

Pronosticar PIB Trimestral en Honduras

Contents

Modelos Univariados	2
1) Importar los datos ajustados estacionalmente	2
2) Inspección de los datos: PIB_HN	2
3) Pruebas de raíz unitaria	3
4) Modelos de pronóstico ARIMA	5
5) Extensión del modelo	12
6) Comparación de modelos	16
7) Resultado de pronóstico dentro de muestra	18
8) Fan chart de la inflación	19
Anexos	22
Anexo 1: Caminata aleatoria	22
Anexo 2: Criterio ADF	23
Anexo 3: Referencias	24

Modelos Univariados

Este trabajo tiene como objetivo desarrollar un ejercicio utilizando el PIB Trimestral como referencia para evaluar diferentes modelos ARIMA y los resultados de pronóstico dentro y fuera de muestra que se obtienen, además de todas las pruebas estadísticas pertinentes. Se observa que, este tipo de ejercicios solamente son útiles para pronosticar resultados de las series en el corto plazo, puesto que los mismos presentan una reversión a la media; por tanto, en caso de requerir un análisis de mediano plazo, se recomienda el uso de modelos determinísticos.

Conviene resaltar que se programó todo el ejercicio en R, incluyendo el presente informe; esto facilita la implementación para otras series económicas de utilidad en el Banco Central de Honduras. La referencia a los códigos básicos se tomó del curso de Macroeconomic Forecasting recibido en junio de 2019 en el Study Centre Gerzensee (Suiza). **Los resultados no representan ninguna postura institucional, solamente se publican con fines de mostrar una aplicación de pronóstico con modelos ARIMA.**

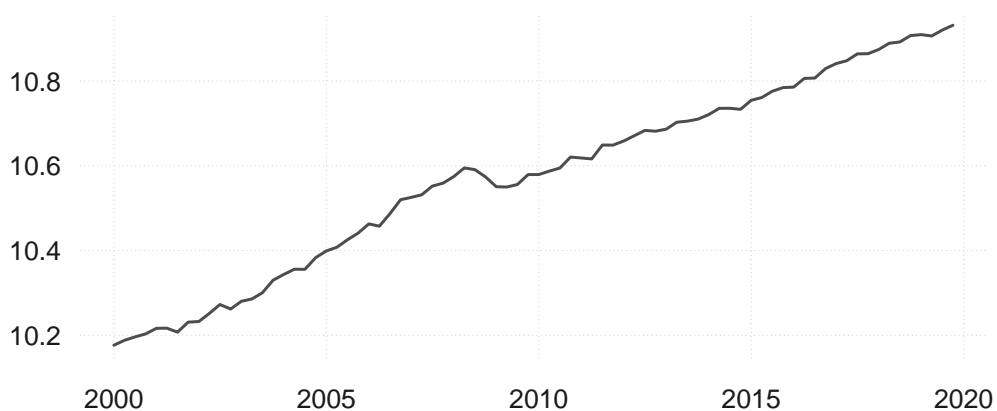
1) Importar los datos ajustados estacionalmente

Se utiliza el PIB trimestral para Honduras y el PIB de Estados Unidos de América.

2) Inspección de los datos: PIB_HN

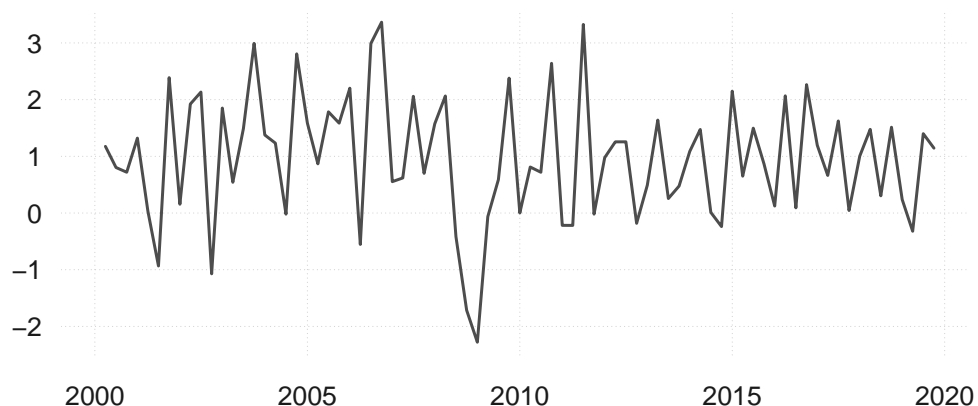
PIB Honduras (en logs)

Valores Constantes, Serie Desestacionalizada



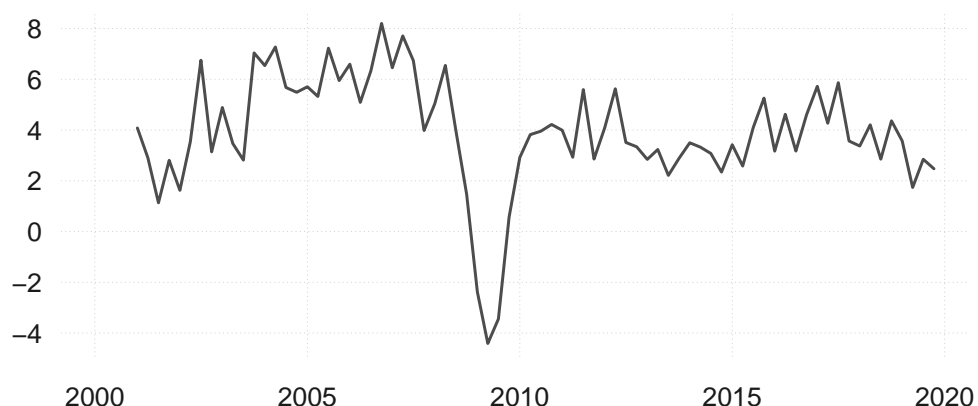
PIB Honduras

Tasa de crecimiento respecto al trimestre anterior



PIB Honduras

Tasa de crecimiento interanual



Conclusiones preliminares: podemos observar que el PIB es no estacionario. Debido a esto, estimaremos el modelo en primeras diferencias. La tasa de crecimiento trimestral (que se aproxima a través de la primera diferencia logarítmica) aparenta ser estacionaria. Pueden observarse tasas de crecimiento atípicas (outliers) en diciembre 2008 y marzo 2009, siendo mayor esta última fecha.

Regularmente, el PIB se reporta en tasas de crecimiento interanual. Al finalizar el ejercicio, se reportará el pronóstico en variaciones interanuales, calculadas a través de la estimación del modelo en primeras diferencias.

3) Pruebas de raíz unitaria

Una explicación respecto a esta prueba, basada en Pindyck & Rubinfeld (2001), se encuentra en los Anexos 1 y 2. Se realizó una prueba de Dickey Fuller Aumentada (ADF, por sus siglas en inglés)¹ para verificar si las series contienen una raíz unitaria. Debido a que las series tienen una tendencia creciente, se agrega una tendencia como un componente determinístico². Los rezagos de la regresión se seleccionan mediante el

¹Anexos 1 y 2

²Anexo 1

criterio de información bayesiano (BIC, por sus siglas en inglés).

Las pruebas sugieren que tenemos que utilizar primeras diferencias para obtener series estacionarias; esto puede verse con los resultados en dlogs (diferenciales logarítmicos), donde se verifica si las primeras diferencias de las series contienen una raíz unitaria. Con la serie transformada en dlogs, se rechaza la hipótesis nula de raíz unitaria. El PIB en niveles es claramente no estacionario, mientras que su tasa de crecimiento (dlogs) es estacionaria. La misma situación se presenta para el PIB de Estados Unidos de América (GDP).

Table 1: Resultados de Prueba CADF

transf	Nombre	Stat	Drift	None	Trend
log	GDP	est.t	-0.4468	3.5255	-1.8367
log	GDP	p.value	0.8943	0.9998	0.6760
log	GDP	sel.lag	1.0000	1.0000	1.0000
log	PIB	est.t	-1.1843	7.1234	-2.0534
log	PIB	p.value	0.6768	1.0000	0.5619
log	PIB	sel.lag	0.0000	0.0000	0.0000
dlog	GDP	est.t	-5.2652	-3.5798	-5.2276
dlog	GDP	p.value	0.0000	0.0005	0.0003
dlog	GDP	sel.lag	0.0000	0.0000	0.0000
dlog	PIB	est.t	-8.3194	-2.0236	-8.4682
dlog	PIB	p.value	0.0000	0.0420	0.0000
dlog	PIB	sel.lag	0.0000	2.0000	0.0000

4) Modelos de pronóstico ARIMA

Debido a que la serie de PIB se incorpora en primeras diferencias, con base en las pruebas de raíz unitaria, los modelos ARIMA que se especificarán en este trabajo pueden ser definidos como un proceso ARMA(p,q)³.

$$Y_t = c + \sum_{i=1}^p \phi_i Y_{t-i} + \sum_{j=0}^q \theta_j \varepsilon_{t-j} \quad (1)$$

utilizando el operador de rezagos:

$$\begin{aligned} \phi(L)Y_t &= c + \theta(L)\varepsilon_t \\ \phi(L) &= 1 - \phi_1 L - \dots - \phi_p L^p \\ \theta(L) &= \theta_0 + \theta_1 L + \dots + \theta_q L^q \end{aligned} \quad (2)$$

Los coeficientes ϕ corresponden al proceso autorregresivo y θ a los de media móvil.

a) Modelo inicial (ARIMA(1,0,0))

Nota: se modela el PIB en diferenciales logarítmicos (dlog) y se incluye una constante. En la función `Arima()`, el primer elemento es la serie que queremos pronosticar/modelar, el orden es un vector con el número de rezagos autorregresivos (p), el número de veces que se aplican diferenciales a la serie (d) y el número de rezagos de media móvil (q). Debido a que la serie ya está expresada en diferencias, se incluye $d = 0$ en todas las aplicaciones.

