

1 Metodología general

En esta sección describimos paso a paso el procedimiento seguido para ajustar tendencias de Guerrero en una familia grande de series financieras y para seleccionar el parámetro de suavización mediante un esquema *train-validation-test* con múltiples criterios de error.

1.1 Formalización del problema y notación

Sea $k \in \{1, \dots, K\}$ el índice de activo (ticker) y sea $t \in \{1, \dots, N_k\}$ el índice temporal (días de negociación). Para cada ticker k observamos una serie de precios de cierre

$$P_{k,t} > 0, \quad t = 1, \dots, N_k,$$

y trabajamos con los *log-precios*

$$Z_{k,t} := \log P_{k,t}.$$

En el estudio actual consideramos $K = 137$ activos, que incluyen: acciones estadounidenses (por ejemplo AAPL, MSFT, AMZN), índices bursátiles (por ejemplo $\hat{\text{GSPC}}$, $\hat{\text{MXX}}$), pares cripto en USD (BTC-USD, ETH-USD, etc.) y otros índices globales (por ejemplo $\hat{\text{N225}}$, $\hat{\text{FTSE}}$), entre otros.

Para cada ticker k fijamos un orden de diferencia $d \in \{1, 2, 3, 4\}$ y un parámetro de suavización $\lambda > 0$. La tendencia de Guerrero $t_k(\lambda) = (t_{k,1}(\lambda), \dots, t_{k,N_k}(\lambda))^\top$ se define como el minimizador del funcional penalizado

$$\hat{t}_k(\lambda) := \arg \min_{t \in \mathbb{R}^{N_k}} \left\{ \sum_{t=1}^{N_k} (Z_{k,t} - t_t)^2 + \lambda \sum_{t=d+1}^{N_k} (\Delta^d t_t)^2 \right\},$$

(1)eq : guerrero – penalized

donde Δ^d denota el operador de diferencias de orden d aplicado componente a componente. En notación matricial podemos escribir

$$\hat{t}_k(\lambda) = S_d(\lambda) \mathbf{Z}_k, \quad S_d(\lambda) := (I_{N_k} + \lambda K_d^\top K_d)^{-1},$$

donde $\mathbf{Z}_k = (Z_{k,1}, \dots, Z_{k,N_k})^\top$ y $K_d \in \mathbb{R}^{(N_k-d) \times N_k}$ es la matriz de diferencias de orden d .

En la práctica Guerrero introduce un índice adimensional de suavidad $S_d(\lambda, N_k) \in (0, 1)$ y se trabaja con s en lugar de λ ; en nuestro código esta cantidad se denota típicamente por `s_unit` o `s_real`.

1.2 Por qué trabajamos con log-precios

En lugar de aplicar la tendencia directamente a los precios $P_{k,t}$, usamos los log-precios $Z_{k,t} = \log P_{k,t}$. Esto tiene varias ventajas estándar en series financieras:

1.3 Esquema de partición: entrenamiento, validación y prueba

Para cada ticker k fijamos una descomposición estricta del horizonte temporal en tres subconjuntos contiguos:

$$\{1, \dots, N_k\} = \underbrace{\{1, \dots, N_k^{\text{tr}}\}}_{\text{entrenamiento}} \cup \underbrace{\{N_k^{\text{tr}} + 1, \dots, N_k^{\text{tr}} + N_k^{\text{va}}\}}_{\text{validación}} \cup \underbrace{\{N_k^{\text{tr}} + N_k^{\text{va}} + 1, \dots, N_k\}}_{\text{prueba}},$$

donde N_k^{tr} , N_k^{va} y N_k^{te} son las longitudes de los segmentos de entrenamiento, validación y prueba.

Notación:

$$\mathbf{Z}_k^{\text{tr}} := (Z_{k,1}, \dots, Z_{k,N_k^{\text{tr}}})^\top, \quad \mathbf{Z}_k^{\text{va}} := (Z_{k,N_k^{\text{tr}}+1}, \dots, Z_{k,N_k^{\text{tr}}+N_k^{\text{va}}})^\top, \quad \mathbf{Z}_k^{\text{te}} := (Z_{k,N_k^{\text{tr}}+N_k^{\text{va}}+1}, \dots, Z_{k,N_k})^\top.$$

La clave es que el ajuste penalizado de Guerrero (??) se realiza *solo* sobre el segmento de entrenamiento, y luego se extiende hacia validación y prueba mediante un polinomio de orden d en diferencias.

