

Egresos del Sector Público

Series Cronológicas

Ignacio Campón

2024-07-05

Introduccion

Los **Egresos Primarios del Sector Público no Monetario**, son una fuente crucial de información proporcionada por el [Ministerio de Economía y Finanzas](#). Esta serie mensual, ofrece una panorámica detallada de los desembolsos financieros del sector público, a lo largo del tiempo.

Los egresos primarios del sector público son fundamentales para entender la dinámica fiscal de un país, ya que representan los gastos esenciales en bienes y servicios no financieros realizados por el gobierno. Estos datos son vitales para analizar la gestión presupuestaria, evaluar políticas económicas y anticipar tendencias futuras en base a modelos estadísticos.

Se explorará la serie temporal con el objetivo de identificar un modelo adecuado que permita reproducirla, facilitando así la capacidad predictiva sobre los egresos del sector público en Uruguay.

Serie temporal: Gasto Público

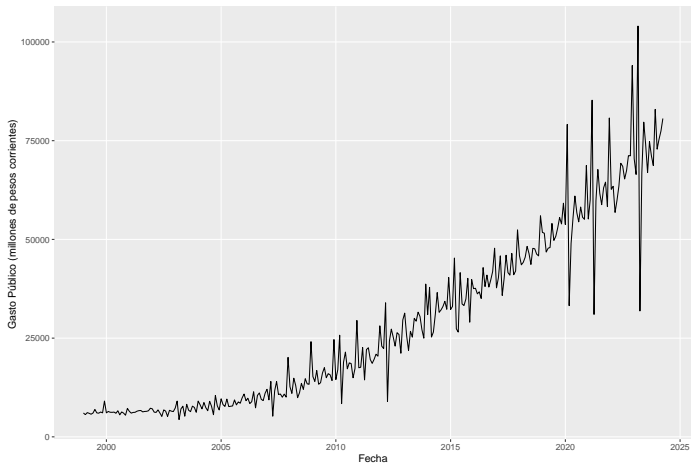


Figura 1: Evolución del Gasto Público mensual en millones de pesos corrientes entre enero 1999 y abril 2024.

Análisis descriptivo

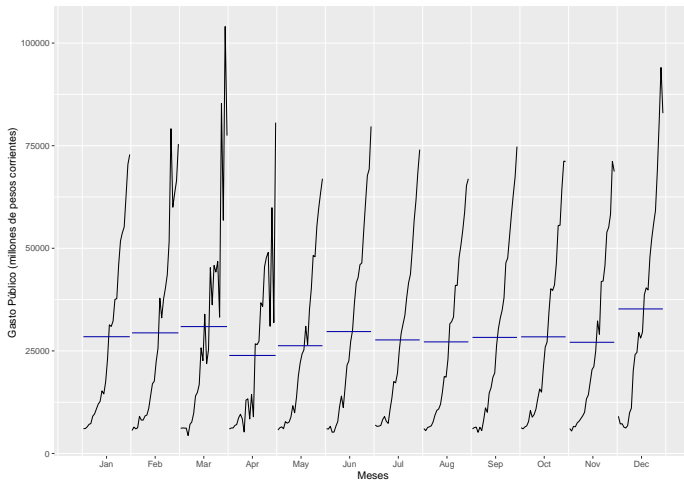


Figura 2: Comportamiento mensual del Gasto Público entre 1999 y 2024.

Análisis descriptivo (2)

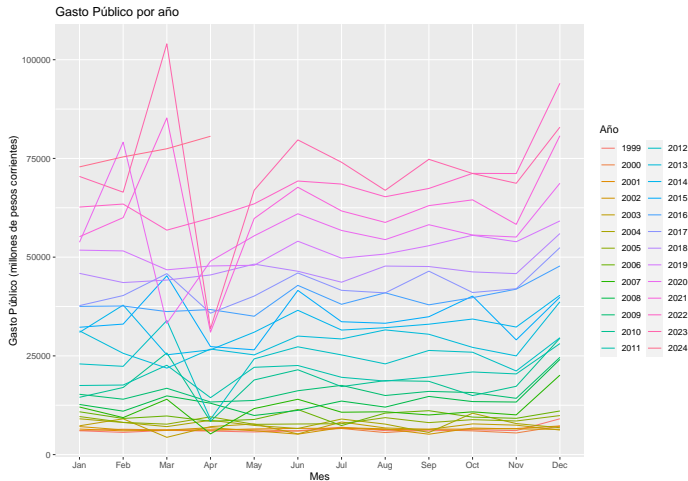


Figura 3: Comportamiento por año del Gasto Público entre 1999 y 2024.

