2005 V33 3: pp. 509–537
REAL ESTATE
ECONOMICS

Appraisal, Agency and Atypicality: Evidence from Manufactured Homes

Dennis R. Capozza,* Ryan D. Israelsen** and Thomas A. Thomson***

The appraisal of the "market value" of homes serving as the collateral for mortgages is a fundamental part of the underwriting process. If a loan should default, however, it is not the retail market value that the lender obtains, but rather the "recovery value." In this research, we show how recovery values differ from market values at origination and explore the reasons for the differences. Using a large sample of chattel mortgages on manufactured homes, we explore the relationship among the selling prices, the book values, and the fitted values from simple hedonic models with spatial autocorrelation. We then address the differences between selling prices at origination and recoveries from repossessed homes. We find that the spread between them varies systematically with home characteristics and especially with "atypicality," that is, with measures of how unusual a home is. Selling prices both at origination and recovery affect borrower defaults.

For decades, property appraisal has been a mainstay of mortgage underwriting and is an essential element of the verification process for lenders. The appraiser provides a professional estimate of the value of the property that can be used to verify that the selling price is representative of current market conditions. The lender uses the appraisal to assess whether the loan will perform, that is, repay in full, and whether the loan will be profitable for the lender. But is the simple point estimate sufficient for assessing the risks associated with the collateral? Can lenders improve loan decisions with information on the expected second moment or with estimates of the likely recovery values from a default?

In this research, we explore these issues with a large data set of chattel mortgages on manufactured homes. Manufactured homes, although little researched, are an increasingly important segment of the housing market. The Manufactured Housing Institute reports that 8% of the U.S. population lives in manufactured homes. In 2001, manufactured homes made up about 15% of residential starts.

*Ross School of Business, University of Michigan, Ann Arbor, MI 48109 or capozza@umich.edu.
**Ross School of Business, University of Michigan, Ann Arbor, MI 48109 or risraels@umich.edu.
***Department of Finance, College of Business University of Texas, San Antonio, TX 78249 or tthomson@utsa.edu.

Chattel mortgages, unlike conventional single-family mortgages, are secured only by the structure or "home" and do not include a lien on the underlying land. In most cases, the borrower does not own the land. Instead, the land is either rented from a third-party owner or from the owner of a manufactured housing "park," which, like a single-family subdivision, often groups similar homes in one location.[1] In parks, the land leases are usually month-to-month.

The foreclosure or repossession process can be much quicker and simpler for manufactured homes because the collateral is mobile, albeit at a nontrivial cost.[2] At most times, dealers provide a ready market for repossessed units, although occasionally they are overstocked with inventory and unwilling to provide the usual level of liquidity to the market.

There are many reasons to question the optimality of existing appraisal procedures for residential mortgages in general and manufactured homes specifically. First, there may be agency issues in the appraisal process. Appraisers, like all parties to real estate transactions, are often subjected to pressures or incentives to shade their estimates so that the transaction will be approved. Loan volume drives the compensation of many participants in the process and inevitably colors all of the participants at least indirectly.

There has been a fair amount of recent experimental research investigating appraisers' incentives and biases. For example, it has been shown that appraisers may be influenced by valuations of others (Diaz 1997) and by their own previous appraisals (Diaz and Wolverton 1998). A number of studies find that clients may influence the valuations.[3]

Even more basic to the appraisal process is whether appraisals are unbiased estimates of market values. It is well known that appraisals lag behind market movements (Quan and Quigley 1989). Appraisals can also be unbiased but inaccurate.[4]

---

[1] It is also possible for units to be placed on owned land and for the lender to have a lien on both the land and the home. These "land homes," then, are very similar to conventional mortgages, but are not included in our data. Other times, manufactured homes are installed on a permanent foundation and become largely indistinguishable from traditional "stick" built homes. Loans on these kinds of manufactured homes are not included in our data.

[2] Repossessed units are often wholesaled to dealers and hauled away to the dealer's sales lot where they are reconditioned and resold for a profit.

[3] See Kinnard, Lenk and Worzala (1997), Levy and Schuck (1999), Wolverton and Gallimore (1999) and Hansz and Diaz (2001).

[4] Cole, Guilkey and Miles (1986) made an *ex post* comparison of the most recent independent appraisal to subsequent sales prices of 144 commercial properties and found

A third issue is whether there is a link between appraisal quality and probability of default. Lenders use appraisals to help evaluate the probability of default. Loan-to-value ratios (LTVs) are widely used as a causal variable in statistical modeling of default probabilities. Noordewier, Harrison and Ramagopal (2001), using a sample of 1,428 residential loans, find evidence that properties that are valued above the sales price of "similar and proximate" properties are more prone to default. However, they conclude that the simple variance of property values is not related to default. Lacour-Little and Malpezzi (2003) find that decreasing appraisal quality is associated with an increase in the probability of default. In addition, they find that over-appraisal is significantly related to default, while under-appraisal is not. We have been unable to find any studies of the relationship of point estimates and dispersion estimates of property value and collateral value, which is ultimately what determines recovery risks.

But are current LTVs, whether based on purchase prices or appraisals, the appropriate driver for default equations? At the time of the default decision, the borrower will typically be distressed and the relevant value may not be a retail value. A forced or distressed sale will occur at a reduced price, which may be more relevant for assessing default probabilities at origination. Liquid properties where an owner can obtain a quick sale at only a small discount from market prices may have lower default probabilities because borrowers will be able to avoid default by selling. If, on the other hand, a property is unusual and difficult to sell, default may be the best alternative. Therefore, understanding the marketability and the liquidity of properties may help evaluate loan performance.

It is certainly the case that appraisals are an inappropriate metric for understanding the expected profitability of loans (Guttentag 1992, Capozza and Thomson 2005). Assessing profitability requires that lenders estimate all the future cash flows on the loan including the recoveries from defaulted loans. For this task, estimates of the recovery values will be more useful than estimates of the market value at origination. Therefore, while appraised values, that is, estimates of the market value at origination, are the most commonly used in underwriting, it may be that lenders should estimate recovery or wholesale values. If recovery values are a better predictor of defaults than appraised values, then appraised values

---

an average absolute difference of about 9%. Using a sample of 500 properties acquired by corporate relocation firms, Dotzour (1988) found that current appraisal methods were unbiased estimates of value and that a large part of the variation in appraisal error could be accounted for with seasonal factors and regional economic conditions. Chinloy, Cho and Megbolugbe (1997) conclude, using a sample of 3.7 million repeat transactions on mortgages bought by Fannie Mae and Freddie Mac, that appraisals are systematically higher than purchase data. However, a drawback of using sales price to evaluate appraisal quality is that low appraisals tend to be correlated with rejected loans, and are therefore not included in the samples.

are superfluous. If, on the other hand, appraised values are a better predictor of defaults, then estimates of both market values at origination and recovery values in default will be needed.

