

Examen Parcial III

César Gamboa Sanabria - Stefany Matarrita Muñoz¹

Preguntas teóricas

1.

La estadística paramétrica hace supuestos sobre las distribuciones de las variables, es decir parte del supuesto que el comportamiento de las variables sigue una función de probabilidad en particular. Si se asume que la función de activación es una función probabilística, por definición las redes neuronales son una técnica paramétrica.

2.

- Cantidad de entradas y salidas: Los nodos de entrada se refieren a las variables que le dan la información a la red, mientras que los nodos de salida se refieren a las estimaciones realizadas por la red.
- Cantidad de capas ocultas: las capas ocultas son las que realizan el procesamiento de la información mediante el aprendizaje de los nodos de entrada, para luego enviarlas a los nodos de salida.
- Función de red/propagación asociada a cada nodo: Esta función se encarga de ponderar las entradas con los pesos sinápticos, para ello existen diferentes funciones como lo es la función Base Lineal o radial.
- Función de activación de cada neurona: esta función se encarga de recibir como insumo los valores de la función red, para así, calcular el estado de activación de la neuronal, y propagar la información. Entre las funciones de activación se encuentra la Sigmoide, Tangente Hipérbolica, Identidad. . .
- Interconexión: se refiere a la forma en que se producen las conexiones entre las diferentes capas de neuronas, por ejemplo puede ser que las salidas de los nodos de otra, o solo con un grupo de nodos.
- La dirección que sigue la información: es la forma en la que fluye la información dentro de la red, puede ser que siempre sea hacia adelante, finalizando en la capa de salida o hacia atrás (Redes de alimentación hacia adelante o atrás respectivamente), inclusive puede que la información se mueva entre neuronas de la misma capa (Redes de alimentación lateral).
- Datos de entrenamiento y validación: el conjunto de datos de entrenamiento es con el que se ajustan los pesos sinápticos para capturar la información que se presenta (aprendizaje), mientras que el conjunto de validación verifica si realmente el modelo entrega resultados aceptables al presentar los patrones de los datos.

3.

La función de agregación o función de red de cada nodo, es la que agrega las entradas de las neuronas y los pesos sinápticos en un solo valor. Una función de red es la Función Base Lineal (FBL), cuya representación matemática para una neurona j con n entradas, viene dada por:

$$f(x, W_j) = \sum_{i=1}^n x_i w_{ij} \quad (1)$$

donde, X es el vector de entradas y W_j es el vector de pesos sinápticos e i es el índice que identifica a cada entrada.

4.

Es la función que define la salida de la neurona dado las entradas que recibió, la salida se usa como entrada para el siguiente nodo, y así sucesivamente hasta que se encuentre una solución. Los valores de salida están en el rango de 0 a 1 sin embargo esto depende de cuál función de activación se seleccione.

$$g(x) = g(f(x, W_j)) \quad (2)$$

5.

La contribución de una entrada de la neurona depende en gran medida de su variabilidad relativa respecto a otras entradas, a mayor variabilidad mayor importancia. Entonces si una de las entradas tiene un rango de 0 a 1, mientras que la otra tiene un rango de 0 a 10000, entonces la contribución de la primera va a ser opacada por la segunda, a pesar de que tenga mayor variabilidad. Por ello es esencial re escalar las entradas, para que la variabilidad sea un reflejo de la importancia de la variable y así poder capturar la información relevante de las entradas.

El vector gradiente evaluado en un punto x del dominio de una función $f(x)$ indica la dirección en la cual el campo f varía más rápidamente y su módulo representa el ritmo de variación de f en la dirección de dicho vector gradiente. Dada la definición anterior, si se toma f como la función red, lo que se busca es encontrar el punto donde f varía más, bajo este objetivo, si se reescala se logrará más rápido que la función encuentre el campo vectorial donde tendrá mayor variabilidad, pues solo se estará tomando la información relevante. Se puede concluir entonces que normalizar hace que el entrenamiento sea más rápido y reduce las posibilidades de estancarse en un óptimo local.

Adecuación de la red

Modelo	MAE	MAPE	RMSE	MASE
T=50	0.8409	0.8444	1.1005	0.6584
T=50 Validacion	0.9998	1.01	1.2839	0.7827
T=100	0.8192	0.8219	1.0958	0.6264
T=100 Validacion	1.044	1.031	1.2667	0.7982
T=200	0.8272	0.8283	1.0692	0.7167
T=200 Validacion	0.8231	0.8316	1.0635	0.7132
T=500	0.8097	0.8116	1.0239	0.7245
T=500 Validacion	0.6166	0.6183	0.8078	0.5517
T=1000	0.7712	0.7726	0.963	0.7072
T=1000 Validacion	0.6949	0.6931	0.7837	0.6372

Al simular el proceso, las medidas de rendimiento son mejores con $T=1000$ para el set de entrenamiento, mientras que en el set de validación dichas medidas son mejores con $T=500$ con excepción del RMSE, que es mejor cuando $T=1000$. En términos generales las medidas de rendimiento parecen disminuir conforme se tienen más periodos.

