



Sector Hábitat en la ciudad de Bogotá

Modelo de pronóstico para las iniciaciones de vivienda en la ciudad de Bogotá.

Entregado

abril 2019

Publicado

mayo 2019

Guillermo Herrera Castaño
Secretario Distrital del Hábitat

Armando Ojeda Acosta
Subsecretario de Planeación y Política

Jorge Alberto Torres Vallejo
Subdirector de Información Sectorial

Cristian Andrés Torres Casallas
Investigador de la Subdirección de Información Sectorial

Contenido

Objetivo del estudio y condiciones de uso	3
Introducción	3
Determinantes del sector de la construcción en Colombia	4
Análisis exploratorio de las series.....	4
Resumen.....	9
Metodologías.	10
Modelos ARMA.....	10
Modelos VAR.....	11
Modelos VEC.....	11
Modelos dinámicos lineales.....	11
Procedimiento para la combinación de pronósticos.....	13
Regla para escoger los modelos	13
Estimaciones y principales resultados.....	14
Esquema ARMA.....	14
Esquema VAR.....	16
Esquema VEC.....	18
Esquema Modelos lineales dinámicos.....	19
Resultados finales.....	20
Lista de referencia.....	¡Error! Marcador no definido.

Objetivo del estudio y condiciones de uso

El objetivo de este documento es determinar qué tipo de modelos de series de tiempo permiten una estimación más precisa de las iniciaciones trimestrales de vivienda en la ciudad de Bogotá y determinar si existe ganancia en el uso de una metodología en particular. Para este fin, durante el estudio se explora el uso de los modelos ARMA, VEC y VAR siguiendo el trabajo de Tsay (2014) y de los modelos lineales dinámicos, siguiendo el enfoque plasmado en Petris, Petrone & Campagnoli (2007).

A partir del procedimiento de evaluación realizado, basado en la metodología aplicada por Fernando & Castaño (1998), se logró identificar que la combinación de varias metodologías permite obtener mejores resultados en la estimación de pronósticos del indicador de iniciaciones del DANE.

La Secretaría Distrital del Hábitat procura que los estudios sectoriales, publicados en la página web <https://habitatencifras.habitatbogota.gov.co/> mantengan altos estándares de calidad. Sin embargo, la entidad no asume responsabilidad alguna desde el punto de vista legal o de cualquier otra índole, por la integridad, veracidad, exactitud, oportunidad, actualización, conveniencia, contenido y/o usos que se den a la información y a los documentos que aquí se presentan.

Los análisis contenidos en los boletines sectoriales son responsabilidad exclusiva de los autores y no comprometen a la Secretaría Distrital del Hábitat. Así mismo la Secretaría Distrital del Hábitat no asume responsabilidad alguna por omisiones de información o por errores en la misma, en particular por las discrepancias que pudieran encontrarse entre la versión electrónica de la información publicada y su fuente original.

Introducción

La variable “*iniciaciones de vivienda*” representa un pilar fundamental en el análisis del sector construcción, ya que posibilita dimensionar la capacidad del mercado formal para atender las necesidades de habitacionales según los perfiles de ingreso de los hogares y además permite medir los niveles de actividad y expectativa del sector construcción en el mediano plazo. En ese sentido, contar con una herramienta estadísticamente robusta para el pronóstico de esta variable en el corto y mediano plazo es relevante en tanto permite comprender los fundamentales de la actividad edificadora.

El desarrollo de este documento contiene una primera parte donde se determinan las variables a ser utilizadas tomando como punto de referencia el trabajo de FEDESARROLLO (2004). En la segunda parte se realiza un análisis exploratorio de las series de tiempo utilizadas en la estimación de los modelos. La tercera parte se centra en la descripción de las metodologías de estimación con base en el trabajo de Tsay (2014) y se aplica la metodología de evaluación a partir del trabajo de Fernando & Castaño (1998). Finalmente, se presentan los principales resultados y conclusiones.

Determinantes del sector de la construcción en Colombia

Para tener una idea inicial sobre las variables a utilizar en la estimación de los modelos¹, y las relaciones que entre ellas existe, se tomó como punto de referencia los trabajos del MVCT (2018) y Fedesarrollo (2004). Estos análisis buscan los determinantes de la actividad edificadora en el país, usando como proxy las variables *Área aprobada en licencias de construcción de vivienda*, *Área aprobada en licencias de construcción totales* y *Área de obras en construcción*. Para explicar el comportamiento de dichas variables, los documentos mencionan los siguientes determinantes:

1. Tasa de colocación de créditos hipotecarios (r), fuente: Banco de Republica.
2. Tasa de inflación (π), fuente: Departamento Administrativo Nacional de Estadística - DANE.
3. Desembolsos hipotecarios ($CredH$) deflactados por índice de Precios al Consumidor – IPC, fuente: DANE, esta se discrimina por constructor e individuales.
4. El Índice de Precios al Consumidor Arriendo - IPCA deflacto con el IPC, fuente: DANE.
5. Índice de la tasa de cambio real ($ITCR$), fuente: Banco de Republica.
6. Índice de Costos de la Construcción – ICCV deflactado por el IPC, fuente: DANE, este indicador no muestra alta influencia en el agregado total pero sí parece ser un determinante fuerte en el mercado VIS, fuente: DANE.
7. Índice de ingresos laborales reales (I), fuente: DANE.
8. Precio de la vivienda nueva – IPVN, fuente: DANE.

El enfoque utilizado en los trabajos ya mencionados para determinar la magnitud y el tipo de relación de cada una de estas variables con los niveles de actividad edificadora en la ciudad nace de la estimación de la forma reducida de un modelo de oferta y demanda, el cual tiene como supuesto fundamental la existencia de un precio de equilibrio en el mercado inmobiliario. En ese sentido, como paso siguiente, se tomaron las variables utilizadas por MVCT y Fedesarrollo y se realizó un análisis exploratorio exhaustivo para entender su dinámica.

Análisis exploratorio de las series.

Esta parte del documento contiene un breve resumen estadístico de las series usadas en el documento y una depuración de las variables de mayor interés, donde se incluyen el *Número licencias de construcción aprobadas* y *Número de iniciaciones de vivienda* a partir de tres enfoques:

1. El primero basado en la metodología ARIMA,
2. El segundo fundamentado en el uso del rango intercuartílico
3. El tercero a partir de un análisis histórico de las dos series teniendo en cuenta intervenciones que modificaron la tendencia del indicador como lo fueron la entrada en vigencia de la Norma Sismo Resistente – NSR10 en 2011, los efectos que sobre el licenciamiento tiene programas como el VIPA en la ciudad y las modificaciones al Plan de Ordenamiento Territorial – POT de la ciudad ocurridos entre 2012 y 2013.

Debido al impacto que tiene la presencia de puntos atípicos al momento de realizar estimaciones o pronósticos, se realizó la depuración de las series de iniciaciones y licenciamiento de vivienda

¹ Es importante mencionar que la mayoría de los análisis fueron realizados sobre los crecimientos anuales de las series definidos como $\log(X_{t-4}) - \log(X_t)$.

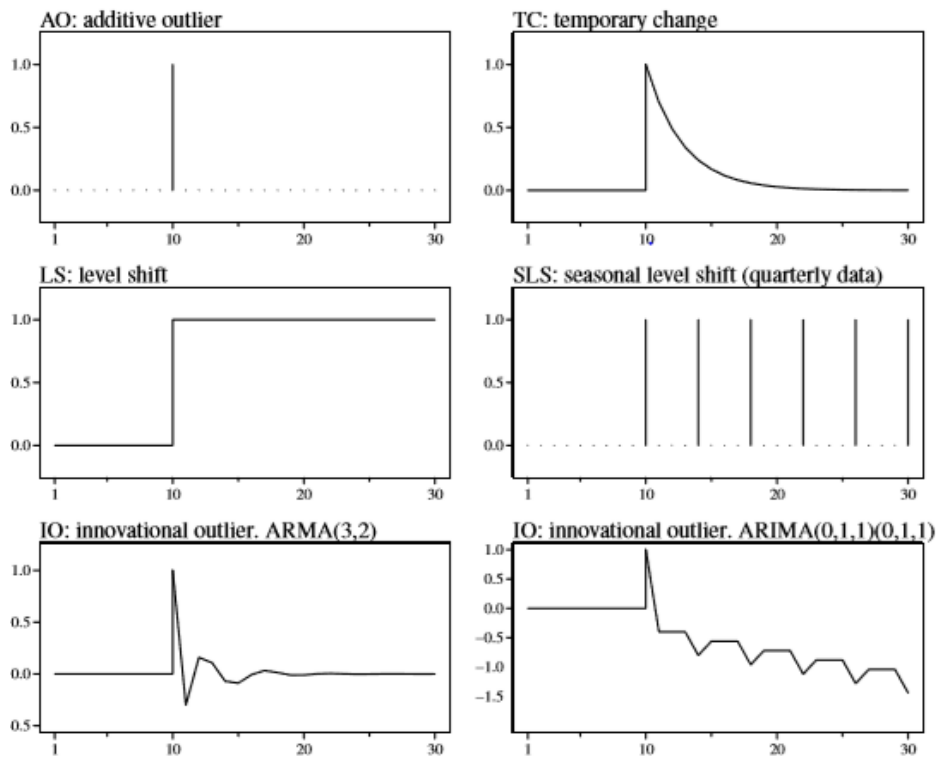
utilizando el procedimiento descrito en el paquete “*tsoutliers*” (Chung & Liu, 1993) del programa estadístico R. El procedimiento, propuesto por Calle (2018), supone que los valores atípicos que no pueden ser explicados por información a priori² pueden ser expresados como $L(B)I(t_j)$, donde $I(t_j)$ representa una variable indicadora que toma el valor de 1 en $t = j$ y 0 en otra parte y $L(B)$ un polinomio de rezago.

Esta representación permite clasificar las perturbaciones de la serie según la forma del polinomio de rezago asociado. Los cinco tipos de atípicos que pueden ser identificados con esta metodología son:

- Outlier Innovador - IO: Para este caso el polinomio de rezago toma la forma $\frac{\theta(B)}{\alpha(B)\phi(B)}$.
- Outlier Aditivo - AO: Para este caso el polinomio de rezago toma la forma 1.
- Cambio de Nivel - LS: Para este caso el polinomio de rezago toma la forma $\frac{1}{(1-B)}$.
- Cambio Temporal - TC: Para este caso el polinomio de rezago toma la forma $\frac{1}{(1-\delta B)}$.
- Cambio Nivel Estacional - SLS: Para este caso el polinomio de rezago toma la forma $\frac{1}{(1-B^s)}$.

