beta_t + \epsilon_t + 1$
$R_{t+1} = \Phi P_t \Phi' + \Sigma_t$
$y_1 \leftarrow y_0$
$\hat{\beta}_1 .. \leftarrow \hat{\beta}_0$
Actualizar
$K_t = R_t X_t^\top (X_t R_t X_t^\top + V_t)^{-1}$
$\hat{E}(\beta_t | Y_t, x) = \hat{E}(\beta_t - 1|x) + K_t[Y_t - X_t \Phi \hat{E}(\beta_t - 1|x)]$
$P_t = R_t - K_t X_t R_t$

Suavizador de Kalman

Estimar un valor particular en un tiempo $t < T$. El filtro de Kalman puede realizar recursiones conocidos como suavizamiento para valores de β_t con $t < T$. La esperanzas lineales se obtienen recursivamente, iniciando al final de la muestra:

Paso 1.

$$\hat{E}(\beta_t|Y_T) = \hat{E}(\beta_t|Y_t) + S_t(\beta_{t+1} - \hat{E}(\beta_{t+1}|Y_t)$$

con su respectiva covarianza en el error:

$$P_{t|T} = P_{t|T} - S_t[P_{t+1|t} - P_{t+1|T}]S_t^{'}$$

El suavizador es:

$$S_t = P_{t|t}\Phi^{'}P_{t+1|t}^{-1}$$

Ejemplo

Estimar una constante aleatoria escalar, como una "lectura de voltaje"de una fuente. Supongamos que tiene un valor constante de aV (voltios), pero, por supuesto, tenemos algunas lecturas ruidosas por encima y por debajo de un voltio. Y suponemos que la desviación estándar del ruido de medición es de 0,1 V.

$$x_k = x_{t-1} + w_t$$

$$z_k = x_t + v_t$$

Tiempo	1	2	3	4	5
Valores	0,39	0,50	0,48	0,29	0,25

Tiempo	6	7	8	9	10
Valores	32	0,34	0,48	0,41	0,45

Ejemplo

Predicción

$$x_t = x_{t-1}$$

$$P_t = P_{t-1}$$

Corrección:

$$K_t = \frac{P_t}{P_t + R}$$

$$\hat{x}_t = \hat{x}_t + K_t(z_t - \hat{x}_t)$$

$$P_t = (1 - K_t)P_t$$

k	z_k	$\hat{x}_{k-1}$	P_k^-	Time Update	Measurement Update	$\hat{x}_k$	P_k
1	0.390	0	1	$\hat{x}_k^- = \hat{x}_{k-1} = 0$ $P_k^- = P_{k-1} = 1$	$K_k = 1 / (1\ 0.1)$ $= 0.909$ $\hat{x}_k = 0.909 . (0.390 - 0)$ $= 0.35$ $P_k = (1 - 0.909) . 1$ $= 0.091$	0.355	0.091
2	0.500	0.355	0.091	$\hat{x}_k^- = 0.355$ $P_k^- = 0.091$	$K_k = 0.091 / (0.091\ 0.1)$ $= 0.476$ $\hat{x}_k = 0.355 . 0.476 . (0.500 - 0.355)$ $= 0.424$ $P_k = (1 - 0.476) . 0.091$ $= 0.048$	0.424	0.048
3	0.480	0.424	0.048			0.442	0.032
4	0.290	0.442	0.032			0.405	0.024
5	0.250	0.405	0.024			0.375	0.020
6	0.320	0.375	0.020			0.365	0.016
7	0.340	0.365	0.016			0.362	0.014
8	0.480	0.362	0.014			0.377	0.012

Máxima verosimilitud: Supuestos

Normalidad en:
Valores iniciales (β_0), los errores de la ecuación de estados (ϵ_i) , ecuación de medición (e_i) y la distribución de los datos Y_n.

$$f(Y_1, ..., Y_t) = f(Y_1)f(Y_2|Y_1), ..., f(Y_t|Y_1, Y_2, .., Y_{t-1})$$

$$Y_t|Y_1, Y_2, .., Y_{t-1} \sim N(\hat{E}(Y_t|x), Cov[Y_t - \hat{E}(Y_t|x))$$

con

$$\hat{e}_t = Y_t - \hat{E}(Y_t|Y_1, .., Y_{t-1})$$

$$Q_t = Cov[Y_t - \hat{E}(Y_t|Y_1, .., Y_{t-1}))$$

Máxima verosimilitud: Condiciones

Hamilton, J.(1994)

- Problema de identificación. Problemática de encontrar los mejores estimadores.
- Vector de estados siga un proceso estacionario.
- Variable β_t, debe tener un comportamiento asintótico.

