

Paolo Sebastiani

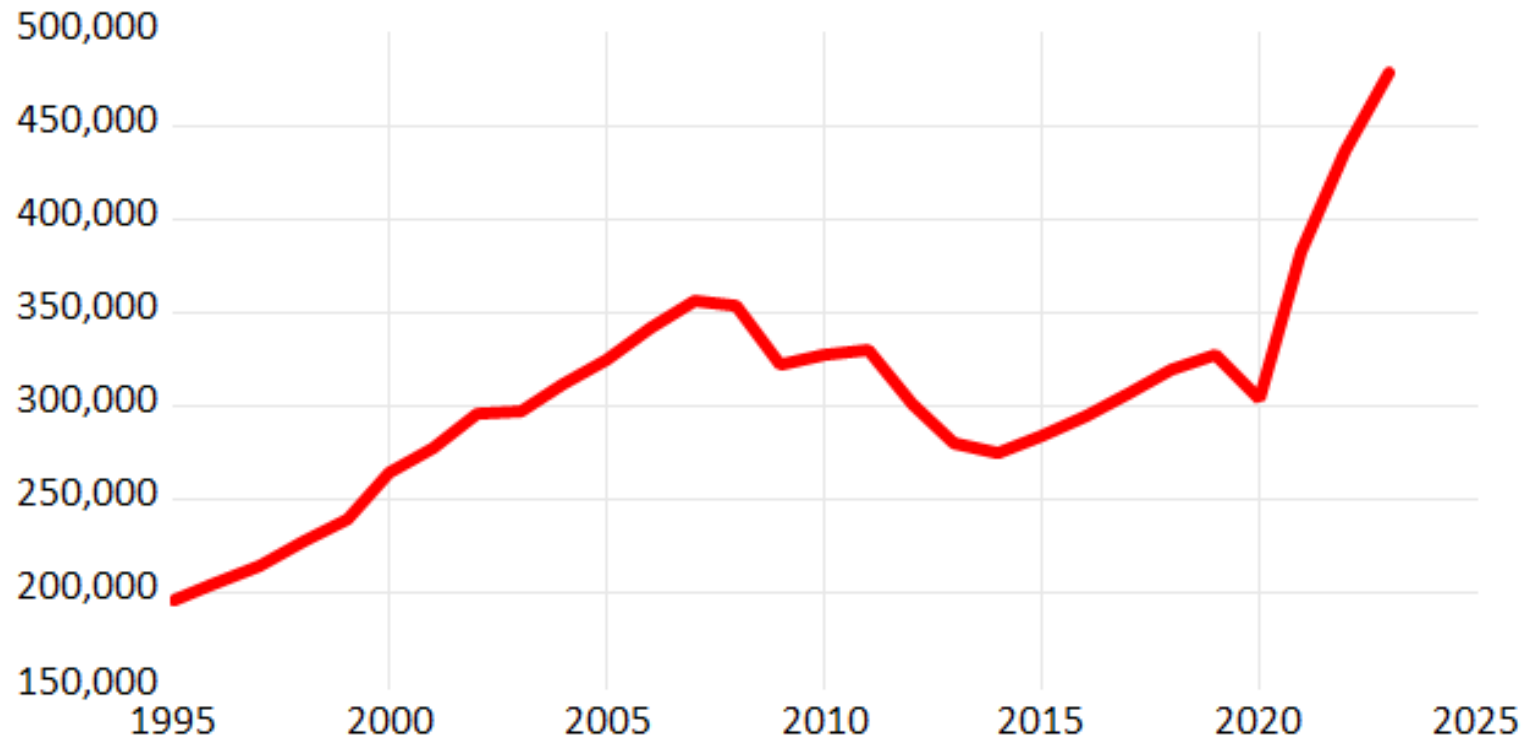
Práctica final

Serie anual

Formación bruta de capital fijo

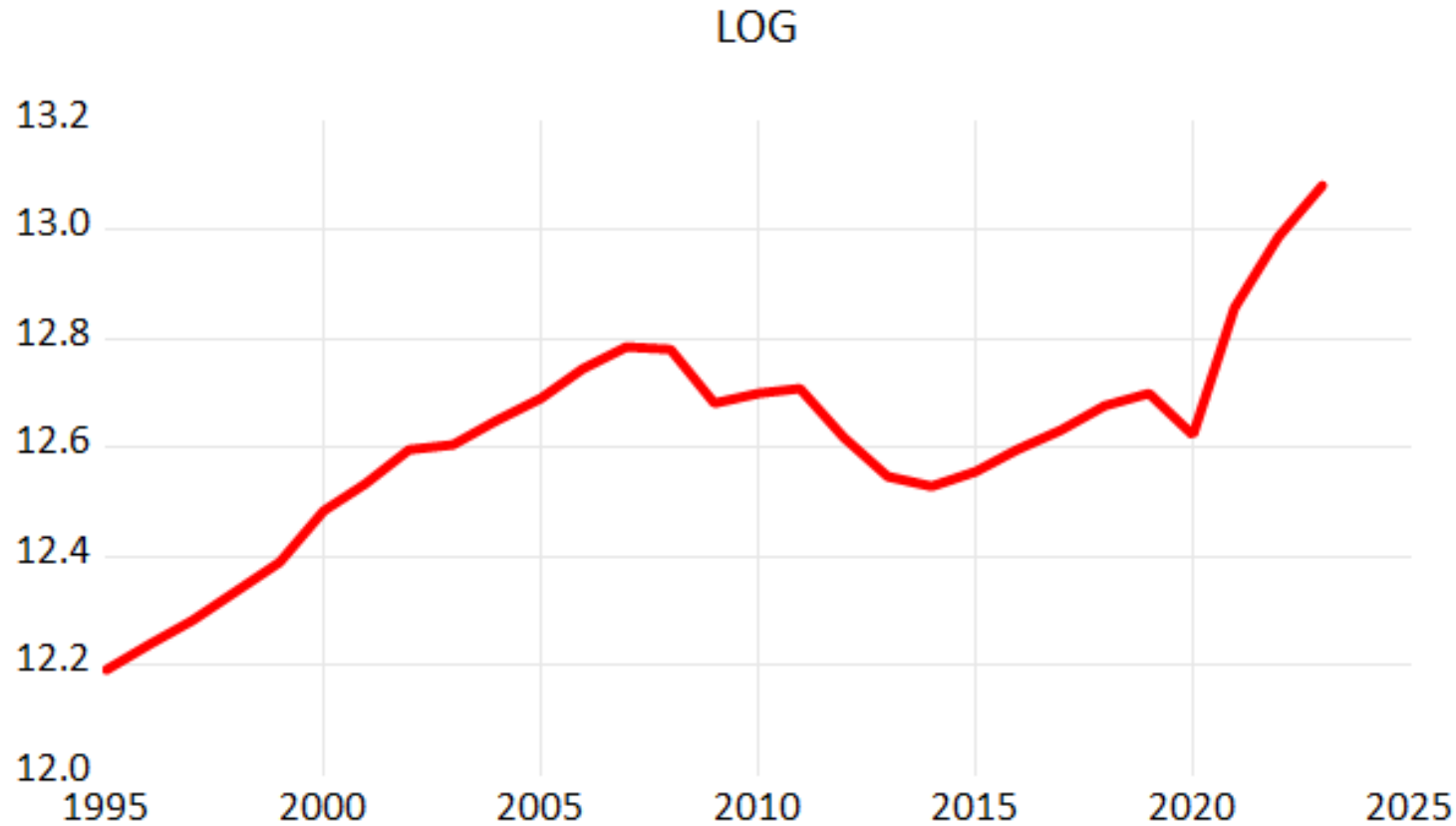
DESCRIPCIÓN DE LA SERIE TEMPORAL

Formación bruta de capital fijo (en millones de euros)



- Vamos a trabajar con la serie de la formación bruta de capital fijo de Italia (en millones de euros, a precios corrientes), fuente: ISTAT
- La periodicidad es anual, y el periodo es 1995-2023.
- Presenta evolutividad: no tiene una media constante, por tanto, no es estacionario.
- Sigue una tendencia alcista hasta 2008 (año en que la Gran Recesión llegó a Italia)
- Desde 2008 se observa una tendencia negativa, que volvió a ser positiva en 2014
- En 2020 se vuelve a producir un descenso (debido a la crisis de los COVID), e inmediatamente después un gran repunte, con una tendencia que presenta una pendiente mayor que en periodos anteriores a la crisis

TRANSFORMACIÓN LOGARÍTMICA DE LA SERIE



- Aplicamos la transformación logarítmica, tanto para reducir la varianza (hacer las series más estables) como para interpretar los datos más fácilmente (por ejemplo, las primeras diferencias pueden interpretarse como valores porcentuales).
- Aunque la transformación logarítmica suele hacer que las tendencias sean más lineales, aquí se observa que las características de la serie en cuanto a su comportamiento/evolución se mantienen

ESTIMACIÓN DE UNA TENDENCIA DETERMINÍSTICA

Dependent Variable: LOG(X)

Method: Least Squares

Date: 02/15/25 Time: 12:36

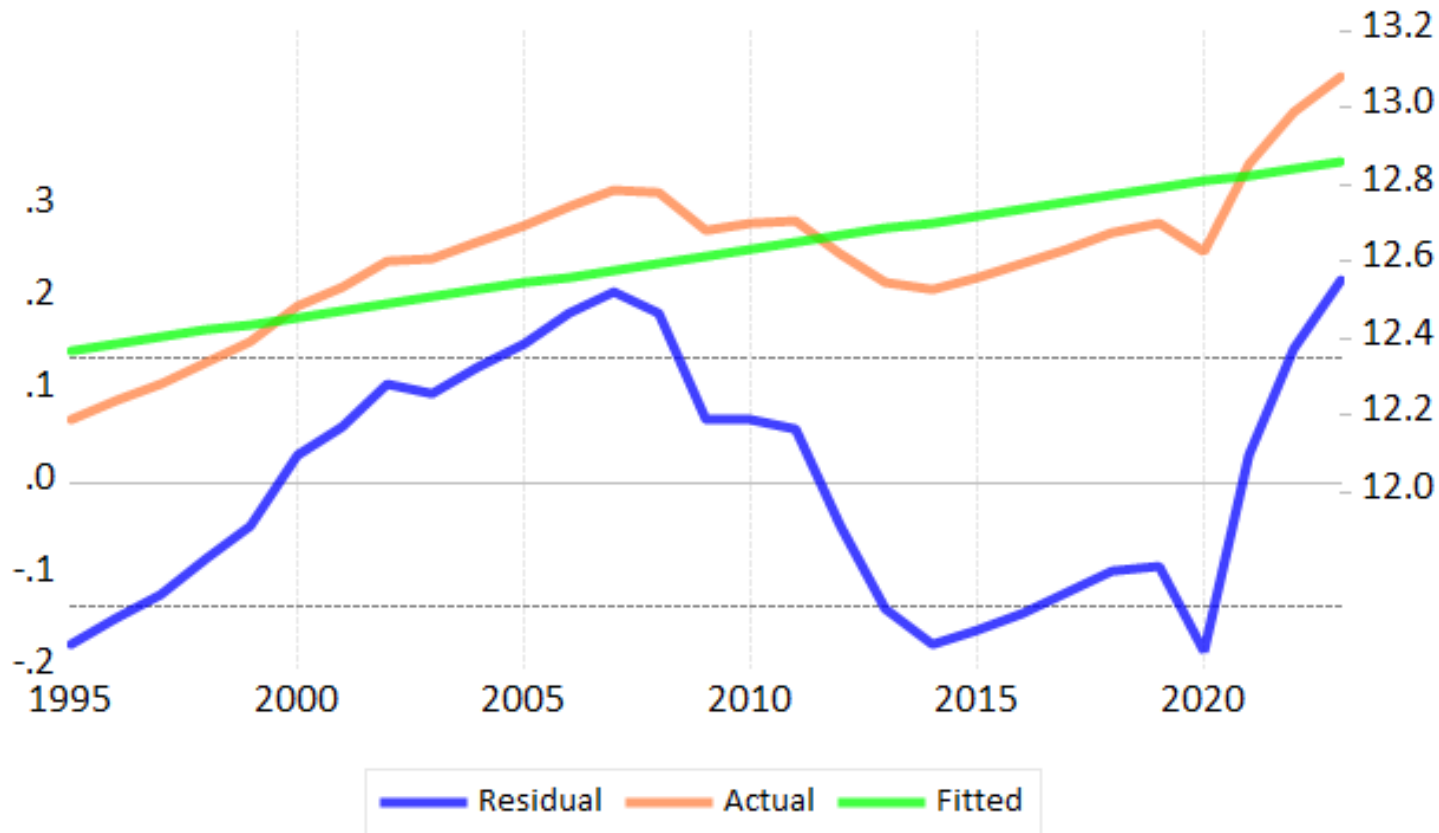
Sample (adjusted): 1995 2023

Included observations: 29 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	12.36597	0.049082	251.9461	0.0000
T	0.017632	0.003009	5.859025	0.0000
R-squared	0.559746	Mean dependent var	12.61282	
Adjusted R-squared	0.543440	S.D. dependent var	0.200668	
S.E. of regression	0.135590	Akaike info criterion	-1.091889	
Sum squared resid	0.496386	Schwarz criterion	-0.997593	
Log likelihood	17.83239	Hannan-Quinn criter.	-1.062357	
F-statistic	34.32817	Durbin-Watson stat	0.251708	
Prob(F-statistic)	0.000003			

