AVM versus Appraisal-Based Underwriting in Re finance Mortgages: The Trade-off Between Noise and Bias.

George Krivorotov * Michael Lacour-Little †‡

First draft: March 1, 2017
This draft: December 22, 2020

Abstract

Extensive research has confirmed that there is significant anchoring bias in purchase mortgage appraisals, with appraisers setting the value at or above the contract price the vast majority of the time. However, understanding appraisal bias in refinance mortgages has been significantly more difficult due to the absence of a sales transaction value, with most research relying on noisy valuations from AVMs or mark-to-marketing prior transactions to provide a baseline of value. To circumvent this problem, this paper utilizes a novel dataset of 9 million recent refinance appraisals from 2013-2017 that contain the initial estimate of value provided by the homeowner/loan officer in the loan application, a value that is not permitted to be shared with the appraiser. We find a limited amount of anchoring bias on this initial estimate of home value, with around 7% of appraisals "hitting the mark" exactly. We analyze some of the potential drivers behind the bias and find that it is concentrated among certain appraisers and occurs more at higher LTVs. Loans with collateral valued by appraisers that routinely hit the mark are observed to have slightly higher default risk and slightly worse risk rank-ordering properties, controlling for all other observables, pointing to appraiser-specific effects on collateral valuation quality. However, even for "biased" loans, appraisal-based LTVs still perform largely similarly to LTVs derived from automated valuation models (AVMs) in risk rank-ordering ability, meaning that the bias from the appraisal and the noise in the AVM largely cancel each other out, in contrast to findings on purchase mortgages.

Keywords: Appraisal Bias, AVM, Refinancing, Mortgage Delinquency

JEL Codes: G21, G28, G51, L85, R31

Disclaimer: The views expressed in this paper are our own and are based on independent research and do not necessarily reflect the views of the Office of the Comptroller of the Currency, the Federal Housing Finance Agency, the U.S. Department of the Treasury, Fannie Mae, or any federal agency, nor has the paper been formally reviewed by any individuals within the

*Retail Credit Risk Analysis Division, Office of the Comptroller of the Currency. george.krivorotov@occ.treas.gov.

†Fannie Mae. michael_lacourlittle@fanniemae.com

‡We would like to acknowledge Eric Rosenblatt, Lan Shi, Michael Eriksen, and Irina Barakova for their invaluable suggestions and comments, and Luke Wong and Robert Barclay for their research assistance.

Office of the Comptroller of the Currency, the U.S. Department of the Treasury, or any federal agency.

1 Introduction

The question of valuation quality is one of the central issues in secured lending - systematic overprediction of asset value is a major risk to the banking industry. Nowhere is this truer than in the retail real estate lending sector. These loans are backed by homes which are largely valued by a licensed appraiser utilizing a standardized appraisal format. Yet these valuations have consistently come under scrutiny for potential biases - most significantly, overvaluation bias which would underestimate loan credit risk and loss amounts. In particular, overvaluation bias could lead to unexpectedly lenient credit underwriting - a loan with overvalued collateral might pass LTV thresholds or be erroneously classified as a low risk loan in a credit risk model, leading it to be acquired by the bank when it otherwise would not have been, or priced lower than it otherwise would have been. The very same biased collateral valuation would also cause the bank to set its loss reserves excessively low due to overpredicted home price and higher recoveries during foreclosure. Given that there is $15 trillion outstanding of mortgage debt as of 2019, any systematic errors of this nature could have a significant consequences on the US economy.

Indeed, in the reckoning after the financial crisis, claims of rigged appraisal processes riddled the mortgage banking sector. Lenders seeking to dig deeper into the credit spectrum allegedly incentivized appraisal providers to give positively biased appraisals to ensure that the loan would go through even for customers with marginal credit quality and dubious collateral. For example, in a well-known case, in November 2007, Washington Mutual was alleged to have pressured its appraisal provider, eAppraiseIT, to use an approved appraiser list or else have business transferred to its competitor. Appraisers claimed that providing favorable appraisals on collateral in WaMu-associated loans would ensure their place on the list, while making low appraisals, even if accurate, would ensure that they would not be selected by the lender any longer. Research has indeed confirmed the occurence of this bias and its contributing impact in the financial crisis- see Ben-David (2011) and Griffin and Maturana (2016).

In the aftermath of this controversy in 2009, the Federal Housing Finance Agency, Freddie Mac, and the New York State Attorney General drafted the Home Valuation Code of Conduct which

addressed the perceived shortcomings of the system. While this expired in 2010, it influenced several appraisal-related provisions in the Dodd-Frank act (see Title XIV, Subtitle F). Since then, the GSEs have also adopted their own Appraiser Independence Requirements. For example, Fannie Mae (2010) mandates:

> *No employee, director, officer, or agent of the Seller, or any other third party acting as joint venture partner, independent contractor, appraisal company, appraisal management company, or partner on behalf of the Seller, shall influence or attempt to influence the development, reporting, result, or review of an appraisal through coercion, extortion, collusion, compensation, inducement, intimidation, bribery, or in any other manner including but not limited to:*
>
> - *Requesting that an appraiser provide an estimated, predetermined, or desired valuation in an appraisal report prior to the completion of the appraisal report, or requesting that an appraiser provide estimated values or comparable sales at any time prior to the appraiser's completion of an appraisal report;*
> - *Withholding or threatening to withhold future business for an appraiser, or demoting or terminating or threatening to demote or terminate an appraiser;*

In addition, Appraisal Independence Requirements require that appraisal functions and loan production functions are entirely independent. This could mean that either appraisals are entirely outsourced to an external Appraisal Management Company, or AMC, or the bank functions involved in selecting appraisers and vetting appraisals have no involvement whatsoever with the functions that underwrite loans.

There has been extensive research that validates the effectiveness of the HVCC and subsequent Appraisal Independence Requirements and other regulations. Broadly, this line of literature has shown that there has been a slight increase in the propensity of the appraiser to value low for HVCC-covered transactions for both refinance and purchase appraisals. Shi and Zhang (2015), using a difference-in-difference identification strategy on 2009 data, find that appraisal bias in refinance mortgages, as measured by the difference between the refinance appraisal value and a subsequent

purchase transaction, has gone down for GSE loans relative to non-GSE loans since the enactment of the rules. Using 2008-2010 data, Ding and Nakamura (2016) similarly find that appraisal bias has gone slightly down since the HVCC, showing that HVCC-covered transactions were at least 2.1 percentage points more likely to have a low value than non-HVCC covered transactions after the enactment of the rule. Finally, Agarwal et al. (2019) find similar positive effects on valuation accuracy post-HVCC on 2008-2010 mortgages.

However, research shows that despite the improvements that the HVCC and related guidelines have brought to the collateral valuations process, appraisal bias in purchases is still widespread. Eriksen et al. (2019) demonstrate that over 90% of recent appraisals from 2013-2017 purchase mortgages anchor at or above the contract price, with appraisers weighting comparables in a skewed manner to hit the mark. Calem et al. (2017), using GSE data from 2013-2015, find that the information loss from using skewed appraisals in purchase mortgages is more severe at LTV notch points - edge points of buckets on which pricing and underwriting policies are typically set. Indeed, they find that LTV values derived from an automated valuations model (AVM) are more informative for credit risk purposes at these points. However, what is less known is the level of bias and information value in *refinance* mortgages in the recent period.

In this paper, we use a unique dataset of GSE refinance loan applications and associated appraisals linked with performance for closed loans to analyze the potential bias in refinance appraisals in the period *after* the HVCC. Instead of focusing on measuring bias using a noisy AVM or mark-to-market (MTM) value which might itself be biased, we instead examine anchoring on the initial estimate of home value provided on the loan application by the borrower/loan officer. We find that around 7.5% of appraisals are exactly equal to this value, which is striking since it is not permitted to be communicated to the appraiser by the aforementioned Appraiser Independence Guidelines, and cannot be explained by coincidence, prior transaction anchoring, or rounding. We decompose some of the reasons behind why appraisers will hit the mark, and find that it is more likely for higher LTV transactions and certain LTV notch points, and initial estimates which have a higher positive difference from an unbiased AVM value. We find that in general, hitting the mark is not associated with an increased likelihood of the borrower getting approved or closing on their loan, contrary to

popular hypotheses on why appraisal bias may occur. Hitting the mark is also not associated with a higher appraiser work volume, confirming results in Tzioumis (2017) on the purchase side. Finally, hitting the mark does not seem to be associated with a markedly higher increased propensity to default in the loans analyzed, meaning that the bias in these valuations is not severe enough to significantly deflate appraised LTV to affect this measure significantly, in contrast to findings in Agarwal et al. (2015). However, we do find small *appraiser*-level effects on these loans. Hitting the mark is highly concentrated among appraisers, and loans valued by these appraisers tend to default slightly more than others. Nevertheless, this bias is not enough to outweigh the noise associated with AVM-based valuations in risk rank-ordering loans, in stark contrast to findings in Calem et al. (2017).