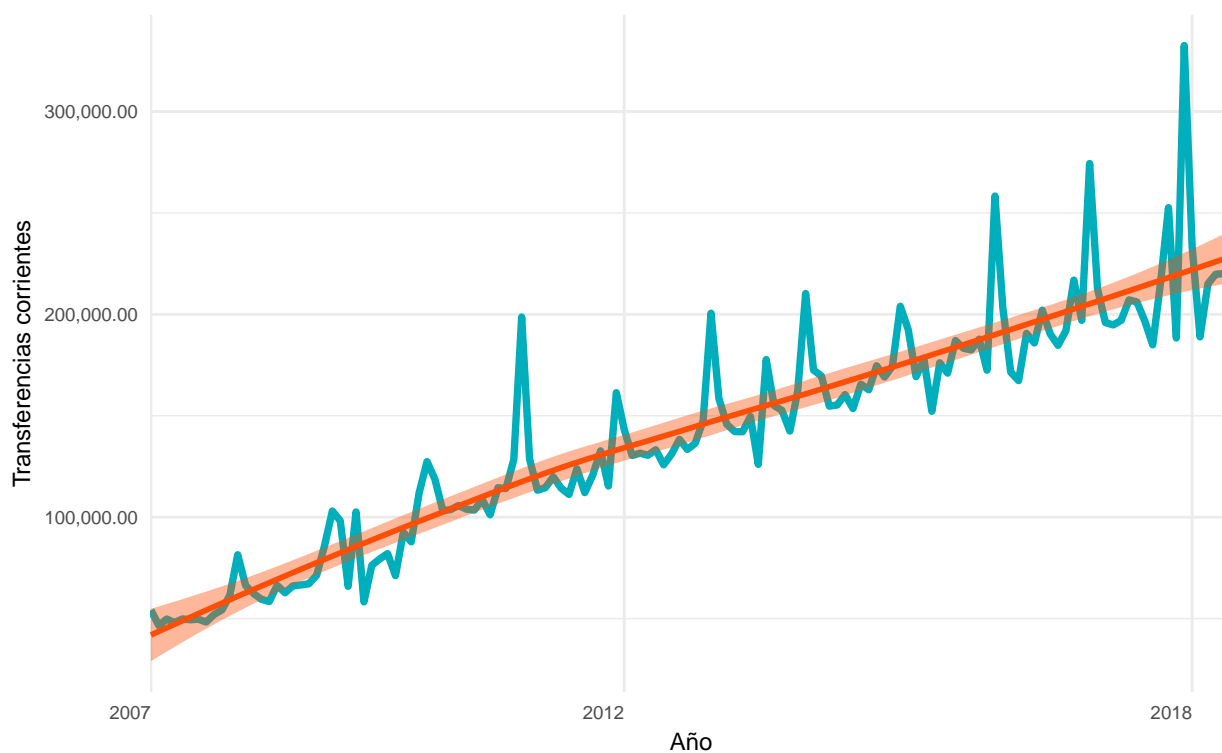
Preguntas prácticas

a

Son erogaciones para financiar fundamentalmente gasto corriente, destinados a personas, entes y órganos del sector público, privado y externo, con el fin de satisfacer necesidades públicas de diversa índole, sin que exista una contraprestación de bienes, servicios o derechos a favor de quien transfiere los recursos.

