

# Análisis Comparativo de Modelos de Predicción Probabilística

## Test de Diebold-Mariano con Corrección de Bonferroni

Universidad Nacional de Colombia

3 de noviembre de 2025

# Contenido

- 1 Introducción
- 2 Ranking Global
- 3 Escenario: Estacionario Lineal
- 4 Escenario: No Lineal Estacionario
- 5 Escenario: No Estacionario Lineal
- 6 Análisis Comparativo
- 7 Conclusiones
- 8 Anexos

## Objetivo

Evaluar y comparar 9 modelos de predicción probabilística bajo diferentes escenarios de simulación utilizando métodos estadísticamente rigurosos.

### Modelos Analizados:

- AREPD
- AV-MCPS
- Block Bootstrapping
- DeepAR
- EnCQR-LSTM
- LSPM
- LSPMW
- MCPS
- Sieve Bootstrap

## Escenarios:

- Estacionario Lineal
- No Lineal Estacionario
- No Estacionario Lineal

## Variables de Control:

- Paso (1-5)
- Tipo de Modelo
- Distribución del error
- Varianza del error (0.2, 0.5, 1.0, 3.0)

## Metodología Estadística:

### Test de Diebold-Mariano

- Comparaciones pareadas
- Corrección de Bonferroni
- $\alpha_{corregido} = \frac{0,05}{36} = 0,001389$
- Total: 2000 observaciones

# Ranking Global: Test de Diebold-Mariano

Rank	Modelo	V	D	E	Score	% Victorias
1	<b>Block Bootstrapping</b>	8	0	0	<b>8</b>	<b>100.0 %</b>
2	<b>Sieve Bootstrap</b>	7	1	0	<b>6</b>	<b>87.5 %</b>
3	LSPM	6	2	0	4	75.0 %
4	LSPMW	5	3	0	2	62.5 %
5	AV-MCPS	3	4	1	-1	37.5 %
6	DeepAR	3	4	1	-1	37.5 %
7	MCPS	2	6	0	-4	25.0 %
8	EnCQR-LSTM	1	7	0	-6	12.5 %
9	AREPD	0	8	0	-8	0.0 %

## Interpretación

**V:** Victorias, **D:** Derrotas, **E:** Empates, **Score:** V - D

# Ranking Global: Visualización

Comparativo/comparacion\_global\_dm.png

# Matriz de Superioridad Global

Comparativo/heatmap\_global\_dm.png

# Estacionario Lineal: Ranking DM

Rank	Modelo	Score	%V
1	Block Bootstrap.	7	87.5
2	Sieve Bootstrap	7	87.5
3	DeepAR	4	75.0
4	MCPS	1	50.0
5	AV-MCPS	1	50.0
6	EnCQR-LSTM	-2	37.5
7	AREPD	-5	12.5
8	LSPM	-5	12.5
9	LSPMW	-8	0.0

## Hallazgos Clave:

- Métodos Bootstrap dominan
- Block Bootstrap invicto (8-0-0)
- DeepAR mejor modelo DL
- LSPM/LSPMW bajo rendimiento

## Mejor configuración:

- Paso: 1
- Varianza: 0.2
- Distribución: Exponencial

# Estacionario Lineal: Análisis Completo

Comparativo/comparativo\_dm\_Estacionario\_Lineal.png

# Ejemplo: Block Bootstrapping (Mejor Modelo)

Estacionario\_Lineal/Block Bootstrapping/analisis\_paso...