From the borrower perspective, both the appraisal (retail values) and the distressed or recovery values should be relevant. If the borrower is subject to financial stress and cannot make payments, the first choice is to sell the property to a retail buyer and prepay the loan, thereby salvaging any equity. When trying to sell to a retail buyer, the liquidity of the property will matter. Illiquid properties will be less likely to sell, reducing the chance that the borrower can repay the loan. The difference between the retail value and the recovery value is a measure of the liquidity of the property similar to the bid–ask spread in other financial markets. If the property does not sell to a retail buyer, the borrower may be forced into foreclosure. Therefore, there are many moving parts in the borrower's decision. Both the appraised value and the estimated recovery value play a role and should be included in the lender's assessment of risks and expected losses.

As part of our analysis, we explore two measures of liquidity and heteroskedasticity in some detail. The first, atypicality, (Haurin 1988) is a measure of how unusual a unit is relative to an average unit. The second, sparsity, measures the number of units that have sold recently within a 15-mile radius. Of the two, atypicality is a more important factor in hetroskedasticity.

In the next section, we describe the data and follow with a discussion of the methodology. The third section presents the results of our empirical analysis, which is divided into eight subsections. We first estimate hedonic models of selling prices that exploit spatial autocorrelation and include our measures of atypicality and sparsity. The next subsection compares the 1-year forecast performance of the various models and shows that the best models have mean squared errors (MSEs) of 6–7%. We then compare the results from using the selling price, the appraised value and the various hedonic models for predicting recovery values for repossessions and show that the simplest hedonic models are the best predictors of recoveries. These results suggest that valuation for recovery is quite different from the valuation of selling prices at origination.

The fourth subsection analyzes the characteristics that are important for the retail sales (at origination) versus the wholesale prices (recoveries). The fifth subsection analyzes the forecast errors and shows that atypical units have larger forecast errors. This suggests that models of automated appraisal should adjust for atypicality and lenders should incorporate this information into profitability estimates. The sixth subsection compares models of default that use selling price, appraisal, automated appraisal and recovery estimates to calculate LTVs.

In the last subsection we document that both manual appraisals and simple statistical hedonic or automated appraisals improve the predictions of default models. Automated and manual appraisals perform about equally well. Estimates of the recovery value also significantly improve risk assessment. The final section summarizes and concludes.

Data

Our data are a large sample of loans from the manufactured housing industry. The data include 195,442 observations of manufactured homes at the time of loan origination and the subsequent recovery information for 9,746 of the homes that defaulted. All of the homes are located on rented land. Variables in the origination database include information about the loan:

Sales price = the purchase price (market price) of the home.
Book or appraised value = the book value or appraised value of the home.[5]
Origination date = the month of loan origination.
Termination date = the final month of data for the loan.
Refinance = a dummy variable indicating that the loan is a refinance of an existing loan.

Variables also include information about the collateral:

Length = the length of the home.
Width = the width of the home.
Model year = the model year of the home.
ZIP code = ZIP code of the location of the home.
Park = a dummy variable indicating that the home is located in a manufactured housing park.
Multi-wide home = a dummy variable indicating that the home is multi-wide.
Resale of repossessed home = a dummy variable indicating that the unit was a former repo.
Manufacturer 1–Manufacturer 13 = dummy variables indicating the manufacturer of the home.

If a unit is repossessed and sold, the following data are available:

Repo sale price = the price for which the repossessed unit sold (the second sales observation).

[5] As in the auto lending industry, the collateral value or "book value" is typically estimated from standard industry guides that provide estimates of value based on market, make, model, year and equipment. Occasionally, however, a full residential appraisal is done.

Number of baths = the number of baths in the repossessed unit.
Condition = the physical condition of the repossessed unit.

In addition to the variables in the database, we created several others. The postal ZIP code data can be assigned a latitude and longitude. From these coordinates for each sale, we calculated the distance between observations and created a variable, *SPARSITY*, which is a measure of proximity to other recent sales. *SPARSITY* is the reciprocal of 1 plus the number of sales within a 15-mile radius during the previous year:

$$SPARSITY_i = \frac{1}{1 + N_i} \tag{1}$$

where N_i is the number of sold homes within 15 miles of the *i*th home. Note that

$$SPARSITY_i \in [0, 1], \forall i, \lim_{N_i \to 0} SPARSITY_i = 1, \quad \text{and}$$
$$\lim_{N_i \to \infty} SPARSITY_i = 0.$$

(As the number of home sales near the *i*th home increases, *SPARSITY* approaches 0, and when there are no homes within 15 miles of the *i*th home, its value is 1.)

Because new units come with a manufacturer's invoice price instead of a manual appraisal, we limit our study to those homes that were previously owned at the time of origination. Table 1 presents summary statistics for the variables.

To account for the difficulty that arises in valuing homes with unusual features, we created a variable, *ATYPICAL*, following the approach outlined in Haurin (1988). We use implicit marginal prices from a semi-log hedonic regression of home sales prices on various characteristics to penalize absolute deviations from the average (see the Appendix for a more complete discussion). Thus, the weights used to generate the *ATYPICAL* variable are percentage influences on value rather than absolute dollar prices.

Results

Hedonic Models with Spatial Autocorrelation

We begin with a standard hedonic regression of the log of real sales price on the available set of characteristics. The results are presented in Table 2, panel A, first column. The most significant independent variables include the age

Table 1 ■ Summary statistics.

Variable	Mean	Std. Dev.	Max.	Min.	*N*
Sale price	$19,757	$10,244	$75,000	$2,300	195,442
Appraised value	$21,783	$11,233	$75,000	$1,000	195,442
Latitude	38	5	66	19	193,843
Longitude	−92	13	−67	−167	193,843
Age of unit at origination (months)	116	73	408	12	195,442
Model year	1985	6	1999	1963	195,442
Length of unit (feet)	65	10	80	40	195,442
Width of unit (feet)	17	5	40	12	195,442
Repo sale price	$10,808	$9,330	$95,000	$200	8,941
Location = park	0.85		1	0	195,442
Multi-wide home	0.24		1	0	195,442
Origination date (1 = January 1987)	101	35	150	1	195,442
Termination date	134	25	150	6	195,442
Loan-to-value at origination	0.89	0.1	1.1	0.2	135,638
Loan term	172	71	360	24	135,638
Loan age	34	25	150	1	195,442
Borrower credit score	650	64	840	430	116,242
Baths per 1,000 SF	1.6	0.4	4.2	0.6	195,442
Baths unknown	0.9	0.2	1.0	0.0	195,442
Spatial lag	9.4	0.3	10.8	8.5	195,442
Sparsity	0.1	0.2	1.0	0.0	195,442
Atypicality	0.5	0.3	1.6	0.1	195,442
Manufacturer 1	0.01		1	0	195,442
Manufacturer 2	0.02		1	0	195,442
Manufacturer 3	0.01		1	0	195,442
Manufacturer 4	0.02		1	0	195,442
Manufacturer 5	0.01		1	0	195,442
Manufacturer 6	0.00		1	0	195,442
Manufacturer 7	0.00		1	0	195,442
Manufacturer 8	0.01		1	0	195,442
Manufacturer 9	0.01		1	0	195,442
Manufacturer 10	0.01		1	0	195,442
Manufacturer 11	0.00		1	0	195,442
Manufacturer 12	0.01		1	0	195,442
Manufacturer 13	0.00		1	0	195,442

Notes: This table displays selected summary statistics for the collateral and the loans. Variable names are self-explanatory. Missing loan amount data limit the number of loan-to-value observations to 135,638. Sparsity is defined as $1/(1 + N)$ where N is the number of homes sold within 15 miles in the previous year. Atypicality is defined in Equation (2) in the text. Spatial lag is a weighted sum of the log of sales prices as defined in Equation (A.1).

of the home, the size of the home and whether the unit is single or multi-wide. The coefficients indicate that each month a home depreciates by 30 basis points or 3.6% per year. Each 100 square feet adds 10% to value. For any given size unit, a multi-wide configuration is worth 35% more. Location in a

Table 2 ■ Hedonic models with spatial autocorrelation, sparsity and atypicality.

