

El filtro de Kalman y su aplicación como método de imputación.

Miranda Belmonte Hairo Ulises,^{1*}
¹ Centro de Investigación en Matemáticas, A.C.
*E-mail: hairo.miranda@cimat.mx.

Resumen

Se desarrolla el filtro de Kalman y se implementa como método de imputación. Con el fin de entender el filtro, se presentan temas como "El mejor predictor lineal", "Modelos estado-espacio", "Estimación de los parámetros del modelo" y "Valores perdidos". Para la aplicación se utilizan métodos para generar valores perdidos de forma aleatoria, con el fin de hacer uso de diversas técnicas de imputación y contrastarlo respecto al suavizador de Kalman. El documento muestra que el filtro de Kalman es una buena opción como método de imputación en los casos de que el porcentaje de valores perdidos no sea lo suficientemente grande.

Introducción

Encontrar valores perdidos en los datos suele ser muy común en el análisis de series de tiempo. Kees, E. y Jan, P. (2011) (12), hablan de la problemática de como las instituciones encargadas de realizar estadísticas oficiales suelen realizar publicaciones con rezagos. Por otro lado, Demirhan, H. et. al. (2018) (9), recalca que en muchos casos al final de las series se encuentran valores perdidos, conociendo a este problema como "ragged edge". Asimismo, recalcan lo poco probable de obtener series completas - ya sea valores perdidos al inicio, en medio o al final de la serie- ,teniendo un profundo efecto en el análisis.

Existen diversos métodos de imputación para solventar esta problemática. Los cuales su exactitud se encuentra en función del porcentaje de valores perdidos que se presenten en la serie (Marcus, W. (2018) (14)). Se desea un buen método dado a que una de las principales razones para el uso de la técnicas en series de tiempo es el realizar pronósticos. De esta manera,

obtener imputaciones precisa tendrían como consecuencia una ligera mejora del pronóstico, ya que muchos de los métodos de estimación en los parámetros -como la función de verosimilitud-, son realizado en base a los datos.

El filtro de Kalman, al encontrarse expresado en modelos estado-espacio, logra lidiar con valores perdidos. A su vez, por medio del suavizador de Kalman, logra realizar un algoritmo recursivo filtrando la serie sujetas a variaciones aleatoria, valores perdidos e inclusive outliers, y realizar predicciones óptimas de variables no observables. Es de esa manera, que se puede pensar en el filtro de Kalman como un método que impute de datos.

El presente trabajo se estructura de la siguiente manera; en la sección 2 se habla sobre la imputación de datos. La sección 3, se abordan distintos temas para la comprensión del filtro de Kalman, como lo son; En la 3.1, se introducen algunas propiedades del mejor predictor lineal. En la 3.2, se presenta a manera general los modelos estado-espacio, los cuales contemplan dinámica en el sistema por medio de componentes no observables, y a su vez, se presentan algunos ejemplos. En la 3.3, se desarrolla el filtro de Kalma. En la 3.4, la estimación de los parámetros del modelo. En la 3.5, se habla un poco del filtro de Kalma y el trato con valores perdidos. En la sección 4, se presenta una aplicación del filtro de Kalman como método de imputación. Finalmente en la sección 5, se da una conclusión en términos de lo eficiente que es el filtro para realizar imputación en los datos.

Sección 2. Imputación de datos

En la mayoría de los casos es bastante usual obtener un conjunto de observaciones que incluyan valores perdidos. Esta ausencia de datos puede implicar no sólo pérdida de información, sino en algunas ocasiones imposibilita el uso métodos estadísticos.

En muchos casos se puede decir que la falta de valores perdidos puede ser por distintas razones. Little y Rubin (1987), mencionan tres de la razones principales. La primera, es conocida

como pérdida de valores completamente aleatoria (MCAR). En la cual, la ausencia de los datos es independiente de cualquier contexto, haciendo que los valores perdidos realmente sea impredecibles. La segunda, pérdida aleatoria (MAR). Esto sucede cuando la ausencia de un dato es independiente del valor de la propia variable en la que falta, pero puede depender de alguna otra variable, haciendo referencia a una estructura de correlación entre los datos. La última, es conocida como pérdida aleatoria (MNAR), donde los datos, no faltan al azar, ya que la ausencia del dato sólo depende de ella misma.

Entre las soluciones al problema se presentan los métodos para la imputación en los datos. A manera general, la imputación de datos es la sustitución de valores no informados por algún otro valor. Existen una variedad muy amplia de métodos que realizan aproximaciones de valores perdidos por métodos de imputación. Entre los más conocidos se encuentran aquellos que remplazan los valores faltantes con la media de la serie. Otros que toman valores previos al valor perdido. Aquellos que utilizan métodos de interpolación lineal. El cual es muy parecido a los que realizan regresión del valor perdido respecto a los valores observado y por medio de predicción realizan la imputación. Regresiones con componentes estocásticos- similar al ya mencionado pero agregando un termino de error aleatorio-, entre otros.

En el caso de las series de tiempo, el obtener una predicciones correcta a la hora de imputar ayuda a mejorar ligeramente el pronóstico de la serie a interés. Es por eso, que en los siguientes apartados se propone el uso del filtro de Kalman como método de imputación, el cual se encuentra bajo la metodología de modelos de series de tiempo, i.e., sistemas dinámicos que se son capaces de realizar pronóstico de componentes no observables mediante el uso de observaciones que son medibles.

Sección 3. El Filtro de Kalman¹

En esta sección se presenta, a manera general, la definición del mejor predictor lineal, el cual será necesario para el entendimiento del filtro de Kalman. Asimismo, se menciona lo que son los modelos de estado-espacio, el filtro de Kalman y su estimación, y por último, un poco de la aplicación a valores perdidos.

Sección 3.1. Mejor predictor lineal

Christesen (1991) (8), recalca la importancia sobre el mejor predictor lineal (Best linear predictor (BLP)) y el mejor predictor lineal no sesgados (Best linear unbiased predictor (BLUP's)) para la teoría del mejor predictor lineal. Se ha utilizado para realizar predicciones en metodologías como series de tiempo, procesos estocásticos, modelos mixtos, análisis de datos espaciales, el filtro de Kalman, entre otros. En general, se ha pensado para el uso del análisis multivariado, y de acuerdo con Hartigan (1996) (11), es considerado como la base de los métodos lineales Bayesianos.

