CAPÍTULO 3: METODOLOGÍA

Como ya se vio, existen varios tipos de pronósticos. Para el propósito de esta tesis se van a considerar aquellos tipos que pertenecen a la categoría de métodos cuantitativos, pero dentro de éstos se excluirán los Métodos de Monitoreo, que son una combinación de métodos cuantitativos y cualitativos. Los métodos cualitativos, los métodos tecnológicos y los Métodos de Monitoreo requieren de conocimientos que están fuera del alcance del estudiante. A continuación se describen aquellos métodos que pertenecen a la categoría de los cuantitativos, citándose una breve definición y características principales. Después, se hace una selección de los métodos a emplear para estimar la demanda de vivienda.

3.1 Series de tiempo

Se define como serie de tiempo los valores que toma una variable en periodos sucesivos de tiempo. En su observación, se pueden distinguir sus cuatro componentes: tendencia, ciclo, estacionalidad y componente aleatorio, que pueden presentarse individualmente o en conjunto. Las series identifican un patrón o comportamiento en los datos pasados y suponen que se repetirán en el futuro, por lo que no sirven para hacer pronósticos de una variable que será influenciada por eventos anteriormente no contemplados. Sirven para hacer pronósticos a inmediato, corto y mediano plazo. Las series de tiempo pueden ser univariadas y multivariadas.

3.1.1 Series de tiempo univariadas

Las univariadas son aquellas series en las que sólo interviene una variable independiente. Sirven para hacer pronósticos a inmediato y corto plazos, pues son periodos en las que la variable a modelar puede ser afectada drásticamente por factores externos con baja probabilidad. Los métodos que pertenecen a las series univariadas son Ingenuo, Descomposición, Suavizamiento, Promedio Móvil Autorregresivo y Bayesiano.

Método Ingenuo

El método ingenuo es el método más sencillo de pronóstico. Su importancia radica en que es el método de referencia para saber la efectividad de los demás métodos de pronóstico. Se divide en Pronóstico Ingenuo 1 y Pronóstico Ingenuo 2. El primero supone que el pronóstico para un periodo futuro es el valor observado de cierto periodo anterior. La ecuación para obtener el pronóstico es

$$F_{t+i} = X_t$$

donde

t = periodo actual

i = número de periodos adelantados pronosticados

F_{t+i} = pronóstico del periodo t + i

X_t = último valor real en el periodo t

Este método es conocido en estadística como caminata aleatoria. El segundo toma en cuenta el factor estacional de las series, a diferencia del primero, por lo que suele ser más exacto. En su aplicación, primero se remueve este factor. En lo subsecuente, la aplicación es igual que el Ingenuo 1.

Método de Descomposición

El método de descomposición separa los componentes de las series de tiempo para identificar un patrón en ellas. La descomposición supone que los datos están estructurados de la siguiente forma:

$$\text{datos} = \text{patrón} + \text{error}$$

en donde el patrón es una combinación de los componentes ciclo, tendencia y estacionalidad. Los datos son descritos por el patrón hallado y un error, el cual es la diferencia entre los valores reales y el patrón. El procedimiento para descomponer una serie es empírico, pero consiste en remover, paso a paso, la estacionalidad, luego la tendencia y finalmente el ciclo. Hay muchos enfoques para separar cada componente de la serie de los otros lo más exactamente posible. Desde el punto de vista estadístico este método posee muchas limitaciones teóricas. La forma general de representar este método de forma matemática es

$$X_t = f(S_t, T_t, C_t, R_t)$$

en donde

X_t = es el valor real de la serie de tiempo en el periodo t

S_t = componente estacional en el periodo t

T_t = componente tendencial en el periodo t

C_t = componente cíclico en el periodo t

R_t = componente aleatorio o error en el periodo t

Existen muchas formas para la fórmula de la relación de los cuatro componentes, pero las más usuales son la multiplicativa ($X_t = S_t \, x \, T_t \, x \, C_t \, x \, R_t$) y la aditiva ($X_t = S_t + T_t + C_t + R_t$).

