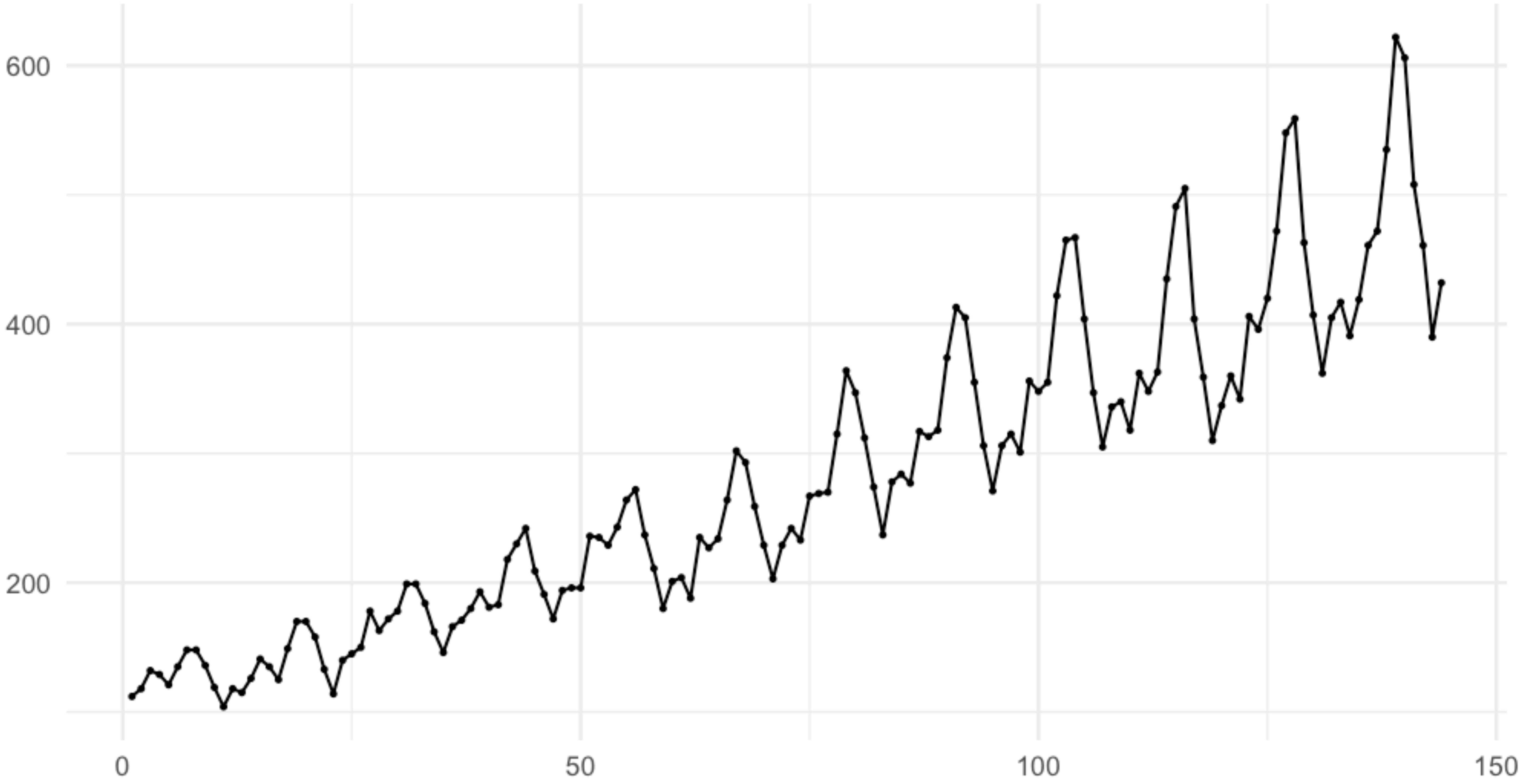
Pasajeros de avión

Imagen tomada de: https://unofficialcuphead.fandom.com/wiki/Cuphead_(character)

Datos

- Datos mensuales de enero de 1949 a diciembre de 1960



Observaciones

- No es estacionaria
- Tiene una tendencia creciente
- Tiene un componente estacional
- Varianza crece (efecto multiplicativo)

Descomposición clásica

- El modelo se puede expresar como

$$X_t = m_t + s_t + Y_t$$

donde

• m_t - es el componente de tendencia

• s_t - es el componente estacional

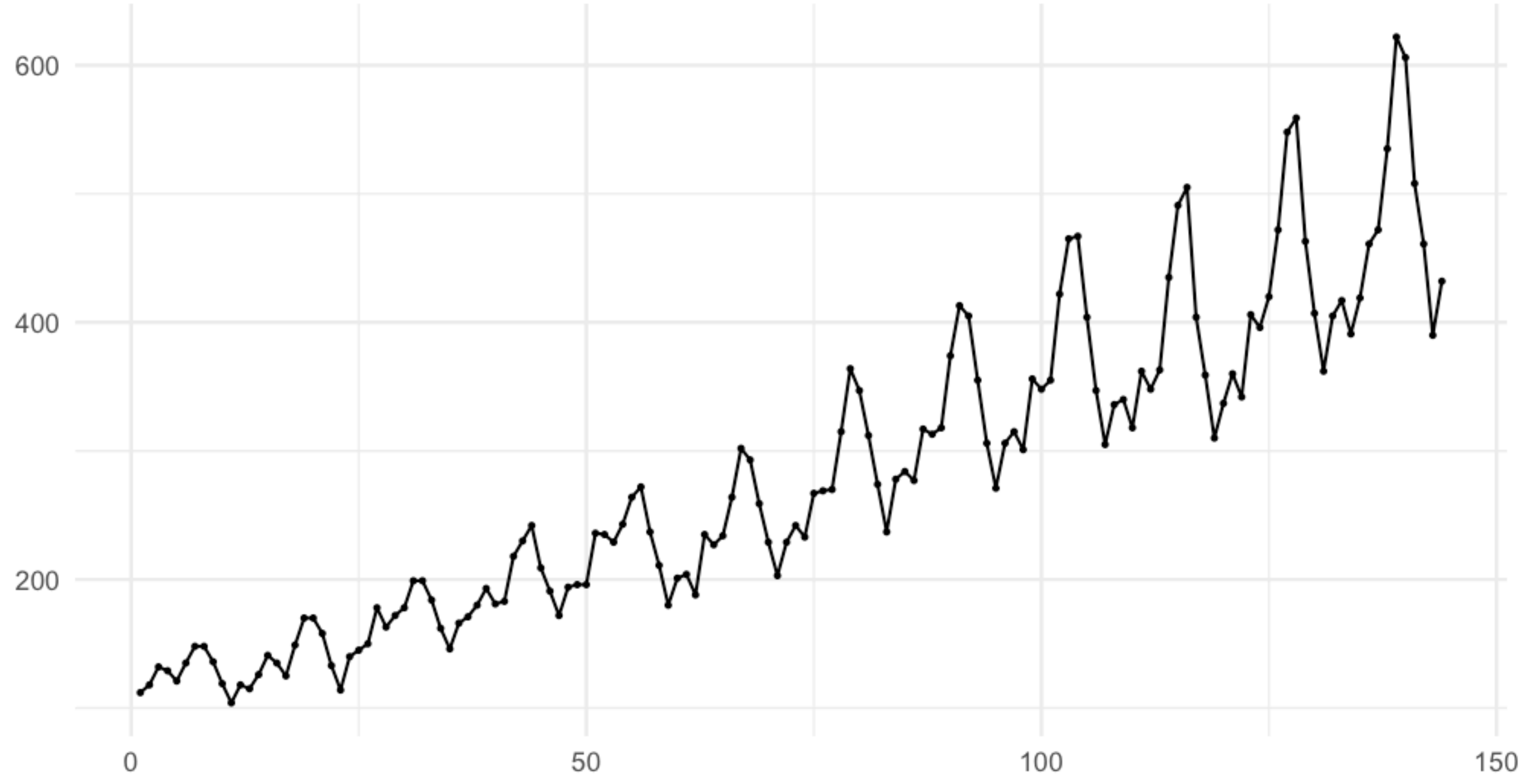
• Y_t - es un componente de ruido aleatorio estacionario

- No necesariamente aparecerán los dos componentes

- El objetivo es estimar m_t y s_t (no se consideran aleatorios)

Datos

- Tendencia creciente y ciclos anuales



Estimación de m_t sin componente estacional

Método 1: Suavizamientos

Suavizamiento mediante promedios móviles

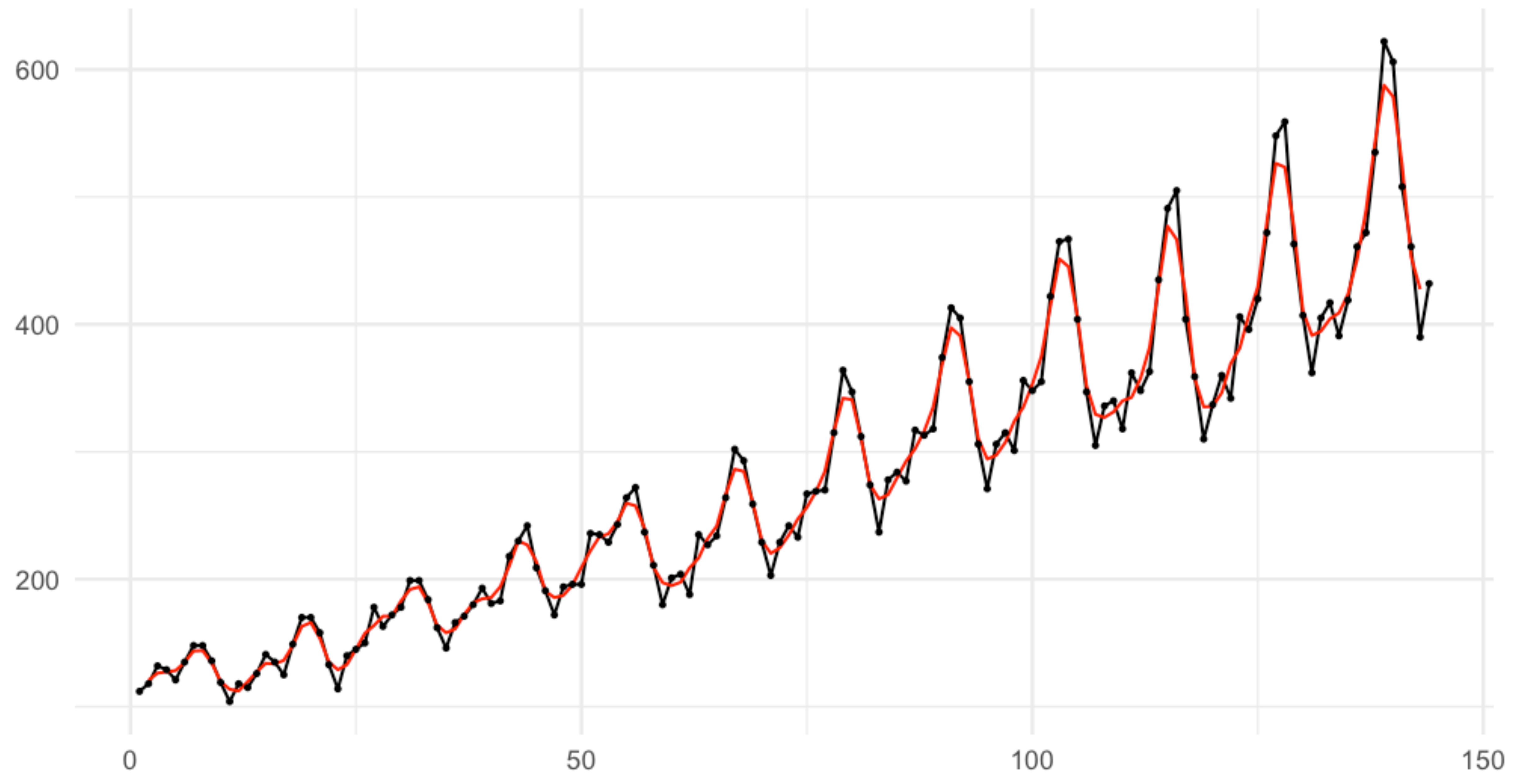
$$\widehat{m}_t = (2q+1)^{-1} \sum_{j=-q}^{q} X_{t-j}, \quad q+1 \leq t \leq n-q$$

- Asumiendo observaciones $X_1, \ldots, X_n$, no se puede calcular para $t \leq q$ ni para $t \geq n-q$

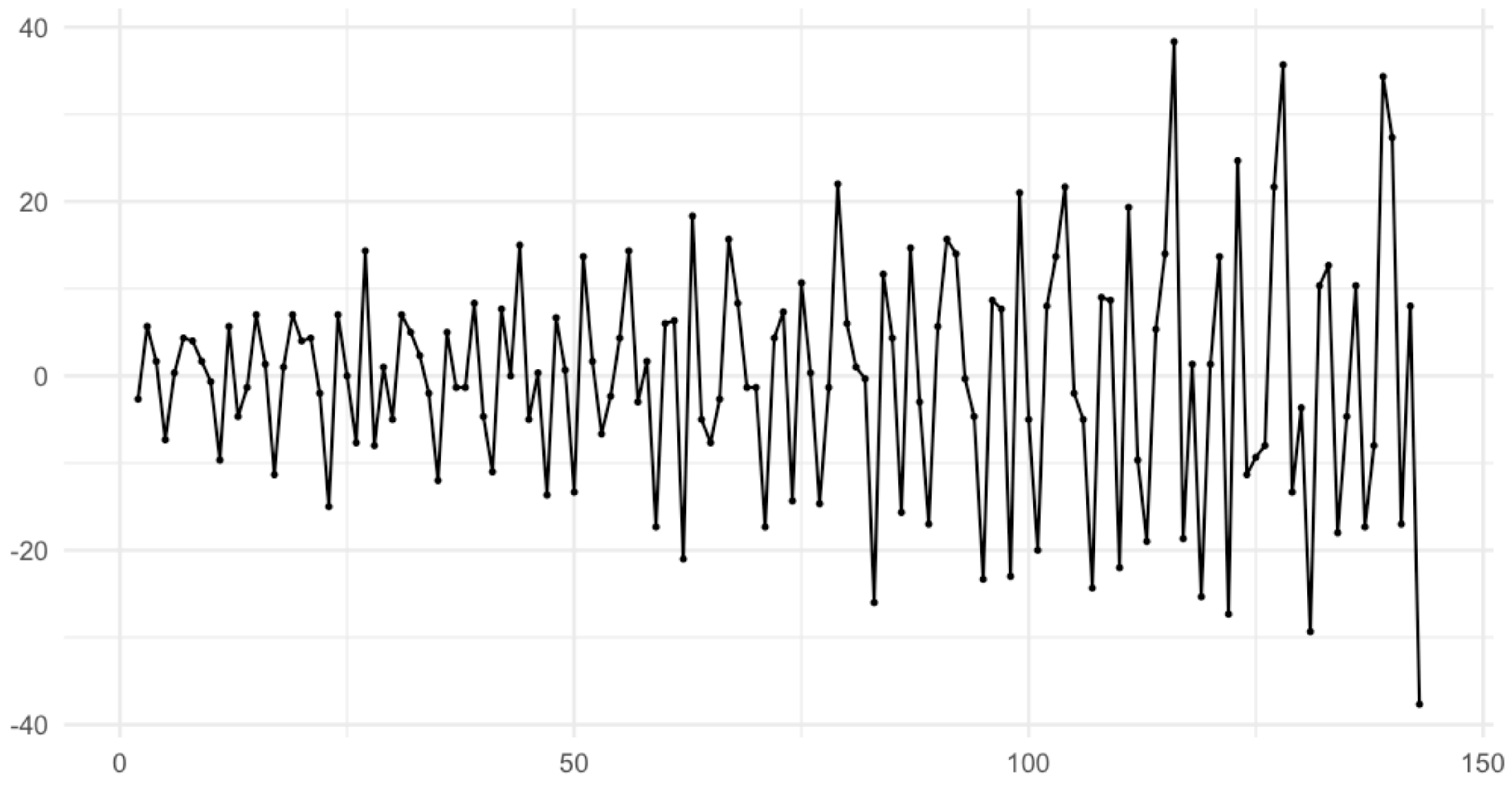
- Bajo ciertas condiciones puede remover la variación estacional

