# SUAVIZAMIENTO DE SERIES DE TIEMPO POR SEGMENTOS DEL RECORRIDO DE LOS DATOS

**Víctor M. Guerrero**
Departamento de Estadística,
Instituto Tecnológico Autónomo de México – ITAM
y
**Eliud Silva**
Facultad de Ciencias Actuariales, Universidad Anáhuac del Norte

**Reunión sobre análisis aplicado de series de tiempo**

**ITAM-Febrero de 2015**

# ÍNDICE

1. INTRODUCCIÓN
2. CONSIDERACIONES INICIALES
3. PROPUESTA METODOLÓGICA
4. ILUSTRACIONES EMPÍRICAS
5. CONCLUSIONES

# 1. INTRODUCCIÓN

- El suavizamiento de datos ordenados (ya sea por edades, en el tiempo o en alguna otra dimensión) se usa para visualizar tendencias subyacentes.

- A los datos ordenados en general se les llamará series de tiempo, aunque la dimensión de referencia de los datos no sea el tiempo.

- Las series observadas por considerar deben contener datos equidistantes en la dimensión que corresponda, sin importar si hay algunos datos faltantes.

- En este contexto, suavizamiento y estimación de tendencias son conceptos que están claramente asociados.

# 1. INTRODUCCIÓN

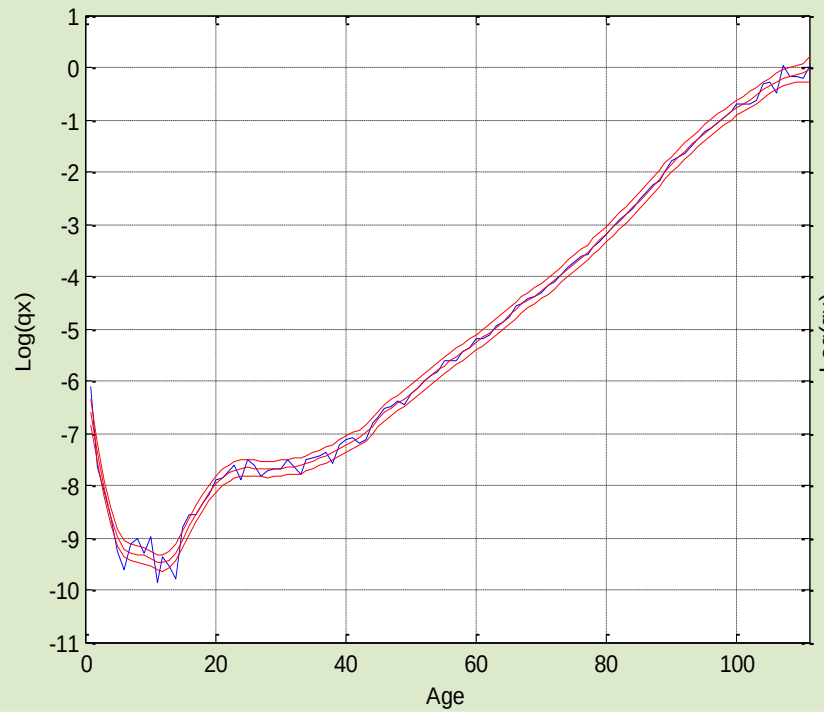
- En problemas actuariales, es común que el analista aplique graduación para suavizar los datos observados, sin que se pierda la interpretación de los mismos.

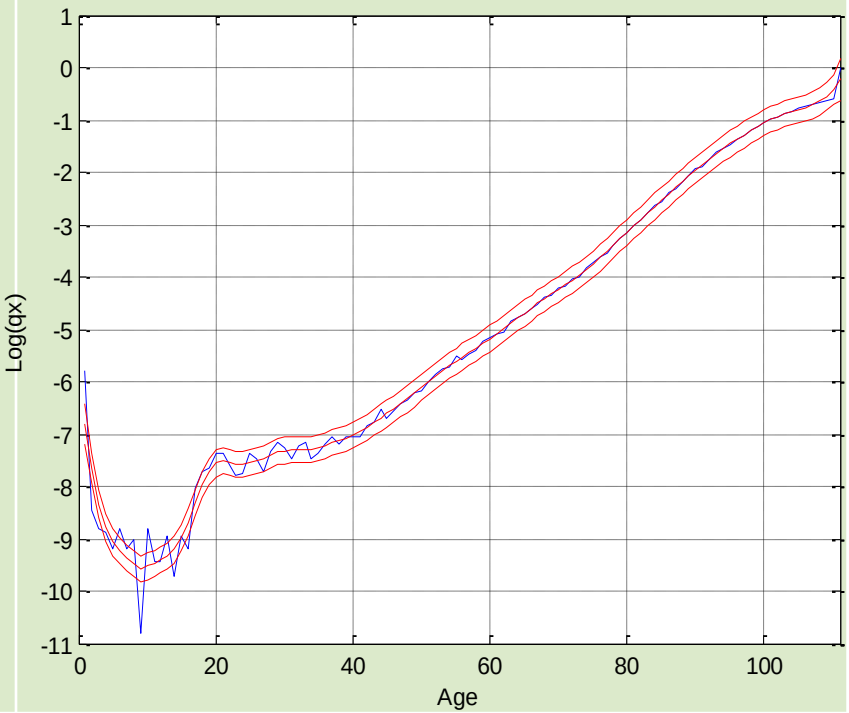
- Al realizar graduación de tasas de mortalidad, se observa que la tasa de mortalidad típicamente muestra tres patrones de variabilidad diferentes: el primero grande, el segundo moderado y el tercero pequeño.

- Ejemplos de esto se muestran en la figura que sigue, con datos de Suecia y de Noruega.

