Fortalecimiento de la Planificación Territorial para el desarrollo sostenible e inteligente del DMQ

Estructuración del modelo de predicción de oferta y demanda del mercado del suelo, vivienda y construcción en el DMQ

Municipio de Quito

Instituto de Investigación de la Ciudad

Elaborado por:
Profesional contratado: Franco Mejía
Modelo de mínimo cuadrado en dos etapas: Christian Pereira

Tabla de contenido

1. Resumen Ejecutivo 4
2. Análisis conceptual 5
2.1. Revisión de metodologías aplicadas 5
2.2. Marco metodológico y teoría 9
2.2.1. Modelos de ecuaciones estructurales 9
2.2.1.1. Modelos de ecuaciones simultáneas 11
2.2.1.1.1. Problema de identificación 11
2.2.1.1.2. Mínimos cuadrados en dos etapas (MC2E) 13
2.2.1.1.3. Mínimos cuadrados en tres etapas (MC3E) 14
2.2.1.2. Modelos de regresión. 15
2.2.1.3. Modelos de trayectoria. 17
3. Integración de datos y fuentes 20
3.1. Exploración de la información 20
3.2. Procesamiento de la información 23
4. Implementación del modelo 26
4.1. Análisis descriptivo 26
4.2. Proceso de selección de variables 36
4.3. Preparación de datos para modelamiento 37
4.4. Calibración del modelo Mínimo cuadrados por tres etapas (3SLS) 38
4.5. Calibración del modelo de Ecuaciones Estructurales (SEM) 41
5. Análisis de performance del modelo y testeo 42
5.1. Diseño del proceso de performance y testeo 42
5.2. Comprobación de supuestos 43
6. Outputs del modelo 44
6.1. Análisis de los resultados del modelo 2SLS y 3SLS 44
6.2. Análisis de los resultados del modelo SEM 46

6.3. Usos de los modelos 47

7. Conclusiones 49

8. Recomendaciones 51

9. Referencias bibliográficas 52

10. Anexos 54

1. Resumen Ejecutivo

El presente documento comprende tres etapas, la primera se compone por una revisión metodológica, la segunda por un análisis descriptivo de las bases de datos entregadas por el Municipio de Quito y procesamiento de la información. Y, por último, la tercera fase tiene como propósito el análisis de las estimaciones realizadas.

La primera parte del proyecto tiene como propósito la definición de la metodología para explicar el comportamiento del precio de la vivienda en el Distrito Metropolitano de Quito (DMQ). Para esto, se realizó una revisión teórica en la cual se sustente la especificación de los modelos a aplicar. A partir del análisis realizado se determinó que las metodologías estadísticas a utilizar son; modelos de ecuaciones estructurales y mínimos cuadrados ordinarios en dos y tres etapas. Dichos métodos permitirán visualizar los diferentes efectos por parte de la oferta y demanda del mercado en el precio de la vivienda. De esta forma, se podrá inferir cuales son los parámetros con mayor relevancia y por ende una visualización de posibles políticas que permitieran generar un mayor impacto en el mercado inmobiliario.

La segunda fase tiene como objetivo analizar la información enviada por parte del Instituto de Investigación de la Ciudad, correspondiente al Catastro del DMQ y la delimitación de observaciones a incorporar en el estudio. El análisis comprende la recopilación de variables de interés, tratamiento de valores nulos, identificación de datos atípicos, integración de bases de datos y primeras inferencias de la información.

Se pretende determinar si la información existente es confiable y pertinente para su uso en el modelo de estimación de factores que afectan al precio de la vivienda. Para ello, este apartado detalla la distribución y tendencia de la información, el estado de las bases de datos y los posibles tratamientos, así como la definición de la población a utilizar en los modelos estadísticos.

La tercera se basa en la implementación de los modelos de mínimos cuadrados de dos y tres etapas y el modelo de ecuaciones estructurales, con la finalidad de observar los efectos de los diferentes factores en el mercado inmobiliario del DMQ. El proceso inicia con la selección de variables, el cual se basó en el cumplimiento de supuestos como una baja relación lineal entre las variables exógenas y el cumplimiento de un nivel mínimo de significancia en la estructura final del modelo.

La preparación de la data para el modelamiento tiene como objetivo definir las bases de desarrollo y de prueba, dicha división permite desarrollar indicadores de performance los cuales determinan la confiabilidad de los resultados.

En este contexto, el objetivo de la presente entrega es la interpretación de las estimaciones y evidenciar qué tan robustos fueron los modelos por medio de indicadores como Índice de Ajuste Comparativo (Comparative Fit Index, CFI), Índice de Ajuste Tucker-Lewis (Tucker-Lewis fit index, TLI), Raíz del Error Cuadrático Medio de Aproximación (Root Mean Square Error of Approximation, RMSA), Raíz Estandarizada del Residuo Cuadrático Medio (Stnadarized Root Mean Square Residual, SRMR) y Logaritmo de verosimilitud (Log-likelihood).

2. Análisis conceptual

2.1. Revisión de metodologías aplicadas

Las investigaciones en el mercado inmobiliario a menudo emplean modelos generales para interpretar el comportamiento de este sector. Estos modelos pueden abordar una serie de variables como el precio de la vivienda, la inversión en construcción residencial o incluso las iniciativas de construcción, tanto en términos cuantitativos como monetarios. Aunque estos modelos persiguen objetivos similares, varían según ciertos criterios, como la profundidad del análisis, las variables consideradas y su estructura dinámica.

Es así como una de las investigaciones de Chesire y Sheppard (1998) estima el precio de la vivienda desde un enfoque de la demanda, acorde a su estructura, elasticidad del ingreso y características de la vivienda. Sus aportes desde el punto de vista del consumidor permiten considerar parámetros como planeamiento del suelo, la plusvalía por barrio, infraestructura pública aspecto que influyen en la demanda por hipotecas. El método utilizado es el uso de modelos hedónicos, los cuales se caracterizan por estimar el precio por medio de variables instrumentales regresadas linealmente. Los resultados de dicha investigación consideran que el aumento de los ingresos genera un incremento de la de demanda de casas con espacios con terreno y en la periferia urbana.

Algunos modelos centran su atención en la oferta y la demanda del mercado, siendo la perspectiva predominante que la oferta se ajustará a la demanda a largo plazo. Gran parte de los estudios también se enfocan en analizar el movimiento en los precios de las viviendas, considerando que en esta variable se verá reflejada cualquier variación en la oferta y demanda. Por ejemplo, Madsen (2012) y Panagiotidis y Printzis (2016) investigan en su estudio el papel del mercado de la vivienda y sus determinantes macroeconómicos, tomando como variable dependiente un índice de precios de la vivienda. Sin embargo, Mikhed y Zemcik (2009), a través de un modelo de valor presente, sugieren que los precios de la vivienda no reflejan consistentemente los fundamentos económicos a lo largo

del tiempo, indicando un posible ajuste prolongado, y que para evaluar ajustes dentro de este mercado con mayor precisión se podría optar por variables adicionales.

Por otro lado, otros estudios asumen que hay un equilibrio en el mercado inmobiliario, sugiriendo que existen mecanismos para contrarrestar cualquier discrepancia en el corto o mediano plazo. Esta idea de equilibrio se refleja en la dinámica de muchos modelos, no solo en los del mercado inmobiliario, sino también en los macroeconométricos que abordan la inversión en vivienda, optando a menudo por una estructura dinámica. Geoff (1999), presta especial importancia a la modelización separada de los lados de la demanda y de la oferta, lo que según el autor permite un entendimiento más profundo de la interacción compleja entre estos dos componentes clave del mercado inmobiliario. Otro artículo con un enfoque similar es el de Berger-Thompson y Ellis (2004) o López (2002), donde se busca una separación rigurosa de la oferta y demanda para comprender las diferencias estructurales en mercados inmobiliarios en diferentes países o regiones.

Siguiendo con los estudios en donde la oferta y demanda son analizados para la definición de los efectos en el precio de la vivienda Huang y Zhang (2012) proponen el análisis del mercado inmobiliario de Singapur mediante la metodología de ecuaciones estructurales. La investigación realiza una crítica a los métodos tradicionales en donde la oferta de vivienda se consideraba como exógena a la demanda debido a que responde a decisiones de los constructores, generando así el supuesto de que el stock de viviendas es fijo. Sin embargo, la investigación demuestra que los shocks en la demanda dependen de las condiciones de la oferta. De esta forma, la demanda estaría representada por variables como ingresos, tasa de interés hipotecaria, índice de precios al consumidor. Por el lado de la oferta, esta se compone de stock de viviendas, precio de las casas, costos de materiales básicos, costos de mano de obra y la tasa de interés hipotecaria.

En lo que respecta a las metodologías y la especificación de los modelos a estudiar también existe una amplia diferencia entre los estudios existentes. Uno de los enfoques tal vez más recurrente es el Modelo de Corrección de Errores de Vectores Cointegrados (VECM). Panagiotidis y Printzis (2016) a través de esta metodología revelan una relación de equilibrio entre el índice de precios de la vivienda y variables clave como el sector minorista y los préstamos hipotecarios. Los hallazgos sugieren que, en el largo plazo, los préstamos hipotecarios y el comercio minorista ejercen la mayor influencia sobre las fluctuaciones del índice de precios de la vivienda. Además, el análisis dinámico señala que los préstamos hipotecarios y el índice de precios al consumidor (IPC) tienen una relación de causalidad con el índice de precios de la vivienda tanto en el corto como en el largo plazo. Estos resultados destacan la importancia crítica de monitorear de cerca el sector bancario y el comercio minorista para comprender y predecir las tendencias de precios de la vivienda en el área de estudio.

Andrews (2010) utiliza un modelo de corrección de errores de vectores para analizar el impacto de los shocks de demanda, los factores estructurales y las políticas fiscales en los precios de la vivienda. Estos enfoques metodológicos ayudan a comprender el complejo y dinámico mercado de la vivienda en su relación con la economía en su conjunto. Por su parte, Adams y Füss (2010) también emplean un análisis de cointegración de panel junto con el VECM, descubriendo que los precios de la vivienda se incrementan en un 0.6% por cada aumento del 1% en la actividad económica, y que la divergencia del equilibrio a largo plazo se ajusta completamente después de aproximadamente 14 años.

Geoff (1999), lleva a cabo un análisis riguroso de la dinámica de mercado de las viviendas en Irlanda. El autor emplea análisis de cointegración y funciones de respuesta a impulsos para investigar las relaciones a largo plazo entre los precios de las viviendas, el stock de viviendas, los ingresos y las tasas de interés hipotecarias. El estudio arroja luz sobre las limitaciones en la oferta de viviendas y subraya la importancia de considerar factores como los costos de construcción y las tasas de interés al analizar la dinámica del mercado.

Otras orientaciones metodológicas que se destacan son: el análisis realizado por Hilbers et al. (2008), que emplea el enfoque del costo para el usuario, señala un buen ajuste del modelo para la mayoría de los países de la UE, capturando así los desarrollos del mercado de la vivienda, Alvarez et al. (2010) utilizan el análisis de correlación cruzada, el filtro de Butterworth y el kernel Epanechnikov para revelar la fuerte relación entre la inversión residencial y el PIB, con la inversión residencial liderando el ciclo económico en periodos de expansión. Iacoviello y Neri (2008) utilizan un enfoque bayesiano de probabilidad para modelar el efecto de la política monetaria en los precios de la vivienda y las inversiones residenciales. Tsatsaronis y Zhu (2004) emplean un marco de autoregresión vectorial estructural (SVAR) para mostrar la dependencia de los precios de la vivienda de factores como la inflación y el crédito. Goodhart y Hofmann (2008) utilizan una regresión de vectores autorregresivos de efectos fijos en paneles para demostrar una relación multidireccional entre los precios de la vivienda, las variables monetarias y la macroeconomía. Davis y Heathcote (2003) emplean un modelo de crecimiento multifactorial de Cobb-Douglas para destacar la volatilidad de la inversión residencial y su relación positiva con el consumo.

Otros estudios como el de Berger-Thompson y Ellis (2004) plantean el estudio del mercado de la vivienda desde un enfoque estructural, utilizando mínimos cuadrados en tres etapas (MC3E). Este estudio proporciona una valiosa visión de los ciclos de inversión en vivienda y su relación con las tasas de interés en un contexto global, incluyendo a los mercados inmobiliarios de Australia, Estados Unidos, Reino Unido y Canadá. Los autores destacan que, si bien gran parte de la ciclicidad en la inversión en vivienda puede atribuirse a la sensibilidad del sector a las tasas de interés, también es

posible que los rezagos de construcción generen una ciclicidad intrínseca en este ámbito. El análisis indica la presencia de posibles restricciones en la oferta de viviendas y subraya la importancia de considerar factores como las limitaciones de terreno y el crecimiento poblacional al examinar la dinámica del mercado. Además, el estudio resalta el papel de los detalles institucionales en el mercado financiero de la vivienda al explicar las diferencias en el comportamiento de la construcción y los precios de la vivienda entre países y a lo largo del tiempo.

Cabe destacar que, aunque existen diferencias en los modelos empleados para la estimación, los estudios revisados muestran consenso en cuanto a las variables seleccionadas para comprender las dinámicas de oferta y demanda de vivienda. López (2002) realiza una recopilación de varios estudios realizados a regiones españolas con la finalidad de estudiar al mercado inmobiliario. De la revisión efectuada, este autor considera que la inversión en viviendas, que abarca el gasto en viviendas y el número de viviendas iniciadas, refleja la interacción de factores tanto de oferta como de demanda. Las variables demográficas, incluyendo el crecimiento de la población y la migración, se reconocen como determinantes clave de la demanda a largo plazo. Además, la renta de las familias se considera un factor principal en los modelos que enfatizan los aspectos de la demanda, mientras que la riqueza de las familias también se identifica como un factor relevante en algunos estudios. Los precios, tanto como un indicador de oferta como de demanda, se relacionan directamente con la inversión en viviendas. Las condiciones de financiación y el coste del crédito también emergen como factores cruciales, con la disponibilidad de préstamos hipotecarios y el racionamiento crediticio influyendo en la actividad de construcción residencial. En la investigación empírica, se observa que los modelos explicativos de precios de viviendas generalmente se basan en las ecuaciones de oferta y demanda, con distintas aproximaciones en la modelización del proceso de ajuste de los precios a su equilibrio. Estos hallazgos señalan la compleja interrelación de diversos factores que afectan la dinámica del mercado de la vivienda en las regiones españolas.

Entre los factores que afectan a la oferta de viviendas, estudios previos como los de Blackley (1999) y Malpezzi y Vandell (2002) incluyen el valor de la construcción residencial privada en términos reales, los costos de los materiales de construcción, los salarios de los trabajadores en el sector, la disponibilidad de terrenos para edificación, las tasas de interés, las expectativas de inflación y el inventario de capital residencial. Además, se ha demostrado que las regulaciones que inciden en el mercado inmobiliario y las tasas de vacantes también son elementos cruciales en el comportamiento de la oferta en este sector. En consonancia con esto, los estudios de Murray (1999) y Malpezzi y Vandel (2002) subrayan el papel de las políticas de vivienda en la oferta de viviendas, que pueden

tener un impacto significativo en la configuración de las áreas donde se desarrollan proyectos de viviendas asequibles.

El inventario de viviendas existente en el mercado, junto con las tasas de vacantes, también se ha establecido como un indicador significativo de la oferta, como lo sugiere Malpezzi & Vandell (2002). Por último, la oferta de viviendas también puede depender de factores específicos del mercado local, como la población, la actividad de mercado y las características particulares de ese mercado, según el enfoque presentado por Goodman (2005).

A partir de las investigaciones realizadas se determinó que para el análisis del mercado inmobiliario se requiere comprender la oferta y demanda de vivienda, con el objetivo de determinar los efectos que tiene cada componente de dichas estructuras en el precio del bien inmobiliario, por lo tanto, se requiere un marco metodológico en donde ambos procesos interactúen y determinen la variable objetivo la cual es el precio de la vivienda en el Distrito Metropolitano de Quito. Para cumplir con dicho propósito los modelos de mínimos cuadrados en dos – tres etapas y ecuaciones estructurales simultaneas son analizados conceptualmente en la siguiente sección.

2.2. Marco metodológico y teoría

Para la estimación de las variables relacionadas al mercado de la vivienda en el Distrito Metropolitano de Quito la metodología a aplicar se basa en dos procesos estadísticos: mínimo cuadrados en dos y tres etapas (2SLS, 3SLS) y ecuaciones estructurales. Los métodos planteados fueron elegidos en base a casos de aplicación en donde la estimación del precio de un bien inmueble está determinada por factores que afectan a la oferta y demanda de la vivienda con la presencia de variables endógenas.

2.2.1. Modelos de ecuaciones estructurales

Los modelos de ecuaciones estructurales (SEM) pueden ser considerados como una generalización o integración de modelos como mínimos cuadrados ordinarios (MCO), análisis de varianza (ANOVA), análisis de regresiones múltiples y análisis de factores. Se puede decir entonces, que los modelos SEM son un conjunto de técnicas estadísticas utilizadas para analizar relaciones causales entre variables observables y no observables. Su uso ha sido enfocado para examinar relaciones complejas entre múltiples variables simultáneamente y para evaluar la validez de las hipótesis teóricas en diversas disciplinas, como sociología, psicología y economía entre otras ciencias sociales.

Los modelos SEM se pueden agrupar en tres distintas familias de técnicas de inferencia, los tres grupos de técnicas se detallan a continuación:

- Análisis de estructuras de covarianza: El objetivo de este tipo de modelos, es la estimación de parámetros utilizando variables observables o latentes, por medio de la minimización en la diferencia entre la matriz de covarianzas de la muestra y la matriz estimada por el modelo, en donde se efectúan hipótesis para determinar cómo y porque existe relación entre las variables, MANOVA
- En el segundo grupo se encuentra los modelos de tipo basado en varianza y modelos de trayectoria de mínimos cuadrados parciales (PLS-PM), este tipo de técnicas se refieren a algoritmos basados en técnicas de regresión y analizan combinaciones ponderadas de variables observables para aproximarse a la hipótesis planteada. Este tipo de técnicas tratan de explicar la matriz de varianzas y covarianzas de la muestra. En este tipo de modelo SEM se encuentra modelos lineales generalizados (GLM) y regresiones múltiples.
- El tercer grupo de modelos SEM son los modelos estructurales causales (SCM) o no paramétricos, en este tipo de técnicas se establecen hipótesis de tipo causal y son representadas por un gráfico dirigido acíclico (DAG) cuando se asume una relación de causalidad unidireccional o un gráfico dirigido cíclico (DCG) cuando ciertas variables tienen efecto entre sí.

Los modelos tradicionales SEM o modelos de análisis estructural de covarianza se pueden dividir en dos metodologías, la primera relacionada con el método analítico de trayectoria o path analytic elaborado por (Huang & Zhang, 2012) y Wright (1934), el cual se basa en técnicas de regresión para estimar efectos directos e indirectos. La segunda metodología se basa en análisis de factores, el objetivo es estimar las variables latentes por medio de las variables observadas.

