



Modelos con Coeficientes Variando

Bladimir Morales Torrez

Julio 2022

Contents

1	Introducción	3
2	Modelo con Coeficientes Variando VCM	3
2.1	El modelo	3
2.2	Casos del modelo	3
2.3	Matricialmente	4
3	Estimación	4
3.1	Mínimos Cuadrados Penalizados	4
3.2	Verosimilitud Penalizada	5
3.3	Vector Score Penalizado	6
3.4	Matriz Hessiana Penalizada	6
3.5	Matriz de Información de Fisher Penalizada	7
3.6	Estimador de Máxima Verosimilitud Penalizada	8
3.7	Valores ajustados	9

3.8	Grados de Libertad Efectivos	9
3.9	Criterio de Información de Akaike	10
4	Inferencia	10
4.1	Esperanza de los estimadores	10
4.2	Matriz de Varianzas-Covarianzas	11
4.3	Matriz de Varianzas-Covarianzas Asintótica	11
5	Aplicación	11
5.1	Datos	11
5.2	Transformación de datos	11
5.3	Análisis exploratorio	11
5.4	Modelo de Coeficientes Variando VCM	13
5.5	Residuos	20
6	Conclusiones	25
	Bibliografía	26

1 Introducción

Los modelos aditivos generalizados (modelos no paramétricos), propuestos por Hastie & Tibshirani (1986) dan a conocer una nueva clase de modelos de regresión, extendiendo y flexibilizando el modelo de regresión clásico, reemplazando la función lineal por una función aditiva suave y no paramétricas, la cual es estimada por el algoritmo de puntuación local (local scoring algorithm).

Es así que, Hastie & Tibshirani (1993), proponen otra generalización al modelo de regresión lineal clásico, denominados modelos de coeficientes variando (Varying-Coefficient Models VCMs), los cuales contienen regresores lineales pero permiten que sus coeficientes cambien suavemente con el valor de otras variables, que se denominan “modificadores de efecto”. A partir de ahí existe bastante bibliografía e investigaciones sobre este tipo de modelos.

2 Modelo con Coeficientes Variando VCM

2.1 El modelo

Supongamos que se tiene una variable aleatoria Y cuya distribución depende de un parámetro η que será el predictor lineal y se relaciona con la media $\mu = \mathbb{E}(Y)$ mediante la función enlace $\eta = g(\mu)$, el modelo lineal generalizado con coeficientes variando (VCGLMs) se puede representar como:

$$\eta_i = \beta_0 + \mathbf{x}_i^{(1)} \beta_1(t_{1i}) + \dots + \mathbf{x}_i^{(s)} \beta_s(t_{si})$$
$$y_i = \eta_i + \varepsilon_i \quad (i = 1, \dots, n), \quad (1)$$

Donde las variables explicativas t_{k_i} cambian el comportamiento de los coeficientes de las covariables $\mathbf{x}_i^{(k)}$ a través de las funciones $\beta_k(\cdot)$ para $k = 1, \dots, s$. La dependencia de β_k en t_{k_i} implica un tipo de interacción entre t_{k_i} y $\mathbf{x}_i^{(k)}$. Las variables t_{k_i} pueden no ser tan diferentes a $\mathbf{x}_i^{(k)}$, pero también puede ser variables como el tiempo.

Si se toma el modelo Gaussiano, donde $g(\mu) = \mu$ y la variable aleatoria y_i tiene distribución normal con media η_i , el modelo (1) es de la forma

$$y_i = \beta_0 + \mathbf{x}_i^{(1)} \beta_1(t_{1i}) + \dots + \mathbf{x}_i^{(s)} \beta_s(t_{si}) + \varepsilon_i \quad (2)$$

donde ε_i es el error aleatorio con $\mathbb{E}(\varepsilon) = 0$ y $Var(\varepsilon) = \sigma^2$

2.2 Casos del modelo

El modelo de coeficientes variando generaliza los siguientes tipos de modelos:

- Si $\beta_k(t_{k_i}) = \beta_k$ es una función constante, entonces ese término es lineal en $\mathbf{x}_i^{(k)}$. Si todos los términos son lineales se reduce a un modelo lineal generalizado clásico.
- Si $\mathbf{x}_i^{(k)} = c$ con c constante entonces el k -ésimo término es $\beta_k(t_{k_i})$ una función no conocida. Si todos los términos tienen esta característica se tiene un modelo aditivo generalizado.
- Una función lineal $\beta_k(t_{k_i}) = \beta_k t_{k_i}$ produce una interacción de la forma $\beta_k \mathbf{x}_i^{(k)} t_{k_i}$.
- En el término del modelo $\mathbf{x}_i^{(k)} \beta_k(t_{k_i})$, cuando \mathbf{x}_i es una variable binaria (0 – 1). Supongamos que hay un término $\beta_0(t_{0_i})$ en el modelo. Esto equivale a tener una curva separada correspondiente a cada uno de los dos valores de \mathbf{x}_i .
- A menudo las t_{k_i} serán la misma variable, como puede ser la edad o el tiempo, que podría modificar los

efectos de $x_i^{(k)}$. Supongamos que, los datos consisten en mediciones repetidas de las variables $y_i, x_i^{(1)}, \dots, x_i^{(s)}$ sobre n puntos de tiempo $t \in (t_1, \dots, t_n)$. Entonces se podría modelar esto a lo largo del tiempo

$$\eta_t = \beta_0(t_i) + x_i^{(1)}(t_{1_i})\beta_1(t_{1_i}) + \dots + x_i^{(s)}(t_{s_i})\beta_s(t_{s_i})$$

Que es llamado como *modelo lineal generalizado dinámico o condicionalmente paramétrico*

f) La variable modificante t_{k_i} puede ser la misma que $x_i^{(k)}$, por simplicidad suponemos que es un modelo lineal normal con un solo término.

$$y_i = x_i \beta(x_i) + \varepsilon_i$$

Este es un modelo común de suavizamiento o regresión no paramétrica de y versus x .

g) Cada t_{k_i} puede tener un valor escalar o vectorial.

h) En todos los casos anteriores, hay muchas formas de modelar las funciones $\beta_k(t_{k_i})$. Por ejemplo, podríamos usar representaciones paramétricas flexibles tales como polinomios, series de Fourier o polinomios por partes. De manera más general se pueden utilizar funciones no paramétricas, mediante el uso de métodos kernel, penalización o formulación bayesiana estocástica.

2.3 Matricialmente

El modelo (1) puede ser representado matricialmente como:

$$\mathbf{y} = \tilde{\mathbf{N}}_1 \beta_1 + \dots + \tilde{\mathbf{N}}_s \beta_s + \varepsilon, \quad (3)$$

donde \mathbf{y} es un vector aleatorio de respuestas observadas ($n \times 1$), $\tilde{\mathbf{N}}_k = \mathbf{X}^{(k)} \mathbf{N}_k$, $\mathbf{X}^{(k)} = \text{diag}_{1 \leq i \leq n}(x_i^{(k)})$ y \mathbf{N}_k es la matriz de incidencia de ($n \times r$) con el (i, l) -ésimo elemento igual al indicador $I(t_{k_i} = t_{k_l}^0)$, con $t_{k_l}^0$ que denota los valores distintos y ordenados de la variable explicativa t_{k_i} y $\beta_k = (\psi_{k_1}, \dots, \psi_{k_r})^T$ es un $(r_k \times 1)$ un vector de parámetros con $\psi_{k_l} = \beta_k(t_{k_l}^0)$, para $l = 1, \dots, r_k$.

3 Estimación

El modelo (1), es general para la mayoría de las aplicaciones, ya que no se imponen restricciones a las funciones de los coeficientes. La estimación no paramétrica sin restricciones de estas funciones probablemente no sería posible, salvo en el caso de diseños especiales. En el caso de los datos observacionales, los t_k pueden tomar valores diferentes para cada observación, por esta razón, se impone restricciones a las funciones de los coeficientes variando, por ejemplo, a trozos con forma paramétrica conocida o funciones suaves pero no paramétricas. Una aproximación puede ser a través de bases paramétricas como funciones polinómicas o trigonométricas. Normalmente, éstas no proporcionan suficiente flexibilidad y adaptabilidad local, por eso es preferible un conjunto de bases splines. Para la estimación se procede igual que con un modelo lineal, sólo que con varias variables definidas por los productos de cada $x_i^{(k)}$, y $\beta_k(t_{k_i})$. Así, las herramientas inferenciales estándar sirven para evaluar conjuntos de coeficientes, puntos de influencia, etc. Las características de las curvas ajustadas pueden ser diferentes con pequeños cambios en las posiciones de los nudos, sobre todo si sólo se tienen pocos.

En este trabajo se mostrará el procedimiento para el modelo (2) ya que para el modelo (1) se incluyen procedimientos diferentes como el algoritmo de tipo NewtonRaphson.

