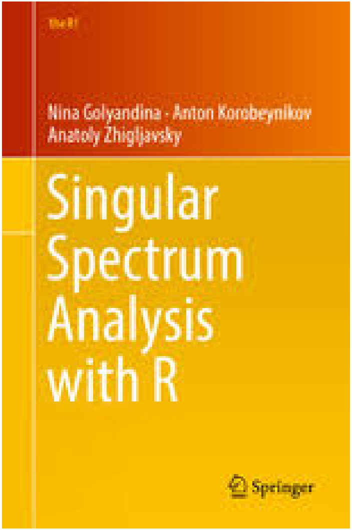
Aprendizaje Supervisado con Series Temporales

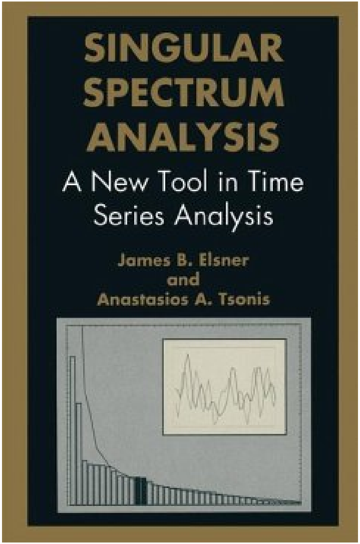
Singular Spectrum Analysis

Gabriel Martos Venturini
gmartos@utdt.edu

UNIVERSIDAD TORCUATO DI TELLA

Bibliografía recomendada





- *Principal Component Analysis*. Jolliffe (§ 12).

Agenda

Agenda

Signular Spectrum Analysis
- Introducción
- El modelo univariante
- Implementación en R

Forecasting y selección de modelos con SSA

Apéndice

Agenda

Motivación

- Las técnicas convencionales de análisis de series temporales (modelos ARIMA y extensiones) requieren de supuestos fuertes (estacionareidad) en los mecanismos de generación de datos.
- Singular Spectrum Analysis (SSA):
  - No necesitamos asumir **estacionareidad**.
  - Descomponemos la serie en diferentes componentes que pueden ser interpretadas como elementos de la tendencia, los ciclos y el ruido.

$$Y(t) = \underbrace{f_1(t) + \cdots + f_m(t)}_{\text{Señal}} + \underbrace{\varepsilon(t)}_{\text{Ruido}}$$

  - Forecasting: Relación de recurrencia lineal.
  - Extensible al contexto multivariante.
  - Drawback: Escasos elementos para construir herramientas de inferencia en torno a las estimaciones que hacemos con el modelo.

Algunas referencias

- de Carvalho, Martos (2021): *Modeling Interval Trendlines: Symbolic Singular Spectrum Analysis for Interval Time Series*. JF.
- de Carvalho, Martos (2020): *Brexit: Tracking and disentangling the sentiment towards leaving the EU*. IJF
  - `ASSA`: Applied Singular Spectrum Analysis.
- Hassani, Soofi, Zhigljavsky (2013): *Predicting inflation dynamics with singular spectrum analysis*. JRSS.
- Hassani, Soofi, Zhigljavsky (2010): *Predicting daily exchange rate with singular spectrum analysis*. NonLinAnalysis.
- de Carvalho, Rodriguez, Rua (2012): *Tracking the US business cycle with a singular spectrum analysis*. Economics Letters.
- Hassani, Heravi, Zhigljavsky. (2009): *Forecasting European industrial production with singular spectrum analysis*. IJF.

Agenda

Etapas de SSA y outputs del modelo

Dado el conjunto train $\{y_1, \ldots, y_n\}$ el *algoritmo* SSA tiene dos etapas:

A Decomposition.

A.1 Embedding (parámetro sensible l).

A.2 Singular value decomposition.

B Reconstruction.

B.1 Grouping (parámetro sensible m).

B.2 Diagonal averaging ("Hankelización").

- Outputs del modelo:
  - Conjunto de "componentes principales" que nos permiten construir una versión *filtrada* de la serie (nowcasting: $\widetilde{\mathbf{y}} = (\widetilde{y}_1, \ldots, \widetilde{y}_n)$).
  - Los elementos necesarios para extrapolar las componentes asociadas a la tendencia y el ciclo hacia el futuro (forecasting).