Generalmente se toma $\delta = 0.7$ y s como la periodicidad de los datos, gráficamente se tiene que:

Ilustración 1 Comportamiento gráfico de los distintos tipos de atípicos.



Fuente: Grafica tomada de Calle (2018).

Si se asume una serie y_t^* observada con m atípicos, con peso w_j para cada uno, se tiene que el modelo ARIMA asociado a esta serie viene dado por:

² Esto permite incluir este efecto dentro del modelo como una variable de intervención.

$$y_t^* = \sum_{j=1}^m w_j L_j(B) I_t(t_j) + \frac{\theta(B)}{\alpha(B)\phi(B)} a_t \quad eq 1$$

Con errores estimados dados por.

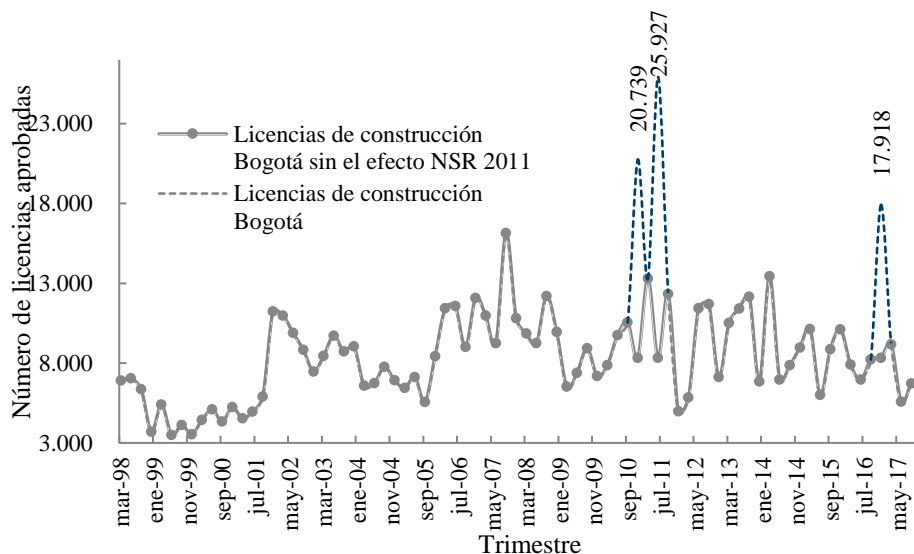
$$\pi(B)^3 y_t^* \equiv \hat{e}_t = \sum_{j=1}^m w_j \pi(B) L_j(B) I_t(t_j) + a_t \quad eq 2$$

Las ecuaciones 1 y 2 son ecuaciones fundamentales de la detección de atípicos. Es importante mencionar que antes de efectuar la mencionada depuración se hizo un análisis gráfico e histórico de las series reemplazando por la mediana aquellos datos considerados atípicos y que su presencia está sustentada por información a priori⁴.

Licencias de construcción Bogotá.

La ilustración 2 muestra el comportamiento de la serie “Licencias aprobadas para construcción de vivienda” antes y después de remover el efecto de la Norma Sismo Resistente de 2011 sobre el VI trimestres de 2010 y el II de 2011, así como el efecto del programa VIPA en el licenciamiento VIS al final de 2016.

Ilustración 2 Serie de licenciamiento antes y después de los efectos NSR y VIPA.



Fuente: DANE. Elaboración SIS - SDHT.

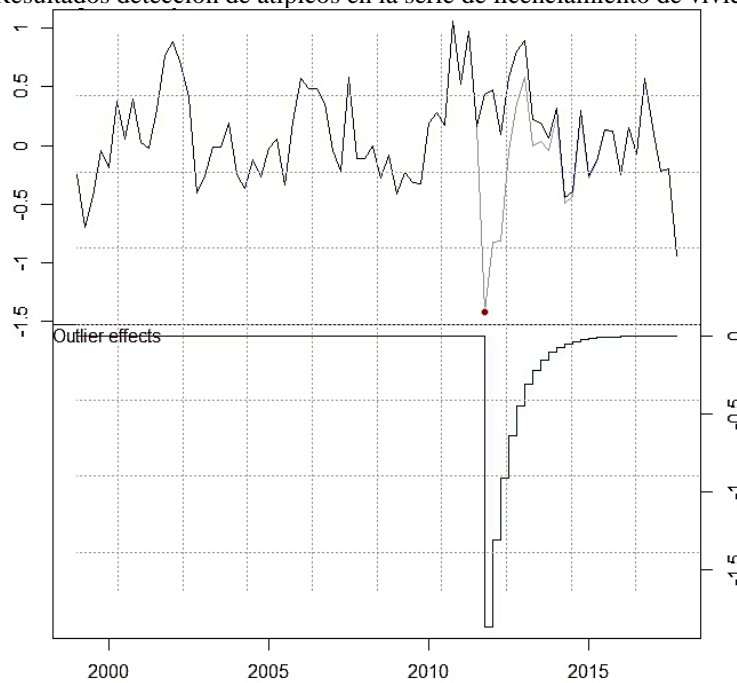
Por su parte, la ilustración 3 muestra los resultados de la aplicación de la metodología descrita en el apartado anterior sobre la serie “Crecimiento anual del licenciamiento” en la ciudad de Bogotá. En este punto se descuenta el efecto NSR. La ilustración incluye la serie original, la serie ajustada y el

³ Los valores para los coeficientes de $\pi(B)$ pueden ser determinados a través de la relación $\pi(B) = \frac{\theta(B)}{\alpha(B)\phi(B)}$

⁴ Un ejemplo de esto es el efecto generado sobre las licencias de construcción aprobadas durante 2011 en la ciudad por el anuncio de la entrada en vigor de la Norma Sismo Resistente, hecho que incentivó una avalancha de licencias con el fin de evitar los costos extras de aprobación en un periodo posterior a la norma.

efecto de los extremos sobre la serie, de manera que se puede apreciar la presencia de un dato atípico en el cuarto trimestre de 2011 del tipo TC.

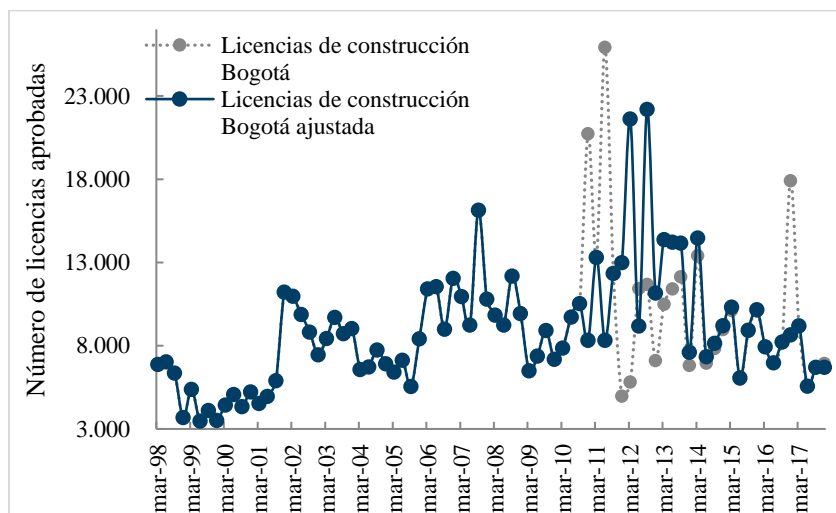
Ilustración 3 Resultados detección de atípicos en la serie de licenciamiento de vivienda en Bogotá.



Fuente: Elaboración SIS - SDHT.

A continuación, la ilustración 4 muestra el resultado del ajuste final realizado sobre la serie original de licencias de construcción de vivienda en la ciudad de Bogotá. Debido a que el atípico encontrado es del tipo TC el ajuste sobre la serie afecta varios periodos posteriores al IV trimestre de 2011 prolongándose hasta I trimestre de 2015.

Ilustración 4 Resultado ajuste serie de licencias de construcción (trimestrales).

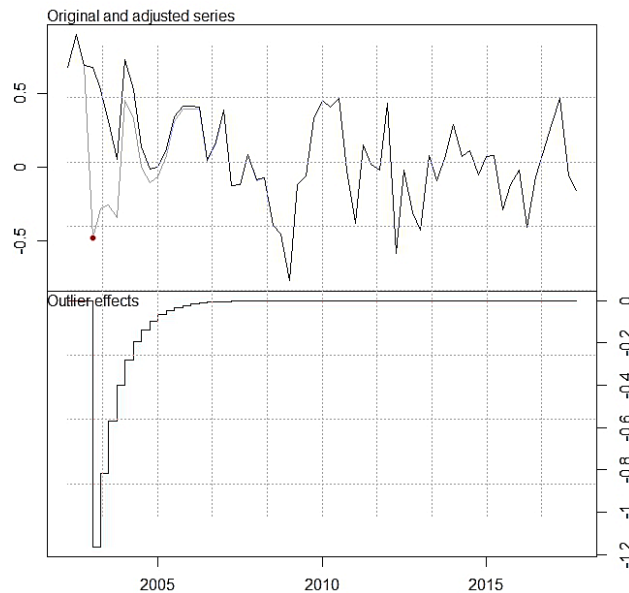


Fuente: DANE. Elaboración SIS - SDHT.

Iniciaciones de vivienda en Bogotá.