In general, the conclusions of this paper are consistent with Shi and Zhang (2015) that find that refinance appraisal bias has decreased markedly since the HVCC. Indeed, the only other paper that explicitly focused on refinance appraisal bias and its effect on default, Agarwal et al. (2015), comes to very different conclusions on the extent and effect of appraisal bias and it is likely that one source of this discrepancy is the different time periods analyzed, with their data being exclusively from 1990-2011, which would consist mostly of mortgages from before the HVCC, and our data being from the period between 2012-2017. In addition, another key difference from other papers is the focus on the anchoring on the initial estimate of home value as a measure of bias, as opposed to a deviation from AVM or MTM values. Indeed, AVM and MTM values by themselves are noisy and sometimes skewed, which causes severe attenuation bias in all regression estimates and can lead to dramatically different conclusions depending on which cut of the data is taken, putting the results of those papers in question. Anchoring on the initial estimate on the other hand is a more definitive marker of bias, since it is not allowed to be communicated to the appraiser. Indeed, the phenomenon in the refinance space can be thought of as similar to anchoring on the contract price in the purchase mortgage space, as documented by Eriksen et al. (2019), although at a smaller scale. This allows us to precisely hone in on these biased appraisals and understand their effect on credit risk. However, these biases are relatively minor, and are not enough to meaningfully deteriorate their performance in providing LTVs for credit risk modeling purposes compared to an AVM.

1.1 AVMs versus Appraisals in the Refinance Mortgage Space

There is an active policy debate on the quality of appraisals versus valuations provided by an automated valuation model, or AVM. Appraisals by licensed appraisers are typically around $300-$400 and are included in the closing costs of the mortgage, paid for by the borrower. In addition, there is a time cost associated with an appraisal as well, further delaying the mortgage to closing. Borrowers and lenders are both disinclined to bear this monetary and time cost.

Indeed, lenders typically compete for customers based on not only rates, but also on closing costs and time to closing. For example, Rocket Mortgage marketing materials state that *"in the eight minutes it takes a space shuttle to reach orbit, Americans will now be able to receive a full mortgage approval online with Rocket Mortgage by Quicken Loans."*[1] In addition, mortgage lenders have been continually expanding their credit box since the financial crisis - e.g. see figure (1).

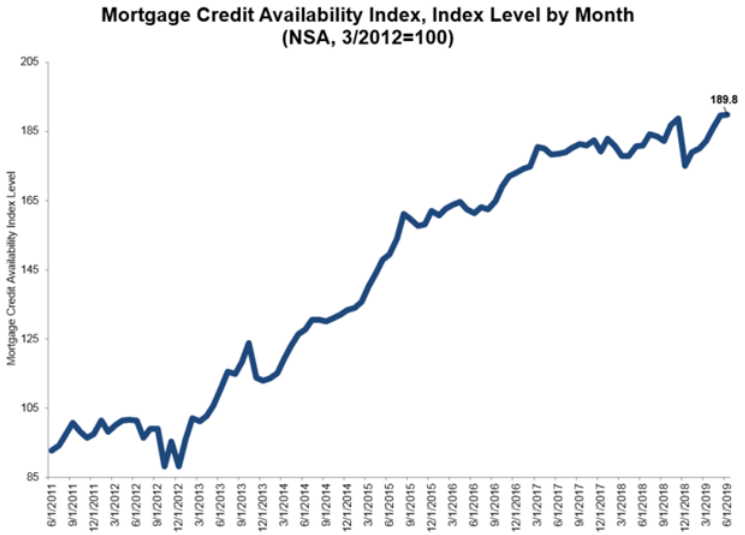

Figure 1: *Source:* Mortgage Bankers Association

One way that lenders have facilitated increasing credit availability and lower origination costs has been to turn to evaluations based on automated valuations model output instead of full appraisals. However, so far, largely due to regulatory concerns, banks have relied on AVM-based evaluations

[1] See Quicken Loans (2015).

mostly for underwriting HELOCs and home equity loans.[2] Indeed, current regulations from the Financial institutions Reform, Recovery, and Enforcement Act of 1989 (FIRREA) only exempts real estate transactions from appraisal if the transaction value is below $250K.[3] This would exclude a large portion of the first mortgage market. However, in response to widespread pressure in the mortgage banking industry, and significant advancements in AVM accuracy and hit rate, the regulatory agencies released a proposal to increase the loan limit for appraisal exemption from $250K to $400K - see FIL-76-2018. The Mortgage Bankers Association, representing over 2,200 member companies and having on its board senior home lending executives from Wells Fargo, Bank of America, UBS, Chase, and so on, followed suite by submitting a public comment strongly in favor of this proposed increase. Given these statements and the aforementioned continual expansion of mortgage credit, it can reasonably be expected that banks will increasingly rely on AVM-based evaluations over appraisals for underwriting first mortgages in this newly opened loan segment.

The consequences of this potential change from appraisal-based underwriting to AVM-based underwriting are still being understood. Calem et al. (2017) consider the information value of using an appraisal value as opposed to an AVM value for purchase mortgages in the face of widespread confirmation bias of the contract price, especially in critical LTV notch points between LTV buckets that determine most underwriting policy and pricing. They find that while the deviation of the appraisal to sales price does provide some information on the probability of default, especially when the appraisal is over the sales price, it consistently underperforms relative to the deviation of the AVM value to the sales price when measured by the in-sample R-squared of the model on the development set. Hence, they find that the bias from the appraisal overwhelms the potentially noisy AVM estimate in providing information on default probability for purchase mortgages in their sample.

This paper further adds to this literature by examining a similar question, but for refinance mortgages, for which bias has hitherto been more difficult to measure. In this case, we find that appraisal quality differs by appraiser, meaning that appraisers that routinely hit the mark will be associated with loans that are slightly more likely to default, conditional on all other observable

[2]See Housing Wire (2017).

[3]See FDIC (2018), 323.3 - "Appraisals required; transactions requiring a state certified or licensed appraiser."

factors. However, even for these biased loans, utilizing an AVM LTV as opposed to an appraisal-derived LTV provides minimal improvement in risk rank-ordering performance. For loans with appraisers that do *not* hit the mark, appraisal LTV is slightly higher in risk rank-ordering performance. Thus, in the case of refinance mortgages, noise from the AVM roughly balances the bias in the appraisals across the board, in sharp contrast to findings on the purchase mortgage side.

2 Data

We utilize a unique appraisal-level dataset taken from the Uniform Collateral Data Portal (UCDP) which includes all lender appraisal submissions that are associated with the universe of a GSE's single family loan casefiles from 2013-2017. Importantly, the dataset includes appraisals associated with loans that did not close, overcoming shortcomings with selection bias that are common in other appraisal-related research.

The appraisals are all full URAR appraisals completed by a licensed appraiser, following the Fannie Mae Form 1004/Freddie Mac Form 70 and standardized UAD format. This means that there is detailed information on the characteristics of the subject home that were recorded during a full interior inspection. This includes basic information such as gross living area, number of bedrooms, bathrooms, basement, and parking spots. It also includes information on the condition and quality of the home, placed into standardized codes according to the UAD format, which is highly informative for home valuation. Since the comparable approach to valuation is required for single family homes, the data also includes analogous information on comparable sales to the subject, as well as the adjustments in each field to make the property equivalent to the subject. Unlike the subject, the property characteristics for the comparable sales are usually sourced from multiple listing service (MLS) postings and exterior photos, since the appraiser is only required to make a full interior inspection of the subject.

The data also includes associated loan application information, such as borrower FICO, loan amount, loan LTV, loan purpose, term, and so on. The data contains *all* appraisal and loan application submissions, so one can see the evolution of the loan's purpose, amount and valuation

from the first submission to the last. Crucially, the valuation given in the first submission of the loan application is always an estimate that is agreed upon between the borrower and the loan officer, rather than an actual appraisal. Hence, we are able to see whether or not the actual appraisal is at all anchored by this initial estimate - theoretically it should not be communicated at all to the appraiser according to the GSE's Appraiser Independence Guidelines.

