



**UNIVERSIDAD DE CHILE**  
FACULTAD DE ECONOMÍA Y NEGOCIOS  
ESCUELA DE INGENIERÍA COMERCIAL

## **PREDICCIÓN DE VARIACIONES DE PRECIO EN EL MERCADO INMOBILIARIO MEDIANTE AUTOMATAS CELULARES**

SEMINARIO DE TÍTULO INGENIERO COMERCIAL, MENCIÓN ECONOMÍA

Álvaro Cepeda Cádiz  
Gastón González Guzmán  
Profesor Guía: Antonino Parisi Fernández PhD

Santiago, Verano 2009

# PREDICCIÓN DE VARIACIONES DE PRECIO EN EL MERCADO INMOBILIARIO MEDIANTE AUTOMATAS CELULARES

Álvaro Cepeda Cádiz<sup>1</sup>

Gastón González Guzmán<sup>2</sup>

Profesor Guía: Antonino Parisi Fernández PhD

## ABSTRACTO

*Los Autómatas Celulares son un novedoso modelo no convencional de optimización. Este se basa en el funcionamiento de sistemas, en particular, en el contagio de sus células. Una aplicación con poca literatura académica, es la predicción de signos en el mercado inmobiliario, donde en esta investigación se busca estudiar la posible predicción del sentido de las variaciones de los precios de los terrenos, y sobre todo, ver si los resultados son producidos por verdadera predicción del modelo, o por efectos del azar. Si bien los resultados son favorables y alentadores, con niveles de predicción por sobre los estándares, hay que ser cuidadoso con la interpretación de estos por las características particulares que tiene el mercado inmobiliario.*

---

<sup>1</sup> Quiero agradecer a mi familia, en particular a mi padre por su revisión, a mis incondicionales amigos, a mis colegas y a Natalia, por el apoyo y sobre todo, por la presión de trabajar en esta tesis. Dedicado a Chelita

<sup>2</sup> Agradezco a mi Familia, por su apoyo incondicional, y por ser un pilar durante estos 5 años, a mis amigos, con los que compartí y trabajé durante tanto tiempo. Finalmente agradezco a Paola, por su apoyo e inmenso cariño, a ti dedico este trabajo, con todo el amor del mundo.

## ÍNDICE

INTRODUCCIÓN	Página 3
MARCO TEÓRICO	Página 5
El Mercado Inmobiliario	Página 5
Autómatas Celulares	Página 6
Aplicaciones a las Finanzas de Autómatas Celulares	Página 9
Algoritmos Genéticos	Página 10
Regla de Mahalanobis	Página 11
DATOS Y METODOLOGÍA	Página 13
Datos	Página 13
Índice Radar Logic Incorporated	Página 13
Datos de Caracterización de Zip Codes	Página 15
Metodología	Página 16
Base del Autómata	Página 16
Porcentaje de Predicción de Signo	Página 18
Optimización del Vector de Características	Página 18
RESULTADOS	Página 20
Test de Acierto Direccional	Página 21
Análisis de Rentabilidad	Página 22
CONCLUSIONES	Página 25
REFERENCIAS	Página 27

## I. INTRODUCCIÓN

La evidencia empírica ha mostrado la imposibilidad de predecir valores con exactitud de diversos activos financieros. Es por esto que en los últimos años, la atención de aquellos dedicados a las finanzas se ha enfocado en predecir la dirección que tienen las variaciones de los precios de los activos, y así poder ganarle al mercado. Dado esto, son muchos los investigadores que han dedicado tiempo y esfuerzos en desarrollar nuevas herramientas predictivas que permitan anticipar los movimientos del mercado.

Durante las últimas décadas, las finanzas han incorporado técnicas provenientes de otras ciencias, como la física y la química, para la administración de carteras con el fin de disminuir el riesgo. Entre éstas destacan los Algoritmos Genéticos, Redes Neuronales, Fractales, Lógica Borrosa y Autómatas Celulares, entre otras.

Este trabajo utiliza como herramienta principal los Autómatas Celulares, los cuales son una técnica matemática que permite modelar comportamientos complejos de distintos sistemas a través de la interacción simple entre individuos de dichos sistemas. Esta técnica nace a finales de la década de los '40, introducida por Ulam, poco tiempo después de aparecidos los computadores. Un autómata celular se define como un modelo discreto, basado en un panel con un conjunto finito de células, donde cada una de ellas puede adoptar un estado de un conjunto finito de estados posibles, dada su interacción con otras células.

En la década de los '70s el matemático John Conway formaliza los autómatas celulares mediante la introducción del *Juego de la Vida*, el cual vio por primera vez la luz pública en la edición de Octubre de 1970 del Scientific American.

Malamud & Turcotte (2000), proponen una definición alternativa, donde los estados de las células van evolucionando en tiempo discreto, de acuerdo a una regla local o un conjunto de reglas, las cuales pueden ser basadas en el estado anterior de la célula, o en el de sus vecinos. En cada período, la regla se le aplica al conjunto de células, entregando una nueva generación de autómatas.

La aplicación común en finanzas de este método matemático, es la predicción del comportamiento de acciones. Sin embargo, una aplicación casi lógica y poco aplicada es predecir variación de precios de bienes raíces (o *real estate*). Decimos casi lógica porque los autómatas son básicamente una teoría de contagio, en particular, entre células vecinas, por lo que es fácil imaginar esta similitud en el mercado inmobiliario, en particular, el precio de los terrenos.

El objetivo principal de este trabajo es predecir los signos de las variaciones de precio del mercado inmobiliario, en particular, en el MSA<sup>3</sup> de Florida, USA a través de Autómatas Celulares. Además, se busca encontrar el modelo que contenga las características (Ingreso por hogar, ubicación geográfica, etc.) de cada zona que mejoren la efectividad del Autómata, esto, a través de una optimización con Algoritmos Genéticos. El conjunto de estas características sirven como

---

<sup>3</sup> Metropolitan Statistical Area

*input* para el Autómata, mediante el uso de la norma de Mahalanobis como regla. En particular, nuestra hipótesis de trabajo es que en el mercado inmobiliario existe contagio de las variaciones de los precios, y que este contagio no se da solamente por la ubicación geográfica de las zonas a analizar. La razón por la que suponemos que el posible contagio no se da sólo por la ubicación geográfica de las propiedades, es que consideramos que los inversionistas observan las características económicas y sociales del entorno de una propiedad para generar sus expectativas de valores futuros, y luego, al llevar a cabo sus operaciones, transmiten tendencias desde unos sectores a otros. Abraham & Hendershott (1996) mostraron, estudiando la reversión a la media, diferencias significativas en la serie de tiempo del precio de las casas, entre ciudades costeras y ciudades interiores, reflejando la influencia de la zona geográfica. Capozza et al (2002) van más allá, y encuentran diferencias significativas en la correlación serial del precio de estos activos, donde estas correlaciones cambian con el tamaño de la ciudad, crecimiento del ingreso, crecimiento de la población y costos de construcción. Las condiciones para no rechazar nuestra hipótesis serían:

- 1) Obtener un porcentaje de predicción de signo (PPS) promedio, superior a 60%.
- 2) Los resultados sean estadísticamente significativos. En particular, que superen el test DA de Pesaran y Timmermann.
- 3) Que el modelo encontrado por el Algoritmo Genético no incluya solamente las variable de ubicación geográfica (Latitud y Longitud)

Especial consideración hay que tener sobre la aplicación de estos modelos a este mercado, ya que el mercado inmobiliario tiene características muy distintas al mercado accionario, donde se centran la mayoría de los estudios que usan los métodos mencionados. El mercado inmobiliario es muy ilíquido, de altos costos de transacción, periodos de transacción muy largos y además, el sector es la base de la reciente crisis financiera mundial.