Promedio móvil $q = 1$

Valores ajustados



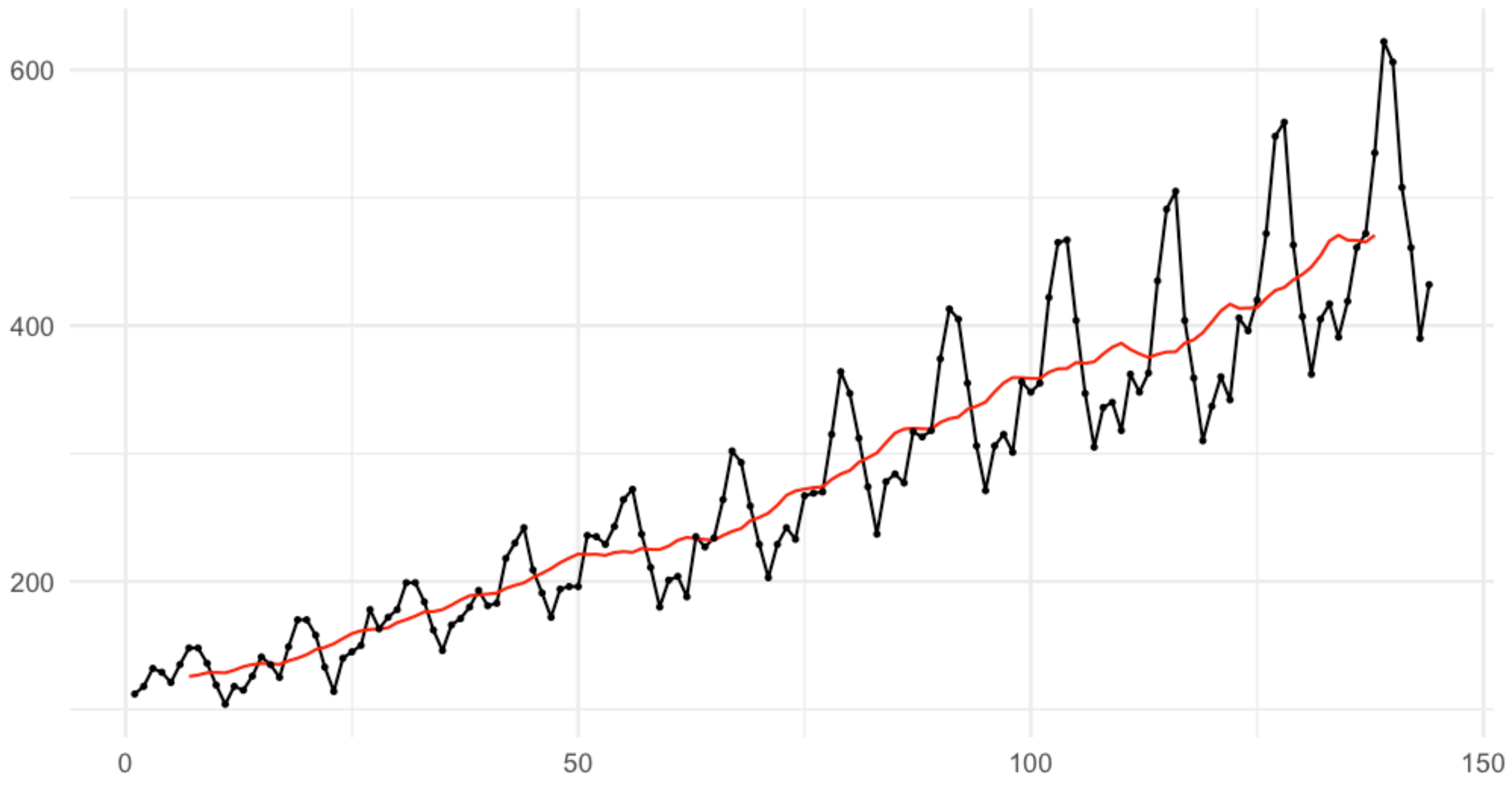
Residuales



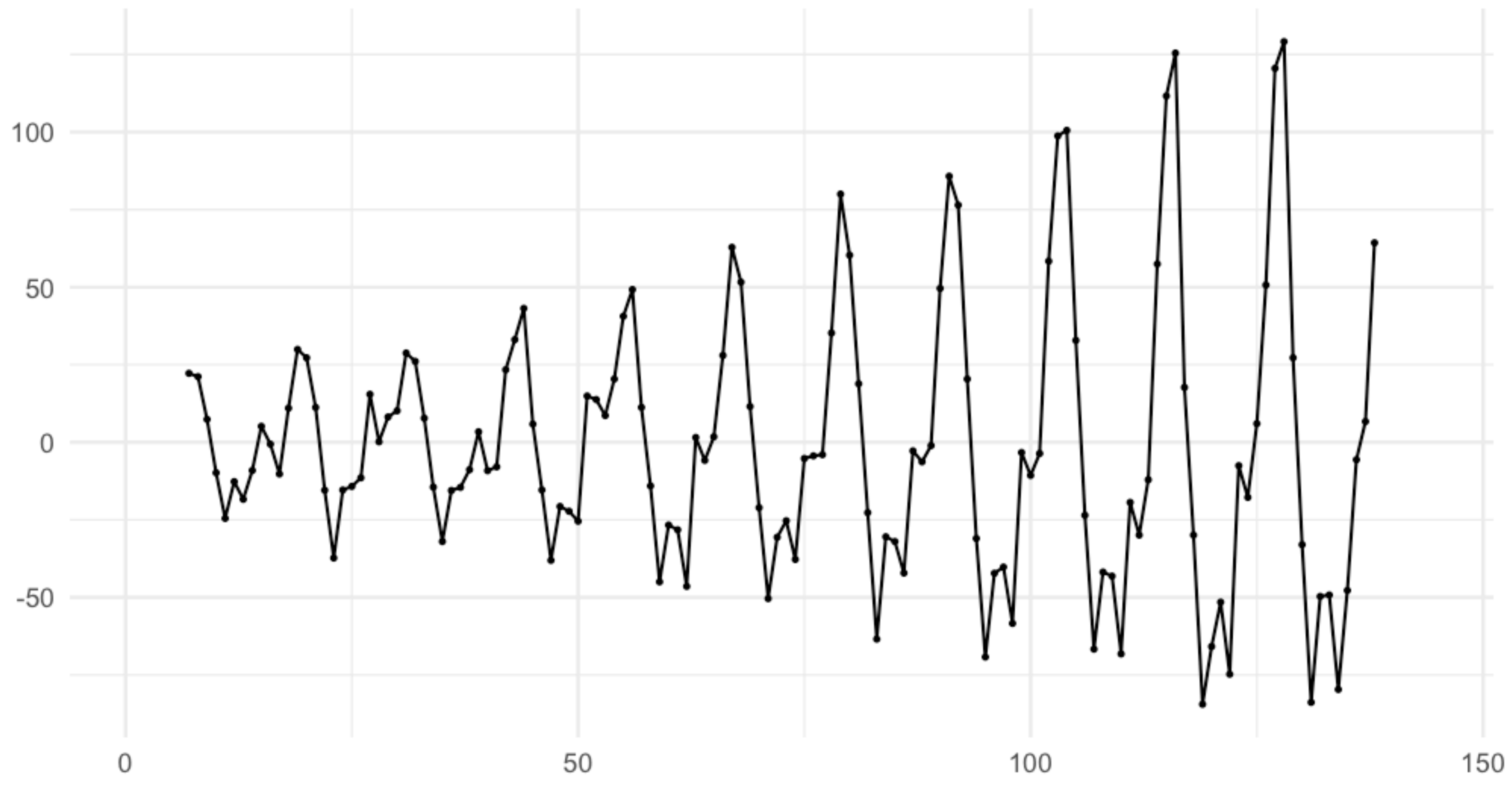
- Sigue sin ser estacionario

Promedio móvil $q = 6$

Valores ajustados



Residuales



- Sigue sin ser estacionario

Método 1: Suavizamientos

Suavizamiento exponencial para $\alpha \in [0,1]$

$$\widehat{m}_t = \alpha X_t + (1-\alpha)\,\widehat{m}_{t-1}, \qquad t = 2,\ldots,n$$
$$\widehat{m}_1 = X_1$$

- Equivalentemente para $t \geq 2$

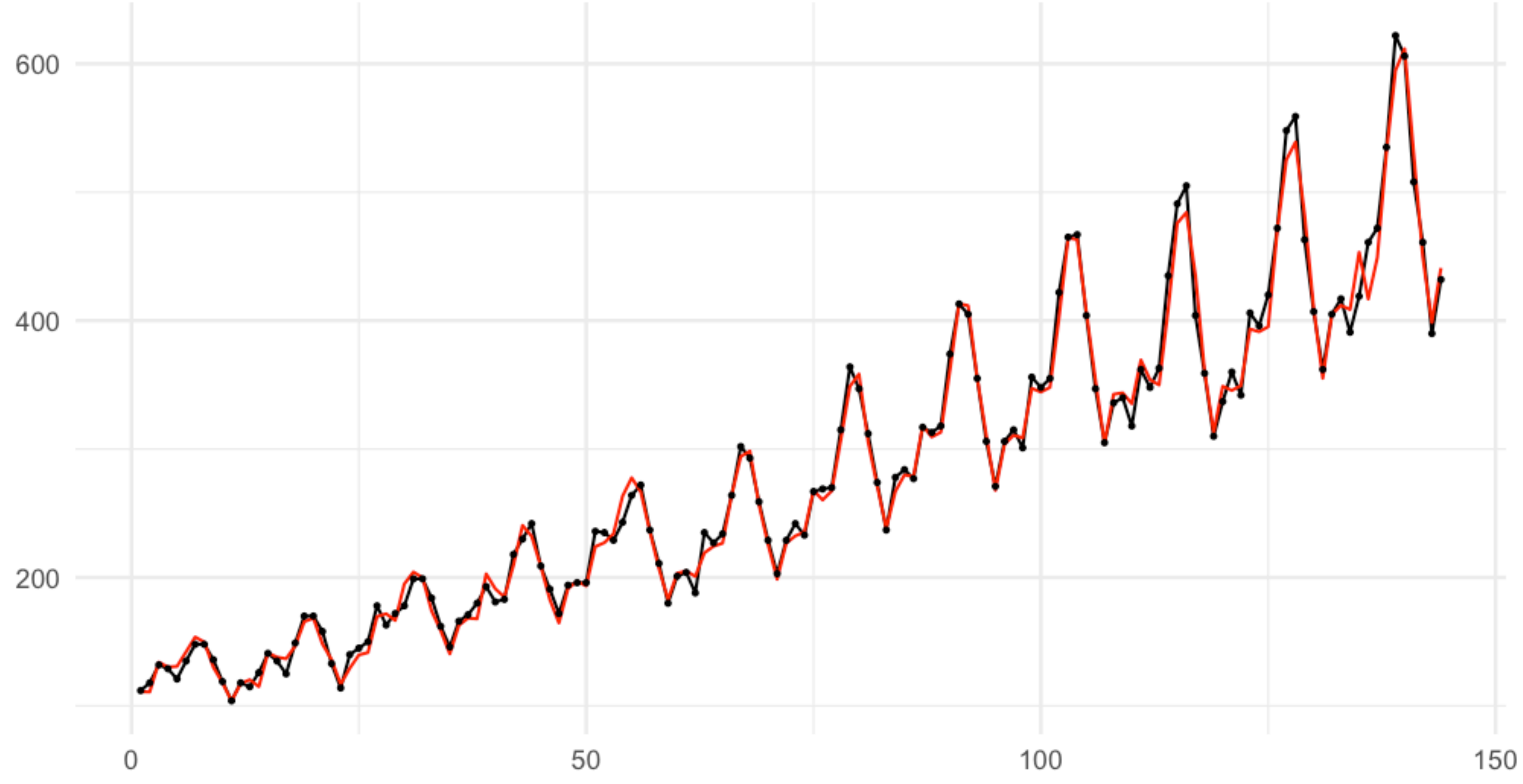
$$\widehat{m}_t = \sum_{j=0}^{t-2} \alpha(1-\alpha)^j X_{t-j} + (1-\alpha)^{t-1} X_1$$

- Versiones más generales para eliminar tendencia, estacionalidad, modelos aditivos y multiplicativos (en R: paquetería forecast y la función es ets)

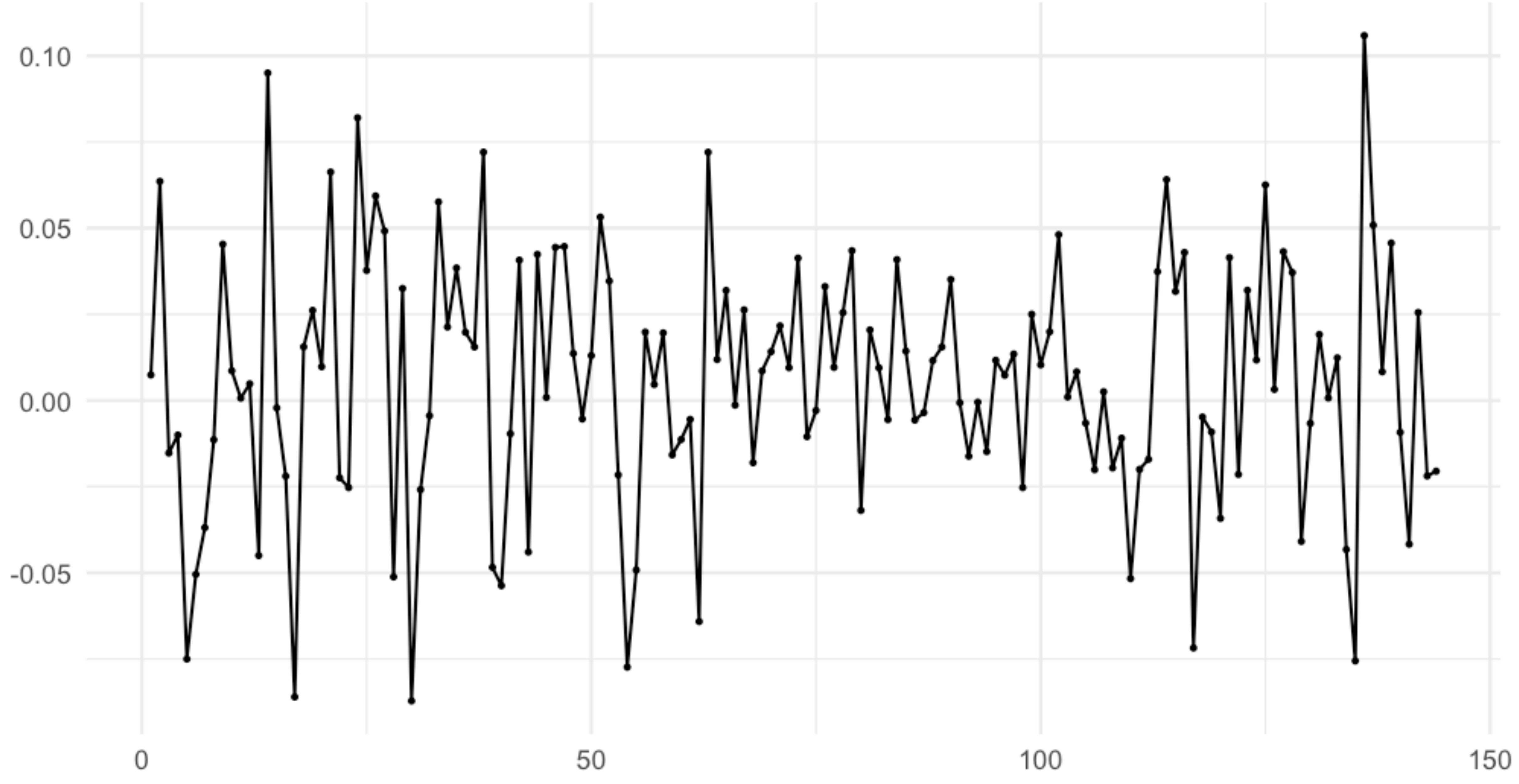
- Eficiente para hacer predicciones a corto plazo pero no siempre es adecuado