3.1 Mínimos Cuadrados Penalizados

Para estimar β_k por minimización se toma en cuenta la siguiente esperanza:

$$\mathbb{E} \left\{ y_i - \sum_{j=1}^s x_i^j \beta_j(t_{j_i}) \right\}^2$$

Se condiciona en t_{k_i} , es una condición suficiente para la solución:

$$\mathbb{E} \left\{ \mathbf{x}_i^k \left[y_i - \sum_{j=1}^s \mathbf{x}_i^j \beta_j(t_{j_i}) \right] \mid t_{k_i} \right\} = 0 \quad k = 1, \dots, s$$

Así se encuentra $\beta_k(\cdot)$, reordenando la ecuación anterior y resolver:

$$\beta_k(t_{k_i}) = \frac{\mathbb{E} \left\{ \mathbf{x}_i^{(k)} \left[y_i - \sum_{j=1, j \neq k}^s \mathbf{x}_i^{(j)} \beta_j(t_{j_i}) \right] \mid t_{k_i} \right\}}{\mathbb{E} \{ \mathbf{x}^{(k)^2} \mid t_{k_i} \}}$$

$$\beta_k(t_{k_i}) = \frac{\mathbb{E} \left\{ \mathbf{x}_i^{(k)^2} \left[y_i - \sum_{j=1, j \neq k}^s \mathbf{x}_i^{(j)} \beta_j(t_{j_i}) \right] / \mathbf{x}_i^{(k)} \mid t_{k_i} \right\}}{\mathbb{E} \{ \mathbf{x}^{(k)^2} \mid t_{k_i} \}}$$

La anterior ecuación es una relación de dos esperanzas condicionales y puede verse como una media ponderada condicional, donde las ponderaciones condicionales vienen dadas por el término $\mathbf{x}_i^{(k)^2}$, el término en el denominador garantiza que las ponderaciones se integren en 1. Existe una ecuación similar para cada función β_k , y el conjunto de k ecuaciones debe resolver simultáneamente la función β_k , como una estimación flexible de una esperanza condicional, esto sugiere que cada función β_k puede estimarse de forma iterativa “de uno en uno” mediante el alisado $\frac{y_i - \sum_{j=1, j \neq k}^s \mathbf{x}_i^{(j)} \beta_j(t_{j_i})}{\mathbf{x}_i^{(k)^2}}$ en t_k , con pesos $\mathbf{x}_i^{(k)^2}$.

Para la estimación de β_k se propone minimizar la suma de cuadrados de los errores penalizados

$$\sum_{i=1}^n \left\{ y_i - \sum_{j=1}^s \mathbf{x}_i^{(j)} \beta_j(t_{j_i}) \right\}^2 + \sum_{j=1}^s \lambda_j \int \beta_j''(t_{j_i})^2 dt_j \quad (4)$$

El primer término mide la bondad del ajuste y el segundo penaliza la rugosidad de cada β_k con un parámetro fijo λ_j .

De manera matricial se define como:

$$\left(\mathbf{y} - \sum_{j=1}^s \tilde{\mathbf{N}}_j \beta_j \right)^\top \left(\mathbf{y} - \sum_{j=1}^s \tilde{\mathbf{N}}_j \beta_j \right) - \sum_{j=1}^s \lambda_j \beta_j^\top \mathbf{K}_j \beta_j \quad (5)$$

donde \mathbf{K} es la matriz definida positiva de $(r \times r)$ que depende de las matrices tridiagonales \mathbf{Q} y \mathbf{R}

3.2 Verosimilitud Penalizada

Bajo el modelo de coeficientes variando de la ecuación (2) se tiene que el log. de verosimilitud de $\mathbf{y}_i \sim (\sum_{j=1}^s \mathbf{x}_i^{(j)} \beta_j(t_{j_i}); \sigma^2)$ es:

$$\ell(\boldsymbol{\theta}) = -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \left\{ y_i - \sum_{j=1}^s \mathbf{x}_i^{(j)} \beta_j(t_{j_i}) \right\}^2 \quad (6)$$

donde $\boldsymbol{\theta} = (\beta_1^\top, \dots, \beta_s^\top, \sigma^2)^\top$ es el vector de parámetros a estimar. Matricialmente se tiene:

$$\ell(\boldsymbol{\theta}) = -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \left(\mathbf{y} - \sum_{j=1}^s \tilde{\mathbf{N}}_j \beta_j \right)^\top \left(\mathbf{y} - \sum_{j=1}^s \tilde{\mathbf{N}}_j \beta_j \right) \quad (7)$$

Luego se tiene la función de log. verosimilitud penalizada que esta definida por:

$$\ell_p(\boldsymbol{\theta}, \lambda) = \ell(\boldsymbol{\theta}) - \sum_{j=1}^s \frac{\lambda_j}{2} \boldsymbol{\beta}_j^\top \mathbf{K}_j \boldsymbol{\beta}_j$$

3.3 Vector Score Penalizado

Considerando λ_k fijo el vector de Score penalizado esta dada por:

$$\mathbf{U}_p(\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{U}_p^{\boldsymbol{\beta}_k}(\boldsymbol{\theta}) \\ \mathbf{U}_p^{\sigma^2}(\boldsymbol{\theta}) \end{pmatrix} = \begin{pmatrix} \frac{\partial \ell_p(\boldsymbol{\theta}, \lambda)}{\partial \boldsymbol{\beta}_k} \\ \frac{\partial \ell_p(\boldsymbol{\theta}, \lambda)}{\partial \sigma^2} \end{pmatrix} \quad (8)$$

Así se tiene:

$$\begin{aligned} \frac{\partial \ell_p(\boldsymbol{\theta}, \lambda)}{\partial \boldsymbol{\beta}_k} &= \frac{\partial}{\partial \boldsymbol{\beta}_k} \left[-\frac{1}{2\sigma^2} \left(\mathbf{y} - \sum_{j=1}^s \tilde{\mathbf{N}}_j \boldsymbol{\beta}_j \right)^\top \left(\mathbf{y} - \sum_{j=1}^s \tilde{\mathbf{N}}_j \boldsymbol{\beta}_j \right) - \sum_{j=1}^s \frac{\lambda_j}{2} \boldsymbol{\beta}_j^\top \mathbf{K}_j \boldsymbol{\beta}_j \right] \\ &= \frac{1}{\sigma^2} \tilde{\mathbf{N}}_k^\top \left(\mathbf{y} - \tilde{\mathbf{N}}_k \boldsymbol{\beta}_k - \sum_{j=1; j \neq k}^s \tilde{\mathbf{N}}_j \boldsymbol{\beta}_j \right) - \lambda_k \mathbf{K}_k \boldsymbol{\beta}_k \end{aligned} \quad (9)$$

$$\begin{aligned} \frac{\partial \ell_p(\boldsymbol{\theta}, \lambda)}{\partial \sigma^2} &= \frac{\partial}{\partial \sigma^2} \left[-\frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \left(\mathbf{y} - \sum_{j=1}^s \tilde{\mathbf{N}}_j \boldsymbol{\beta}_j \right)^\top \left(\mathbf{y} - \sum_{j=1}^s \tilde{\mathbf{N}}_j \boldsymbol{\beta}_j \right) \right] \\ &= -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \left(\mathbf{y} - \sum_{j=1}^s \tilde{\mathbf{N}}_j \boldsymbol{\beta}_j \right)^\top \left(\mathbf{y} - \sum_{j=1}^s \tilde{\mathbf{N}}_j \boldsymbol{\beta}_j \right) \end{aligned} \quad (10)$$

3.4 Matriz Hessiana Penalizada

Esta definida como las segundas derivadas del vector de Score.

$$\frac{\partial^2 \ell_p(\boldsymbol{\theta}, \lambda)}{\partial \boldsymbol{\beta}_k \partial \boldsymbol{\beta}_k^\top} = \frac{\partial}{\partial \boldsymbol{\beta}_k^\top} \left[\frac{1}{\sigma^2} \tilde{\mathbf{N}}_k^\top \left(\mathbf{y} - \tilde{\mathbf{N}}_k \boldsymbol{\beta}_k - \sum_{j=1; j \neq k}^s \tilde{\mathbf{N}}_j \boldsymbol{\beta}_j \right) - \lambda_k \mathbf{K}_k \boldsymbol{\beta}_k \right]$$

* Si $k = k'$

$$\frac{\partial^2 \ell_p(\boldsymbol{\theta}, \lambda)}{\partial \boldsymbol{\beta}_k \partial \boldsymbol{\beta}_k^\top} = -\frac{1}{\sigma^2} \tilde{\mathbf{N}}_k^\top \tilde{\mathbf{N}}_k - \lambda_k \mathbf{K}_k \quad (11)$$

* Si $k \neq k'$

$$\frac{\partial^2 \ell_p(\boldsymbol{\theta}, \lambda)}{\partial \boldsymbol{\beta}_k \partial \boldsymbol{\beta}_k^\top} = -\frac{1}{\sigma^2} \tilde{\mathbf{N}}_k^\top \tilde{\mathbf{N}}_k \quad (12)$$

$$\begin{aligned}
\frac{\partial^2 \ell_p(\boldsymbol{\theta}, \lambda)}{\partial \sigma^2 \partial \boldsymbol{\beta}_k^\top} &= \frac{\partial}{\partial \sigma^2} \left[\frac{1}{\sigma^2} \tilde{\mathbf{N}}_k^\top \left(\mathbf{y} - \tilde{\mathbf{N}}_k \boldsymbol{\beta}_k - \sum_{j=1; j \neq k}^s \tilde{\mathbf{N}}_j \boldsymbol{\beta}_j \right) - \lambda_k \mathbf{K}_k \boldsymbol{\beta}_k \right] \\
&= -\frac{1}{\sigma^4} \tilde{\mathbf{N}}_k^\top \left(\mathbf{y} - \tilde{\mathbf{N}}_k \boldsymbol{\beta}_k - \sum_{j=1; j \neq k}^s \tilde{\mathbf{N}}_j \boldsymbol{\beta}_j \right)
\end{aligned} \tag{13}$$

$$\begin{aligned}
\frac{\partial^2 \ell_p(\boldsymbol{\theta}, \lambda)}{\partial \sigma^2 \partial \sigma^2} &= \frac{\partial}{\partial \sigma^2} \left[-\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \left(\mathbf{y} - \sum_{j=1}^s \tilde{\mathbf{N}}_j \boldsymbol{\beta}_j \right)^\top \left(\mathbf{y} - \sum_{j=1}^s \tilde{\mathbf{N}}_j \boldsymbol{\beta}_j \right) \right] \\
&= \frac{n}{2\sigma^4} - \frac{1}{2\sigma^6} \left(\mathbf{y} - \sum_{j=1}^s \tilde{\mathbf{N}}_j \boldsymbol{\beta}_j \right)^\top \left(\mathbf{y} - \sum_{j=1}^s \tilde{\mathbf{N}}_j \boldsymbol{\beta}_j \right)
\end{aligned} \tag{14}$$

