

PONTIFICIA UNIVERSIDAD JAVERIANA

Facultad de Ciencias

Departamento de Matemáticas

Carrera: Ciencia de Datos

Trabajo de grado

Estimación de Series de Tiempo no Gaussianas

Juan José Rodríguez Cubillos

Tutor: Carlos Alberto Cardozo Delgado

Bogotá D.C., Colombia
Octubre 2025

1. Agradecimientos

Primero que todo quiero agradecer al profesor Carlos, cuyo apoyo constante y beneficioso me ayudaron a llevar a cabo este proyecto. A la mayoría de profesores que me han dictado clases en estos 8 semestres, pues con su saberes y experiencias nutrieron mi saber y me prepararon para este proyecto y próximos que vendrán mas adelante. A mi familia gracias por el apoyo constante no solo en la universidad sino en cada una de mis ideas y proyectos, este trabajo es solo el comienzo de la cosecha de frutos de todo el esfuerzo realizado hasta ahora. Finalmente, pero no menos importante, a mis amigos, les agradezco todo lo que vivido con ustedes en estos casi 4 años, repetiría cada decisión que he tomado en esta vida solo para terminar conociéndolos y pasar por todo lo que hemos pasado hasta ahora, para ustedes mi mayor agradecimiento, respeto y admiración, y ojala poder seguir creciendo juntos como profesionales y como personas.

2. Introducción

El análisis y pronóstico de series de tiempo no gaussianas es una temática de creciente interés en diversas disciplinas como la economía, la ingeniería, la meteorología y las ciencias sociales. Muchos de los modelos tradicionales utilizados para la estimación y predicción de estos procesos, como los modelos autorregresivos de media móvil (ARMA), asumen que las innovaciones o perturbaciones de la serie siguen una distribución normal. Sin embargo, en la práctica, muchas series de tiempo presentan características que violan este supuesto, tales como colas pesadas, asimetría o comportamientos más complejos que no pueden ser adecuadamente modelados con distribuciones gaussianas. En este contexto, es necesario explorar y desarrollar modelos alternativos que puedan abordar las características no gaussianas de las series temporales.

El planteamiento del problema radica en la falta de modelos adecuados para tratar series de tiempo cuyos residuos no sigan distribuciones gaussianas, lo que puede generar estimaciones sesgadas e imprecisas si se usan los métodos convencionales. En particular, en aplicaciones financieras, meteorológicas y de ingeniería, la presencia de colas pesadas y asimetría en los datos es un fenómeno común, lo que motiva la necesidad de explorar distribuciones más flexibles, como la Birnbaum-Saunders (BS) y la Sinh-Arcsinh (SAS). Estos modelos no gaussianos han demostrado ser más adecuados para capturar los comportamientos reales de las series temporales que presentan colas pesadas o asimetría.

El objetivo principal de este trabajo es desarrollar y evaluar un enfoque estadístico para modelar y pronosticar series de tiempo utilizando distribuciones no gaussianas. Específicamente, se propone el uso de modelos AR(1) y

ARMA(1,1) con innovaciones que siguen una distribución Sinh-Arcsinh (SAS). Este modelo se justifica por su capacidad para capturar tanto la asimetría como las colas pesadas observadas en muchas series de tiempo reales. Se evalúa la eficiencia de este modelo a través de simulaciones de series de tiempo con innovaciones que siguen distribuciones BS, y se comparan los resultados con los obtenidos por los métodos convencionales que asumen normalidad. La metodología emplea la verosimilitud condicional, lo que permite una estimación eficiente de los parámetros, optimizando el rendimiento computacional en comparación con la verosimilitud total.

La justificación de este estudio radica en la necesidad de adaptar los modelos tradicionales a las características complejas de las series de tiempo no gaussianas. El uso de la verosimilitud condicional, que toma en cuenta la densidad de las innovaciones, permite un enfoque computacionalmente eficiente sin perder la capacidad de estimar adecuadamente los parámetros. Los modelos propuestos no solo ofrecen una alternativa más robusta a los modelos convencionales, sino que también proporcionan un marco teórico más flexible para el análisis de series de tiempo en contextos donde los datos presentan distribuciones no gaussianas.

A través de simulaciones, este busca demostrar cómo los modelos SAS-AR(1) y SAS-ARMA(1,1) pueden ser utilizados de manera efectiva en el análisis de series de tiempo, superando las limitaciones de los modelos tradicionales. Además, se implementan en R utilizando funciones y paquetes ya disponibles, lo que facilita su adopción en aplicaciones prácticas y su integración en el análisis de datos reales. Los resultados de este estudio tienen implicaciones importantes para los analistas que trabajan con series temporales en disciplinas como las finanzas, la ingeniería y la meteorología, donde los supuestos de normalidad no son adecuados y el modelado de colas pesadas y asimetría es esencial para obtener pronósticos precisos y confiables.

En resumen, este trabajo contribuye al campo del análisis de series de tiempo al ofrecer una metodología innovadora y eficiente para modelar procesos no gaussianos mediante la incorporación de distribuciones flexibles como la SAS, mejorando así la capacidad predictiva y la fiabilidad de las estimaciones en una amplia gama de aplicaciones.

3. Marco Teórico

3.1. Series de tiempo

Las series de tiempo son observaciones cuantitativas indexadas en el tiempo. Su análisis es esencial en diversos campos como economía, ingeniería, meteorología y ciencias sociales. Según Brockwell y Davis (2016), una serie de tiempo puede representarse como una realización de un proceso estocástico $\{X_t\}_{t\in\mathbb{Z}}$, y su modelado se basa frecuentemente en supuestos de estacionariedad. En este trabajo se hablara formalmente de los modelos AR(1) y

ARMA(1,1)

3.1.1. Modelo AR(1)

Un modelo autorregresivo de orden 1 se define como

$$X_t \;=\; \phi X_{t-1} + \varepsilon_t, \tag{1}$$

donde ϕ es el coeficiente autorregresivo y ε_t son las innovaciones del modelo. Bajo los supuestos gaussianos clásicos se tiene:

- La media marginal es $E(X_t) = 0$.
- La varianza marginal es

$$(X_t) \;=\; \frac{\sigma^2}{1-\phi^2}, \tag{2}$$

  donde $\sigma^2 = (\varepsilon_t)$.
- Las innovaciones $\{\varepsilon_t\}$ son independientes e idénticamente distribuidas (i.i.d.) como

$$\varepsilon_t \sim N(0, \sigma^2). \tag{3}$$

- La función de autocorrelación depende sólo del rezago h y viene dada por

$$\rho(h) \;=\; \phi^{\,h}, \qquad h = 0, 1, 2, \ldots \tag{4}$$

- Cuando $\sigma^2 = 1-\phi^2$ la densidad marginal de cada X_t es normal estándar, hecho que será útil más adelante.

3.1.2. Modelo ARMA(1,1)

Un modelo ARMA(1,1) se define como

$$X_t \;=\; \phi X_{t-1} + \varepsilon_t + \theta\, \varepsilon_{t-1}, \tag{5}$$

donde ϕ es el coeficiente autorregresivo, θ el coeficiente de media móvil y ε_t las innovaciones. Bajo normalidad gaussiana se tiene:

- La media marginal es $E(X_t) = 0$.
- La varianza marginal es

$$(X_t) \;=\; \sigma^2\, \frac{1 + \theta^2 + 2\theta\phi}{1-\phi^2}, \tag{6}$$

  donde $\sigma^2 = (\varepsilon_t)$.

- Las innovaciones $\{\varepsilon_t\}$ son i.i.d. como

$$\varepsilon_t \sim N(0, \sigma^2). \tag{7}$$

- La autocorrelación de orden $h = 0$ (correlación entre X_t y X_t en el modelo ARMA(1,1) estandarizado) es

$$\rho(1) \;=\; \frac{(\theta + \phi)(1 + \theta\phi)}{1 + \theta^2 + 2\theta\phi}. \tag{8}$$

En ambos modelos se asume que $\{X_t\}$ es estacionaria en sentido débil, es decir, la media y la varianza son constantes en el tiempo y la autocovarianza (y por tanto la autocorrelación) depende sólo del rezago y no de un tiempo t específico.

3.1.3. Incumplimiento de la normalidad

Cuando las innovaciones Z_t no son gaussianas, los métodos tradicionales pueden producir estimaciones sesgadas o ineficientes. En aplicaciones económicas e industriales es frecuente observar asimetría y colas pesadas en las innovaciones, lo que afecta la eficiencia y la validez de la inferencia gaussiana. Esto ha motivado el uso de distribuciones alternativas con control explícito de forma (por ejemplo, Jones y Pewsey, 2009; Contreras-Espinoza et al., 2023) y el desarrollo de herramientas robustas en control de procesos y modelación (Leiva et al., 2014).

Entre las soluciones propuestas se encuentran:

- Distribuciones con colas pesadas (por ejemplo, t-Student).
- Distribuciones asimétricas (por ejemplo, Gamma, log-normal, skew-t).
- Estimadores robustos, como M-estimadores o métodos basados en cuantiles.

Sin embargo, para un tratamiento sistemático de series de tiempo no gaussianas interesa disponer de familias *flexibles* que:

1. Permitan controlar *por separado* la asimetría y el peso de cola.
2. Sean compatibles con la estructura dinámica de modelos ARMA.
3. Sean computacionalmente manejables al construir verosimilitudes (exactas o condicionales).

En este contexto, la familia *sinh–arcsinh* (SAS) ofrece una alternativa especialmente atractiva. Aplicada a una normal estándar, la transformación SAS genera una familia de cuatro parámetros (dos de localización/escala y dos de forma) capaz de producir tanto colas más pesadas como más ligeras

que la normal, con asimetría controlable mediante un parámetro de sesgo y otro de cola (Jones y Pewsey, 2009). Versiones extendidas como la SAS "two-piece" refuerzan la idea de que los parámetros de forma pueden asociarse de manera clara a sesgo y curtosis.

Para destacar las ventajas de la SAS frente a otras alternativas no gaussianas de uso frecuente, se presenta a continuación una comparación cualitativa con la t-Student asimétrica (skew-t) y la Distribución de Valores Extremos Generalizada (GEV).

Cuadro 1: Comparación cualitativa entre SAS, skew-t y GEV para innovaciones de series de tiempo.

Familia	**Parámetros de forma**	**Asimetría y curtosis**	**Complejidad computacional**	**Comportamiento en los extremos**
SAS	Dos parámetros de forma (ε, δ), además de localización y escala.	Permite controlar casi de forma separada la asimetría (ε) y el peso de cola (δ); puede generar colas más pesadas o más ligeras que la normal.	La densidad se escribe en forma cerrada a partir de la normal; simulación y evaluación usan transformaciones elementales (sinh, cosh, asinh).	Produce colas flexibles en ambos lados (ligeras o pesadas) sin especializarse únicamente en máximos o mínimos; es adecuada para modelar toda la distribución de la serie.
Skew-t	Parámetro de asimetría y grados de libertad, además de localización y escala.	Permite asimetría y colas pesadas, pero los grados de libertad controlan simultáneamente curtosis y colas; el desacople entre sesgo y peso de cola es menos claro que en SAS.	La densidad es más costosa de evaluar (funciones t y, en dimensión mayor, matrices de covarianza); en modelos dinámicos suele requerir algoritmos numéricos más complejos.	Presenta colas polinomiales adecuadas para extremos, pero los grados de libertad pequeños pueden generar inestabilidad numérica y estimación más sensible al tamaño muestral.
GEV	Un parámetro de forma ξ, además de localización y escala.	Asimetría y colas vienen determinadas simultáneamente por ξ; no hay una separación limpia entre sesgo y curtosis.	La verosimilitud en ξ es altamente no lineal; la estimación puede ser difícil e inestable en muestras pequeñas o cuando los datos apenas exploran la cola.	Es la distribución límite para máximos de bloques; genera colas muy pesadas o truncadas según el signo de ξ, ideal para análisis de extremos pero menos natural para describir toda la trayectoria de la serie.

En este trabajo se privilegia el uso de la SAS porque:

- Se construye como una transformación monótona de un proceso ARMA

gaussiano, lo que permite heredar propiedades de estacionariedad y estructura de autocorrelación del modelo base.

- Ofrece un control claro y separado de asimetría y colas, algo menos transparente en familias como la skew-t o la GEV.
- Su densidad y su jacobiano tienen expresión cerrada y relativamente sencilla, lo que facilita la construcción de verosimilitudes condicionales y reduce la carga computacional frente a alternativas basadas en distribuciones de extremos o mezclas escalares más complejas.

3.2. Modelos ARMA con Distribución Birnbaum-Saunders

La distribución Birnbaum-Saunders surge en el análisis de fatiga de materiales y se puede representar a partir de una normal estándar, facilitando simulación e inferencia y proporcionando asimetría positiva en soporte (0,) (Leiva et al., 2011). En ingeniería de calidad, el uso de índices de capacidad adaptados a BS mejora la interpretación cuando los datos no son normales (Leiva et al., 2014). Extensiones recientes incluyen modelación espacial con diagnóstico de influencia local en datos agrícolas (García-Papani et al., 2018).

Se define por la transformación:

$$Y = \beta \left(\frac{\alpha X}{2} + \sqrt{\left(\frac{\alpha X}{2} \right)^2 + 1} \right)^2, \quad X \sim N(0,1),$$

con función de densidad:

$$f_Y(y;\alpha,\beta) = \frac{1}{2\alpha\beta\sqrt{2\pi}} \left(\left(\frac{\beta}{y} \right)^{1/2} + \left(\frac{\beta}{y} \right)^{3/2} \right) \exp\left(-\frac{1}{2\alpha^2} \left(\frac{y}{\beta} + \frac{\beta}{y} - 2 \right) \right), \quad y > 0.$$

Para series dependientes, se han propuesto construcciones ARMA con márgenes BS que preservan la estructura dinámica y permiten ajustar asimetría y colas en datos positivos (Rahul et al., 2018). Este modelo transforma un proceso ARMA subyacente con innovaciones gaussianas, generando una serie marginalmente BS.

3.3. Distribución Sinh-Arcsinh (SAS)

La familia SAS se obtiene mediante una transformación que introduce dos parámetros de forma: uno de asimetría y otro de colas, logrando de manera parsimoniosa desde colas ligeras hasta pesadas con asimetría controlable (Jones Pewsey, 2009). En series de tiempo, los modelos GAS con densidades SAS actualizan parámetros condicionales vía el score, capturando variaciones

temporales en asimetría y curtosis; se ha documentado mejor ajuste a extremos frente a alternativas gaussianas o t-clásicas en aplicaciones empíricas (Contreras-Espinoza et al., 2023).

La distribución se construye a partir de la transformación:

$$Z = \sinh(\delta \sinh^{-1}(X) - \epsilon), \quad X \sim F,$$

donde ϵ controla la asimetría y δ el peso de las colas.

