Modelo de Detección de Fraude por Sobrevaluación del Valor de la Vivienda

Julieta Alemán H.
Javier Gutiérrez G.
Gabriel Gómez R.

Sociedad Hipotecaria Federal, S.N.C.
22 de enero de 2008

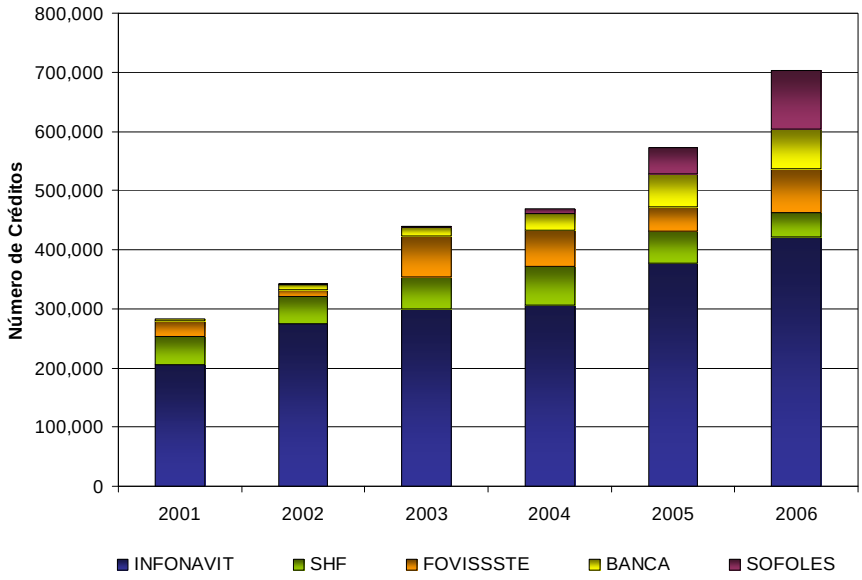
SOCIEDAD HIPOTECARIA FEDERAL

Las opiniones expresadas son únicamente responsabilidad de los autores y no necesariamente representan el punto de vista de la Sociedad Hipotecaria Federal.

1. Introducción

Durante los últimos años el mercado hipotecario en México ha experimentado un crecimiento sin precedentes, esto después de haberse colapsado durante la crisis de 1995. A partir de 2001, las nuevas condiciones macroeconómicas, bajas tasas de interés y un entorno cambiario e inflacionario estables, sentaron las bases para la reactivación del financiamiento a la vivienda tanto público como privado. De 2001 a 2006 el número de créditos otorgados tanto por los organismos públicos como por la banca privada creció más de 100 por ciento[1].

Gráfica 1
Número de Créditos Hipotecarios Originados



Fuente: CONAVI.

Este *boom* dentro del mercado de la vivienda ha creado una serie de presiones, principalmente, sobre los desarrolladores de vivienda y los intermediarios financieros quienes se enfrentan al reto de mantener tasas de crecimiento elevadas dentro de un mercado cada vez más competido. En el caso de Estados Unidos lo que se hizo para aminorar estas presiones fue relajar los estándares de originación de manera abierta, lo que a su vez ha ocasionado la crisis hipotecaria del mercado *subprime*. En México no se ha abierto el mercado a estándares de bajo perfil crediticio como en Estados Unidos; sin embargo, debido a que las presiones persisten se han empezado a detectar una serie de malas prácticas que tienen como objetivo mantener el ritmo de crecimiento del mercado como: 1) mayor dependencia del intermediario financiero con respecto al desarrollador, en particular, los grandes desarrolladores quienes pueden ejercer mayor presión dado el volumen de negocio que llevan a los intermediarios, 2) avalúos inflados con el objetivo de

[1] Fuente: CONAVI.

ofrecer promociones “enganche cero” de forma velada y fraudulenta, 3) fabricación de documentos de comprobación de ingresos.

Actualmente, se cuenta con evidencia puntual sobre este tipo de prácticas; sin embargo, no se puede decir con claridad en que magnitud han permeado al mercado. Lo que si se sabe es que a partir de 2005 se ha observado un deterioro en la cartera crediticia hipotecaria. Este deterioro como ya se mencionó parece tener su origen tanto por el lado de la oferta de vivienda, con los grandes desarrolladores y los intermediarios financieros, así como por el lado de la demanda de crédito. Es decir, existen algunas fallas y problemas de incentivos en ambos mercados que han repercutido en la calidad de la cartera crediticia, aumentando la probabilidad de incumplimiento de la misma.

Por el lado de la oferta de vivienda se han observado malas prácticas, particularmente, por parte de los grandes desarrolladores.[2] La gran mayoría de estos desarrolladores se enfrentan a un problema de incentivos ya que por el hecho de ser empresas que cotizan en bolsa requieren mantener cierto crecimiento de sus ventas para mantener sus utilidades y el precio de su acción. Para estos desarrolladores lo importante es colocar sus viviendas a un buen precio y en un periodo de tiempo relativamente corto de tal forma que las ventas se vean reflejadas en sus estados de resultados. Para ello, el mercado de financiamiento a la vivienda representa el complemento perfecto ya que les permite colocar sus viviendas a través de créditos y recuperar su inversión de manera inmediata sin importar la calidad crediticia del comprador de la vivienda, puesto que ellos no enfrentan los costos de largo plazo de una mala originación. Sin embargo, es importante tener en cuenta que también el intermediario financiero tiene su parte dentro del proceso que da origen a un fraude ya que él es quien se encarga de revisar el proceso de originación de un crédito y el problema radica en que para él también resulta importante captar una gran cantidad de créditos de tal forma que se pueda mantener competitivo dentro del mercado de financiamiento.

De esta forma lo que se observa es que parece existir una colusión entre promotores de grandes desarrolladores e intermediarios financieros. Sin embargo, es importante dejar claro que en la fase en la que se encuentra el mercado estas prácticas no se pueden considerar como políticas generalizadas entre todos los grandes desarrolladores, sino más bien como un problema específico en algunas combinaciones de plaza-desarrollador-intermediario financiero. Debido a esto se considera que todavía es buen momento para corregir este tipo de prácticas sin tener que enfrentar costos equiparables a los observados en Estados Unidos.

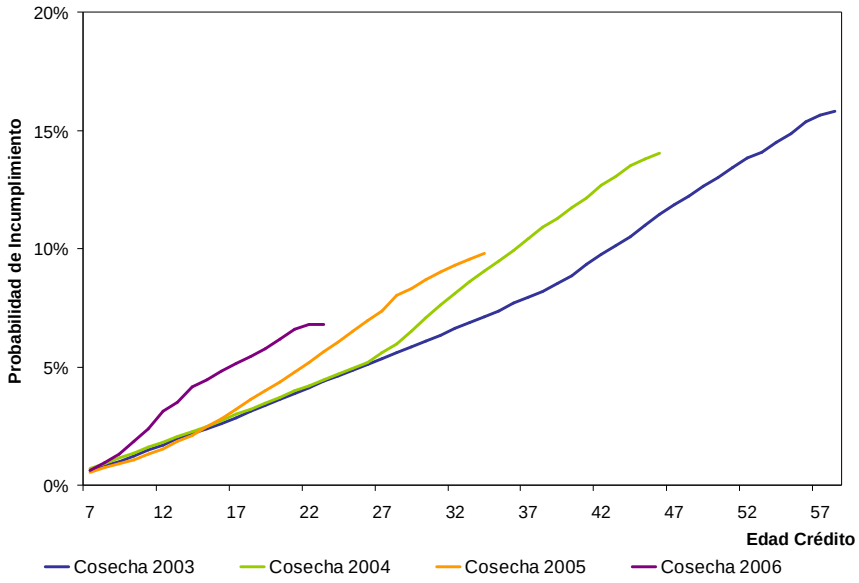
Las siguientes gráficas muestran el deterioro que se ha venido observando, así como el efecto que los grandes desarrolladores han tenido dentro del sector. La Gráfica 2 presenta las curvas de incumplimiento sectoriales por año de originación (cosecha)[3]. La Gráfica 3 muestra las mismas curvas de incumplimiento pero incorpora las curvas de incumplimiento

[2] Los grandes desarrolladores representan cerca de una cuarta parte del sector y están conformados por: ARA, GEO, HOGAR, HOMEX, SARE, URBI. Estos 6 cotizan en bolsa, asimismo, existen otros que no cotizan pero que manejan un volumen de ventas importante como SADASI, VIVEICA, CASAS BETA.