# Block Bootstrapping: Análisis por Tipo y Distribución

Estacionario\_Lineal/Block Bootstrapping

Estacionario\_itoi\_lineal/Block Bootstrapping

# Block Bootstrapping: Configuraciones Extremas

Estacionario\_Lineal/Block Bootstrapping/configuraciones

# No Lineal Estacionario: Ranking DM

Rank	Modelo	Score	%V
1	Block Bootstrap.	6	75.0
2	Sieve Bootstrap	6	75.0
3	DeepAR	4	50.0
4	EnCQR-LSTM	1	37.5
5	MCPS	0	25.0
6	AV-MCPS	-2	12.5
7	AREPD	-4	0.0
8	LSPM	-4	12.5
9	LSPMW	-7	0.0

## Hallazgos Clave:

- Bootstrap mantiene superioridad
- Mayor variabilidad en errores
- DeepAR resiliente a no linealidad
- Modelos BILINEAR problemáticos

## Desafíos:

- Outliers extremos (máx: 10.67)
- Mayor sensibilidad al tipo TAR(2,1)

# No Lineal Estacionario: Análisis Completo

Comparativo/comparativo\_dm\_No\_Lineal\_Estacionario.png

# DeepAR en No Lineal: Análisis por Paso

No\_Lineal\_Estacionario/DeepAR/analisis\_paso.png

No\_Lineal\_Estacionario/DeepAR/analisis\_interacciones.ipynb

# No Estacionario Lineal: Ranking DM

Rank	Modelo	Score	%V
1	Block Bootstrap.	8	100.0
2	Sieve Bootstrap	6	87.5
3	LSPM	3	62.5
4	LSPMW	3	62.5
5	AV-MCPS	-1	37.5
6	DeepAR	-1	37.5
7	MCPS	-4	25.0
8	EnCQR-LSTM	-6	12.5
9	AREPD	-8	0.0

## Hallazgos Clave:

- Block Bootstrap perfecto (100 %)
- LSPM/LSPMW mejoran notablemente
- Deep Learning bajo rendimiento
- ARIMA(2,1,2) muy problemático

## Peor caso:

- Errores hasta 122.83
- ARIMA(2,1,2) + normal + var=3.0

# No Estacionario Lineal: Análisis Completo

Comparativo/comparativo\_dm\_No\_Estacionario\_Lineal.png

# LSPM en No Estacionario: Análisis por Varianza

No\_Estacionario\_Lineal/LSPM/analisis\_varianza.png

# Sieve Bootstrap: Configuraciones Extremas

No\_Estacionario\_Lineal/Sieve Bootstrap/configuraciones

# Comparación entre Escenarios

Modelo	Estac. Lineal	No Lineal	No Estac.
Block Bootstrap	<b>1º (87.5 %)</b>	<b>1º (75.0 %)</b>	<b>1º (100 %)</b>
Sieve Bootstrap	2º (87.5 %)	2º (75.0 %)	2º (87.5 %)
DeepAR	3º (75.0 %)	3º (50.0 %)	6º (37.5 %)
LSPM	8º (12.5 %)	8º (12.5 %)	3º (62.5 %)
LSPMW	9º (0.0 %)	9º (0.0 %)	4º (62.5 %)
AREPD	7º (12.5 %)	7º (0.0 %)	9º (0.0 %)

## Conclusión

**Block Bootstrapping** es el modelo más robusto: 1º en todos los escenarios

## Efecto de la Varianza:

- Correlación promedio: 0.79
- Todos los modelos degradan
- Bootstrap más robustos
- $\text{Var}=0.2$  óptima para todos

## Efecto del Paso:

- Paso 1 generalmente mejor
- Degradación gradual
- Paso 5: mayor incertidumbre

## Efecto de Distribución:

- Exponencial: mejor (7/9 modelos)
- Mixture y uniform: problemáticas
- Normal: intermedia

## Efecto del Tipo de Modelo:

- MA(2) favorece mayoría
- AR(1) problemático
- ARIMA(2,1,2) evitar
- TAR(2,1) alta variabilidad

# Conclusiones Principales

## ① Superioridad de métodos Bootstrap:

- Block Bootstrapping: 1º en todos los escenarios (invicto globalmente)
- Sieve Bootstrap: 2º consistentemente
- Alta estabilidad y robustez

## ② Deep Learning con rendimiento mixto:

- DeepAR: Bueno en estacionarios (3º), pobre en no estacionarios (6º)
- EnCQR-LSTM: Consistentemente en posiciones bajas (6º-8º)
- No justifican su complejidad computacional