La estructura del documento es la siguiente: En la segunda sección se presenta un Marco Teórico donde se explicará una breve reseña del mercado inmobiliario, se presenta en detalle la regla de Mahalanobis y las dos herramientas matemáticas a utilizar. La sección III presenta los Datos y la Metodología, en la sección IV se entregan los resultados y la sección V concluye sobre los anteriores.

## II. MARCO TEÓRICO

### II.A. El Mercado Inmobiliario

El objetivo de este apartado, no es dar una descripción detallada y completa del mercado de bienes raíces, si no de presentar algunas consideraciones importantes que se deben tener en cuenta al ver la muestra y los resultados de este trabajo, donde el enfoque va en el auge y caída del mercado desde el año 2000 a la fecha.

El mercado del “*real estate*”, y su profunda alianza con el mercado de la construcción, es un indicador fundamental del dinamismo de la economía ya que refleja condiciones crediticias, patrones de consumo y expectativas de crecimiento. Sin embargo, no se puede mencionar mucho más de este sector, sin hacer referencia al origen de la actual crisis financiera mundial, origen que se dio en este mercado, en particular, en los créditos hipotecarios *subprime*, categoría usada para prestatarios con mal historial crediticio, alto riesgo financiero y bajo (o incierto) nivel de ingresos.

Desde el año 2000 empezó a surgir una especie de *boom* inmobiliario, caracterizado por el alza del precio de las casas. Sin embargo, luego del atentado a las Torres Gemelas en Septiembre del 2001, y ante el temor de una recesión por parte de la Fed y la administración, el aumento de los precios se vio acompañado por recortes de las tasas, facilitando, e incentivando, la inversión en este sector. Esto convirtió al sector en el foco de los inversionistas, tanto experimentados como *amateurs*, donde la mayoría de las estrategias era comprar una vivienda mediante un crédito hipotecario muy favorable, dejar subir el precio, liquidar el activo, pagar el crédito y quedar con un excedente.

En Estados Unidos, los índices nacionales de precios reales de casas y rentas reales se movían en conjunto hasta el año 2000. Desde ahí, los precios reales de casas subieron a un nivel de 70% por sobre las rentas equivalentes. La pregunta es, ¿Qué llevó a esta alza sostenida del precio de las viviendas?, según el excelente artículo de Martin Feldstein (2007), aparte de expectativas irracionales sobre el precio de las viviendas, existen razones fundamentales: el crédito se volvió barato y fácil de obtener, esto por los recortes de tasa de la Fed, que indujeron a una caída en las tasas hipotecarias, y una amplia oferta de hipoteca con bajas tasas “seductoras”<sup>4</sup> temporales. Por otra parte, el dinero hipotecario también se volvió abundante, como resultado de varios cambios institucionales. El nacimiento de las hipotecas *subprime* fueron resultado de cambios legislativos y el vasto uso de modelos estadísticos de valoración de riesgo usados por los prestamistas. Además, la securitización de estas deudas indujo una disminución de estándares por prestamistas que no poseían las hipotecas que creaban.

Quizás, uno de los problemas más llamativos, es la interpretación de los expertos de lo que estaba pasando, donde se dijo que los bajos *spreads* crediticios eran respuesta a que los mercados se habían vuelto menos riesgosos. Muchos administradores de portfolio comenzaron a tomar riesgos que en otras situaciones no lo habrían hecho, los inversionistas se sentían satisfechos por

---

<sup>4</sup> N del T, “teaser” rates

la aparente transferencia del riesgo, y los inversionistas menos sofisticados compraban estos productos estructurados (paquetes crediticios, bonos securitizados) sin identificar el riesgo existente.

El comienzo de la crisis, como ya se mencionó, fue la crisis en el mercado *subprime*. Estos prestatarios aprovecharon las condiciones crediticias para incorporarse al *boom* hipotecario, donde la mayoría de ellos negoció con *brokers* hipotecarios, y no con bancos directamente.<sup>5</sup> La caída (que muchos esperaban que explotara la burbuja) de los precios de las viviendas llevó a estos inversionistas a caer en *default* crediticio, no pudiendo cubrir los pagos mensuales del préstamo hipotecario (como sería en situaciones normales), siendo esto transferido a los *brokers*, quienes además, vendían esto al mercado financiero, donde la securitización por “*tranchers*” permitió combinar los riesgos y llegar a distintos tipos de inversionistas, los que incluían bancos americanos y europeos. El hecho de no haber puesto atención a estos indicios de problemas financieros, dado que el volumen de estas transacciones son una fracción pequeña del mercado crediticio global, no fue considerado preocupante para la economía nacional de Estado Unidos, fue lo que llevó a la propagación de la crisis hacia otros países. Expertos y representantes de gobierno afirmaron que serían efectos temporales o que tendrían un impacto limitado en el mercado de capitales y la economía. Quiebras en bancos suizos, inversionistas sin respaldo a sus activos, etc. fue a su vez, lo que llevó la crisis hipotecaria a crisis financiera a nivel global, siendo Octubre del 2008 uno de los meses más duros en la historia económica mundial.

## II.A. Autómatas Celulares

Como se mencionó en la sección anterior, los Autómatas Celulares, son una técnica matemática basada en un panel con un conjunto finito de células, donde cada una de ellas puede adoptar cierto estado de un conjunto finito de estados posibles. Los estados de las células van evolucionando en tiempo discreto, de acuerdo a una regla local o un conjunto de reglas, las cuales pueden ser basadas en el estado anterior de la célula, o en el de sus vecinos. En cada período, la regla se le aplica al conjunto de células, entregando una nueva generación de autómatas.

John Conway plantea un modelo de autómatas bidimensional. Para entender este autómata celular, debemos considerar un tablero de ajedrez sin límites, donde cada cuadrado tiene ocho vecinos y dos estados posibles, blanco (viva) o negro (muerta). Por ejemplo, si se sigue una regla de 3-3, quiere decir que una celda viva en el período “0”, pasa a ser muerta en el período “1”, si y sólo si, 3 o más de sus vecinos están muertas en “0”. Dado que la regla es 3-3, es la misma condición para pasar de muerta a viva.

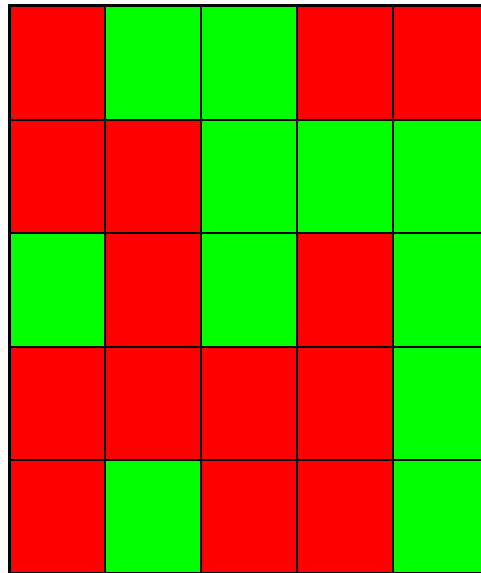
A manera de ejemplo, la situación descrita se presenta en Figura 1 y Figura 2. La Figura 1 muestra la primera generación de células, consideremos las celdas verdes como las vivas y las

---

<sup>5</sup> De hecho, esa fue una tendencia. Los *brokers* fueron desplazando a los bancos. En la actualidad, los bancos han perdido participación en algunos sectores. Por ejemplo, solo el 20% de la deuda corporativa no financiera es mantenida por bancos (Chari et al, 2008)

rojas muertas. Siguiendo la regla mencionada de 3-3, la segunda generación deberá ser como muestra la Figura 2. Para ejemplificar aún más, consideremos la celda central la cual está viva y rodeada de 2 celdas vivas y 6 muertas, esto implica que para la segunda generación, la celda deberá estar muerta.