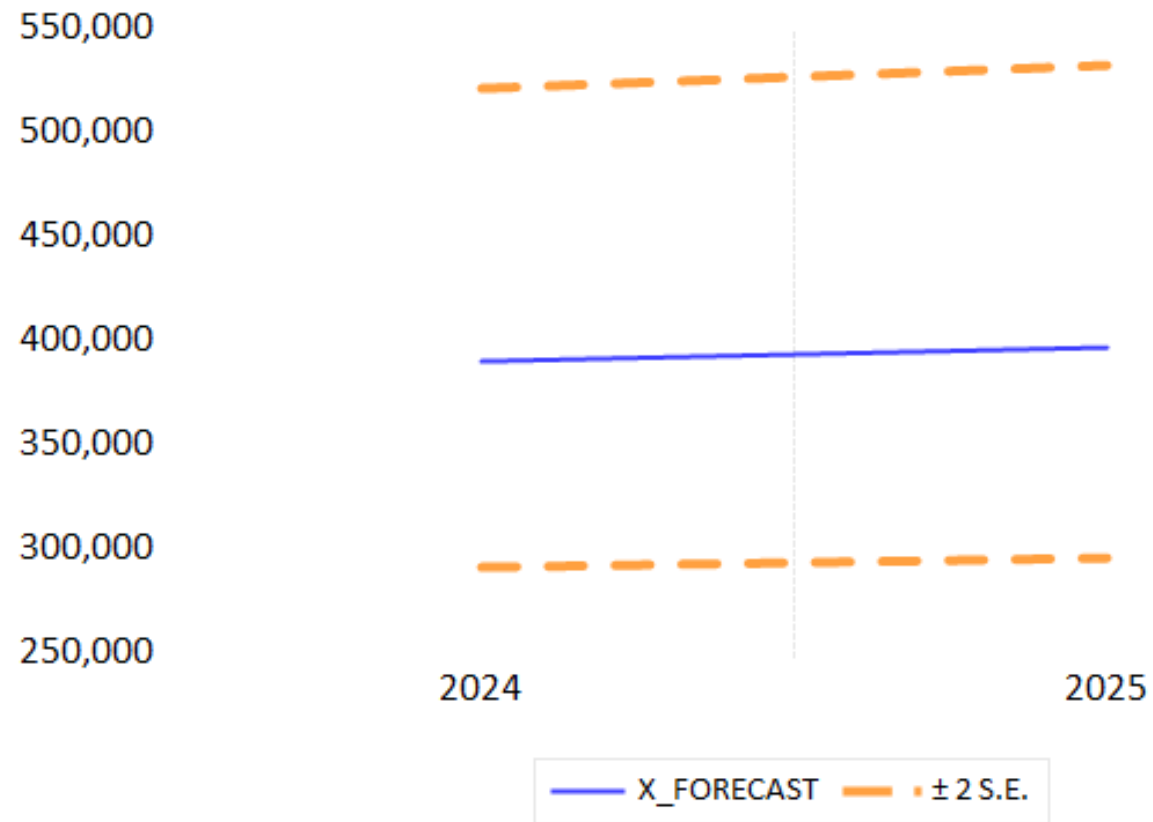
- Estimemos ahora una tendencia lineal del tipo $\log X_t = \alpha + \beta t + u_t$, donde la tendencia será: $T = \alpha + \beta t$ (α es la constante, β es la pendiente y t es la variable tiempo que va de 1 a 31)
- Los coeficientes ($\alpha \approx 12,37$ y $\beta \approx 0,018$) son estadísticamente significativos
- El índice R^2 es 0,56: ni bajo ni alto
- El estadístico F es bastante elevado (es decir, el modelo conjuntamente es significativo)
- Nos dan la desviación típica de los residuos (0,1356)
- Nos da el estadístico Durbin-Watson, que mide correlación de orden 1, y es igual a 0,25. Dado que el valor está muy lejos de 2, entonces los residuos están autocorrelacionados (correlación de primer orden), y se violan los supuestos de MLG

ESTIMACIÓN DE UNA TENDENCIA DETERMINÍSTICA - RESIDUOS

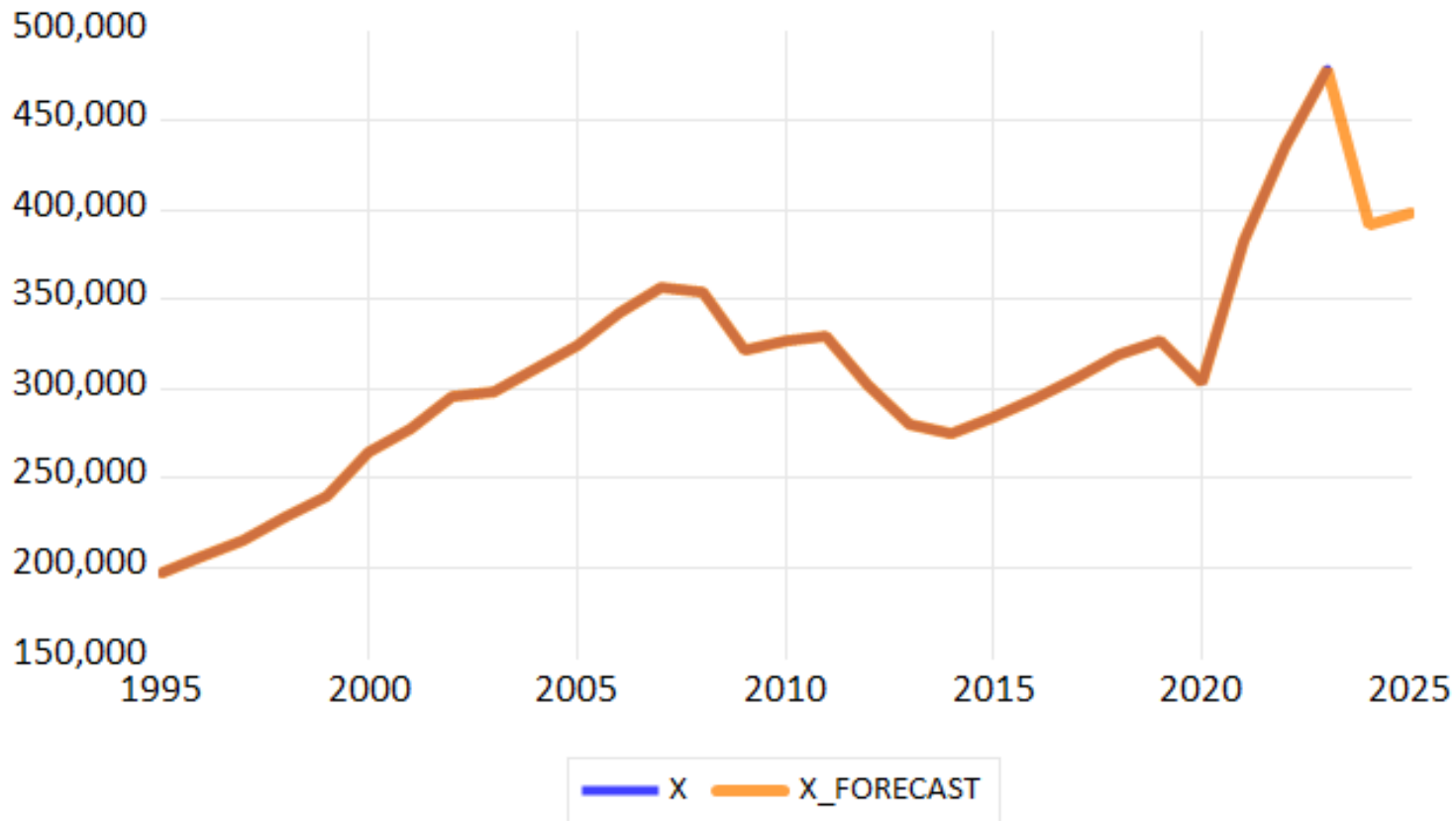


- Se observa claramente que la evolución tendencial estimada (en verde) no ajusta la serie (en naranja)
- Si se analizan los residuos, es evidente que no cumplen las hipótesis del modelo de regresión: la media no es cero y se sitúan sistemáticamente por encima o por debajo de cero (es decir, no son estacionarios y entonces están autocorrelacionados)
- Todo esto indica claramente que el modelo determinístico lineal no es válido en este caso

ESTIMACIÓN DE UNA TENDENCIA DETERMINÍSTICA - PREDICCIONES



ESTIMACIÓN DE UNA TENDENCIA DETERMINÍSTICA - PREDICCIONES



- La predicción pronostica una fuerte caída en 2024, lo que no parece tener una base económica racional
- La razón es que, como podemos ver en el gráfico de dos diapositivas antes, al final la serie estaba muy por encima del modelo, por lo que debe haber la caída
- Además, desde 2025 parece continuar de forma lineal (con un crecimiento igual al β estimado, es decir 0,018)

ESTIMACIÓN DE UNA TENDENCIA DETERMINÍSTICA - PREDICCIONES

	2022	2023	2024	2025
CON DATOS HASTA 2021	-1,29%	+1,8%		
CON DATOS HASTA 2022		-11,9%	+1,8%	
CON DATOS HASTA 2023			-18,3%	+1,8%
REAL	+14,02%	+9,77%		

- Se observa que los valores predichos siguen la tendencia alcista dictada por el β estimado (es decir, +1,8%)
- Comparando los valores previstos con los valores reales (disponibles hasta 2023), está claro que el modelo es malo porque se equivoca totalmente
- En 2024, como ya hemos dicho, se prevé una caída enorme de la economía para volver a ajustar el modelo



















Modelos estocásticos

SERIE ANUAL EN LOGARITMOS - CORRELOGRAMA

Date: 03/04/25 Time: 19:16

Sample (adjusted): 1995 2023

Included observations: 29 after adjustments

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
		1	0.759	0.759	18.484	0.000
		2	0.519	-0.134	27.440	0.000
		3	0.317	-0.066	30.915	0.000
		4	0.213	0.080	32.545	0.000
		5	0.069	-0.199	32.723	0.000
		6	-0.048	-0.049	32.814	0.000
		7	-0.145	-0.062	33.675	0.000
		8	-0.203	-0.069	35.433	0.000
		9	-0.229	-0.008	37.787	0.000
		10	-0.191	0.070	39.509	0.000

- No hay un punto de corte evidente, aunque la función de autocorrelación muestral se vuelve no significativa a partir del tercer retardo
- Considerando entonces que el decaimiento de la función de autocorrelación es bastante rápido, podríamos pensar que la serie es estacionaria
- Sin embargo, observando el gráfico de la serie, no parece estacionaria de ninguna manera, por lo que es mejor recurrir a una prueba estadística formal (es decir, un contraste de raíz unitaria)