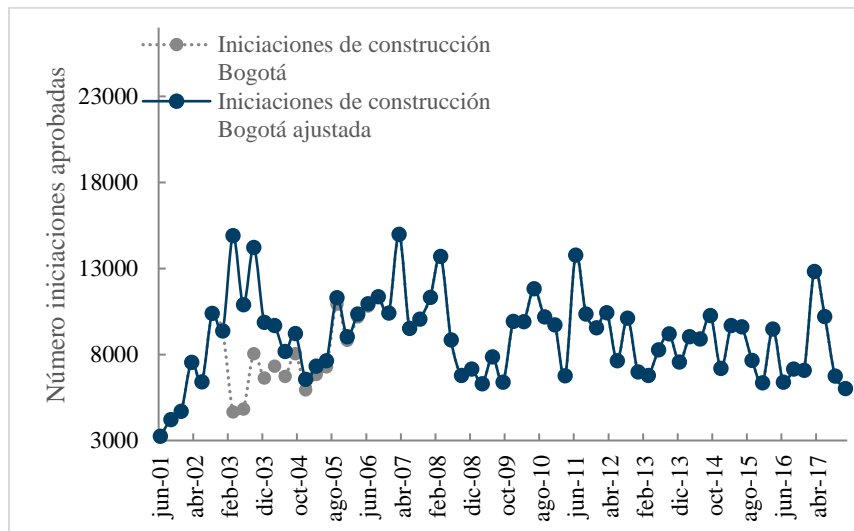
La aplicación de la metodología descrita sobre la variable iniciaciones de vivienda arroja evidencia sobre la presencia un dato atípico en el I trimestre de 2003 de tipo TC ver ilustración 4. Al parecer la serie de iniciaciones no fue afectada de manera importante por los cambios normativos y programas de política de 2011 y 2016 como a la serie de licenciamiento, esto podría explicarse por el hecho de que la decisión de iniciar una obra depende principalmente de un factor comercial, es decir, de la velocidad con la que se vende el proyecto durante su periodo de preventa y por ende del entorno macroeconómico. Lo que hace que no necesariamente la serie iniciaciones refleje los mismos choques de la serie licencias, esto por los dos años de vigencia que se ofrecen entre la licencia de construcción y la iniciación.

Ilustración 5 Resultados detección de atípicos en la serie de iniciaciones de vivienda en Bogotá.



Fuente: Elaboración SIS - SDHT.

Ilustración 6 Resultados serie de iniciaciones original y ajustada de vivienda en Bogotá



Fuente: DANE. Elaboración SIS - SDHT.

Para los cálculos y análisis finales se usaron las series depuradas, removiendo los atípicos reportados por las tres metodologías análisis con información a priori, análisis estructural de las series y por último el uso del rango entre cuartiles.

Resumen.

Tabla 1 Estadística de resumen principales variables.

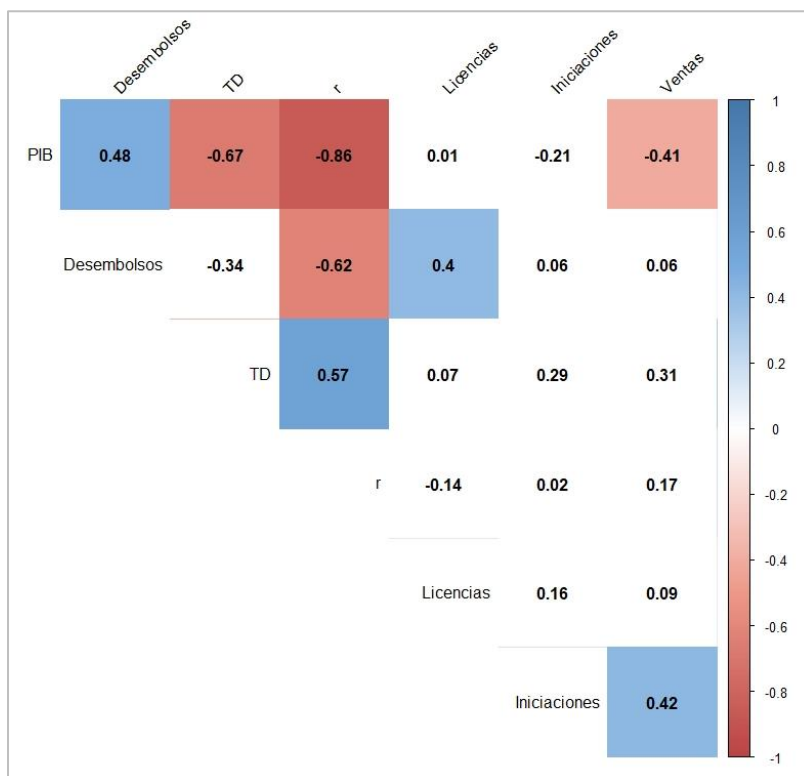
	Licencias	Iniciaciones	TD	PIB	r	Desembolsos	Ventas
Min. :	5.556	6.027	7.793	19.791	11,19%	3.983	2.204
1st Qu. :	7.555	7.286	9.873	24.461	12,42%	8.730	6.867
Median. :	9.014	9.496	11.448	27.971	13,29%	10.584	8.197
Mean. :	9.775	9.305	11.434	28.592	14,08%	9.925	8.384
3rd Qu. :	11.011	10.347	12.645	32.951	16,14%	11.420	9.988
Max. :	22.206	14.992	17.003	37.186	18,10%	13.964	13.544
RIQ. :	3.456	3.061	2.772	8.490	3,72%	2.690	3.121
SD	3.445	2.086	1.79	4.736	2,07%	2.324	2.198
Asimetría	1.649	0.645	0.014	0.0458	0.55	-0.909	0.033

Fuente: DANE, Banco de la Republica y Galería Inmobiliaria. Elaboración SIS - SDHT.

La tabla 1 muestra las estadísticas resumen de las principales variables involucradas en el documento, en ella se verifican los altos grados de asimetría en las variables licencias de construcción y desembolsos para adquisición de vivienda las cuales presentaron los valores más extremos en el coeficiente de asimetría 1.64 y -0.90 respectivamente. Así mismo, al analizar la media y la mediana de las series, se puede apreciar la cercanía entre los niveles de actividad de las variables licencias de construcción, iniciaciones de vivienda, ventas y desembolsos de crédito para compra de vivienda, lo cual es aún más evidente si se analiza la matriz de correlaciones contemporáneas de las series (ilustración 7), donde se encuentran fuertes relaciones contemporáneas positivas entre los pares de variables iniciaciones – ventas 42% y licencias – desembolsos 40%.

Con el fin de determinar si existe o no evidencia estadística sobre algún grado de asociación no contemporánea entre las variables licencias, ventas, desembolsos e iniciaciones de vivienda se calculan y grafican las correlaciones cruzadas para estas variables hasta tres rezagos atrás. Estos resultados pueden ser analizados en el Anexo 10 donde se puede encontrar fuerte evidencia para plantear a las variables ventas, licencias y desembolsos como buenos predictores de las iniciaciones de vivienda en la ciudad de Bogotá. Se destacan los fuertes niveles de correlación presentes entre las variables licencias e iniciaciones en el segundo rezago 19%, ventas e iniciaciones en el segundo rezago 57% y entre las variables desembolsos e iniciaciones en el primer rezago 16%.

Ilustración 7 Matriz de correlaciones principales variables.



Fuente: DANE, Banco de la Republica y Galería Inmobiliaria. Elaboración SIS - SDHT.

Metodologías.

A continuación, se hace una breve descripción de las metodologías aplicadas en la construcción de la batería de modelos a usar en el pronóstico de la variable iniciaciones de vivienda.

Modelos ARMA.

Seguendo a (Shumway. R & Stoffer. D, 2011) una serie de tiempo $\{x_t; t = 0, 1, 2 \dots\}$ puede ser descrita a través de un proceso ARMA con orden autoregresivo p y de media móvil q , si esta es estacionaria y puede ser representada adecuadamente a través de:

$$x_t = \alpha + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p} + w_t + \theta_1 w_{t-1} + \dots + \theta_q w_{t-q}$$

Con $\phi_p \neq 0$, $\theta_p \neq 0$, $\sigma_w^2 > 0$ y media dada por $\mu = \frac{\alpha}{(1-\phi_1-\phi_2-\dots-\phi_p)}$. Para el caso en que $q = 0$ la representación anterior es conocida como un modelo autoregresivo de orden p y cuando $p = 0$ se le denomina modelo de media móvil con orden q . La forma más común y útil de representar un modelo ARMA es a través de la utilización de los polinomios de rezago de tal manera que el modelo viene dado por $\phi(B)x_t = \theta(B)w_t$ donde $\phi(B) = 1 - B - B^2 - \dots - B^p$ y $B^p x_t = x_{t-p}$.

Modelos VAR.

Los modelos de vectores autoregresivos o VAR son la herramienta de mayor uso en la literatura económica para realizar análisis y pronósticos sobre series de tiempo multivariadas. Siguiendo (Tsay, 2014) el comportamiento de una serie multivariada Z_t está regido por un modelo VAR si

$$Z_t = \phi_0 + \sum_{i=1}^p \phi_i Z_{t-i} + a_t$$

Donde ϕ_0 es un vector de constantes de tamaño q , ϕ_i es un conjunto de matrices $(k \times k)$ para $i > 0$ y a_t es una secuencia de vectores aleatorios independientes e idénticamente distribuidas con media cero y matriz de varianzas y covarianzas Σ_a la cual es definida positiva. Este modelo puede ser representado como $\phi(B)Z_t = \phi_0 + a_t$, donde $\phi(B) = I_t - \sum_{i=1}^p \phi_i B^i$.

Modelos VEC.

Como es bien conocido el concepto de raíz unitaria⁵ juega un papel muy importante en el análisis de series de tiempo, especialmente cuando estas representan grandes agregados económicos. En la literatura económica una serie de tiempo es integrada de orden d o $I(d)$ si $(1 - B)^d z_t$ es estacionaria e invertible donde $d > 0$.

Supongamos una serie de tiempo multivariada z_t tal que cada una de las series que la componen z_{it} es $I(1)$ pero existe una combinación lineal no trivial de las misma $y_t = \beta' z_t$ tal que y_t es un proceso $I(0)$, en este caso β es denominado el vector de cointegración⁶. Uno de los ejemplos más comunes es la literatura económica es la ecuación cuantitativa del dinero, la cual asume una relación a largo plazo entre el producto de la masa monetaria y su velocidad con la multiplicación entre la producción total y sus niveles de precio. La gran ventaja de este enfoque es que permite estimar los efectos de corto y largo plazo de cada una de las variables analizadas sobre las demás, permitiendo la velocidad de ajuste del conjunto de series para llegar a su punto de equilibrio de largo plazo.

Modelos dinámicos lineales.

Los modelos lineales dinámicos son un caso particular de un modelo estado espacio, el cual define una serie de tiempo como el resultado de la interacción de un sistema dinámico perturbado por un ruido aleatorio, permitiendo a este enfoque ser usado en análisis tanto univariados como multivariados, donde existen patrones irregulares, no estacionariedad o cambios estructurales en las series. En cuanto a la generación de pronósticos esta metodología usa la solución de la ecuación integral *Wiener – Hopf* propuesta por (Kalman, 1960) y conocida como el filtro de Kalman. Todo el desarrollo básico de estos modelos aplicado a las series de tiempo aparece en los artículos (Akaike, 1974), (Harrison & Stevens, 1976) y (West & Harrison, 1997). La idea fundamental detrás de los

⁵ Una serie de tiempo con una raíz unitaria puede ser representada como: $Z_t = Z_{t-1} + a_t$ donde a_t es una secuencia de variables aleatoria independientes e idénticamente distribuidas – iid, este proceso describe una variable cuyo comportamiento está regido completamente por el comportamiento de una camina aleatoria lo cual lo hace casi impredecible.