## Logaritmo de las Tasas de Mortalidad para ambos sexos en 2009 y tendencias con **suavidad de 75%.** Las bandas son de ± 2 desviaciones estándar.

| Suecia | Noruega |
|---|---|

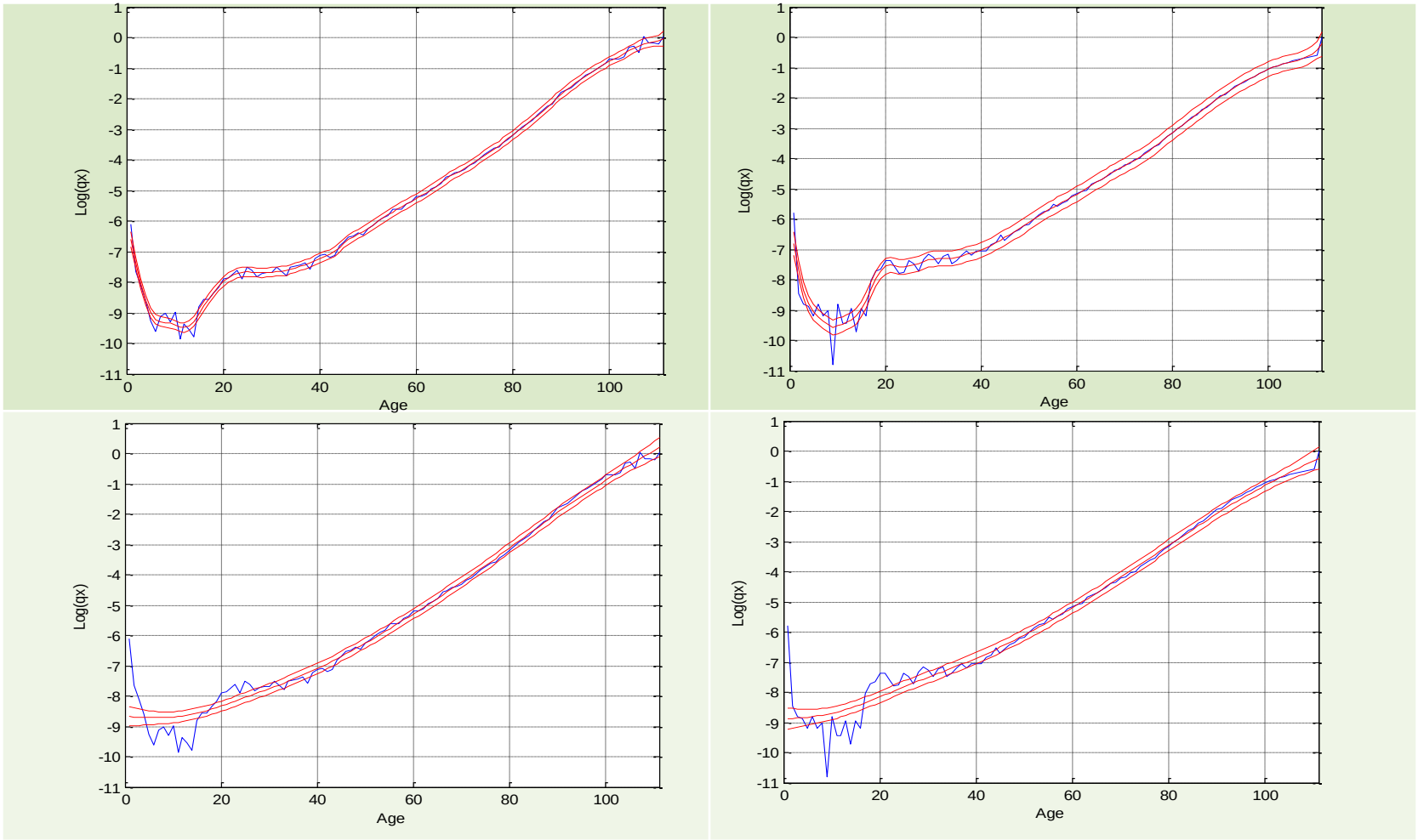



- El porcentaje de suavidad es el mismo en ambos casos y eso permite la comparabilidad.
- Los patrones de tendencia son similares globalmente, pero no por segmentos de las series.

## Tendencias con **suavidades de 75% y 90%.**

| Suecia | Noruega |
|---|---|




- Al crecer la suavidad, las tendencias se asemejan, pero se pierde la interpretación demográfica de los datos.

# 1. INTRODUCCIÓN

- Al estudiar series de tiempo económicas, es común que exista algún cambio estructural en la tendencia, al igual que en la variabilidad.

- Un modelo estadístico con cambio estructural de la serie se puede obtener al realizar un análisis de intervención, como en Box *et al*. (1994).

- Con dicha metodología de análisis se tiene en cuenta el rompimiento de la tendencia, pero no el cambio en la variabilidad.

- Una estimación de la tendencia subyacente, que tiene en cuenta el cambio en la variabilidad, se obtiene al dividir la serie en los segmentos que se requieran y aplicar después el procedimiento que aquí se sugiere.

# 1. INTRODUCCIÓN

- El objetivo de este trabajo consiste en simplificar la inspección de los datos, después de eliminar de la serie las fluctuaciones que sean indeseables, en el contexto de un análisis exploratorio.

- Se parte del hecho de que se puede medir la suavidad de la tendencia y para ello se hace uso de unos índices de suavidad que aquí se deducen.

- Los índices permiten fijar las suavidades en valores deseados y, en consecuencia, se pueden controlar las suavidades de las tendencias estimadas resultantes.

- Sólo es válido comparar tendencias si sus suavidades son iguales. Esto es semejante a lo que ocurre al comparar intervalos de confianza, los cuales deben tener el mismo nivel de confianza para ser comparables.

- La graduación actuarial propuesta por Whittaker (1923) surge al minimizar la función

$$M(\lambda) = (\mathbf{Y} - \boldsymbol{\tau})' W (\mathbf{Y} - \boldsymbol{\tau}) + \lambda\, \boldsymbol{\tau}' K' K \boldsymbol{\tau}$$

donde

$\mathbf{Y} = (Y_1, \ldots, Y_N)'$ es un vector de valores observados;

$\boldsymbol{\tau} = (\tau_1, \ldots, \tau_N)'$ contiene los valores graduados;

$W = \mathrm{diag}\,(w_1, \ldots, w_N)$ es una matriz de ponderaciones.

