

# Filtros Lineales y Regresión Rezagada

Felipe Gallo Aristizábal  
Universidad de Antioquia  
Facultad de Ciencias Exactas y Naturales  
Pregrado de Estadística  
felipe.galloa@udea.edu.co

1 de noviembre de 2018

---

## Resumen

*En el presente artículo, se presentara el uso de filtros lineales aplicados a series de tiempo, cuya finalidad es extraer algún tipo de señal de interés de la serie. Existen una gran variedad de filtros, los cuales se usan de acuerdo a qué es lo que queremos lograr con estos, es decir, que frecuencias o señales de la serie se esta interesado en analizar, y en los cuales podemos identificar tres grandes tipos de filtros: Pasa baja, Pasa alta y Pasa Banda, de los cuales se hará una explicación a lo largo del documento. También, se presenta el uso de la regresión rezagada como una posibilidad de establecer una relación entre dos series ante una correlación alta entre estas.*

Palabras clave: Filtros lineales, Regresión rezagada.

---

## 1. Introducción

El objetivo principal de este documento es explicar un poco acerca de los filtros lineales y como se implementan para el procesamiento de señales con el fin de estudiar más a fondo las características de estas. Para ser más concretos, se hablará de dos filtros en específico y se ilustraran sus usos en un ejemplo con la base de datos SOI disponible en el software R. También se tratará de explicar cual es la idea con la rgresión rezagada al relacionar dos series de tiempo dando algunos fundamentos teoricos que posteriormente se aplicaran a las bases de datos SOI y Recruitment también con el software R

## 2. Objetivo

Conocer la aplicabilidad de los filtros lineales a la hora de procesar señales y obtener características deseables en estas. A su vez, conocer el uso de la regresión rezagada para poder establecer una relación entre dos series de tiempo dado que se presenta una alta correlación entre estas.

## 3. Filtros Lineales

En el procesamiento de señales, un filtro es un dispositivo o proceso que elimina algunos componentes o características no deseados de una señal. El filtrado es una clase de procesamiento de señales y la característica definitoria de los filtros es la supresión total o parcial de algún aspecto de la señal. Muy a menudo, esto significa eliminar algunas frecuencias o bandas de frecuencia. Los filtros son ampliamente utilizados en electrónica y telecomunicaciones, en radio, televisión, grabación de audio, radar, sistemas de control, síntesis de música, procesamiento de imágenes y gráficos de computadora.

Ahora, un caso particular de los filtros son los filtros lineales, los cuales procesan señales de entrada que varían en el tiempo para producir señales de salida, sujetas a la restricción de la linealidad, es decir, se aplica un operador lineal a la señal que se encuentra en estudio con el fin de eliminar frecuencias no deseadas.

Supongamos entonces que  $x_t$  es un proceso estacionario con una función de autocovarianza absolutamente sumable  $\gamma_x(h)$  y correspondiente espectro

$$f_x(\omega) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma_x(h) e^{-i\omega h} \quad (1)$$

considerando el siguiente filtro lineal de  $x_t$  dado por

$$\begin{aligned} y_t &= \sum_{j=-\infty}^{\infty} a_j x_{t-j} \\ &= \sum_{j=-\infty}^{\infty} a_j B^j x_t \\ &= a(B) x_t \end{aligned} \quad (2)$$

donde

$$a(B) = \sum_{j=-\infty}^{\infty} a_j B^j \quad y \quad \sum_{j=-\infty}^{\infty} |a_j| < \infty$$

Tenemos que

$$\begin{aligned}
\gamma_{xy}(h) &= \text{cov}(x_{t+h}, y_t) \\
&= \text{cov} \left( \sum_k a_k x_{t+h-k}, \sum_j a_j x_{t-j} \right) \\
&= \sum_k \sum_j a_k \gamma_{xx}(h-k+j) a_j \\
&= \sum_k \sum_j a_k \left[ \int_{-1/2}^{1/2} e^{2\pi i \omega (h-k+j)} f_{xx}(\omega) d\omega \right] a_j \\
&= \int_{-1/2}^{1/2} \left( \sum_k a_k e^{-2\pi i \omega k} \right) \left( \sum_j a_j e^{2\pi i \omega j} \right) e^{2\pi i \omega h} f_{xx}(\omega) d\omega \\
&= \int_{-1/2}^{1/2} e^{2\pi i \omega h} |A_{xy}(\omega)|^2 f_{xx}(\omega) d\omega
\end{aligned} \tag{3}$$

donde sabemos que por la relación biunívoca entre la función de autocovarianza y la densidad espectral, el espectro de la serie de salida esta dado por

$$f_{yy}(\omega) = |A_{xy}(\omega)|^2 f_{xx}(\omega) \tag{4}$$

donde  $|A_{xy}(\omega)|$  se conoce como la función de filtro o la función de transferencia. Esta función de filtro se puede utilizar para medir el efecto de aplicar un filtro en una serie. Como ya se mencionó anteriormente, en muchos estudios, un objetivo importante es diseñar un buen filtro para que la serie o señales de salida satisfagan ciertas propiedades deseadas. Así, hay varios tipos de filtros lineales en lo que respecta a su respuesta en frecuencia::