Suavizamiento exponencial

Valores ajustados



Residuales



- Ya parece ser un ruido blanco

Método 2: Operador de diferencias

Diferencia de lag 1

$$\nabla X_t = X_t - X_{t-1} = (1 - B)X_t$$

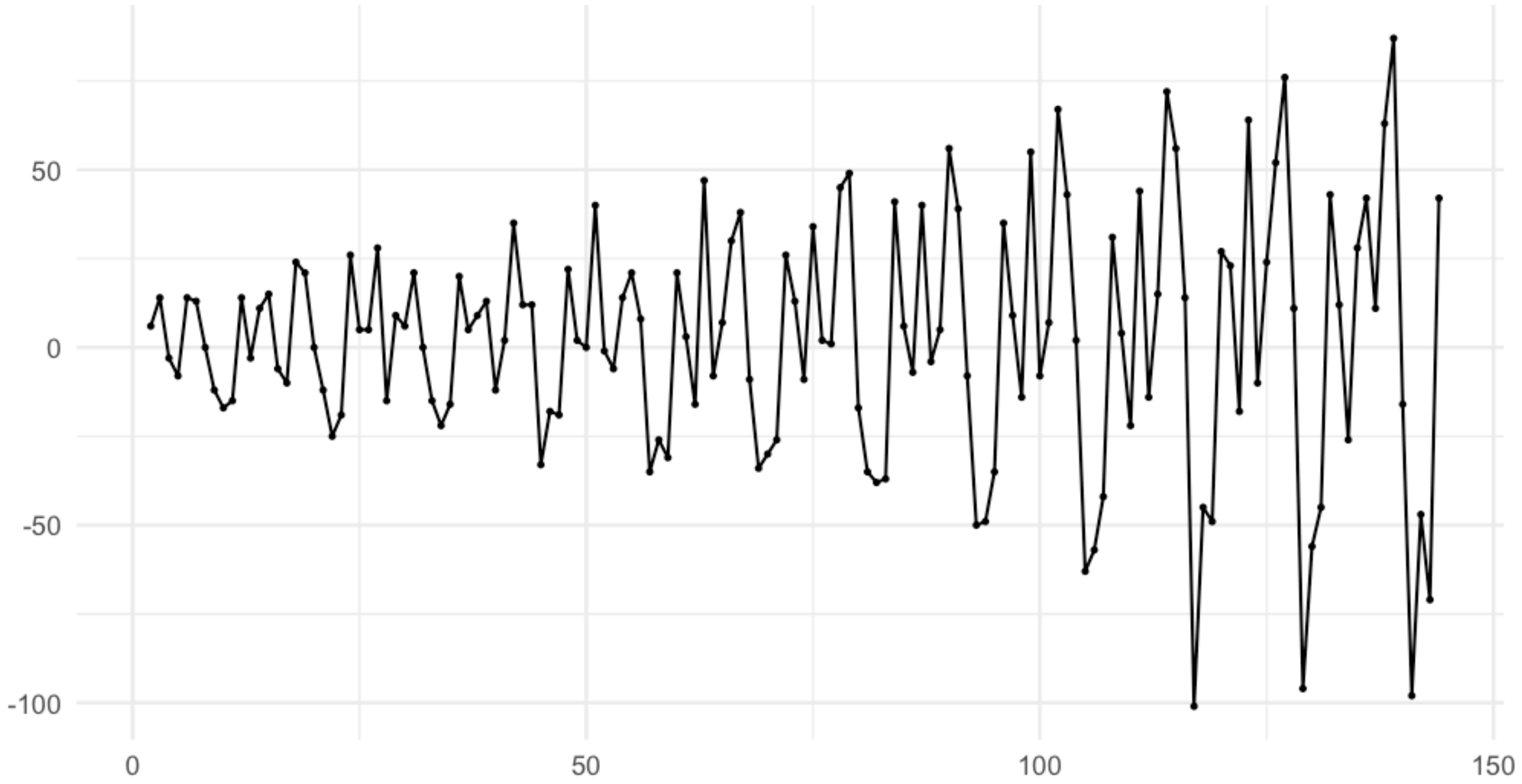
donde $BX_t = X_{t-1}$

- Álgebra de ∇ y B
  - $B^j X_t = X_{t-j}$
  - $\nabla^j X_t = \nabla(\nabla^{j-1} X_t)$ con $\nabla^0 X_t = X_t$
- Por ejemplo:

$$\nabla^2 X_t = (1 - B)(1 - B)X_t = X_t - 2BX_t + B^2 X_t = X_t - 2X_{t-1} + X_{t-2}$$

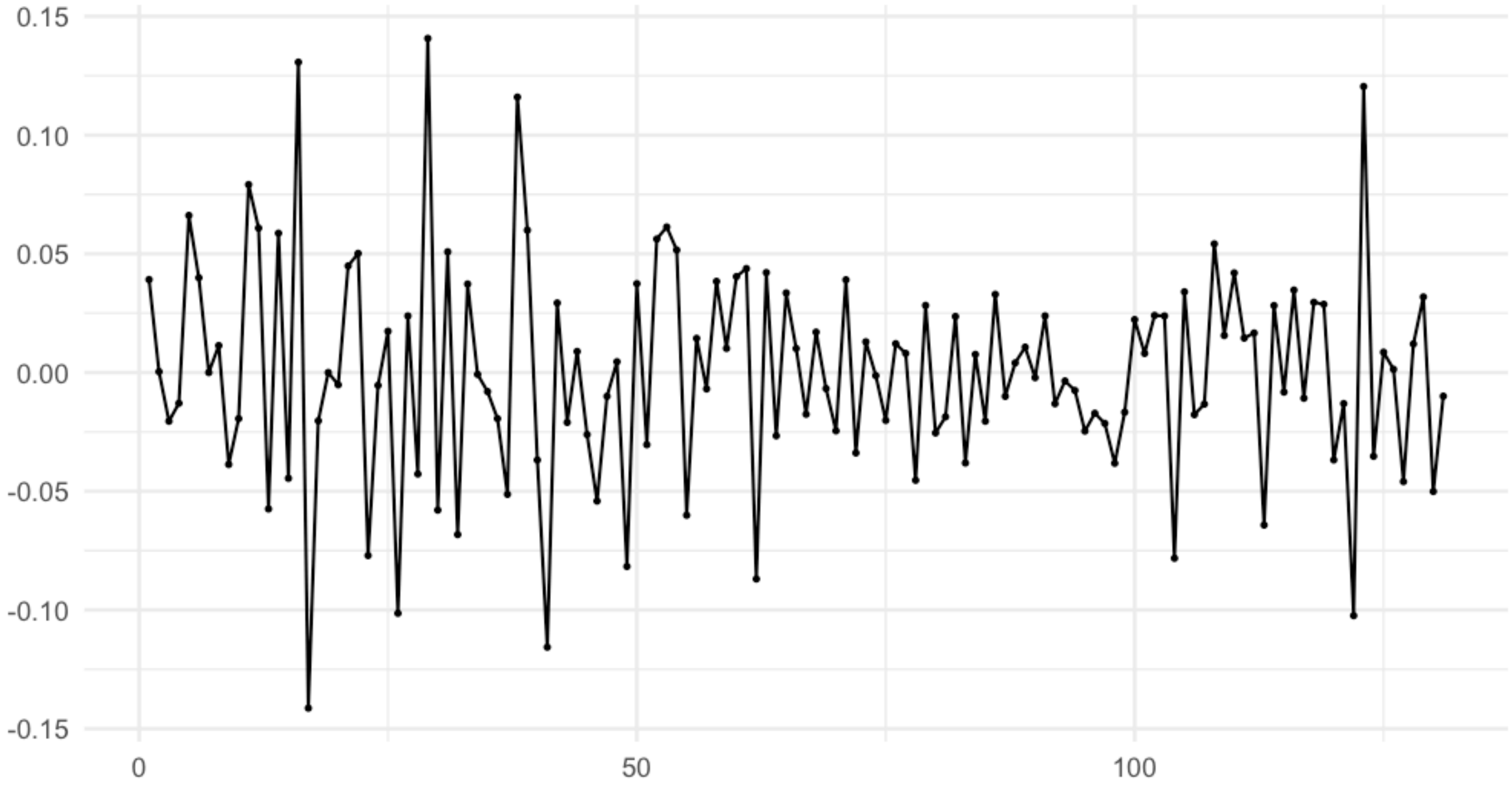
Diferencia de lag 1

- No parece ser estacionario



Diferencias mejoradas

- Se puede mejorar estabilizando la varianza con logaritmo y combinando operadores



Estimación de m_t con componente estacional s_t

Método 1: Suavizamientos

- Método 'clásico' (función decompose en R)

1. Estimar m_t con un filtro de promedios móviles
2. Construir serie sin tendencia, $d_t = X_t - m_t$
3. Estimar la estacionalidad de d_t promediando por año en cada punto estacional
4. Construir la serie $y_t = x_t - \widehat{m}_t - \hat{s}_t$

- Brockwell y Davis (no hay función en R que lo haga)

1. Estimar s_t mediante los promedios por punto estacional
2. Construir serie sin estacionalidad, $d_t = X_t - s_t$
3. Estimar la tendencia de d_t
4. Construir la serie $y_t = x_t - \widehat{m}_t - \hat{s}_t$

Método 1: Suavizamientos

Opciones más avanzadas y modernas

- Suavizamiento exponencial mediante modelos de espacio de estados
  - Función ets de R
  - Hyndman (2008). Forecasting with exponential smoothing: The state space approach.

- Suavizamiento Loess (Locally Estimated Scatterplot Smoothing)
  - Función stl de R
  - Cleveland et al. (1990). STL: A Seasonal-Trend Decomposition Procedure Based on Loess.

Método 2: Operador diferencias

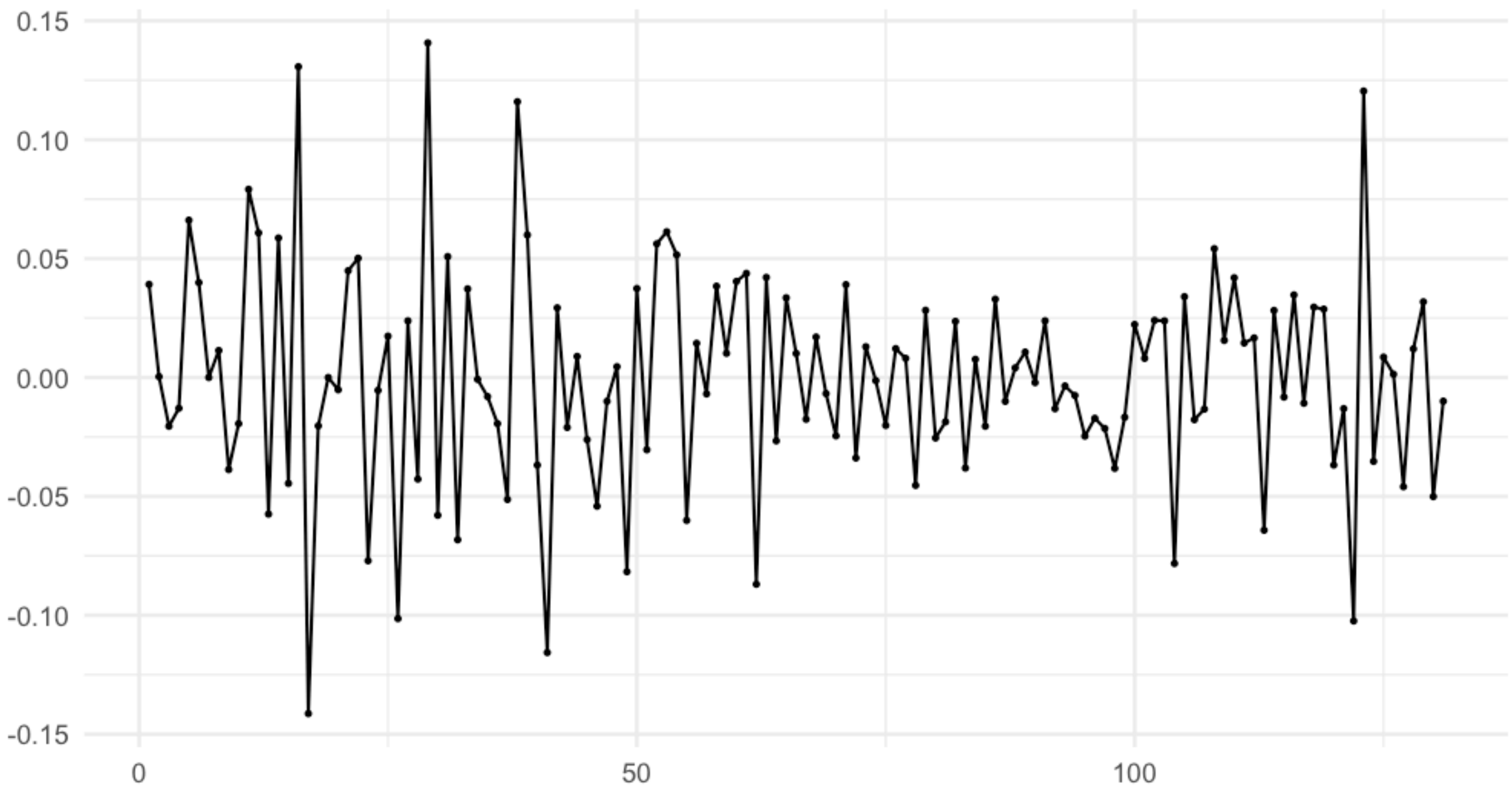
- Para remover estacionalidad utilizar el operado:

$$\nabla_d X_t = (1 - B^d)X_t = X_t - X_{t-d} = m_t - m_{t-d} + Y_t - Y_{t-d}$$

- Eliminar el nuevo componente de tendencia $m_t - m_{t-d}$ con algún operador ∇^k

