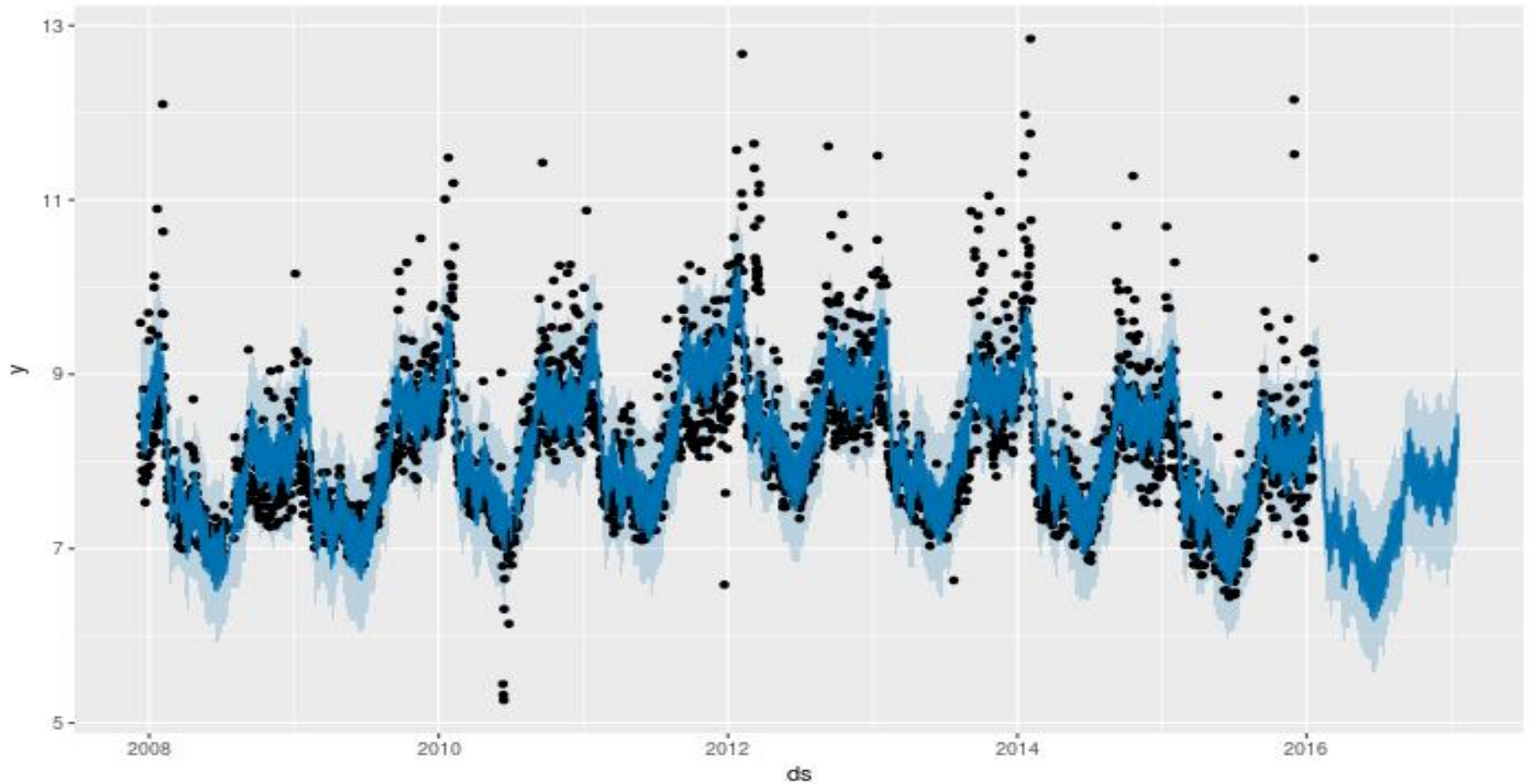


Introducción a las series cronológicas



Oscar Centeno Mora

Índice

1

Importancia de
pronosticar series
de tiempo

4

Tipos de series de
tiempo

2

¿Qué es una serie
de tiempo?

5

Componentes y
análisis de las
series de tiempo

3

Objetivo de la
serie temporal

6

Análisis visual de
la serie de
tiempo

Índice

7

Longitud de la
serie de tiempo

8

Etapas de análisis
de una serie de
tiempo

9

Medidas de
rendimiento

Índice

1

Importancia de
pronosticar series
de tiempo

Importancia de pronosticar una ST




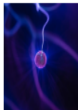








- El análisis de las series temporales juega un rol esencial tanto para conocer un determinado fenómeno como para pronosticar hacia el futuro.
- Los millones y billones que gastan las empresas y otros para predecir por ejemplo el movimiento financiero, la demanda salarial, el comportamiento y la demanda de la población, etc., son tan solo unos pocos casos de argumenta para apostarle al análisis de las series temporales.
- “By **time series** analysis, we build models depicting the cutting tool states, coacervate information from dynamic data and construct feature vectors for discrimination.”



Importancia de pronosticar una ST

← → ↻ | Seguro | <https://www.kaggle.com/competitions>

Aplicaciones M M y Book download Nueva carpeta (?) Test de velocidad Google Traducción Inicio

	Predict Future Sales Final project for "How to win a data science competition" Coursera course Playground · 5 months to go	Kudos 1,197 teams
	New York City Taxi Fare Prediction Can you predict a rider's taxi fare? Playground · a month to go ·  tabular data, regression	Knowledge 435 teams
	Flavours of Physics: Finding $\tau \rightarrow \mu\mu$ (Kernels Only) Identify a rare decay phenomenon Playground · a month to go	Knowledge 33 teams
	Store Item Demand Forecasting Challenge Predict 3 months of item sales at different stores Playground · a month to go ·  time series, tabular data	Knowledge 205 teams
	Costa Rican Household Poverty Level Prediction Can you identify which households have the highest need for social welfare assistance? Playground · a month to go ·  tabular data, multiclass classification	Swag 280 teams
	Forest Cover Type (Kernels Only) Use cartographic variables to classify forest categories Playground · a month to go ·  ecology, forestry, tabular data	Knowledge 197 teams
	What's Cooking? (Kernels Only) Use recipe ingredients to categorize the cuisine Playground · a month to go ·  food and drink, text data, multiclass classification	Knowledge 254 teams



Importancia de pronosticar una ST

<https://www.slideshare.net/DevieMohan1/endofyear-fintech-thoughts-and-predictions>

There is definitely a banking investment race...



Prediction 3: Gamification, RegTech and Tax Advisory will see big investments from banks

Índice

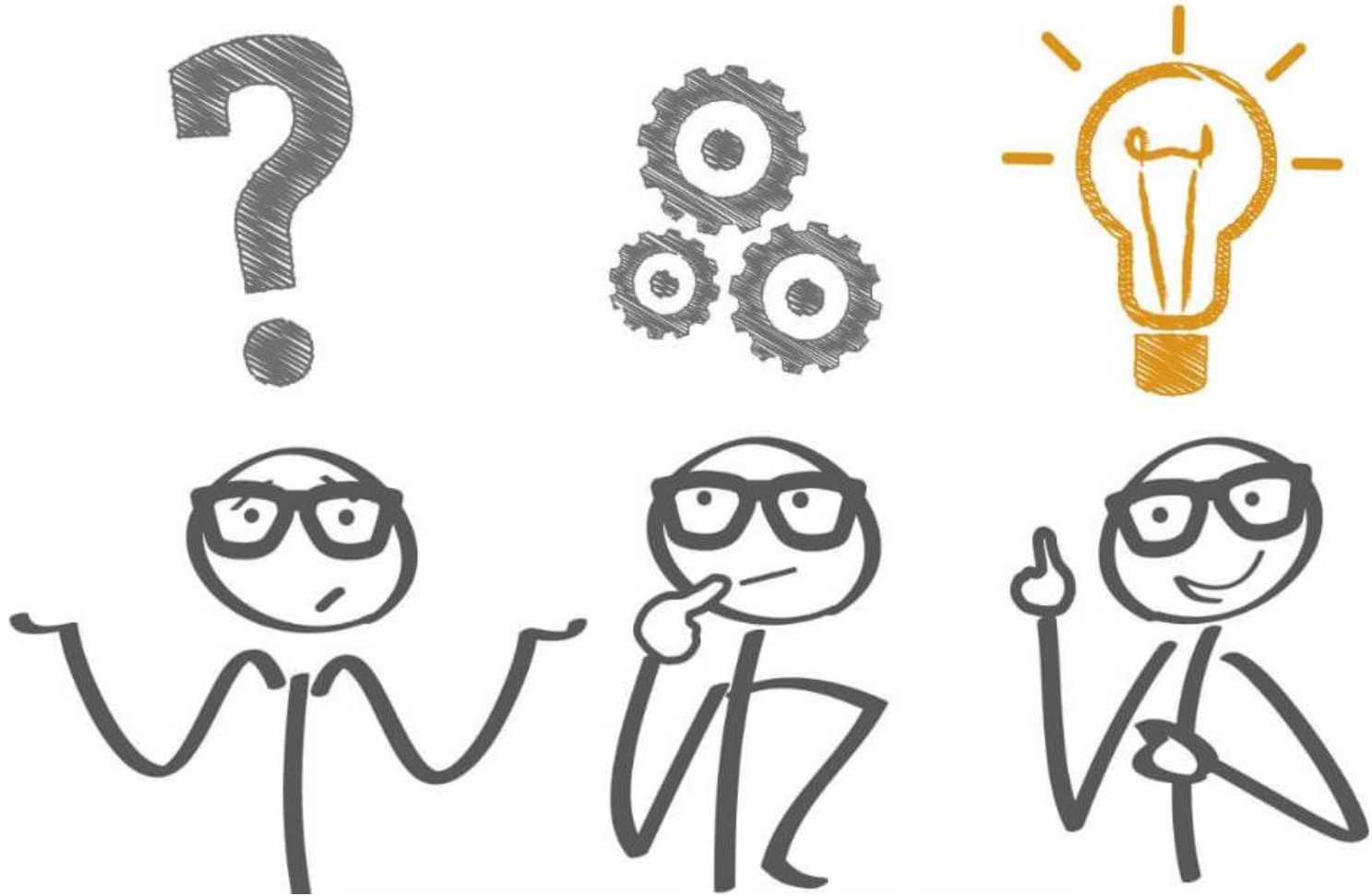
1

Importancia de
pronosticar series
de tiempo

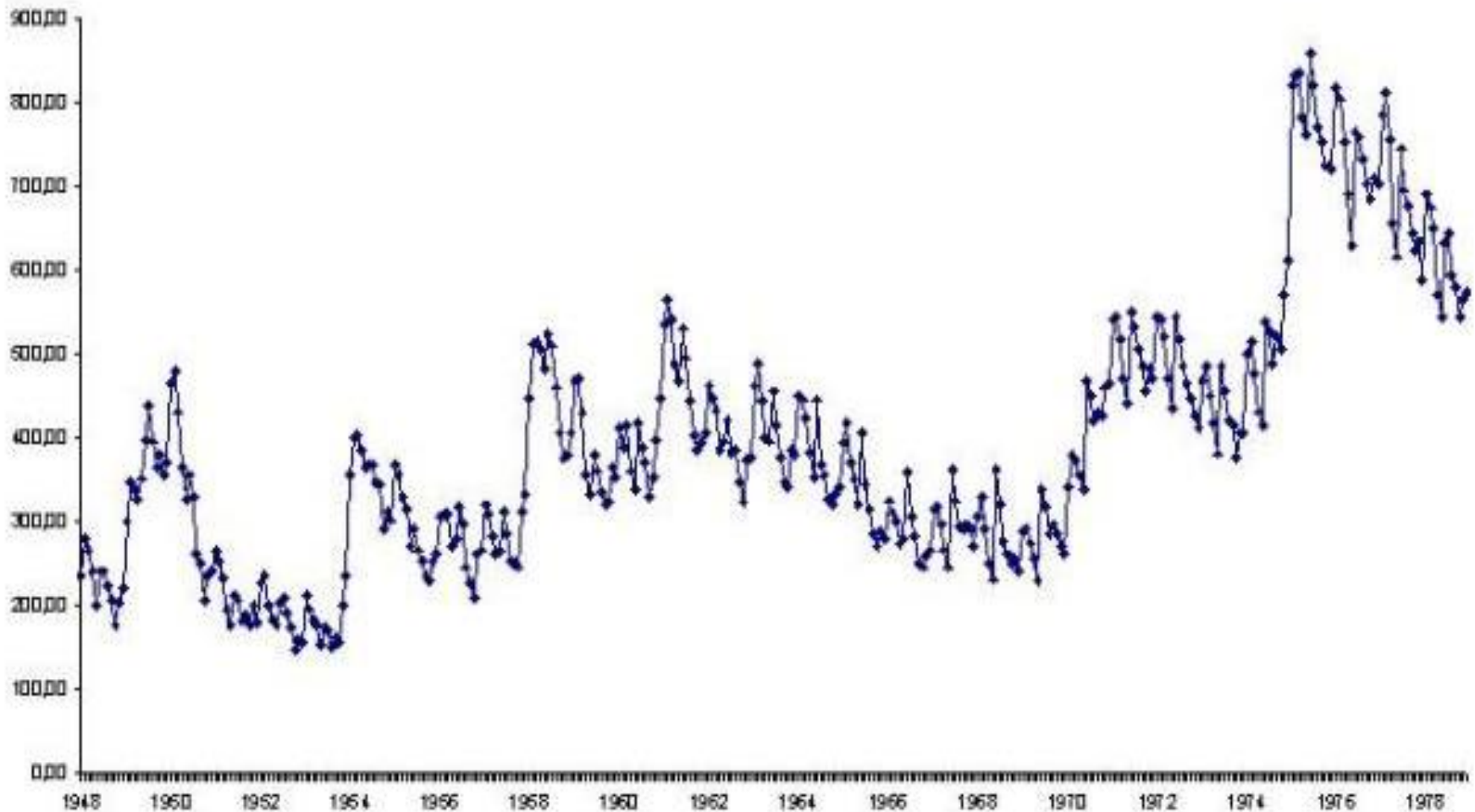
2

¿Qué es una serie
de tiempo?

¿Qué es una serie cronológica o serie de tiempo?



¿Qué es una serie cronológica o serie de tiempo?



¿Qué es una serie cronológica o serie de tiempo?

- Las series de tiempo son observaciones sobre un determinado fenómeno efectuadas en el tiempo, en lapsos ojala equivalente, o con intervalos de igual valor.
- Ejemplos: exportaciones mensuales, ventas diarias de un producto, casos semanales de sida, temperatura promedio diaria, tasa anual de mortalidad, numero mensual de divorcios.
- Para un determinado tiempo t , que se considera el tiempo actual, se dice que una serie se constituye de tres momentos: pasado (rezagos, $Y_{t-1}, Y_{t-2}, \dots, Y_{t-k}$), presente (Y_t), y los pronósticos ($Y_{t+1}, Y_{t+2}, \dots, Y_{t+h}$).

¿Qué es una serie cronológica o serie de tiempo?

- La ecuación de una serie temporal univariada, con lapsos entre los tiempos dichos equidistantes o iguales, se presenta como:

$$Y_{t-k}, \dots, Y_{t-2}, Y_{t-1}, Y_t, Y_{t+1}, Y_{t+2}, \dots Y_{t+h}$$

The diagram illustrates the components of a univariate time series equation. The equation $Y_{t-k}, \dots, Y_{t-2}, Y_{t-1}, Y_t, Y_{t+1}, Y_{t+2}, \dots Y_{t+h}$ is shown with blue curly braces underneath. The first brace, under the terms $Y_{t-k}, \dots, Y_{t-2}, Y_{t-1}$, points to a blue rounded rectangle labeled "Rezagos (k)". The second brace, under the terms $Y_{t+1}, Y_{t+2}, \dots Y_{t+h}$, points to a blue rounded rectangle labeled "Adelantos (h)".

Rezagos (k)

Adelantos (h)

- Toda serie cronológica posee una parte determinística y estocástica.

¿Qué es una serie cronológica o serie de tiempo?