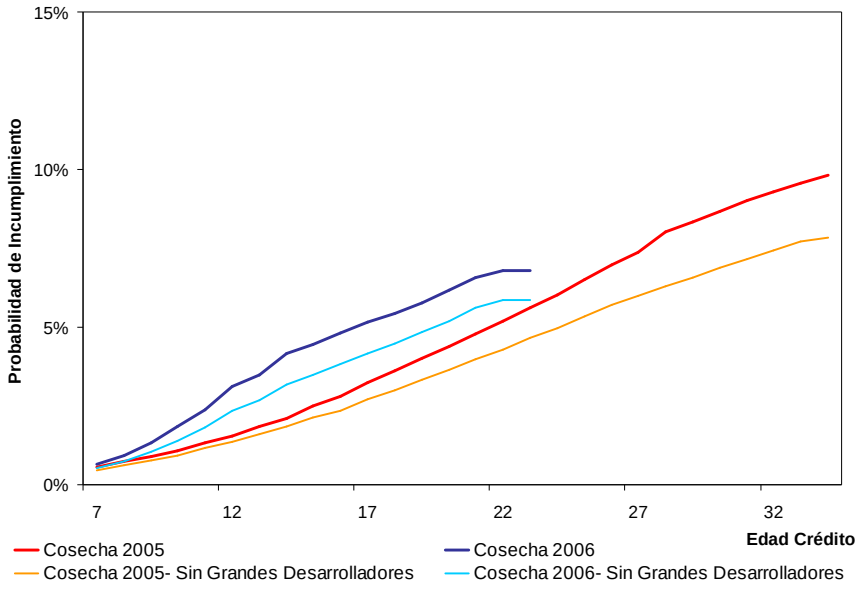
[3] Una curva de incumplimiento muestra la probabilidad de que un crédito incumpla dada una edad determinada.

del sector excluyendo a los grandes desarrolladores. Como se puede apreciar, las curvas que excluyen a los grandes desarrolladores se encuentran por debajo de las que incluyen a todos los participantes del sector, indicando una mejor calidad crediticia. Esto implica que si bien se ha observado un deterioro en la originación de créditos a partir de 2005, éste tiene causas específicas que deben tomarse en cuenta para erradicar el problema.

Gráfica 2
Curvas Incumplimiento Acumuladas Sector



Gráfica 3
Curvas Incumplimiento Sector vs.
Grandes Constructores



El problema del los grandes desarrolladores tiene su contraparte por el lado de la demanda de crédito. En particular, en México, el enganche parece ser una de las mayores restricciones a las que se enfrentan los hogares para poder acceder a un crédito. Esto ocasiona que el promotor de vivienda tenga el incentivo a inflar el valor de la vivienda por un monto equivalente al enganche requerido por las instituciones financieras, de tal forma que el solicitante no tenga que realizar ningún desembolso. Por ejemplo, si se quiere adquirir una vivienda de $150,000 el enganche mínimo es de $15,000. Si el individuo no cuenta con el enganche lo que hace el promotor es coludirse con el perito valuador y elevar el valor de la vivienda a $165,000 de tal forma que el monto total del crédito solicitado sea igual a $150,000 lo que representa el valor total de la vivienda. De esta forma el individuo no realiza ningún desembolso y obtiene un crédito por el monto total del valor de su vivienda. A través de esta operación el individuo obtiene una vivienda y el promotor obtiene una ganancia rápida de la venta sin tener que enfrentar los costos futuros de la mala calidad crediticia del individuo.

Como se puede apreciar, en el caso de México se presenta una combinación de 2 clases de fraudes (Fannie Mae, 2006) que ocasionan que la calidad de la cartera se haya venido deteriorando durante los últimos dos años. Por un lado, se tiene a los grandes desarrolladores junto con los intermediarios financieros para los cuales lo importante es mantener el ritmo de crecimiento del sector. Y por otro lado, se tiene a un solicitante de crédito que desea obtener una vivienda pero que por no cumplir con los requisitos establecidos por las instituciones crediticias como lo son el enganche, un empleo formal, cierto nivel de ingresos, una razón pago-ingreso determinada, entre otros, decide falsificar u omitir información ya que de esta manera se hace sujeto de crédito.

De esta forma la sobrevaluación de las viviendas parece ser el canal a través del cual opera el fraude cometido tanto por los desarrolladores e intermediarios como por los individuos que desean adquirir una vivienda. Al elevar el valor de la vivienda se cubre el monto requerido por las instituciones financieras como enganche y de esta forma se otorga el crédito; sin embargo, el problema radica en que el crédito es más riesgoso que aquel en el que *ceteris paribus* si se pago un enganche. En el caso de SHF esta sobrevaluación se ha hecho evidente al realizar avalúos posteriores a la fecha de originación para algunos créditos que reclamaron el pago de la garantía (GPI). En estos casos se observó que el precio de la vivienda es aproximadamente 40[4] por ciento menor al precio de venta.

De acuerdo a esto, el objetivo del presente documento es presentar un modelo de detección de fraudes, particularmente de avalúos inflados, que sirva como una herramienta dentro del análisis de originación de la cartera crediticia. Es importante mencionar que este trabajo representa una primera aproximación al problema y que su principal interés es promover la discusión académica de tal forma que se pueda contribuir a erradicar las malas prácticas observadas dentro del mercado de la vivienda. La metodología propuesta consiste en utilizar de manera conjunta un modelo que estime el valor de la vivienda a partir de las características de la misma; y un modelo que calcule la dispersión del valor de la vivienda respecto al valor promedio dentro del desarrollo al que pertenece. El primer modelo compara el valor estimado contra el valor del avalúo y determina si el error de estimación se encuentra dentro de un intervalo "estándar" definido dentro del mismo modelo. Del mismo modo el segundo modelo determina si la dispersión en el valor de la vivienda se encuentra dentro de un intervalo "estándar", que una vez más, es definido dentro del mismo modelo.

El presente documento se encuentra organizado de la siguiente manera: en la sección 2 se presenta el marco teórico del modelo de detección de fraudes propuesto, en donde se muestran los supuestos y las metodologías propuestas para detectar un fraude; en la sección 3 se describe la base de datos utilizada y se presentan las estimaciones del modelo, así como los criterios propuestos para determinar si un crédito puede considerarse como "sospechoso" de fraude; en la sección 4 se presentan algunas aplicaciones del modelo que tiene como principal objetivo mostrar la efectividad del mismo, y finalmente la sección 5 presenta algunas conclusiones y sugiere posibles líneas de investigación.

2. Marco Teórico

En esta sección se presenta la metodología propuesta para detectar avalúos inflados. En la primera parte se presenta el modelo de regresión hedónica, basado en las características cualitativas y constitutivas de las viviendas. En la segunda parte se presenta el modelo de dispersión cuyo objetivo es determinar si el precio de las viviendas se encuentra dentro de un intervalo "estándar" respecto al promedio de las viviendas que pertenecen al mismo desarrollo.

[4] Dato obtenido con información de una muestra de 18 créditos de SHF que reclamaron el pago de la garantía y para los cuales se detectó alguna anomalía en el proceso de originación.

2.1 Estimación del Valor de la Vivienda a partir del Modelo de Regresión Hedónica

Para estimar el valor de la vivienda dentro del primer modelo se utiliza la "regresión hedónica", la cual representa una metodología utilizada comúnmente para ajustar los índices de precios o dentro del mercado de bienes raíces para realizar avalúos. La regresión hedónica descompone el bien para el cual se quiere calcular el precio en todas sus características constitutivas y estima el valor de cada una éstas.

$$P = f(x) \qquad (2.1)$$

en donde, P es el precio de la vivienda y,

$$x = (x_1, x_2, ..., x_k) \qquad (2.2)$$

representa el vector de características de la vivienda.

A través de la regresión hedónica se obtienen los precios implícitos:

$$\frac{\partial p}{\partial x_k}(x) = \frac{\partial f}{\partial x_k}(x) \qquad k = (1, ..., k) \qquad (2.3)$$

Es decir, cuanto cambia el precio *p* de un bien, ceteris paribus, si este bien es dotado de una unidad adicional de x_k.

El elemento fundamental de la regresión hedónica es que controla el efecto de los cambios en la calidad de los productos sobre su precio partiendo del supuesto de que existe un mercado para cada característica. Waugh (1928) y Court (1939) fueron los primeros en utilizar esta metodología para mostrar la relación existente entre el precio de un bien y las características del mismo. A partir de la década de los sesentas gracias a los trabajos de Zvi Griliches (1961), Lancaster (1971) y Rosen (1974) se estableció la relación teórica entre la función hedónica y las funciones de utilidad y producción estandarizándose así su uso en el ajuste de los índices de precios. Más tarde en la década de los ochentas Blackley, (1986) y Thibodeau, T.G. (1989) implementaron esta metodología para la estimación de los precios de vivienda.