ARIMA(1,0,0): Evaluación de los coeficientes

Table 2: Coeficientes del ARIMA(1,0,0)

	Betas	Error_Estandar	t_Estadistic	p_value
ar1	-0.0458490	0.1117331	-0.4103439	0.6826787
intercept	0.0095568	0.0011719	8.1551247	0.0000000

Table 3: Estadísticos del ARIMA(1,0,0)

sigma_2	logLik	AIC	AICc	BIC
0.0001205	245.3616	-484.7232	-484.4032	-477.6149

Puede verse que solamente la constante es estadísticamente significativa.⁴(<https://otexts.com/fpp2/accuracy.html>).

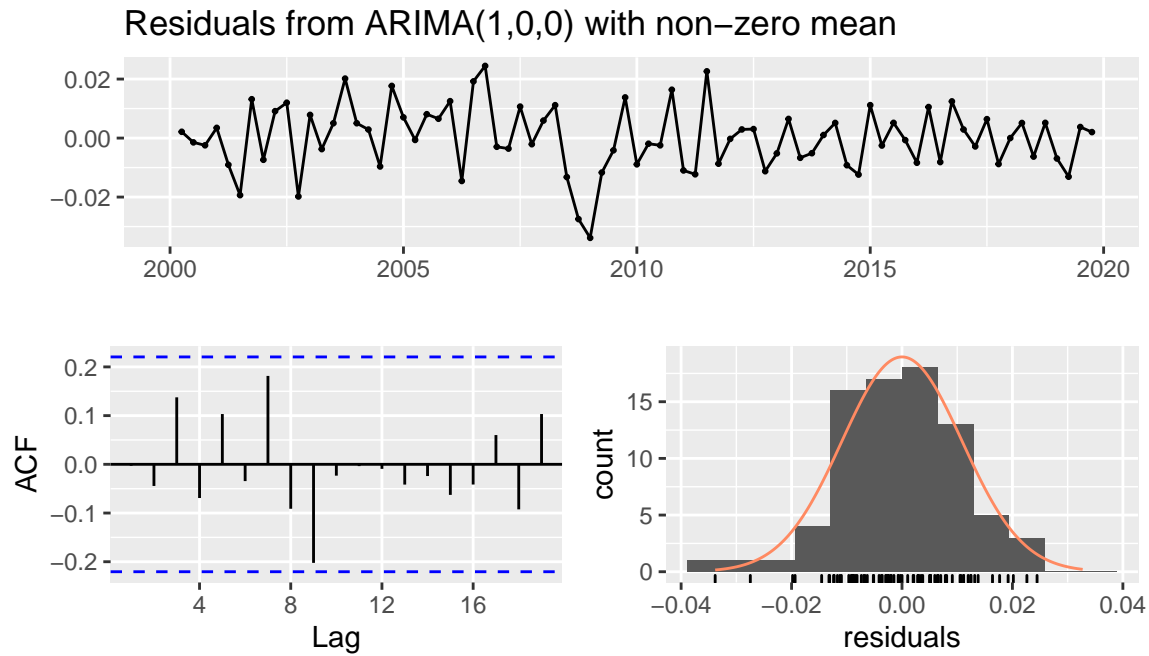
ARIMA(1,0,0): Análisis de los residuales

³Hayashi (2000)

⁴Evaluating forecast accuracy

Table 4: Estadísticos de residuales ARIMA(1,0,0): Training set error measures

ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
1e-06	0.0108369	0.0086114	-881.9949	1303.383	0.6860476	-0.0028669

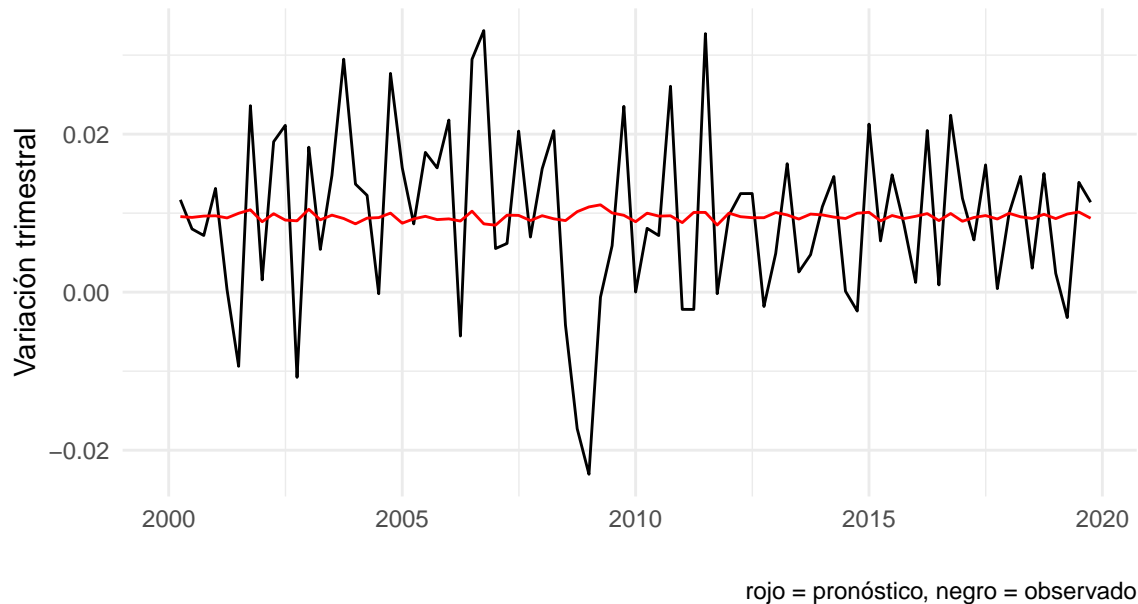


Ljung-Box test

data: Residuals from ARIMA(1,0,0) with non-zero mean
 $Q^* = 6.8691$, $df = 6$, $p\text{-value} = 0.3331$

Model df: 2. Total lags used: 8

Evaluación del modelo ARIMA(1,0,0)



Box-Ljung test

```
data: Model_ar1$residuals
X-squared = 3.087, df = 4, p-value = 0.5434
```

Jarque Bera Test

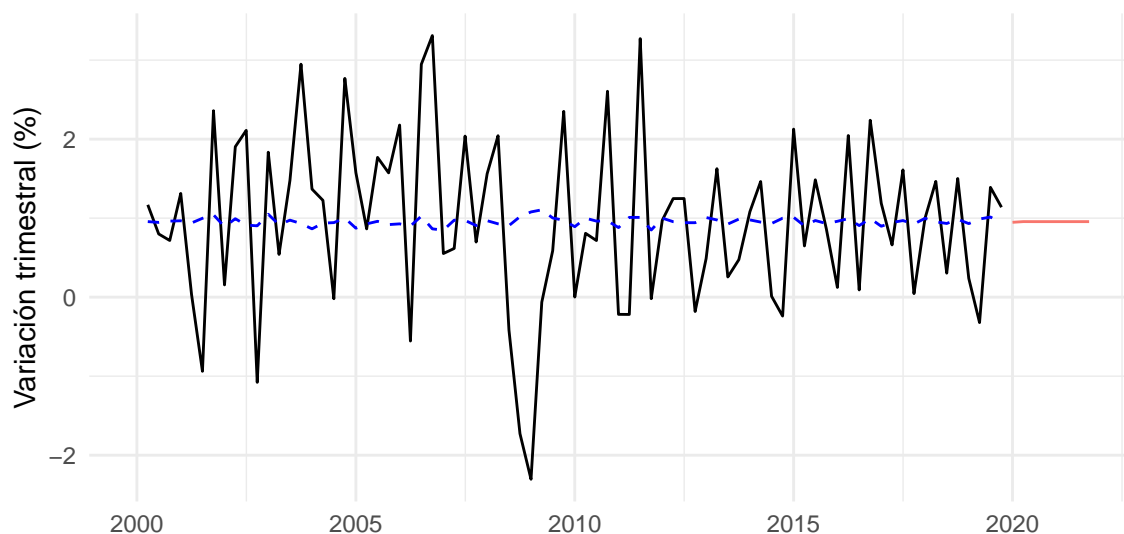
```
data: Model_ar1$residuals[!is.na(Model_ar1$residuals)]
X-squared = 1.5448, df = 2, p-value = 0.4619
```

Tanto la prueba de Ljung-Box como la de Box-Ljung muestran que los residuos del modelo ARIMA(1,0,0) no tienen autocorrelación. Adicionalmente, la prueba de Jarque-Bera implica normalidad en los residuos.

ARIMA(1,0,0): Pronóstico fuera de muestra

h corresponde al número de períodos a ser estimados (dos años).