- Una serie cronológica debería siempre estar compuesta de dos partes:
- La serie estocástica: una parte conocida (sistemática) susceptible de predecir y de una parte totalmente desconocida (aleatoria).
- La serie determinística: el pronóstico es una ecuación matemática sin error, dado que no se posee más que el componente determinístico. Es una variable que está determinada o fija y que no cambia de una muestra a otra.

¿Qué es una serie cronológica o serie de tiempo?

¿Entendemos bien por qué la serie es estocástica al inicio, y determinística en la proyección?



¿Qué es una serie cronológica o serie de tiempo?

- Una serie cronológica puede ser de tipo continua o discreta.
- Una serie es continua si los valores se obtienen para todo tiempo t en un intervalo de tiempo. Ejemplo: registro continuo de la temperatura en un laboratorio...
- Una serie es discreta si sus observaciones se obtienen sólo en momentos particulares, usualmente *equiespaciados*. Ejemplo: número diario de pacientes en un hospital; ganancia semanal de una empresa.
- El presente curso utiliza únicamente series discretas equiespaciadas

¿Qué es una serie cronológica o serie de tiempo?

- Finalmente, una serie de tiempo posee dos características esenciales:
 1. Los valores están ordenados o presentados en el tiempo.
 2. Existe una dependencia o correlación entre los valores de la serie en el tiempo.
- De no presentarse estos dos últimos, no se estaría en presencia de una serie cronológica.

Índice

1

Importancia de
pronosticar series
de tiempo

2

¿Qué es una serie
de tiempo?

3

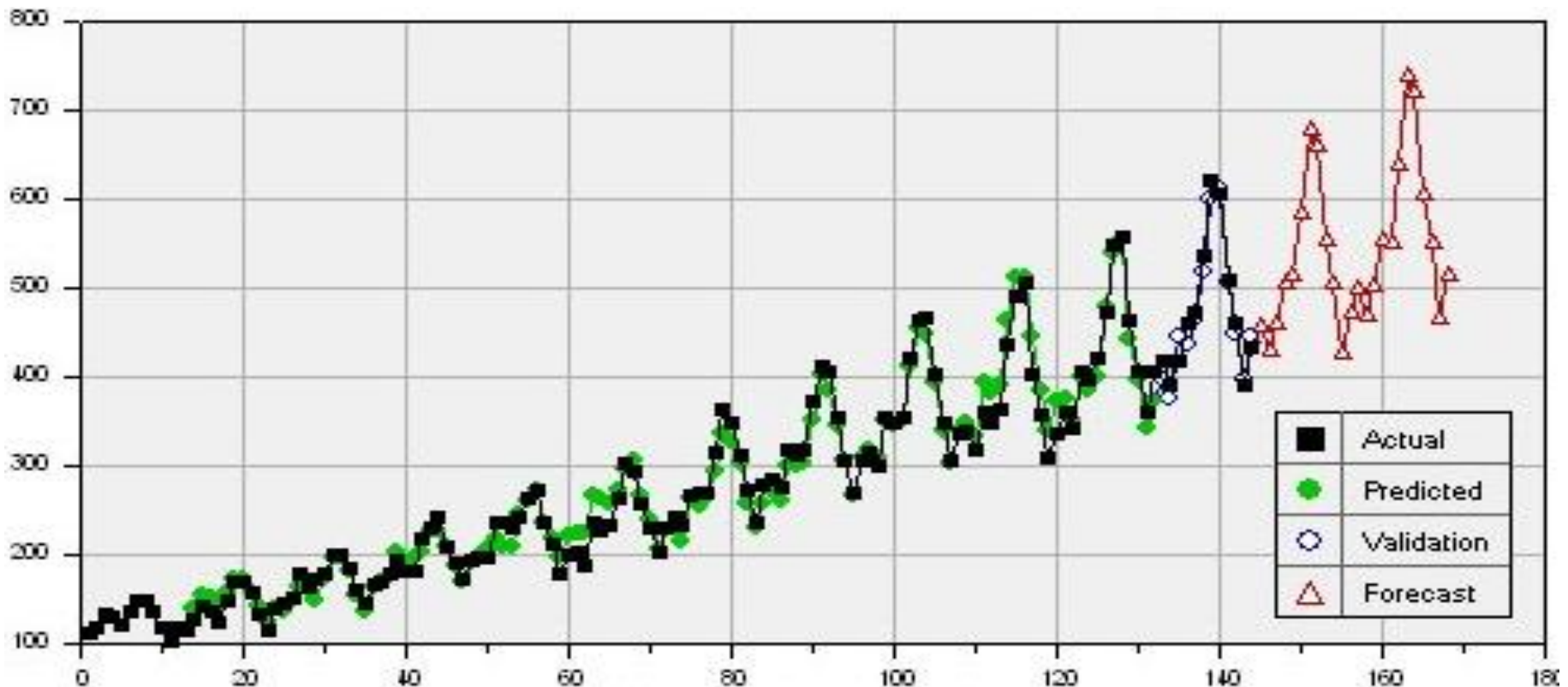
Objetivo de la
serie temporal

Objetivo del análisis por series temporales

- Un análisis por series temporales podría buscar:
 1. Descripción de la serie
 2. Adecuar un modelo o técnica estocástica
 3. Pronostico en un periodo h de la serie.
- El análisis de la serie debe preguntarse sobre el tipo de serie que se está analizando, los tipos de datos, y el período de referencia en la adecuación del mejor modelo estocástico.
- De igual forma, dependiendo de la serie, el pronóstico debe considerar ciertas restricciones.

Objetivo del análisis por series temporales

- ¿Para qué han utilizado el análisis de una ST o SC?



Índice

1

Importancia de
pronosticar series
de tiempo

4

Tipos de series de
tiempo

2

¿Qué es una serie
de tiempo?

3

Objetivo de la
serie temporal

Tipos de análisis de series de tiempo

¿Cuáles análisis temporales conocemos?

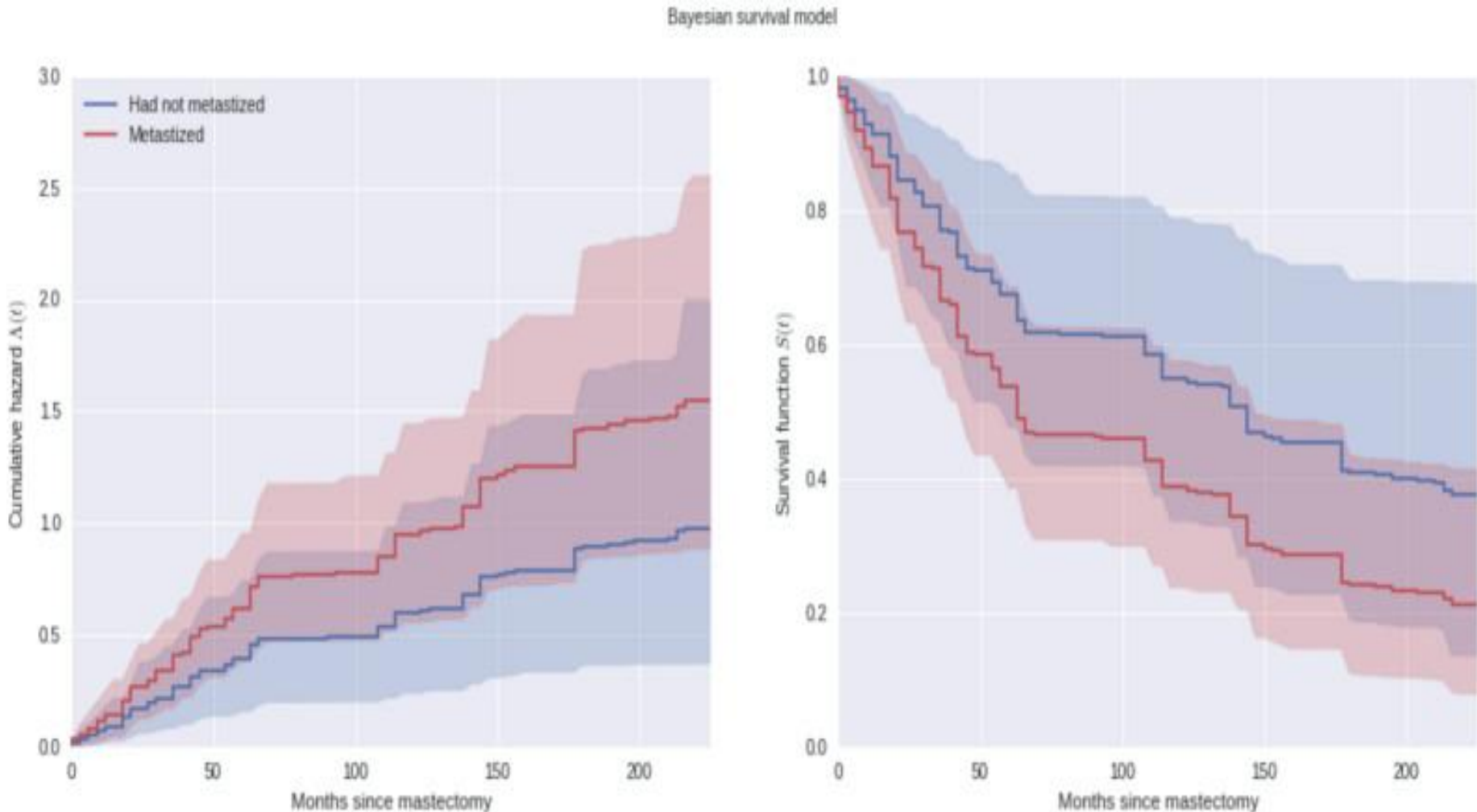


Tipos de análisis de series de tiempo

- En la carrera de estadística o economía, por lo general se estudian los siguientes modelos de tiempo:
 1. Análisis de sobrevivencia
 2. Datos panel
 3. Información geo-espacial en el tiempo
 4. Regresiones en el tiempo
 5. Series de tiempo univariadas (días)
 6. Series de tiempo univariadas (trimestres, meses, años)
 7. Series de tiempo multivariadas