Diferencias mejoradas

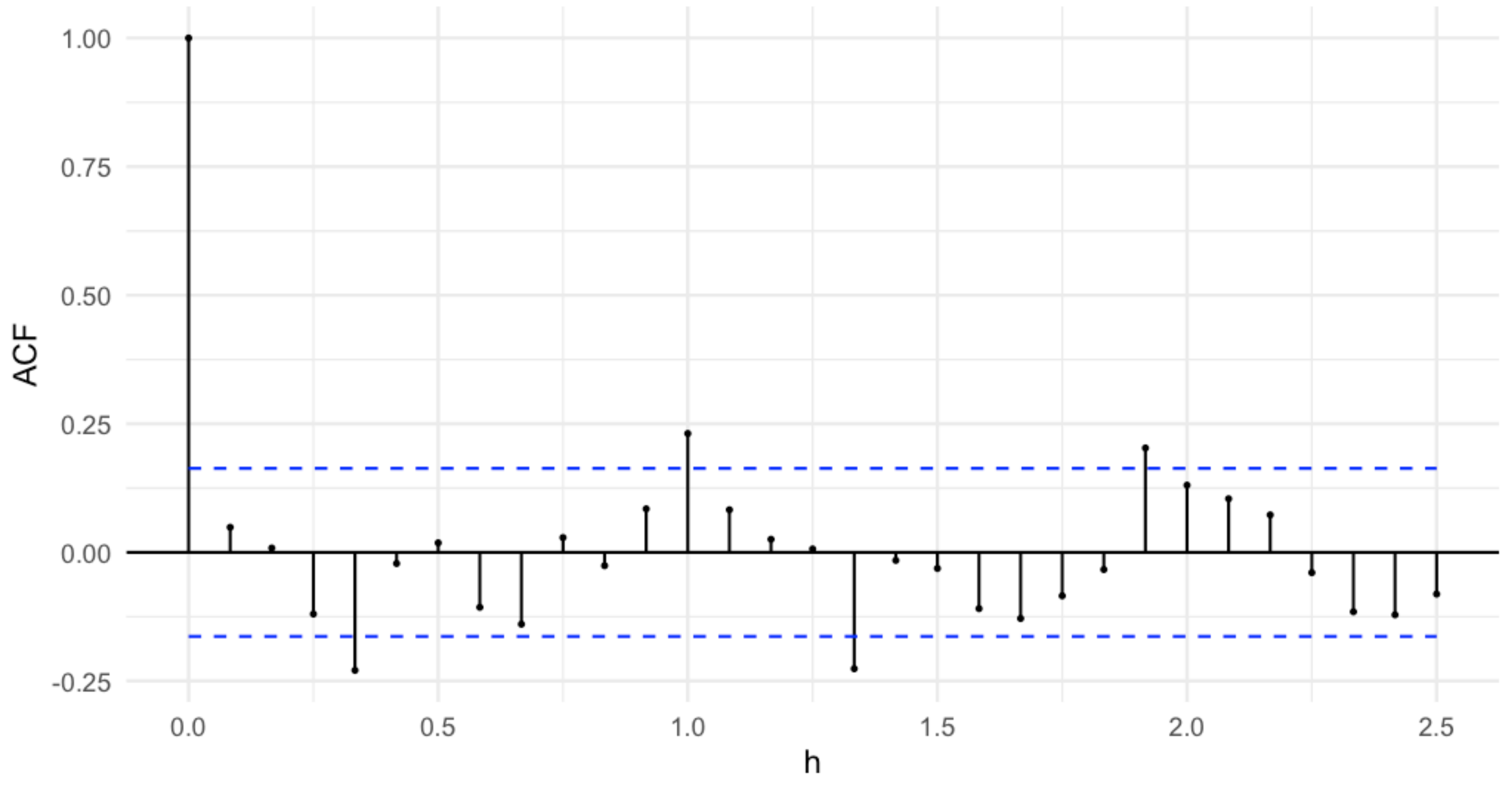
- Aplicar logaritmo a la serie y utilizar ∇_{12} para el componente estacional y ∇ para el componente de tendencia



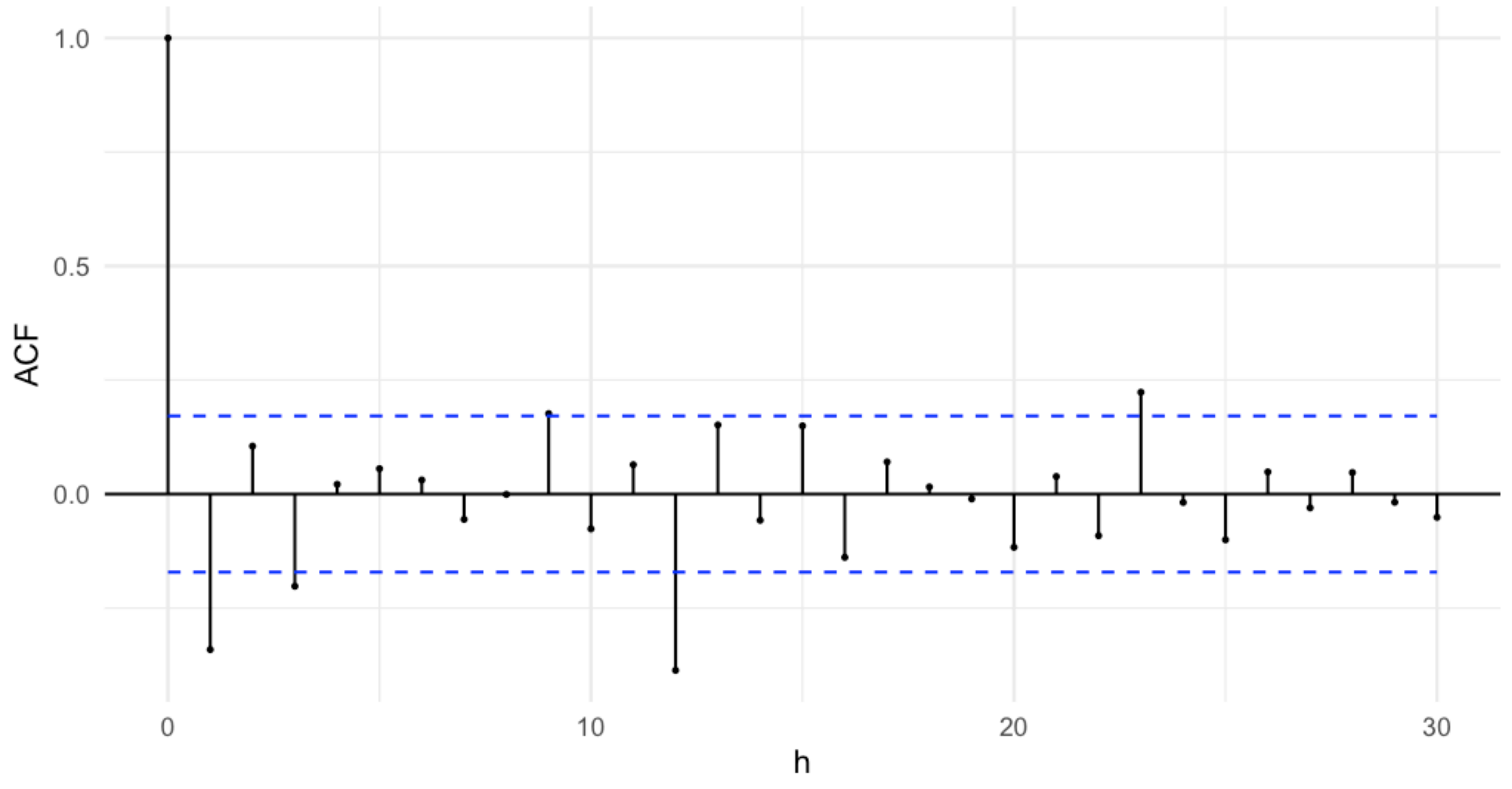
Análisis de residuales

ACF

- Para n suficientemente grande las autocorrelaciones muestrales $\widehat{\rho}(j)$ de los residuales deben ser $\mathcal{N}(0,n^{-1})$
- Graficar el acf y checar que no haya muchos valores fuera de las bandas $(\pm 1.96/\sqrt{n})$ o un valor extremadamente grande



Suavizamiento exponencial



Operador diferencias

Prueba de Portmanteau

- Para n suficientemente grande las autocorrelaciones muestrales $\hat{\rho}(j)$ de los residuales deben ser $\mathcal{N}(0,n^{-1})$

- Bajo la hipótesis nula

$$Q = n \sum_{j=1}^{h} \hat{\rho}(j)^2 \sim \chi_h^2$$

- Rechazamos para valores grandes de Q

- No es muy utilizada en la práctica

Prueba de Ljung-Box

- Modificación de la prueba de Portmanteau, donde bajo la hipótesis nula

$$Q_{LB} = n(n+2)\sum_{j=1}^{h}\frac{\hat{\rho}(j)^2}{n-j} \sim \chi^2_h$$

- Rechazamos para valores grandes de Q_{LB}

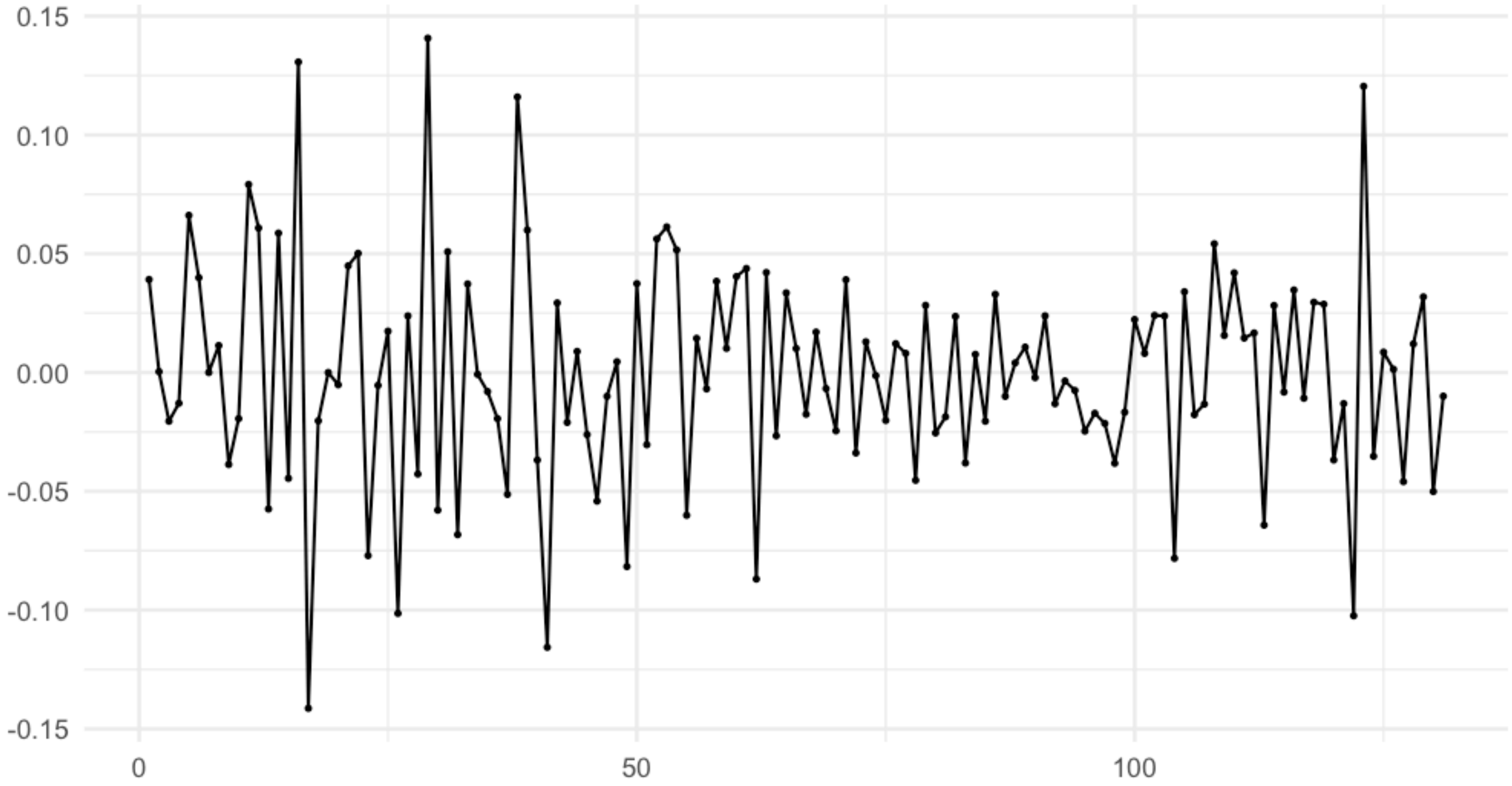
Residuales	Q_{LB}	**p-valor**
Suavizamiento Exponencial	0.34999	0.5541
Operador Diferencias	15.596	7.843E-05

- Los residuales del suavizamiento ya son ruido estacionario pero los del operador de diferencias no