Metodología

El desarrollo esta basado en la metodología de Box y Jenkins, en la construcción de modelos ARIMA para el análisis de series de tiempo. Podemos dividir esta metodología en cuatro grandes etapas:

- ▶ identificación
- ▶ estimación
- ▶ diagnostico
- ▶ predicción

Metodología (2)

La identificación del modelo, esta basada en las funciones de autocorrelación y autocorrelación parcial y a partir de ello, podremos obtener diferentes modelos candidatos. Tras la estimación, un diagnostico con determinadas etapas es realizado; y por último, aplicaremos predicciones sobre los mismos y se evaluará su desempeño.

Logaritmo de la serie

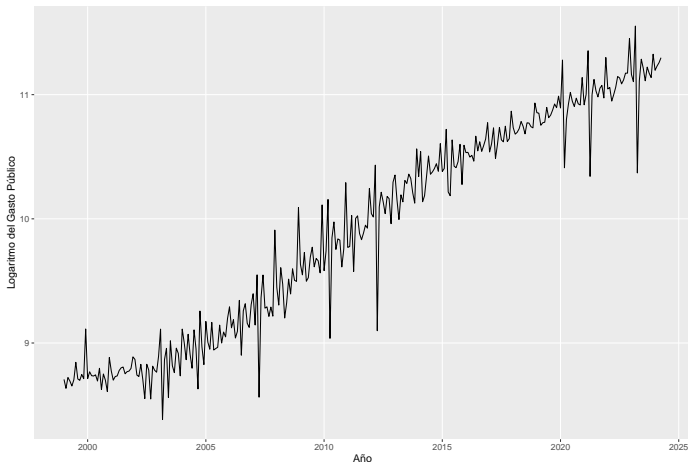


Figura 4: Evolución del logaritmo del Gasto Público mensual entre 1999 y 2024.

Funciones de Autocorrelación

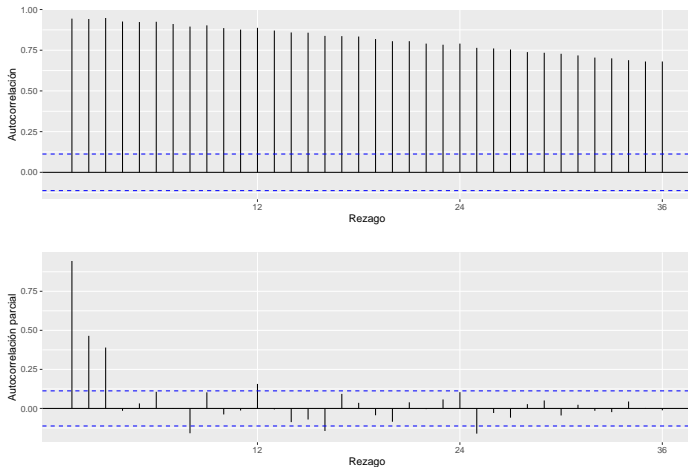


Figura 5: Funciones de Autocorrelación y Autocorrelación Parcial estimadas del logaritmo del Gasto Público.

Test de Dickey-Fuller

Este test contrasta en su hipótesis nula que el proceso contiene una raíz unitaria y por lo tanto que es no estacionario, mientras que la alternativa es que no presenta raíz unitaria y por lo tanto, se cumple el supuesto de estacionariedad.

La siguiente ecuación es la regresión auxiliar con tendencia determinística para la solución paramétrica de DF (1979):

$$\Delta Y_t = a_0 + a_2 t + \gamma Y_{t-1} + \sum_{j=1}^k \beta_j \Delta Y_{t-j} + \epsilon_t$$

donde Y_t es la serie de tiempo, t es el tiempo, Δ es el operador de diferencia, a_0 y a_2 son los coeficientes de la regresión, γ es el coeficiente de la raíz unitaria, β_j son los coeficientes de los rezagos y ϵ_t es el término de error.

