Evolución regional y futuro a corto plazo del despliegue del IPv6

Autor: Emiro A. Molina-Cuevas, Ph.D.
Coordinación/revisión: Guillermo Cicileo
Edición: Área de Comunicaciones
Supervisión general: Área de Tecnología

lacnic
www.lacnic.net

Resumen 2
Introducción 3
Modelos estocásticos del despliegue del IPv6 7
Exploración y ajuste de modelos con componentes no observados 9
Pronóstico 13
Referencias 16

Resumen

El objetivo de este trabajo es proporcionar pronósticos sobre la evolución del despliegue del IPv6 en las tres principales subregiones de Latinoamérica que atiende LACNIC: Centroamérica, Suramérica y el Caribe. Concretamente, busca calcular el momento aproximado en que el despliegue del IPv6 superará el 50% de los usuarios de internet en esas subregiones.

La metodología consiste en ajustar modelos estocásticos de componentes inobservados (UCM en inglés) a las estimaciones mensuales del porcentaje de usuarios del IPv6 en cada subregión. Una vez obtenido un modelo con un ajuste y propiedades adecuadas, este es utilizado para realizar los pronósticos.

En la primera sección se discute la evolución mensual del porcentaje de usuarios en las tres subregiones. En la segunda, se introducen los modelos estocásticos de componentes inobservados. En la tercera, el ajuste de dichos modelos a los datos observados. En la cuarta sección se presentan los pronósticos.

De acuerdo con estas proyecciones, en Centroamérica los usuarios del IPv6 alcanzarían el 51% hacia diciembre de 2023, con un valor proyectado del 51,45 % con un error (RECM) del 6,54 %.

En Suramérica, se alcanzaría el 51% hacia finales de enero de 2027, con un valor proyectado del 51,07% y un error del 4,74%. En el Caribe, los usuarios del IPv6 alcanzarían el 51% hacia octubre de 2028, con un valor proyectado del 51,31% y un error del 20,26%.

Las estimaciones mensuales fueron obtenidas a partir de los datos diarios proporcionados por LACNIC, que se encuentran en el repositorio disponible en Open Data desde el sitio ‹stats.labs.lacnic.net›. Provienen de las estadísticas de Google IPv6 ubicadas en ‹https://www.google.com/intl/en/ipv6/statistics.html›.

Los porcentajes de usuarios de internet en la región fueron tomados del repositorio de datos de la CEPAL, CEPALSTAT, del Internet World Stats (https://www.internetworldstats.com/stats15.htm) y de los Digital Reports (https://datareportal.com/).

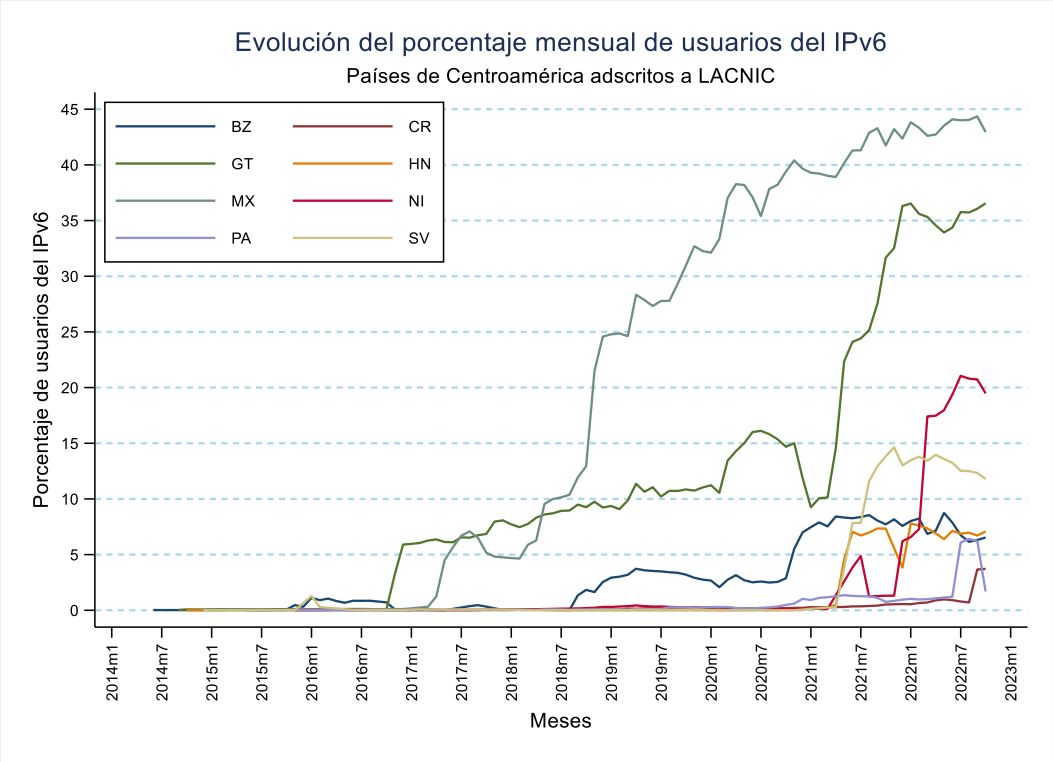
Introducción

El despliegue del protocolo IPv6 en Latinoamérica es marcadamente heterogéneo, incluso si consideramos por separado las tres regiones principales (Centroamérica, Suramérica y el Caribe).

A modo de ejemplo, el gráfico 1 nos permite apreciar las diferencias en la evolución de la adopción del protocolo en los países de la región de Centroamérica adscritos a LACNIC (medido como el porcentaje de usuarios que utilizaron el protocolo para acceder a los servicios de Google en cada país).

En algunos países, el despliegue exhibe un crecimiento sostenido (por ejemplo, en México), mientras que otros presentan cambios notables en las tasas de crecimiento (Guatemala). Algunos presentan cambios de nivel (como Belice) y otros un crecimiento muy precario.

Gráfico 1



Códigos: BZ – Belice; CR – Costa Rica; GT – Guatemala; HN – Honduras; MX – México; NI – Nicaragua; PA – Panamá; SV – El Salvador.

Los gráficos 2 a 4 muestran el despliegue mensual por país en cada una de estas tres regiones. En este trabajo, estamos interesados en la evolución regional de la adopción del protocolo IPv6. Particularmente, es de interés predecir el momento aproximado en el que podríamos esperar que el despliegue del IPv6 supere el 50% en cada región.