A comparación con el mejor predictor (Best predictor (BP)), definido como la esperanza condicional $E(y|x)$, en el cual es necesario el conocimiento de la distribución conjunta de las variables aleatorias, el predictor lineal $\hat{E}(y|x)$, solo requiere de la existencia de la media y la varianza.

El mejor predictor lineal, no es la esperanza condicional, ni tampoco es su estimación, pero lo que sí es, es la función que minimiza el error cuadrático medio a la hora de ser el predictor lineal de y . En el caso en el que el proceso o la distribución de las variables aleatorias provengan de una distribución normal multivariada, el mejor predictor, esperanza condicional, se le considera como el mejor predictor lineal.

El criterio para la selección del mejor predictor lineal se encuentra dado por aquella función

¹Sección y ecuaciones basada en: Christensen, R (1991) (8) y 1987 (7))

que minimice el error cuadrático medio (MSE).² Una definición más formal se encuentra dada a continuación:

Definición 3.1.1

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 como esperanza lineal.

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

Donde β es la solución de $V_{xx}\beta = V_{xy}$, haciendo que $\mu_y + \beta'(x - \mu_x)$, sea el mejor predictor lineal. No obstante, se trabajo bajo del supuesto que β es conocida. Sin embargo, en los casos en el cual no conocemos β , la teoría del mejor predictor lineal desarrolla el mejor predictor lineal insesgado, el cual toma importancia en modelos con efectos aleatorios (i.e., β se asume no conocido y aleatorio), y debilitando el supuesto de conocer β , se obtienen predicciones débiles. (Christesen, 1987) (7).

Una vez sabiendo que $\mu_y + \beta'(x - \mu_x)$ es el mejor predictor lineal de y , se presentan algunas propiedades³. Asimismo, en el siguiente cuadro se presenta algo de notación.

Notación

$$\begin{array}{ccc} V_{xy} = Cov(x, y) & V_{yy} = Cov(y) & \\ V_{xx} = Cov(x) & Ey = \mu_y & Ex = \mu_x \end{array}$$

La esperanza lineal es una función de x.

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

² El error cuadrático medio, es una función de una muestra, por lo tanto un estimador, el cual mide el promedio de los errores al cuadrado, i.e., la diferencia entre el estimador y lo que estima

³Para ver sus demostraciones vea: Christensen (1991) (8), Cap. 6

Error de predicción de la matriz de varianza-covarianza

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

Proposición 3.1.1. La esperanza lineal es un operador lineal. Sea A una matriz $m \times n$ y a un vector $m \times 1$:

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

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 será ella misma. Sea y un vector:

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

Proposición 3.1.3. Transformaciones lineales nonsingular del predictor, no cambiarán la esperanza lineal. Sea A una matriz $m \times n$ y a un vector $m \times 1$.

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

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

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

Estas propiedades generalizan los resultados del mejor predictor lineal y el mejor predictor. Las cuales se utilizan para el desarrollo del filtro de Kalman. No obstante, se requieren propiedades adicionales con el uso de dos vectores como predictores (el cual es el caso del filtro de Kalman).

$$y' = \begin{bmatrix} y'_1 & y'_2 \end{bmatrix}$$

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

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

$$E(y_i) = \mu_i$$

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

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

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 .

$$\overline{Cov(e(y_1|x), x) = 0}$$

Proposición 3.1.7.

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

Proposición 3.1.8.

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

Proposición 3.1.9.

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

Sección 3.2. Modelo Estado Espacio⁴

Los modelos de estado espacio son generalizaciones para la modelación de series de tiempo multivariadas. En términos generales, se pretende medir variable no observables mediante aquellas que son medibles. Esto indica, que los modelos de series de tiempo multivariados,

⁴Sección basada en Lütkepohl (2005) (13)

sea y_1, \dots, y_T , dependen de un estado no observable, sea z_t , el cual sigue un comportamiento estocástico e introduce la dinámica al proceso.

La relación entre las series no observables y las que si lo son entran en interacción por medio de dos ecuaciones. La primera de ellas se define como la ecuación de medición.

$$Y_t = X_t \beta_t + e_t \quad (1)$$

Donde:

Y_t	es un vector ($k \times 1$) de series observables o endógenas
β_t	es el vector de estados ($n \times 1$)
X_t	Es una matriz ($k \times n$) de parámetros de la ecuación (1)
e_t	Es un vector de los errores de la ecuación

La segunda ecuación captura estructura de dependencia en el tiempo. A su vez, como se mencionó, el componente no observable crea una estructura dinámica en la ecuación de medición. Esto indica, que la ecuación de estado (o transición), es modelada con autoregresivo de orden uno (i.e. con un rezago).

Donde:

$$\beta_t = \Phi \beta_{t-1} + \epsilon_{t-1} \quad (2)$$

Donde:

β_t	vector ($p \times 1$) de los estados
Φ	Matriz ($p \times p$) de parámetros de la ecuación de estado
ϵ_t	Es un vector ($p \times 1$) de los errores de la ecuación

Finalmente se debe de tener en cuenta ciertos supuestos en el modelo. Las matrices que contienen a los parámetros se asumen como conocidos en el tiempo t . Sin embargo, puede cambiar con el tiempo y ser no conocidos, y por ende deben ser estimados. Por otro lado, se sume que los errores de las ecuaciones son ruido blanco, i.e., que no se encuentra correlacionados, tienen media cero y con una matriz de varianza-covarianza que puede variar con el tiempo.

Algunos supuestos:

$$\begin{array}{c}
 \hline
 E(e_t) = 0 \\
 E(\epsilon_t) = 0 \\
 Cov(e_i, e_j) = 0; i \neq j \\
 Cov(\epsilon_i, \epsilon_j) = 0; i \neq j \\
 Cov(e_i, \epsilon_i) = 0; \forall i, j \\
 Cov(e_i) = V_t \\
 Cov(\epsilon_i) = \sum_t \\
 \hline
 \end{array}$$

Asimismo, se tienen que asumir valores iniciales, ya que como veremos más adelante, el filtro de Kalman es un algoritmo recursivo desarrollado en un modelo de estado espacio.

$$\begin{array}{c}
 \hline
 E(\beta_0) = \beta_0 \\
 Cov(\beta_0) = P_0 \\
 Cov(\beta_0, e_i) = 0 \\
 Cov(\beta_0, \epsilon_i) = 0 \\
 \hline
 \end{array}$$

De acuerdo con Christensen (1991) (8), menciona que principalmente el modelo estado espacio suele implementarse para la predicción del vector no observable β_t . En el cual el filtro de Kalman toma un papel importante al ser un proceso recursivo de predicción y corrección en el pronóstico de β_t .