Las variables latentes son un conjunto de variables que no pueden ser directamente cuantificadas, son indirectamente medidas, y por lo tanto estimadas por medio de un conjunto de variables que son observables y medidas utilizando tests, encuestas, escalas, entre otros. Como ejemplo la confianza del consumidor es una variable latente. Las variables observables son un conjunto de variables que son utilizadas para la estimación de las variables latentes.

La presente investigación del mercado inmobiliario de Quito una vez hecho el análisis de las diferentes familias de modelos de ecuaciones estructurales, establece que el método más optimo y que cumple con el objetivo de identificar los efectos de diversas variables observables en el precio de la vivienda son los métodos SEM tradicionales, específicamente los modelos de trayectoria o *path analytic*, en el

siguiente apartado se detalla en profundidad dicha metodología y sus ventajas frente a otros métodos tradicionales de las técnicas SEM.

2.2.1.1. Modelos de ecuaciones simultáneas

De acuerdo con la revisión bibliográfica efectuada, en ámbitos como el mercado inmobiliario, es posible que se requiera utilizar modelos que consideren múltiples relaciones simultáneas. Brooks y Tsolacos (2010) señalan que, aunque ecuaciones individuales pueden ajustarse de manera eficaz a series de datos históricos específicos del sector inmobiliario, es posible integrarlas en modelos de sistemas de ecuaciones simultáneas, especialmente cuando la fundamentación teórica sugiere que las conexiones causales podrían tener una naturaleza bidireccional o multidireccional. Sin embargo, es crucial subrayar que el desempeño de dichos modelos podría verse comprometido debido a la complejidad inherente en la dinámica de un sistema compuesto por múltiples ecuaciones.

No obstante, como señala Ho y Cuervo (1999), un modelo de demanda y oferta estructural puede ser más apropiado si el objetivo es establecer relaciones causales para el análisis estructural; determinar elasticidades y multiplicadores para el análisis de políticas; y hacer previsiones para fines de planificación.

Por ello, considerando la finalidad del presente estudio, se analiza la demanda y oferta del mercado de viviendas utilizando un modelo de ecuaciones simultáneas con un análisis de la trayectoria de series temporales del mercado de viviendas.

Podemos escribir el sistema de ecuaciones simultaneas descrito como $Y = X'\beta + u$, donde Y es un vector de variables endógenas, X' es una matriz de variables explicativas que podría contener algunas variables endógenas y u es el término de perturbación estocástica. No obstante, si los regresores de una o más ecuaciones están correlacionados con u, es decir, $E(X, u) \neq 0$, el estimador de mínimos cuadrados ordinarios proporcionará resultados sesgados e inconsistentes. Este sesgo, llamado sesgo de endogeneidad puede corregirse aplicando el enfoque de variable instrumental del método generalizado de momentos (GMM), mínimos cuadrados en dos etapas (MC2E), mínimos cuadrados en tres etapas (MC3E), máxima verosimilitud de información completa (FIML), máxima verosimilitud de información limitada (LIML), entre otros.

2.2.1.1.1. Problema de identificación

Cualquier modelo estructural donde por cada ecuación exista una variable endógena puede ser reformulado en una forma reducida. Esta transformación permite establecer un vínculo directo entre

las variables endógenas y los determinantes exógenos, simplificando el análisis de las relaciones entre las mismas.

Por ejemplo, tomemos como punto de partida un sistema compuesto por dos ecuaciones:

$$y_1 = \beta_0 + \beta_1 y_2 + \beta_1 X_1 + u_1 \quad (1)$$

$$y_2 = \alpha_0 + \alpha_1 y_1 + \alpha_2 X_2 + u_2 \quad (2)$$

donde y_1 y y_2 son variables mutuamente dependientes o endógenas, X_1 y X_2 son variables exógenas y u_1 y u_2, los términos de perturbación estocásticas.

La forma reducida[1] para este sistema se expresa de la siguiente manera:

$$y_1 = \Pi_0 + \Pi_1 X_1 + \Pi_2 X_2 + \nu_1 \quad (3)$$

$$y_2 = \Pi_3 + \Pi_4 X_1 + \Pi_5 X_2 + \nu_2 \quad (4)$$

donde Π_i son los coeficientes en la forma reducida y ν_i son los nuevos términos de error expresados como combinaciones lineales de los términos de error originales u_i.

Al observar las ecuaciones en su forma reducida, es evidente la eliminación de las variables endógenas, dejándolas únicamente como variables dependientes. Los términos de error introducidos, denotados como ν_i, reflejan las desviaciones respecto al valor de equilibrio del mercado. Es esencial subrayar que estos errores no necesariamente coinciden con los del modelo estructural, pero capturan la misma esencia: las perturbaciones no observadas que afectan a las variables endógenas.

Cabe recalcar que la utilidad de manejar una ecuación en su forma reducida depende de que el modelo se encuentre identificado. La identificación se refiere a la posibilidad de estimar de manera única y coherente los parámetros estructurales del modelo, es decir, se refiere a la capacidad de estimar los parámetros de su forma estructural (ecuación original) en función de los parámetros de su forma reducida de manera única.

La condición de orden y la condición de rango son dos criterios utilizados para determinar si una ecuación en un sistema de ecuaciones simultáneas está identificada.

La condición de orden es una condición necesaria pero no suficiente para la identificación de una ecuación y se basa en el número de variables excluidas de una ecuación en relación con el número

[1] La forma reducida se obtiene a partir de un concepto de equilibrio de mercado. Despejar las variables endógenas y su posterior reemplazo en las ecuaciones originales permite obtener la forma reducida para cada variable endógena o ecuación.

total de ecuaciones en el sistema, es decir, una ecuación en un sistema se considera "justamente identificada" si el número de variables excluidas de esa ecuación es igual al número total de ecuaciones menos uno. Si hay más variables excluidas que este número, la ecuación se considera sobreidentificada. Si hay menos, se considera no identificada o subidentificada.

La condición de rango es una condición necesaria y suficiente para la identificación, se refiere al rango de una matriz de coeficientes formada por las variables excluidas de una ecuación en particular. Si esta matriz tiene rango completo, la ecuación se considera identificada.

Por tanto, dependiendo de cómo se encuentre identificado nuestro sistema de ecuaciones se aplicarán distintos métodos para su resolución. En el caso de que una ecuación se encuentre exactamente identificada, un método posible es el de mínimos cuadrados indirectos, mientras que, si la misma ecuación se encuentra sobreidentificada, el método a usar será el de mínimos cuadrados en dos etapas.

2.2.1.1.2. Mínimos cuadrados en dos etapas (MC2E)

La estimación de mínimos cuadrados en dos etapas, como su nombre lo indica, implica la realización de dos pasos. Dentro de la primera etapa se plantea la forma reducida del modelo, es decir, existe una ecuación para cada variable endógena y esta se encuentra en función de variables exógenas o predeterminadas. Se toman como ejemplo las ecuaciones (1) y (2) planteadas anteriormente, mismas que se encuentran sobreindentificadas.

Una vez obtenida la forma reducida, se estima cada una de las ecuaciones por mínimos cuadrados ordinarios:

$$\hat{\Pi} = \begin{pmatrix} \hat{\Pi}_0 \\ \hat{\Pi}_1 \\ \hat{\Pi}_2 \end{pmatrix} = (X'X)^{-1}(X'Y_1) \tag{5}$$

Donde $\hat{\Pi}$ corresponderá a los coeficientes estimados en su forma reducida, la matriz X incluirá las variables predeterminadas, X_1, X_2 y Y_1 corresponderá a un vector con los diferentes valores de y_1 para cada observación.

Finalmente, podemos encontrar $\hat{y}_1$ (el valor esperado de y_1), el cual sería equivalente a las variables predeterminadas o exógenas utilizadas en su cálculo:

$$E[y_1] = \hat{y}_1 = \hat{\Pi}_0 + \hat{\Pi}_1 X_1 + \hat{\Pi}_2 X_2 + E[v_1]$$

$$\hat{y}_1 = \hat{\Pi}_0 + \hat{\Pi}_1 X_1 + \hat{\Pi}_2 X_2 \tag{6}$$

La etapa 2 del modelo de MC2E comprende el estimar cada una de las ecuaciones estructurales utilizando mínimos cuadrados ordinarios, pero reemplazando cualquier variable endógena con los valores ajustados encontrados en el paso anterior.

Es decir, se reemplazan las variables estimadas en la ecuación original estructural, de manera que:

$$y_2 = \alpha_0 + \alpha_1 y_1 + \alpha_2 X_2 + \mu_2$$

$$y_2 = \alpha_0 + \alpha_1(\hat{y}_1 + v_1) + \alpha_2 X_2 + \mu_2$$

$$y_2 = \alpha_0 + \alpha_1 \hat{y}_1 + \alpha_2 X_2 + \mu_2^* \quad (7)$$

donde $\mu_2^* = \mu_2 + \alpha_1 v_1$ y finalmente se procede a estimar por Mínimos Cuadrados Ordinarios, lo cual dará estimaciones consistentes de la variable y_2.

2.2.1.1.3. Mínimos cuadrados en tres etapas (MC3E)

Los mínimos cuadrados en tres etapas proporcionan un tercer paso en el proceso de estimación del MC2E que permite covarianzas distintas de cero entre los términos de error en las ecuaciones estructurales. Es asintóticamente más eficiente que el método de MC2E, ya que este último ignora cualquier información que pueda estar disponible sobre las covarianzas de error (y también cualquier información adicional que pueda contenerse en las variables endógenas de otras ecuaciones).

El proceso de estimación en este modelo es similar al aplicado dentro de MC2E, es decir, en la primera etapa, cada variable endógena se regresa en todas las variables predeterminadas (exógenas así como endógenas rezagadas) para obtener valores predichos para las variables endógenas y en la segunda etapa se realiza una estimación de la matriz de correlación cruzada de la ecuación residual. Como tercer paso, a diferencia del MC2E, se puede usar estimaciones de mínimos cuadrados generalizados (GLS) de los parámetros usando la matriz de covarianza (paso 2) y los valores predichos (paso 1) en lugar de las variables endógenas especificadas como explicativas en la ecuación.

Entonces, considerando que este método implica la utilización de matriz de covarianzas consideraremos una representación matricial de un sistema de ecuaciones:

$$y_{1t} = \hat{Y}_1 \beta_1 + X_1 \gamma_1 + u_{1t}^* \quad (8)$$

$$y_{2t} = \hat{Y}_2 \beta_2 + X_2 \gamma_2 + u_{2t}^* \quad (9)$$

El sistema de ecuaciones que se planteó para el MC2E puede ser replanteado, tomando en cuenta que contábamos únicamente con dos variables endógenas, y_1 y y_2; es decir:

$$y_j = \hat{Y}_j\beta_j + X_j\gamma_j + u_j^*$$

$$y_j = \hat{Z}_j\delta_j + u_j^* \quad (10)$$

O desde el punto de vista matricial, la ecuación representaría a $y \epsilon \mathbb{R}^{2Tx1}$; $Z \epsilon \mathbb{R}^{2Tx2}$; $\delta \epsilon \mathbb{R}^{2x1}$ $u \epsilon \mathbb{R}^{2Tx1}$.

De esta forma, $\hat{Z}_j = [Y_j X_j]$, $\delta_j = [\beta_j \gamma_j]'$, $u_j \sim (0, \sigma_j^2 I_T)$, $E(u_i, u_j') = \sigma_{ij} I_T$

Con la expresión anterior, podemos determinar el valor de $\hat{\delta}$ como sigue:

$$\hat{\delta} = (Z^*\Sigma^{-1} \otimes (X'X)^{-1})Z^*)^{-1}(Z^*\Sigma^{-1} \otimes (X'X)^{-1})y) \quad (11)$$

Si consideramos que la matriz H es igual a $X(X'X)^{-1}X'$, podemos reestructurar la ecuación anterior y representarla de forma que:

$$\hat{\delta} = (\widehat{W}Z)^{-1}(\widehat{W}y) \quad (12)$$

Donde $\widehat{W}$ será una matriz con todas las variables exógenas del modelo.

Por tanto, del proceso de estimación de MC3E podemos mencionar que es posible encontrar el estimador a partir de calcular los residuos del parámetro de tres etapas, es decir, hallando $\hat{\Sigma}$, y calculando nuevamente los parámetros. Dicho procedimiento se deberá realizar hasta cuando se logre que $\hat{\delta}$ logre converger a su valor de equilibrio.

2.2.1.2. Modelos de regresión.

Los modelos de regresión múltiple predicen variables dependientes utilizando múltiples variables independientes u observables. En medidas estándar, una regresión lineal simple para predecir la variable Y por medio de una variable independiente X podría representarse de la siguiente manera.

$$\widehat{z_y} = \beta z_x \quad (13)$$

Donde β es el coeficiente estandarizado de la regresión. La relación entre el coeficiente de regresión no estandarizado b y el coeficiente estandarizado β, se establece bajo la siguiente ecuación.

$$\beta = \frac{\sum z_x z_y}{\sum z_x^2} = b\frac{s_x}{s_y} = r_{xy} \tag{14}$$

En donde s_x y s_y son la desviación standard de las variables X y Y respectivamente. Para dos variables independientes la ecuación se representaría de la siguiente manera.

$$\widehat{z_y} = \beta_1 z_1 + \beta_2 z_2 \tag{15}$$

Y la versión parcial estandarizada para los coeficientes β_1 y β_2 son obtenidos por medio de:

$$\beta_1 = \frac{r_{y1} - r_{y2}r_{12}}{1 - r_{12}^2} \; y \; \beta_2 = \frac{r_{y2} - r_{y1}r_{12}}{1 - r_{12}^2} \tag{16}$$

La correlación entre la variable dependiente Y y la estimada $\hat{Y}$ tiene el nombre de correlación múltiple de coeficiente, y es escrito de la siguiente manera:

$$R_{y\hat{y}} = R_{y12} \tag{17}$$

La ecuación indica que la variable dependiente está siendo estimada por dos variables independientes, por lo tanto, la correlación múltiple al cuadrado de los coeficientes es calculado como:

$$R^2{}_{y\hat{y}} = R^2{}_{y12} = \beta_1 r_{y1} + \beta_2 r_{y2} \tag{18}$$

La correlación cuadrada múltiple de los coeficientes indica la cantidad de varianza que es explicada por las variables independientes. R^2 también es interpretado como una medida de bondad de ajuste en el análisis de una regresión múltiple. Con el propósito de distinguir entre explicación y predicción de las variables en el modelo de regresión, dos medidas se pueden utilizar. El primero es el análisis de correlación múltiple (MCA), el cual el parámetro de interés es la correlación que existe entre las variables independientes y la variable dependiente, adicional teniendo cuenta la dimensión de la muestra n y el número de predictores p, se obtiene.

$$R_{Adj}^2 = R^2 - \frac{p}{n - p - 1}(1 - R^2) \tag{19}$$

La segunda medida es el análisis de regresión múltiple (MRA), se utiliza el mismo índice R^2 pero calculado de la siguiente manera.

$$R_{Adj^*}^2 = R^2 - \frac{2p}{n - p}(1 - R^2) \tag{20}$$

Para la estimación de la significancia de R^2 se puede realizar un test de hipótesis por medio del estadístico F, por lo tanto, la significancia del valor R^2 estaría determinado por:

$$F = \frac{R^2/p}{(1-R^2)/n-p-1} \tag{21}$$

Una vez observada la composición de una regresión se puede decir que la ecuación es un modelo teórico especificado por el criterio del desarrollador. La especificación del modelo involucra desarrollar una matriz de varianzas y covarianzas de una muestra que explique el propósito de la investigación, por ende, las métricas R^2 y el estadístico F permiten corroborar la bondad de ajuste del modelo de regresión. Dichos procedimientos son primordiales para dar validez a los resultados.

Una vez el modelo esta correctamente especificado se sigue un proceso de identificación el cual consiste por medio de la matriz de varianza y covarianza obtener los parámetros de la ecuación. Dicho procedimiento caracteriza a los modelos de regresión línea como modelos de completa identificación debido a que todos los parámetros de la regresión son estimados. El procedimiento en el cual se obtienen los pesos de la regresión se denomina estimación del modelo.

Existen diferentes métodos de estimación el más común es mínimo cuadrados ordinarios (MCO) el cual selección los pesos por medio de la minimización de los errores al cuadrado. Entre los métodos están el de máxima verosimilitud (ML), mínimos cuadrados en dos etapas (2SLS), mínimos cuadrados ponderados diagonal (DWLS) y métodos más robustos en Mplus (MLM, MLMV, WLSM, WLSMV). El uso de los diferentes métodos dependerá del análisis de los datos.

2.2.1.3. Modelos de trayectoria.

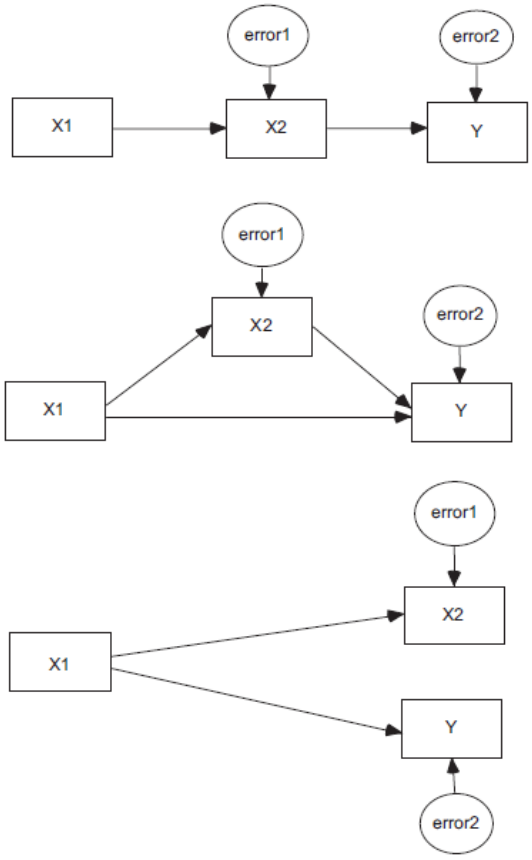
Los modelos de trayectoria son una extensión de los modelos de regresión al incluir varias ecuaciones. Los modelos de trayectoria especifican los efectos directos, indirectos y su correlación entre las variables observables. Es decir, los modelos de trayectoria simultáneamente resuelven varias regresiones múltiples especificadas previamente. Adicional este tipo de modelos permite la comparación entre la matriz original de correlaciones y la matriz implícita que es producida en base la especificación del método de trayectoria.

Los modelos de trayectoria se adhieren a un convenio de gráficos comunes que son utilizados en los modelos SEM. Los rectángulos representan las variables observables y los errores de la predicción son identificados por óvalos. Las líneas entre variables observadas denotan efectos directos. Una

flecha curva de dos variables independientes indica covarianza, incluso sugiere que dicha relación puede estar influenciada por factores externos a la definición del modelo de trayectoria. Cada variable dependiente requiere un término de error, el cual indica efectos no explicados por cada ecuación, para mayor información sobre el diagrama de los modelos de trayectoria dirigirse al anexo 2.