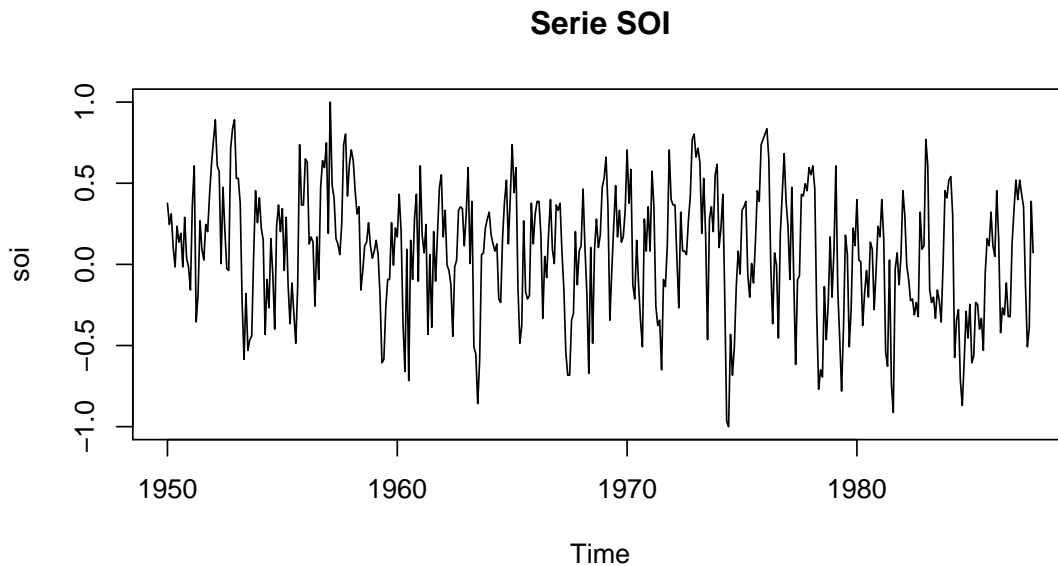
- *Filtro pasa bajo* es un filtro que pasa señales con una frecuencia más baja que una frecuencia de corte seleccionada y atenúa las señales con frecuencias más altas que la frecuencia de corte. La respuesta de frecuencia exacta del filtro depende del diseño del filtro.
- *Filtro pasa alto* es un filtro que pasa señales con una frecuencia más alta que una determinada frecuencia de corte y atenúa las señales con frecuencias más bajas que la frecuencia de corte. La cantidad de atenuación para cada frecuencia depende del diseño del filtro.
- *Filtro pasa banda* es un filtro que pasa señales con frecuencias dentro de un cierto rango y atenúa las que estan fuera de éste.

- *Filtro banda de parada* es un filtro que atenúa las señales con frecuencias dentro de un cierto rango.

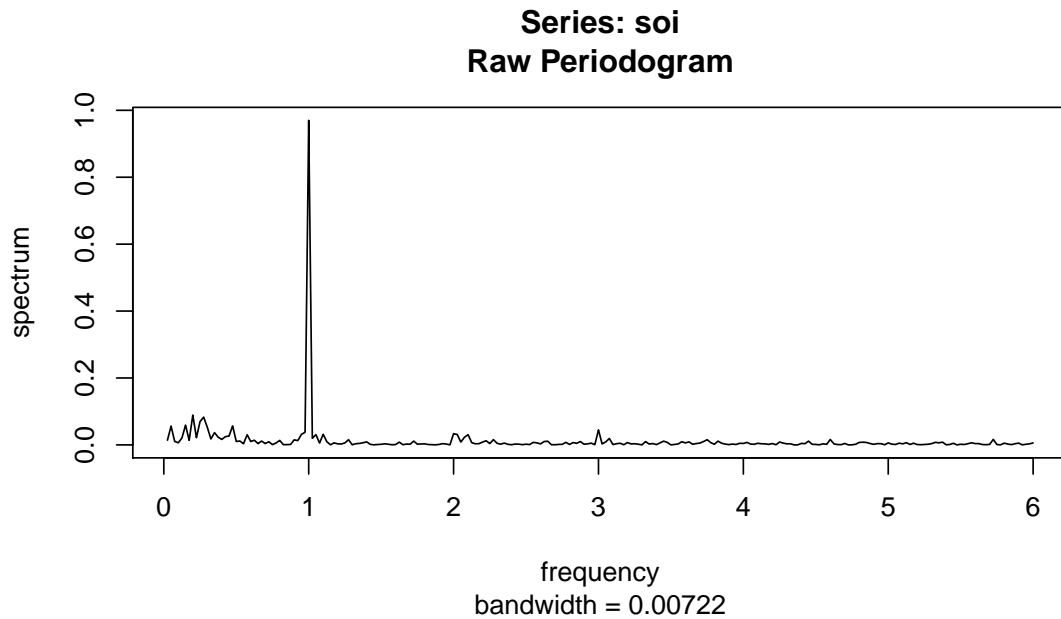
Si bien el tipo de filtros lineales es amplio, en el presente documento solo se tratara los dos primeros tipos de filtros mencionados ya que los demás surgen como una combinación de estos.

### 3.1. Filtro pasa baja - Medias Móviles

Consideremos los datos de SOI, que es el índice que mide la Oscilación del Sur al correlacionar valores de presión atmosférica obtenidos en el Pacífico occidental con los del Pacífico central. Este es medido por un período de 453 meses que abarca los años 1950-1987. Esta serie, se comporta de la siguiente manera



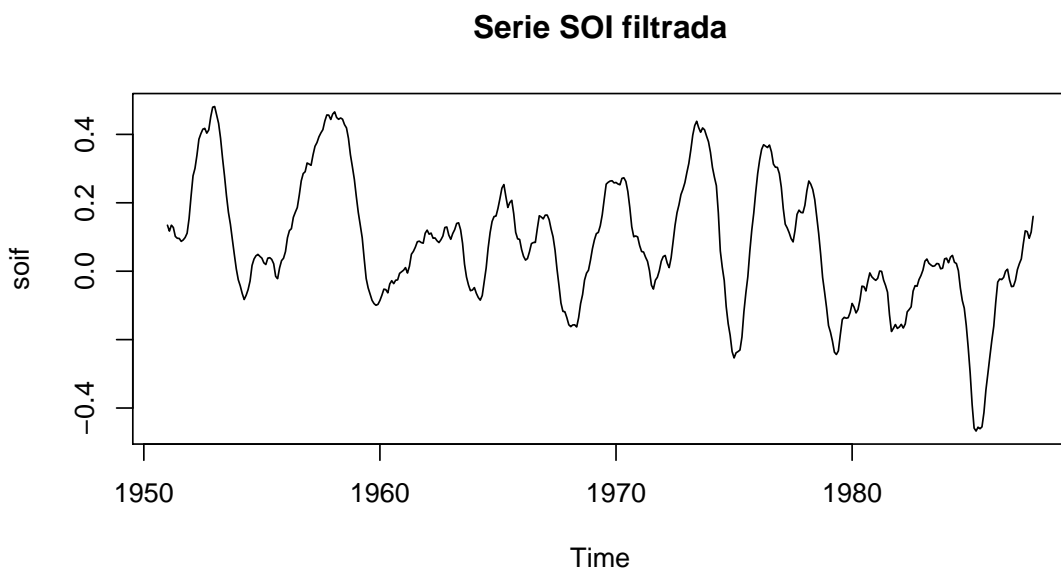
El espectro es de la siguiente forma



Como podemos ver hay una sola frecuencia significativa que aporta a la varianza del proceso y el período de esta frecuencia es de un año. Ahora, consideremos el filtro de medias móviles simétrico anual, de la siguiente forma:

$$y_t = \frac{1}{24}(x_{t-6} + x_{t+6}) + \frac{1}{12} \sum_{r=-5}^5 x_{t-r}$$

el cual se conoce como un suavizador de Daniell modificado con  $m = 6$ . Al aplicar este filtro, la serie queda de la siguiente manera