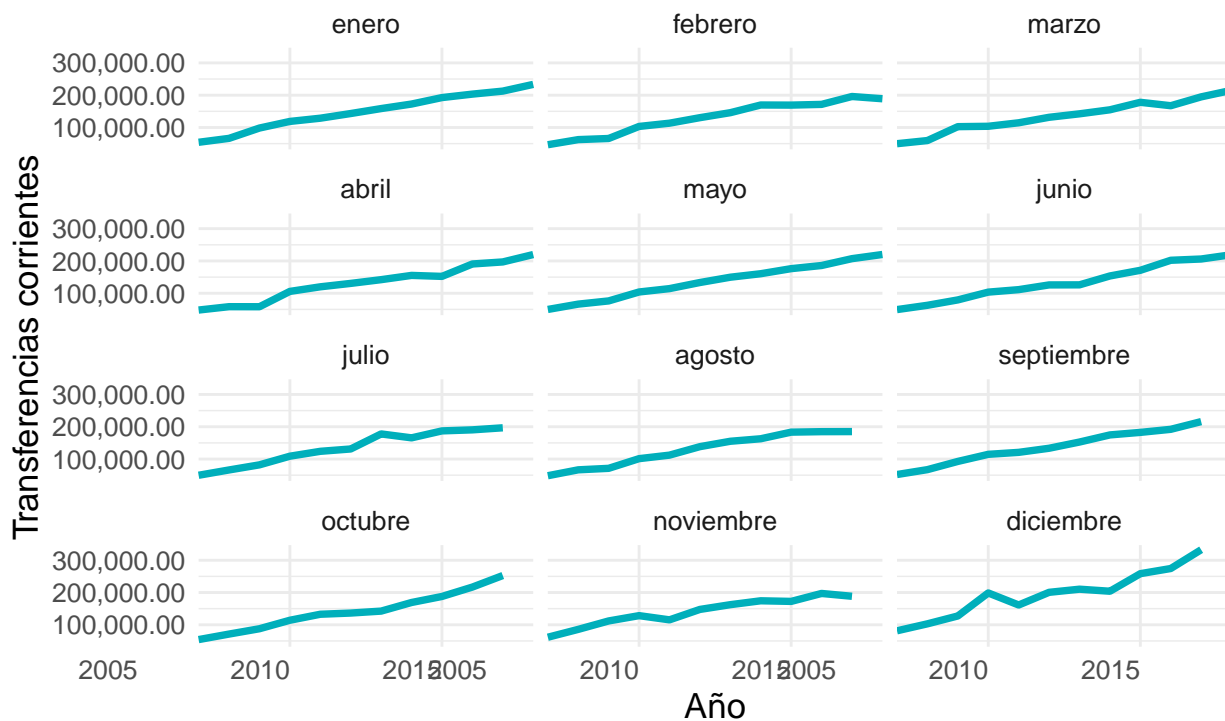
b

Gráfico 1: Transferencias corrientes en el periodo 2007 – 2018



En el gráfico 1, al hacer un suavizamiento Loess puede observarse la tendencia creciente a lo largo de todo el periodo

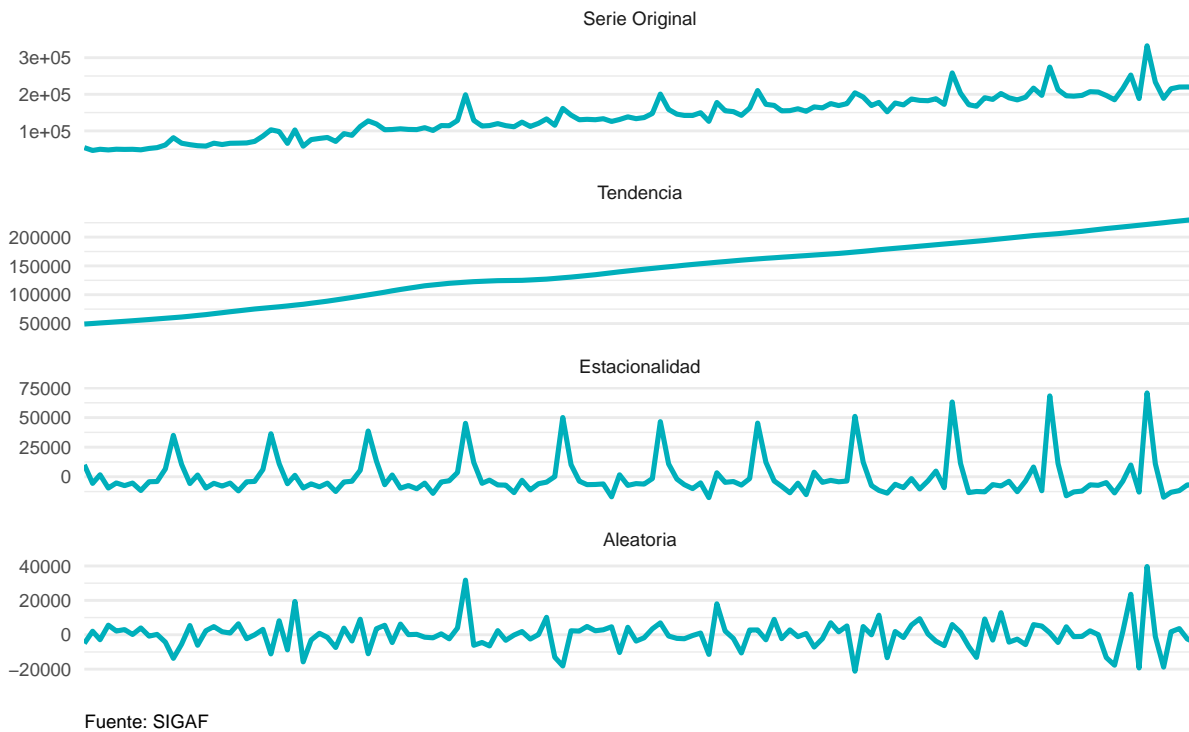
Gráfico 2: Transferencias corrientes en el periodo 2007 – 2017



Fuente: SIGAF

El gráfico 2 muestra cómo hay un crecimiento sostenido de las transferencias corrientes durante todo el periodo, este aumento se da de una manera casi constante. Los cambios más notables se dan en Noviembre y Diciembre, especialmente en este último mes, pues incluso la tasa de cambio es más elevada.

Gráfico 3: Descomposición de la serie de transacciones corrientes en el periodo 2007 – 2018



En el gráfico 3 se muestra la descomposición de la serie en sus distintos componentes. Pueden observarse, además de un crecimiento a lo largo del tiempo, los picos y las caídas en la parte estacional. El componente aleatorio muestra indicios de que la variabilidad de la serie no es homogénea, sino que cambia conforme pasa el tiempo, pues durante un tiempo se mantuvo relativamente estable pero luego presenta algunos cambios.

c.