- La matriz $K$ permite expresar la $d$-ésima diferencia de los datos.

- El método de Whittaker, con $d = 2$ y $W = I$, fue usado por Hodrick y Prescott (1980, 1997) para estimar tendencias y ciclos de series económicas. Se le conoce en econometría como filtro de Hodrick y Prescott (HP).

- El problema de minimización se plantea como

$$\min_{\tau_t} \sum_{t=1}^{N} \frac{1}{\sigma_1^2}(Y_t - \tau_t)^2 + \sum_{t=3}^{N} \frac{1}{\sigma_0^2}(\tau_t - 2\tau_{t-1} + \tau_{t-2})^2$$

con

$Y_t$ la variable económica observada al tiempo $t$;

$\tau_t$ el valor de la tendencia;

$\sigma_1^2$ la varianza del componente cíclico;

$\sigma_0^2$ la varianza del crecimiento de la tendencia;

$\lambda = \sigma_1^2 / \sigma_0^2$ el balance entre la fidelidad a los datos originales y la suavidad de la tendencia.

- El estimador de la tendencia resulta ser

$$\hat{\boldsymbol{\tau}} = (\mathrm{I}_{\mathrm{N}} + \lambda \mathrm{K}'\mathrm{K})^{-1}\mathbf{Y}$$

en donde lo único que se requiere para calcular la tendencia estimada es conocer el parámetro de suavizamiento $\lambda$ , pues los demás elementos están dados.

- En Guerrero (2008) se presenta un método para elegir ese parámetro en el caso del filtro HP y que permite al usuario imponer de antemano un porcentaje de suavidad deseado para la tendencia.

- Se usa un índice que relaciona a la suavidad con el parámetro de suavizamiento, de manera que la elección de dicho parámetro se vuelve tan sencillo como resolver una ecuación no-lineal.

# 2. CONSIDERACIONES INICIALES

- Nuestra propuesta surge de la idea de que es importante conocer y medir el efecto que tiene la selección de $\lambda$ sobre el grado de suavidad de la tendencia.

- No utilizamos criterios automáticos (e. *g*. AIC, BIC, Validación Cruzada, etc.) para elegirlo, aunque los procedimientos respectivos sean fáciles de usar. Hastie y Tibshirani (1999) muestran algunas de sus deficiencias.

- Tampoco usamos Máxima Verosimilitud, para no perder de vista la suavidad deseada (preferimos calibrar $\lambda$ ). Proietti (2005) y Kauerman (2005) indican que la suavidad que se logra con MV no es suficiente.

- El índice de suavidad es una reparametrización de lo que se conoce como grados de libertad efectivos en algunas técnicas estadísticas (e. *g*. en Modelos Mixtos y en Suavizado con "*Splines*") y que es útil para elegir $\lambda$ .

- Respecto a la segmentación de los datos, o sea, para elegir el valor de $N_1$ que divide la serie en dos segmentos, se sugiere seguir alguno de los caminos siguientes:

1) Fijar un valor razonable *a priori*, teniendo en cuenta el fenómeno que se estudia.

Por ejemplo, para tasas de mortalidad se podría considerar una segmentación del recorrido de las edades, de acuerdo con las diferentes etapas de la vida.

2) Usar un procedimiento basado en los datos, que optimice algún criterio estadístico.

Aquí se propone mantener la idea de sencillez que caracteriza a un análisis exploratorio y minimizar el error estándar de la tendencia estimada.

Se considera el caso de dos segmentos en detalle. Los resultados se extienden sin dificultad al caso de tres o más segmentos.

- La función a minimizar ahora es

$$\min_{\tau_t} \sum_{t=1}^{N_1} \frac{1}{\sigma_1^2}(Y_{1,t} - \tau_t)^2 + \sum_{t=N_1+1}^{N} \frac{1}{\sigma_2^2}(Y_{2,t} - \tau_t)^2 + \sum_{t=3}^{N} \frac{1}{\sigma_0^2}(\tau_t - 2\tau_{t-1} + \tau_{t-2})^2$$

donde $\sigma_1^2$ es la varianza de las desviaciones respecto a la tendencia del primer segmento de datos y $\sigma_2^2$ la del segundo tramo.

$N_1$ y $N_2 = N - N_1$ son las longitudes de los segmentos

y

$\{Y_{1,t}\}$, $\{Y_{2,t}\}$ son las series de datos en cada segmento.

- Un modelo de componentes no-observables, del tipo *señal más ruido,* para este problema es el siguiente

$$\begin{aligned} Y_{1,t} &= \tau_t + \eta_{1,t} \\ Y_{2,t} &= \tau_t + \eta_{2,t} \end{aligned} \quad \text{con} \quad \begin{aligned} \eta_{1,t} &\sim (0, \sigma_1^2) \\ \eta_{2,t} &\sim (0, \sigma_2^2) \end{aligned} \quad \text{para} \quad \begin{aligned} t &= 1, \ldots, N_1 \\ t &= N_1 + 1, \ldots, N \end{aligned}$$

$$\tau_t = 2\tau_{t-1} - \tau_{t-2} + \varepsilon_t \text{ con } \varepsilon_t \sim (0, \sigma_0^2) \text{ para } t = 3, \ldots, N,$$

donde los errores aleatorios:

- no están autocorrelacionados,
- ni se encuentran correlacionados entre sí.

- En notación matricial, el modelo se puede escribir como

$$\begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{pmatrix} = \begin{pmatrix} \boldsymbol{\tau}_1 \\ \boldsymbol{\tau}_2 \end{pmatrix} + \begin{pmatrix} \boldsymbol{\eta}_1 \\ \boldsymbol{\eta}_2 \end{pmatrix} \quad \text{y} \quad \mathrm{K}\begin{pmatrix} \boldsymbol{\tau}_1 \\ \boldsymbol{\tau}_2 \end{pmatrix} = \boldsymbol{\varepsilon}$$

donde $\mathbf{Y}_i$, $\boldsymbol{\tau}_i$ y $\boldsymbol{\eta}_i$ son vectores de dimensión $\mathrm{N}_i$, $i = 1, 2$, con $\boldsymbol{\varepsilon} \sim (\mathbf{0}, \sigma_0^2 \mathrm{I}_{\mathrm{N-2}})$ y $\boldsymbol{\eta}_i \sim (\mathbf{0}, \sigma_i^2 \mathrm{I}_{\mathrm{N}_i})$, donde los vectores de errores respectivos no están correlacionados entre sí.