El estado-espacio puede ser conveniente para modelar sumas de procesos estocásticos o las consecuencias en la medición de errores (Hamilton, 1994 (10)). A su vez, Hamilton indica que cualquier $AR(p)$ puede ser expresado como un $AR(1)$, con el fin de obtener un mejor resumen de la dinámica del sistema.

En Aoki, M.(1987), se mencionan propiedades peculiares de los modelos estado-espacio. El primero es que estos modelos son estables. Esto es, que el sistema dinámico es asintótico y el efecto de valores iniciales tiende a desaparecer en el tiempo. Esto indica que los valores propios de la matriz rezagada de la ecuación de estado se encuentran dentro del círculo unitario (Φ) . Por otro lado, hace mención de que la representación en estado espacio genera modelos

parsimoniosos, i.e. que las dimensiones se reducen por parte del vector de estados⁵.

Ejemplo 3.2.1

Un ejemplo del modelo estado espacio es el siguiente. 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:

$$\beta_t = i_t - \pi_t^e - \mu \quad (3)$$

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

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

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).

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

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 3.2.2

En este ejemplo se muestra la representación en forma matricial. Considere un proceso $MA(1)$

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

En forma de estado-espacio se tiene:

Vector de estados:

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

⁵Para profundizar en el tema vea: Aoki M. (1987) Properties of State Space Models. In: State Space Modeling of Time Series. Universitext. Springer, Berlin, Heidelberg. ó Aoki M. (1990) Properties of State Space Models. In: State Space Modeling of Time Series. Universitext. Springer, Berlin, Heidelberg

La ecuación de medición:

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

Los modelos de estado espacio y filtro de Kalman fueron introducidos por Kalman (1960). Como ya se mencionó, el filtro de Kalman, se considera una herramienta útil en el uso de los modelos de estado-espacio, debido a que genera predicciones del vector de estado con las series que pueden ser observables. Estas predicciones se encuentran bajo la perspectiva del mejor predictor lineal, o lo que viene siendo, la esperanza lineal.

Sección 3.3. Filtro de Kalman⁶

El filtro de Kalman fue introducido por Rudolf Kalman en 1960 en su artículo "New Approach to Linear Filtering and Prediction Problems". Fue utilizado en el proyecto Apollo, el cual requería la estimación en la trayectoria de naves espaciales tripuladas con dirección a la luna y su regreso a la tierra.

El filtro es utilizado para estimar estados no observables en base a series de mediciones observables con ruido. De esta manera, funciona como filtro al tener este componente "ruidoso" -o algún otro inconveniente en los datos- en la serie observable, e inclusive con eso, producir estimaciones estadísticamente óptimas.

El filtro de Kalman, de acuerdo con Christensen, R.(1991) (8), se puede definir como un algoritmo que predice el vector de estados en base a los datos. Siendo un proceso recursivo de predicción de β_t basado en las variaciones de β_{t-1} . Islas, A. y Cortez A. (2009) (1) lo definen como un algoritmo matemático que opera mediante un mecanismo de predicción y corrección. Asimismo, mencionan que se utiliza para estimar variables de estado que no son observables en un sistema dinámico lineal a partir de variables observables y en base a un

⁶Sección basada en: Christensen, R (1991 (8))

modelo estado-espacio.

En el trabajo de Matthew, B. et.al. (2017) (15), menciona que el filtro de Kalman actúa como filtro, pero tienen en cuenta que es una operación muy difícil de entender. Por otro lado, mencionan que el filtro de Kalman toma información en el cual se reconoce que cuenta con algún error o ruido y trata de reducir la incertidumbre.

Lütkepohl, H. (2005) (13), en cambio, recalca la relevancia del filtro de Kalman como una herramienta extremadamente útil en el análisis de los modelos estado espacio. Ya que dado un vector de variables observables Y_t , puede realizar estimaciones del vector de estados β_t y medir la precisión de las estimaciones. Asimismo, en el caso de que la matriz de los coeficientes de la ecuación de medición X_t y la ecuación de estado Φ sean desconocidos, el filtro de Kalman es de gran importancia a la hora de estimar los parámetros.

Como se mencionó en la sección anterior, se busca realizar predicción del vector de estados por medio de la aproximación del mejor predictor lineal (ó esperanza lineal). De esta forma, se asume como conocido las matrices de coeficientes, y se define al vector x como el rezago de los valores observados, $(Y_{t-1}, Y_{t-2}, \dots, Y_1)'$. De la misma manera, se hará uso de las propiedades mencionadas en la sección 3.1., tanto las que solo hacen uso de un predictor como aquellas adicionales que toman en cuenta dos predictores.⁷

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. No obstante, se desarrolla el filtro bajo el contexto de la esperanza lineal. Entonces, deseamos predecir $\hat{E}(\beta_t|Y_t, x)$, y de acuerdo a la definición de esperanza lineal:

$$\hat{E}(\beta_t|Y_t, x) = \hat{E}(\beta_t|x) + Cov(\beta, Y_t - \hat{E}(Y_t|x))[Cov(Y_t - \hat{E}(Y_t|x))]^{-1}[Y_t - \hat{E}(Y_t|x)] \quad (5)$$

donde:

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

⁷Para un desarrollo más explícito vea Christensen, R. (1991) (8)

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

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

Utilizando la ecuación de estado y la ecuación de medición.

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

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

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

Donde se contempla los siguientes supuestos:

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

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

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

donde:

$$R_t = \Phi P_{t-1}\Phi' + \Sigma_t \quad (12)$$

El cual es la forma de predecir el vector del estado, dado las observaciones ,el rezago de el vector no observable y la covarianza del error de predicción.