1.4 Polinomio global en diferencias y pronóstico

Dado un d y un índice de suavidad s , el *solver espectral de Guerrero* (`solver.fit_for_s`) devuelve, a partir de la serie de entrenamiento, un vector de tendencia ajustada en entrenamiento $\widehat{t}_k^{\text{tr}}(s)$ y un escalar $m_k(s)$ que codifica el comportamiento en diferencias de orden d al final de la ventana de entrenamiento. A partir de ello construimos un polinomio global $\mathbf{p}_k^{(d,s)}$ sobre todo el horizonte $\{1, \dots, N_k\}$ con la rutina `build_polynomial_from_train_tail`:

$$\mathbf{p}_k^{(d,s)} := (p_{k,1}^{(d,s)}, \dots, p_{k,N_k}^{(d,s)})^\top.$$

Este polinomio satisface:

- $p_{k,t}^{(d,s)} \approx \widehat{t}_{k,t}^{\text{tr}}(s)$ en el segmento de entrenamiento;
- las diferencias de orden d de $p_{k,t}^{(d,s)}$ son constantes (o siguen la estructura impuesta por $m_k(s)$) más allá del final del entrenamiento, de modo que en validación y prueba la “tendencia pronosticada” es simplemente la extrapolación polinómica.

En notación compacta, para cada segmento $A \in \{\text{tr}, \text{va}, \text{te}\}$ definimos los errores residuales

$$e_{k,t}^A(d, s) := p_{k,t}^{(d,s)} - Z_{k,t}, \quad t \in A.$$

También consideramos el segmento combinado

$$A = \text{tv} := \text{tr} \cup \text{va}$$

para estudiar el ajuste sobre entrenamiento más validación de forma conjunta.

1.5 Búsqueda sobre una rejilla de suavidad y mínimos locales

El parámetro de suavización s no es observable y no viene dado por la teoría; debe elegirse a partir de los datos. Nuestro enfoque es:

1. Fijar un orden de diferencia $d \in \{1, 2, 3, 4\}$.
2. Fijar una rejilla de valores de suavidad

$$\mathcal{S} = \{s_1, s_2, \dots, s_M\} \subset (0, 1),$$

suficientemente fina (por ejemplo espaciamiento uniforme en escala de s o en escala logarítmica de λ).

3. Para cada $s \in \mathcal{S}$:
 - (a) Ajustar la tendencia de Guerrero en entrenamiento, obteniendo $\hat{\mathbf{t}}_k^{\text{tr}}(s)$, $m_k(s)$ y $\lambda(s)$.
 - (b) Construir el polinomio global $\mathbf{p}_k^{(d,s)}$.
 - (c) Calcular funciones de pérdida tipo RMSE en los distintos segmentos, como se detalla más abajo.
4. Sobre cada función $s \mapsto \text{pérdida}(s)$ se exploran *mínimos locales* en la rejilla: es decir, valores s^* tales que, en la rejilla, su correspondiente pérdida es menor que la de sus vecinos inmediatos. Esto es coherente con el hecho de que la función de pérdida en s no es estrictamente convexa y puede presentar múltiples mínimos.

En paralelo se ha obtenido una derivación analítica de la derivada de la pérdida de validación respecto a λ , lo que permitiría buscar raíces de $f'(\lambda) = 0$ por métodos de optimización unidimensional. Sin embargo, debido a la posible multimodalidad de $f(\lambda)$ y a que el solver de Guerrero ya opera cómodamente en rejillas de s , optamos por una búsqueda directa sobre \mathcal{S} y la detección de mínimos locales.

1.6 Métricas de error: RMSE simples y ponderadas

Para cada ticker k , orden d y suavidad s , definimos los errores cuadráticos medios (RMSE) no ponderados en los segmentos $A \in \{\text{tr}, \text{va}, \text{te}, \text{tv}\}$:

$$\text{RMSE}_k^A(d, s) := \sqrt{\frac{1}{N_k^A} \sum_{t \in A} (e_{k,t}^A(d, s))^2}, \quad N_k^A := |A|.$$

Además introducimos versiones ponderadas temporalmente para dar mayor peso a los errores cercanos al final de cada segmento. Para un segmento genérico A con índices consecutivos $\{t_1, \dots, t_{N_A}\}$ definimos pesos linealmente crecientes

$$\tilde{w}_j^{(A)} := j, \quad j = 1, \dots, N_A,$$

y los normalizamos a suma uno:

$$w_j^{(A)} := \frac{\tilde{w}_j^{(A)}}{\sum_{u=1}^{N_A} \tilde{w}_u^{(A)}} = \frac{2j}{N_A(N_A + 1)}.$$

Nótese que $w_{N_A}^{(A)} = N_A \cdot w_1^{(A)}$, de modo que el último punto del segmento tiene N_A veces más peso que el primero.

