Credit Supply and House Prices: Evidence from Mortgage Market Segmentation Online Appendix

Manuel Adelino* Duke University and NBER
Antoinette Schoar MIT and NBER
Felipe Severino Dartmouth College

November 26, 2020

1 Robustness and refinements: Additional tests

1.1 Restrict LTV choices

We want to test that our estimates are not driven by borrowers who have very unusual LTV levels, namely, those with an LTV below 50% or above 80%. Borrowers with these LTV choices are likely to either have access to abundant equity to put up when buying a home or be very constrained and need a very high LTV. By limiting our sample to borrowers who choose a first lien LTV between 50% and 80%, we capture those 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% and 79.5% in the year that the CLL is in effect because they stick with a conforming loan, even though their house costs more than 125% 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 borrowers who choose an 80% 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 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% of the CLL threshold. Table 1 shows the results for Fama-MacBeth coefficients from year-by-year regressions, much like we described in Section 5.1 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 2 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 US$5,000 or US$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 solely driven by buyers who choose to obtain a conforming mortgage and put up additional equity from other sources.

1.3 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% 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 onward, 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

To clean the raw data received from Dataquick, we perform the following modifications to the data: Table 0 shows the number of observations deleted in each step of the data

Table 0: Data cleaning description.

Criterion	*Deleted observations*	*Remaining observations*
Initial data		**11,820,578**
Transaction value equal to zero	1,365,335	10,457,243
Missing ZIP code	16,987	10,440,256
Missing square feet	1,499,110	8,941,146
First loan greater than the transaction value	350,923	8,590,223
House ¿500 sq ft	46,616	8,543,607
Transaction ¡$1,200,000 and ¿$300,000,00	381,151	8,162,456
Company owned obs. based on Dataquick flag	451,113	7,711,343
Company owned obs. based on owner/seller/buyer information	740,028	6,971,315
Value per square feet yearly outliers	141,305	6,830,010
Same property, date, and buyer/seller information	11,521	6,818,489
Same property and date and no seller information	364	6,818,125
Same property, date, and transaction value	41,777	6,776,348
Same property, date, and A sell to B and B sell to C	22,207	6,754,141
Special Transaction, based on Dataquick flag	609	6,753,532
Same property and date, multiple sales in a day	244	6,753,288
Clean data		**6,753,288**
Remove single-family houses	1,745,009	5,008,279
Transaction ¡$600 M and ¿$130 M	1,051,758	3,956,521
Whole sample for hedonic regressions		**3,956,521**
Transactions outside the 10k band for each year	3,706,036	250,485
Transactions used twice (treatment in year t and control in year t+1	+10,919	261,404
Regression sample		**261,404**

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.

preparation and a basic description of each criterion used to drop 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 11,820,578 observations.

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 less than 500 sq ft, as well as records with transaction values less than $3,000 and greater than $1,200,000.
- Value per square foot outliers per year: We drop observations that are above the 99th 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 a 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 ÔÔcash" 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 restrict the sample for the hedonic regressions to transactions that are between $130,000 and $600,000. 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 Section 3.1 of the main paper.

Regression Sample

- 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 US$10,000 band for each year: Based on the threshold value for each year that we describe in the empirical 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. Therefore, we keep records with transaction values between $290,000 and $310,000 that happened between 1999 and

[1] The FIPS county code is a five-digit Federal Information Processing Standard (FIPS) code that 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.

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, the transaction will be the treatment in year t and the controls in year $t+1$, and therefore used twice in the empirical strategy.

This gives us our regression sample with 261,404 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: the value per square foot and the 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 the style of building (architectural style and structural style), and additional indicators for type of construction, exterior material, heating and cooling, heating and cooling mechanisms, type of roof, view, attic, basement, and garage.

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

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

· *Bedrooms:* This characteristic is divided into four categories (dummies) depending on the number of bedrooms: one, two, three, or more than three bedrooms.

· *Bathrooms:* This characteristic is divided into four categories depending on the number of bathrooms: one, one and a half, two, or more than two bathrooms.

· *Rooms:* This characteristic is divided into five categories (dummies) depending on the number of rooms: one, two, three, four, or 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, the construction code was divided into 10 different categories that indicate the material used on the framework of the building. In this case, we create 10 dummies based on this categorization.

- *Heating and cooling:* This category was further divided into four subcategories: 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 highlight the type of cooling or heating system that is being used in the house.