PIB Honduras: Pronóstico ARIMA(1,0,0)

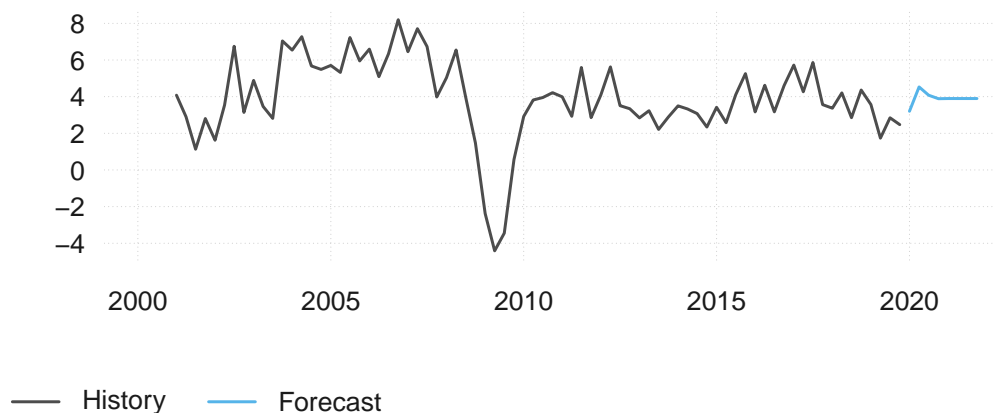


rojo = pronóstico, azul = estimado, negro = observado

Algunas veces se necesita presentar el pronóstico en variaciones interanuales. Una manera de hacerlo es transformando las series históricas y el pronóstico de las tasas de crecimiento y graficar el resultado. El modelo se estima en dlogs, luego tenemos que transformar la serie para volverla a niveles antes de calcular las tasas de crecimiento.

Honduras: PIB y Pronóstico ARIMA(1,0,0)

Variación interanual (%)



Los resultados sugieren que este modelo sencillo puede mejorarse. Los residuales muestran valores atípicos, además que el pronóstico converge rápidamente al promedio de los datos observados y no puede incorporarse fenómenos como choques aleatorios (ej: efectos del COVID-19 o una crisis interna similar a la de 2008-2009).

En los ejercicios siguientes, se trata de seleccionar el orden de los rezagos con un criterio de información y se incluyen variables exógenas adicionales que pueden ser importantes para explicar el crecimiento del PIB en Honduras.

b) Procedimiento automático para seleccionar el modelo de pronóstico

Una estrategia similar para seleccionar un modelo de pronóstico es mediante el uso de la función `auto.arima`. Esta función selecciona automáticamente el modelo de acuerdo con algún criterio. En este ejemplo se usa el AIC.⁵

ARIMA automático: Evaluación de los coeficientes

Table 5: Coeficientes del ARIMA automático

	Betas	Error_Estandar	t_Estadistic	p_value
ar1	-0.9066397	0.1134672	-7.990322	0.0000000
ar2	-0.1725819	0.1111626	-1.552518	0.1245866
ma1	0.9683044	0.0437542	22.130568	0.0000000
intercept	0.0095793	0.0010923	8.769501	0.0000000

Table 6: Estadísticos del ARIMA automático

sigma_2	logLik	AIC	AICc	BIC
0.0001093	249.6487	-489.2974	-488.4755	-477.4502

Todos los coeficientes (excepto el correspondiente al AR(2)) son estadísticamente significativos al 5%.

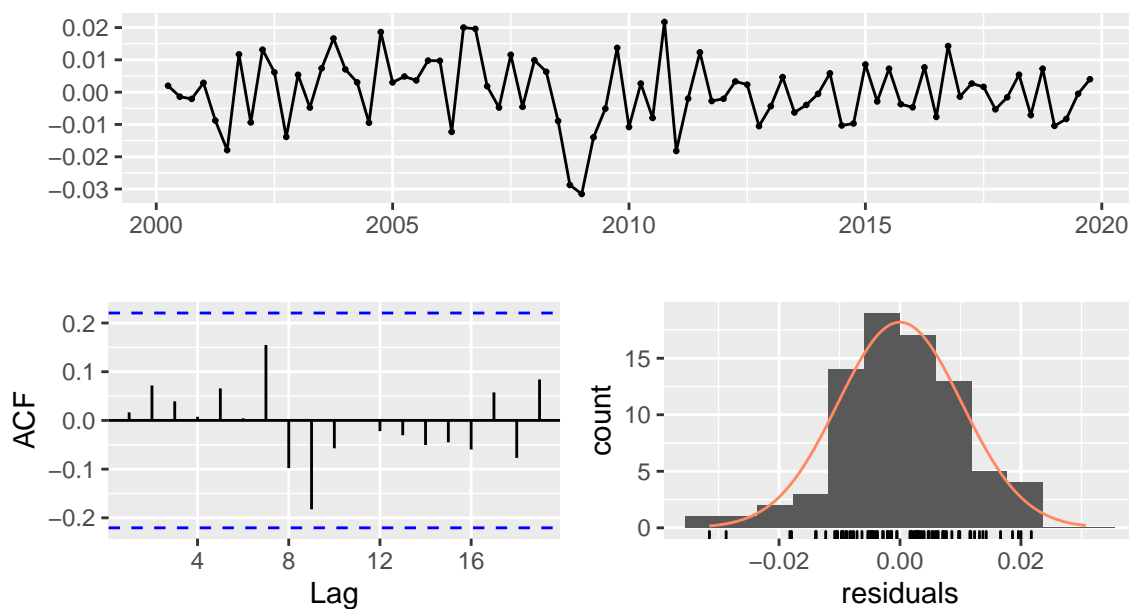
ARIMA automático: Análisis de los residuales

Table 7: Estadísticos de residuales ARIMA automático: Training set error measures

ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
-1.41e-05	0.0101855	0.0081016	-1115.759	1450.788	0.6454329	0.0162983

⁵ Algunas veces esta selección automática con AIC proporciona resultados inconsistentes.

Residuals from ARIMA(2,0,1) with non-zero mean

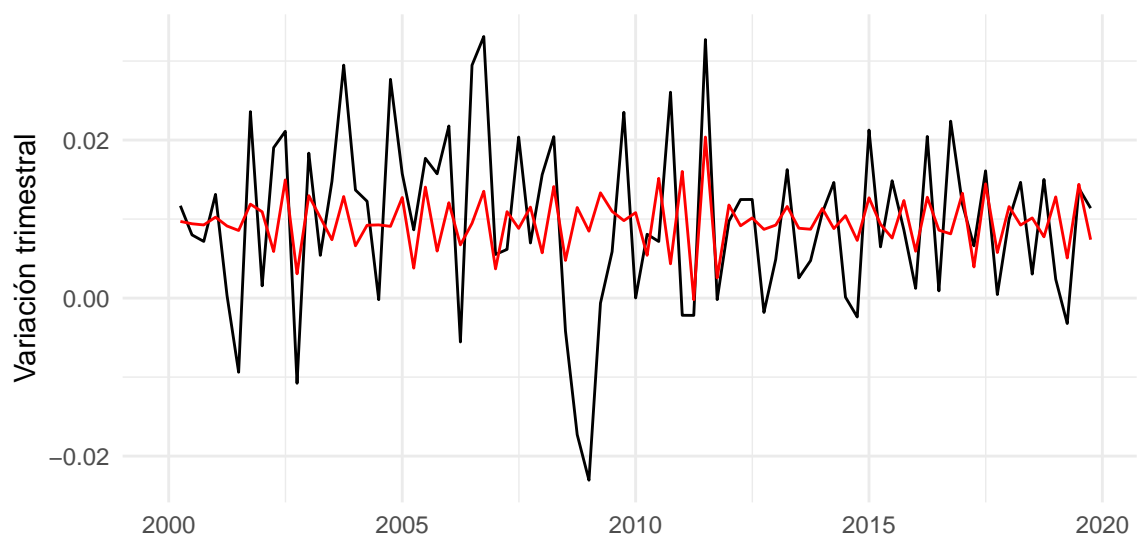


Ljung-Box test

data: Residuals from ARIMA(2,0,1) with non-zero mean
 Q* = 3.9544, df = 4, p-value = 0.4122