Panel A: Dependent Variable = *Log of Real Purchase Price*.
Includes Manufacturer Indicators (Not Reported)

	Baseline		Base + Spatial		Base + Sparsity + Atypicality		Base + Spatial + Sparsity + Atypicality	
Variable	Coefficient	t Value	Coefficient	t Value	Coefficient	t Value	Coefficient	t Value
Intercept	8.74	664.0	4.51	165.7	8.75	664.4	4.54	166.5
Home age	(0.003)	(263.4)	(0.003)	(289.6)	(0.003)	(248.1)	(0.003)	(269.4)
Area	0.001	163.3	0.001	181.7	0.001	153	0.001	171.8
Multi-wide	0.35	147.1	0.26	118.1	0.32	123.6	0.25	100.6
Baths/area	65.75	9.2	51.81	7.8	66.65	9.3	52.62	7.9
Baths unknown	−0.02	−1.8	−0.01	−0.6	−0.02	−1.9	−0.01	−0.7
Location = park	0.16	81.6	0.12	64.4	0.156	78.3	0.117	62.5
Spatial rho			0.45	174.1			0.45	172.9
Atypicality					0.11	21.4	0.08	16.7
Sparsity					(0.07)	(16.7)	(0.04)	(9.7)
Manuf. Dum.								
Adjusted R^2	0.59		0.65		0.59		0.65	
Root MSE	0.306		0.284		0.306		0.284	

Table 2 ■ continued.

Panel B: Dependent Variable = *Log of Real Purchase Price*.
Includes Manufacturer Indicators (Not Reported) and State Indicators (Not Reported)

Base + States			Base + States + Spatial		Base + States + Sparse + Atypicality		Base + States + Spatial + Sparse + Atypicality	
	Coefficient	t Value	Coefficient	t Value	Coefficient	t Value	Coefficient	t Value
Intercept	8.60	694.8	6.10	210.6	8.62	696.8	6.16	212.4
Home age	(0.003)	(331.7)	(0.003)	(332.5)	−0.003	(311.0)	−0.003	(310.7)
Area	0.001	189.6	0.001	192.3	0.001	178.7	0.001	181.6
Multi-wide	0.22	93.6	0.21	92.2	0.2	81.2	0.2	80.3
Baths/area	64.83	10.1	61.95	9.9	65.31	10.2	62.54	10
Baths unknown	0.02	1.7	0.02	2.3	0.02	1.5	0.02	2.1
Location = Park	0.06	32.8	0.05	29.0	0.05	28.1	0.05	25.3
Spatial rho			0.27	95.3			0.26	93.4
Atypicality					0.08	16.2	0.07	15.2
Sparsity					(0.10)	(29.0)	(0.08)	(23.0)
State dummies								
Manuf. Dum.								
Adjusted R^2	0.68		0.69		0.68		0.69	
Root MSE	0.273		0.267		0.272		0.266	

Notes: This table presents the results of regressions of the log of the real purchase price of used homes on eight combinations of variables. The regressions all include dummy variables for the manufacturer of the home. Rho is the weighted average value of the 15 nearest homes sold within the last year. Sparsity is 1/[(No. of sales within 15-mile radius during the previous year) + 1]. Atypicality is defined in the text. State dummy variables are included in panel B. Alabama is the excluded state. There are 190,441 observations used in the regressions covering the years 1988–1999.

manufactured housing park is worth 16% more. Extra baths add to value, but are not highly significant, perhaps because this variable is sparsely populated in the database. Manufacturer indicator variables were included in all the equations in Table 2, but are not reported. In the second equation of panel A, we adjust for spatial autocorrelation by including a weighted average of the surrounding sales as described in the Appendix. This spatial variable is highly significant and reduces the MSE. In the third equation of panel A, the *SPARSITY* and *ATYPICAL* variables are included, but not spatial autocorrelation. *SPARSITY* and *ATYPICAL* are significant, but add little to the fit and accuracy of the model.

In panel B, we present identical regressions but now including state indicator variables. The addition of the 50 state indicators does improve the fit relative to panel A. The second equation in panel B shows that spatial autocorrelation is still an economically important and statistically significant variable even when state indicators are included.

Out-of-Sample Forecast Performance

In Table 3, we present the results of our test of out-of-sample forecast accuracy of the eight models in Table 2. We first train the model on an initial sample of data, and then use the estimates to forecast sales prices 1 year ahead. Each year the model is "retrained" using an additional year of data. Table 3 displays the MSEs by year and by model from this experiment.

In each year, the most accurate models include both the state indicators and the spatial autocorrelation variable. This suggests that local factors play an important role in the price of manufactured homes. One curious result in Table 3 is the tendency for the MSEs for all the models to become larger over time despite being trained on extra years of data. This suggests that the prices of these homes became more disperse and more difficult to explain over time.

The Value of Recovered Units

In Table 4, we begin to explore the relationship between purchase prices and recovery liquidation prices. We do so by regressing the log of the real recovery price on one of several estimates of the value of the home at origination—the actual purchase price, the book or appraised value and six of the hedonic estimates from Table 2. Also included in each equation are indicator variables for the condition of the home (poor, fair, good) and for the sales channel (refinance, wholesale, cash, redemption). The loan age at the time of default is included as a measure of the depreciation since origination.

Table 3 ■ One-year ahead forecast performance of the hedonic models.

Forecast Year		1994	1995	1996	1997	1998	1999
N		22,086	28,859	32,776	35,016	33,384	11,460
Model 1	Hedonic (baseline)	8.5%	8.2%	8.5%	8.6%	9.4%	9.5%
Model 2	Hedonic + Spatial	7.2%	7.8%	8.1%	8.2%	8.6%	8.6%
Model 3	Hedonic + Sparsity + Atypicality	8.6%	8.2%	8.4%	8.6%	9.3%	9.5%
Model 4	Hedonic + Spatial + Sparsity + Atypicality	7.2%	7.8%	8.1%	8.2%	8.5%	8.5%
Model 5	Hedonic + States	6.9%	7.2%	7.4%	7.5%	8.1%	8.0%
Model 6	Hedonic + States + Spatial	6.5%	7.0%	7.2%	7.2%	7.6%	7.5%
Model 7	Hedonic + States + Sparsity + Atypicality	6.9%	7.1%	7.4%	7.4%	8.0%	8.0%
Model 8	Hedonic + States + Spatial + Sparsity + Atypicality	6.5%	7.0%	7.2%	7.1%	7.6%	7.5%

Notes: This table presents mean squared errors resulting from forecasting the log of real purchase price of a home for 1994, 1995, 1996, 1997, 1998 and 1999 by using the models presented in Table 2 for the years 1988 through 1993, 1988 through 1994, ... 1988 through 1998, respectively. MSE is the mean squared error of the predicted log of the real cash price of the home versus the actual log of real cash price. *N* is the number of predictions per year. Dependent Variable = *Log of Real Purchase Price*.