La *RMSE ponderada* para el segmento A se define entonces como

$$\text{RMSE}_{k,w}^A(d, s) := \sqrt{\sum_{j=1}^{N_A} w_j^{(A)} (e_{k,t_j}^A(d, s))^2}.$$

En la implementación se han considerado explícitamente las siguientes métricas para cada combinación (k, d, s) :

- $\text{RMSE}_k^{\text{tr}}(d, s)$, $\text{RMSE}_k^{\text{va}}(d, s)$, $\text{RMSE}_k^{\text{te}}(d, s)$, $\text{RMSE}_k^{\text{tv}}(d, s)$;
- $\text{RMSE}_{k,w}^{\text{tr}}(d, s)$, $\text{RMSE}_{k,w}^{\text{va}}(d, s)$, $\text{RMSE}_{k,w}^{\text{te}}(d, s)$, $\text{RMSE}_{k,w}^{\text{tv}}(d, s)$,

es decir, versiones simples y ponderadas en entrenamiento, validación, entrenamiento+validación y prueba.

1.7 Selección de candidatos y evaluación en prueba

Para cada ticker k y orden d , y para cada métrica de error de las enumeradas, obtenemos un conjunto de mínimos locales en la rejilla \mathcal{S} :

$$\mathcal{S}_{k,d}^{(\text{métrica})} = \{s \in \mathcal{S} : s \text{ es mínimo local de la métrica en la rejilla}\}.$$

La filosofía es:

- Cada métrica encapsula una noción distinta de “buen suavizamiento” (por ejemplo, buen ajuste puro en validación, buen compromiso entrenamiento–validación, énfasis en el final de la muestra, etc.).
- En lugar de comprometernos con una única métrica, recolectamos todos los s que son prometedores para *alguna* de ellas.

Una vez identificados estos candidatos s^* :

1. Se recalculan las tendencias y polinomios $\mathbf{p}_k^{(d,s^*)}$.
2. Se computan las métricas de prueba $\text{RMSE}_k^{\text{te}}(d, s^*)$ y $\text{RMSE}_{k,w}^{\text{te}}(d, s^*)$, que miden el desempeño completamente fuera de muestra.
3. Se analizan las distribuciones de estas métricas sobre tikers, órdenes d y criterios de selección para estudiar la robustez del método y posibles patrones comunes.

1.8 Resumen simbólico

Resumiendo de forma compacta, para cada ticker k , orden d y suavización s :

$$\hat{\mathbf{t}}_k^{\text{tr}}(d, s) = \arg \min_{\mathbf{t}} \left\{ \sum_{t \in \text{tr}} (Z_{k,t} - t_t)^2 + \lambda(s) \sum_{t \in \text{tr}} (\Delta^d t_t)^2 \right\},$$

$$\mathbf{p}_k^{(d,s)} = \text{polinomio global de orden } d \text{ anclado en entrenamiento},$$

$$e_{k,t}^A(d, s) = p_{k,t}^{(d,s)} - Z_{k,t}, \quad A \in \{\text{tr}, \text{va}, \text{tv}, \text{te}\},$$

$$\text{RMSE}_k^A(d, s) = \sqrt{\frac{1}{N_k^A} \sum_{t \in A} (e_{k,t}^A(d, s))^2},$$

$$\text{RMSE}_{k,w}^A(d, s) = \sqrt{\sum_{t \in A} w_t^{(A)} (e_{k,t}^A(d, s))^2}.$$

La tarea central de selección de suavización consiste en estudiar, para cada ticker y orden, la familia de funciones $s \mapsto \text{RMSE}(\cdot, d, s)$ sobre una rejilla \mathcal{S} , localizar mínimos locales y evaluar su robustez y desempeño en el segmento de prueba.