Modelo	AIC	AICc	BIC	MAE	MAPE	RMSE	MASE
ARIMA(4,1,2)(3,1,3)[12]	2202.62	2206.81	2236.62	6749.76	5.27	10079.53	0.38
ARIMA(4,1,2)(3,1,3)[12] Validacion				11505.15	4.92	19332.37	0.65
ARIMA(4,1,0)(4,1,0)[12]	2197.6	2199.58	2221.14	7162.3	5.52	10209.75	0.4
ARIMA(4,1,0)(4,1,0)[12] Validacion				11795.31	5.17	18516.66	0.66
ARIMA(1,1,1)(1,1,0)[12]	2199.58	2199.99	2210.04	7670.97	5.91	11424.42	0.43
ARIMA(1,1,1)(1,1,0)[12] Validacion				11807.03	5.23	17187.26	0.66
ARIMA(0,1,1)(1,1,0)[12]	2201.24	2201.48	2209.08	7813.45	6	11628.76	0.44
ARIMA(0,1,1)(1,1,0)[12] Validacion				11973.34	5.29	17315.36	0.67

El cuadro anterior muestra los 4 mejores modelos ARIMA, siendo el ARIMA(4,1,2)(3,1,3) el mejor de todos. Ahora, considerando un arima con intervención se tiene:

Gráfico 4: Intervenciones sugeridas



Fuente: SIGAF



Fuente: SIGAF

Modelo	AIC	AICc	BIC	MAE	MAPE	RMSE	MASE
ARIMA(3,1,0)(4,1,0)[12]	2207.27	2210.81	2238.65	7271.63	5.6	10543.48	0.41
ARIMA(3,1,0)(4,1,0)[12] Validacion				11983.18	5.33	18407.49	0.67
ARIMA(4,1,0)(0,1,1)[12]	2203.41	2205.85	2229.56	7368.65	5.72	10574.35	0.41
ARIMA(4,1,0)(0,1,1)[12] Validacion				12585.31	5.4	20229.82	0.71
ARIMA(1,1,1)(1,1,0)[12]	2206.43	2207.99	2227.35	7714.27	5.96	11361.38	0.43
ARIMA(1,1,1)(1,1,0)[12] Validacion				12749.13	5.58	18032.63	0.72
ARIMA(3,1,0)(0,1,0)[12]	2240.85	2242.41	2261.77	9743.24	7.46	13754.5	0.55
ARIMA(3,1,0)(0,1,0)[12] Validacion				12950.19	6.08	16718.11	0.73

Al incorporar las intervenciones, el mejor modelo es un ARIMA(3,1,0)(4,1,0), el cual es ligeramente superior al modelo sin intervenciones.

d.

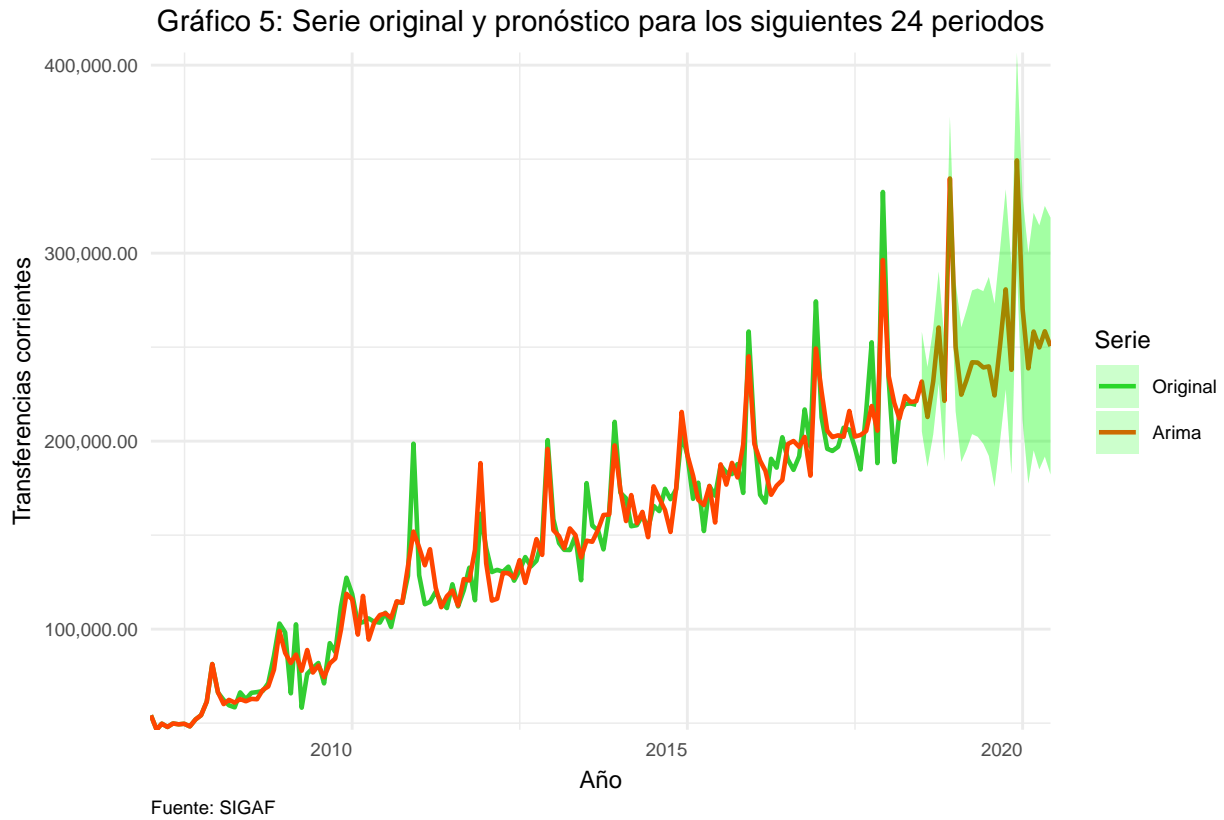
Modelo	AIC	AICc	BIC	MAE	MAPE	RMSE	MASE
NNAR(2,1,2)[12]	0	0	0	8656.7021	6.7029	12487.2732	0.486
NNAR(2,1,2)[12] Validacion				15423.8145	6.6621	22247.2702	0.8659

e.

