CREDIT SUPPLY AND HOUSE PRICES: EVIDENCE FROM MORTGAGE MARKET SEGMENTATION ONLINE APPENDIX

Manuel Adelino*   Antoinette Schoar   Felipe Severino

Duke University   MIT and NBER   MIT

August 28, 2012

1 Robustness and Refinements - Additional Tests

1.1 Restrict LTV Choices

We want to test that our estimates are not driven by borrowers with very unusual LTV levels, namely those with LTV below 50 percent and above 80 percent. Borrowers with those choices of LTV are likely to either have access to abundant equity to put up when buying a home, or to be very constrained and need a very high LTV. By limiting our sample to include only borrowers who choose a first lien LTV between 50 and 80 percent, we capture the transactions that should be most affected by the conforming loan limit. In particular, this subsample includes the group of borrowers that end up with an LTV between 77 percent and 79.5 percent in the year that the CLL is in effect because they stick with a conforming loan, even though their house costs more than 125 percent of the CLL. This choice of LTV is very common for the "Above the Threshold" group of borrowers in the year that the limit is in effect, but very infrequent everywhere else in the distribution of transactions. Also, this subsample includes all the borrowers that choose an 80 percent LTV, the most frequent choice in the data. This means getting a jumbo loan for transactions "Above the Threshold" and a conforming loan for transactions below that threshold. Finally, the transactions that are excluded from this sample should be least affected by the conforming loan limit, either

*Corresponding author: 100 Fuqua Drive, Durham, NC 27708. manuel.adelino@duke.edu.

because their LTVs are very low, in which case they are never affected by the limit anyway, or alternatively, because they have high LTVs and thus obtain jumbo loans in the year in which the limit is in effect whether the price of the transactions is above or below the 125 percent of the CLL threshold.

Table 1 shows the results for Fama-MacBeth coefficients from year-by-year regressions, much like we described in the Main Results section of the paper, except using only transactions with an LTV between 0.5 and 0.8. The results are quantitatively similar to those we obtain for the whole sample, which means that our main results are not being driven by very low or very high LTVs. This reinforces our interpretation that our main results are caused by the CLL and not some other spurious factor. The magnitude of the coefficients is very similar to the ones in the previous table, but we lose statistical significance for the coefficient of interest when we use the "Value Residual" measure as the left-hand side measure.

1.2   Different Bands

Table 3 shows that the result is very stable as we move away from the threshold of CLL/0.8. In fact, the point estimates are indistinguishable from each other whether we use a band of USD 5,000 or USD 10,000, which suggests that the difference in the cost of credit is likely to be similar for these two sets of buyers relative to buyers below the threshold. This is further evidence that the result is not driven solely by buyers who choose to obtain a conforming mortgage and put up additional equity from other sources.

1.3   Timing of the Control Group

We run a final robustness test in which, instead of comparing the year in which the limit is in effect with the subsequent year, we compare it to the previous year. In this way, we are comparing houses that are never eligible for an 80 percent conforming loan (those above the threshold) to transactions that initially are not eligible, but become eligible once the limit changes. The research design is the same as before, but we shift the window of analysis back one year. Table 2 shows the Fama-MacBeth coefficients for this specification. The point estimates are smaller than the ones in Table 2, but they are in the same direction and remain statistically significant for the first years in the sample.

1.4   Value per Square Foot by ZIP Code Income

In Figure 2, we split ZIP codes by their median income in order to consider the effect of the conforming loan limit on the distribution of value per square foot on the whole sample

of transactions. We plot the average value per square foot as a function of the distance of each transaction to the threshold of 125 percent of the CLL. We can see that for the ZIP codes in the lowest quartile of the income distribution, the average value per square foot is monotonically increasing for up to conforming loan limit threshold, and from this point onwards the distribution becomes flat. This pattern is not visible for zip codes with higher median incomes, where the distribution seems monotonically increasing both below and above the threshold.