3.5 Matriz de Información de Fisher Penalizada

Se define a la matriz de información de Fisher penalizada como

$$\mathcal{I}_p(\boldsymbol{\theta}) = -\mathbb{E} \left\{ \frac{\partial^2 \ell_p(\boldsymbol{\theta}, \lambda)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} \right\}$$

donde:

$$\mathcal{I}_p(\boldsymbol{\theta}) = \begin{pmatrix} \mathcal{I}_p^{\boldsymbol{\beta}_1 \boldsymbol{\beta}_1^\top}(\boldsymbol{\theta}) & \dots & \mathcal{I}_p^{\boldsymbol{\beta}_1 \boldsymbol{\beta}_s^\top}(\boldsymbol{\theta}) & \mathcal{I}_p^{\boldsymbol{\beta}_1 \sigma^2}(\boldsymbol{\theta}) \\ \vdots & \ddots & \vdots & \vdots \\ \mathcal{I}_p^{\boldsymbol{\beta}_s \boldsymbol{\beta}_1^\top}(\boldsymbol{\theta}) & \dots & \mathcal{I}_p^{\boldsymbol{\beta}_s \boldsymbol{\beta}_s^\top}(\boldsymbol{\theta}) & \mathcal{I}_p^{\boldsymbol{\beta}_s \sigma^2}(\boldsymbol{\theta}) \\ \mathcal{I}_p^{\sigma^2 \boldsymbol{\beta}_1}(\boldsymbol{\theta}) & \dots & \mathcal{I}_p^{\sigma^2 \boldsymbol{\beta}_s}(\boldsymbol{\theta}) & \mathcal{I}_p^{\sigma^2 \sigma^2}(\boldsymbol{\theta}) \end{pmatrix}$$

Es así que aplicamos a (11), (12), (13) y (14) se tiene:

- Para $k = k'$

$$\begin{aligned}
\mathcal{I}_p^{\boldsymbol{\beta}_k \boldsymbol{\beta}_k^\top}(\boldsymbol{\theta}) &= -\mathbb{E} \left\{ -\frac{1}{\sigma^2} \tilde{\mathbf{N}}_k^\top \tilde{\mathbf{N}}_k - \lambda_k \mathbf{K}_k \right\} \\
&= \frac{1}{\sigma^2} \tilde{\mathbf{N}}_k^\top \tilde{\mathbf{N}}_k + \lambda_k \mathbf{K}_k
\end{aligned} \tag{15}$$

* Para $k \neq k'$

$$\begin{aligned}
\mathcal{I}_p^{\boldsymbol{\beta}_k \boldsymbol{\beta}_{k'}^\top}(\boldsymbol{\theta}) &= -\mathbb{E} \left\{ -\frac{1}{\sigma^2} \tilde{\mathbf{N}}_k^\top \tilde{\mathbf{N}}_{k'} \right\} \\
&= \frac{1}{\sigma^2} \tilde{\mathbf{N}}_k^\top \tilde{\mathbf{N}}_{k'}
\end{aligned} \tag{16}$$

$$\begin{aligned}
\mathcal{I}_p^{\beta_k \sigma^2}(\boldsymbol{\theta}) &= -\mathbb{E} \left\{ -\frac{1}{\sigma^4} \tilde{\mathbf{N}}_k^\top \left(\mathbf{y} - \tilde{\mathbf{N}}_k \beta_k - \sum_{j=1; j \neq k}^s \tilde{\mathbf{N}}_j \beta_j \right) \right\} \\
&= \frac{1}{\sigma^4} \tilde{\mathbf{N}}_k^\top \left(\mathbb{E}[\mathbf{y}] - \sum_{j=1}^s \tilde{\mathbf{N}}_j \beta_j \right) \\
&= \mathbf{0}_{(\mathbf{r} \times \mathbf{1})}
\end{aligned} \tag{17}$$

$$\begin{aligned}
\mathcal{I}_p^{\sigma^2 \sigma^2}(\boldsymbol{\theta}) &= -\mathbb{E} \left\{ \frac{n}{2\sigma^4} - \frac{1}{2\sigma^6} \left(\mathbf{y} - \sum_{j=1}^s \tilde{\mathbf{N}}_j \beta_j \right)^\top \left(\mathbf{y} - \sum_{j=1}^s \tilde{\mathbf{N}}_j \beta_j \right) \right\} \\
&= \frac{n}{2\sigma^4}
\end{aligned}$$

3.6 Estimador de Máxima Verosimilitud Penalizada

Si consideramos λ y σ^2 fijos, para hallar el estimador de máxima verosimilitud penalizada se debe igualar el vector score penalizado a cero.

Así se tiene para β_k :

$$\mathbf{U}_p^{\beta_k}(\boldsymbol{\theta}) = \frac{\partial \ell_p(\boldsymbol{\theta}, \lambda)}{\partial \beta_k} = 0$$

Entonces de (9):

$$\frac{1}{\sigma^2} \tilde{\mathbf{N}}_k^\top \left(\mathbf{y} - \tilde{\mathbf{N}}_k \beta_k - \sum_{j=1; j \neq k}^s \tilde{\mathbf{N}}_j \beta_j \right) - \lambda_k \mathbf{K}_k \beta_k = 0$$

luego,

$$\tilde{\mathbf{N}}_k^\top \left(\mathbf{y} - \sum_{j=1; j \neq k}^s \tilde{\mathbf{N}}_j \beta_j \right) = (\tilde{\mathbf{N}}_k^\top \tilde{\mathbf{N}}_k + \lambda_k \mathbf{K}_k) \beta_k$$

Finalmente se tiene:

$$\hat{\beta}_k = (\tilde{\mathbf{N}}_k^\top \tilde{\mathbf{N}}_k + \lambda_k \mathbf{K}_k)^{-1} \tilde{\mathbf{N}}_k^\top \left(\mathbf{y} - \sum_{j=1; j \neq k}^s \tilde{\mathbf{N}}_j \beta_j \right) \tag{18}$$

$$\tag{19}$$

Para finalizar la estimación se puede utilizar algún método de estimación como el Algoritmo Scoring Fisher o el Algoritmo Backfitting.

Para σ^2 , se tiene:

$$\mathbf{U}_p^{\sigma^2}(\boldsymbol{\theta}) = \frac{\partial \ell_p(\boldsymbol{\theta}, \lambda)}{\partial \sigma^2} = 0$$

Entonces de (9):

$$-\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \left(\mathbf{y} - \sum_{j=1}^s \tilde{\mathbf{N}}_j \beta_j \right)^\top \left(\mathbf{y} - \sum_{j=1}^s \tilde{\mathbf{N}}_j \beta_j \right) = 0$$

luego,

$$\frac{1}{2\sigma^4} \left[-n\sigma^2 + \left(\mathbf{y} - \sum_{j=1}^s \tilde{\mathbf{N}}_j \beta_j \right)^\top \left(\mathbf{y} - \sum_{j=1}^s \tilde{\mathbf{N}}_j \beta_j \right) \right] = 0$$

finalmente,

$$\hat{\sigma}^2 = \frac{1}{n} \left(\mathbf{y} - \sum_{j=1}^s \tilde{\mathbf{N}}_j \hat{\beta}_j \right)^\top \left(\mathbf{y} - \sum_{j=1}^s \tilde{\mathbf{N}}_j \hat{\beta}_j \right) \quad (20)$$

3.7 Valores ajustados

Dado el vector $\hat{\mathbf{y}} = (\hat{y}_1, \dots, \hat{y}_n)^\top$ se define los valores ajustados

$$\begin{aligned} \hat{\mathbf{y}} &= \tilde{\mathbf{N}}_k \hat{\beta}_n \\ &= \tilde{\mathbf{N}}_k (\tilde{\mathbf{N}}_k^\top \tilde{\mathbf{N}}_k + \lambda_k \mathbf{K}_k)^{-1} \tilde{\mathbf{N}}_k^\top \left(\mathbf{y} - \sum_{j=1; j \neq k}^s \tilde{\mathbf{N}}_j \beta_j \right) \\ &= \mathbf{H}(\lambda) \left(\mathbf{y} - \sum_{j=1; j \neq k}^s \tilde{\mathbf{N}}_j \beta_j \right) \end{aligned}$$

donde $\mathbf{H}(\lambda) = \tilde{\mathbf{N}}_k (\tilde{\mathbf{N}}_k^\top \tilde{\mathbf{N}}_k + \lambda_k \mathbf{K}_k)^{-1} \tilde{\mathbf{N}}_k^\top$ es la matriz de proyección de dimensión $(n \times n)$, que no es idempotente.