$\mu = E[x]$

Métodos de Suavizamiento

Estos métodos tienen como propósito obtener valores ponderados para una serie y después extrapolarlos para obtener un pronóstico. Existen dos tipos de suavizamiento: Promedios Móviles Simples y Suavizamiento Exponencial. Los Promedios Móviles Simples utilizan un número de valores que son promediados, pero estos promedios se van moviendo con cada observación nueva. Este método da igual ponderación a todas las observaciones contempladas y ninguna a las anteriores. La ecuación para obtener el pronóstico es

$$F_{t+1} = S_t = (X_t + X_{t-1} + \ldots + X_{t-N+1}) / N$$

en donde

F_{t+1}= pronóstico para el tiempo t + 1

S_t = valor suavizado en el tiempo t

X_i = valor actual en el tiempo i

i = periodo de tiempo

N = número de observaciones en el promedio

Los métodos de Suavización Exponencial se caracterizan por ponderar con mayor peso a las observaciones más recientes y con menor a las más antiguas. Dentro de estos métodos existen varios tipos: Único, Doble Lineal, Lineal y Estacional y Amortiguado de Tendencia. El *Suavizamiento Exponencial Único* es la forma general de los suavizamientos exponenciales. Es útil sólo para aquellas series que tienen un patrón horizontal, por lo que ignora las tendencias, ciclos y estacionalidades. Se basa en un valor observado y un pronóstico, ambos del periodo anterior, para hacer el pronóstico del periodo siguiente. La ecuación para obtener el pronóstico es

$$F_{t+1} = \alpha X_t + (1- \alpha)F_t$$

en donde

F_{t+1} = pronóstico para el periodo t+1

α = primer factor de ponderación, $0 < \alpha < 1$

X_t = valor real en el tiempo t

$1\text{-}a$ =segundo factor de ponderación, $0 < 1\text{-}\alpha < 1$

F_t = pronóstico para el periodo t

La *Doble Suavización Exponencial Lineal o de Holt* contempla las tendencias en las series, sean a la baja o a la alta, pero no los ciclos ni las estacionalidades. La ecuación para obtener el pronóstico es

$$F_{t+m} = S_t + T_t\mathrm{m}$$

donde

F_{t+m} = pronóstico para el periodo t + m

m = número de periodos adelantados al periodo t

El término *T* es la estimación suavizada de la tendencia y su fórmula es

$$T_t = \beta(S_t - S_{t-1}) + (1- \beta)T_{t-1}$$

en donde

T_t = tendencia suavizada en la serie de datos

S_t = equivalente del valor suavizado exponencial único

β = factor de ponderación análogo a α, $0 < \beta < 1$

Finalmente, el término S es el valor suavizado de la serie desestacionalizada y su fórmula es

$$S_t = \alpha X_t + (1-\alpha)(S_{t-1} + T_{t-1})$$

El *Suavizamiento Exponencial Lineal y Estacional o de Winters* contempla la tendencia y la estacionalidad de las series, pero no el ciclo. La ecuación para obtener el pronóstico es

$$F_{t+m} = (S_t + T_t m) I_{t-L+m}$$

y las formulas de los términos que la conforman son

$$S_t = \alpha(X_t / I_{t-L}) + (1-\alpha)(S_{t-1} + T_{t-1})$$

$$T_t = \beta(S_t - S_{t-1}) + (1-\beta)T_{t-1}$$

$$I_t = \gamma(X_t / S_t) + (1-\gamma) I_{t-L}$$

en donde

T = valor suavizado de la tendencia

S = valor suavizado de la serie desestacionalizada

I = valor suavizado del factor estacional

L = duración de la estacionalidad

El *Método de Brown* es un caso especial del Doble Suavizamiento Exponencial Lineal, tiene una segunda atenuación del nivel atenuado simple. La ecuación del modelo es

$$F_{t+m} = a_t + mb_t$$

donde m es el número de periodos del horizonte del pronóstico y a_t y b_t se calculan como sigue:

$$a_t = 2S_t – S^{(2)}_t$$

$$b_t =(\alpha/(1-\alpha))(S_t – S^{(2)}_t)$$

Aquí, S_t se calcula mediante suavización exponencial simple y $S^{(2)}_t$ es la suavización exponencial simple de S_t.