2   Data Manipulation

2.1   Data Cleaning

In order to clean the raw data received from Dataquick, we perform the following modifications to the data:

Table 0: Data Cleaning Description

Criterion	Deleted Observations	Remaining Observations
Initial data		**11,884,730**
Transaction value equal to zero	1,365,429	10,519,301
Missing zipcode	18,766	10,500,535
Missing square feet	1,509,732	8,990,803
Mislabeled year	5	8,990,798
First loan greater than transaction value	353,552	8,637,246
House of less than 500 square feet	47,059	8,590,187
Transaction greater than 1,2 MM and smaller than 30 M	381,786	8,208,401
Company owned observation based on Dataquick flag	451,295	7,757,106
Company owned obs based on owner/seller/buyer information	746,754	7,010,352
Simple duplicated transactions	0	7,010,352
Value per square feet yearly outliers	142,079	6,868,273
Same property, date and buyer/seller information	11,577	6,856,696
Same property, and date and no seller information	364	6,856,332
Same property, date and transaction value	41,855	6,814,477
Same property, date and A sell to B and B sell to C	22,258	6,792,219
Special Transaction, based on Dataquick flag	609	6,791,610
Same property and date, multiple sales in a day	248	6,791,362
Clean data		**6,791,362**
Remove single-family houses	1,751,670	5,039,692
Transaction greater than 600 M and smaller than 130 M	1,056,117	3,983,575
Whole sample for hedonic regressions		**3,983,575**
Transactions outside the 10k band for each year	3,742,840	240,735
Transactions used twice (treatment in year t and control in year t+1	+21,936	262,671
Regression sample		**262,671**

Note: This table enumerates the steps taken in the data cleaning process and gives the number of observations that are dropped in each step, as well as the remaining observations after each step.

Table 0 shows the number of observations deleted in each step of the data preparation and a basic description of the criterion used to drop those observations from the sample. In the following paragraphs, we categorize each step and describe the criteria we used in detail, providing additional information about the data construction. We start with 11,884,730 ob-

servations.

Missing observations and outliers

- · We drop records with missing transaction value, house size, zip code, property unique identifier, or mislabeled year.

- · We drop a record if the house size is smaller than 500 square feet, as well as records with transaction values smaller than three thousand and greater than one million and two hundred thousand dollars.

- · Value per square foot outliers per year: We drop observations that are above the ninety-ninth percentile for the value per square foot variable or below the first percentile each year.

Company owned observations

- · We drop observations that Dataquick identifies as being bought by a corporation.

- · Company owned observations based on owner/seller/buyer information: If the owner, seller, or buyer names contain LLC, CORP, or LTD, the observation is removed from the sample.

Duplicate transactions

- · Simple duplicated transactions: Remove records for which all the property information is the same.

- · Same property, date, and buyer/seller information: Drop observations that are duplicated based on transaction value, date, and buyer/seller information.

- · Same property and date, no seller information: Drop observations for which the property unique identifier and date are the same and have no seller information.

- · Same property, date, and transaction value: Drop observations for which property unique identifier, date, and transaction value are the same.

- · Same property and date and A sells to B and B sells to C: If person A sells to B and B sells to C in the same date, we keep the most recent transaction.

· Special transaction, based on Dataquick flag: This flag allows us to identify records that are not actual transactions. For example, if a transaction was only an ownership transfer without a cash transfer, this field is populated, allowing us to delete this transaction.

· Same property and date, multiple sales in a day: If a property is sold more than twice during the same day, we keep only one transaction.

Additional information

· We merge the Metropolitan Statistical Area (MSA) classification obtained from the Census Bureau definition, using FIPS unique code identifier by county[1].

· Change the second lien amount to missing if the first loan amount is equal to the second loan amount, or if the second loan amount is greater than the transaction value.

· Change the second lien amount to missing if combined loan to value (CLT) is greater than two and loan to value (LTV) is equal to one.

· Change house age to missing if house age, calculated using transaction year minus year built, is smaller than zero.

This procedure gives us our clean sample with 6,791,362.

Whole Sample for Hedonic Regression Sample

· We further restricted the sample for the hedonic regressions to transactions that are between one hundred and thirty thousand and six hundred thousand dollars. This selection aims to avoid that the estimates from the hedonic regression be driven by transactions that are far from the region of interest.

This gives us our whole sample with 3,983,575 observations that are summarized in the Summary Statistics section of the paper.

Regression Sample

---

[1]FIPS county code is a five-digit Federal Information Processing Standard (FIPS) code which uniquely identifies counties and county equivalents in the United States, certain U.S. possessions, and certain freely associated states. The first two digits are the FIPS state code and the last three are the county code within the state or possession.

· Non-single-family houses: Our identification strategy relies on the change in the conforming loan limit for single-family houses, therefore, we restrict our attention to this type of house.