Table 4 ■ Forecasting collateral value at origination: Where's the beef?

Panel A: Dependent Variable = *Log of Real Recovery Price*

Independent Variable	Purchase Price		Actual Appraisal			Baseline			Baseline + Spatial		
Variable	Estimate	*t* Value	Variable	Estimate	*t* Value	Variable	Estimate	*t* Value	Variable	Estimate	*t* Value
Intercept	−1.74	−5.2		−0.85	−2.5		−9.32	−21.0		−6.22	−15.0
Log purchase price	1.13	33.5	Log appraisal	1.03	30.8	Baseline	1.95	42.14	Base + Spatial	1.61	37.6
Loan age	−0.02	−27.6		−0.02	−25.6		−0.02	−24.1		−0.02	−27.0
Poor condition	−0.46	−3.7		−0.50	−4.1		−0.79	−6.8		−0.62	−5.2
Fair	0.28	2.3		0.27	2.2		−0.02	−0.2		0.13	1.1
Good	0.55	4.4		0.55	4.4		0.29	2.4		0.46	3.8
Refi by lender	0.54	0.4		0.13	0.1		0.33	0.3		0.55	0.4
Wholesale	0.21	0.4		0.17	0.3		0.08	0.2		0.13	0.3
Cash sale	0.34	1.4		0.38	1.5		0.35	1.5		0.30	1.2
Redemption	0.52	2.4		0.50	2.3		0.39	1.9		0.44	2.1
8,071 Observations											
Adjusted R^2	0.28			0.27			0.33			0.30	
Root MSE	1.35			1.36			1.30			1.33	

Table 4 ■ continued.

Panel B: Dependent Variable = *Log of Real Recovery Price* with State Indicators Included

Independent Variable	Base + Spatial + Sparse + Atypical		Base + States			Base + States + Spatial			Base + States + Spatial + Sparse + Atypical		
Variable	Estimate	t Value	Variable	Estimate	t Value	Variable	Estimate	t Value	Variable	Estimate	t Value
Intercept	−4.96	−12.2		−5.91	−14.4		−4.95	−12.4		−2.71	−7.2
Base + Spatial	1.48	35.4		1.58	37.3	Base	1.48	35.9	Base + States	1.24	32.1
+ Sparse			Base			+ States			+ Sparse		
+ Atypicality			+ States			+ Spatial			+ Atyp		
Loan age	−0.02	−26.7		−0.02	−26.7		−0.02	−27.2		−0.02	−26.3
Poor condition	−0.63	−5.2		−0.63	−5.3		−0.59	−4.9		−0.60	−5.0
Fair	0.14	1.2		0.12	1.0		0.16	1.4		0.18	1.5
Good	0.47	3.8		0.41	3.3		0.47	3.8		0.50	4.0
Refi by lender	0.51	0.4		0.18	0.1		0.33	0.2		0.26	0.2
Wholesale	0.10	0.2		0.22	0.4		0.19	0.4		0.13	0.3
Cash sale	0.30	1.3		0.27	1.1		0.26	1.1		0.27	1.1
Redemption	0.46	2.2		0.47	2.2		0.47	2.2		0.49	2.3
8,071 Observations											
Adjusted R^2	0.29			0.30			0.29			0.27	
Root MSE	1.34			1.33			1.33			1.35	

Notes: This table presents the results of the log of the real recovery price regressed on the condition of the home at time of repossession, the finance type for the repossessed unit, the maximum age of the loan (*LOANAGE*) and either the log of the real purchase price at origination, the log of the real appraised value of the home at origination or one of six automated appraisal models estimated in Table 2. Dummy variables indicate that at the time of recovery, the unit is in poor, fair or good condition, respectively. The case in which the condition of the unit is unknown is the excluded category. Dummy variables that indicate the sales channel for the repossessed home are refinance, wholesale, cash or redemption, respectively. The case in which the sales channel is unknown is the excluded category in the regressions. There were 8,071 observations of homes that were repossessed and for which we had data for all of the necessary variables. The data cover the time period from 1988 to 1999. Adjusted R^2 and root mean squared error for each of the regressions is presented, as well as coefficient estimates and t values.

The most remarkable feature of Table 4 is that the simplest model with the worst 1-year-ahead forecasting performance in Table 3, the baseline model, is the most accurate for predicting recovery prices. This suggests that price-setting behavior in recoveries is quite different from retail pricing at origination. That is, the characteristics that the buyers of repossessions value are different from those valued by retail purchasers.

Other coefficients are similar across the equations in Table 4. Condition of the units has the expected effects, as does sales channel.

Why Does Value Change in a Default?

Table 5a probes the relationship between retail sales and recovery sales further by displaying hedonic regressions for retail prices at origination next to similar equations for recovery prices. Because the homes associated with loans that default may be different from the entire sample of homes, only the 8,071 observations for which both the origination price and the recovery price are available are included in Tables 5a–5c. Several items are noteworthy. First, age is a more important factor in a recovery sale because the implied depreciation rate is three or more times larger than the recovery sale regressions (−0.01 versus −0.003). Second, there are sign changes on the baths variable. Additional baths, apparently, are a liability in a recovery. This is consistent with a world where baths are in poor condition at the time of recovery or even vandalized, making rehabilitation costs higher. Third, location in a manufactured housing park is advantageous for recovered units. Units in parks may be easier to resell. Fourth, size is more important in recovery. Recovery buyers place more weight on square footage than retail buyers.

The most remarkable result, however, is the sign flip on atypicality. Unusual units trade for a premium at origination but for a very large discount in a recovery. This is very strong evidence of the effect of liquidity on the spread between origination and recovery prices. Unusual units will be more difficult to sell and on average be held longer in inventory before a sale. At the retail level, dealers will want to sell unusual units at a premium to compensate for the greater inventory expense. On the other hand, at the recovery level, dealers, expecting to have to hold unusual units for an extended period, will only purchase at a discount to compensate for the inventory costs.

Cross-Equation Tests for Equality of Coefficients

In Table 6, we test for the significance of the differences between the coefficients in the retail price (at origination) and the recovery price regressions. Only the

Table 5a ■ Purchase price hedonics versus recovery hedonics: Why does value change in default?