En los modelos de trayectoria la especificación requiere examinar la relación entre múltiples variables, se puede definir diferentes relaciones entre un conjunto de variables y generar hipótesis con diferentes parámetros. A medida de ejemplo en un modelo con tres variables, pueden establecerse varias trayectorias en base a diferentes hipótesis.

Ilustración 1: Modelo de trayectoria con tres variables



Fuente: (Randall E. Schumacker, 2016)

La Ilustración 1 muestra las posibles relaciones en un modelo de tres variables. El primer modelo $X1$ influencia $X2$ el cual influencia Y. En el segundo modelo, presenta una trayectoria adicional entre $X1$ y Y, en donde $X1$ tiene tanto una influencia directa e indirecta en Y. El efecto directo es que $X1$ tiene una influencia directa en Y, mientras que el efecto indirecto es que $X1$ influencia a Y mediante $X2$.

En el caso del tercer modelo, $X1$ influencia a los dos $X2$ y Y, sin embargo $X2$ no tiene ninguna relación con Y.

Para determinar el tipo de interacción, es decir, la especificación del modelo debe estar sustentado en bases teóricas e investigaciones previas que justifiquen la relación entre las variables. Una de las limitaciones del modelo de trayectoria es que no existe una metodología para especificar el modelo, el propósito es determinar los efectos de cada variable una vez que el modelo ha sido especificado bajo consideraciones teóricas.

Los coeficientes en los modelos de trayectoria son derivados por medio de los valores de la correlación de Pearson momento producto y de la estandarización parcial de los coeficientes de regresión, como ejemplo el segundo modelo de la Ilustración 1 el coeficiente de cada trayectoria se puede definir mediante ρ, para cada una de las flechas de $X1$ a Y y $X2$.

$$\beta_1 = \rho_{Y1} \qquad (22)$$
$$\beta_2 = \rho_{Y2}$$

Y la trayectoria entre $X1$ y $X2$ se denota como:

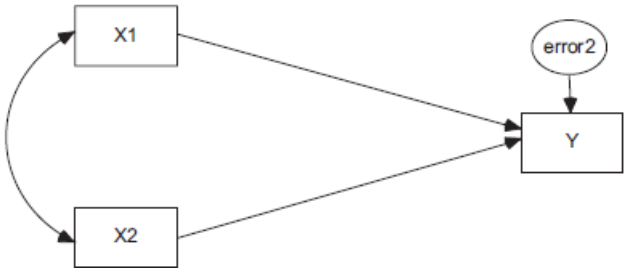
$$r_{X1,X2} = \rho_{12} \qquad (23)$$

A diferencia de los modelos de regresión múltiple donde la variable dependiente es regresada en un solo análisis con todas las variables independientes, en el análisis de trayectoria más de una regresión múltiple es analizada, dependiendo en relación de las diferentes variables. Por lo tanto, los coeficientes de trayectoria con estimados mediante un set de variables independientes.

En los modelos de trayectoria se puede realizar una descomposición de la matriz de correlación, implica que la matriz de correlación original puede ser completamente reproducida si todos los parámetros del modelo son especificados. Como ejemplo consideremos el siguiente modelo.

Ilustración *2* muestra que $X1$ y $X2$ predicen Y, existen dos efectos directos, desde $X1$ a Y y desde $X2$ a Y. Adicionalmente, existen efectos indirectos debido a la correlación entre $X1$ y $X2$. $X1$ indirectamente influencia Y mediante $X2$, de la misma forma $X2$ tiene un efecto indirecto a Y por medio de $X1$. La correlación entre estos tres parámetros se puede descomponer en las siguientes ecuaciones.

Ilustración 2: Modelo de tres variables



Fuente: (Randall E. Schumacker, 2016)

$$r_{12} = \rho_{12} \quad (24)$$

$$r_{y1} = \rho_{y1} + \rho_{12}\rho_{y2} \quad (25)$$

$$r_{y2} = \rho_{y2} + \rho_{12}1 \quad (26)$$

Donde r representa la correlación de los valores observados y ρ los valores de los coeficientes de la trayectoria. La ecuación (24) es simplemente la correlación entre $X1$ y $X2$ (CO). En la ecuación (26), la correlación entre $X1$ y Y es una función entre un efecto directo de $X1$ en Y y efecto indirecto entre $X1$ y Y mediante $X2$. En la ecuación es el mismo comportamiento que en la ecuación (26) excepto que esta ves es la variable $X2$.

La metodología SEM busca comparar la matriz de correlación de la muestra con la matriz de correlación implícita estimada por modelo de trayectoria. El método de comparación es por medio del test $\mathcal{X}^2 = S - \sum$, dado que al momento de la especificación del modelo no todas las posibles relaciones serán especificadas, se busca que la matriz estimada se asemeje a la matriz de varianzas y covarianzas de la muestra.

3. Integración de datos y fuentes

3.1. Exploración de la información

La información enviada por parte del Instituto de Investigación de la Ciudad corresponde al Catastro del Distrito Metropolitano de Quito (DMQ). Esta base de datos abarca seis tablas, que incluyen:

- Características de la construcción (CONSTRUCCION)
- Construcciones en propiedad horizontal (PH y BLOQUE PH)
- Información del predio (PREDIO)
- Información del lote (LOTE)

- Variables relacionadas con el propietario (PROPIETARIO)

La Tabla 1 muestra la estructura de información de la base de datos proporcionada.

Tabla 1 Estructuras de las bases de datos

Nombre tabla	No Observaciones	Columnas
Construcción	2,329,570	24
Predio	1,016,352	24
Lote	464,421	12
PH	26,204	18
Propietario	1,155,280	14

Fuente: Catastro DMQ

La primera tabla de interés serían *CONSTRUCCIÓN*, la cual tiene 2.32 millones de observaciones y 24 variables conforme el siguiente detalle:

Tabla 2 Descripción Base Construcción.

Variable	mean	std	min	25%	50%	75%	max
Numero_pisos	1.256	0.68981292	1	1	1	1	323
Alicuota_ph	1.114	3.88125054	0	0	0.0376	0.457	200
Area_piso_metros	53.942	307.304222	0	11.27	26.99	64.86	172558.66
Area_total	72.852	367.701758	0	11.52	30.9	79.05	172558.66
Area_declarada	24.738	853.398495	0	0	4.73	23	1239759.3
Valoracion	19,991.670	134212.736	0	1209.4825	5856.62	21473.53	69981564.7

Fuente: Catastro DMQ

De la descripción de las variables continuas de la base de datos de construcción se puede observar que el 75% de las construcciones en el Distrito Metropolitano de Quito posee 1 piso, adicionalmente, es posible la existencia de datos erróneos dado que se encontró una observación con 323 pisos. De la misma manera, la variable 'área_declarada', 'área_piso_metros' y 'área_total' muestran una alta dispersión en la información, esto se puede observar por medio de la desviación estándar la cual es significativamente elevada (por encima de la media), en los tres casos.

Tabla 3 Descripción Base Lote.

Variable	mean	Std	Min	25%	50%	75%	max
Area_terreno_escritura	8,552.05	239,832.50	-	200.00	307.06	864.57	70,000,000.00
Frente_total	34.47	91.56	-	10.00	15.48	31.38	10,078.72

Fuente: Catastro DMQ

La Tabla 3 muestra la descripción estadística de las variables continuas de la base de datos Lote. Se evidencia que pueden existir valores atípicos por la dispersión de la información de la información (ver *std*) y los valores máximos. Adicionalmente, se observa que el 75% de los datos presenta un valor

menor o igual 864.57 metros cuadrados, para el caso de la variable 'area_terreno_escritura'. Para la variable frente total, el 75% de las observaciones presentan un valor menor o igual a 31.38 metros cuadrados.

De la base de propiedad horizontal se puede observar en la Tabla 4 que el promedio de avalúo total es de USD 1.003 millones, el 75% de dicha población posee un área menor o igual 1,745 metros cuadrados, así también, se observa que el 50% de las PH no posee terreno comunal.

Tabla 4 Descripción Base Propiedad Horizontal (PH).

Variable	mean	std	min	25%	50%	75%	max
Area_terreno_comunal	303.1	2552.14028	0	0	0	87.6	309488.9
Area_bloque	2061.1	4528.27169	0	507.2	849.5	1745.4	169400.4
Alicuota_total	124.7	768.879095	0.64	100	100	100	100000
Cat_ph_avaluo_terreno	302102.9	1365312.1	0	47818.6	111000	264008.3	183774318.8
Cat_ph_avaluo_construccion	0	0	0	0	0	0	0
Cat_ph_avaluo_total	1003144.4	2563041.3	0	184788.8	343761.7	817802.2	153206725.3

Fuente: Catastro DMQ

Finalmente, la tabla *Predio* posee 1.01 millón de observaciones y 24 variables. Esta tabla muestra también la necesidad de un tratamiento de datos atípicos, esto debido a los significativos valores máximos observados (*max*) y el alto valor en la desviación standard (*std*). En cuanto a las variables de interés, la valoración total en promedio se encuentra en USD 91 mil no obstante el 75% de los predios posee un valor menor a 90 mil dólares lo que indica un sesgo en la distribución. El área de terreno en promedio es de 3,906 metros cuadrados y cerca de la mitad de la población presenta menos o igual a 115 metros cuadrados.

Tabla 5 Descripción Base Predio.

Variable	mean	std	min	25%	50%	75%	max
Cat_aiva_valor	267.25	337.97	0	94	145	282	2210
Frente	15.73	63.093	0	0.35	3.34	12.88	10078.72
Area_terreno	3906.62	162170.29	0	12.54	115	283.69	70000000
Area_construccion	155.73	886.88	0	11.52	73.13	175.64	257572
Valoracion_construccion	42350.93	376410.53	0	1493.67	20336.36	50602.91	257844803.3
Valoracion_terreno	46248.22	1112690.46	0	3367.19	14249.16	30830.39	770123027.5
Valoracion_adicionales	540.71	98956.14	0	0	0	0	98254843.1
Valoracion_total	91481.62	1435622.04	0.08	11115.88	43924.28	90504.86	1126222674

Fuente: Catastro DMQ

3.2. Procesamiento de la información

Una vez revisada la información disponible y su estructura en cada una de las tablas relacionadas al catastro del DMQ se procede con su depuración y reducción de dimensionalidad[2].

Se contempla como temporalidad para el análisis descriptivo y para la elaboración de los modelos la correspondiente al 2013 – 2022. No se incluye la información correspondiente al año 2012, dado que algunas variables se establecen en diferencias al momento de incluirlo en el modelo. Adicionalmente, la información de la variable 'valoración_total' presenta una variable positiva elevada en este año; dado las variaciones propias de la naturaleza de esta variable y el período total de información disponible para esta variable (2010 – 2023), el proceso de corrección de la tendencia a nivel de parroquia podría incluir demasiado ruido en la serie temporal. De esta forma, se considera que la opción más viable es la de segmentar el período de estudio a partir del año 2013. Por otro lado, la información macroeconómica no está disponible al año 2023, por lo que este año no es incluido en el estudio.

El análisis de las viviendas tomará como unidad de análisis a cada predio. Para ello, en primer lugar, se utiliza la tabla *Construcción,* la cual constituye la base del análisis temporal de la población a incluir en el análisis al permitir definir la evolución en el número de construcciones que se han integrado a la ciudad a través del tiempo.

No obstante, es crucial enfatizar que frecuentemente diversas construcciones pueden integrarse en una única unidad residencial, por tanto, se procede con la eliminación de duplicados de la tabla *Construcción* conservando los registros con valores más altos en las variables 'valoración' y 'área_total'. De esta forma, se pasa de 2,329,570 registros a 902,706 registros (construcciones).

Una vez obtenida la información necesaria de la tabla *Construcciones* se procede con su integración con la tabla *Predio*. Una de las principales novedades identificadas es el porcentaje de valores perdidos en las columnas '*uso'* y '*tipo_construccion*' de la tabla *Construcciones* y en la columna '*uso_principal'* de la tabla *Predio*. El 14.6% de los registros tiene valores en blanco en alguno de estos campos y 13.7% en los tres campos a la vez.

Este 13.7% de registros, equivalente a 140,045 predios, no posee valores en los campos. año de construcción y área de construcción. Adicionalmente, 77% no muestra información en destino

[2] En este apartado se presenta una versión resumida del procesamiento llevado a cabo. Para un mejor entendimiento se recomienda la revisión del *Anexo 1,* mismo que comprende todo el procesamiento, desde la información del Catastro en su estado original hasta la composición de tabla final a utilizarse tanto en formato html como .ipnyb para facilitar su réplica.

económico y del 23% restante, solo un 7% corresponde a un destino económico 'Habitacional'. Asimismo, alrededor del 40% de estos registros no poseen información de ubicación.

Por tanto, considerando que el año de construcción será fundamental en nuestro análisis, estos registros (140.045) quedan excluidos del estudio.

Una vez eliminados estos registros, la Tabla 6, Tabla 7 y Tabla *8*, detallan las diferentes categorías encontradas en las variables 'uso', 'uso_principal' y 'tipo_construccion'. Como se puede observar aún existen registros que mantienen tipos o usos fuera del ámbito de estudio. Por tal motivo, se excluyen un 1.83% de los registros que muestran valores diferentes a las categorías casa, departamento o vivienda en los campos antes detallados.

Tabla 6 Tipos de construcciones en Catastro DMQ

Tipo construcción	Número	Porcentaje
Casa	389,779	68.76
Departamento	171,322	30.22
Patio / jardín	1,306	0.23
Parqueadero	906	0.16
Almacen	666	0.12
Lavanderia / secadero	574	0.10
Galpón	559	0.10
Otros	1,485	0.26
Nulos/missing	295	0.05
Total	**566,892**	**100**

Fuente: Catastro DMQ

Tabla 7 Tipos de uso para construcciones en Catastro DMQ

Uso	Conteo	Porcentaje
Casa	388,489	68.53
Departamento	171,228	30.20
Almacén, comercio menor, local	624	0.11
Parqueadero descubierto con acabado	532	0.09
Vivienda	492	0.09
Casa barrial	451	0.08
Lavandería, secadero abierto	439	0.08
Otros	4,631	0.82
Nulos/missing	6	0.001
Total	566,892	100

Fuente: Catastro DMQ

Finalmente se elimina 6 registros que no poseen información en año de construcción.

Considerando que la información a analizarse será a nivel parroquial y también se realizarán pruebas con información agregada a nivel de Administración Zonal, se integra la información con la tabla *Lote,* desde donde se obtiene las dos variables antes mencionadas.

También se integra la tabla resultante con la tabla *Propiedad* para la obtención del número de ventas.

Tabla 8 Tipos de uso para construcciones en Catastro DMQ

Uso principal	Conteo	Porcentaje
Casa	389,873	68.77
Departamento	172,296	30.39
Almacén, comercio menor, local	976	0.17
Vivienda	492	0.09
Casa barrial	386	0.07
Parqueadero cubierto	316	0.06
Patio, jardín abierto césped	264	0.05
Otros	2,026	0.36
Nulos/missing	*263*	*0.05*
Total	566,892	100

Fuente: Catastro DMQ

El proceso para la obtención del número de ventas desde la tabla Propiedad resulta un poco más complejo.

En primer lugar, en esta tabla pueden existir varios registros para cada predio. Por tal motivo se consideran únicamente los registros que en el campo 'cat_t_adquisicion_nombre' presentan la categoría 'compraventa'. Si los registros no poseen esta categoría se considera el último registro bajo el supuesto que otra categoría también implica el traspaso de este bien mueble.

Adicionalmente existen un 9% de registros, una vez eliminados duplicados, que no presentan información en los campos 'anio_escritura' o 'anio_inscripcion'. Para imputar un valor en estos registros se considera la diferencia promedio existente entre el 'anio_escritura' y 'anio_construccion'. Luego, una vez que se cuente con el anio_construccion se utiliza la diferencia antes obtenida para el cálculo del anio_escritura.

De esta forma, la información final a utilizarse en el modelo consta de 556,340 registros.

Para el valor del precio por metro cuadrado no fue posible contar con el histórico del Catastro bajo la misma estructura que la correspondiente al año 2023 sino únicamente con algunas variables de la tabla *Predio*. El principal inconveniente es la ausencia de la variable 'area_construccion' que permita obtener el precio catastral por metro cuadrado.

Por tanto, para la obtención del precio se utilizó el ID de los predios para los que se cuenta con el año de construcción y demás variables para fitrarlas en este histórico y utilizar el área de construcción registrado en 2023. De esta forma se obtuvo el histórico del precio para el universo de las viviendas estudiadas dado que la inclusión.

Considerando que variaciones en el '*área_total*' así como inconsistencias en la información podrían alterar los resultados del precio por metro cuadrado promedio, previo al calculo del precio promedio se eliminan los registros considerando las siguientes condiciones:

1. Se asume una distribución normal de la variable y se eliminan los registros por fuera de dos desviaciones (alrededor del 5% de datos) de la mediana del precio.

La información de las variables macroeconómica incluyendo créditos BIESS no fue objeto de procesos y fue utilizada directamente en el estudio conforme se muestran en cada institución utilizada para el efecto.

4. Implementación del modelo

Para la implementación del modelo se tomó en consideración la revisión bibliográfica sobre evaluación del mercado inmobiliario analizado en la sección 2. A partir de dicho análisis se determinó que la mejor manera de evaluar el mercado inmobiliario en el Distrito Metropolitano de Quito es por medio de la representación de tres ecuaciones, las cuales representan, la oferta, demanda y su influencia en el precio por metro cuadrado. Adicional, se toma en consideración aspectos macroeconómicos que influencian la dinámica del punto de equilibrio del mercado.

El propósito de la investigación es evidenciar los diferentes factores económicos que influencian en el mercado inmobiliario de Quito, por ende, es necesario contemplar sus características desde la oferta y la demanda. La forma de representar dicha interacción posee limitaciones, ya que se necesita de una estructura estadística capaz de establecer relaciones entre diversas ecuaciones o varias variables dependientes. Una de las estrategias planteadas por las investigaciones analizadas en el apartado anterior es por medio de un modelo de ecuaciones estructurales o utiliza mínimo cuadrados ordinarios en dos o tres etapas. El usar ambas metodologías se debe a una práctica de contrastar la confiablidad de las estimaciones, y mitigar posibles restricciones como el no cumplir con supuestos estadísticos de autocorrelación o heteroscedasticidad los cuales son resueltos por el modelo de ecuaciones estructurales SEM.

En los siguientes apartados se detallan los pasos realizados hasta llegar a la estructura final del modelo. El procedimiento considerado se basa en condiciones estadísticas que permiten obtener estimaciones consistentes y confiables para su uso en el desarrollo de posibles políticas a través de una interpretación adecuada de los diferentes factores que influyen en el mercado inmobiliario de Quito.