Modelos ARMA

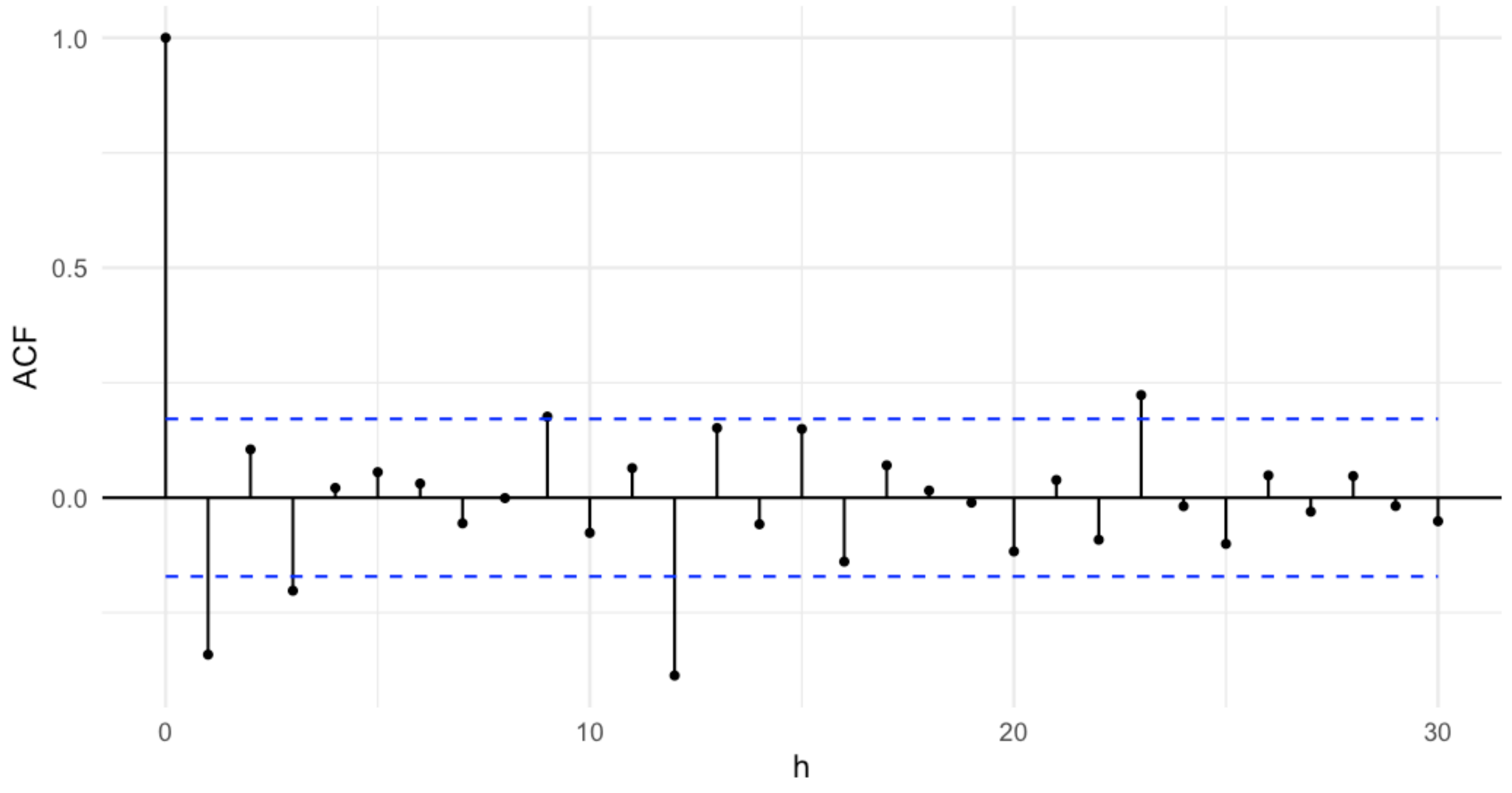
Diferencias mejoradas

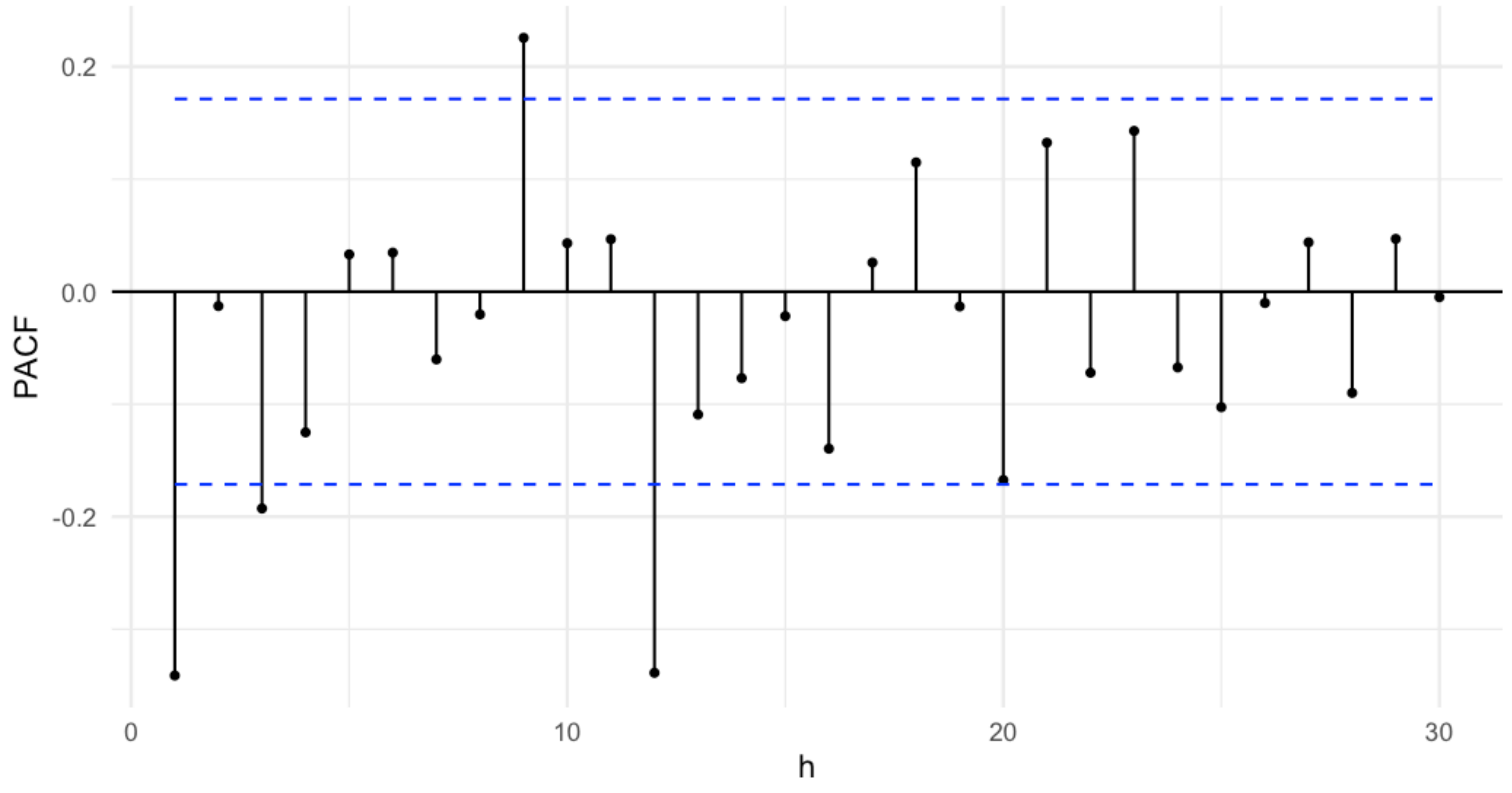
- Aplicar logaritmo a la serie y utilizar ∇_{12} para el componente estacional y ∇ para el componente de tendencia



Modelos ARMA

- A partir del ACF y PACF, ¿qué modelos ARMA se pueden ajustar?





Modelos ARMA

- ARMA(12,12)
- ARMA(1,1)
- AR(12)
- AR(1)
- MA(12)
- MA(1)

ARMA(1,1)

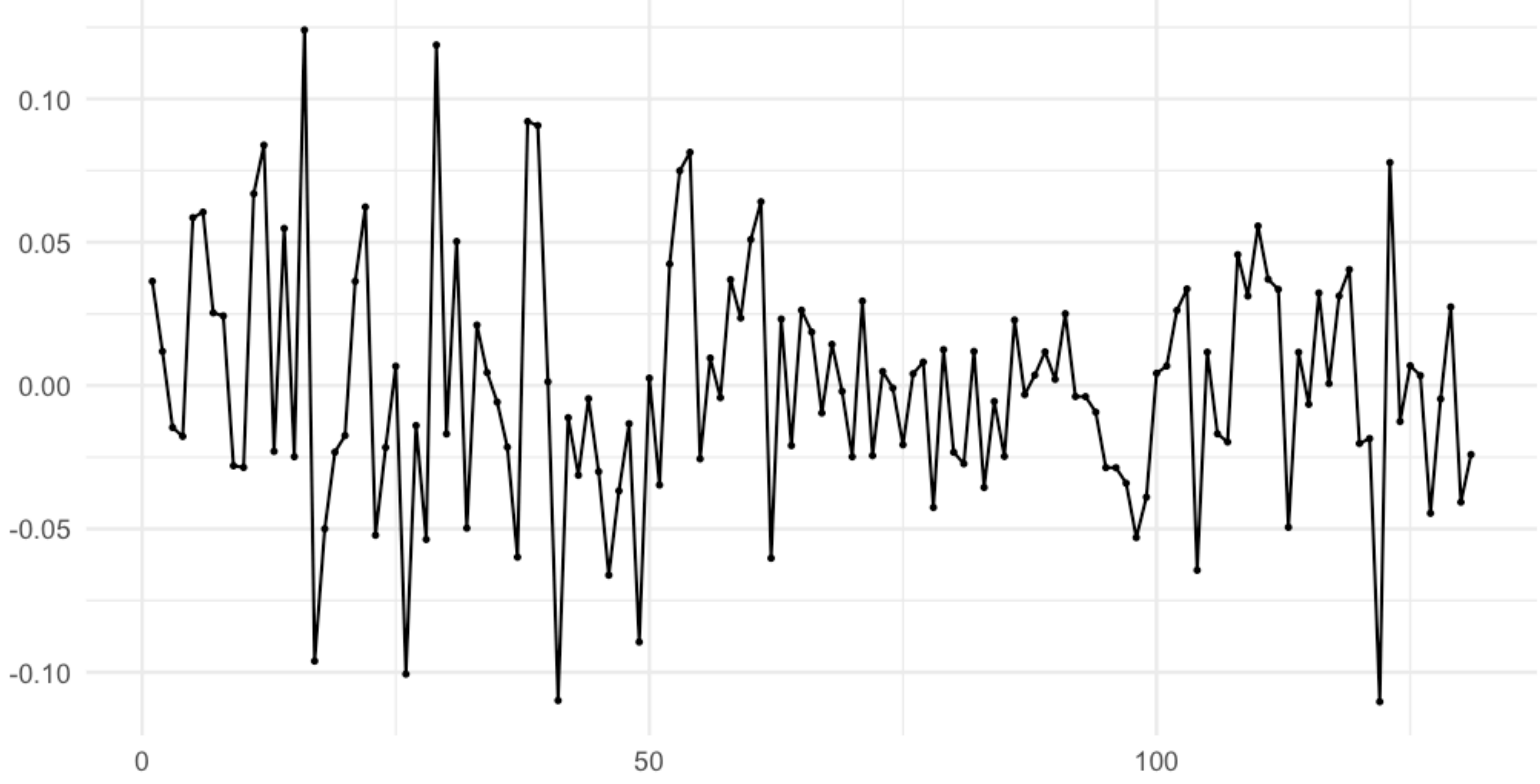
- Los parámetros estimados son

$$\phi = 0.1448$$
$$\theta = -0.5190$$
$$\sigma^2 = 0.001824$$

- AIC de -446.27

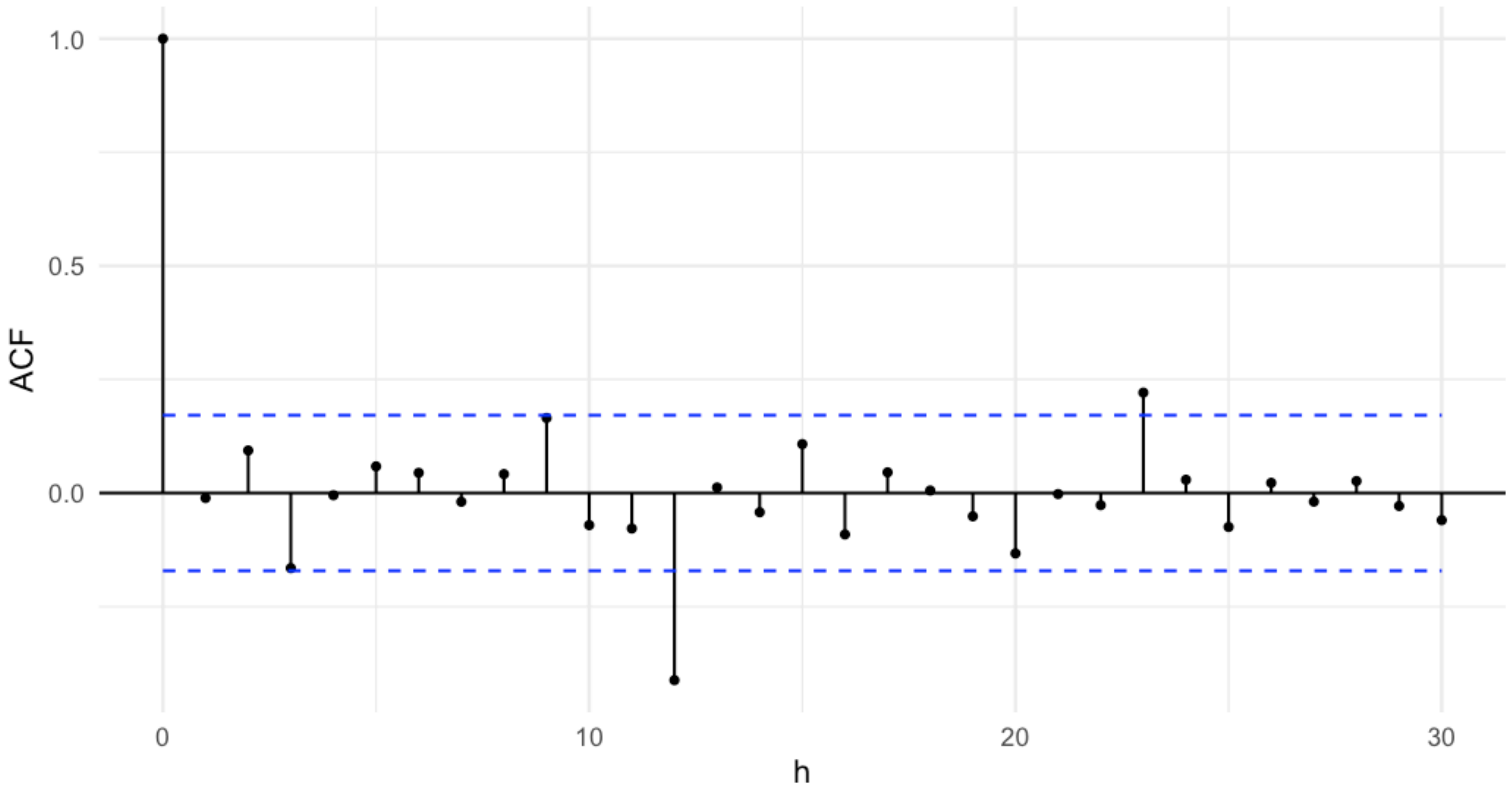
ARMA(1,1)

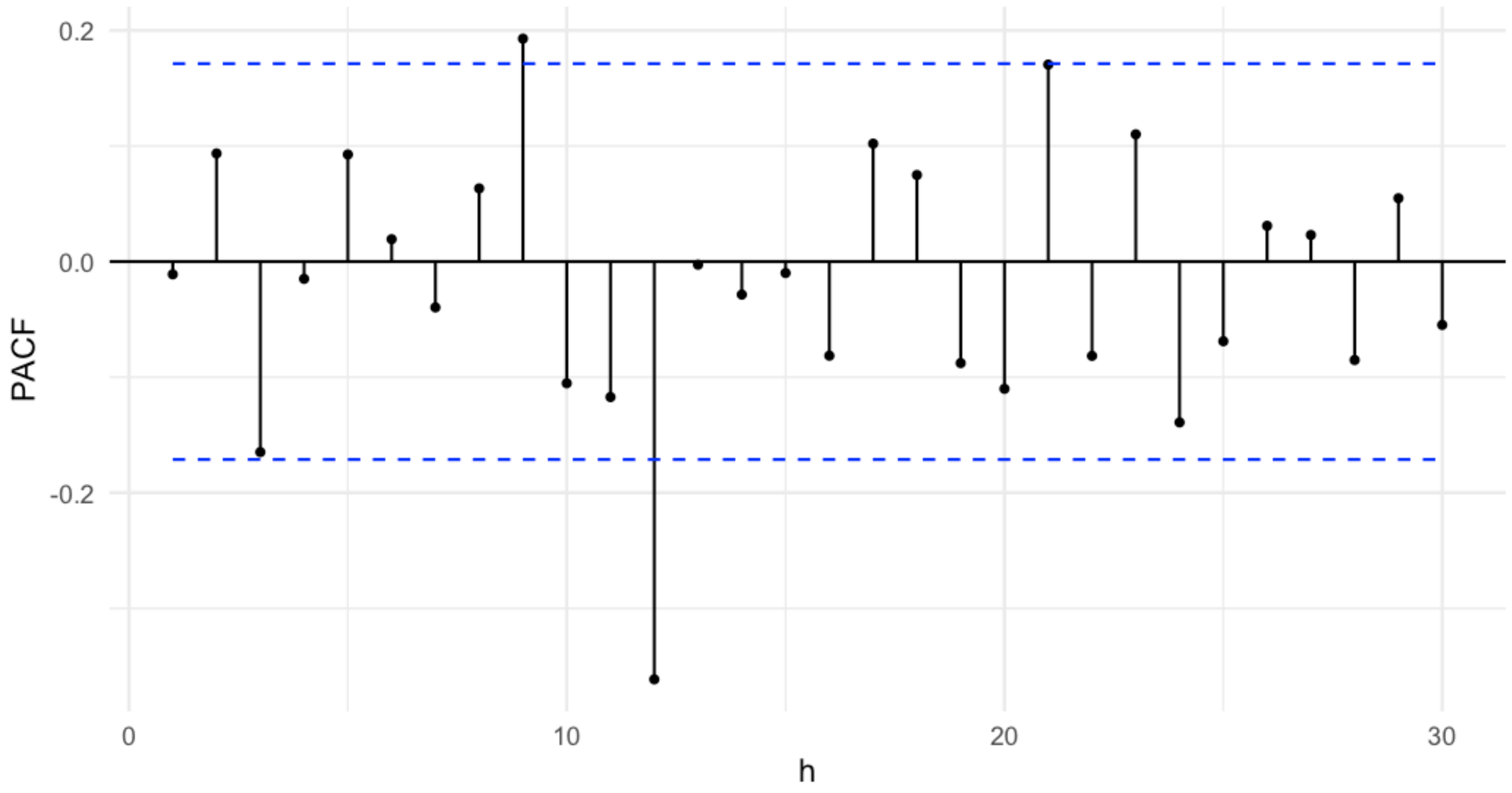
- Residuales



ARMA(1,1)