- Al aplicar Mínimos Cuadrados Generalizados se obtiene el Mejor Estimador Lineal e Insesgado de la tendencia, que está dado por

$$\begin{pmatrix} \hat{\boldsymbol{\tau}}_1 \\ \hat{\boldsymbol{\tau}}_2 \end{pmatrix} = \begin{pmatrix} I_{N_1} + \lambda_1 (K'_1 K_1 + k'_1 k_1) & \lambda_1 k'_1 k_2 \\ \lambda_2 k'_2 k_1 & I_{N_2} + \lambda_2 (K'_2 K_2 + k'_2 k_2) \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{pmatrix}$$

donde $K_1$ y $K_2$ son de dimensión $(N_1\text{-}2)\times N_1$ y $(N_2\text{-}2)\times N_2$ respectivamente, mientras que

$$k_1 = \begin{pmatrix} 0_{2\times(N_1-2)} & \begin{matrix} 1 & -2 \\ 0 & 1 \end{matrix} \end{pmatrix} \quad \text{y} \quad k_2 = \begin{pmatrix} \begin{matrix} 1 & 0 \\ -2 & 1 \end{matrix} & 0_{2\times(N_2-2)} \end{pmatrix}$$

- Con la estimación simultánea y la presencia de estas dos matrices se evita la discontinuidad en la unión de los dos segmentos.

- Además, la matriz de varianza-covarianza del estimador está dada por

$$\Gamma = \mathrm{Var}\begin{pmatrix} \hat{\boldsymbol{\tau}}_1 \\ \hat{\boldsymbol{\tau}}_2 \end{pmatrix} = \left[ \begin{pmatrix} \sigma_1^{-2} I_{N_1} & 0 \\ 0 & \sigma_2^{-2} I_{N_2} \end{pmatrix} + \sigma_0^{-2} \begin{pmatrix} K'_1 K_1 + k'_1 k_1 & k'_1 k_2 \\ k'_2 k_1 & K'_2 K_2 + k'_2 k_2 \end{pmatrix} \right]^{-1}$$

- Para estimar las varianzas de los errores, se usa la Suma de Cuadrados Residual.

- Los estimadores insesgados que se obtienen son

$$\hat{\sigma}_0^2 = \mathbf{d}'\mathbf{d}/(N-2) \quad \text{y} \quad \hat{\sigma}_i^2 = \lambda_i \hat{\sigma}_0^2 \quad \text{para} \quad i = 1, 2,$$

donde los valores de las $\lambda_i$ se suponen conocidos.

## Medidas de participación en la precisión e índices de suavidad

- La inversa de la matriz de varianza-covarianza, $\Gamma$, es una matriz de precisión que se puede escribir como

$$\Gamma^{-1} = \begin{pmatrix} \sigma_1^{-2} I_{N_1} & 0 \\ 0 & \sigma_2^{-2} I_{N_2} \end{pmatrix} + \sigma_0^{-2} \begin{pmatrix} K'_1 K_1 + c_1 k'_1 k_1 & c_2 k'_1 k_2 \\ c_3 k'_2 k_1 & c_4 k'_2 k_2 \end{pmatrix} + \sigma_0^{-2} \begin{pmatrix} (1-c_1) k'_1 k_1 & (1-c_2) k'_1 k_2 \\ (1-c_3) k'_2 k_1 & K'_2 K_2 + (1-c_4) k'_2 k_2 \end{pmatrix}$$

con $0 \le c_i \le 1$ para $i = 1, 2$.

- La primera matriz del lado derecho, $B_0$, está ligada al modelo de señal más ruido, mientras que las otras dos matrices, $B_1$ y $B_2$, se asocian con el suavizamiento de cada uno de los segmentos.

- Se eligen $c_1 = 1$, $c_2 = c_3 = N_1/N$ y $c_4 = 0$, para que $B_1$ y $B_2$ tengan propiedades semejantes a las de $B_1 + B_2$, como se requiere para poder medir suavidad.

## Medidas de participación en la precisión e índices de suavidad

- Las medidas propuestas son

$$\mathrm{tr}[B_1(B_0+B_1+B_2)^{-1}]/N = \mathrm{tr}[\sigma_0^{-2}\begin{pmatrix} K'_1K_1+k'_1k_1 & \frac{N_1}{N}k'_1k_2 \\ \frac{N_1}{N}k'_2k_1 & 0 \end{pmatrix}(B_0+B_1+B_2)^{-1}]/N$$

y

$$\mathrm{tr}[B_2(B_0+B_1+B_2)^{-1}]/N = \mathrm{tr}[\sigma_0^{-2}\begin{pmatrix} 0 & \frac{N_2}{N}k'_1k_2 \\ \frac{N_2}{N}k'_2k_1 & K'_2K_2+k'_2k_2 \end{pmatrix}(B_0+B_1+B_2)^{-1}]/N$$

- Cuantifican la proporción de precisión atribuible a la suavidad del segmento respectivo:

(i) toman valores en [0, 1]; (ii) junto con la proporción de $B_0$, suman a la unidad; (iii) son invariantes ante transformaciones lineales no-singulares; y (iv) se comportan linealmente.