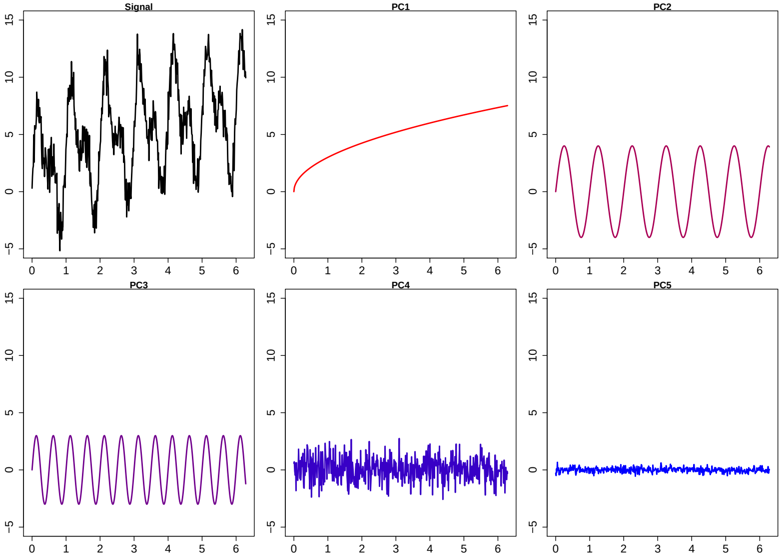


Figure: "l" impacta en la forma en que construimos estas 'componentes principales'. "m" indica cuantas componentes utilizamos para extrapolar.

A.1 Embedding

- Reorganizamos $\{y_1, \ldots, y_n\}$ en una matriz de **trayectorias**:

$$\mathbf{X}_{l \times k} = \begin{bmatrix} y_1 & y_2 & \cdots & y_k \\ y_2 & y_3 & \cdots & y_{k+1} \\ \vdots & \vdots & \ddots & \vdots \\ y_l & y_{l+1} & \cdots & y_n \end{bmatrix}.$$

- l es un parámetro de localidad $(k \equiv n - l + 1)$.
  - Algunos resultados teóricos sugieren que $l \approx n/2$.
- Elegir l por “VC” si el objetivo es hacer predicciones.
- Notar que $\mathbf{X}$ es una matriz Hankel.
  - A toda matriz de trayectorias le corresponde una serie temporal y visceversa (hay una relación biunívoca entre $\{y_1, \ldots, y_n\}$ y $\mathbf{X}$).

A.2 SVD

- Sean $\sigma_1 \geq \ldots \geq \sigma_l$ las raíces de los autovalores de $\mathbf{X}\mathbf{X}^T$
- $\mathbf{u}_1, \ldots, \mathbf{u}_l$ los correspondientes autovectores de $\mathbf{X}\mathbf{X}^T$.

$$\mathbf{X} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T \equiv \sum_{i=1}^{d} \mathbf{Z}_i,$$

(Ver detalles en Apénice)

- donde $\mathbf{Z}_i = \sigma_i \mathbf{u}_i \mathbf{v}_i^T$, $\mathbf{v}_i = \mathbf{X}^T \mathbf{u}_i / \sigma_i$, y

$$d = \max\{i \in \{1, \ldots, l\} : \sigma_i > 0\} = \mathrm{rank}(\mathbf{X}).$$

- $\mathbf{X}$ se descompone como una suma de d matrices de rango 1.
- Cada matriz $\mathbf{Z}_i$ para $i = 1, \ldots, d$ esta asociada a una 'componente principal' de la serie original $\{y_1, \ldots, y_n\}$.