⁶ Generalmente los procesos de cointegración implican relaciones de largo plazo entre las series de tiempo analizadas.

modelos estado espacio es considerar el vector aleatorio $Y_t = (Y_{(1,t)}, Y_{(2,t)}, \dots, Y_{(m,t)})^7$ como una función ruidosa e incompleta de algún proceso no observable $(\theta_t, t = 1, 2, 3, \dots)$, donde θ_t puede verse como una variable auxiliar que facilita la tarea de encontrar la ley de probabilidad de Y_t y la cual sigue una dinámica Markoviana. Formalmente se tiene que un modelo estado espacio puede ser planteado de la siguiente manera:

- Sea $(\theta_t, t = 1, 2, 3, \dots)$ una cadena de Markov, es decir, su valor en t depende exclusivamente de θ_{t-1} , así su ley probabilidad queda especificada totalmente a través de una densidad inicial $p_0(\theta_0)$ y una densidad de transición $p(\theta_t|\theta_{t-1})$ de θ_t condicionada a θ_{t-1} .
- Los Y_t 's son independientes condicionados sobre θ_t , y los Y_t dependen solamente de θ_t , con lo cual para algún $n \geq 1$, $(Y_1, Y_2, \dots, Y_n)|\theta_1, \theta_2, \dots, \theta_n$ tiene una densidad condicional conjunta dada por $\prod_{t=1}^n f(y_t|\theta_t)$.

Cuando la variable de análisis es continua a este esquema se le denomina un modelo estado espacio y un modelo de oculto de Markov si la variable es discreta. Uno de los resultados más importantes de este enfoque es que Y_t es independiente de su pasado condicionado a los valores de θ_t , es decir, θ_t resume toda la información relevante de la serie analizada y es suficiente para su pronóstico. Dentro de este marco metodológico un modelo lineal dinámico viene especificado por el siguiente sistema de ecuaciones:

$$\begin{aligned} Y_t &= F_t \theta_t + v_t, & v_t &\sim N(0, V_t) \\ \theta_t &= G_t \theta_{t-1} + w_t, & w_t &\sim N(0, W_t) \end{aligned}$$

Donde G_t y F_t son matrices conocidas y v_t y w_t son secuencias de ruido blanco independientes entre sí y su propio pasado, con media cero y matrices de covarianzas V_t y W_t respectivamente, adicionalmente se supone que $\theta_0 \sim N_p(m_0, C_0)$. Aunque estos modelos tienen una gran flexibilidad, al momento de efectuar un pronóstico el investigador se enfrenta al problema de realizar una especificación adecuada del modelo a usar, aunque existen casos donde es posible construir un modelo estado espacio donde el estado tiene una interpretación intuitiva y el conocimiento de los expertos permite conocer las matrices de transición esta no es regla.

Para los análisis univariados se seguirá lo planteado en (Giovanni, 2018), enfoque que permite ejecutar un análisis bayesiano dentro del marco de los dlm y además da vía libre a la adición de modelos para representar una misma serie de tiempo. Si se dispone de k modelos independientes para las mismas observaciones m dimensionales, donde el i -ésimo modelo esta dado por:

$$\begin{cases} y_t^{(i)} = F_t^{(i)} \theta_t^{(i)} + v_t^{(i)}, & v_t^{(i)} \sim N(0, V^{(i)}) \\ \theta_t^{(i)} = G_t^{(i)} \theta_{t-1}^{(i)} + w_t^{(i)}, & w_t^{(i)} \sim N(0, W^{(i)}) \end{cases}$$

Con $m_0^{(i)}$ y $C_0^{(i)}$ la media y la varianza del estado inicial en el modelo i , para este caso se tiene que el vector de estados puede variar de dimensión para cada modelo, de esta manera cada modelo puede representar una característica particular de la serie: tendencia estocástica, componente periódico, una tendencia determinística, entre otras, así la observación en el periodo t viene dada por $y_t = y_t^1 + y_t^2 + \dots + y_t^k$.

⁷ Un ejemplo de esto pueden ser los precios de m acciones en el periodo t

Procedimiento para la combinación de pronósticos.

Para la estimación de los ponderados usados en la combinación lineal de los pronósticos realizados sobre la serie de iniciaciones de vivienda, se sigue el trabajo realizado por (Granger, C.W.J & Ramanathan, R, 1984), el cual sugiere usar un modelo de regresión lineal de la forma:

$$y_t = \beta_0' + \beta_1' f_1 + \beta_2' f_2 + \dots + \beta_k' f_k \text{ para } t = N + 1, N + 2, \dots, N + h$$

Donde y_t representa los valores dentro de muestra de la serie no usados en la estimación y f_i el i -ésimo pronóstico de la serie. Granger & Ramanathan (1984) muestran que si se cumplen los supuestos del modelo de regresión lineal clásico este procedimiento arroja pronósticos insesgados con error cuadrático medio menor que el de cualquier pronóstico individual involucrado. Algunas derivaciones a este procedimiento surgen de la necesidad de superar el incumplimiento de los supuestos del modelo de regresión clásico⁸, estas derivaciones permiten la presencia de errores autocorrelacionados, coeficientes cambiantes en el tiempo, presencia de colinealidad entre los diferentes pronósticos y presencia de datos atípicos.

Regla para escoger los modelos

Siguiendo a Fernando & Castaño (1998) el primer paso en la selección de modelos radica en evaluar las propiedades de calidad básicas de los pronósticos (no sesgo, errores de pronóstico con media cero y normalidad en los desfases). Para evaluar la no presencia de sesgos en los pronósticos se emplea la ecuación de (Holden & Peel, 1989).

$$y_t = a + b f_t^j + u_t$$

Donde f_t^j representa el pronóstico del modelo j para el periodo t , para esta representación se acepta la hipótesis de no sesgo si: $a = 0$ y $b = 1$. Con el fin de evaluar la superioridad de un modelo sobre los demás⁹ se usa la prueba de "Encompassing", la cual permite determinar si una combinación de pronósticos es conveniente o no, siguiendo la formalización realizada por Chong & Hendry (1986) el contraste para dos modelos viene dado por:

$$y_t = a + b_1 f_t^1 + b_2 f_t^2 + u_t$$

Donde si $\begin{bmatrix} a \\ b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$ se dice que el modelo 1 incorpora toda la información relevante del pronóstico

del modelo 2 y si $\begin{bmatrix} a \\ b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ se dice que el modelo 2 incorpora toda la información relevante del pronóstico del modelo 1. Por último, para verificar la significancia estadística de la ganancia aportada

⁸ Por ejemplo, para el caso en que existen relaciones de colinealidad, la literatura sugiere la aplicación de métodos como los componentes principales, y la regresión de raíces latentes.

⁹ Esta hipótesis es evaluada al verificar si algún pronóstico incorpora toda la información relevante de los demás

por cada metodología se evalúan los cambios en el Error Cuadrático Medio – ECM el cual viene dado por:

$$ECM = \frac{1}{h} \sum_{t=T}^{T+h} (y_{t+h} - \hat{y}_{t+h})^2$$

Donde h representa el número de pasos hacia adelante que toma el pronóstico realizado. Para esta prueba se usa el estadístico planteado por (Grnager & Newbold, 1977).

Estimaciones y principales resultados.

El objetivo de este documento implica el análisis del vector $Y_t = (Y_{(1,t)}, Y_{(2,t)}, Y_{(3,t)}, \dots, Y_{(m,t)})$, el cual representa el registro para la ciudad de Bogotá de las variables trimestrales

- Número de licencias de construcción de vivienda, fuente DANE.
- Número de iniciaciones de construcción de vivienda, fuente DANE.
- Tasa de colocación de créditos hipotecarios (r), fuente Banco de Republica.
- Tasa de inflación (π), fuente Departamento Administrativo Nacional de Estadística - DANE.
- Desembolsos hipotecarios ($CredH$) deflactados por índice de Precios al Consumidor – IPC, fuente DANE, esta se discrimina por constructor e individuales.
- El Índice de Precios al Consumidor Arriendo - IPCA deflacto con el IPC, fuente DANE.
- Índice de Costos de la Construcción – ICCV deflactado por el IPC, fuente DANE, este indicador no muestra alta influencia en el agregado total pero sí parece ser un determinante fuerte en el mercado VIS.
- Tasa de desempleo (D).
- Precio de la vivienda nueva – IPVN.

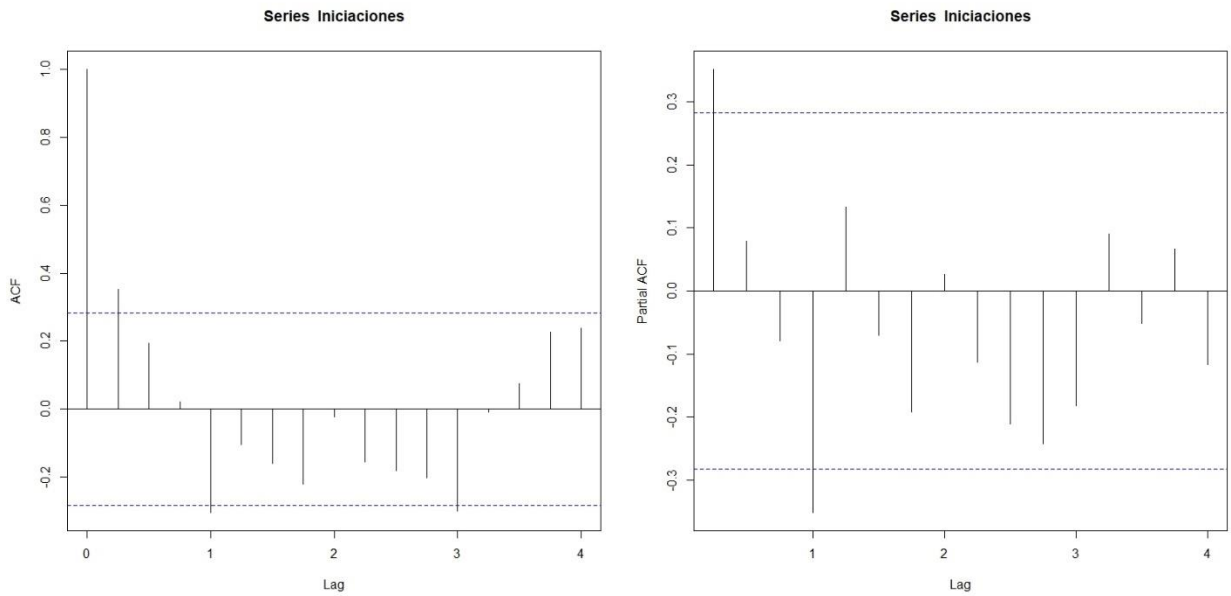
La construcción de pronósticos desde el marco teórico de los modelos estado espacio en este documento tiene dos enfoques: el primero se centra en el análisis univariado de la serie de interés iniciaciones de vivienda; y el segundo, se centra en la comprensión de la estructura de correlación a través del tiempo de las variables analizadas.