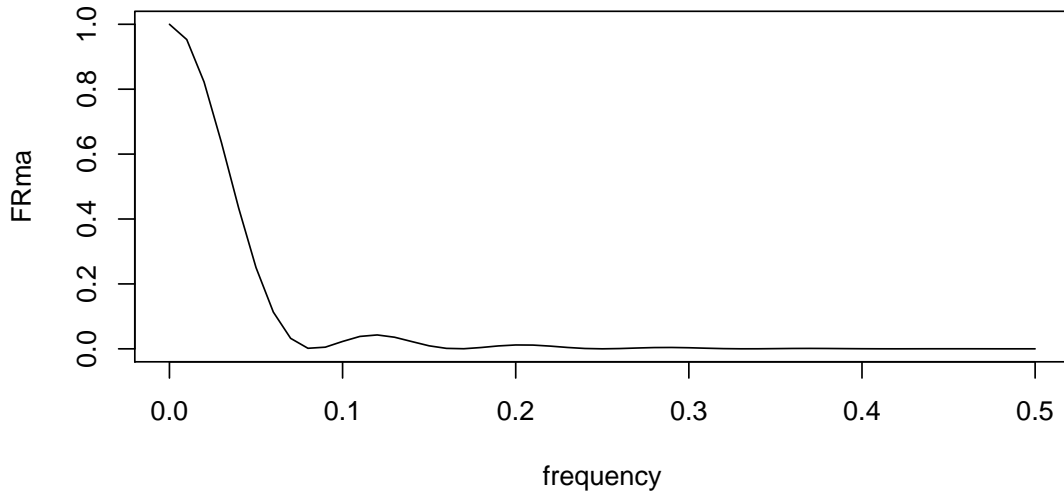


Como se ve, el filtro ha suavizado la serie porque retiene las frecuencias más bajas

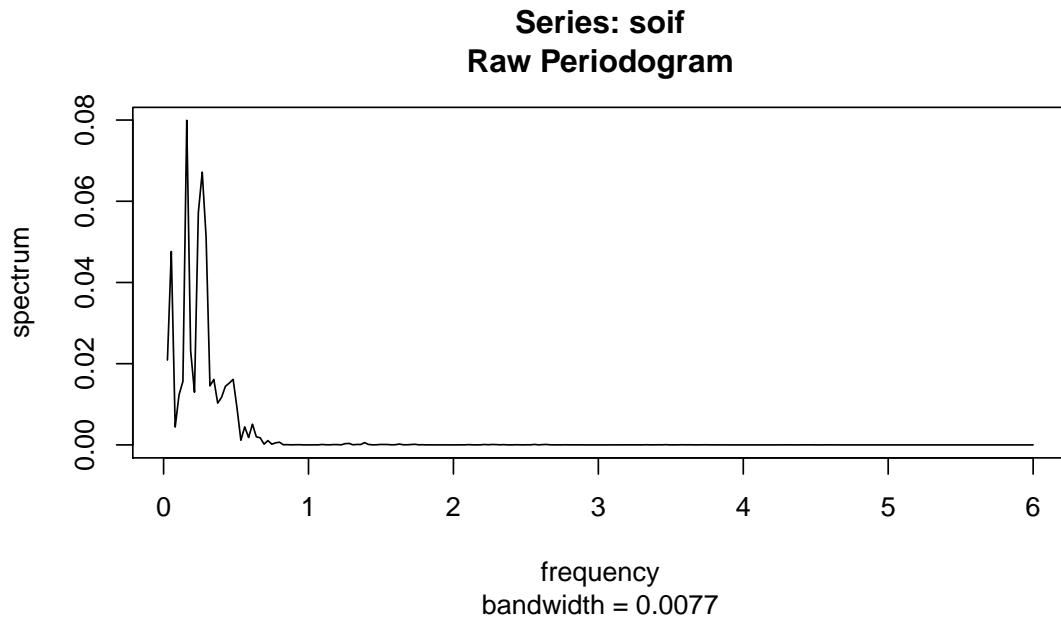
y tiende a atenuar las frecuencias más altas, esto debido a que el multiplicador del espectro o respuesta de la frecuencia al cuadrado,  $|A_{xy}(\omega)|^2$ , es grande para frecuencias más bajas y pequeño para las frecuencias más altas. Esta respuesta de la frecuencia esta dada de la siguiente forma

$$A_{xy}(\omega) = \frac{1}{12} \left[ 1 + \cos(12\pi\omega) + 2 \sum_{k=1}^5 \cos(2\pi\omega k) \right]$$

De manera gráfica tenemos que la respuesta de frecuencia cuadrada es



Esto nos muestra que podemos esperar que este filtro corte la mayor parte del contenido de frecuencia por encima de  $1/12 = 0.083$ . En particular, esto reduce las componentes anuales con períodos de 12 meses y mejora la frecuencia de el niño, que es algo menor. Así, miremos que sucedio con el espectro



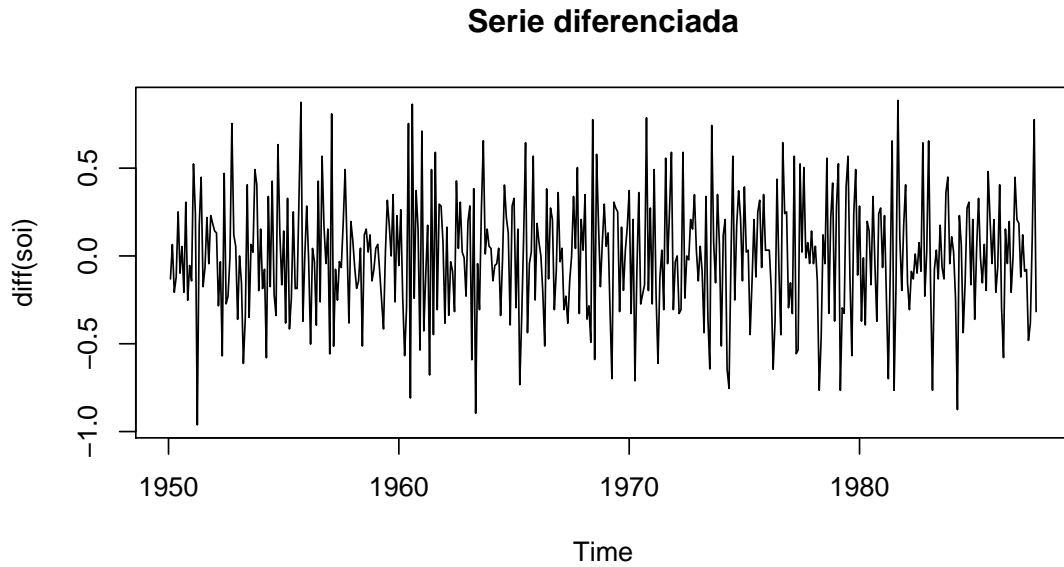
Se puede ver claramente, que las frecuencias más significativas son las más bajas, donde aparecen frecuencias como la del niño.