Su densidad es:

$$f_{\epsilon,\delta}(x) = \frac{\delta}{\sqrt{2\pi(1+x^2)}} \cosh(\delta \sinh^{-1}(x) - \epsilon) \exp\left(-\frac{1}{2} \sinh^2(\delta \sinh^{-1}(x) - \epsilon)\right).$$

A continuación se puede evidenciar el comportamiento de la función en función de sus parámetros:

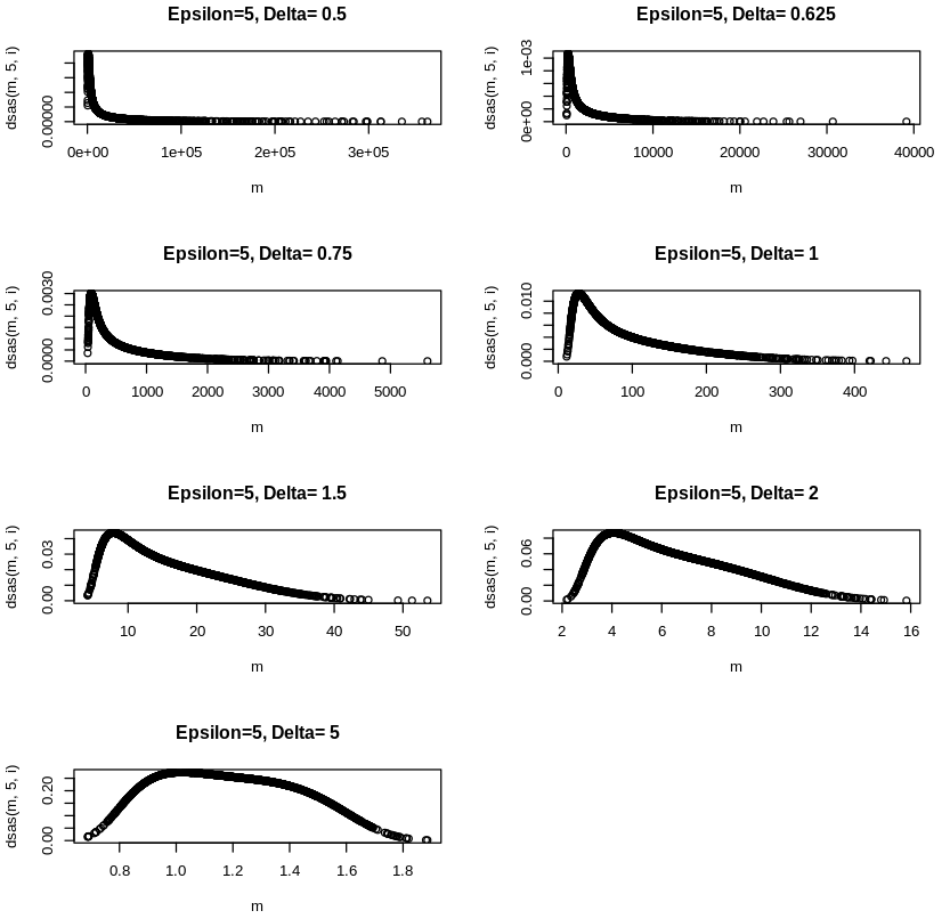


Figura 1: Efecto del parámetro de cola δ en la forma de la distribución SAS con asimetría fija $\varepsilon = 5$. Cada curva corresponde a un valor distinto de δ, ilustrando cómo un aumento en δ reduce la pesadez de las colas y concentra la masa de probabilidad, manteniendo la asimetría positiva.

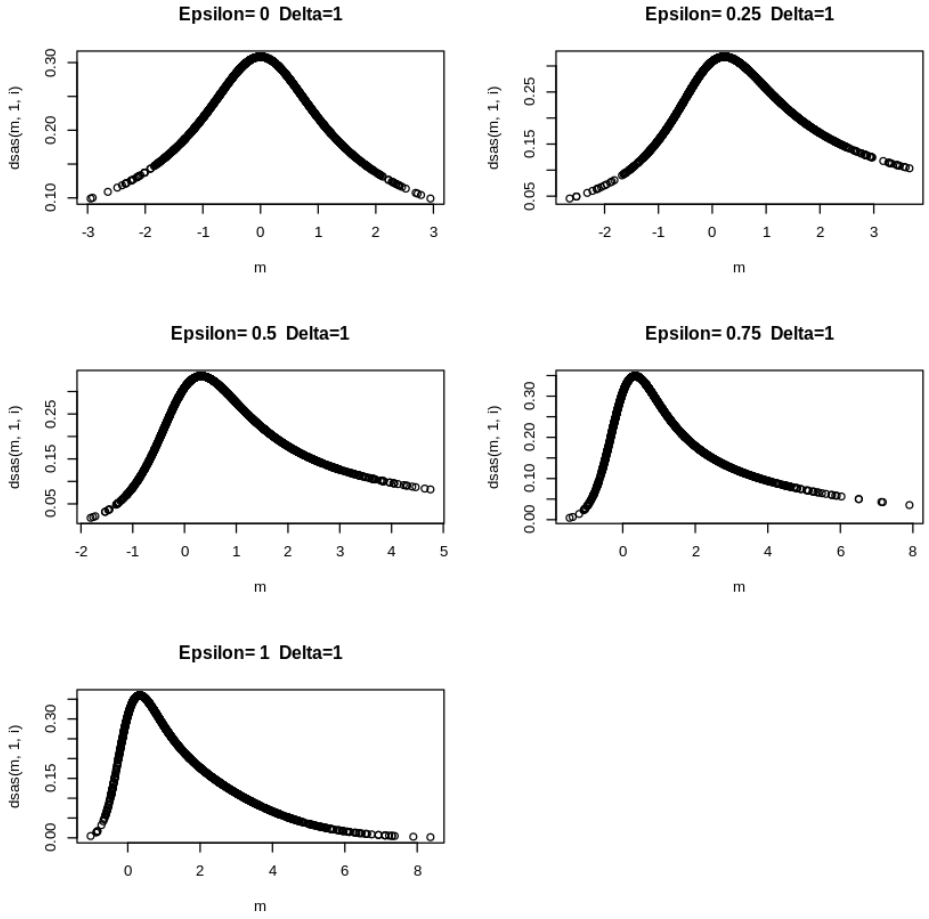


Figura 2: Efecto del parámetro de asimetría ε en la distribución SAS con peso de cola fijo $\delta = 1$. Al incrementar ε la distribución se sesga hacia la derecha, desplazando la masa hacia valores positivos y generando colas desbalanceadas.

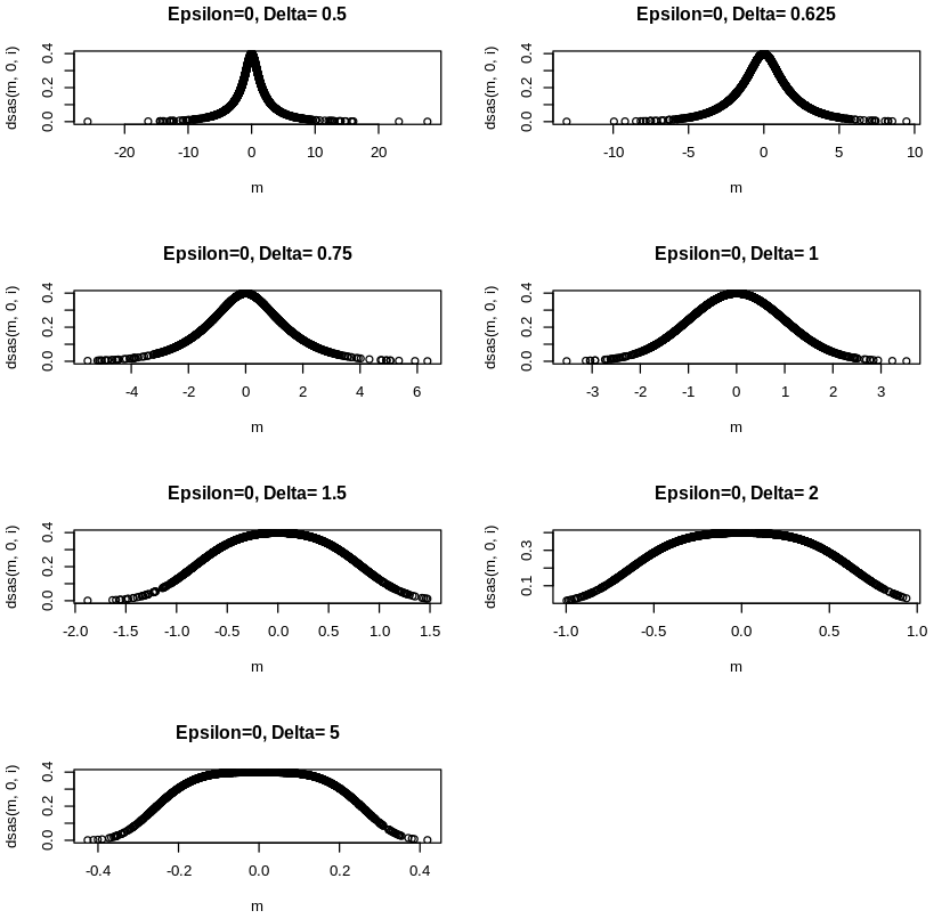


Figura 3: Efecto del parámetro de cola δ en la distribución SAS simétrica con $\varepsilon = 0$. Valores pequeños de δ producen colas pesadas y mayor probabilidad de extremos en ambos signos; valores grandes de δ generan colas ligeras y una distribución muy concentrada alrededor de cero.

- **Rol de ε (asimetría).** En la Fig. 2, con $\delta = 1$ constante, al aumentar ε de 0 a 1 la distribución pasa de casi simétrica a *fuertemente sesgada a la derecha*: la *cola derecha* se alarga, la izquierda se contrae y la masa se desplaza hacia valores positivos. Se observa el patrón media $>$ mediana $>$ moda para $\varepsilon > 0$.
- **Rol de δ (colas y escala).** Con ε fijo, δ controla la *pesadez de colas* y la *concentración*:
  - *Asimetría fija y positiva* (Fig. 1): con δ pequeño (p. ej., 0.5–0.75) aparecen colas muy pesadas y una cola derecha larguísima; conforme δ crece a 1.5–2 la masa se concentra y la cola derecha se

aligera; con $\delta = 5$ la distribución es compacta alrededor de $\approx$ 1–1.6, manteniendo la asimetría.

- *Simetría exacta* (Fig. 3): todas las curvas son simétricas con modo en 0. Para δ pequeño las colas son gruesas y el rango es amplio; al aumentar δ la varianza efectiva cae y la curva se hace más "aplanada y compacta" en torno al centro; con $\delta = 5$ el soporte efectivo es muy corto.

- **Momentos y forma.** δ bajo $\Rightarrow$ alta curtosis (pico pronunciado + colas gruesas) y mayor probabilidad de valores extremos; δ alto $\Rightarrow$ colas ligeras, menor curtosis y fuerte concentración central. A δ fijo, al crecer ε aumentan asimetría y varianza efectiva por la extensión de la cola derecha.

Posibles implicaciones para los modelos

- **Escenarios "estresantes".** Usar $\delta < 1$ (colas pesadas), con $\varepsilon = 0$ (simétrico) o $\varepsilon > 0$ (asimétrico), genera outliers frecuentes y grandes. Esto es ideal para comparar *robustez* de estimadores frente a Normal y Birnbaum–Saunders y suele incrementar RMSE y MAPE cuando el modelo no captura bien las colas.

- **No normalidad moderada.** Con $\delta \in [1, 2]$ se obtienen colas moderadas; variando $\varepsilon \in [0, 1]$ se controla el grado de asimetría sin explosión de extremos, escenario adecuado para datos económicos reales.

- **Cercanía a Gauss.** Con $\delta \gg 1$ las colas son ligeras y la dispersión es pequeña; si además $\varepsilon \approx 0$, el comportamiento se acerca al gaussiano, aunque el SAS mantiene flexibilidad y puede mejorar la verosimilitud y, por ende, el AIC.

3.3.1. Caso multivariado

La función de densidad d-variada de la distribución sinh-arcsinh es:

$$f_{\epsilon,\delta} = \{(2\pi)^d |R|\}^{-\frac{1}{2}} \prod_{i=1}^{d} (1 + x_i^2)^{-\frac{1}{2}} \delta C_{\epsilon,\delta}(x_i) \exp\{\frac{-S_{\epsilon,\delta}(X)^T R^{-1} S_{\epsilon,\delta}(X)}{2}\}$$

Cuyas marginales por construcción son sinh-arcsinh con los mismos parámetros ϵ, δ.

3.4. Verosimilitud condicional

En modelos ARMA gaussianos, la *verosimilitud exacta* se construye a partir de la densidad conjunta de $(X_1, \ldots, X_T)$, lo que implica trabajar con la matriz de covarianzas completa o con algoritmos equivalentes (por ejemplo,

filtros tipo Kalman o factorizaciones de matrices de tipo Toeplitz; véase, entre otros, Brockwell y Davis, 2016). Esto puede ser costoso computacionalmente, especialmente cuando se extiende el modelo a innovaciones no gaussianas, donde la estructura conjunta deja de ser puramente normal.

Una alternativa clásica es la *verosimilitud condicional* o *aproximada*, que condiciona en un número finito de términos iniciales y sólo utiliza densidades de transición de un paso. En nuestro caso, para una serie $\{X_t\}_{t=1}^{T}$ transformada a márgenes SAS, la verosimilitud condicional se escribe como

$$L_c(\phi, \theta, \varepsilon, \delta) \;=\; f_{\varepsilon,\delta}(x_1) \prod_{t=2}^{T} f_{\varepsilon,\delta}\big(x_t \mid x_{t-1}\big), \tag{9}$$

donde $f_{\varepsilon,\delta}(x_t)$ denota la densidad marginal SAS y $f_{\varepsilon,\delta}(x_t \mid x_{t-1})$ la densidad condicional de transición de orden 1, obtenida a partir de la densidad bivariada SAS:

$$f_{\varepsilon,\delta}(x_t \mid x_{t-1}) \;=\; \frac{f_{\varepsilon,\delta}\big(x_{t-1}, x_t\big)}{f_{\varepsilon,\delta}(x_{t-1})}. \tag{10}$$

La log-verosimilitud condicional asociada es

$$\ell_c(\phi, \theta, \varepsilon, \delta) \;=\; \log f_{\varepsilon,\delta}(x_1) \;+\; \sum_{t=2}^{T} \log f_{\varepsilon,\delta}\big(x_t \mid x_{t-1}\big). \tag{11}$$

Trade-off entre eficiencia computacional y estadística

En modelos ARMA gaussianos se sabe que la estimación basada en la verosimilitud condicional es, bajo condiciones estándar de estacionariedad e invertibilidad, *asintóticamente equivalente* a la estimación por verosimilitud exacta: la diferencia entre ambas log-verosimilitudes es de orden acotado y los estimadores resultantes comparten el mismo límite normal. Sin embargo, en muestras pequeñas o cuando los parámetros están cerca de la frontera de estacionariedad, la verosimilitud exacta tiende a ser más eficiente (menor varianza y menor sesgo), como se ha documentado en estudios de simulación para modelos ARMA.