Esquema ARMA.

Para la selección de los modelos se empleó el Criterio de Información Bayesiano – BIC, por sus siglas en inglés¹⁰ y el análisis grafico de las funciones de autocorrelación simple y parcial ilustración 8, el primer criterio arroja un modelo ARMA con dos rezagos en la media móvil y un rezago estacional como el mejor candidato (Modelo 1), mientras que el análisis grafico arroja un modelo ARMA con dos rezagos autorregresivos como el modelo más adecuado para modelar la serie (Modelo 2).

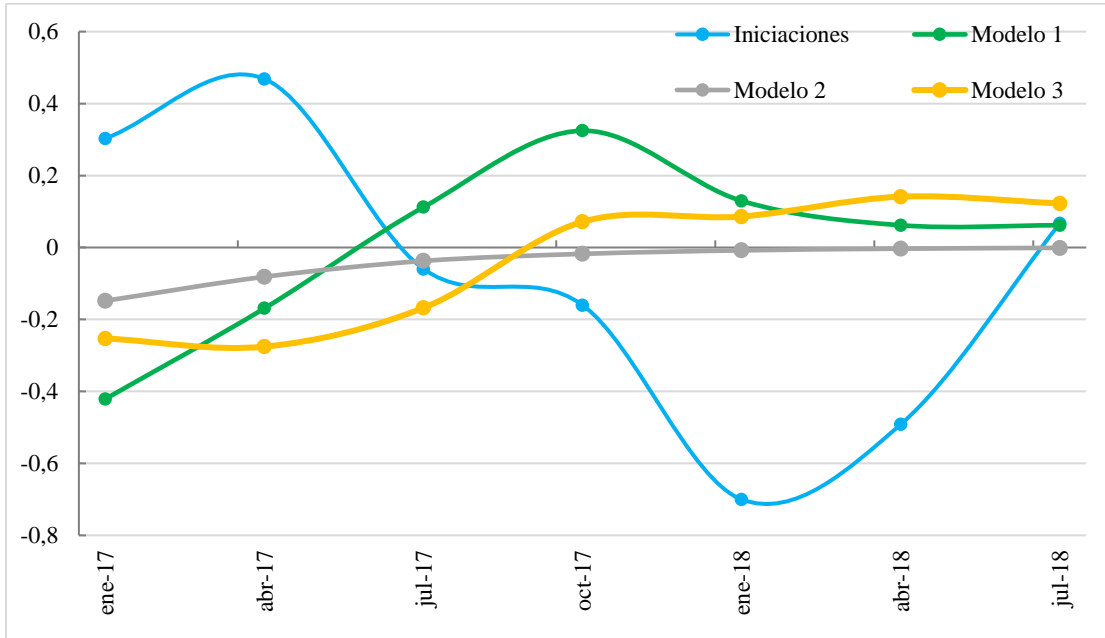
¹⁰ Esta aplicación se realizó a través de la función *auto.arima* del programa estadístico R tomado de: <https://cran.r-project.org/web/packages/forecast/forecast.pdf>

Ilustración 8 Función de autocorrelación simple y parcial.



Fuente: DANE. Elaboración SIS - SDHT.

Ilustración 9 Crecimiento iniciaciones de vivienda Vs pronósticos ARMA.



Fuente: DANE. Elaboración SIS - SDHT.

Al analizar los residuales de pronóstico de los dos modelos se tiene que no existe evidencia estadística para rechazar las hipótesis nulas de normalidad ¹¹ y media igual a cero en ninguno de los dos casos. Sin embargo, en cuanto a la hipótesis de no sesgo, ningún modelo pasa la prueba. Adicional a los modelos ya analizados se probaron algunas otras especificaciones de las cuales se seleccionó un modelo ARMA con 4 rezagos autorregresivos el cual supera los supuestos de normalidad, media cero, el modelo continuó sin pasar el supuesto de no presencia de sesgo (Modelo 3).

Esquema VAR.

Para el enfoque VAR se testearon dos escenarios¹²: uno donde se usaron todas las series de tiempo disponibles y cuya importancia dentro de la actividad edificadora esta ya verificada y un segundo escenario donde se tomaron las series de *Iniciaciones*, *Licencias*, *Ventas* y *PIB*.

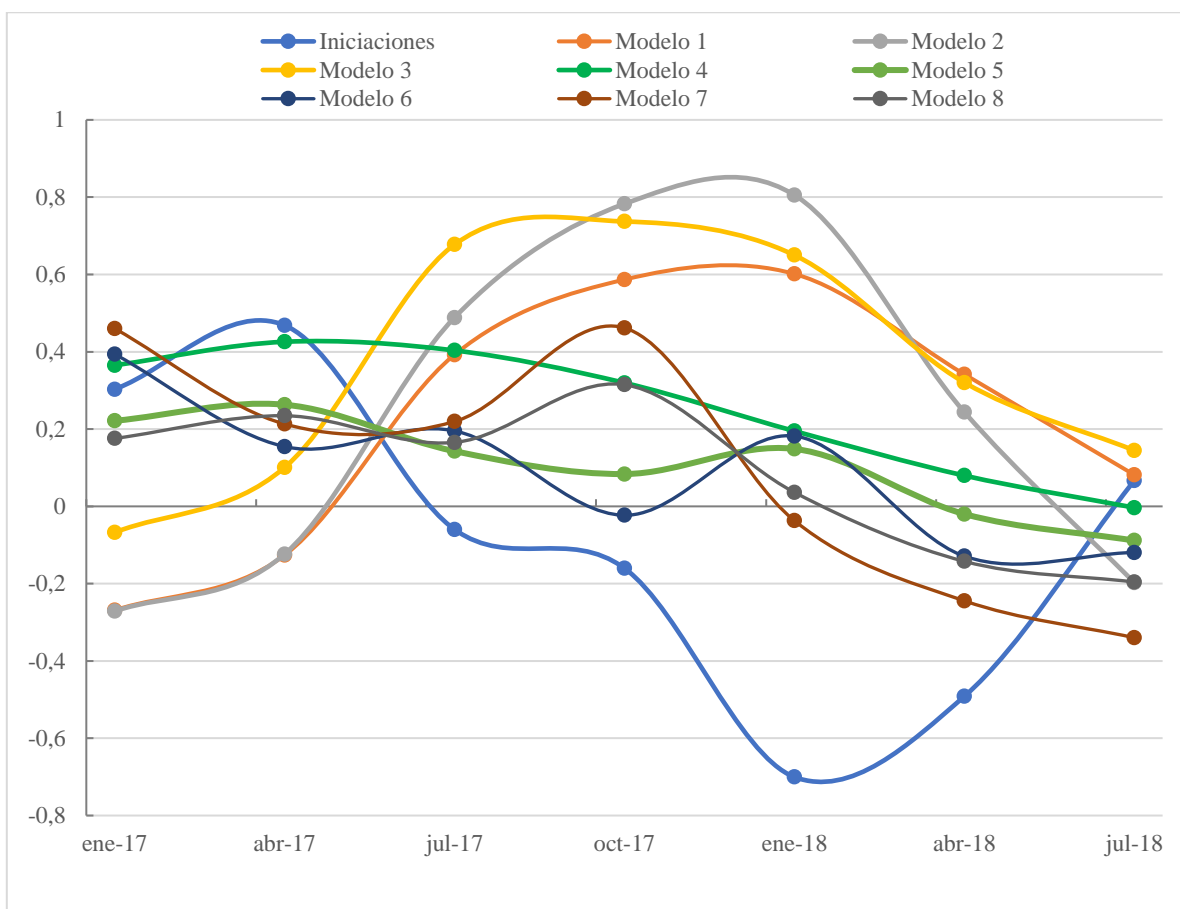
Escenario 1: En esta parte se seleccionaron 8 modelos con diferentes combinaciones entre las características determinísticas (tendencia y constante) y la parte autorregresiva del modelo.

- Modelo 1: modelo con dos rezagos autorregresivos sin constante ni tendencia (Pasa supuesto de media cero y normalidad, falla en el supuesto de no presencia de sesgo).
- Modelo 2: modelo con dos rezagos autorregresivos, constante y sin tendencia (Pasa supuesto de media cero y normalidad, falla en el supuesto de no presencia de sesgo).
- Modelo 3: modelo con dos rezagos autorregresivos con tendencia y sin constante (Pasa supuesto de media cero y normalidad, falla en el supuesto de no presencia de sesgo).
- Modelo 4: modelo con un rezago autorregresivo, tendencia y sin constante (Pasa normalidad y no presencia de sesgo, falla en el supuesto de media cero).
- Modelo 5: modelo construido con las variables Licencias, Iniciaciones, PIB, Tasa de desempleo y Ventas, tres rezagos autorregresivos, tendencia y sin constante (Pasa normalidad, no presencia de sesgo y media cero).
- Modelo 6: modelo construido con las variables Licencias, Iniciaciones, PIB, Tasa de desempleo y Ventas, tres rezagos autorregresivos, tendencia y constante (Pasa normalidad, no presencia de sesgo y media cero).
- Modelo 7: modelo construido con las variables Licencias, Iniciaciones, PIB y Ventas, tres rezagos autorregresivos, tendencia y constante (Pasa normalidad y media cero, falla en el supuesto de no presencia de sesgo).
- Modelo 8: modelo construido con las variables Licencias, Iniciaciones, PIB, Desembolsos y Ventas, cuatro rezagos autorregresivos, tendencia y constante (Pasa normalidad y media cero, falla en el supuesto de no presencia de sesgo).

¹¹ Para este caso se usó la prueba de Shapiro.