### 3.2. Filtro pasa alta - Primera diferencia

Consideremos la misma serie que la usada para el filtro pasa bajo. Así, aplicando el filtro de la primera diferencia, dado por

$$y_t = (1 - B)x_t$$

obtenemos que la serie queda de la siguiente forma



Como se ve, el filtro no suavizado la serie porque retiene las frecuencias más altas y tiende a atenuar las frecuencias las bajas, esto debido a que el multiplicador del espectro o respuesta de la frecuencia al cuadrado,  $|A_{xy}(\omega)|^2$ , es grande para frecuencias más altas y pequeño para las frecuencias más bajas. Esta respuesta de la frecuencia esta dada de la siguiente forma

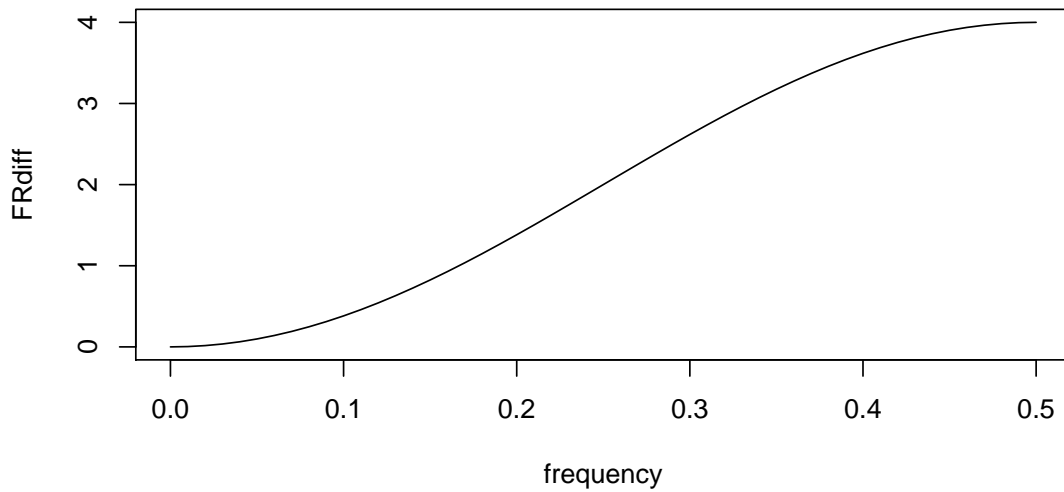
$$A_{xy}(\omega) = 1 - e^{-2\pi i \omega}$$

y la respuesta de la frecuencia cuadrada se convierte

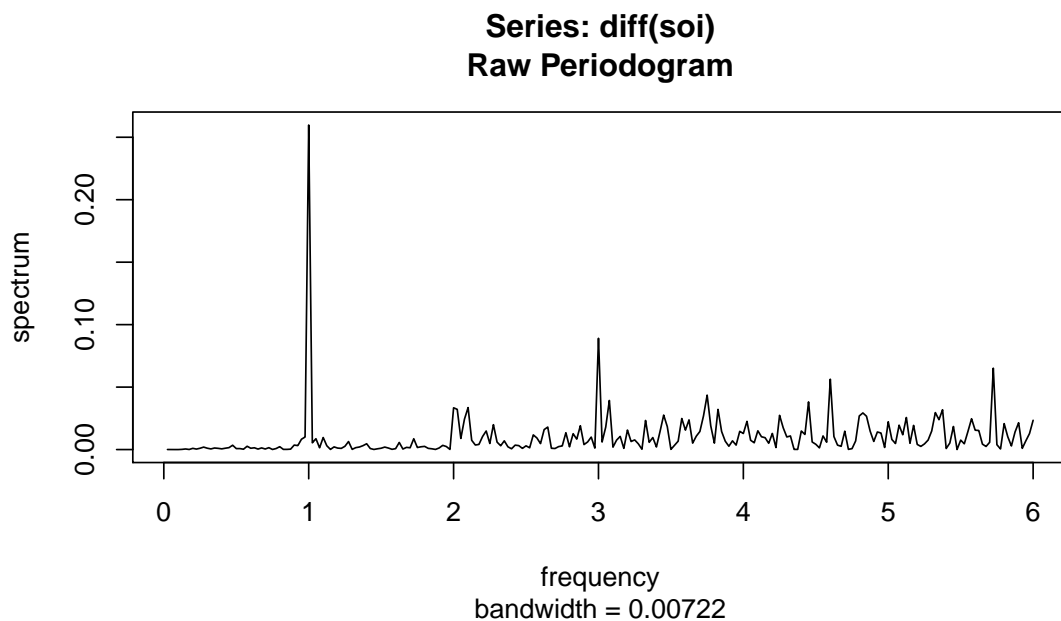
$$\begin{aligned} |A_{xy}(\omega)|^2 &= (1 - e^{-2\pi i \omega})(1 - e^{2\pi i \omega}) \\ &= 2[1 - \cos(2\pi \omega)] \end{aligned}$$

De manera gráfica tenemos que la respuesta de frecuencia cuadrada es





Lo cual nos indica que el filtro le da más peso a las frecuencias más altas y esto se vera reflejado en el espectro del proceso, donde identificaremos facilmente que las frecuencias altas seran las que mas aporten a la varianza del proceso. Así, observemos que sucedio con el espectro



Se nota claramente que las frecuencias altas empezaron a ser relevantes en el espectro y aportan más a la varianza del proceso, algo que es obvio.