Modelo	AIC	AICc	BIC	MAE	MAPE	RMSE	MASE
ARIMA(3,1,0)(4,1,0)[12]	2207.27	2210.81	2238.65	7271.63	5.6	10543.48	0.41
ARIMA(3,1,0)(4,1,0)[12] Validacion				11983.18	5.33	18407.49	0.67
NNAR(2,1,2)[12]	0	0	0	8656.7021	6.7029	12487.2732	0.486
NNAR(2,1,2)[12] Validacion				15423.8145	6.6621	22247.2702	0.8659

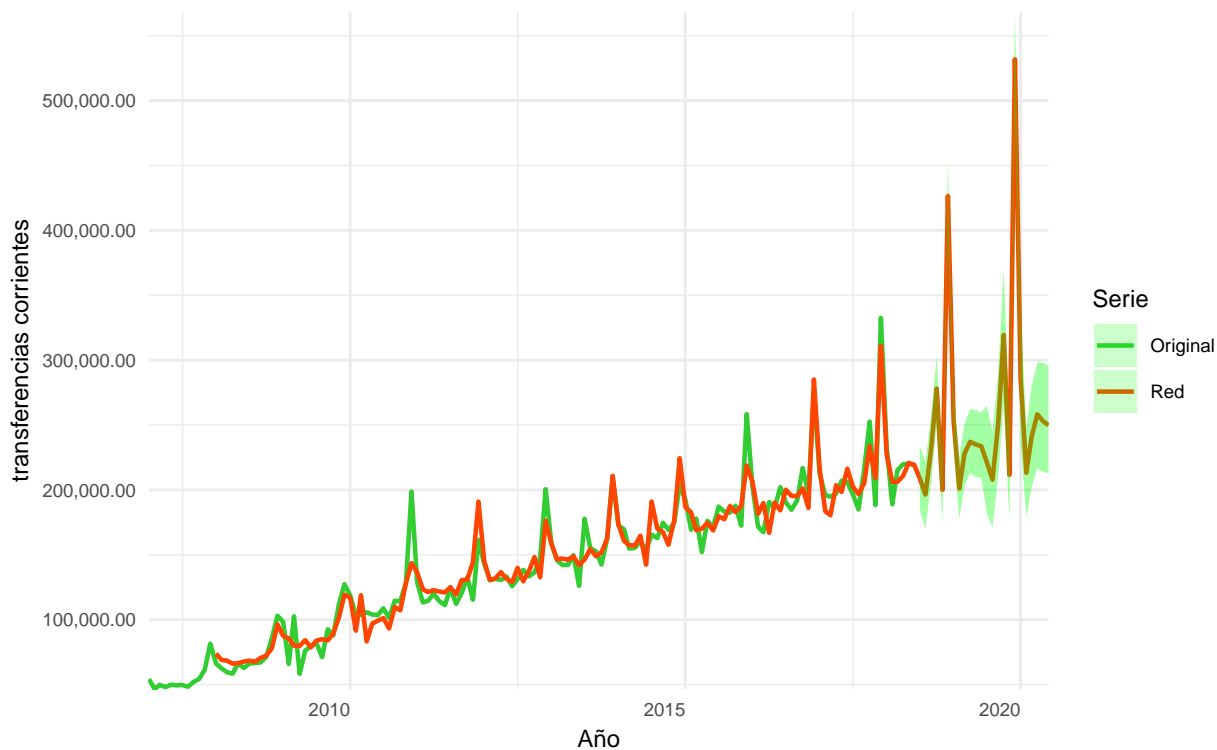
Al comparar las medidas d erendimiento para ambos modelos, los mejores resultados los ofrece el ARIMA con intervención escalonada, tanto en el set de entrenamiento como en el de validación.

f.



	Pronostico	LI	LS
Jul 2018	231705.8	205182.0	258229.6
Ago 2018	212857.6	186276.1	239439.2
Sep 2018	231809.0	203497.9	260120.1
Oct 2018	260464.3	230657.8	290270.8
Nov 2018	221437.9	189386.3	253489.4
Dic 2018	339753.3	306816.4	372690.3
Ene 2019	250333.5	215849.1	284817.9
Feb 2019	224695.0	188926.6	260463.4
Mar 2019	232370.0	195286.0	269454.0
Abr 2019	241965.1	203751.5	280178.6
May 2019	241802.2	202357.6	281246.8
Jun 2019	239195.0	198621.4	279768.7
Jul 2019	239749.8	192149.1	287350.5
Ago 2019	224314.5	175563.5	273065.5
Sep 2019	251903.1	200828.3	302977.9
Oct 2019	280740.6	227531.0	333950.2
Nov 2019	237936.6	182265.7	293607.5
Dic 2019	349345.6	291957.3	406733.9
Ene 2020	270365.7	210930.2	329801.2
Feb 2020	238707.5	177399.1	300016.0
Mar 2020	258311.9	195143.7	321480.1
Abr 2020	249849.4	184951.5	314747.4
May 2020	258460.9	191805.2	325116.5
Jun 2020	250554.7	182221.8	318887.6