En el caso de este trabajo se utiliza el modelo de regresión hedónica para estimar el valor de la vivienda a partir de sus características. La forma funcional elegida fue la de doble logaritmo, ya que de esta forma el valor estimado de los coeficientes puede ser interpretado como la elasticidad entre la variable independiente y el valor de la vivienda, además, de esta forma nos aseguramos de que el precio estimado siempre sea positivo. En la siguiente ecuación se muestra la especificación utilizada:

$$p_i = \beta_0 \cdot \prod_{k=1}^{K} x_{i,k}^{\beta_k} \cdot \prod_{j=1}^{J} \exp\left(\gamma_j \cdot d_{i,j}\right) \cdot \exp\left(\varepsilon_i\right) \tag{2.4}$$

o bien,

$$\ln p_i = \ln \beta_0 + \sum_{k=1}^{K} \beta_k \ln x_{i,k} + \sum_{j=1}^{J} \gamma_j \cdot d_{i,j} + \varepsilon_i \tag{2.5}$$

con precios hedónicos,

$$\frac{\partial p}{\partial x_k} = \frac{\beta_k}{x_k} p \tag{2.6}$$

en donde,

p_i es el valor de la vivienda i,
$x_{i,k}$ es el valor de la característica k para la vivienda i
$d_{i,j}$ es el valor de la variable dicotómica j correspondiente a la vivienda i
β_0, β_k y γ_j son los coeficientes estimados de la regresión,
ε_i es el error de estimación el cual tiene una distribución normal con media cero y varianza σ^2.

Los coeficientes se obtienen mediante la aplicación de mínimos cuadrados a la ecuación (2.5).

El valor de vivienda estimado es [5]

$$\hat{p}_i = \beta_0 \cdot \prod_{k=1}^{K} x_{i,k}^{\beta_k} \cdot \prod_{j=1}^{J} \exp\left(\gamma_j \cdot d_{i,j}\right) \cdot E\left[\exp\left(\varepsilon_i\right)\right] \tag{2.7}$$

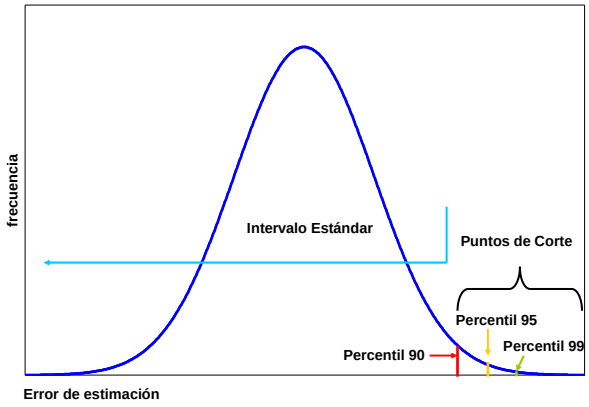
Una vez que se calcula el valor estimado de la vivienda a partir de sus características constitutivas y cualitativas se calcula el error de estimación respecto al valor del avalúo en términos porcentuales.

$$ErrorEstimación = \frac{ValorAvalúo - ValorEstimado}{ValorEstimado} * 100 \tag{2.8}$$

[5] Dado que la estimación se realiza en logaritmos, para regresar la variable a niveles a partir de los coeficientes estimados es necesario incorporar el siguiente factor de ajuste, $E\left[\exp\left(\varepsilon_i\right)\right] = \exp\left(\frac{\sigma^2}{2}\right)$, ya que los errores se distribuyen log normal.

Ahora bien, para establecer un criterio que permita conocer si la diferencia entre el valor obtenido a partir del modelo y el valor del avalúo se debe a sobrevaluaciones o es un error estadístico es necesario conocer la distribución del error. Para efectos de la metodología de detección de fraudes se decidió establecer puntos de corte basados en los percentiles de la distribución del error que sirvan para identificar comportamientos anormales que deben ser sujetos a revisiones más exhaustivas.

Gráfica 4
Distribución del Error de Estimación: Puntos de Corte (percentiles)



La elección de los puntos de corte va a depender de la capacidad operativa de las áreas de negocio, es decir, de aquellas que realizan el análisis de los expedientes para otorgar un crédito. De la gráfica anterior se puede apreciar que si se toma, por ejemplo, el percentil 90 como punto de corte todos los créditos que presentan un error de estimación por arriba de este percentil se considerarían como "sospechosos" y sujetos a una revisión más exhaustiva.

2.3 Modelo de Dispersión del Valor de la Vivienda Intra-Desarrollo

En esta sección se presenta la metodología utilizada para medir la dispersión del valor de la vivienda dentro de un mismo desarrollo. La motivación detrás de este modelo es que el valor de la vivienda controlando por la superficie de la misma no debe presentar grandes variaciones dentro del mismo conjunto habitacional, ya que es de esperarse que aún cuando no todas las casas sean iguales, éstas pertenezcan al mismo tipo, por ejemplo, que dentro de un mismo conjunto todas sean viviendas de interés social. El supuesto del modelo de que exista una baja dispersión en el valor de la vivienda intra-desarrollo parece razonable y congruente si se considera que es poco probable que dentro de un mismo desarrollo se construya una vivienda de tipo residencial junto a una de tipo económico.

Para calcular la dispersión del valor de la vivienda se debe controlar esencialmente por dos variables: la superficie construida y la superficie del terreno; ya que si se parte del supuesto de que el resto de las características cualitativas de las viviendas dentro del mismo desarrollo son muy similares entre si éstas dos variables son las que podrían ocasionar variaciones en el valor de la vivienda.

De acuerdo a esto se construyeron dos medidas de dispersión. La primera variable toma en cuenta el tamaño de la vivienda y corresponde al valor del metro cuadrado de la misma:

$$PVivienda_{m^2} = \frac{ValorAvalúo}{SuperficieConstruida} \tag{2.9}$$

La segunda variable controla el costo de la vivienda por su tamaño y por el tamaño del terreno y corresponde al costo del metro cuadrado de construcción más el costo del metro cuadrado del terreno:

$$PFísico_{m^2} = \frac{ValorConstrucción}{SuperficieConstruida} + \frac{ValorTerreno}{SuperficieTerreno} \tag{2.10}$$

Para calcular la dispersión se construyen las siguientes variables:

$$desv_PVivienda_{m^2} = PVivienda_{m^2} - \overline{PVivienda_{m^2}} \tag{2.11}$$

y

$$desv_PFísico_{m^2} = PFísico_{m^2} - \overline{PFísico_{m^2}} \tag{2.12}$$

en donde,

$\overline{PVivienda_{m^2}}$ [6]= valor promedio del precio de la vivienda por metro cuadrado dentro del mismo desarrollo.

$\overline{PFísico_{m^2}}$ [1]= valor promedio del valor físico de la vivienda por metro cuadrado dentro del mismo desarrollo.

Una vez que se tienen las desviaciones respecto al valor promedio del desarrollo se calcula el porcentaje que representa dicha desviación del valor promedio de la vivienda. Posteriormente, al igual que con el modelo de estimación se obtiene la distribución para cada una de las medidas de desviación construidas en términos porcentuales. Esto con el objeto de conocer el comportamiento y establecer puntos de corte que sean utilizados como criterios de decisión para determinar que créditos deben ser sujetos de una revisión más exhaustiva de acuerdo al modelo de dispersión.

[6] Para calcular el valor promedio se toman todas aquellas viviendas que pertenecen a un desarrollo con al menos 50 casas.

3. Estimación Modelo de Detección de Fraudes

En esta sección se estiman los modelos presentados en la sección anterior. Para ello, primero se presenta una descripción de la base de datos utilizada, posteriormente, se presentan los resultados de las estimaciones tanto del modelo de estimación del valor de la vivienda como del modelo de dispersión. Finalmente, se establecen los criterios de decisión para determinar si un crédito puede considerarse como "sospechoso" y ser sujeto de una revisión más exhaustiva.

3.1 Datos

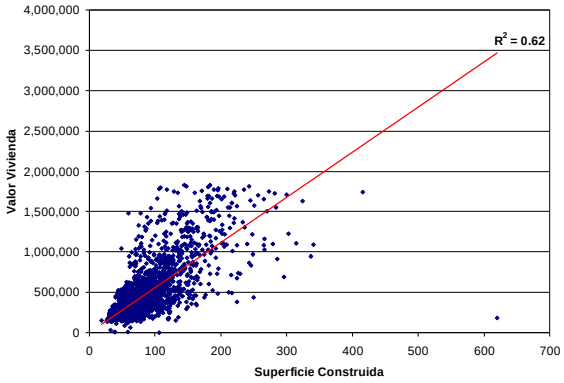
Para la estimación de los modelos se utilizó la base de avalúos que administra SHF, la cual cuenta con información de todo el sector. Dicha base empezó a recolectarse a partir de febrero de 2005 y a la fecha cuenta con más de 1,300,000 registros. Para efectos de este trabajo no se utilizó la muestra completa sino que se extrajo una muestra aleatoria de 250,000 registros. Asimismo, cabe mencionar que dentro de la muestra se impuso la restricción de que el valor de la vivienda no superara las 500,000 udis, ya que éste representa el monto máximo para el cual SHF otorga garantías. La base está conformada por 37 variables las cuales proveen información sobre la identificación, la ubicación geográfica, las características cualitativas y el valor de la vivienda (ver Anexo A).