3.8 Grados de Libertad Efectivos

Indican el número efectivo de parámetros considerados en el modelo, definido por:

$$GLE(\lambda) = \text{traza}\{\mathbf{H}(\lambda)\}$$

Observaciones:

- * Si $\lambda \rightarrow 0$ se está priorizando el ajuste del modelo con la interpolación de las n observaciones, $GLE(\lambda) \rightarrow n$.
- * Si $\lambda \rightarrow \infty$ el modelo equivale al modelo lineal clásico, $GLE(\lambda) \rightarrow s$.
- * El dominio de GLE se definen en $[s, n]$

Para el VCM se tiene:

$$\begin{aligned}
GLE(\lambda_k) &= \text{traza}\{\mathbf{H}(\lambda_k)\} \\
&= \text{traza}\left\{\tilde{\mathbf{N}}_k (\tilde{\mathbf{N}}_k^\top \tilde{\mathbf{N}}_k + \lambda_k \mathbf{K}_k)^{-1} \tilde{\mathbf{N}}_k^\top\right\} \\
&= \text{traza}\left\{\tilde{\mathbf{N}}_k^\top \tilde{\mathbf{N}}_k (\tilde{\mathbf{N}}_k^\top \tilde{\mathbf{N}}_k + \lambda_k \mathbf{K}_k)^{-1}\right\}
\end{aligned}$$

Sea $\mathbf{B}_k = \tilde{\mathbf{N}}_k^\top \tilde{\mathbf{N}}_k$ y $\mathbf{L}_k = \mathbf{B}_k^{-1/2} \mathbf{K}_k \mathbf{B}_k^{-1/2}$. Entonces se tiene:

$$\begin{aligned}
GLE(\lambda_k) &= \text{traza}\left\{\mathbf{B}_k (\mathbf{B}_k + \lambda_k \mathbf{K}_k)^{-1}\right\} \\
&= \text{traza}\left\{\mathbf{B}_k^{1/2} \mathbf{B}_k^{1/2} (\mathbf{B}_k^{1/2} \mathbf{B}_k^{1/2} + \lambda_k \mathbf{B}_k^{1/2} \mathbf{L}_k \mathbf{B}_k^{1/2})^{-1}\right\} \\
&= \text{traza}\left\{\mathbf{B}_k^{1/2} \mathbf{B}_k^{1/2} \mathbf{B}_k^{-1/2} (\mathbf{I}_r + \lambda_k \mathbf{L}_k)^{-1} \mathbf{B}_k^{-1/2}\right\} \\
&= \text{traza}\left\{\mathbf{B}_k^{1/2} \mathbf{B}_k^{-1/2} \mathbf{B}_k^{1/2} \mathbf{B}_k^{-1/2} (\mathbf{I}_r + \lambda_k \mathbf{L}_k)^{-1}\right\} \\
&= \text{traza}\left\{(\mathbf{I}_r + \lambda_k \mathbf{L}_k)^{-1}\right\} \\
&= \sum_{j=1}^r \frac{1}{1 + \lambda_k L_{k_j}}
\end{aligned}$$

3.9 Criterio de Información de Akaike

Es un criterio de bondad de ajuste para calcular un modelo agrupado, que consiste en minimizar la función:

$$AIC(\lambda) = -2\ell_p(\boldsymbol{\theta}, \lambda) + 2\{s + 1 + GLE(\lambda)\}$$

4 Inferencia

4.1 Esperanza de los estimadores

Se tiene de (18) que $\hat{\beta}_k = (\tilde{\mathbf{N}}_k^\top \tilde{\mathbf{N}}_k + \lambda_k \mathbf{K}_k)^{-1} \tilde{\mathbf{N}}_k^\top (\mathbf{y} - \sum_{j=1; j \neq k}^s \tilde{\mathbf{N}}_j \beta_j)$, donde $\mathbb{E}[\mathbf{y}] = \sum_{j=1}^s \tilde{\mathbf{N}}_j \beta_j$, así,

$$\begin{aligned}
\mathbb{E}[\hat{\beta}_k] &= \mathbb{E}\left[(\tilde{\mathbf{N}}_k^\top \tilde{\mathbf{N}}_k + \lambda_k \mathbf{K}_k)^{-1} \tilde{\mathbf{N}}_k^\top \left(\mathbf{y} - \sum_{j=1; j \neq k}^s \tilde{\mathbf{N}}_j \beta_j\right)\right] \\
&= (\tilde{\mathbf{N}}_k^\top \tilde{\mathbf{N}}_k + \lambda_k \mathbf{K}_k)^{-1} \tilde{\mathbf{N}}_k^\top \left(\mathbb{E}[\mathbf{y}] - \sum_{j=1; j \neq k}^s \tilde{\mathbf{N}}_j \beta_j\right) \\
&= (\tilde{\mathbf{N}}_k^\top \tilde{\mathbf{N}}_k + \lambda_k \mathbf{K}_k)^{-1} \tilde{\mathbf{N}}_k^\top \left(\sum_{j=1}^s \tilde{\mathbf{N}}_j \beta_j - \sum_{j=1; j \neq k}^s \tilde{\mathbf{N}}_j \beta_j\right) \\
&= (\tilde{\mathbf{N}}_k^\top \tilde{\mathbf{N}}_k + \lambda_k \mathbf{K}_k)^{-1} \tilde{\mathbf{N}}_k^\top \tilde{\mathbf{N}}_k \beta_k
\end{aligned} \tag{21}$$

Es un estimador sesgado, sin embargo es posible escoger el λ_k para disminuir el sesgo de los estimadores. *Observación.* Si $\lambda_k \rightarrow 0$ se tendrá un estimador insesgado $\mathbb{E}(\hat{\beta}_k) \rightarrow \beta$

4.2 Matriz de Varianzas-Covarianzas

Se tiene que $Var[\mathbf{y}] = \sigma^2 \mathbf{I}$, así,

$$Var[\hat{\beta}_k] = Var \left[(\tilde{\mathbf{N}}_k^\top \tilde{\mathbf{N}}_k + \lambda_k \mathbf{K}_k)^{-1} \tilde{\mathbf{N}}_k^\top \left(\mathbf{y} - \sum_{j=1; j \neq k}^s \tilde{\mathbf{N}}_j \beta_j \right) \right]$$

4.3 Matriz de Varianzas-Covarianzas Asintótica

La matriz de Varianzas-Covarianzas Asintótica de los estimadores del método de máxima verosimilitud penalizada están dados por la inversa de la matriz de información de Fisher penalizada,

$$\hat{Cov}(\hat{\theta}) \approx \mathcal{I}_p^{-1}(\hat{\theta})$$

5 Aplicación

5.1 Datos

Se utilizará los datos presentados en el libro de Harezlak et al. (2018) de los rendimientos diarios de las acciones de la General Electric Company y el índice de Standard & Poor's 500 (S&P 500). Los datos están en el `data.frame` `capm` en el paquete `HRW`. La variable `Date` se tiene desde el 1 de noviembre de 1993 al 31 de marzo de 2003, por lo que hay más de nueve años de datos.

Mencionar también que existe ejemplos detallados de este modelo en Ruppert et al. (2003).

5.2 Transformación de datos

Si P_t es el precio de una acción en el día t , entonces el rendimiento logarítmico de ese día es $\log\left(\frac{P_t}{P_{t_1}}\right)$. El exceso de rendimiento logarítmico es el rendimiento logarítmico menos la tasa de interés libre de riesgo, generalmente considerada como la tasa de facturación del Tesoro a corto plazo.

El siguiente código calcula el rendimiento logarítmico excedente tanto de las acciones de General Electric Company como del índice S&P 500:

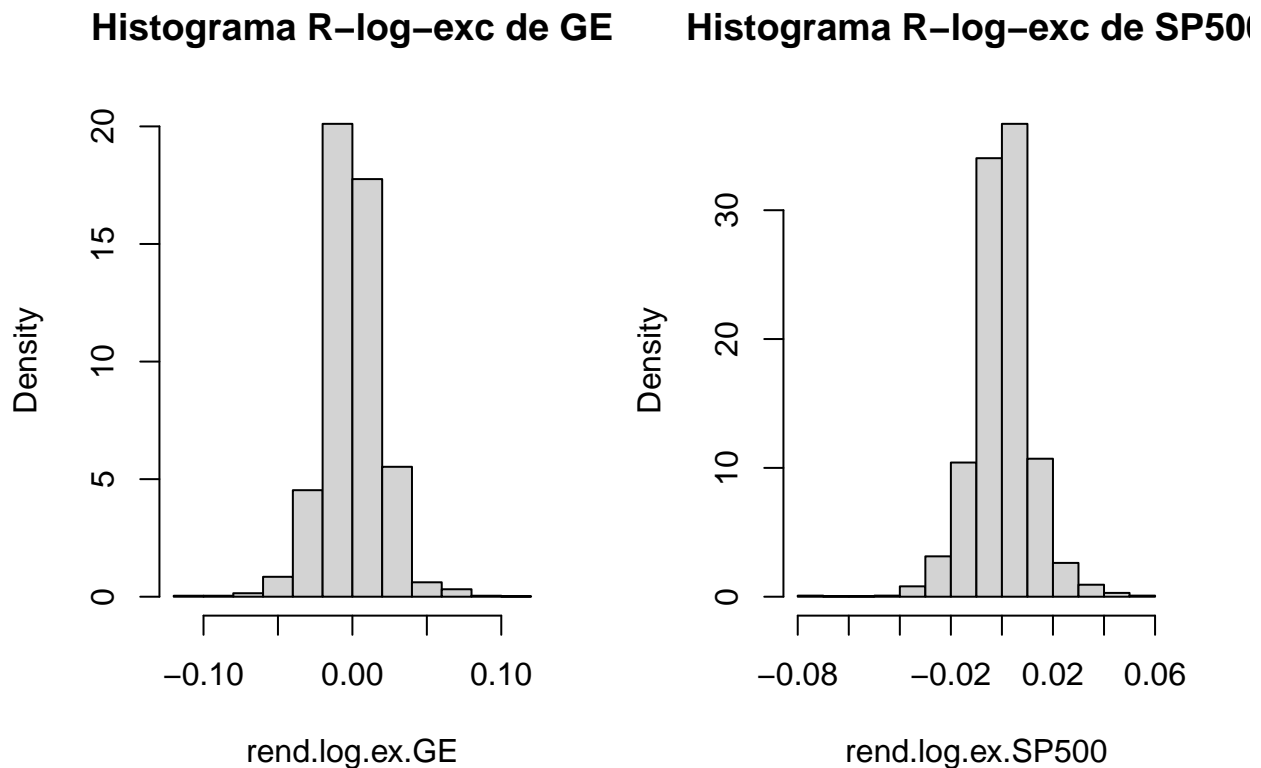
```
library(HRW)

## Warning: package 'HRW' was built under R version 4.1.3

data(capm)
n <- dim(capm)[1]
#riesgo.libre=riskfree
riesgo.libre <- capm$Close.tbill[2:n]/(100*365)
#elrGE=rend.log.ex.GE
rend.log.ex.GE <- diff(log(capm$Close.ge))-riesgo.libre
rend.log.ex.SP500 <- diff(log(capm$Close.sp500))-riesgo.libre
```