SERIE ANUAL EN LOGARITMOS – CONTRASTE DE RAÍCES UNITARIAS

Null Hypothesis: LOG_X has a unit root

Exogenous: None

Lag Length: 0 (Automatic - based on SIC, maxlag=6)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	2.516526	0.9960
Test critical values: 1% level	-2.650145	
5% level	-1.953381	
10% level	-1.609798	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation

Dependent Variable: D(LOG_X)

Method: Least Squares

Date: 03/04/25 Time: 19:19

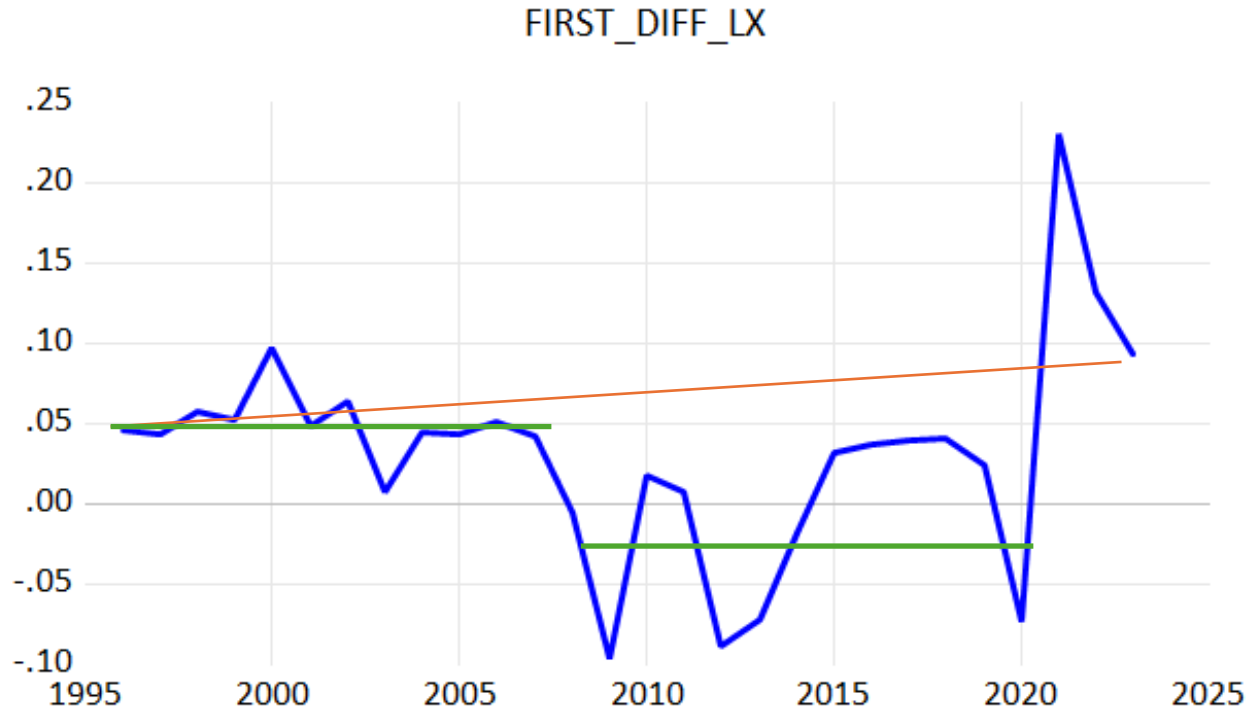
Sample (adjusted): 1996 2023

Included observations: 28 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
LOG_X(-1)	0.002512	0.000998	2.516526	0.0181
R-squared	-0.001715	Mean dependent var		0.031762
Adjusted R-squared	-0.001715	S.D. dependent var		0.066487
S.E. of regression	0.066544	Akaike info criterion		-2.546847
Sum squared resid	0.119559	Schwarz criterion		-2.499268
Log likelihood	36.65585	Hannan-Quinn criter.		-2.532301
Durbin-Watson stat	1.351385			

- Este es el resultado del “test de Dickey-Fuller aumentado”, en cuyo modelo tanto la constante como la tendencia temporal están ausentes (porque no son significativas)
- El valor del t-estadístico es igual a +2,52
- Mucho mayor que valores críticos
- El P-valor es muy alto (es decir, 0,996), entonces aceptamos que hay una raíz unitaria
- Trabajemos entonces con la primera diferencia de la serie en logaritmos

SERIE ANUAL EN LOGARITMOS DIFERENCIADA (PRIMERA DIFERENCIA)



- La primera diferencia de los logaritmos es una aproximación a las tasas de crecimiento
- Al tomar una primera diferencia, se ha reducido la tendencia alcista que presentaba la serie
- La media a lo largo de toda la serie no parece ser constante, pero hay dos medias constantes en dos subperíodos (en verde)
- Así que a simple vista la serie sigue pareciendo no estacionaria, sin embargo, es mejor confiar en la visualización del correlograma y más aún en el contraste de raíz unitaria
- Sin embargo, es importante observar que hay dos medias (en verde) y un periodo atípico muy significativo. Esto debe tenerse en cuenta para la modelización























SERIE ANUAL EN LOGARITMOS DIFERENCIADA (PRIMERA DIFERENCIA)

CORRELOGRAMA COMPLETO

Date: 03/04/25 Time: 19:28

Sample (adjusted): 1996 2023

Included observations: 28 after adjustments

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
		1	0.308	0.308	2.9539	0.086
		2	0.131	0.040	3.5082	0.173
		3	0.081	0.033	3.7283	0.292
		4	0.156	0.131	4.5794	0.333
		5	0.087	0.000	4.8566	0.434
		6	0.066	0.022	5.0206	0.541
		7	-0.053	-0.101	5.1312	0.644
		8	-0.179	-0.183	6.4721	0.595
		9	-0.374	-0.328	12.649	0.179
		10	-0.228	-0.067	15.084	0.129
		11	-0.069	0.075	15.318	0.168
		12	-0.320	-0.305	20.678	0.055

- Observando el correlograma, parece que la serie es ahora un paseo aleatorio, ya que no hay autocorrelación significativa (y las Q-stats son muy pequeñas)
- Sin embargo, esto parece bastante extraño desde un punto de vista económico
- Por tanto, puede que los valores atípicos influyan demasiado















SERIE ANUAL EN LOGARITMOS DIFERENCIADA (PRIMERA DIFERENCIA)

CORRELOGRAMA SOLO HASTA 2019

Date: 03/11/25 Time: 10:16

Sample (adjusted): 1996 2019

Included observations: 24 after adjustments

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
		1	0.533	0.533	7.7126	0.005
		2	0.253	-0.043	9.5303	0.009
		3	0.233	0.161	11.144	0.011
		4	0.251	0.097	13.114	0.011
		5	0.059	-0.188	13.228	0.021
		6	-0.055	-0.049	13.331	0.038
		7	-0.217	-0.274	15.053	0.035
		8	-0.204	0.009	16.683	0.034

- Al quitar los últimos cuatro años se observa que ahora hay una correlación serial (el ultimo p-valor es menor que 0,05)
- Por lo tanto, es probable que los valores extremos de 2020 (Covid) y 2021 (debido a la enorme recuperación de las inversiones tras la crisis Covid) influyen mucho en la serie

SERIE ANUAL EN LOGARITMOS DIFERENCIADA (PRIMERA DIFERENCIA)

CONTRASTE DE RAÍZ UNITARIA

Null Hypothesis: FIRST_DIFF_LX has a unit root
 Exogenous: None
 Lag Length: 0 (Automatic - based on SIC, maxlag=6)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-3.076587	0.0034
Test critical values: 1% level	-2.653401	
5% level	-1.953858	
10% level	-1.609571	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
 Dependent Variable: D(FIRST_DIFF_LX)
 Method: Least Squares
 Date: 03/04/25 Time: 19:40
 Sample (adjusted): 1997 2023
 Included observations: 27 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
FIRST_DIFF_LX(-1)	-0.556471	0.180873	-3.076587	0.0049
R-squared	0.266515	Mean dependent var		0.001750
Adjusted R-squared	0.266515	S.D. dependent var		0.078715
S.E. of regression	0.067415	Akaike info criterion		-2.519577
Sum squared resid	0.118163	Schwarz criterion		-2.471583
Log likelihood	35.01429	Hannan-Quinn criter.		-2.505306
Durbin-Watson stat	2.108289			

- Realizando ahora un ADF test sobre la primera diferencia de los logaritmos (de nuevo, tendencia y constante no significativas), podemos rechazar la hipótesis nula
- Por tanto, podemos concluir que esta serie no tiene raíz unitaria y, por tanto, es estacionaria (como hemos visto en el correlograma)
- Es decir, no se necesita una segunda diferencia

Estimación de modelos autorregresivos

ESTIMACIÓN

AR(1)

Dependent Variable: D(LOG(X),1)
Method: ARMA Conditional Least Squares (Gauss-Newton / Marquardt steps)
Date: 03/17/25 Time: 18:27
Sample (adjusted): 1997 2023
Included observations: 27 after adjustments
Convergence achieved after 1 iteration
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	0.443529	0.180873	2.452155	0.0212
R-squared	0.008247	Mean dependent var		0.031237
Adjusted R-squared	0.008247	S.D. dependent var		0.067694
S.E. of regression	0.067415	Akaike info criterion		-2.519577
Sum squared resid	0.118163	Schwarz criterion		-2.471583
Log likelihood	35.01429	Hannan-Quinn criter.		-2.505306
Durbin-Watson stat	2.108289			
Inverted AR Roots	.44			

El coeficiente es significativo al 5%

AR(2)

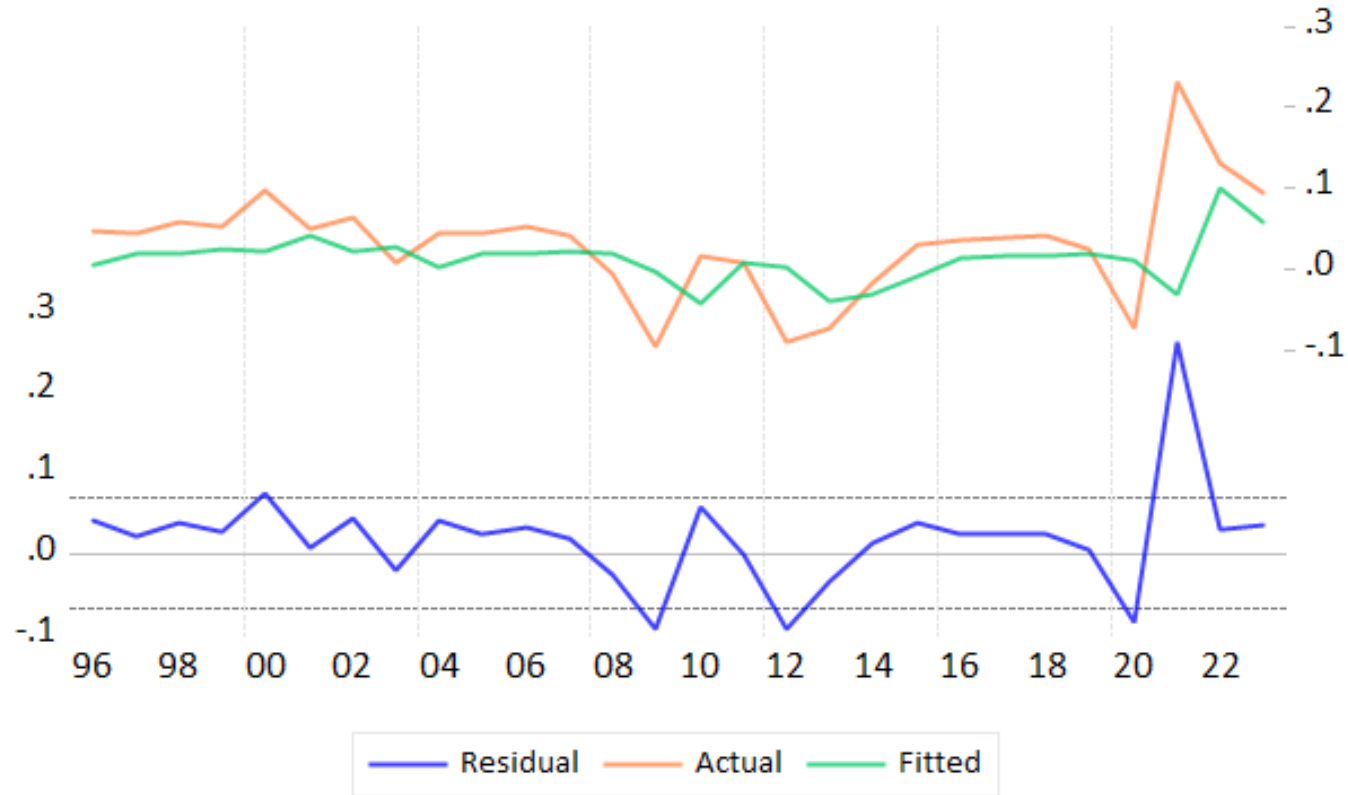
Dependent Variable: D(LOG(X),1)
Method: ARMA Conditional Least Squares (Gauss-Newton / Marquardt steps)
Date: 03/17/25 Time: 18:29
Sample (adjusted): 1998 2023
Included observations: 26 after adjustments
Convergence achieved after 1 iteration
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	0.385852	0.203143	1.899405	0.0696
AR(2)	0.138790	0.215377	0.644407	0.5254
R-squared	0.028216	Mean dependent var		0.030788
Adjusted R-squared	-0.012275	S.D. dependent var		0.068994
S.E. of regression	0.069416	Akaike info criterion		-2.423596
Sum squared resid	0.115646	Schwarz criterion		-2.326820
Log likelihood	33.50675	Hannan-Quinn criter.		-2.395728
Durbin-Watson stat	2.036387			
Inverted AR Roots	.61	-.23		