Como se mencionó, el filtro de Kalman puede verse como un algoritmo de predicción y corrección. A continuación, se presenta el algoritmo:

Comenzar recursión.

Dado uns valores iniciales pronostico del vector de estados, dado observación de y ó x .

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

con el error cuadrático medio asociado.

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

Paso 1. Predicción ($1 < t < T$)

El primer paso consiste generar pronóstico del estado hacia adelante, tomando en cuenta la información que se tiene. Dado los valores iniciales, predecir para $t = 2, 3, \dots, T$.

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

con la matriz del error cuadrático medio que se le asocia a cada estimación.

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

Ahora, se pronostica los valores de Y_t

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

con la covarianza del error de predicción (ECM)

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

Paso 2. Corrección o Actualización El segundo paso consiste en generar una actualización y mejora en la estimación del estado. En este paso, es importante el cálculo de la ganancia de Kalman. La ganancia de Kalman se define como un factor de ponderación que se selecciona con el fin de minimizar la covarianza del error de la nueva estimación del estado. Siendo así, el siguiente paso es medir el proceso para obtener una predicción de Y_t , y con ella realizar la corrección (actualización) de la 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

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

Para el paso final, se calcula la covarianza del error para la actualización de la estimación

$$P_t = R_t - K_tX_tR_t$$

Paso 3. Pronóstico ($t > T$) 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):

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

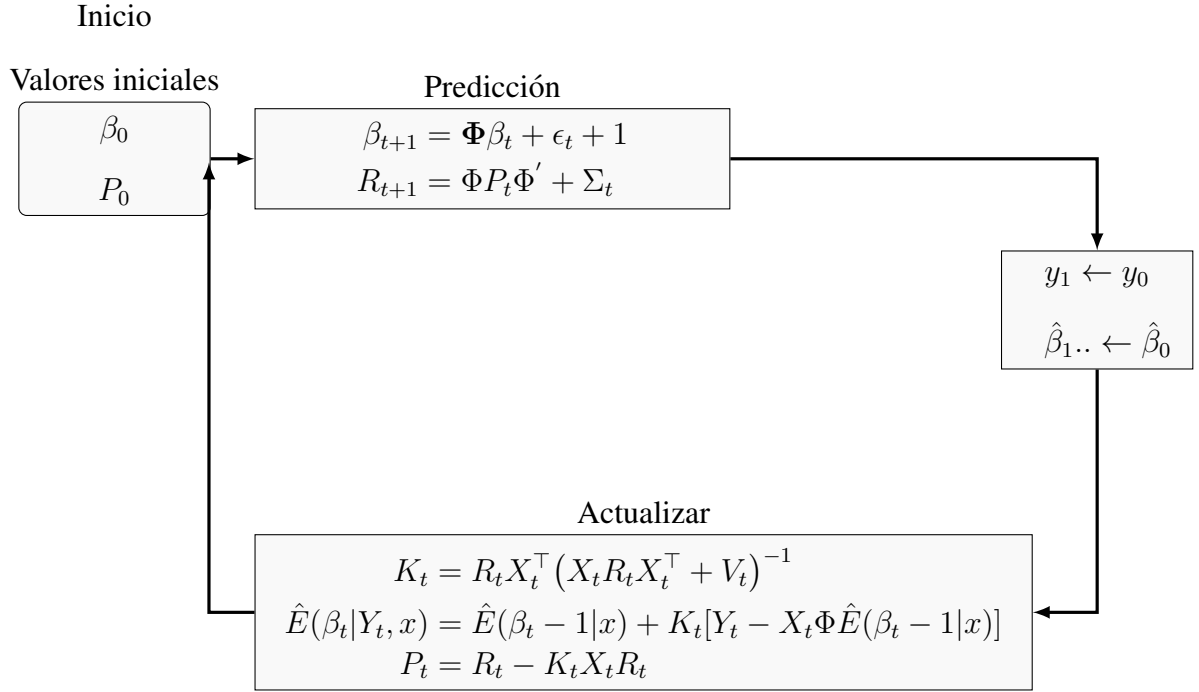
De forma similar, para la ecuación de medición:

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

Con la covarianza del error correspondiente (MSE)

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

A continuación se resume el algoritmo de Kalman:



El filtro de Kalman ha sido presentado como un algoritmo que realiza pronóstico del vector de estados como una función lineal dado las observaciones. Asimismo se puede realizar pronósticos de forma recursiva, pero el filtro, ha sido pensado como un proceso para eliminar el ruido, en lugar de pronosticar. Es por eso que los resultados de los pronósticos pueden ser un poco inapropiados.

En algunos casos, el vector de estado puede tener alguna interpretación, en estos casos el valor no observable es de interés. Por ejemplo, querer 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:

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

Sección 3.4. Estimación de parámetros

La estimación de máxima verosimilitud se basa en el supuesto de normalidad en; los valores iniciales (β_0) , los errores (innovaciones) de la ecuación de estados (ϵ_i) , ecuación de medición (e_i) y la distribución de los datos Y_n . Al tener en cuenta el supuesto de normalidad, la función de densidad conjunta se puede escribir como el producto de las densidades condicionales:

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

Al asumir el valor inicial cada densidad conjunta se distribuye normal con media y varianza dada por el filtro de Kalman.

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

con

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

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

No obstante, de acuerdo con Hamilton, J.(1994) (10), indica que previo a expresar la función de verosimilitud y estimar los estadísticos de máxima verosimilitud (MV), se requieren condiciones necesarias para que el filtro de Kalman evalúe la función. La primera hace relación al problema de identificación. Haciendo referencia a la problemática de encontrar los mejores estimadores, debido a que este problema se suele presentar en modelos de múltiple ecuaciones,

y al ser un modelo de estado-espacio, puede incurrir en esta problemática. La segunda es que el proceso del vector de estados siga un proceso estacionario. Esto es, que los valores propios de Φ se encuentren en el círculo unitario. Finalmente, indica que la variable β_t , debe tener un comportamiento asintótico.