5.3 Análisis exploratorio

```
par(mfrow=c(1,2))
hist(rend.log.ex.GE,freq = FALSE,main="Histograma R-log-exc de GE")
hist(rend.log.ex.SP500,freq = FALSE,main="Histograma R-log-exc de SP500")
```



```
dev.off
```

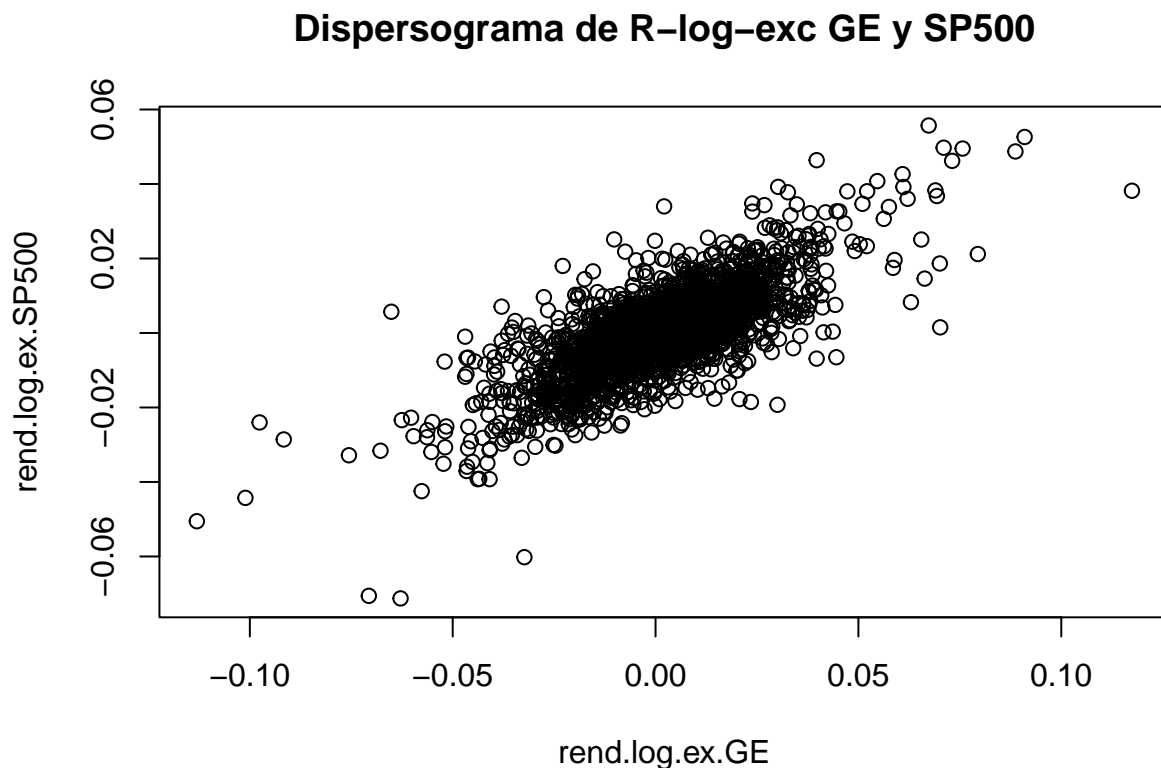
```
## function (which = dev.cur())
## {
##   if (which == 1)
##     stop("cannot shut down device 1 (the null device)")
##   .External(C_devoff, as.integer(which))
##   dev.cur()
## }
## <bytecode: 0x00000000136922d0>
## <environment: namespace:grDevices>
```

Donde `capm$Close.sp500` y `capm$Close.ge` son los precios de cierre diarios y `capm$Close.tbill` es la tasa diaria de facturación del Tesoro expresada como porcentaje. En el código, `capm$Close.tbill` primero se divide por 100 para convertir de un porcentaje a una fracción y luego se divide por 365 para convertir de una tasa anual a una tasa diaria. Por eso, `rend.log.ex.GE` y `rend.log.ex.SP500` son los rendimientos logarítmicos excedentes diarios de las acciones de General Electric y en el índice S&P 500, respectivamente.

5.4 Modelo de Coeficientes Variando VCM

Uno de los supuestos del modelo de fijación de precios de activos de capital en finanzas es que los rendimientos logarítmicos excedentes medios de una acción dependen linealmente de los rendimientos logarítmicos excedentes del mercado. El rendimiento del índice S&P 500 será utilizado como aproximación de la rentabilidad del mercado. Por lo general, el modelo de regresión lineal simple, se ajusta solo a datos recientes, ya que se espera que la pendiente cambie lentamente a lo largo del tiempo. Una alternativa es ajustar un modelo de coeficiente variando a todos los datos, con el intercepto y pendiente, que dependa de la fecha. Hacer esto proporcionaría evidencia como y con qué rapidez está cambiando la pendiente.

```
plot(rend.log.ex.GE,rend.log.ex.SP500,main="Dispersograma de R-log-exc GE y SP500")
```



La teoría económica predice que el intercepto es cero, por lo que se espera que el intercepto estimado sea no significativo. Primero, ajustamos un modelo de regresión lineal simple a todos los datos usando:

```
library(mgcv)
```

```
## Warning: package 'mgcv' was built under R version 4.1.3
```

```
## Loading required package: nlme
```

```
## This is mgcv 1.8-40. For overview type 'help("mgcv-package")'.
```

```

#?gam
fitSLR <- gam(rend.log.ex.GE ~ rend.log.ex.SP500)
summary(fitSLR)

##
## Family: gaussian
## Link function: identity
##
## Formula:
## rend.log.ex.GE ~ rend.log.ex.SP500
##
## Parametric coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.0002971  0.0002628   1.131   0.258
## rend.log.ex.SP500 1.2441346  0.0227404  54.710 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## R-sq.(adj) =  0.559   Deviance explained = 55.9%
## GCV = 0.00016324   Scale est. = 0.0001631   n = 2362

```

```
fitSLR$aic
```

```
## [1] -13892.27
```

Para verificar si un modelo lineal es apropiado, ajustamos el modelo de regresión no paramétrico del rendimiento logarítmico excedente de las acciones de General Electric que es una función suave e invariable en el tiempo del rendimiento logarítmico excedente del índice S&P 500.

```

fitNPR <- gam(rend.log.ex.GE ~ s(rend.log.ex.SP500),method = "GCV.Cp")
summary(fitNPR)

```

```

##
## Family: gaussian
## Link function: identity
##
## Formula:
## rend.log.ex.GE ~ s(rend.log.ex.SP500)
##
## Parametric coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.0004786  0.0002612   1.832   0.067 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##              edf Ref.df      F p-value
## s(rend.log.ex.SP500) 7.121   8.19 373.4 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##

```

```
## R-sq.(adj) = 0.564   Deviance explained = 56.6%
## GCV = 0.00016169   Scale est. = 0.00016113   n = 2362
```

```
fitNPR$aic
```

```
## [1] -13914.81
```

Este modelo en comparación con el modelo lineal simple a través de una prueba F .

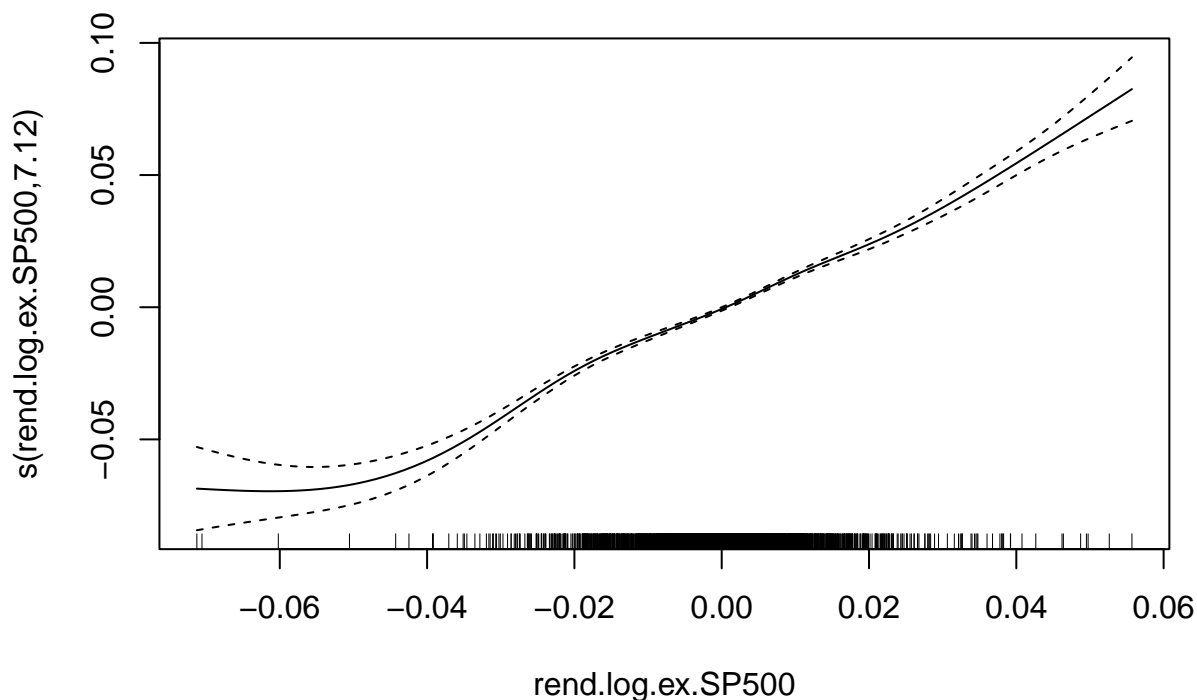
```
anova(fitSLR,fitNPR,test = "F")
```

```
## Analysis of Deviance Table
##
## Model 1: rend.log.ex.GE ~ rend.log.ex.SP500
## Model 2: rend.log.ex.GE ~ s(rend.log.ex.SP500)
##   Resid. Df Resid. Dev    Df Deviance      F   Pr(>F)
## 1     2360.0     0.38492
## 2     2352.8     0.37929  7.1897 0.0056266  4.8567 1.51e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

La prueba F rechaza el modelo de regresión lineal simple. Sin embargo, el tamaño de la muestra es bastante grande ya que hubo 2362 días de retornos y por lo tanto, hay un poder sustancial para detectar desviaciones de la linealidad.

La gráfica de la regresión no paramétrica `fitNPR` se desvía solo ligeramente de una línea recta, al menos sobre el rango de la mayor parte de los datos.