¹² Para cada uno de los escenarios se probaron modelos sin constante ni tendencia, modelos con constante, modelos con solo tendencia y modelos con ambas características. Para la selección de los rezagos para cada una de las especificaciones se usó la función *VARselect* (Ptaff 2008) del programa R.

Ilustración 10 Crecimiento iniciaciones de vivienda Vs pronósticos VAR.



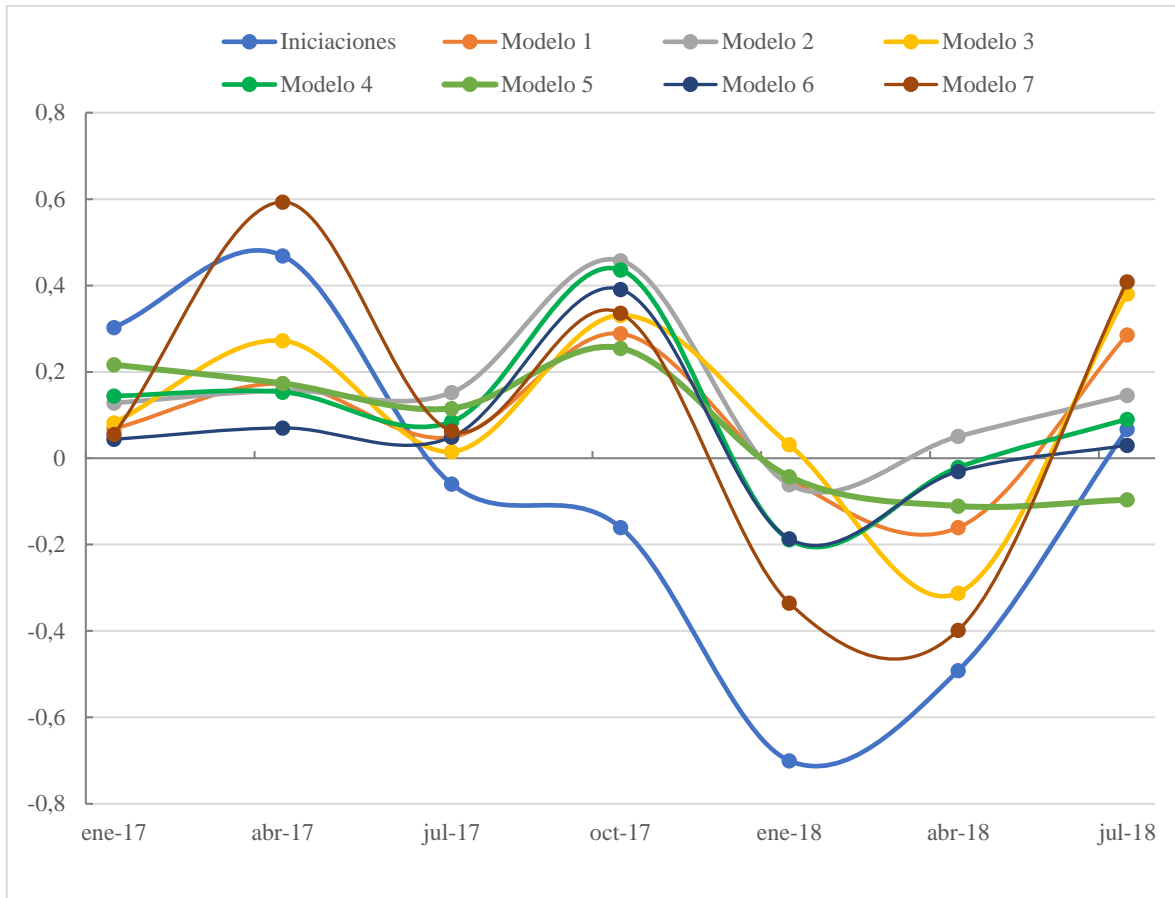
Fuente: DANE. Elaboración SIS - SDHT.

Escenario 2: Para esta parte solo se consideran las variables de Iniciaciones, Licencias, Ventas y PIB.

- Modelo 1: modelo con seis rezagos autorregresivos, tendencia y sin constante (Pasa supuesto de media cero, normalidad y de no presencia de sesgo).
- Modelo 2: modelo con cinco rezagos autorregresivos, constante y sin tendencia (Pasa supuesto de media cero, normalidad y de no presencia de sesgo).
- Modelo 3: modelo con seis rezagos autorregresivos, constante y tendencia (Pasa supuesto de media cero, normalidad y de no presencia de sesgo).
- Modelo 4: modelo con cinco rezagos autorregresivos, constante y tendencia (Pasa supuesto de media cero, normalidad y de no presencia de sesgo).
- Modelo 5: modelo con cuatro rezagos autorregresivos, constante y tendencia (Pasa supuesto de media cero, normalidad y de no presencia de sesgo).
- Modelo 6: modelo con cinco rezagos autorregresivos, constante y sin tendencia (Pasa supuesto de media cero, normalidad y de no presencia de sesgo).

- Modelo 7: modelo con ocho rezagos autorregresivos, constante y tendencia (Pasa supuesto de media cero, normalidad y de no presencia de sesgo), este fue construido con la información de las variables PIB, Iniciaciones y Ventas.

Ilustración 11 Crecimiento iniciaciones de vivienda Vs pronósticos VAR.



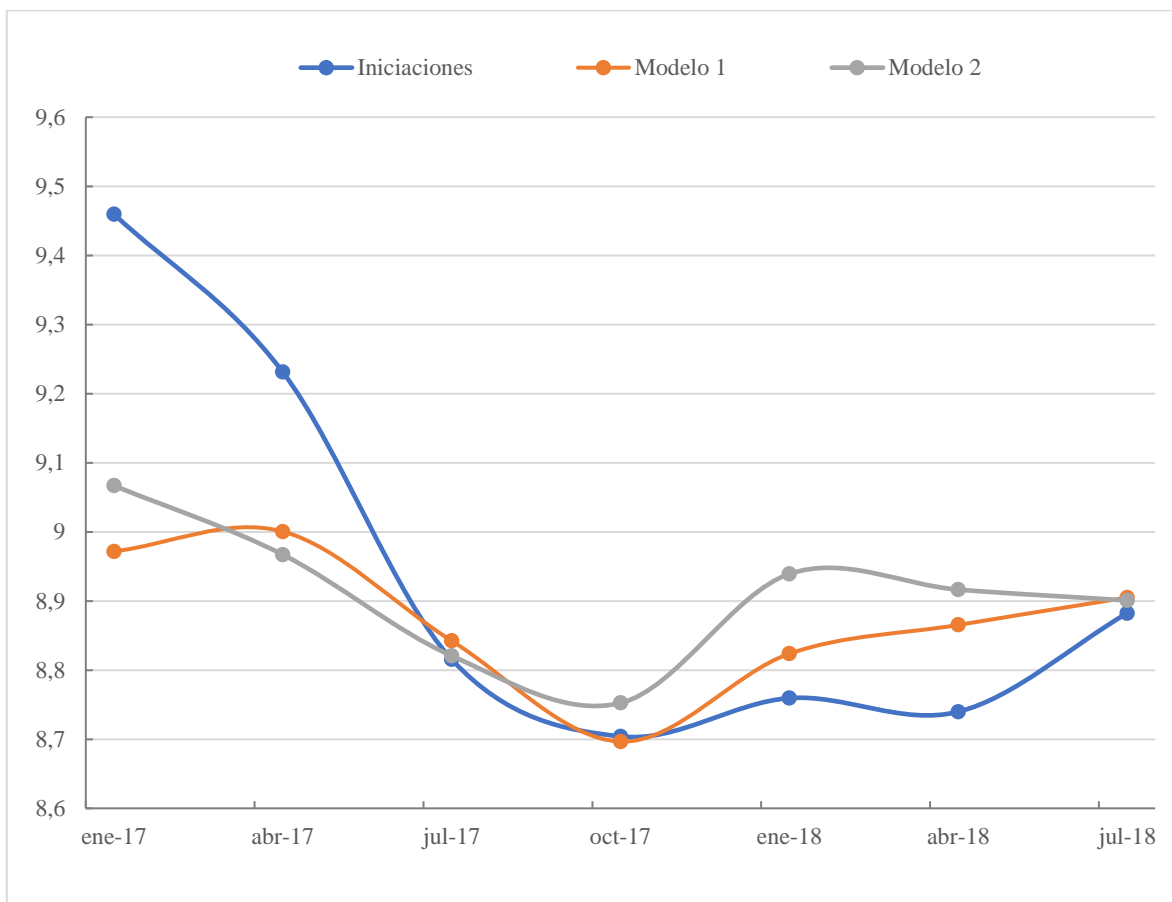
Fuente: DANE. Elaboración SIS - SDHT.

Esquema VEC

Se emplea debido a la necesidad de aprovechar la relevante relación de largo plazo que existe entre el licenciamiento de vivienda, las ventas y el posterior inicio de la obra.

- Modelo 1: modelo con seis rezagos autorregresivos y una relación de cointegración (Pasa supuesto de media cero, normalidad, pero incumple el supuesto de no presencia de sesgo).
- Modelo 2: modelo con seis rezagos autorregresivos y una relación de cointegración (Pasa supuesto de media cero, normalidad e incumple el supuesto de no presencia de sesgo) solo tienen en cuenta las variables de iniciaciones y ventas.

Ilustración 12 Logaritmo de las iniciaciones de vivienda Vs pronósticos VAR.