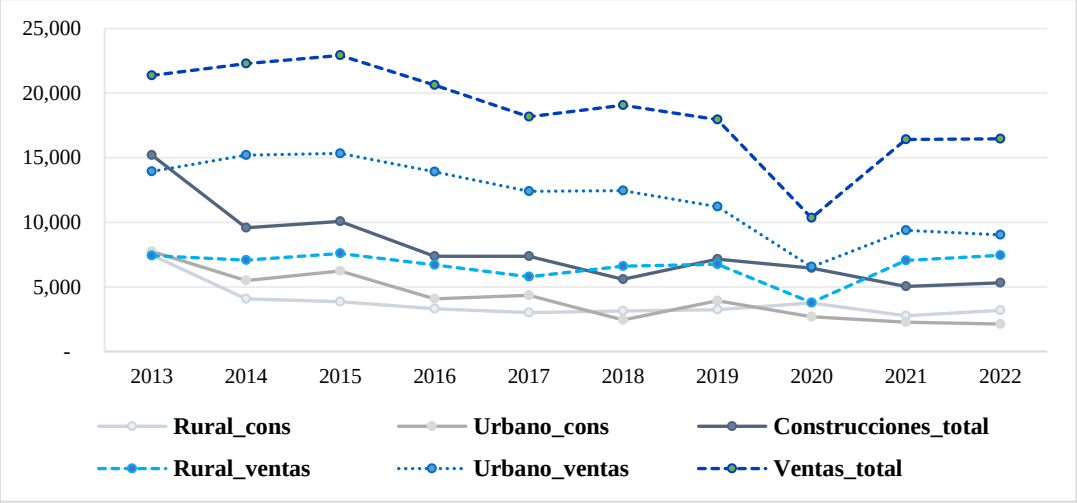
4.1. Análisis descriptivo

Previo a la implementación del modelo y, con la finalidad de tener un panorama más claro de la situación del mercado inmobiliario en el Distrito Metropolitano de Quito, se realiza un análisis descriptivo de las principales variables a incluirse.

Como se mencionó en la sección anterior, el período de estudio abarca los años 2013 – 2022. El Gráfico 1 muestra la evolución en el número de construcciones nuevas en el DMQ. La tendencia decreciente es bastante marcada, con una reducción en el número de construcciones desde el año 2013 al año 2022 del 65.1%, siendo esta reducción más marcada en el sector urbano con una reducción del 72.6% en comparación con el 57.2% de la zona rural. Durante gran parte del período analizado el número de nuevas construcciones en las parroquias urbanas ha sido ligeramente mayor, no obstante, desde el año 2020 esta tendencia cambia, siendo las construcciones en la zona rural 1.4 veces más que las del sector urbano.

A priori, no es posible establecer la causa de esta disminución. Si bien, puede atribuirse a diversos factores económicos y regulatorios que influyen en el mercado inmobiliario, la fluctuación en la oferta y demanda de propiedades, cambios en las políticas de urbanismo, o la variabilidad en la inversión en el sector de la construcción pueden ser algunas de las causas subyacentes de esta tendencia. Por otro lado, la reducción en el número de construcciones nuevas también podría reflejar una saturación en ciertas zonas o un cambio en las preferencias de los consumidores hacia ciertas zonas del DMQ.

Gráfico 1 Evolución del número de nuevas viviendas en el DMQ

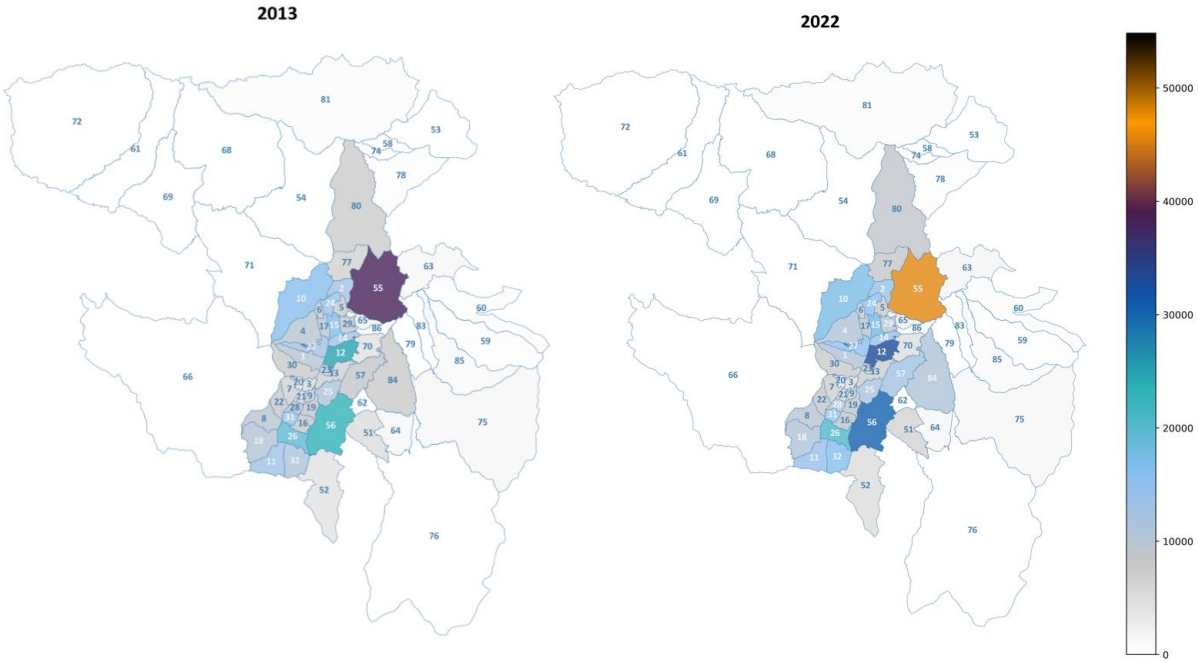


Fuente: Catastro DMQ

Es evidente que, la tendencia antes indicada puede no aplicarse en algunas parroquias mientras en otras seguramente esta realidad está mucho más acentuada. El Gráfico *2* muestra de forma inicial el número de viviendas existente en cada parroquia en los años 2013 y 2022.

Para el año 2013 la parroquia con mayor número de viviendas es Calderón (55) con 38,392 viviendas seguido de Iñaquito (12) y Conocoto (56) con 24,751 y 22,986 viviendas respectivamente. Le siguen Quitumbe (26) y la Kennedy (15) con 19,294 y 16,433 viviendas. Estas cinco parroquias representan el 27.4% del total de predios del DMQ. Para el año 2022 estás cinco parroquias siguen siendo las que concentran mayor participación de viviendas en el mismo orden, representando el 29.1% del total de predios del DMQ.

Gráfico 2 Stock de viviendas en el DMQ



Fuente: Catastro DMQ

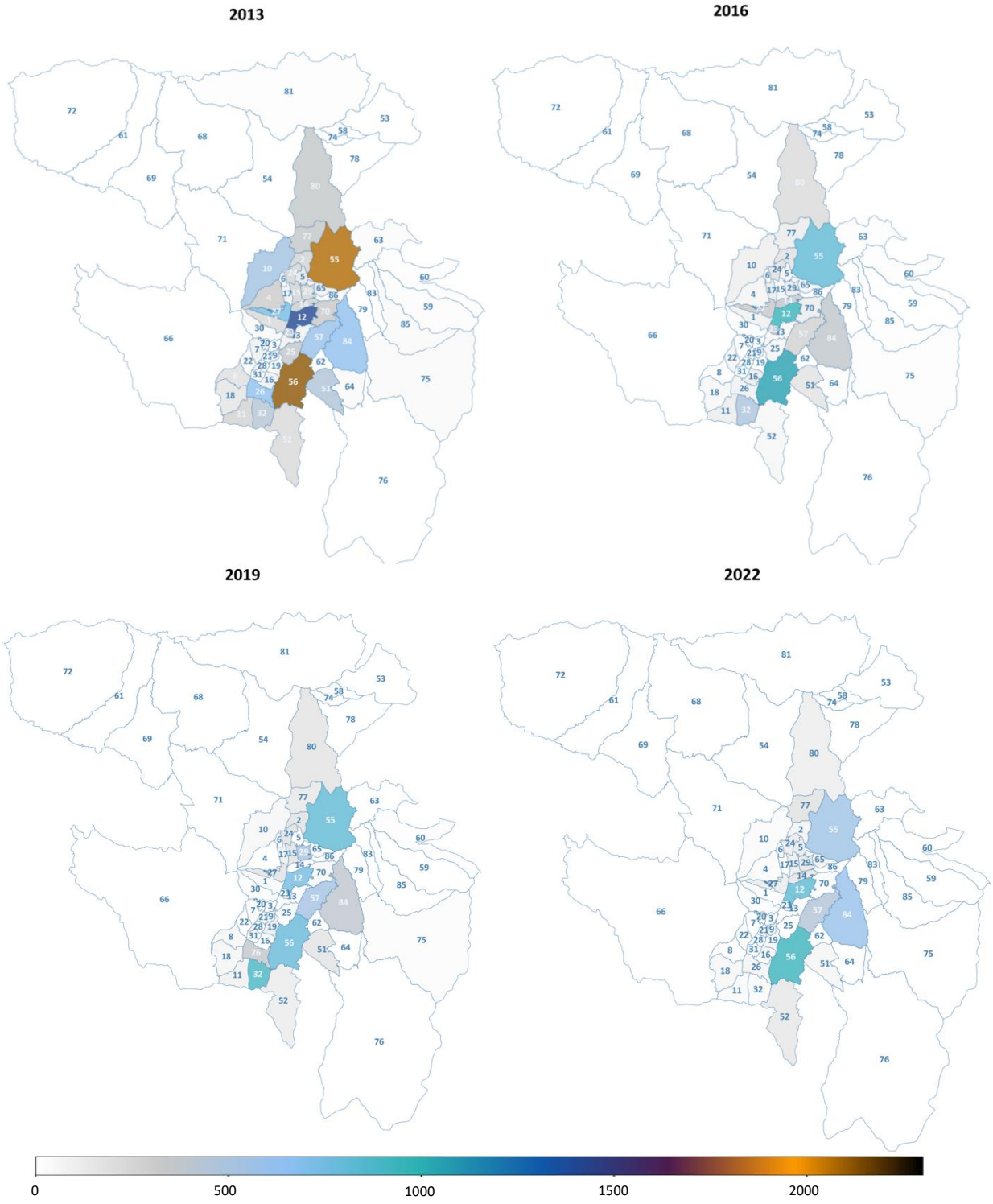
Pese a ello, hay algunas parroquias que han experimentado mayor crecimiento en este período. Entre el 2013 y el 2022 las parroquias con mayor crecimiento son Tumbaco (84) y Cumbayá (57) con el 53.9% y 50.7%. Le siguen Turubamba (32), Iñaquito (12) y Conocoto (56) con el 40.3%, 33.6% y 33.1% respectivamente.

Las parroquias urbanas de menor crecimiento son Solanda y San Bartolo, permaneciendo con un número de viviendas casi estable, registrando variaciones de alrededor del 1% en el período analizado.

Por otro lado, si analizamos el número de nuevas construcciones, encontramos que la parroquia con mayor número de construcciones en 2013 fue Conocoto (56) con 2120 viviendas seguido de Calderón (55) con 2072 viviendas e Iñaquito (12) con 1406 viviendas. Estas tres parroquias, en este año, concentraron el 37% del total de nuevas construcciones. Para el año 2022, Conocoto se mantiene como la parroquia con mayor número de nuevas construcciones con 938 seguido de Iñaquito con 825

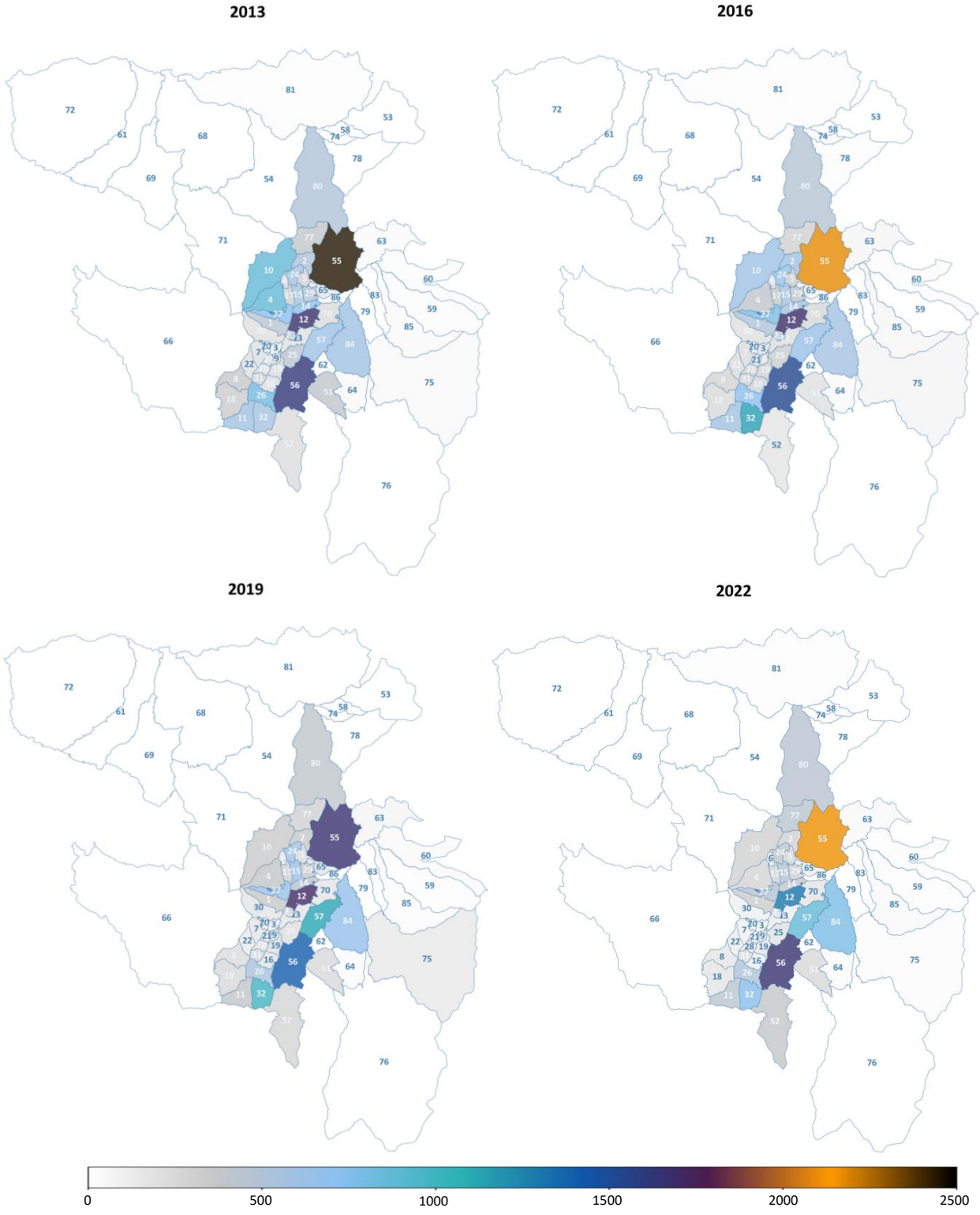
y Tumbaco (84) con 596 construcciones. Juntas representaron el 44% de las construcciones de ese año.

Gráfico 3 Nuevas viviendas por parroquia



Fuente: Catastro DMQ

Gráfico 4 Número de ventas por parroquia



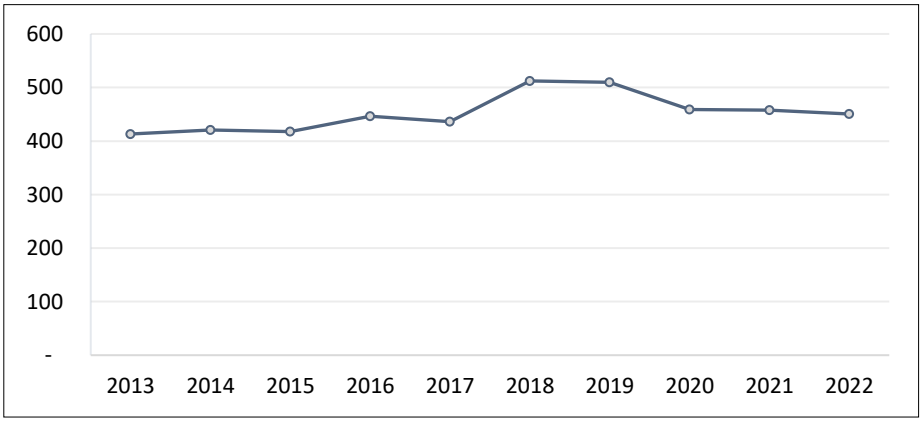
Fuente: Catastro DMQ

El número de ventas lógicamente muestra un comportamiento similar. Calderón, Iñaquito y Conocoto son las parroquias que registran un mayor número de ventas entre el año 2013 – 2022 representando el 28% del total de ventas. Las parroquias antes mencionadas se mantienen también año a año como las parroquias con mayor número de ventas de viviendas. Sin embargo, en 2013, El Condado (10) y Cochapamba (4) aparecen como la 4 y 5 parroquia con mayor número de ventas seguidas de Quitumbe (26) y Rumipamba (27). Para el año 2022, las parroquias con mayor numero ventas luego de las 3 ya mencionadas (Calderón, Iñaquito y Conocoto) son Cumbayá y Tumbaco seguido de Turubamba y Quitumbe.

Entre el 2013 y 2019 Cumbayá experimenta el mayor crecimiento (81.4%) en el número de ventas experimentando una importante reducción entre 2019 y 2022 durante y postpandemia. Entre el 2013 y 2022 existen fuertes decrecimientos en el número de ventas anuales. El más notorio en parroquia urbanas se observa en Cochapamba (4), El Condado (10) y La Libertad (20) con el 68.8%, 65.3% y 64.3%. La única parroquia Urbana con un crecimiento positivo en este periodo es Turubamba con el 30.7%.

Paralelamente, existe una variabilidad limitada en el valor del metro cuadrado catastral. Si bien la información limita el entendimiento de este fenómeno, el incremento en el valor puede estar influenciado por factores como la escasez de terrenos disponibles en zonas de alta demanda, el aumento de los costos de materiales y mano de obra, o la disminución en el crecimiento económico acentuado desde la pandemia.