```
plot(fitNPR)
```



Además, el modelo no paramétrico de R^2 – *ajustado*, 56,4%, en comparación con el del modelo lineal, 55,9%. Cuando se recurre a modelos con coeficientes variando, como suposición se dirá que la regresión de los rendimientos logarítmicos excedentes de General Electric sobre los rendimientos logarítmicos excedentes del mercado es lineal, pero la intersección y la pendiente varían con el tiempo.

Ahora, se ajusta y se grafica el modelo de coeficientes variando. En el modelo de coeficientes variando, el intercepto y la pendiente de la regresión lineal del logaritmo de retorno excedente de General Electric con el excedente de rendimiento logarítmico del índice S&P 500 son funciones suaves en el tiempo (años), almacenada como la variable t :

```
dayNums <- (1:(n-1))/(n-1)
startTime <- 1993 + 11/12 #Noviembre
endTime <- 2003 + 3/12 #Marzo
t <- startTime + (endTime - startTime)*dayNums
fitVCM <- gam(rend.log.ex.GE ~ s(t) + s(t,by = rend.log.ex.SP500), method = "GCV.Cp")
summary(fitVCM)
```

```
##
## Family: gaussian
## Link function: identity
##
## Formula:
## rend.log.ex.GE ~ s(t) + s(t, by = rend.log.ex.SP500)
##
## Parametric coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.0003279 0.0002619 1.252 0.211
```



```
##
## Approximate significance of smooth terms:
##              edf Ref.df      F p-value
## s(t)          1.000  1.000   0.001   0.974
## s(t):rend.log.ex.SP500 9.244  9.818 312.226 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.565   Deviance explained = 56.6%
## GCV = 0.00016178   Scale est. = 0.00016101   n = 2362
```

```
fitVCM$aic
```

```
## [1] -13913.49
```

Comparamos con el modelo lineal de coeficientes constantes mediante una prueba F .

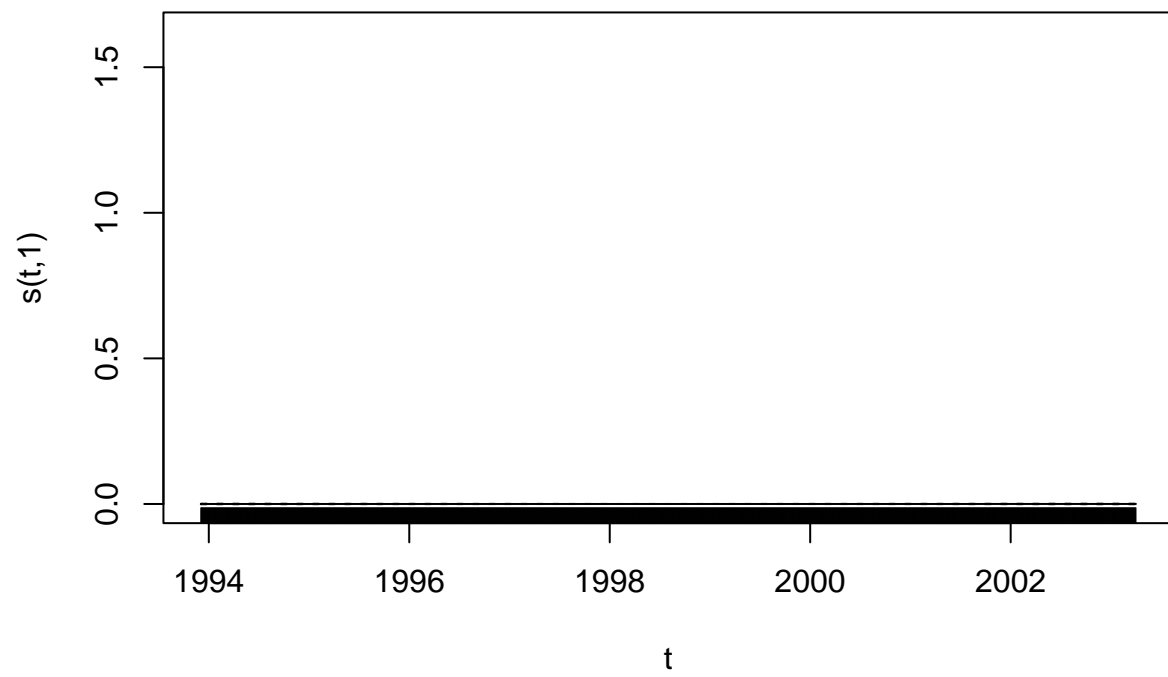
```
anova(fitSLR,fitVCM,test = "F")
```

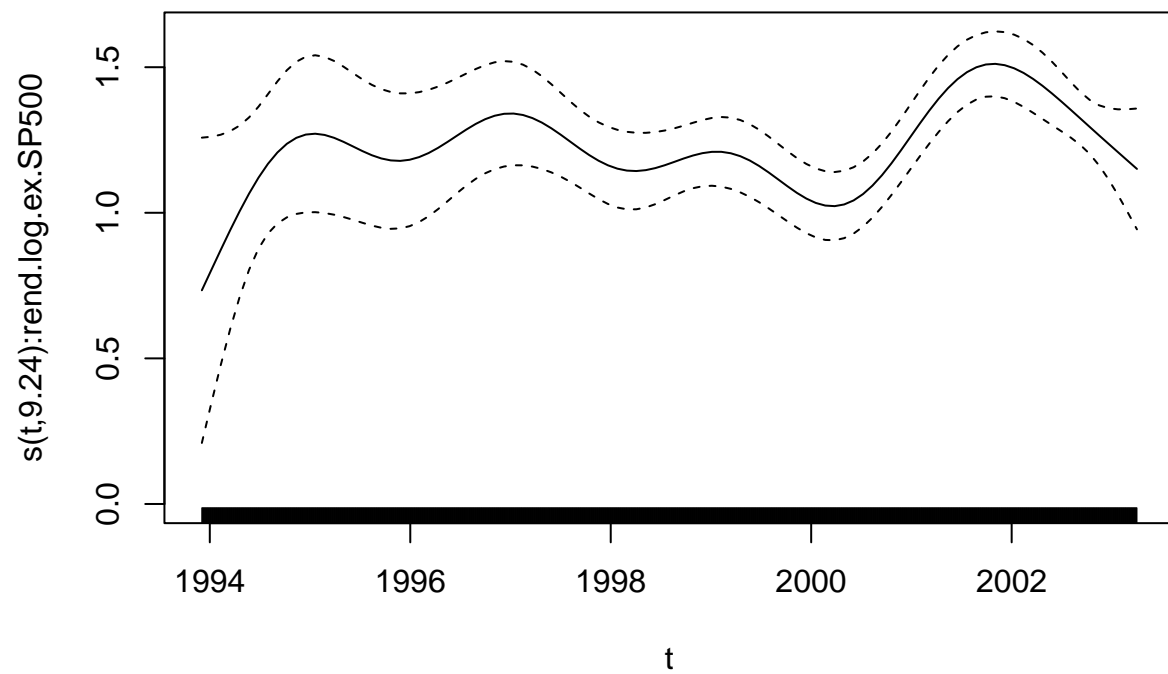
```
## Analysis of Deviance Table
##
## Model 1: rend.log.ex.GE ~ rend.log.ex.SP500
## Model 2: rend.log.ex.GE ~ s(t) + s(t, by = rend.log.ex.SP500)
##   Resid. Df Resid. Dev    Df Deviance      F    Pr(>F)
## 1      2360.0      0.38492
## 2      2350.2      0.37850 9.8183 0.0064164 4.0588 1.744e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

De la misma manera el modelo VCM es significativo, lo cual concluye que supuestamente el mejor modelo es el de coeficientes variando.