Los coeficientes no son significativamente distintos de cero (al 5%)

VALIDACIÓN (1/3)

Gráfico de los residuos del modelo AR(1)



- En 2020 se produjo la crisis de Covid, que provocó una fuerte caída del stock bruto de capital fijo
- Sin embargo, el valor extremadamente atípico entre los residuos es el de 2021
- Probablemente la razón es que estamos trabajando con primeras diferencias y un AR(1), y hay que tener en cuenta que en 2021 hubo un «rebote» excepcional desde la caída de 2020

















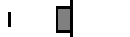

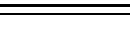
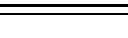
VALIDACIÓN (2/3)

Correlograma de los residuos del modelo AR(1)

Date: 03/17/25 Time: 17:38

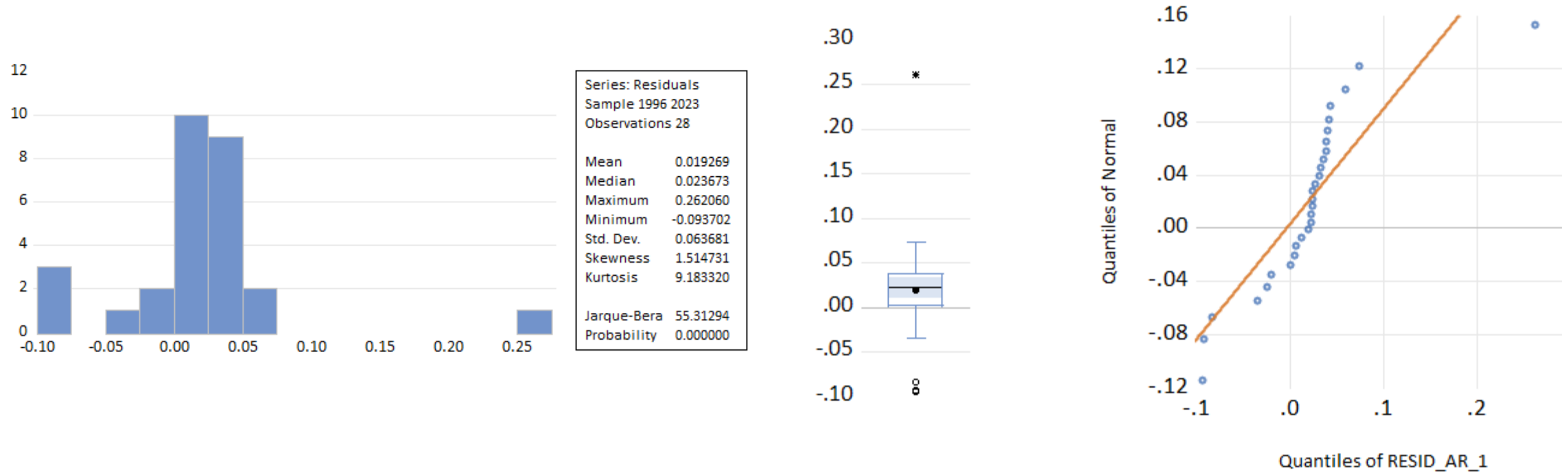
Sample (adjusted): 1996 2023

Q-statistic probabilities adjusted for 1 ARMA term

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
		1 -0.131	-0.131	0.5310	
		2 -0.009	-0.027	0.5337	0.465
		3 0.033	0.028	0.5695	0.752
		4 0.105	0.115	0.9540	0.812
		5 0.007	0.038	0.9557	0.916
		6 0.075	0.086	1.1692	0.948
		7 -0.014	0.001	1.1766	0.978
		8 -0.024	-0.039	1.2006	0.991
		9 -0.283	-0.317	4.7406	0.785
		10 -0.084	-0.222	5.0663	0.828

- Los residuos son coherentes con la hipótesis de independencia
- Todas las correlaciones están dentro de las bandas de confianza
- Todas las estadísticas Q tienen asociados un valor p superior a 0,05

VALIDACIÓN (3/3)



Conclusiones:

- Normalidad: Los residuos no son normales, el test de JB es superior a 6
- Atípicos: Tanto el boxplot como el gráfico QQ reflejan la existencia de atípicos
- En particular, el boxplot confirma que el mayor valor atípico es positivo (es decir, el de 2021)

No se valida el modelo, porque los residuos no tienen distribución normal y haya típicos relevantes

MODELOS CON CORRECCIÓN DE ATÍPICOS – ATÍPICO 2020

AR(1) + atípico 2020

Dependent Variable: D(LOG(X),1)
Method: ARMA Conditional Least Squares (Gauss-Newton / Marquardt steps)
Date: 03/17/25 Time: 18:29
Sample (adjusted): 1997 2023
Included observations: 27 after adjustments
Convergence achieved after 4 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(I2020)	-0.134293	0.020366	-6.593936	0.0000
AR(1)	0.747493	0.143070	5.224653	0.0000
R-squared	0.598691	Mean dependent var	0.031237	
Adjusted R-squared	0.582639	S.D. dependent var	0.067694	
S.E. of regression	0.043733	Akaike info criterion	-3.350246	
Sum squared resid	0.047814	Schwarz criterion	-3.254258	
Log likelihood	47.22833	Hannan-Quinn criter.	-3.321704	
Durbin-Watson stat	2.024723			
Inverted AR Roots	.75			

- Los dos coeficientes son significativos al 5%
- La desviación típica residual ha pasado de 0,067 a 0,043.

AR(2) + atípico 2020

Dependent Variable: D(LOG(X),1)
Method: ARMA Conditional Least Squares (Gauss-Newton / Marquardt steps)
Date: 03/17/25 Time: 18:31
Sample (adjusted): 1998 2023
Included observations: 26 after adjustments
Convergence achieved after 7 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(I2020)	-0.134746	0.021612	-6.234861	0.0000
AR(1)	0.733047	0.209493	3.499142	0.0019
AR(2)	0.015887	0.229972	0.069082	0.9455
R-squared	0.598923	Mean dependent var	0.030788	
Adjusted R-squared	0.564047	S.D. dependent var	0.068994	
S.E. of regression	0.045554	Akaike info criterion	-3.231654	
Sum squared resid	0.047730	Schwarz criterion	-3.086489	
Log likelihood	45.01150	Hannan-Quinn criter.	-3.189852	
Durbin-Watson stat	2.001382			
Inverted AR Roots	.75	-.02		

- El coeficiente del primer retardo si es significativo
- El coeficiente del segundo retardo no es significativo

MODELOS CON CORRECCIÓN DE ATÍPICOS – ATÍPICO 2020

Cuadro comparativo de los estadísticos más relevantes

	Akaike (AIC)	Schwarz (BIC)	Log-likelihood	Standard error
AR(1) + I2020	-3,35	-3,25	47,23	0,044
AR(2) + I2020	-3,23	-3,09	45,01	0,045

- Elegimos claramente el primer modelo porque tiene todos los coeficientes significativos
- Además, el análisis estadístico muestra que este modelo presenta:
 - Valores inferiores en los criterios de información (AIC y BIC)
 - El valor más alto para la función de log-verosimilitud
 - Una menor desviación estándar de los residuos

MODELOS CON CORRECCIÓN DE ATÍPICOS – ATÍPICO 2020

Gráfico de los residuos del modelo AR(1) + atípico 2020

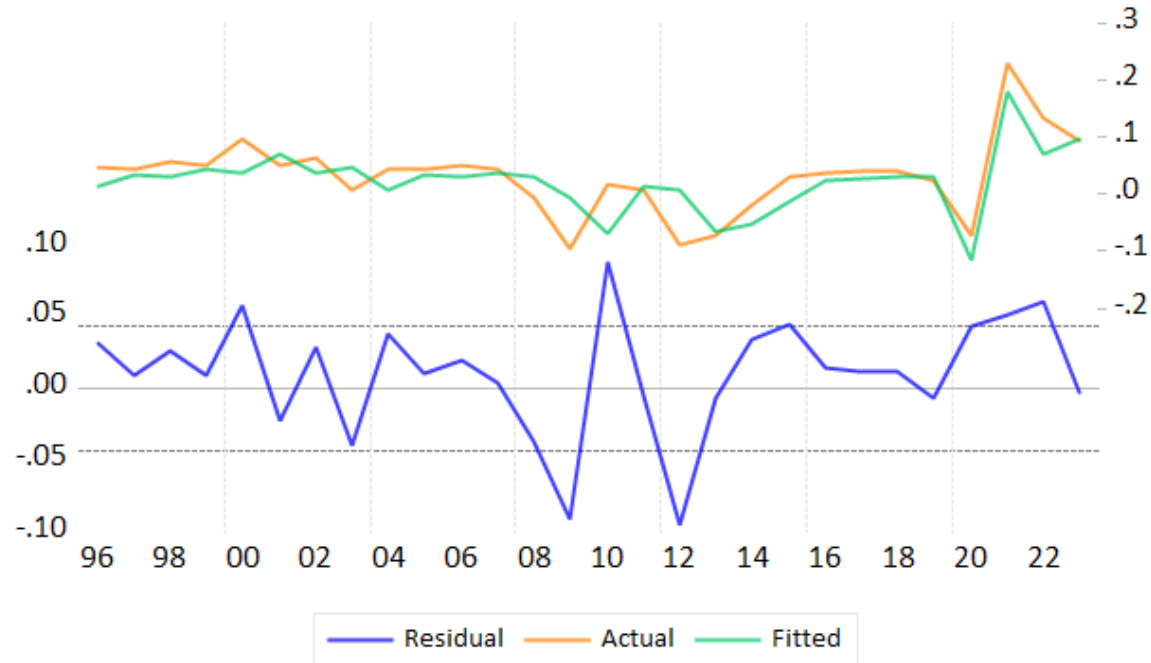


Tabla de datos

obs	Actual	Fitted	Residual	Residual Plot
1997	0.04292	0.03374	0.00918	
1998	0.05750	0.03152	0.02598	
1999	0.05152	0.04222	0.00929	
2000	0.09657	0.03783	0.05874	
2001	0.04845	0.07092	-0.02247	
2002	0.06406	0.03558	0.02848	
2003	0.00735	0.04704	-0.03969	
2004	0.04413	0.00540	0.03874	
2005	0.04280	0.03241	0.01039	
2006	0.05126	0.03143	0.01983	
2007	0.04208	0.03764	0.00444	
2008	-0.00638	0.03090	-0.03728	
2009	-0.09647	-0.00468	-0.09178	
2010	0.01732	-0.07084	0.08816	
2011	0.00764	0.01272	-0.00508	
2012	-0.08922	0.00561	-0.09483	
2013	-0.07258	-0.06552	-0.00706	
2014	-0.01868	-0.05330	0.03461	
2015	0.03084	-0.01372	0.04455	
2016	0.03706	0.02264	0.01441	
2017	0.03939	0.02721	0.01217	
2018	0.04098	0.02892	0.01205	
2019	0.02339	0.03009	-0.00670	
2020	-0.07321	-0.11717	0.04396	
2021	0.23031	0.17925	0.05106	
2022	0.13119	0.07046	0.06072	
2023	0.09318	0.09634	-0.00316	