- Residuales

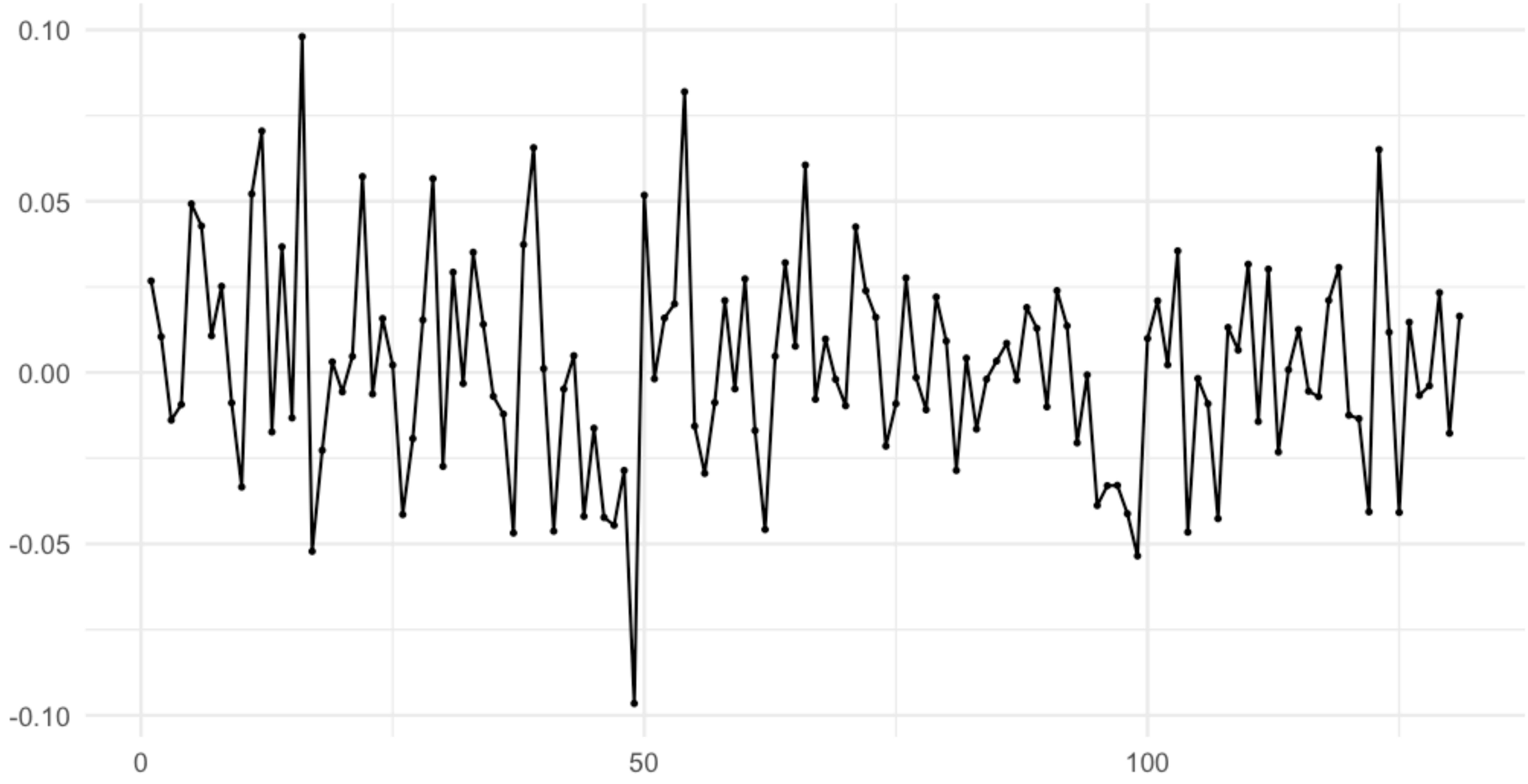




- p-value de Ljung-Box de 0.8985

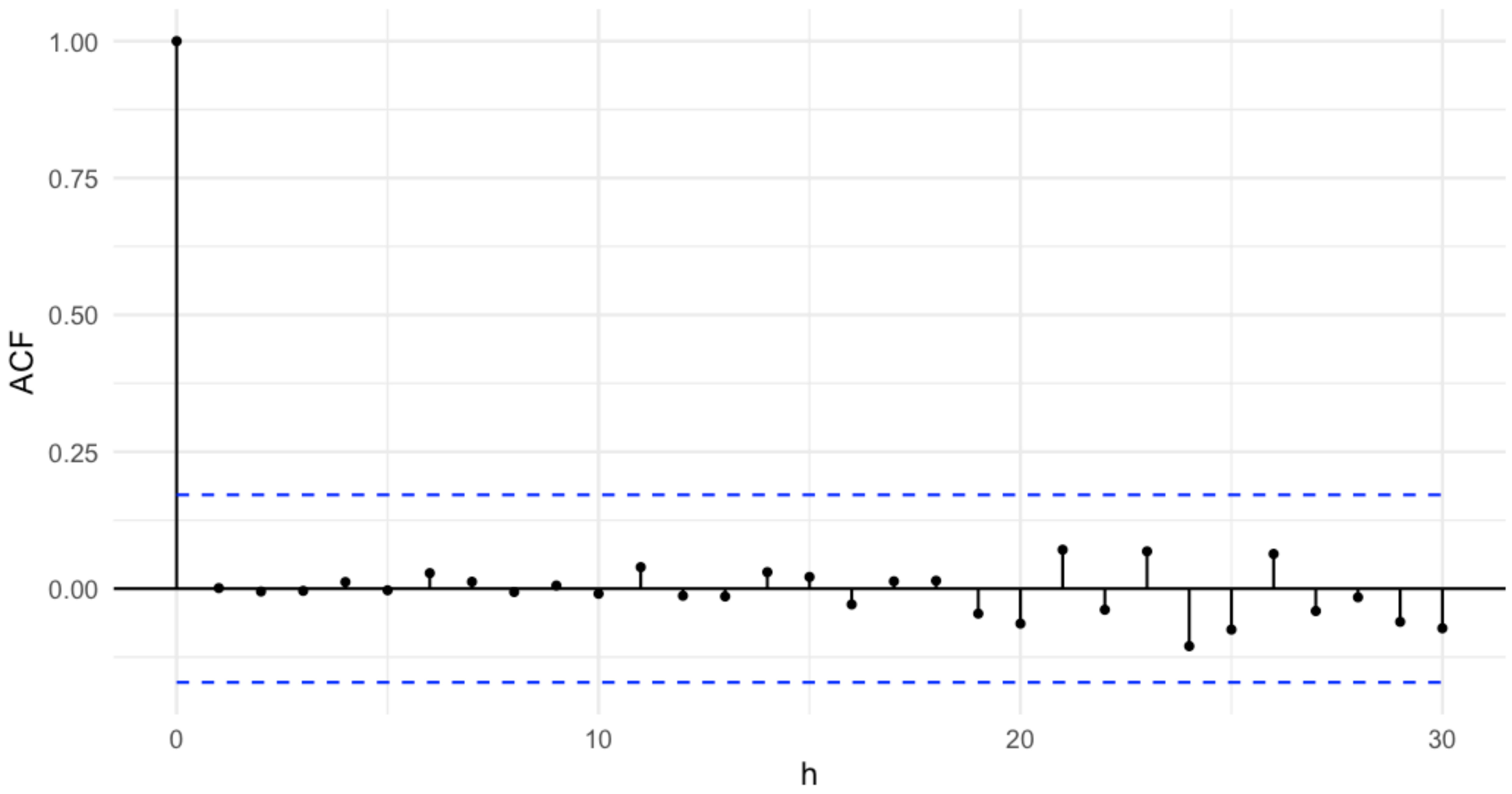
ARMA(12,12)

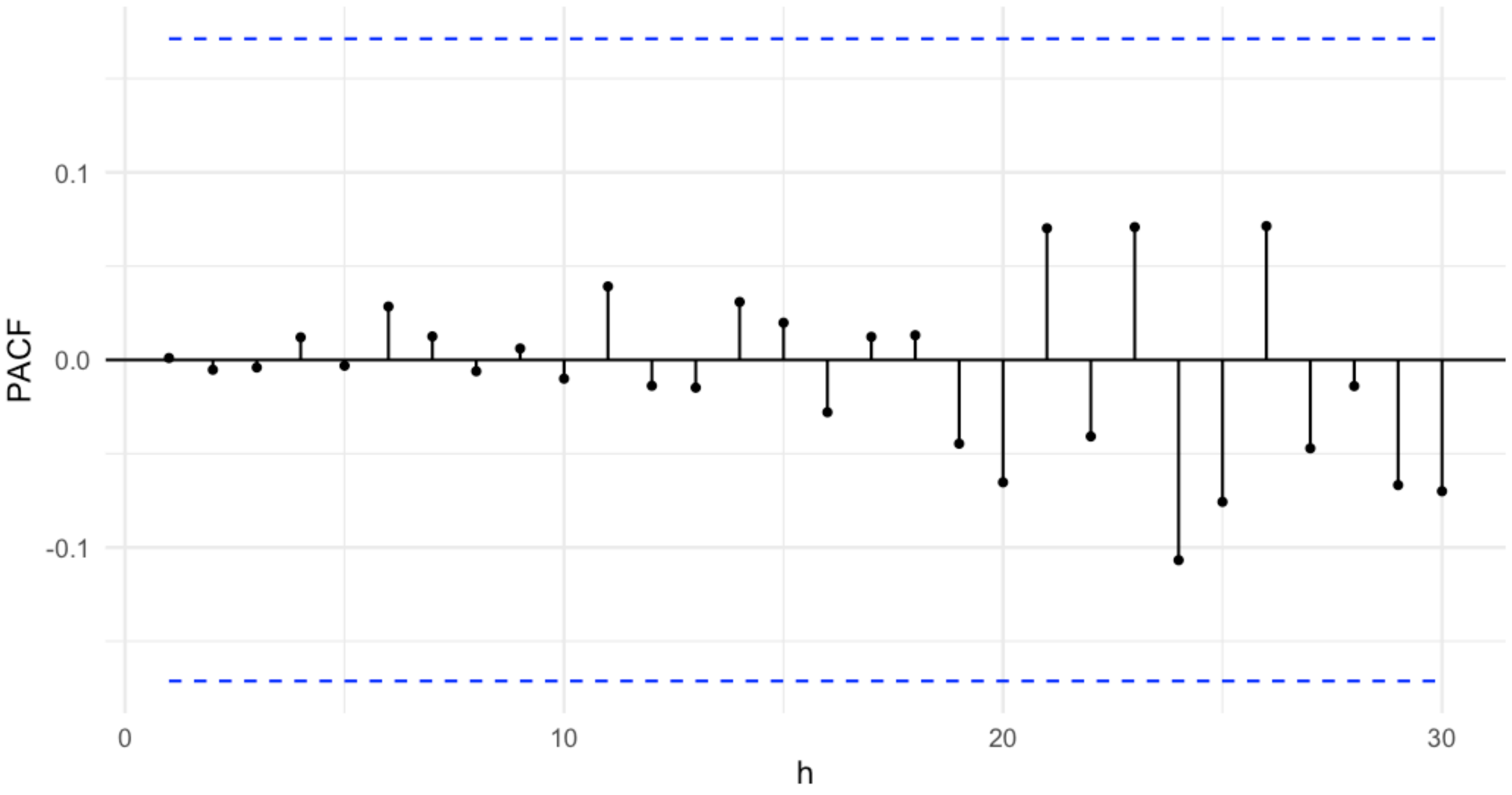
- ¡Muchos parámetros (posiblemente innecesarios)!
- AIC de -467.34
- Residuales



ARMA(12,12)

- ACF y PACF

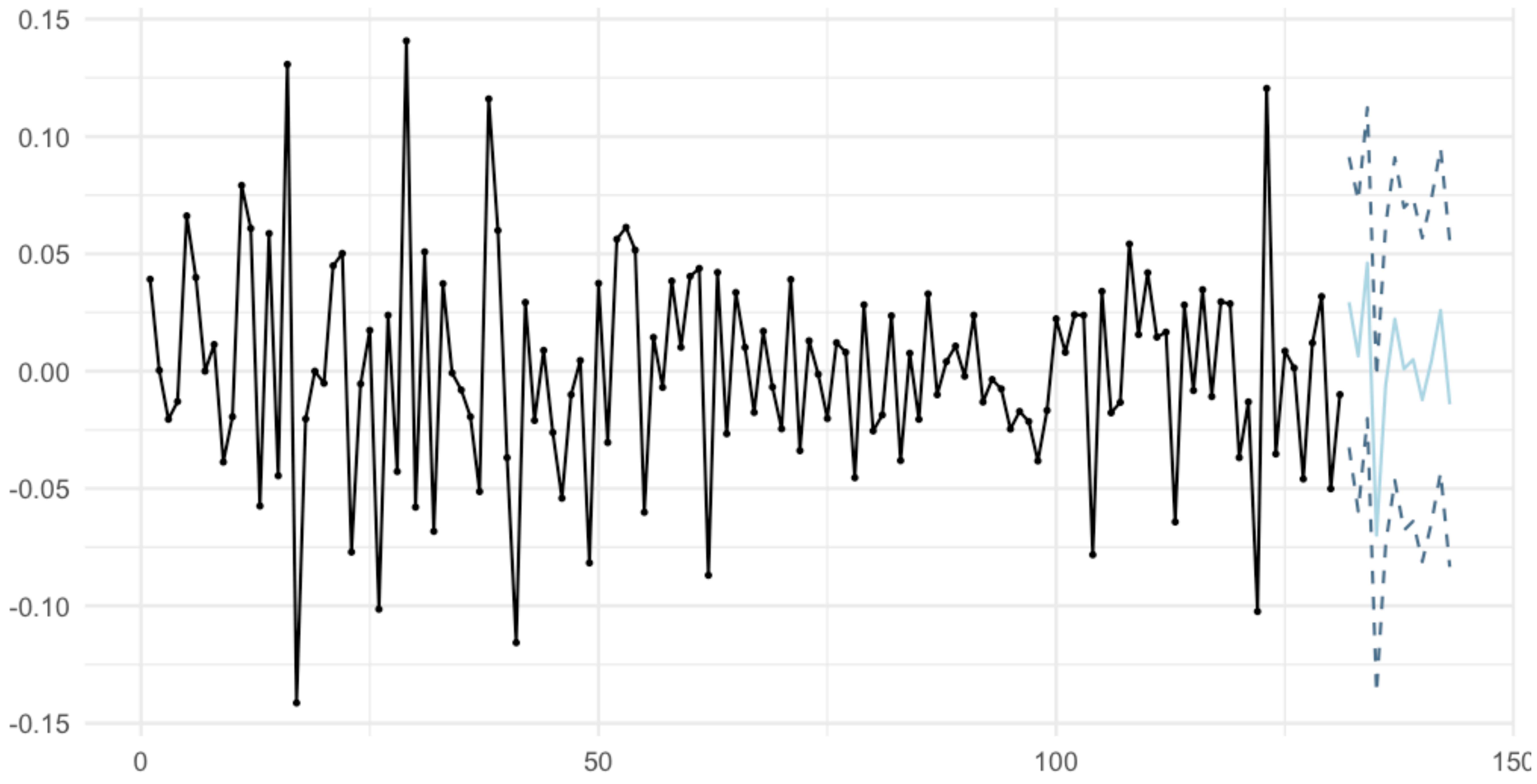




- p-value de Ljung-Box de 0.9907

ARMA(12,12)