Finally, this data includes underwriting findings from the GSE's automated underwriting system, and a loan ID that enables closed loans to be linked to public GSE single family performance data. This allows us to see how loan performance is affected by factors in the appraisal, and examine which valuation methods provide the best rank ordering of risk. This subsequently provides insights on whether an AVM or an appraisal is has less bias or noise for the purposes of providing an LTV to underwrite loans. Additionally, the data has loss information in it as well, potentially allowing us to understand the scale to which errors in underwriting would affect losses, as well as the level to which losses may be mispredicted from using one or the other valuation methods.

3 The Refinance Appraisal

While there is variability across lenders in practices, which also depends on if the lender primarily sells their mortgages to the GSEs, generally the refinance mortgage and appraisal process tends to follow the below process.

1. Borrowers, in collaboration with loan officers or brokers, complete a loan application that includes loan purpose (cash-out or limited cash-out), loan amount, and estimated home value (which determines LTV on the first pass).

    (a) Lenders will sometimes have bounds on the initial estimated home value, but these will be relatively lax. For example, loans will be automatically rejected if the estimated home value is less than half or more than twice an AVM value.

2. A licensed appraiser is hired by the lender (but paid for by the borrower) to inspect the property. Appraisers can in principle communicate to the borrower, but are officially not

allowed to be influenced by any third party. Sometimes, the appraiser does not even encounter the borrower at the property, although this depends.

3. The borrower can appeal their appraisal to the lender if they suspect that the appraiser is committing errors such as not taking into account the property's home improvements or comparable pending sales in the neighborhood. If the appraiser does not subsequently modify the value, the borrower can order a new appraisal from a different appraiser.

4. Once the appraisal process is complete, the number given in the final appraisal is used as the final estimate of value for LTV. The borrower can change their loan purpose/amount to request more or less cash-out. This can ensure that the loan goes through with the right price even if the valuation is lower than expected. The borrower can also withdraw their application if the valuation was too low to have the refinance make financial sense.

In this paper, we will focus on the difference between the first estimate of value given by the *borrower/loan officer* and the last estimate of value given by the *appraiser* and largely abstract from the possible intermediate appraisals that may have happened in between. We also examine the evolution of loan submissions, particularly how loan amount may have changed in between the first loan application and the final submission to the lender.

There are multiple types of appraisals, which vary greatly in terms of detail and accuracy.[4] In general, the two major categories are a) the desktop appraisal, which does not involve a physical inspection of the property, but instead relies upon tax records, multiple listing services, and external photographs of the property, and b) the full appraisal, which involves a full walk-through of the home. Typically, for any loan above $250K, and first mortgages in general, lenders will require a full appraisal to be done, and we will focus only on these kinds of appraisals.

For a full appraisal, it is widespread practice to use a standardized appraisal form, known as the Uniform Residential Appraisal Report (URAR), in order to conform with GSE requirements, corresponding to Fannie Mae Form 1004/Freddie Mac Form 70,. For loans that are subsequently sold to Fannie Mae or Freddie Mac, the appraisal is submitted to the Uniform Collateral Data Portal

[4]This is in addition to other non-appraisal evaluation methods, such as a BPO.

(UCDP) in a standardized format. This includes standardized codes for possibly subjective fields such as Condition and Quality, greatly improving the data quality for use in subsequent AVMs.[5] The appraisal is then subject to a review for possible defects from the Fannie Mae Collateral Underwriter or the Freddie Mac Loan Collateral Advisor systems, including overvaluation. In addition, regular loan quality reviews by the GSEs periodically uncover collateral and other loan defects that may require the lender to repurchase the loan.

There are multiple ways that an appraiser can value a home, including the comparables approach to value, the cost approach to value, and income approach to value, commonly used for income-producing properties such as 2-4 units homes, multifamily homes, or commercial properties. In the case of a single family home or condominium, the only method required by the GSEs is the comparables approach and it is used in the vast majority of cases. An appraiser is required to select at least three recent comparable sales in the same neighborhood as the subject. The appraiser then adjusts the home values to take into account observable differences between the subject property and the comparables. For example, if a comparable home is 500 square feet larger than the subject property, the comparable's value is adjusted down by 500 multiplied by what the appraiser believes the approximate price per square foot in the neighborhood is. Similarly, the comparable home is adjusted for differences in bedrooms, bathrooms, geographic distance from the subject, time distance from the subject, and so on. The adjusted values of the comps, although they have been adjusted down to be equivalent to the subject, still vary, sometimes significantly, from each other. The final step in the process is known as reconciliation, where these values are weighted together to produce the final estimate of value. Eriksen et al. (2019) find that appraisers primarily use the reconciliation process to set the valuation of the subject home equal to the contract price through unequal weighting of adjusted comparables.

[5]Prior to the adoption of standardized codes for Condition and Quality in 2011 and 2014, there was significant ambiguity in existing appraisal reports on their meaning.

4 Level of "Hitting the Mark" in Refinance Appraisals

As mentioned previously, the phenomenon of hitting the mark is well known in the purchase mortgage space, where appraisers anchor estimates to be at or above the contract price.[6] However, in the refinance mortgage space, it is less clear what the appraiser could actually anchor their estimate on due to the absence of a transaction associated with the mortgage. Previous studies have primarily relied on using mark-to-market or AVM estimates of home value to measure appraisal bias in the refinance space. However, AVM and mark-to-market estimates themselves are highly noisy, and could very well be less accurate than appraiser estimates in the first place, making their use as a baseline of value problematic.

This study circumvents this issue by examining the first estimate of value provided by the borrower/loan officer. Limited cash-out refinances are typically an exclusively financial decision undertaken with the goal of lowering rates - borrowers are unconstrained by the desire for relocation that exists in purchase transactions for owner-occupied homes. Since these rates are largely determined by LTV, it would be in the borrowers interest to have a valuation of their home such that the loan falls into a favorable LTV bucket - and if the LTV of the refinance is such that rates are not enough of a savings to cover closing costs, the borrower could very well withdraw from the loan. Thus, it is plausible that the first estimate of value is a statement of borrower's desire.

The borrower would have an interest in communicating this value to the appraiser to inform them on their goals for the refinance, and potentially, the minimum value that they would accept in order to close on the loan. If appraiser independence was compromised and was influenced by loan production, appraisers would be incentivized to go at or above this value to ensure that the loan goes through. In addition, there may be a supply-side effect, where an appraiser would not go under this number because of concerns on the approval of the mortgage from the lender/GSE's risk score.

In the sample of around 2 million unique loan casefiles from limited cash-out refinances (LCORs) from between 2013 and 2017, we show the deviation of the final appraisal to the first estimate of

[6] See Eriksen et al. (2019)

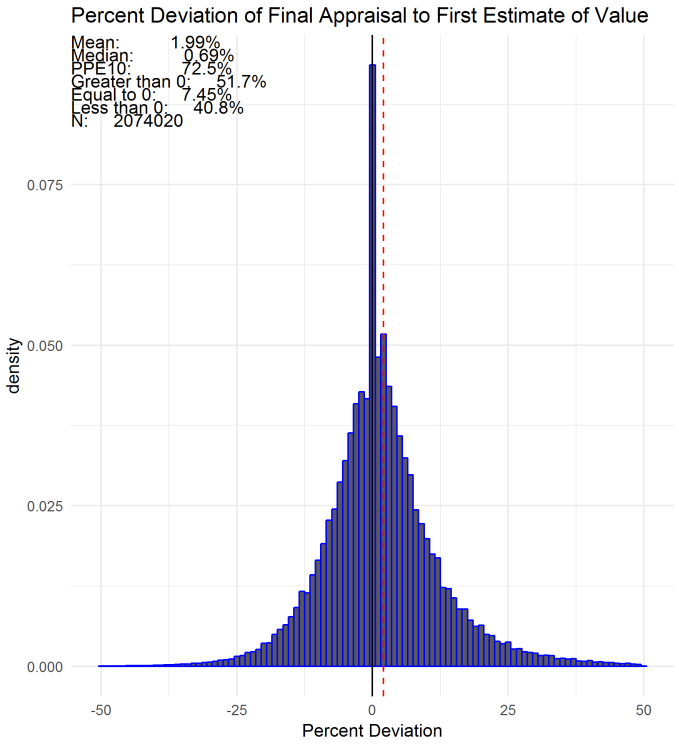

Figure 2: Restricted to Limited Cash-Out Refinance.

value in figure (2). Around 7.5% of these appraisals exactly hit the estimate that is put on the loan application by the borrower/loan officer, with around 60% going at or above the initial estimate of value. In robustness analysis available in appendix A, we conduct a simulation to show that this cannot be explained by coincidence or rounding. In figure (14), we also find that the rate of hitting the mark stays approximately the same when omitting all refinances that had recent sales associated with them, ruling out the possibility that both the initial estimate and the appraisal could be anchored on a third value, a prior recent transaction.