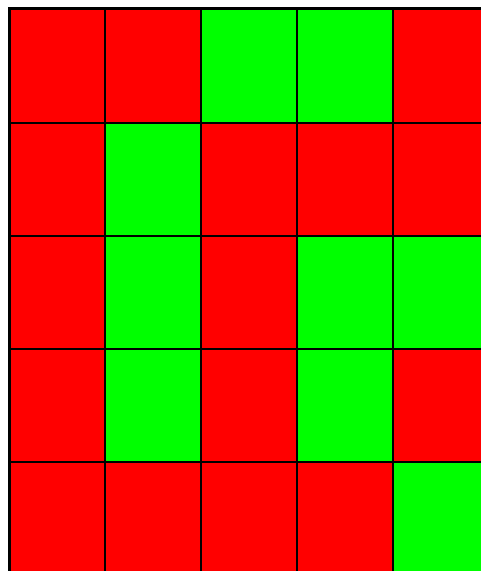
FIGURA 1



Primera Generación

Fuente: Elaboración Propia

FIGURA 2



Segunda Generación

Fuente: Elaboración Propia

Como se puede apreciar en la Figura 2, la celda central está muerta en la segunda generación, respondiendo a la regla planteada.



De esta manera, la conformación del llamado “tablero de ajedrez” cambia en cada periodo, y las generaciones futuras se encuentran determinadas por la primera matriz, es decir de la primera generación. De todas maneras, y aunque las reglas sean simples, los resultados que genera este modelo pueden replicar sistemas complejos. Así, un autómata celular es un sistema dinámico discreto, cuya evolución es consecuencia de reglas locales.

Un punto importante del “Juego de la Vida” es que es un juego de “cero jugadores”, es decir, la evolución está determinada por un estado inicial, que no necesita de la alimentación de un jugador humano. Así esta segunda generación, evoluciona en saltos discretos de tiempo y en todas las direcciones.

En la medida que se vayan obteniendo más generaciones se irán generando diferentes estructuras; algunas veces obtendremos figuras más simétricas, y otras veces una imagen cualquiera sin una forma definida; además, siempre pueden ocurrir cambios inesperados en donde una figura puede pasar a ser una figura completamente diferente. Tal como ocurre en la “Teoría de la Evolución de las Especies”, en la que están fundamentados los Autómatas Celulares, en algunos casos la sociedad simplemente muere (sin importar el número de generaciones previas), y en otros casos, se alcanzan patrones estables de “sociedad” que corresponden a figuras estables dentro de la expansión. Estos últimos patrones son llamados por Conway como “Aún Vivas”, y corresponden a patrones que se mantienen estables en el tiempo o que oscilan por siempre. Incluso patrones que inicialmente no parecen simétricos pueden tender a serlo en el desarrollo de nuevas generaciones, lo que claramente les da un grado mayor de estabilidad.

Los patrones que se van formando dentro de los autómatas, al igual que estos, tienen nombres definidos. Existen una serie de patrones organizados por tipo, primero que nada están los “Aún Vivas” que son patrones estables que se mantienen en el tiempo, luego están los “Osciladores” que repiten patrones<sup>6</sup>, y por último los están “Naves Espaciales” que se van trasladando a través de del espacio.

Ya en los ochentas, Stephen Wolfram comienza a desarrollar el área mediante estudios teóricos para luego aplicarlos a códigos computacionales sencillos, que buscan encontrar patrones complejos del comportamiento de la naturaleza. Dicho de otra forma, un sistema cuyo funcionamiento parece simple, es tan simple como cualquier sistema computacional, según el cual, reglas simples, generan resultados complejos. Es decir, muchas veces, los sistemas complejos tienen un comportamiento contra intuitivo. Una forma de interpretar y aplicar lo anterior, es decir que la complejidad de los sistemas se debe simplemente a gran cantidad de componentes simples que interactúan de forma simultánea. Así nace el principio de “Equivalencia Computacional”, que plantea que las herramientas computacionales son tan complejas como el sistema a estudiar.

---

<sup>6</sup> Estos patrones pueden ser considerados como un set de “Aún Vivas”.

Lo recién planteado no es trivial, ya que nos indica que habría límites a la misma ciencia, ya que muchas cuestiones serían imposibles de responder, debido a que la única manera de descubrir los resultados de procesos de complejidad, sería simplemente, dejar que ocurran.

Esta “nueva ciencia” tiene en cuenta la propia irreductibilidad computacional, en que las reglas que subyacen a un sistema, no siempre pueden predecir el comportamiento que resulta de dicho sistema. Por ejemplo, según Wolfram, el problema del libre albedrío podría analizarse desde esta perspectiva, como el resultante de una complejidad, dada una serie de reglas simples a nivel básico, que luego deriva en cálculos irreductibles y complejos, que difícilmente pueden ser simplificados por lo que, finalmente, son considerados el “producto de la libre voluntad”.

Entonces la estructura de un autómata celular viene dada por:

- Un espacio de  $n$  dimensiones, subdividido en un número de sub espacios homogéneos que se conocen como celdas o células.
- Cada celda puede estar en algunos de los estados definidos por un conjunto finito de posibilidades.
- Cada celda en todo período o generación debe tener un estado asignado.
- Cada celda debe tener una vecindad conocida, determinada por las celdas que la rodean.
- Es necesario que exista además, al menos una regla que determine el cambio o permanencia de estado de las celdas.
- Debe existir un “reloj virtual”, que provoque cambios de acuerdo a la regla a todas las celdas al mismo tiempo.
- Finalmente y a modo de resumen, todas las celdas deben cumplir las siguientes características:
  - Tener el mismo conjunto de estados posibles.
  - Tener la misma forma o función de vecindad
  - Someterse a la(s) misma(s) regla(s) de evolución.

### *II.A.1. Aplicaciones en Finanzas de Autómatas Celulares*

Dentro de las aplicaciones en Finanzas, destacan tres artículos publicados. El primero es de Zhou et al (2004), que modela el comportamiento de un Mercado Financiero ficticio mediante el modelamiento con Autómatas Celulares. El asume que las celdas que componen la matriz son inversionistas, ellos solo tenían tres opciones de comportamiento comprar, vender o no hacer nada (Estados de las celdas), y además tenían la capacidad de aprender de sus estrategias pasadas y las de sus vecinos. La decisión de cada inversionista se desarrolla en dos partes, en la primera, cada uno decide una estrategia, de acuerdo a su historial de inversión pasada y el nivel de precios (relacionado con su aversión al riesgo), luego, en la segunda etapa comienza a examinar las

estrategias de sus vecinos, y pueden cambiar su decisión de acuerdo al resto, esto representa que los inversionistas, tienden a seguir a los “ganadores”.

El segundo artículo fue escrito por PhD. Antonino Parisi y PhD. Franco Parisi (2004), y mediante autómatas celulares ellos logran modelar el comportamiento de índices bursátiles norteamericanos, logrando resultados estadísticamente significativos.

Finalmente Bartolazzy y Thomas (2006) mostraron que los Autómatas Celulares son capaces de capturar algunas características del comportamiento de los mercados, como la no linealidad y su agrupamiento de volatilidades. Todo lo anterior nos permite pensar que existe la posibilidad de reproducir mercados y generar estrategias de inversión rentables.

## II.B. Algoritmos Genéticos

Los Algoritmos Genéticos (de ahora en adelante AG) son un método de optimización de estructura de modelos basado la Teoría de la Evolución de las Especies, planteada por Charles Darwin, donde se afirma que son sólo los más aptos quienes van sobreviviendo al paso del tiempo y además son estos quienes pueden reproducirse.