El *Suavizamiento Exponencial Amortiguado de Tendencia* disminuye la tendencia lineal que se extrapola conforme avanzan los periodos de tiempo. Utiliza los mismos parámetros que la ecuación de Holt excepto por el término de amortiguación. La ecuación para obtener el pronóstico es

$$F_{t+m} = S_t + \Sigma \phi_i T_t$$

Donde las ecuaciones de *S* y *T* tienen las mismas interpretaciones que el modelo anterior y sus fórmulas son

$$S_t = \alpha X_t + (1- \alpha)(S_{t-1} + T_{t-1})\phi$$

$$T_t = \beta(S_t - S_{t-1}) + (1- \beta)T_{t-1}\phi$$

Como nota final cabe mencionar que hay muchas clases de suavizamiento, pero se citaron los que se consideran más importantes (Makridakis y Weelwright., 2000).

Promedio Móvil Autorregresivo (ARMA)

Esta clase de métodos fueron propuestos desde hace mucho tiempo, pero se alcanzó un auge y desarrollo recientemente en su aplicación debido al avance de las computadoras. De acuerdo a la teoría de estos métodos, existen tres clases de series de tiempo que son descritos por una ecuación distinta y son los modelos autorregresivos (AR), los modelos de promedio móvil (MA) y los de promedio móvil autorregresivos mixtos (ARMA)

Los modelos autorregresivos son aquellos cuya variable dependiente depende de valores anteriores de la misma variable y su estructura es parecida a los modelos de regresión. La ecuación general de estos modelos es la siguiente:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} + e_t$$

donde e_t es el error o término residual que no puede ser explicado por el modelo y hay p parámetros a determinar.

Los modelos de promedio móvil son aquellos cuya variable dependiente depende de valores anteriores de su error. La ecuación general de estos modelos es la siguiente:

$$Y_t = e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \ldots - \theta q e_{t-q}$$

donde e_t es el error o término residual que no puede ser explicado por el modelo y hay q parámetros a determinar.

Los modelos mixtos son una combinación lineal de los dos anteriores y en estos la variable dependiente depende tanto de valores anteriores de la misma variable como de valores anteriores de su error. . La ecuación general de estos modelos es la siguiente:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \ldots - \theta q e_{t-q} \qquad (3.1)$$

donde e_t es el error o término residual que no puede ser explicado por el modelo, hay p y q parámetros a determinar y no necesariamente p = q.

Existen distintos enfoques que aplican esta clase de métodos, entre ellos están la metodología Box-Jenkins, modelos ARARMA de Parzen, el Filtrado Adaptativo, el Filtrado del Procedimiento de Estimación Adaptativo (AEP), el Filtrado de Kalman y el FORSYS de Lewandowsky.

La metodología *Box-Jenkins* necesita una estacionariedad en los datos para poder hacer un pronóstico. Consiste en 3 etapas: la primera, identificar el modelo tentativo más apropiado de predicción; la segunda, ajustar el modelo los datos históricos, estimar parámetros y determinar si es adecuado; la tercera, obtener un pronóstico para un futuro periodo. Su ventaja es que proporciona pronósticos muy acertados a inmediato y corto plazo y su desventaja es que es un método difícil de entender.

La metodología ARARMA necesita una estacionariedad en los datos para poder hacer un pronóstico. Clasifica las series de tiempo en 3 categorías: las de memoria larga (tienen tendencia, estacionalidad y otras características no estacionarias), las de memoria corta (estacionarias) y las que tienen perturbaciones “blancas” o aleatoriedad. Consta de tres etapas: la primera, identificación de la serie de acuerdo a criterios específicos; la segunda, transformación de la serie por los modelos adecuados que pueden ser AR primordialmente;

la tercera, selección y aplicación de un modelo para pronosticar que puede ser AR, MA o ARMA, que se logra con distintos criterios establecidos.