Dependent Variable =	Log of Real Purchase Price		Log of Real Recovery Price		Log of Real Appraisal Price		Log of Real Recovery Price		Log of Real Purchase Price		Log of Real Recovery Price	
Variable	Coefficient	*t* Ratio	Coefficient	*t* Ratio	Coefficient	*t* Ratio	Coefficient	*t* Ratio	Coefficient	*t* Ratio	Coefficient	*t* Ratio
Intercept	8.86	247.5	8.34	57.8	8.88	246.4	8.35	57.6	2.44	24.7	4.96	10.1
Home age	−0.003	−38.4	−0.01	−34.0	−0.003	−38.6	−0.01	−34.0	−0.003	−48.4	−0.01	−34.3
Loan age			−0.02	−28.2			0.02	−28.1			−0.02	−29.1
Area	0.0003	12.7	0.0015	16.1	0.0003	12.9	0.00	16.1	0.0006	29.6	0.0017	17.1
Multi-wide	0.59	41.9	0.62	11.1	0.59	41.6	0.62	11.0	0.29	24.2	0.49	8.2
Baths/area	78.56	8.9	−85.48	−2.5	77.80	8.8	−85.99	−2.5	53.08	7.6	−98.03	−2.9
Baths unknown	0.15	5.4	−0.20	−1.8	0.15	5.3	−0.20	−1.8	0.07	3.3	−0.24	−2.2
Location = park	0.23	19.5	0.45	9.7	0.22	18.5	0.45	9.4	0.12	12.2	0.39	8.3
Atypicality	0.24	8.0	−0.79	−6.7	0.24	7.9	−0.79	−6.7	0.13	5.5	−0.85	−7.2
Sparsity					−0.08	−4.4	−0.05	−0.7	0.01	0.8	0.01	0.1
Spatial rho									0.68	68.2	0.36	7.2
Adjusted R^2	0.54		0.30		0.54		0.30		0.71		0.31	
Root MSE	0.34		1.33		0.34		1.33		0.3		1.3	

Notes: This table compares recovery hedonic model coefficients with those of origination price hedonic appraisals for various combinations of independent variables. The first model is a regression of the log of the real purchase price on home age, size of the home (*area*), an indicator that the unit is multi-section, baths per area, an indicator that the unit is located in a manufactured home park, dummy variables to indicate the manufacturer and measures of atypicality of the unit and the sparseness of surrounding sales. Only the 8,071 observations for which both origination price and recovery price were available are included in this table.

Table 5b ■ Cross-equation test for equality of coefficients.

	Model 1		Model 2		Model 3		Model 4	
Variable	Coefficient	*t* Ratio	Coefficient	*t* Ratio	Coefficient	*t* Ratio	Coefficient	*t* Ratio
Intercept	2.30	30.3	0.31	2.2	0.32	2.3	−2.31	−4.8
Home age			0.01	25.5	0.01	25.4	0.01	25.3
Loan age			0.03	35.1	0.03	35.1	0.03	32.1
Area	−0.0011	−15.7	−0.0012	−12.4	−0.0012	−12.4	−0.0011	−11.2
Multi-wide			−0.10	−1.8	−0.10	−1.8	−0.20	−3.5
Baths/area			160.80	4.8	160.27	4.8	150.92	4.5
Baths unknown			0.35	3.3	0.35	3.3	0.32	3.0
Location = Park			−0.23	−5.1	−0.24	−5.1	−0.28	−5.9
Atypicality			1.04	9.1	1.04	9.1	0.99	8.7
Sparsity					−0.05	−0.8	−0.01	−0.1
Spatial rho							0.28	5.7
Adjusted R^2	0.03		0.25		0.25		0.26	
Root MSE	1.47		1.29		1.29		1.29	

Notes: In this table, we subtract the log of the real recovery price from the log of the real origination price and regress this difference on the independent variables. There are 8,071 observations from 1988 through 1999. Only the 8,071 observations for which both origination price and recovery price were available are included in this table.

Table 5c ■ Recovery price hedonic versus origination price hedonic (weighted least squares).

	Origination Price Hedonic		Recovery Price Hedonic		Origination Price Hedonic		Recovery Price Hedonic	
Dependent Variable =	Log Real Purchase Price		Log of Real Recovery Price		Log Real Purchase Price		Log of Real Recovery Price	
Variable	Coefficient	*t* Ratio	Coefficient	*t* Ratio	Coefficient	*t* Ratio	Coefficient	*t* Ratio
Intercept	8.87	240.3	8.23	60.0	2.47	24.8	3.99	8.9
Home age	−0.003	−39.8	−0.009	−37.1	−0.003	−48.5	−0.009	−37.2
Loan age			−0.02	−31.5			−0.02	−33.0
Area	0.0004	14.0	0.00	16.0	0.001	29.9	0.002	17.5
Multi	0.55	36.2	0.55	9.7	0.28	22.2	0.38	6.5
Baths/area	52.46	6.4	−110.74	−4.1	45.72	6.8	−119.90	−4.4
Bath unknown	0.09	3.3	−0.27	−3.0	0.06	2.7	−0.29	−3.2
Park	0.21	18.8	0.44	12.4	0.12	12.1	0.39	10.8
Atypicality	0.32	9.9	−0.38	−3.3	0.15	6.2	−0.48	−4.2
Sparsity					0.00	0.3	−0.04	−0.8
Spatial rho					0.67	67.1	0.45	9.9
Adjusted R^2	0.49		0.30		0.69		0.30	
Root MSE	1.01		1.30		1.03		1.32	

Notes: This table compares recovery price hedonic model coefficients with those of origination price hedonics for various combinations of independent variables. All regressions use weighted least squares with atypicality as the weights as described in the text. There are 8,071 observations for the years 1988–1999.

8,071 observations for which both origination price and recovery price were available are included in this table.

The most significant differences are for the depreciation rates (*Home Age* and *Loan Age*) and the unit size (*Area*). All the coefficient differences are significant at the usual levels except for *Sparsity*. These results are additional evidence that collateral values in a default are very different from values at origination. In particular, collateral values at default are not a simple additive or multiplicative transformation of values at origination. Some characteristics are more valuable, others less. In some cases, characteristics that are an asset at origination become a liability in a default.

Precision Regressions: What Determines Forecast Error?

Given the findings above that the prices (coefficients) of characteristics can vary or even change sign depending on the type of sale, we expect that the residuals from the hedonic regression might be heteroskedastic in one or more of the characteristics. Atypicality[6] is a primary candidate as a possible determinant of the absolute size of residual errors.

In Table 6, we explore residual variation with Glejser regressions (see Appendix) of transformed residuals on *atypicality* and *sparsity*. In Table 6, we regress the transformed residuals from the first equation in Table 2 on measures of atypicality. *Atypical* (composite) is defined as:

$$\begin{aligned} Atypical = {} & |P_{area}(Area - \overline{Area})| \\ & + |P_{age}(Age - \overline{Age})| + |P_{width}(DMulti - \overline{Dmulti})| \end{aligned} \tag{2}$$

where P_i are the coefficients from Model 1 in Table 2, and the bars indicate mean values. Thus, highly atypical homes are units where the weighted sums of the absolute values of these characteristics are far from their sample means. The weights are simply the implicit price weights from the Model 1 in Table 2 regression.

The results indicate that most of the explainable residual variation depends on the *Atypical* composite variable. Breaking this variable into its components or adding the sparsity measure does not improve the fit substantially. In the next

[6] Goodman and Thibodeau (1998) show that error variance tends to increase with home age and to length of time between sales, presumably because of an increased likelihood of undocumented renovations. Accordingly, we include age in our measure of atypicality.