En particular, para el modelo de detección de fraudes se utilizaron las siguientes variables: el valor comparativo de mercado que es equivalente al valor del avalúo, la superficie construida, la superficie del terreno, el valor físico construcción, el valor físico terreno, la clase de inmueble y la proximidad urbana, éstas dos últimas son variables categóricas.

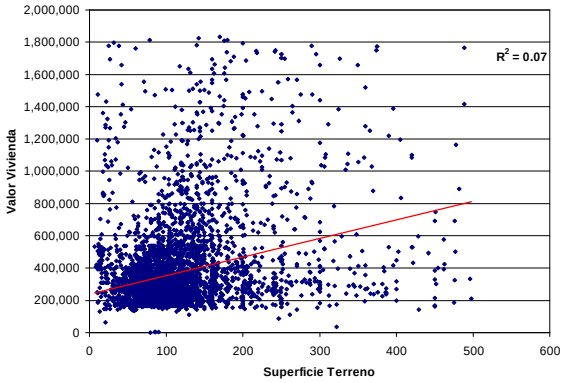
En el caso de las variables categóricas para poder introducirlas dentro del modelo de estimación del valor de la vivienda es necesario realizar una codificación con variables dicotómicas. Dicha codificación se explica a detalle en el Anexo B.

En las siguientes gráficas se presenta la relación que existe entre las variables utilizadas en el modelo y el valor de la vivienda.

Gráfica 5
Valor de la Vivienda vs.
Superficie Construida

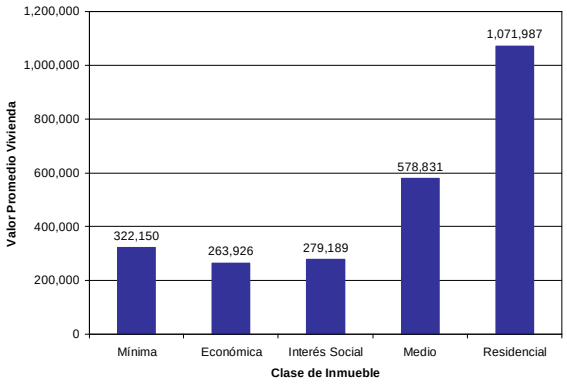


Gráfica 6
Valor de la Vivienda vs.
Superficie Terreno

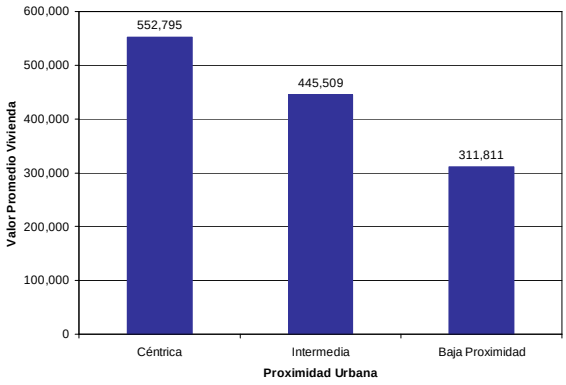


Como se puede apreciar la superficie construida guarda una estrecha relación con el valor de la vivienda, mientras que el tamaño del terreno no parece ser una determinante del mismo. Como se verá más adelante esta variable no resulta ser estadísticamente significativa dentro del modelo de estimación del valor de la vivienda por lo que sólo se utiliza dentro del modelo de dispersión.

Gráfica 7
Valor de la Vivienda Promedio vs.
Clase de Inmueble



Gráfica 8
Valor de la Vivienda Promedio vs.
Proximidad Urbana



En el caso de las variables categóricas se observa que el valor de la vivienda aumenta conforme mejora la clase de vivienda, así como, mientras mejor es la ubicación geográfica respecto a los centros urbanos.

3.2 Estimación del Valor de la Vivienda a través de la Regresión Hedónica

Como se explicó en la sección 2 el primer modelo propuesto consiste en estimar el valor de la vivienda a partir de sus características cualitativas utilizando la metodología de la

regresión hedónica. A continuación se muestra la especificación de la regresión hedónica y la forma como se determinan los puntos de corte correspondientes.

3.2.1 Especificación de la Regresión Hedónica

La siguiente ecuación presenta la especificación propuesta para la estimación[7].

$$\ln(ValorVivienda) = \ln \beta_0 + \beta_1 \ln(SuperficieConstruida) + \gamma_1 ClaseInmueble + \gamma_2 \Pr oximidadUrbana + error$$

Ahora bien, debido a la evidencia mostrada a través del desempeño de las curvas de incumplimiento sobre el comportamiento de los grandes desarrolladores se propone estimar una segunda especificación en la que se controle por el efecto de los grandes desarrolladores con el propósito de determinar si tal efecto es estadísticamente significativo. Para ello, se incorpora una variable dicotómica[8] adicional a la especificación antes mencionada con el fin de capturar dicho efecto. La siguiente ecuación muestra esta especificación.

$$\ln(ValorVivienda) = \ln \beta_0 + \beta_1 \ln(SuperficieConstruida) + \gamma_1 ClaseInmueble + \gamma_2 \Pr oximidadUrbana + \gamma_7 GD$$

en donde, *GD* es la variable dicotómica para identificar a los grandes desarrolladores.

La estimación se realizó utilizando Mínimos Cuadrados Ordinarios (MCO) con la corrección propuesta por Newey-West (1987) para la matriz de varianza-covarianza ante problemas de autocorrelación y heteroscedasticidad[9]. Los resultados se presentan a continuación.

Regresión Hedónica: Estimación del Valor de la Vivienda

Muestra: 250,000
Variable dependiente: Ln(ValorVivienda)

	Constante	lnSupCons	DCéntrica	DIntermedia	DMínima	DEconómica	DInterésSocial	Medio	GD	R^2 ajustada
Estimación I	8.642***	0.979***								0.646
	(0.017)	(0.004)								
Estimación II	9.408***	0.819***	0.062***	-0.041***	-0.154***	-0.238***	-0.141***	0.099***		0.688
(variables categóricas)	(0.025)	(0.006)	(0.005)	(0.003)	(0.008)	(0.005)	(0.004)	(0.005)		
Estimación III	9.293***	0.839***	0.059***	-0.038***	-0.152***	-0.231***	-0.136***	0.087***	0.143***	0.704
(controlando por GD)	(0.024)	(0.005)	(0.005)	(0.003)	(0.008)	(0.005)	(0.004)	(0.005)	(0.005)	

*** Estadísticamente significativo al 1%.
Los datos entre paréntesis representan las desviaciones estándar.

[7] Las variables categóricas Clase de Inmueble y Proximidad Urbana se introducen de acuerdo a la codificación dicotómica detallada en el Anexo B.
[8] Para esta variable la codificación es 0,1.
[9] La muestra presenta heteroscedasticidad. Para corregir este problema se probaron otras especificaciones, así como la corrección por Mínimos Cuadrados Ponderados, sin embargo, debido a que no hubo una mejora satisfactoria se optó por incorporar la corrección propuesta por Newey-West a la matriz de varianza-covarianza.

En la primera especificación, únicamente se incorporó la superficie de construcción como variable independiente, con el objetivo de medir su importancia como variable explicativa.

Debido a que las variables se encuentran en logaritmos, los coeficientes de las mismas representa elasticidades respecto al valor de la vivienda, es decir, en esta primera estimación la superficie de construcción presenta una elasticidad de 0.97% con el valor de la vivienda. La R^2 ajustada bajo esta especificación es de 65%.

En la segunda especificación se incorporan las variables categóricas, clase de inmueble y proximidad urbana. Para esta especificación todas las variables resultaron estadísticamente significativas, nuevamente, la superficie construida es la que explica en mayor medida el valor de la vivienda. Asimismo, las variables categóricas presentan los signos esperados, por ejemplo, céntrica presenta un signo positivo mientras que intermedia presenta un signo negativo lo cual era de esperarse ya que ante una mejor ubicación aumenta el valor de la vivienda. Por otro lado, la R^2 ajustada del modelo aumenta a 69%.

Finalmente, se tiene que en la tercera especificación la variable dicotómica que controla el efecto de sobrevaluación de los grandes desarrolladores resultó ser significativa y la elasticidad entre esta variable y el valor de la vivienda es igual a 0.14. El resto de las variables independientes presentan coeficientes similares a los de la especificación II, manteniendo su significancia y los signos esperados. Asimismo, el ajuste de la especificación mejora ligeramente a 70%. Esta última especificación es la que se utilizará para el modelo de detección de fraudes.

Es importante mencionar que se probaron otras variables dentro del modelo tales como: la superficie del terreno, la vida útil remanente de las viviendas y el porcentaje de infraestructura, pero ninguna de éstas mejoró significativamente el ajuste de la especificación por lo que se decidió no incorporarlas para mantener el modelo simple.