Mikusheva, A. (2017) (16), deja en claro que no solo es una reexpresión de los resultados usuales del estadístico de máxima verosimilitud, ya que en este caso no fue requerido que las observaciones sean independientes e idénticamente distribuidas. Lo anterior implica que el filtro de Kalman converga cuando el tamaño de la muestra tienda a infinito, en el sentido de que la matriz Φ_t convega en probabilidad a Φ , es decir, que sea constante en el tiempo (estacionario). Este resultado es igual para la matriz de covarianza del error de pronóstico del estado P_t .

Finalmente, en caso de que el supuesto de normalidad no se cumpla, ya sea en los errores o en la distribución de la muestra, uno puede realizar la estimación por cuasi-verosimilitud. White (1982), realiza dicha estimación por el método cuasi-verosimilitud, encontrando que los estimadores siguen siendo consistentes y asintóticamente normales, pero con varianza distinta⁸. Una vez mencionado lo necesario para estimar los estadísticos de máxima verosimilitud con el filtro de Kalman, la función de verosimilitud se plantea de la siguiente manera:

$$L(\Phi, V_n, \Sigma_n) = (\Phi_{t=1}^n (2\phi)^{-q_t/2}) (\Phi_{t=1}^n (2\phi)^{-q_t/2}) (\Phi_{t=1}^n (|Q_t|)^{-1/2}) \exp\left[\left(\frac{-1}{2}\right) (\Sigma_{t=1}^n \hat{e}_t Q_t^{-1} \hat{e}_t)\right]$$

Aplicando logaritmos a la función de verosimilitud:

$$\log L(\Phi, V_n, \Sigma_n) = -\log(2\phi) \sum_{t=1}^n \frac{q_t}{2} - \frac{1}{2} \sum_{t=1}^n \log(|Q_t|) - \frac{1}{2} \sum_{t=1}^n \hat{e}_t Q_t^{-1} \hat{e}_t$$

No obstante, gran parte de la literatura argumentan sobre la dificultad de maximizar la función de verosimilitud. De esta manera, se utilizan métodos numéricos para realizar la maximización. Entre ellas se encuentran; El algoritmo EM⁹, El método Newton-Raphson, el algoritmo del

⁸Vea: White, H. (1982). Maximum Likelihood Estimation of Misspecified Models (17)

⁹Vea: Lütkepohl, (2005) (13)

gradiente, el algoritmo score¹⁰, entre otros.

Sección 3.5. Valores perdidos

Es bien sabido que valores perdidos son muy usuales en series de tiempo, y a la hora de realizar alguna representación de los datos lo usual es representar el modelo en forma estado-espacio¹¹. Cipra,T. y Romera,R. (1997) (6), hacen mención sobre lo útil que es el filtro de Kalman bajo ciertas modificaciones, permitiendo el tratamiento de los datos con valores perdidos o valores outliers.

Como se mencionó en la sección de estado-espacio, el vector Y_t es de dimensión fija ($qx1$), que a pesar de tener dimensión fija puede tomar diferentes valores en el tiempo. No obstante, el supuesto de que en cada tiempo se tiene las $qx1$ no siempre se cumple. En algunas ocasiones el vector de observaciones se compone de valores perdidos, los cuales son permitidos en los modelos estado-espacio.

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.

$$\begin{aligned} W_t &= \begin{bmatrix} Y_t \\ Y_{tM} \end{bmatrix} \\ Z_t &= \begin{bmatrix} X_t \\ X_{tM} \end{bmatrix} \\ \epsilon_t &= \begin{bmatrix} e_t \\ e_{tM} \end{bmatrix} \\ Cov(\epsilon_t) &= \begin{bmatrix} V_t & V_{tM} \\ V_{Mt} & V_{MM} \end{bmatrix} \end{aligned}$$

En este caso, uno puede solo observar solo algunos componentes en el tiempo, o un agregarlos de los datos observables (Cipra,T. y Romera,R. (1997) (6)). Asimismo. esto se expresa en el

¹⁰Vea:Lütkepohl, (2005) (13)

¹¹En el software estadístico R. A la hora de modelar algún ARIMA, lo representa en forma estado-espacio para evitar los valores perdidos.

último componente del vector (Y_{tM}, X_{tM}, e_{tM}) . De esta manera, en el filtro de Kalman, el tener observaciones perdidas no afectan a la recursion, ya que ajustando la ecuación de medición simplemente ignora los valores y continua el algoritmo. La ecuación de medición sería la siguiente:

$$Y_t = X_t\beta_t + e_t$$

La cual varía su dimensión.¹²

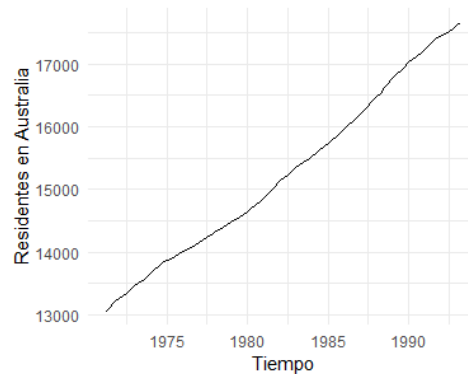
Sección 4. Aplicación del Filtro de Kalman: Imputación de datos

El filtro de Kalman puede utilizarse como método de imputación. Al implementar el suavizador de Kalman, lo que se está haciendo es interpolar el proceso. Como vimos en el apartado anterior, la recursión permite valores perdidos y con ello se puede pensar al suavizador de Kalman como un método de imputación de datos.

En esta sección se implementa el filtro de Kalman como método de imputación. Se utiliza el suavizador de Kalman y se contrasta contra otros métodos de imputación. Se toma en cuenta diferentes métricas de error en la predicción, con el fin de evaluar el desempeño del filtro de Kalman respecto a otros.

La serie de tiempo se toma del conjunto de datos de R, la cual es utilizada en Brockwell y Davis (1996) (5). La serie se compone de datos trimestrales del número de residentes en Australia de Marzo de 1971 a Marzo de 1994.

¹² ver un desarrollo más amplio de datos incompletos, vea: Cipra,T. y Romera,R. (1997) (6)



Se utiliza el paquete *imputeTestbench* realizado por Marcus W. Beck (2017). Entre sus funciones se encuentra "impute error", que simula valores perdidos en las serie por medio de MCAR (missing completely at random) y MAR (missing at random). El primero, asume que todas las observaciones tienen la misma probabilidad de ser seleccionadas y se sugiere su uso en series no autocorrelacionadas. El segundo realiza la selección por bloques tal que la probabilidad de ser seleccionado dependa de si la observación se encuentra cercana a una ya evaluada.