Gráfico 6: Serie original y pronóstico para los siguientes 24 periodos



	Pronostico	LI	LS
Jul 2018	206217.6	180454.7	230535.6
Ago 2018	194320.2	167230.0	218553.5
Sep 2018	231019.3	205412.2	255582.7
Oct 2018	276216.2	251284.4	301403.1
Nov 2018	196890.1	171952.7	220535.8
Dic 2018	453303.9	420406.2	481039.2
Ene 2019	251648.2	226915.1	277404.6
Feb 2019	202556.8	178854.0	226693.8
Mar 2019	222969.8	199976.8	249297.0
Abr 2019	236069.2	210243.5	258775.6
May 2019	233176.6	208548.7	259106.3
Jun 2019	229937.8	205557.0	253602.1
Jul 2019	214675.7	179994.3	252119.0
Ago 2019	202516.5	168433.9	239703.6
Sep 2019	249029.0	209473.7	290124.7
Oct 2019	317062.3	261131.6	375357.1
Nov 2019	204350.6	172355.0	238718.0
Dic 2019	526620.8	471444.6	584038.3
Ene 2020	284492.1	236858.9	354258.4
Feb 2020	213817.4	182409.9	245809.2
Mar 2020	227029.1	191942.1	271222.4
Abr 2020	256081.7	213542.5	297640.7
May 2020	249927.2	204861.5	295491.0
Jun 2020	239646.9	201712.4	285467.0

Si bien el modelo de redes neuronales ofrece un intervalo de confianza más estrecho para el pronóstico, es el ARIMA con intervención el que ofrece resultados más adecuados en base a los datos de la serie, pues logra captar de mejor manera los picos estacionales y no sobreestima los valores de éstos en el pronóstico.

g.

Modelo	AIC	AICc	BIC	MAE	MAPE	RMSE	MASE
ARIMA(2,1,0)(1,1,0)[12]	1231.55	1232.33	1239.65	7265.46	6.82	11610.05	0.39
ARIMA(2,1,0)(1,1,0)[12] Validacion				6852.93	4.38	9493.5	0.37
ARIMA(1,1,0)(1,1,0)[12]	1233.83	1234.29	1239.9	7626.74	7.17	12030.9	0.41
ARIMA(1,1,0)(1,1,0)[12] Validacion				7233.35	4.62	9763.71	0.39
ARIMA(0,1,1)(1,1,0)[12]	1226.53	1226.99	1232.6	6812.13	6.41	11352.1	0.37
ARIMA(0,1,1)(1,1,0)[12] Validacion				7240.76	4.55	9830.19	0.39
ARIMA(3,1,0)(0,1,1)[12]	1222.41	1223.61	1232.54	5428.62	5.14	9032.2	0.29
ARIMA(3,1,0)(0,1,1)[12] Validacion				7298.65	4.62	9841.76	0.39

El cuadro anterior muestra los 4 mejores modelos ARIMA, siendo el ARIMA(2,1,0)(1,1,0) el mejor de todos. Ahora, considerando un arima con intervención se tiene:

Modelo	AIC	AICc	BIC	MAE	MAPE	RMSE	MASE
ARIMA(1,1,0)(1,1,0)[12]	1237.76	1238.96	1247.88	7619.12	7.16	12017.59	0.41
ARIMA(1,1,0)(1,1,0)[12] Validacion				9464.26	5	14010.87	0.51
ARIMA(0,1,0)(2,1,0)[12]	1260.82	1262.02	1270.95	8430.95	8.15	14657.34	0.45
ARIMA(0,1,0)(2,1,0)[12] Validacion				10255.76	5.1	15878.25	0.55
ARIMA(0,1,1)(0,1,1)[12]	1222.24	1223.44	1232.36	5320.64	5.06	8982	0.29
ARIMA(0,1,1)(0,1,1)[12] Validacion				10477.89	5.33	16239.15	0.56
ARIMA(2,1,3)(2,1,0)[12]	1234.26	1239.15	1254.51	6532.67	6.15	10537.15	0.35
ARIMA(2,1,3)(2,1,0)[12] Validacion				10670.4	5.45	15224.58	0.57