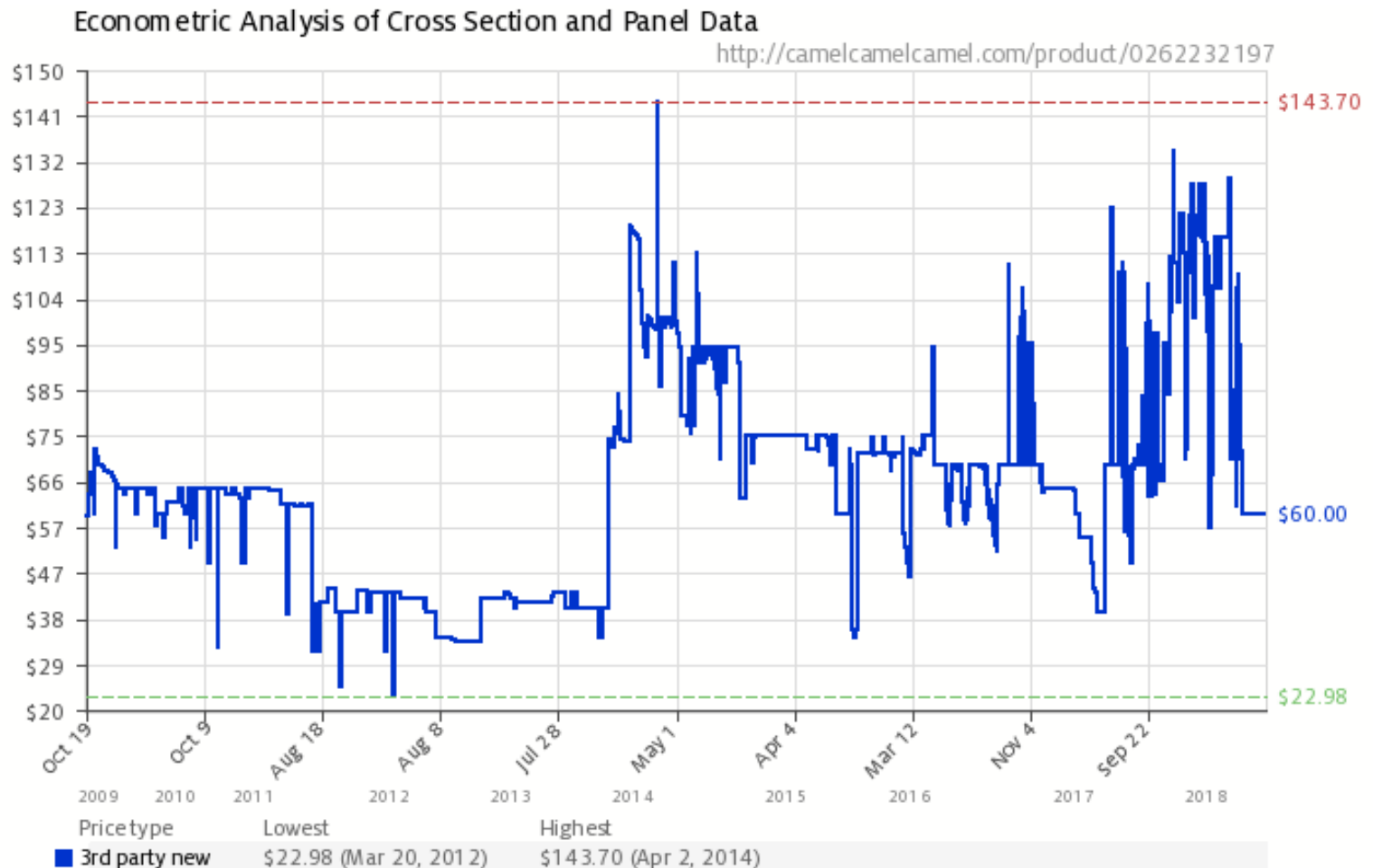
Tipos de análisis de series de tiempo

Análisis de sobrevivencia



Tipos de análisis de series de tiempo

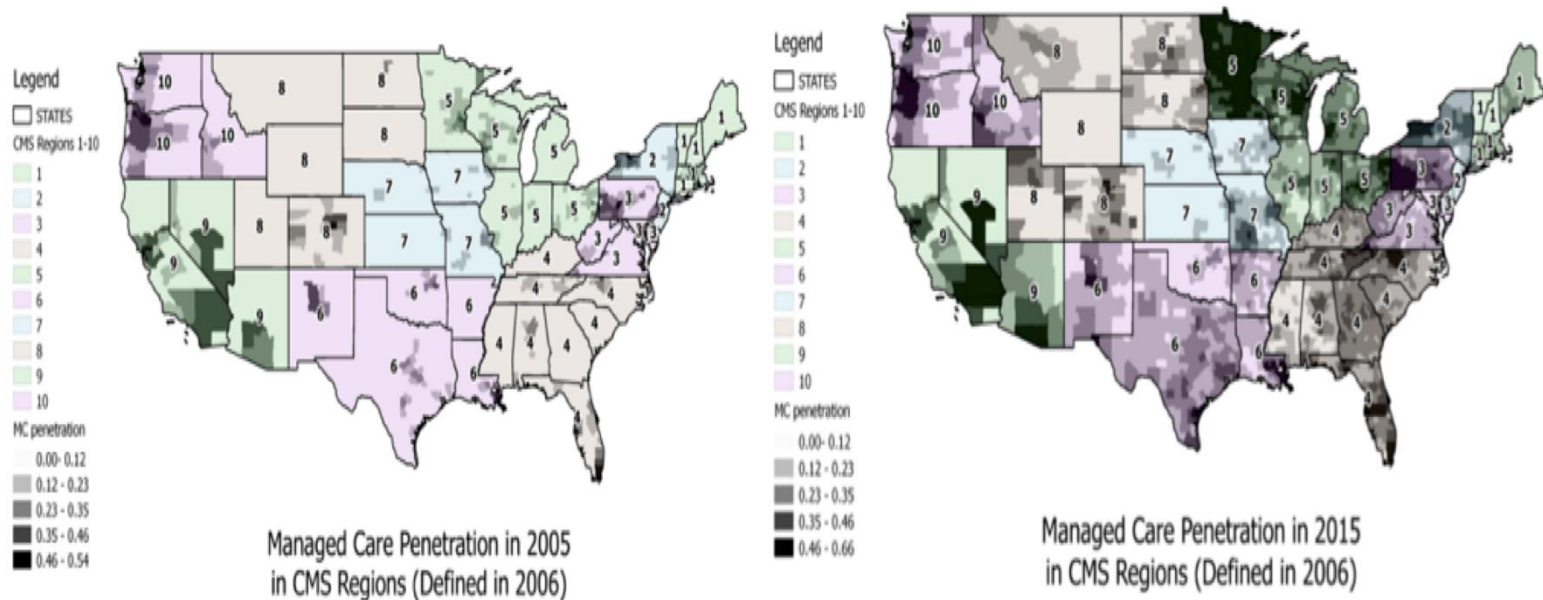
Datos panel



Tipos de análisis de series de tiempo

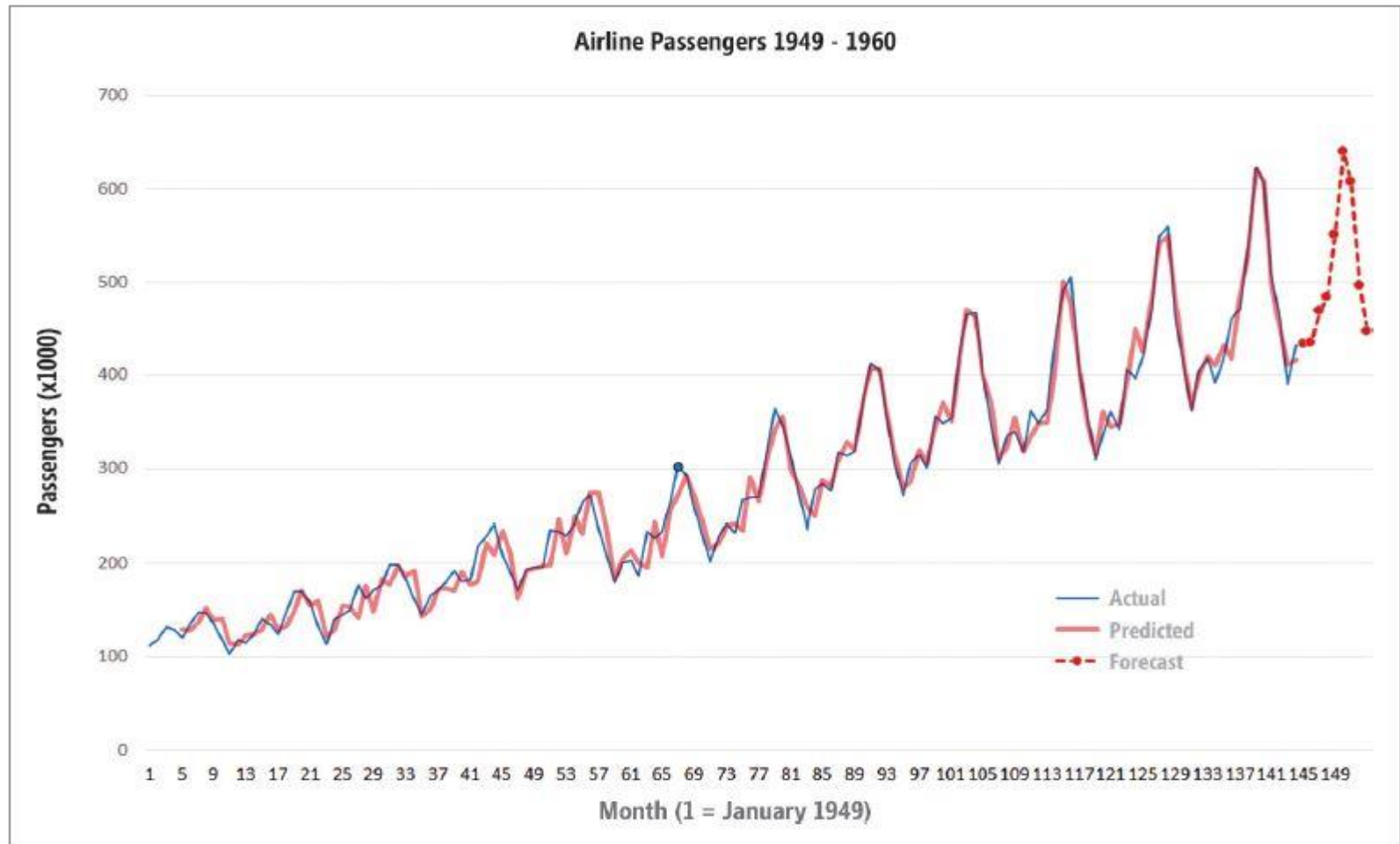
Información geo-espacial en el tiempo

Figure 1: Medicare Managed Care Penetration Across the Ten CMS Regions, Before and After the MMA



Tipos de análisis de series de tiempo

Regresión en el tiempo

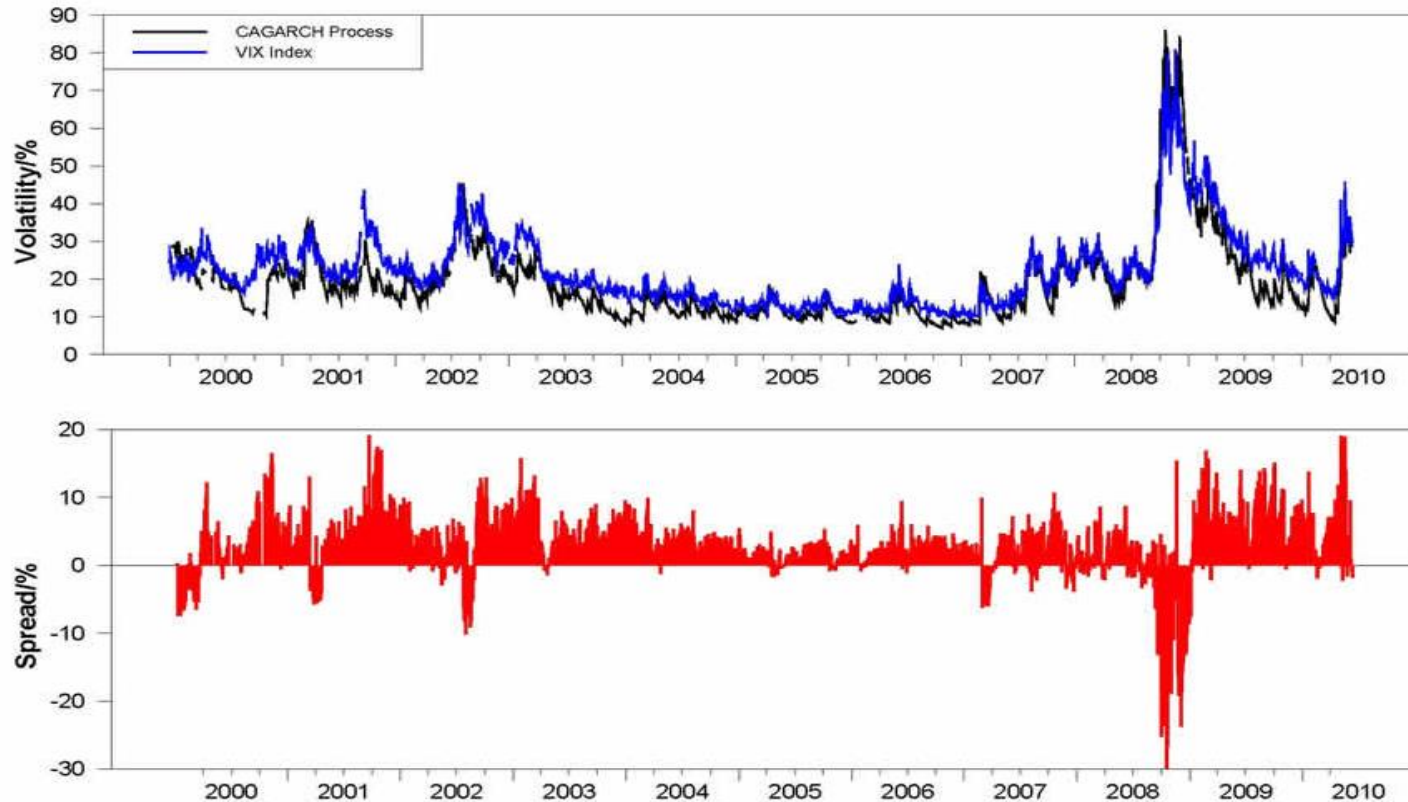


Tipos de análisis de series de tiempo

Series de tiempo univariadas (días)

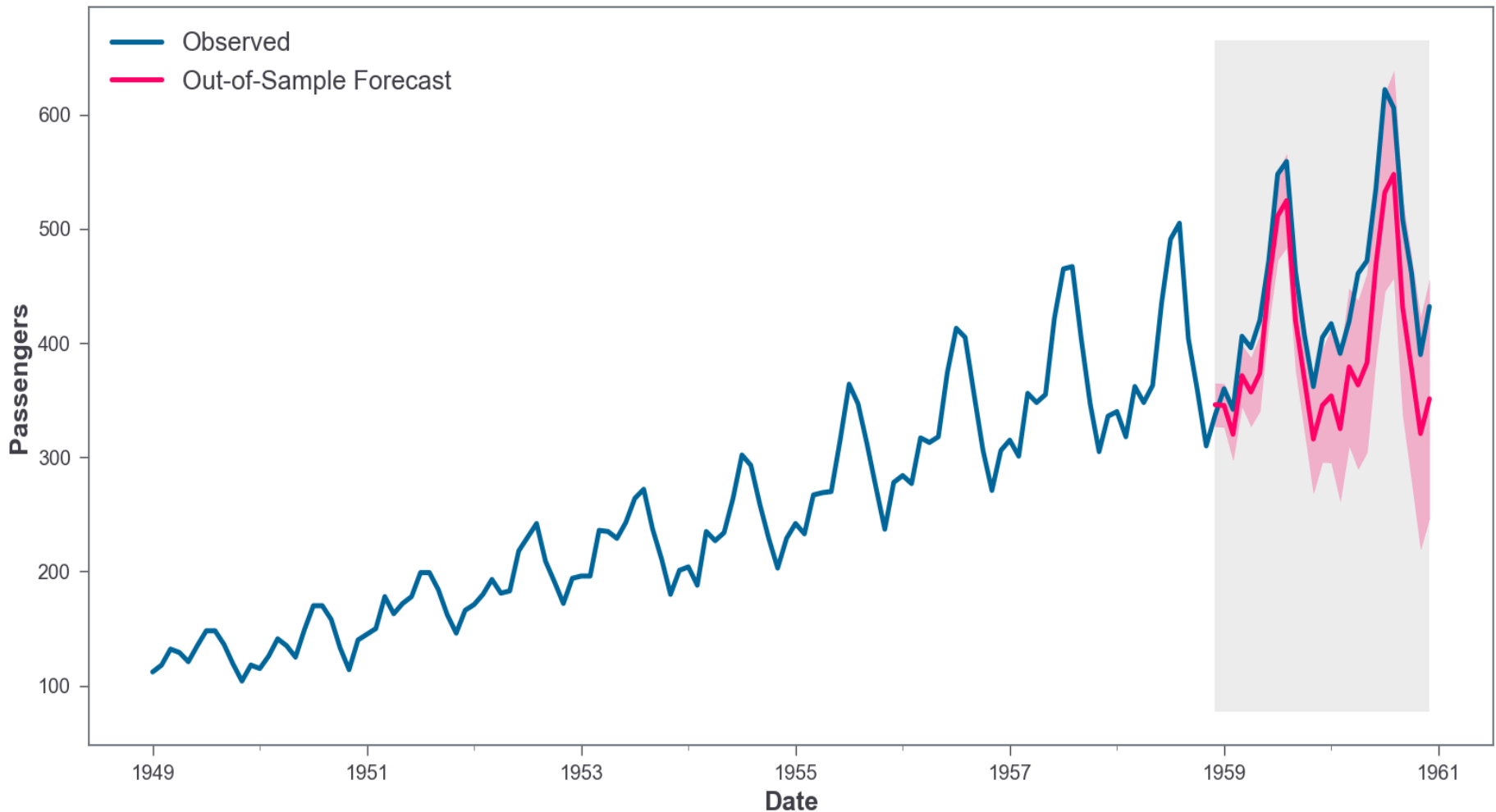
Comparison of CAGARCH(1,1) Model and VIX Index

Resolution: Daily; Range: 2000-01-03 -- 2010-06-14



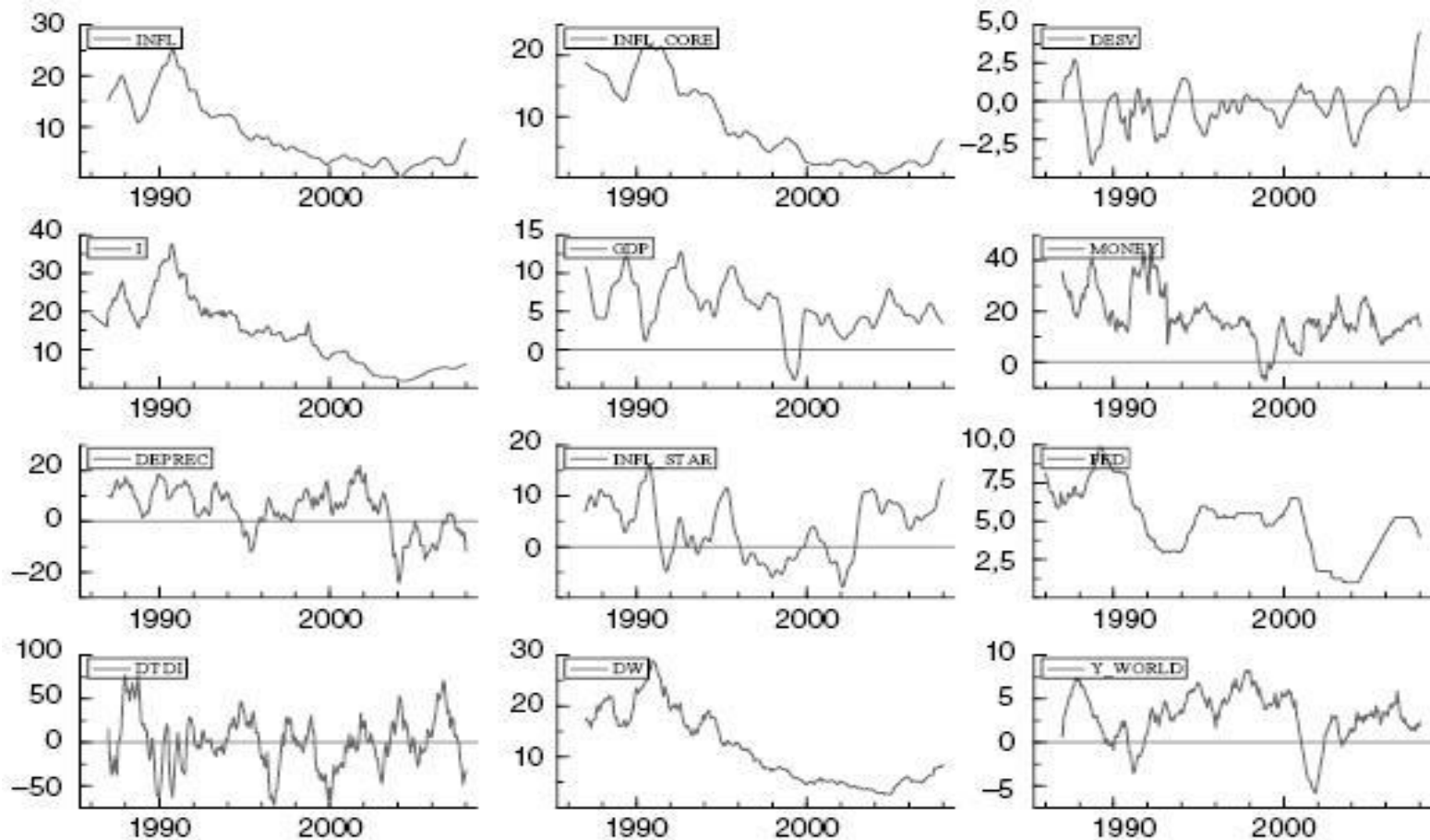
Tipos de análisis de series de tiempo

Series de tiempo univariadas (trimestres, meses, años)



Tipos de análisis de series de tiempo

Series de tiempo multivariadas



Índice

1

Importancia de
pronosticar series
de tiempo

4

Tipos de series de
tiempo

2

¿Qué es una serie
de tiempo?