Gráfico 5 Evolución del valor del metro cuadrado (catastral) en construcciones

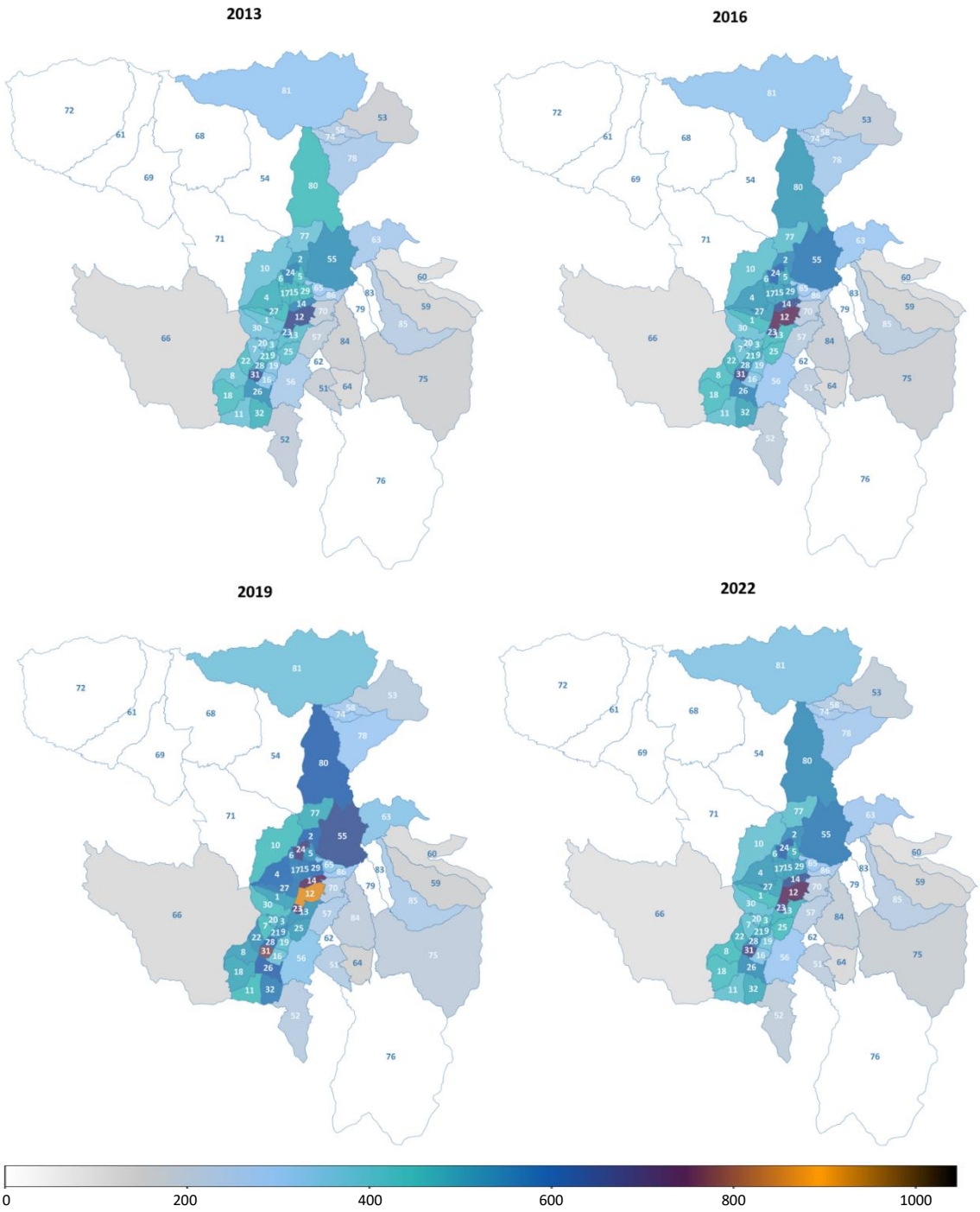


Fuente: Catastro DMQ

Es interesante notar cómo el aumento del valor del metro cuadrado contrasta con la disminución en el número de construcciones y predios, lo cual podría sugerir una mayor inversión en propiedades de

alto valor, una revalorización del mercado existente o un persistente incremento en los precios para este sector.

Gráfico 6 Variación en el precio catastral por metro cuadrado



Fuente: Catastro DMQ

Además, esta tendencia creciente en los precios podría indicar un mercado cada vez más orientado hacia propiedades de lujo o la gentrificación de ciertas áreas, lo que posiblemente esté desplazando a segmentos del mercado más económicos. Asimismo, esta dinámica refleja posiblemente una diversificación en las inversiones inmobiliarias, donde los inversores buscan propiedades con mayor potencial de valorización a largo plazo.

A la par del número de viviendas, también es importante analizar el precio catastral por metro cuadrado promedio por parroquia. Como se aprecia en el Gráfico 6 Iñaquito (12) se mantiene en todos los años analizados como la parroquia con el precio por metro cuadrado más alto, seguido de Solanda (31), Jipijapa (14) y Mariscal Sucre (23).

Llama la atención la presencia de Solanda por encima de parroquias como Jipijapa o Rumipamba. Si bien, durante el procesamiento de la información se analizó de forma cuidadosa la presencia de registros atípicos, en el caso de Solanda el valor promedio no parece estar afectado por un valor atípico, sino que es la tendencia en todos los registros que pertenecen a esta parroquia. Esta situación compromete la calidad de información presentada en el catastro del DMQ.

En el año 2013, las parroquias urbanas con el precio más bajo son La Argelia (16), La Ferroviaria (19) y la Libertad (20) con valores alrededor del 330 dólares por metro cuadrado. Mientras que las parroquias rurales con el valor más bajo fuero El Quinche (60), Lloa (66) y La Merced (64) con valore inferiores a los 125 dólares. La parroquias rurales con el valor más alto fueron Calderón (55) con un valor de alrededor 525 dólares seguido de San Antonio de Pichincha (80).

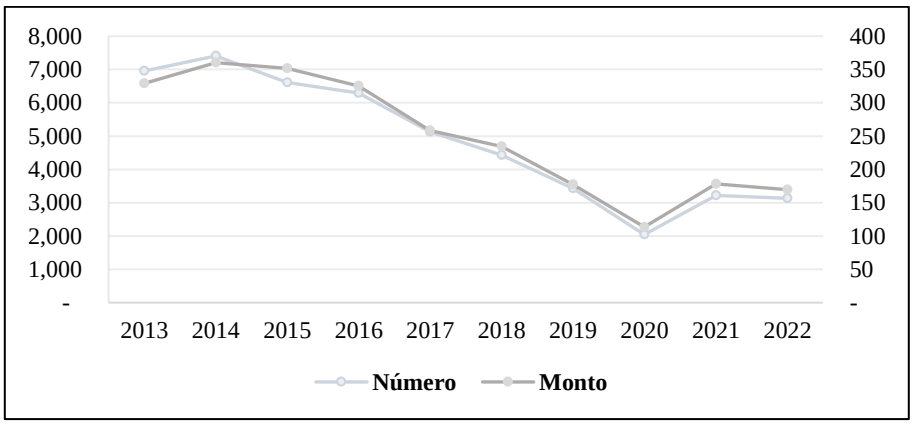
Para el año 2022, esta realidad es similar, con Calderón y San Antonio de Pichincha como las parroquias rurales con mayor valor en el precio por metro cuadrado y Lloa , El Quinche y La Merced como las parroquias rurales con el valor más bajo, lo mismo para las parroquias urbanas.

En cuanto a la variación en el precio, entre el año 2013 al 2022, las parroquias urbanas con mayor incremento están Jipijapa y Mariscal Sucre con el 20.1% y 15.7% respectivamente, mientras las parroquias rurales con mayor crecimiento son San Antonio de Pichincha y Atahualpa con el 19.3% y 18.2%. Como referencia, en este período, las parroquias de Cumbaya y Tumbaco tuvieron incrementos del 14.9% y 14.6%, Conocoto del 13.1% e Iñaquito del 12.0%.

La Ecuatoriana (18) y Quitumbe (26) son las parroquias urbanas con menor crecimiento, por debajo del 3%, mientras en las parroquias rurales están Lloa (66), La Merced (64) y Yaruquí (85).

Finalmente, se presenta una breve descripción de la evolución del monto otorgado por el Banco del Instituto Ecuatoriano de Seguridad Social. El financiamiento del BIESS en el 2013 representó el 79.4% del total de créditos otorgados en el DMQ para la adquisición de viviendas, mientras que para el 2022 este porcentaje fue de 68.5%. Esto muestra la relevancia de esta institución en la financiación de vivienda en el DMQ frente al sector privado. La evolución del financiamiento del BIESS muestra una contracción en el número de créditos otorgados y en monto. El número de operaciones entre el 2013 y 2022 se redujo en un 54.9% y en términos de volumen de crédito este se redujo en un 48.3%.

Gráfico 7 Evolución del número y monto de créditos hipotecarios otorgados por el BIESS en el DMQ

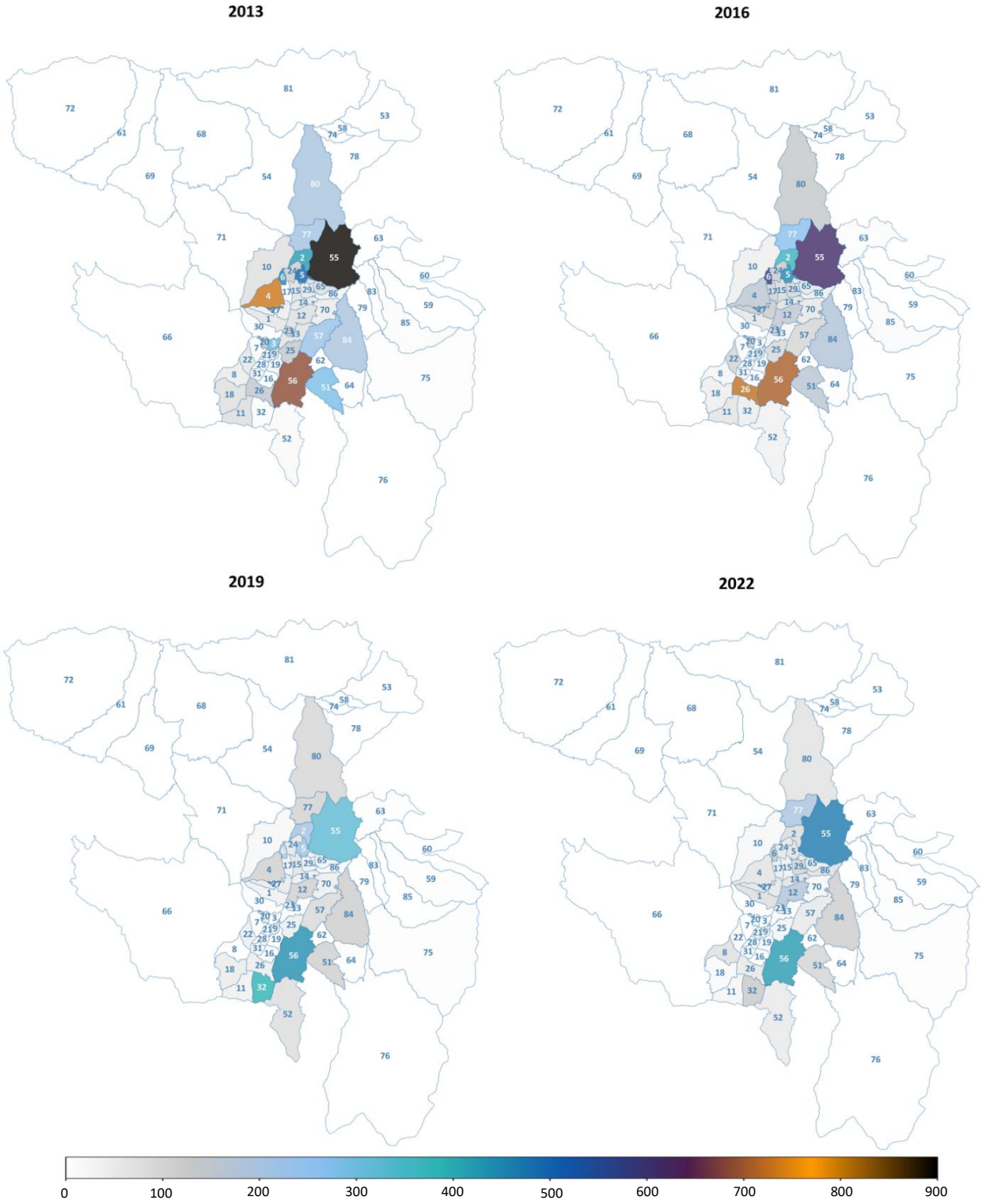


Fuente: Catastro DMQ

Al realizar la revisión por parroquias[3] en el DMQ, se observa la concentración en las parroquias de Conocoto (56) y Calderón (55). Para el año 2022 estas parroquias representan el 14.1% y 13.1% respectivamente, por tanto, agrupan el 27,2% del total de créditos del BIESS. Para el mismo período le sigue Pomasqui (77) con el 6.7% e Iñaquito (12) con el 6.5%.

[3] Conforme los manuales de reporte de la Superintendencia de Bancos, la estructura de información actual no contempla la parroquia de construcción de una vivienda, no obstante, se tiene la ubicación de la garantía siendo esta última la considerada para la estratificación de la información.

Gráfico 8 Número de créditos hipotecarios otorgados por BIESS en DMQ



Fuente: Superintendencia de Bancos

En resumen, la información analizada por si sola es insuficiente. El mercado inmobiliario del DMQ muestra un panorama complejo, donde, por ejemplo, la disminución en el número de construcciones y predios podría estar dando paso a un mercado más enfocado en la calidad y valorización a largo plazo de las propiedades, en lugar de la cantidad. No obstante, rechazar o aceptar este tipo de hipótesis requieren de la integración con fuentes de información macroeconómica, financiera entre otras. La aplicación de los modelos estadísticos propuestos permitirá comprender esta dinámica, así como las razones detrás del comportamiento de estas variables y proporcionarán una comprensión más profunda de las fuerzas subyacentes que moldean tendencias clave, como la variación en el número de predios y la valoración de terrenos y construcciones. Al desentrañar estos patrones, se podrán identificar las palancas críticas que impulsan el comportamiento del mercado y, en consecuencia, se facilitará la toma de decisiones informadas basadas en datos tanto para inversores como para formuladores de políticas

4.2. Proceso de selección de variables

El primer paso para el desarrollo de los modelos estadísticos es la depuración y estructura de las variables a utilizar. Para ello, el proceso de selección de variables se basó en criterios teóricos en los cuales se observaba la interacción de variables demográficas, de crédito, macroeconómicas y del mercado inmobiliario. Adicionalmente, una vez determinado las variables cuya representación teórica cumple con el propósito del modelo, se consideró su significancia estadística con un nivel de confianza del 95%. En el caso de las variables endógenas, su determinación se basó en el análisis de las investigaciones mencionadas en la sección 2. A continuación, se detallan las variables dependientes:

- *num_ventas*: Número de ventas efectuadas en el año
- *total_construcciones:* Número total de construcciones realizadas en el año
- *precio_m2*: Precio de la vivienda por metro cuadrado

La variable “Número de Ventas” representa la demanda y oferta, bajo un punto de vista de equilibrio en la estructura de ecuaciones. Se intenta medir factores macroeconómicos y demográficos que influyen en la demanda de bienes inmuebles y de la misma manera determinar cuáles son las condiciones del mercado que se encuentran detrás de las variaciones en la oferta inmobiliaria en Quito. A partir de dichas estimaciones, se pretende analizar la relación que existe entre la oferta y la demanda sobre el precio catastral por metro cuadrado de viviendas terminadas en el DMQ. Cabe recalcar que varias de las estimaciones también se realizaron utilizando el total de construcciones como proxy de la oferta y demanda del mercado inmobiliario.

Para la selección de las variables exógenas el análisis se basó en su nivel de significancia e interpretación en la teoría económica. En total, se eligieron ocho variables exógenas que incluyen en la oferta y demanda del mercado inmobiliario de Quito.

Tabla 9 Estructura Oferta y Demanda

Demanda	Oferta
Monto de crédito del segmento vivienda (BIESS)	Monto de crédito del segmento comercial
Índice de precios al consumidor	Índice de precios de la construcción
Desempleo	Tasa activa referencial
Valor AIVA	Salario promedio en la construcción
Crecimiento poblacional	Stock de viviendas
Tasa efectiva ponderada vivienda (BIESS)	

Fuente: Elaboración propia

La Tabla 9 muestra la estructura de la oferta y demanda en el sistema de ecuaciones utilizado para representar las condiciones del mercado inmobiliario de Quito. En el lado de la demanda, la variable del volumen de crédito del segmento de vivienda excluyó créditos destinados a la remodelación o ampliación de viviendas existentes e incluye a banco públicos, privados y al BIESS con una desagregación a nivel de parroquia según domicilio del usuario. En la oferta, el volumen de crédito del segmento comercial incluye únicamente al sector de la construcción, específicamente los créditos destinados a la construcción de edificaciones y viviendas e incluye a bancos públicos y privados.

Las variables macroeconómicas como el IPC (Índice de precios al consumidor), IPCO (Índice de precios de la construcción, tasa activa referencial, desempleo y crecimiento población fueron obtenidas con una temporalidad anual y a nivel cantonal.

En el siguiente apartado se detalla los procedimientos realizados en la información utilizada para la calibración del modelo.

4.3. Preparación de datos para modelamiento

Para la preparación de la muestra empleada en el modelo se utilizaron diferentes fuentes de información, dado los múltiples factores que influyen en el mercado inmobiliario. Se describe a continuación las fuentes y usos de la información:

Tabla 10 Fuentes y uso de información

USO	Fuentes
Monto de Crédito	Superintendencia de Bancos
Tasa efectiva ponderada	Superintendencia de Bancos
IPC	Banco Central del Ecuador
Desempleo	Banco Central del Ecuador
Tasa Activa Referencial	Banco Central del Ecuador

AIVA	Municipio de Quito
Número de Ventas	Municipio de Quito
Número de Construcciones	Municipio de Quito
Precio m2	Municipio de Quito
IPCO	INEC
Salario construcción	INEC
Crecimiento Poblacional	INEC

Fuente: Elaboración propia

Las fuentes utilizadas en las estructuras finales de los modelos estadísticos, se encuentra descrita en la Tabla 10. La base final contempla 550 observaciones agregadas a nivel de parroquia y cuenta con un histórico de información desde el 2012 hasta el año 2022 con una frecuencia anual. No obstante, como se mencionó previamente, para la estimación se considera únicamente desde el año 2013.

4.4. Calibración del modelo Mínimo cuadrados por tres etapas (3SLS)

Nelson y Plosser (1982) advierten que los datos de series temporales macroeconómicas a menudo presentan un proceso de raíz unitaria lo que puede conducir a generar estimaciones espurias. Por lo tanto, se realiza la verificación de esta conforme el proceso recomendado por Levin, Lin y Chu (2002), proceso adaptado para modelos de panel considerando la estacionariedad conjunta.

Los resultados de las pruebas se muestran en la Tabla 11. Como se indica, todas las series de datos son estacionarias con excepción de las variables 'IPC' e 'IPCO' para las que es necesario un proceso de diferenciación.

Tabla 11 Prueba de estacionariedad

Variables	Test levinlin p value
numero_ventas	0.0038
precio_m2	0.0000
biess_monto_creditos	0.0008
biess_num_creditos	9.29 e-09
crec_poblacional	0.0000
Ipc	0.0922
desempleo	0.0000
aiva_valor	0.0003
salario	0.0000
stock_construcciones	3.87 e-09
ipco_indice_general	0.9900
o_monto_cred_privado	2.71 e-06
tasa_efectiva_ponderada	1.64 e-79
tasa_activa_referencial	0.0000

Por otro lado, también es necesario constatar la no presencia de multicolinealidad. Una vez aplicado el Factor de Inflación de la Varianza se encuentran los siguientes resultados:

Tabla 12 Variance Inflation Factor (VIF)

Variable	VIF
ln_crec_poblacion	9.84
ln_aiva_valor	6.97
ln_salario	5.04
diff_ipc	3.64
diff_ipco_indice_general	2.83
ln_desempleo	2.46
ln_tasa_activa_referencial	2.24
ln_tasa_efectiva_ponderada	2.08
diff_stock_construcciones	1.79
ln_biess_monto_creditos	1.53

Los resultados observados en la Tabla 12 muestran una alta correlación de las variables 'crecimiento poblacion' y 'aiva valor', no obstante puede ser un valor tolerable considerando que serán estimada también en diferentes ecuaciones.