- *Garage and Garage Carport:* This dummy accounts for houses that have a garage surface greater than 50 sq ft. For those transactions without the information, a missing dummy is created for this category. Finally, we create a dummy that indicates whether or not a house has a garage carport.

- *Renovation:* This continuous variable accounts for the number of years since the last renovation. 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 sq ft, and another dummy to account for missing information about the attic in the houses.

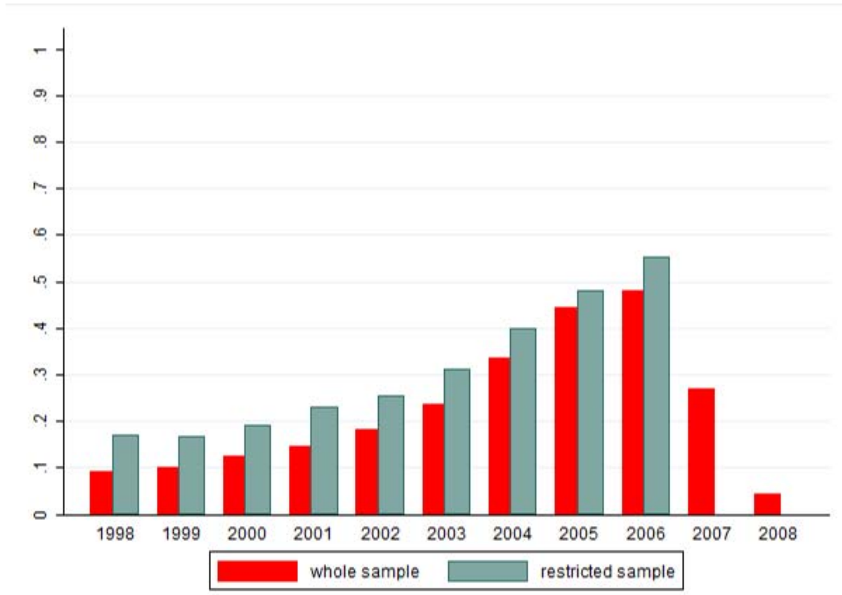
- *Basement Finished and Unfinished:* For the finished basement information, we create a dummy for houses with a basement size greater than 100 sq ft and another dummy to account for missing information about the finished basement. The same procedure is used to incorporate the information about unfinished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

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

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

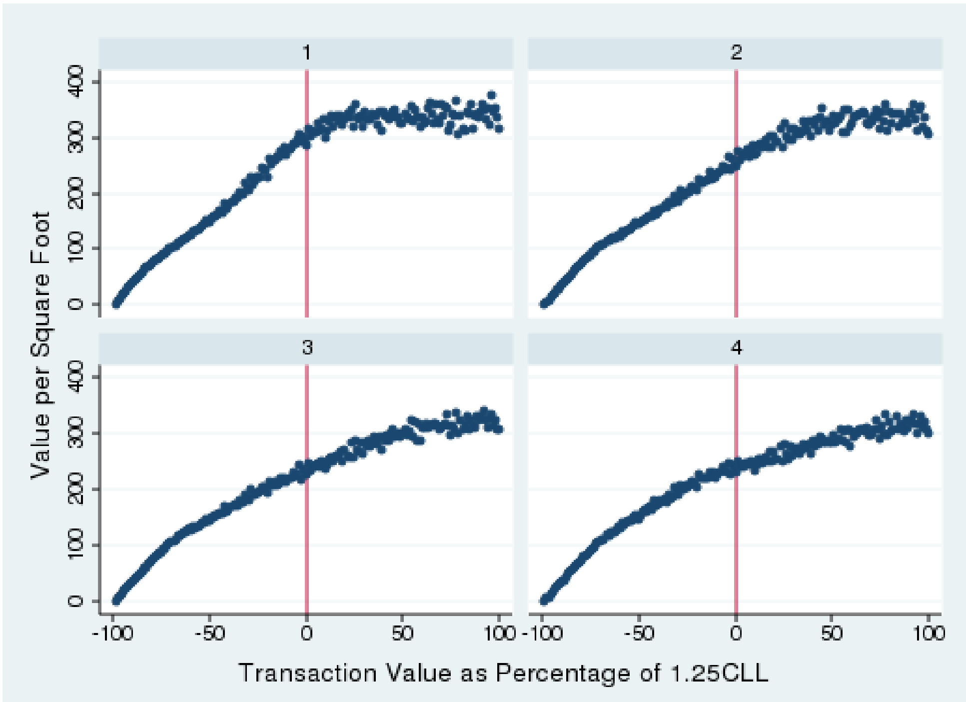
When a record is missing the interior square feet, the lot size, the number of bedrooms or bathrooms, or information on a house's age, we exclude this observation from the hedonic regressions. This explains the difference between the number of observations for the value per square foot hedonic regressions (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.