- Ahora los mayores valores atípicos parecen ser los relacionados con la crisis de 2008-2012
- En concreto, muestran la famosa forma de “W” de la economía del sur de Europa durante los años 2008-2014
- El residuo de 2012 es el mayor: es 2,17 veces la desviación típica de los residuos, que aún no es mucho

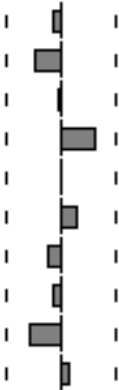
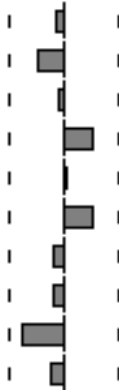
MODELOS CON CORRECCIÓN DE ATÍPICOS – ATÍPICO 2020

Correlograma de los residuos del modelo AR(1) + atípico 2020

Date: 03/17/25 Time: 18:09

Sample (adjusted): 1996 2023

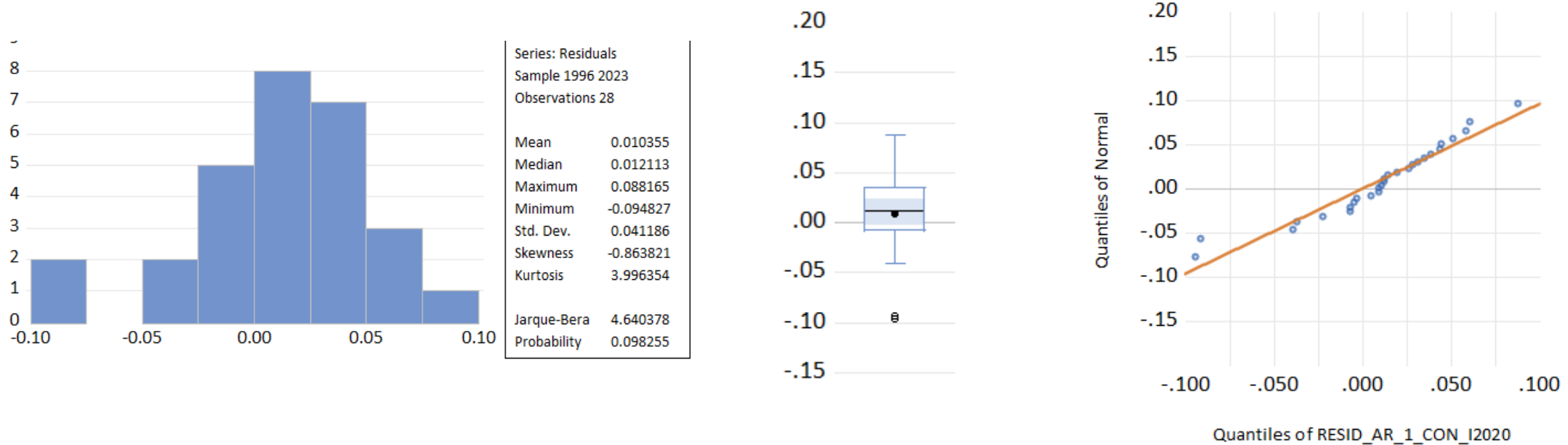
Q-statistic probabilities adjusted for 1 ARMA term and 1 dynamic regressor

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob*
		1 -0.055	-0.055	0.0949	
		2 -0.174	-0.178	1.0755	0.300
		3 -0.024	-0.047	1.0950	0.578
		4 0.226	0.198	2.8893	0.409
		5 -0.006	0.011	2.8906	0.576
		6 0.114	0.196	3.3901	0.640
		7 -0.095	-0.070	3.7540	0.710
		8 -0.053	-0.067	3.8742	0.794
		9 -0.226	-0.292	6.1317	0.632
		10 0.046	-0.098	6.2305	0.717

*Probabilities may not be valid for this equation specification.

- Los residuos son coherentes con la hipótesis de independencia
- Todas las correlaciones están dentro de las bandas de confianza
- La ultima Q-stat tiene asociado un p-valor superior a 0,05

MODELOS CON CORRECCIÓN DE ATÍPICOS – ATÍPICO 2020



Conclusiones:

- Normalidad: Según el test de JB, los residuos siguen una distribución normal, ya que el test JB es menor que 6
- Atípicos: Tanto el boxplot como el gráfico QQ reflejan la existencia de 2 valores atípicos
- En particular, el boxplot confirma que los mayores valores atípicos son las caídas de la crisis de 2008-2012
- Hay que decir, sin embargo, que no superan 3 veces la desviación típica, por lo que podríamos no considerarlos valores atípicos
- Sin embargo, veamos qué ocurre si añadimos otra variable dummy referida a esos años de crisis

MODELOS CON CORRECCIÓN DE ATÍPICOS – ATÍPICO 2009 Y ATÍPICO 2020

Después de varios intentos, el dummy I2009 es el que más mejora el modelo. Probablemente esté relacionado con el valor atípico de 2008 por las mismas razones que el valor atípico de 2021 está relacionado con la crisis de 2020.

AR(1) + atípicos 2009 y 2020

Dependent Variable: D(LOG(X),1)				
Method: ARMA Conditional Least Squares (Gauss-Newton / Marquardt steps)				
Date: 03/17/25 Time: 18:35				
Sample (adjusted): 1997 2023				
Included observations: 27 after adjustments				
Convergence achieved after 5 iterations				
Coefficient covariance computed using outer product of gradients				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(I2009)	-0.054586	0.016647	-3.279005	0.0032
D(I2020)	-0.133999	0.016619	-8.063010	0.0000
AR(1)	0.831423	0.126765	6.558754	0.0000
R-squared	0.719345	Mean dependent var	0.031237	
Adjusted R-squared	0.695958	S.D. dependent var	0.067694	
S.E. of regression	0.037327	Akaike info criterion	-3.633778	
Sum squared resid	0.033439	Schwarz criterion	-3.489796	
Log likelihood	52.05601	Hannan-Quinn criter.	-3.590965	
Durbin-Watson stat	2.071302			
Inverted AR Roots	.83			

- Los tres coeficientes son significativos al 5%
- La desviación típica residual ha pasado de 0,043 a 0,037

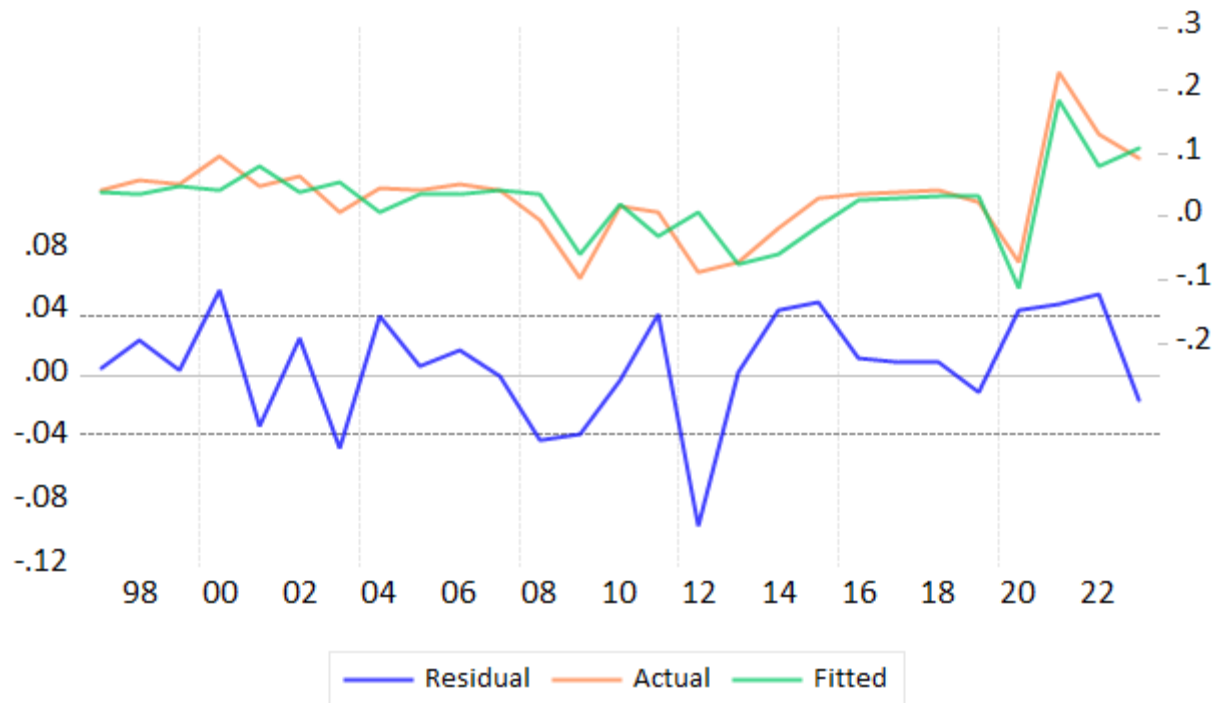
AR(2) + atípicos 2009 y 2020

Dependent Variable: D(LOG(X),1)				
Method: ARMA Conditional Least Squares (Gauss-Newton / Marquardt steps)				
Date: 03/17/25 Time: 18:36				
Sample (adjusted): 1998 2023				
Included observations: 26 after adjustments				
Convergence achieved after 15 iterations				
Coefficient covariance computed using outer product of gradients				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(I2009)	-0.057319	0.018139	-3.159926	0.0045
D(I2020)	-0.136222	0.018374	-7.413993	0.0000
AR(1)	0.770364	0.215979	3.566851	0.0017
AR(2)	0.081609	0.238953	0.341528	0.7359
R-squared	0.720206	Mean dependent var	0.030788	
Adjusted R-squared	0.682052	S.D. dependent var	0.068994	
S.E. of regression	0.038903	Akaike info criterion	-3.514828	
Sum squared resid	0.033297	Schwarz criterion	-3.321275	
Log likelihood	49.69276	Hannan-Quinn criter.	-3.459091	
Durbin-Watson stat	1.973593			
Inverted AR Roots	.86	-.09		

- El coeficiente del primer retardo si es significativo
- El coeficiente del segundo retardo no es significativo

MODELOS CON CORRECCIÓN DE ATÍPICOS – ATÍPICO 2009 Y ATÍPICO 2020

Gráfico de los residuos del modelo AR(1) + atípicos 2009 y 2020



- Ahora el modelo parece ligeramente mejorado (es decir, los residuos presentan un comportamiento aleatorio, con media constante)
- Sin embargo, se mantiene el valor atípico de 2012, igual a 2,56 veces la desviación típica de los residuos