Gráfico 2

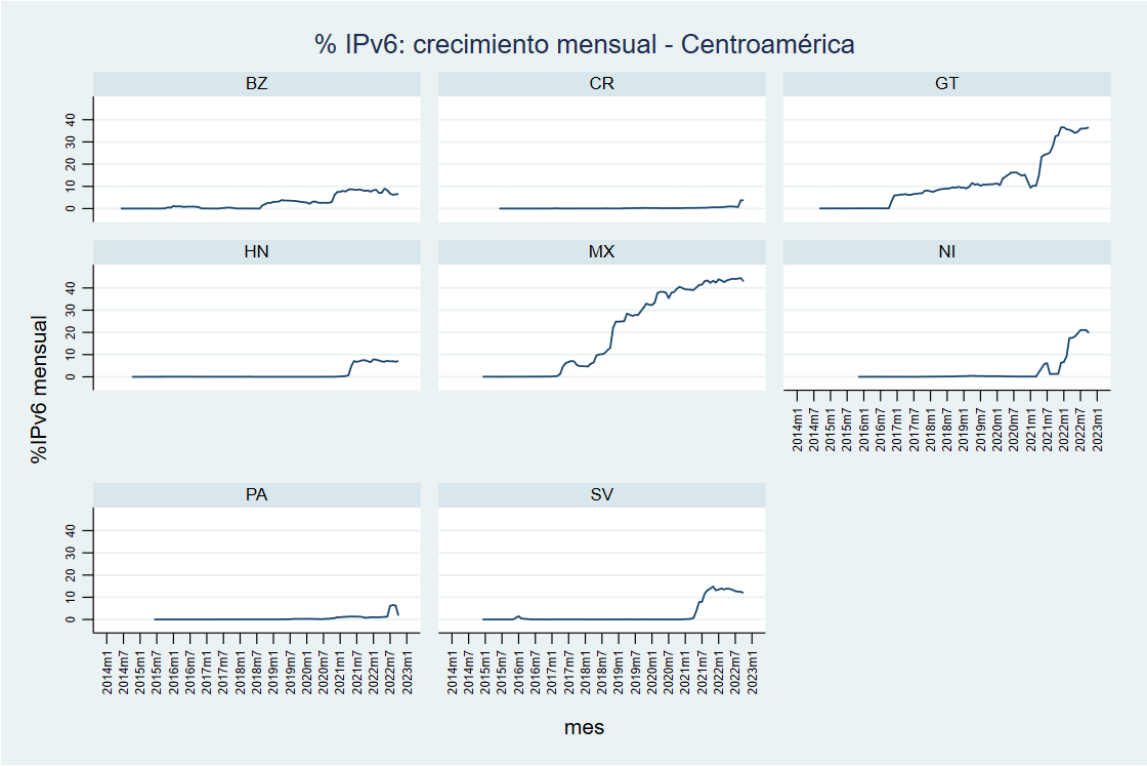


Gráfico 3

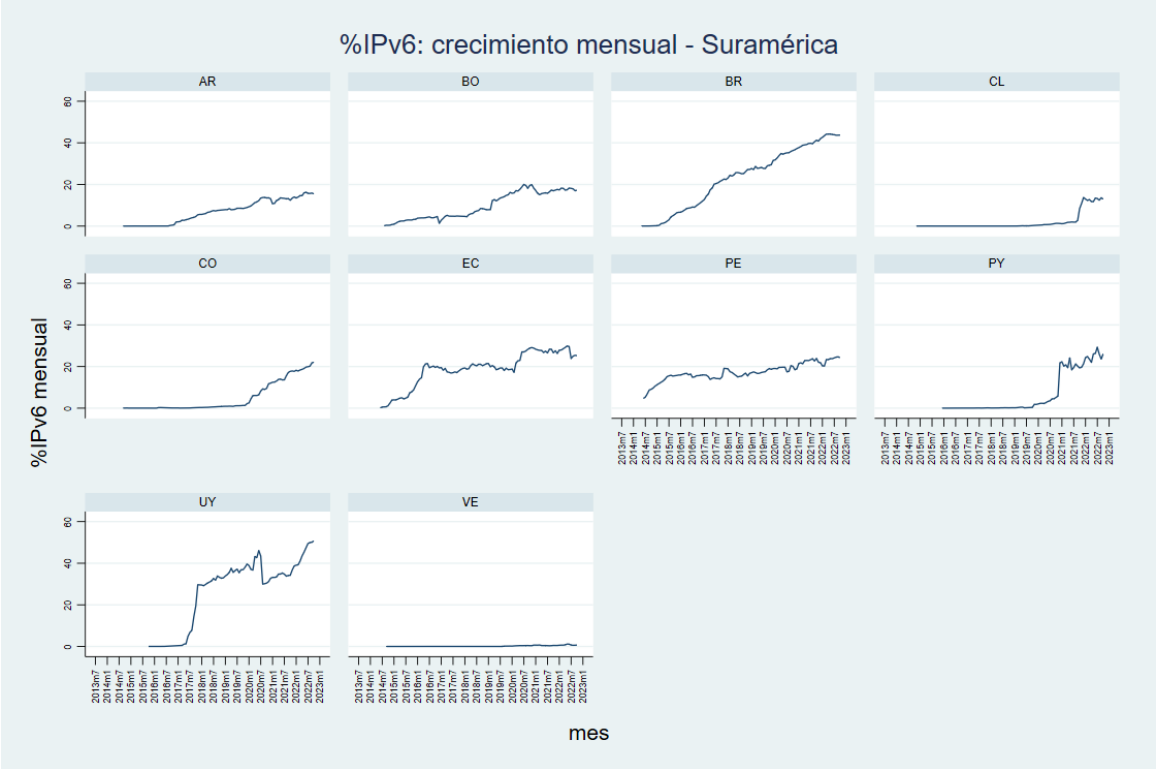
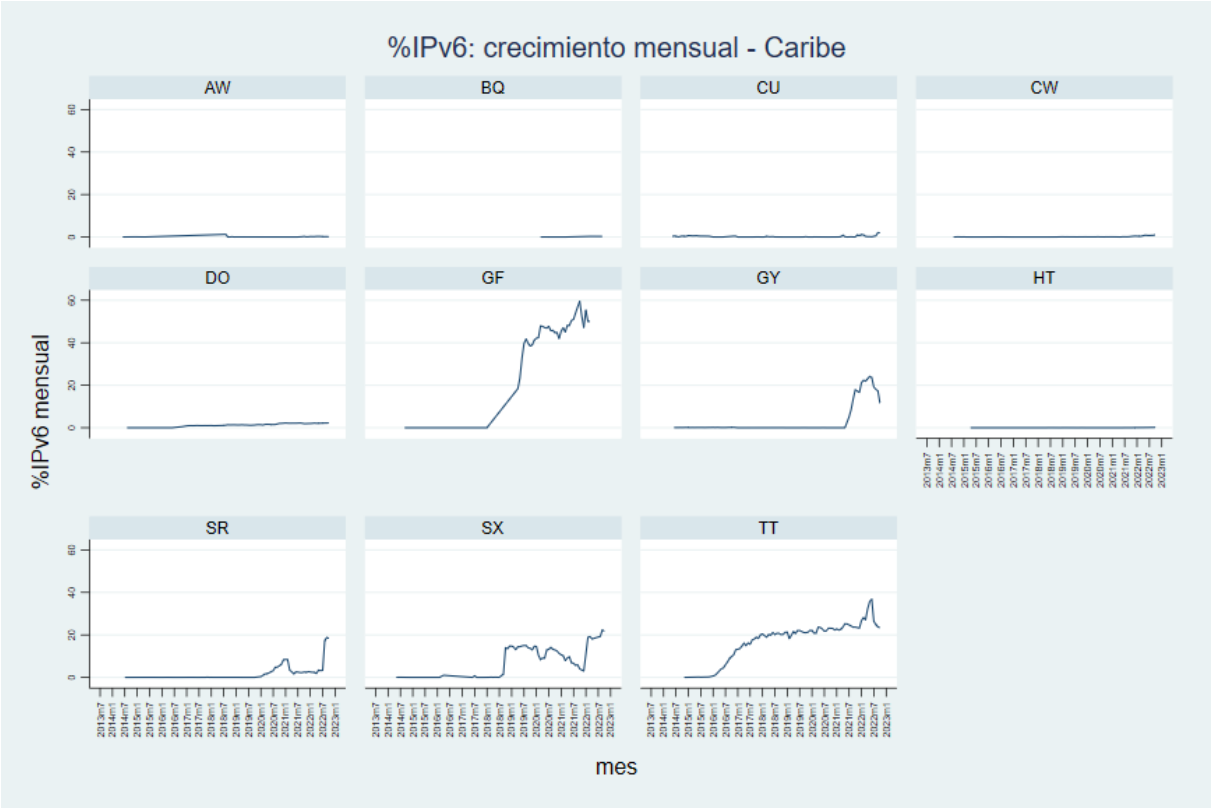
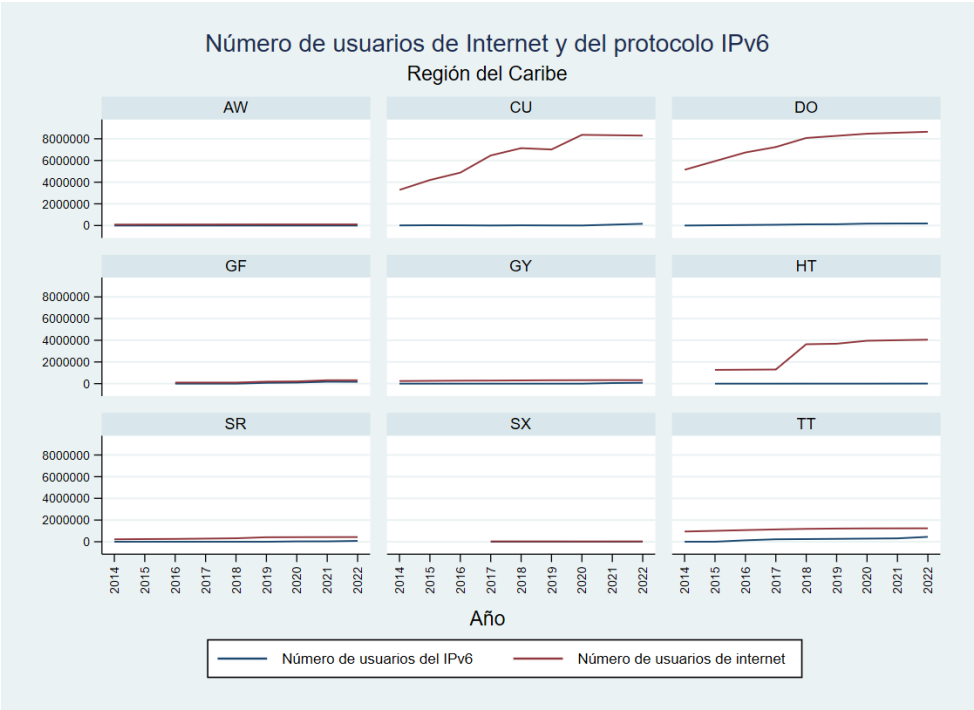