En el contexto de modelos SAS–AR(1) y SAS–ARMA(1,1), la situación es más delicada:

- La **verosimilitud exacta** requeriría especificar la densidad conjunta SAS de dimensión T, es decir, la imagen de un vector normal $(Z_1, \ldots, Z_T)$ correlacionado bajo una transformación no lineal componente a componente. Evaluar esta densidad implicaría trabajar con matrices de covarianza de tamaño $T \times T$ y jacobianos de dimensión T, lo cual vuelve el problema prácticamente intratable para T moderados o grandes.

- La **verosimilitud condicional** sólo necesita densidades marginales y bivariadas SAS (casos $d = 1$ y $d = 2$), que sí admiten expresiones cerradas relativamente simples a partir de la normal estándar y su extensión bivariada. Esto reduce de forma importante el costo computacional y permite realizar simulaciones Monte Carlo y búsquedas en rejilla sobre los parámetros.

Desde el punto de vista de eficiencia estadística, el uso de ℓ_c en lugar de la verosimilitud exacta tiene dos implicaciones principales:

1. **Pérdida de información en los primeros términos.** Al condicionar en X_1 (y, en general, en un número finito de términos iniciales), la contribución exacta de estos datos a la verosimilitud total se aproxima por la densidad marginal SAS en lugar de por su verdadera densidad conjunta. En series largas (T grande), el impacto relativo de esta aproximación es pequeño, pues la suma en (11) domina. En series cortas, en cambio, la pérdida de información puede ser relevante y reflejarse en mayor sesgo o varianza de los estimadores.

2. **Equivalencia asintótica esperada.** Por analogía con el caso gaussiano, y dado que el proceso SAS se construye como transformación monótona de un ARMA gaussiano, es razonable esperar que, bajo condiciones regulares, la verosimilitud condicional y la exacta sean asintóticamente equivalentes también en el escenario SAS: la diferencia entre ambas log-verosimilitudes sería acotada en T y los estimadores compartirían el mismo límite normal. Un análisis teórico riguroso de esta equivalencia excede el alcance de este trabajo, pero la evidencia numérica de las simulaciones sugiere que, para tamaños muestrales moderados o grandes, el sesgo adicional es reducido.

En resumen, la elección de verosimilitud condicional en los modelos SAS–AR(1) y SAS–ARMA(1,1) responde a un *compromiso explícito*: se sacrifica una parte de la eficiencia estadística potencial en muestras pequeñas a cambio de una ganancia sustancial en tractabilidad computacional. Dado que el objetivo principal de este trabajo es explorar la viabilidad de estos modelos y estudiar su comportamiento por simulación, esta concesión metodológica resulta razonable. En aplicaciones empíricas con series muy cortas o con fuerte dependencia, el desarrollo de aproximaciones más cercanas a la verosimilitud exacta SAS queda como una línea relevante de investigación futura.

Trade-off entre eficiencia computacional y estadística

En modelos ARMA gaussianos, se sabe que la estimación basada en la verosimilitud condicional es *asintóticamente equivalente* a la estimación por verosimilitud exacta: la diferencia entre ambas log-verosimilitudes es de orden

$O(1)$ y los estimadores resultantes comparten el mismo límite normal bajo condiciones estándar de estacionariedad e invertibilidad (Brockwell y Davis, 2016; Yao y Brockwell, 2006).:contentReference[oaicite:10]index=10 Sin embargo, en muestras pequeñas o cuando los parámetros están cerca de la frontera de estacionariedad, la verosimilitud exacta tiende a ser más eficiente (menor varianza y menor sesgo), tal como se ha documentado en estudios de simulación para modelos ARMA (Hauser, 1999).:contentReference[oaicite:11]index=11

En el contexto de modelos SAS–AR(1) y SAS–ARMA(1,1), la situación es más delicada:

- La **verosimilitud exacta** requeriría especificar la densidad conjunta SAS de dimensión T, es decir, la imagen de un vector normal $(Z_1, \ldots, Z_T)$ correlacionado bajo una transformación no lineal componente a componente. Evaluar esta densidad implicaría trabajar con matrices de covarianza de tamaño $T \times T$ y jacobianos de dimensión T, lo cual vuelve el problema prácticamente intratable para T moderados o grandes.

- La **verosimilitud condicional** sólo necesita densidades marginales y bivariadas SAS (casos $d = 1$ y $d = 2$), que sí admiten expresiones cerradas relativamente simples a partir de la normal estándar y su extensión bivariada. Esto reduce dramáticamente el costo computacional y permite realizar simulaciones Monte Carlo y búsquedas en rejilla sobre los parámetros.

Desde el punto de vista de eficiencia estadística, el uso de ℓ_c en lugar de la verosimilitud exacta tiene dos implicaciones:

1. **Pérdida de información en los primeros términos.** Al condicionar en X_1 (y, en general, en un número finito de términos iniciales), la contribución exacta de estos datos a la verosimilitud total se aproxima por la densidad marginal SAS en lugar de por su verdadera densidad conjunta. En series largas (T grande), el impacto relativo de esta aproximación es pequeño, pues la suma en (11) domina. En series cortas, en cambio, la pérdida de información puede ser no despreciable y reflejarse en mayor sesgo o varianza de los estimadores.

2. **Equivalencia asintótica esperada.** Por analogía con el caso gaussiano, y dado que el proceso SAS se construye como transformación monótona de un ARMA gaussiano, es razonable esperar que, bajo condiciones regulares, la verosimilitud condicional y la exacta sean asintóticamente equivalentes también en el escenario SAS: la diferencia entre ambas log-verosimilitudes sería acotada en T y los estimadores compartirían el mismo límite normal. Un análisis teórico riguroso de esta equivalencia excede el alcance de este trabajo, pero la evidencia

numérica de las simulaciones sugiere que, para tamaños muestrales moderados o grandes, el sesgo adicional es reducido.

En resumen, la elección de verosimilitud condicional en los modelos SAS–AR(1) y SAS–ARMA(1,1) responde a un *compromiso explícito*: se sacrifica una parte de la eficiencia estadística potencial en muestras pequeñas a cambio de una ganancia sustancial en tractabilidad computacional. Dado que el objetivo principal de este trabajo es explorar la viabilidad de estos modelos y estudiar su comportamiento por simulación, esta concesión metodológica resulta razonable. En aplicaciones empíricas con series muy cortas o con fuerte dependencia, el desarrollo de aproximaciones más cercanas a la verosimilitud exacta SAS queda como una línea relevante de investigación futura.

4. Construcción del Modelo SAS-AR(1)

Se parte de un modelo AR(1) gaussiano donde:

$$Z_t = \phi Z_{t-1} + e_t$$

Dado que la distribución sinh-arcsinh proviene de transformar una normal estándar, es necesario que las marginales de la serie Z_t sean normales estándar y eso nos lleva a que las innovaciones(e_t deber ser:

$$e_t \sim N(0, 1 - \phi)$$

De este modo se garantiza que cada z_t en la serie tenga distribución marginal $N(z_t|0, 1)$. Después se procede a aplicar la transformación inversa a $S^{-1}_{\epsilon,\delta}(x_t)$. De este modo se obtiene la serie SAS-AR(1).

Luego se observa como se comporta la serie en función de los parámetros ϵ y δ dejando a un ϕ fijo en 0.5.

Epsilon fijo y delta variable (`are5`). En la primera lámina se mantiene $\varepsilon = 5$ (asimetría positiva marcada) y se hace crecer δ desde 0.5 hasta 5. Con $\delta = 0.5$ se observan valores extremos muy grandes y esporádicos que dominan la escala del gráfico; estos picos ocurren principalmente hacia el lado positivo por la asimetría inducida por $\varepsilon = 5$. A medida que δ aumenta a 0.625, 0.75 y 1, dichos extremos se vuelven menos frecuentes e intensos: la serie conserva el sesgo a la derecha, pero la dispersión decrece visiblemente. Cuando $\delta \geq 1.5$ (paneles con $\delta = 1.5, 2, 5$), las colas se aligeran y la trayectoria queda contenida en un rango estrecho; persiste la asimetría, aunque ya no produce explosiones numéricas. En todos los paneles la dependencia temporal es similar—consistente con un AR(1) con $\phi = 0.5$—y lo que cambia es la forma marginal: δ regula el peso de cola y, por tanto, la frecuencia/intensidad de observaciones extremas, mientras que ε fija la dirección del sesgo.

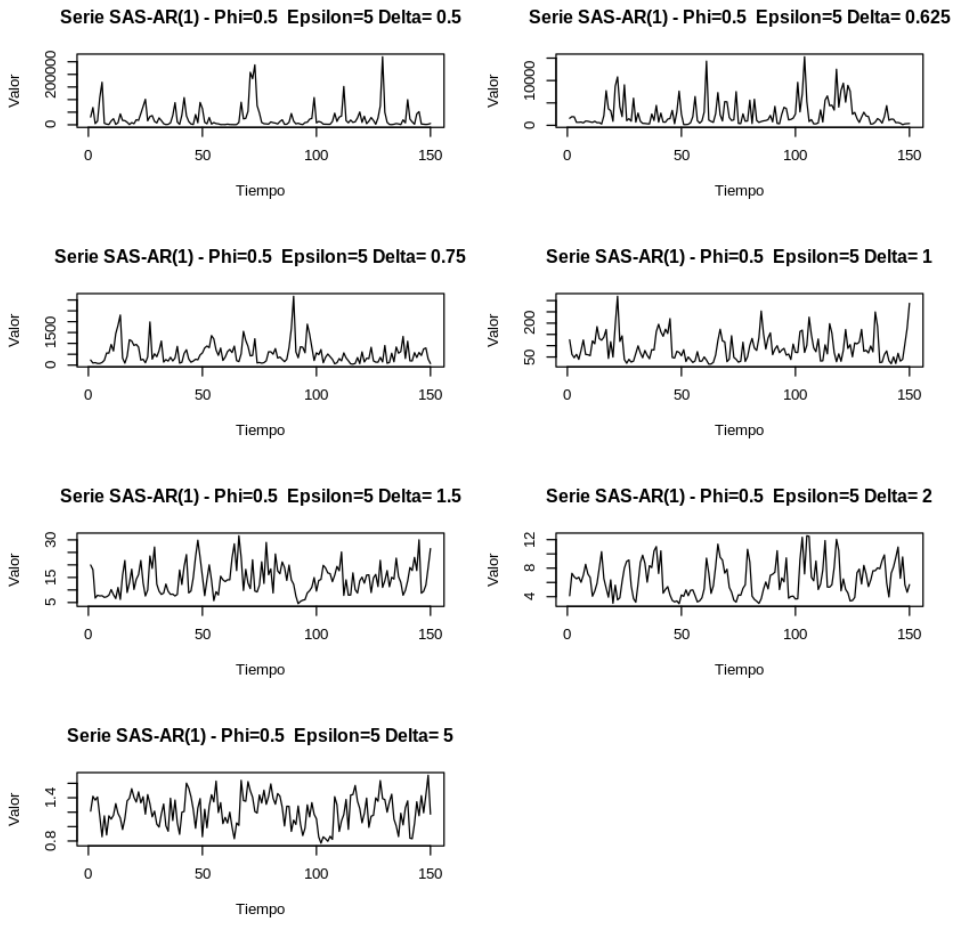


Figura 4: Serie SAS–AR(1) con $\phi = 0.5$ y $\varepsilon = 5$ para varios valores de δ.

Delta fija y epsilon variable (`ard1`). En la segunda lámina se fija $\delta = 1$ (peso de cola comparable al normal) y se incrementa la asimetría ε de 0 a 1. Con $\varepsilon = 0$ la serie es visualmente simétrica alrededor de cero: los máximos y mínimos tienen magnitudes comparables. Conforme ε crece (0.25, 0.5, 0.75, 1), la masa de la distribución se desplaza hacia la derecha y aumentan la frecuencia y tamaño de valores positivos grandes; los mínimos, en cambio, se mantienen relativamente acotados. La estructura de autocorrelación (suavidad y persistencia) permanece estable entre paneles, lo que evidencia que las diferencias provienen de la asimetría marginal controlada por ε y no de cambios en la dinámica AR(1).

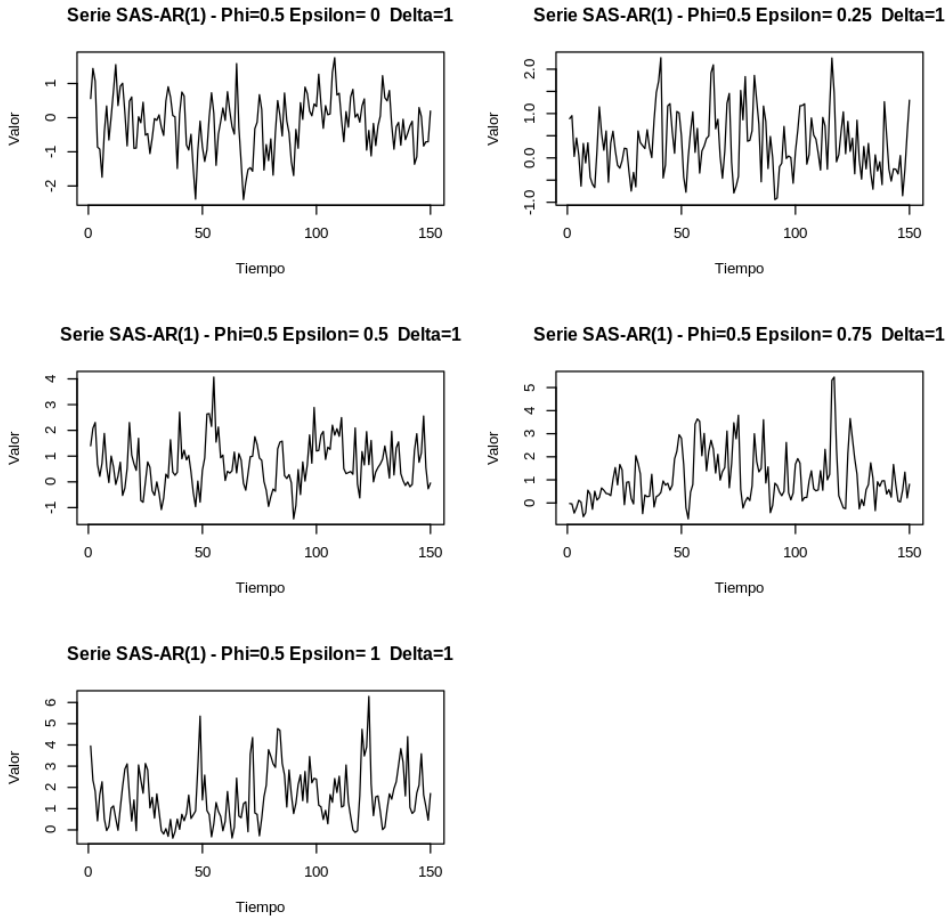


Figura 5: Serie SAS–AR(1) con $\phi = 0.5$ y $\delta = 1$ para varios valores de ε.