## 4. Regresión Rezagada

En estadística y econometría, un modelo de regresión rezagado es un modelo para datos de series de tiempo en el que se utiliza una ecuación de regresión para predecir los valores actuales de una variable dependiente en base a los valores actuales de una variable explicativa y los valores rezagados de esta variable explicativa. Entonces, la idea es extender la regresión clásica al análisis de modelos de regresión rezagados de la forma

$$y_t = \sum_{r=-\infty}^{\infty} \beta_r x_{t-r} + v_t \quad (5)$$

donde  $v_t$  es un proceso de ruido estacionario,  $x_t$  es la serie de entrada observada, y  $y_t$  es la serie de salida observada. Estamos interesados en estimar los coeficientes filtro  $\beta_r$  relacionando los valores rezagados adyacentes de  $x_t$  con la serie de salida  $y_t$ . En general, habrá más de una serie de entrada posible y podremos imaginar un vector  $q \times 1$  de series de conducción. El modelo (5) es útil bajo varios escenarios diferentes, que corresponden a diferentes suposiciones que pueden hacerse sobre los componentes.

Supongamos que las series tanto de entrada como de salida, tienen media cero y estacionarios conjuntamente con el proceso vectorial  $2 \times 1$   $(x_t, y_t)'$  que tienen una matriz espectral de la forma

$$f(\omega) = \begin{pmatrix} f_{xx}(\omega) & f_{xy}(\omega) \\ f_{yx}(\omega) & f_{yy}(\omega) \end{pmatrix} \quad (6)$$

Aquí,  $f_{xy}(\omega)$  es el espectro cruzado que relaciona la entrada  $x_t$  con la salida  $y_t$ , y  $f_{xx}(\omega)$  y  $f_{yy}(\omega)$  son los espectros de entrada y salida, respectivamente y están dados por

$$\begin{aligned} f_{xy}(\omega) &= \sum_{h=-\infty}^{\infty} \gamma_{xy}(h) e^{-2\pi i \omega h} \\ f_{xx}(\omega) &= \sum_{h=-\infty}^{\infty} \gamma_{xx}(h) e^{-2\pi i \omega h} \\ f_{yy}(\omega) &= \sum_{h=-\infty}^{\infty} \gamma_{yy}(h) e^{-2\pi i \omega h} \end{aligned}$$

En general, observamos dos series, consideradas como entrada y salida y buscamos funciones de regresión  $\{\beta_t\}$  que relacionen las entradas con las salidas. Supongamos que todas las funciones de autocovarianza satisfacen las condiciones de similitud

absoluta de la forma

$$\sum_{h=-\infty}^{\infty} |h| |\gamma(h)| < \infty \quad (7)$$

Entonces, minimizando el error cuadrático medio

$$MSE = E \left( y_t - \sum_{r=-\infty}^{\infty} \beta_r x_{t-r} \right)^2 \quad (8)$$

conduce a las condiciones de ortogonalidad habituales

$$E \left[ \left( y_t - \sum_{r=-\infty}^{\infty} \beta_r x_{t-r} \right) x_{t-s} \right] = 0 \quad (9)$$

para todo  $s = 0, \pm 1, \pm 2, \dots$ . Tomando esperanzas dentro conduce a las ecuaciones normales

$$\sum_{r=-\infty}^{\infty} \beta_r \gamma_{xx}(s-r) = \gamma_{yx}(s) \quad (10)$$

para  $s = 0, \pm 1, \pm 2, \dots$

Una solución aproximada de dominio de frecuencia es más fácil en este caso por que los cálculos solo dependerán de espectros y espectros cruzados que pueden estimarse a partir de datos de muestra y además no tendrá que invertirse ninguna matriz. Ahora, la parte izquierda de (10) pueden escribirse en la forma

$$\int_{-1/2}^{1/2} \sum_{r=-\infty}^{\infty} \beta_r e^{2\pi i \omega (s-r)} f_{xx}(\omega) d\omega = \int_{-1/2}^{1/2} e^{2\pi i \omega s} \beta(\omega) f_{xx}(\omega) d\omega$$

donde

$$\beta(\omega) = \sum_{r=-\infty}^{\infty} e^{-2\pi i \omega r} \quad (11)$$

es la transformada de Fourier de los coeficientes de regresión  $\beta_t$ . Ahora, porque  $\gamma_{yx}(s)$  es la transformada inversa del espectro cruzado  $f_{yx}(\omega)$  podríamos escribir el sistema de ecuaciones en el dominio de la frecuencia, usando la singularidad de la transformada de Fourier como

$$\beta(\omega) f_{xx}(\omega) = f_{yx}(\omega) \quad (12)$$

que luego se convierten en los análogos de las ecuaciones normales usuales. Entonces podemos tomar

$$\hat{\beta}(\omega_k) = \frac{\hat{f}_{yx}(\omega_k)}{\hat{f}_{xx}(\omega_k)} \quad (13)$$

que luego el estimador de la transformada de Fourier de los coeficientes de regresión evaluada en algún subconjunto de frecuencias fundamentales  $\omega_k = k/M$  con  $M < n$ . Generalmente, asumimos la suavidad  $\beta(\cdot)$  en intervalos de la forma  $\{\omega_k + l/n; l = -m, \dots, 0, \dots, m\}$ , con  $L = 2m + 1$ . La transformada inversa de la función  $\hat{\beta}(\omega)$  daría  $\hat{\beta}_t$  y observamos que la aproximación de tiempo discreto se puede tomar como

$$\hat{\beta}_t = M^{-1} \sum_{k=0}^{M-1} \hat{\beta}(\omega_k) e^{2\pi i \omega_k t} \quad (14)$$

para  $t = 0, \pm 1, \pm 2, \dots, (M/2 - 1)$ . Si usáramos (14) para definir  $\hat{\beta}_t$  para  $|t| \geq M/2$ , terminaríamos con una secuencia de coeficientes que es periódica con un período de  $M$ . En práctica, fijamos  $\hat{\beta}_t = 0$  para  $|t| \geq M/2$  en su lugar.