Como señala Gujarati & Porter (2009), si no existe un esquema de ecuaciones simultáneas o presencia del problema simultaneidad, se puede generar estimadores consistentes y eficientes bajo el método de Mínimos Cuadrados Ordinarios, caso contrario, es necesario utilizar métodos que permitan abordar la presencia de endogeneidad como el de mínimos cuadrados de dos etapas (2SLS) o tres etapas (3SLS). Por tanto, es necesario comprobar si la variable precio, establecida como endógena, está correlacionada con el término de error.

No obstante, previo a verificar la existencia de endogeneidad, es necesario validar una especificación adecuada para el modelo. Considerando la estructura de la información, con una parte temporal (2013 a 2022) y a nivel de parroquias, las pruebas aplicadas están especificados para modelos de datos de panel. Estos resultados se detallan en el Anexo 3, y evidencian, bajo una estimación lineal considerando datos de panel existe presencia de heterocedasticidad y autocorrelación. Considerando estimaciones que cubren estas distorsiones[4], se determina, mediante la prueba de Hausman, que es pertinente capturar efectos fijos.

Es necesario hacer hincapié en que las suposiciones en la elección de cualquiera de los dos modelos limitan los resultados. Es decir, al optar por un modelo de efectos aleatorios, se asumiría que las diferencias no observadas entre parroquias (como características socioeconómicas, calidad de las escuelas, tasas de criminalidad, etc.) no están correlacionadas con las variables endógenas en nuestro

[4] Comando xtgls (panel correlated) del software Stata cubren problemas de heterocedasticidad y correlación serial en modelo de efectos aleatorios y xtreg (vce)

modelo (número de ventas, precio). Esta suposición podría ser irreal considerando que, dentro del mercado inmobiliario, factores no observados a nivel de barrio pueden influir significativamente en la oferta y la demanda.

Por otro lado, al elegir un modelo de efectos fijos, se asume que hay características específicas de cada barrio que podrían estar correlacionadas con las variables independientes y que estos efectos no varían en el tiempo. Esto podría incluir factores como la ubicación geográfica de la parroquia, historia, infraestructura local, y otros atributos únicos que no cambian a lo largo del período de estudio. Al usar efectos fijos, se controla explícitamente estas diferencias invariables entre barrios, lo que tampoco podría cumplirse en algunos sectores del Distrito Metropolitano de Quito.

Por tal motivo, se considera realizar las siguientes estimaciones:

- Un modelo de efectos fijos. Sin embargo, dado que se está trabajando con un modelo de panel con una extensión en las unidades transversales mucho mayor al factor temporal, una estimación bajo este método implica una mayor pérdida de grados de libertad y una estimación menos eficiente que con efectos aleatorios.
- Un modelo de efectos mixtos, considerando que es importante evaluar las diferencias no observadas en las variables incluidas y en las características de cada parroquia.

No obstante, como se mencionó en el marco metodológico. Bajo un enfoque de modelo de panel no se soluciona el problema de endogeneidad corroborado en esta sección.

La prueba de endogeneidad de Hausman, conforme el proceso detallado en Gujarati & Porter (2009) arroja los siguientes resultados:

Tabla 13 Prueba de endogeneidad de Hausman

	Parameter	Std. Err.	T-stat	P-value
Residuos	0.4166	0.0568	7.339	0.0000

Los resultados muestran que los residuos obtenidos de la estimación de la forma reducida en el sistema de ecuaciones planteado resultan significativos a menos del 1%. Adicionalmente, las pruebas aplicadas confirman un problema de simultaneidad en la variable precio por metro cuadrado para las ecuaciones de oferta y demanda de viviendas (Tabla 14).

Tabla 14 Test Wu-Hausman y Durbin

		Parameter	P-value
Ecuación Oferta	Durbin	25.91	0.0000
	Wu-Hausman	27.61	0.0000
Ecuación Demanda	Durbin	88.00	0.0000
	Wu-Hausman	117.58	0.0000

Por tal motivo, es necesario la estimación de las ecuaciones bajo un enfoque de 2SLS o 3SLS. Por tal motivo, se utiliza el proceso expuesto en Gujarati & Porter (2009) para la estimación del método de 2SLS y 3SLS a través de la estimación de la variable endógena desde una forma reducida del sistema de ecuaciones.

4.5. Calibración del modelo de Ecuaciones Estructurales (SEM)

La calibración del modelo de ecuaciones estructurales se basó en el análisis de diferentes procesos de especificación en las interacciones de las variables exógenas y endógenas en total se realizaron diez iteraciones con diferentes variables y estructuras. Adicionalmente, se aplicaron transformaciones a las variables, con la finalidad de generar indicadores confiables y una correcta especificación del modelo.

La calibración se basó en obtener una adecuada significancia en las variables exógenas y que los indicadores de performance cumplan con los umbrales definidos en la sección 5. Adicionalmente, no debe existir una alta correlación entre las variables lo que se corroboró en la Tabla 12.

En el caso de la variable de tasa de crecimiento poblacional se eliminó del análisis debido a un valor VIF mayor a 5 e indicadores de calibración como RMSEA y CFI observados en la Tabla 15. A continuación, se presentan las ecuaciones finales.

$$\log(Ventas) = \log(precio\ m2) + \log(Credito) + \log(diff_IPC) + \log(BIESS) + \log(Desempleo) \quad (27)$$

$$\log(Total_construcc) = \log(precio\ m2) + \log(Credito) + \log(diff_IPCO) + \log(diff_Salario) \quad (28)$$

$$\log(precio_m2) = \log(Número\ de\ Ventas) + \log(Total\ Construcciones) \quad (29)$$

Para obtener significancia en las variables y mitigar problemas de correlación se aplica la transformación logarítmica. Adicionalmente, para el proceso de minimización de error se utilizó el método Wishart Maximum Likelihood (MLW), disponible en las opciones de la librería de "semopy" bajo el lenguaje de programación de Python.

Una vez definida las ecuaciones finales para los modelos de dos etapas y SEM, se procede al análisis de resultados y comprobación de su correcta especificación.

5. Análisis de performance del modelo y testeo

5.1. Diseño del proceso de performance y testeo

Para comprobar que los modelos calibrados estén correctamente especificados se utilizaron algunos indicadores. Para el caso del modelo SEM se utilizan los siguientes:

- Raíz del Error Cuadrático Medio de Aproximación (RMSEA, por sus siglas en ingles)
- Índice de Ajuste Comparativo (Comparative Fit Index o CFI, por sus siglas en inglés)
- Criterio de información bayesiano (BIC, por sus siglas en ingles)

Los indicadores antes mencionados se utilizan como medidas de bondad de ajuste con la finalidad de determinar si las predicciones del modelo son confiables y la estructura de las ecuaciones definidas son correctamente especificadas. Adicionalmente, para los dos primeros indicadores se establecen umbrales, que se detallan a continuación:

Tabla 15 Indicadores de Performance

KPI	Umbral
RMSEA	≤ 0.10
CFI	≥0.90

Fuente: (Ding-Geng Chen, 2024)

Los resultados obtenidos en el modelo seleccionado se detallan a continuación:

Tabla 16 Resultados KPI

Modelo	CFI	RMSEA	BIC	LogLik
SEM	0.953599	0.048253	88.022920	0.157968

Acorde a los resultados observados en la Tabla *16* el modelo SEM utilizado para el análisis del mercado inmobiliario de la ciudad de Quito cumple con los criterios definidos en la Tabla 22. Adicionalmente, al observar el indicador CFI, se puede decir que el 0.96% de la varianza y covarianza de la data observada se encuentra explicada por la estructura del modelo SEM propuesta.

En el caso de los modelos de 2SLS y 3SLS se utilizaron como criterios de ajuste los siguientes:

- Criterio de información bayesiano (BIC, por sus siglas en ingles)
- Logaritmo natural de la función de verosimilitud (log-likelihood)
- R^2 overall y variantes, común en datos de panel o modelos 2SLS y 3SLS.

A continuación, se detallan los valores obtenidos en los modelos seleccionados:

Tabla 17 Resultados 2SLS y 3SLS

Modelo		R-squared (Overall):	Log-likelihood
2SLS	Ecuacion Demanda	0.5733	-1219.7
	Ecuacion Oferta	0.8821	-1026.1
3SLS	Ecuacion Demanda	0.641	-924.19
	Ecuacion Oferta	0.855	-1107.83

El ajuste de los mejores modelos según los criterios de ajuste mostrados en la Tabla 17, muestra que los modelos de 2SLS y 3SLS parecen proporcionar un ajuste similar, variando dependiendo de la ecuación. El modelo 3SLS, en particular, muestra el mejor ajuste global para ambas ecuaciones (demanda y oferta), lo cual es esperado dado que el 3SLS está diseñado para manejar mejor la interdependencia entre ecuaciones en sistemas simultáneos.

5.2. Comprobación de supuestos

Dentro del modelo econométrico 2SLS o 3SLS, es necesario el cumplimiento de algunos supuestos para garantizar la confiabilidad de los resultados, entre ellos presencia de endogeneidad y validez de los instrumentos.

La presencia de endogeneidad fue expuesta en la sección anterior, por lo que a continuación el análisis se centra en la validez de los instrumentos. Entre los instrumentos utilizados se encuentra la estimación residual de la variable endógena (precio), obtenida al regresarla contra otras variables exógenas del modelo, es decir, de la forma reducida, conforme lo indicado por Gujarati & Porter (2009).

Se consideran otras variables exógenas tanto del planteamiento de la oferta y la demanda. Los resultados obtenidos se detallan a continuación:

Tabla 18 Kleibergen-Paap rk LM statistic

	Statistic	P-value
Ecuación Demanda	45.683	0.000
Ecuación Oferta	148.328	0.000

Considerando que un modelo solo puede ser estimado mediante 2SLS o 3SLS si se encuentra identificado o sobreidentificado, la prueba de identificación de Kleigergen-Paap rk un valor p bajo permite rechazar la hipótesis nula de que el modelo esta subidentificado.

Tabla 19 Test para instrumentos 2SLS – Ecuación demanda

	Ecuación Demanda	Ecuación Oferta
Cragg-Donald Wald F statistic	19.633	289.88
Kleibergen-Paap rk Wald F statistic	29.397	449.60
Stock-Yogo weak ID test critical values:	*5% max. IV s. 19.93* *10% max. IV s. 11.59* *10% max. IV s. 8.75* *15% max. IV s. 7.25*	*10% max. IV s. 24.58* *15% max. IV s. 13.96* *20% max. IV s. 10.26* *25% max. IV s. 8.31*

Los resultados observados tanto para el estadístico F de Cragg-Donald y Kleibbergen-Paap es que son mayores al valor crítico de Stock-Yogo más altos, lo que indicaría que los instrumentos utilizados para suplir la variable endógena son fuertes y por tanto, el modelo está especificado de forma correcta.

Tabla 20 Hansen J statistic

	Statistic	P-value
Ecuación Demanda	2.395	0.1217
Ecuación Oferta	34.797	0.000

Considerando que las ecuaciones presentaron heterocedasticidad, no se utiliza la prueba de Sargan y en su reemplazo se utiliza la Prueba de Hansen, que es robusta a este problema. Los resultados muestran que para la ecuación de la demanda los instrumentos no se encuentran correlacionados con el término de error, de forma que se válida el instrumento. Para el caso de la oferta los instrumentos podrían no ser válidos. No obstante, se utiliza la estimación de la variable endógena conforme el proceso expuesto por Gujarati & Porter (2009) para sustituir el instrumento.

6. Outputs del modelo

6.1. Análisis de los resultados del modelo 2SLS y 3SLS

Los resultados a continuación corresponden a dos enfoques; el modelo 2SLS y un modelo bajo 3SLS. En cuanto al primer modelo, se realizó la estimación utilizando un modelo de efectos fijos en dos etapas. Entre los resultados relevantes, como se detalla en Tabla 21, es el coeficiente positivo y significativo en la variable precio por metro cuadrado en la ecuación de oferta, mientras la misma variable presenta signo negativo y es significativo en la ecuación de la demanda.

Tabla 21 Estimación de la ecuación de oferta y demanda bajo un modelo 2SLS de efectos fijos

	Parámetro	Error	P-value
Ecuación Demanda			
ln_precio_m2	-7.401	1.456	0.000
ln_biess_monto_creditos	0.097	0.033	0.004
ln_crec_poblacional	21.611	4.651	0.000
ln_tasa_efectiva_c	11.647	2.024	0.000
ln_desempleo	-1.044	0.170	0.000
Ecuación Oferta			
ln_precio_m2	3.386	0.407	0.000
diff_ipco_indice_general	2.168	0.478	0.000
diff_construcciones_acum	3.857	1.322	0.004
ln_o_monto_cred_privado	0.026	0.009	0.004
ln_aiva_valor	-1.445	0.132	0.000
ln_precio_m2; instrumentos (oferta) : ln_precio_m2_hat ln_tasa_activa_referencial; instrumentos (demanda):ln_precio_m2_hat diff_ipc ln_tasa_efectiva_c			

De igual forma, se destaca que la variable de crédito también tiene un impacto positivo tanto en la oferta como en la demanda, aunque menos significativo que la variable precio. No obstante, es importante destacar que incrementos en el 10% en la otorgación de créditos por parte de BIESS únicamente se traducirá en un incremento del 1% en el número de ventas de viviendas, lo que implicaría que un importante segmento realiza la adquisición sin este tipo de financiamiento o con financiamiento desde otras instituciones financieras.

Otras variables macroeconómicas como el desempleo y el crecimiento poblacional también influyen en la demanda de viviendas en la dirección esperada. Es interesante notar que el crecimiento poblacional es determinante en el comportamiento de la demanda. Se destaca también un signo no esperado en el Índice de Precios a la Construcción del que se esperaba una influencia negativa, al igual que la tasa efectiva ponderada con que se otorgan los créditos en el BIESS.

La presencia de coeficientes significativos para variables como el precio por metro cuadrado, el crédito privado y los indicadores económicos sugiere que estos factores juegan un papel importante en la dinámica del mercado de viviendas.

Cabe recalcar que el resultado del valor esperado en el número de viviendas para cada parroquia variará dado que se trata de una estimación de efectos fijos. Por tal motivos, la constante (correspondiente a la parroquia de código 01) así como las diferencias del resto de parroquias se detallan en el Anexo 7 para una posible estimación a este nivel.

De igual forma, en el modelo de 3SLS se presentan coeficientes con magnitudes y signos similares al obtenido bajo 2SLS (Tabla 22). Acorde a las estimaciones realizadas por el modelo, los coeficientes mantienen una adecuada interpretación económica, se observa que para el caso de la oferta representa una relación negativa los costos de producción medida por el índice IPCO y por el incremento en el salario promedio de la construcción, mientras que se incentiva la construcción cuando existe un incremento del precio por metro cuadrado o un mayor volumen del crédito productivo.

Respecto a la demanda, los resultados muestran que tanto el crecimiento de la población, el volumen de crédito hipotecario presenta un efecto positivo en el número de ventas. Por otro lado, el precio por metro cuadrado presenta un efecto negativo. Las variables tasa efectiva ponderada, desempleo e IPC no presentan coeficientes significativos y/o signos acordes a lo esperado desde la literatura.

Tabla 22 Estimación de la ecuación de oferta y demanda bajo un modelo lineal 3SLS

	Parámetro	Error	P-value
Ecuación Oferta			
log_precio_m2	3.9221	0.3961	0.0000
diff_ipco_indice_general	-1.5033	1.8627	0.4196
ln_salario	-2.9259	0.3759	0.0000
ln_o_monto_cred_privado	0.0892	0.0117	0.0000
Ecuación demanda			
ln_precio_m2	-0.8635	0.5369	0.1078
diff_ipc	2.9361	3.8585	0.4467
ln_tasa_efectiva_c	2.5926	0.7632	0.0007
ln_desempleo	-0.1139	0.5368	0.8319
ln_crec_poblacional	0.1224	0.1837	0.5050
ln_biess_monto_creditos	0.197	0.0243	0.0000
ln_precio_m2; instrumentos (oferta) : ln_precio_m2_hat ln_tasa_activa_referencial; instrumentos (demanda):ln_precio_m2_hat diff_ipc ln_tasa_efectiva_c			

6.2. Análisis de los resultados del modelo SEM

A partir de los procesos de calibración y testeo de las especificaciones del modelo se obtuvo la siguiente estructura de ecuaciones con sus respectivos coeficientes:

Tabla 23: Estructura del modelo final - SEM

L – Val	R - Val	Dirección	Magnitud
ln_numero_ventas	ln_precio_m2	-	5.81763341
ln_numero_ventas	diff_ln_ipc	-	20.00197138
ln_numero_ventas	ln_biess_monto_creditos	+	0.299449369
ln_numero_ventas	diff_ln_desempleo	-	1.26371473
ln_numero_ventas	diff_ln_aiva_valor	-	0.426596154
ln_total_construcciones	ln_precio_m2	+	3.349029735

ln_total_construcciones	ln_o_monto_cred_privado	+	0.011127459
ln_total_construcciones	diff_ln_ipco_indice_general	-	10.56879778
ln_total_construcciones	diff_ln_salario	+	10.9659999
ln_precio_m2	ln_numero_ventas	+	0.343712597
ln_precio_m2	ln_total_construcciones	-	0.103785652
ln_numero_ventas	ln_numero_ventas	+	6.820920479
ln_total_construcciones	ln_total_construcciones	+	3.392414152

Las estimaciones del modelo SEM se muestran en la Tabla 23. Los resultados muestran que el modelo mantiene una interpretación económica en la dirección de sus coeficientes, es decir, se puede inferir que existe una relación negativa significativa en los costos de producción de la oferta inmobiliaria, medida por medio de los índices IPCO y tasa de interés activa referencial. De la misma manera se observa que para el caso de la demanda el crecimiento poblacional y el monto de crédito incentivan el número de ventas en el Distrito Metropolitano de Quito.

6.3. Usos de los modelos

A partir de los resultados obtenidos en los modelos de mínimos cuadrados en dos y tres etapas y de ecuaciones estructurales, se puede concluir que el mercado inmobiliario del Distrito Metropolitano de Quito ha experimentado ciclos de auge y caída durante los últimos diez años y se evidencia una tendencia decreciente en el nivel de ventas durante periodo del 2013 al 2022. A partir de dichos eventos, se puede inferir que el comportamiento registrado en el precio de los bienes inmuebles en la ciudad se ve influenciado por la representación de la demanda. A continuación, se detalla los principales hallazgos de las estimaciones realizadas.