Epsilon fijo en cero y delta variable (`are0`). La tercera lámina aísla el efecto del peso de cola con $\varepsilon = 0$ (márgenes simétricas) y δ de 0.5 a 5. Para $\delta < 1$ (paneles con 0.5, 0.625, 0.75) se aprecian colas pesadas: hay mayor

rango y presencia de picos en ambos signos, lo que incrementa notablemente la variabilidad observada. Al aumentar δ a 1, 1.5, 2 y 5 las colas se aligeran y la serie se "comprime", reduciendo la aparición de valores extremos sin introducir sesgo alguno. Como antes, la persistencia temporal característica de $\phi = 0.5$ se mantiene inalterada; lo que varía es la probabilidad de extremos determinada por δ.

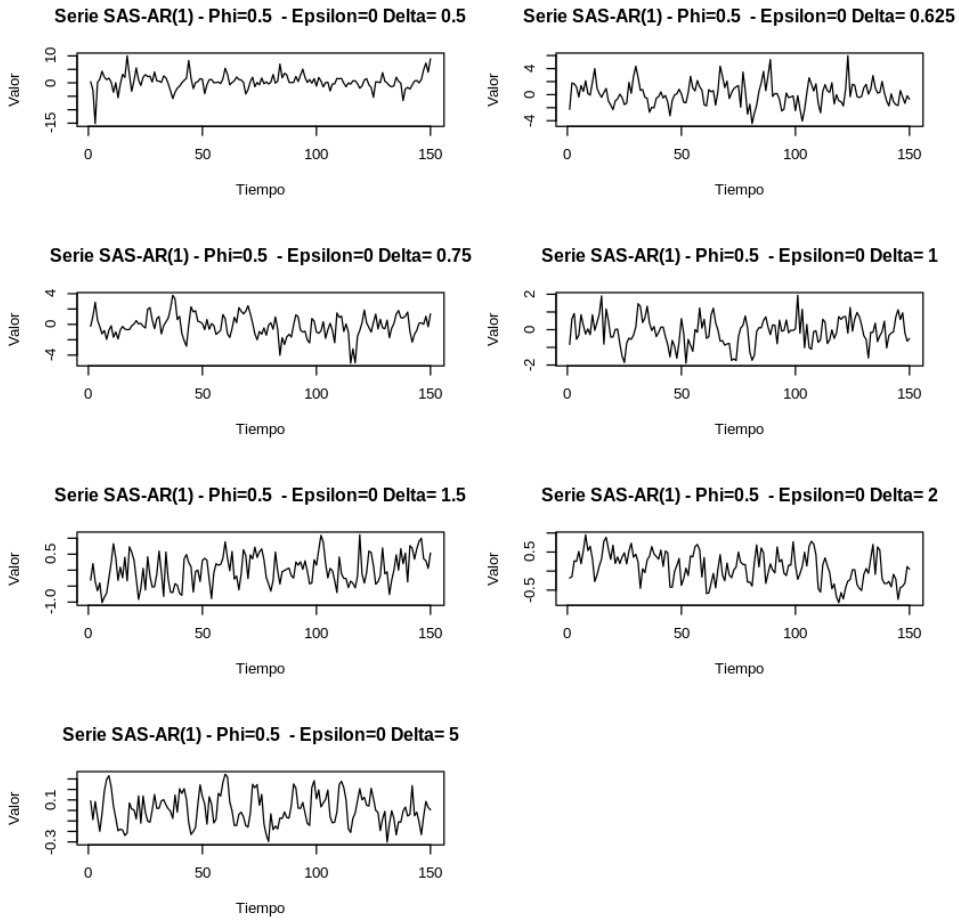


Figura 6: Serie SAS–AR(1) con $\phi = 0.5$ y $\varepsilon = 0$ para varios valores de δ.

En las tres láminas la dinámica AR(1) (misma suavidad y decaimiento de la autocorrelación) es constante porque $\phi = 0.5$ es fijo. Las diferencias visuales obedecen exclusivamente a la forma marginal SAS: ε controla la dirección e intensidad de la asimetría (desbalance entre colas), mientras que δ controla el peso de cola (frecuencia e intensidad de observaciones extremas). En conjunto, los paneles muestran que, aun con igual dependencia temporal, la elección de la margen SAS permite modelar series con sesgo pronunciado y/o colas pesadas, modulando de manera separada ambos rasgos.

4.0.1. Estimación de parámetros

Paso 1 (Cociente). Se construye la densidad de transición 1–paso como cociente entre la densidad conjunta bivariada y la marginal del rezago (cambio de variable a la escala normal).

$$f(Y_t \mid Y_{t-1}) = \frac{f(Y_{t-1}, Y_t)}{f(Y_{t-1})} = \phi\big(s_t;\, \rho\, s_{t-1},\, 1-\rho^2\big)\ \underbrace{\frac{\delta\, \cosh(A_t)}{\sqrt{1+Y_t^2}}}_{J_t},$$

donde

$$A_t = \delta\ \operatorname{asinh}(Y_t) - \varepsilon, \qquad s_t = \sinh(A_t),$$

y $\phi(\cdot; \mu, \sigma^2)$ es la densidad normal en la escala transformada s.

Paso 2 (Verosimilitud). Por markovianidad de orden 1, la verosimilitud condicional factoriza como el producto de la marginal inicial y las transiciones 1–paso.

$$L(\varepsilon, \delta, \rho \mid y_{1:T}) = f(y_1; \varepsilon, \delta) \prod_{t=2}^{T} f(y_t \mid y_{t-1}; \varepsilon, \delta, \rho).$$

Paso 3 (Log–verosimilitud). Tomar logaritmo, separar la contribución marginal y la suma de transiciones; omitir constantes aditivas no dependientes de los parámetros. El resultado es:

$$\ell(\varepsilon, \delta, \rho) = \big[\ \log J_1 - \tfrac{1}{2} s_1^2\ \big] + \sum_{t=2}^{T} \left\{ -\tfrac{1}{2} \log(1-\rho^2) + \log J_t - \frac{\big(s_t - \rho s_{t-1}\big)^2}{2(1-\rho^2)} \right\},$$

con

$$\log J_t = \log \delta + \log\big(\cosh A_t\big) - \tfrac{1}{2} \log(1 + Y_t^2).$$

Paso 4 (Score en ρ). Derivar ℓ respecto a ρ usando que solo los términos de transición dependen de ρ; la condición de primer orden se reduce a la ecuación normal en la escala transformada s.

$$\frac{\partial \ell}{\partial \rho} = \sum_{t=2}^{T} \left[\frac{\rho}{1-\rho^2} + \frac{(s_t - \rho s_{t-1})\, s_{t-1}}{1-\rho^2} + \frac{\rho\, (s_t - \rho s_{t-1})^2}{(1-\rho^2)^2} \right] \implies \sum_{t=2}^{T} s_t s_{t-1} \;=\; \rho \sum_{t=2}^{T} s_{t-1}^2.$$

Paso 5 (Scores en ε y δ). Derivar ℓ usando regla de la cadena: entran las derivadas de s_t y de $\log J_t$. Los resultados (sustituidos y simplificados) son:

$$\frac{\partial \ell}{\partial \varepsilon} = \Big[-\tanh A_1 + s_1 \cosh A_1 \Big] + \sum_{t=2}^{T} \left\{ -\tanh A_t - \frac{s_t - \rho s_{t-1}}{1-\rho^2} \Big(-\cosh A_t + \rho \cosh A_{t-1} \Big) \right\},$$

$$\frac{\partial \ell}{\partial \delta} = \Big[\frac{1}{\delta} + \operatorname{asinh}(Y_1)\tanh A_1 - s_1\ \operatorname{asinh}(Y_1)\cosh A_1\Big] + \sum_{t=2}^{T}\left\{\frac{1}{\delta} + \operatorname{asinh}(Y_t)\tanh A_t - \frac{s_t - \rho s_{t-1}}{1-\rho^2}\left(\operatorname{asinh}(Y_t)\cosh A_t - \rho\ \operatorname{asinh}(Y_{t-1})\cosh A_{t-1}\right)\right\}. \tag{12}$$

5. Construcción del Modelo SAS-ARMA(1,1)

Se parte de un modelo ARMA(1,1) gaussiano donde:

$$Z_t = \phi Z_{t-1} + e_t + \theta e_{t-1}$$

Dado que la distribución sinh-arcsinh proviene de transformar una normal estándar, es necesario que las marginales de la serie Z_t sean normales estándar y eso nos lleva a que las innovaciones(e_t deber ser:

$$e_t \sim N(0, \frac{1+\theta^2+2\theta\phi}{1-\phi^2})$$

De este modo se garantiza que cada z_t en la serie tenga distribución marginal $N(z_t|0,1)$. Después se procede a aplicar la transformación inversa a $S^{-1}_{\epsilon,\delta}(x_t)$. De este modo se obtiene la serie SAS-ARMA(1,1).

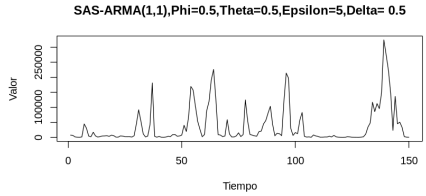
Luego se observa como se comporta la serie en función de los parámetros ϵ y δ dejando a un ϕ fijo en 0.5 y θ fijo en 0.5 igualmente.

5.1. Análisis conjunto de trayectorias SAS-ARMA(1,1)

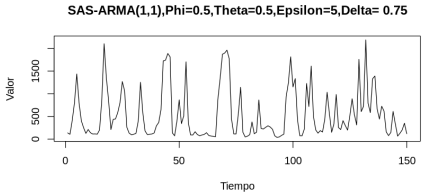
Consideramos series simuladas bajo SAS-ARMA$(1,1)$ con $\phi = \theta = 0.5$. Las diferencias entre paneles se deben a la distribución SAS de las innovaciones: ε controla la *asimetría* ($\varepsilon = 0$ simetría; $\varepsilon > 0$ sesgo a la derecha) y δ el *peso de cola/curtosis* ($\delta < 1$ colas pesadas; $\delta > 1$ colas ligeras). La persistencia temporal (ARMA) se mantiene; cambia la forma de los choques.

Efecto de δ con $\varepsilon = 5$

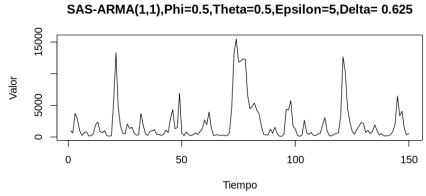
Al aumentar δ de 0.5 a 1 las colas se aligeran: disminuyen extremos y varianza; con $\delta = 5$ la serie queda comprimida y casi sin outliers, aunque se conserva la asimetría a la derecha.



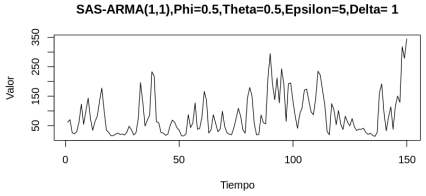
(a) $\varepsilon = 5,\ \delta = 0.5$



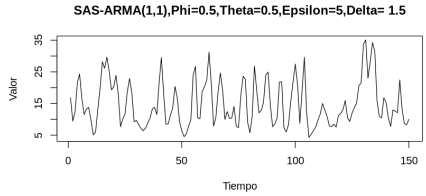
(b) $\varepsilon = 5,\ \delta = 0.625$

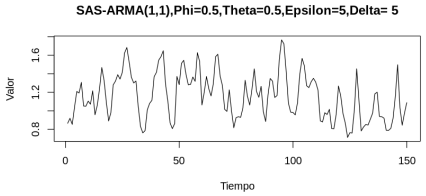


(c) $\varepsilon = 5,\ \delta = 0.75$



(d) $\varepsilon = 5,\ \delta = 1$





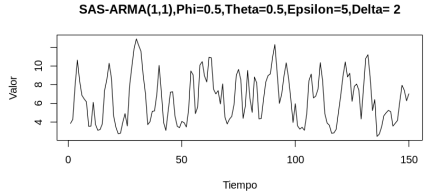
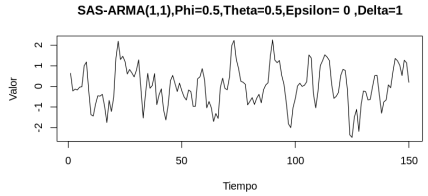


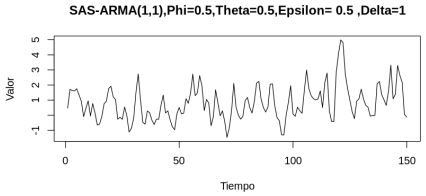
Figura 7: Efecto de δ con $\varepsilon = 5$: al crecer δ disminuyen la varianza y los picos, manteniendo asimetría positiva.

Efecto de ε con $\delta = 1$

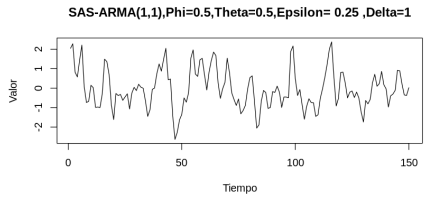
Con $\delta = 1$ (peso de cola medio), incrementar ε aumenta la asimetría: la serie pasa más tiempo en niveles positivos y los máximos superan a los mínimos.

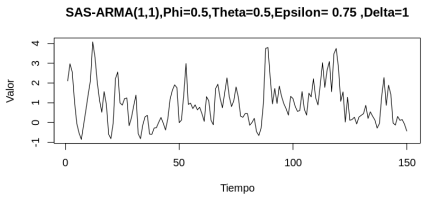


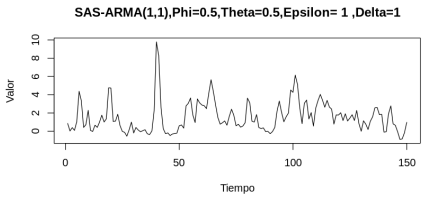
(a) $\varepsilon = 0,\ \delta = 1$



(b) $\varepsilon = 0.25,\ \delta = 1$





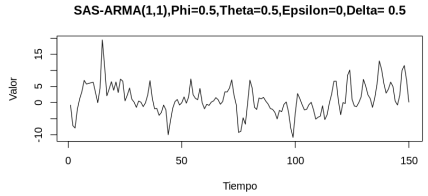


(c) $\varepsilon = 0.5,\ \delta = 1$

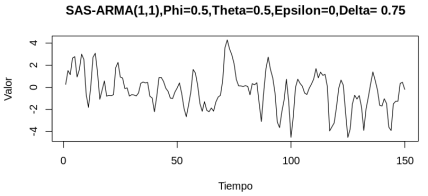
Figura 8: Efecto de ε con $\delta = 1$: mayor sesgo a la derecha conforme ε aumenta.

Efecto de δ con $\varepsilon = 0$

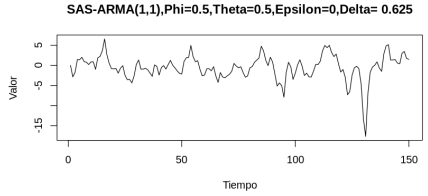
Con innovaciones simétricas, $\delta < 1$ induce outliers grandes en ambos signos y varianza alta; al acercarse a $\delta = 1$ la trayectoria es más regular; con $\delta \gg 1$ queda muy concentrada.