Los AG buscan la autoselección dentro de las potenciales variables que pueden determinar el comportamiento de alguna variable. Cada una de estas posibles candidatas a ser variables explicativas dentro del modelo, corresponden a un gen, que según las características de un individuo puede ser activo o recesivo. Entonces cualquier modelo puede ser representado mediante su “código genético”, que nos dice que genes se encuentran activos y cuáles no en su organismo. Por ejemplo:

FIGURA 3

$X_1$	$X_2$	$X_3$	$X_4$	$X_5$
1	1	0	0	1
1	0	1	0	1

Fuente: Elaboración Propia

En la Figura 3 podemos ver un posible modelo que cuenta con cinco potenciales variables explicativas, e inmediatamente después dos “individuos” o modelos con su código genético, entonces esto quiere decir que el primer individuo, considera las variables 1, 2 y 5 como genes activos (variables a utilizar en la estimación) y las variables 3 y 4 como genes recesivos análogamente el segundo considera las variables 1, 3 y 5 como genes activos y las variables 2 y 4 como genes recesivos.

Como el objetivo de los AG, es filtrar a los mejores modelos para que estos sigan adelante es necesario un criterio que permita ordenar a los individuos, con el fin de ver cuáles son los “más aptos”, para esto típicamente se utiliza el criterio del PPS<sup>7</sup>.

Finalmente son los modelos más aptos los que dan vida a la siguiente generación mediante su reproducción (combinación aleatoria de estos modelos). Un punto importante de los AG es que al igual que los seres humanos, comparten una característica fundamental en su proceso reproductivo, la mutación. Es decir cada individuo recibe sus genes, por un lado de sus padres y por otro del azar. En los seres humanos esta cualidad ha demostrado ser una herramienta eficaz contra la lucha por ejemplo de enfermedades, y en los AG también ha demostrado ser de gran ayuda, ya que modelos que permiten mutación logran mejores resultados que aquellos que transmiten directamente los genes de los padres.

Esta herramienta de optimización ha probado su tremenda utilidad, en especial en modelos que permiten una gran cantidad de posibles variables explicativas para un fenómeno. Es importante destacar que la cantidad de posibles modelos para un variable que tenga n potenciales variables explicativas es igual  $2^n$  y claramente, en algunos casos eso pueden ser muchas alternativas como para pensar en probarlas todas. Por ejemplo para la construcción de modelos ARIMA para predecir la dirección de la variación de precios de acciones norteamericanas, se ha utilizado 18 variables explicativas. Es decir las combinaciones posibles son 262.144, pero el AG ha probado que luego de 8 generaciones de individuos (aproximadamente 800 alternativas), los resultados convergen y se encuentra un óptimo.

### *II.C. Norma de Mahalanobis*

Esta medida de distancia fue creada por el matemático y estadístico indio Prasanta Chandra Mahalanobis en el año 1936, y la gran diferencia que tiene con la norma de distancia euclidiana, que es la más utilizada, es que considera la correlación entre los componentes del vector de características a analizar. La ventaja que tiene el incluir la matriz de varianza y covarianza en el análisis, es que permite ponderar las distancias por sus varianzas, no como en el caso de la norma euclidiana, que sólo las pondera linealmente. De esta forma podemos eliminar posibles fuentes de sesgo que se produzcan en el cálculo de la distancia entre los activos, de los que no nos hayamos dado cuenta. Entonces la norma de Mahalanobis se define como:

$$NM = \sqrt{\frac{1}{(x_i - x_j)^T \Sigma^{-1} (x_i - x_j)}} = ((x_i - x_j)^T \Sigma^{-1} (x_i - x_j))^{-\frac{1}{2}}$$

---

<sup>7</sup> Este criterio es explicado más adelante en la sección de metodología.

Donde  $x_i$  es el vector de características del individuo  $i$ ,  $x_j$  es el vector de características del individuo  $j$  y  $\Sigma^{-1}$  es la inversa de la matriz de varianza y covarianza de las características de los individuos.

### III. DATOS Y METODOLOGÍA

#### III.A. Datos

Los datos a utilizar en el estudio corresponden a datos mensuales del índice de precios por pie cuadrado para la MSA de Florida, USA, que proporciona Radar Logic Incorporated<sup>®</sup>. Estos vienen subdivididos por zip codes, y son para un período de 8 años, comprendido entre el 28-04-2000 y el 21-11-2008. Los códigos del área de nuestro interés van del 33004, hasta el 33498, obteniendo así las zonas geográficas de menor tamaño disponible. Esto nos da un total de 182 códigos postales. Contamos con 104 observaciones para cada zip code, de los cuales eliminamos 51 ZIP Codes por información incompleta, por lo tanto la muestra final está compuesta de 131 ZIP Codes

La razón de utilizar estos datos sobre los precios registrados en las últimas transacciones es precisamente la dificultad de obtener datos exactos sobre precios de las mismas propiedades. Además, el mercado inmobiliario se caracteriza por ser poco líquido, altamente pro cíclico con los movimientos de la economía y además, tiene un volumen de transacciones bajo en comparación a otros activos.

Por otro lado, utilizaremos datos económicos y de uso de suelo, entre otros, provenientes de USA Zip Codes Data Business Deluxe, con el fin de caracterizar la muestra. Estos datos son explicados con más detención en la definición de las variables en la subsección siguiente.

#### III.A.1 Índice Radar Logic Incorporated<sup>®</sup>

Para sobrellevar las dificultades que tienen la obtención e interpretación de los datos del mercado inmobiliario, Radar Logic busca entender la naturaleza de proceso de valoración del “real estate”, y cómo reflejar la actividad total de un área específica, en un día en particular, a un índice que confiablemente represente los cambios en el mercado día a día. El estudio que ellos realizaron los condujo a lo que ellos nombraron como la *Triple Power Law*, o la ley de tres poderes, con la que se caracterizan las transacciones del mercado de bienes raíces. Las razones de porque se utiliza este índice para este trabajo es porque presenta las siguientes ventajas:

1. Refleja la actividad diaria del mercado, sin los rezagos y distorsiones inherentes en promedios mensuales o trimestrales.
2. Captura las tendencias y patrones del mercado sin suavizar o rezagar la realidad del mercado. Es decir, los patrones estacionarios son mostrados sin rezago o disolución y cambios repentinos del mercado son reflejados en cambios repentinos en el índice.
3. Es un índice de “información completa”, ya que no excluye cualquier transacción válida, como posibles *outliers*. Además se vuelve resistente a fraudes y errores de datos.

Para caracterizar este tipo de transacciones que se efectúan en un área, necesitamos una medida que nos permita comparar entre propiedades grandes y pequeñas. Si sólo miramos los precios con los que las casas cambian de propietario, son una representación limitada por la información que ignoran, más aún, la uniformidad del activo, en este caso, una casa, puede cambiar en el tiempo por posibles renovaciones. Otras cosas que pueden afectar, son la tenencia promedio del activo<sup>8</sup> o la imposibilidad de incluir ventas de nuevas casas.

En vez de observar el precio de una casa, se mira el precio de una casa por pie cuadrado, de forma de hacer las transacciones comparables. Esto es una práctica aceptada dentro del mercado inmobiliario, también utilizado por la mayoría de los constructores en Estados Unidos, y menos formalmente, por quienes buscan nuevas propiedades. Sin embargo, es importante destacar que si bien las transacciones se vuelven más comparables, no son un *commodity*, ya que siguen existiendo grandes diferencias entre casa y casa.

Por lo tanto, lo que plantea Radar Logic Incorporated<sup>®</sup> es caracterizar una distribución de transacciones. Por ejemplo, tomemos un *commodity* como el petróleo. Si el precio del barril es de US\$60, esto representa que podemos vender barriles de petróleo a ese precio (corrigiendo por ciertos factores como costos de transporte, calidad, etc.), en otras palabras, por muy redundante que suene, el precio realmente refleja el valor al que se puede transar el activo en ese momento del tiempo. Esto no sucede necesariamente en el mercado inmobiliario, ya que al ver las transacciones de propiedades en un día en particular, observamos una distribución de precios, incluso cuando miramos el precio por pie cuadrado. La razón, es que además de la superficie, hay otros atributos que afectan el valor de un bien inmueble. Por lo tanto, el índice debe ser construido usando una composición de precios de todas las diferentes propiedades para las cuales haya transacciones registradas. Esta composición debe ser muy meticulosa, ya que hay muchas medidas simples que no son precisas en reflejar el valor de mercado del activo.