Gráfico 4



Como puede apreciarse, en varios países el protocolo tiene un despliegue inferior al 3%, a pesar de tener un significativo número de usuarios de internet. El gráfico 5 ilustra la situación del Caribe.

Gráfico 5



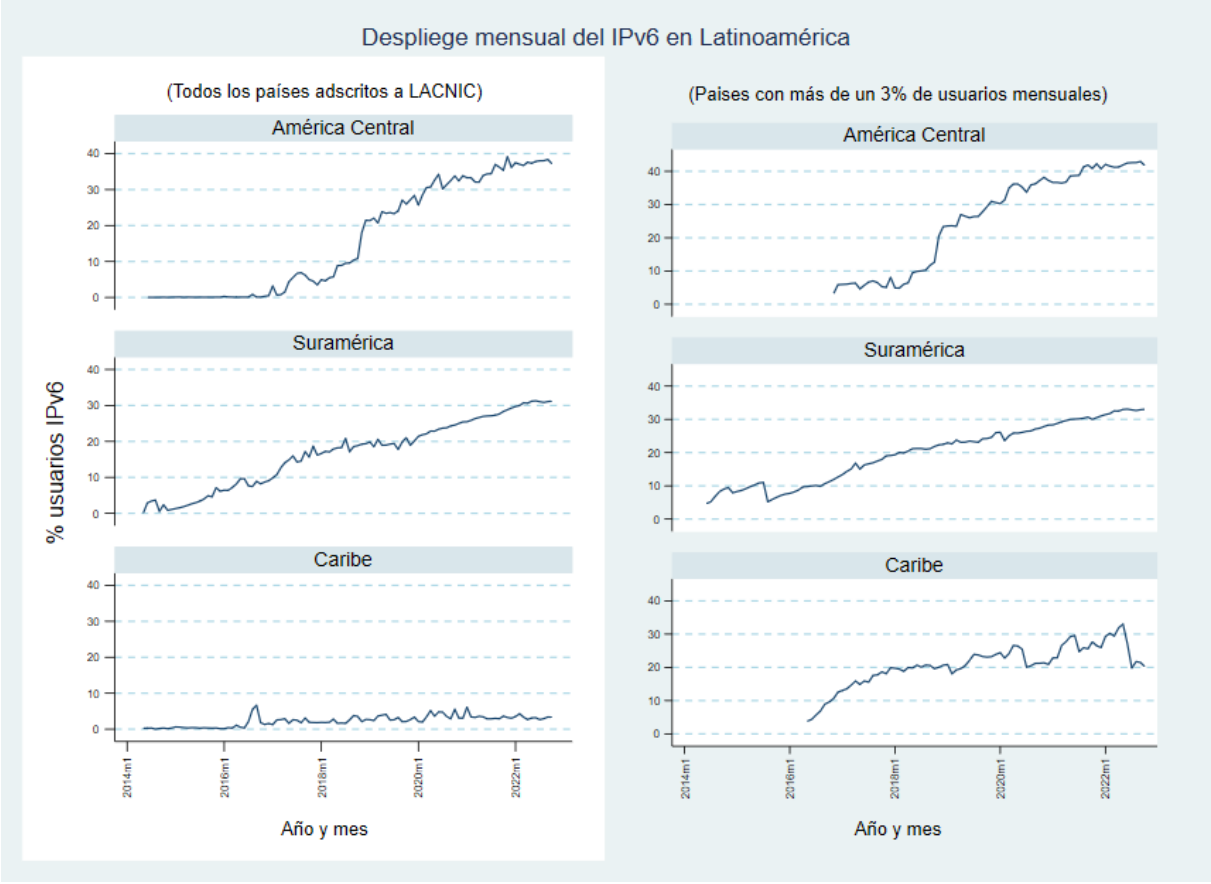
Para calcular el porcentaje de usuarios en la región, debemos primero estimar el número de usuarios de internet que utilizan el protocolo en cada país, totalizarlos y dividir entre el número de usuarios de internet en la región.