## Medidas de participación en la precisión e índices de suavidad

- Se escriben los índices de suavidad en términos de las matrices que usan las lambdas, o sea,

$$B_{1,\lambda} = B_1 B_0^{-1} = \begin{pmatrix} \lambda_1(K_1'K_1 + k_1'k_1) & \frac{N_1}{N}\lambda_2 k_1' k_2 \\ \frac{N_1}{N}\lambda_1 k_2' k_1 & 0 \end{pmatrix}$$

y

$$B_{2,\lambda} = B_2 B_0^{-1} = \begin{pmatrix} 0 & \frac{N_2}{N}\lambda_2 k_1' k_2 \\ \frac{N_2}{N}\lambda_1 k_2' k_1 & \lambda_2(K_2'K_2 + k_2'k_2) \end{pmatrix}$$

- Se definen así los índices de suavidad como

$$S_i(\lambda_1, \lambda_2; N) = tr[B_{i,\lambda}(I_N + B_{1,\lambda} + B_{2,\lambda})^{-1}]/N$$

y cada uno se interpreta como proporción de precisión atribuible al suavizamiento del segmento i de los datos, con $i = 1, 2$.

## Medidas de participación en la precisión e índices de suavidad

- Se cumple que, para $i = 1, 2$,

$$S_i(\lambda_1, \lambda_2; N) \to N_i/N \quad \text{conforme} \quad \lambda_i \to \infty$$

así que, para encontrar el valor de lambda, el porcentaje deseado de suavidad, digamos $100S_i\%$, se debe multiplicar por $N_i/N$ e igualarlo a su respectivo índice de suavidad.

- O sea, si se fija el porcentaje deseado de suavidad para cada segmento en $100S_i\%$, se tiene que resolver con un método numérico el sistema de ecuaciones

$$S_i(\lambda_1, \lambda_2; N) = (N_i/N)100S_i\% \quad \text{con} \quad \lambda_i > 0$$

para $i = 1, 2$, simultáneamente.

## Medidas de participación en la precisión e índices de suavidad

- El porcentaje de suavidad total que se alcance estará dado por el promedio ponderado

$$100S\% = [N_1(100S_1\%) + N_2(100S_2\%)]/N$$

- De acuerdo a la experiencia lograda hasta el momento, la elección del porcentaje global de suavidad se realiza mejor con base en el conocimiento que se tiene del fenómeno en estudio.

- Además, conviene saber lo que es razonable esperar que ocurra al variar el porcentaje de suavidad (o sea, mayor suavidad implica menos fidelidad a los datos observados y viceversa).

## Ejemplo 1.
## Tasas de mortalidad en Suecia y Noruega.

- Los primeros dos conjuntos de datos que aquí se estudian se refieren a las Tasas de Mortalidad Total (ambos sexos) para Suecia y Noruega en 2009.

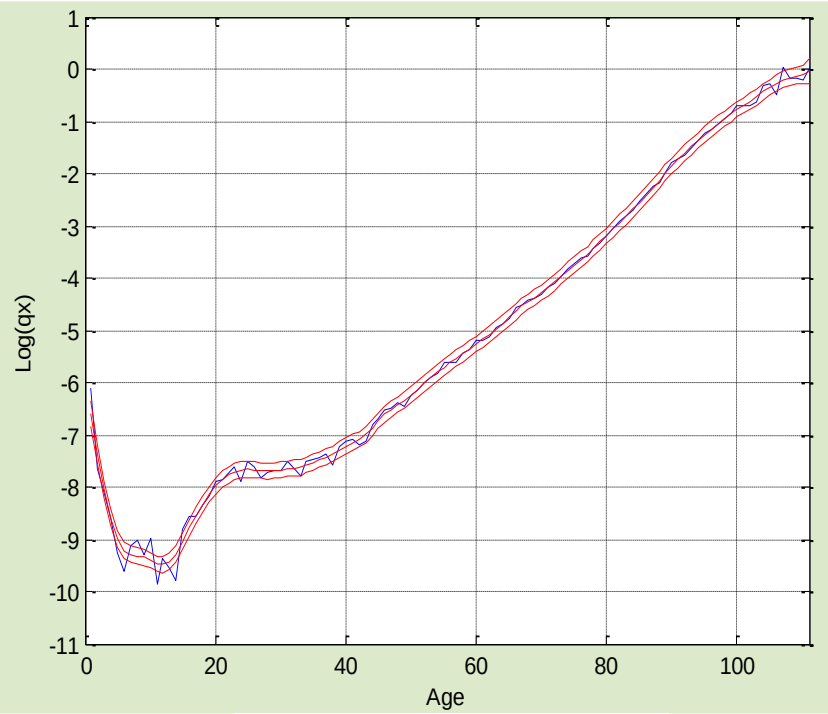
- Los datos provienen de *www.humanmortality.org* y están expresados como $\log(q_x)$, donde $q_x$ es la tasa de mortalidad a edad $x$.

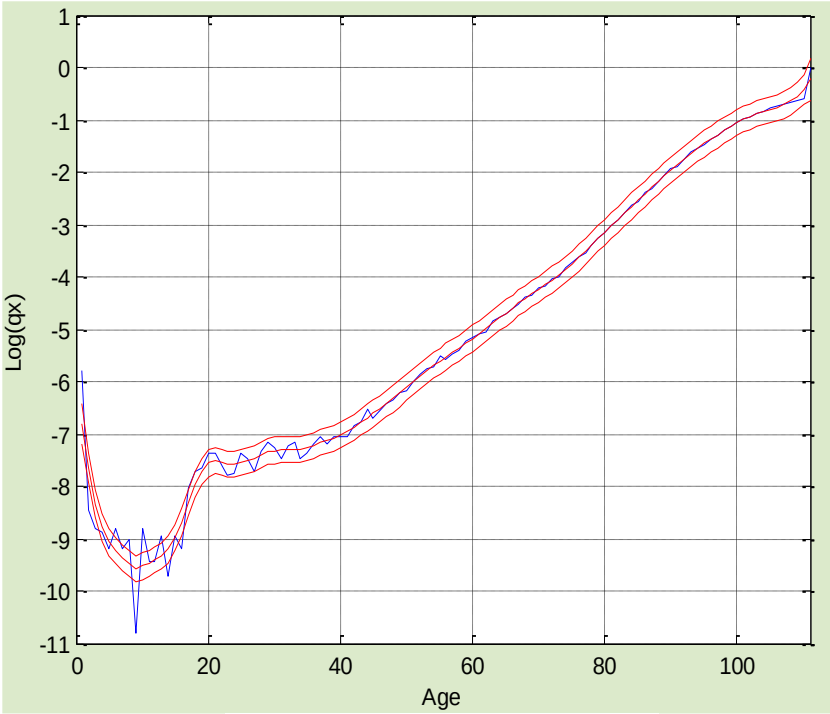
- El objetivo de esta aplicación es mostrar que, **aun en países que producen excelentes estadísticas vitales, se requiere aplicar suavizamiento a los datos** que generan.