Correlograma de los residuos

Date: 03/17/25 Time: 18:40

Sample (adjusted): 1997 2023

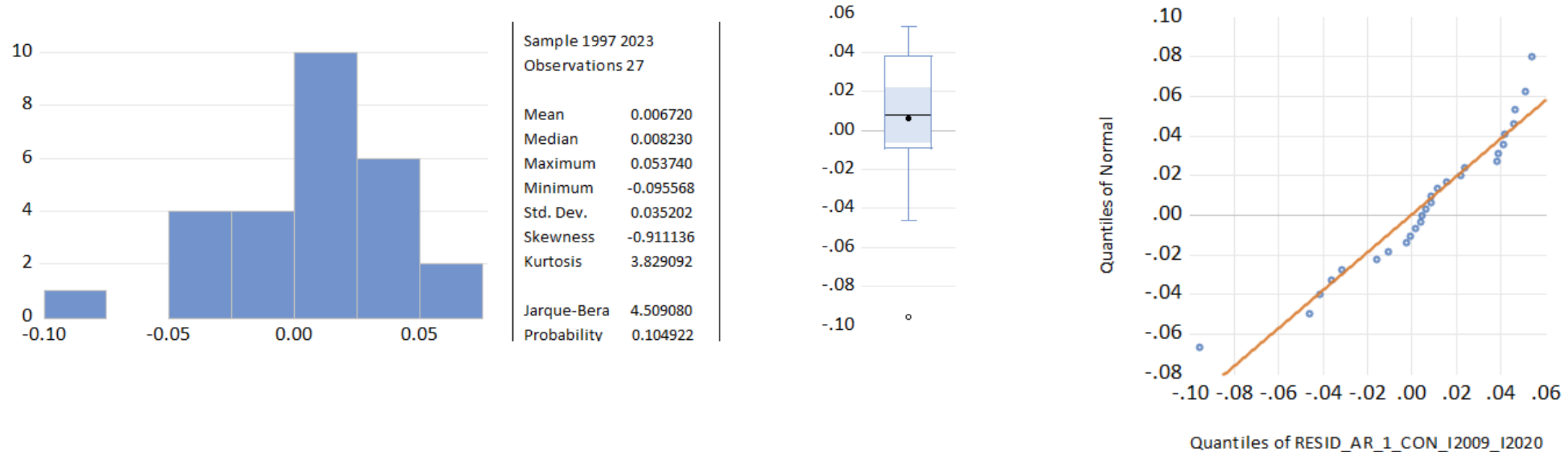
Q-statistic probabilities adjusted for 1 ARMA term and 2 dynamic regressors

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob*
		1 -0.083	-0.083	0.2067	
		2 0.022	0.015	0.2212	0.638
		3 -0.212	-0.210	1.6836	0.431
		4 0.198	0.173	3.0204	0.388
		5 0.003	0.032	3.0208	0.554
		6 0.032	-0.012	3.0590	0.691
		7 0.138	0.233	3.8099	0.702
		8 -0.292	-0.348	7.3199	0.396
		9 0.015	0.012	7.3303	0.501
		10 -0.169	-0.123	8.6408	0.471

*Probabilities may not be valid for this equation specification.

- Los residuos son coherentes con la hipótesis de independencia
- Todas las correlaciones están dentro de las bandas de confianza
- La ultima Q-stat tiene asociado un p-valor superior a 0,05

MODELOS CON CORRECCIÓN DE ATÍPICOS – ATÍPICO 2009 Y ATÍPICO 2020



Conclusiones:

- Normalidad: Los residuos siguen una distribución normal, ya que el test JB es menor que 6
- Atípicos: Tanto el boxplot como el gráfico QQ reflejan la existencia de un valor atípico
- En particular, el boxplot confirma que el mayor valor atípico es negativo (es decir, el de 2012)

MODELOS CON CORRECCIÓN DE ATÍPICOS – ATÍPICO 2009 Y ATÍPICO 2020

Cuadro comparativo de los estadísticos más relevantes

	Akaike (AIC)	Schwarz (BIC)	Log-likelihood	Standard error
AR(1) + I2020	-3,35	-3,25	47,23	0,044
AR(2) + I2020	-3,23	-3,09	45,01	0,045
AR(1) + I2009 + I2020	-3,63	-3,49	52,05	0,037
AR(2) + I2009 + I2020	-3,51	-3,32	49,7	0,04

- Elegimos el tercero modelo porque tiene todos los coeficientes significativos y porque, analizando estas estadísticas, tiene:
 - Criterios de información más bajos
 - Mayor valor de la función de verosimilitud
 - Menor error estándar (desviación típica residual)

REESTIMACIÓN – MODELO CON 3 DUMMY

¿Qué ocurre si intentamos tratar ese valor atípico de 2012 añadiendo otra variable dummy?

MODELOS CON CORRECCIÓN DE ATÍPICOS – ATÍPICOS 2009, 2011 y 2020

AR(1) + atípicos 2009, 2011 y 2020

Dependent Variable: D(LOG(X),1)
Method: ARMA Conditional Least Squares (Gauss-Newton / Marquardt steps)

Date: 03/18/25 Time: 09:34

Sample (adjusted): 1997 2023

Included observations: 27 after adjustments

Convergence achieved after 5 iterations

Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(I2009)	-0.061832	0.014077	-4.392392	0.0002
D(I2011)	0.043771	0.014076	3.109740	0.0049
D(I2020)	-0.133878	0.013875	-9.649170	0.0000
AR(1)	0.889759	0.112016	7.943125	0.0000
R-squared	0.800725	Mean dependent var	0.031237	
Adjusted R-squared	0.774732	S.D. dependent var	0.067694	
S.E. of regression	0.032129	Akaike info criterion	-3.902141	
Sum squared resid	0.023743	Schwarz criterion	-3.710165	
Log likelihood	56.67890	Hannan-Quinn criter.	-3.845056	
Durbin-Watson stat	1.979069			
Inverted AR Roots	.89			

- Los cuatro coeficientes son significativos al 5%
- La desviación típica residual ha pasado de 0,037 a 0,032

AR(2) + atípicos 2009, 2011 y 2020

Dependent Variable: D(LOG(X),1)
Method: ARMA Conditional Least Squares (Gauss-Newton / Marquardt steps)

Date: 03/18/25 Time: 09:36

Sample (adjusted): 1998 2023

Included observations: 26 after adjustments

Convergence achieved after 11 iterations

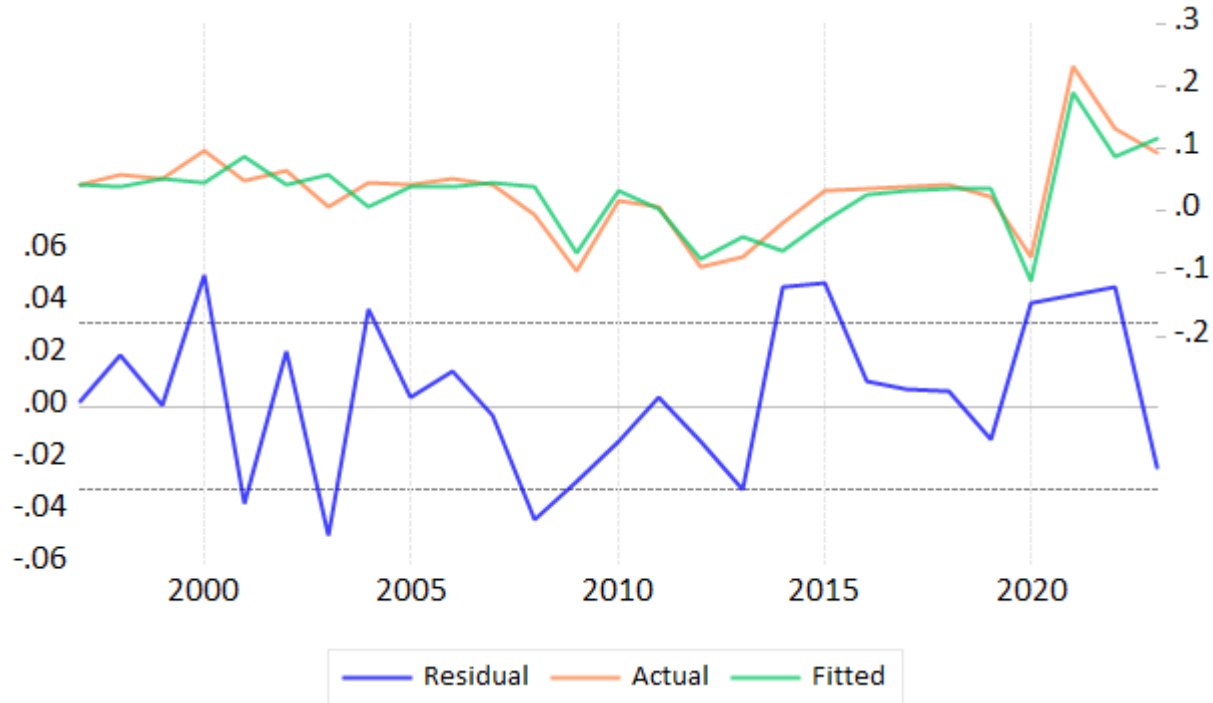
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(I2009)	-0.061833	0.014742	-4.194399	0.0004
D(I2011)	0.043774	0.014803	2.957018	0.0075
D(I2020)	-0.133883	0.014754	-9.074141	0.0000
AR(1)	0.888499	0.225105	3.947046	0.0007
AR(2)	0.000118	0.250762	0.000472	0.9996
R-squared	0.800523	Mean dependent var	0.030788	
Adjusted R-squared	0.762527	S.D. dependent var	0.068994	
S.E. of regression	0.033621	Akaike info criterion	-3.776262	
Sum squared resid	0.023738	Schwarz criterion	-3.534320	
Log likelihood	54.09140	Hannan-Quinn criter.	-3.706591	
Durbin-Watson stat	1.964469			
Inverted AR Roots	.89	-.00		

- El coeficiente del primer retardo si es significativo
- El coeficiente del segundo retardo no es significativo

MODELOS CON CORRECCIÓN DE ATÍPICOS – ATÍPICOS 2009, 2011 Y 2020

Gráfico de los residuos del modelo AR(1) + atípicos 2009, 2011 Y 2020



- Se observa un buen ajuste entre la serie real y la estimada y como los residuos del modelo elegido presentan un comportamiento aleatorio, con media constante
- Hay algunos valores fuera del intervalo, pero todos son inferiores a 2 veces la desviación típica

Correlograma de los residuos

Date: 03/17/25 Time: 18:40

Sample (adjusted): 1997 2023

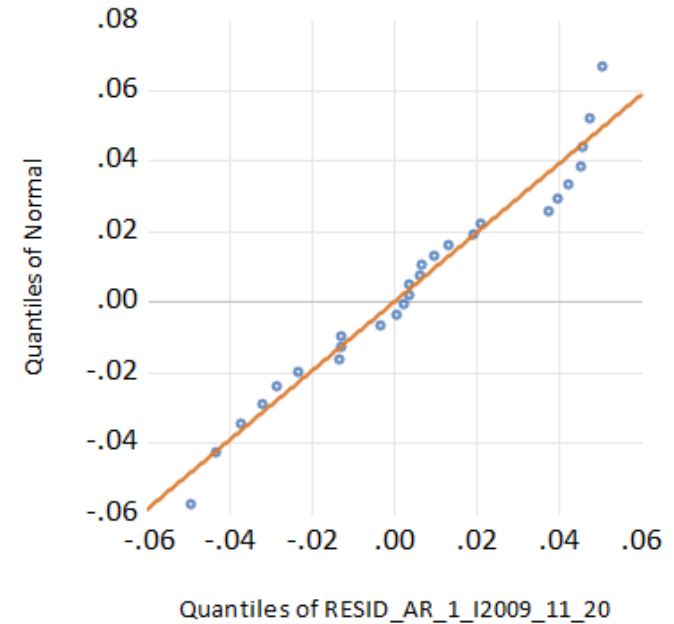
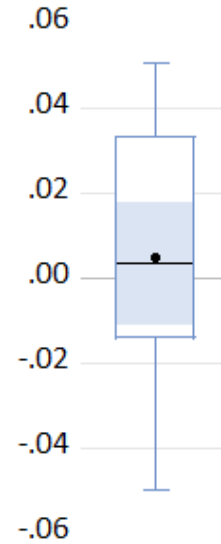
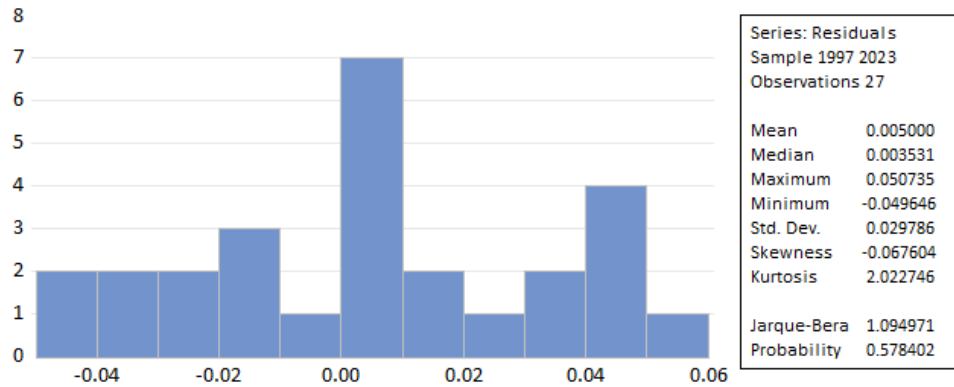
Q-statistic probabilities adjusted for 1 ARMA term and 2 dynamic regressors