Model df: 4. Total lags used: 8

Evaluación del modelo ARIMA automático



rojo = pronóstico, negro = observado

Box-Ljung test

data: Model_Auto\$residuals

X-squared = 0.95808, df = 4, p-value = 0.9161

Jarque Bera Test

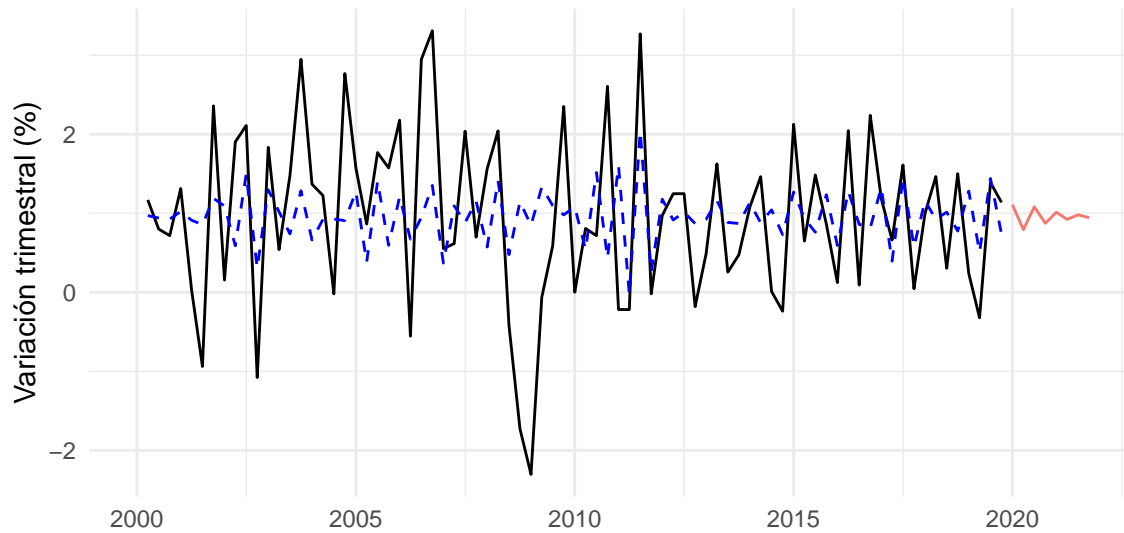
```
data: Model_Auto$residuals[!is.na(Model_Auto$residuals)]
```

X-squared = 2.6446, df = 2, p-value = 0.2665

Al igual que con el modelo ARIMA(1,0,0), los residuos no presentan autocorrelación y se ajustan a una distribución normal, con valores atípicos significativos. Nuevamente, se transforman los resultados a tasas de crecimiento interanual.

ARIMA automático: Pronóstico fuera de muestra

PIB Honduras: Pronóstico ARIMA automático



rojo = pronóstico, azul = estimado, negro = observado

Puede observarse que el pronóstico converge con menor velocidad respecto al modelo ARIMA(1,0,0).

Variaciones interanuales:

Honduras: PIB y Pronóstico ARIMA

Variación interanual (%)

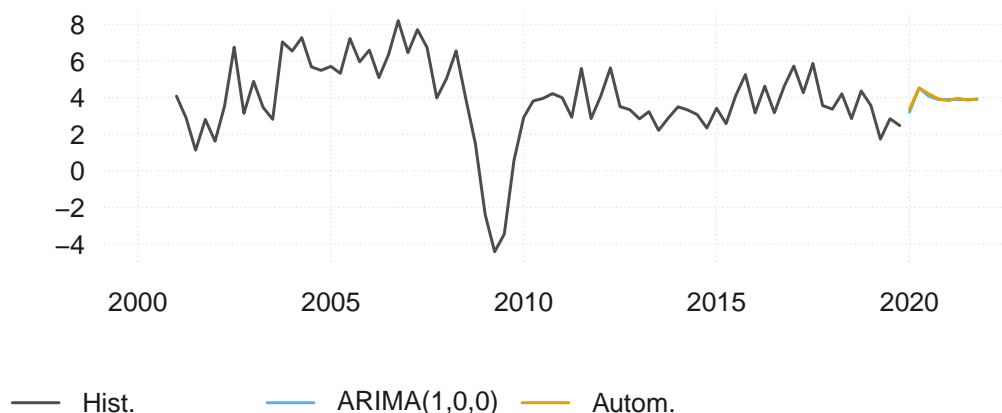


Table 8: Comparación de resultados de pronóstico: Variaciones Interanuales

Fecha	ARIMA(1,0,0)	Automático
2020-01-01	3.203224	3.367647
2020-04-01	4.530621	4.525679
2020-07-01	4.077984	4.205252
2020-10-01	3.888411	3.930016
2021-01-01	3.897095	3.832820
2021-04-01	3.896697	3.968383
2021-07-01	3.896715	3.862239
2021-10-01	3.896714	3.935063

Los resultados del modelo automático son bastante similares a los del modelo ARIMA(1,0,0). Puede observarse que, en todos los casos, los resultados de pronóstico fuera de muestra para un período mayor a seis meses convergen hacia valores similares a la media histórica, por tanto los mismos no son eficientes si se toman en cuenta los últimos datos observados y también si se muestra que la serie puede ser explicada por un proceso más determinístico, evaluando variables explicativas como posibles regresores.

5) Extensión del modelo

a) Control por dummies

Con el fin de que la media de los errores en el procedimiento automático no se vea afectada por valores extremos (variaciones trimestrales mayores a 1.5%), se incluyen dummy en dicho modelo para la siguiente fecha:

- Primer trimestre 2009 (D09a);
- Segundo trimestre 2009 (D09a);

Table 9: Coeficientes del ARIMA, incluyendo dummies

	Betas	Error_Estandar	t_Estadistic	p_value
ar1	-1.0496039	0.1048954	-10.006199	0.0000000

	Betas	Error_Estandar	t_Estadistic	p_value
ar2	-0.3696799	0.1036767	-3.565700	0.0006236
ma1	0.9999998	0.0433748	23.054870	0.0000000
intercept	0.0102271	0.0008426	12.137621	0.0000000
D09a	-0.0355877	0.0067707	-5.256154	0.0000012
D09b	-0.0170904	0.0067404	-2.535528	0.0132272

Table 10: Estadísticos del ARIMA, incluyendo dummies

sigma_2	logLik	AIC	AICc	BIC
8.38e-05	260.1834	-506.3668	-504.7893	-489.7806

Table 11: Estadísticos de residuales ARIMA incl. Dummies: Training set error measures

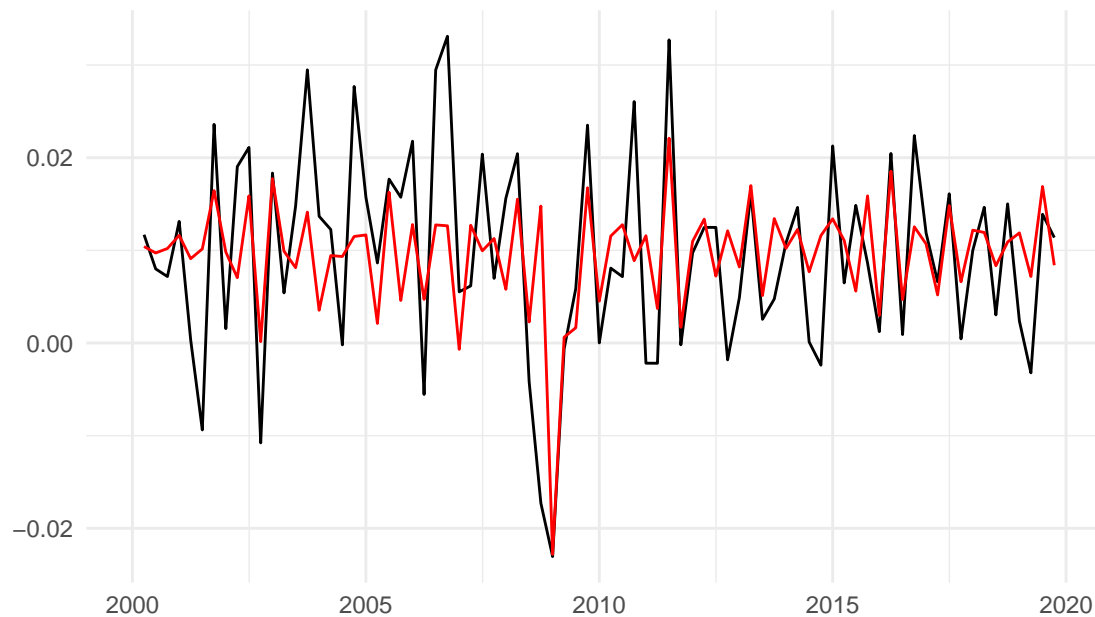
ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
-2.3e-06	0.0087972	0.0068175	-479.8207	754.6297	0.5431329	-0.0106213



Ljung-Box test

data: Residuals from Regression with ARIMA(2,0,1) errors
 Q* = 11.897, df = 3, p-value = 0.007745

Model df: 6. Total lags used: 9



Box-Ljung test

```
data: Model_ArimaDummy$residuals
X-squared = 1.3672, df = 4, p-value = 0.8499
```

Jarque Bera Test

```
data: Model_ArimaDummy$residuals[!is.na(Model_ArimaDummy$residuals)]
X-squared = 6.4372, df = 2, p-value = 0.04001
```

Aunque el comportamiento de los residuos es menos aceptable tomando en cuenta los estadísticos de autocorrelación y normalidad, se ha mejorado respecto a estimación dentro de muestra al eliminar los valores atípicos más grandes.

b) Economía Estadounidense y Dummies

Adicionalmente, se incluye como variable explicativa el PIB de Estados Unidos de América (GDP) en un segundo modelo:

Table 12: Coeficientes del ARIMA, incluyendo dummies

	Betas	Error_Estandar	t_Estadistic	p_value
ar1	-1.1338792	0.1030951	-10.998376	0.0000000
ar2	-0.4246880	0.1007029	-4.217237	0.0000661
ma1	0.9999990	0.0518750	19.277079	0.0000000
intercept	0.0071342	0.0012061	5.914981	0.0000001
dGDP_USA	0.5733016	0.1748166	3.279446	0.0015557
D09a	-0.0195547	0.0079605	-2.456469	0.0162529
D09b	-0.0154535	0.0062161	-2.486047	0.0150550