Test de Dickey-Fuller (2)

Los contrastes de hipótesis que se realizan son los siguientes tres:

- ▶ Estadístico de prueba, τ_3 contrasta la hipótesis nula de que hay raíz unitaria, contra la alternativa de que no hay raíz unitaria, es decir es un proceso estacionario.

$$\mathbf{H}_0 : \gamma = 0$$

$$\mathbf{H}_1 : \gamma < 1$$

- ▶ Estadístico de prueba, ϕ_2 contrasta la hipótesis nula de que no tiene raíz unitaria, constante ni tendencia, contra la alternativa de que una de ellas es distinta de 0.

$$\mathbf{H}_0 : \gamma = a_0 = a_2 = 0$$

$$\mathbf{H}_1 : \text{Al menos uno de ellos es distinto de 0.}$$

Test de Dickey-Fuller (3)

- Estadístico de prueba, ϕ_3 contrasta la hipótesis nula de que no tiene raíz unitaria ni tendencia, contra la alternativa de que una de ellas es distinta de 0.

$$\mathbf{H}_0 : \gamma = a_2 = 0$$

$$\mathbf{H}_1 : \text{Al menos uno de ellos es distinto de 0.}$$

El estadístico τ_3 bajo la hipótesis nula sigue una distribución t-Student, mientras que los estadísticos ϕ_2 y ϕ_3 de prueba conjunta siguen una distribución F de Fisher.

Test de Dickey-Fuller (4)

Se aplica el test de Dickey-Fuller sobre el logaritmo del gasto público. El test con tendencia y 12 rézagos significativos obtuvo los siguientes resultados para un nivel de significación al 5%. El estadístico $\tau_3 = -2.33$ y un $p - valor = -3.42$, no se rechaza H_0 y por lo tanto la serie tiene una raíz unitaria. Luego, $\phi_2 = 20.29$ y un $p - valor = 4.71$, se rechaza H_0 , por lo tanto al menos uno de $\gamma = a_0 = a_2$ es distinto de 0, es decir o presenta constante, o presenta tendencia, porque ya se contrastó que $\gamma = 0$. Por último, $\phi_3 = 3.13$ y un $p - valor = 6.30$, no se rechaza H_0 , entonces γ o a_0 es distinto de 0.

Por lo tanto, dado los 3 contrastes se concluye que la serie tiene raíz unitaria, no tiene tendencia y presenta constante. Esto sugiere que la serie no es estacionaria y se debe de aplicar una diferencia regular para obtener un proceso estacionario.

Diferencia regular

$$(1 - L) \text{Ln}(\text{gasto})$$

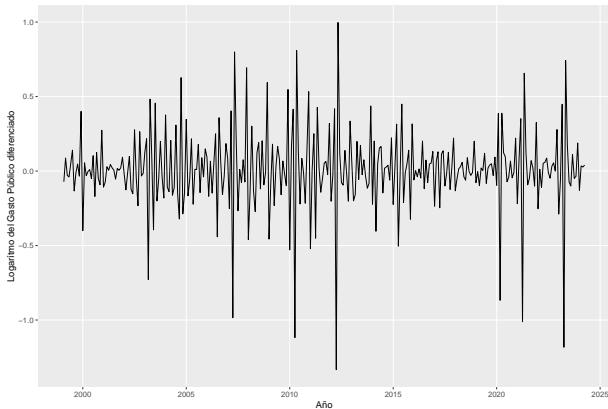


Figura 6: Evolución de la primera diferencia regular del logaritmo del Gasto Público mensual entre 1999 y 2024.

Funciones de Autocorrelación serie diferenciada

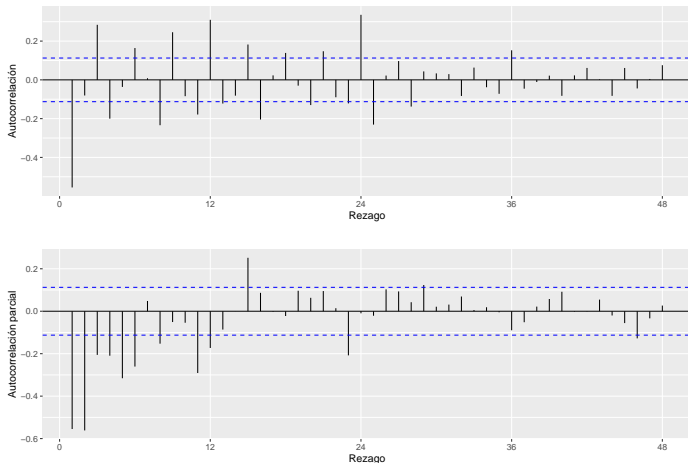


Figura 7: Funciones de Autocorrelación y Autocorrelación Parcial estimadas para la primera diferencia regular del logaritmo del Gasto Público.