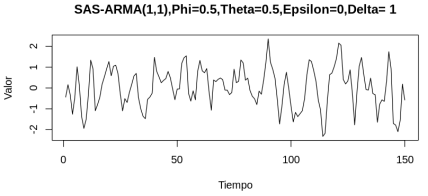
(a) $\varepsilon = 0,\ \delta = 0.5$



(b) $\varepsilon = 0,\ \delta = 0.625$



(c) $\varepsilon = 0,\ \delta = 0.75$



(d) $\varepsilon = 0,\ \delta = 1$

Figura 9: Efecto de δ con $\varepsilon = 0$: de colas pesadas (alta variabilidad y outliers) a colas ligeras (trayectorias comprimidas).

5.2. Estimación de parámetros

Log-verosimilitud y *scores* del modelo SAS-ARMA(1,1)

Paso 1: Definiciones (familia SAS y notación). Sea

$$S_{\varepsilon,\delta}(x) = \sinh\big(\delta\ \mathrm{asinh}(x) - \varepsilon\big), \qquad C_{\varepsilon,\delta}(x) = \cosh\big(\delta\ \mathrm{asinh}(x) - \varepsilon\big).$$

Definimos para cada t:

$$z_t := S_{\varepsilon,\delta}(x_t), \quad C_t := C_{\varepsilon,\delta}(x_t), \quad A_t := \delta\ \mathrm{asinh}(x_t) - \varepsilon, \quad D := 1 - \rho^2, \quad r_t := z_t - \rho z_{t-1}\ (t \geq 2).$$

Paso 2: Construcción de la log-verosimilitud. La verosimilitud dada se escribe como

$$L(\rho,\varepsilon,\delta)=\frac{\delta C_1}{\sqrt{1+x_1^2}}\,\varphi(z_1;0,1)\prod_{t=2}^{T}\frac{\delta C_t}{\sqrt{1+x_t^2}}\,\varphi\big(z_t;\rho z_{t-1},\,1-\rho^2\big),$$

por lo que, ignorando constantes aditivas ($-\frac{1}{2}\log 2\pi$), la log-verosimilitud es

$$\ell(\rho,\varepsilon,\delta)=\sum_{t=1}^{T}\Big[\log\delta+\log C_t-\tfrac{1}{2}\log(1+x_t^2)\Big]-\tfrac{1}{2}z_1^2+\sum_{t=2}^{T}\left\{-\tfrac{1}{2}\log D-\frac{r_t^2}{2D}\right\}.$$

Paso 3: Derivadas elementales que se usarán. Por la regla de la cadena, usando $z_t=\sinh(A_t)$ y $C_t=\cosh(A_t)$,

$$\frac{\partial z_t}{\partial\varepsilon}=-C_t,\quad \frac{\partial C_t}{\partial\varepsilon}=-z_t,\qquad \frac{\partial z_t}{\partial\delta}=\operatorname{asinh}(x_t)\,C_t,\quad \frac{\partial C_t}{\partial\delta}=\operatorname{asinh}(x_t)\,z_t.$$

Para los términos normales condicionales, con $y_t\equiv z_t$:

$$\frac{\partial\ell_t^N}{\partial y_t}=-\frac{r_t}{D},\qquad \frac{\partial\ell_t^N}{\partial y_{t-1}}=\frac{\rho\,r_t}{D},\qquad \frac{\partial\ell_t^N}{\partial\rho}=\frac{\rho}{D}+\frac{r_t y_{t-1}}{D}-\frac{\rho\,r_t^2}{D^2}\quad (t\geq 2).$$

Paso 4: *Score* respecto a ε. Se diferencian los términos de jacobiano ($\log C_t$), el término normal inicial ($-\frac{1}{2}z_1^2$) y los normales condicionales ($t\geq 2$). El resultado compacto es

$$\frac{\partial\ell}{\partial\varepsilon}=-\sum_{t=1}^{T}\frac{z_t}{C_t}+z_1C_1+\sum_{t=2}^{T}\left[\frac{r_t\,C_t}{D}-\frac{\rho\,r_t\,C_{t-1}}{D}\right].$$

Paso 5: *Score* respecto a δ. Se derivan (i) los jacobianos en $\log\delta$ y $\log C_t$, (ii) el término normal inicial y (iii) los términos condicionales. Se obtiene

$$\begin{aligned}\frac{\partial\ell}{\partial\delta}=\frac{T}{\delta}&+\sum_{t=1}^{T}\operatorname{asinh}(x_t)\,\frac{z_t}{C_t}-z_1\,\operatorname{asinh}(x_1)\,C_1\\&+\sum_{t=2}^{T}\left[-\frac{r_t\,\operatorname{asinh}(x_t)\,C_t}{D}+\frac{\rho\,r_t\,\operatorname{asinh}(x_{t-1})\,C_{t-1}}{D}\right].\end{aligned}$$

Paso 6: *Scores* respecto a ϕ y θ (vía la regla de la cadena). Aquí los parámetros entran sólo a través de

$$\rho(\phi,\theta)=\frac{(\theta+\phi)(1+\theta\phi)}{1+\theta^2+2\theta\phi}.$$

Primero, las derivadas parciales de ρ:

$$\frac{\partial \rho}{\partial \phi} = \frac{2\phi^2\theta^2 + 2\phi\theta^3 + 2\phi\theta + \theta^4 + 1}{\big(1 + \theta^2 + 2\theta\phi\big)^2}, \qquad \frac{\partial \rho}{\partial \theta} = \frac{(\phi - 1)(\phi + 1)(\theta - 1)(\theta + 1)}{\big(1 + \theta^2 + 2\theta\phi\big)^2}.$$

Luego, aplicando la regla de la cadena y sumando sobre $t \geq 2$:

$$\frac{\partial \ell}{\partial \phi} = \left[\sum_{t=2}^{T} \left(\frac{\rho}{D} + \frac{r_t z_{t-1}}{D} - \frac{\rho\, r_t^2}{D^2}\right)\right] \cdot \frac{\partial \rho}{\partial \phi},$$

$$\frac{\partial \ell}{\partial \theta} = \left[\sum_{t=2}^{T} \left(\frac{\rho}{D} + \frac{r_t z_{t-1}}{D} - \frac{\rho\, r_t^2}{D^2}\right)\right] \cdot \frac{\partial \rho}{\partial \theta}.$$

Con las expresiones anteriores. Estas ecuaciones están listas para implementarse en un método de optimización numérico. Sin embargo tanto para este modelo como el SAS-AR(1) no fue posible realizar la optimización y por lo mismo la obtención de estimadores.

Dado que fue posible la obtención de estimadores para los parametros, para obtener aproximaciones a los mismos, se hicieron pruebas bajo distintos valores de ϕ y θ (para el caso del SAS-ARMA(1,1), donde se probaban multiples combinaciones de parametros de modo que se maximisara la log-verosimilitud. Los valores de ϵ y δ fueron fijados en 0 y 0.5 respectivamente.

6. Simulaciones

El objetivo del estudio de simulación es evaluar el comportamiento de los estimadores en modelos SAS–AR(1) y SAS–ARMA(1,1) construidos mediante verosimilitud condicional y búsqueda en rejilla, así como explorar las dificultades de identificabilidad detectadas en el caso ARMA(1,1).

En una primera etapa se consideró un proceso SAS–AR(1). Para cada configuración se generaron $N = 1000$ muestras simuladas de longitud $T \in \{500, 1000\}$, fijando los parámetros de forma en $\varepsilon = 0$ y $\delta = 0.5$, y variando el valor verdadero de ϕ en un conjunto de valores representativo tanto de persistencia positiva como negativa:

$$\phi \in \{-0.90, -0.70, -0.50, -0.25, 0, 0.25, 0.50, 0.70, 0.90\}.$$

Para cada réplica se estimaron los parámetros mediante búsqueda exhaustiva en una malla discreta de valores para $(\phi, \varepsilon, \delta)$, y se almacenaron las estimaciones

$$\widehat{\phi}^{(i)}, \quad \widehat{\varepsilon}^{(i)}, \quad \widehat{\delta}^{(i)}, \qquad i = 1, \ldots, N.$$

El desempeño se resumió mediante la media muestral de cada estimador y su error cuadrático medio (MSE), definidos, por ejemplo, para ϕ como

$$\overline{\widehat{\phi}} = \frac{1}{N}\sum_{i=1}^{N} \widehat{\phi}^{(i)}, \qquad \mathrm{MSE}_\phi = \frac{1}{N}\sum_{i=1}^{N} \big(\widehat{\phi}^{(i)} - \phi\big)^2.$$

Este diseño permite responder a preguntas del tipo: (i) ¿qué tan bien recupera la metodología el parámetro autorregresivo para distintos niveles de dependencia? y (ii) ¿en qué medida los parámetros de forma ε y δ se mantienen estabilizados cuando se fijan en la simulación?

En una segunda etapa se analizaron procesos SAS–ARMA(1,1), manteniendo fijos los parámetros de forma $(\varepsilon, \delta) = (0, 0.5)$ y variando pares (ϕ, θ) en un grid que cubre valores positivos y negativos. Nuevamente se generaron $N = 1000$ réplicas para cada combinación de parámetros y se obtuvo, por búsqueda en rejilla, el cuádruplo estimado

$$\big(\widehat{\phi}^{(i)}, \widehat{\theta}^{(i)}, \widehat{\varepsilon}^{(i)}, \widehat{\delta}^{(i)}\big), \qquad i = 1, \ldots, N,$$

resumiendo el desempeño mediante medias y MSE. Este esquema permite estudiar tanto la estabilidad de los parámetros de forma como el *acoplamiento fuerte* entre ϕ y θ, observado en las tablas de resultados y discutido en las conclusiones.

Las subsecciones siguientes describen en más detalle el pseudocódigo de las funciones implementadas en `R` y presentan los resultados numéricos obtenidos.

6.1. Pseudocódigo de las funciones en `R` para el modelo SAS-AR(1)

A continuación se describe, de forma concisa y paso a paso, el procedimiento lógico (pseudocódigo) implementado en cada función usada para simular, transformar, evaluar la log-verosimilitud, estimar por búsqueda en malla y realizar un experimento Monte Carlo bajo un modelo `SAS-AR(1)`.

`sas_ar_1(t, phi, epsilon, delta)`
Objetivo: Generar una serie temporal con estructura AR(1) en el dominio normal y transformarla al dominio *sinh–arcsinh* (SAS).

1. Fijar $\rho \leftarrow \phi$.
2. Generar innovaciones gaussianas $\{e_{\text{norm},t}\}_{t=1}^{T} \sim \mathcal{N}(0,\, 1-\rho^2)$.
3. Inicializar vector `z` de longitud `t`; asignar $z_1 \leftarrow \rho \cdot 0 + e_{\text{norm}}$.
4. Para $i = 2, \ldots, t$: actualizar $z_i \leftarrow \rho\, z_{i-1} + e_{\text{norm},i}$.
5. Transformar a SAS: $x \leftarrow \sinh\!\left(\dfrac{\operatorname{asinh}(z) + \epsilon}{\delta}\right)$.
6. Devolver x.

`s(x, epsilon, delta)`
Objetivo: Transformar datos desde la escala SAS a la escala normal intermedia s usada en la verosimilitud.

1. Calcular $a \leftarrow \delta \operatorname{asinh}(x) - \epsilon$.
2. Devolver $s \leftarrow \sinh(a)$.

`ll_ar(x, phi, epsilon, delta)`
Objetivo: Evaluar la log-verosimilitud (hasta constante aditiva) del modelo AR(1) bajo la transformación SAS.

1. Calcular $a_t \leftarrow \delta \operatorname{asinh}(x_t) - \epsilon$ para todo t.
2. Calcular el *Jacobiano* punto a punto: $j_t \leftarrow \dfrac{\delta \cosh(a_t)}{\sqrt{1 + x_t^2}}$.
3. Obtener $s_t \leftarrow \sinh(a_t)$ (serie en la escala normal intermedia).
4. Fijar $\rho \leftarrow \phi$.
5. Armar la parte AR(1):
   - Término inicial: $v_1 \leftarrow \log j_1 - \frac{1}{2} s_1^2$.
   - Suma para $t = 2, \ldots, T$:

$$v_{\text{resto}} \leftarrow \sum_{t=2}^{T} \Big(-\tfrac{1}{2} \log(1 - \rho^2) + \log j_t - \tfrac{(s_t - \rho s_{t-1})^2}{2(1-\rho^2)} \Big).$$

6. Devolver $v \leftarrow v_1 + v_{\text{resto}}$.

`ar1_est(x)`
Objetivo: Estimar (ϕ, ϵ, δ) por búsqueda exhaustiva en malla (*grid search*) maximizando `ll_ar`.

1. Construir mallas: $\epsilon \in \{-2, -1.8, \ldots, 2\}$, $\delta \in \{0.1, 0.3, \ldots, 1.9\}$, $\phi \in \{-0.9, -0.7, \ldots, 0.9\}$.
2. Inicializar `max` con un valor muy pequeño y `best_params` vacío.
3. Para cada tripleta (ϕ, ϵ, δ) en la malla:
   a) Calcular $\ell \leftarrow$ `ll_ar`$(x, \phi, \epsilon, \delta)$.
   b) Si $\ell >$ `max`: actualizar `max` y `best_params` $\leftarrow (\phi, \epsilon, \delta)$.
4. Devolver `best_params`.

`montecarlo_ar1_tab1(t, phi, epsilon, delta, n)`
Objetivo: Estudiar desempeño de los estimadores por Monte Carlo; reporta medias y MSE.

1. Inicializar vectores para almacenar n estimaciones de $\widehat{\phi}, \widehat{\epsilon}, \widehat{\delta}$.
2. Para $i = 1, \ldots, n$:
   a) Simular $x^{(i)} \leftarrow$ `sas_ar_1`$(t, \phi, \epsilon, \delta)$.
   b) Estimar $(\widehat{\phi}^{(i)}, \widehat{\epsilon}^{(i)}, \widehat{\delta}^{(i)}) \leftarrow$ `ar1_est`$(x^{(i)})$.
   c) Almacenar las tres estimaciones en sus vectores.
3. Calcular MSE: $\text{MSE}_\phi \leftarrow \text{mean}\{(\widehat{\phi} - \phi)^2\}$, $\text{MSE}_\epsilon \leftarrow \text{mean}\{(\widehat{\epsilon} - \epsilon)^2\}$, $\text{MSE}_\delta \leftarrow \text{mean}\{(\widehat{\delta} - \delta)^2\}$.
4. Calcular medias muestrales: $\overline{\widehat{\phi}}, \overline{\widehat{\epsilon}}, \overline{\widehat{\delta}}$.
5. Devolver vector resumen: $\big(\overline{\widehat{\phi}}, \text{MSE}_\phi, \overline{\widehat{\epsilon}}, \text{MSE}_\epsilon, \overline{\widehat{\delta}}, \text{MSE}_\delta\big)$.