While this is significantly less than the anchoring that is shown in the purchase mortgage case, where there is around half hitting the mark and over 90% at or above the contract price, it is still a large amount. As mentioned previously, unlike the contract price, the initial estimate of value is not on the standard URAR appraisal form and theoretically should not be communicated to the appraiser at all, and so the fact that many appraisals are anchored on it demonstrates a

significant breach in policy. In the following sections, we will examine the reasons of why this may be happening, and the potential consequences for underwriting purposes.

5 Motivations Behind "Hitting the Mark"

5.1 Loan Cost

For loans that are sold to GSEs, as of 2018 comprising around half of originations, risk-based pricing depends on a combination of LTV and FICO known as the LLPA, or loan-level price adjustment. This is added on top of a non-risk-based rate which is determined by market conditions and lock-in time. LLPA is determined in a dual matrix where rate is determined by which bucket of LTV and FICO one is in, with the crucial LTV notch points being at 60, 70, 75, 80, 85, 90, 95, and 97. Jumps in risk-based pricing along the LTV dimension are typically steeper in lower FICO buckets, to take into account the increased marginal risk a high LTV loan has for higher credit risk individuals than for lower credit risk individuals - see figure (3) for an illustration of LLPA rates for limited cash-out refinances. Rates also change based on a) if CLTV is not equal to LTV due to second liens, b) if the loan is cash-out, c) the level of MI coverage if above 80 LTV, and d) there are other special features of the mortgage, such as being an adjustable-rate mortgage (ARM), or being intended to purchase a manufactured home, second home, or investment property.

Note that LLPA rates actually decline after 80 LTV. This is due to the requirement for private mortgage insurance (PMI) paid for by borrowers that partially reimburses the GSE for losses in loans above 80 LTV. Borrowers tend to avoid PMI, and many prominent websites offering advice on mortgage finance discourage getting loans over 80 LTV because of this requirement.[7] This can be seen in the large discontinuity in the density of limited cash-out refinance (LCOR) mortgage applications between $\leq$80 LTV and >80 LTV in figure (4), which shows the distribution of LTVs upon first application for refinance mortgages. we leave out cash-out refinances since they are not underwritten by the GSEs if above 80 LTV.

In figure (4), also note the discontinuous nature of the distribution of LTV in the first applica-

[7] See Investopedia (2019) for such an example from a prominent investment advice website.

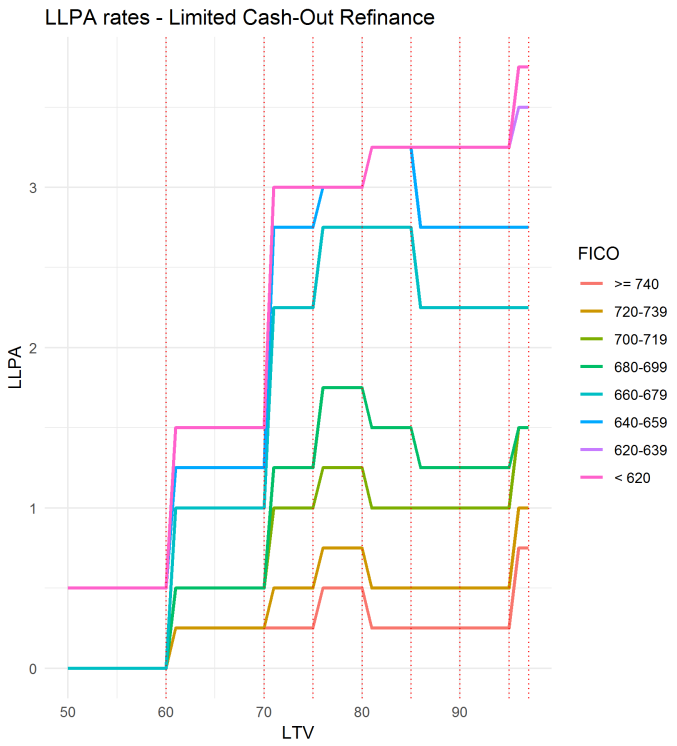

Figure 3: Red dotted lines indicate LTV buckets.
Source: Fannie Mae (2019).

tion - borrowers tend to bunch into LTV notch points. This graph displays only limited cash-out refinances, or LCORs, where the loan value must be within 2% of the old loan. This means the major degree of freedom for borrowers is on the home valuation side. In order for LTVs to cluster on these points, it is likely there is at least some influence of the LTV on the estimation of the value.

Empirically, we can examine the level to which borrowers save by being in any particular LTV bucket by simulating loan cost for when the valuation of the home places them in a higher LTV bucket than where their first estimate of value placed them. Figure (5) shows a visualization of the average loan savings (or costs) by going into a higher LTV bucket.

For most mortgages in this dataset, there is a significant cost for jumping to a higher bucket. However, surprisingly, we find that for many loans in the 76-80 LTV bucket, there would be a

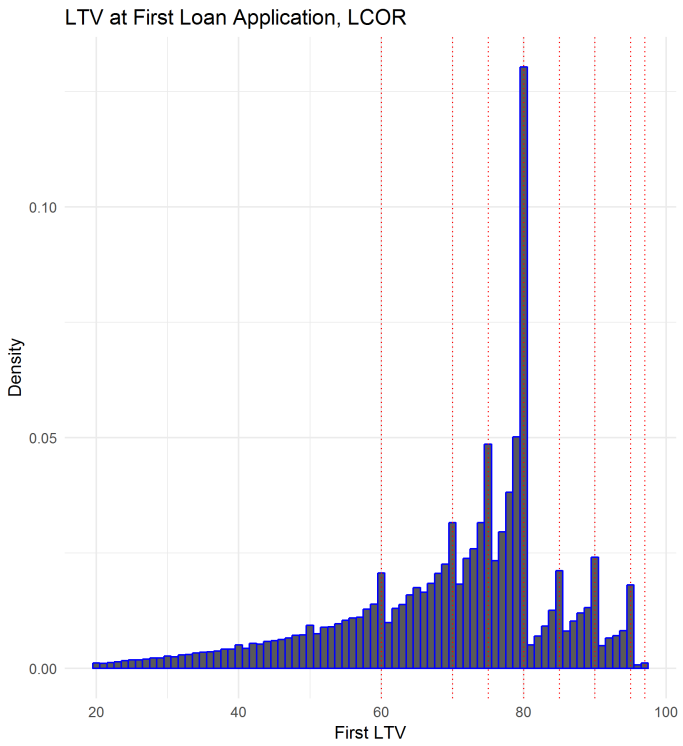

Figure 4: Red dotted lines indicate LTV buckets.

decrease in loan costs in this situation. This is because the cost of PMI is relatively low compared to the LLPA decrease that would occur above 80 LTV. PMI rates have been in the vicinity of .5%-1.5% of the original loan amount per year. Moreover, one only needs to pay these rates when below 20% equity, and so most borrowers with 81-85 LTV loans would only need to pay PMI for less than a year.