- Esto hace evidente que el suavizamiento es imprescindible para datos de otros países.

## Figura 1. Logaritmo de las tasas de mortalidad (ambos sexos) en 2009 y tendencias con suavidad del 75%.

| Suecia | Noruega |
|---|---|



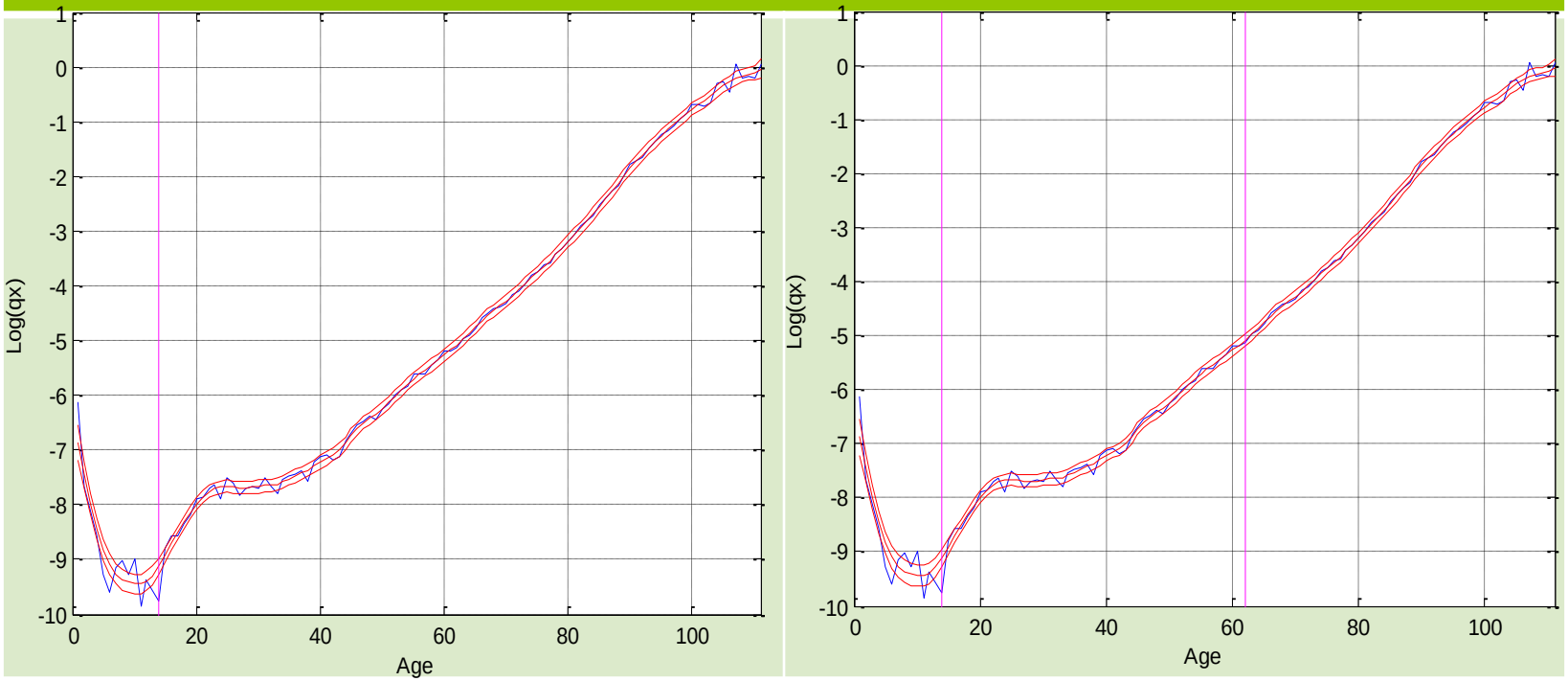

$\lambda = 5.6$ y $\hat{\sigma}_{\varepsilon} = 0.0662$

$\lambda = 5.6$ y $\hat{\sigma}_{\varepsilon} = 0.1074$

- Igual constante de suavizamiento y mayor varianza en Noruega.
- El comportamiento de los datos observados y de sus tendencias difiere en edades tempranas y medianas.

Figura 2. . **Segmentación empírica.** Tendencias con 75% de suavidad global: dos segmentos con corte en edad 14 y suavidades de 80% y 74.28% (izquierda); tres segmentos con cortes en edades 14 y 62, y suavidades de 80%, 75% y 73.57% (derecha).