6.2. Pseudocódigo de las funciones en `R` para el modelo SAS-ARMA(1,1)

Se presenta el flujo lógico (pseudocódigo) de cada función empleada para simular, transformar, evaluar la log-verosimilitud, estimar parámetros por búsqueda en malla y ejecutar un estudio Monte Carlo bajo un esquema `SAS-ARMA(1,1)`.

`sas_arma_11(t, phi, theta, epsilon, delta)`
Objetivo: Generar una serie con dinámica ARMA(1,1) en la escala normal intermedia y luego transformarla a la familia *sinh–arcsinh* (SAS).

1. Convertir $\phi, \theta, \epsilon, \delta, t$ a numéricos.
2. Calcular el parámetro compuesto
$$\rho \leftarrow \frac{(\phi + \theta)\,(1 + \phi\theta)}{1 + \theta^2 + 2\phi\theta}.$$
3. Generar innovaciones gaussianas $\{e_{\text{norm},t}\}_{t=1}^{T} \sim \mathcal{N}(0, 1 - \rho^2)$.
4. Inicializar $\{z_t\}$ de longitud $T + \text{burnin}$ y fijar $z_1 \leftarrow \phi \cdot 0 + e_{\text{norm},1} + \theta \cdot 0$.
5. Para $i = 2, \ldots, T + \text{burnin}$: actualizar el proceso ARMA(1,1)
$$z_i \leftarrow \phi\, z_{i-1} + e_{\text{norm},i} + \theta\, e_{\text{norm},i-1}.$$
6. Transformar a SAS: $x \leftarrow \sinh\!\left(\dfrac{\text{asinh}(z) + \epsilon}{\delta}\right)$.
7. Eliminar el *burn-in*: conservar $x_{(\text{burnin}+1):(\text{burnin}+T)}$.
8. Devolver la serie x.

`s(x, epsilon, delta)`
Objetivo: Mapear observaciones en escala SAS a la escala normal intermedia s.

1. Calcular $a \leftarrow \delta \operatorname{asinh}(x) - \epsilon$.
2. Devolver $s \leftarrow \sinh(a)$.

`ll_arma(x, phi, theta, epsilon, delta)`
Objetivo: Evaluar la log-verosimilitud (hasta constante) usando la transformación SAS y una agregación AR(1) vía ρ.

1. Para cada x_t, calcular $a_t \leftarrow \delta \operatorname{asinh}(x_t) - \epsilon$.
2. Jacobiano punto a punto: $j_t \leftarrow \dfrac{\delta \cosh(a_t)}{\sqrt{1 + x_t^2}}$.
3. Escala intermedia: $s_t \leftarrow \sinh(a_t)$.
4. Calcular
$$\rho \leftarrow \frac{(\phi + \theta)\,(1 + \phi\theta)}{1 + \theta^2 + 2\phi\theta}.$$
5. Armar la verosimilitud tipo AR(1):
$$v_1 \leftarrow \log j_1 - \tfrac{1}{2} s_1^2,$$
$$v_{\text{resto}} \leftarrow \sum_{t=2}^{T} \Big(-\tfrac{1}{2} \log(1 - \rho^2) + \log j_t - \tfrac{(s_t - \rho s_{t-1})^2}{2(1-\rho^2)} \Big).$$
6. Devolver $v \leftarrow v_1 + v_{\text{resto}}$.

`arma11_est(x)`
Objetivo: Estimar $(\phi, \theta, \epsilon, \delta)$ por búsqueda exhaustiva en malla (*grid search*) maximizando `ll_arma`.

1. Definir mallas:
$$\epsilon \in [-2, 2] \text{ cada } 0.2, \qquad \delta \in [0.1, 1.9] \text{ cada } 0.2, \qquad \phi, \theta \in [-0.9, 0.9] \text{ cada } 0.2.$$
2. Inicializar `max` con un valor muy pequeño y `best_params` vacío.
3. Para cada cuádrupla $(\phi, \theta, \epsilon, \delta)$ en la malla:
   a) Evaluar $\ell \leftarrow$ `ll_arma`$(x, \phi, \theta, \epsilon, \delta)$.
   b) Si $\ell >$ `max`: actualizar `max` y `best_params` $\leftarrow (\phi, \theta, \epsilon, \delta)$.
4. Devolver `best_params`.

`montecarlo_arma11(t, phi,` $\theta, \epsilon, \delta, n$)
Objetivo: Evaluar el desempeño de los estimadores por simulación Monte Carlo; reporta medias y MSE.

1. Inicializar vectores para n réplicas de $\widehat{\phi}, \widehat{\theta}, \widehat{\epsilon}, \widehat{\delta}$.
2. Para $i = 1, \ldots, n$:
   a) Generar $x^{(i)} \leftarrow$ `sas_arma_11`$(t, \phi, \theta, \epsilon, \delta)$.
   b) Estimar $(\widehat{\phi}^{(i)}, \widehat{\theta}^{(i)}, \widehat{\epsilon}^{(i)}, \widehat{\delta}^{(i)}) \leftarrow$ `arma11_est`$(x^{(i)})$.
   c) Almacenar las estimaciones en sus vectores correspondientes.
3. Calcular los errores cuadráticos medios (MSE):

$$\mathrm{MSE}_\phi = \mathrm{mean}\{(\widehat{\phi} - \phi)^2\}, \quad \mathrm{MSE}_\theta = \mathrm{mean}\{(\widehat{\theta} - \theta)^2\},$$
$$\mathrm{MSE}_\epsilon = \mathrm{mean}\{(\widehat{\epsilon} - \epsilon)^2\}, \quad \mathrm{MSE}_\delta = \mathrm{mean}\{(\widehat{\delta} - \delta)^2\}.$$

4. Calcular medias muestrales: $\overline{\widehat{\phi}}$, $\overline{\widehat{\theta}}$, $\overline{\widehat{\epsilon}}$, $\overline{\widehat{\delta}}$.
5. Devolver el vector resumen

$$\big(\overline{\widehat{\phi}}, \mathrm{MSE}_\phi, \overline{\widehat{\theta}}, \mathrm{MSE}_\theta, \overline{\widehat{\epsilon}}, \mathrm{MSE}_\epsilon, \overline{\widehat{\delta}}, \mathrm{MSE}_\delta\big).$$

Tabla de resultados SAS-AR(1)

Columnas: ϕ verdadero; promedio de la estimación $\hat{\phi}$ y su MSE; $\hat{\epsilon}$ y su MSE; $\hat{\delta}$ y su MSE. Se muestran por separado $t = 500$ y $t = 1000$.

Cuadro 2: SAS-AR(1) con $t = 500$ (1000 réplicas). En negrita se resaltan los casos con $|\phi| \geq 0.5$, donde la recuperación de ϕ es prácticamente perfecta y el MSE es mínimo.

ϕ	$\hat{\phi}$	MSE($\hat{\phi}$)	$\hat{\epsilon}$	MSE($\hat{\epsilon}$)	$\hat{\delta}$	MSE($\hat{\delta}$)
0.90	**0.90**	**0.0000**	0.002	0.0020	0.50	0.0000
0.70	**0.70**	**0.0000**	-0.002	0.0020	0.50	0.0000
0.50	**0.50**	**0.0000**	-0.004	0.0016	0.50	0.0000
0.25	0.274	0.0051	0.000	0.0000	0.50	0.0000
0.00	-0.002	0.0100	0.000	0.0000	0.50	0.0000
-0.25	-0.276	0.0049	0.000	0.0000	0.50	0.0000
-0.50	**-0.50**	**0.0000**	0.000	0.0000	0.50	0.0000
-0.70	**-0.70**	**0.0000**	0.000	0.0000	0.50	0.0000
-0.90	**-0.90**	**0.0000**	0.000	0.0000	0.50	0.0000

Cuadro 3: SAS-AR(1) con $t = 1000$ (1000 réplicas). En negrita se resaltan nuevamente los casos con $|\phi| \geq 0.5$, donde el estimador de ϕ es prácticamente insesgado y el MSE es mínimo.

ϕ	$\hat{\phi}$	MSE($\hat{\phi}$)	$\hat{\epsilon}$	MSE($\hat{\epsilon}$)	$\hat{\delta}$	MSE($\hat{\delta}$)
0.90	**0.90**	**0.0000**	-0.002	0.0000	0.50	0.0000
0.70	**0.70**	**0.0000**	0.000	0.0000	0.50	0.0000
0.50	**0.50**	**0.0000**	0.000	0.0000	0.50	0.0000
0.25	0.292	0.0033	0.000	0.0000	0.50	0.0000
0.00	0.020	0.0100	0.000	0.0000	0.50	0.0000
-0.25	-0.290	0.0035	0.000	0.0000	0.50	0.0000
-0.50	**-0.50**	**0.0000**	0.000	0.0000	0.50	0.0000
-0.70	**-0.70**	**0.0000**	0.000	0.0000	0.50	0.0000
-0.90	**-0.90**	**0.0000**	0.000	0.0000	0.50	0.0000

Análisis e interpretación

Estabilidad de $\hat{\epsilon}$ y $\hat{\delta}$. En todas las configuraciones, $\hat{\epsilon}$ es prácticamente cero con MSE nulo o casi nulo, y $\hat{\delta}$ se mantiene exactamente en 0.5 con MSE $= 0$.

Comportamiento de $\hat{\phi}$ y MSE. Para $|\phi| \geq 0.5$, $\hat{\phi}$ es esencialmente insesgado y el MSE es prácticamente 0. En $|\phi| = 0.25$ aparece un sesgo leve de magnitud ($0.25 \to 0.274$; $-0.25 \to -0.276$ con t=500, y $0.25 \to 0.292$; $-0.25 \to -0.290$ con t=1000), con MSE pequeño (del orden de 3×10^{-3} a 5×10^{-3}). En $\phi = 0$ la estimación fluctúa alrededor de cero (MSE ≈ 0.01), reflejando variabilidad local y la resolución de la rejilla.

Efecto del tamaño muestral. Al pasar de $t = 500$ a $t = 1000$ disminuyen los MSE donde eran no nulos (por ejemplo, en ± 0.25), y se mantienen nulos donde ya lo eran. El patrón de pequeño sesgo en $|\phi| = 0.25$ persiste, coherente con efectos de discretización más que de falta de información.

Tabla de resultados SAS-ARMA(1,1)

Columnas: valor verdadero y promedio de la estimación de ϕ y θ ("$\hat{\phi}$" y "$\hat{\theta}$"), junto a sus MSE; además, $\hat{\epsilon}$ y $\hat{\delta}$ y sus MSE.

Análisis e interpretación

Estabilidad de $\hat{\epsilon}$ y $\hat{\delta}$. En todas las configuraciones, $\hat{\epsilon}$ permanece muy próximo a 0 con MSE minúsculos, y $\hat{\delta}$ se mantiene en 0.5 con MSE prácticamente nulo, indicando que estos parámetros no presentan problemas de identificación en la rejilla utilizada.

Cuadro 4: SAS-ARMA(1,1) con $t = 500$ (1000 réplicas). En negrita se resaltan los escenarios donde la identificación conjunta de ϕ y θ es más estable (MSE reducido en ambos parámetros).

ϕ	$\hat{\phi}$	MSE($\hat{\phi}$)	θ	$\hat{\theta}$	MSE($\hat{\theta}$)	$\hat{\epsilon}$	MSE($\hat{\epsilon}$)	$\hat{\delta}$	MSE($\hat{\delta}$)
0.90	**0.8996**	**0.0002**	0.00	0.0208	0.0108	−0.0016	0.0025	0.5002	0
0.70	0.8014	0.0296	0.25	−0.0004	0.2393	0.0006	0.0018	0.5000	0
0.50	**0.5050**	**0.0010**	0.50	**0.5178**	**0.0212**	0.0002	0.0012	0.5000	0
0.25	0.4998	0.1316	0.70	0.1938	0.4755	0.0010	0.0009	0.5000	0
0.00	0.3208	0.1674	0.90	0.2978	0.5332	0.0004	0.0002	0.5000	0
−0.25	−0.5446	0.1556	−0.90	−0.1716	0.7859	0.0000	0.0000	0.5000	0
−0.50	**-0.5302**	**0.0540**	−0.70	**-0.6428**	**0.0853**	0.0000	0.0000	0.5000	0
−0.70	**-0.7010**	**0.0002**	−0.50	**-0.5418**	**0.0260**	0.0000	0.0000	0.5000	0
−0.90	**-0.9000**	**0.0000**	−0.25	**-0.2754**	**0.0070**	0.0000	0.0000	0.5000	0

Cuadro 5: SAS-ARMA(1,1) con $t = 1000$ (1000 réplicas). En negrita se muestran nuevamente los escenarios con mejor identificación de los parámetros dinámicos ϕ y θ.

ϕ	$\hat{\phi}$	MSE($\hat{\phi}$)	θ	$\hat{\theta}$	MSE($\hat{\theta}$)	$\hat{\epsilon}$	MSE($\hat{\epsilon}$)	$\hat{\delta}$	MSE($\hat{\delta}$)
0.90	**0.9000**	**0.0000**	0.00	0.0232	0.0100	−0.0004	0.0004	0.5000	0
0.70	0.8404	0.0318	0.25	−0.1206	0.2471	0.0006	0.0000	0.5000	0
0.50	**0.5012**	**0.0002**	0.50	**0.5218**	**0.0084**	0.0002	0.0001	0.5000	0
0.25	0.5302	0.1567	0.70	0.1252	0.5583	−0.0004	0.0001	0.5000	0
0.00	0.3780	0.2014	0.90	0.0200	0.6406	0.0000	0.0000	0.5000	0
−0.25	−0.6100	0.2079	−0.90	−0.0478	1.0123	0.0000	0.0000	0.5000	0
−0.50	**-0.5314**	**0.0064**	−0.70	**-0.6232**	**0.0750**	0.0000	0.0000	0.5000	0
−0.70	**-0.7000**	**0.0000**	−0.50	**-0.5308**	**0.0153**	0.0000	0.0000	0.5000	0
−0.90	**-0.9000**	**0.0000**	−0.25	**-0.2898**	**0.0036**	0.0000	0.0000	0.5000	0

Comportamiento de $\hat{\phi}$. Cuando $\phi > 0$ y $\theta > 0$, $\hat{\phi}$ tiende a inflarse (p. ej., t=500: ϕ=0.25, θ=0.7 $\Rightarrow \hat{\phi} \approx 0.50$; t=1000: $\hat{\phi} \approx 0.53$). Con signos negativos coincidentes, la estimación de ϕ es muy precisa (p. ej., ϕ=−0.7 $\Rightarrow \hat{\phi} \approx -0.70$ y MSE casi nulo).

Comportamiento de $\hat{\theta}$. Para $\theta > 0$ (en especial con $\phi > 0$), $\hat{\theta}$ se atenúa marcadamente hacia 0 y el MSE crece (p. ej., t=500: θ=0.7 $\Rightarrow \hat{\theta} \approx 0.19$ con MSE$\approx$ 0.48; θ=0.9 $\Rightarrow \hat{\theta} \approx 0.30$ con MSE$\approx$ 0.53). Con $\phi < 0$ y $\theta < 0$ la identificación mejora sustancialmente (p. ej., t=1000: θ=−0.25 $\Rightarrow \hat{\theta} \approx -0.29$, MSE$\approx$ 0.0036).