Si incorporáramos al cálculo los países con un número significativo de usuarios de internet que aún no adoptan el protocolo o cuyo porcentaje de despliegue se ha mantenido estancado en niveles muy bajos, las tasas de crecimiento regional se verían muy afectadas y no reflejarían el despliegue que ha tenido el protocolo en los países. Por esta razón, en este trabajo nos restringiremos a los países en los que el despliegue supera el 3%.

El gráfico 6 contrasta la evolución del porcentaje de usuarios del IPv6 por regiones cuando incluimos todos los países de cada región y cuando tomamos únicamente los países de cada región cuyo despliegue es superior al 3%.

El mayor impacto se aprecia en la región del Caribe, donde, al incluir a países en los cuales el protocolo es muy poco utilizado, a pesar de que los usuarios de internet se han venido incrementando, la región exhibe un estancamiento relativo, con un crecimiento que no supera el 10%. En contraste, al incluir solo los países con un despliegue superior al 3%, se puede apreciar una tasa de crecimiento moderado, que llega a alcanzar el 30%.

Gráfico 6



Técnicamente, estos datos constituyen procesos estocásticos en el tiempo o *series temporales*. Un aspecto a considerar es la longitud de las series y el nivel de agregación a utilizar. Las series de datos diarias son inconvenientes en este caso, dada la alta volatilidad que presentan y la cantidad de datos a procesar.

Por otra parte, una predicción de precisión diaria es innecesaria en este caso. Una agregación anual arrojaría series muy cortas para los propósitos de predicción. Escogimos trabajar con un nivel de agregación mensual, el porcentaje mensual de usuarios, lo que aporta series de longitud moderada, menor volatilidad y horizontes razonables.

Modelos estocásticos del despliegue del IPv6

Como muestra el gráfico 6, la evolución del despliegue regional del IPv6 presenta un comportamiento irregular, característico de las series temporales. La literatura para el análisis y predicción de este tipo de datos es amplia y varias familias de modelos han sido propuestos.

Una forma estándar de pensar en las series temporales es como la suma (o multiplicación) de componentes no observables directamente, tales como tendencias, estacionalidades, ciclos y un componente totalmente aleatorio. Un ejemplo de este enfoque lo constituyen los métodos de promedios móviles ponderados exponencialmente, muy utilizados en la literatura temprana de la predicción de series temporales.

Sin embargo, estos métodos producen predicciones sin utilizar modelos de generación de los datos observados, lo que es inconveniente para la interpretación. Con el advenimiento de los modelos integrados autorregresivos y de promedios móviles (ARIMA: Box and Jenkins, 1976) se introduce un proceso iterativo para construir modelos de series temporales y una serie de herramientas de diagnóstico para la identificación y discriminación entre modelos.

Sin embargo, para la aplicación en pronósticos en el mundo real, los modelos ARIMA muchas veces no superan los obtenidos mediante promedios móviles ponderados exponencialmente.

Los modelos de componentes no observados seleccionan las mejores características de ambas metodologías:[1] comparten el mismo marco estocástico que los modelos ARIMA, pero, al mismo tiempo, se basan en una descomposición (tendencias, estacionalidades, ciclos) y tienden a funcionar muy bien para el pronóstico de las series temporales.

En este trabajo utilizaremos estos modelos y conservaremos la notación en inglés (UCM) para referirnos a ellos.

Denominemos p_t al porcentaje mensual de usuarios del IPv6 en el mes t. Suponemos que p_t, la serie de tiempo observable, puede representarse, en general, como la suma de componentes que representan una tendencia, ciclos, estacionalidad y ruido blanco:

$$p_t = \underbrace{\mu_t}_{tendencia} + \underbrace{\psi_t}_{ciclo} + \underbrace{\gamma_t}_{estacionalidad} + \underbrace{\varepsilon_t}_{ruido\ blanco}$$

[1] Unobserved Components Models, UCM's en inglés. Ver, por ejemplo, Pelagatti, 2016.

La tendencia representa el movimiento a largo plazo de la serie. En el caso del despliegue regional del IPv6, se aprecia una tendencia clara y creciente (aunque irregular: gráfico 6) para Centroamérica y Suramérica. Cuando solo incluimos países con más de 3% de despliegue, también se aprecia una tendencia ligeramente creciente en el caso del Caribe.

Los ciclos son comportamientos de largo alcance (por ejemplo, recesiones económicas) y se modelan generalmente mediante componentes sinusoidales que pueden tener carácter determinístico o estocástico.

Las estacionalidades se refieren a comportamientos periódicos de más corta duración. Se especifican como el número de períodos de tiempo necesarios para que se complete la estacionalidad: por ejemplo, *cuatro* para la estacionalidad trimestral, *doce* para la estacionalidad mensual. Los componentes estacionales también pueden ser determinísticos o estocásticos y, generalmente, se modelan por medio de variables indicadoras.[2]

Comenzamos asumiendo que los datos pueden explicarse con el componente correspondiente a la tendencia. Siguiendo a Harvey (1978), un modelo general para la tendencia μ_t *puede* escribirse estructuralmente como:

$$p_t = \mu_t + \varepsilon_t$$
$$\mu_t = \mu_{t-1} + \beta_{t-1} + \eta_t$$
$$\beta_t = \beta_{t-1} + \xi_t$$

donde μ_t es el nivel de la serie en el tiempo t, β_t es el incremento de nivel entre los tiempos t y $t+1$ (la pendiente) *y* ε_t, η_t, ξ_t son sucesiones independientes de ruido blanco (variables aleatorias independientes con media cero y varianzas fijas $\sigma^2_{À}$, σ^2_{η} y σ^2_{ξ} respectivamente).