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob*
		1 -0.083	-0.083	0.2067	
		2 0.022	0.015	0.2212	0.638
		3 -0.212	-0.210	1.6836	0.431
		4 0.198	0.173	3.0204	0.388
		5 0.003	0.032	3.0208	0.554
		6 0.032	-0.012	3.0590	0.691
		7 0.138	0.233	3.8099	0.702
		8 -0.292	-0.348	7.3199	0.396
		9 0.015	0.012	7.3303	0.501
		10 -0.169	-0.123	8.6408	0.471

*Probabilities may not be valid for this equation specification.

- Los residuos son coherentes con la hipótesis de independencia
- Todas las correlaciones están dentro de las bandas de confianza
- La ultima Q-stat tiene asociado un p-valor superior a 0,05

MODELOS CON CORRECCIÓN DE ATÍPICOS – ATÍPICOS 2009, 2011 Y 2020



Conclusiones:

- Normalidad: Los residuos siguen una distribución normal, ya que el test JB es menor que 6
- Atípicos: No hay atípicos relevantes

MODELOS CON CORRECCIÓN DE ATÍPICOS – ATÍPICO 2009 Y ATÍPICO 2020

Cuadro comparativo de los estadísticos más relevantes

	Akaike (AIC)	Schwarz (BIC)	Log-likelihood	Standard error
AR(1) + I2020	-3,35	-3,25	47,23	0,044
AR(2) + I2020	-3,23	-3,09	45,01	0,045
AR(1) + I2009 + I2020	-3,63	-3,49	52,05	0,037
AR(2) + I2009 + I2020	-3,51	-3,32	49,7	0,04
AR(1) + I2009 + I2011 + I2020	-3,9	-3,71	56,68	0,032
AR(2) + I2009 + I2011 + I2020	-3,77	-3,53	54,09	0,033

- Elegimos el quinto modelo porque tiene todos los coeficientes significativos y porque, analizando estas estadísticas, tiene:
 - Criterios de información más bajos
 - Mayor valor de la función de verosimilitud
 - Menor error estándar (desviación típica residual)

REESTIMACIÓN – MODELO CON CORRECCIÓN DE ATÍPICOS Y CONSTANTE

Ahora estimamos de nuevo el último modelo elegido incluyendo la constante, ya que en los gráficos iniciales podría haber existido una media distinta de cero

MODELOS CON CORRECCIÓN DE ATÍPICOS Y CONSTANTE

AR(1) + c + atípicos 2009, 2011 y 2020

Dependent Variable: D(LOG(X),1)
 Method: ARMA Conditional Least Squares (Gauss-Newton / Marquardt steps)
 Date: 03/18/25 Time: 09:46
 Sample (adjusted): 1997 2023
 Included observations: 27 after adjustments
 Convergence achieved after 5 iterations
 Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.039327	0.035608	1.104443	0.2813
D(I2009)	-0.061797	0.014614	-4.228582	0.0003
D(I2011)	0.043804	0.014610	2.998118	0.0066
D(I2020)	-0.134024	0.014403	-9.304933	0.0000
AR(1)	0.822185	0.132799	6.191185	0.0000
R-squared	0.808648	Mean dependent var	0.031237	
Adjusted R-squared	0.773857	S.D. dependent var	0.067694	
S.E. of regression	0.032192	Akaike info criterion	-3.868640	
Sum squared resid	0.022799	Schwarz criterion	-3.628670	
Log likelihood	57.22664	Hannan-Quinn criter.	-3.797285	
F-statistic	23.24285	Durbin-Watson stat	1.931538	
Prob(F-statistic)	0.000000			
Inverted AR Roots	.82			

- Los cuatro coeficientes son significativos al 5%, pero la constante no
- La desviación típica residual es la misma que el modelo sin constante

AR(2) + c + atípicos 2009, 2011 y 2020

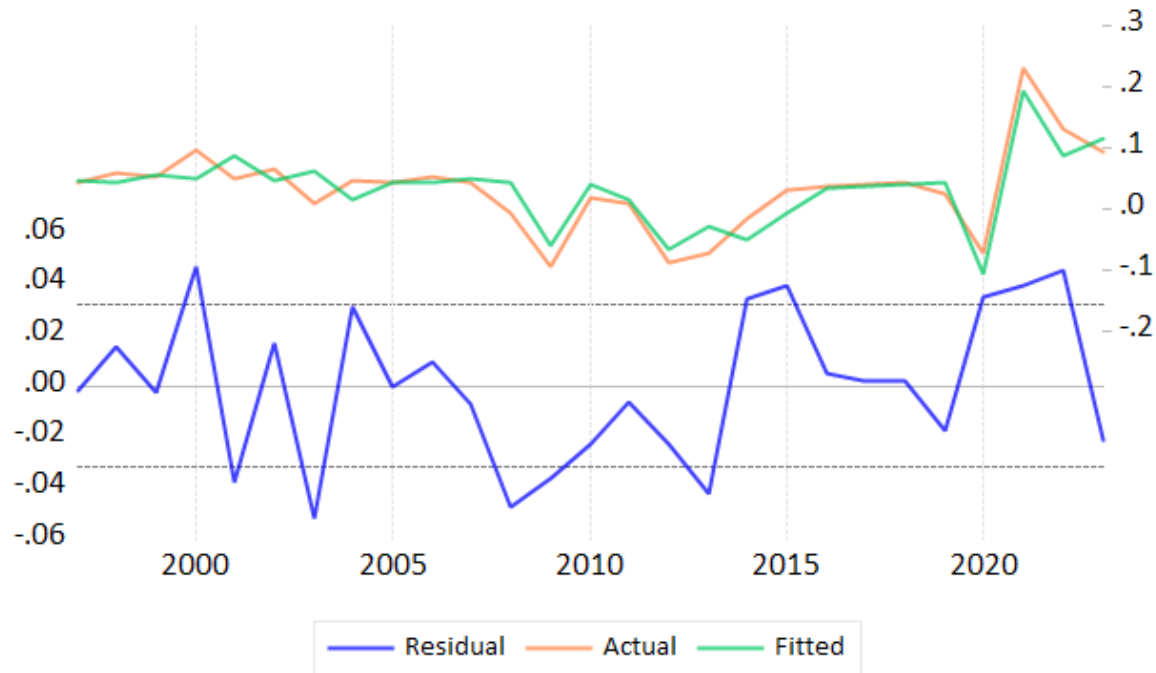
Dependent Variable: D(LOG(X),1)
 Method: ARMA Conditional Least Squares (Gauss-Newton / Marquardt steps)
 Date: 03/18/25 Time: 09:47
 Sample (adjusted): 1998 2023
 Included observations: 26 after adjustments
 Convergence achieved after 10 iterations
 Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.038239	0.036364	1.051585	0.3055
D(I2009)	-0.061348	0.014997	-4.090782	0.0006
D(I2011)	0.043182	0.015050	2.869203	0.0095
D(I2020)	-0.132940	0.014922	-8.909263	0.0000
AR(1)	0.855468	0.229519	3.727224	0.0013
AR(2)	-0.045535	0.254283	-0.179072	0.8597
R-squared	0.808698	Mean dependent var	0.030788	
Adjusted R-squared	0.760873	S.D. dependent var	0.068994	
S.E. of regression	0.033738	Akaike info criterion	-3.741184	
Sum squared resid	0.022766	Schwarz criterion	-3.450854	
Log likelihood	54.63540	Hannan-Quinn criter.	-3.657580	
F-statistic	16.90936	Durbin-Watson stat	1.987505	
Prob(F-statistic)	0.000001			
Inverted AR Roots	.80	.06		

- El coeficiente del primer retardo si es significativo
- El coeficiente del segundo retardo no es significativo, y la constante tampoco

MODELOS CON CORRECCIÓN DE ATÍPICOS Y CONSTANTE

Gráfico de los residuos del modelo AR(1) + c + atípicos 2009, 2011 Y 2020



- Se observa un buen ajuste entre la serie real y la estimada y como los residuos del modelo elegido presentan un comportamiento aleatorio, con media constante
- Hay algunos valores fuera del intervalo, pero todos son inferiores a 2 veces la desviación típica

Correlograma de los residuos

Date: 03/18/25 Time: 09:49

Sample (adjusted): 1997 2023

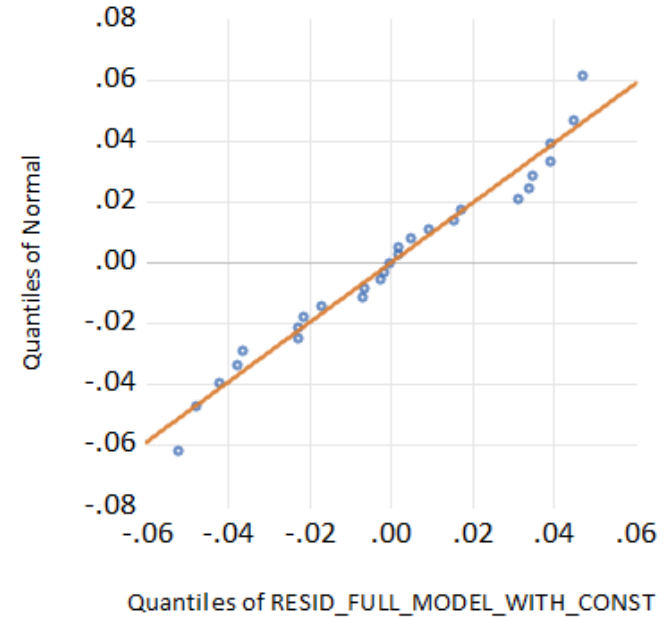
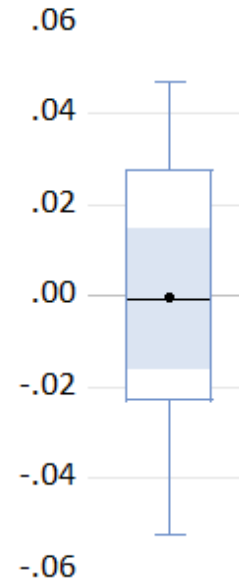
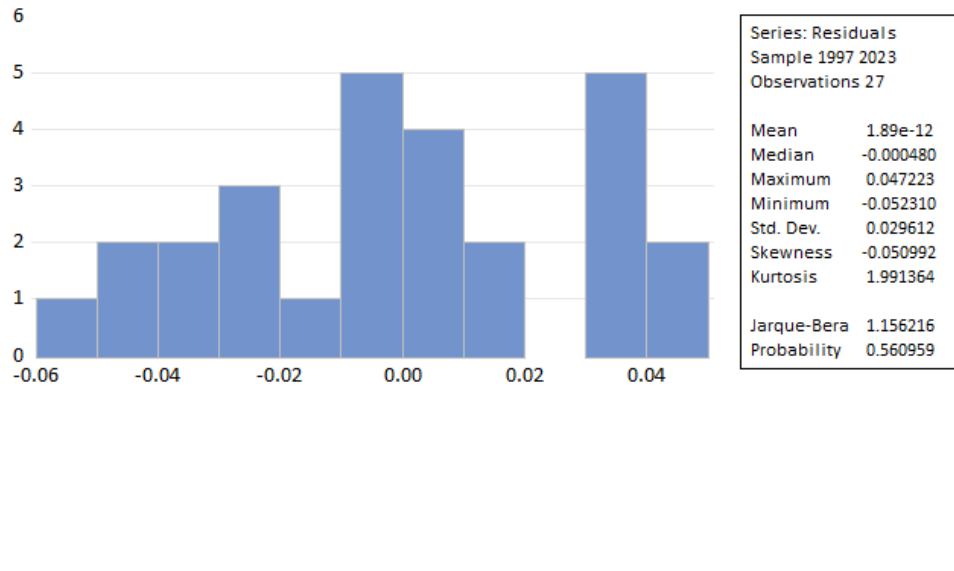
Q-statistic probabilities adjusted for 1 ARMA term and 3 dynamic regressors

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob*
		1 0.024	0.024	0.0171	
		2 0.189	0.188	1.1314	0.287
		3 -0.253	-0.271	3.2154	0.200
		4 0.101	0.100	3.5657	0.312
		5 0.037	0.137	3.6136	0.461
		6 0.093	-0.031	3.9384	0.558
		7 0.043	0.065	4.0104	0.675
		8 -0.150	-0.157	4.9324	0.668
		9 -0.188	-0.212	6.4701	0.595
		10 0.036	0.180	6.5311	0.686

*Probabilities may not be valid for this equation specification.