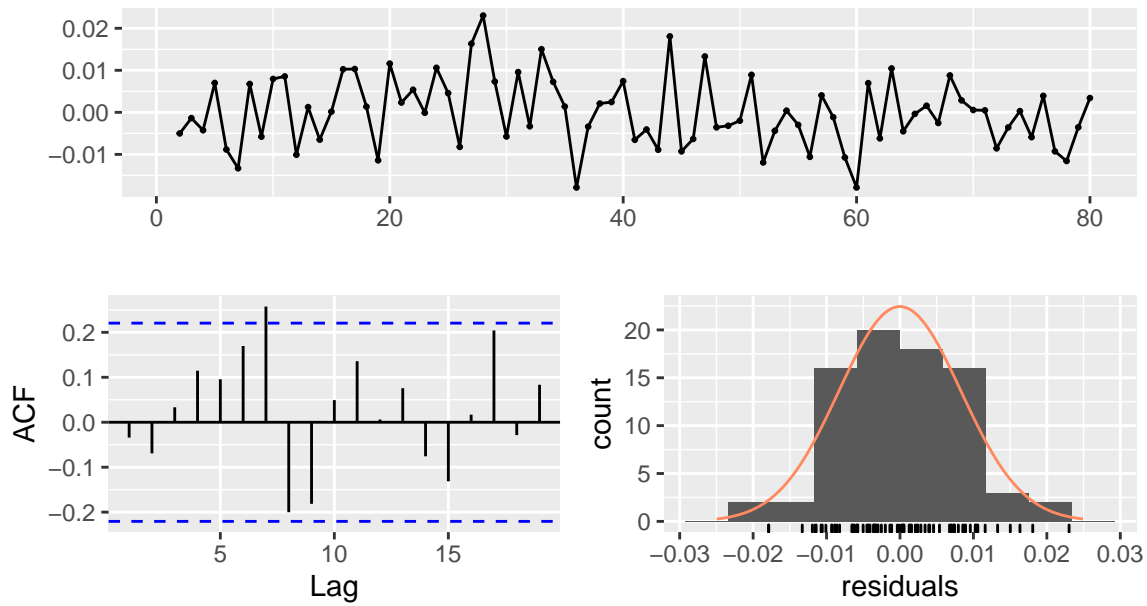
Table 13: Estadísticos del ARIMA, incluyendo dummies

sigma_2	logLik	AIC	AICc	BIC
7.49e-05	265.1436	-514.2872	-512.23	-495.3316

Table 14: Estadísticos de residuales ARIMA incl. Regresores:
Training set error measures

ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
-1.2e-05	0.0082602	0.0066943	-713.3399	1016.809	0.5333182	-0.0338943

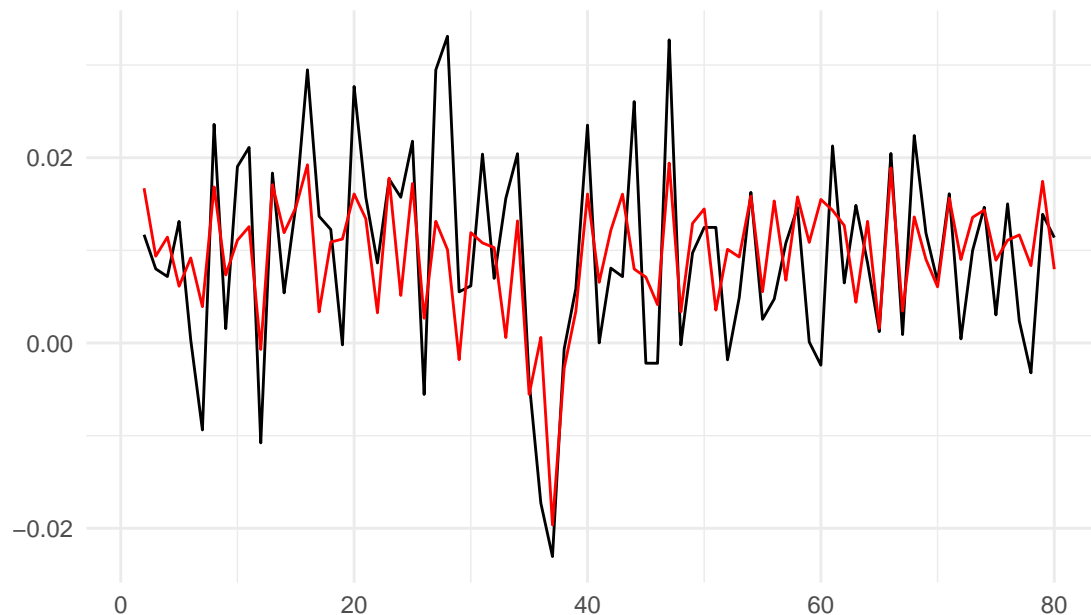
Residuals from Regression with ARIMA(2,0,1) errors



Ljung-Box test

data: Residuals from Regression with ARIMA(2,0,1) errors
Q* = 17.748, df = 3, p-value = 0.0004959

Model df: 7. Total lags used: 10



Box-Ljung test

```
data: Model_ArimaX$residuals
X-squared = 2.4988, df = 4, p-value = 0.6449
```

Jarque Bera Test

```
data: Model_ArimaX$residuals[!is.na(Model_ArimaX$residuals)]
X-squared = 1.0762, df = 2, p-value = 0.5838
```

6) Comparación de modelos

Prueba de Ljung-Box

Esta prueba se utiliza para verificar correlación serial en los residuales⁶.

$$Q^* = T(T+2) \sum_{k=1}^h (T-k)^{-1} \hat{\rho}_k^2$$

- T = número de observaciones;
- h = máximo rezago a ser considerado;
- $\hat{\rho}_k$ es la autocorrelación para el rezago k ;

Valores altos de Q^* sugieren que la autocorrelación no proviene de una serie que se comporta como ruido blanco. Q^* tiene una distribución χ^2 con $(h - K)$ grados de libertad, donde K es el número de parámetros en el modelo. Si Q^* es no significativo (p-values relativamente grandes), podemos concluir que los residuos no pueden distinguirse de un proceso ruido blanco.

Table 15: Ljung-Box test

Modelos	Q	df	p_value
ARIMA(1,0,0)	6.869088	6	0.3331238
ARIMA Automático	3.954418	4	0.4122098

⁶<https://otexts.com/fpp2/residuals.html>

Modelos	Q	df	p_value
ARIMA con Dummies	11.896826	3	0.0077451
Arima con Regresores	17.747550	3	0.0004959

De acuerdo con estos resultados, el modelo cuyos residuales se comportan más como un proceso ruido blanco es el ARIMA seleccionado de manera automática.

Prueba de Box-Ljung

Este tests examina la hipótesis nula de independencia en una serie de tiempo dada; se trata de comprobar si un grupo de autocorrelaciones son simultáneamente cero. Esta prueba es asintóticamente equivalente a la prueba anterior; si el p-value es mayor a 0.05, los residuales son independientes y los errores pueden considerarse una aproximación a un proceso ruido blanco. Los parámetros para su cálculo son los mismos a los usados en la prueba anterior:

$$Q^* = T \sum_{k=1}^K \hat{\rho}_k^2$$

El estadístico Q está distribuido (aproximadamente) como una χ^2 con K grados de libertad. De esta manera, si el valor calculado de Q^* es mayor que el nivel crítico (digamos, 5%), podemos asegurar con un 95% de confianza que los coeficientes de autocorrelación verdaderos ρ_1, \dots, ρ_k no todos son cero.

Conviene resaltar que para estas pruebas, no existe una guía clara para la elección de K . Si K es pequeño, existe el peligro de no tomar en cuenta autocorrelaciones de mayor orden, pero si K es demasiado grande relativo al tamaño de la muestra, la distribución en muestras finitas está propensa a deteriorarse, divergiendo de la distribución χ^2 .⁷

Table 16: Prueba de Box-Ljung

Modelos	ChiSqrt	df	p_value
ARIMA(1,0,0)	3.0870372	4	0.5433671
ARIMA Automático	0.9580848	4	0.9160837
ARIMA con Dummies	1.3671895	4	0.8498771
Arima con Regresores	2.4987951	4	0.6448516

Al igual que con la prueba anterior, el modelo cuyos residuales se aproximan más a un proceso ruido blanco es el ARIMA automático.