Diferencia estacional

$$(1 - L^{12})\ln(gasto)$$

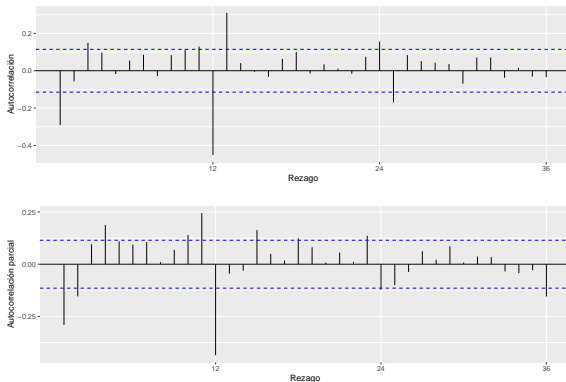


Figura 8: Funciones de Autocorrelación y Autocorrelación Parcial estimadas para la primera diferencia estacional del logaritmo del Gasto Público (1999-2024).

Identificación

Se identifican dos modelos posibles, **SARIMA(0,0,1)(0,1,1)** y **SARIMA(0,0,1)(1,1,0)**.

Estos modelos surgen de los rezagos significativos presentes en las funciones. Con respecto a la parte regular de modelo, se puede observar una rápida convergencia de la serie en la FAC, mientras que en la FACP dicha convergencia no es tan rápida, lo cual sugiere un modelo que contiene solo parte MA; el rezago 1 significativo de la FAC sugiere $q = 1$ y $p = d = 0$ en un modelo SARIMA(p,d,q)(P,D,Q). Con respecto a la parte estacional, los rezagos múltiplos de 12 son los relevantes en la FAC y FACP, ambas funciones presentan el primer rezago (rezago 12) significativo, y una convergencia medianamente rápida. Esto sugiere probar dos posibilidades $P = 1$ o $Q = 1$ siendo $D = 1$.

Diferencia regular y estacional

$$(1 - L)(1 - L^{12})Ln(gasto)$$

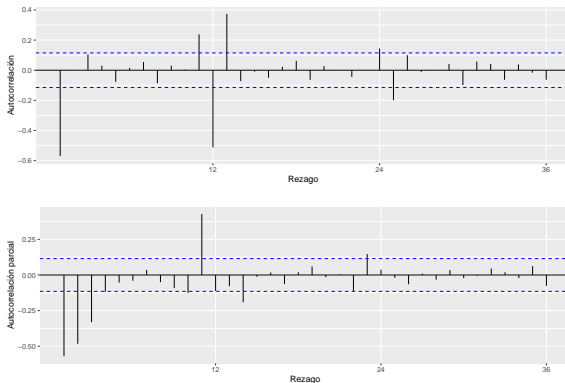


Figura 9: Funciones de Autocorrelación y Autocorrelación Parcial estimadas para la primera diferencia regular de la primera diferencia estacional del logaritmo del Gasto Público.

Identificación (2)

La convergencia de las funciones es mas consistente que en la serie con diferencia estacional únicamente. A partir de estas funciones se identifican 2 modelos adicionales, **SARIMA(0,1,2)(1,1,0)** y **SARIMA(0,1,1)(0,1,1)**.

En este caso al haber ambas diferencias los parámetros $d = 1$ y $D = 1$ son fijos. Con respecto a la parte regular, la FACP presenta los primeros tres rezagos significativos, sin embargo se descarta un modelo con parte AR de orden tan grande dado que la FAC no presenta un ruido tan grande; de haber parte AR, la FAC no presentaría una convergencia tan rápida como tiene a partir del rezago 2. En cambio, la convergencia de la FACP es más lenta, lo cual sugiere un modelo con parte MA de orden 1 e incluso 2, por lo tanto $q = 1, 2$. Con respecto a la parte estacional, las conclusiones son las mismas a la de la figura 8, la FAC y la FACP presenta el primer rezago significativo dando lugar a dos posibilidades $P = 1$ o $Q = 1$.

Variables regresoras

La etapa de identificación ha identificado 4 modelos posibles para modelar la serie de Gasto Público. Estos modelos son en esta parte estimados y comparados. Adicionalmente es incorporada al modelo una variable regresora, los días de turismo, con el objetivo de capturar la variabilidad de la serie generada en los marzos y abril de cada año, que coincida con el pago de los aguinaldos de los empleados públicos. La figura 10 representa la evolución del logaritmo del Gasto Público entre enero 2009 y diciembre 2012, con las líneas verticales punteadas marcando los meses de marzo y abril de cada año.