The small financial consequences (and often benefits) of getting a valuation that places one in an LTV bucket above 80 poses a puzzle to the sharp discontinuity in the density of applicants below and above 80 LTV. One possible explanation is that while the cost of getting an appraisal that places one above 80 LTV is low, borrowers believe it is high, possibly due to advice given by loan officers or other external parties. Another possibility is a supply-side effect - due to bank internal policies and risk scores, or GSE internal policies and risk scores, a loan above 80 LTV is much less likely to be approved than one with at least a 20% equity position. we will later examine this

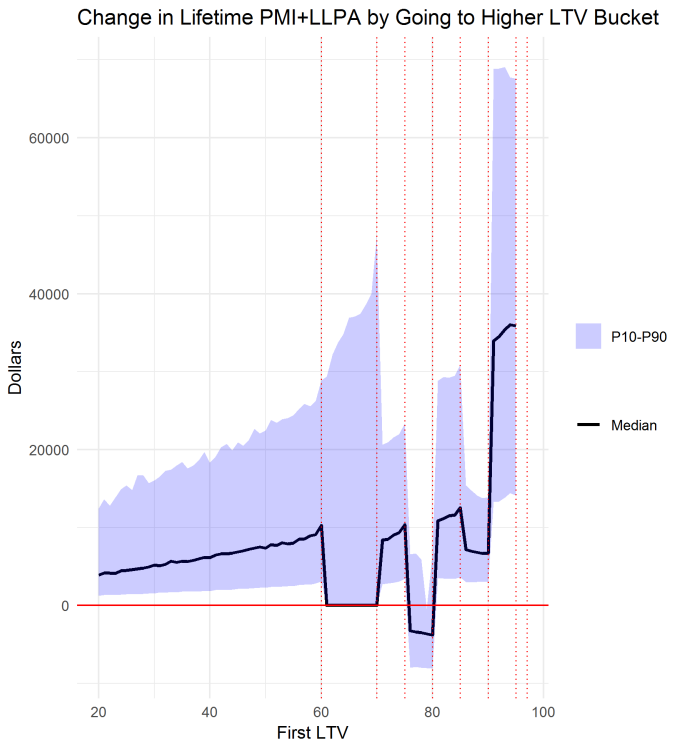

Figure 5: Red dotted lines indicate LTV buckets. Data is restricted to customers that are applying for an LCOR at first application. Valuation is using the borrower/loan officer's first estimate of value on the loan application, prior to appraisal. Line indicates median per LTV, while shaded area indicates the area between the 10th and 90th percentiles, per LTV. Note that this assumes that loans do not default or prepay - however the shape of the graph is robust to taking 2,4, and 7 year (average life of mortgage loan) loan cost as well.

question as well, and find a small effect of LTV on loan approval above 80 LTV.

In any case, for most loans, there is still a significant monetary cost to having a loan fall in a higher LTV bucket due to a lower-than-expected home valuation, and it is common knowledge that borrowers actively avoid low appraisals. One can observe this through examining the reaction of borrowers to low evaluations. If borrowers did not change their loan amounts in response to the change in the appraisal, figure (6) would show the distribution of LTV.

Note that this no longer has the bunching at LTV notch points that is visible in the first LTV histogram. However, once one takes into account the borrowers' adjustment of their loan amounts and look at the distribution of LTV as measured by the final loan amount over the final valuation

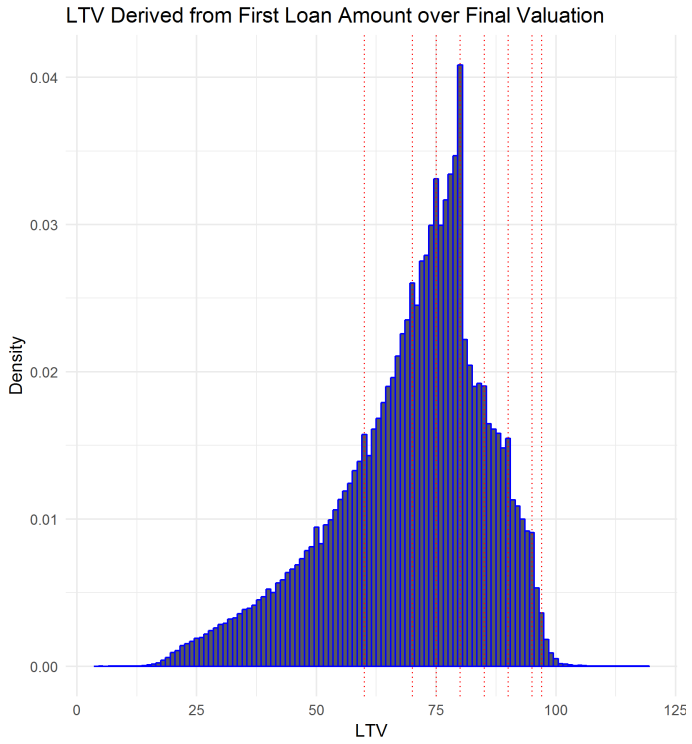

Figure 6: Red dotted lines indicate LTV buckets. Data is restricted to customers that are applying for an LCOR at first application. Loan amount is from the first loan application, while valuation is from the final appraisal.

in figure (7), this once again has the concentration at notch points that is visible in figure (6). This suggest borrowers act strategically when it comes to the LTV of their refinances, although in many cases, the direct financial benefits do not seem clear.

5.2 Loan Approval

A major possible risk from an appraisal coming in low could be that the loan does not go through at all. A home's valuation may lead to an LTV that, combined with other risk factors, would lead the underwriter, whether it is the bank or GSE, to reject the application. Hence, a borrower applying for a refinance would actively seek a valuation that would put them in a favorable LTV range.

However, risk models across the banks and GSEs differ greatly in their treatment of LTV -

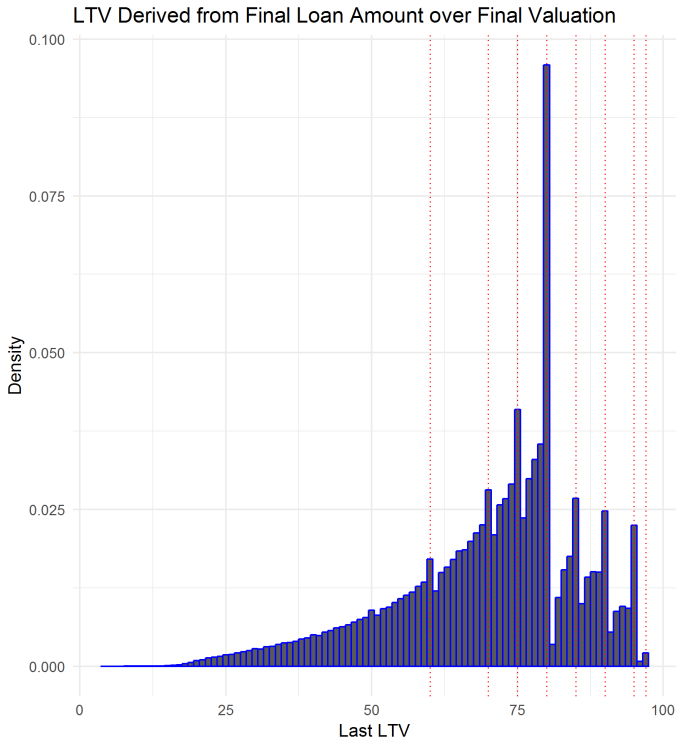

Figure 7: Red dotted lines indicate LTV buckets. Data is restricted to customers that are applying for an LCOR at first application. Loan amount is from the first loan application, while valuation is from the final appraisal.

unlike the LLPA, it is oftentimes much more complicated than a simple dual-matrix of binned LTV and FICO scores. In fact, risk models might not bin LTV at all, meaning that a decrease from 76 to 75 LTV is just as meaningful as a decrease from 75 to 74. This implies that borrower bunching behavior at notch points could be hard to explain in this context.

To tackle this question, we run a regression predicting loan approval given observable loan and collateral factors, with results outlined in the first column of table (1). First, we observe in figure (8) that approval rate at final loan application underwriting is weakly negatively dependent on final LTV, except for a marked decrease above 95 LTV, and smaller decreases in approval rate above the notch points. [8] This controls for all other observable factors, such as FICO, loan term, and

[8]To give an understanding of the scale of the results- Since approval rates hover around 90%, or log-odds of around 2, a decrease in log odds of -.5 would decrease the probability of approval to around 85%, not insignificant, but unlikely to explain all of the bunching in the data.