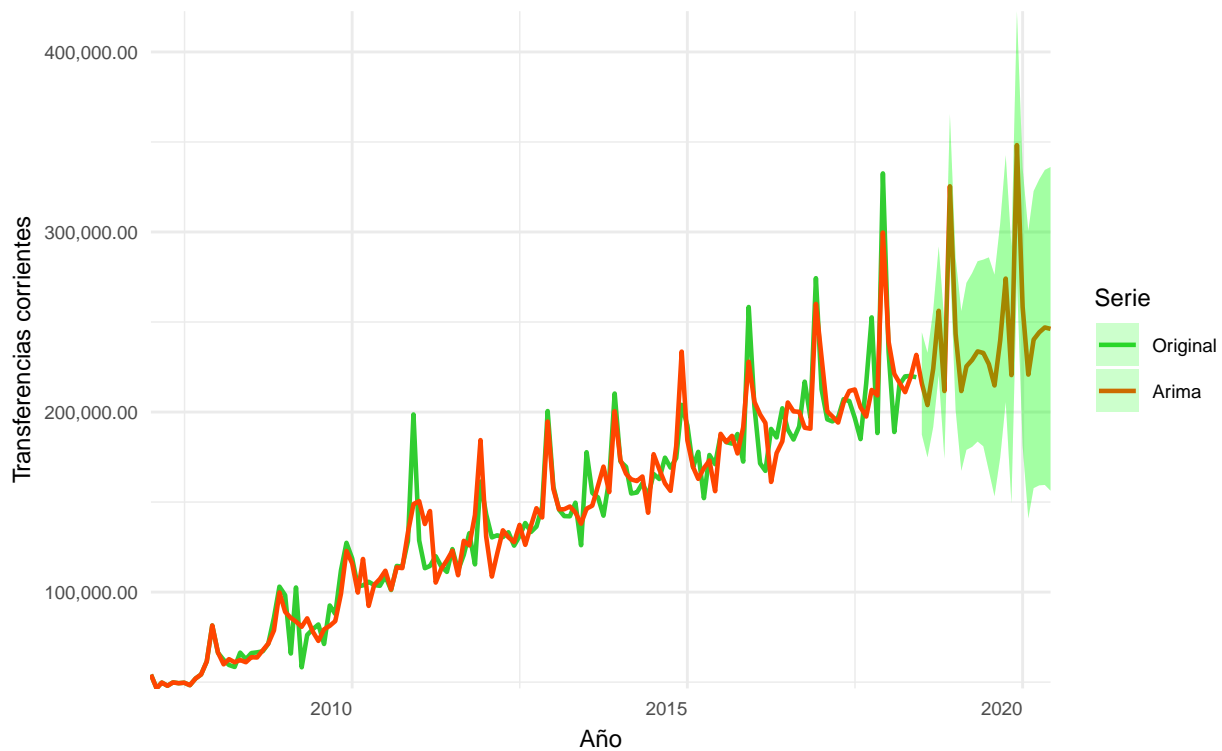
Al incorporar las intervenciones, el mejor modelo es un $ARIMA(3,1,0)(4,1,0)$, el cual es inferior al $ARIMA$ sin intervenciones.

Modelo	AIC	AICc	BIC	MAE	MAPE	RMSE	MASE
NNAR(2,1,2)[12]	0	0	0	6637.6311	6.7201	9201.3872	0.3559
NNAR(2,1,2)[12] Validacion				21474.666	13.1651	24752.4614	1.1514

Modelo	AIC	AICc	BIC	MAE	MAPE	RMSE	MASE
ARIMA(2,1,0)(1,1,0)[12]	1231.55	1232.33	1239.65	7265.46	6.82	11610.05	0.39
ARIMA(2,1,0)(1,1,0)[12] Validacion				6852.93	4.38	9493.5	0.37
NNAR(2,1,2)[12]	0	0	0	6637.6311	6.7201	9201.3872	0.3559
NNAR(2,1,2)[12] Validacion				21474.666	13.1651	24752.4614	1.1514

Los mejores resultados los ofrece el $ARIMA$ sin intervención.

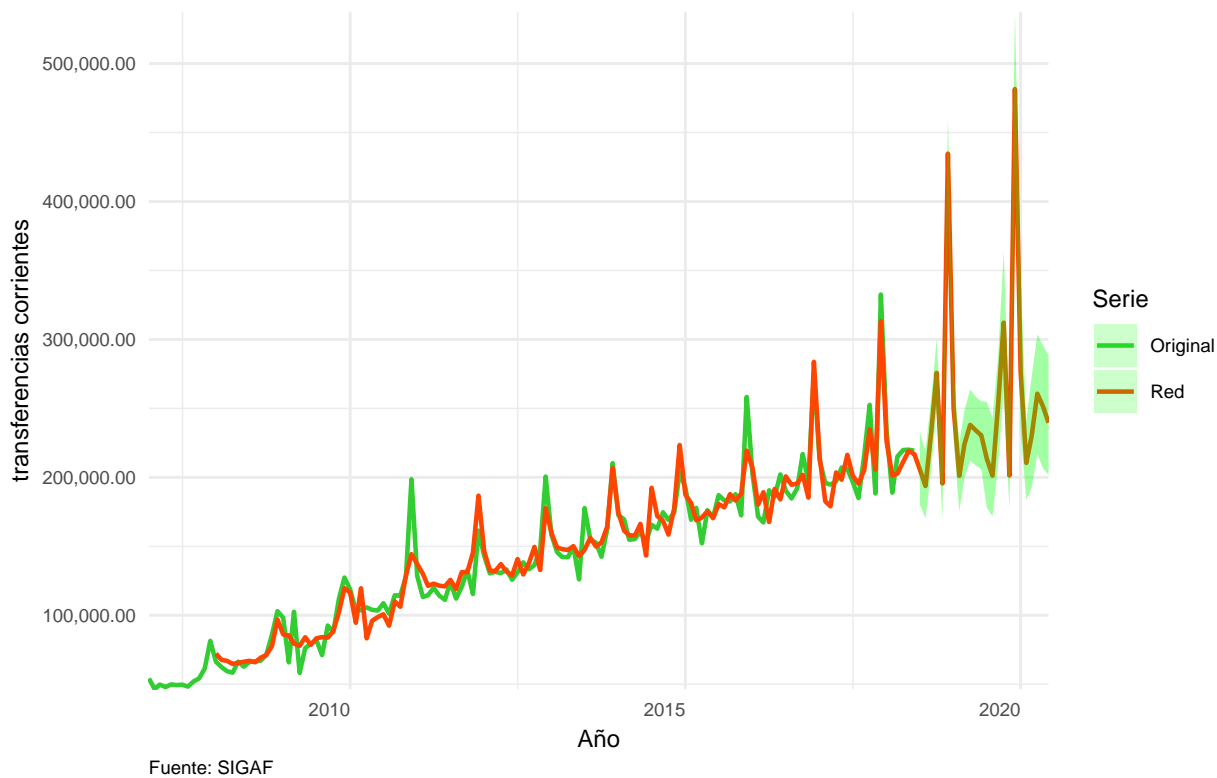
Gráfico 7: Serie original y pronóstico para los siguientes 24 periodos