· Transactions outside the USD 10,000 band for each year: Based on the threshold value for each year that we describe in the Identification Strategy subsection, we define a relevant transaction band around that threshold. For example, in 1999 the house threshold (1.25 of the conforming loan limit) is 300,000 dollars. Therefore, we keep records with transaction values between 290,000 and 310,000 dollars that happened between 1999 and 2000. This subsample will be the sample used to run the differences-in-differences specification using the 1999 threshold. For years when transaction bands overlapped, transaction will be treatment in year t and controls in year t+1, and therefore used twice in the empirical strategy

This gives us our regression sample with 262,671 observations

2.2   Variable Construction

In this appendix, we describe in more detail the variables used in the hedonic regressions. The hedonic regressions use two left-hand side variables: value per square foot and price of each transaction. As we pointed out when we describe the hedonic regression in the paper (Section 3.2), we use a similar set of controls as those used in Campbell, Giglio, and Pathak (2010), and we add a few more characteristics.

The variables we use are interior square feet (linearly, high and low square feet dummies), lot size, bedrooms, bathrooms, total rooms, house age (linearly and squared), type of house, an indicator for whether the house was renovated, an indicator for fireplace and parking, indicators for style of building (architectural style and structural style), and additional indicators for type of construction, exterior material, heating and cooling, heating and cooling mechanism, type of roof, view, attic, basement, and garage.

While interior square feet, lot size, and age are included as continuous variables, all the other controls are included as indicator variables.

· *Type of house:* This variable is 1 if the house is a single-family house and 0 if it is a condo or a multifamily property.

· *Bedrooms:* This characteristic is divided into four categories (dummies): one bedroom, two bedrooms, three bedrooms, and more than three bedrooms.

· *Bathrooms:* This characteristic is divided into four categories: one bathroom, one and a half bathrooms, two bathrooms, and more than two bathrooms.

· *Rooms:* This characteristic is divided into five categories (dummies): one room, two rooms, three rooms, four rooms, and more than four rooms.

· *Building Shape, Architectural Code, Structural Code, Exterior Material, Construction Code, Roof Code, View Code:* These characteristics were divided based on the numeric categorization of the original field. For example, construction code was divided into 10 different categories that indicated the material used on the framework of the building. In this case, we created 10 dummies based on this categorization.

· *Heating and cooling:* This information was divided into four categories: only heating, only cooling, both heating and cooling, and heating-cooling information missing. The last variable was created to avoid dropping transactions for which the information was not available.

· *Heating and cooling type:* These characteristics were divided based on the numeric categorization of the original field. In this case, they discriminate the type of cooling or heating system that is being used in the house.

· *Garage and Garage Carport:* A dummy is created to account for houses that have garage surface greater than 50 square feet. For those transactions without the information, a missing dummy is created for this category. Finally, we used additional information to create a dummy that indicates if the houses have a garage carport or not.

· *Renovation:* This variable accounts for the number of years since the last renovation. Based on this continuous variable, five categories (dummies) are defined: missing renovation if the renovation date is missing or renovation period is negative, last renovation in less than 10 years, renovated between 10 and 20 years, renovated between 20 and 30 years , and last renovation in more than or equal to 30 years.

· *Attic:* This characteristic is accounted for using a dummy for houses with an attic greater than 50 square feet, and another dummy to account for missing information about the attic in the houses.

· *Basement Finished and Unfinished:* For the finished basement information, we created a dummy for houses with basement size greater than 100 square feet, and another dummy to account for missing information about the finished basement. The same procedure is used to incorporate the information about unfinished basement.

We use both the price of a transaction as well as the value per square foot as our dependent variables. By estimating these regressions by year and by Metropolitan Statistical Areas (MSA), we allow the coefficients on the characteristics to vary along these two dimensions. We included monthly indicator variables to account for seasonality in the housing market, as well as zip code fixed effects. The set of controls X_i is composed of all the variables described above, but in the case of the value per square foot regression, we exclude the interior square feet continuous variables.

$$LHS_i = \gamma_0 + \Gamma X_i + month_i + zipcode_i + \varepsilon_i$$

When a record is missing the interior square feet, the lot size, the number of bedrooms or bathrooms, or information on a houses age, we do not include this observation in the hedonic regressions. This explains the difference between the number of observations for the value per square foot hedonic regressions (where we exclude interior square footage) and the transaction value residual in our main regression results.