**Suecia, ambos sexos, 2009**

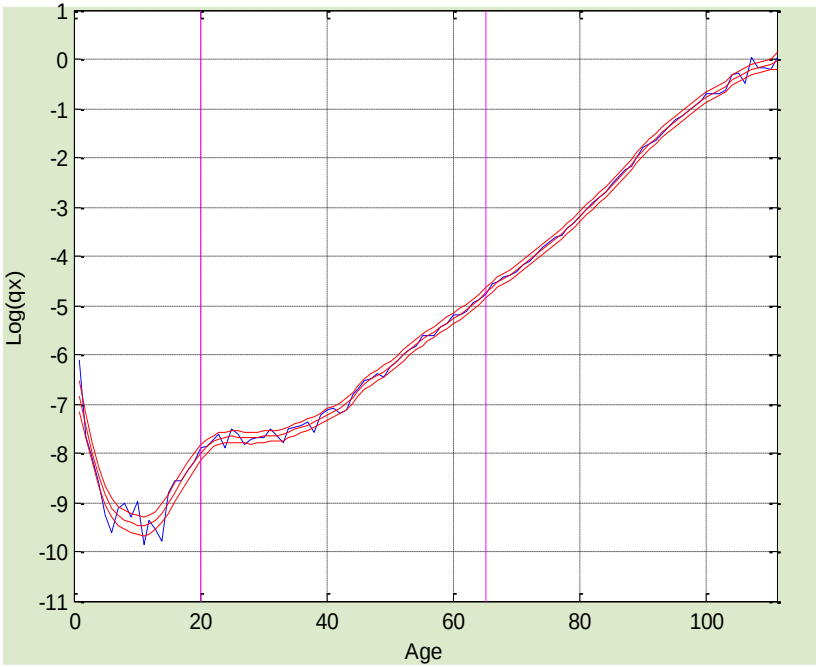


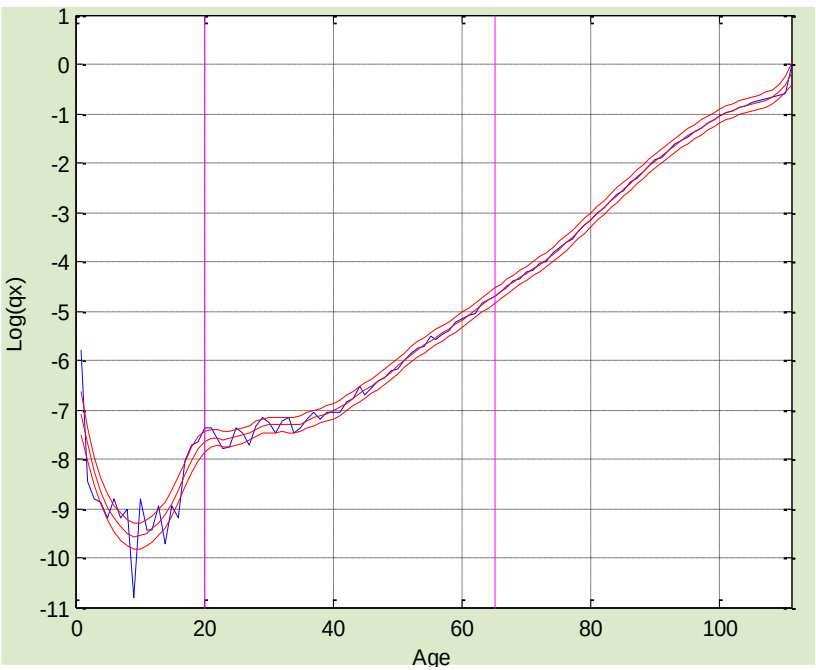

$\lambda_1 = 19.0,\ \lambda_2 = 4.8$ y $\hat{\sigma}_\varepsilon = 0.0519$ $\lambda_1 = 18.5,\ \lambda_2 = 3.8,\ \lambda_3 = 3.4$ y $\hat{\sigma}_\varepsilon = 0.0552$

- Disminuye la varianza al pasar de 1 a 2 segmentos, pero no al pasar de 2 a 3.

Figura 3. **Segmentación *a priori*.** Tendencias con 75% de suavidad global y tres segmentos, con cortes en edades 20 y 65, y suavidades de 80%, 75% y 72.83%.

| Suecia | Noruega |
|---|---|




$\lambda_1 = 16.0,\ \lambda_2 = 3.7,\ \lambda_3 = 3.1$ y $\hat{\sigma}_\varepsilon = 0.0563$ $\quad$ $\lambda_1 = 16.0,\ \lambda_2 = 3.7,\ \lambda_3 = 3.1$ y $\hat{\sigma}_\varepsilon = 0.0772$

- Mismas constantes de suavizamiento y mayor varianza en Noruega.
- El patrón de las tendencias en edades tempranas ya luce similar.

## Ejemplo 2.
## Remesas familiares trimestrales de EUA a México.

- Para este ejemplo de carácter económico se usan $N = 68$ datos de la serie trimestral de remesas familiares de EUA a México, que cubren del trimestre 1 de 1995 al trimestre 4 de 2011 (*www.inegi.org.mx/sistemas/bie*).

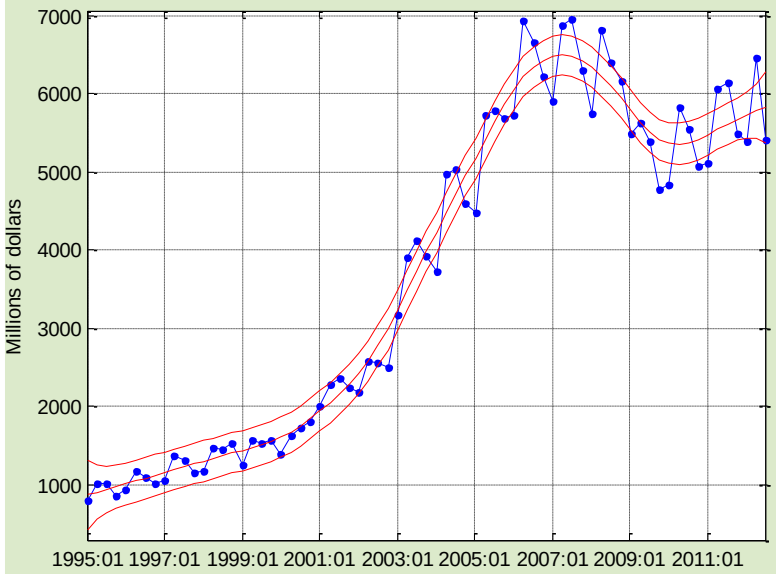
- Como no era claro el motivo para el cambio en variabilidad, se aplicó el procedimiento para elegir el punto de corte que minimiza el error estándar de la tendencia estimada.

- Se impuso 100S% = 85% como porcentaje de suavidad global y se localizó el punto de corte óptimo en $N_1 = 27$ (el trimestre 3 de 2001), cuando ocurrió el ataque terrorista a Nueva York.