Fuente: SIGAF

	Pronostico	LI	LS
Jul 2018	215899.8	187433.5	244366.0
Ago 2018	203868.0	174815.8	232920.2
Sep 2018	224043.7	191598.7	256488.6
Oct 2018	256243.8	220478.0	292009.5
Nov 2018	211654.2	174018.6	249289.8
Dic 2018	325411.7	285196.9	365626.5
Ene 2019	243461.4	201098.9	285823.9
Feb 2019	211649.6	167282.5	256016.7
Mar 2019	225370.0	178973.4	271766.5
Abr 2019	228907.0	180639.6	277174.3
May 2019	233680.9	183590.5	283771.3
Jun 2019	232813.6	180958.7	284668.4
Jul 2019	226634.0	167344.9	285923.2
Ago 2019	214748.8	153198.0	276299.6
Sep 2019	239847.1	174750.6	304943.6
Oct 2019	274136.1	205576.2	342696.0
Nov 2019	220507.4	149168.2	291846.5
Dic 2019	348269.0	273891.5	422646.6
Ene 2020	258402.5	181229.8	335575.2
Feb 2020	220761.6	140909.7	300613.5
Mar 2020	240228.9	157732.8	322725.0
Abr 2020	244305.4	159281.9	329329.0
May 2020	246984.1	159495.5	334472.6
Jun 2020	246182.7	156294.1	336071.3

Gráfico 8: Serie original y pronóstico para los siguientes 24 periodos



	Pronostico	LI	LS
Jul 2018	208151.1	184716.0	231586.2
Ago 2018	196295.8	170944.8	220910.8
Sep 2018	229664.1	203984.9	253989.7
Oct 2018	279826.2	254047.4	306130.9
Nov 2018	200338.5	174915.1	223850.8
Dic 2018	464661.9	438153.0	488323.6
Ene 2019	262945.8	235829.2	288485.6
Feb 2019	200084.0	175071.7	223933.3
Mar 2019	226522.4	202071.5	251777.1
Abr 2019	234898.8	210788.0	257741.7
May 2019	234022.4	209927.5	257505.9
Jun 2019	232723.1	207600.9	258274.5
Jul 2019	219864.6	185469.2	258185.9
Ago 2019	207229.1	173510.2	245423.0
Sep 2019	246142.0	206067.3	291737.8
Oct 2019	336176.3	280195.4	405259.9
Nov 2019	213982.6	181732.6	251580.4
Dic 2019	577382.2	517929.0	620146.2
Ene 2020	335473.3	272149.1	406127.6
Feb 2020	215040.4	179865.9	252494.5
Mar 2020	237251.2	197238.0	279896.2
Abr 2020	253699.8	217269.4	297446.3
May 2020	251582.7	211835.5	298879.6
Jun 2020	248869.9	210722.2	297802.9

Al tomar la mitad de los datos para el conjunto de entrenamiento y la otra mitad para validación, nuevamente las redes neuronales ofrecen un intervalo de confianza más estrecho, si embargo el problema de la sobreestimación en el pronóstico persiste. El ARIMA(3,1,0)(4,1,0), que es sin intervención, ofrece un mejor ajuste a los datos históricos y brinda un pronóstico más acorde a lo observado en toda la serie.

h.

El ajuste hecho a la base de entrenamiento muestra que la tendencia de las transacciones corrientes han ido aumentando con el tiempo, mismo comportamiento que se espera al realizar el pronóstico. Dada la situación fiscal del país, una intervención en este tipo de transferencias podría reducir aún más el gasto y destinar esas cantidades a necesidades más específicas y requeridas en la actualidad; aunque no es posible reducir las transferencias corrientes a su mínima expresión, sí es posible intervenirla para tener un mayor control fiscal en el país.