5

Componentes y
análisis de las
series de tiempo

3

Objetivo de la
serie temporal

Componentes de una serie cronológica

- Una serie cronológica, de forma clásica, se constituye de 4 componentes: tendencia (T), movimientos cíclicos (C), variación estacional (E), y movimientos irregulares (I).
- Una posible ecuación estaría dada como:

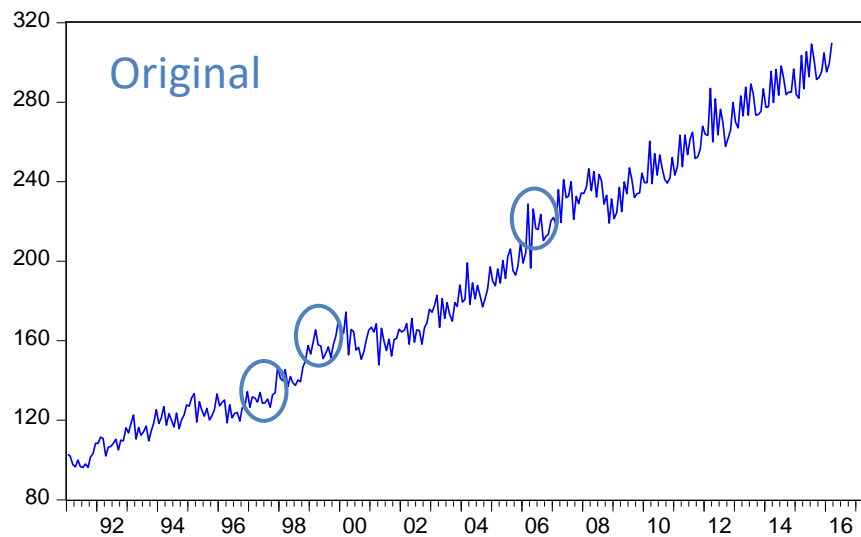
$$SC = T + C + E + I$$

- La tendencia (T) es un movimiento de larga duración que persiste durante un período de tiempo, largo o corto. Se trata de ver si existe algún crecimiento, decrecimiento o constancia en la serie.

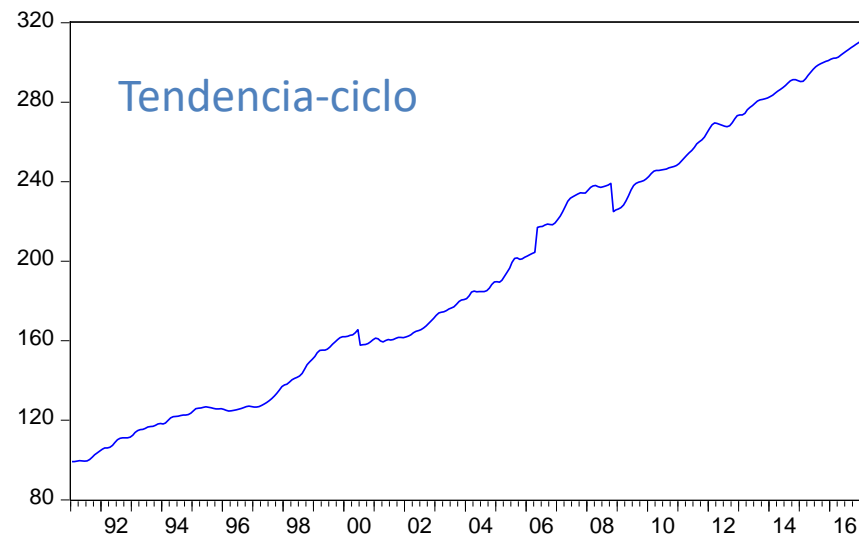
Componentes de una serie cronológica

- Los movimientos cíclicos (C), son oscilaciones alrededor de la tendencia producidos por periodos alternativos de prosperidad y depresión. Tiene una duración de 2 a 10 años, medido de máximo a máximo o de mínimo a mínimo. Pueden ser periódicos o no.
- La variación estacional (E) se le atribuye a los los movimientos que se producen dentro del año y que se repiten de un año a otro. Se observa en algunas series de periodicidad mayor al año (mensual, trimestral, semanal, etc). Se produce por cambios climáticos, cambios en las costumbres, festividades, inicio de clases, navidad, etc. Es también un ciclo pero corto.
- Los movimientos irregulares (I), son las oscilaciones erráticas o accidentales que obedecen a variadas causas. No siguen ningún patrón específico de comportamiento y por tanto son impredecibles. Es lo que queda de la serie luego de haber eliminado la tendencia, el ciclo y el estacional. Ejemplos: huelgas, terremotos, subida precios del petróleo, etc.

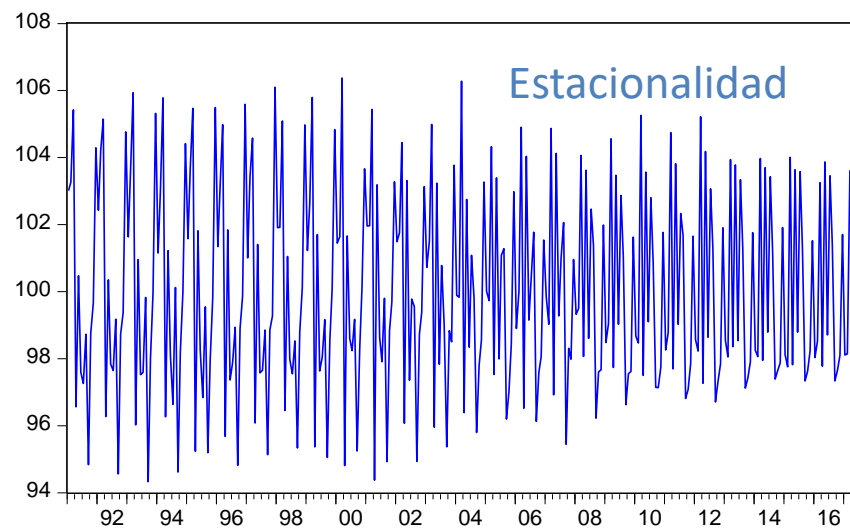
IMAE



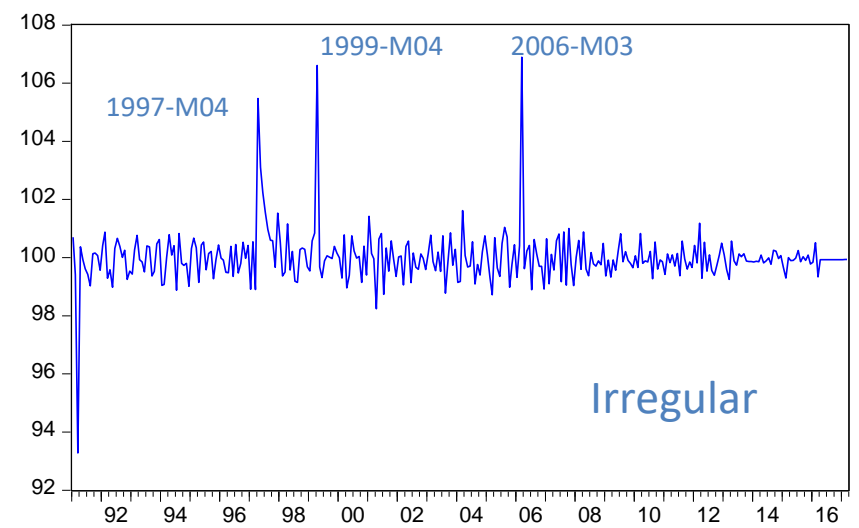
Final trend-cycle



Final seasonal component/factor



Final irregular component/factor



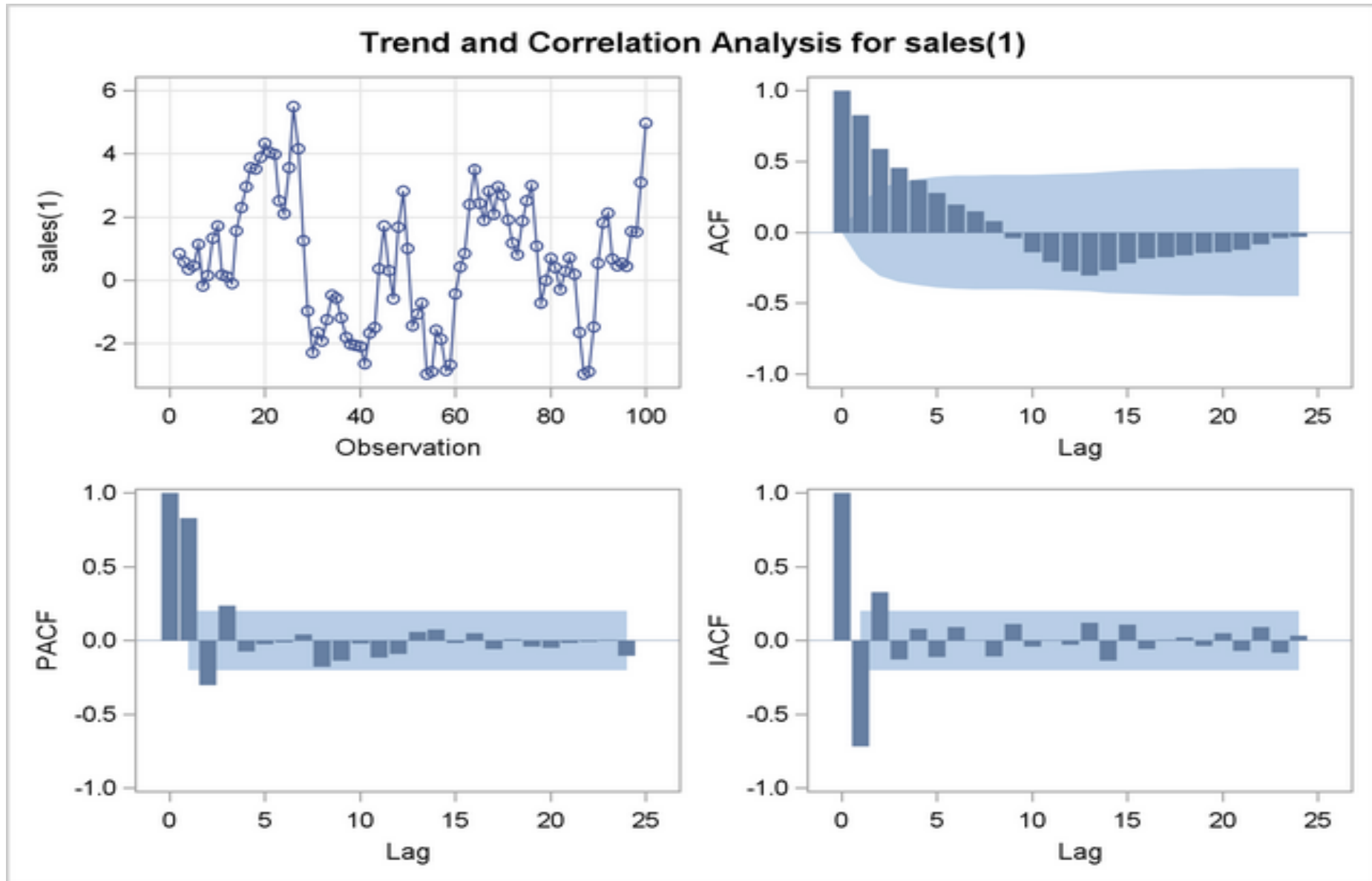
Componentes de una serie cronológica

- Las técnicas estudiadas en el presente curso tratarán tanto de identificar la composición de la serie, y pronosticar la serie en cuestión.
- Si bien existen métodos para únicamente saber la composición y hacer una descripción de la serie, el objetivo del curso se centra más en estudiar métodos que estimen y pronostiquen los valores.
- Si están interesados en un análisis más descriptivo de la serie, el Census Bureau de Estados Unidos ha desarrollado una metodología para mejorar los componentes de las series de tiempo: <https://www.census.gov/econ/currentdata/>

Otra forma de analizar la serie cronológica

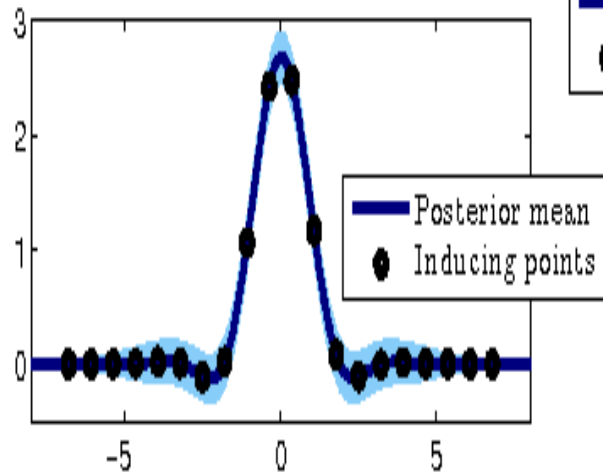
- El análisis por la descompocisión de una serie fueron los primeros en establecerse.
- Una segunda generación es el análisis de la serie por sus observaciones pasadas, o sus auto correlaciones. Este es el más utilizado para los modelos paramétricos.
- Una tercera generación es el análisis mediante la adecuación de las relaciones y las capas ocultas de las auto-correlaciones. Estos son los modelos no paramétricos.
- Estos tres tipos son los modelos que estudiaremos en el presente curso.

Otra forma de analizar la serie cronológica

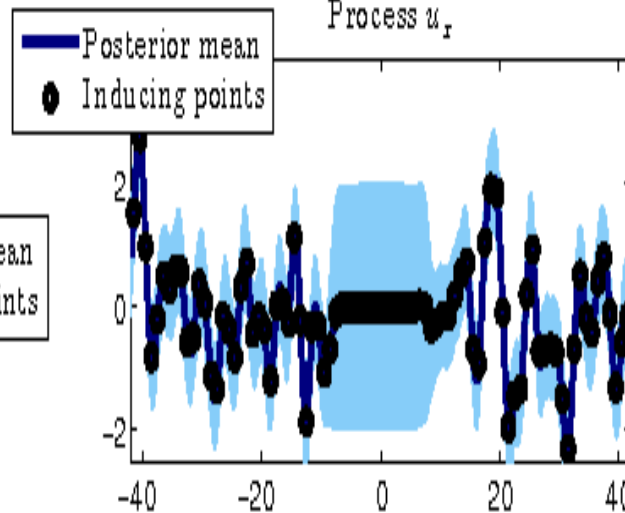


Otra forma de analizar la serie cronológica

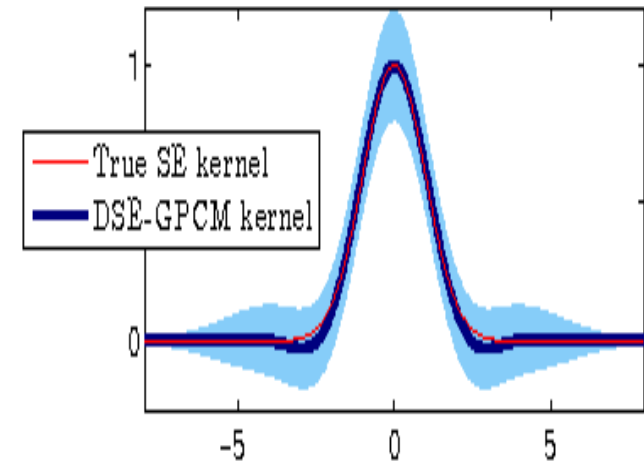
Filter $h(t)$



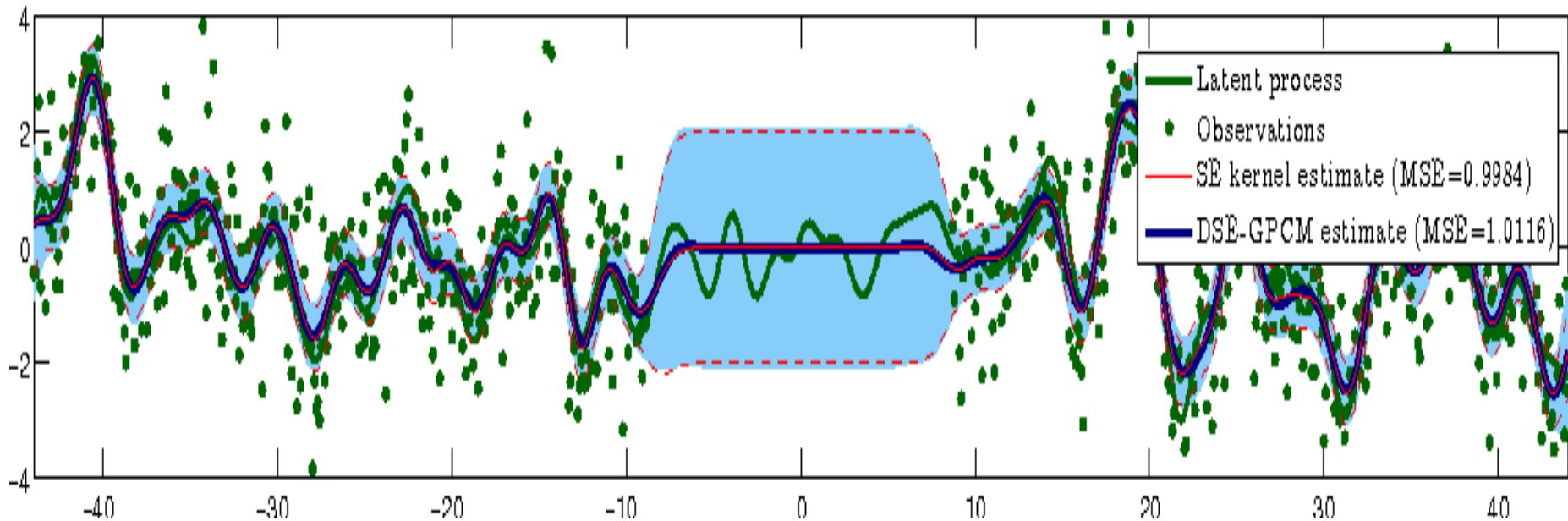
Process u_t



Kernels (normalised). Discrepancy: 5.4%



Observations, latent process and kernel estimates



Índice

1

Importancia de
pronosticar series
de tiempo

4

Tipos de series de
tiempo

2

¿Qué es una serie
de tiempo?