Figure 1: Fraction of Transactions with a Second Lien Loan by Year
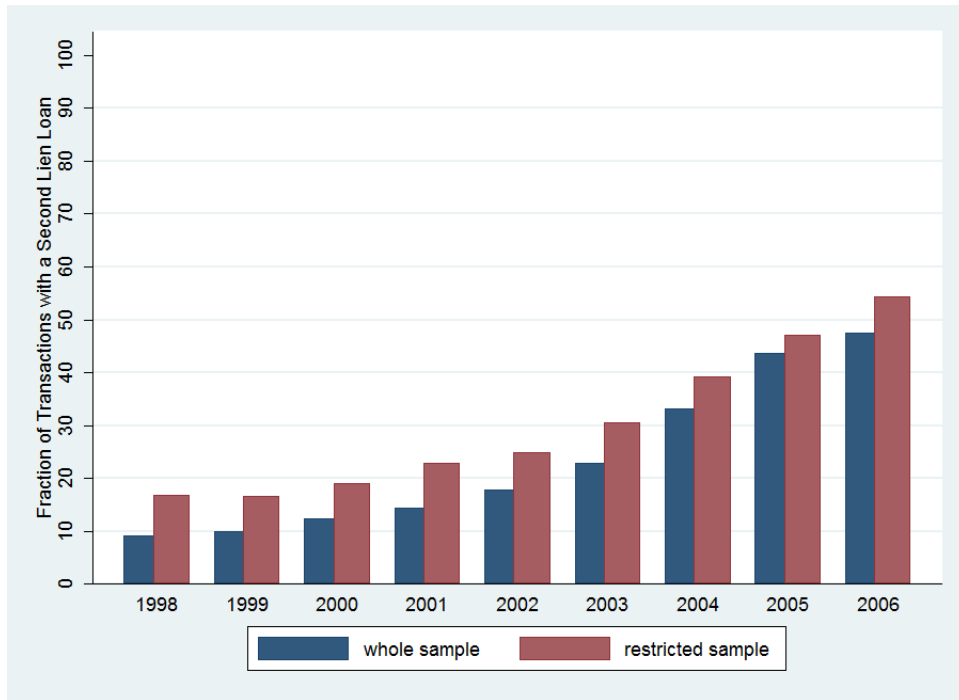
Note: This figure shows the average fraction of transactions with a second lien loan by year for the whole sample and the restricted sample used in the regression. Years 2007 and 2008 are excluded from the regression sample because there was no change on the conforming loan limits on those years

Figure 2: Value per Square Foot by House Value and by ZIP Code Income
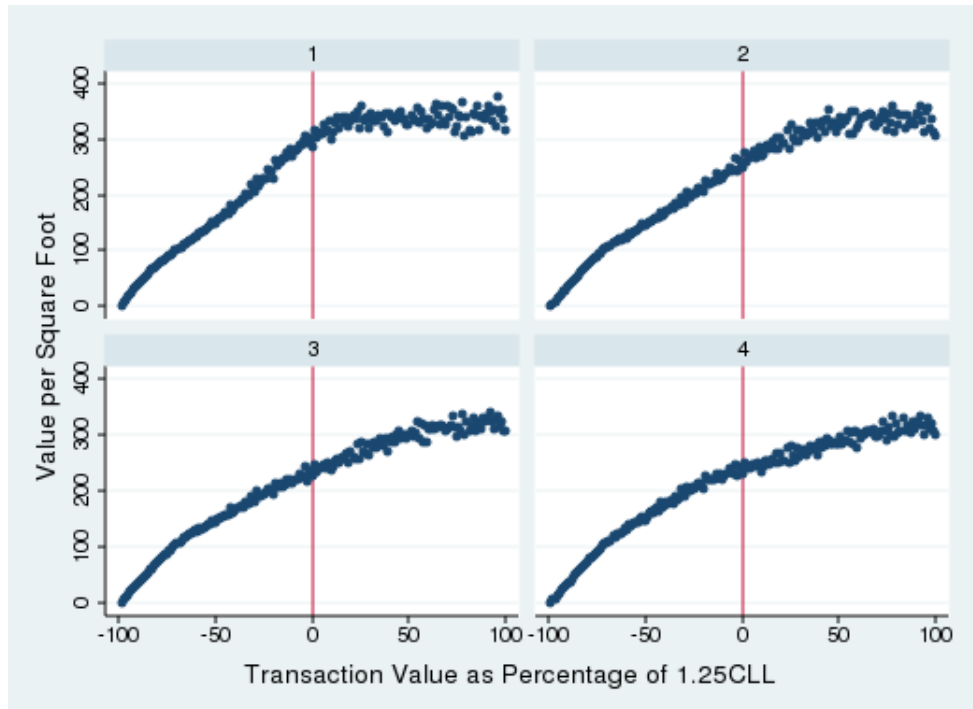
Note: This figure shows the average value per square foot plotted against the value of the house. We split ZIP codes into quartiles according to their median income, where 1 includes the ZIP codes in the lowest income quartile and 4 includes the ZIP codes with the highest median income. We use the average of the median yearly income over the whole sample to place ZIP codes into the quartiles. The x-axis is represented as one minus the transaction value as a percentage of each year's threshold of 125 percent of the conforming loan limit (e.g. if the threshold is 200,000, a transaction of 150,000 will appear as -25 percent). The vertical red line is the threshold and the transactions for all years are centered around that value.