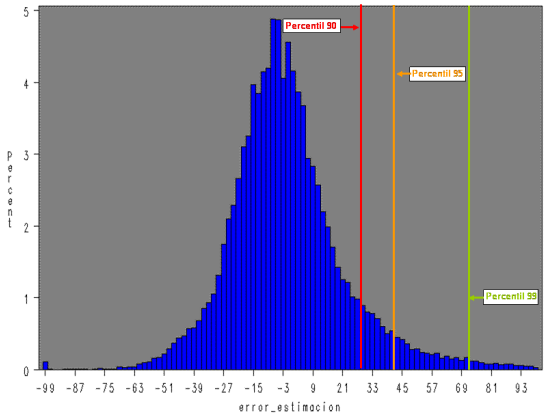
Asimismo, la base de avalúos actual no cuenta con información más específica de las viviendas tal como el número de habitaciones, el número de baños, los lugares de estacionamientos, entre otras, las cuales podrían resultar útiles para mejorar el ajuste de la especificación al representar información que intuitivamente debe de impactar en el valor de una vivienda. Se espera que en un futuro cercano la base de avalúos cuente con dichas variables.

3.2.2 Distribución del Error de Estimación y Puntos de Corte

Para efectos del modelo de detección de fraudes la estimación del valor de la vivienda a través de la regresión hedónica nos ayuda a determinar en que grado el valor estimado difiere del valor del avalúo que en este caso es igual al valor de la vivienda. De acuerdo a esto se calcula el error de estimación con base en la ecuación (2.8).

En la siguiente gráfica se presenta la distribución del error de estimación para la muestra utilizada en el modelo.

Gráfica 9
Distribución del Error de Estimación



Cuantiles	
Cuantil	Estimado
100% Max	99.6
99%	73.5
95%	41.1
90%	27.1
75% Q3	9.3
50% Mediana	-4.0
25% Q1	-15.6
10%	-26.6
5%	-35.0
1%	-52.2
0% Min	-100.0

La distribución del error de estimación presenta un ligero sesgo hacia la derecha (coeficiente de asimetría = 0.65) lo que indica que en general se observan más sobrevaluaciones que subvaluaciones. Como se puede apreciar el 50% de la distribución se encuentra concentrada en valores del error de estimación que van de -16% a 11%. Asimismo, se observa que las sobrevaluaciones mayores a 27% representan el percentil 90, es decir, éstas se podrían considerar observaciones atípicas. En particular, sobrevaluaciones mayores a 41% representan tan sólo el 5% de la distribución, y más aún, sobrevaluaciones mayores a 74% representan menos del 1%.

La distribución presentada en la Gráfica 8 nos permite establecer puntos de corte para establecer un criterio que nos ayude a determinar si alguna sobrevaluación podría considerarse como “sospechosa”. Para efectos de este trabajo las observaciones que se encuentran por arriba del percentil 90 se consideran como atípicas, lo que nos lleva a fijar en dicho percentil un punto de corte y a considerar como observaciones “sospechosas” todas aquellas que presenten un error de estimación mayor a 27%. Asimismo, este número parece congruente con la evidencia recolectada por SHF en la que se detectaron fraudes con sobrevaluaciones de 40% en promedio.

3.3 Estimación de la Dispersión del Valor de la Vivienda Intra-Desarrollo

El modelo de dispersión está basado en el supuesto de que la variabilidad en el valor de las viviendas dentro de un mismo desarrollo debe ser muy pequeña, ya que las características cualitativas tales como la proximidad urbana, el nivel de infraestructura urbana, la calidad de las viviendas debe ser muy similar. Las diferencias radican esencialmente en la superficie de construcción y en el tamaño del terreno, por lo que una vez que se controla por estas variables el valor de la vivienda por metro cuadrado no debe diferir demasiado entre una vivienda y otra.

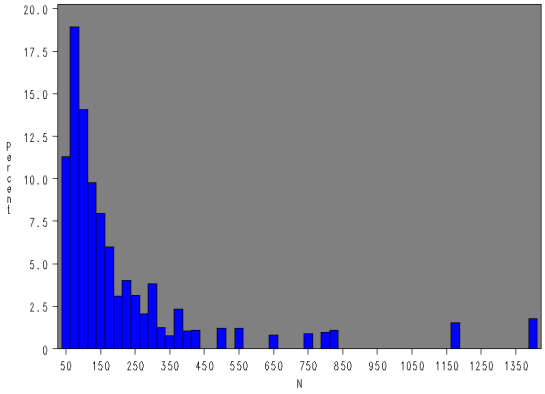
3.3.1 Construcción de las Medidas de Dispersión

Para poder aplicar el modelo es necesario contar con una muestra en la que todas las viviendas pertenezcan a un desarrollo. Asimismo, para garantizar que la comparación entre

el valor de las viviendas sea robusto se establece la restricción de que las viviendas elegidas dentro de la muestra deben pertenecer a desarrollos con al menos 50 viviendas. De acuerdo a estas restricciones la muestra que se utilizó para estimar la dispersión intra-desarrollo consiste en todos aquellos registros que cumplen con estas restricciones dentro de la muestra de 250,000 créditos que se utilizó para el modelo de la regresión hedónica. Al depurar la base para cumplir con las restricciones del modelo se obtuvo una muestra de 75,330 registros de vivienda que se encuentran agrupadas en 706 desarrollos.

En la siguiente gráfica se muestran algunas estadísticas de la muestra utilizada.

Gráfica 10
Distribución de Desarrollos con "N" Número de Viviendas



Cuantiles	
Cuantil	**Estimado**
100% Max	1411
99%	1411
95%	796
90%	419
75% Q3	234
50% Mediana	125
25% Q1	80
10%	61
5%	55
1%	50
0% Min	50

Como se puede observar el 75% de los desarrollos que se encuentran en la muestra tienen entre 50 y 234 viviendas, lo cual representa un número aceptable para realizar los cálculos deseados. Dentro de la muestra, la media del número de viviendas por desarrollo es de 216.

Para calcular las medidas de dispersión se debe realizar una correcta identificación de los desarrollos por lo cual se utiliza el nombre del mismo, el código postal y la clave del estado al que pertenecen. Esto con el fin de evitar errores debido a que existen desarrollos en la muestra con el mismo nombre pero en diferentes estados.

Ahora bien, para calcular las medidas de dispersión se utilizan las expresiones dadas por (2.9) y (2.10). La expresión (2.9) representa el valor de la vivienda por metro cuadrado de construcción. Esta medida refleja en gran proporción el valor de la vivienda ya que como se pudo apreciar en el modelo de la regresión hedónica la superficie de construcción es la variable con mayor poder explicativo del valor de la vivienda. La expresión dada por (2.10) toma en cuenta tanto a la superficie de construcción como a la superficie de terreno. Esta medida es importante ya que se observó que en algunos casos la superficie del terreno si ocasionaba variaciones importantes en el valor de la vivienda por lo que al incorporar esta medida se está controlando por dicho efecto. Posteriormente, se deben de calcular los valores medios de estas medidas para cada desarrollo, y de esta forma se pueden calcular (2.11) y (2.12).

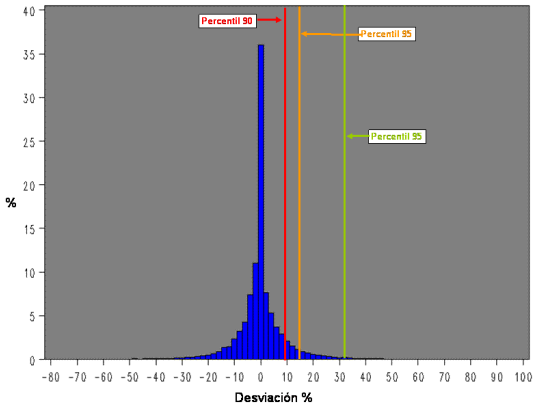
Al contar con las desviaciones tanto del valor de la vivienda y del valor físico de la misma por metro cuadrado se debe calcular cuanto representa dicha desviación en términos porcentuales del valor medio del desarrollo para posteriormente calcular las distribuciones respectivas.

3.3.2 Distribución de las Medidas de Dispersión

Siguiendo el mismo criterio que para el modelo de la regresión hedónica, se busca conocer la distribución de las medidas de dispersión con el objeto de establecer puntos de corte que nos ayuden a determinar si una vivienda presenta una dispersión mayor que el resto de las viviendas del desarrollo al que pertenece.

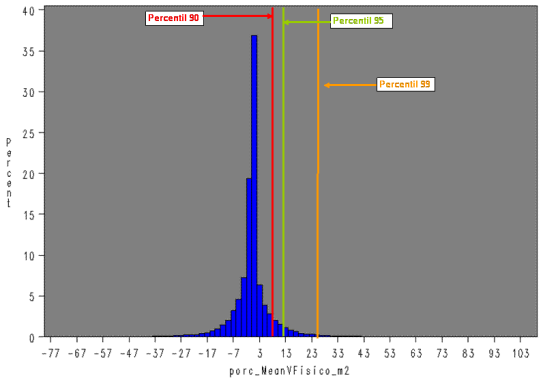
En las siguientes gráficas se presentan las distribuciones de las desviaciones porcentuales tanto del valor de la vivienda por metro cuadrado como del valor físico por metro cuadrado.