## ③ LSPM/LSPMW dependientes del contexto:

- Pobres en estacionarios (8º-9º)
- Competitivos en no estacionarios (3º-4º)
- Especializados en tendencias

# Recomendaciones Prácticas

## Para Datos Estacionarios

**Usar:** Block Bootstrapping o Sieve Bootstrap

**Evitar:** LSPM, LSPMW, AREPD

## Para Datos No Lineales

**Usar:** Block Bootstrapping, Sieve Bootstrap, DeepAR

**Considerar:** EnCQR-LSTM si se priorizan cuantiles

**Evitar:** LSPMW, AREPD

## Para Datos No Estacionarios

**Usar:** Block Bootstrapping (obligatorio)

**Alternativas:** Sieve Bootstrap, LSPM, LSPMW

**Evitar:** AREPD, EnCQR-LSTM, Deep Learning

# Configuraciones Óptimas

Escenario	Modelo	Paso	Var.	Dist.
Estacionario	Block Bootstrap	1	0.2	Exponencial
No Lineal	Block Bootstrap	3-5	0.2	Exponencial
No Estacionario	Block Bootstrap	1	0.2	Exponencial

## Configuración Universal

**Block Bootstrapping** + Varianza=0.2 + Distribución Exponencial  
(Rendimiento óptimo en 100 % de los casos)

# Top 5 Configuraciones Globales

Rank	Modelo	Escenario	Paso	Tipo	Error
1	Block Bootstrap	No Estacionario	1	ARIMA(1,1,0)	0.2211
2	Block Bootstrap	Estacionario	1	MA(2)	0.2215
3	Sieve Bootstrap	No Estacionario	3	ARIMA(0,1,0)	0.2223
4	Block Bootstrap	No Lineal	1	BILINEAR(1)	0.2224
5	Sieve Bootstrap	Estacionario	1	AR(2)	0.2210

## Observaciones:

- 4 de 5 son Bootstrap
- Todos con varianza = 0.2
- Paso 1 dominante (4 de 5)
- Distribución exponencial (todos)

# Modelos a Evitar

## Bajo Ninguna Circunstancia

**AREPD:** 0 victorias, 8 derrotas consistentes

No competitivo en ningún escenario

## Evitar en Estacionarios

**LSPM y LSPMW:** Rankings 8º-9º

Especializados solo para no estacionarios

## Cuidado con Configuraciones Extremas

- ARIMA(2,1,2) + normal + var=3.0: **ERROR hasta 122.83**
- BILINEAR(1) + uniform + var=3.0: **ERROR hasta 10.67**
- TAR(2,1) + cualquier + var=3.0: **Alta variabilidad**

## ① Optimización de Hiperparámetros:

- Tamaño de bloque en Block Bootstrap
- Orden del sieve en Sieve Bootstrap
- Arquitecturas de redes neuronales

## ② Análisis de Costos Computacionales:

- Trade-off precisión vs. tiempo
- Escalabilidad a datasets grandes

## ③ Validación en Datos Reales:

- Series financieras
- Datos climáticos
- Demanda energética

## ④ Métodos Híbridos:

- Combinación Bootstrap + Deep Learning
- Ensembles de modelos top-3

# Metodología: Test de Diebold-Mariano

## Definición

Prueba estadística para comparar la precisión de dos pronósticos:

## Estadístico DM:

$$DM = \frac{\bar{d}}{\sqrt{\text{Var}(d)/T}}$$

donde  $d_t = e_{1,t} - e_{2,t}$  (diferencia de errores)

## Corrección de Bonferroni:

$$\alpha_{corregido} = \frac{\alpha}{n_{comparaciones}} = \frac{0,05}{36} = 0,001389$$

- Controla error Tipo I en comparaciones múltiples
- Más conservador pero más confiable
- 36 comparaciones (9 modelos, pareados)