Table 1: Effect of the CLL on House Valuation Measures, Constrained Sample (0.5<LTV≤0.8)

Panel A: Value Per Square Foot

	All years	1998-2001	2002-2005
Above Threshold	0.956**	1.584***	0.328
	(0.462)	(0.556)	(0.650)
Year CLL	-24.627***	-15.935***	-33.319***
	(4.386)	(2.576)	(5.726)
Above Threshold x	-1.257***	-1.610**	-0.904
Year CLL	(0.422)	(0.646)	(0.576)
No. Obs.	190,450	75,304	115,146

Panel B: Log of Transaction Value Residual from Hedonic Regressions

	All years	1998-2001	2002-2005
Above Threshold	0.0118***	0.0145***	0.0090***
	(0.0014)	(0.0017)	(0.0007)
Year CLL	0.0367***	0.0335***	0.0398***
	(0.0038)	(0.0040)	(0.0066)
Above Threshold x	-0.0017	-0.0019	-0.0015*
Year CLL	(0.0011)	(0.0022)	(0.0008)
No. Obs.	183,643	71,843	111,800

Panel C: Value Per Square Foot Residual from Hedonic Regressions

	All years	1998-2001	2002-2005
Above Threshold	1.565***	1.958***	1.172***
	(0.298)	(0.356)	(0.431)
Year CLL	3.431***	3.470***	3.392***
	(0.550)	(0.417)	(1.113)
Above Threshold x	-0.931***	-1.085***	-0.777**
Year CLL	(0.260)	(0.413)	(0.360)
No. Obs.	183,789	71,917	111,872

Note: This table shows Fama Macbeth coefficients computed from year by year regressions that use three alternative measures of valuation as the dependent variable in each of the three panels. The hedonic regressions that produce the residuals for panels B and C are described in Section 3.2. The sample for each year's regression includes transactions within +/- USD 10,000 of that year's conforming loan limit, as well as transactions in the same band in the subsequent year. Unlike the main regression table in the paper, the sample for these regressions is constrained to transactions with an LTV between 0.5 and 0.8. All year by year regressions include ZIP code fixed effects. Above the Threshold refers to transactions up to USD 10,000 above the conforming loan limit divided by 0.8 (i.e. the transactions that were "ineligible" to be bought with a conforming loan at a full 80 percent LTV) and Year CLL is the year in which the conforming loan limit is in effect.

Table 2: Effect of CLL on Alternative Valuation Measures - Different Control Group

Panel A: Value Per Square Foot

	All Transactions			0.5<LTV≤0.8 Transactions		
	All years	*1999-2002*	*2003-2006*	*All years*	*1999-2002*	*2003-2006*
Below Threshold	0.012	-0.005	0.029	0.522*	0.628	0.417
	(0.236)	(0.282)	(0.423)	(0.270)	(0.412)	(0.404)
Pre-Year CLL	-23.739***	-15.890***	-31.588***	-25.061***	-16.995***	-33.127***
	(4.391)	(2.489)	(6.534)	(4.636)	(2.666)	(7.057)
Below Threshold X	-0.375	-0.817	0.068	-0.555	-0.812***	-0.298
Pre-Year CLL	(0.473)	(0.549)	(0.783)	(0.434)	(0.233)	(0.884)
No. Obs.	227,325	93,612	133,713	168,865	66,072	102,793