El Filtrado del Procedimiento de Estimación Adaptativo (AEP) no necesita estacionariedad. El usuario especifica un modelo autorregresivo y una tendencia que puede ser de cualquier tipo. La ecuación de este modelo es la siguiente:

$$X_t = [a_{0t} + a_{1t}t + \dots + a_{qt}t^q + \phi_{1t}X_{t-1} + \phi_{2t}X_{t-2} + \dots + \phi_{pt}X_{t-p}] \times [I_{1t} \times I_{2t} \times \dots \times I_{Lt}] + e_t$$

El modelo se divide en 3 partes: los términos con a especifican la tendencia, los términos con ϕ son los autorregresivos y los términos con I especifican la estacionalidad.

El Filtrado de Kalman es conocido también como Predicción Bayesiana. Es un modelo de estimación adaptativo que se actualiza de acuerdo a las probabilidades que proporciona el usuario, lo cual implica una dificultad y una desventaja. El modelo básico es el siguiente:

$$Z_t = \mu_t + S_{i,t} + e_t$$

$$\mu_t = \mu_{t-1} + \beta_t + \delta\mu_t$$

$$\beta_t = \beta_{t-1} + \delta\beta_t$$

$$S_{i,t} = S_{i,t-1} + \delta S_{i,t}$$

en donde e_t, $\delta\mu_t$ *y* $\delta\beta_t$ están normalmente distribuidas con medias 0 y varianzas v_e, v_μ *y* v_β respectivamente. Además, $Z_t = \log X_t$ y μ_t, β_t y S_t son los logaritmos del nivel, la tendencia y la estacionalidad de la serie.

El método FORSYS (Forcasting System) es empleado en Europa en empresas grandes. Utiliza la misma ecuación que la del Suavizamiento Exponencial de Winters, excepto que los parámetros α, β y γ no son fijos, sino que se modifican con un procedimiento llamado OPS. Tiene capacidad de tomar en cuenta sucesos importantes para las predicciones y es útil para pronósticos a corto plazo.

3.1.2 Series de tiempo multivariadas

Las multivariadas son aquellas series en las que intervienen dos o más variables independientes. Por lo anterior, la forma en que son analizadas guarda similitud con el método de regresión múltiple. Estas series nos permiten hacer pronósticos de mayor fiabilidad a mediano y largo plazos, pues consideran factores que podrían afectar significativamente a la variable dependiente en lapsos de tiempo medianos y largos, donde hay mayor probabilidad de que éstos ocurran. Los métodos que pertenecen a las series multivariadas son Modelos de Función de Transferencia, Modelos VAR y Filtrado de Kalman Multivariable. La ecuación general de los modelos multivariables es

$$Y_t = \delta_1 Y_{t-1} - \delta_2 Y_{t-2} - \ldots - \delta_r Y_{t-r}$$

$$+ \omega_0 X_{t-b} - \omega_1 X_{t-b-1} - \ldots - \omega_s X_{t-b-s}$$

$$+ \zeta_0 Z_{t-c} - \zeta_1 Z_{t-c-1} - \ldots - \zeta_s Z_{t-c-v}$$

$$+ \ldots +$$

$$+ \xi_0 W_{t-c} - \xi_1 W_{t-c-1} - \ldots - \xi_s W_{t-c-u} + e_t$$

donde Y, X, Z y W son variables independientes y los parámetros r, s, v y u determinan el número de variables necesarias para una ecuación específica y deben determinarse.

Modelos de Función de Transferencia

Estos métodos presentan una dificultad media para su aplicación en la vida real y generalmente se aplican en ingeniería. Consta de tres pasos. El primero es convertir las variables dependiente e independientes en aleatorias usando un modelo ARMA univariable en cada variable. El segundo, es identificar un modelo que relacione las variables dependiente e independientes utilizando los residuos de cada serie, una vez que se han eliminado las correlaciones entre series debida a los componentes de las mismas. El tercero es modelar los errores sistemáticos que quedan mediante un modelo ARMA univariable.

Modelos VAR

Estos modelos son los Vectoriales Autorregresivos y tienen una dificultad baja para su aplicación en la vida real, por lo que suelen emplearse en las áreas económicas y de negocios. Requieren una fuerte preparación teórica de cómo funciona la variable dependiente y cómo se ve afectada por variables independientes, pues con base en ello se identifican las variables, se estiman los coeficientes de las mismas y se identifican los rezagos. La estimación de los modelos VAR se puede basar en cualquiera de los siguientes métodos: Enfoque Bayesiano de Litterman, Método de Parzen, Modelos VAR con Coeficiente Aleatorio de Sims y el Error de Predicción Final de Hsiao.