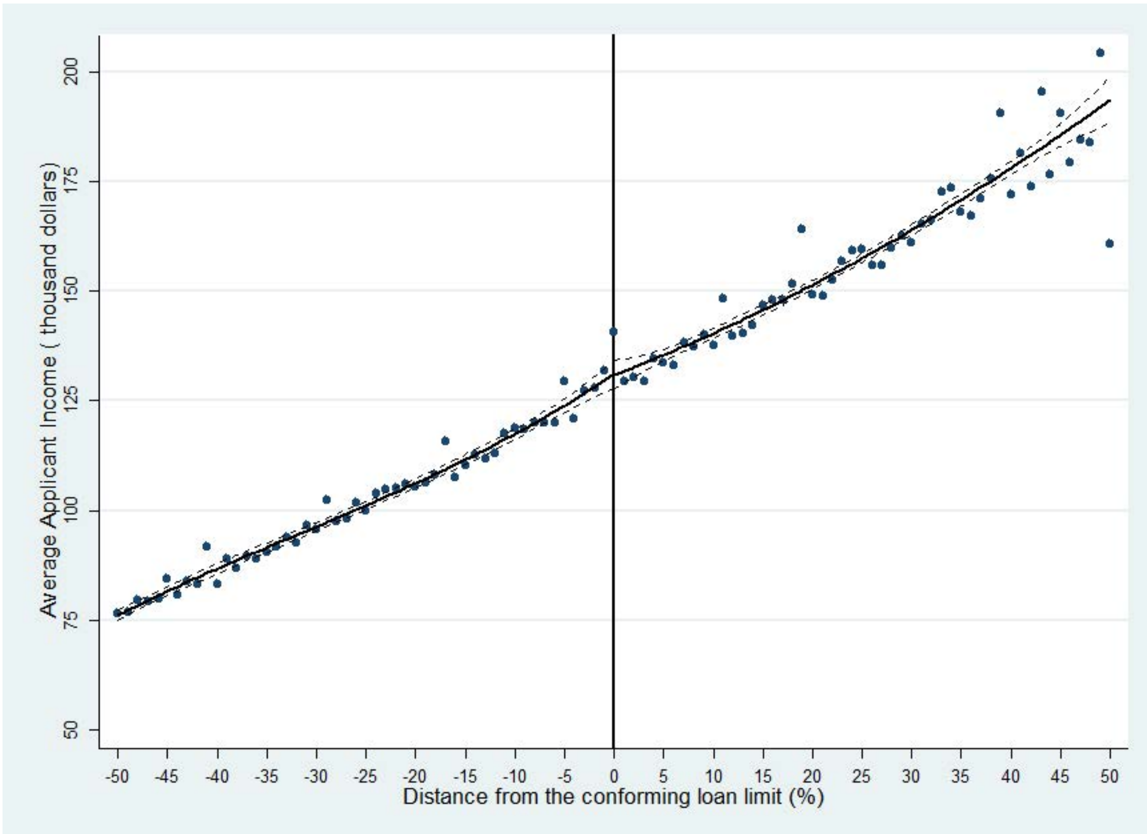
This figure charts the average fraction of transactions with a second lien loan by year for the whole sample and the restricted sample used in the regression. The conforming loan limits do not change in years 2007 and 2008, so we exclude these years from the regression sample.

Figure 2: Value per square foot by house value and by ZIP code income.



This figure plots the average value per square foot against the value of the house. We split ZIP codes into quartiles by their median income, where one includes the ZIP codes in the lowest income quartile and four 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 represents one minus the transaction value as a percentage of each year's threshold of 125% of the conforming loan limit (e.g., if the threshold is 200,000, a transaction of 150,000 will appear as -25%). The vertical red line represents the threshold, and the transactions for all years are centered on that value.

Figure 3: Income as a percentage of the CLL threshold.