Variable regresora: turismo

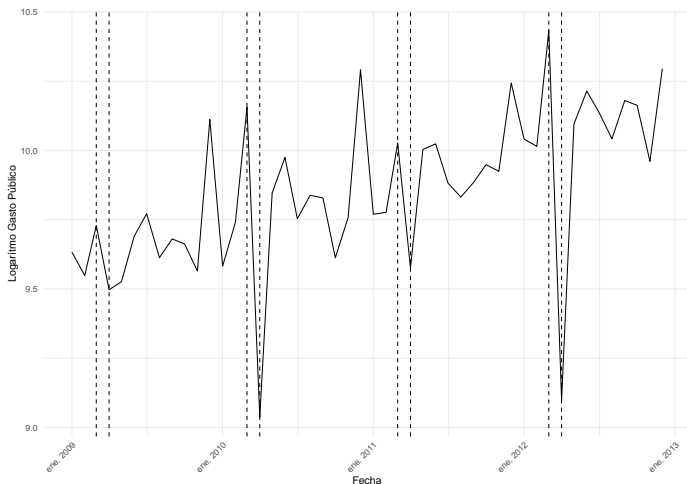


Figura 10: Evolución del Logaritmo del Gasto Público mensual entre enero 2009 y diciembre 2012.

Variable regresora: turismo

Dependiendo de cuando caen los días de turismo, determina el mes en el que se efectuará el pago de los aguinaldos de los empleados públicos. Esto puede generar un pago anticipado de los aguinaldos en determinado mes, generando un alto gasto, y un gasto mucho menor al mes siguiente. El cuadro 1 muestra los días de turismo de cada año entre 2009-2012:

	2009	2010	2011	2012
Marzo	0	3	0	0
Abril	7	4	7	7

Cuadro 1: Días de turismo en marzo y abril de 2009 a 2012.

Estimación

El cuadro 2, es un resumen de la estimación de los 4 modelos SARIMA identificados, proporciona la varianza del modelo y la significancia de sus componentes.

Modelo	SARIMA(p,d,q)(P,D,Q) + turismo	σ^2	Significancia de los componentes
1	(0,0,1)(0,1,1)	0.0493	MA 1 no significativo
2	(0,0,1)(1,1,0)	0.0484	MA 1 no significativo
3	(0,1,2)(1,1,0)	0.0240	Todos significativos
4	(0,1,1)(0,1,1)	0.0271	Todos significativos

Cuadro 2: Estimación de modelos SARIMA.

Los dos modelos que incluyen únicamente la diferencia estacional, aparentan ser peores que los modelos con diferencias regular y estacional en términos de varianza del modelo, a su vez el componente MA de orden 1 en la parte regular de los dos primeros modelos, suele ser rechazada al 5% de nivel de significación.

Diagnóstico

En esta etapa, los 4 modelos estimados son diagnosticados para verificar si cumplen con los supuestos de ruido blanco es decir incorrelación y normalidad de los residuos.

El análisis de las funciones de autocorrelación y autocorrelación parcial de los residuos permite observar la autocorrelación de los mismos, y acompañado del test de Ljung-Box, se verifica la incorrelación.

Por otro lado los tests de Shapiro-Wilks y Jarque-Bera, permiten verificar la normalidad de los residuos.

Diagnóstico: Residuos y residuos estandarizados

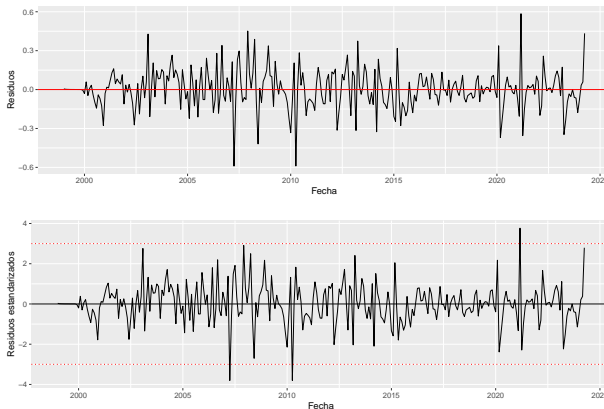


Figura 11: Residuos normales y estandarizados de un modelo $\text{SARIMA}(0,1,2)(1,1,0)$ para el logaritmo del Gasto Público.