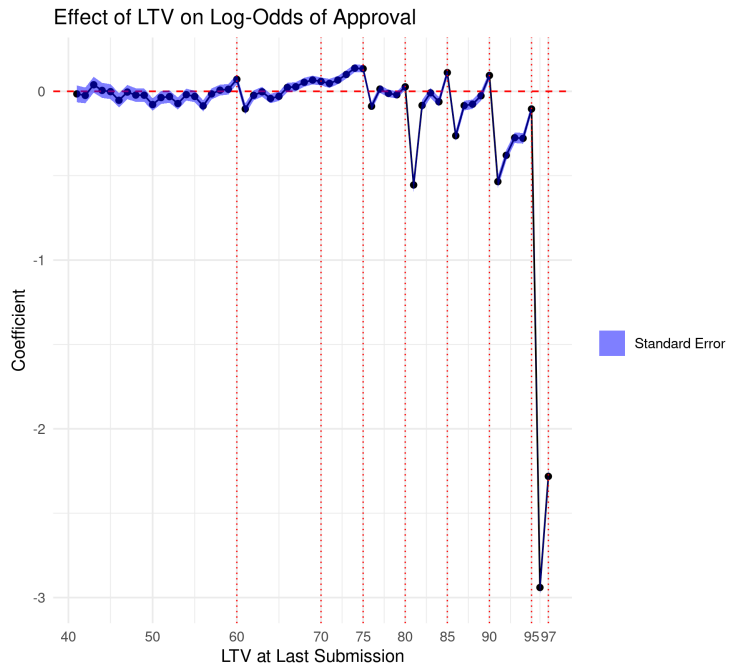

Figure 8: Results from approval regression

some features of the appraisal, such as whether or not it hits the mark, and how far off from an AVM value the first estimate lies. Approval rates slightly decline above notch points, but it does not appear to be the whole story explaining bunching behavior.

Also note from the regression that there is a slight tendency for more approvals when the appraiser goes above the mark than below - but surprisingly, those loans with appraisals that hit the mark have the same approval rate as the ones where the appraiser goes far below. This is counter to the intuition that hitting the mark would help a loan get approval in a marginal case. However, this may be due to a selection effect - it is possible that appraisers hit the mark in marginal cases which are more likely to be declined for reasons that are not observed in this data.

5.3 Borrowers Closing

While approval rates for high LTV loans do not explain borrower bunching behavior, it could be that borrowers themselves are pulling out of loans that they feel are no longer good deals due to a

	Approval	**Closing Given Approval**
Intercept	**2.96**	**-2.12e-01**
	*(2.57e-02)***	*(1.29e-02)***
Final Loan Term	**-1.36e-03**	**-3.81e-04**
	*(4.13e-05)***	*(1.73e-05)***
Pc. Dev. of First Estimate to AVM	**-8.64e-04**	**-3.14e-03**
	*(3.01e-04)***	*(1.43e-04)***
Pc. Dev. of Final Appraisal to First Estimate	**1.95e-02**	**5.47e-03**
	*(3.82e-04)***	*(1.82e-04)***
Pc. Dev. of First estimate to AVM * Pc. Dev	**2.43e-04**	**1.15e-04**
. of Final Appraisal to First Estimate	*(6.36e-06)***	*(4.39e-06)***
Pc. Dev. of Final Loan Amount	**-1.17e-02**	**-2.98e-03**
to First Loan Amount	*(6.18e-04)***	*(3.05e-04)***
Dummy for No Change in Loan Amount Between	**-2.23e-01**	-3.07e-03
First and Last Submission.	*(6.54e-03)***	*(3.04e-03)*
Dummy for Hitting the Mark	**-5.97e-01**	**-2.64e-01**
	*(9.73e-03)***	*(5.59-03)***
ARM Dummy	**-1.44**	**-6.41**
	*(9.86e-03)***	*(9.04e-01)***
640 $\leq$ FICO <660	5.85e-02	3.23e-02
	*(2.32e-02)**	*(1.32e-02) **
660$\leq$ FICO <680	**1.43e-01**	-2.03e-02
	*(2.17e-02)***	*(1.23e-02)*
680 $\leq$ FICO <700	**2.66e-01**	**-5.04e-02**
	*(2.05e-02)***	*(1.16e-02)***
700 $\leq$ FICO <720	**3.72e-01**	**-4.15e-02**
	*(2.02e-02)***	*(1.14e-02)***
720 $\leq$ FICO <740	**4.76e-01**	**3.99e-02**
	*(2.02e-02)***	*(1.13e-02)*
740 $\leq$ FICO	**8.58e-01**	**2.04e-01**
	(1.88e-02)***	*(1.06e-02)***
Dummies for LTV	*See figure (8)*	
Somer's D	0.334	0.143

Table 1: Regressions are logistic functional form. Limited to LCORs. *** stands for $< .001$ P-value, ** stands for $< .01$ P-value, * stands for $< .05$ P-value. Bolded values have a P-value of less than .01.

low appraisal. This is observed on the purchase side, as shown in Eriksen et al. (2019), with most valuations going under the contract price either having a renegotiation of the purchase price or the deal being cancelled entirely. For example, the March 2017 Realtor's Confidence Index survey indicated that the second most common issue among contracts with a delayed settlement were related to appraisal issues.[9]

However, what is less known is the effect of appraisals going under the borrower's preferred value on their propensity to close on the refinance. Like mentioned previously, it is plausible that borrowers' closing probability would be highly elastic to valuation level, since this affects loan cost, and limited cash-out refinances, not being associated with relocation, are typically a purely financial decision. To do this, we run a model to predict the propensity of a borrower to close on a loan given they have an approve/eligible score from the automated underwriting system. The results are encapsulated in table (1).

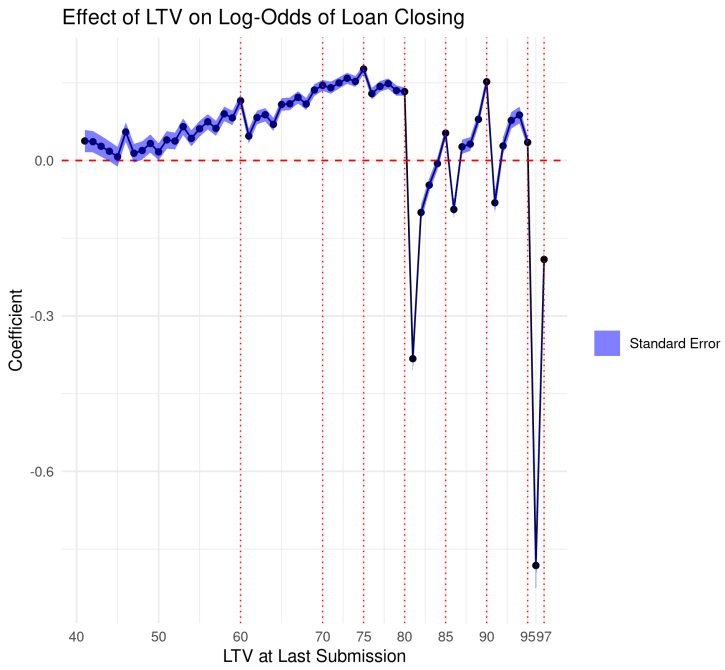

Figure 9: Results from closure regression

[9] See Yun (2017).

In figure (9), we can see a mildly positive effect on closing from going up in the LTV scale, possibly due to a selection on unobservables. However, what is noticeable is the drop in closing at the notch points above 80 LTV, similar to what is seen on the approval side. This suggests that borrower proclivity to avoid >80 LTV loans at the notch points could be associated with a combination of a decline in approval and closing rates.

Again, the effect of hitting the mark is negative, like in the case of accept/decline. This is counterfactual to the intuition that hitting the mark would encourage borrowers to close on a loan, and is contrary to findings in Agarwal et al. (2015). However, this could again be due to a selection effect. In particular, appraisers could be hitting the mark on appraisals that were marginal in an unobservable way in the first place.

6 Factors Predicting "Hitting the Mark"

So far, there has been suggestive evidence that conditional on loan attributes, deviation of the appraisal from the initial estimate, and deviation of the final loan amount to the first loan amount, a loan above certain LTV notch points is less likely to be approved, and given approval, is less likely to close. Irrespective of the reasons why this occurs, this indicates that borrowers and lenders intent on getting the loan through would prefer LTV not go above critical LTV notch points.

However, as seen in figure (2), only 7.5% of loans have appraisals that hit the mark, which is actually less than the number of loans initially at notch points due to the borrower's valuation. To understand the risk of the valuation bias in these appraisals, it is instructive to examine the factors that are most instrumental in predicting its occurrence.