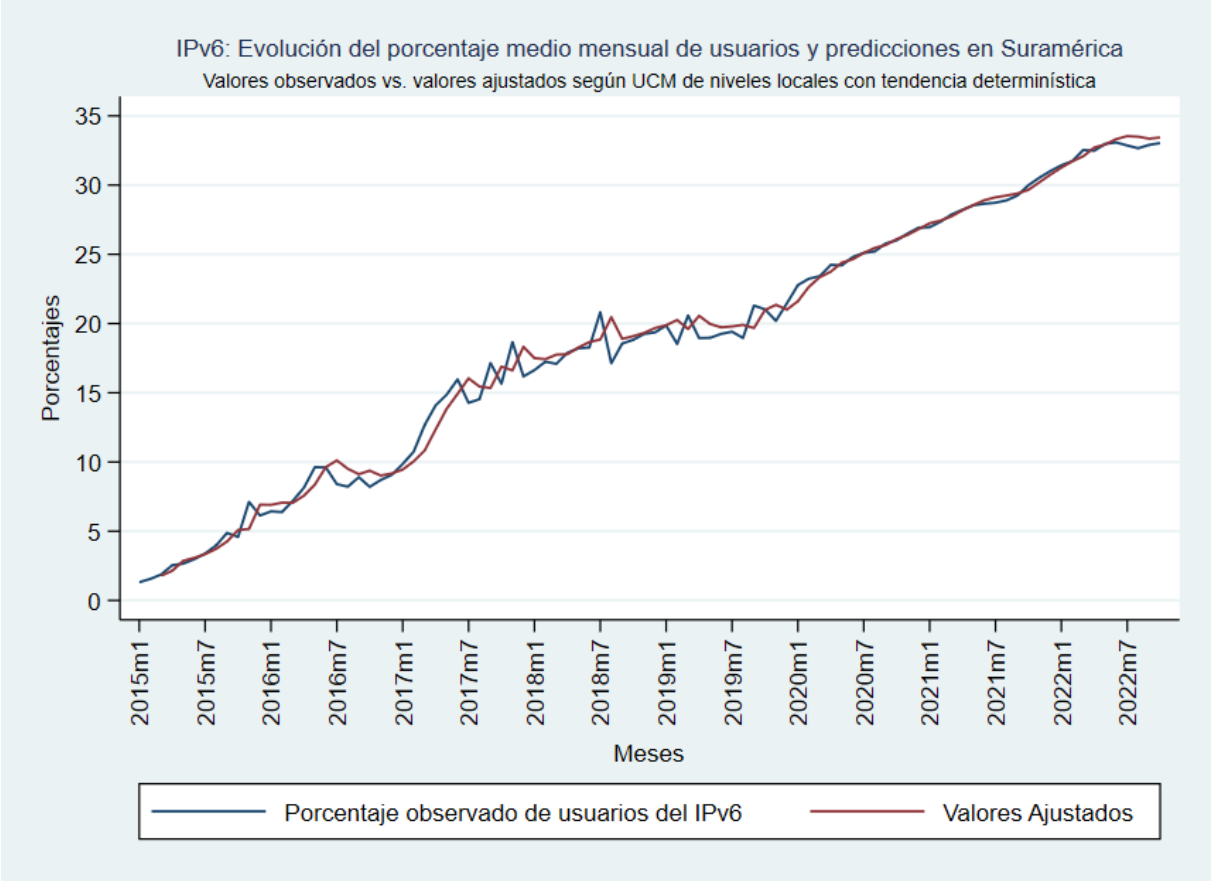
Modificando los valores de las varianzas se obtienen diferentes casos. Por ejemplo, $\sigma^2_{\xi} = 0$ implica una pendiente determinística constante: $\beta_t = \beta_{t-1} = \beta$. Tal modelo es denominado de niveles locales con pendiente determinística. Si además colocamos $\sigma^2_{\eta} = 0$ obtenemos la regresión lineal: $p_t = \mu_0 + \beta_0 t + \varepsilon_t$.

[2] Véase, por ejemplo, Pelagatti (2016), cap.3, o Baum and Hurn (2021), Sec. 9.4, para un tratamiento detallado.

Exploración y ajuste de modelos con componentes no observados

Mediante el procedimiento `ucm` del programa estadístico STATA (v17) exploramos varios modelos para la evolución del IPv6 en las regiones. En el caso de Suramérica (países con más de un 3% de usuarios), el mejor correspondió a un modelo de niveles locales con tendencia determinística. El gráfico 7 muestra los valores observados del porcentaje mensual del IPv6 y los valores correspondientes a dicho modelo.

Gráfico 7



El modelo ajustado se obtiene fijando $\sigma^2_\xi = 0$. Equivalentemente, podemos representarlo mediante las ecuaciones:

$$p_t = \mu_t + \varepsilon_t$$
$$\mu_t = \mu_{t-1} + \beta_t + \eta_t$$
$$\beta_t = \beta_{t-1} = \beta$$

Los valores estimados para los parámetros son

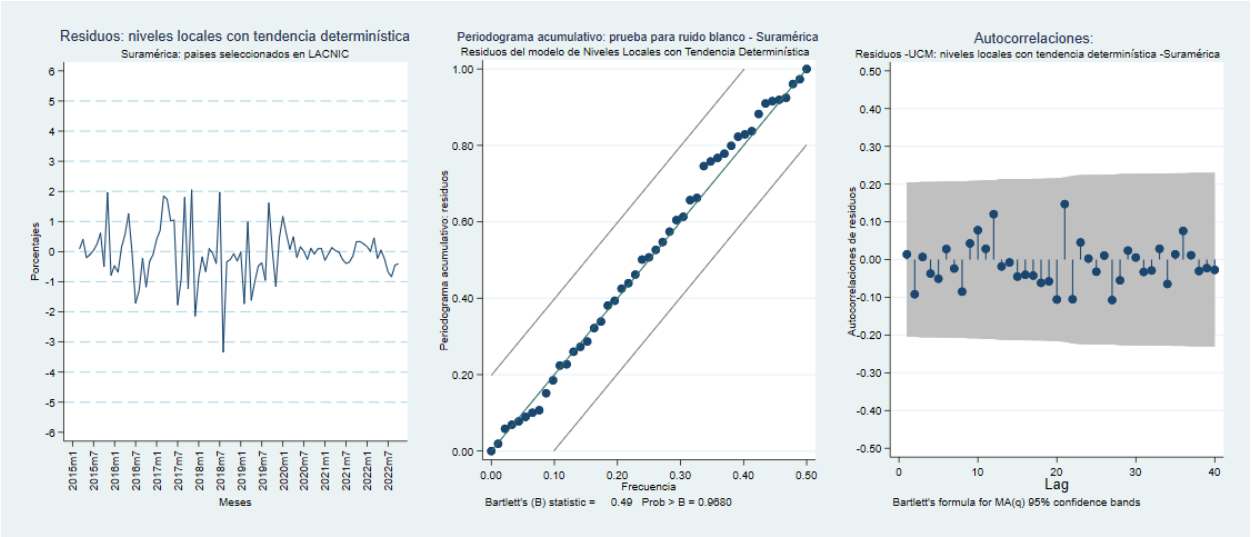
$$\sigma^2_{\grave{A}} = 0.265;\ \sigma^2_\eta = 0.329;\ \beta = 0.346$$

Substrayendo los valores ajustados a las observaciones, obtenemos los residuos del modelo. Al aplicarles la prueba Q de Portmanteau, obtenemos un valor de 34.9027, el cual no es significativo al compararlo con la distribución Chi-cuadrado respectiva, lo cual indica que no es posible diferenciar los residuos del ruido blanco.