El análisis de los datos recolectados, permitió encontrar importantes patrones de la distribución de los precios, y es a partir de esa distribución que ellos construyen el índice. Dado que la distribución es continua, el índice evita problemas de concentración<sup>9</sup> que tengan los datos. Por ejemplo, imaginemos una distribución cuya mediana fuese \$100, pero que no registra transacciones entre \$95 y \$105. Si usáramos los datos, se podría obtener un valor bajo \$95 o sobre \$105, pero no \$100. Además la distribución ha mostrado ser menos sensible a outliers, que el uso de medianas sobre los datos. Aún más, es muy difícil manipular el índice introduciendo transacciones fraudulentas, o ajustar datos de cierre para un día en particular.

Una consideración importante que se debe hacer sobre cualquier índice que refleje de forma precisa los datos de mercado, es que debe tener una cantidad significativa de movimiento del día a día, especialmente cuando el número de transacciones registrada es relativamente pequeñas. En particular, este índice se basa en datos de todo un año para obtener la distribución

---

<sup>8</sup> Amihud y Mendelson (1986), mostraron que la tenencia promedio influye en el valor del activo.

<sup>9</sup> N del T, "Granularity" grado de disolución de una composición

de un solo día, con esto se genera parámetros de forma de la distribución, donde los datos del día en cuestión entregan la posición de la distribución y finalmente el valor del índice para ese día.

La distribución que se utiliza finalmente es la que maximiza la verosimilitud de los datos. Esta distribución se obtiene computando tanto los parámetros de forma y posición para 365 días consecutivos, incluyendo el día actual. El uso de 365 días garantiza que no existan efectos estacionarios que puedan distorsionar los parámetros de forma, mientras que todavía permite que puedan evolucionar en el tiempo, mientras van cambiando las condiciones del mercado. Además, como la posición cambia día a día, no hay rezagos en el cambio del índice al cambio en precios.

Información detallada de la construcción del índice puede ser encontrada en el documento de Kagarlis et al (2007)<sup>10</sup>, dado que el tema central de este trabajo no es la construcción de dicho índice.

### *III.A.2 Datos de Caracterización de Zip Codes*

Para caracterizar los Zip Codes, se utilizan datos del USA Zip Codes Data Business Deluxe. Esto para tener en cuenta variables, aparte de la geográfica, que puedan ser importantes para la determinación de precios del pie cuadrado. Las variables que consideraremos (y que tienen datos disponibles) son:

- i) Ubicación Geográfica: Es la variable más lógica como medida de distancia, esto por dinámica del contagio y su semejanza a la aplicación de los autómatas celulares. Esto consiste en latitud (LAT) y longitud (LONG)
- ii) Población: Representaremos el nivel de población de cada zip code como un ranking de deciles del porcentaje de población con respecto al total del MSA de Florida (D\_POP). La idea del ranking es disminuir la posible volatilidad de la variable, y que al presentarla como porcentaje del total del MSA, los datos resultan más comparables. Estos son datos estimados, en base al censo del 2000
- iii) Hogares: Esto es la cantidad total de hogares por zip code, lo que podría considerarse como una aproximación del tamaño del zip code (HOG). Esto tiene importancia ya que los movimientos de zonas con mayor nivel de hogares podría influir más fuerte que zonas menos habitadas.
- iv) Personas por Hogar: Este es el promedio estimado del número de personas por hogar (PPH), a partir del censo del 2000
- v) Valor Promedio: Valor promedio de las casas en cada zip code (VPR).
- vi) Ingreso por Hogar: Ingreso promedio por hogar (INGH), también estimado en base al censo del año 2000. Esta también es una aproximación importante del tamaño económico de la zona.

---

<sup>10</sup> The Radar Logic Index White Paper. Link: <http://www.radarlogic.com/white%20papers/The%20Radar%20Logic%20Index%20Methodology%20White%20Paper.pdf>



- vii) Edad: Mediana de la edad de todas las personas que viven en la zona geográfica respectiva (MED), estimado a partir del censo del año 2000.
- viii) Zona: Con esto buscamos diferenciar si la zona es mayoritariamente comercial o residencial. Para lograrlo usaremos 2 proxys distintas. Primero, tenemos el “Establishment”, es el número total de establecimientos de negocios para cada zip code, rankeado por deciles(D\_EST). La otra, calculamos un ratio entre la cantidad de casillas postales de negocios sobre casillas postales de hogares, también ordenado por deciles (D\_NEG).

### III.B Metodología

Para la aplicación de Autómatas Celulares, vamos a definir diferentes zonas geográficas del Metropolitan Statistical Area de Miami, del estado de Florida, como células, en particular, las zonas representadas por distintos ZIP Codes (códigos postales). De los 131 ZIP Codes disponibles, se toman las variaciones porcentuales de mes a mes como principal Input para el modelamiento.

Los Autómatas exigen un ordenamiento de las células, el cual puede ser al azar, o por algún patrón o características, donde la ubicación geográfica surge lógicamente como orden de las células. Sin embargo, hay características que van más allá de la ubicación geográfica que podrían influir en el precio de los bienes raíces. Por ejemplo, si una zona es residencial, con cierto nivel de ingresos, sin que necesariamente sean zonas colindantes, puede influenciar sobre otra de características similares. Es por esto, que la distancia que tendrá un ZIP code a otro se basará en la norma de Mahalanobis, donde se pueden incluir varias características, que en este caso en particular, son las variables definidas en la sección III.A.2. Esto permite una caracterización y ubicación más realista de las células del Autómata.

#### III.B.1 Base del Autómata

Para poder generar el autómata, lo primero es calcular las distancias entre cada uno de los Zip Codes. La distancia (valor de la norma de Mahalanobis) entre los Zip Codes “i” y “j” queda representada por  $DISTM_{ij}$ . Para la muestra de este trabajo:

$$DISTM_{ij} = ((\vec{x}_i - \vec{x}_j)^T * \Sigma^{-1} * (\vec{x}_i - \vec{x}_j))^{-1/2}, \text{ con}$$

$$\vec{x}_i = \begin{pmatrix} LAT_i \\ LONG_i \\ D\_POP_i \\ HOG_i \\ PPH_i \\ VPR_i \\ INGH_i \\ MED_i \\ D\_EST_i \\ D\_NEG_i \end{pmatrix}$$

Donde  $\vec{x}_i$  es el vector de características del Zip Code "i",  $\Sigma$  es la matriz (10x10) de Varianza-Covarianza de las características entre todos los Zip Codes, por lo tanto  $DISTM_{ij}$  es un escalar. Luego, estas distancias son normalizadas y usadas como proxys de la influencia que tiene un Zip Code sobre los otros, definiendo

$$w_{ij} = \frac{DISTM_{ij}}{\sum_{j \neq i}^N DISTM_{ij}}$$

Donde  $w_{ij}$  es la distancia que usaremos como input de acá en adelante.