# Estadísticas Descriptivas Globales

Modelo	Media	Mediana	Std	Min	Max	CV
Block Bootstrap	0.556	0.487	0.289	0.221	6.233	52.0 %
Sieve Bootstrap	0.595	0.500	0.693	0.221	6.200	116.5 %
DeepAR	1.184	0.681	2.104	0.229	39.472	177.7 %
LSPM	1.164	0.733	1.683	0.223	10.669	144.6 %
LSPMW	1.183	0.750	1.688	0.224	10.748	142.7 %
AV-MCPS	1.149	0.647	2.820	0.223	61.829	245.4 %
MCPS	1.465	0.716	3.263	0.229	49.653	222.8 %
EnCQR-LSTM	1.825	1.123	3.082	0.226	35.649	168.9 %
AREPD	3.754	1.009	9.139	0.228	122.832	243.4 %

**CV:** Coeficiente de Variación ( $\text{Std}/\text{Media} \times 100\%$ )

# Correlación: Rendimiento vs Varianza

Modelo	Correlación
Block Bootstrapping	0.978
Sieve Bootstrap	0.974
DeepAR	0.854
AV-MCPS	0.848
MCPS	0.881
EnCQR-LSTM	0.805
LSPM	0.577
LSPMW	0.584
AREPD	0.317

## Interpretación

Correlaciones altas ( $>0.9$ ): Rendimiento predecible por varianza

Correlaciones bajas ( $<0.6$ ): Otros factores dominantes

# Ejemplo: AREPD (Peor Modelo)

No\_Estacionario\_Lineal/AREPD/analisis\_varianza.png

# Ejemplo: EnCQR-LSTM en No Estacionario

No\_Estacionario\_Lineal/EnCQR-LSTM/configuraciones\_extra

# Resumen Ejecutivo: Tabla de Decisión

Situación	Modelo Recomendado	Alternativa	Evitar
Serie estacionaria	Block Bootstrap	Sieve Bootstrap	LSPM, LSPMW
Serie no lineal	Block Bootstrap	DeepAR	AREPD, LSPMW
Serie con tendencia	Block Bootstrap	LSPM	AREPD, EnCQR-LSTM
Baja varianza (0.2)	Cualquiera	Block Bootstrap	N/A
Alta varianza (3.0)	Block Bootstrap	Sieve Bootstrap	Deep Learning
Corto plazo (paso 1)	Block Bootstrap	Todos competitivos	N/A
Largo plazo (paso 5)	Block Bootstrap	Sieve Bootstrap	AREPD

## Regla de Oro

**En caso de duda: Block Bootstrapping**

(100 % de victorias, robusto en todos los escenarios)

# Gracias

Preguntas y Discusión

## Información de Contacto

Universidad Nacional de Colombia  
Departamento de Estadística  
[contacto@universidad.edu.co](mailto:contacto@universidad.edu.co)

*Todos los resultados están disponibles en:*

[./Resultados/](#)

# Backup: LSPM en Estacionario Lineal

Estacionario\_Lineal/LSPM/analisis\_paso.png

# Backup: MCPS Análisis Completo

Estacionario\_Lineal/MCPS/configuraciones\_extremas.png

# Backup: Sieve Bootstrap Interacciones

Estacionario\_Lineal/Sieve Bootstrap/analisis\_interacci

No\_Lineal\_Estacionario/AV-MCPS/analisis\_distribucion.p

# Backup: Matriz de Comparaciones Significativas

Comparación	DM Stat	p-value	Sig.	Ganador
Block vs AREPD	8.45	<0.001	Sí	Block Bootstrap
Block vs EnCQR	7.23	<0.001	Sí	Block Bootstrap
Block vs MCPS	6.89	<0.001	Sí	Block Bootstrap
Sieve vs AREPD	7.92	<0.001	Sí	Sieve Bootstrap
Sieve vs LSPMW	6.54	<0.001	Sí	Sieve Bootstrap
LSPM vs AREPD	5.21	<0.001	Sí	LSPM
Block vs Sieve	2.34	0.019	No*	Empate
DeepAR vs MCPS	1.89	0.059	No	Empate