(BackUp slide)

- ¿Qué representa la matriz $\mathbf{X}\mathbf{X}^T$?

$$\mathbf{X}\mathbf{X}^T = \begin{bmatrix} \sum_{j=1}^{k} y_j^2 & \sum_{j=1}^{k} y_j y_{j+1} & \cdots & \sum_{j=1}^{k} y_j y_{j+l-1} \\ \sum_{j=1}^{k} y_{j+1} y_j & \sum_{j=1}^{k} y_{j+1}^2 & \cdots & \sum_{j=1}^{k} y_{j+1} y_{j+l-1} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{j=1}^{k} y_{j+l-1} y_j & \sum_{j=1}^{k} y_{j+l-1} y_{j+1} & \cdots & \sum_{j=1}^{k} y_{j+l-1}^2 \end{bmatrix}.$$

- $\mathbf{X}\mathbf{X}^T$ es una estimación "local" de la función de autocovarianza.
- La descomposición en valores y vectores propios de $\mathbf{X}\mathbf{X}^T$ nos permite descomponer la serie como una suma de componentes que pueden interpretarse como tendencia, ciclos y ruido.
    - PCA para series temporales.

B.1 Agrupamiento

- Objetivo: Separar la señal del ruido.
- Para $\mathcal{I} \subset \{1, \ldots, d\}$ computamos

$$\mathbf{X}_{\mathcal{I}} = \sum_{i \in \mathcal{I}} \mathbf{Z}_i.$$

- Generalmente ocurre que $\mathcal{I} = \{1, \ldots, m\}$ con $m \ll d$.
- Criterios prácticos para elegir m:
  - $m_c = \min\{q : \sum_{i=1}^{q} \sigma_i / \sum_{j=1}^{d} \sigma_j \geq c\}$, con $c \in [0.7, 0.9]$.
  - Scree plot: $\{(i, \sigma_i)\}_{i=1}^{d}$ para definir el corte (criterio del "codo").
  - Criterio estadístico para automatizar la selección de componentes.
  - VC: Minimizamos el error de predicción fuera de la muestra.
- En general $\mathbf{X}_{\mathcal{I}}$ no será una matriz Hankel.

B.2 Hankelización

Promediamos sobre las antidiagonales de $\mathbf{X}_{\mathcal{I}}$ (Hankelizamos, ver BackUp) para construir la versión filtrada de la serie $\widetilde{\mathbf{y}} = (\widetilde{y}_1, \ldots, \widetilde{y}_n)$:

$$\widetilde{\mathbf{y}} = \overline{\mathbb{D}}(\mathbf{X}_{\mathcal{I}}) = \left(\frac{1}{|\mathcal{A}_1|} \sum_{(i,j)\in\mathcal{A}_1} x_{ij}^{l}, \ldots, \frac{1}{|\mathcal{A}_n|} \sum_{(i,j)\in\mathcal{A}_n} x_{ij}^{l} \right),$$

$|A|$ denota la cardinalidad del conjunto A, y la secuencia de conjuntos:

$$\mathcal{A}_c = \{(i,j) : i+j = c+1, i \in \{1, \ldots, l\}, j \in \{1, \ldots, k\}\},\ c = 1, \ldots, n,$$

definen los elementos pertenecientes a las n antidiagonales de $\mathbf{X}_{\mathcal{I}}$.

- El output del modelo $\widetilde{\mathbf{y}}$ es una versión *filtrada* de la serie original.
- Con $\widetilde{\mathbf{y}}$, los elementos de la descomposición, y el supuesto de recurrencia lineal construimos las predicciones (a continuación).

(BackUp)

Hankelizar = hacer promedios antidiagonales. Si

$$\mathbf{X}_{\mathcal{I}} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1k} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2k} \\ a_{31} & a_{32} & a_{33} & \cdots & a_{3k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{l1} & a_{l2} & a_{l3} & \cdots & a_{lk} \end{bmatrix}.$$

entonces:

$$\widetilde{y}_1 = a_{11}, \quad \widetilde{y}_2 = \frac{a_{12} + a_{21}}{2}, \quad \widetilde{y}_3 = \frac{a_{13} + a_{22} + a_{31}}{3}, \ldots, \quad \widetilde{y}_n = a_{lk}.$$

Criterio estadístico para determinar $\mathcal{I}$ (BackUp)

El *periodograma* asociado a $\mathbf{y} = (y_1, \ldots, y_n)$ se corresponde con $\{(\omega_j, \mathbb{I}(\omega_j))\}_{j=1}^{J}$, donde $\omega_j = 2\pi j/n$ para $j = 1, \ldots, J = \lfloor n/2 \rfloor$.

$$\mathbb{I}(\omega_j) = \frac{1}{n} \left| \sum_{t=1}^{n} y_t \exp^{-it\omega_j} \right|^2 .$$

Targeted grouping based on the Kolmogorov–Smirnov statistic
Set $i = 1$ and execute the steps:

Step 1. Compute residuals $\varepsilon = \mathbf{y} - \widetilde{\mathbf{y}}$, yield from SSA on $\mathcal{I} = \{1, \ldots, i\}$.