- Predicción

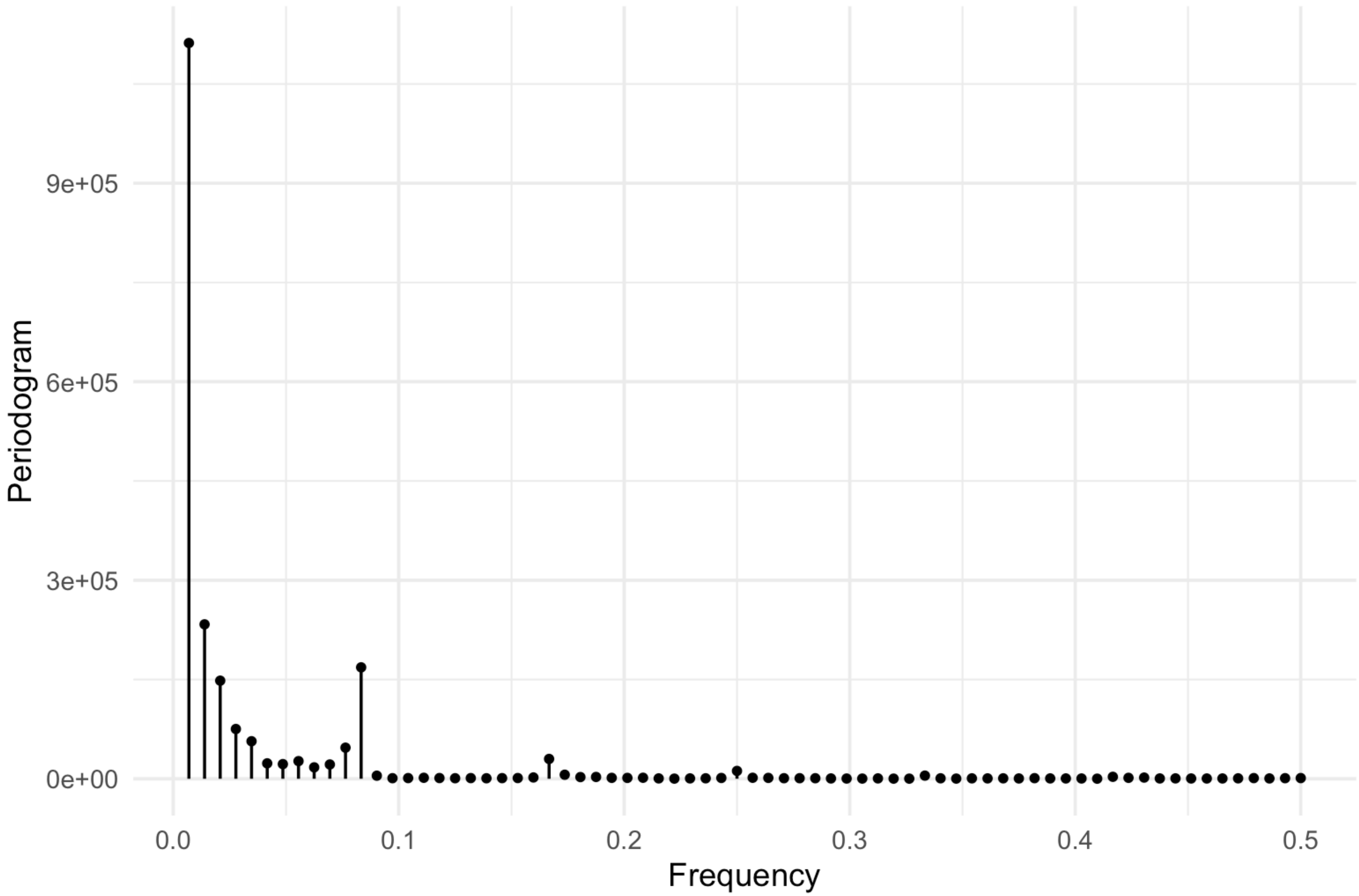


- ¡Se tienen que revertir todas las transformaciones! (Proceso tedioso)

Modelos SARIMA

Periodograma

- Se calcula el periodograma para evaluar el ciclo de la serie



Periodograma

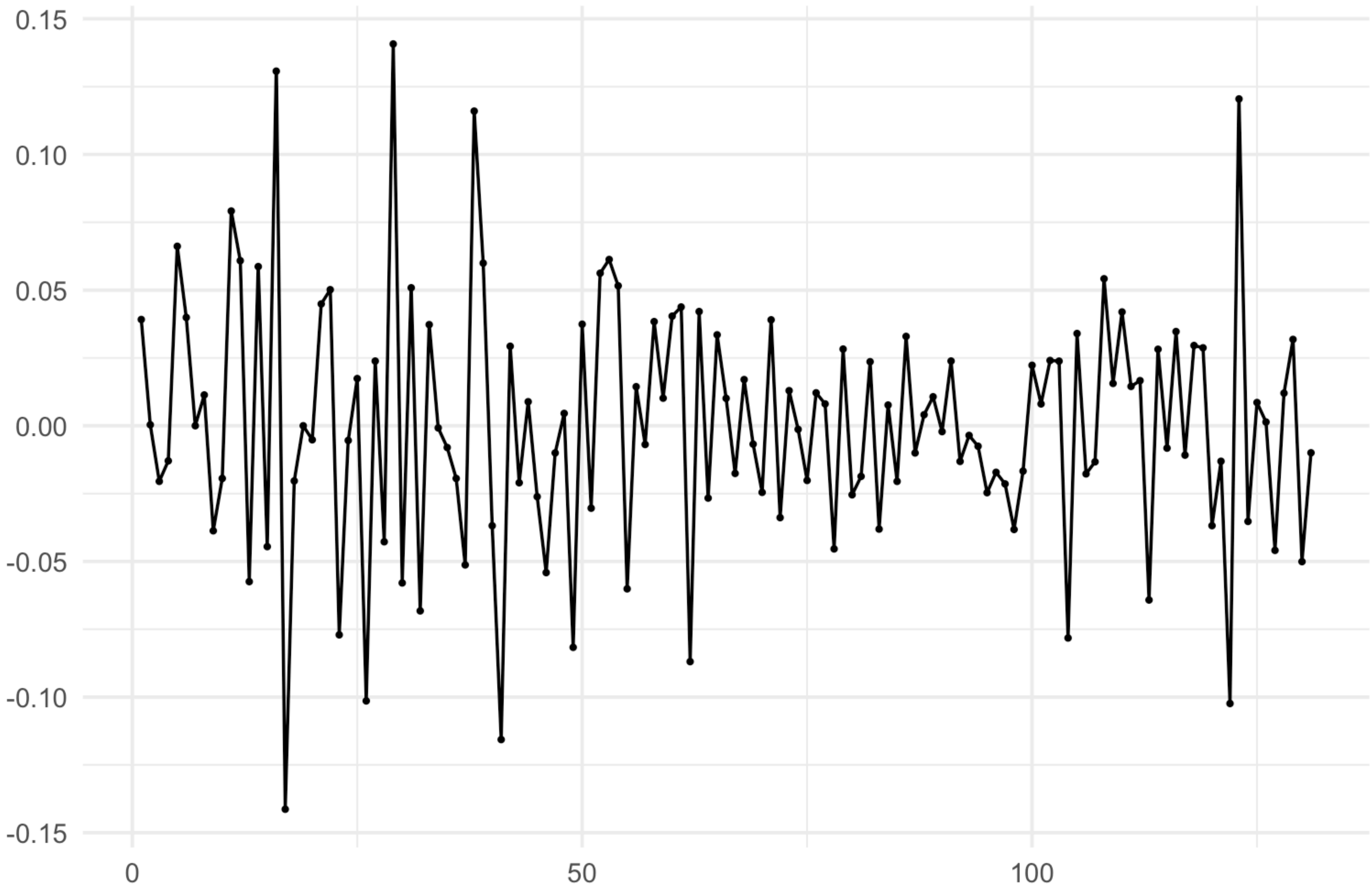
- Las 4 frecuencias dominantes están en las posiciones 1,2,12,3

Posición	Frecuencia	Ciclo
1	0.006944444	144
2	0.01388889	72
12	0.08333333	12
3	0.02083333	48

- Tiene picos en frecuencias bajas (se puede deber a la tendencia)

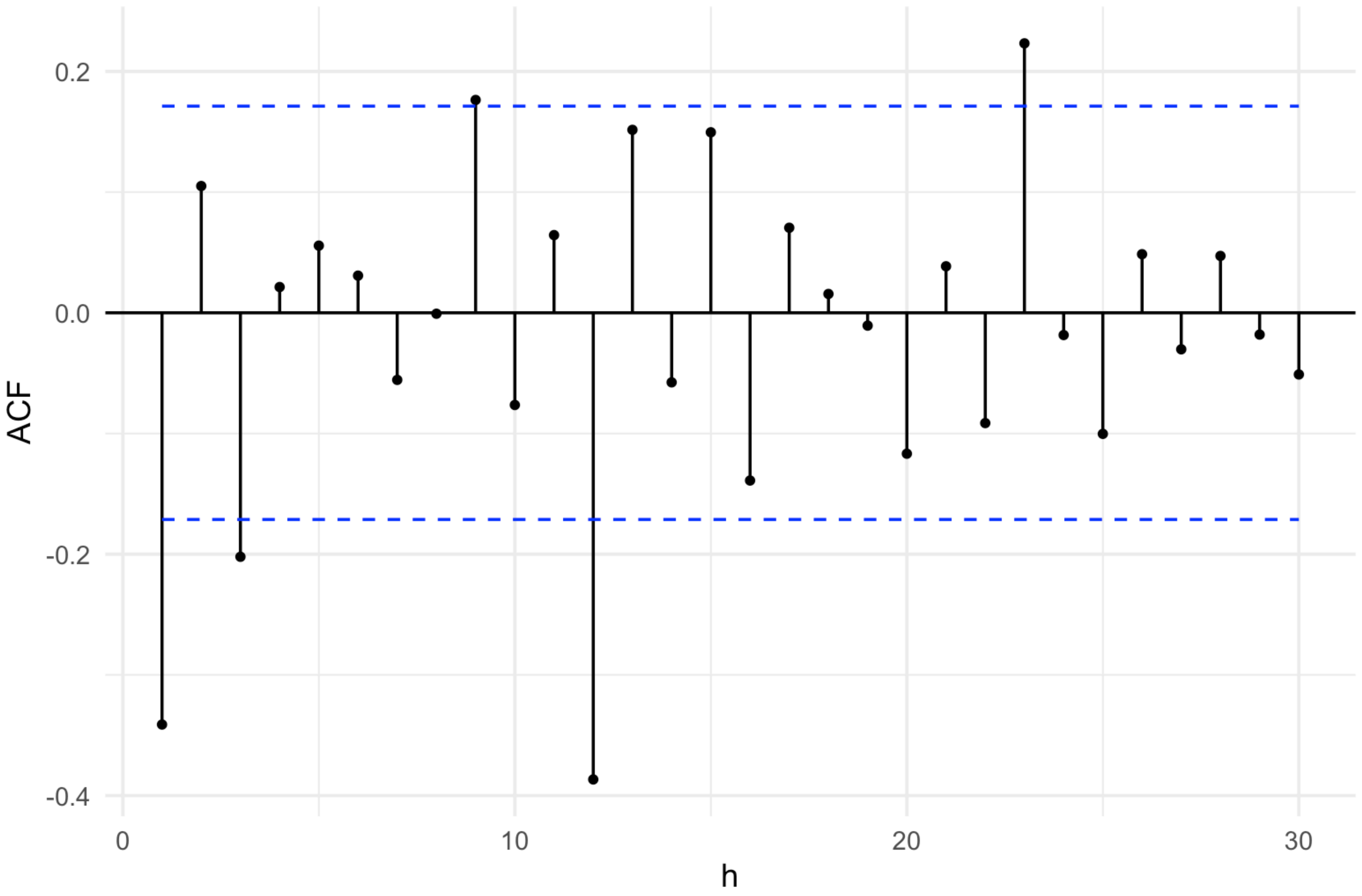
Diferenciando

- Generamos el proceso $Y_t = (1 - B)(1 - B^{12})X_t$



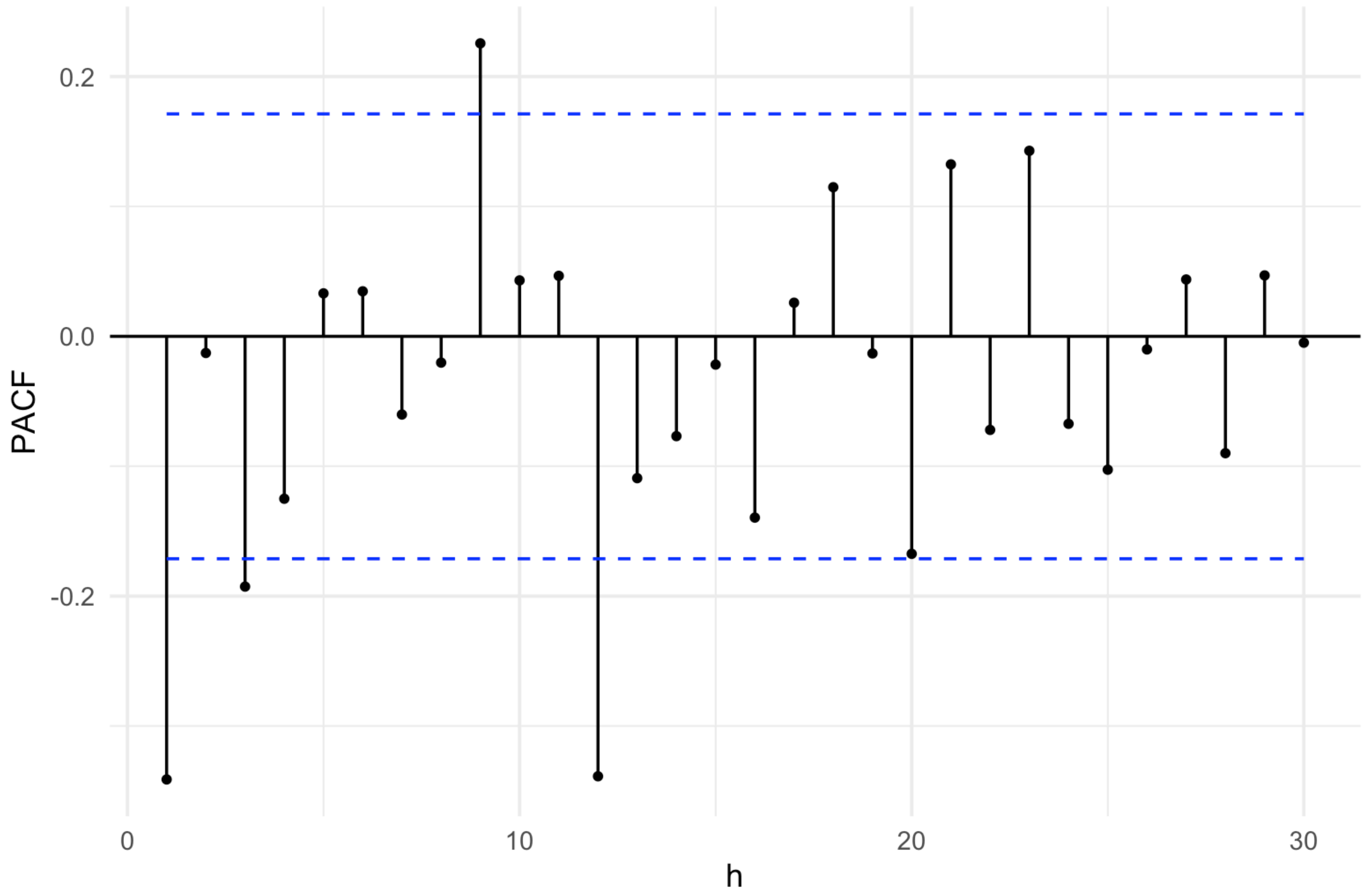
ACF

- ¿Qué valores para q y Q ? (Sol: $q = Q = 1$)



PACF

- ¿Qué valores para p y P? (Sol: $p = P = 1$)