Filtrado de Kalman Multivariable

Estos son los modelos generales de todos los modelos multivariables y son los que requieren de menos cálculo, debido a su actualización recursiva de los parámetros conforme surgen nuevos datos. Este método es muy difícil de aplicar, debido a que requiere determinar adecuadamente la estructura del modelo y además, información como varianza y covarianza de los errores no están disponibles de forma inmediata en las empresas Por lo anterior, no existen o hay muy pocos ejemplos de aplicación de los mismos para el área de los negocios.

3.2 Métodos causales o explicativos

Los métodos causales tratan de encontrar un modelo que describa una relación entre variables: la variable dependiente y las variables independientes. De esta forma, se puede hacer un análisis de la situación y por lo tanto, se pueden hacer análisis de cómo impactará un cambio a la variable dependiente. Los métodos que pertenecen a esta categoría son tres: Regresión Lineal Simple, Regresión Lineal Múltiple y la Econometría. Cabe destacar que la Regresión Simple es un caso de la Regresión Múltiple y ésta, es un caso especial de los Métodos Econométricos.

Regresión Lineal Simple

El supuesto inicial de la Regresión Lineal Simple es que existe una relación lineal entre la variable dependiente, en este caso Y, y la variable independiente, en este caso X. Así, la relación se puede expresar como

$$\hat{Y}_i = \beta_0 + \beta_1 X_i \text{, i = 1, 2, … , n}$$

donde

$\hat{Y}_i$ = Valor estimado de la variable dependiente Y en el i-ésimo modelo

X_i = Valor de la variable independiente en el i-ésimo modelo

y β_0 y β_1 son los parámetros a estimar en esta ecuación de regresión, usando el método de los mínimos cuadrados. Si la ecuación que explica la relación entre variables es no lineal, ésta se puede transformar para hacerla lineal. Existen varios supuestos que deben comprobarse antes de afirmar que una ecuación es adecuada para describir la variable dependiente real. La Regresión Lineal Simple tiene la ventaja de poder hacer pronósticos tomando en cuenta el comportamiento de una variable independiente que impacta a la dependiente. Se pueden obtener intervalos de confianza de la variable estimada. Es un método relativamente sencillo que no requiere de cálculos sofisticados.

Regresión Lineal Múltiple

La Regresión Lineal Múltiple es la generalización de la Simple y, a diferencia de ésta, toma en cuenta el comportamiento de distintas variables independientes para determinar su efecto en la variable dependiente. La ecuación general de este método es

$$\hat{Y}_i = \beta_0+\beta_1 X_{1i}+\beta_2 X_{2i}+\ldots+\beta_k X_{ki}\,, \quad i = 1, 2, \ldots , n; j = 0, 1, \ldots, k \qquad (3.2)$$

donde

$\hat{Y}_i$ = Valor estimado de la variable dependiente Y en el i-ésimo modelo

X_{ki} = Valor de la k-ésima variable independiente en el i-ésimo modelo

y donde los β_j son los parámetros a estimar en las ecuaciones de regresión, usando el método de los mínimos cuadrados. Si la ecuación que explica la relación entre variables es no lineal, ésta se puede transformar para hacerla lineal. Existen varios supuestos que deben comprobarse antes de afirmar que una ecuación es adecuada para describir la variable dependiente real. La Regresión Lineal Múltiple tiene la ventaja de poder hacer pronósticos más acertados que la Simple, pues toma en cuenta el comportamiento de varias variables independientes que impactan a la dependiente. Se pueden obtener intervalos de confianza de la variable estimada. Sin embargo, es un método laborioso que requiere de cálculos sofisticados, por lo que no se aconseja su uso a menos de que se disponga de una computadora.