Evaluación del pronóstico, residuales dentro de muestra

Los siguientes estadísticos fueron tomados en cuenta para evaluar si los modelos observan un buen pronóstico dentro de muestra, considerando un modelo ARIMA con un término de error igual a $\varepsilon_t = Y_t - \hat{Y}_t$:

- Error promedio (ME)

$$ME = \sum_{t=1}^T \varepsilon_t$$

- Raíz del Error Cuadrático Medio (RMSE)

$$RMSE = \sqrt{\frac{\sum_{t=1}^T \varepsilon_t^2}{T}}$$

- Error Absoluto Medio (MAE)

$$MAE = \frac{\sum_{t=1}^T |\varepsilon_t|}{T}$$

⁷Hayashi (2000), pgs.142-144

- Error Porcentual Medio (MPE)

$$MPE = \frac{100\%}{T} \sum_{t=1}^T \frac{Y_t - \hat{Y}_t}{Y_t}$$

- Error Absoluto Porcentual Medio (MAPE)

$$MAPE = \frac{100\%}{T} \sum_{t=1}^T \left| \frac{Y_t - \hat{Y}_t}{Y_t} \right|$$

- Error Absoluto Escalado Medio (MASE)

$$MASE = \frac{\frac{1}{J} \sum_j |\varepsilon_j|}{\frac{1}{T-1} \sum_{t=2}^T |Y_t - Y_{t-1}|}$$

Table 17: Evaluación de Residuales: Pronóstico dentro de Muestra

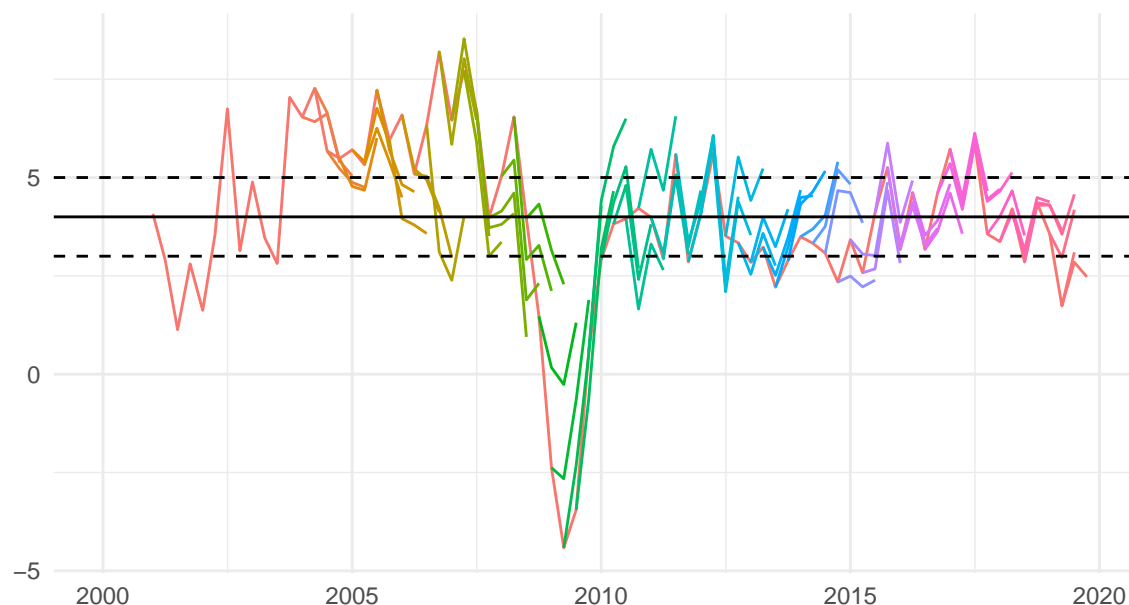
Model	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
ar1	1.0e-06	0.010837	0.008611	-881.9949	1303.3830	0.686048	-0.002867
automático	-1.4e-05	0.010185	0.008102	-1115.7595	1450.7883	0.645433	0.016298
ARIMA incl. Dummies	-2.0e-06	0.008797	0.006818	-479.8207	754.6297	0.543133	-0.010621
ARIMA incl. Regresores	-1.2e-05	0.008260	0.006694	-713.3399	1016.8091	0.533318	-0.033894

En la tabla puede observarse que el modelo que tiene los menores valores de los estadísticos (mejor modelo) es el que incluye como regresores las dummies y GDP.

7) Resultado de pronóstico dentro de muestra

Se estima el modelo con menor RMSE usando ventanas expandibles (rolling windows) con datos pasados (primer trimestre 2004 a último trimestre 2019) y produciendo un pronóstico (“pseudo-out-of Sample”). Esta es una manera de mostrar la eficiencia del pronóstico con datos observados⁸.

Pronóstico de inflación dentro de muestra: 4 trimestres



⁸Para estos resultados se omite la inclusión de las dummies dentro del RegArima

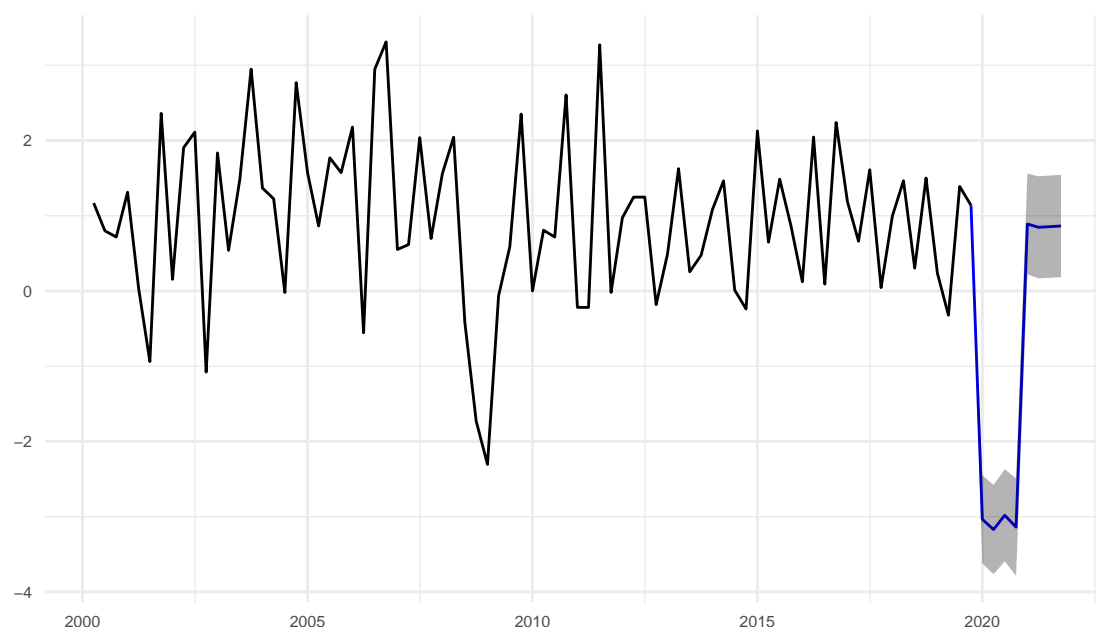
8) Fan chart de la inflación

Este ejercicio implica la construcción de un modelo de pronóstico y evaluación mediante fanchart. Adicionalmente, se calcula la probabilidad de eventos particulares utilizando simulación de Montecarlo.

8.1) Estimación de un modelo ARMA y fan chart

Se estima el modelo en diferencias logarítmicas. Alternativamente, se puede estimar en tasas de crecimiento interanual (ts_pc). Note que la función autoplot automáticamente produce un intervalo de pronóstico. Podemos especificar manualmente las probabilidades que queremos cubrir (la opción de niveles).

Con el fin de simular un impacto por el efecto del COVID-19 durante 2020, se realiza un pronóstico asumiendo una caída de la variable GDP de 0.5% en cada trimestre de este año, adicional a un impacto esperado de la economía interna similar a los dos primeros trimestres de 2009 ($D09a = D09b = 1$); adicionalmente, para 2021 se asume un crecimiento trimestral de GDP para 2021 de 0.25% (1% anual) y que la economía interna vuelva a sus niveles de producción normales ($D09a = D09b = 0$).



	Point Forecast	Lo 50	Hi 50	Lo 70	Hi 70	Lo 80	Hi 80	Lo 95	Hi 95
2020 Q1	-3.0344786	-3.6216240	-2.447333	-3.9366971	-2.132260	-4.1500732	-1.918884	-4.740633	-1.3283
2020 Q2	-3.1708111	-3.7641338	-2.577488	-4.0825216	-2.259101	-4.2981427	-2.043479	-4.894916	-1.4467
2020 Q3	-2.9807391	-3.5932185	-2.368260	-3.9218862	-2.039592	-4.1444690	-1.817009	-4.760510	-1.2009
2020 Q4	-3.1383595	-3.7854756	-2.491243	-4.1327300	-2.143989	-4.3679003	-1.908819	-5.018780	-1.2579
2021 Q1	0.8902708	0.2206176	1.559924	-0.1387306	1.919272	-0.3820911	2.162633	-1.055639	2.8361
2021 Q2	0.8460881	0.1680005	1.524176	-0.1958738	1.888050	-0.4422995	2.134476	-1.124331	2.8165
2021 Q3	0.8545658	0.1748006	1.534331	-0.1899739	1.899106	-0.4370093	2.146141	-1.120728	2.8298
2021 Q4	0.8637170	0.1838946	1.543539	-0.1809107	1.908345	-0.4279668	2.155401	-1.111743	2.8391

Interpretación: El gráfico muestra la media de pronóstico (línea sólida) y el intervalo de pronóstico. Los números asociados con el intervalo de pronóstico representan la cobertura. Los intervalos representan una medida de la confianza que tenemos en el pronóstico: si son muy grandes, el pronóstico puntual no es muy exacto y vice versa.