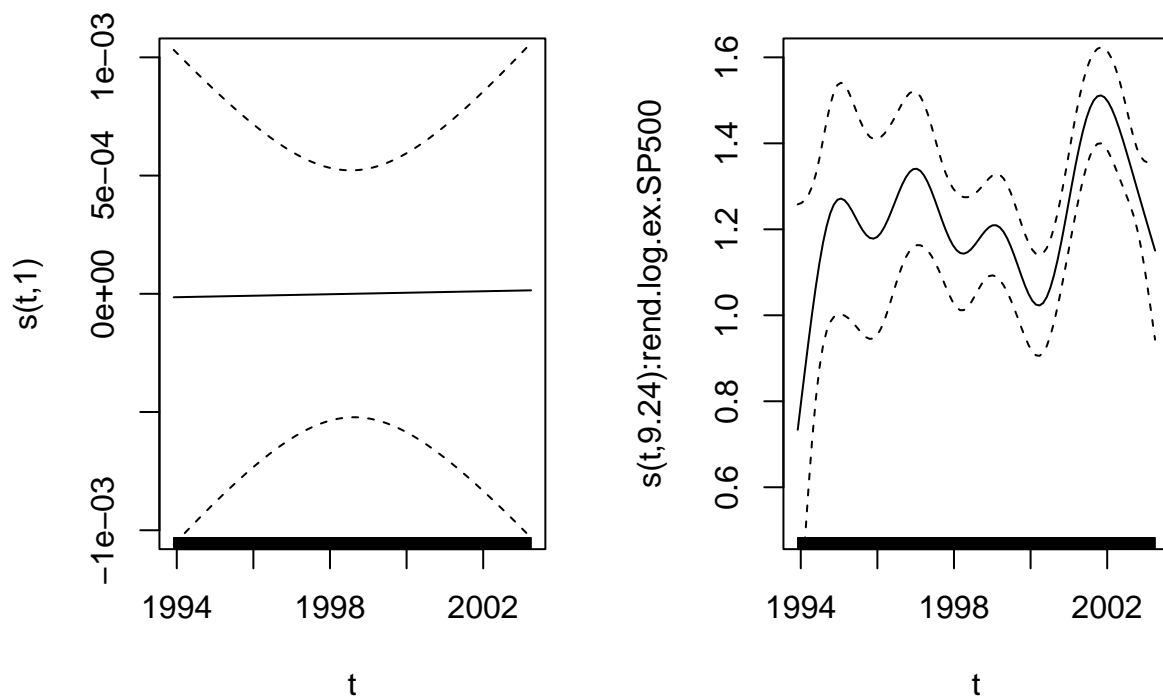
Las funciones de intersección y pendiente se representan en la siguiente gráfica. El argumento seleccionado de `gam()` se usa para seleccionar qué componente trazar. El gráfico de código `fitVCM` sin el uso del seleccionador, trazaría tanto la intersección como la pendiente pero usaría la misma escala en el eje y , lo cual no permitiría especificar valores separados de `ylab` para las dos parcelas. El argumento `seWithMean = TRUE` se usa para especificar que los intervalos de confianza deben incluir la incertidumbre sobre la media.

```
plot(fitVCM)
```





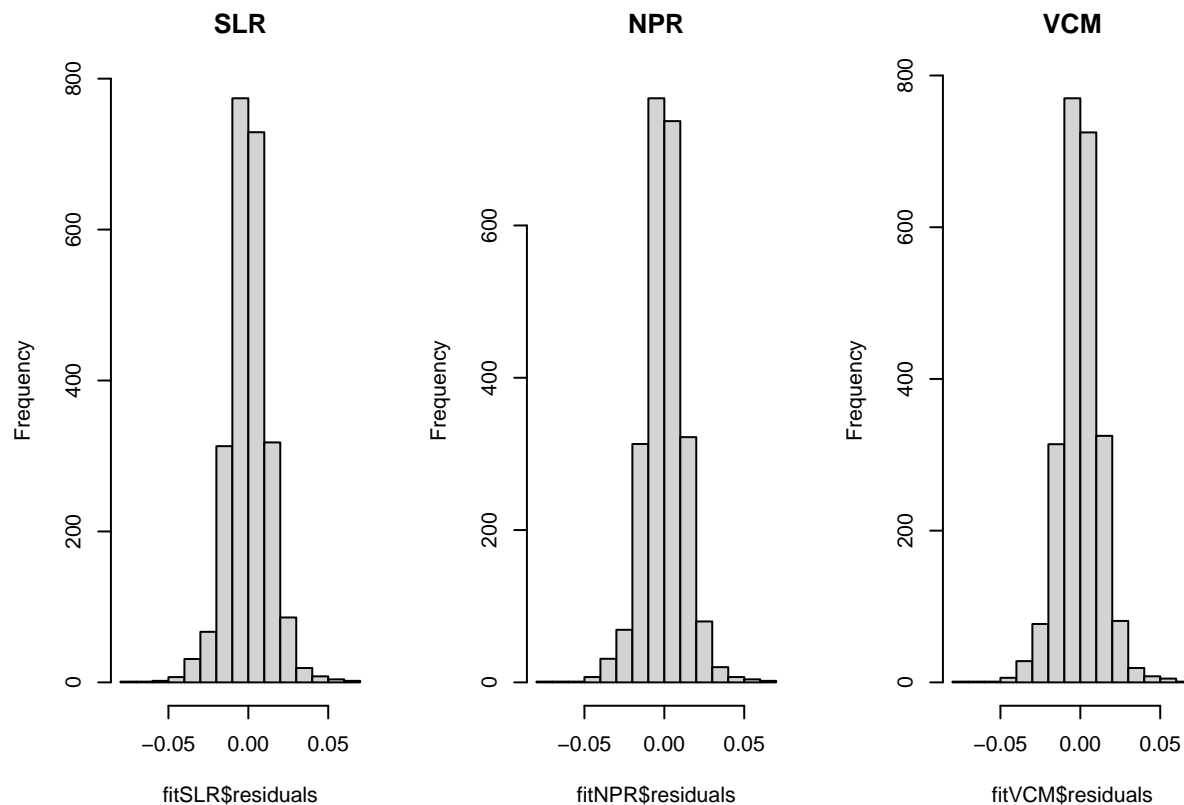
```
par(mfrow=c(1,2))
plot(fitVCM,select = 1,ylim = c(-0.001,0.001),seWithMean = TRUE)
plot(fitVCM,select = 2,ylim = c(0.5,1.6),seWithMean = TRUE)
```



La estimación de la intersección sugiere que es cero o casi nula, durante todo el período, lo que concuerda con la teoría económica. La pendiente estimada varía entre aproximadamente 1,0 y 1,4 durante este período. El modelo de regresión lineal simple se rechaza en favor del modelo de coeficientes variando, aunque el último proporciona solo una pequeña mejora en el ajuste medido por el $R^2 - \text{ajustado}$.

5.5 Residuos

```
par(mfrow=c(1,3))
hist(fitSLR$residuals,main="SLR")
hist(fitNPR$residuals,main="NPR")
hist(fitVCM$residuals,main="VCM")
```



```
library(PerformanceAnalytics)
```

```
## Warning: package 'PerformanceAnalytics' was built under R version 4.1.3
```

```
## Loading required package: xts
```

```
## Loading required package: zoo
```

```
##
```

```
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':
```

```
##
```

```
## as.Date, as.Date.numeric
```

```
##
```

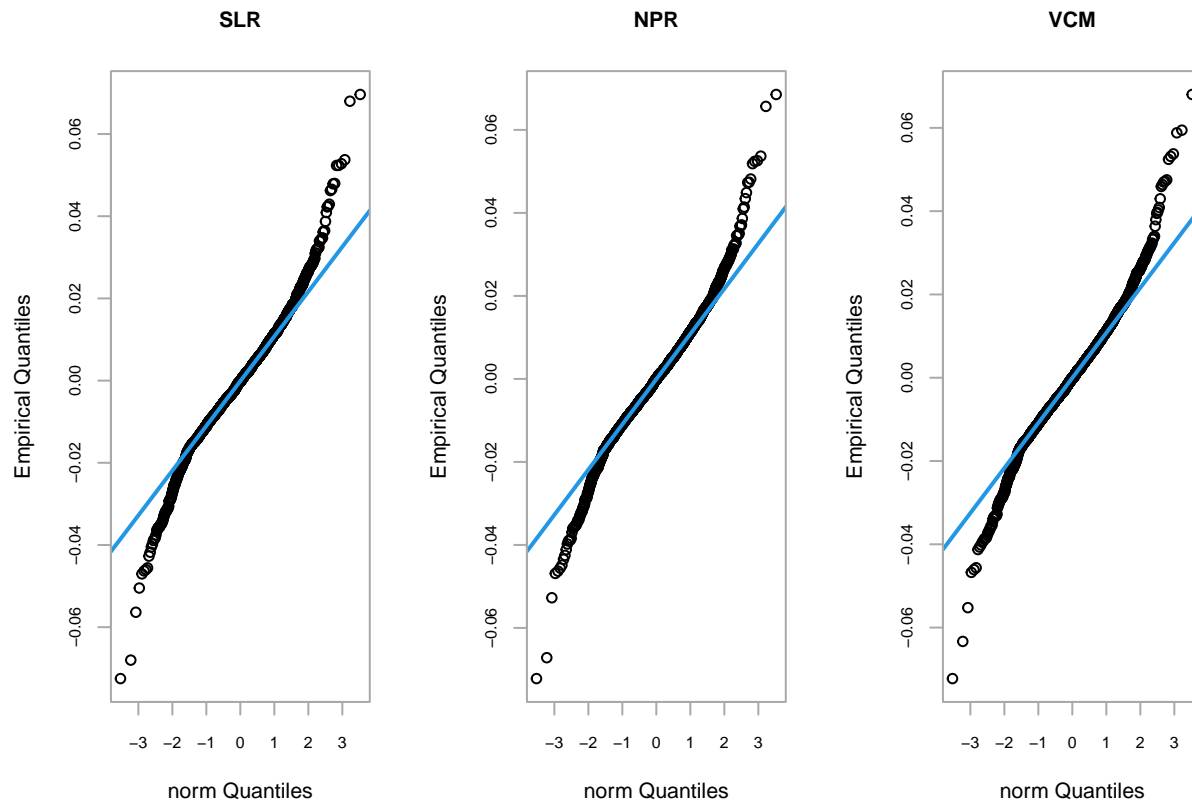
```
## Attaching package: 'PerformanceAnalytics'
```

```
## The following object is masked from 'package:graphics':
```

```
##
```

```
## legend
```