The horizontal axis represents the difference between loan amounts and the conforming loan limit as a percentage of the conforming loan limit. The figure plots the average mortgage applicant income, which we compute from HMDA mortgage applications. We aggregate these proportions into 1% bins, and each dot in the figure represents the share of unused mortgages for each bin. We also plot third-degree polynomials (to the left and right of the conforming loan limit) as well as the 95% confidence intervals (dashed lines). Data are extracted from the HMDA, 1998–2006.

Table 1: Effect of the CLL on house valuation measures, constrained sample $(0.5<\text{LTV}\leq 0.8)$.

.

Value per sq ft	*log of Trans value residual*	*Value per sq ft residual*
0.999**	0.012***	1.618***
(0.464)	(0.001)	(0.304)
-24.643***	0.037***	3.494***
(4.390)	(0.004)	(0.573)
-1.292***	-0.002	-0.984***
(0.449)	(0.001)	(0.272)
189,594	182,852	182,998

This table shows Fama-Macbeth coefficients, which are computed from year-by-year regressions that use three alternative measures of valuation as the dependent variable. The hedonic regressions that produce the residuals are described in Section 3.2. The sample for each year's regression includes transactions within +/- US$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 in the paper, these regressions use a sample that 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 US$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% LTV), and Year CLL is the year in which the conforming loan limit is in effect.

Table 2: Effect of the CLL on valuation: Alternative bands

Panel A: Value per square foot

	10K	*0k to 5K*	*5K to 10K*
Above threshold	1.281***	0.971	1.414**
	(0.483)	(0.716)	(0.550)
Year CLL	-22.891***	-23.035***	-23.217***
	(4.068)	(4.003)	(4.199)
Above threshold x	-1.173***	-1.064*	-1.185**
Year CLL	(0.269)	(0.561)	(0.588)
No. obs.	261,404	133,466	127,938

Panel B: Logarithm of transaction value residual from hedonic regressions

	10K	*0k to 5K*	*5K to 10K*
Above threshold	0.0129***	0.0070	0.0181**
	(0.0012)	(0.0018)	(0.0015)
Year CLL	0.0392***	0.0392***	0.0391***
	(0.0042)	(0.0046)	(0.0038)
Above threshold x	-0.0017***	-0.0016*	-0.0021**
Year CLL	(0.0007)	(0.0010)	(0.0014)
No. obs.	250,281	127,846	122,435

Panel C: Value per square foot residual from hedonic regressions

	10K	*0k to 5K*	*5K to 10K*
Above threshold	1.754***	1.265*	2.122***
	(0.357)	(0.697)	(0.393)
Year CLL	4.147***	4.085***	3.987***
	(0.647)	(0.686)	(0.767)
Above threshold x	-0.669***	-0.723	-0.635***
Year CLL	(0.249)	(0.513)	(0.246)
No. obs.	250,607	128,013	122,594

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 all transactions within +/- US$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 US$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% LTV), and Year CLL is the year in which the conforming loan limit is in effect.

Table 3: Summary statistics predicted value assignment.

Panel A. House characteristics

	Regression sample, N = 244,476		
	Mean	SD	Median
Transaction value (US$1,000)	413.82	89.53	419.90
Loan to value	0.75	0.14	0.80
House size (sq ft)	2,006	705	1,906
Lot size (sq ft)	12,094	17,965	7,425
Number of rooms	7.36	1.66	7.00
Number of bedrooms	3.38	0.79	3.00
Number of bathrooms	2.18	1.04	2.00
House age (years)	34.83	27.45	34.00

Panel B. House valuation

	Regression sample, N = 244,476		
	Mean	SD	Median
Value per sq ft (US$/sq ft)	235.24	103.67	214.33
Value per sq ft residual (US$/sq ft)	-0.23	48.75	-0.99
log of transaction value residual (US$)	0.02	0.18	0.03

Panel A shows the descriptive statistics for the transactions used on our alternative treatment assignment regressions in our data from 1998 to 2008. The data were extracted from deeds records by Dataquick. Panel B shows the different valuation measures we use in the regression analysis. Value per square foot is the transaction amount divided by the size of the house measured in square feet. Both residual measures are obtained from hedonic regressions run by year and by metropolitan area for the value per square foot and transaction value on a set of detailed house characteristics. Section 2 of the main paper explains more information about the construction of the residuals.