Modelo

- X_t es un proceso SARIMA $(1,1,1) \times (1,1,1)_{12}$

- $Y_t = (1 - B)(1 - B^{12})X_t$ es un proceso ARMA causal tal que

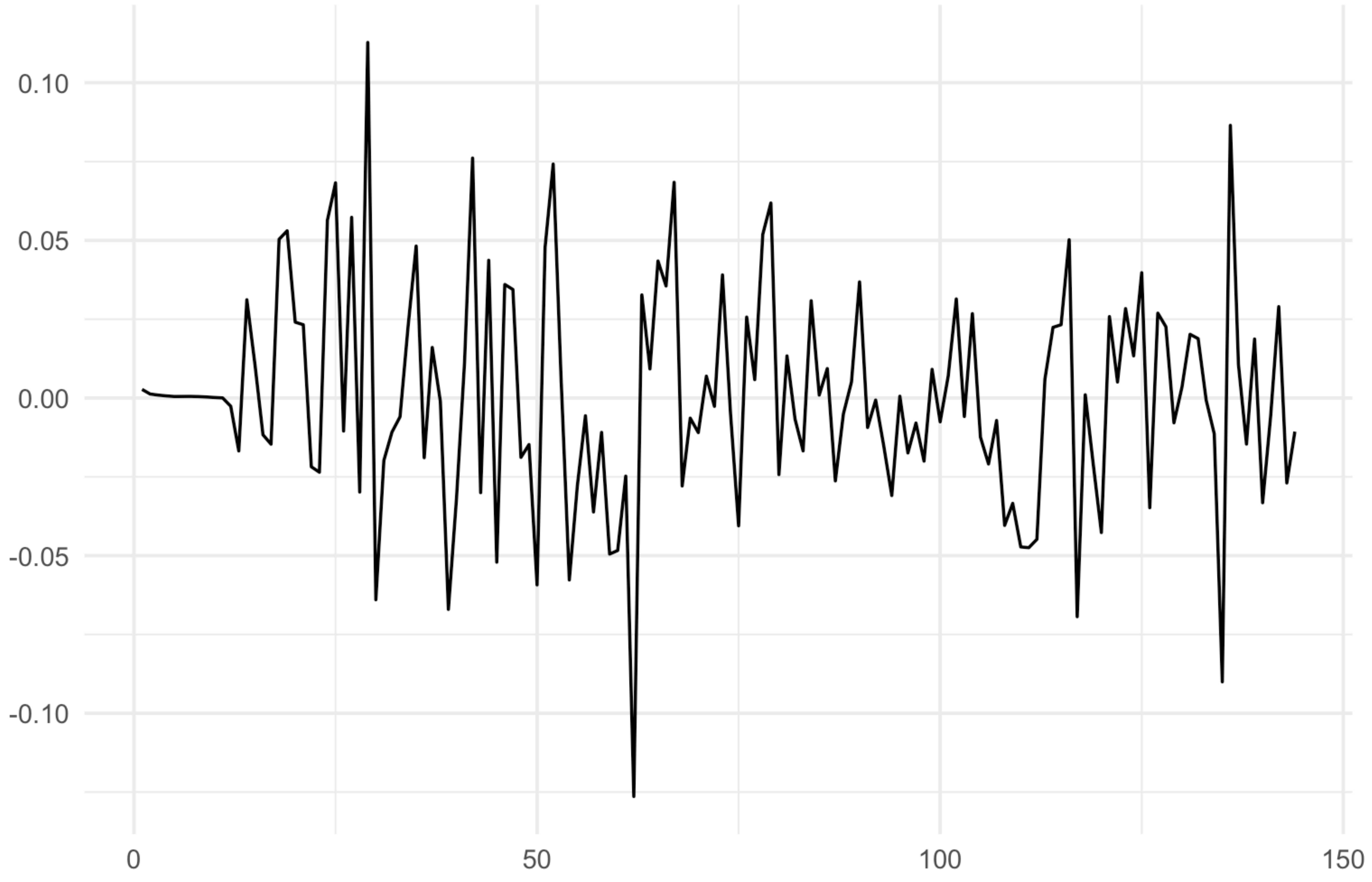
$$(1 - \phi B)(1 - \Phi B^{12})Y_t = (1 + \theta B)(1 + \Theta B^{12})Z_t$$

ϕ	Φ	θ	Θ	σ
0.1666	-0.099	-0.5615	-0.4973	0.03714835

- AICc de -479.83 y BIC -465.93

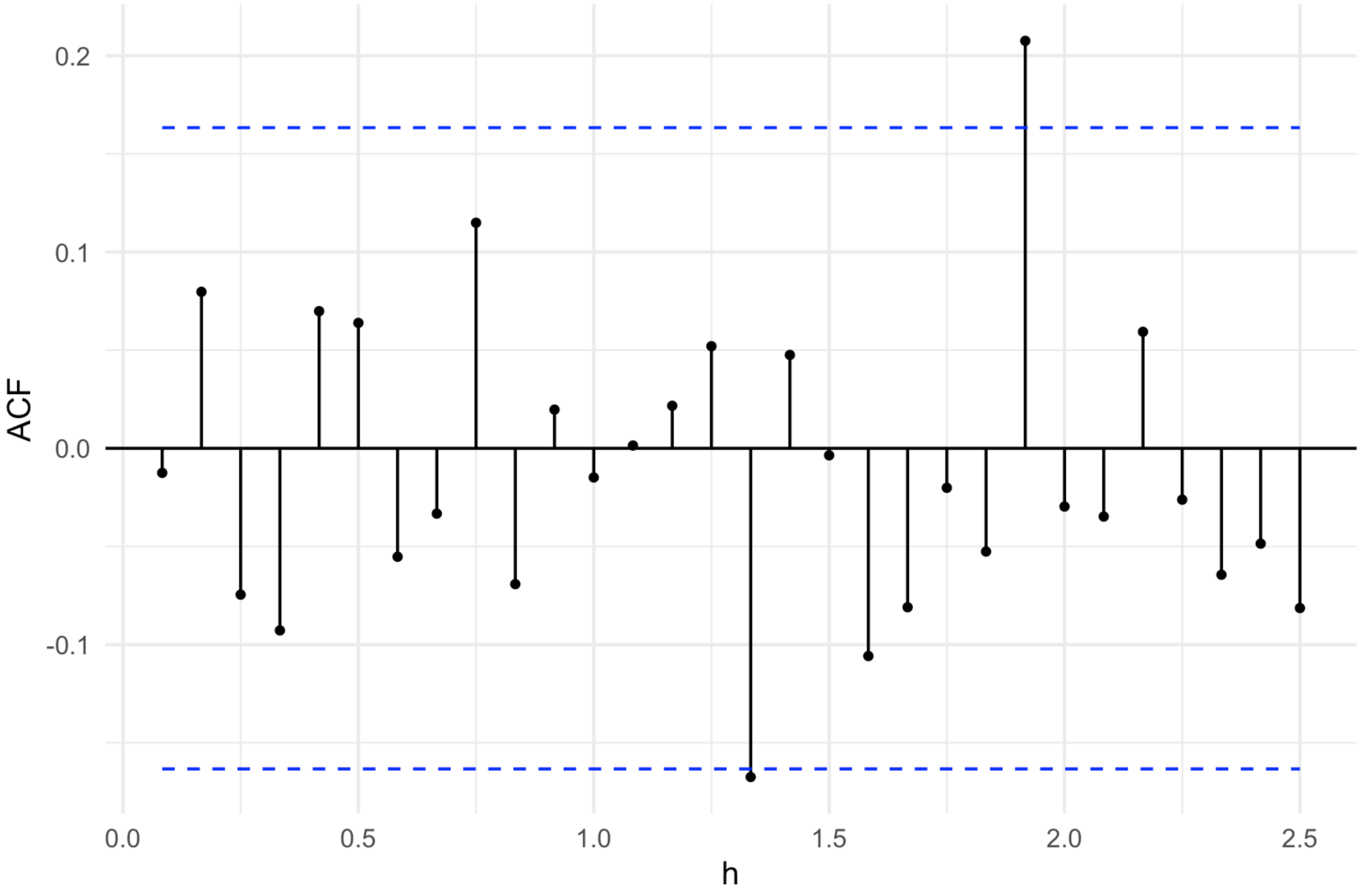
Residuales

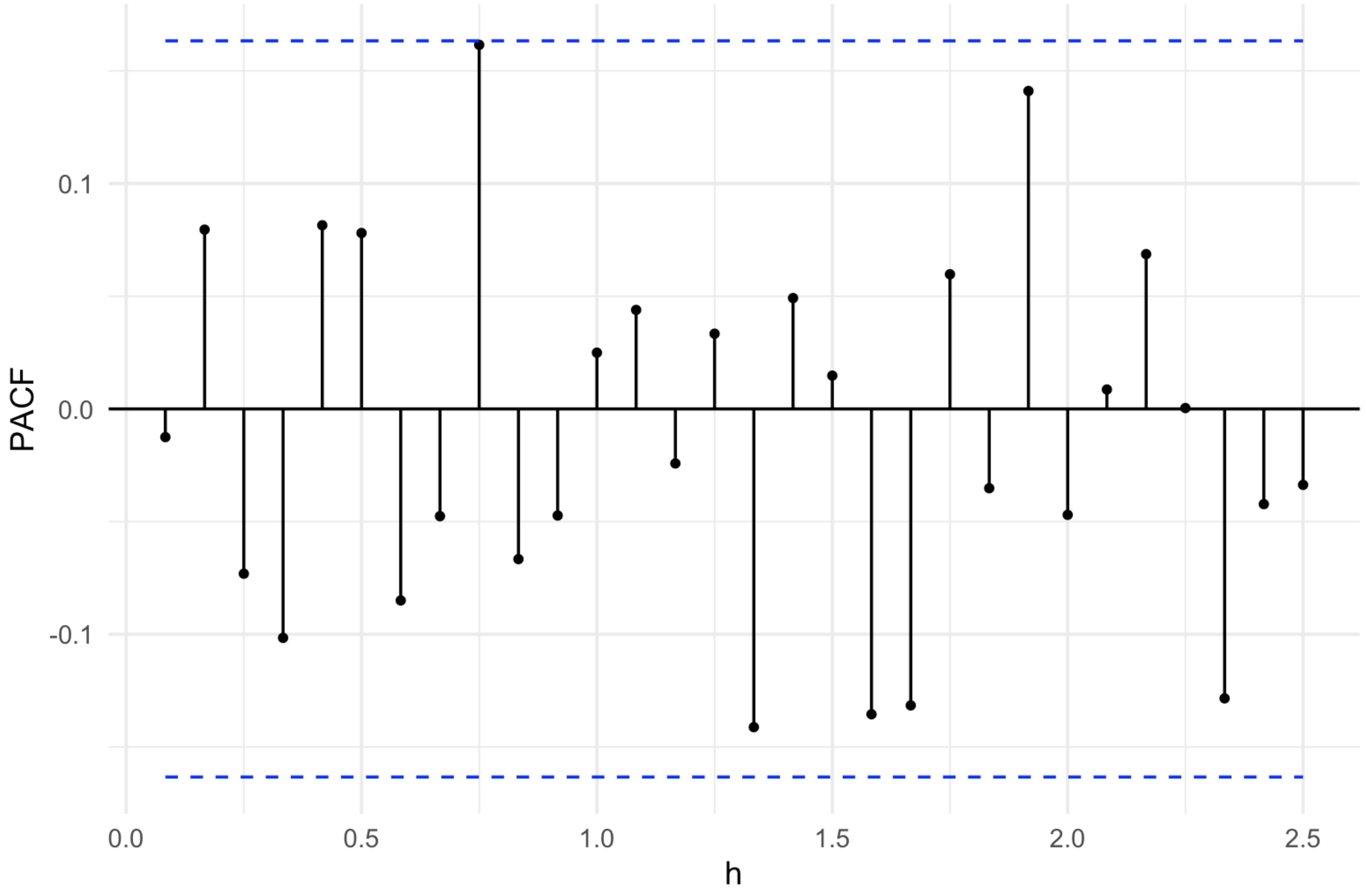
- Parece ser estacionario



Residuales

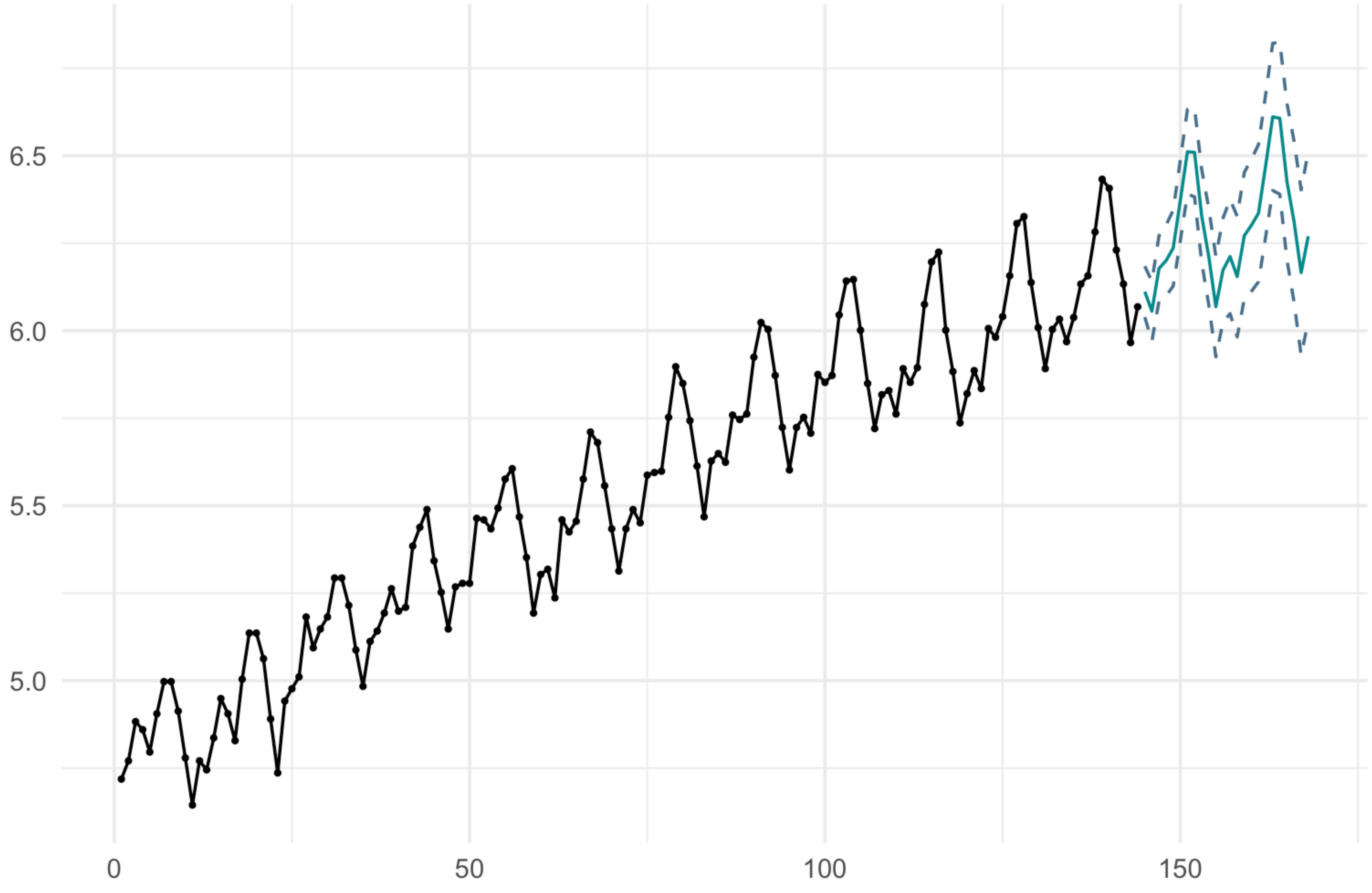
- Prueba de Ljung-Box : p-valor = 0.6863





Predicción

- En escala logarítmica



Predicción

- Regresando a escala original