5

Componentes y
análisis de las
series de tiempo

3

Objetivo de la
serie temporal

6

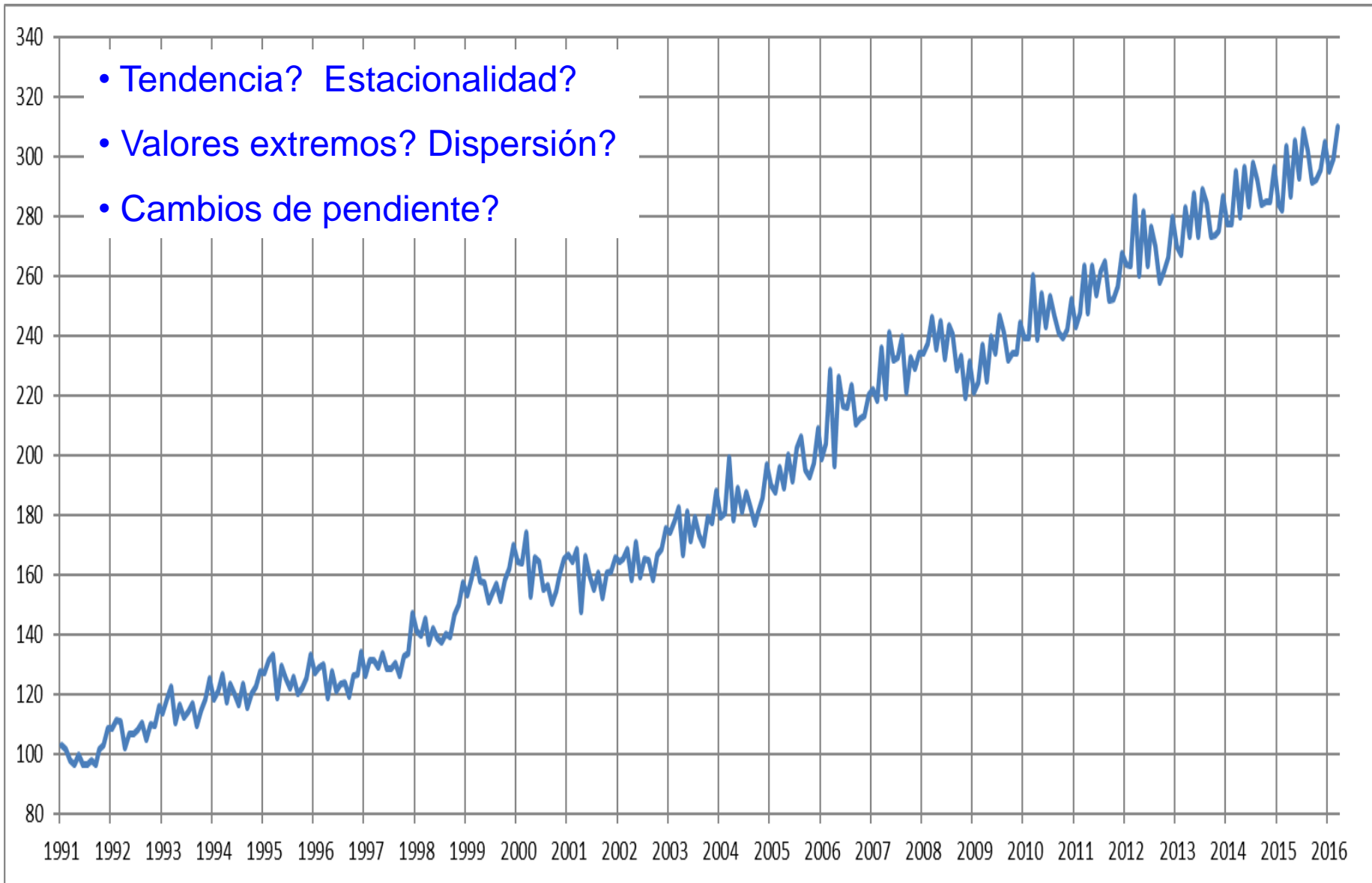
Análisis visual de
la serie de
tiempo

Análisis visual

- La investigación científica asume como una de sus primeras tareas, identificar las cosas (características o factores) que participan en un fenómeno. ¿Suenan normales, no creen?
- Los gráficos son la forma más efectiva de identificar efectos de eventos que inciden en los datos. De ser posible, estos eventos deben ser ajustados o incluidos en el modelo.
- Un gráfico permite visualizar: total de datos, tendencia, valores extremos, dispersión, cambios estructurales, cambios de pendiente, estacionalidad. Los gráficos en las series de tiempo son, de forma personal, las herramientas más poderosas para atender las necesidades de las series temporales.

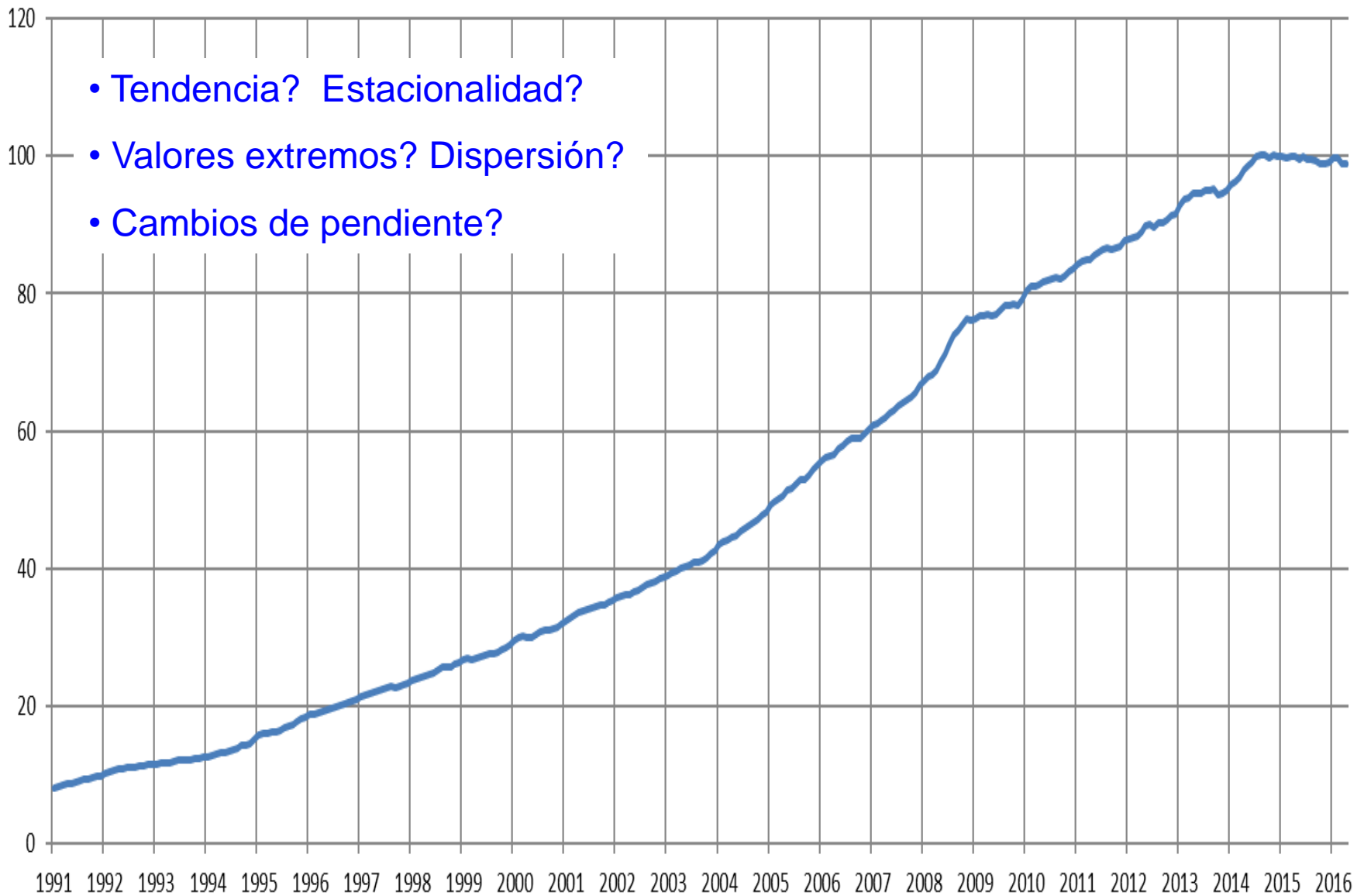
Análisis visual

- Tendencia? Estacionalidad?
- Valores extremos? Dispersión?
- Cambios de pendiente?



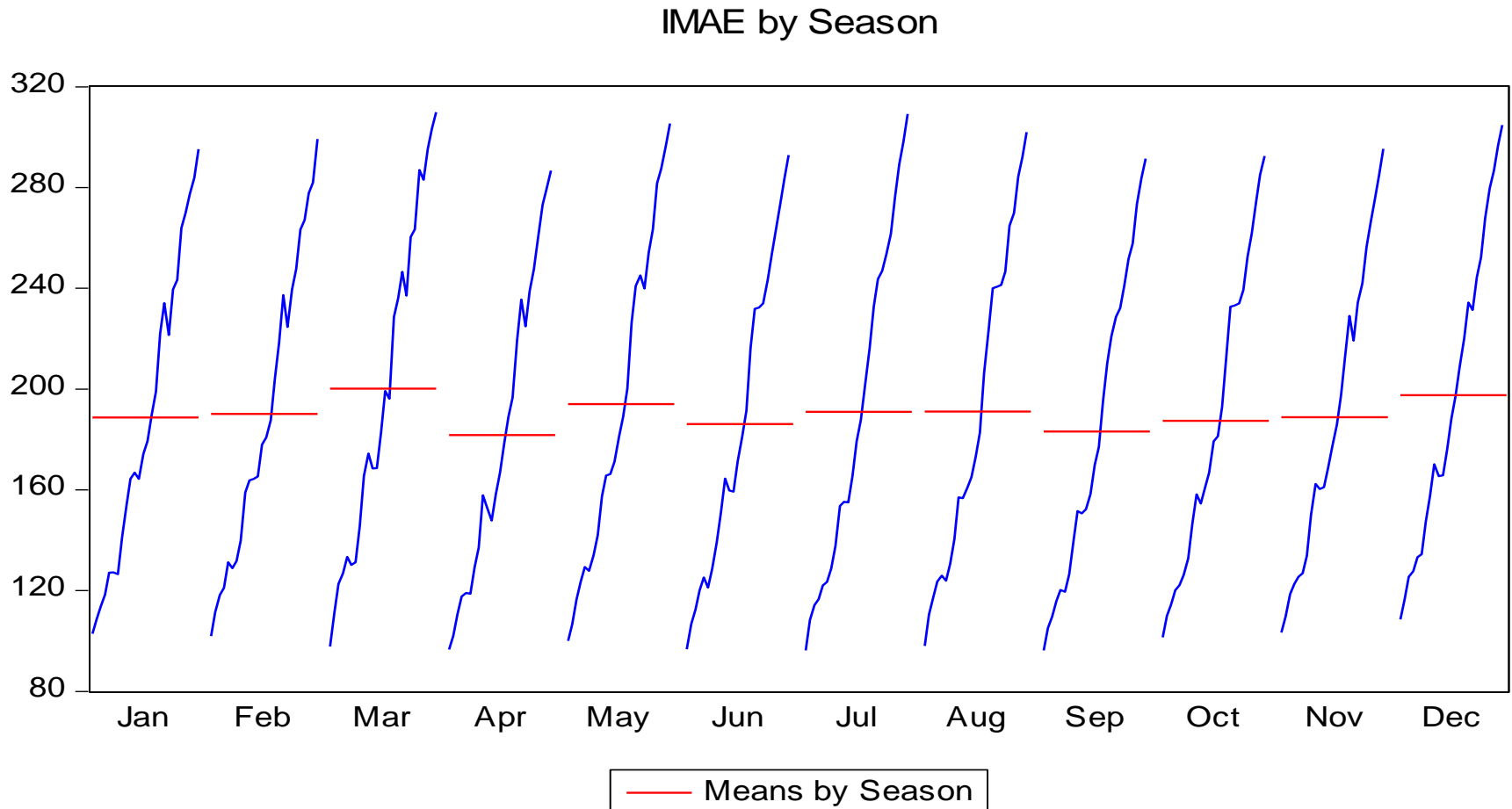
Análisis visual

- Tendencia? Estacionalidad?
- Valores extremos? Dispersión?
- Cambios de pendiente?