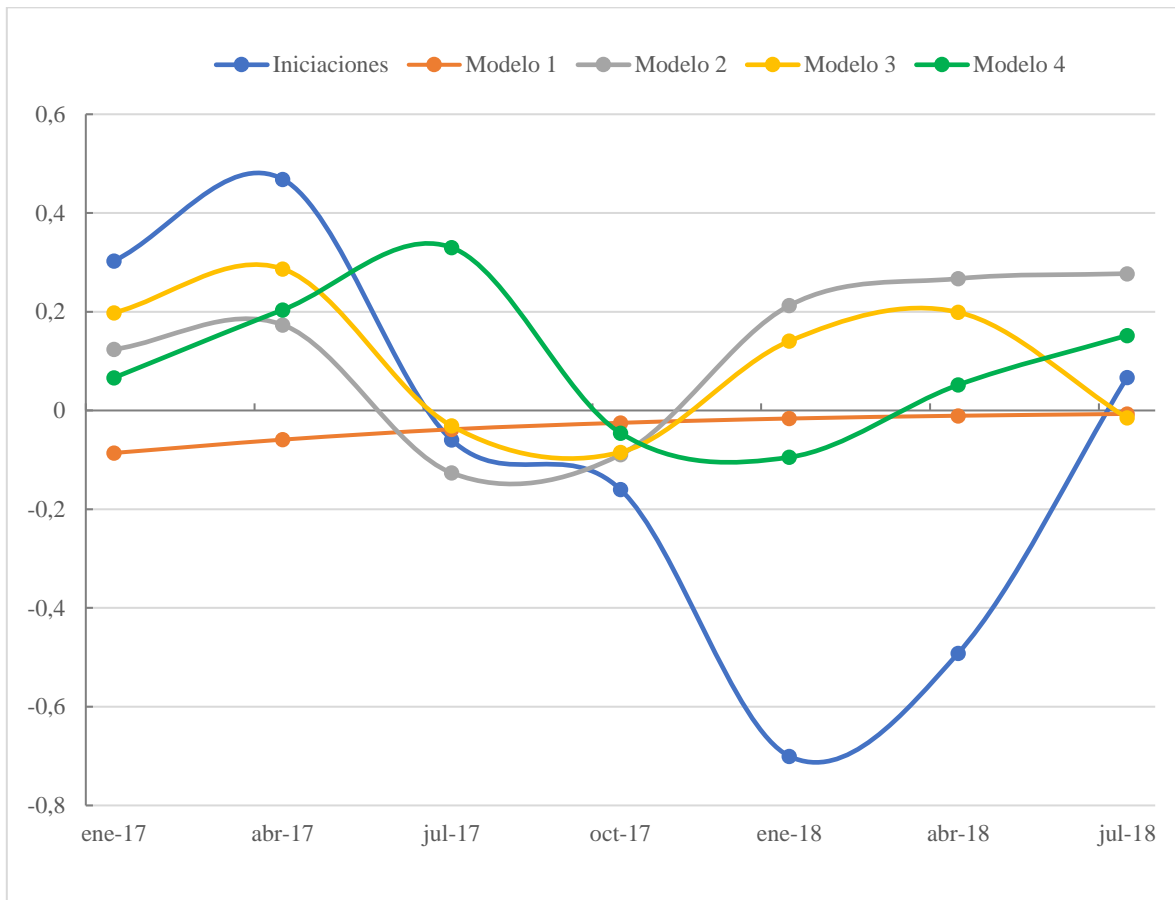


Fuente: DANE. Elaboración SIS - SDHT.

Esquema Modelos lineales dinámicos

Para esta sección se probaron diferentes representaciones de los modelos Arma tanto univariados (Modelos del 1 al 3) como multivariados (Modelo 4) todos dentro de la representación de los modelos lineales dinámicos. Se tomaron los promedios de pronóstico de cada modelo.

Ilustración 13 Crecimiento iniciaciones de vivienda Vs pronósticos modelo lineal dinámico.



Fuente: DANE. Elaboración SIS - SDHT.

Resultados finales

Después de seleccionar los modelos con mejor comportamiento se decide trabajar el proceso de combinación de pronósticos con los siguientes modelos.

Esquema VAR con todas las variables relevantes.

- Modelo 5: modelo construido con las variables Licencias, Iniciaciones, PIB, Tasa de desempleo y Ventas, tres rezagos autorregresivos, tendencia y sin constante (Pasa normalidad, no presencia de sesgo y media cero).
- Modelo 6: modelo construido con las variables Licencias, Iniciaciones, PIB, Tasa de desempleo y Ventas, tres rezagos autorregresivos, tendencia y constante (Pasa normalidad, no presencia de sesgo y media cero).
- Modelo 7: modelo construido con las variables Licencias, Iniciaciones, PIB y Ventas, tres rezagos autorregresivos, tendencia y constante (Pasa normalidad y media cero, falla en el supuesto de no presencia de sesgo).
- Modelo 8: modelo construido con las variables Licencias, Iniciaciones, PIB, Desembolsos y Ventas, cuatro rezagos autorregresivos, tendencia y constante (Pasa normalidad y media cero, falla en el supuesto de no presencia de sesgo).

Esquema VAR con las variables Indicaciones, Licencias, Ventas y PIB.

- Modelo 4: modelo con cinco rezagos autorregresivos, constante y tendencia (Pasa supuesto de media cero, normalidad y de no presencia de sesgo).
- Modelo 5: modelo con cuatro rezagos autorregresivos, constante y tendencia (Pasa supuesto de media cero, normalidad y de no presencia de sesgo).
- Modelo 6: modelo con cinco rezagos autorregresivos, constante y sin tendencia (Pasa supuesto de media cero, normalidad y de no presencia de sesgo).
- Modelo 8: modelo construido con las variables Licencias, Iniciaciones, PIB, Desembolsos y Ventas, cuatro rezagos autorregresivos, tendencia y constante (Pasa normalidad y media cero, falla en el supuesto de no presencia de sesgo).

Esquema VEC

- Modelo 1: modelo con seis rezagos autorregresivos y una relación de cointegración (Pasa supuesto de media cero, normalidad, pero incumple el supuesto de no presencia de sesgo).
- Modelo 2: modelo con seis rezagos autorregresivos y una relación de cointegración (Pasa supuesto de media cero, normalidad e incumple el supuesto de no presencia de sesgo) solo se tienen en cuenta las variables de iniciaciones y ventas.

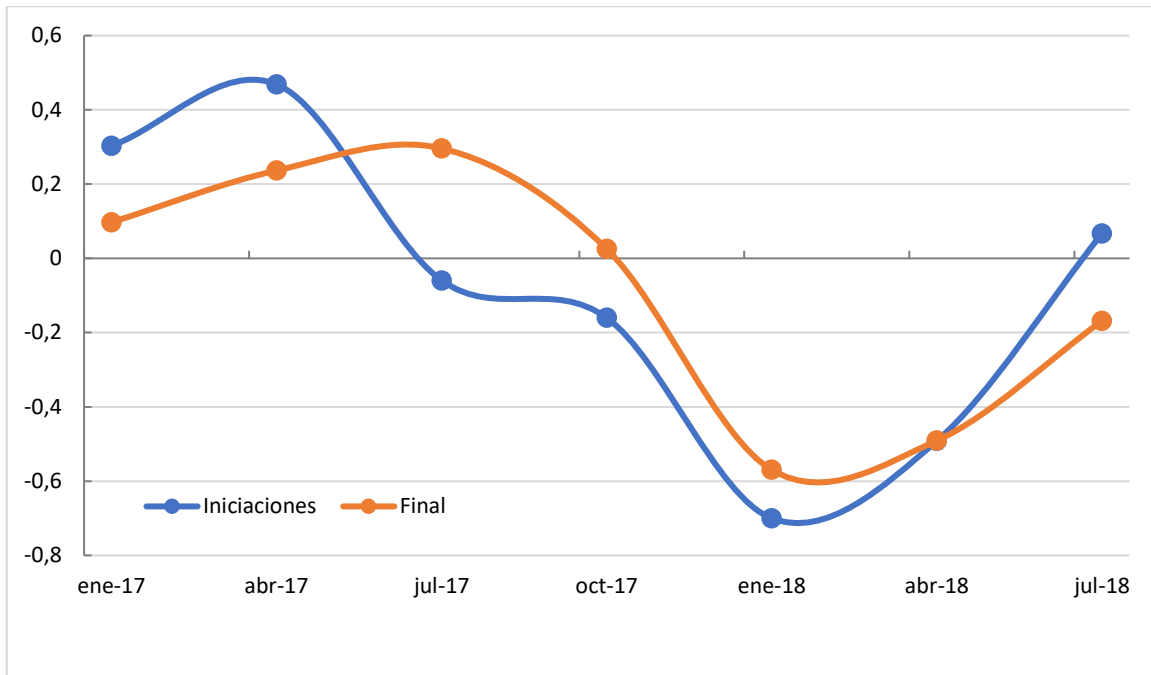
Esquema modelos lineales dinámicos

- Modelo 1: Modelo arma univariado.
- Modelo 4: Modelo ARMA bivariado.

Después de trabajar con distintas combinaciones de los resultados se decidió que, dada la información disponible, el horizonte de pronóstico (7 periodos) y las metodologías testeadas, el mejor modelo para combinar y generar un pronóstico final es:

$$y_{t+h} = -0.4239 + 0.3861y_{t+h}^{\text{Modelo VAR 5}} + 0.5812y_{t+h}^{\text{Modelo VAR 6}} \\ + 0.8651y_{t+h}^{\text{Modelo VAR 4 escenario 2}} + 0.33112y_{t+h}^{\text{Modelo VAR 8 escenario 2}} \\ + 1.34690y_{t+h}^{\text{Modelo lineal dinamico bivariado}}$$

Ilustración 14 Crecimiento iniciaciones de vivienda Vs pronóstico combinado.



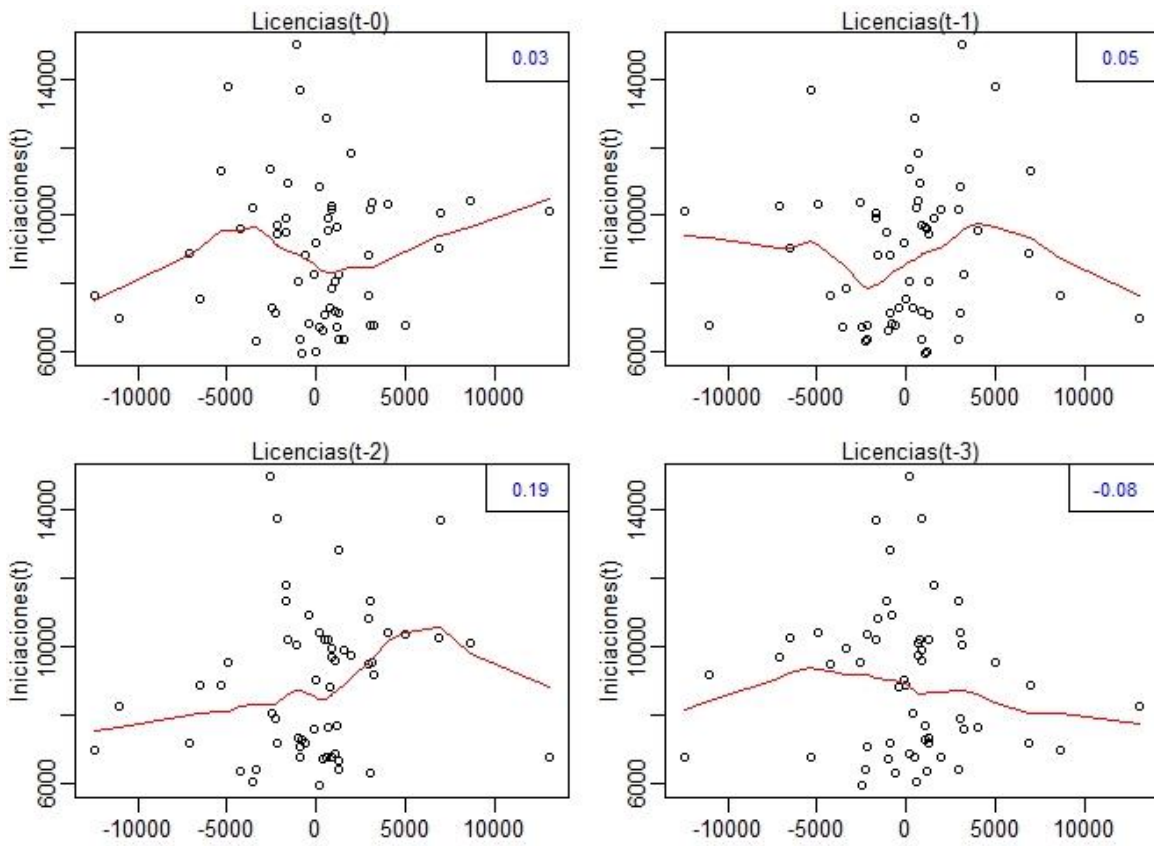
Fuente: DANE. Elaboración SIS - SDHT.