#### 4.1. Ragresión Rezagada - Datos SOI y Recruitment

Para ilustrar el uso de la regresión rezagada se considera la serie SOI (Southern Oscillation Index Data) Índice de Oscilación del Sur la cual contiene la diferencia en la presión barométrica al nivel de mar entre Tahití y Darwin. Se dan datos mensuales por un período de 453 meses que abarca los años 1950-1987. La segunda serie considerada comprende los datos de Recruitment (número de peces nuevos) por un período de 453 meses que abarca los años 1950-1987. Entre estas dos series hay una alta coherencia, lo cual sugiere una relación de regresión rezagada entre las dos series. Una dirección natural para la implicación en esta situación está implícita porque creemos que la temperatura de la superficie del mar o SOI debería ser la entrada y la serie Recruitment debería ser la salida. Con esto en mente, dejemos  $x_t$  ser la serie SOI y  $y_t$  la serie Recruitment.

Aunque pensamos naturalmente en SOI como la entrada y Recruitment como el resultado, dos configuraciones de entrada-salida son de interés con SOI como entrada, el modelo es

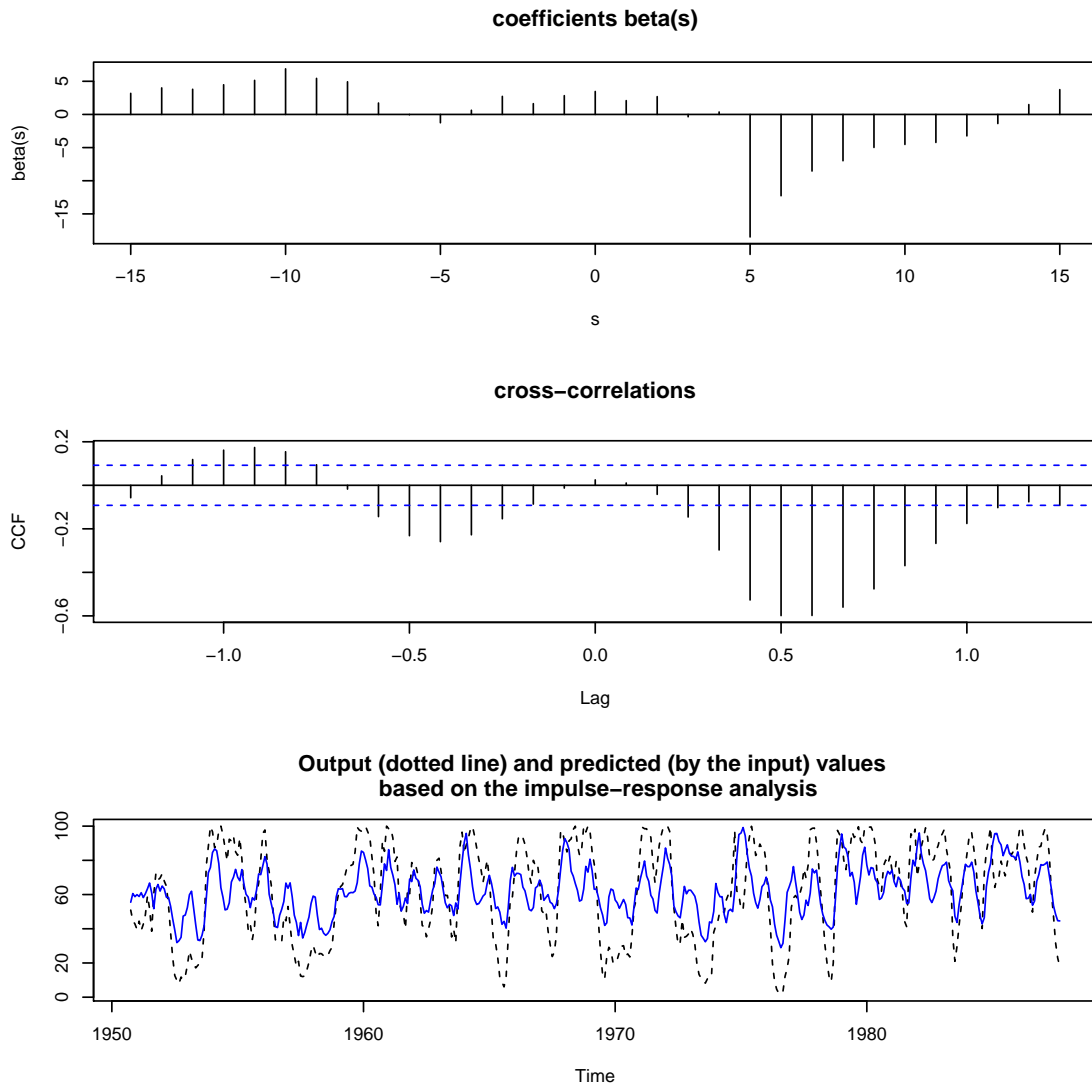
$$y_t = \sum_{r=-\infty}^{\infty} a_r x_{t-r} + w_t$$

$$x_t = \sum_{r=-\infty}^{\infty} b_r y_{t-r} + v_t$$

donde  $w_t$  y  $v_t$  son procesos ruido blanco. Aunque no existe una explicación ambiental plausible para el segundo de estos dos modelos, mostrar ambas posibilidades ayuda a establecer un modelo de función de transferencia parcimonioso.

Basados en el guion LagRec en *astsa*, la regresión estimada o la función de respuesta de impulso para SOI, con  $M = 32$  y  $L = 15$  es

```
## INPUT: soi OUTPUT: rec    L = 15    M = 32
##
## The coefficients beta(0), beta(1), beta(2) ... beta(M/2-1) are
##
## 3.463141 2.088613 2.688139 -0.3515829 0.3717705 -18.47931 -12.2633
## -8.539368 -6.984553 -4.978238 -4.526358 -4.223713 -3.239262 -1.372815
## 1.489903 3.744727
##
##
## The coefficients beta(0), beta(-1), beta(-2) ... beta(-M/2+1) are
##
## 3.463141 2.835444 1.628129 2.726815 0.6330645 -1.256092 -0.05458373
## 1.722774 4.925481 5.440352 6.877381 5.141606 4.479202 3.796848 4.004762
## 3.184184
## The positive lags, at which the coefficients are large
## in absolute value, and the coefficients themselves, are:
##      lag s      beta(s)
## [1,]      5 -18.479306
## [2,]      6 -12.263296
## [3,]      7  -8.539368
## [4,]      8  -6.984553
```



```
##
## The prediction equation is
##  $\text{rec}(t) = \alpha + \sum_s [\text{beta}(s) \cdot \text{soi}(t-s)]$ , where  $\alpha = 65.96584$ 
## MSE = 414.0847
```