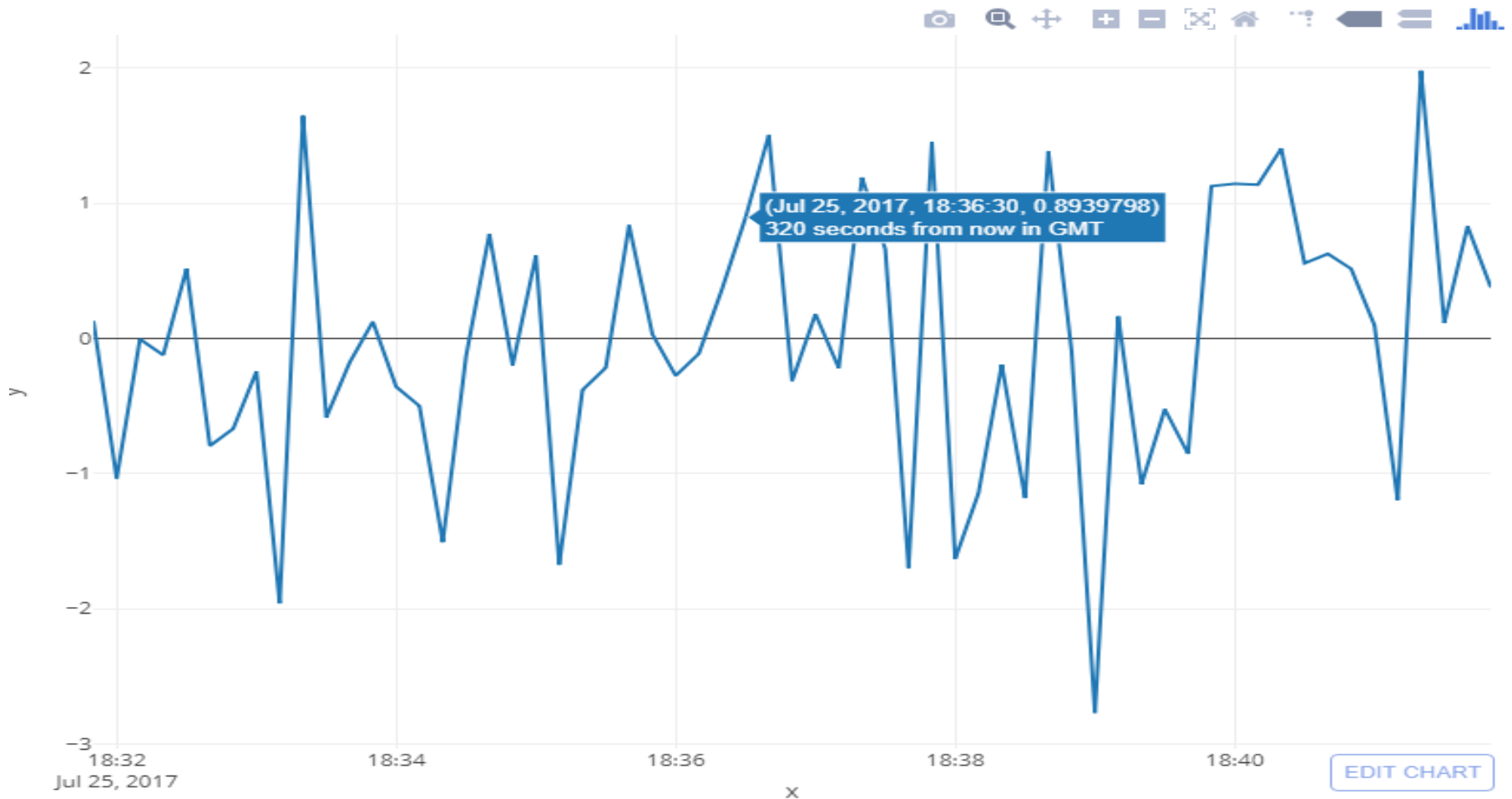
Análisis visual: el gráfico estacional

- Hay ciertos gráficos como el estacional que permite visualizar los diferentes patrones estacional en los datos.
- Acá se observa cómo en determinados momentos la serie se desvía de su patrón estacional. ¿Qué se constata?



Análisis visual de las series de tiempo

El análisis de las series temporales puede mejorarse con gráficos interactivos.
<https://plot.ly/r/time-series/>



Análisis visual de las series de tiempo

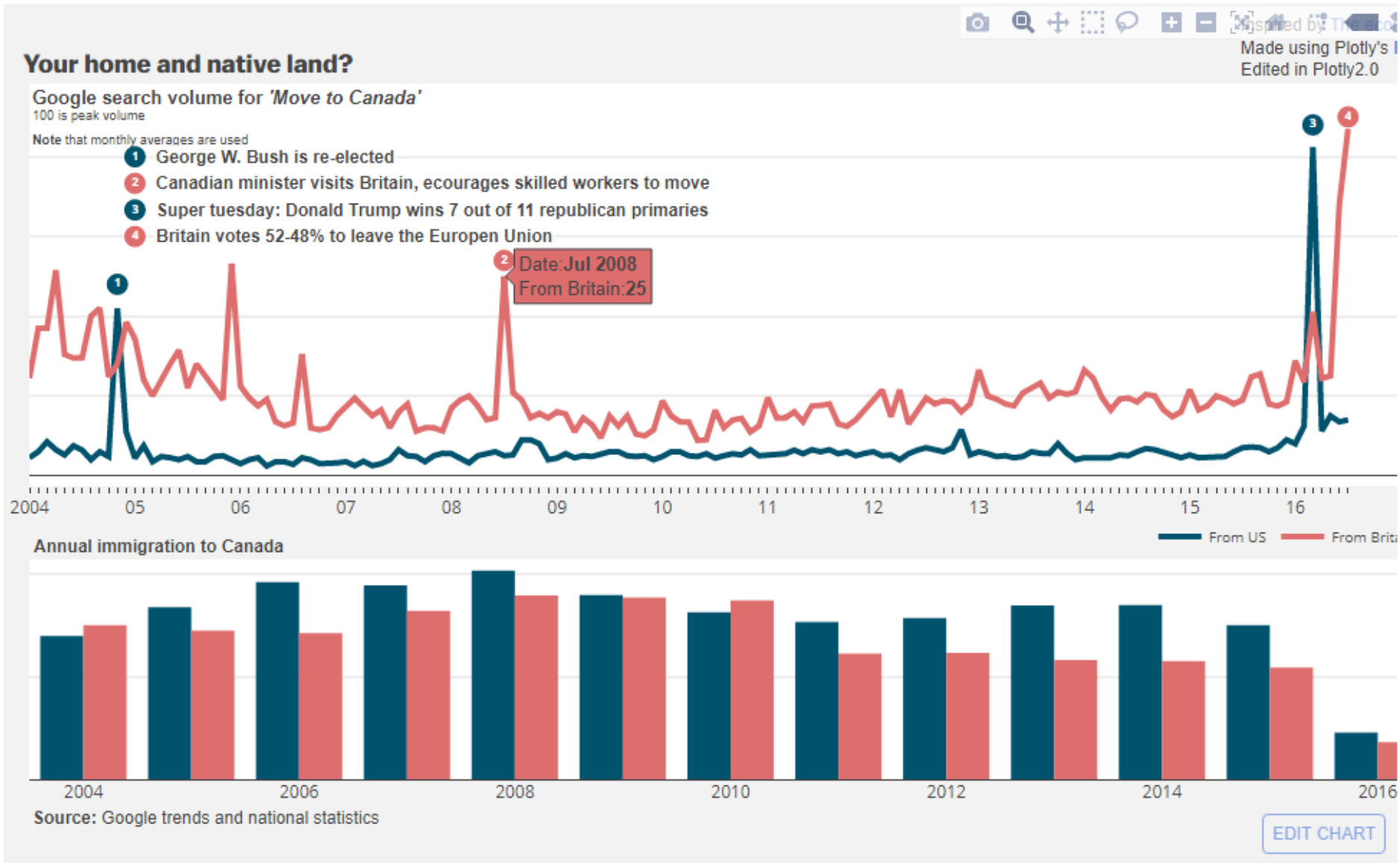
<https://plot.ly/python/time-series/>



EDIT CHART

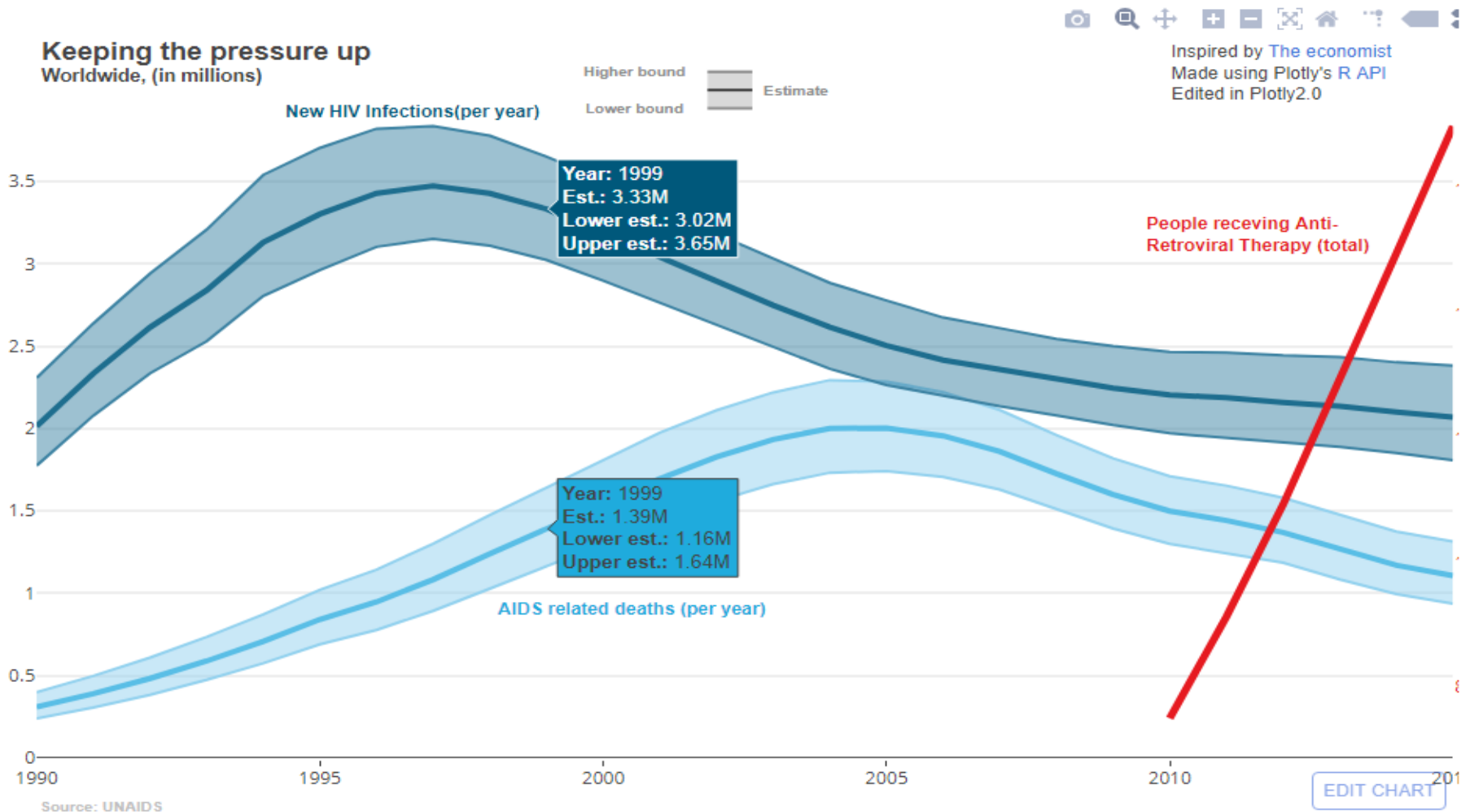
Análisis visual de las series de tiempo

<https://moderndata.plot.ly/time-series-charts-by-the-economist-in-r-using-plotly/>



Análisis visual de las series de tiempo

<https://moderndata.plot.ly/time-series-charts-by-the-economist-in-r-using-plotly/>



Análisis visual de las series de tiempo

- En R se pueden utilizar las librerías y las funciones de

-plot()

-lattice → xyplot

-ggplot

-ggplot2

-ploty

-plotly, etc...

- Python también es un excelente gráficador de series de tiempo...

Índice

7

Longitud de la
serie de tiempo

Longitud de la serie

- Depende del objetivo del estudio:
 1. Para análisis de ciclos se requieren series muy largas (más de 30 años)
 2. Para modelos univariantes se sugiere no menos de 5 años
 3. Para modelos de regresión no menos de 15 datos
 4. Para calcular la correlación entre dos variables no menos de 35 datos
- Para los modelos paramétricos se prefieren series de más de 50 datos. Para los modelos no paramétricos agrandar los datos por lo menos a 75%. Los modelos no paramétricos trabajan muy bien bajo una serie bastante extensa.
- Sin embargo, la única forma de saber si la cantidad de datos es suficiente o no, es llevando a cabo análisis de simulación y probar la sensibilidad de las estimaciones de los estimadores.

Longitud de la serie

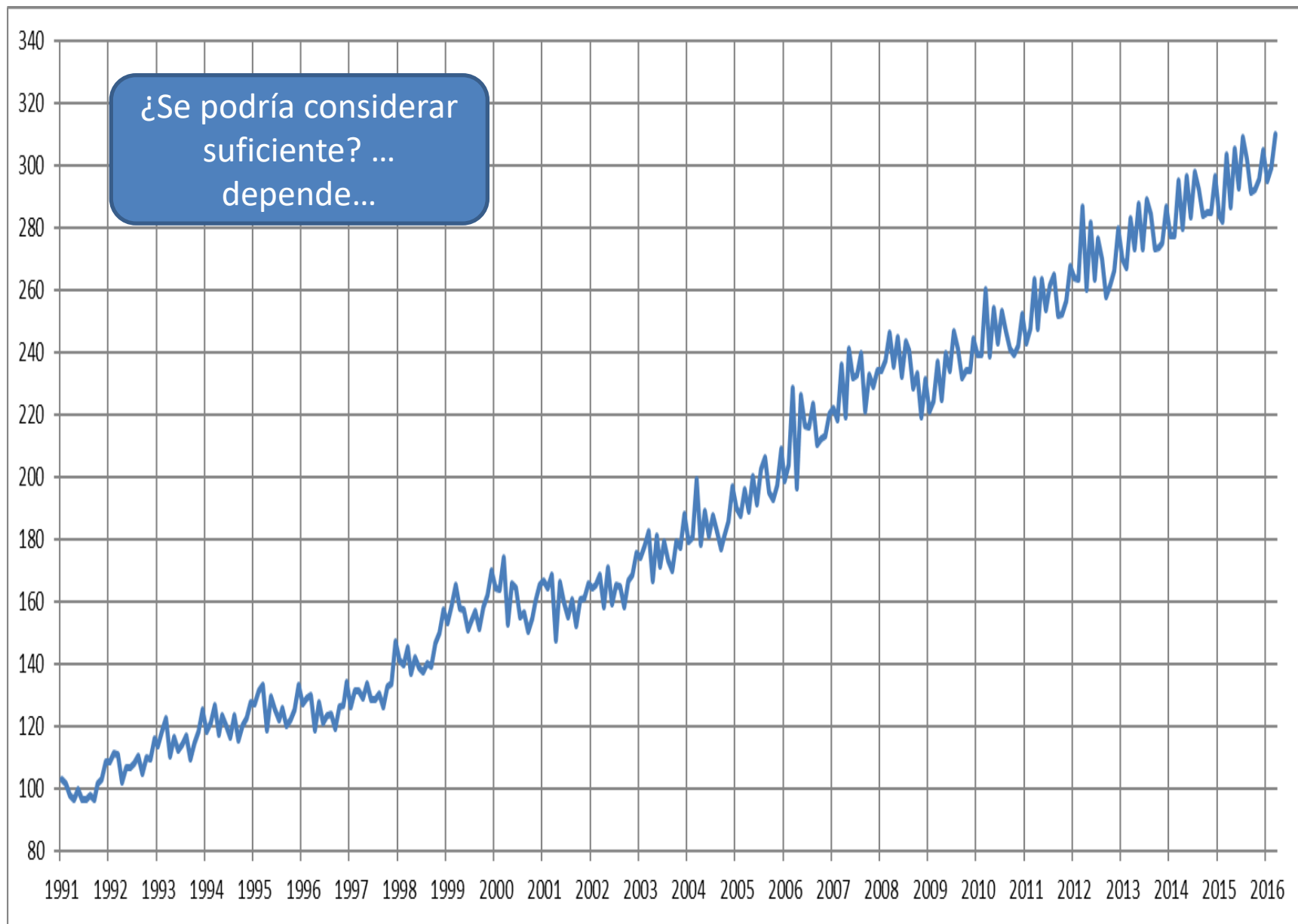
- El aporte a las series de tiempo de “La crítica de Lucas” sugiere el número de datos de una serie de tiempo que se deben utilizar, reduciéndolo a aquel periodo de datos que lucen homogéneos

Critica de Lucas: sostiene que, bajo la hipótesis de expectativas racionales, los parámetros estimados a partir de un modelo econométrico no se mantendrían. La ocurrencia de cambios de política llevaría a los agentes a modificar sus comportamientos, a fin de adecuarse a la nueva realidad. Lucas, Robert Jr.(1976), “Econometric policy evaluation: A critique”, *Conference Series on Public Policy*.