Conclusiones

Con este pequeño ejercicio de testeo se logró consolidar un proceso más o menos estructurado de producción de pronósticos, el cual implica la combinación de varias metodologías de estimación de pronósticos, lo que hace de este proceso una herramienta flexible permitiendo probar y agregar nuevas metodologías que aporten información no capturada por las ya probadas. Se comprobó que al igual que en muchos campos, los niveles de actividad del sector construcción son un proceso lo suficientemente complejo como para requerir la combinación de varias herramientas estadísticas al momento de intentar predecir su comportamiento.

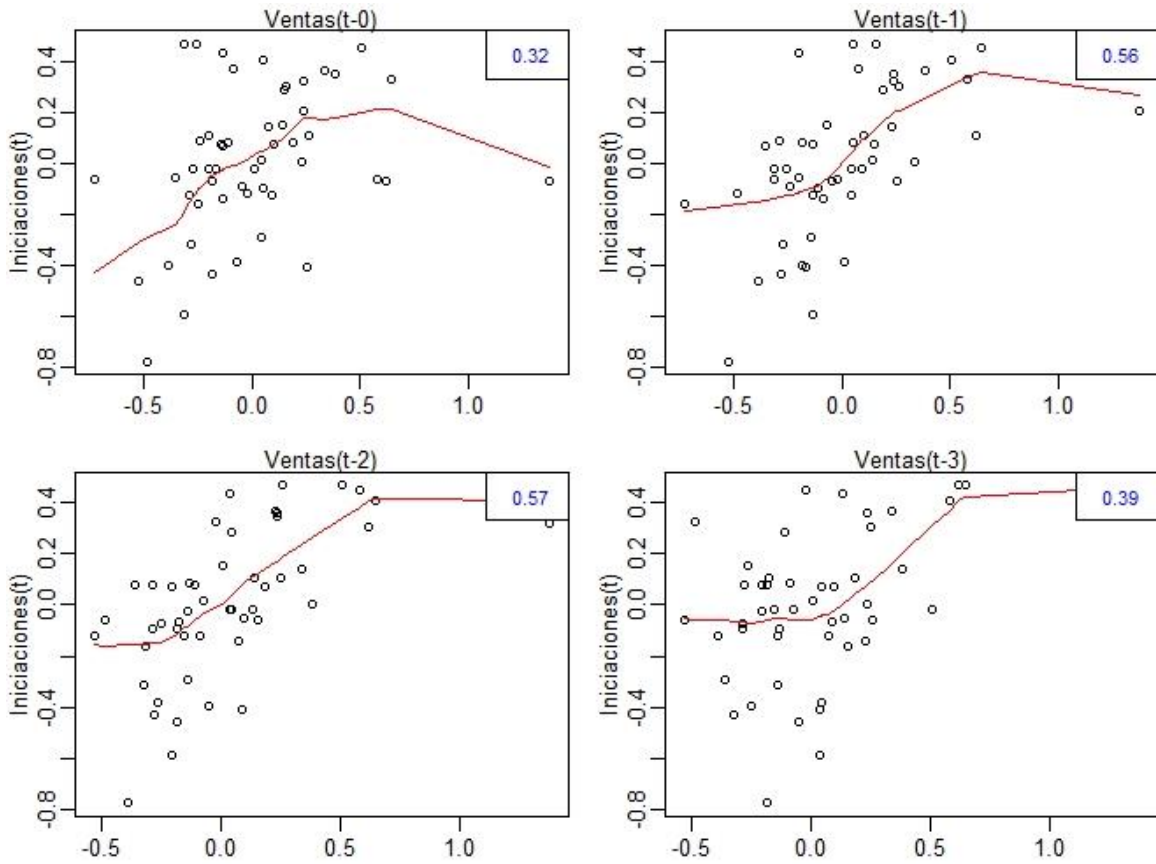
Anexo 1.

Ilustración 15 Correlación cruzadas entre los crecimientos anuales de las licencias e Iniciaciones de vivienda en Bogotá.



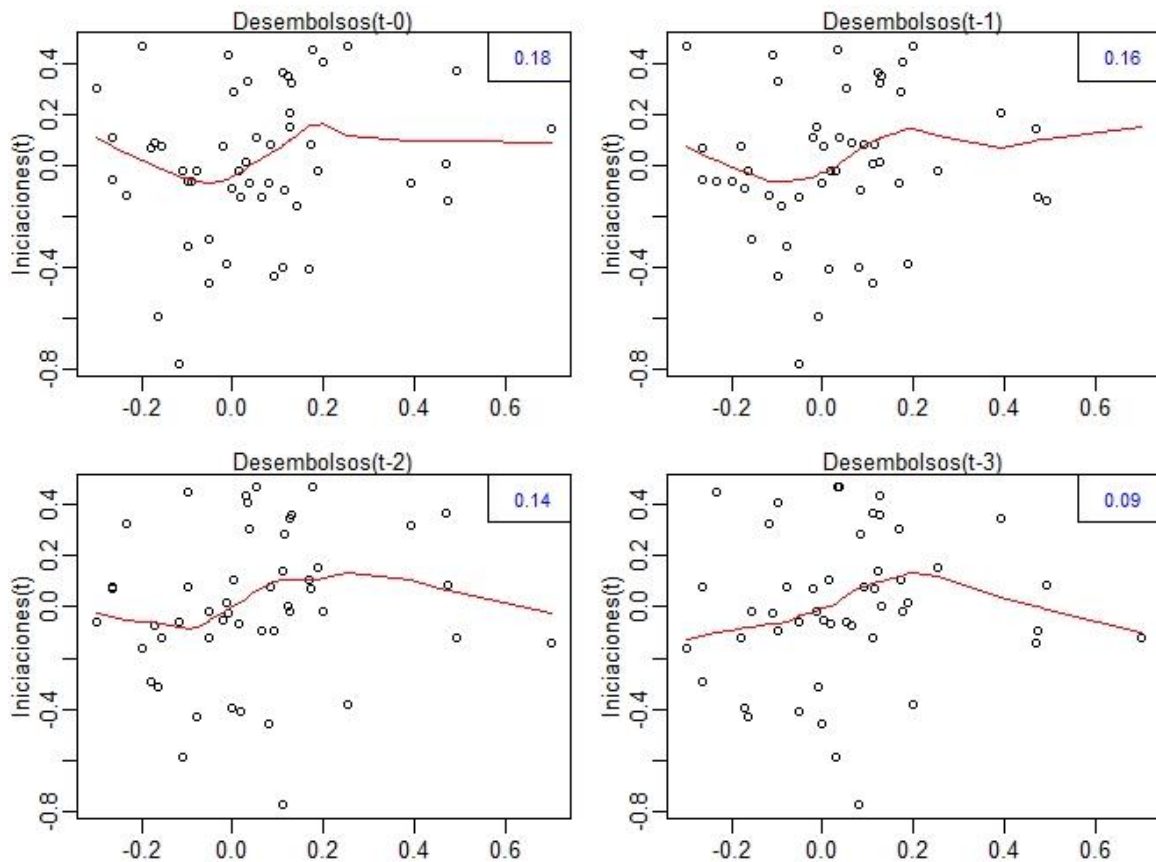
Fuente: DANE. Elaboración SIS - SDHT.

Ilustración 16 Correlación cruzadas entre los crecimientos anuales de las ventas e Iniciaciones de vivienda en Bogotá.



Fuente: DANE y Galería Inmobiliaria. Elaboración SIS - SDHT.

Ilustración 17 Correlación cruzadas entre los crecimientos anuales de los desembolsos e Iniciaciones de vivienda en Bogotá.



Fuente: DANE. Elaboración SIS - SDHT.

Akaike, H. (1974). Markovian representation of stochastic processes and its application. *Annals of the Institute of Statistical Mathematics*, 363–387.

Calle, J. L. (20 de 03 de 2018). *Jalobe*. Obtenido de Jalobe: <https://jalobe.com/doc/tsoutliers.pdf>

Chung Chen, L.-M. L. (1993). Joint Estimation of Model Parameters and Outlier Effects in. *Journal of the American Statistical Association*, 284–297.

Diebold, F., & Mariano, R. (1995). Comparing Predictive Accuracy. *Journal of Business and Economic Statistics*, 253-263.

Giovanni, P. (07 de 06 de 2018). *cran.r-project*. Obtenido de cran.r-project: <https://cran.r-project.org/web/packages/dlm/vignettes/dlm.pdf>

Harrison, P., & Stevens, C. (1976). Bayesian forecasting (with discussion). *J. Royal Statist. Soc*, 205–247.

Kalman, R. (1960). A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 35 - 45.

MVCT, M. d. (20 de 03 de 2018). *Asocapitales*. Obtenido de Asocapitales:
<http://www.asocapitales.co/documentos/145.pdf>

West, M., & Harrison, J. (1997). *Bayesian forecasting and dynamic models*. New York.: Springer.