Una vez que la función genera valores perdidos de forma aleatoria, realiza la imputación por distintos métodos. Entre ellos se encuentran los siguientes; Última observación realizada (na.locf), remplazando el valor perdido con el valor reciente que no es valor perdido. Imputación por el valor medio (na.mean), remplaza los valores perdidos con el valor medio de la serie (se puede utilizar media o moda). Por último, interpolación (na.approx) genera valores para remplazar los perdidos por medio de valores interpolados.

Para utilizar el filtro de Kalman como método de imputación, se utiliza el paquete *ImputeTS*, creado por Steffen Moritz(2018). Se genera una función que reciba la serie de tiempo y por medio de "na.kalman", se realiza la imputación mediante el suavizador de Kalman. El modelo de predicción se realiza implementando un modelo ARIMA, expresado en estado-espacio, cuyo orden se selecciona es en base al criterio Akaike que se encuentra por default en la función "auto.arima". Una vez que ya se tiene el método de imputación de Kalman, se manda a llamar

con la función de "impute error" del paquete imputeTestbench.

Al tener disponibles estos métodos de imputación, "impute error" calcula los estadísticos que cuantifican la métrica del error (los cuales se desean valores bajos). Entre estos se encuentran los siguientes:

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)

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

A continuación, se presentan dos gráficos en los cuales se generan valores perdidos (40% y 80% de los datos) de forma aleatoria e imputando con los métodos previamente mencionados. Asimismo, se realiza un contraste de manera gráfica con los valor originales de la serie. Se puede observar, que con 40 por ciento de valores perdidos, la imputando con el suavizador de Kalman y una interpolación lineal (na.approx) realizan con precisión la imputación. No obstante, si uno pretende imputar tomando el valore previo (na.locf), se pueden observar ligeras desviaciones de la imputación con su valor original. En el caso de imputar por el método de la media, se observa que el meter el valor medio a la serie distorsiona su tendencia.

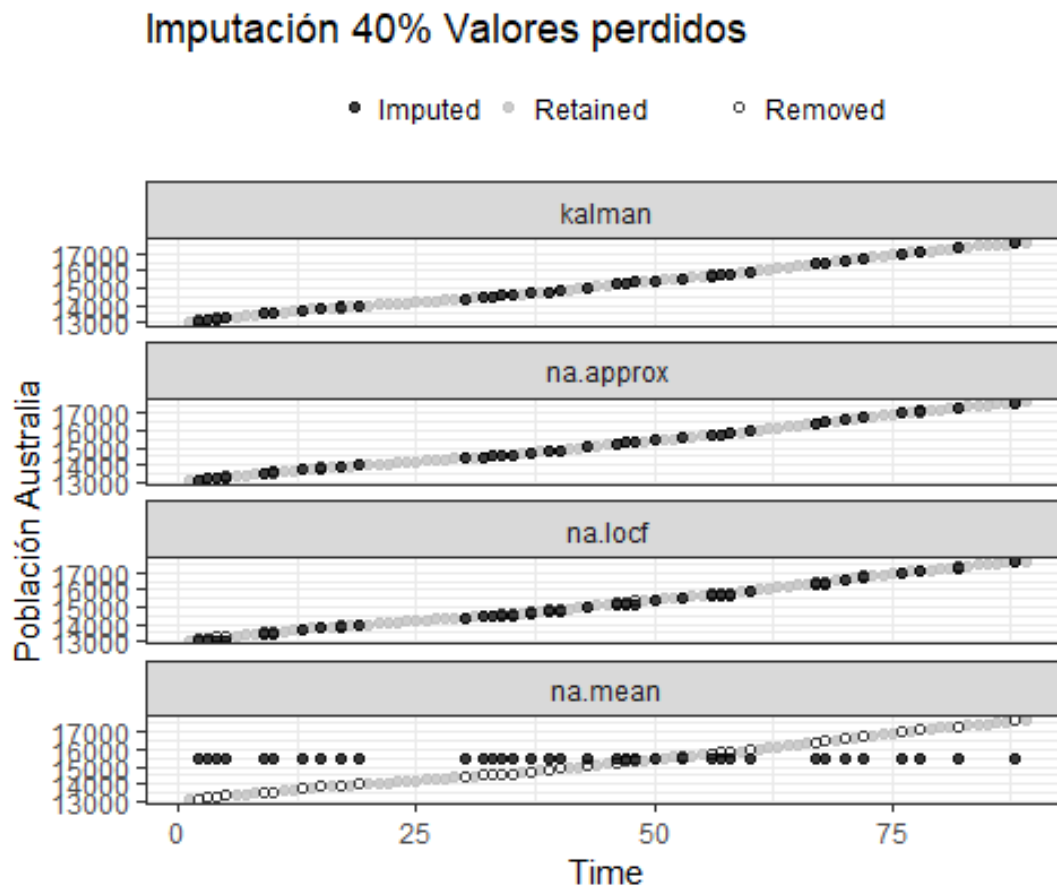


Figure 1: Círculos negros: valores imputados. Círculos grises: valores retenidos. Círculos blancos: valores originales

En el caso de generar valores perdidos en el 80 por ciento, se observa que sólo la aproximación lineal (na.approx) imputa de una forma muy precisa. En contraste al realizar la imputación con el suavizador de Kalman, que comienza a obtener diferencias en los valores imputados. Para los otros dos métodos, queda claro que comienzan a realizar imputaciones inadecuadas.