Step 2. Compute the periodogram of ε to test the null hypothesis of white noise using the Kolmogorov–Smirnov test based on the statistic:

$$\sqrt{J} \max\{|C(\omega_j) - j/J|\}_{j=1}^{J}, \quad C(\omega_j) = \frac{\sum_{i=1}^{j} \mathbb{I}(\omega_i)}{\sum_{i=1}^{J} \mathbb{I}(\omega_i)}.$$

If the null is rejected then make $i = i + 1$ and repeat Steps 1 and 2. Otherwise stop.

Agenda

Signular Spectrum Analysis

- Introducción
- El modelo univariante
- Implementación en R

Forecasting y selección de modelos con SSA

Apéndice

Nowcasting: Brexit poll tracker data

Datos composicionales: $\{(p_t, s_t, u_t)\}_{t=1}^{n}$ (leave–stay–undecided).

Table: Summary statistics for the 16 houses included in the poll tracker.

Poll-house	Mean Nmb Resp	n–pools
BMG Research	1464	8
ComRes	2258	16
Greenberg QR Research	2327	2
Harris	2114	1
ICM	1883	47
Ipsos MORI	909	12
Lord Ashcroft Polls	20058	1
Opinium	1970	13
ORB	1294	19
Panelbase	1547	2
Pew Research Center	1006	2
Populus	3368	2
Survation	1879	22
TNS	1414	12
YouGov	1882	113

https://ig.ft.com/sites/brexit-polling/

Paquete ASSA 2.0 en R

```
y = tsframe(dates, y)
sst(y, l = "automatic", m = "automatic")

Arguments:
y: tsframe format data.
l: window length; "automatic" sets the default option
              l = ceiling(y$n + 1) / 2.

m: number of leading eigentriples. An automatic criterion
   based on the cumulative periodogram of the residuals
   is provided by default.
```

- ASSA (2.0 en adelante) tiene implementados varios modelos de SSA (uni y multi variantes); los métodos automáticos de selección de modelos y las rutinas necesarias para hacer forecasting (next).
- Competencia: Rssa (Golyandina et al).

Nowcasting: SSA vs MSSA

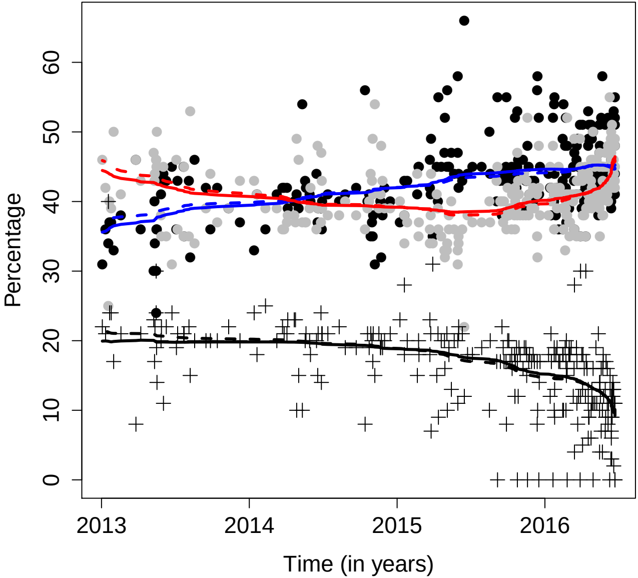


Figure: SSA vs MSSA (automatic) trendline estimations.