Máxima verosimilitud: Estimación

Función de verosimilitud:

$$L(\Phi, V_n, \Sigma_n = (\Phi_{t=1}^{n}(2\phi)^{-q_t/2})$$

$$(\Phi_{t=1}^{n}(2\phi)^{-q_t/2})(\Phi_{t=1}^{n}(|Q_t|)^{-1/2})exp[(\frac{-1}{2})(\Sigma_{t=1}^{n}\hat{e}_t Q_t^{-1}\hat{e}_t]$$

Aplicando logaritmos a la función de verosimilitud:

$$logL(\Phi, V_n, \Sigma_n = -log(2\phi)\Sigma_{t=1}^{n}\frac{q_t}{2}$$

$$-\frac{1}{2}\Sigma_{t=1}^{n}log(|Q_t|) - \frac{1}{2}\Sigma_{t=1}^{n}\hat{e}_t Q_t^{-1}\hat{e}_t$$

Máxima verosimilitud: Casos especiales

White (1982)

- Cuasi-verosimilitud, encontrando que los estimadores siguen siendo consistentes y asintóticamente normales, pero con varianza distinta

Máxima verosimilitud: Mazimixar

- Métodos númericos para realizar la maximización.
- El algoritmo EM.
- Método Newton-Raphson.
- El algoritmo del gradiante
- El algoritmo score, entre otros.

Valores perdidos

El manejo de valores perdidos en modelos estado-espacio es el siguiente:
Sea $W_t = Z_t\beta_t + \epsilon_t$ con W_t el vector $(qx1)$ observable en el tiempo t y $Cov(\epsilon_t) = V$. Entonces, si se tiene valores perdidos podemos partir los componentes de la serie de la siguiente manera.

$$W_t = \begin{bmatrix} Y_t \\ Y_{tM} \end{bmatrix}; Z_t = \begin{bmatrix} X_t \\ X_{tM} \end{bmatrix}$$

$$\epsilon_t = \begin{bmatrix} e_t \\ e_{tM} \end{bmatrix}; Cov(\epsilon_t) = \begin{bmatrix} V_t & V_{tM} \\ V_{Mt} & V_{MM} \end{bmatrix}$$

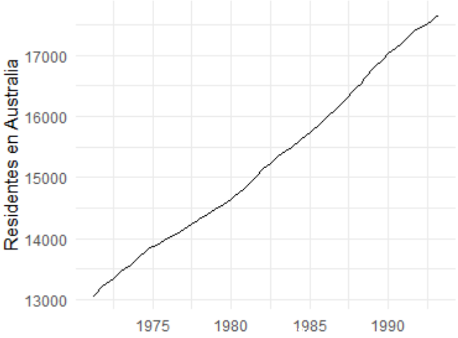
La ecuación de medición sería la siguiente:

$$Y_t = X_t\beta_t + e_t$$

Datos

Datos trimestrales del número de residentes en Australia de Marzo de 1971 a Marzo de 1994.

P. J. Brockwell and R. A. Davis. Introduction to Time Series and Forecasting. Springer-Verlag, 1996. ISBN.



Valores perdidos

Paquete *imputeTestbench* por Marcus W. Beck (2017).
Funciones:

- ïmpute error”: simula valores perdidos
- MCAR (missing completely at random) :observaciones misma probabilidad de ser seleccionadas, series no autocorrelacionadas.
- MAR (missing at random): por bloques, probabilidad de ser seleccionado dependa de si la observación se encuentra cercana a una ya evaluada.

Imputación

- Última observación realizada (na.locf).
- Valor medio (na.mean).
- Interpolación (na.approx).

Paquete ImputeTS creado por Steffen Moritz(2018).

- "na.kalman". imputación mediante el suavizador de Kalman.
- Modelo ARIMA, en estado-espacio, orden en base al críterio Akaike función .ªuto.arima".

Raíz del error cuadrático medio (RMSE)

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \hat{x}_i)^2}{n}}$$

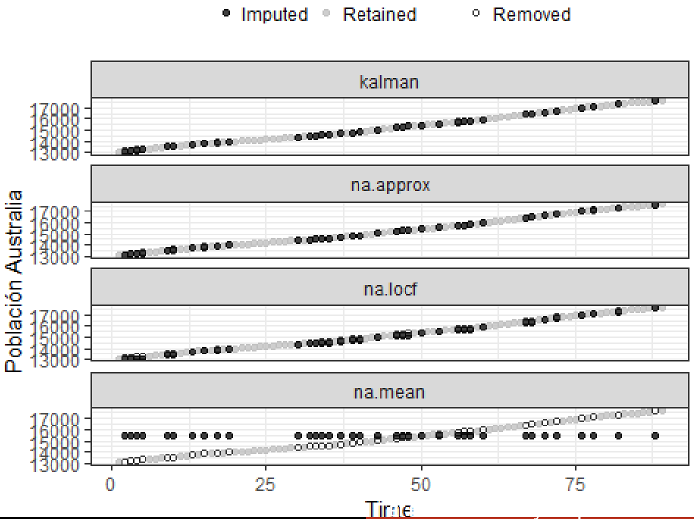
La media absoluta del porcentaje de error (MAPE)