Interacción ϕ–θ. Los patrones anteriores indican un acoplamiento fuerte entre ϕ y θ: con ambos positivos, múltiples pares (ϕ, θ) producen verosimilitudes similares, dando lugar a *crestas* donde la rejilla favorece $\hat{\phi}$ altos y $\hat{\theta}$ bajos. Con signos negativos coincidentes, la superficie es más curvada y los parámetros se separan mejor.

Al pasar de t=500 a t=1000, las zonas bien identificadas reducen aún más el MSE, pero el sesgo $\hat{\phi} \uparrow$, $\hat{\theta} \downarrow$ en el cuadrante positivo persiste, lo que sugiere que domina la estructura de la superficie de verosimilitud inducida por la búsqueda en rejilla.

7. Conclusiones y Trabajo Futuro

El objetivo de este trabajo fue estudiar modelos de series de tiempo con innovaciones no gaussianas basadas en la familia *sinh–arcsinh* (SAS), acopladas a dinámicas AR(1) y ARMA(1,1), utilizando verosimilitud condicional y simulaciones de Monte Carlo. En términos generales, los resultados muestran que:

- La familia SAS permite incorporar, de forma parsimoniosa, *asimetría* y *colas pesadas* en las innovaciones, manteniendo una construcción relativamente sencilla a partir de procesos ARMA gaussianos.
- En el caso SAS–AR(1), la estimación por búsqueda en rejilla recupera adecuadamente el parámetro autorregresivo ϕ cuando $|\phi|$ es moderado o alto ($|\phi| \geq 0.5$), y los parámetros de forma ε y δ permanecen estables en los escenarios considerados.
- En el caso SAS–ARMA(1,1), las simulaciones revelan un *acoplamiento fuerte* entre los parámetros ϕ y θ, que se traduce en sesgos sistemáticos: $\hat{\phi}$ tiende a inflarse y $\hat{\theta}$ a atenuarse hacia cero en el cuadrante donde ambos son positivos, mientras que la identificabilidad mejora cuando ϕ y θ son negativos.

En conjunto, estos resultados indican que la propuesta SAS–AR(1)/SAS–ARMA(1,1) es conceptualmente atractiva para modelar no normalidad en series de tiempo, pero que su implementación práctica requiere un tratamiento más cuidadoso de la estimación.

Discusión del acoplamiento fuerte en SAS–ARMA(1,1)

El fenómeno de *acoplamiento fuerte* entre ϕ y θ no es exclusivo del modelo SAS: incluso en ARMA(1,1) gaussiano, la autocorrelación de orden uno depende de la combinación

$$\rho(1) \;=\; \frac{(\theta+\phi)(1+\theta\phi)}{1+\theta^2+2\theta\phi},$$

de modo que distintos pares (ϕ, θ) pueden generar valores muy similares de $\rho(1)$. Cuando la verosimilitud (o la verosimilitud condicional) se apoya de manera predominante en la estructura de segundo orden (autocovarianzas de pocos rezagos), la superficie de log-verosimilitud tiende a presentar *crestas* a lo largo de curvas donde $\rho(1)$ es casi constante. Esto dificulta la identificación individual de ϕ y θ.

En el modelo SAS–ARMA(1,1), esta dificultad se ve amplificada por dos factores:

- La transformación SAS es monótona componente a componente; por lo tanto, la estructura de dependencia de segundo orden (autocorrelaciones) que proviene del proceso gaussiano subyacente se preserva en gran medida. En consecuencia, la verosimilitud condicional basada en densidades bivariadas SAS sigue siendo muy sensible a combinaciones de (ϕ, θ) que producen correlaciones similares, reforzando las crestas en la superficie de log-verosimilitud.
- El uso de verosimilitud condicional y una búsqueda en rejilla relativamente gruesa favorece soluciones que maximizan la verosimilitud a lo largo de dichas crestas, desplazando el máximo hacia regiones donde ϕ es mayor y θ menor (o viceversa), sin que ello implique una clara mejora en el ajuste dinámico.

No parece tratarse de una limitación intrínseca de la familia SAS como tal, sino de una combinación entre:

1. La conocida identificabilidad delicada de los parámetros ARMA(1,1).
2. La elección de una verosimilitud condicional basada en transiciones de un paso.
3. La estrategia de estimación por búsqueda en rejilla, que no explota derivadas ni información de segundo orden.

Una vía natural para mitigar este problema es adoptar enfoques de *estimación bayesiana* con *priores débilmente informativos* sobre ϕ y θ. Priorizando, por ejemplo, regiones donde los parámetros se separan en términos de su contribución a la autocorrelación (o reparametrizando en función de $\rho(1)$ y de un segundo parámetro de dependencia), podría reducirse la confusión entre modos de la verosimilitud y estabilizar las estimaciones en presencia de crestas.

Limitaciones del estudio

A la luz de los resultados obtenidos, es importante dejar explícitas las principales limitaciones de este trabajo:

- **Modelos de bajo orden.** El análisis se restringe a modelos SAS–AR(1) y SAS–ARMA(1,1). No se exploran órdenes más altos (ARMA(p,q) con $p, q > 1$), donde podrían aparecer patrones de identificabilidad diferentes o una separación más clara entre componentes AR y MA.

- **Ausencia de volatilidad cambiante en el tiempo.** Las simulaciones suponen innovaciones con varianza constante. No se incorporan estructuras de heterocedasticidad condicional (por ejemplo, GARCH o modelos de volatilidad estocástica), que son fundamentales en aplicaciones financieras y podrían interactuar de manera importante con la asimetría y las colas pesadas de la SAS.

- **Uso de verosimilitud condicional y búsqueda en rejilla.** La estimación se basa en verosimilitud condicional y en una búsqueda exhaustiva en malla. Esto evita los problemas de inestabilidad observados con algoritmos de optimización estándar, pero introduce discretización en el espacio de parámetros y puede subestimar la incertidumbre. No se logra un estimador máximo verosímil continuo ni se deriva la matriz de información observada.

- **Foco en escenarios simulados específicos.** Las simulaciones consideran valores particulares de $(\phi, \theta, \varepsilon, \delta)$ y tamaños muestrales fijos. No se realiza un estudio sistemático sobre un rango más amplio de configuraciones ni se analizan datos reales, lo que limita la generalización directa de los hallazgos.

- **Comparación acotada con otras familias.** Aunque se discuten y comparan conceptualmente SAS, skew-t y GEV, el trabajo no incluye una comparación empírica detallada (por ejemplo, mediante AIC, BIC o medidas de desempeño predictivo) entre estas familias en el contexto de series de tiempo.

Reconocer estas limitaciones es clave para situar adecuadamente el alcance de las conclusiones y para motivar investigaciones futuras más amplias.

Líneas de trabajo futuro

A partir de lo anterior, se identifican varias direcciones concretas y prometedoras para investigaciones futuras:

- **Estimación bayesiana y regularización.** Desarrollar esquemas bayesianos para modelos SAS–ARMA(1,1) (y extensiones) utilizando priores débilmente informativos sobre ϕ, θ, ε y δ. Esto podría ayudar a resolver el problema de acoplamiento fuerte mediante regularización, exploración completa de la superficie posterior y cuantificación más adecuada de la incertidumbre.

- **Extensión a modelos GARCH y volatilidad condicional (SAS–GARCH).** Integrar innovaciones SAS en modelos GARCH o en estructuras de volatilidad cambiante en el tiempo, permitiendo capturar simultáneamente asimetría, colas pesadas y *clústeres de volatilidad*. Esta línea es especialmente relevante en finanzas, donde las tres características suelen aparecer de forma conjunta.

- **Modelos de mayor orden y casos multivariados.** Extender el análisis a modelos SAS–ARMA(p,q) con $p, q > 1$, así como a modelos multivariados con márgenes SAS y estructuras de correlación más ricas. Ello permitiría estudiar si el problema de identificabilidad se atenúa o se transforma cuando la dinámica es más compleja.

- **Mejoras numéricas en la optimización.** Diseñar algoritmos de optimización específicos para la verosimilitud condicional SAS, combinando métodos de gradiente (o cuasi-Newton) con técnicas globales (por ejemplo, recocido simulado o algoritmos evolutivos) que exploren de forma más eficiente las crestas de la superficie de log-verosimilitud.

- **Comparación empírica con otras familias no gaussianas.** Aplicar los modelos SAS–AR(1) y SAS–ARMA(1,1) a conjuntos de datos reales (financieros, meteorológicos o industriales) y compararlos empíricamente con modelos basados en skew-t, Birnbaum–Saunders u otras distribuciones no gaussianas, utilizando criterios de información y medidas de desempeño en pronóstico.

- **Aproximaciones a la verosimilitud exacta SAS.** Explorar aproximaciones intermedias entre la verosimilitud condicional y la exacta, tales como verosimilitudes *compuestas* (por ejemplo, basadas en pares de observaciones) o esquemas de tipo *composite likelihood*, que puedan recuperar parte de la información perdida sin incurrir en el costo completo de la verosimilitud conjunta.

En síntesis, este trabajo muestra que la familia SAS ofrece un marco flexible y coherente para incorporar no normalidad en modelos de series de tiempo, pero también revela desafíos importantes de estimación e identificabilidad, especialmente en el caso ARMA(1,1). Abordar estos desafíos mediante nuevas técnicas de inferencia, extensiones a modelos con volatilidad condicional y aplicaciones empíricas constituye un camino natural para consolidar la utilidad práctica de los modelos SAS en el análisis de series de tiempo no gaussianas.

Anexo. Derivación detallada de los *scores* de la log-verosimilitud en el modelo SAS–AR(1)

A.1. Preliminares y notación

La familia SAS (canónica) se define por la transformación

$$S_{\varepsilon,\delta}(y) \;=\; \sinh\Big(\delta\,\operatorname{asinh}(y)-\varepsilon\Big), \qquad C_{\varepsilon,\delta}(y) \;=\; \cosh\Big(\delta\,\operatorname{asinh}(y)-\varepsilon\Big).$$

En lo que sigue fijamos la notación, para cada t:

$$A_t \equiv \delta\,\operatorname{asinh}(Y_t)-\varepsilon, \qquad s_t \equiv \sinh(A_t), \qquad C_t \equiv \cosh(A_t), \qquad J_t \equiv \frac{\partial s_t}{\partial Y_t} = \frac{\delta\, C_t}{\sqrt{1+Y_t^2}}.$$

El proceso latente en la escala s es un AR(1) gaussiano:

$$s_t \,\big|\, s_{t-1} \sim \mathcal{N}(\rho\, s_{t-1},\, 1-\rho^2), \qquad |\rho|<1,$$

de modo que, por cambio de variable, la densidad de transición de Y_t dado Y_{t-1} es

$$f(Y_t \mid Y_{t-1};\varepsilon,\delta,\rho) \;=\; \varphi\Big(s_t;\, \rho s_{t-1},\, 1-\rho^2\Big) \;\times\; J_t,$$

donde $\varphi(\cdot;\mu,\sigma^2)$ es la densidad normal.

A.2. Log-verosimilitud condicional

La verosimilitud condicional factoriza como $L = f(y_1)\prod_{t=2}^{T} f(y_t \mid y_{t-1})$. Ignorando constantes aditivas, la *log*-verosimilitud queda

$$\ell(\varepsilon,\delta,\rho) \;=\; \underbrace{\Big(\log J_1 - \tfrac{1}{2}s_1^2\Big)}_{\text{marginal } t=1} + \sum_{t=2}^{T}\left[-\tfrac{1}{2}\log(1-\rho^2) + \log J_t - \frac{(s_t-\rho s_{t-1})^2}{2(1-\rho^2)}\right].$$

Para simplificar, denotamos

$$D := 1-\rho^2, \qquad r_t := s_t - \rho s_{t-1} \quad (t\ge 2).$$

A.3. Derivadas elementales (regla de la cadena)

Como $s_t=\sinh(A_t)$ y $C_t=\cosh(A_t)$ con $A_t=\delta\,\operatorname{asinh}(Y_t)-\varepsilon$, se tiene

$$\frac{\partial s_t}{\partial \varepsilon} = -C_t, \qquad \frac{\partial s_t}{\partial \delta} = \operatorname{asinh}(Y_t)\, C_t,$$

$$\frac{\partial \log J_t}{\partial \varepsilon} = \frac{\partial}{\partial \varepsilon}\big(\log\delta + \log C_t - \tfrac{1}{2}\log(1+Y_t^2)\big) = -\tanh(A_t),$$

$$\frac{\partial \log J_t}{\partial \delta} = \frac{1}{\delta} + \operatorname{asinh}(Y_t)\,\tanh(A_t).$$

Para los términos normales condicionales ($t\ge 2$), vistos como funciones de (s_t, s_{t-1}, ρ),

$$\frac{\partial}{\partial s_t}\left(-\frac{r_t^2}{2D}\right) = -\frac{r_t}{D}, \qquad \frac{\partial}{\partial s_{t-1}}\left(-\frac{r_t^2}{2D}\right) = \frac{\rho\, r_t}{D}.$$

A.4. Score respecto a ε

Separando contribuciones de $t = 1$ y $t \geq 2$:

$$\frac{\partial \ell}{\partial \varepsilon} = \underbrace{\left(\frac{\partial \log J_1}{\partial \varepsilon} - \frac{1}{2}\,\frac{\partial (s_1^2)}{\partial \varepsilon}\right)}_{\text{marginal}} + \sum_{t=2}^{T} \left[\underbrace{\frac{\partial \log J_t}{\partial \varepsilon}}_{\text{jacobiano}} + \underbrace{\frac{\partial}{\partial \varepsilon}\left(-\frac{r_t^2}{2D}\right)}_{\text{parte normal}} \right].$$

$$\frac{\partial (s_1^2)}{\partial \varepsilon} = 2 s_1 \,\frac{\partial s_1}{\partial \varepsilon} = 2 s_1 (-C_1) = -2 s_1 C_1, \quad \Rightarrow \quad -\tfrac{1}{2}\,\frac{\partial (s_1^2)}{\partial \varepsilon} = +\, s_1 C_1,$$

$$\frac{\partial \log J_1}{\partial \varepsilon} = -\tanh(A_1).$$

Para $t \geq 2$,

$$\frac{\partial}{\partial \varepsilon}\left(-\frac{r_t^2}{2D}\right) = -\frac{r_t}{D}\left(\frac{\partial s_t}{\partial \varepsilon} - \rho\,\frac{\partial s_{t-1}}{\partial \varepsilon}\right) = -\frac{r_t}{D}\big(-C_t + \rho\, C_{t-1}\big) = \frac{r_t\, C_t}{D} - \frac{\rho\, r_t\, C_{t-1}}{D}.$$

Por tanto,

$$\boxed{\frac{\partial \ell}{\partial \varepsilon} = \Big[-\tanh(A_1) + s_1 C_1 \Big] + \sum_{t=2}^{T} \left\{ -\tanh(A_t) + \frac{r_t\, C_t}{D} - \frac{\rho\, r_t\, C_{t-1}}{D} \right\}.}$$