Table 6 ■ Precision regressions: What determines forecast error?

Variable	Coefficient	t Ratio	Coefficient	t Ratio	Coefficient	t Ratio	Coefficient	t Ratio	Coefficient	t Ratio
Intercept	0.20	28.9	0.19	27.7	0.20	26.3	0.20	27.6	0.34	100.4
Atypical age			0.0006	7.5	0.0006	7.5				
Atypical size			0.0002	8.8	0.0002	8.8				
Atypical width (Dmulti)			0.19	13.9	0.19	13.8				
Atypical (composite)	0.34	20.9					0.34	20.8		
Sparsity					−0.002	−0.2	−0.01	−0.4	−0.03	−2.4
Adjusted R^2	0.05		0.06		0.06		0.05		0.00	
Root MSE	0.26		0.26		0.26		0.26		0.26	

Notes: This table presents the results of regressions of the square root of $\pi/2$ times the absolute value of the residuals from Model 1 in Table 2 on various combinations of atypicality variables and sparsity. There are 8,071 observations. *Atypical Age* $= |P_{age}(Age -$ Mean $Age)|$, *Atypical Area* $= |P_{area}(Area -$ Mean $Area)|$ and *Atypical Width* $= |P_{width}(Dmulti -$ Mean $Dmulti)|$, where the P_i are the implicit prices of *age*, *area* and *Dmulti*, respectively, derived from a price hedonic equation. *Atypical* (composite) = *Atypical Age* + *Atypical Area* + *Atypical Width*. *Sparsity* = 1/[(No. of sales within 15 miles during the previous year) + 1]. The table presents the adjusted R^2 and root mean squared errors for each of the regressions.

Dependent variables: sqrt(pi/2) ∗ abs(residuals from Table 2, Model 1) 8,071 observations.

section, we use these results to re-estimate selected Table 2 and 5a–c regressions with weighted least squares (WLS) using *Atypical* as the weights.

Weighted Least Squares Estimates

Table 5c presents the selected WLS regressions. Weighting the observations does affect the coefficients, presumably because the WLS estimates are more efficient. However, the qualitative results and the conclusion remain unchanged. There are still significant differences between the origination price and the recovery price coefficients.

Predicting Defaults: Actual Versus Appraisal Versus Hedonic Prices

We have now established definitively that the recovery values are different from origination values and are not a simple multiplicative transform of origination values. Lenders attempting to estimate the profitability of a loan, which depends directly on these recovery values rather than the origination values, will be best served by having estimates of recovery values independent from the appraisal at origination. In this section, we begin to analyze default probabilities using the various estimates of collateral value. In default equations, important determinants of loss rates are the initial and the subsequent LTVs (Capozza, Kazarian and Thomson 1997, 1998). The denominator of LTV is a measure of the value of the collateral and one would expect that neutral assessments of property value at origination would improve the power of LTV in a default equation. For example, we know that transaction prices in real estate are the result of bargaining between the buyer and the seller. A buyer who obtains a more favorable price may be less likely to default. Therefore, estimates of collateral value like a manual appraisal should be better for predicting defaults, because they should remove the bias from bargaining power.

On the other hand, the default decision, which is made at a later and perhaps distressed time in the life of the loan, may be driven by the liquidation values of the property. The recovery estimates should provide a better measure of recovery values.

Table 7 displays our results on the effect of the various measures of LTV. The first three columns use the actual purchase price, the manual appraisal price and the fitted origination price values, respectively, as the denominator in LTV. The analysis uses logistic regression to estimate the equations. We provide two measures of fit: pseudo-correlation, which is the correlation between the actual and the fitted values, and percent concordant that measures the ability of the models to correctly classify the observations. Both the manual or the book

Table 7 ■ Default equations with LTV from actual prices, manual appraisals and origination price hedonic.

	LTV Based on											
	Actual Purchase Price		Manual Appraisal		Fitted Origination Price		Manual Appraisal		Fitted Origination Price		Fitted Origination Price and Fitted Recovery Price	
	Coefficient	Chi-Square	Coefficient	Chi-Square	Coefficient	Chi-Square	Coefficient	Chi-Square	Coefficient	Chi-Square	Coefficient	Chi-Square
Intercept	−1.07	37	−0.94	41	−0.29	4	−1.57	74	−1.40	58	−1.58	69
Length of exposure	0.04	2,586	0.04	2,588	0.04	2,583	0.04	2,585	0.04	2,608	0.04	2,563
LTV (origination price)	2.01	223	1.77	583	0.96	452	2.58	328	2.28	259	2.58	296
LTV_plus							0.29	3	0.74	16	0.44	7
LTV_minus							−2.74	138	−1.77	122	−2.81	138
LTV (Recovery price)											0.09	179
Bureau score	−0.07	1,305	−0.07	1,212	−0.07	1,279	−0.07	1,205	−0.07	1,286	−0.07	1,230
No Bureau score	−4.20	999	−4.05	939	−4.15	985	−4.04	922	−4.18	982	−4.11	946
Sparsity					0.03	0			0.03	0	−0.02	0
Atypicality					0.02	0			0.15	12	−0.35	38
Pseudo-correlation	0.18		0.19		0.19		0.20		0.19		0.20	
Percent condordant	71.6		72.2		72.1		72.4		72.3		72.7	

Notes: This table uses logistic regression to estimate the relationship between defaults and three measures of loan-to-value ratios (LTVs). *Length of exposure* is the time from origination to the end of the observation period. *LTV_plus (minus)* is an indicator that the appraisal-based LTV is above (below) the purchase price-based LTV. *Bureau score* is the borrower credit score provided by one of the credit reporting agencies. *No bureau score* is an indicator equal to one when the bureau score is missing. *Sparsity* and *atypicality* are as defined in the text. 9,198 loans defaulted, 126,268 did not default; there was a 6.8% default rate.

appraisal and the hedonic estimates are slightly better predictors of defaults than the actual purchase prices as expected. The manual and the hedonic appraisals are roughly comparable although the concordance is slightly (0.1) higher for the manual appraisal.

In the penultimate two columns, we add variables that measure whether the appraisals are below (*LTV_plus*) or above (*LTV_minus*) the purchase price. In both cases, the results indicate that buyers who pay more than the neutral appraisal prices default at significantly higher rates. This implies that having a neutral assessment of the value of the property at origination, whether manual or book appraisal or statistical appraisal, is indeed beneficial for predicting defaults.

In the last two columns, we test whether the estimated recovery values affect the borrowers default decision. Earlier, we presented evidence that atypical houses sell for lower recovery values. We can also expect atypical houses to take longer to market for sale. In a default situation, there is a limited time before the borrower's equity is foreclosed. Therefore, while the recovery price is paramount in determining the losses on loans, it may also have an effect on the default probability. These last two columns are consistent with this hypothesis. Both the coefficient on the LTV from the origination price hedonic and the LTV using the recovery price hedonic are highly significant.

These results support the use of appraisals at origination. Both the manual and the statistical (automated) appraisals improve the predictive power of default equations. The last result suggests that in addition to retail value appraisals, lenders would be well served to have estimates of recovery values from a foreclosure when assessing default probabilities.