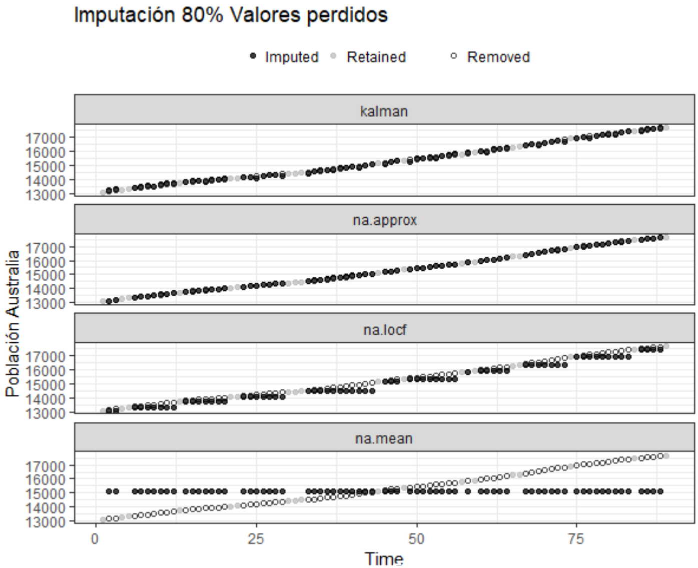
$$MAPE = 100 * \frac{\sum_{i=1}^{n} |(x_i - \hat{x}_i)/x_i|}{n}$$

Error absoluto medio (MAE)

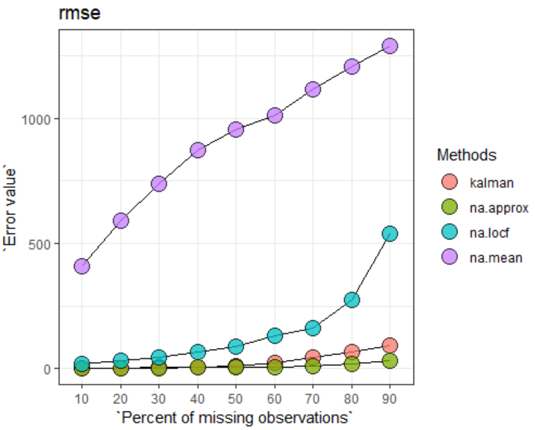
$$MAE = \frac{\sum_{i=1}^{n} |(x_i - \hat{x}_i)|}{n}$$

Imputación 40% Valores perdidos

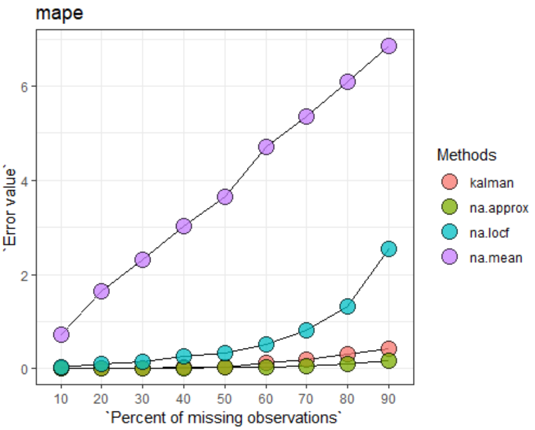



Imputación 80% Valores perdidos
Imputed
Retained
Removed
kalman
na.approx
na.locf
na.mean
Población Australia
17000
16000
15000
14000
13000
0
25
50
75
Time

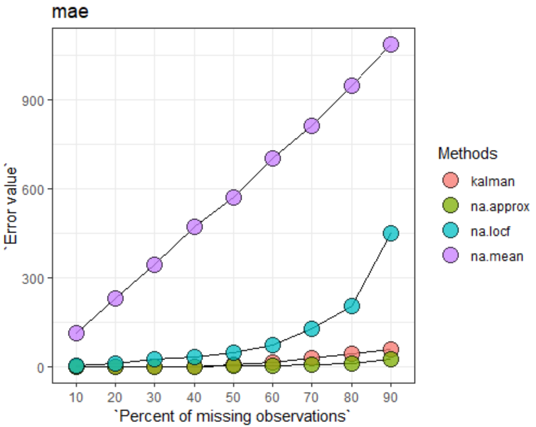
Resultados


rmse
`Error value`
1000
500
0
10
20
30
40
50
60
70
80
90
`Percent of missing observations`
Methods
kalman
na.approx
na.locf
na.mean

Resultados


mape
`Error value`
`Percent of missing observations`
0
2
4
6
10
20
30
40
50
60
70
80
90
Methods
kalman
na.approx
na.locf
na.mean

Resultados


mae
`Error value`
900
600
300
0
10
20
30
40
50
60
70
80
90
`Percent of missing observations`
Methods
kalman
na.approx
na.locf
na.mean

Conclusiones

Para la mayoría de los casos, el suavizador de Kalman es un buen método de imputación. Sin embargo, cuando se tiene un gran número de observaciones perdidas, el error de medición es más elevado que el que se calcula en el método de interpolación lineal, debido que por su naturaleza recursiva, al realizar el suavizamiento de Kalman, y tener gran número de valores perdidos, prácticamente estaría extrapolando información, y como vimos en el caso de hacer pronostico, los resultados pueden ser un poco inapropiados.

Gracias por su atención