El gráfico 8 muestra los residuos y, adicionalmente, el peridiograma y las autocorrelaciones correspondientes. El aspecto del primer diagrama (los residuos a lo largo del tiempo), en efecto, no muestra patrones sistemáticos; los valores en el peridiograma y en las autocorrelaciones nunca escapan de las bandas de confianza y las pruebas estadísticas asociadas (pruebas de Bartlett) tampoco indican violaciones a la hipótesis de que los residuos se comporten como ruido blanco.

No hay, por tanto, indicaciones de ciclos, estacionalidades u otras variaciones sistemáticas. Concluimos que el modelo explica suficientemente la evolución del despliegue del IPv6 en la región de Suramérica.

Gráfico 8



En el caso de Centroamérica, el mejor modelo UCM resultó ser una caminata aleatoria con deriva, el cual viene dado por las ecuaciones:

$$p_t = \mu_t$$
$$\mu_t = \mu_{t-1} + \beta_t + \eta_t$$
$$\beta_t = \beta_{t-1} = \beta\ .$$

Equivalentemente, en este caso se fijan las varianzas $\sigma^2_{\grave{A}} = 0$, $\sigma^2_{\xi} = 0$. Los valores estimados resultantes fueron:

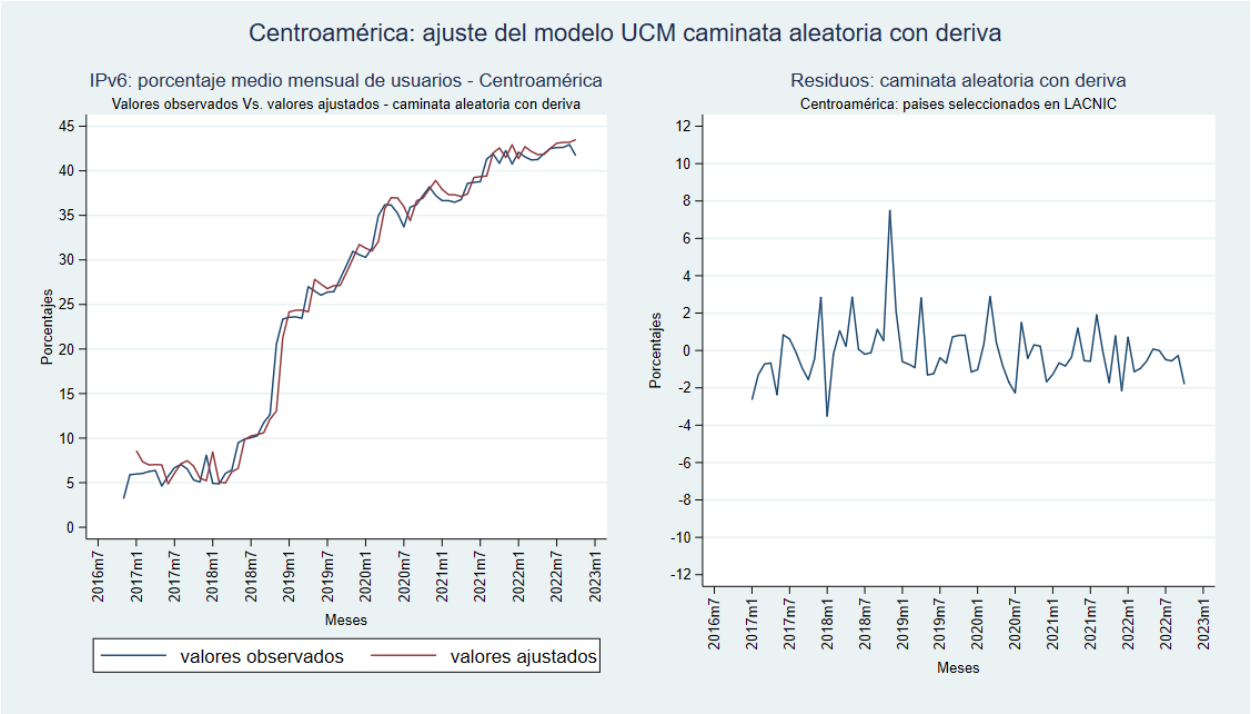
$$\sigma^2_{\eta} = 2.348;\ \beta = 0.568$$

Nótese que en este caso la pendiente es determinística.

El gráfico 9 nos muestra dos diagramas. En el primero se aprecian los valores observados del porcentaje mensual del IPv6 en Centroamérica (países seleccionados adscritos a LACNIC con más de un 3% de despliegue).

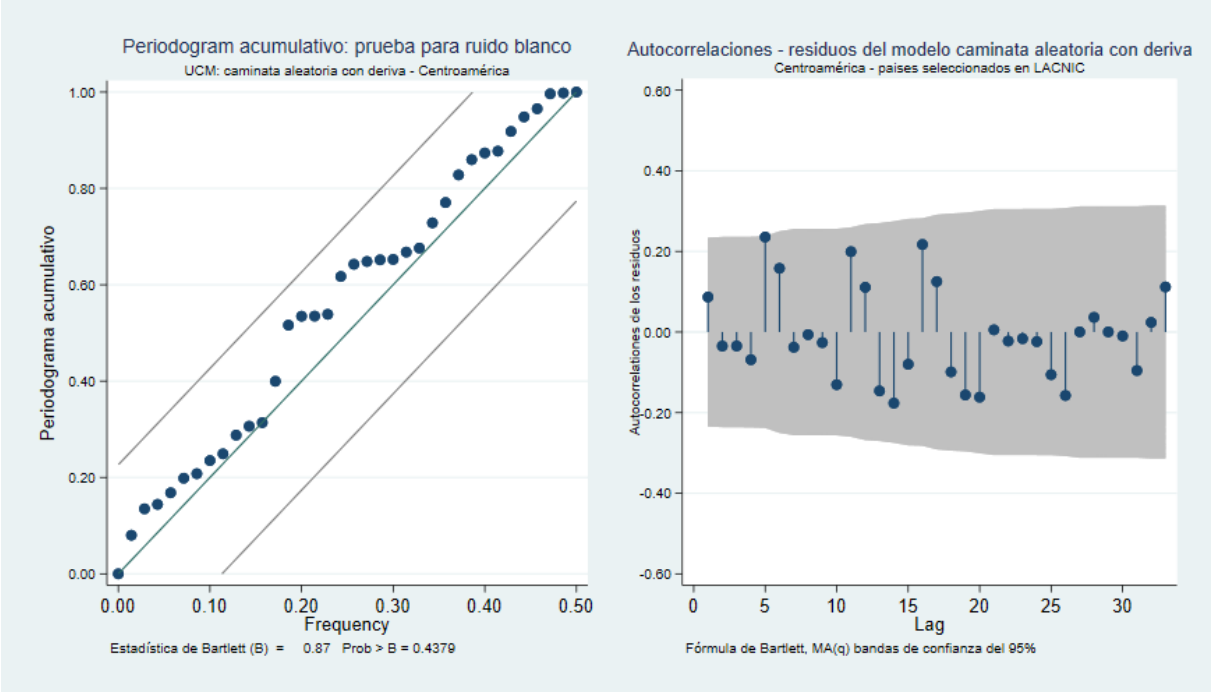
En el segundo, apreciamos los residuos del modelo, los cuales no muestran desviaciones sistemáticas del patrón del ruido blanco. Aplicando la prueba Q de Portmanteau obtenemos un valor de 37.903, el cual no es significativo al compararlo con la distribución Chi-cuadrado respectiva, indicando que no es posible diferenciar los residuos del ruido blanco.