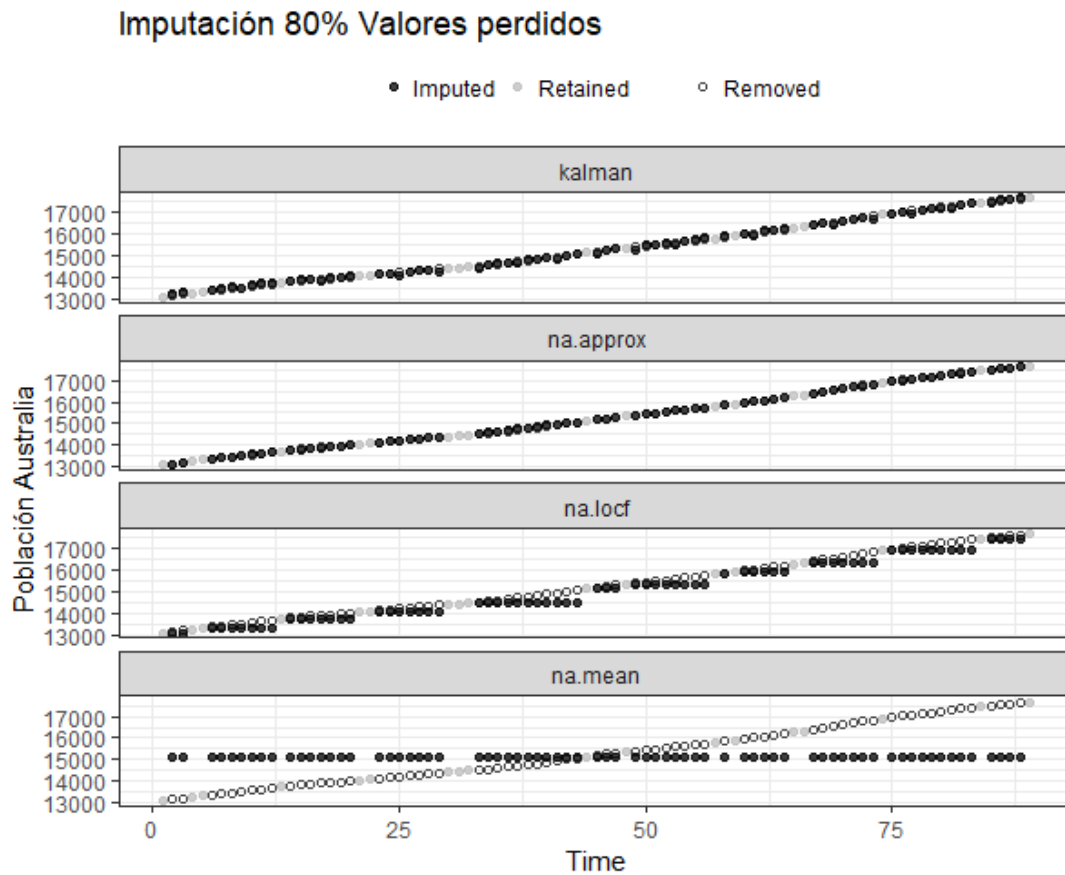
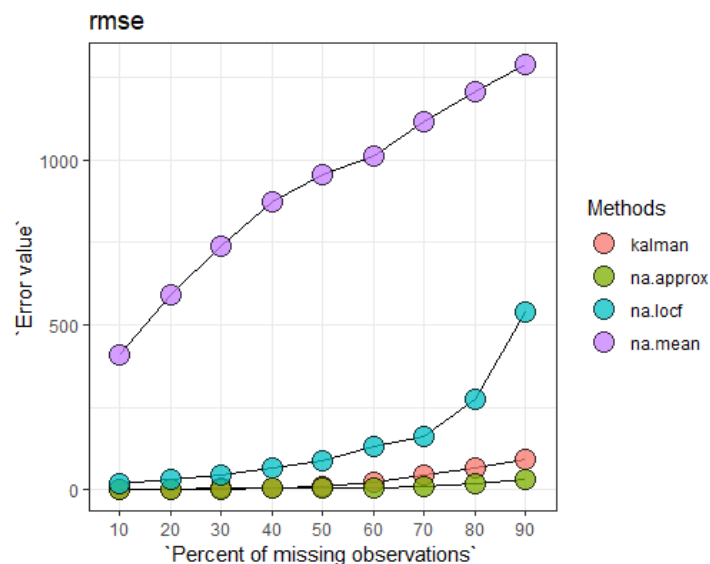


Figure 2: Círculos negros: valores imputados. Círculos grises: valores retenidos. Círculos blancos: valores originales

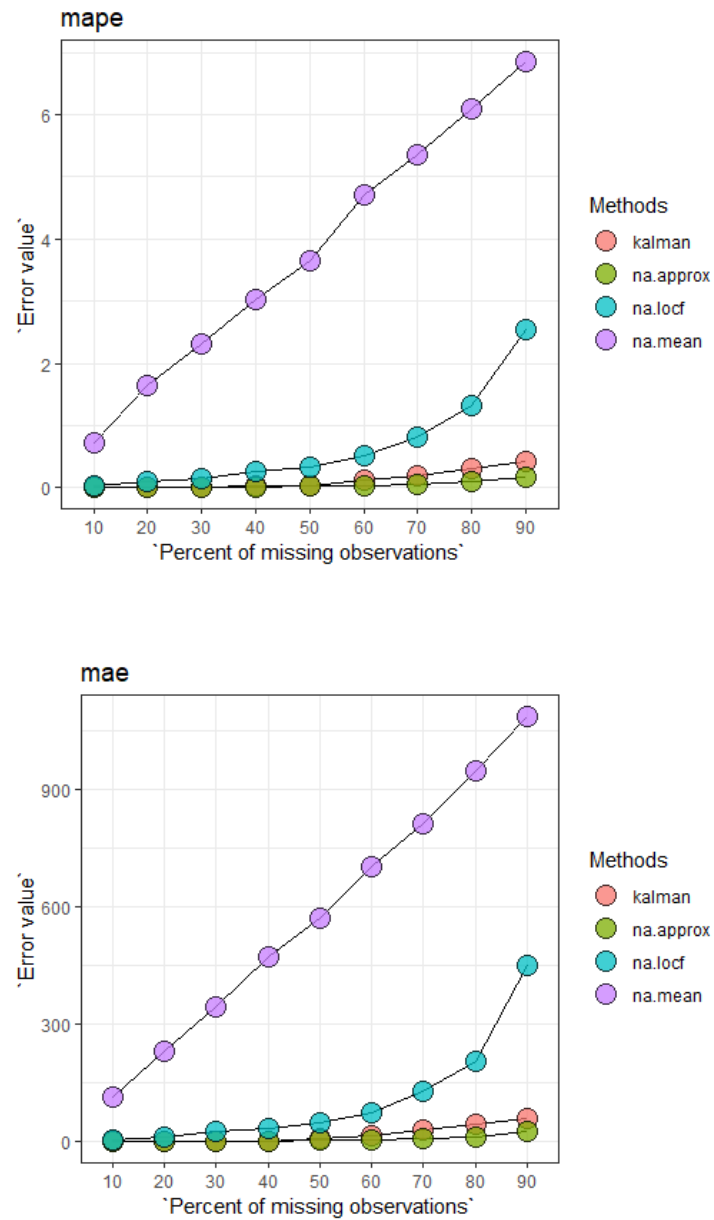
Como se observó, el filtro de Kalman compite de manera adecuada contra los métodos de imputación que realizan interpolación lineal. Sin embargo, al realizar imputaciones en series que contienen una cantidad grande de valores perdidos, por la naturaleza recursiva del filtro y su forma de realizar la predicción de valores, no sería el mejor método para imputar.