Conclusions

In this research, we have explored the relationship among purchase prices, appraised values and recovery prices for a large sample of manufactured homes in the United States. Our purpose has been to try to understand whether manual appraisals are an adequate and cost-effective adjunct to the underwriting process for mortgage lenders.

First, we show that simple hedonic models of prices at origination provide surprisingly good explanation of purchase prices, especially when spatial autocorrelation is included in the analysis. The year-ahead forecast performance of these hedonic models improves with sophistication but erodes over time. Our second set of results finds that recovery values are very different from origination prices. Recoveries are not a simple multiplicative transformation of origination prices. One important result is the role of atypicality, which reverses sign in the recovery price regressions. Illiquid homes sell for higher prices at origination but for large discounts in a recovery. There is also heteroskedasticity in the hedonic regressions related to the atypicality of the home. Atypical homes have a large bid–ask spread, and their prices are more difficult to explain.

Finally, we test whether defaults are best explained by purchase prices, appraised values, or recovery values. We find that indeed manual or book appraisals are a valuable adjunct to the underwriting process. There is additional information content to an appraisal for explaining defaults when the appraisal (whether manual or hedonic) price is above the purchase price, but not when below. We also find that expected recovery values have an effect on the default decision.

We thank the editor, Tom Thibodeau, the anonymous reviewers and Kelly Pace for their helpful comments and assistance.

References

Capozza, D.R., D. Kazarian and T.A. Thomson. 1997. Mortgage Default in Local Markets. *Real Estate Economics* 25(4): 631–655.

——. 1998. The Conditional Probability of Default. *Real Estate Economics* 26(3): 359–390.

Capozza, D.R. and T.A. Thomson. 2005. Optimal Stopping and Losses on Subprime Mortgages. *Journal of Real Estate Finance and Economics* 30(2): 115–131.

Chinloy, P.T., M. Cho and I.F. Megbolugbe. 1997. Appraisals, Transaction Incentives, and Smoothing. *Journal of Real Estate Finance and Economics* 14(1): 89–122.

Cole, R., D. Guilkey and M. Miles. 1986. Toward and Assessment of the Reliability of Commercial Appraisals. *The Appraisal Journal* 54(3): 422–432.

Diaz, J. III. 1997. An Investigation into the Impact of Previous Expert Value Estimates on Appraisal Judgement. *Journal of Real Estate Research* 13(1): 45–55.

Diaz, J. III and M.L. Wolverton. 1998. A Longitudinal Examination of the Appraisal Smoothing Hypothesis. *Real Estate Economics* 26(2): 349–358.

Dotzour, M.G. 1988. Quantifying Estimation Bias in Residential Appraisal. *Journal of Real Estate Research* 3(3): 1–12.

Glejser, H. 1969. A New Test for Heteroskedasticity. *Journal of the American Statistical Association* 64: 316–323.

Goodman, A.C. and T.G. Thibodeau. 1998. Dwelling Age Heteroskedasticity in Repeat Sales House Price Equations. *Real Estate Economics* 26(1): 151–171.

Guttentag, J.M. 1992. When Will Residential Mortgage Underwriting Come of Age? *Housing-Policy-Debate* 3(1): 143–156.

Hansz, J.A. and J. Diaz III. 2001. Valuation Bias in Commercial Appraisal: A Transaction Price Feedback Experiment. *Real Estate Economics* 29(4): 553–565.

Haurin, D. 1988. The Duration of Marketing Time of Residential Housing. *American Real Estate and Urban Economics Association Journal* 16(4): 396–410.

Kinnard, W.N., M.M. Lenk and E.M. Worzala. 1997. Client Pressure in the Commercial Appraisal Industry: How Prevalent is It? *Journal of Property Valuation and Investment* 15(3): 233–244.

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

Levy, D. and E. Schuck. 1999. The Influence of Clients on Valuations. *Journal of Property Valuation and Investment* 17(4): 380–400.

Noordewier, T.G., D.M. Harrison and K. Ramagopal. 2001. Semivariance of Property Value Estimates as a Determinant of Default Risk. *Real Estate Economics* 29(1): 127–159.

Pace, R.K., R. Barry, O.W. Gilley and C.F. Sirmans. 2000. A Method for Spatial–Temporal Forecasting with an Application to Real Estate Prices. *International Journal of Forecasting* 16: 229–246.

Quan, D.C. and J.M. Quigley. 1989. Inferring an Investment Return Series for Real Estate from Observations on Sales. *American Real Estate and Urban Economics Association Journal* 17(2): 218–230.

Schwert, G.W. and P.J. Seguin. 1990. Heteroskedasticity in Stock Returns. *Journal of Finance* 45: 1129–1155.

Wolverton, M. and P. Gallimore. 1999. Client Feedback and the Role of the Appraiser. *Journal of Real Estate Research* 18(3): 415–432.

Appendix

Methodology

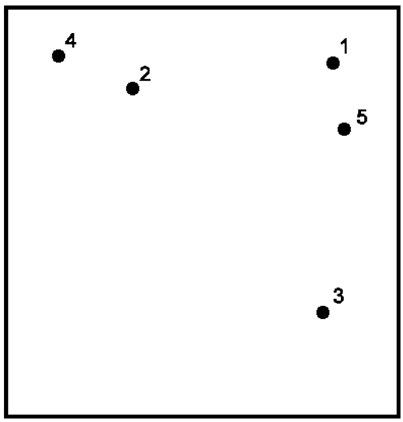
We use the following mixed regressive spatial autoregressive model:

$$\begin{aligned} Y &= \rho W Y + X\beta + \varepsilon \\ \varepsilon &\sim N(0, \sigma^2 I) \end{aligned} \tag{A.1}$$

where Y is an $n \times 1$ vector of dependent variables, β is a $K \times 1$ vector of parameters associated with the exogenous variables X ($n \times K$ matrix), and ρ is the coefficient of the spatially lagged dependent variable. The $N \times N$ matrix W is a spatial weight matrix associated with a spatial autoregressive process in the dependent variable.

Structure of the Spatial Weight Matrix

To illustrate the spatial weight matrix, consider the following figure showing the spatial location of five observations. For simplicity, assume that we are only concerned about the nearest neighbor.