Agenda

Recurrencia lineal y predicciones

- RL: $y_{n+1} = s_{n+1} + \varepsilon_{n+1}$, con $s_{n+1} = \phi_1 s_n + \cdots + \phi_{l-1} s_{n+1-l}$.
  - Golyandina and Zhigljavsky, 2013, Sect 3.8.
- Elegimos (l, m) y estimamos: $\{\widetilde{y}_n, \ldots, \widetilde{y}_1\}$ para forcastear:

$$\widehat{y}_{n+1} = a_1 \widetilde{y}_n + \cdots + a_{l-1} \widetilde{y}_{n+1-l}.$$

- El vector $\mathbf{a} = (a_1, \ldots, a_{l-1})$ (estimaciones de ϕ) se obtiene

$$\mathbf{a} = \frac{1}{1-\nu^2} \sum_{i=1}^{m} \mathbf{u}_i[l] \mathbf{u}_i[-l] \in \mathbb{R}^{l-1}, \;\; \nu^2 = \sum_{i=1}^{m} \mathbf{u}_i[l]^2.$$

- Implementado en la versión 2.0 de `ASSA`.
  - En algunas versiones de Windows, R necesita que instales `Rtools` antes de descargar e instalar `ASSA` 2.0 (necesitas compilar código en C y Fortran para poder correr las rutinas internas del paquete).

Forecasting con `ASSA` (versión 2.0 en adelante)

```
predict(modelo, p)

Arguments

modelo: modelo estimado (sst, msst, msstc, isst, misst).

p: horizonte temporal del forecast.
```

- El comando `predict` toma todos los hiperparámetros (l, m) seteados en el modelo estimado como imputs para la predicción.
- Competencia: `Rssa` (desarrollado por Golyandina et al).

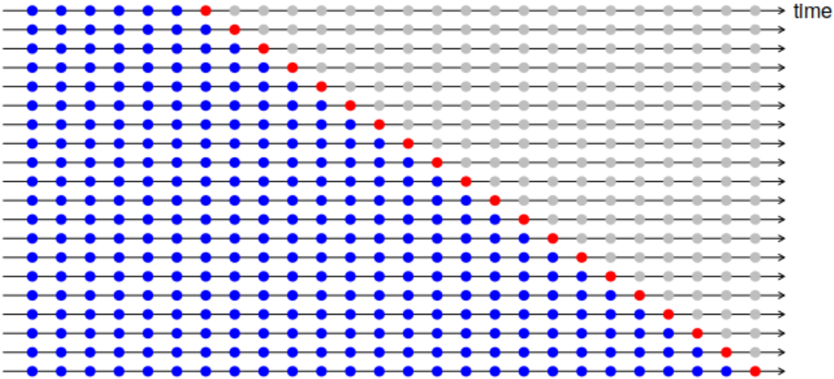
Selección de modelos: Hold–out set

- División de la data:

- Construimos una grilla de posibles valores para (l, m).
- Aprendemos las componentes sobre Train.
- Forcasteamos sobre Test y computamos MSE o MAE.
- Elegimos (l^*, m^*) minimizando las cantidades anteriores.
- Métrica alternativa: Errores escalados.
  - $\mathrm{ECE}_t = [(y_t - \widetilde{y}_t)/y_t]^2$.
  - $\mathrm{EAE}_t = |(y_t - \widetilde{y}_t)/y_t|$.
- Aconsejable considerar horizonte de forecast reducidos.

Selección de modelos: Rolling forecasting origin

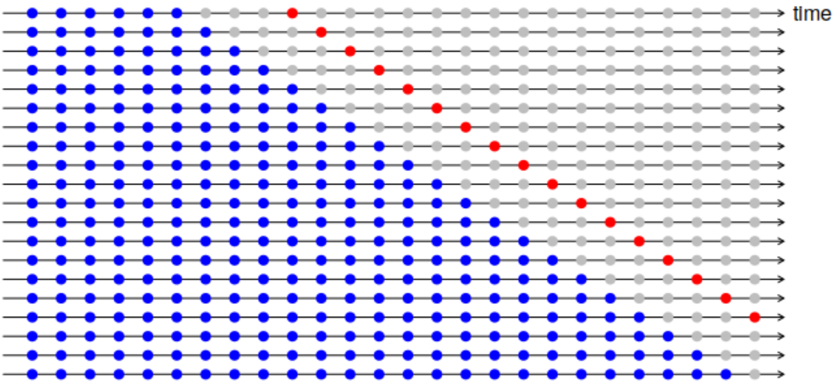
- Ventana creciente para train:



- Promediamos los errores de predicción a 1 período.