Vamos a definir también una variable dummy llamada "Estado Célula" (EC), la cual toma valor uno si una célula está viva y cero si está muerta. En términos de nuestros datos:

$$EC_{it} = \begin{cases} 1 & \text{si } Ret_{it} > 0 \\ 0 & \text{si } Ret_{it} \leq 0 \end{cases}$$

Donde  $Ret_{it}$  es el retorno del Zip Code "i" en el período "t". Para definir la regla de contagio, nosotros desarrollamos una adaptación del Juego de la Vida de Conway. Entonces, si el retorno del pie cuadrado en un Zip Code es positivo (célula viva), sumamos la influencia de todos los Zip Codes que tengan retornos negativos (célula muerta) para esa fecha, esta variable la denominaremos "Influencia Celular" (IC), donde:

$$IC_{it} = \begin{cases} \sum_{j \neq i}^n w_{ij} (1 - EC_{jt}) & \text{si } EC_{it} > 0 \\ \sum_{j \neq i}^n w_{ij} EC_{jt} & \text{si } EC_{it} \leq 0 \end{cases}$$

Para desarrollar el Autómata se hace necesario definir Valores Críticos para los estados de las células, es decir, un Valor Crítico para células muertas (VC-), y uno para células vivas (VC+). Entonces, al igual que en el Juego de la Vida, la celda " $X_{it}$ " cambiará de estado sí, y sólo sí, la función  $IC_{it}$  supera el Valor Crítico asociado a su estado actual.

Ambos Valores Críticos son obtenidos de un proceso de optimización que busca maximizar el PPS conjunto de todos los Zip Codes. El proceso de optimización consiste en la generación de números aleatorios con 10.000 iteraciones.

### III.B.2 Porcentaje de Predicción de Signo

Una forma de medir el desempeño de una herramienta predictiva, es medir su PPS, o porcentaje de predicción de signo. Para calcular el PPS, se compara el signo de la variación proyectada con el signo de la variación observada real en el periodo “t” para un Zip Code dado, comenzando desde  $t = 2$ . Si los signos de la variación proyectada y la variación observada son los mismos, entonces se anota un 1, lo cual representa un acierto o hit, en caso contrario, se anota un 0, lo que señala un error de predicción del modelo. Por lo tanto, se define

$$p_{it} = \begin{cases} 1 & \text{si } Ret_{it} * \widehat{Ret}_{it} > 0 \\ 0 & \text{tol} \end{cases}$$

$$PPS_i = \frac{1}{T} \sum_{t=1}^T p_{it}$$

$$PPS = \frac{1}{N} \sum_{i=1}^N PPS_i$$

Donde  $p_{it}$  representa el acierto del modelo, tomando valor  $p_{it} = 1$  si es un acierto, y  $p_{it} = 0$  si es un error de predicción para el Zip Code “i” en el periodo “t”.  $PPS_i$  es el porcentaje de predicción de signo promedio para el Zip Code “i”, y  $PPS$  es el porcentaje de predicción de signo promedio del modelo, siendo este último el que se busca optimizar.

### III.B.3 Optimización del Vector de Características

Si bien tiene lógica usar características que van más allá de la ubicación geográfica para ver la influencia de un sector a otro, no necesariamente todas las variables propuestas en la sección III.A.2 son relevantes. Por lo tanto, se procede a la optimización del vector de características, más específicamente a la selección de sus componentes, para dejar sólo aquellos que sean necesarios.

Para lograr lo anterior se utilizó la herramienta de Algoritmos Genéticos que prueba, en una primera etapa, estructuras generadas de manera aleatoria. La única restricción que nosotros imponemos sobre la generación de modelos aleatorios, es que el primero, incluye todas las variables propuestas. El resto de los modelos que compone la generación inicial es creado mediante números aleatorios.

La razón por la que usamos Algoritmos Genéticos es que, dado que nosotros proponemos 10 posibles componentes para el vector de características, existen muchos posibles modelos. Los AG solucionan este problema probando sólo una fracción de estos.

Para el desarrollo de nuestro modelo, ejecutaremos 5 generaciones, cada una con 100 individuos. Como mencionamos anteriormente la primera generación es aleatoria a excepción del primer modelo que incluye todas las variables.

Luego de probarse<sup>11</sup> esta generación inicial, los 20 mejores modelos darán vida a la generación siguiente. Para la reproducción, cada uno de los padres aportará el 50% de los genes, para generar 80 modelos más, la selección de los padres es aleatoria dentro de los modelos “clasificados”, es decir, cualquiera de los 20 mejores se puede reproducir con cualquier otro del mismo grupo. Posteriormente se da paso al proceso de mutación, que para nuestro propósito, es que si un gen estaba desactivado se active o el caso contrario. La probabilidad de mutación definida para esto es de un 8,375%. Al terminar la mutación, los modelos están listos para poder ser probados nuevamente y repetir el proceso.

---

<sup>11</sup> Por probarse se entiende la ejecución del AC y la optimización de los Valores Críticos, para cada uno de los 100 individuos

#### IV. RESULTADOS

Después de cinco generaciones, el mejor modelo presenta un PPS de 64,2778%, con valores críticos de 0,236319 para los valores positivos y 0,386885 para los valores negativos. Lo interesante es ver primero, como van mejorando los resultados a medida que el Algoritmo tiene más generaciones, y a la vez, que el uso de cinco generaciones parece ser adecuado por la convergencia de los resultados. Esto se puede ver en la Tabla 1, donde se muestran los 3 mejores resultados (en términos de PPS) para las 5 generaciones, incluyendo el modelo y los valores críticos.

TABLA 1

	N°	LAT	LONG	D_POP	HOG	PPH	VPR	INGH	MED	D_EST	D_NEG	VC(+)	VC(-)	PPS
GEN I	1	0	0	1	1	0	1	0	1	0	1	0,1532884	0,3937847	<b>0,641962</b>
	2	0	0	1	1	0	0	0	0	0	1	0,0973562	0,3914004	<b>0,641814</b>
	3	0	0	1	1	1	1	1	0	1	1	0,2081483	0,3926123	<b>0,641592</b>
GEN II	1	0	0	1	1	1	1	1	0	1	1	0,2954402	0,40141	<b>0,642037</b>
	2	0	0	0	0	1	0	0	1	1	1	0,2777885	0,3950276	<b>0,642037</b>
	3	0	0	1	1	0	1	0	1	0	1	0,1532884	0,3937847	<b>0,641962</b>
GEN III	1	0	0	1	1	0	1	0	1	0	1	0,2363188	0,3868848	<b>0,642778</b>
	2	0	0	1	1	0	1	0	1	0	1	0,3424609	0,3880126	<b>0,642333</b>
	3	0	0	1	1	0	1	1	0	1	1	0,125124	0,4004344	<b>0,642185</b>
GEN IV	1	0	0	1	1	0	1	0	1	0	1	0,2363188	0,3868848	<b>0,642778</b>
	2	0	0	1	1	0	0	1	1	0	1	0,2438393	0,3868614	<b>0,642778</b>
	3	0	0	1	1	0	1	0	1	0	1	0,1801363	0,3866317	<b>0,64263</b>
GEN V	1	0	0	1	1	0	1	0	1	0	1	0,2363188	0,3868848	<b>0,642778</b>
	2	0	0	1	1	0	0	1	1	0	1	0,2438393	0,3868614	<b>0,642778</b>
	3	0	0	1	1	0	1	0	1	0	1	0,188366	0,3868463	<b>0,642778</b>

FUENTE: Elaboración Propia

Como se puede apreciar en la Tabla 1, existe una convergencia hacia un PPS de 0,642778 desde la tercera generación, así como también estos fueron mejorando de generación en generación. Sobre las variables, los unos en cada variable representan que la variable es usada por el modelo como una característica de cada Zip Code, y por ende, forman parte de la distancia entre cada célula según la norma de Mahalanobis. El modelo finalmente incluye el decil de población, el número de hogares, el valor promedio de las viviendas, la media de la edad y el decil de la variable de tipo de zona D\_NEG para cada Zip Code. Además, el modelo con mejores resultados, presenta sus PPS's máximos para los Zip Codes 33155, 33407 y 33414, alcanzando un nivel de aciertos del 72,8%. Estos resultados son sumamente positivos, ya que apoyan la tesis de que existe un grado de predictibilidad en el movimiento de los precios para el mercado de Real Estate.