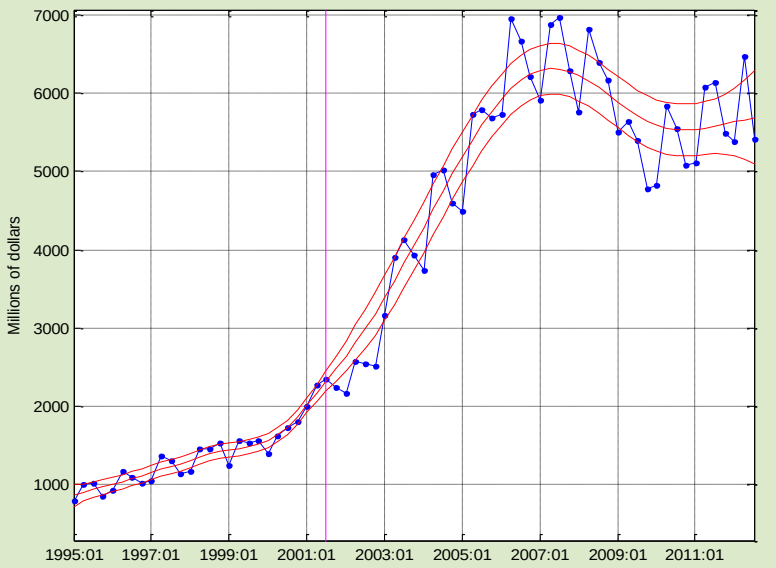
- Con suavidad global de 85% y ya que hay menor variabilidad en el primer segmento, se cambió ligeramente la suavidad a $100S_1\% = 84\%$, con lo cual se dedujo $100S_2\% = 85.6\%$.

Figura 4. Tendencias con suavidad global de 85% sin segmentar (izquierda) y segmentada. **Punto de corte elegido empíricamente**, localizado en el trimestre 3 de 2001.

**Remesas familiares trimestrales de EUA a México**




$\lambda = 49$ y $\hat{\sigma}_0 = 50.1306$




$\lambda_1 = 8.55, \lambda_2 = 265.98$ y $\hat{\sigma}_0 = 33.3777$

- Se percibe un ligero cambio en la tendencia estimada y es notoria la diferente amplitud de los intervalos.
- El error estándar de la tendencia estimada se redujo.

# 5. CONCLUSIONES

- La metodología sugerida proporciona mucha flexibilidad para estimar tendencias con diferentes grados de suavidad, para los segmentos de datos que se hayan definido.

- Esta propuesta puede ser útil, por ejemplo, en ministerios de población o de salud, así como en el sector asegurador, donde se requiera ver con mayor detalle la tendencia para un cierto segmento de los datos.

- De igual manera, la técnica puede servir para estimar tendencias en situaciones en las que la calidad de la información sea notoriamente distinta por segmentos del recorrido de los datos.

# 5. CONCLUSIONES

- La propuesta es valiosa cuando se desea tomar alguna decisión basada en datos que están oscurecidos por fluctuaciones sin importancia, de forma que la estimación y visualización de su tendencia, al igual que de su variabilidad, son esenciales.

- La metodología tiene como propósito básico servir como herramienta de análisis exploratorio. Es de esperar que el analista efectúe posteriormente un análisis confirmatorio y que busque explicaciones para sus hallazgos, basadas en la teoría del fenómeno en estudio.

- Con la segmentación *a priori*, no se busca reducir la varianza del error de estimación.

- La segmentación empírica sí busca reducir dicha varianza y si no lo logra, se debe tomar ese hecho como una indicación de que no es apropiado segmentar.

# REFERENCIAS


- Box, G. E. P., Jenkins, G. W. & Reinsel, G. C. (1994). *Time Series Analysis. Forecasting and Control*, 3rd edition. Englewood Cliffs, N. J.: Prentice Hall.
- Cantoni, E. & Hastie, T. (2002). Degrees of freedom tests for smoothing splines. *Biometrika*, 89, 251-263.
- Carriere, J. (1992). Parametric models for life tables. *Transactions of the Society of Actuaries*, 44, 77-99.
- Fledelius, P., Guillen, M., Jens, P. N. & Petersen, K. (2004). A comparative study of parametric and non-parametric estimators of old-age mortality in Sweden. *Journal of Actuarial Practice*, 11, 101-126.
- Guerrero, V. M. (2008). Estimating trends with percentage of smoothness chosen by the user. *International Statistical Review*, 76, 187-202.
- Guerrero, V. M. & Silva, E. (2010). Non-parametric and structured graduation of mortality rates. *Population Review*, 49, 13-26.

# REFERENCIAS


- Helligman, L. & Pollard, J. H. (1980). The age pattern of mortality. *Journal of the Institute of Actuaries*, 107, 49-80.
- Hodrick, R. & Prescott, E. (1997). Post-war U.S. business cycles: An empirical investigation. *Journal of Money, Credit and Banking*, 29, 1-16.
- Lee, R. & Carter, L. (1992). Modeling and Forecasting U.S. Mortality. *Journal of the American Statistical Association*, 87, 659-675.
- Tuljapurkar, S. & Edwards, R. (2011). Variance in death and its implications for modeling and forecasting mortality. *Demographic research*, 21, 497-526.
- London, D. (1985). *GRADUATION: The Revision of Estimates*. Connecticut: ACTEX Publications, Winsted and Abington.
- Ruppert,. D., Wand, M. P. & Carroll, R. J. (2003). *Semiparametric Regression*. Cambridge: Cambridge University Press.

# REFERENCIAS


- Theil, H. (1963). On the use of incomplete prior information in regression analysis. *Journal of the American Statistical Association*, 58, 401-414.
- Thiele, P. (1871). On a mathematical formula to express the rate of mortality through the whole of life. *Journal of the Institute of Actuaries*, 16, 313-329.
- Wand, M. P. (1999). On the optimal amount of smoothing in penalized spline regression *Biometrika*, 86, 936-940.
- Whittaker, E. T. (1923). On a new method of graduation. *Proceedings of the Edinburgh Mathematical Society*, 41, 63-75.
- Zhang, C. (2003). Calibrating the degrees of freedom for automatic data smoothing and effective curve checking. *Journal of the American Statistical Association*, 98, 609-628.