Tengamos en cuenta el pico negativo en un retraso de 5 pasos en la parte superior de la figura; en este caso, SOI es la serie de entrada. La caída después del retraso parece ser aproximadamente exponencial y un posible modelo es

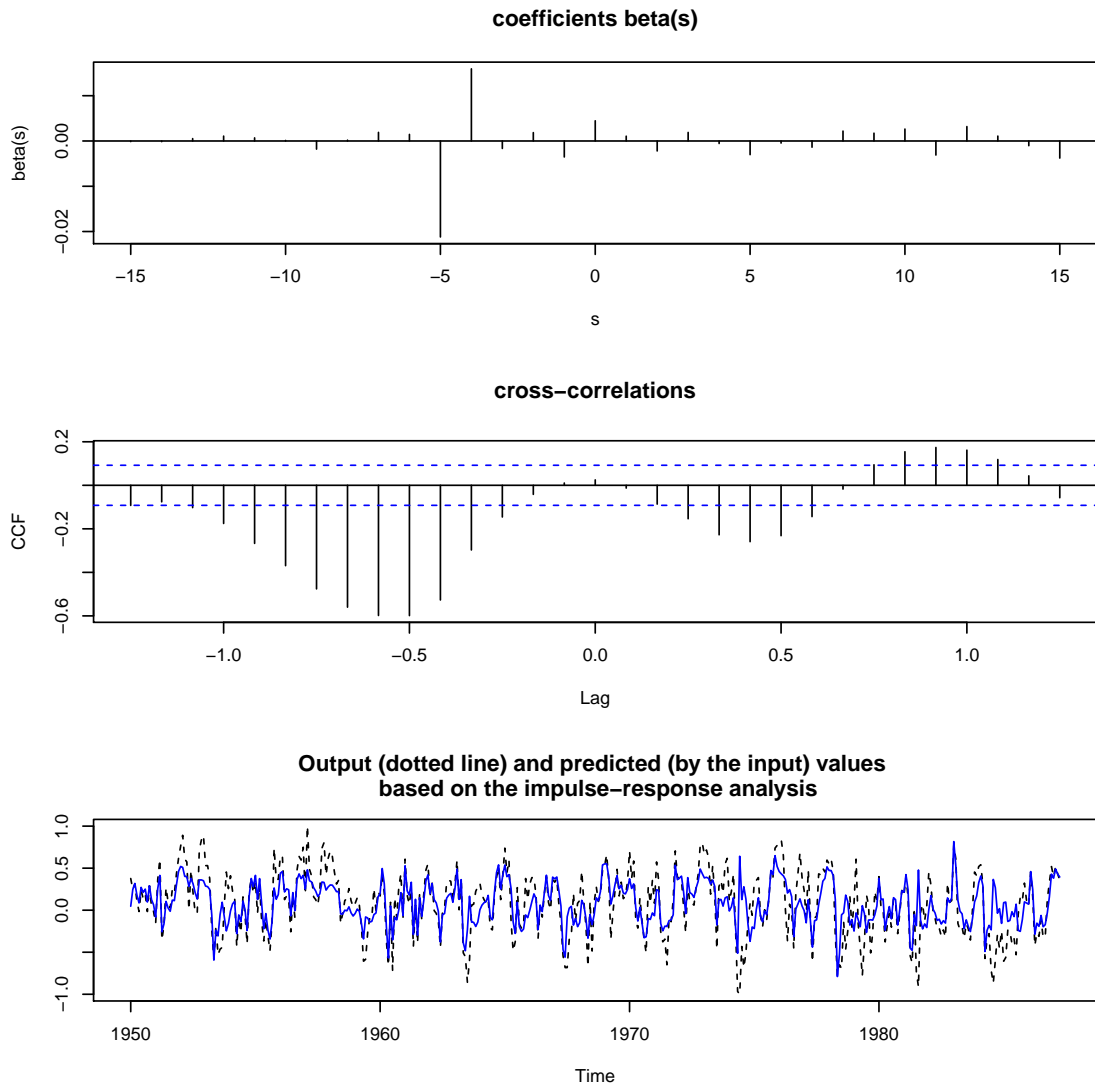
$$y_t = 66 - 18.5x_{t-5} - 12.3x_{t-6} - 8.5x_{t-7} - 7x_{t-8} + w_t$$

Si examinamos la relación inversa, es decir, un modelo de regresión con la serie de Recrutment  $y_t$  como la entrada, implica un modelo mucho más simple

```

## INPUT: rec OUTPUT: soi    L = 15    M = 32
##
## The coefficients beta(0), beta(1), beta(2) ... beta(M/2-1) are
##
## 0.004461216 0.001061309 -0.002206203 0.001890203 -0.0005578082
## -0.003021943 -0.0004643641 -0.00137669 0.002186134 0.001726505 0.002637365
## -0.003106259 0.003181714 0.001085057 -0.001040998 -0.003780144
##
##
## The coefficients beta(0), beta(-1), beta(-2) ... beta(-M/2+1) are
##
## 0.004461216 -0.003535525 0.001851014 -0.001654371 0.01593167 -0.02120013
## 0.001450279 0.001901667 0.000181094 -0.001796688 0.0001158145 0.0006828962
## 0.00107619 0.0005235587 -0.000145583 -0.0001296385
## The negative lags, at which the coefficients are large
## in absolute value, and the coefficients themselves, are:
##      lag s      beta(s)
## [1,]      3 0.01593167
## [2,]      4 -0.02120013

```



```
##
## The prediction equation is
##  $soi(t) = \alpha + \sum_s [ \beta(s) \cdot rec(t+s) ]$ , where  $\alpha = 0.4080661$ 
## MSE = 0.07023683
```