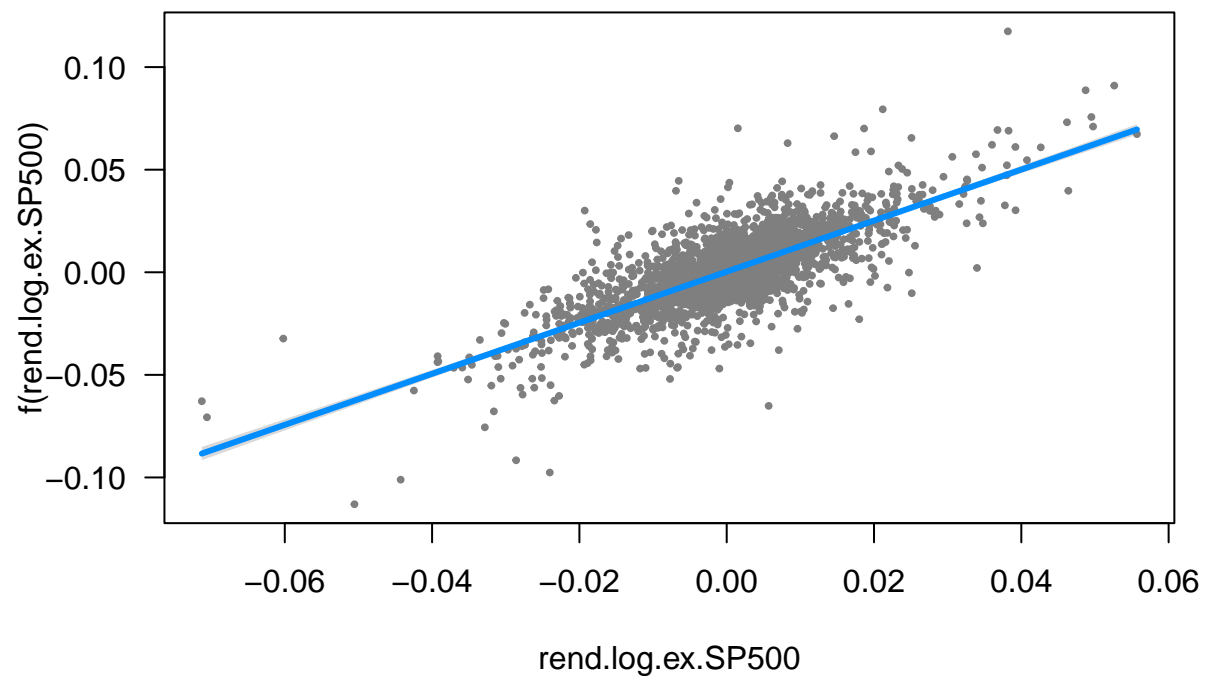
```
par(mfrow=c(1,3))
chart.QQPlot(fitSLR$residuals,main="SLR")
chart.QQPlot(fitNPR$residuals,main="NPR")
chart.QQPlot(fitVCM$residuals,main="VCM")
```



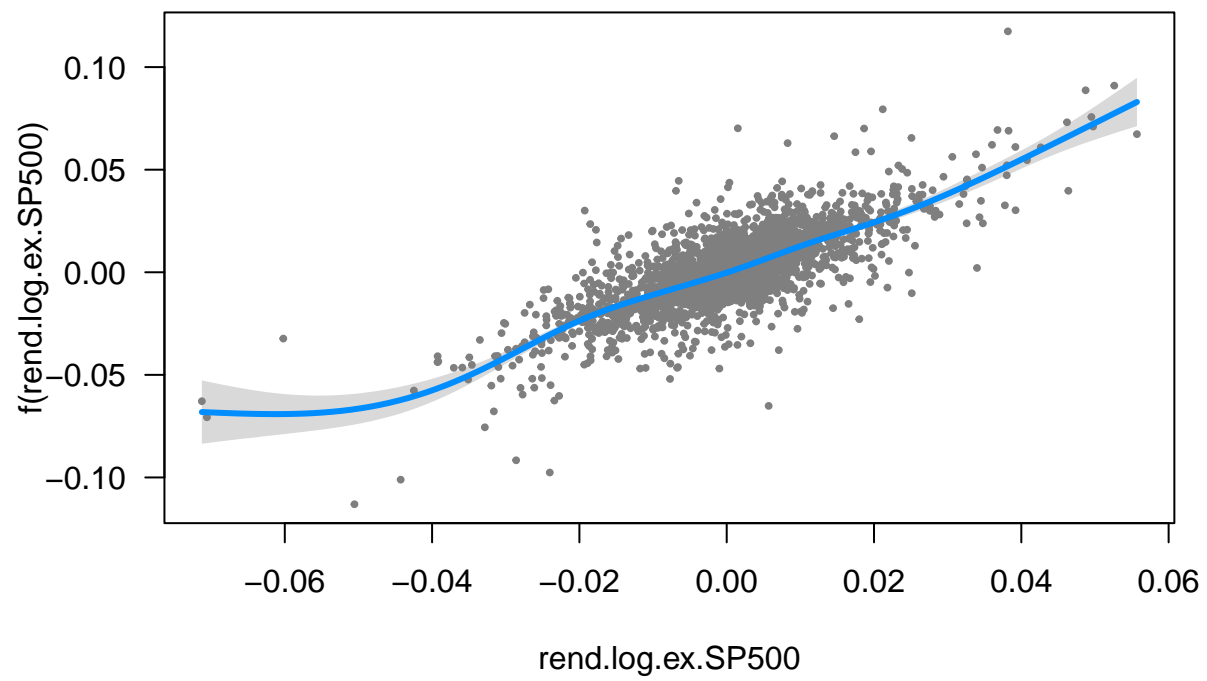
```
library(visreg)
```

```
## Warning: package 'visreg' was built under R version 4.1.3
```

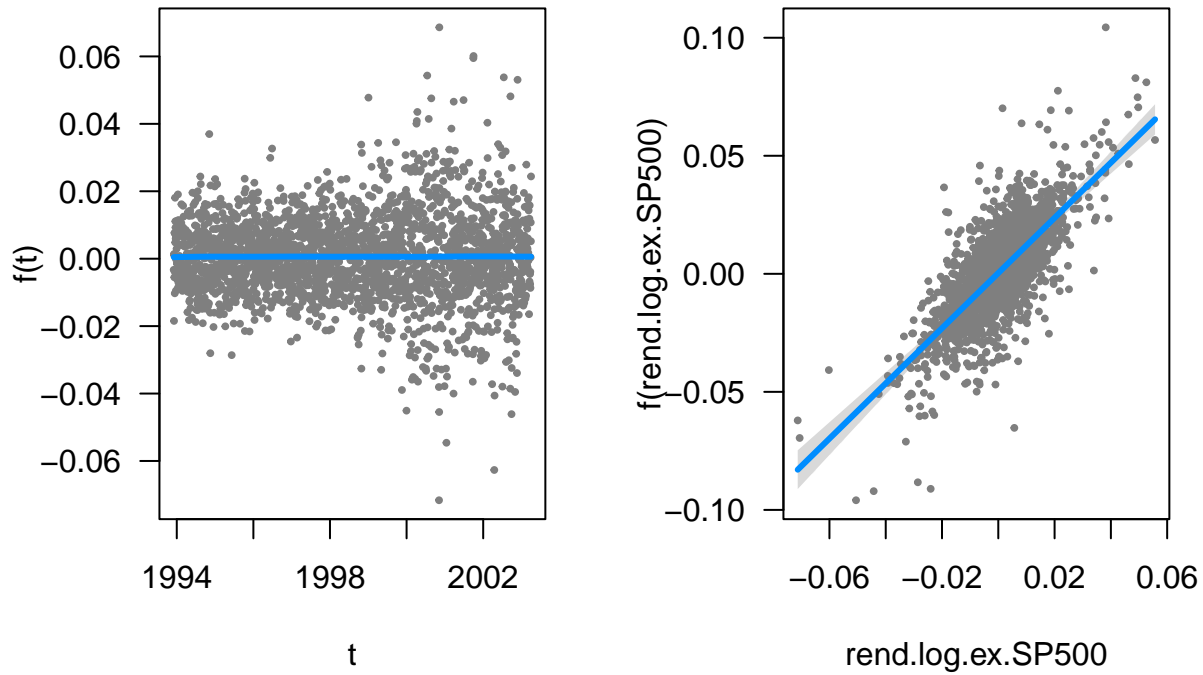
```
visreg(fitSLR)
```



```
visreg(fitNPR)
```



```
par(mfrow=c(1,2))  
visreg(fitVCM)
```

6 Conclusiones

- En general los modelos de regresión con coeficientes variando es una alternativa de modelación de datos, cuando una covariable en interacción con otra variable puede ayudar a explicar de mejor forma el modelo, teniendo siempre en cuenta el principio de parsimonia.
- Es bastante flexible en cuanto al estudio de datos que se requiera realizar, se podría utilizar en datos de corte transversal y así también en datos longitudinales (panel).
- En este trabajo se presento de manera inicial el modelo lineal generalizado de coeficientes variando, lo cual nos proporciona diferentes modelos con distribuciones pertenientes a la familia exponencial.
- También se puede contar con modelos de coeficientes variando asimétricos (skew).

Bibliografia

- Harezlak, J., Ruppert, D., & Want, M. (2018). *Semiparametric regression with r*. Springer Science+Business Media, LLC, part of Springer Nature. <https://doi.org/10.1007/978-1-4939-8851-8>
- Hastie, T., & Tibshirani, R. (1986). Generalized Additive Models. *Statistical Science*, 1(3), 297–310. <https://doi.org/10.1214/ss/1177013604>
- Hastie, T., & Tibshirani, R. (1993). Varying-coefficient models. *Journal of the Royal Statistical Society. Series B (Methodological)*, 55(4), 757–796.
- Ruppert, D., Want, M., & Carroll, R. (2003). *Semiparametric regression*. Cambridge University Press. <https://doi.org/10.1007/978-0-511-06683-2>