Diagnóstico: Función de autocorrelación y autocorrelación parcial de los residuos

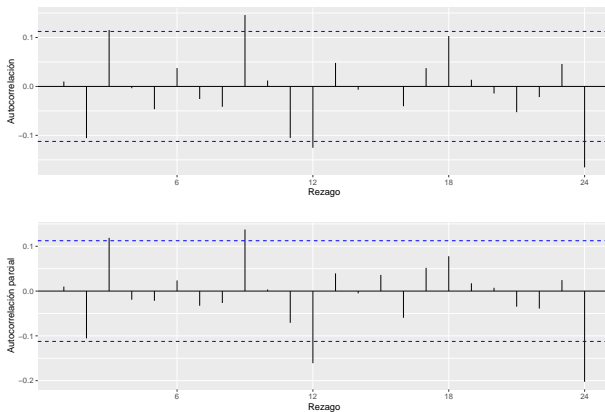


Figura 12: Funciones de Autocorrelación y Autocorrelación Parcial estimadas de los residuos de un modelo SARIMA(0,1,2)(1,1,0) para el logaritmo del Gasto Público.

Diagnóstico: Tests de Ljung-Box y Shapiro-Wilks

Veánse los resultados de la incorrelación y normalidad de los residuos en el cuadro 3.

Modelo	SARIMA(p,d,q)(P,D,Q) + turismo	Incorrelación	Normalidad
1	(0,0,1)(0,1,1)	No	No
2	(0,0,1)(1,1,0)	No	No
3	(0,1,2)(1,1,0)	No	No
4	(0,1,1)(0,1,1)	No	No

Cuadro 3: Diagnostico de modelos SARIMA.

Diagnóstico: Detección de outliers

Los outliers detectados para cada modelo son las siguientes observaciones:

- ▶ Modelo 1

- ▶ AO: 51, 100, 136, 160, 195, 255, 267, 280, 304.
- ▶ LS: 117, 159.
- ▶ TC: 254

- ▶ Modelo 2

- ▶ AO: 51, 100, 136, 160, 195, 255, 267, 280, 304.
- ▶ TC: 159, 254, 291.

- ▶ Modelo 3

- ▶ AO: 100, 160, 292.

- ▶ Modelo 4

- ▶ AO: 51, 100, 136, 160, 195, 255, 268, 279, 292.

Reestimación

Modelo	SARIMA(p,d,q)(P,D,Q) + turismo + outliers	σ^2	Significancia de los componentes
1	(0,0,1)(0,1,1)	0.0227	s_MA 1 no significativo
2	(0,0,1)(1,1,0)	0.0226	Todos significativos
3	(0,1,2)(1,1,0)	0.0205	Regresor turismo no significativo
4	(0,1,1)(0,1,1)	0.0156	Todos significativos

Cuadro 4: Reestimación de modelos SARIMA con intervención de outliers.

Reestimación (2)

Modelo	SARIMA(p,d,q)(P,D,Q) + outliers	σ^2	Significancia de los componentes
1	(0,0,1)(0,1,0) + turismo	0.0226	Todos significativos
3	(0,1,2)(1,1,0)	0.0205	Todos significativos

Cuadro 5: Reestimación de modelos SARIMA con intervención de outliers.

Diagnóstico: Residuos y residuos estandarizados

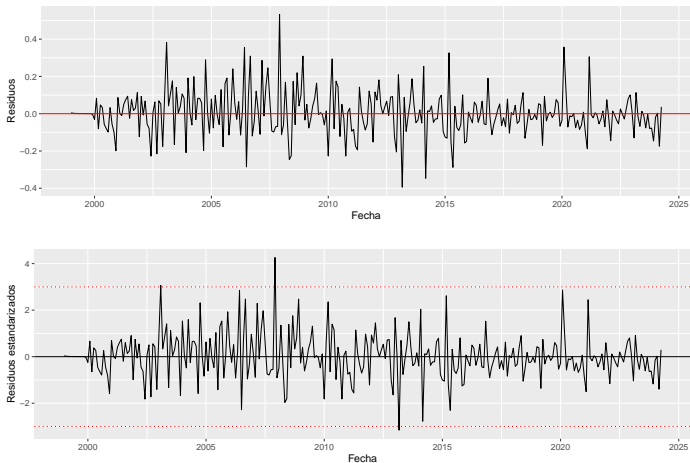


Figura 13: Residuos normales y estandarizados de un modelo $\text{SARIMA}(0,1,1)(0,1,1)$ intervenido para el logaritmo del Gasto Público.