Gráfico 9



El gráfico 10 muestra el peridiograma y las autocorrelaciones. De nuevo, las cantidades graficadas se encuentran entre las respectivas bandas de confianza y las pruebas de Bartlett no son significativas. Concluimos que el modelo no requiere agregar ciclos o estacionalidades y lo adoptamos para explicar la evolución del despliegue del IPv6 en Centroamérica.

Gráfico 10



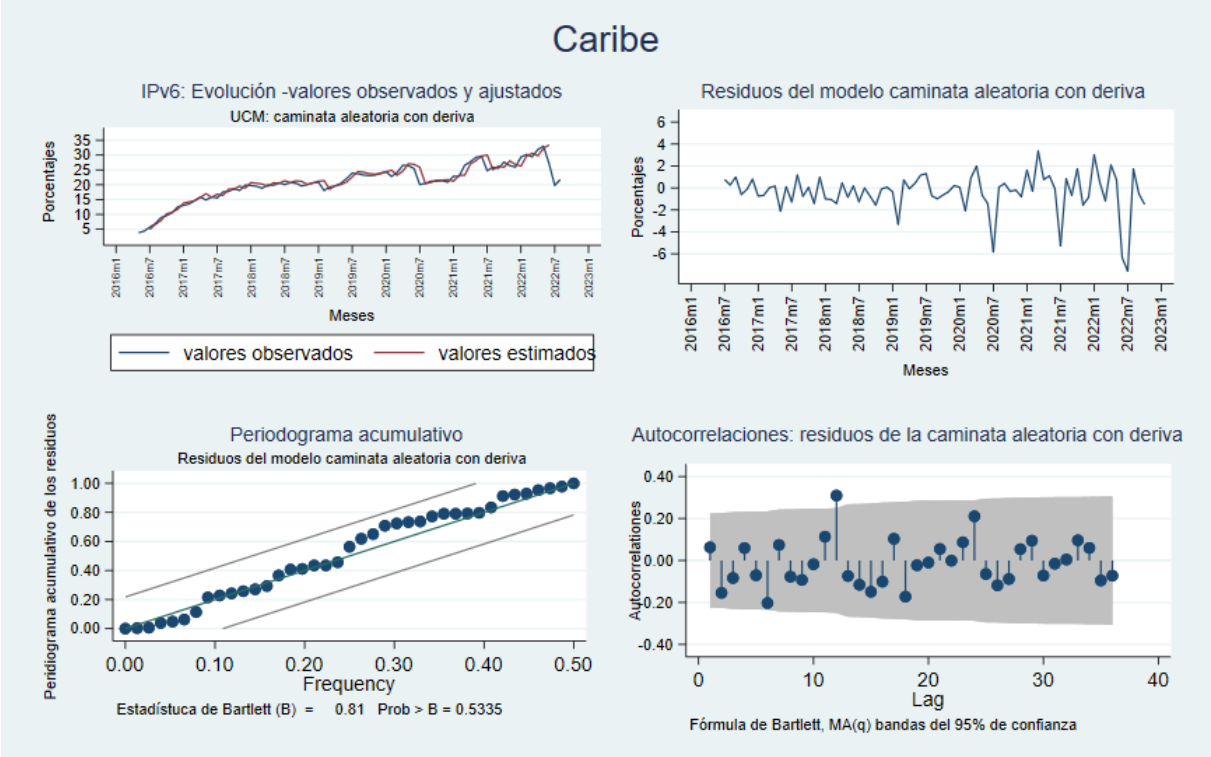
Para los países del Caribe con un despliegue mayor al 3%, el modelo que mejor se ajustó fue la caminata aleatoria con deriva, ya descrito al discutir el caso de Centroamérica.

El gráfico 11 muestra los valores ajustados y observados, los residuos, peridiograma y autocorrelaciones correspondientes a dicho ajuste. En este caso, los valores estimados de los parámetros son:

$$\sigma^2_\eta = 2.645;\ \beta = 0.319$$

Nótese que la autocorrelación correspondiente al retraso 12 ("lag" en inglés) está ligeramente por encima de las bandas de confianza, lo que podría indicar una estacionalidad mensual. Sin embargo, las otras pruebas no soportan esta idea y un intento de agregar estacionalidades al modelo no arrojó coeficientes significativos. Concluimos que el modelo explica razonablemente la evolución del porcentaje de usuarios del IPv6 en el Caribe.

Gráfico 11



Pronóstico

El principal propósito de modelar la evolución regional del despliegue del IPv6 en este trabajo es pronosticar su comportamiento y así contribuir a la toma de decisiones efectivas para la promoción del uso del protocolo en la región.

De particular interés es conocer con alguna precisión en qué momento el protocolo será adoptado por al menos el 50% de los usuarios en la región. Utilizaremos el término *proyección* para referirnos a estos pronósticos, los cuales se encuentran fuera del rango de los datos observados.

Es importante destacar que estas son proyecciones estadísticas y, como tales, se basan en el comportamiento histórico de las observaciones. En otras palabras, asumen que las variables de interés se comportarán a corto o mediano plazo de manera similar al comportamiento exhibido en su pasado reciente.

Un pronóstico estadístico es una estimación cuantitativa sobre el valor futuro más probable de una variable aleatoria, con un rango de valores futuros probables y un nivel de confianza específico.