Finalmente, podemos ver que los mejores modelos, excluyen ambas variables asociadas a la ubicación geográfica de las células. Nosotros ofrecemos dos posibles interpretaciones para este fenómeno. Primero, que la medida de ubicación utilizada, no sea la correcta. Con esto nos

referimos a que las diferencias o distancias entre 2 Zip Codes vecinos, puede ser muy similar a las diferencias entre 2 Zip Codes que estén en los extremos de la muestra. Esto quiere decir, que las diferencias no son lo suficientemente claras, como para marcar una diferencia. De todas formas este problema debió solucionarse, mediante el uso de la Norma de Mahalanobis como medida de distancia, ya que por el uso de la matriz de varianza y covarianza, corrige por la variabilidad de la característica, pero aún así el error persiste.

La segunda razón posible, es que el no haber definido una vecindad restringida para nuestras celdas, diluimos el efecto que tienen ciertas características, sobre la distancia, ya que al ser tantos “vecinos”, la influencia que puedan tener unos pocos, sus verdaderos vecinos, es muy baja.

#### IV.A Test de Acierto Direccional

Es importante asegurarse que los resultados que arroja el modelo son por una verdadera capacidad predictiva del mismo, y no por el azar. Para esto, realizamos el Test de Acierto Direccional (DA, de su sigla en inglés) de Pesaran y Timmermann (1992). Los autores prueban la hipótesis nula de que los signos de los retornos reales son independientes de los signos de los retornos proyectados, por lo tanto, el rechazo de la hipótesis nula, se puede afirmar que existe evidencia de que el modelo tiene una capacidad predictiva estadísticamente significativa.

El test, al igual que la medición por PPS, compara el signo de la proyección, con el signo de la rentabilidad observada. Entonces, mientras los signos coincidan, aumenta la efectividad del Automata, en caso contrario, aumenta el error de predicción.

Luego, se calcula el “Ratio de Éxito” para cada Zip Code, que de acuerdo a nuestra notación corresponde a  $PPS_i$ . Vamos a definir una función  $I(x)$ , donde:

$$I(x) = \begin{cases} 1 & \text{si } x > 0 \\ 0 & \text{si } x \leq 0 \end{cases}$$

Además, definimos  $P_i$  y  $\hat{P}_i$  como

$$P_i = \frac{1}{T} * \sum_{t=1}^T I_{it} (Ret_{it})$$

$$\hat{P}_i = \frac{1}{T} * \sum_{t=1}^T I_{it} (\widehat{Ret}_{it})$$

Donde el ratio de éxito en el caso de independencia de los retornos (SRI), es:

$SRI_i = P_i * \hat{P}_i + (1 - P_i) * (1 - \hat{P}_i)$ , donde  $P_i$  y  $\hat{P}_i$  son el porcentaje de positividad para las observaciones y las predicciones respectivamente. El test se compone de las varianzas de las variables  $SRI_i$  y  $PPS_i$ . Por lo tanto, necesitamos encontrar dichas varianzas. Para la primera:

$$VAR(SRI_i) = \frac{T(2\hat{P}_i - 1)^2 P_i(1 - P_i) + T(2P_i - 1)^2 \hat{P}_i(1 - \hat{P}_i) + 4P_i \hat{P}_i(1 - P_i)(1 - \hat{P}_i)}{T^2}$$

$$VAR(PPS_i) = \frac{SRI_i(1 - SRI_i)}{T}$$

Por lo que finalmente, el test DA de Pesaran & Timmermann se define como

$$DA_i = \frac{PPS_i - SRI_i}{\sqrt{VAR(PPS_i) - VAR(SRI_i)}}$$

Dado que el test sigue una distribución normal estándar, los valores de rechazo de la hipótesis nula, al 5% de significancia son 1,96 y -1,96.

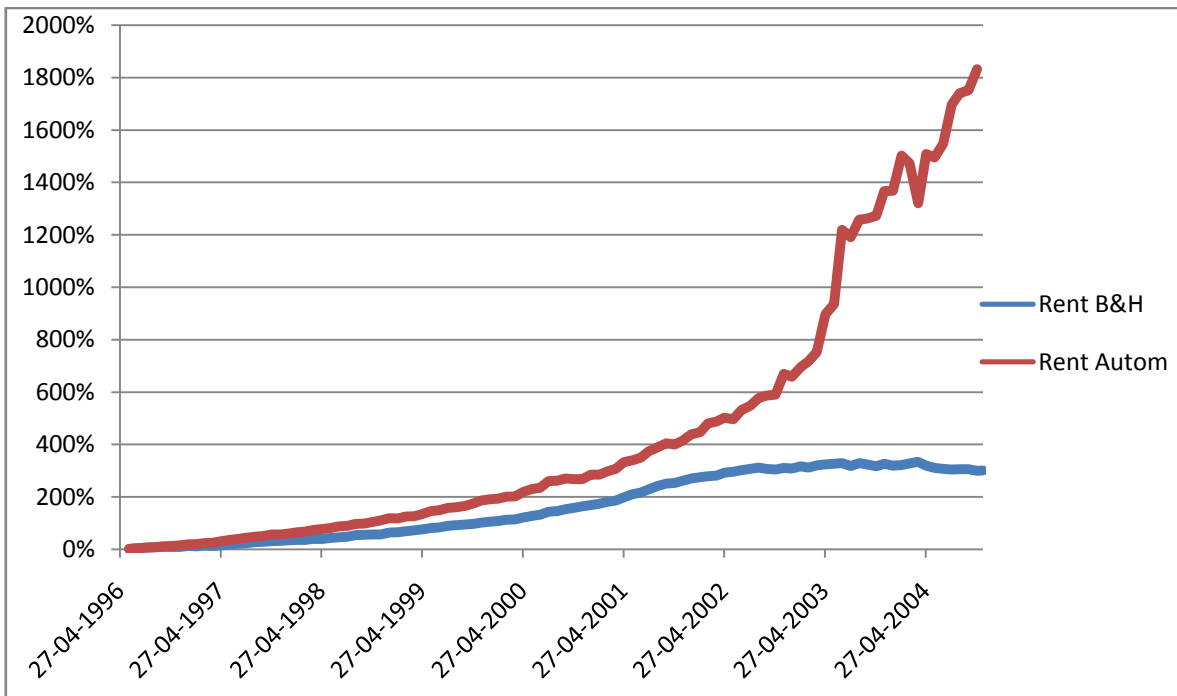
Finalmente los resultados del test para nuestro mejor modelo, son bastante alentadores, rechazando la hipótesis nula para el 81,68% de los casos, es decir, en la gran mayoría de los Zip Codes que se trato de modelar, existe una capacidad predictiva estadísticamente significativa.

#### IV.B Análisis de Rentabilidad

Una completa interpretación de los resultados de este tipo de estudios, requiere hacer un análisis de rentabilidad de seguir la estrategia recomendada por el modelo. Para realizar el análisis de rentabilidad, nosotros asumimos que disponíamos de un capital inicial de US\$100.000, y que lo vamos a invertir en un portfolio de igual ponderación, para cada uno de los Zip Codes, Eso quiere decir, que disponemos de US\$763,85 para invertir en cada área. Luego, asumimos que las predicciones generadas por nuestro modelo, son recomendaciones de compra o de venta. Es decir si el modelo predice un alza se considera como una recomendación de compra, en caso de predecir bajas, se asume como recomendación de venta.

A partir de esto se diseñó una estrategia similar a la que se hace, cuando se trabaja con acciones, y se consideró como *benchmark*, una estrategia pasiva, o *Buy & Hold*, que consiste en hacer la inversión en el primer período y luego mantenerla hasta el último período en que todo es liquidado. Los resultados de ambas estrategias se muestran a continuación en la Figura 4.