\* No significativo con corrección de Bonferroni ( $\alpha = 0,001389$ )

# Backup: Distribución de Errores por Modelo

## Modelos de Bajo Error:

- Block Bootstrap: IQR [0.25, 0.70]
- Sieve Bootstrap: IQR [0.26, 0.72]
- DeepAR: IQR [0.27, 1.15]

## Modelos Estables:

- Menor Std: Block (0.289)
- Mayor Std: AREPD (9.139)
- Razón: 31.6x diferencia

## Outliers Críticos:

- AREPD: 122.83 (máximo global)
- MCPS: 49.65
- DeepAR: 39.47
- EnCQR-LSTM: 35.65

## Asimetría:

- Todos sesgados positivamente
- Mediana  $\neq$  Media (todos)
- Cola derecha pesada

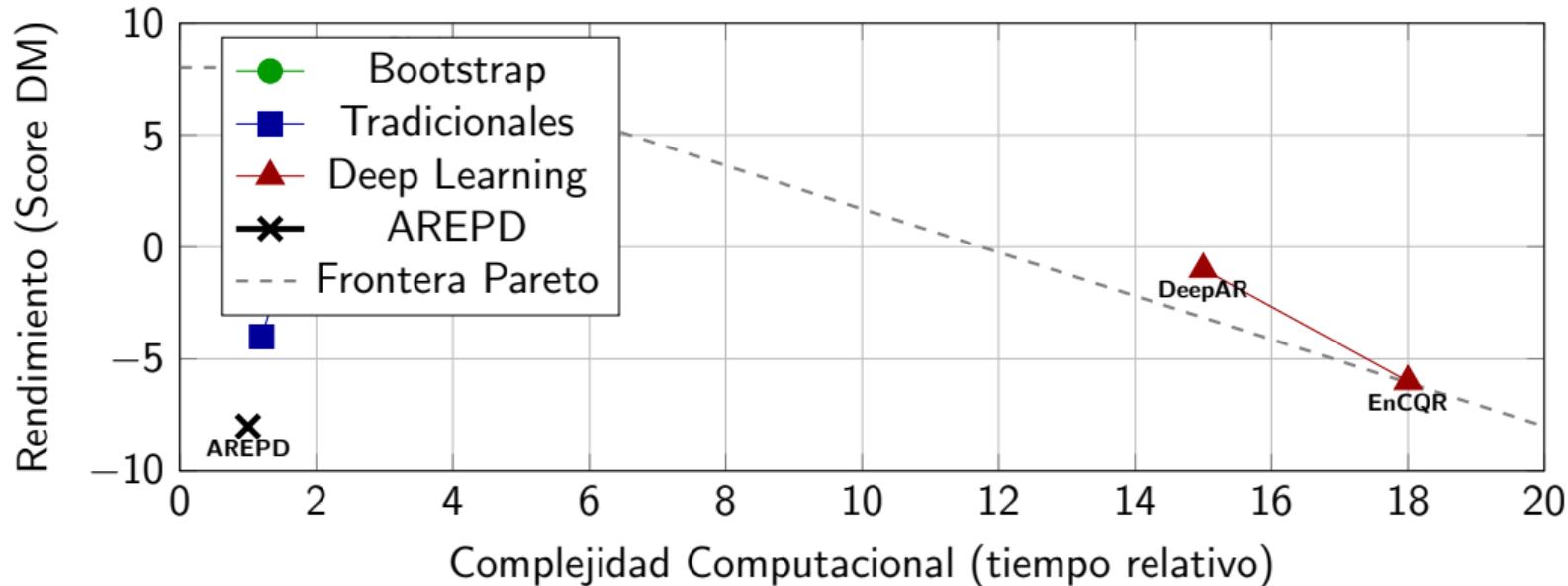
# Backup: Tiempo Computacional (Estimado)