dependiendo de solo dos coeficientes

$$x_t = 0.41 + 0.016y_{t+4} - 0.02y_{t+5} + v_t$$

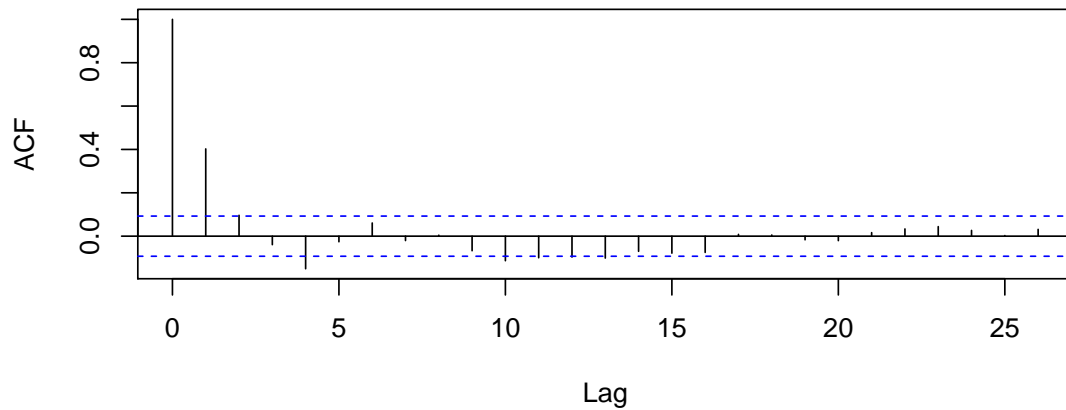
multiplicando ambos lados por  $50B^2$  y reorganizando, tenemos

$$(1 - 0.8B)y_t = 20.5 - 50B^5x_t + \epsilon_t$$

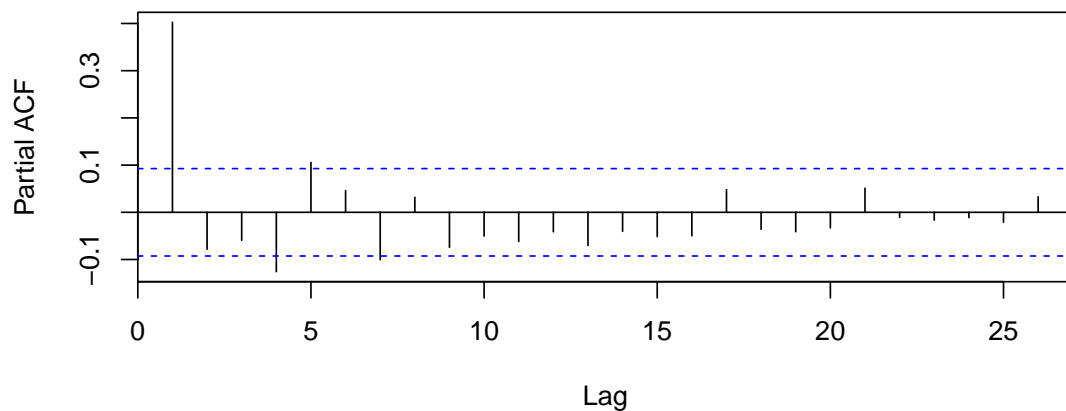
Finalmente, verificamos si el ruido,  $\epsilon_t$ , es blanco, donde vemos que al parecer el ruido sigue un proceso autorregresivo. Así, se tiene que



**Series u\$residuals**



**Series u\$residuals**



```
##
## Call:
## arima(x = fish[, 1], order = c(1, 0, 0), xreg = fish[, 2:3])
##
## Coefficients:
##          ar1  intercept      RL1      SL5
##          0.4489   14.6838   0.7902  -20.9988
## s.e.    0.0495    1.5605   0.0229    1.0812
##
## sigma^2 estimated as 49.57:  log likelihood = -1510.14,  aic = 3030.28
```

nuestro último modelo ajustado y parsimonioso es

$$y_t = 12 + 0.8y_{t-1} - 21x_{t-5} + \epsilon_t \quad y \quad \epsilon_t = 0.45\epsilon_{t-1} + w_t$$

donde  $w_t$  es ruido blanco con  $\sigma_w^2 = 50$ .

El ejemplo muestra que podemos obtener un estimado limpio para las funciones de transferencia que relacionan las dos series si la correlación cruzada  $\hat{\rho}_{xy}^2(\omega)$  es grande. La razón es que podemos escribir el error cuadrático medio reducido como

$$MSE = E \left[ \left( y_t - \sum_{r=-\infty}^{\infty} \beta_r x_{t-r} \right) y_t \right] = \gamma_{yy}(0) - \sum_{r=-\infty}^{\infty} \beta_r \gamma_{xy}(-r)$$

usando el resultado sobre la ortogonalidad de los datos y el término del error en el teorema de proyección. Entonces, sustituir las representaciones espectrales de las funciones de autocovarianza y covarianza cruzada e identificar la transformada de Fourier (11) en el resultado conduce a

$$\begin{aligned} MSE &= \int_{-1/2}^{1/2} [f_{yy}(\omega) - \beta(\omega) f_{xy}(\omega)] d\omega \\ &= \int_{-1/2}^{1/2} f_{yy}(\omega) [1 - \rho_{yx}^2(\omega)] d\omega \end{aligned} \quad (15)$$

donde  $\rho_{yx}^2(\omega)$  es solo la coherencia dada por

$$\rho_{yx}^2(\omega) = \frac{|f_{yx}(\omega)|^2}{f_{xx}(\omega) f_{yy}(\omega)}$$

La similitud es de (15) con el error cuadrático medio usual que resulta de predecir  $y$  desde  $x$  es obvio. Es ese caso, tendríamos

$$E(y - \beta x)^2 = \sigma_y^2(1 - \rho_{xy}^2)$$

para variables distribuidas conjuntamente  $x, y$  con media cero, varianzas  $\sigma_x^2$  y  $\sigma_y^2$ , y covarianza  $\sigma_{xy} = \rho_{xy}\sigma_x\sigma_y$ . Porque el error cuadrático medio en (15) satisface  $MSE \geq 0$  con  $f_{yy}(\omega)$  una función no negativa, se deduce que la coherencia satisface

$$0 \leq \rho_{xy}^2(\omega) \leq 1$$

para todo  $\omega$ . Además, se puede mostrar que la coherencia cuadrada es uno, cuando la salida está relacionada linealmente con la relación filtro (5), y no hay ruido, es decir,  $v_t = 0$ . Por lo tanto, la coherencia múltiple da una medida de la asociación o correlación entre las series de entrada y de salida como una función de la frecuencia.

## Referencias

- [1] Shumway, Robert H. ; Stoffer, David S. Time series analysis and its applications. With R examples. Third edition. Springer Texts in Statistics. Springer, New York, 2011. xii+596 pp. ISBN: 978-1-4419-7864-6
- [2] Wei, William W. S. Time series analysis. Univariate and multivariate methods. Second edition. Addison Wesley/Pearson, Boston, MA, 2006. xxii+614 pp. ISBN: 0-321-32216-9