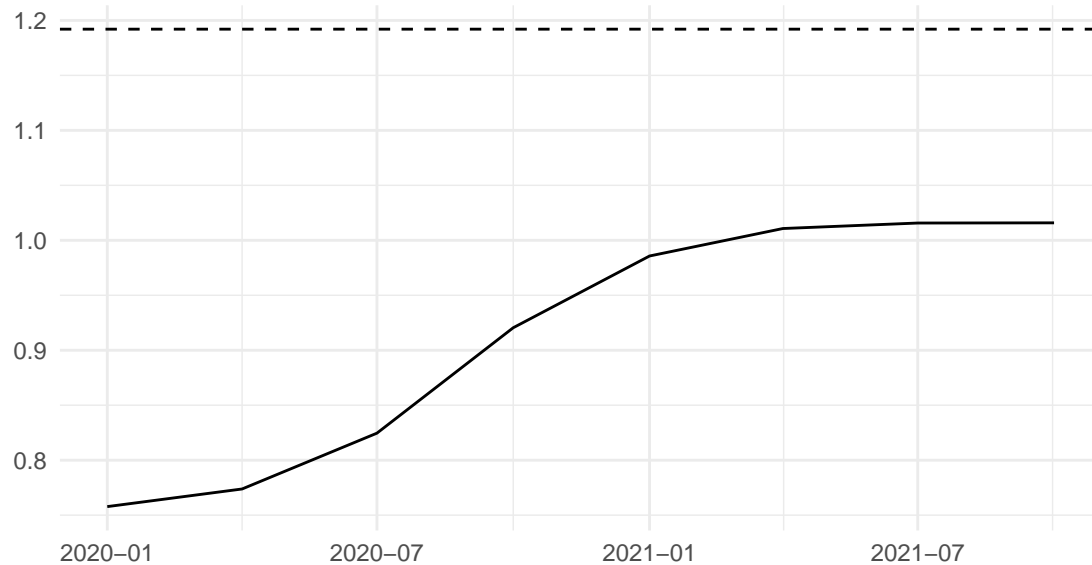
Aunque es difícil de observar en muestras cortas, el intervalo de pronóstico se agranda para horizontes más largos (debido a que la varianza del error de pronóstico también se incrementa).

8.2) Calcular la varianza del error de pronóstico VEP y simular la densidad de pronóstico

a) Extraer la varianza del error de pronóstico y comparar con la varianza incondicional

Esta es una función útil para calcular la varianza del error de pronóstico asumiendo que dichos errores están normalmente distribuidos.

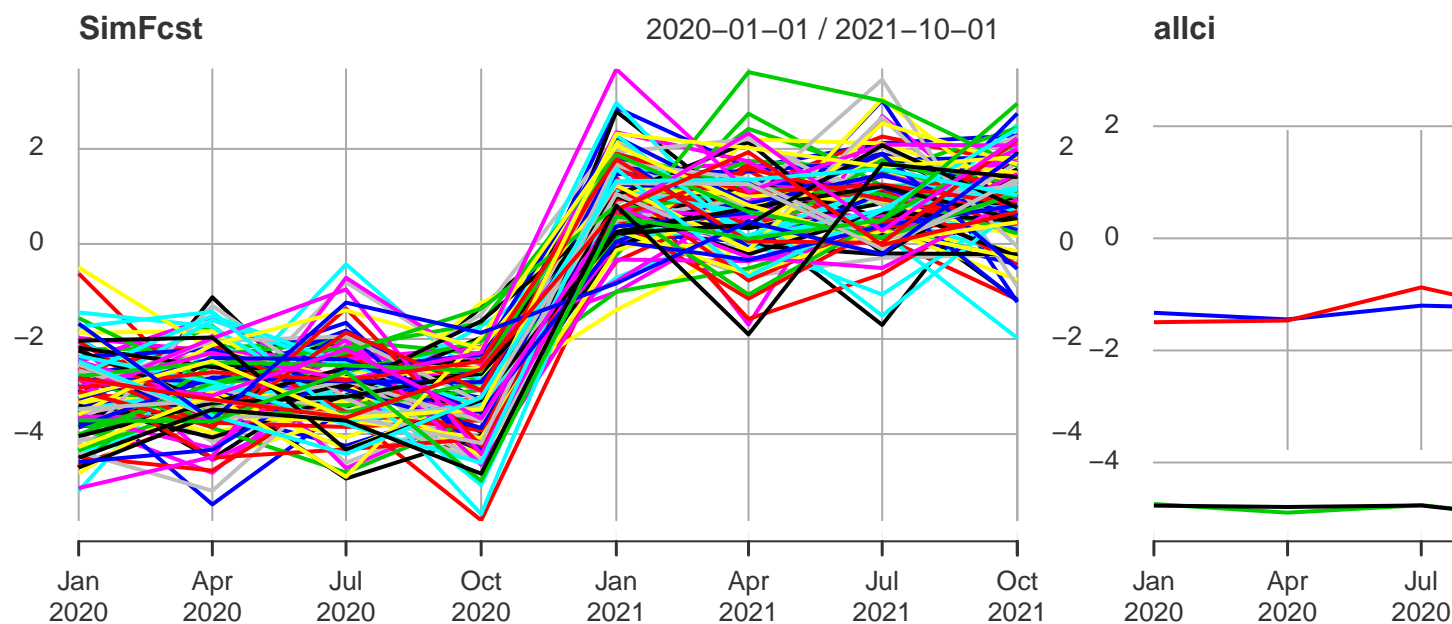
Varianza del error de pronóstico y varianza incondicional



b) Simular la densidad de pronóstico y comparar con el intervalo de pronóstico analítico obtenido

Se realiza una simulación de la densidad de pronóstico asumiendo que el error de pronóstico está distribuido como una normal. Esto implica que $y_{t+h} \sim N(y_{t+h|t}, \text{sig}^2)$, esto es, el valor futuro del crecimiento del PIB está distribuido como una normal con media igual al pronóstico puntual y varianza igual a la varianza del error de pronóstico.

Podemos graficar las simulaciones. Cada línea es un posible resultado. El promedio será igual a la media de pronóstico y podemos calcular percentiles que sean iguales a los intervalos de pronóstico.



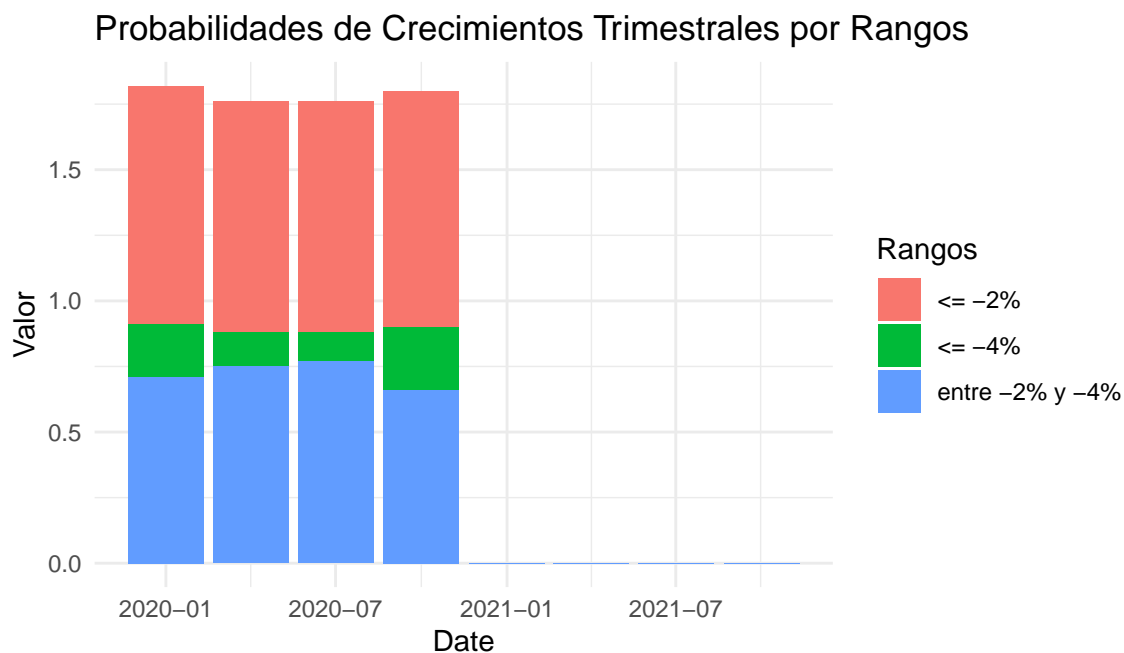
c) Calcular probabilidad de crecimiento dentro de un rango del PIB en Q1 2020

La ventaja real de simular la incertidumbre del pronóstico es que podemos fácilmente computar la probabilidad de eventos particulares.