Gráfica 11
Distribución de la Desviación Porcentual respecto al Valor de la Vivienda Promedio



Cuantiles	
Cuantil	Estimado
100% Max	100
99%	31
95%	14
90%	9
75% Q3	2
50% Mediana	0
25% Q1	-3
10%	-8
5%	-14
1%	-28
0% Min	-80

Gráfica 12
Distribución de la Desviación Porcentual respecto al Valor Físico Promedio



Cuantiles	
Cuantil	Estimado
100% Max	100
99%	25
95%	11
90%	7
75% Q3	1
50% Mediana	0
25% Q1	-2
10%	-6
5%	-11
1%	-24
0% Min	-77

.o

Como se puede apreciar en ambas distribuciones se observa un alto grado de concentración hacia la media. En el primer caso se tiene que entre el percentil 25 y el 75 el intervalo de las desviaciones va de -3% a 2%; mientras que con la medida del valor físico el mismo intervalo va de -2% a 1%. La forma de estas distribuciones confirma los supuestos del modelo de una baja dispersión en el valor de las viviendas intra-desarrollo. En particular, se tiene que el 95% de la distribución no rebasa desviaciones superiores a 14% en el caso del valor de la vivienda por metro cuadrado y del 11% en el caso del valor físico.

De acuerdo a estos resultados para establecer un criterio de decisión que contribuya a detectar viviendas cuyos avalúos se encuentren por arriba del promedio, se utilizarán los puntos de corte de las distribuciones de las dos medidas de dispersión de manera conjunta. Es decir, todas aquellas desviaciones del valor de la vivienda por metro cuadrado respecto al promedio intra-desarrollo que se sitúen por arriba de 9% junto con aquellas observaciones que también presenten una desviación del valor físico por metro cuadrado respecto al promedio intra-desarollo por arriba del 7%.

La importancia de tomar las dos medidas de dispersión de forma conjunta radica en el hecho de que de esta manera nos aseguramos de elegir viviendas que en caso de presentar un valor mayor en términos relativos al resto del conjunto al que pertenecen no se deba a una mayor superficie de terreno. O bien, que en caso de presentar un desviación del valor físico mayor a la del conjunto, está se refleje en el valor de mercado.

3.4 Regla de Decisión del Modelo de Detección de Fraudes

Antes de establecer la regla de decisión del modelo de detección de fraudes, es importante aclarar que el modelo propuesto tiene como principal objetivo ser una herramienta adicional dentro del análisis de originación de créditos o de garantías, de tal forma que la detección de créditos con el modelo propuesto no quiere decir que *de facto* representen un fraude, sino que por presentar características diferentes por cierto rango a las del promedio de la población deben ser sujetos a una revisión más exhaustiva con el fin de evitar posibles fraudes. Asimismo, cabe mencionar que el modelo se puede utilizar tanto con fines

preventivos (antes de la originación) como con fines de depuración y mantenimiento de la calidad de la cartera en el caso de las garantías, es decir, aquellos créditos que se hayan originado a través de un fraude y que por alguna razón deseen hacer efectiva la garantía no tendrán derecho al cobro de la misma.

Ahora bien, el modelo de detección de fraudes propuesto en este trabajo consiste en utilizar de manera conjunta el modelo de la regresión hedónica y el modelo de dispersión. Como se explicó en las secciones anteriores los criterios de decisión de ambos modelos están basados en la distribución del error de estimación en el caso del primer modelo y en las distribuciones de las desviaciones de las medidas de dispersión respecto al promedio intra-desarrollo en el caso del segundo modelo.

En particular, los criterios de decisión se fijan de acuerdo a los percentiles de las distribuciones antes mencionadas. Dichos percentiles pueden ser elegidos en función de la capacidad operativa de las áreas de negocio. Para efectos de este trabajo se toma como punto de corte los percentiles 90. Esto significa que los créditos que presenten un error de estimación con el modelo hedónico mayor a **27%**, y que a su vez, presenten una desviación del valor de la vivienda por metro cuadrado respecto al promedio intra-desarrollo mayor a **9%** y una desviación del valor físico por metro cuadrado respecto al promedio intra-desarrollo mayor a **7%** deben ser sujetos a una revisión más exhaustiva por parte del área de negocio. Esta regla de decisión es aplicable únicamente a aquellas viviendas que pertenezcan a desarrollos con al menos 50 viviendas.

Para el caso de las viviendas que pertenecen a desarrollos con un menor número de viviendas o que no pertenecen a ningún desarrollo, sólo será posible aplicar el modelo de la regresión hedónica y el criterio de decisión será el mismo que en el caso anterior, es decir, si las viviendas presentan un error de estimación superior al **27%**, entonces serán sujetas a una revisión más exhaustiva.

4. Aplicaciones del Modelo de Detección de Fraudes

El objetivo de esta sección es medir la efectividad y utilidad del modelo de detección de fraudes propuesto. Para ello, se presentan dos ejercicios: el primero consiste determinar la efectividad del modelo contra una "lista negra" de valuadores realizada por SHF; y el segundo, consiste en corroborar estadísticamente la hipótesis de que los grandes desarrolladores tienden a sobrevaluar el valor de la viviendas.

4.1 Efectividad del modelo

La lista negra de valuadores de SHF comenzó a elaborarse hace unos cuantos meses y se encuentra actualmente en una fase preliminar ya que sólo cuenta con información para cuatro estados: Estado de México, Jalisco, Nuevo León y San Luis Potosí. Dentro de estos estados se han detectado a un total de 31 controladores y 37 peritos, los cuales representan aproximadamente el 5% del total de valuadores registrados en el país.

Lista Negra

	Estado de México	Jalisco	Nuevo León	San Luis Potosi	Total
Controladores	10	11	7	3	**31**
Peritos	10	13	12	2	**37**

Para elaborar la lista negra se siguió la técnica de "Mystery Shopping" la cual consiste en visitar desarrollos y fingir que se desea adquirir una vivienda. Para conformar la lista se han visitado desarrollos en los que se han presentado reclamaciones de garantías. Durante el ejercicio de "Mystery Shopping", se acreditó que en dichos desarrollos se ofrecían promociones con el fin de que no se cubriera el enganche y/o los gastos de escrituración. La información recabada fue confirmada con los reportes de avalúos que las Unidades de Valuación (UVs) están obligadas a enviar a SHF y en la mayoría de los casos, con la información relativa al otorgamiento de créditos y/o garantías por parte de SHF. Los valores de los avalúos reportados por las UVs coinciden con exactitud – y en muchos casos son mayores- a las promociones obtenidas de los desarrolladores.

Cabe destacar que en todos los conjuntos visitados, el desarrollador ofreció al "Mystery shopper" inflar el avalúo con el fin de que no se cubrieran los enganches y/o los gastos de escrituración. Dadas las coincidencias – de hecho de valuadores distintos y de UVs distintas- se puede establecer que los valuadores reportaron en los avalúos, valores predeterminados por el desarrollador, en los que no se cubrió el enganche y en muchos casos los gastos de escrituración. Además, SHF realizó un estudio de mercado de las zonas respectivas que permiten confirmar que los avalúos realizados por las personas que integran la lista están inflados.

Para medir la efectividad del modelo respecto a lo observado en la lista negra se utilizó una base de datos de 29,324 registros con créditos que cuentan con la garantía por incumplimiento (GPI) de SHF. De esta base 7,462 registros cumplen con el supuesto de que los créditos pertenecen a desarrollos con al menos 50 viviendas, por lo que sólo a estos registros se les pudo aplicar el modelo de detección de fraudes en su versión conjunta (regresión hedónica y modelo de dispersión). Al resto de los créditos sólo se les aplica el modelo de regresión hedónica. El ejercicio consiste en aplicar el modelo de detección de fraudes a dicha base y determinar cuantos de los valuadores de los créditos identificados por el modelo como "sospechosos" coinciden con una lista negra de valuadores elaborada por la SHF. Es importante tener en cuenta que la lista negra se encuentra actualmente en una fase preliminar, por lo que existen varios estados para los cuales no se pueden realizar comparaciones.

El criterio de decisión utilizado para aplicar tanto el modelo de regresión hedónica como el de dispersión es el percentil 90. En la siguiente tabla se muestran los resultados obtenidos para la muestra de los 7,462 registros.