- A partir de los resultados obtenidos en ambos modelos se infiere que el monto de crédito privado, enfocado en productos hipotecarios tiene una relación positiva, es decir que en el caso de un incremento del 10% en créditos hipotecarios, el número de ventas ascendería en un 2.9% en el caso del modelo de ecuaciones estructurales y en el caso de 3SLS existe la misma relación positiva con una elasticidad del 1.97%. Por lo tanto, dado el nivel de significancia estadística y su magnitud en la demanda, aumentar la dinámica de mercado hipotecario es una forma efectiva para incentivar la demanda de bienes inmuebles.
- Dentro del análisis de variables macroeconómicas no existe una relación clara desde el IPC, dado que su impacto y dirección varía dependiendo el modelo analizado. En contraste, el crecimiento poblacional influye de manera positiva en el nivel de ventas. Otras variables como el desempleo también registran una dirección adecuada e influye de manera negativa en la demanda de viviendas.

- Al analizar los costos que influyen en la oferta de vivienda, se puede inferir que los costos de producción (IPCO) influyen de manera significativa en la cantidad de construcciones. De esta forma, dada la especificación de 5% en el Impuesto al Valor Agregado realizada recientemente al sector de la construcción tendrá un impacto positivo en la oferta de viviendas.
- Por otro lado, se evidencia una relación positiva entre el volumen de crédito privado y el nivel de construcciones, es decir que ante un aumento del 10% en créditos productivos, la oferta se incrementaría en un 2% según el modelo SEM, y en el caso del modelo 3SLS se esperaría un incremento del 0.8%. De la misma manera, una posible política para dinamizar el mercado inmobiliario desde el lado de la oferta sería promover la construcción de bienes por medio de la asignación de recursos a créditos productivos.
- El modelo SEM presenta limitaciones al momento de realizar un análisis por parroquia, data su estructura metodológica, no se puede realizar una desagregación a nivel de segmentos o grupos, los resultados se basan en una estimación global de la variabilidad de la data sin contemplar la heterogeneidad entre las parroquias, por lo tanto, se sugiere contemplar o complementar los resultados con modelo que presenten la opción de un análisis con mayor desagregación, como es el caso de Mínimos Cuadrados en dos etapas estimado bajo efectos fijos o regresión multivariable con estimadores de máxima verosimilitud (MLE).Los efectos fijos en el modelo 2SLS permiten establecer una base de referencia para el promedio esperado de viviendas por parroquia. Esto posibilita la estimación de cómo variaría el número de ventas bajo diferentes escenarios, considerando las variables incluidas en el modelo. De esta manera, podemos prever la tendencia en el número de ventas en función de las variaciones en las variables explicativas

7. Conclusiones

Las siguientes conclusiones se obtuvieron del análisis metodológico y revisión de casos empíricos.

- Considerando la posición en el mercado inmobiliario que representaría el Municipio de Quito, se opta por una especificación del modelo en la cual tanto variables relacionadas con la demanda y oferta de vivienda sean contempladas, por ende, se requiere una estructura estadística que permita visualizar el efecto de los parámetros de ambas funciones en el precio de la vivienda, esta última variable como medida de análisis que genera equilibrio en el mercado inmobiliario.
- A pesar de considerar una especificación del modelo que se asemeje al contexto socioeconómico a implementar, los supuestos como la variable precio contemplada como indicador de equilibrio del mercado, limita el análisis de posibles distorsiones que existen en la compra y venta de la vivienda, elementos como procesos regulatorios o aspectos institucionales en el mercado de la vivienda no son considerados como parte del análisis del modelo.
- La literatura sobre el mercado inmobiliario muestra una gran diversidad de estudios que varían en términos del número de ecuaciones estimadas, las variables seleccionadas y el tratamiento de la dinámica del mercado. Estos estudios suelen adoptar un enfoque de equilibrio, empleando una ecuación en forma reducida basada en la intersección de oferta y demanda, o analizando los componentes en múltiples ecuaciones con relaciones causales multidireccionales centradas principalmente en el número de viviendas, precios e inversión inmobiliaria. A pesar de la variedad, hay una notable consistencia en las variables explicativas: en el lado de la demanda, se enfatizan el ingreso disponible per cápita, el stock de capital residencial per cápita, variables demográficas y la tasa de coste de uso de la vivienda; mientras que, en la oferta, los precios y los costos de construcción, particularmente los costos laborales, son recurrentes.
- Debido a la necesidad de considerar la función de demanda y oferta de vivienda en la especificación del modelo estadístico, se opta por utilizar el modelo de ecuaciones estructurales (SEM) y métodos de ecuaciones simultáneas, ya que permite relacionar múltiples variables a diferencias de otros métodos estadísticos en los cuales se limita el uso de varias variables dependiente e independientes como es el caso de mínimos cuadrados ordinarios.

- Considerando la presencia de al menos tres variables endógenas: el precio de la vivienda además de las cantidades de ofertada y demanda de vivienda, y al menos tres ecuaciones en forma de oferta y demanda y equilibrio, es importante abordar una posible correlación entre las variables endógenas y los términos de error de las ecuaciones. Por lo tanto, se plantea a la par del modelo de trayectoria el uso del estimador de variables instrumentales de mínimos cuadrados en dos y tres etapas que, además de permitir establecer relaciones causales para el análisis estructural y multiplicadores para el análisis de políticas, evita problemas estadísticos involucrados con el uso de variables explicativas endógenas y correlación de errores entre las ecuaciones a incluir.
- Una vez realizado el análisis de las tablas entregadas por el Instituto de Investigaciones de la Ciudad, si bien contienen un número importante de valores faltantes y observaciones que podría considerarse como valores atípicos, estos no representarán un problema a la hora de estimar los modelos estadísticos propuestos. No obstante, la tendencia en el precio catastral si representa un problema al momento de realizar la estimación, Esta variable, al ser endógena, tiene incidencia directa en el sistema de ecuaciones planteado y podría comprometer la magnitud con que cada variable termina influyendo en el mercado inmobiliario.
- La información contenida en el Catastro del DMQ provee variables fundamentales para el análisis y aplicación de la metodología propuesta. No obstante, conforme se detalló en la sección 2.2, esta información por si sola resulta insuficiente. El comportamiento de las variables descritas podría obedecer a varios fenómenos y únicamente la integración con otras fuentes de información del sector real y financiero de la economía permitían una lectura completa.

8. Recomendaciones

Durante la presente investigación se evidenciaron limitaciones en la calidad de la información así también posibles consideraciones respecto a la confiabilidad de la valoración de la vivienda, a continuación, se detallan la posibles mejoras o aspectos a tener en cuenta para posibles usos de la metodología implementada:

- Existe limitaciones respecto al precio o valoración de la vivienda en la base de catastro, se recomienda mejorar la metodología, por medio de modelos estadísticos, como el uso de modelos hedónicos, en donde se contemple las características de la vivienda, características del sector, ubicación geográfica, entre otros. Al mismo tiempo es importante realizar comparativos con las tendencias observadas en el mercado, de manera que los datos en el catastro guarden relación con la realidad. Esto considerando que valores erróneos en este campo condicionan e incluso invalidan los resultados obtenidos en esta investigación.
- Se evidenció que las valoraciones de la propiedad por parroquia pueden no ser confiables conforme se detalla en el Anexo 8, para determinar si existe inconsistencias, se recomienda un escrutinio de la fuente de información.
- Se desconoce si la información de la base de datos del catastro contiene una significativa representación del mercado inmobiliario en las zonas rurales del Distrito Metropolitano de Quito (DMQ), se requiere realizar un análisis sobre la dimensión de la información disponible por el Municipio de Quito, con el propósito de establecer los límites de los hallazgos de las posibles investigaciones, a su vez definir estrategias sobre recopilación de información.
- Durante la construcción de los presentes métodos estadísticos no se tuvo en consideración la distribución por edad de la población en el DMQ, el cambio en la estructura de los hogares, entre otras. Acorde a las investigaciones realizadas referente a la demanda de vivienda, un análisis sobre las características demográficas contribuye a generar políticas de manera efectiva. El análisis permitirá observar características de la vivienda en demanda y permitirá complementar los resultados de la presente investigación.

9. Referencias bibliográficas

Adams, Z., & Füss, R. (2010). Macroeconomic determinants of international housing markets. *Housing Economics*, 38-50.

Alvarez, L., & Cabrero, A. (2010). Does housing really lead the business cycle? *Banco de España Working Papers*.

Andrews, D. (2010). Real House Prices in OECD Countries: The Role of Demand Shocks and Structural and Policy Factors. *OECD Economics Department Working Papers, 831*.

Blackley, D. (1999). The long-run elasticity of new housing supply in the United States: empirical evidence for 1950 to 1994. *Journal of Real Estate Finance and Economics*, 25-42.

Brooks, C., & Tsolacos, S. (2010). *Real Estate Modelling and Forecasting.* New York: Cambridge University Press.

Davis, M., & Heathcote, J. (2003). Housing and the business cycle. *SSRN*. Obtenido de https://ssrn.com/abstract=528102

Ding-Geng Chen, Y.-F. Y. (2024). *Structural Equation Modeling Using R/SAS.* New York: CRC Press.

Ellis, L., & Berger-Thomson, L. (2004). Housing construction cycles and interest rates. *Reserve Bank of Australia Economic Group*.

Geoff, K. (1999). Modelling the demand and supply sides of the housing market: evidence from Ireland. *Economic Modelling*, 289-409.

Godman, A. (2005). The Other Side of Eight Mile: Suburban Population and Housing Supply. *Real Estate Economics*, 539–569.

Goodhart, C., & Hofmann, B. (2008). House prices, money, credit, and the macroeconomy. *Oxford Review of Economic Policy*, 180-205.

Gujarati, D., & Porter, D. (2009). *Basic econometrics.* México: McGraw-Hill.

Hilbers, P., Hoffmaister, A., Banerji, A., & Shi, H. (2008). House Price Developments in Europe: a comparison. *International Monetary Fund Working Papers*.

Ho, D., & Cuervo, J. (1999). A cointegration approach to the price dynamics of private housing—A singapore case study. *Journal of Property Investment & Finance*, 35-60.

Huang, W., & Zhang, Y. (2012). Structural change modeling of singapore private housing price in simulaitaneous equation model. *Journal of Financial Rik Management*, 7 - 14.

Iacovello, M., & Neri, S. (2008). Housing market spillovers: Evidence from an estimated DSGE model. *American Economic Journal*.

Kline, R. B. (2023). *Principles and Practive of Structural Equation Modeling.* Nueva York: The Guilford Press.

Levin, A., Lin, C., & Chu, C.-S. (2002). Unit root tests in panel data : asymptotic and finite-sample properties. *Journal of Econometrics*, 1-24.

Lopez, M. C. (2002). Modelos econométricos del mercado de la vivienda en las regiones españolas. *Working Paper Series Economic Development. nº 59*.

Madsen, J. (2012). A behavioural model of house prices. *Journal of Economic Behavior & Organization*, 21-38.

Malpezzi, S., & Vandell, K. (2002). Does the low-income housing tax credit increase the supply of housing? *Journal of Housing Economics*, 360-380.

Meen, G. (2002). An overview of sigues in housing modelling and policy. *The University of Reading paper series*.

Mikhed, V., & Zemcik , P. (2009). Do house prices reflect fundamentals? Aggregate and panel data evidence. *Housing Economics*, 140-149.

Murray, M. (1999). Subsidized and unsubsidized housing stocks 1935 to 1987: Crowding out and cointegration. *Journal of Real Estate Finance and Economics*, 107-124.

Panagiotidis, T., & Printzis, P. (2016). On the macroeconomic determinants of the housing market in Greece: a VECM approach. *International Economics and Economic Policy*, 387–409.

Paul Cheshire, S. S. (1998). Estimating the demanda for housing, land, and neighbourhood characteristics. *Oxford Bulletin of economis and statistics*, 60.

Randall E. Schumacker, R. G. (2016). *A Beginner´s Guide to Structural Equation Modeling.* Nueva York: Routledge.

Tsatsaronis, K., & Zhu, H. (2004). What drives housing price dynamics: cross-country evidence. *BIS quarterly review*.

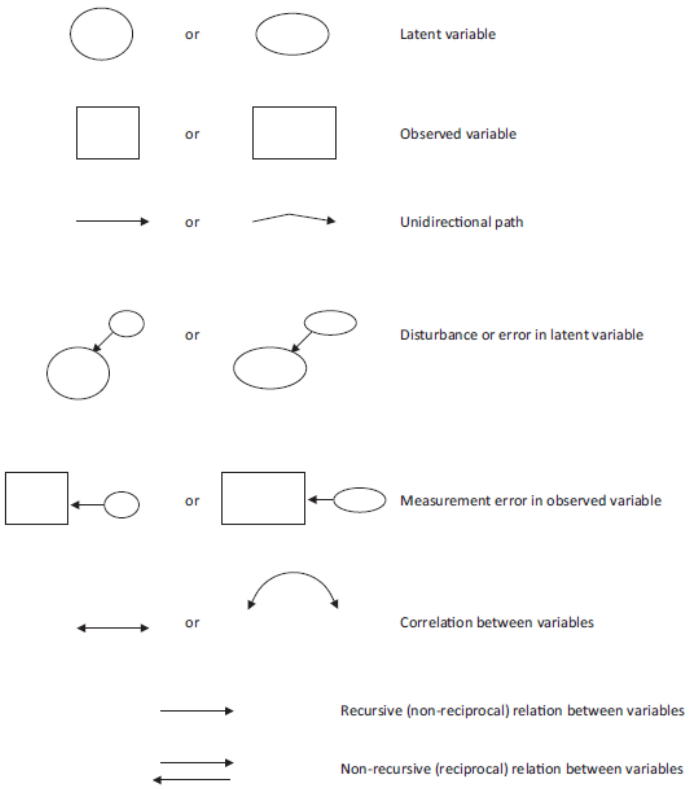
Wright, S. (1934). The method of path coefficients. *Annals of Mathematical Statistics*, 161 - 215.

10. Anexos

Anexo 1. Procesamiento de información del Catastro DMQ (archivo .ipynb)

Anexo 2. Representación de formas en modelo de trayectoria

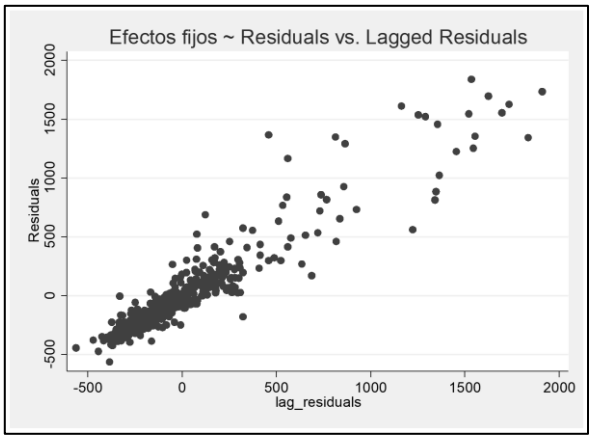
Ilustración 3: Símbolos del diagrama del modelo de trayectoria



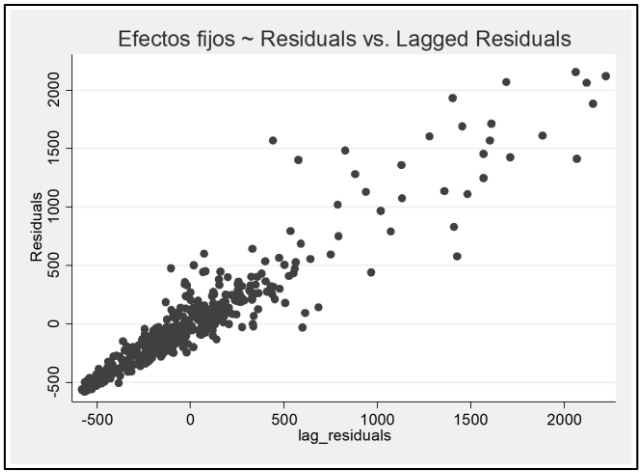
Fuente: (Randall E. Schumacker, 2016)

Anexo 3. Resultados estimadores de especificación de modelo de panel

- Revisión de heterocedasticidad – Ecuación de demanda, efectos fijos



- Revisión de heterocedasticidad – Ecuación de oferta, efectos fijos



- Revisión de heterocedasticidad en efectos fijos– Wald test

	Modified Wald test	P-value
Ecuación Demanda	32911.35	0.000
Ecuación Oferta	25355.31	0.000

- Revisión de heterocedasticidad efectos aleatorios – Comparativa de errores estándar

numero_ventas	Std. Err.	z	Robust Std. Err.	z
precio_m2	0.15	1.27	0.28	0.69
biess_monto_creditos	0.00	9.01	0.00	3.00
ipc	2.82	3.46	2.90	3.37
desempleo	968.49	-3.33	1,039.79	-3.10
crec_poblacional	0.00	-4.62	0.00	-2.80
tasa_efectiva_p	10.89	1.82	16.90	1.17
_cons	266.02	3.20	388.11	2.19

- Revisión de correlación efectos aleatorios – Ecuación de demanda

	Wooldridge test	
	F	Prob > F
Ecuación Oferta	6.423	0.0147
Ecuación Demanda	12.762	0.0008

- Test de Hausman

	Parameter	P value
Chi 2	6,97	0.1377

Anexo 4. Revisión de literatura sobre estimación de modelos del mercado inmobiliario

Artículo	Metodología	Demanda	Oferta
Structural Change Modeling of Singapore Private Housing Price in Simultaneous Equation model; Weinhong, H. & Yang, Z. (2012)	Modelo de ecuaciones estructurales, Mínimos cuadrados de 3 etapas, VECM.	• *Precios de las viviendas* Precio de las propiedades sector residencial privado (RPPI), precio de viviendas públicas (HDB). • *Ingresos* Producto Nacional Bruto (PNB) real • *Tasas de interés* Tasa Efectiva Anual Hipotecaria • *CPI* Índice de precios al consumidor	• *Stock de viviendas* Se calculó una medida del stock de viviendas utilizando datos sobre finalizaciones de vivienda (metodología de inventario perpetuo) • *Costos* Materiales básicos de construcción Costo de mano de obra.
Modelling the demand and supply sides of the housing market: evidence from Ireland; Kenny, G. (2000)	Cointegración, Modelo de Error de Corrección Vectorial (VECM)	• *Precios de las viviendas:* Promedio ponderado del precio promedio a nivel nacional de viviendas nuevas y de segunda mano deflactado por IPC • *Ingresos* Producto Nacional Bruto (PNB) real • *Tasas de interés* Una tasa representativa que se aplica a préstamos hipotecarios para sociedades de construcción. • *Stock de viviendas* Se calculó una medida del stock de viviendas utilizando datos sobre finalizaciones de vivienda (metodología de inventario perpetuo) • *Población* Una estimación base del número de hogares mediante el Censo de Población.	• *Índice compuesto de costos* El índice compuesto de costos se calcula ponderando conjuntamente el Índice de Costo de Construcción de Viviendas y un índice del precio de la tierra. Incluye otras variables de costo: mano de obra, tierra, materiales, costos de interés, etc.
Housing construction cycles and Interest rates; Berger, L. & Ellis, L (2004)	Mínimos cuadrados en 3 etapas para datos de panel.	• *Número* de hogares (dependiente): Se asume una tasa constante de viviendas vacías. • Factores macroeconómicos A corto plazo ingresos, precios de viviendas (totales) y tasas de interés. • El stock de viviendas Se trasmite a fluctuaciones en la tasa de vacantes, calculada como la diferencia entre el stock (logarítmico) de viviendas y el número (logarítmico) de hogares. • Las tasas de interés Se incluyen tasas reales como las nominales • Inflación	• Suministro total de viviendas (dependiente): Puede derivarse como el producto de calidad y cantidad, resultando en una ecuación de suministro de flujo. • Precios de viviendas • Precios del terreno • Costos de producción (salarios, tasas de interés y costos de materiales). • Stock de viviendas
Modelos econométricos del mercado de la vivienda en las regiones españolas; Lopez, C. (2002)	---	• *Variables demográficas* A largo plazo, la población es un factor reconocido por unanimidad interpretado en términos de demanda potencial. • *Variables de renta de las familias y riqueza* Se considera la renta como uno de los principales determinantes de la inversión en viviendas a largo plazo. • *Los precios* Influye en la inversión en vivienda en oferta y demanda. • *Variables financieras* Disponibilidad y el coste del crédito. Aparecen tanto si se da prioridad a las condiciones de oferta como en demanda • *Otras variables* La tasa de paro, variable que recoge el nivel de confianza de los consumidores o la incertidumbre.	