Para tener una medición de la métrica de error más puntual y menos gráfica, a continuación, se presentan tres gráficos que miden la métrica del error al imputar a distintos porcentajes de valores perdidos. El primero, mide la raíz del error cuadrático medio (RMSE) de la imputación realizada por los cuatro métodos. El segundo, la media absoluta del porcentaje de error (MAPE). El tercero y último, el error absoluto medio (MAE).

Al utilizar la raíz del error cuadrático medio, a primera impresión se observa que la imputación por la media (na.mean) genera valores del RMSE muy elevados. En el caso de querer utilizar el valor previo del valor no observable (na.locf), se observa que al tener pocos valores perdidos en la serie (entre 10 a 20 por ciento) el valor del RMSE se parece al que se obtiene con Kalman y la interpolación lineal (na.aprox). En el caso de imputar con Kalman, el método se aproxima a la interpolación lineal, no obstante, con series que cuentan porcentajes muy altos de valores perdidos el RMSE es ligeramente mayor que el RMSE del método de interpolación.



Para el caso de la media absoluta del porcentaje del error (MAPE) y el error medio absoluto (MAE) se encuentran resultados similares a los de utilizar la raíz del error cuadrático medio.



En general, se observa tomar el valor medio de una serie y remplazarlos en los valores perdidos no es lo mejor que se pueda hacer. En cambio, tomar un valor previo al valor perdido puede funcionar simple y cuando la serie no cuenta con muchos valores perdidos. En el caso de utilizar Kalman, como método de imputación, sus errores de predicción son considerables, i.e., se puede decir que es un buen método de imputación para la mayoría de los casos.

Sección 5. Conclusión

Como se mencionó al inicio, el filtro de Kalman es utilizado para estimar estados no observables por medio de un algoritmo de predicción y corrección en base a los datos, los cuales en su mayoría pueden presentar imperfecciones. Entre ellas, series que contienen "ruido", i.e. variaciones aleatorias, valores faltantes, u outliers. De esta manera, el filtro de Kalman actúa como "filtro" para trabajar con esa información y realizar predicciones estadísticamente óptimas.

En este documento, se presentó el algoritmo de Kalman de forma general con motivos de dar el lector la intuición del filtro y su desarrollo. Se dio una introducción a lo que es el mejor predictor lineal, los modelos estado-espacio y una aplicación del filtro como método de imputación.

En el caso de la aplicación se encuentra que, 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 suavizado de Kalman, y tener gran número de valores perdidos, prácticamente estaría extrapolando información, y como vimos en el caso de hacer pronóstico, los resultados pueden ser un poco inapropiados.

Sección 6. Bibliografía

Bibliografía

1. Alejandro, I. y Cortez, W. (2009). NAIRU y política monetaria en México (1987-2004). Ed. Universidad de Guadalajara, 2009.
2. Aoki M. (1987) Properties of State Space Models. In: State Space Modeling of Time Series. Universitext. Springer, Berlin, Heidelberg

3. Aoki M. (1987) Properties of State Space Models. In: State Space Modeling of Time Series. Universitext. Springer, Berlin, Heidelberg
4. Aoki M. (1990) Properties of State Space Models. In: State Space Modeling of Time Series. Universitext. Springer, Berlin, Heidelberg
5. Brockwell and R. A. Davis (1996) Introduction to Time Series and Forecasting. Springer
6. Cipra, T. y Romera, R. (1997). Kalman filter with outliers and missing observations, Sociedad de Estadística e Investigación Operativa Test (1997) Vol. 6, No. 2, pp. 379-395.
7. Christensen, Ronald (1987). Plane Answers to Complex Questions. The theory of linear models. Ed. Springer, 19987.
8. Christensen, Ronald (1991). Linear Model for Multivariate, Time Series, and Spatial Data. Ed. Springer, 1991.
9. Demirhan, H. y Renwick, Z. (2018) Missing value imputation for short to mid-term horizontal solar irradiance data. Applied Energy 225 (2018) 998–1012.
10. Hamilton, J. (1994). Time Series Analysis, Princeton University. 1994
11. Hartigan, J. A. (1969). Linear Bayesian Methods. Journal of the Royal Statistical Society, Series B, 31, 446-454, 1996.
12. Kees, E. Jan, P. (2011). Forecasting with real-time macroeconomic data: The ragged-edge problem and revisions, Journal of Macroeconomics 33 (2011) 784–792.
13. Lütkepohl, H (2005). New Introduction to Multiple Time series Analysis. Ed. Springer, 2005.

14. Marcus, W. Neeraj, B., Gualberto, A. , Kishore, K. (2017). R Package imputeTestbench to Compare Imputation Methods for Univariate Time Series, The R Journal Vol. 10/1, July 2018
15. Matthew, B. et.al. (2017). A KALMAN FILTERING TUTORIAL FOR UNDERGRADUATE STUDENTS, International Journal of Computer Science and Engineering Survey Vol.8, No.1, 2017.
16. Mikusheva, A.(2017) course materials for 14.384 Time Series Analysis, Fall 2007. MIT OpenCourseWare (<http://ocw.mit.edu>), Massachusetts Institute of Technology
17. White, H. (1982). Maximum Likelihood Estimation of Misspecified Models,Econometrica, Vol. 50, No. 1. (Jan., 1982), pp. 1-25.