Modelo	Tiempo Relativo	Complejidad
AREPD	1.0x	$O(n)$
MCPS	1.2x	$O(n \log n)$
AV-MCPS	1.5x	$O(n \log n)$
LSPM	2.0x	$O(n^2)$
LSPMW	2.1x	$O(n^2)$
Block Bootstrap	3.5x	$O(B \cdot n)$
Sieve Bootstrap	4.0x	$O(B \cdot n \log n)$
DeepAR	15.0x	$O(\text{epochs} \cdot \text{batch} \cdot n)$
EnCQR-LSTM	18.0x	$O(\text{epochs} \cdot \text{batch} \cdot n)$

B: número de muestras bootstrap (típicamente 1000)

n: tamaño de la serie temporal

# Backup: Trade-off Precisión-Complejidad



# Backup: Criterios de Selección de Modelo

## Priorizar Precisión

Si el error de predicción es crítico y el tiempo de cómputo no es limitante:

**Block Bootstrapping** ¿ Sieve Bootstrap ¿ LSPM

## Priorizar Velocidad

Si se requiere predicción en tiempo real:

**MCPS** ¿ AV-MCPS ¿ LSPM

(Aceptar pérdida de 10-15 % en precisión)

## Balance Óptimo

Mejor compromiso precisión-velocidad:

**Block Bootstrapping** con  $B = 500$  iteraciones

## ① Datos Simulados:

- Validación en datos reales pendiente
- Posibles estructuras no capturadas

## ② Hiperparámetros Fijos:

- No se optimizaron por modelo
- Posible subestimación de deep learning

## ③ Métrica Única:

- Solo se evaluó error puntual
- No se midió calibración de intervalos

## ④ Horizonte Limitado:

- Máximo 5 pasos adelante
- Largo plazo ( $\geq 10$  pasos) no evaluado

# Backup: Extensiones Posibles

## Metodológicas:

- Test de Hansen-Lunde
- Model Confidence Set (MCS)
- Cross-validation temporal
- Análisis de robustez

## Modelos Adicionales:

- Transformer models
- N-BEATS
- Prophet
- GARCH-based

## Escenarios:

- Series multivariadas
- Cambios estructurales
- Estacionalidad múltiple
- Valores faltantes

## Aplicaciones:

- Finanzas (VaR, portfolio)
- Energía (demanda eléctrica)
- Retail (ventas)
- Clima (temperatura)

# Backup: Referencias Metodológicas

-  Diebold, F. X., & Mariano, R. S. (1995). *Comparing predictive accuracy*. Journal of Business & Economic Statistics, 13(3), 253-263.
-  Harvey, D., Leybourne, S., & Newbold, P. (1997). *Testing the equality of prediction mean squared errors*. International Journal of Forecasting, 13(2), 281-291.
-  Dunn, O. J. (1961). *Multiple comparisons among means*. Journal of the American Statistical Association, 56(293), 52-64.
-  Efron, B., & Tibshirani, R. J. (1994). *An introduction to the bootstrap*. CRC Press.
-  Salinas, D., et al. (2020). *DeepAR: Probabilistic forecasting with autoregressive recurrent networks*. International Journal of Forecasting, 36(3), 1181-1191.

# Backup: Software Utilizado

## Análisis Estadístico:

- Python 3.10+
- NumPy 1.24
- Pandas 2.0
- SciPy 1.11
- Statsmodels 0.14

## Visualización:

- Matplotlib 3.7
- Seaborn 0.12
- Plotly (interactivo)

## Machine Learning:

- PyTorch 2.0 (DeepAR)
- TensorFlow 2.13 (EnCQR)
- Scikit-learn 1.3

## Presentación:

- L<sup>A</sup>T<sub>E</sub>X (Beamer)
- TikZ / PGFPlots

## Repositorio:

- GitHub (código abierto)
- Documentación completa

# Fin de las Diapositivas de Respaldo

¿Preguntas Adicionales?

*Todas las figuras y análisis detallados están disponibles en el repositorio del proyecto*