Clearly, 1 and 5 are nearest neighbors to each other as are 2 and 4. 5 is the closest neighbor to 3. In a standardized spatial weight matrix, each row must sum to 1. The *ij*th matrix entry will be a 1 if the nearest neighbor to the *i*th observation is the *j*th observation. Otherwise, it will be a 0. In our example,

$$W = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Note that W might not be symmetric. In this case, 5 is the closest neighbor to 3, but the reciprocal is not true. Note also that the diagonal is zero. This is because an observation cannot be its own nearest neighbor. Using this W, we get:

$$WY = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{bmatrix} = \begin{bmatrix} y_5 \\ y_4 \\ y_5 \\ y_2 \\ y_1 \end{bmatrix}$$

Now let us see what happens when we consider the two nearest neighbors and weight them equally. The second nearest neighbor to 1 is 2 and vice versa. The second nearest neighbor to 3 and 4 is 1, and the second nearest neighbor to 5 is 3. Now, W looks like:

$$W = \begin{bmatrix} 0 & 1/2 & 0 & 0 & 1/2 \\ 1/2 & 0 & 0 & 1/2 & 0 \\ 1/2 & 0 & 0 & 0 & 1/2 \\ 1/2 & 1/2 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 \end{bmatrix}$$

Note that the rows still sum to 1. Multiplying this matrix by Y, we get:

$$WY = \begin{bmatrix} 0 & 1/2 & 0 & 0 & 1/2 \\ 1/2 & 0 & 0 & 1/2 & 0 \\ 1/2 & 0 & 0 & 0 & 1/2 \\ 1/2 & 1/2 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{bmatrix} = \begin{bmatrix} (1/2)(y_2 + y_5) \\ (1/2)(y_1 + y_4) \\ (1/2)(y_1 + y_5) \\ (1/2)(y_1 + y_2) \\ (1/2)(y_1 + y_3) \end{bmatrix}$$

So, WY is just an average of the dependent variable of the two nearest neighbors. If we give the nearest neighbor a weight of 2/3, and the second nearest neighbor a weight of 1/3, we get:

$$WY = \begin{bmatrix} 0 & 1/3 & 0 & 0 & 2/3 \\ 1/3 & 0 & 0 & 2/3 & 0 \\ 1/3 & 0 & 0 & 0 & 2/3 \\ 1/3 & 2/3 & 0 & 0 & 0 \\ 2/3 & 0 & 1/3 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{bmatrix} = \begin{bmatrix} (1/3)(y_2 + 2y_5) \\ (1/3)(y_1 + 2y_4) \\ (1/3)(y_1 + 2y_5) \\ (1/3)(y_1 + 2y_2) \\ (1/3)(2y_1 + y_3) \end{bmatrix}$$

which is a weighted average of the dependent variable of the two nearest neighbors.

In the current study, we restricted the rows of the spatial weight matrix, W, to sum to 1. Because we only wanted to model the influence of previously sold neighboring homes, we chose the 15 nearest observations occurring within the previous year and weighted them (as in Pace, Barry, Gilley and Sirmans 2000) via:

$$W_{ij} = \begin{cases} \dfrac{0.75^l}{\sum_{l=1}^{15} 0.75^l} & \text{if } j \text{ is the } l\text{th nearest observation to } i, \text{ and } l \leq 15 \\ 0 & \text{otherwise} \end{cases}$$

Because most of the entries of W are zero, we were able to use sparse matrix routines, which saved a considerable amount of computational space. We dropped the first year of observations after creating the spatial weight matrix. We did this because the estimator could perform poorly initially, as it would have a very small selection of previously sold neighbors to use in the first predictions.

Maximum Likelihood Estimation:

Equation (A.1) can be represented as:

$$Ay = X\beta + \varepsilon \tag{A.2}$$

where $A = I - \rho W$.

Solving for ε, we get:

$$Ay - X\beta = \varepsilon. \tag{A.3}$$

This is a nonlinear expression where ε is a vector of independent normal error terms. Although the error term has a well-behaved joint distribution, it cannot be observed, and the likelihood function has to be based on y. Therefore, it is necessary to introduce the concept of a Jacobian, which allows us to derive the joint distribution for the y from that for the ε. The Jacobian for the transformation of the vector of random variables ε into the vector of random variables y is:

$$J = \det(\partial\varepsilon/\partial y) = |A| = |I - \rho W|. \tag{A.4}$$

The density function of a joint normal distribution with mean zero and variance $\sigma^2 I$ is:

$$f(\varepsilon) = \frac{1}{(2\pi)^{N/2}|\sigma^2 I|^{1/2}} \exp\left\{-\frac{\varepsilon'\varepsilon}{2\sigma^2}\right\}. \tag{A.5}$$

Using the change of variable technique with Equations (A.3), (A.4) and (A.5), we can find the density function for y:

$$\begin{aligned} f(y) = {} & \frac{1}{(2\pi)^{N/2}|\sigma^2 I|^{1/2}} \\ & \times \exp\left\{-\frac{1}{2\sigma^2}(Ay - X\beta)'(Ay - X\beta)\right\} \cdot |I - \rho W|. \end{aligned} \tag{A.6}$$

Taking natural logarithms of both sides, we get the following log-likelihood function:

$$L = -\left(\frac{N}{2}\right)\ln(2\pi) - \left(\frac{N}{2}\right)\ln(\sigma^2) - \frac{1}{2\sigma^2}(Ay - X\beta)'(Ay - X\beta) + \ln|I - \rho W|. \tag{A.7}$$

In general, maximizing the log-likelihood of y would amount to minimizing the sum of squared errors, $(Ay - X\beta)'/(Ay - X\beta)$, were it not for the term involving the determinant of the Jacobian. However, in our case W is lower triangular with zeros on the diagonal, so $A = I - \rho W$ is lower triangular with ones on the diagonal. The determinant of a lower triangular matrix is simply the product of the elements of its diagonal. Therefore, the log-determinant of A is 0. So, due to the structure of the spatial weight matrix, ordinary least squares is the maximum likelihood estimate.

Atypicality

Following Haurin (1988), we created a variable to account for the difficulty that arises in valuing homes with unusual features. We use implicit marginal prices from a price hedonic regression of home sales prices on various characteristics to penalize absolute deviations from the average. We then aggregate these values. Our measure of atypicality for the ith home is as follows:

$$ATYP_i = |P_{AREA}(AREA_i - \overline{AREA})| + |P_{AGE}(AGE_i - \overline{AGE})| + |P_{DMULTI}(DMULTI - \overline{DMULTI})| \tag{A.8}$$

where $P_{AREA} = 0.00063446$, $P_{AGE} = -0.00257$ and $P_{DMULTI} = 0.34933$ are implicit marginal prices obtained from a regression of log of real purchase price on $AREA$, AGE and $DMULTI$, and where $\overline{AREA}$, $\overline{AGE}$ and $\overline{DMULTI}$ are the sample means for area, home age and for the multi-wide dummy variable, respectively.

Glejser Test

We use the Glejser test (Glejser 1969, Schwert and Seguin 1990) to detect heteroskedasticity in our model. In the first stage, we estimate the model:

$$\begin{aligned} y_i &= \rho w_i Y + x_i\beta + \varepsilon_i \\ \varepsilon_i &\sim N\left(0, \sigma_i^2\right) \end{aligned} \qquad i = 1, 2, \ldots, n \tag{A.9}$$

We then use the residuals from this regression, $e = Y - \hat{Y}$, in a Glejser regression:

$$(\pi/2)^{1/2}|e| = X\beta + v. \tag{A.10}$$

We use $(\pi/2)^{1/2}|e|$ as an estimate of the standard deviation of e. If $e_i \sim N(0, \sigma_i^2)$, then $E[(\pi/2)^{1/2}|e_i|] = \sigma_i$. So, we are regressing the estimated variance on our independent variables. We find that the variable *atypicality* is the source of the heteroskedasticity, and we run a weighted least squares regression (WLS) of the model in Equation (A.9) using *atypicality* as the weight.