Premio Nobel en Economía en 1995. Nació en 1937 (78 años)

Existe un “trade-off” entre el tamaño de la muestra y la estabilidad del modelo

Longitud de la serie



Índice

7

Longitud de la
serie de tiempo

8

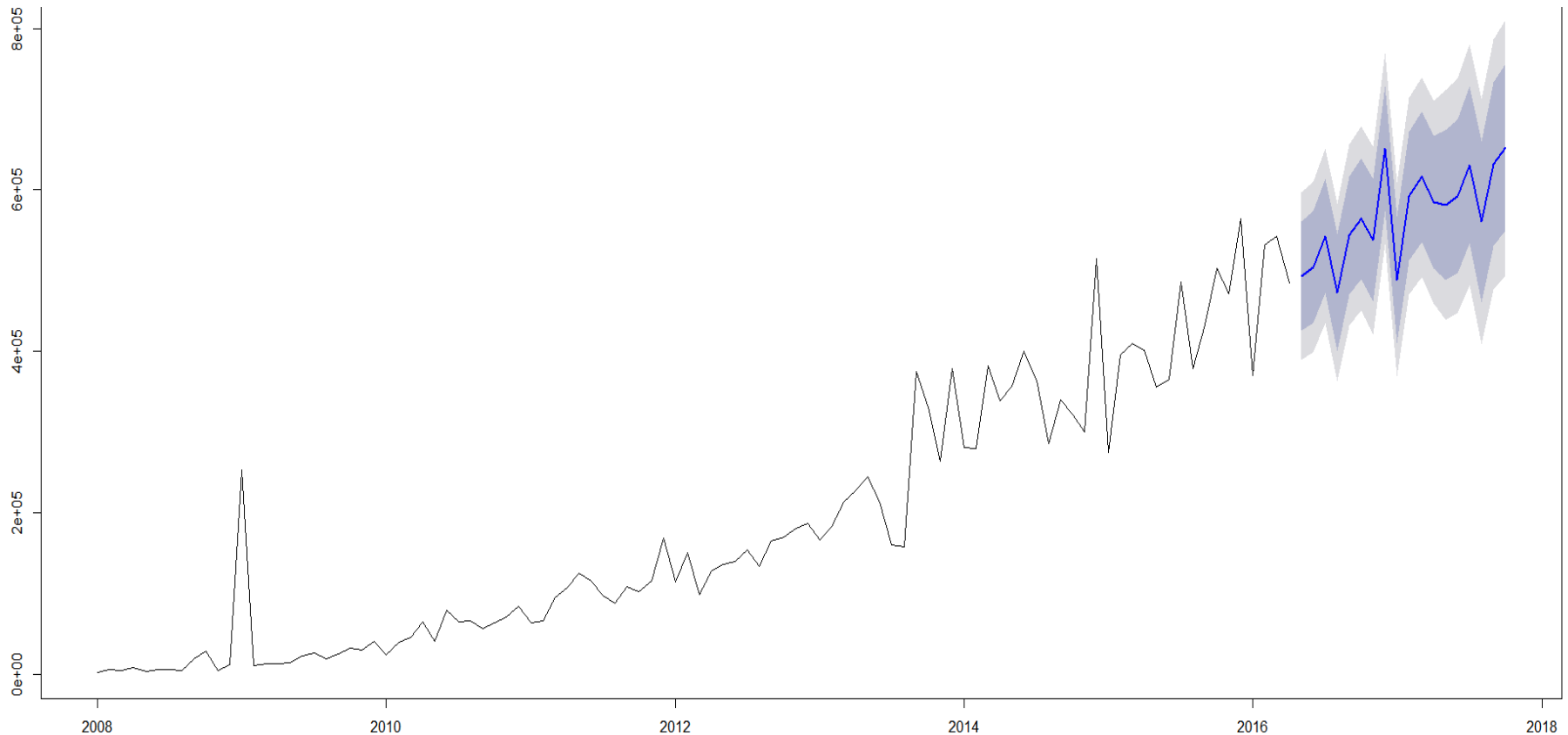
Etapas de análisis
de una serie de
tiempo

Etapas en el análisis de la serie

- El análisis de una serie de tiempo de proceder, sin ser restrictivo, procede a groso modo de la siguiente forma:
 1. Descripción de la serie temporal.
 2. Selección de datos: rango de modelización, rango de estimación, pronóstico.
 3. Estimación de los modelos.
 4. Medidas de rendimiento (set de entrenamiento).
 5. Estimación al set de validación.
 6. Medidas de rendimiento (set de validación).
 7. Selección del mejor método de estimación
 8. Pronóstico.

Etapas en el análisis de la serie

1. Descripción de la serie temporal: visualización de la serie y conocimiento de fenómeno de estudio.



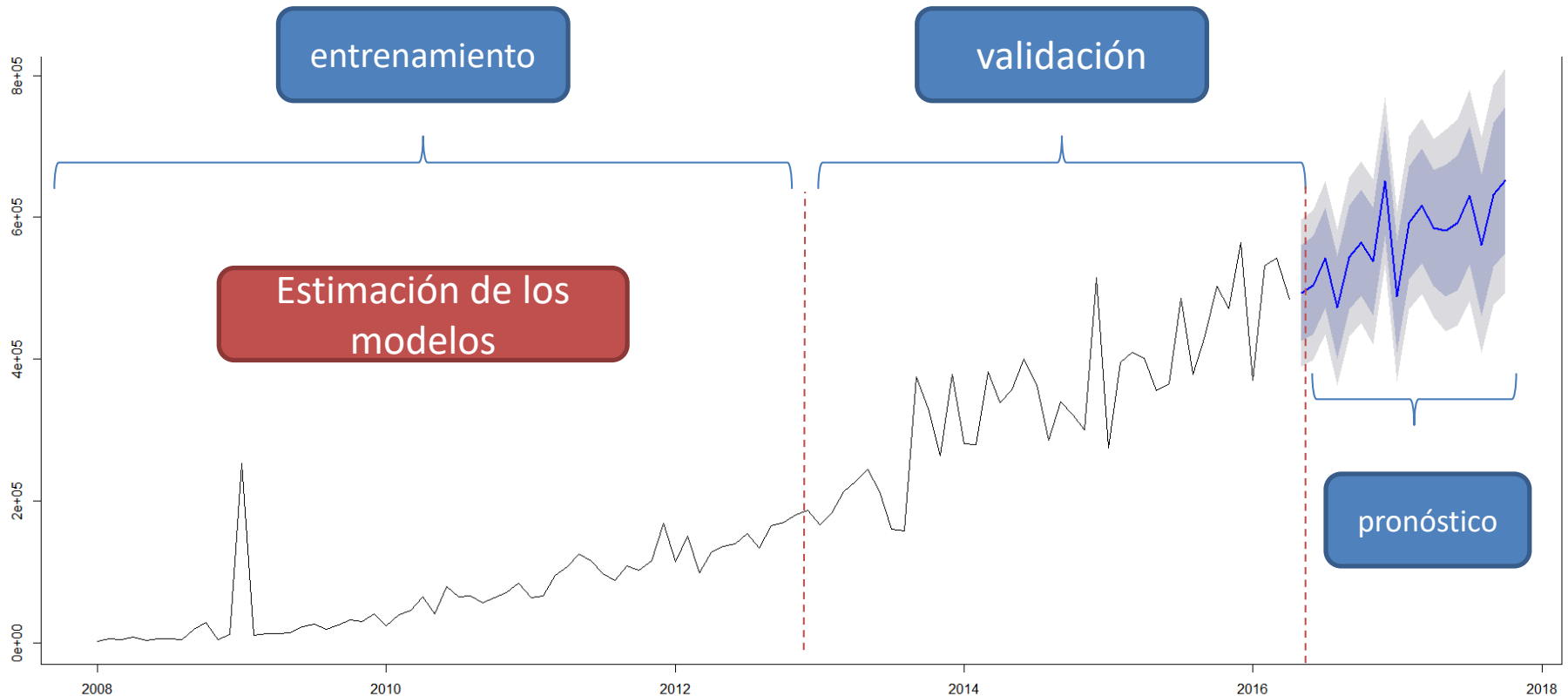
Etapas en el análisis de la serie

2. Selección de datos: rango de modelización (set de entrenamiento), rango de estimación (set de validación), pronóstico (para un horizonte $h > 1$, luego del set de datos).



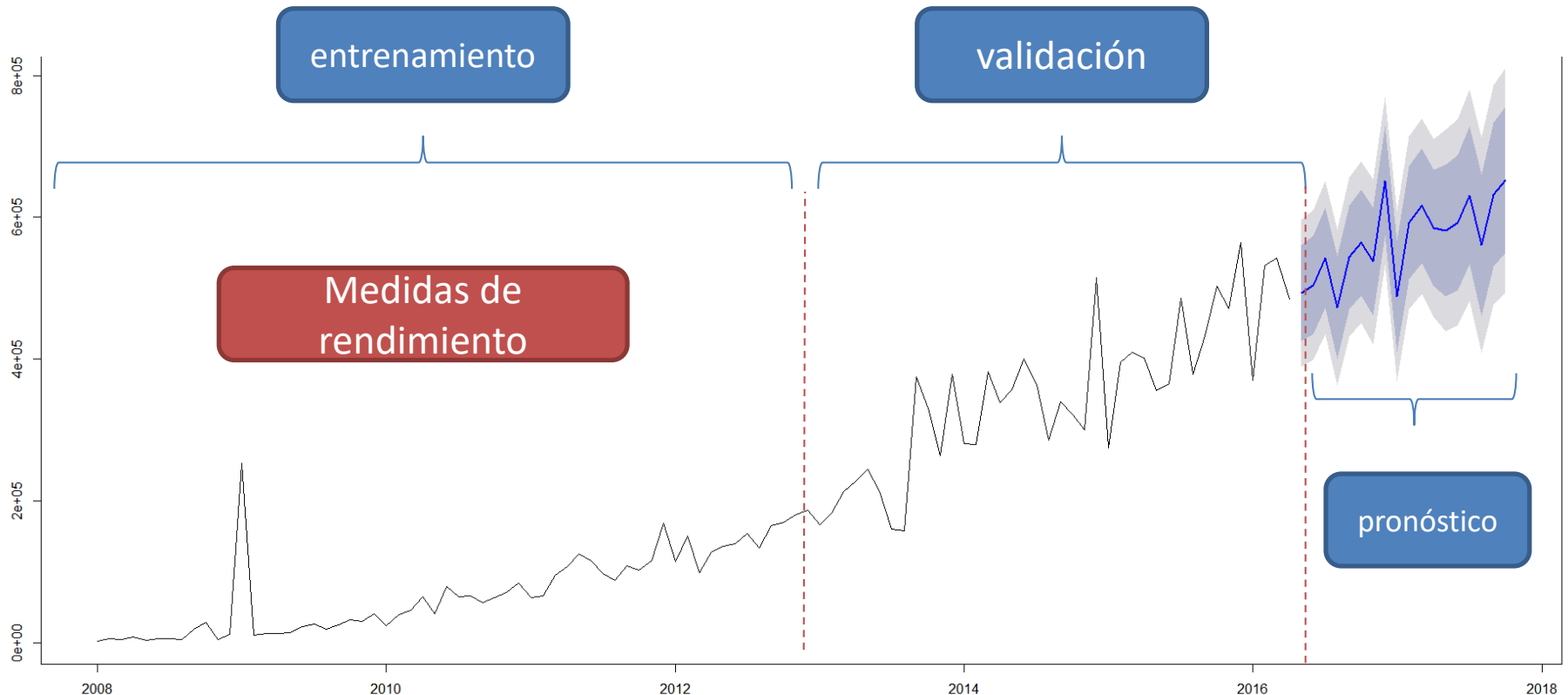
Etapas en el análisis de la serie

3. Estimación de los modelos: trabajando sobre el set de entrenamiento, se estiman el o los modelos.



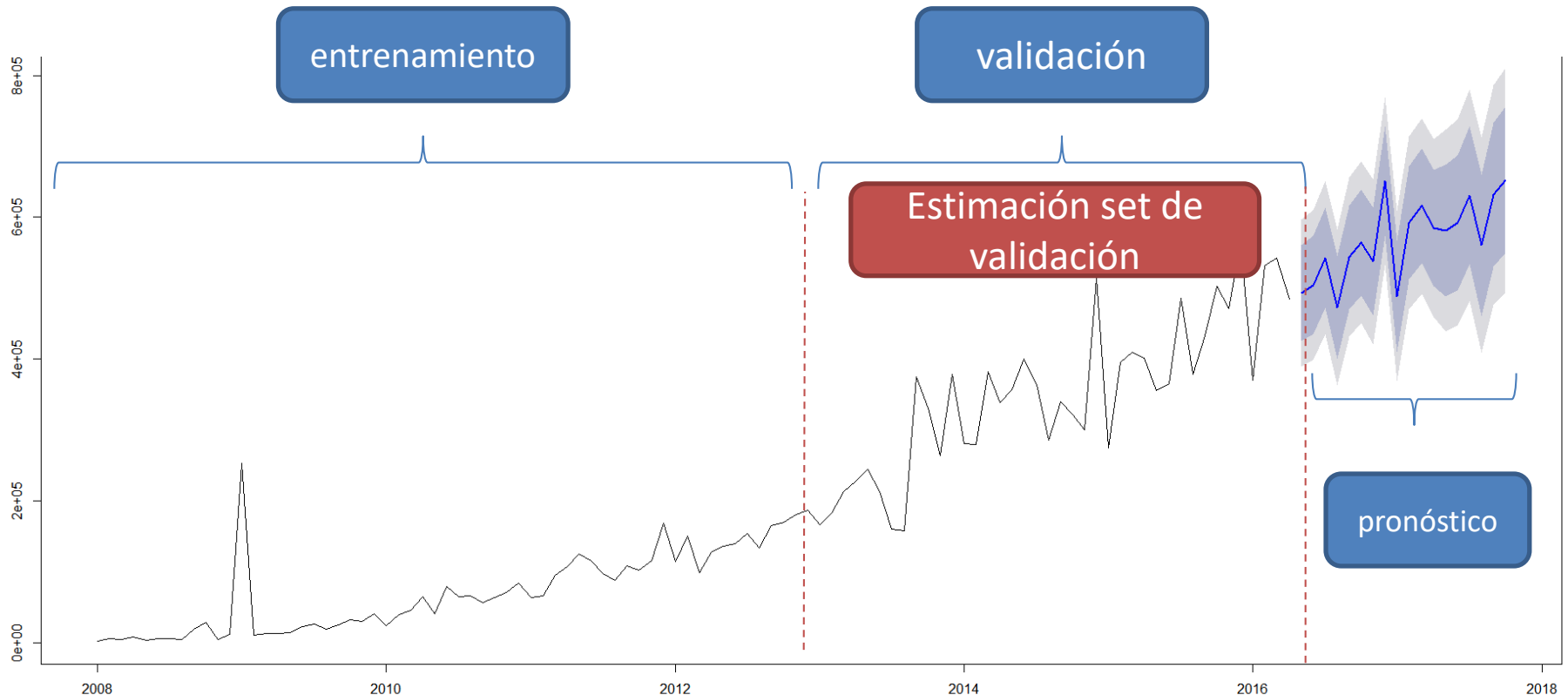
Etapas en el análisis de la serie

4. Medidas de rendimiento (set de entrenamiento): se obtienen las medidas de AIC, AICc ,BIC, RMSE, MAD, etc., para comparar los modelos en el set de entrenamiento.