[1] 0.91

[1] 0.2

[1] 0.71



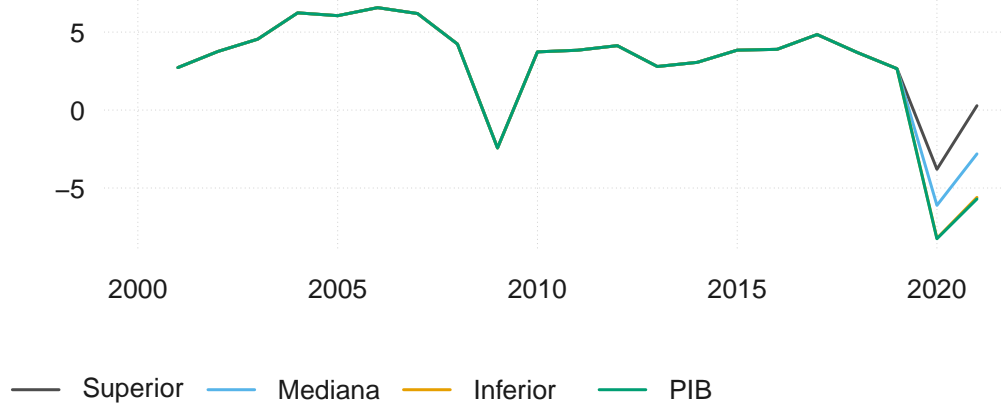
d) Probabilidad que el crecimiento anual del PIB en 2020 sea menor a 5%

Paso 1: Calcular el nivel para cada simulación y agregarlo a una frecuencia anual

Paso 2: Calcular la probabilidad

PIB Honduras

Tasa de crecimiento anual



Los resultados indican que la probabilidad que la tasa de crecimiento anual del PIB sea menor a -5% a finales de 2020 es 81% y para 2021 de 6%

Anexos

Los Anexos 1 y 2 se basan en Pyndick & Rubinfeld (2001), capítulo 16, pg.515 en adelante.

Anexo 1: Caminata aleatoria

El ejemplo más simple de una serie de tiempo estocástica es el proceso de caminata aleatoria; en este proceso, cada cambio sucesivo en y_t es extraído en forma independiente de una distribución de probabilidad con media cero. Por tanto, y_t está determinado por

$$Y_t = Y_{t-1} + \varepsilon_t \quad (\text{A1.1})$$

con $E(\varepsilon_t) = 0$ y $E(\varepsilon_t, \varepsilon_s) = 0$ para $t \neq s$

Consideremos el caso en el que se requiere un pronóstico para un proceso de caminata aleatoria. El pronóstico está dado por

$$\hat{Y}_{T+1} = E(Y_{T+1}|Y_T, \dots, Y_1) = Y_T + E(\varepsilon_{T+1}) = Y_T \quad (\text{A1.2})$$

El pronóstico dos períodos adelante es:

$$\hat{Y}_{T+2} = E(Y_{T+2}|Y_T, \dots, Y_1) = E(Y_{T+1} + \varepsilon_{T+2}) = E(Y_T + \varepsilon_{T+1} + \varepsilon_{T+2}) = Y_T \quad (\text{A1.3})$$

Del mismo modo, el pronóstico l periodos adelante también es Y_T .

Aunque el pronóstico \hat{Y}_{T+1} será el mismo sin importar cuán grande sea l , la varianza del error de pronóstico crecerá conforme l se haga mayor. Para un periodo el error de pronóstico está dado por:

$$e_1 = Y_{T+1} - \hat{Y}_{T+1} = Y_T + \varepsilon_{T+1} - Y_T = \varepsilon_{T+1} \quad (\text{A1.4})$$

y su varianza es $E(\varepsilon_{T+1}^2) = \sigma_\varepsilon^2$. Para el pronóstico de dos periodos

$$e_2 = Y_{T+2} - \hat{Y}_{T+2} = Y_T + \varepsilon_{T+1} + \varepsilon_{T+2} - Y_t = \varepsilon_{T+1} + \varepsilon_{T+2} \quad (\text{A1.5})$$

y su varianza es:

$$E[(\varepsilon_{T+1} + \varepsilon_{T+2})^2] = E(\varepsilon_{T+1}^2) + E(\varepsilon_{T+2}^2) + 2E(\varepsilon_{T+1}\varepsilon_{T+2}) \quad (\text{A1.6})$$

Dado que ε_{T+1} y ε_{T+2} son independientes, el tercer término de la ecuación (A1.6) es cero y la varianza del error es $2\sigma_\varepsilon^2$. De igual forma, para el pronóstico del periodo l , la varianza del error es $l\sigma_\varepsilon^2$. Así, el *error estándar del pronóstico* se incrementa con la raíz cuadrada de l . Por tanto se pueden obtener *intervalos de confianza* para nuestros pronósticos y estos intervalos se volverán más amplios conforme se incremente el horizonte del pronóstico.

Anexo 2: Criterio ADF

Suponga que creemos que una variable Y_t , la cual ha estado creciendo con el tiempo, puede describirse por la siguiente ecuación:

$$Y_t = \alpha + \beta \cdot t + \rho \cdot Y_{t-1} + \varepsilon_t \quad (\text{A2.1})$$

Una posibilidad es que Y_t ha estado creciendo debido a que tiene una tendencia positiva ($\beta_1 > 0$) pero sería estacionaria después de eliminar la tendencia (es decir, $\rho < 1$).

Otra posibilidad es que Y_t ha estado creciendo debido a que sigue una caminata aleatoria con un rumbo positivo (es decir, $\alpha > 0$, $\beta = 0$ y $\rho = 1$); en este caso trabajamos con ΔY_t .

La eliminación de la tendencia no haría estacionaria la serie y la inclusión de Y_t en una regresión (aún si se elimina la tendencia) podría conducir a resultados imprecisos.

Podríamos pensar que la ecuación (A2.1) pudiera ser estimada con OLS y que la estadística t en $\hat{\rho}$ se puede usar para probar si $\hat{\rho}$ es significativamente diferente de 1. Sin embargo, si el valor verdadero de ρ en efecto es 1, el estimador OLS está sesgado hacia cero. Por tanto, al usar OLS de esta manera puede conducirnos a rechazar incorrectamente la hipótesis de la caminata aleatoria.

Dickey y Fuller derivaron la distribución para el estimador $\hat{\rho}$ que se cumple cuando $\rho = 1$ y generaron estadísticas para una prueba F simple de la hipótesis de caminata aleatoria; es decir, la hipótesis de que $\beta = 0$ y $\rho = 1$. Suponga que Y_t puede describirse con la ecuación (A2.1)

Usando OLS, ejecutamos primero la regresión sin restricción

$$Y_t - Y_{t-1} = \alpha + \beta \cdot t + (\rho - 1) \cdot Y_{t-1} + \varepsilon_t \quad (\text{A2.2})$$

y luego la regresión restringida

$$Y_t - Y_{t-1} = \alpha \quad (\text{A2.3})$$

Luego calculamos la razón F estándar⁹, donde ESS_R y ESS_{UR} son las sumas de cuadrados de los residuales en las regresiones restringida y sin restricciones, respectivamente, N es el número de observaciones, k es el número de parámetros estimados en la regresión sin restricción y q es el número de restricciones de parámetro para probar si las restricciones ($\beta = 0$, $\rho = 1$) se cumplen; esta razón no está distribuida como una distribución F estándar bajo la hipótesis nula, de acuerdo con la distribución calculada por Dickey y Fuller, los valores críticos son mucho mayores que los de la tabla F estándar.

⁹ $F = \frac{(N-k)(ESS_R - ESS_{UR})}{q(ESS_{UR})}$

Un problema con la ecuación (A2.1) es que hace la suposición implícita de que no hay correlación serial de ninguna clase en el término de error ε_t . Nos gustaría permitir, con frecuencia, una correlación serial en ε_t y todavía probar por una raíz unitaria. Esto puede hacerse con la prueba de Dickey-Fuller aumentada. Esta prueba se lleva a cabo expandiendo la ecuación (A2.1) para incluir cambios rezagados en Y_t en el lado derecho de la ecuación:

$$Y_t = \alpha + \beta \cdot t + \rho \cdot Y_{t-1} + \sum_{j=1}^p \lambda_j \Delta Y_{t-j} + \varepsilon_t \quad (\text{A2.4})$$

donde $\Delta Y_t = Y_t - Y_{t-1}$.

La prueba de raíz unitaria se realiza así:

Usando OLS, primero se ejecuta la regresión sin restricción

$$Y_t - Y_{t-1} = \alpha + \beta \cdot t + (\rho - 1) \cdot Y_{t-1} + \sum_{j=1}^p \lambda_j \Delta Y_{t-j} \quad (\text{A2.5})$$

y luego la regresión restringida

$$Y_t - Y_{t-1} = \alpha + \sum_{j=1}^p \lambda_j \Delta Y_{t-j} \quad (\text{A2.6})$$

Entonces, se calcula una razón F estándar para probar si se cumplen las restricciones ($\beta = 0$, $\rho = 1$). Una vez más, debemos usar la distribución tabulada por Dickey y Fuller.

Debe tomarse en cuenta que la prueba ADF solo nos permite rechazar (o dejar de rechazar) la hipótesis de que una variable *no* es una caminata aleatoria. Una falla en rechazar (en especial en un nivel de significancia alto) solo proporciona una evidencia débil a favor de la hipótesis de la caminata aleatoria.

Anexo 3: Referencias

<https://www.rapidtables.com/code/text/unicode-characters.html>

<https://www.andrewheiss.com/blog/2018/03/08/amelia-broom-huxtable/>

http://haozhu233.github.io/kableExtra/awesome_table_in_html.html

https://plotly-r.com/plotly_book.pdf

<https://bookdown.org/yihui/rmarkdown/>

Hayashi, F., “Econometrics”, First Ed., Princeton University Press, 2000.

Pindyck, RS, Rubinfeld, D., “Econometría Modelos y Pronósticos”, 4ta. Ed., McGraw-Hill, 2001.