A.5. Score respecto a δ

Análogamente,

$$\frac{\partial \ell}{\partial \delta} = \underbrace{\left(\frac{\partial \log J_1}{\partial \delta} - \frac{1}{2}\,\frac{\partial (s_1^2)}{\partial \delta}\right)}_{\text{marginal}} + \sum_{t=2}^{T} \left[\underbrace{\frac{\partial \log J_t}{\partial \delta}}_{\text{jacobiano}} + \underbrace{\frac{\partial}{\partial \delta}\left(-\frac{r_t^2}{2D}\right)}_{\text{parte normal}} \right].$$

$$\frac{\partial (s_1^2)}{\partial \delta} = 2 s_1 \,\frac{\partial s_1}{\partial \delta} = 2 s_1 \,\operatorname{asinh}(Y_1)\, C_1, \qquad -\tfrac{1}{2}\,\frac{\partial (s_1^2)}{\partial \delta} = -\, s_1 \,\operatorname{asinh}(Y_1)\, C_1,$$

$$\frac{\partial \log J_1}{\partial \delta} = \frac{1}{\delta} + \operatorname{asinh}(Y_1)\,\tanh(A_1).$$

Para $t \geq 2$,

$$\frac{\partial}{\partial \delta}\left(-\frac{r_t^2}{2D}\right) = -\frac{r_t}{D}\left(\frac{\partial s_t}{\partial \delta} - \rho\,\frac{\partial s_{t-1}}{\partial \delta}\right) = -\frac{r_t}{D}\left(\operatorname{asinh}(Y_t) C_t - \rho\,\operatorname{asinh}(Y_{t-1}) C_{t-1}\right).$$

Como $\frac{\partial \log J_t}{\partial \delta} = \frac{1}{\delta} + \operatorname{asinh}(Y_t)\tanh(A_t)$,

$$\boxed{\begin{aligned}\frac{\partial \ell}{\partial \delta} &= \left[\frac{1}{\delta} + \operatorname{asinh}(Y_1)\tanh(A_1) - s_1\ \operatorname{asinh}(Y_1)\, C_1\right] \\ &\quad + \sum_{t=2}^{T}\left\{\frac{1}{\delta} + \operatorname{asinh}(Y_t)\tanh(A_t) - \frac{r_t\ \operatorname{asinh}(Y_t)\, C_t}{D}\right. \\ &\qquad \left. + \frac{\rho\, r_t\ \operatorname{asinh}(Y_{t-1})\, C_{t-1}}{D}\right\}.\end{aligned}}$$

A.6. Resumen de resultados (para implementación)

$$\boxed{\begin{aligned}\frac{\partial \ell}{\partial \varepsilon} &= \Big[-\tanh(A_1) + s_1 C_1\Big] + \sum_{t=2}^{T}\left\{-\tanh(A_t) + \frac{r_t\, C_t}{D} - \frac{\rho\, r_t\, C_{t-1}}{D}\right\}, \\ \frac{\partial \ell}{\partial \delta} &= \Big[\frac{1}{\delta} + \operatorname{asinh}(Y_1)\tanh(A_1) - s_1\ \operatorname{asinh}(Y_1)\, C_1\Big] \\ &\quad + \sum_{t=2}^{T}\left\{\frac{1}{\delta} + \operatorname{asinh}(Y_t)\tanh(A_t) - \frac{r_t\ \operatorname{asinh}(Y_t)\, C_t}{D}\right. \\ &\qquad \left. + \frac{\rho\, r_t\ \operatorname{asinh}(Y_{t-1})\, C_{t-1}}{D}\right\}.\end{aligned}}$$

Anexo B. Derivación detallada de los *scores* de la log-verosimilitud en el modelo SAS–ARMA(1,1)

B.1. Estructura del modelo y notación

Sea la transformación SAS (canónica):

$$S_{\varepsilon,\delta}(x) = \sinh\big(\delta\ \operatorname{asinh}(x) - \varepsilon\big), \qquad C_{\varepsilon,\delta}(x) = \cosh\big(\delta\ \operatorname{asinh}(x) - \varepsilon\big).$$

Para $t = 1, \ldots, T$ definimos

$$z_t := S_{\varepsilon,\delta}(x_t), \quad C_t := C_{\varepsilon,\delta}(x_t), \quad A_t := \delta\ \operatorname{asinh}(x_t) - \varepsilon.$$

En la escala z, consideramos el vector bivariado (z_{t-1}, z_t) como normal con correlación

$$\rho \equiv \rho(\phi, \theta) = \frac{(\phi + \theta)\,(1 + \phi\,\theta)}{1 + 2\,\phi\,\theta + \theta^2},$$

que es la autocorrelación de orden 1 del ARMA(1,1). Así,

$$(z_{t-1}, z_t)\ \sim\ \mathcal{N}_2\left(\mathbf{0},\ \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right), \qquad |\rho| < 1.$$

Densidades involucradas. La densidad marginal SAS de x_t y la bivariada inducida para (x_{t-1}, x_t) se obtienen por cambio de variable desde las normales estándar (univariada y bivariada). El jacobiano univariado es

$$J_t \;=\; \left|\frac{\partial z_t}{\partial x_t}\right| = \frac{\delta\, C_t}{\sqrt{1+x_t^2}}.$$

B.2. Verosimilitud condicional (pareada a rezago 1)

Usamos la factoría

$$L(\varepsilon,\delta,\phi,\theta) = f(x_1;\varepsilon,\delta)\prod_{t=2}^{T} f(x_t \mid x_{t-1};\varepsilon,\delta,\phi,\theta),$$

donde

$$f(x_t \mid x_{t-1}) = \varphi\big(z_t;\ \rho\, z_{t-1},\ 1-\rho^2\big)\ \times\ J_t, \qquad \varphi(\cdot;\mu,\sigma^2) \text{ es densidad normal.}$$

Ignorando constantes aditivas ($-\frac{1}{2}\log 2\pi$), la log-verosimilitud es

$$\ell(\varepsilon,\delta,\phi,\theta) = \sum_{t=1}^{T}\Big[\log\delta + \log C_t - \tfrac{1}{2}\log(1+x_t^2)\Big] - \tfrac{1}{2}z_1^2 + \sum_{t=2}^{T}\left\{-\tfrac{1}{2}\log D - \frac{r_t^2}{2D}\right\},$$

con la notación

$$D := 1-\rho^2, \qquad r_t := z_t - \rho\, z_{t-1} \quad (t \ge 2).$$

B.3. Derivadas elementales (regla de la cadena)

De $z_t = \sinh(A_t)$ y $C_t = \cosh(A_t)$ se obtienen:

$$\frac{\partial z_t}{\partial \varepsilon} = -C_t, \qquad \frac{\partial z_t}{\partial \delta} = \operatorname{asinh}(x_t)\, C_t,$$

$$\frac{\partial \log C_t}{\partial \varepsilon} = -\frac{z_t}{C_t} = -\tanh(A_t), \qquad \frac{\partial \log C_t}{\partial \delta} = \operatorname{asinh}(x_t)\,\frac{z_t}{C_t} = \operatorname{asinh}(x_t)\ \tanh(A_t).$$

Además,

$$\frac{\partial \log J_t}{\partial \varepsilon} = \frac{\partial \log C_t}{\partial \varepsilon} = -\tanh(A_t), \qquad \frac{\partial \log J_t}{\partial \delta} = \frac{1}{\delta} + \operatorname{asinh}(x_t)\ \tanh(A_t).$$

Para los términos normales condicionales ($t \ge 2$), vistos como función de (z_t, z_{t-1}, ρ):

$$\frac{\partial}{\partial z_t}\left(-\frac{r_t^2}{2D}\right) = -\frac{r_t}{D}, \qquad \frac{\partial}{\partial z_{t-1}}\left(-\frac{r_t^2}{2D}\right) = \frac{\rho\, r_t}{D},$$

$$\frac{\partial}{\partial \rho}\left[-\tfrac{1}{2}\log D - \frac{r_t^2}{2D}\right] = \frac{\rho}{D} + \frac{r_t\, z_{t-1}}{D} - \frac{\rho\, r_t^2}{D^2}.$$

B.4. *Score* respecto a ε

Separando $t = 1$ y $t \geq 2$:

$$\frac{\partial \ell}{\partial \varepsilon} = \underbrace{\left(\frac{\partial \log J_1}{\partial \varepsilon} - \frac{1}{2} \frac{\partial (z_1^2)}{\partial \varepsilon} \right)}_{\text{marginal}} + \sum_{t=2}^{T} \left[\underbrace{\frac{\partial \log J_t}{\partial \varepsilon}}_{\text{jacobiano}} + \underbrace{\frac{\partial}{\partial \varepsilon} \left(-\frac{r_t^2}{2D} \right)}_{\text{normal}} \right].$$

Como $\partial(z_1^2)/\partial\varepsilon = 2z_1(-C_1)$,

$$-\frac{1}{2} \frac{\partial (z_1^2)}{\partial \varepsilon} = + z_1 C_1, \qquad \frac{\partial \log J_1}{\partial \varepsilon} = - \tanh(A_1).$$

Para $t \geq 2$,

$$\frac{\partial}{\partial \varepsilon} \left(-\frac{r_t^2}{2D} \right) = -\frac{r_t}{D} \left(\frac{\partial z_t}{\partial \varepsilon} - \rho \, \frac{\partial z_{t-1}}{\partial \varepsilon} \right) = -\frac{r_t}{D} \left(-C_t + \rho \, C_{t-1} \right) = \frac{r_t C_t}{D} - \frac{\rho \, r_t C_{t-1}}{D}.$$

Por tanto,

$$\boxed{\frac{\partial \ell}{\partial \varepsilon} = \Big[- \tanh(A_1) + z_1 C_1 \Big] + \sum_{t=2}^{T} \left\{ - \tanh(A_t) + \frac{r_t \, C_t}{D} - \frac{\rho \, r_t \, C_{t-1}}{D} \right\}.}$$

B.5. *Score* respecto a δ

Análogamente,

$$\frac{\partial \ell}{\partial \delta} = \left(\frac{\partial \log J_1}{\partial \delta} - \frac{1}{2} \frac{\partial (z_1^2)}{\partial \delta} \right) + \sum_{t=2}^{T} \left[\frac{\partial \log J_t}{\partial \delta} + \frac{\partial}{\partial \delta} \left(-\frac{r_t^2}{2D} \right) \right].$$

Como $\partial(z_1^2)/\partial\delta = 2z_1 \, \operatorname{asinh}(x_1) C_1$,

$$-\frac{1}{2} \frac{\partial (z_1^2)}{\partial \delta} = - \, z_1 \, \operatorname{asinh}(x_1) \, C_1, \qquad \frac{\partial \log J_1}{\partial \delta} = \frac{1}{\delta} + \operatorname{asinh}(x_1) \, \tanh(A_1).$$

Para $t \geq 2$,

$$\frac{\partial}{\partial \delta} \left(-\frac{r_t^2}{2D} \right) = -\frac{r_t}{D} \left(\operatorname{asinh}(x_t) C_t - \rho \, \operatorname{asinh}(x_{t-1}) C_{t-1} \right).$$

Con $\partial \log J_t / \partial \delta = \frac{1}{\delta} + \operatorname{asinh}(x_t) \tanh(A_t)$,

$$\begin{aligned} \frac{\partial \ell}{\partial \delta} = & \left[\frac{1}{\delta} + \operatorname{asinh}(x_1) \tanh(A_1) - z_1 \, \operatorname{asinh}(x_1) \, C_1 \right] \\ & + \sum_{t=2}^{T} \left\{ \frac{1}{\delta} + \operatorname{asinh}(x_t) \tanh(A_t) - \frac{r_t \, \operatorname{asinh}(x_t) \, C_t}{D} + \frac{\rho \, r_t \, \operatorname{asinh}(x_{t-1}) \, C_{t-1}}{D} \right\}. \end{aligned} \tag{13}$$

B.6. *Scores* respecto a ϕ y θ (regla de la cadena)

Los parámetros (ϕ, θ) entran sólo mediante $\rho(\phi, \theta)$. Definamos

$$A := (\phi + \theta)(1 + \phi\theta), \qquad B := 1 + 2\phi\theta + \theta^2, \qquad \rho = \frac{A}{B}, \quad D = 1 - \rho^2.$$

Derivadas parciales de ρ:

$$\frac{\partial \rho}{\partial \phi} = \frac{(1 + 2\phi\theta + \theta^2)\, B - 2\theta\, A}{B^2}, \qquad \frac{\partial \rho}{\partial \theta} = \frac{(1 + 2\phi\theta + \phi^2)\, B - (2\phi + 2\theta)\, A}{B^2}.$$

Además, de B.3,

$$\frac{\partial \ell}{\partial \rho} = \sum_{t=2}^{T} \left(\frac{\rho}{D} + \frac{r_t\, z_{t-1}}{D} - \frac{\rho\, r_t^2}{D^2} \right).$$

Por regla de la cadena:

$$\boxed{\frac{\partial \ell}{\partial \phi} = \frac{\partial \ell}{\partial \rho} \cdot \frac{\partial \rho}{\partial \phi}, \qquad \frac{\partial \ell}{\partial \theta} = \frac{\partial \ell}{\partial \rho} \cdot \frac{\partial \rho}{\partial \theta}.}$$

Referencias

[1] Brockwell, P. J., & Davis, R. A. (2016). *Introduction to time series and forecasting* (3.ª ed.). Springer. https://doi.org/10.1007/978-3-319-29854-2

[2] Jones, M. C., & Pewsey, A. (2009). Sinh–arcsinh distributions. *Biometrika, 96*(4), 761–780. https://doi.org/10.1093/biomet/asp053

[3] Thekkedath, R., Balakrishnan, N., & Balakrishna, N. (2018). Time series with Birnbaum–Saunders marginal distributions. *Applied Stochastic Models in Business and Industry, 34*(4), 562–581. https://doi.org/10.1002/asmb.2324

[4] ContrerasEspinoza, S., CaamañoCarrillo, C., & ContrerasReyes, J. E. (2023). Generalized autoregressive score models based on sinh–arcsinh distributions for time series analysis. *Journal of Computational and Applied Mathematics, 423*, 114975. https://doi.org/10.1016/j.cam.2022.114975

[5] Leiva, V., Marchant, C., Saulo, H., Aslam, M., & Rojas, F. (2014). Capability indices for Birnbaum–Saunders processes applied to electronic and food industries. *Journal of Applied Statistics, 41*(9), 1881–1902. https://doi.org/10.1080/02664763.2014.897690

[6] Leiva, V., Athayde, E., Azevedo, C., & Marchant, C. (2011). Modeling wind energy flux by a Birnbaum–Saunders distribution with an unknown shift parameter. *Journal of Applied Statistics, 38*(12), 2819–2838. https://doi.org/10.1080/02664763.2011.570319

[7] GarcíaPapani, F., UribeOpazo, M. A., Leiva, V., & Barros, M. (2017). Birnbaum–Saunders spatial modelling and diagnostics applied to agricultural engineering data. *Stochastic Environmental Research and Risk Assessment, 31*(1), 105–124. https://doi.org/10.1007/s00477-015-1204-4