FIGURA 4



FUENTE: Elaboración Propia

Como podemos observar en el gráfico anterior, la rentabilidad de la estrategia pasiva es de un 300%, lo que muestra el fuerte crecimiento que han experimentado los precios de los bienes raíces en el estado de Florida durante esta década, principalmente por el desarrollo de nuevas áreas urbanas, con propiedades de alto valor, en sectores que hace 10 años eran principalmente pantanos, de todas formas, la crisis por la que estamos atravesando actualmente, ha reducido las rentabilidades en este sector. Este hecho ya se alcanza a observar en la serie, con la que desarrollamos este estudio, pero la situación ha sido más drástica aún desde el último trimestre del año 2008 hasta la fecha, período para el que no poseemos datos.

De todas formas la estrategia propuesta por nosotros alcanza una rentabilidad de 1905%, mucho mayor que la registrada por el Buy & Hold (aproximadamente 6 veces), y si bien este resultado es sumamente alentador, debe ser mirado con mucho cuidado y con ojos críticos, principalmente por 2 razones. Al asumir que la liquidación de los activos, es similar a la de las acciones, nosotros pasamos por alto dos características muy importantes de este mercado. Primero, los costos de transacción para este tipo de operaciones son mucho más altos, que para el mercado de las acciones. Lo anterior, sin lugar a dudas mermaría nuestra rentabilidad en un porcentaje significativo. Lamentablemente, dado que no disponíamos de ninguna estimación confiable de estos costos, no los incluimos dentro de nuestro análisis, y decidimos hacer mención a este problema.



En segundo lugar, y quizás el mayor problema de este análisis es que no considera la iliquidez del mercado inmobiliario, es decir, nosotros asumimos que al momento de poner en venta los activos, ya disponemos del dinero, para hacer nuevas transacciones, similarmente a lo que ocurre con las acciones, pero sabemos que no es así. Las propiedades requieren estar un tiempo a la venta, para recibir ofertas y finalmente encontrar la mejor. Se podría decir que el proceso de subasta *Wallrasiana* es más lento en este mercado por su falta de liquidez. Por lo mismo, existen períodos de transacción más largos, que no permiten disponer del dinero. Es más incluso cuando un trato es cerrado, hay un plazo en que se demora el propietario en recibir el dinero por parte de quien compra la propiedad. Este hecho sin lugar a duda también reduciría nuestra rentabilidad de manera significativa. De todas formas tenemos la seguridad de que dados los PPS alcanzados, al momento de hacer el ajuste necesario para considerar los puntos antes expuestos, debiéramos mantener una rentabilidad superior a la del Buy & Hold.

## V. CONCLUSIONES

Los resultados encontrados para la predicción de dirección de movimientos de precios para el mercado inmobiliario, mediante el uso de los Autómatas Celulares y la Norma de Mahalanobis, fueron satisfactorios. Tanto en PPS como en Rentabilidad. Lo anterior, se mejoró mediante la introducción de Algoritmos Genéticos, que logró perfeccionar la estructura de los modelos, para mejorar los resultados.

De todas formas lo anterior nos llevó a un resultado poco intuitivo, en el modelo de mejores resultados no estaba incluida la ubicación geográfica, más bien dicho, el contagio estaba dado principalmente por características sociales y económicas, como el decil de ingresos o la edad promedio de las personas que vivían en esa zona. Lo anterior se puede deber a dos razones, según lo creemos nosotros. La primera, y quizás la más obvia, es que la medida de ubicación escogida, simplemente no sea buena, ya que las variaciones parecían muy pequeñas, este problema debiera haberse solucionado por el uso de la Distancia de Mahalanobis, en lugar de una Distancia Euclidiana, pero es posible que aún persista. La segunda causa, puede deberse a que no establecimos una vecindad restringida para nuestras células, lo anterior, tiene como consecuencia, que la influencia que ejercen sus vecinas por separado, sea muy dispersa, y que en ese proceso se pierda información relevante para el contagio.

El mercado inmobiliario presenta grandes diferencias con respecto al mercado accionario, principal foco de aplicación de modelos de predicción en finanzas. Por ejemplo, los costos de transacción en el mercado de valores pueden ser supuestos como nulos o insignificantes, situación que en el mercado inmobiliario se podría considerar como un supuesto heroico. El volumen de transacciones es muy bajo en el mercado inmobiliario con respecto a su par ya mencionado, entregando datos menos completos para hacer estimaciones. Finalmente, el periodo de tiempo que toma llevar a cabo una transacción promedio de alguna vivienda es significativamente mayor que un proceso de compra y venta de acciones, que son prácticamente automáticas.

Finalmente, son estos y otros aspectos relativos al mercado lo que nos llama a ser cautelosos con los resultados, que a su vez, otorga más credibilidad a los mismos, donde un PPS promedio de 64,28% y una rentabilidad de un 1905% deja de ser una tasa astronómica. Hay que considerar, que teniendo datos del año 2000 al 2008, estos consideran el boom y el colapso del mercado de bienes raíces, y que aún así, una estrategia de *Buy & Hold* logra una rentabilidad del 300%. Además los resultados son estadísticamente significativos para sobre el 80% de los Zip Codes cuyos precios se trato de predecir, según el test de Pesaran y Timmerman, dejando el azar de lado.

En conclusión, este documento muestra que el contagio de las variaciones de precio se da por razones más allá de la ubicación geográfica, y que estas variaciones pueden ser predichas a un nivel satisfactorio. Es decir, el autómata propuesto basado en la norma de Mahalanobis como regla de distancia resulta satisfactorio. El perfeccionamiento del modelo, considerando las

características propias del mercado del *Real Estate*, o definiendo vecindades restringidas para las células con el fin de mejorar la influencia del contagio, queda para investigaciones posteriores.

## REFERENCIAS

- Abraham, J. & Hendershott, P. (1996) "Bubbles in Metropolitan Housing Markets", *Journal of Housing Research* 7, 191-207.
- Amihud, Y. & Mendelson, H. (1986) "Asset Pricing and the Bid-Ask Spread", *JFE* vol 17, December 1986, p. 223-249
- Capozza, Hendershott, Mack, Mayer (2002) "Determinants of Real House Price Dynamics", NBER Working Paper 9262, October 2002.
- Chari, Christiano, Kehoe (2008) "Facts and Myths About the Financial Crisis of 2008" Federal Reserve Bank of Minneapolis, Working Paper n° 666, October 2008.
- Feldstein, M. (2007) "Housing, Housing Finance and Monetary Policy". Texto basado en la conferencia del 1 de Septiembre de 2007 del Banco de la Reserva Federal de Kansas City. September 13, 2007
- Kagalis, Peterson, Eberlein, Fiddaman (2007) "The Radar Logic Daily Index", Radar Logic Inc, Ventana Systems Inc. March 28, 2007
- Mahalanobis (1936) "On the Generalized Distance in Statistics", *Proc. Nat. Inst. Of Science of India* vol. 2, 1936
- Malamud & Turcotte (2000) "Cellular-automata models applied to natural hazards", *IEEE Computing in Science and Engineering*, v. 2, no. 3, p. 42-51.
- Parisi, A., F. Parisi & E. Cornejo (2004). "Algoritmos genéticos y modelos multivariados recursivos en la predicción de índices bursátiles de América del Norte: IPC, TSE, Nasdaq y DJI". *El Trimestre Económico*, N°284, vol. LXXI (4), p.p. 789-809. México.
- Pesarann M.H. & Timmermann, A. (1992) "A Simple Nonparametric Test of Predictive Performance", *Journal of Business and Economic Statistics* 10, 461-465.
- Wolfram, S. (2002). "A New Kind of Science". Wolfram Media, Inc., 1192 pp.
- Zhou, Zhou, Wang, Tang & Liu (2004) "Modeling Stock Market Based on Genetic Cellular Automata", *International Journal of Modern Physics*, October 2007