Panel B: Transaction Value Residual from Hedonic Regressions

	All Transactions			0.5<LTV≤0.8 Transactions		
	All years	*1999-2002*	*2003-2006*	*All years*	*1999-2002*	*2003-2006*
Below Threshold	-0.0099***	-0.0106***	-0.0092***	-0.0086***	-0.0087***	-0.0085***
	(0.0010)	(0.0010)	(0.0017)	(0.0011)	(0.0018)	(0.0015)
Pre-Year CLL	0.0346***	0.0342***	0.0350***	0.0342***	0.0334***	0.0350***
	(0.0045)	(0.0037)	(0.0089)	(0.0045)	(0.0042)	(0.0088)
Below Threshold X	0.0000	-0.0019	0.0019	-0.0011	-0.0031	0.0009
Pre-Year CLL	(0.0016)	(0.0021)	(0.0023)	(0.0016)	(0.0023)	(0.0020)
No. Obs.	217,410	88,416	128,994	162,584	62,897	99,687

Panel C: Value Per Square Foot Residual from Hedonic Regressions

	All Transactions			0.5<LTV≤0.8 Transactions		
	All years	*1999-2002*	*2003-2006*	*All years*	*1999-2002*	*2003-2006*
Below Threshold	-0.903***	-0.881***	-0.925	-0.524**	-0.446**	-0.603
	(0.289)	(0.197)	(0.593)	(0.208)	(0.206)	(0.395)
Pre-Year CLL	3.215***	3.019***	3.411**	2.852***	2.591***	3.112**
	(0.712)	(0.529)	(1.436)	(0.699)	(0.547)	(1.392)
Below Threshold X	-0.175	-0.605**	0.256	-0.467	-0.915***	-0.020
Pre-Year CLL	(0.351)	(0.245)	(0.625)	(0.315)	(0.130)	(0.560)
No. Obs.	217,804	88,613	129,191	162,788	62,997	99,791

Note: Table shows Fama McBeth coefficients computed from year by year regressions that use three alternative measures of valuation as the dependent variable in each of the three panels. The sample includes all transactions within USD 10,000 of each year's conforming loan limit, as well as transactions of the same amount in the *previous year* (unlike the previous tables where we use the subsequent year). In this table we include the results for all transactions, as well as those for the sample that is restricted to having an LTV between 0.5 and 0.8. Below the Threshold refers to transactions up to USD 10,000 below the conforming loan limit at year t divided by 0.8 (i.e. the transactions that were "eligible" to be bought with a conforming loan at a full 80 percent LTV in year t , but were "ineligible" in year t-1) and Pre-Year CLL is the previous year in which the conforming loan limit is in effect. This specification makes the interaction coefficient directly comparable to the main regression on signs and magnitudes.

Table 3: Effect of the CLL on House Valuation Measures Different Band Samples

Panel A: Value Per Square Foot

	10K	0k to 5K	5K to 10K
Above Threshold	1.261**	0.969	1.406***
	(0.494)	(0.722)	(0.544)
Year CLL	-22.869***	-23.008***	-23.194***
	(4.047)	(3.988)	(4.177)
Above Threshold x	-1.162***	-1.064*	-1.181**
Year CLL	(0.264)	(0.556)	(0.581)
No. Obs.	262,671	134,117	128,554

Panel B: Log of Transaction Value Residual from Hedonic Regressions

	10K	0k to 5K	5K to 10K
Above Threshold	0.0129**	0.0071	0.0180***
	(0.0013)	(0.0019)	(0.0013)
Year CLL	0.0387***	0.0384***	0.0389***
	(0.0041)	(0.0045)	(0.0038)
Above Threshold x	-0.0017***	-0.0015*	-0.0023**
Year CLL	(0.0008)	(0.0011)	(0.0016)
No. Obs.	251,431	128,429	123,002

Panel C: Value Per Square Foot Residual from Hedonic Regressions

	10K	0k to 5K	5K to 10K
Above Threshold	1.733***	1.255*	2.110***
	(0.360)	(0.700)	(0.387)
Year CLL	4.103***	4.052***	3.946***
	(0.644)	(0.678)	(0.763)
Above Threshold x	-0.651***	-0.712	-0.623***
Year CLL	(0.238)	(0.508)	(0.238)
No. Obs.	251,764	128,601	123,163

Note: This table shows Fama MacBeth coefficients computed from year by year regressions that use three alternative measures of valuation as the dependent variable in each of the three panels. The hedonic regressions that produce the residuals for panels B and C are described in Section **??**. The sample for each year's regression includes all transactions within +/- USD 10,000 of that year's conforming loan limit, as well as transactions in the same band in the subsequent year. All year by year regressions include ZIP code fixed effects. Above the Threshold refers to transactions up to USD 10,000 above the conforming loan limit divided by 0.8 (i.e. the transactions that were "ineligible" to be bought with a conforming loan at a full 80 percent LTV) and Year CLL is the year in which the conforming loan limit is in effect.