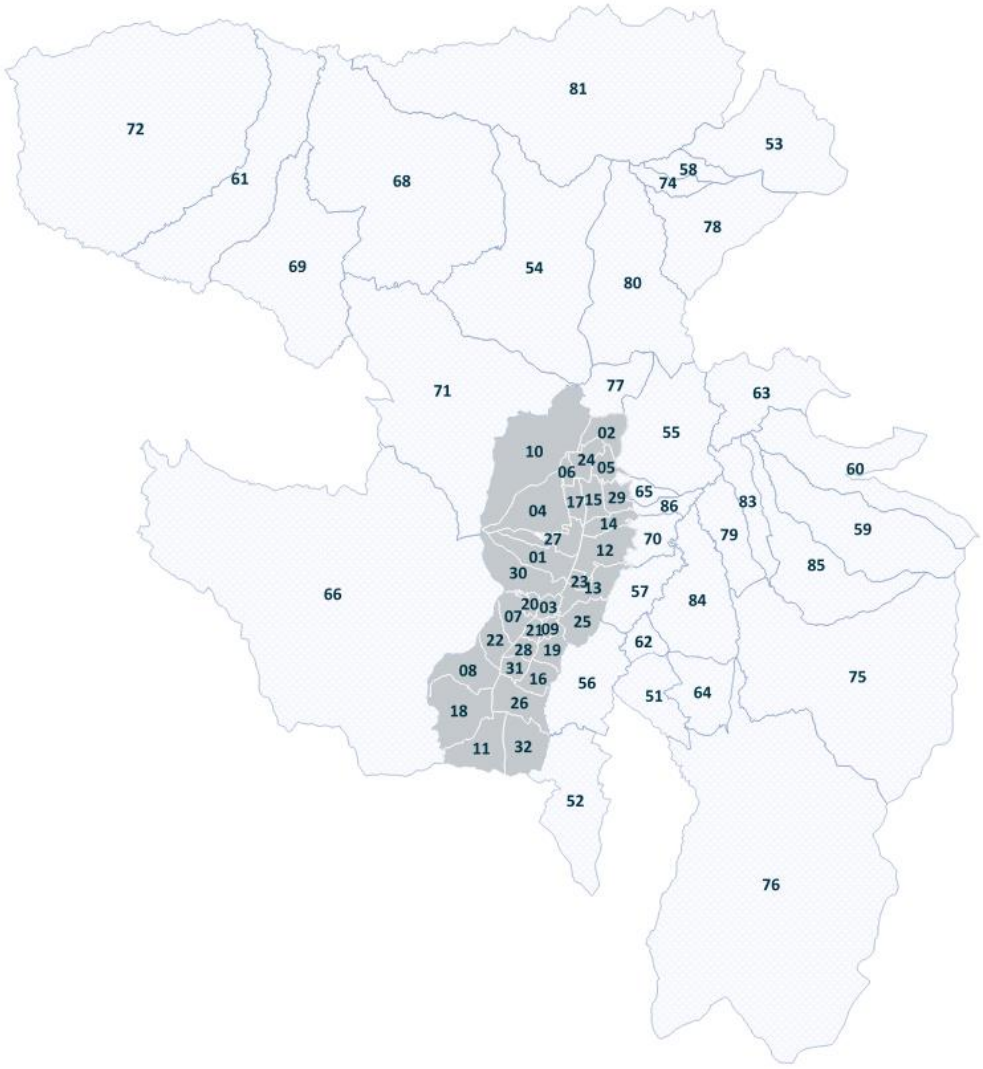
Artículo	Metodología	Demanda	Oferta
A dynamic model of demand for houses and neighborhoods; Bayer, P. & McMillan, R. (2016)	Modelo hedónico	Se proporcionan información sobre tres características a nivel de hogar: • Ingresos y riqueza relacionada con la vivienda. Riqueza como la diferencia entre el valor actual de la vivienda del hogar y el monto inicial de la hipoteca. El valor actual de la vivienda se define como el precio de venta en el año en que se compró la casa y un valor imputado de la vivienda en años posteriores.	• *Los precios* Influye en la inversión en vivienda en oferta y demanda. • *Variables financieras* Disponibilidad y el coste del crédito. Aparecen tanto si se da prioridad a las condiciones de oferta como en demanda • *Costes de construcción* En algunos casos aparece solamente uno de los componentes de dichos costes, los costes salariales. No es habitual, pero en algunos modelos interviene también el coste del suelo.
Forecasting and assessing euro area house prices through the lens of key fundamentals; Gattini, L. & Hiebert, P. (2010)	Modelo de corrección de errores (VECM)	Desde el lado de la demanda, McQuinn y OReilly (2007) proponen un modelo teórico de determinación del precio de las viviendas que está impulsado por: • Cambios en los ingresos y las tasas de interés • Tasa de crecimiento poblacional • Desde el lado de la oferta, la inversión en vivienda puede considerarse como: • Una función de los fondos internos de los hogares (utilizando el ingreso actual como una aproximación para el ingreso permanente) • Costo de financiamiento externo (la tasa de interés) Los precios de las viviendas como una medida del valor del colateral, en el espíritu de un marco acelerador financiero modificado.	
Housing market spillovers: Evidence from an estimated DSGE model, Iacovello, M., & Neri, S. (2008)	Modelo de equilibrio general dinámico estocástico con enfoque de verosimilitud bayesiana	Se considera como variable dependiente al precio de la vivienda y a las inversiones residenciales. Entre las variables incluidas están: • Tasa de interés real anual • Participación del consumo • Participación de la inversión empresarial • Participación de la inversión en vivienda • Riqueza inmobiliaria • Capital empresarial en el sector no inmobiliario • Capital empresarial en el sector inmobiliario • Valor del terreno Se destaca el lento progreso tecnológico del sector de la construcción como determinante del alza en el precio de las viviendas.	
A behavioural model of house prices, Madsen, J. (2012)	Modelo de equilibrio de amortización de precios de vivienda	Se consideran como determinantes del precio de las viviendas: • El número neto de posibles propietarios de viviendas • Suma del ingreso disponible de los potenciales compradores de casas. • Stock de viviendas, igual al número de compradores de casas multiplicado por el tamaño promedio de la casa que están considerando comprar. • Disposición de los bancos a prestar, lo cual a su vez depende de las políticas monetarias, las regulaciones bancarias, los precios pasados de las casas y las perspectivas económicas. • Regulaciones bancarias • Perspectivas económicas • Costos de adquisición de viviendas • Monto de alquileres vigentes Los precios de las casas son independientes de los alquileres y a largo plazo, los precios de las casas están impulsados por los costos de adquisición.	
Does the low-income housing tax credit increase the supply of housing?, Malpezzi, S., & Vandell, K. (2002)	Mínimos Cuadrados Ordinarios con variables instrumentales	Se incluyen como determinantes de la demanda (número de unidades de vivienda de bajos ingresos construidas): • Vector de medidas de ingresos y pobreza, • Variables socio-demográficas, • Oferta de subsidios del lado de la demanda, como vales de vivienda • Población.	Se incluyen como determinantes de la oferta (número de viviendas privadas y programas de vivienda pública): • Población • Costos de construcción o adquisición y de la tierra incluido impuestos • Una medida del entorno regulatorio, y • Condiciones del mercado de vivienda (por ejemplo, tasas de vacancia). • Relación entre los precios de vivienda y los costos efectivos de adquisición.
The Other Side of Eight Mile: Suburban Population and Housing Supply, Godman, A. (2005).	Mínimos Cuadrados Indirectos y Mínimos Cuadrados de tres etapas	Variable dependiente: Demanda de mercado de unidades de vivienda en una ubicación particular Variables independientes: • Precio de alquiler de los servicios de vivienda • El ingreso per cápita • Población metropolitana	Variable dependiente: Oferta de mercado de unidades de vivienda Variables independientes: • El valor del stock de vivienda • Factores regionales que incluyen: costos de factores, clima o grado de sindicalización del mercado laboral,

Artículo	Metodología	Demanda	Oferta
		El equilibrio a largo plazo relaciona los alquileres de mercado para los flujos de servicios de vivienda con los valores de mercado para los stocks de vivienda por el costo de uso, que incluye los efectos de intereses perdidos, depreciación de activos, impuestos sobre la propiedad y ganancias de capital esperadas.	que usualmente se caracterizarían con variables binarias de ciudad, estado y/o región.
Do house prices reflect fundamentals? Aggregate and panel data evidence, Mikhed, V. & Zemcik , P. (2009)	Modelo de valor presente y cointegración en datos de panel	Se utiliza como variable dependiente el precio de las viviendas y como determinante de este: • Alquiler real de la vivienda • Tasa hipotecaria • Ingreso personal • Costo de construcción • Riqueza del mercado de acciones • Población.	
Subsidized and unsubsidized housing stocks 1935 to 1987: Crowding out and cointegration, Murray, M. (1999)	Cointegración, Modelo de Error de Corrección Vectorial (VECM)	Variable dependiente: Stocks de vivienda no subvencionada per cápita Variables independientes: • Vivienda subvencionada acumulada per cápita • Vivienda pública acumulada per cápita • Vivienda subvencionada financiada privadamente acumulada per cápita • Ingreso disponible real per cápita • Tasa de interés real (tasa de bonos menos inflación) • Población urbana, en porcentaje • Población mayor de 21 años, en porcentaje	
On the macroeconomic determinants of the housing market in Greece: a VECM approach, Panagiotidis, T., & Printzis, P. (2016)	VECM, Mínimos cuadrados de dos etapas, Cointegracion	Variable dependiente: Número de viviendas Variables independientes: • Índice de Precios de la Vivienda, • Índice de Precios al Consumidor, • Índice de Producción Industrial • Volumen de Comercio Minorista (RETAIL), • Tasa de interés de préstamos, • Tasa de crecimiento anual de las hipotecas • Tasa de crecimiento de la oferta monetaria M1 • Tasa de Desempleo	
What drives housing price dynamics: cross-country evidence, Tsatsaronis, K., & Zhu, H. (2004)	Vector Autoregresivo (VAR)	Variable dependiente: Precio de la vivienda, Variables independientes: • Crecimiento en el ingreso disponible de los hogares, • Cambios graduales en la demografía (como el tamaño relativo de las generaciones mayores y menores), • Características permanentes del sistema fiscal que podrían incentivar la propiedad de viviendas en oposición a otras formas de acumulación de riqueza • Nivel promedio de tasas de interés (posiblemente relacionado con el comportamiento a largo plazo de la inflación). • La disponibilidad y el costo del terreno • Costo de construcción e inversiones en la mejora de la calidad del stock existente de viviendas.	
Macroeconomic determinants of international housing markets. Zeno Adams, Roland Fuss	Panel de cointegración	Variable dependiente: Precio de la vivienda Variables independientes: • Actividad económica, la investigación mide la actividad económica por medio las variables empleo y tasa de crecimiento de producción industrial. • Tasa de interés a largo plazo: Se determina que tiene un efecto negativo al aumentar los costos de financiamiento de viviendas. • Costos de construcción: Incrementos en los costos de producción, como materiales o contratación de mano de obra	

Artículo	Metodología	Demanda	Oferta
Does housing really lead the Business Cycle. Luis J. Álvarez, Alberto Cabrero	Filtros Butterworht, Epanichnekov	La investigación busca determinar la relación que existe entre el mercado inmobiliario y los movimientos del PIB, la serie de tiempo va desde el Q1 de 1980 hasta Q4 2008, las variables utilizadas son las siguientes: • Consumo privado • Consumo público • Inversión en equipamiento • Inversión en residencias • Inversiones no residenciales • Exportaciones • Importaciones.	
Real House Prices in OECD Countries: The role of demand shocks and structural and policy factors. Dan Andrews	Panel de efectos fijos	La investigación determina los shocks que posiblemente afectan al precio del mercado inmobiliario, entre los que se encuentran, respuesta de la oferta ante variaciones en el precio, alivios impositivos ante deudas hipotecarias, costos de transacción, desregulación del mercado hipotecario, condiciones del mercado laboral y cambios demográficos. A continuación, se detallan las variables independientes: • Ingreso disponible para vivienda • Tasa de interés a largo plazo • Costos de construcción • Stock per cápita • Costo de arrendamiento • Tasa de inflación • Tasa de migración • Tasa de crecimiento poblacional.	
The long-run elasticity of new housing supply in the United States: Empirical Evidence for 1950 to 1994	Mínimos cuadrados en dos etapas	La variable dependiente es el precio real de la vivienda. Las variables independientes se detallan a continuación: • Tasa de interés real a largo plazo de los bonos del tesoro • Tasa de consumo personal • Tasa de crecimiento poblacional de adultos en Estados Unidos.	La variable dependiente es el precio real de la vivienda. Las variables independientes se detallan a continuación: • Precio real de los materiales de construcción • Promedio de hora de trabajo • Salario de trabajadores en el sector de la construcción.
Housing and the Business Cycle, Davis, M., & Heathcote, J	Modelo de crecimiento stocástico multisector	La investigación realiza un proceso estocástico definido por medio de un conjunto de precios de los posibles estados en momento en el tiempo t, el objetivo es evidenciar el punto de equilibrio entre consumo e inversión en bienes, la variable dependiente es el precio de vivienda, las variables independientes se definen a continuación: • Arrendamiento inmobiliario • Tasa de crecimiento de hipotecas • Consumo privado • Producto interno bruto • Indicador de inversión en residencias (RESI)	
Housing construction cycles and interest rates, Laura Berger Thomson and Luci Ellis.	Mínimos cuadrados en tres etapas	El modelo se basó en dos ecuaciones una para demanda y otra para la oferta: Variable independiente: Cantidad de casas demandada Variables independientes: • Índice de precios al consumidos • Tasa de interés • Ingresos anuales • Precio de la vivienda	El modelo se basó en dos ecuaciones una para demanda y otra para la oferta: Variable independiente: Oferta de vivienda Variables independientes: • Tasa de construcción • Costos de construcción • Precio por estructura.

Artículo	Metodología	Demanda	Oferta
Modelling the demand and supply sides of the housing market: evidence from Ireland, Geoff Kenny.	Modelo Vector de corrección de error.	El modelo se basa en observar la existencia de cointegración entre la variable de demanda y oferta, las series temporales utilizadas en la parte de la demanda son: • Ingresos anuales • Tasa de interés en el largo plazo	El modelo se basa en observar la existencia de cointegración entre la variable de demanda y oferta, las series temporales utilizadas en la parte de la oferta son: • Costos promedios de la industria de construcción. • Tasas de interés en el largo plazo.
House prices, money, credit and the macroeconomy	Modelo de datos de Panel con efectos fijos	Los datos de panel tienen una periodicidad trimestral desde 1973 Q1 hasta 2006 Q4, las variables utilizadas para el modelo de panel con efectos fijos son: • Precio de la vivienda. • Dinero en circulación M1 • Volumen de crédito • PIB • Índice de precios al consumidor	

Anexo 6. Distribución de parroquias en el DMQ utilizada en el análisis



Belisario quevedo	1	La concepcion	17	Alangasi	51	Nanegal	68
Carcelen	2	La ecuatoriana	18	Amaguaã‘a	52	Nanegalito	69
Centro historico	3	La ferroviaria	19	Atahualpa	53	Nayon	70
Cochapamba	4	La libertad	20	Calacali	54	Nono	71
Comite del pueblo	5	La magdalena	21	Calderon	55	Pacto	72
Cotocollao	6	La mena	22	Conocoto	56	Perucho	74
Chilibulo	7	Mariscal sucre	23	Cumbaya	57	Pifo	75
Chillogallo	8	Ponceano	24	Chavezpamba	58	Pintag	76
Chimbacalle	9	Puengasi	25	Checa	59	Pomasqui	77
El condado	10	Quitumbe	26	El quinche	60	Puellaro	78
Guamani	11	Rumipamba	27	Gualea	61	Puembo	79
Iã‘aquito	12	San bartolo	28	Guangopolo	62	San antonio	80
Itchimbia	13	San isidro del inca	29	Guayllabamba	63	San jose de minas	81
Jipijapa	14	San juan	30	La merced	64	Tababela	83
Kennedy	15	Solanda	31	Llano chico	65	Tumbaco	84
La argelia	16	Turubamba	32	Lloa	66	Yaruqui	85
						Zambiza	86

Anexo 7. Efectos fijos obtenidos en el modelo 2SLS

Constante	-112.4967						
Cód. parroquia	Efecto fijo	Cód. parroquia	Efecto fijo	Cód. parroquia	Efecto fijo	Cód. parroquia	Efecto fijo
2	-0.074896	15	0.4103059	27	0.5980471	63	-1.509234
4	0.0665432	16	-0.2268284	29	0.1045493	64	-2.030142
5	-0.4884917	17	-0.1268183	30	-0.4149844	65	-1.254678
6	-0.7521374	18	-0.3235254	31	-0.0071072	70	-0.5829462
7	-0.4438024	19	-0.2925653	32	0.7872978	75	-1.832546
8	-0.2326755	20	-1.681988	51	-0.3867012	77	-0.154855
9	-0.7340595	21	-0.6460409	52	-0.4294904	80	0.0940109
10	0.3560065	22	-0.7319603	55	1.441818	84	0.454625
11	0.2670883	23	-0.0598334	56	1.197745	85	-1.742252
12	1.439337	24	0.3655052	57	0.536469		
13	-0.5413631	25	-0.3055268	59	-2.934929		
14	0.2846782	26	0.4631447	60	-2.452389		

Anexo 8. Revisión de precios comercial y catastral 2018 y 2022

Parroquia	Precio catastral		Variación	Precio comercial		Variación
	2018	**2022**		**2018**	**2022**	
Iñaquito	870.48	743.56	14.60%	1,607	1,442	10.30%
Jipijapa	745.78	655.69	-12.10%	1,458	1,252	-14.10%
Rumipamba	691.28	609.94	-11.80%	1,102	922	-16.30%
Cumbayá	247.55	219.87	-11.90%	1,726	1,689	-0.02%