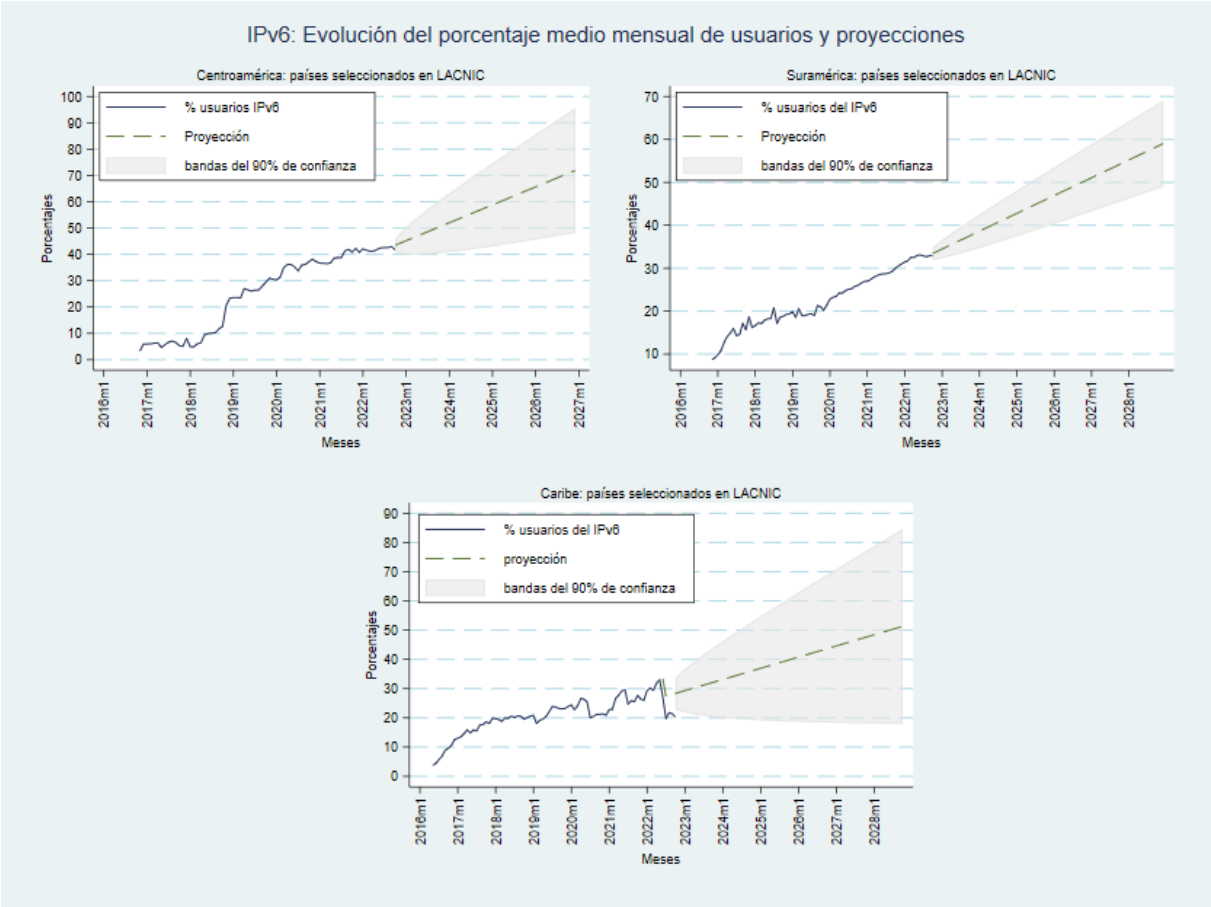
En nuestro caso, hemos escogido un nivel de confianza del 90%. Todo pronóstico estadístico está sujeto a error y es de importancia estimar su magnitud. En este estudio, el error en la predicción viene dado por la raíz cuadrada del error cuadrático medio (RECM), el cual es utilizado para construir los intervalos de confianza.[3]

El gráfico 12 presenta las proyecciones correspondientes a los modelos presentados en este trabajo. Las bandas alrededor de las proyecciones representan intervalos de confianza del 90% para cada pronóstico. Al ser proyecciones fuera del rango de las observaciones, estas tienen carácter dinámico: la proyección en el momento $t+1$ se basa en la proyección en el momento t.

A medida que nos alejamos del rango de los datos mayor es, por tanto, el error de pronóstico y mayor es la longitud de los intervalos de confianza.

[3] Véase, por ejemplo, Pelagatti, 2016, cap.1.

Gráfico 12



De acuerdo con estas proyecciones, en Centroamérica los usuarios del IPv6 alcanzarían el 51% hacia diciembre de 2023, con un valor proyectado del 51,45 % con un error (RECM) del 6,54 %. En Suramérica se alcanzaría el 51% hacia finales de enero de 2027, con un valor proyectado del 51,07% y un error del 4,74%.

En el Caribe, los usuarios del IPv6 alcanzarían el 51% hacia octubre de 2028, con un valor proyectado del 51,31% y un error del 20,26%, un alto error de predicción, ya que estamos proyectando más de 70 observaciones (meses) fuera del rango de los datos observados.

La tabla 1 recoge varias proyecciones desde enero de 2023 hasta llegar a la meta, omitiendo varios valores por brevedad.

Tabla 1: proyecciones del porcentaje medio de usuarios del IPv6

Centroamerica				
			Intervalo del 90%	
fecha	proyección	RECM	cota inf.	cota sup.
2023m1	34.476	1.283	32.365	36.587
2023m2	45.775	3.547	39.940	51.609
2023m3	46.342	3.911	39.909	52.776
2023m4	46.910	4.252	39.916	53.905
2023m5	47.478	4.575	39.952	55.004
2023m6	48.046	4.884	40.012	56.079
2023m7	48.613	5.181	40.092	57.135
2023m8	49.181	5.467	40.188	58.174
2023m9	49.749	5.746	40.298	59.200
2023m10	50.317	6.017	40.420	60.213
2023m11	50.884	6.281	40.553	61.216
2023m12	51.452	6.540	40.694	62.210
2024m1	52.020	6.794	40.844	63.196
2024m2	52.588	7.044	41.001	64.174

Suramérica				
			Intervalo del 90%	
fecha	proyección	RECM	cota inf.	cota sup.
2023m1	34.476	1.283	32.365	36.587
2023m5	35.859	1.693	33.074	38.643
2023m10	37.587	2.127	34.089	41.085
2024m1	38.625	2.364	34.737	42.512
2024m6	40.353	2.734	35.856	44.851
2025m1	42.773	3.220	37.477	48.070
2025m6	44.502	3.551	38.661	50.344
2026m1	46.922	4.000	40.343	53.501
2026m6	48.651	4.312	41.559	55.743
2026m11	50.380	4.619	42.783	57.977
2026m12	50.725	4.680	43.028	58.422
2027m1	51.071	4.740	43.274	58.868
2027m2	51.417	4.801	43.520	59.313
2027m3	51.763	4.861	43.767	59.758

El Caribe				
			Intervalo del 90%	
fecha	proyección	RECM	cota inf.	cota sup.
2023m1	29.317	4.505	21.907	36.726
2024m1	33.142	7.959	20.052	46.233
2025m1	36.968	10.808	19.189	54.746
2026m1	40.793	13.444	18.680	62.906
2027m1	44.618	15.972	18.347	70.889
2028m1	48.444	18.436	18.120	78.768
2028m3	49.081	18.842	18.089	80.074
2028m4	49.400	19.045	18.074	80.726
2028m5	49.719	19.247	18.060	81.378
2028m6	50.038	19.450	18.046	82.029
2028m7	50.356	19.652	18.032	82.680
2028m8	50.675	19.853	18.019	83.331
2028m9	50.994	20.055	18.006	83.981
2028m10	51.313	20.256	17.994	84.631

Referencias

- HARVEY, A. C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge: Cambridge University Press.
- BAUM, C.F., HURN, S. (2021). *Environmental Econometrics Using Stata*. Texas: A Stata Press Publication, StataCorp LLC.
- BOX, G. E. P., G. M. JENKINS (1970). *Time Series Analysis: Forecasting and Control.* San Francisco: Holden Day.
- PELAGATTY, M.M. (2016). Time Series Modelling with Unobserved Components. Boca Raton: CRC Press.
- STATACORP. (2021). *Stata: Release 17. Statistical Software*. College Station, TX: StataCorp LLC.