Diagnostico: Función de autocorrelación y autocorrelación parcial de los residuos

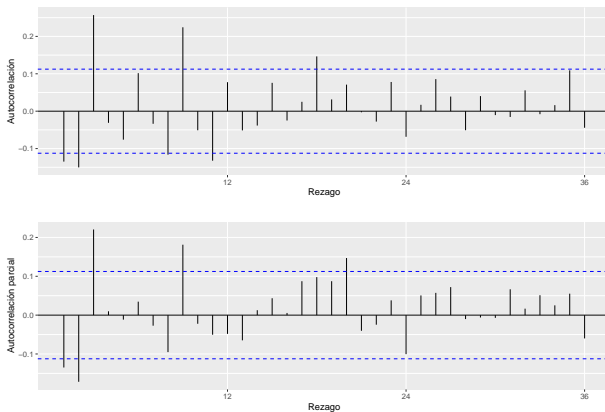


Figura 14: Funciones de Autocorrelación y Autocorrelación Parcial estimadas de los residuos de un modelo SARIMA(0,1,1)(0,1,1) intervenido para el logaritmo del Gasto Público.

Diagnostico: Tests de Ljung-Box y Shapiro-Wilks

Modelo	SARIMA(p,d,q)(P,D,Q) + outliers	Incorrelación	Normalidad
1	(0,0,1)(0,1,0) + tursimo	No	No
2	(0,0,1)(1,1,0) + turismo	No	No
3	(0,1,2)(1,1,0)	No	No
4	(0,1,1)(0,1,1) + turismo	No	No

Cuadro 6: Diagnosticos de modelos SARIMA.

Diagnóstico: Histograma de los residuos

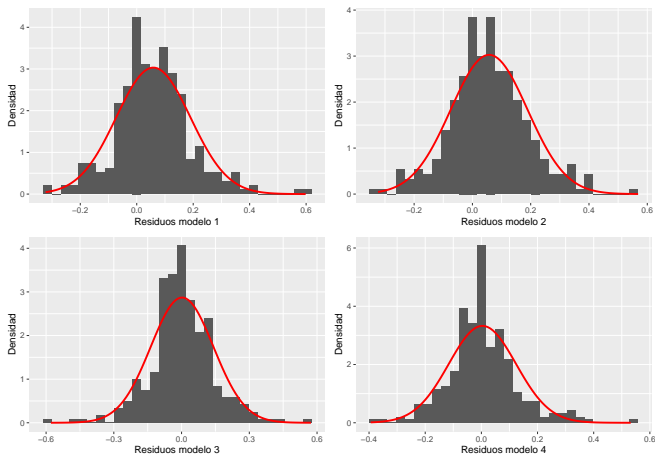


Figura 15: Histograma de los residuos de los modelos reestimados SARIMA intervenido para el logaritmo del Gasto Público. La línea roja corresponde a una densidad normal con media y desvío muestrales igual al de los residuos.

Predicción

Puntos a tener en cuenta:

- ▶ Función de pérdida: simétrica.
- ▶ Horizonte de predicción: 12 meses.
- ▶ Conjunto de entrenamiento: hasta abril de 2023.
- ▶ Conjunto de testeo: desde mayo de 2023.
- ▶ Variables exógenas: turismo y outliers.

Predicción (2)

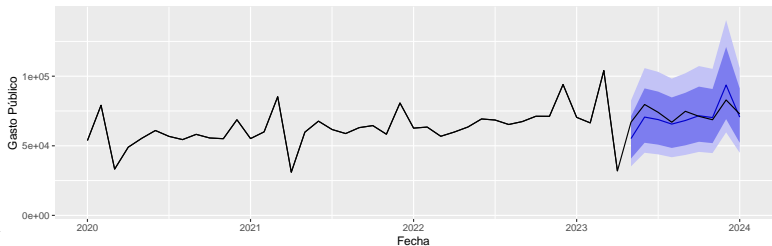
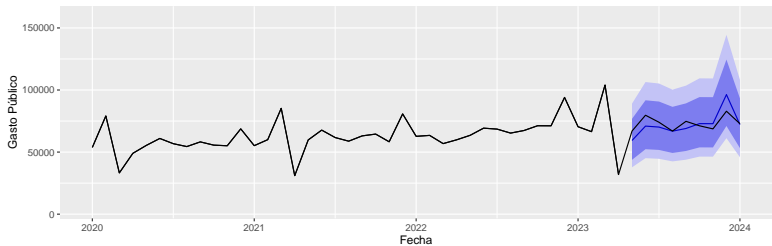
Las métricas de error que se utilizarán para comparar los modelos son:

- ▶ Error absoluto medio (MAE): mide la magnitud absoluta promedio del desvío de la predicción del valor real.
- ▶ Error porcentual absoluto medio (MAPE): mide la magnitud promedio del desvío de la predicción del valor real en términos porcentuales.
- ▶ Error medio absoluto escalado (MASE): mide la magnitud promedio del desvío de la predicción del valor real, escalada por el MAE del conjunto de entrenamiento.