The variable that we first turn to is LTV at first application, which can proxy in for a variety of factors. First, as mentioned in section (4), a higher LTV loan will be more expensive because of LLPA and mortgage insurance requirements, and these will have starker consequences at certain LTV notch points. Second, higher LTV is associated with higher chance of both loan approval and borrower closure as seen in the previous section. Third, a higher LTV may be associated with other unobserved attributes of the borrower - for example, a borrower that decides to refinance at

a higher LTV may have less marginal interest rate savings from reducing LLPA, which implies that the smaller savings were still significant for them, either indicating that they were cash-strapped or that they had a small opportunity cost of time in applying for a refinance.

In another specification, we also include the potential loan price increase from going into the higher LTV bucket as a predictor, to attempt to disentangle these effects from other components which may affect the propensity to hit the mark that would be captured by changes in LTV. Noting the potentially different perception customers may have about paying private mortgage insurance versus its actual cost as shown in figure (5), we then separate these into two components, interest cost, and MI cost, to examine the difference in potential change on propensity to hit the mark.

Another key variable under examination is the deviation of the first valuation input into the loan application from an AVM value. This attempts to answer how close the first estimate of loan value to an unbiased third-party valuation. In figure (16), we conduct an analysis of the bias and noise features of this AVM as well as a mark-to-market value on the set of purchase mortgages to evaluate their efficacy as a measure of value and find they are approximately unbiased. If the propensity to hit the mark increases for loans with higher overvaluation in the initial estimate, this could have especially sharp consequences for the overvaluation risk associated with the anchoring bias evident in refinance appraisals.

We estimate models for both the propensity to hit the mark and the propensity to value above the mark, that is, strictly greater than the first estimate of value. In this, we attempt to understand if bias is limited to anchoring on the first value, or whether appraisers will value the home even higher in an attempt to close the loan. Finally, we control for a variety of property-level and individual characteristics, such as loan term, gross living area (GLA), condition, quality, and state-level fixed effects, which do not significantly alter coefficient estimates.

Regressions are summarize in table (3) for "hitting the mark," and table (4) for going "above the mark." First, there is a highly significant and positive effect of potential loan price changes on the propensity to hit the mark. Regression (2) demonstrates that there is an additional effect on top of that from having a first LTV in the 76-80 bucket, above which the mortgage would fall under 20% equity and would require private mortgage insurance and might be less likely to go through.

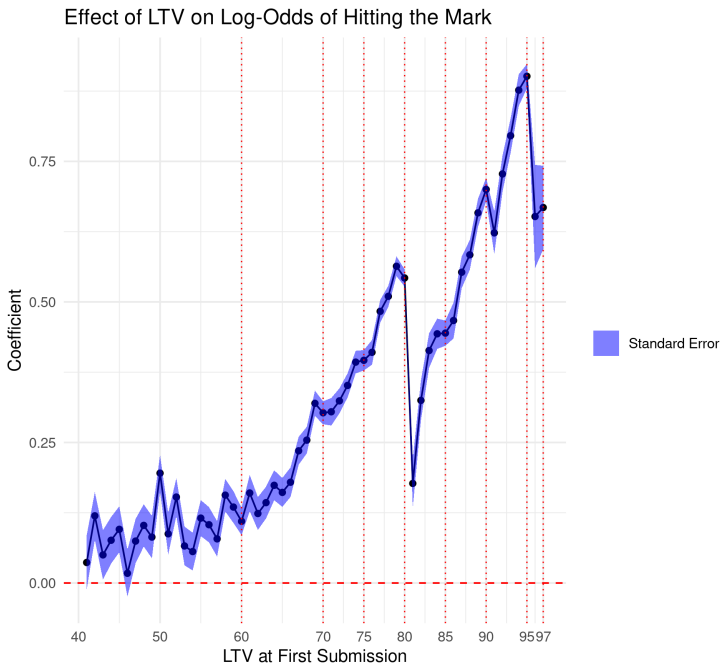

Figure 10: LTV fixed effects for "hitting the mark" from regression (5).

In fact, in regression (4), which separates the effect of the change in PMI from the change in LLPA in the change in loan cost, while both coefficients are significant, the dollar effect of potential PMI increases is several orders of magnitude larger than for potential changes in LLPA. One possible explanation behind this difference is that PMI is paid earlier in the loan life, while LLPA payments are spread through the entire life of the loan. Consumers may place less importance on LLPA payments in the future because of either high personal discount rates or an expectation that they will default or prepay earlier than the loan term, which is not an unrealistic assumption given it has been found that the average life of a 30 year loan is around 7 years. However it is unlikely this is the only explanation since the PMI effect is orders of magnitude larger than the LLPA effect in regression (4). Instead, it appears that consumers have a strong aversion to PMI above the pure dollar effect.

Instead of examining the effect of the potential dollar change in LLPA or PMI from changing an LTV bucket, we can also adopt a more flexible specification and examine LTV fixed effects. These

would incorporate most information on potential loan price and PMI changes when controlling for other factors such as loan term and FICO, but would also include other potential LTV-related effects, such as the effects of loan approval, or factors related to the borrower's liquidity. This can be seen in figure (10). Note that hitting the mark increases significantly and monotonically as LTV at first submission increases, except for at discontinuities just above critical LTV notch points, most notably 81 LTV and 91 LTV. This may have to do with a selection on unobservables, as borrowers who initially choose a valuation and loan amount with an 81 or 91 LTV can easily modify their loan amount or plausible initial valuation estimate to be at 80 LTV, and are seemingly unconcerned with potential consequences on the approval or pricing of the loan.

Another factor that is highly predictive of hitting the mark is the deviation of the first estimate to an AVM value. Note that the largest single jump in Somer's D occurs when it is introduced in regression (3). Hitting the mark occurs more frequently when the first value is above an unbiased AVM result, which has significant implications for the level of bias in the final valuation in these loans. In table (6), we show that this result persists when using a mark-to-market value on recent sales. In table (4), where we run a model that predicts appraisers going "above the mark," the relationship is the opposite. In those cases, the first valuation being below the AVM is highly predictive of the appraiser going above the first value, indicating that hitting the mark might have significant risks associated with overvaluation.

Finally, hitting the mark is especially common in lower FICO ranges, even when controlling for first LTV, and ARMs generally have more appraisals that hit the mark, although this effect ceases to be significant when controlling for loan term. In general, hitting the mark is associated with typical drivers of credit risk. In a later section, we will examine more closely how it will affect the probability of default, as well as how using these potentially biased valuations will affect estimates of credit risk, especially compared with a noisier, but potentially less biased, AVM value.

In regressions predicting going above the mark, we find that the relationships found in hitting the mark no longer hold. In particular, going above the mark seems to be largely explained by borrowers that make a first estimate of value below what the AVM would suggest, which is consistent with unbiased appraiser valuation, although the relationship with LTV is similar as with hitting the

mark. In addition, we will also show in a later section that going above the mark is not highly concentrated among specific appraisers, suggesting it has less to do with intentional bias. Hence, we focus on hitting the mark as the most concentrated source of bias for refinance appraisals.

7 Appraiser-Specific Effects

Now, we examine whether idiosyncratic variability between appraisers can partly explain the phenomenon of hitting the mark. In particular, is hitting the mark concentrated among some appraisers that violate appraisal independence principles, or is it evenly spread between appraisers and depends more on the circumstance that the appraiser is faced with.

To that end, we examine measures of concentration for appraiser volume, percentage hitting the mark, and percentage going above the mark, in particular, looking at Gini coefficient and Lorenz curves as particularly instructive measures.

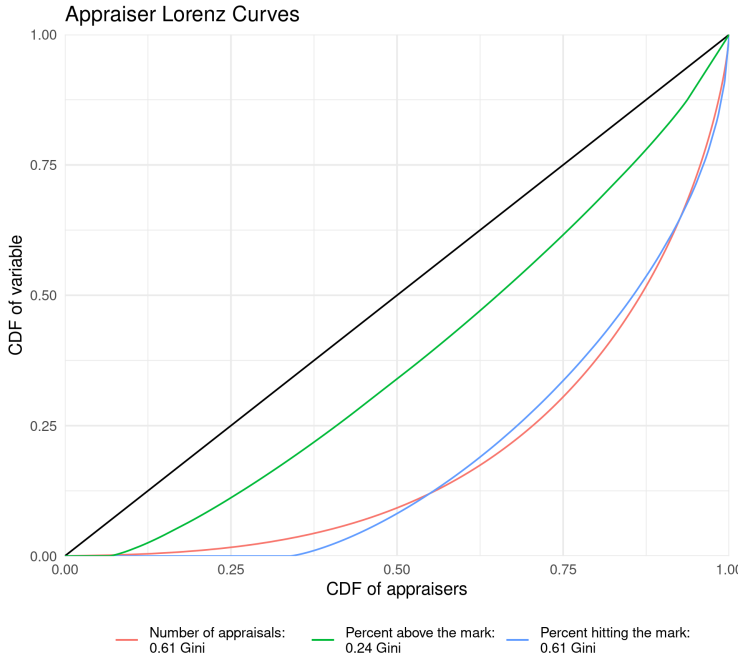

Figure 11: Concentration of Hitting the Mark among Appraisers

As can be seen in figure (11), appraisal volume is highly concentrated among certain appraisers. In addition, among appraisers, the percent hitting the mark is also highly concentrated. Notice from the Lorenz curve that almost 40% of appraisers do not hit the mark at all, although this graph includes appraisers who had only completed one appraisal on file.