Econometría

Los Métodos Econométricos son sistemas de ecuaciones lineales de regresión múltiple. Estas ecuaciones tienen muchas variables interdependientes, por lo que si unas tienen a ciertas variables independientes cuando otras tienen a éstas como las dependientes. Esto representa la afirmación básica de estos modelos, la cual es que todo depende de todo en el mundo real. Teóricamente se pueden desarrollar una gran cantidad de ecuaciones con variables interdependientes para un modelo, pero las limitantes de los datos, los sistemas de cómputo y los problemas de estimación restringen el número de estas ecuaciones. Pasos similares para realizar los cálculos se necesitan tanto en la Econometría como en la Regresión Múltiple, sólo que en ésta requieren de mayor refinamiento. La aplicación de estos modelos es en el área económica y empresarial, pero esta aplicación es sumamente complicada debido a la complejidad de los cálculos, el dominio experto de temas económicos, los altos costos asociados y los modelos patentados. Por esta razón es que la mayoría de las empresas no pueden desarrollar estos cálculos, sólo las grandes empresas privadas y las empresas gubernamentales más sólidas pueden darse este lujo.

3.3 Tabla de los métodos cuantitativos más usuales

A continuación, se resume en la siguiente tabla los métodos cuantitativos más usuales empleados en la vida diaria. Condensa datos mencionados en la sección anterior y aporta algunos nuevos. Se excluyeron aquellos de los cuales no se cuenta con la suficiente información como para hacer una comparación.

Tabla 3.1. Comparación de los Métodos Cuantitativos más usuales.

Método	Dificultad de implementación	Horizonte de tiempo	Caraterísticas básicas
Ingenuo	Baja	Plazo inmediato	No modela los cambios externos
Descomposición	Baja	Inmediato, corto y mediano plazo	No modela los cambios externos
Suavización de Promedios Móviles	Baja	Inmediato y corto plazo	No modela los cambios externos
Suavización Exponencial Única	Baja	Inmediato y corto plazo	No modela los cambios externos
Doble Suavización Exponencial Lineal	Baja	Inmediato, corto y mediano plazo	No modela los cambios externos
Suavización Exponencial Lineal y Estacional	Baja	Inmediato, corto y mediano plazo	No modela los cambios externos
Suavizamiento Exponencial Amortiguado de Tendencia	Baja	Inmediato, corto y mediano plazo	No modela los cambios externos
Método de Box-Jenkins	Alta	Inmediato y corto plazo	No modela los cambios externos
Regresión Simple	Baja	Inmediato, corto y mediano plazo	Modela limitadamente los cambios externos
Regresión Múltiple	Alta	Inmediato, corto, mediano y largo plazo	Modela moderadamente los cambios externos
Métodos Econométricos	Experta	Corto, mediano y largo plazo	Modela ampliamente los cambios externos

Fuente: Elaboración propia

La definición de los periodos de tiempo más usual es la siguiente. Se entiende como plazo inmediato el periodo de tiempo que comprende de un día a un mes, como corto plazo de dos a tres meses, como mediano plazo de cuatro meses a un año y largo plazo más de un año (Makridakis y Weelwright, 2000, p.18). Las observaciones necesarias son el número mínimo requerido por un método para aplicar el mismo adecuadamente.

3.4 Selección y justificación de los 2 métodos específicos

Para hacer el pronóstico de toda variable, se pueden usar distintos tipos de pronósticos. Por supuesto, es necesario siempre escoger a aquellos métodos que sean los apropiados para realizar dicho fin, pues cada método tiene ciertas características que lo diferencían de los demás. Los puntos para la selección de los 2 métodos a usar en la presente tesis son:

- El horizonte de tiempo del pronóstico de la demanda es de 1 a 6 bimestres, por lo que se trata de un pronóstico a corto y mediano plazo. Esto implica que los métodos a usar pueden ser a corto, mediano o largo plazo, dependiendo del criterio del pronosticador.
- La variable a pronosticar es la demanda de vivienda, lo cual puede ser tratado de forma equivalente como la venta de vivienda.
- El número de observaciones con las que se cuenta es de 50.
- La experiencia de los pronosticadores, lo cual es una referencia importante.

Utilizando estos criterios, los 2 métodos más apropiados son la Doble Suavización Exponencial Lineal y el Método de Box-Jenkins. Estos métodos se describirán más a detalle y se compararán y complementarán con base en algún criterio en el siguiente capítulo, para hacer el pronóstico definitivo.