Un valor más cercano a cero indica una mejor predicción para las tres métricas.

Predicción de los modelos 1 y 2

\begin{figure}[H]

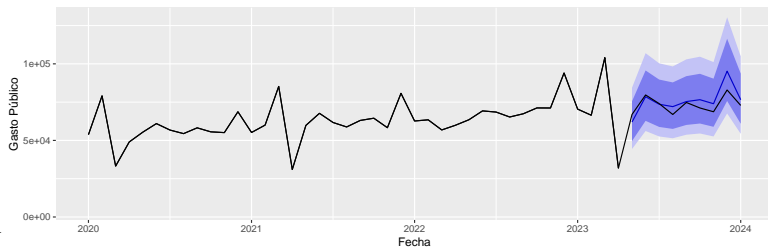
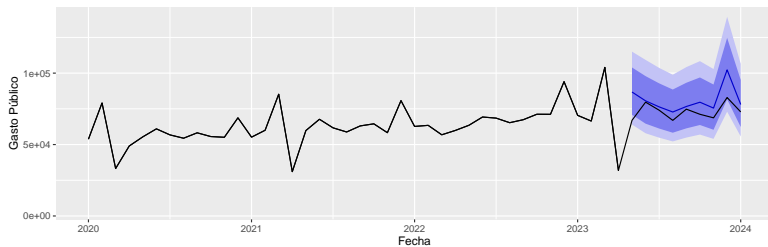


{

}

Predicción de los modelos 3 y 4

`\begin{figure}[H]`



{

}

Predicción: Métricas de error

Modelo	Conjunto	MAE	MAPE	MASE
1	Training	3541	12.85	0.868
1	Test	10870	13.98	2.664
2	Training	3565	13.10	0.874
2	Test	9909	12.84	2.429
3	Training	2757	10.92	0.676
3	Test	9333	12.45	2.288
4	Training	2729	10.87	0.669
4	Test	6829	8.93	1.674

Cuadro 7: Resultados de error de predicción para los diferentes modelos.

Conclusiones

En base a los resultados obtenidos, se puede concluir que el modelo 4 (entre los presentados) es el que presenta un mejor desempeño en los datos. Este modelo es un SARIMA(0,1,1)(0,1,1) con un regresor turismo y la inclusión de outliers detectados.

Se ha logrado llegar a resultados positivos, mas alla de la falta de cumplimiento de los supuestos de los residuos respecto a la incorrelación y normalidad de los residuos. Los puntos atípicos presentes en la serie, han sido correctamente identificados y tratados en los modelos, lo cual ha permitido mejorar la calidad de las estimaciones.

Comentarios Finales

Hubiese estado bueno que al identificar los outliers aparte de haber mejorado la estimación, hubiese ayudado al cumplimiento de la incorrelación y normalidad de los residuos.

La predicción realizada en el conjunto de testeo, respondió correctamente, presentando un error de predicción dentro de las expectativas. El modelo 3 presenta una predicción rara de la primer observación que solo queda contenida en el intervalo de predicción al 95% de confianza, se desconoce el motivo de este comportamiento.

A efectos futuros, se sugiere probar las siguientes combinaciones de modelos que quedaron pendientes de estimar y diagnosticar, SARIMA(p,d,q)(P,D,Q):

▶ (0,1,2) (0,1,1)

▶ (0,1,1) (1,1,0)

Bibliografía

Box, G. E. P., G. M. Jenkins, y G. C. Reinsel. 1994. *Time Series Analysis; Forecasting and Control*. 3rd ed. Englewood Cliff, New Jersey: Prentice Hall.

Ministerio de Economía y Finanzas. 2024. «Información y Resultados del Sector Público».

<https://www.gub.uy/ministerio-economia-finanzas/datos-y-estadisticas/estadisticas/informacion-resultados-del-sector-publico>.

¡Muchas Gracias!