No. Créditos Detectados por el Modelo de Detección de Fraudes

Muestra: 7462		
Punto de Corte: percentil 90		
Modelo	**No. Créditos**	**% (total)**
Modelo Regresión Hedónica	1,299	17%
Modelo Dispersión	444	6%
Regla de Decisión Conjunta	129	2%

Distribución Créditos "Sospechosos" por Estado

Estado	No. Créditos
Edo. de México	75
Nuevo León	16
Baja California	12
Jalisco	11
Sonora	8
Tamaulipas	4
Morelos	1
Nayarit	1
Sinaloa	1

De los resultados anteriores se puede decir que el modelo de regresión hedónica detecta un mayor número de créditos como "sospechosos". Esto se debe en gran medida a que en dicho modelo la superficie del terreno no resultó estadísticamente significativa por lo que no se controla por cambios en el tamaño del terreno, y en algunos casos, éste si ocasiona variaciones en el valor de la vivienda. Por su parte, el modelo de dispersión detecta 444 créditos, un número casi 3 veces menor al del modelo de la regresión hedónica. Esto puede explicarse por el hecho de que al tomar en cuenta dos medidas el criterio de decisión se vuelve más estricto. Finalmente, la regla de decisión conjunta determina 129 créditos, el 2% de la muestra, y de estos créditos el mayor número se ubica en el Estado de México seguido de Nuevo León y Baja California.

Al comparar los nombres de los valuadores de estos 129 créditos con la lista negra de SHF se obtienen los siguientes resultados:

- El 76.7% de los avalúos detectados por el modelo fueron realizados por un valuador o controlador que se encuentra en la "lista negra".

- De los 23 valuadores detectados 6 se encuentran en la lista negra, tres de ellos en el Estado de México y tres en Nuevo León.

- De los 21 controladores detectados 6 se encuentran en la lista negra. Uno en Jalisco, cuatro en el Estado de México y uno en Nuevo León.

- Otros estados donde el modelo de sobrevaluación detecto avalúos sospechosos fueron: Baja California, Morelos, Nayarit, Sinaloa, Sonora y Tamaulipas.

Como se mencionó la lista negra de SHF sólo abarca 4 estados, por lo que se espera que un alto porcentaje de los avalúos detectados en los otros estados estén efectivamente sobrevaluados.

Si se analizan los créditos detectados para la muestra en la que únicamente se puede utilizar el modelo de regresión hedónica se tiene que de una muestra de 21,912 créditos el modelo

detectó 5,119 registros (23%) como créditos "sospechosos". En la siguiente tabla se muestra la distribución de dichos créditos por estado.

Distribución Créditos "Sospechosos" por Estado

Estado	No. Créditos	Estado	No. Créditos	Estado	No. Créditos
Baja California	1,244	Tamaulipas	83	Tabasco	19
Jalisco	873	Querétaro	77	Campeche	7
Distrito Federal	629	Veracruz	77	Coahuila	6
Edo de México	472	Aguascalientes	56	Colima	5
Nuevo León	408	Puebla	51	Chiapas	5
Sinaloa	211	Morelos	49	Hidalgo	5
Baja California Sur	184	Michoacán	36	Yucatán	5
Chihuahua	180	Guerrero	26	Zacatecas	3
Gunajuato	119	Oaxaca	26	Durango	2
Sonora	112	Nayarit	23		
Quintana Roo	105	San Luis Potosi	21		

Nuevamente, aparecen Baja California, Jalisco, Estado de México y Nuevo León como los principales estados en los que se presentan anomalías en los avalúos. Asimismo, en el caso del modelo de regresión el Distrito Federal también presenta un número importante de créditos "sospechosos".

Al igual que en el caso de la regla conjunta resulta interesante medir la efectividad del modelo por si solo para detectar avalúos inflados, por lo que una vez más se compara la lista de controladores y peritos de estos créditos contra la lista negra elaborada por SHF. De este análisis se obtienen los siguientes resultados:

- El 34% de los avalúos detectados por el modelo de regresión hedónica fueron realizados por un valuador o controlador que se encuentra en la "lista negra". Este número resulta menor a lo observado en el caso del modelo en su versión conjunta, lo cual puede explicarse por el hecho de que Baja California aparece como el estado con mayor número de créditos "sospechosos"; sin embargo, este estado no se encuentra actualmente en la lista negra.

- En el caso del Estado de México y Nuevo León más del 80% de los avalúos detectados fueron realizados por un perito o controlador en la lista negra.

- Para Jalisco este número representó el 47% de los créditos detectados.

- El 15% de los controladores y el 10% de los peritos que se encuentran en la muestra de los 5,119 créditos "sospechosos" pertenecen a la lista negra. Este porcentaje es bajo debido una vez más a que la lista negra sólo incluye cuatro estados.

- Sin embargo, si se realiza el ejercicio inverso y se mide el número de valuadores y peritos que se encuentran en la lista negra y que fueron detectados por el modelo este porcentaje es igual a 80%.

De acuerdo a los resultados presentados, el modelo detecta un porcentaje considerable de los valuadores y peritos que se encuentran en la lista negra, ahora bien, si se toma en cuenta el hecho de que ésta no abarca todos los estados en los que se tienen créditos, es de esperarse que los porcentajes de detección del modelo respecto a la lista negra aumenten.

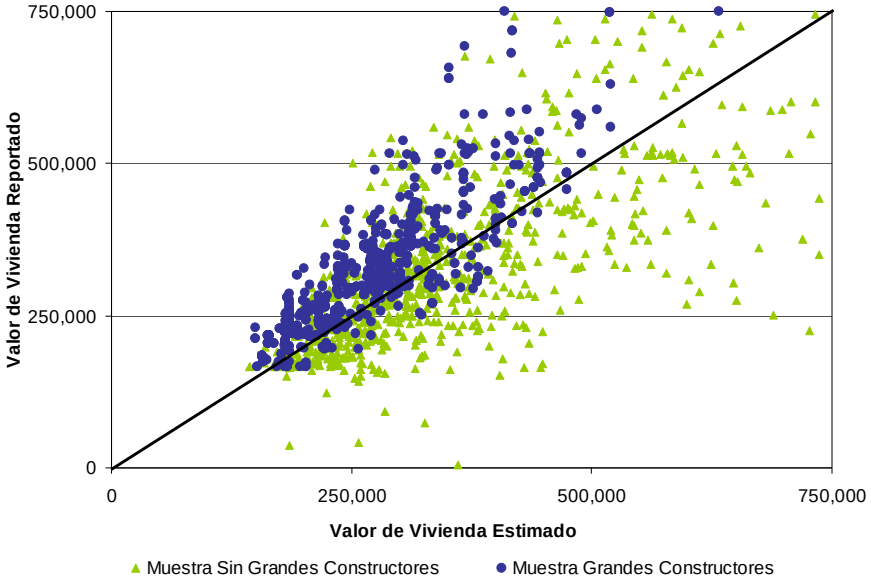
4.2 Información Heurística: El efecto de los Grandes Desarrolladores

Una manera de mejorar el modelo de regresión es incorporando variables con base en información heurística. Si estas variables resultan significativas, no sólo se mejora el modelo sino que además se comprueba la información de manera empírica. Un ejemplo de ésto, es la incorporación de la información de que los grandes desarrolladores tienden a realizar malas prácticas de originación, lo que probablemente incluya la sobrevaluación de viviendas.

El propósito de esta sección es mostrar cómo se puede incorporar información heurística al modelo, utilizando como ejemplo el caso de los grandes desarrolladores y de paso determinar si la sobrevalución es una práctica de los grandes desarrolladores.

En la sección 3.2 se presentan las estimaciones del modelo de regresión hedónica, en ellas se muestra que la variable dicotómica que controla el efecto de los grandes constructores resulta estadísticamente significativa. En la siguiente gráfica se presentan el valor de la vivienda reportado contra el valor de la vivienda estimado con la especificación que incorpora la variable dicotómica para los grandes desarrolladores. Esto se realiza para dos muestras, una que sólo incluye créditos de grandes desarrolladores y otra que los excluye de tal forma que se pueda diferenciar el comportamiento de las dos clases de créditos.

Gráfica 13
Valor Vivienda Reportado vs Valor Vivienda Estimado



En el caso de que no hubiera sobrevaluación, se esperaría que los puntos estuvieran centrados alrededor de la línea de 45°. Sin embargo, como se puede apreciar los avalúos reportados por los grandes constructores (representados por círculos color negro) se encuentran en su mayoría por encima de esta línea, lo que indica que están sistemáticamente sobrevaluados. La diferencia entre el valor estimado y el valor del avalúo es en promedio de 45,000 pesos, y representa aproximadamente un 17% más sobre el valor estimado de la vivienda.

Esta evidencia cuadra con lo que SHF ha detectado a través del "Mystery Shopping", es decir, los avalúos se inflan con el fin de cubrir el monto equivalente al enganche (más de 10%) de tal manera que el solicitante del crédito no tenga que realizar ningún desembolso.