- Los residuos son coherentes con la hipótesis de independencia
- Todas las correlaciones están dentro de las bandas de confianza
- La última Q-stat tiene asociado un p-valor superior a 0,05

MODELOS CON CORRECCIÓN DE ATÍPICOS Y CONSTANTE



Conclusiones:

- Normalidad: Los residuos siguen una distribución normal, ya que el test JB es menor que 6
- Atípicos: No hay atípicos relevantes

CUADRO COMPARATIVO DE LOS ESTADÍSTICOS MÁS RELEVANTES

	Akaike (AIC)	Schwarz (BIC)	Log-likelihood	Standard error
AR(1) + I2020	-3,35	-3,25	47,23	0,044
AR(2) + I2020	-3,23	-3,09	45,01	0,045
AR(1) + I2009 + I2020	-3,63	-3,49	52,05	0,037
AR(2) + I2009 + I2020	-3,51	-3,32	49,7	0,04
AR(1) + I2009 + I2011 + I2020	-3,9	-3,71	56,68	0,032
AR(2) + I2009 + I2011 + I2020	-3,77	-3,53	54,09	0,033
AR(1) + I2009 + I2011 + I2020 + C	-3,86	-3,63	57,22	0,032
AR(2) + I2009 + I2011 + I2020 + C	-3,74	-3,45	54,63	0,033

- Elegimos de nuevo el quinto modelo porque tiene todos los coeficientes significativos y porque, analizando estas estadísticas, tiene:
 - Criterios de información más bajos
 - Segundo mayor valor de la función de verosimilitud (el modelo con el mayor valor tiene la constante no significativa)
 - Menor error estándar (desviación típica residual)

Predicción

PREDICCIÓN DOS PERIODOS POR DELANTE (SUBMUESTRA)

Dependent Variable: D(LOG(X),1)

Method: ARMA Conditional Least Squares (Gauss-Newton / Marquardt steps)

Date: 05/01/25 Time: 18:18

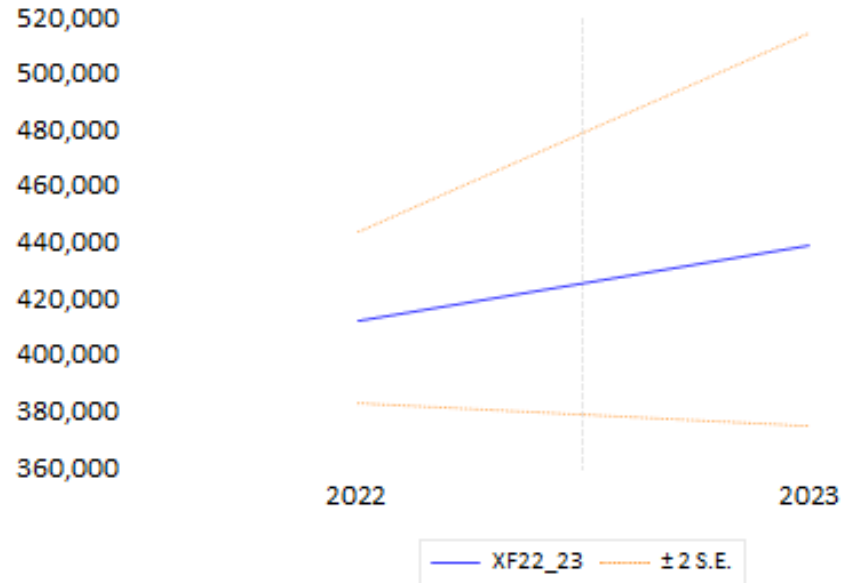
Sample (adjusted): 1997 2021

Included observations: 25 after adjustments

Convergence achieved after 5 iterations

Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(I2009)	-0.061818	0.014005	-4.413986	0.0002
D(I2011)	0.043785	0.014002	3.127110	0.0051
D(I2020)	-0.143314	0.015240	-9.403615	0.0000
AR(1)	0.856218	0.134600	6.361207	0.0000
R-squared	0.801390	Mean dependent var	0.024761	
Adjusted R-squared	0.773017	S.D. dependent var	0.065913	
S.E. of regression	0.031403	Akaike info criterion	-3.938189	
Sum squared resid	0.020709	Schwarz criterion	-3.743169	
Log likelihood	53.22736	Hannan-Quinn criter.	-3.884099	
Durbin-Watson stat	2.050647			
Inverted AR Roots	.86			



Forecast: XF22_23
Actual: X
Forecast sample: 2022 2023
Included observations: 2
Root Mean Squared Error 32711.49
Mean Absolute Error 31784.73
Mean Abs. Percent Error 6.881416
Theil Inequality Coef. 0.036994
Bias Proportion 0.944140
Variance Proportion 0.055860
Covariance Proportion -0.000000
Theil U2 Coefficient 0.927299
Symmetric MAPE 7.137030

- Dado que el último dato disponible de la serie es 2023, se estima de nuevo el modelo hasta 2021
- Se hacen predicciones para 2022 y 2023
- Como se espera en teoría, los intervalos de confianza aumentan con el horizonte de predicción

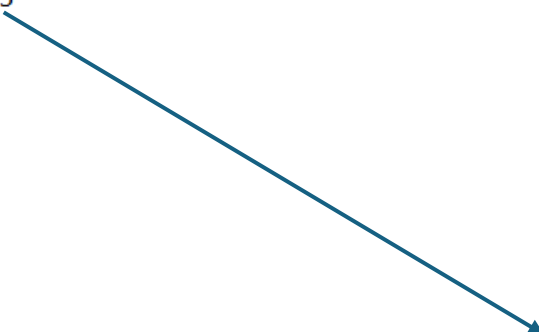
PREDICCIÓN DOS PERIODOS POR DELANTE (SUBMUESTRA)

Sin cte	Predicción	Valor real	Diferencia	Dif al cuadrado	Límite inferior	Límite superior
2022	412293,05	436346,5	24053,45	578568598,72	383560,6	443177,86
2023	439444,60	478960,6	39516,00	1561514183,37	376106,2	513449,58

- Los datos reales se sitúan dentro de los intervalos de confianza del 95%
- Ahora se puede observar que los intervalos de confianza se amplían a medida que aumenta el horizonte de predicción

Suma de
cuadrados/h
raíz

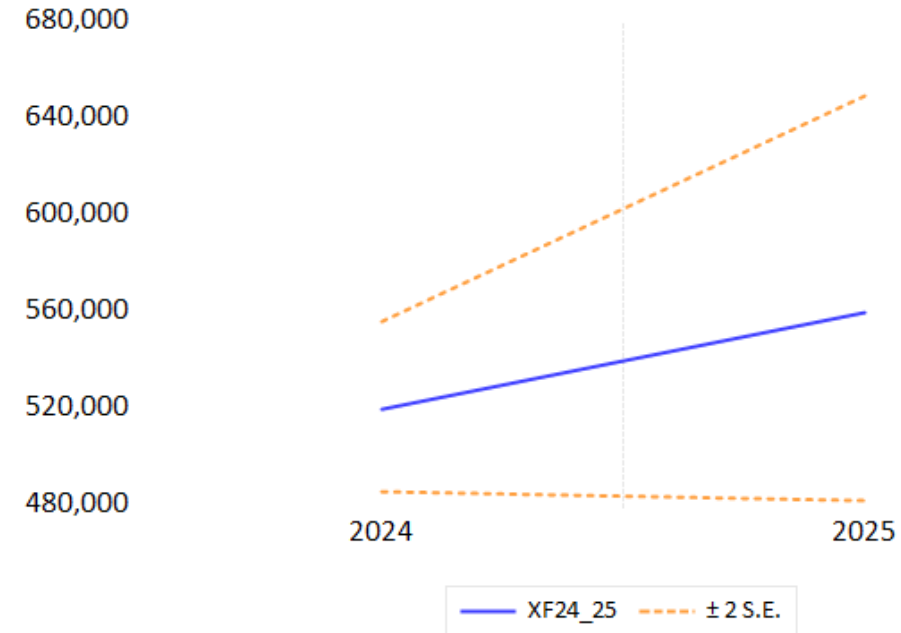
1070041391,0
32711,5


$$ECM = \sqrt{\frac{\sum_{i=1}^h (X_{t+i} - \hat{X}_{t+i})^2}{h}}$$

PREDICCIÓN DOS PERIODOS POR DELANTE (FUERA DE LA MUESTRA)

Dependent Variable: D(LOG(X),1)
Method: ARMA Conditional Least Squares (Gauss-Newton / Marquardt steps)
Date: 03/18/25 Time: 09:45
Sample (adjusted): 1997 2023
Included observations: 27 after adjustments
Convergence achieved after 5 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(I2009)	-0.061832	0.014077	-4.392392	0.0002
D(I2011)	0.043771	0.014076	3.109740	0.0049
D(I2020)	-0.133878	0.013875	-9.649170	0.0000
AR(1)	0.889759	0.112016	7.943125	0.0000
R-squared	0.800725	Mean dependent var	0.031237	
Adjusted R-squared	0.774732	S.D. dependent var	0.067694	
S.E. of regression	0.032129	Akaike info criterion	-3.902141	
Sum squared resid	0.023743	Schwarz criterion	-3.710165	
Log likelihood	56.67890	Hannan-Quinn criter.	-3.845056	
Durbin-Watson stat	1.979069			
Inverted AR Roots	.89			



- Ahora se utiliza toda la muestra disponible, es decir, 1997-2023
- Se hacen predicciones para 2024 y 2025
- Como se espera en teoría, los intervalos de confianza aumentan con el horizonte de predicción

PREDICCIÓN DOS PERIODOS POR DELANTE (FUERA DE LA MUESTRA)

Sin constante	Real	Predicción	Límite inferior	Límite superior	Tasa de crecimiento
2019	327062,5				
2020	303972,5				-7,1%
2021	382698,4				25,9%
2022	436346,5				14,0%
2023	478960,6				9,8%
2024		520363,51	487024,4	555984,8	8,6%
2025		560.201,72	484.027,55	648.363,86	7,7%

- Los datos reales se sitúan dentro de los intervalos de confianza del 95%
- Ahora se puede observar que los intervalos de confianza se amplían a medida que aumenta el horizonte de predicción

PREDICCIÓN DOS PERIODOS POR DELANTE (FUERA DE LA MUESTRA)