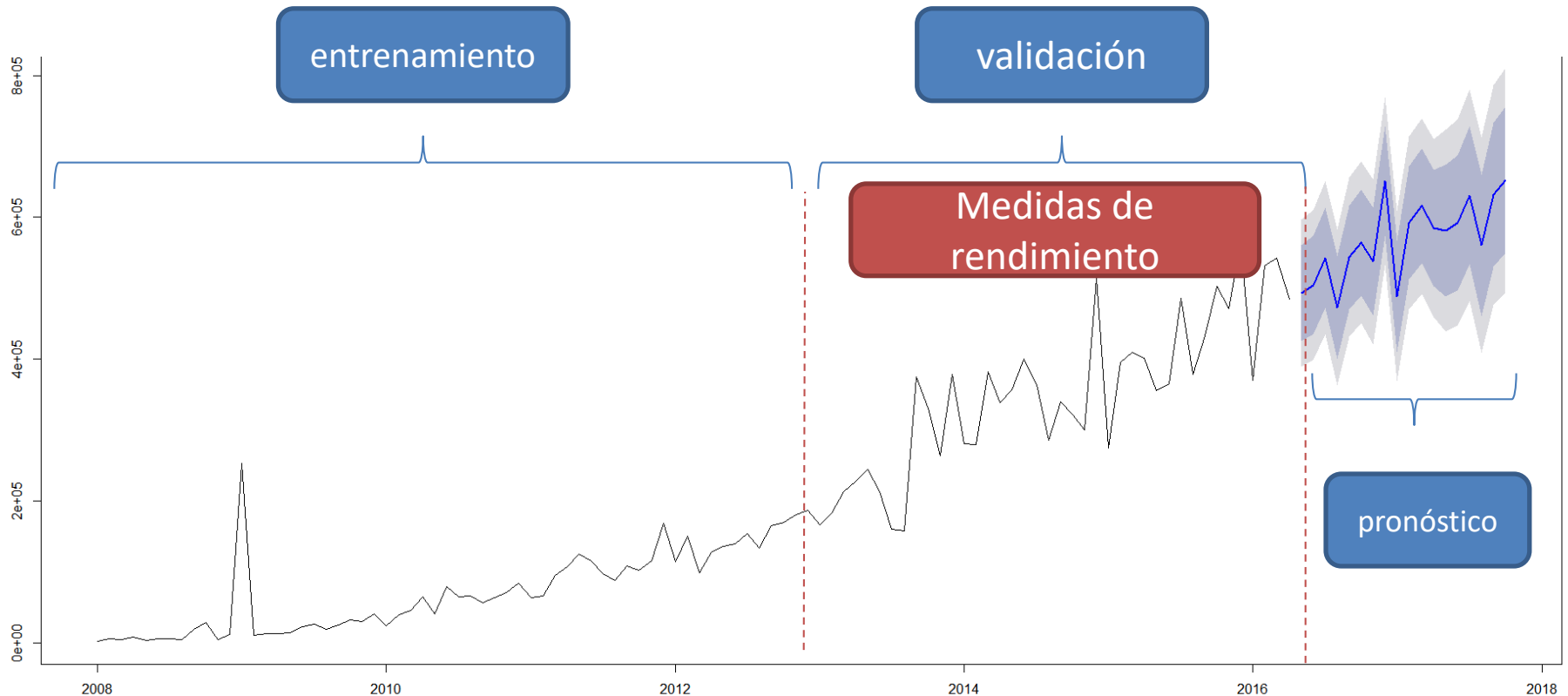
Etapas en el análisis de la serie

5. Estimación al set de validación: los modelos escogidos se utilizan para estimar los valores al set de validación



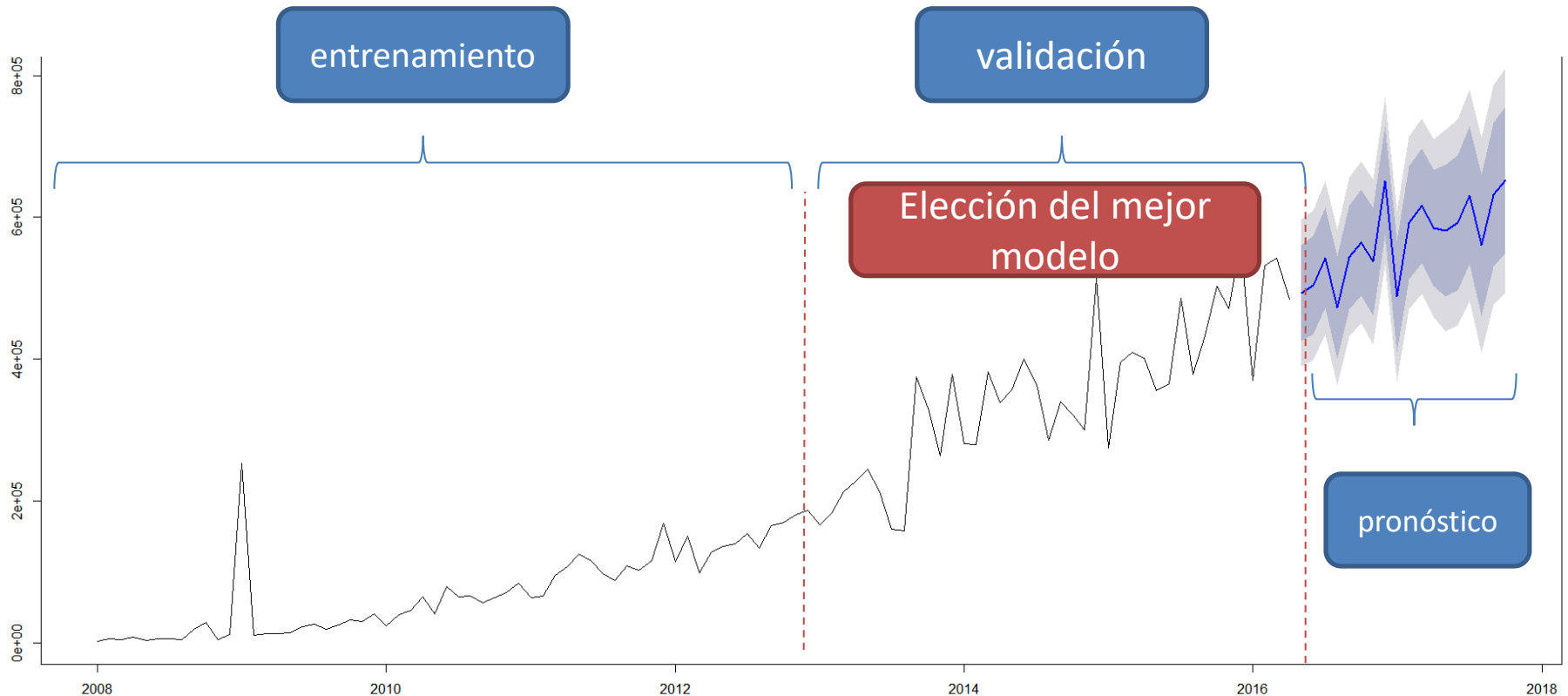
Etapas en el análisis de la serie

6. Medidas de rendimiento (set de validación):



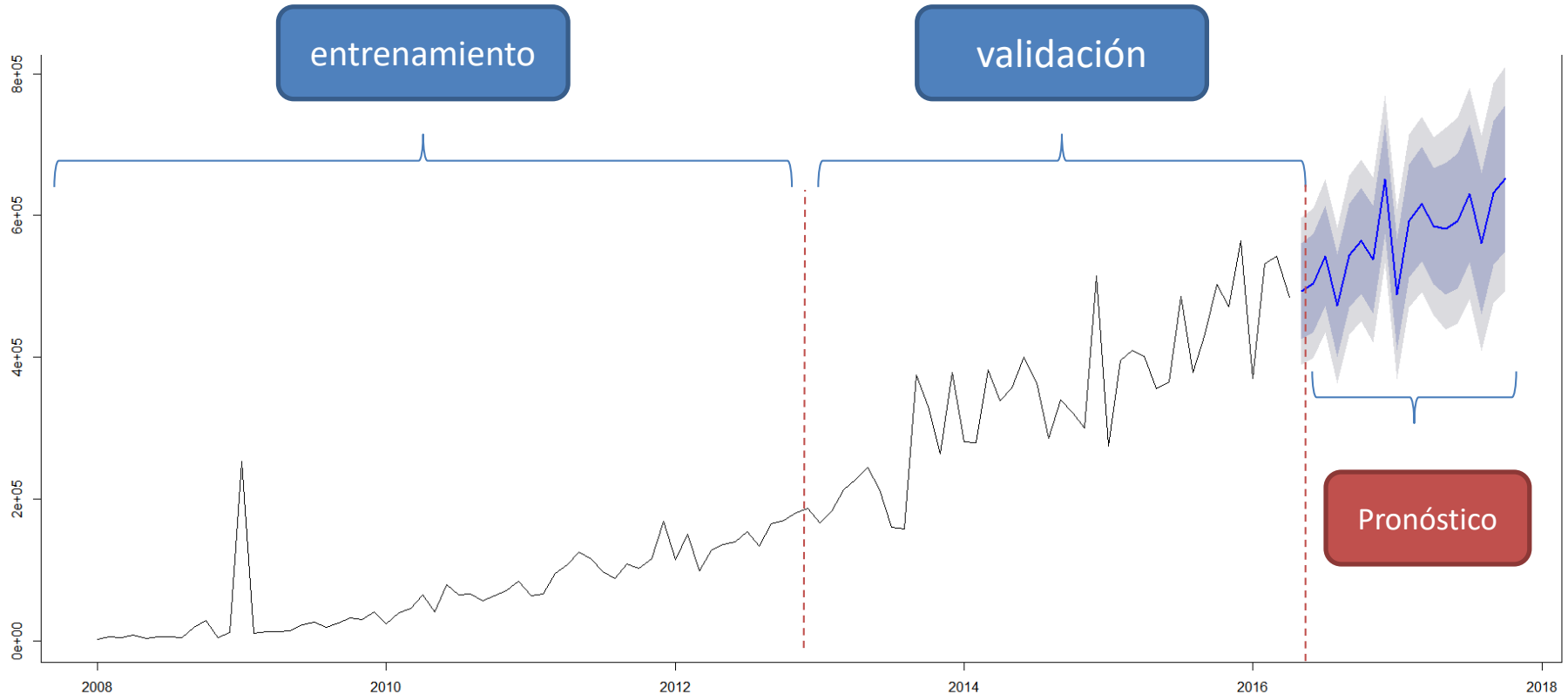
Etapas en el análisis de la serie

7. Selección del mejor método de estimación: según las comparaciones de las medidas de rendimiento tanto para el set de entrenamiento + set de validación, se elige el mejor modelo llevar a cabo el pronóstico.



Etapas en el análisis de la serie

8. Pronóstico: se pronostica para el período $t+1, \dots, t+h$.

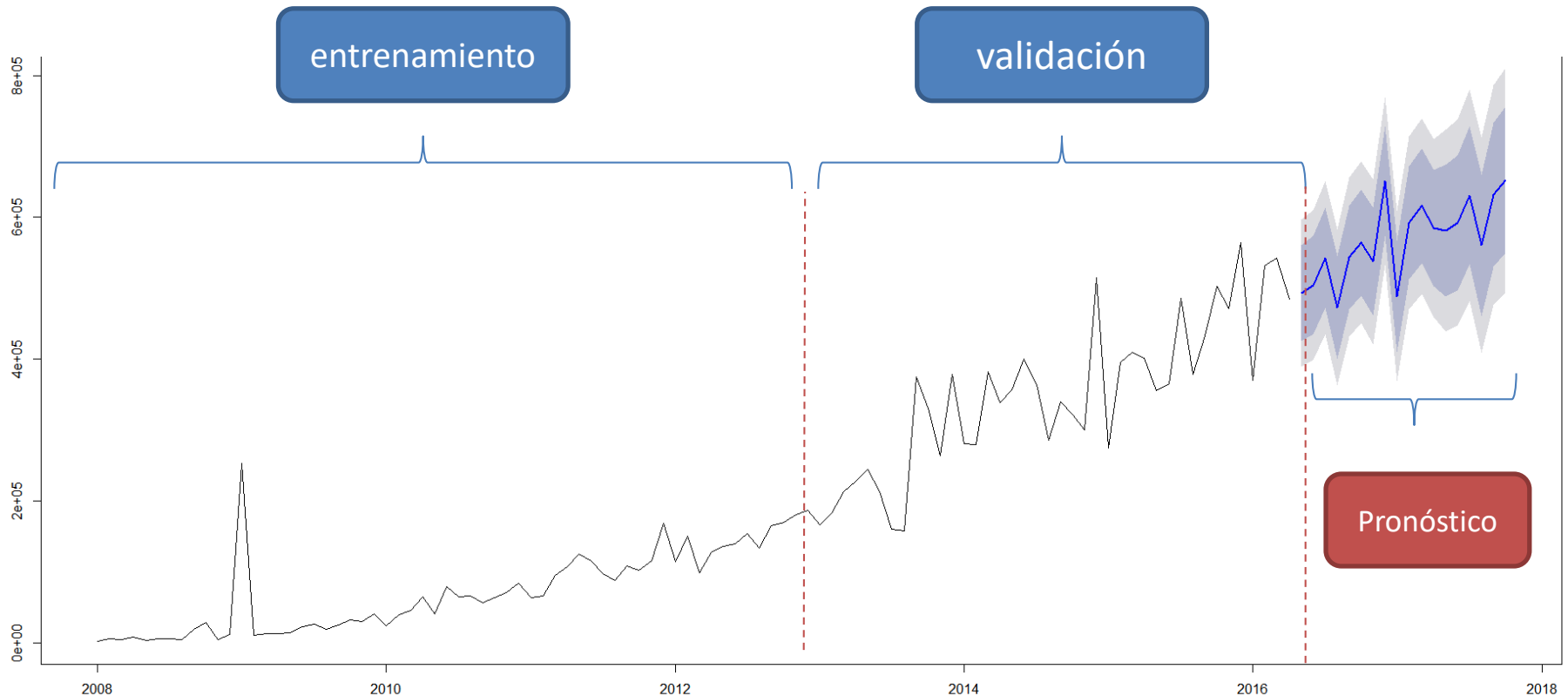


Etapas en el análisis de la serie



Etapas en el análisis de la serie

8. Pronóstico: este se debe de interpretar. La razón de ser del análisis es pronosticar hasta el periodo ***H***. Se debe de explicar o describir el pronóstico, contextualizar y finalmente dar la valoración de lo que sucederá en el futuro.



Índice

7

Longitud de la
serie de tiempo

8

Etapas de análisis
de una serie de
tiempo

9

Medidas de
rendimiento

Medidas de rendimiento

- Las medidas clásicas son ... (y_t son los valores observados y f_t son los valores estimados).

Estadísticos con
escala dependiente

$$\text{MAE} = n^{-1} \sum_{t=1}^n |y_t - f_t|$$

$$\text{MSE} = n^{-1} \sum_{t=1}^n (y_t - f_t)^2$$

$$\text{RMSE} = \sqrt{n^{-1} \sum_{t=1}^n (y_t - f_t)^2}$$

Estadísticos con
escala dependiente

$$\text{MAPE} = 100n^{-1} \sum_{t=1}^n |y_t - f_t| / |y_t|$$

Medidas de rendimiento

- Una medidas bastante clásica bastante utilizada el MASE (Mean Absolute Scaled Error)

$$\text{MASE} = n^{-1} \sum_{t=1}^n |y_t - f_t| / q$$

Valor de q sin
estacionalidad

Valor de q
estacionalidad

$$q = (n - 1)^{-1} \sum_{t=2}^n |y_t - y_{t-1}|$$

$$q = (n - m)^{-1} \sum_{t=m+1}^n |y_t - y_{t-m}|$$

Medidas de rendimiento

- Finalmente, otras medidas de rendimiento son los asociados a los criterios de información:

$$AIC = -2\log L(\hat{\theta}) + 2k$$

$$AIC_c = -2\log L(\hat{\theta}) + 2k + (2k+1) / (n-k-1)$$

$$BIC = -2\log L(\hat{\theta}) + k\log n$$

Medidas de rendimiento

- ¿Cuál es la diferencia principal entre los métodos de rendimiento clásicos y aquellos por el criterio de información?
- ¿Cuál es el criterio para decir que un método clásico de rendimiento es mejor que otro?
- ¿Cuál es el criterio para decir que un método de información de rendimiento es mejor que otro?



¿Dónde encontrar series temporales?

Mi sitio preferido: <https://datamarket.com/data/list/?q=provider%3Atdl>

Series económicas: <http://www.economicswbinstitute.org/ecdata.htm>

Kaggle: <https://www.kaggle.com/datasets>

Banco Central: <https://www.bccr.fi.cr/SitePages/default.aspx>

Inec: <http://www.inec.go.cr/>

Sin embargo, deben de conocer el contexto o las características de las series....

THE END