- Método de validación adecuado cuando vas a utilizar el modelo para hacer predicciones (de corto plazo) de manera dinámica.

- Si queremos considerar horizontes predictivos más largos:



- Puede resultar conveniente utilizar diferentes modelos (l's y m's distintos) para hacer predicciones a corto, mediano y largo plazo.

- A estos procedimientos NO los deberíamos llamar de "Validación Cruzada", sino más de bien de validación por fuera de la muestra.

Caso estudio

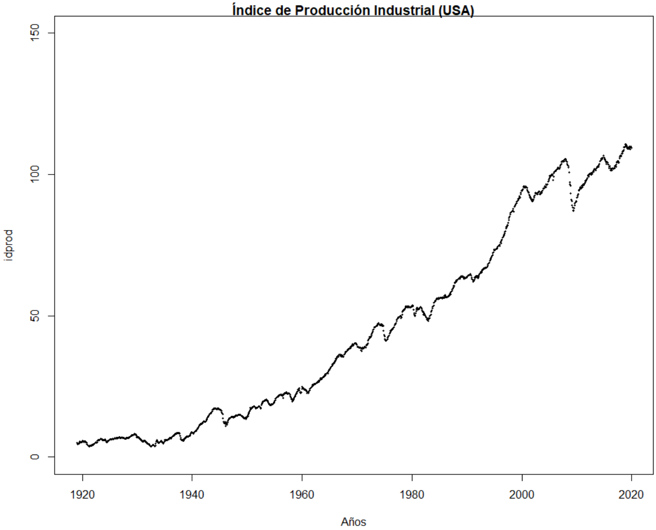


Figure: Índice de producción industrial, datos mensuales (período 1919–2019).


"That's all Folks!"

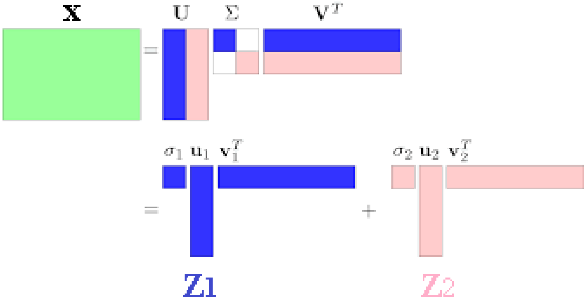
Agenda

Descomposición en valores singulares (BackUp)

- Descomponemos $\mathbf{X} \in \mathbb{R}^{l \times k}$ como suma de matrices de rango 1:

$$\mathbf{X} = \mathbf{U}_{l \times d} \mathbf{\Sigma}_{d \times d} \mathbf{V}^T_{d \times k}.$$

Ejemplo para el caso en donde $\text{rank}(\mathbf{X}) = 2$.



- $\mathbf{\Sigma}$ contiene las raíces de los autovalores no nulos de $\mathbf{X}\mathbf{X}^T$ o $\mathbf{X}^T\mathbf{X}$.
- $\mathbf{U}$ y $\mathbf{V}$ contienen los autovectores de $\mathbf{X}\mathbf{X}^T \in \mathbb{R}^{l \times l}$ y $\mathbf{X}^T\mathbf{X} \in \mathbb{R}^{k \times k}$.

Aproximaciones de rango pequeño (BackUp)

- Consideremos que $\mathbf{\Sigma}$ tiene ordenados los elementos en su diagonal principal de la siguiente manera: $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_d$.
- Para $r \leq d = \text{rank}(\mathbf{X})$, $\mathbf{X}_r \equiv \sum_{i=1}^{r} \sigma_i \mathbf{u}_i \mathbf{v}_i^T$ es la mejor aproximación de $\mathbf{X}$ con una matriz de rango r en el sentido:

$$\mathbf{X}_r = \arg\left\{ \min_{\mathbf{A}} \|\mathbf{X} - \mathbf{A}\|_F, \text{ s.a. } \text{rank}(\mathbf{A}) = r \right\}$$

- El resultado también vale para otras normas matriciales.
- Esta es la idea que subyace al hacer SVD y descomponer la matriz de trayectorias en SSA en una suma de matrices de rango 1.
- Volver a la slide §–12