Estos resultados muestran la importancia de controlar por el efecto de los grandes constructores dentro del modelo de detección de fraudes. Asimismo, siguiendo este procedimiento se pueden incorporar más información heurística que podría ayudar a mejorar el desempeño del modelo tal como: la diferenciación por regiones, ya sea de acuerdo a una clasificación basada en la ubicación geográfica o por alguna característica cualitativa como: zona fronteriza, zona turística, zona en expansión o por nivel per cápita.

Por otro lado, se tiene que la inclusión de un mayor número de variables descriptivas de la vivienda dentro de la base de datos de avalúos, cómo son el número de baños, lugares de estacionamiento, número de recamaras, etc, también podría ayudar de manera significativa a mejorar el poder explicativo de la regresión hedónica y a controlar por un mayor número de factores el grado de dispersión en el valor de las viviendas.

5. Conclusiones

Durante los últimos años la cartera hipotecaria en México ha venido mostrando un deterioro importante, el cual tiene relación con malas prácticas de originación. De acuerdo a la evidencia recolectada por SHF, la sobrevaluación de avalúos con el fin de obtener créditos y lograr que el acreditado no tenga que aportar el monto equivalente al enganche se ha convertido en una práctica común que ha impactado de manera negativa en la calidad de la cartera crediticia observándose un aumento en la probabilidad de incumplimiento.

En este trabajo se presentan dos modelos que utilizados ya sea de manera conjunta o individual se pueden utilizar como una herramienta adicional dentro del análisis de originación de créditos para detectar y prevenir posibles fraudes. El primer modelo consiste en la estimación del valor de la vivienda a partir de sus características cualitativas y el segundo modelo en el cálculo de las desviaciones del valor de la vivienda por metro cuadrado y del valor físico de la vivienda por metro cuadrado respecto a los valores promedio dentro del desarrollo al que pertenecen las viviendas.

Las primeras pruebas de eficiencia han mostrado que el modelo arroja resultados satisfactorios: de 129 créditos catalogados como "los más sospechosos" de sobrevaluación se encontró que el 77% fueron realizados por un valuador que se encuentra en la "lista

negra” y que el 80% de los valuadores en la lista negra fueron detectados por el modelo. Asimismo, con base en el modelo se pudo comprobar que los grandes desarrolladores de vivienda están sistemáticamente sobrevaluando el valor de las viviendas.

Como se mencionó el modelo representa un primer acercamiento al problema de la sobrevaluación de la vivienda por lo cual existen aún varios pasos a seguir con el fin de mejorar el desempeño del mismo. Uno de estos consiste en incorporar un mayor número de variables a la regresión hedónica que proporcionen información cualitativa sobre la vivienda como lo son el número de baños, recámaras, estacionamientos, si se cuenta con estancia comedor, jardín, patio, etc. Todo esto se podrá realizar una vez que la base de avalúos comience a capturar dichas variables, las cuales se espera resulten significativas dentro de la regresión y de esta manera permitan mejorar el ajuste de la misma. Otra posible mejora al modelo sería dividir por zona geográfica o por alguna característica específica a la zona como si es fronteriza, turística, industrial, en expansión. Esto con el fin de mejorar la precisión del modelo. Asimismo, es importante mencionar que un número importante de mejoras pueden surgir conforme se observe el desempeño del modelo en operación.

Bibliografía

Blackley, D.M., J.R. Follain, and H. Lee (1986), "An Evaluation of Hedonic Price Indices for Thirty-Four Large SMSAs", *AREUEA Journal*, 14, pp. 179-205.

Court, Andrew. (1939), "Hedonic Price Indexes with Automotive Examples", *The Dynamics of Automobile Demand*, pp. 99-117, General Motors Corporation.

Fannie Mae, (2006), Mortgage Fraud Overview.

Franklin, Joel P. and Paul Wadell, (2002), "A Hedonic Regression of Home Prices in King County, Washington, using Activity-Specific Accessibility Measures", *University of Washington*, working paper.

Greene, William. (1993). *Econometric Analysis*, 2a, ed., United States of America, Prentice Hall.

Griliches, Zvi. (1961), "Hedonic Price Indexes for Automobiles: An Econometric Analysis of Quality Change", *The Price Statistics of Federal Gov*ernment, New York, NBER.

Lancaster, K. (1971). "Consumer Demand: A New Approach" Columbia University Press, New York

Lacour-Little, Michael and S. Malpezzi, (2003). "Appraisal Quality and Residential Mortgage Default: Evidence from Alaska", *Journal of Real Estate and Economics*, 27:2, pp. 211-233.

Nair, Bhaskaran, (2006) "Use of Hedonic Regression Methods for Quality Adjustments In Statistics NZ", Inflation Measures Division, Statistics New Zealand, working paper.

Newey WK, and West KD (1987). "A Simple, Positive-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix", *Econometrica*, 55, pp. 703-708.

Rosen, S. (1974), "Hedonic Prices and Implicit Markets: Product Differentiation and Pure Competition", *Journal of Political Economy*, 82, pp. 34-49.

Thibodeau, T.G. (1989), "Housing Price Indexes from the 1974-1983 SMSA Annual Housing Surveys", *AREUEA Journal*, 17, pp. 100-117.

Waugh, Frederick, V. (1928), "Quality Factors Influencing Vegetables Prices", *Journal of Farm Economics*, 10:2, pp.185-196.

Anexo A:
Variables que conforman la Base de Datos de Avalúos

No.	Tipo de Variables	Nombre
1	Identificación	CP
2		CUV
3		Clave Banco
4		Fecha Avalúo
5		Fecha Registro
6		No. Registro Conjunto Infonavit
7		No. Registro Infonavit
8		Nombre Conjunto Infonavit
9		Nombre Constructor
10		No. Predial
11		Propósito
12		Status
13		Usuario Registro
14		Volante Sistema
15	Ubicación Geográfica	Calle Número
16		Colonia
17		Clave Estado
18		Clave Municipio
19	Características Vivienda: variables continúas	Año Terminación Obra
20		Porcentaje Infraestructura
21		Superficie Accesoria
22		SuperficieAsentada
23		Superficie Construida
24		Superficie Terreno
25		Superficie Vendible
26		Unidades Rentables
27		Unidades Rentables Generales
28		Vida Útil Remanente
29	Características Vivienda: variables categóricas	Clase Inmueble
30		Proximidad Urbana
31		Tipología
32	Valor Vivienda	Importe Valor Concluido
33		Valor Comparativo Mercado
34		Valor Físico Construcción
35		Valor Físico Instalaciones
36		Valor Físico Terreno
37		Otros Valores

Anexo B:
Codificación Dicotómica

Para incorporar variables categóricas dentro de un modelo de regresión lineal es necesario realizar una codificación dicotómica en la que se transforma la variable categórica en varias variables dicotómicas. De esta forma, en lugar de tener un solo coeficiente para la variable categórica, se tiene un coeficiente por cada categoría de la variable.

Existen diferentes formas de codificar las variables dicotómicas. Para efectos de interpretación, se eligió una codificación en la que los coeficientes de las variables dicotómicas representen la desviación del valor medio del promedio de dichos coeficientes.

La transformación de las variables categórica en variables dicotómicas se realiza siguiendo la siguiente regla: si la variable categórica tiene *k* categorías o niveles, entonces de deben de construir *k-1* variables dicotómicas. Por ejemplo, en el caso de la variable "proximidad urbana" la cual cuenta con tres niveles: céntrica, intermedia y baja proximidad[10]; se tienen que construir 2 variables dicotómicas. En la siguiente tabla se muestra la forma de codificación elegida para este trabajo tomando la variable proximidad urbana.

Codificación Dicotómica

Proximidad Urbana		
	Variables Dicotómicas	
Niveles	**Dcéntrica**	**Dintermedia**
Céntrica	1	0
Intermedia	0	1
Baja proximidad	-1	-1

Si la proximidad urbana corresponde a la categoría céntrica entonces las variables dicotómicas *"Dcéntrica"* y *"Dintermedia"* toman el valor de 1 y cero respectivamente. Si la categoría es intermedia entonces tomas el valor de 0 y 1 respectivamente, y si la categoría es baja proximidad ambas toman el valor de -1.

Finalmente, las variables que se introducen en la especificación de la regresión hedónica son *"Dcéntrica"* y *"Dintermedia"* para el caso de proximidad urbana, y para el caso de clase de inmueble se introducen *"Dmínima"*, *"Deconómica"*, *"DinterésSocial"* y *"Dmedio"*.

[10] En la base de avalúos la variable de baja proximidad cuenta con 5 niveles (céntrica, intermedia, periférica, de expansión y rural); sin embargo, para efectos de este trabajo se decidió agrupar a los tres últimos niveles en uno solo denominado baja proximidad debido a que no se observó diferencias significativas en las medias del valor de la vivienda para cada uno de estos niveles. Para el caso de clase de inmueble se realizó una agrupación similar ya que la variable original cuenta con siete niveles (mínima, económica, interés social, medio, semilujo, residencial y residencial plus) y en las estimaciones realizadas se agruparon los tres últimos como residencial quedando 5 niveles.