In robustness analysis in figure (15), we find that this figure drops to around 15% when restricting to appraisers that have done at least 10 appraisals.[10] However, the relative scale of concentration of anchoring bias remains high, with a Gini of 0.39 in hitting the mark in this restricted sample, as compared to the Gini of overall volume which is at 0.46.

In contrast, going above the mark is relatively evenly distributed among appraisers, both on aggregate and when restricting to appraisers that have done at least 10 appraisals. This provides further suggestive evidence that going above the mark is largely correcting for underprediction of home value by the borrower at the first valuation, rather than a definitive sign of appraisal bias.

In addition, we examine whether appraiser work volume is correlated with frequency of hitting the mark. This is commonly cited as one possible motivation on why appraisers would provide biased valuations. In particular, the hypothesis is that banks or AMCs would provide additional volume to appraisers who ensure that deals go through. As can be seen in table (3), appraiser work volume is not significant when trying to predict the probability of hitting the mark, similar to findings in Tzioumis (2017) on the purchase mortgage side. In addition, the correlation coefficient of appraiser work volume and the percentage of appraisals that hit the mark is largely close to 0 and insignificant. However, due to the large amount of unobservables related to the incentives of the appraisers, the effect of hitting the mark on appraisal volume is still largely an open question, since selection effects could be driving the insignificant correlation. For example, it could be that even if hitting the mark causes an increase in appraiser volume, appraisers only do this when business is poor, meaning that the relationship between volume and hitting the mark is 0 or even negative.

[10] This is around 65% of the 50,610 unique appraisers on file in the dataset.

8 AVM versus Appraisal LTV in Predicting Loan Default

With results on *when* the appraisal hits the mark and the possible motivations behind this behavior, we can now turn to the consequences of this source of bias on credit risk. We will answer the question of whether the bias in the valuation that arises from hitting the mark in appraisals will outweigh the error that results from noisy valuations that do not suffer from this anchoring bias, such as AVMs or a mark-to-marketed home value. In particular, we will focus on the different valuation methods' abilities to rank-order realized risk as indicated by typically-used measures of mortgage delinquency.

In addition, we conduct a swap-set analysis that will measure whether using one valuation method for a risk-based underwriting would lead to a mortgage portfolio with higher losses than another. Since this is credit policy-specific, we will measure these losses across the range of possible risk-based cut-offs.

We also examine whether factors related to hitting the mark by themselves are indicative of higher risk of a loan. This is accomplished by examining whether the variable has any significance in the risk model, along with an analysis of whether there are appraiser-specific effects in the credit risk of a loan.

We first estimate a mortgage credit risk model that is typically used in underwriting loans. The dependent variable and definition of risky event is whether or not the customer goes 90 days or more past due in the first 2 years on book, including charge-offs, bankruptcies, foreclosures, and other forms of liquidation. This performance window is used largely because it would allow the entire sample of appraisals from 2013-2017 to be included. In addition, it is consistent with what financial institutions typically use for originations models. In robustness analysis available upon request, we find that restricting to the 2013-2014 vintages and using a 4 year performance window yields largely the same results. We also estimate mortgage prepayment and repurchase models simulataneously since these are also significant events, with around 20% of loans prepaying within the performance window and would adversely skewed coefficients if not included.

This model includes most factors that would typically be included in such a model such as LTV,

DTI, loan amount, and FICO. This data does not include detailed credit bureau information on the customer's tradelines which are commonly used in such models, but still maintains a rank-ordering in the range that one would typically see in a well-functioning acquisitions mortgage model. See table (5) for an overview of results. We also evaluate the rank-ordering ability of the model using Somer's D on a hold-out set, to mitigate the risk of overfitting. In robustness analysis available upon request, I show that using LTV bins versus a linear LTV variable give comparable qualitative results.

First note that the directionalities of the coefficient estimates are as expected, with FICO score decreasing causing a highly significant increase in log-odds of default, and LTV increases causing a significant increase in the likelihood of default for all models shown. In regression (2), which uses an AVM-derived LTV as opposed to an LTV derived from the appraised value, the risk rank-ordering of the model is slightly worse, with this relationship holding across multiple hold-out samples. In addition, note that the coefficient on the AVM-derived LTV is significantly smaller than the appraised LTV, consistent with attenuation bias due to the inherent noisiness of the estimate.

In regression (3), we evaluate whether an appraisal hitting the mark directly impacts the riskiness of the loan, and find that the coefficient is not significant. This does not have any implications on the quality of these appraisals themselves. Instead, it means that these appraisals are not associated with loans that have a higher propensity to default. It is conceivable that these appraisals have a low information value, but they happen to be associated with loans that do not have a higher chance of going delinquent.

In addition, neither the deviation of the appraisal or the first estimate from the AVM value is significant in predicting the riskiness of a loan. This may be surprising, since a borrower or appraiser that inflates the valuation above a certain amount could be attempting to close a marginal credit quality loan. However, there may also be attenuation bias from the noise of the AVM when it is used as a baseline of value, which reduces the significance of these estimates.

However, in regressions (4)-(6) we find that loans with appraisers that have a positive probability of hitting the mark tend to have a higher likelihood of default, although this effect is not significant for all specifications. This may mean indicate that appraisers with biased valuations tend to be

	Holdout Somer's D	
Valuation method	**Appraisal LTV Model**	**AVM LTV Model**
Appraisal LTV	0.6823	0.6783
AVM LTV (entire set)	0.6772	0.6757
AVM LTV (hit the mark)	0.6810	0.6777
AVM LTV (notch points)	0.6813	0.6776
AVM LTV (high LTV)	0.6816	0.6781
AVM LTV (biased appraisers)	0.6828	0.6823

Table 2: Valuation method only used on the *scored* population, with the model being estimated on either AVM LTV or Appraisal LTV. Appraisal LTV is used as default for scoring, with AVM LTV used in certain subsets as indicated.

the ones that push marginal loans through, with appraisals that hit the mark a sign of the biased appraiser rather than a risk factor in its own right. This emphasizes the heterogeneity of valuation quality between *appraisers*.

We now turn to the information value of using an AVM-derived LTV relative to an appraisal-derived LTV for risk rank-ordering purposes. We do this by taking customers in a holdout set not included in the training set of the model, and scoring them with either the AVM LTV or the Appraisal LTV using the same model. The main choice to be made would be which LTV to use when training the model, and we evaluate the results using both possibilities.

Table (2) shows that using an appraised LTV when scoring customers slightly outperforms the AVM LTV for both the model estimated using the AVM LTV and the model estimated with the appraised LTV. Significantly, using the appraised LTV to score customers on a model that was trained on an AVM LTV still slightly outperforms using the AVM LTV for which it was designed. In addition, the model trained on the appraised LTV still outperforms the AVM-trained model using every combination of LTVs in the holdout set.

We also examine whether replacing the appraised LTV with an AVM LTV can improve performance for loans with potentially risky appraisals. Interestingly, using an AVM LTV for appraisals that hit the mark does not provide an improvement in performance over using the appraisal itself, mirroring the insignificant coefficient on the variable in regression (3). However, again, mirroring results in regression (4)-(6), using an AVM value for appraisers that tend to hit the mark slightly

improves risk rank-ordering performance, once again suggesting that while hitting the mark by itself might not be problematic from this perspective, the loss of impartiality of the appraiser does affect the quality of the rest of their valuations. This is the only situation where using an AVM-provided LTV provides lift over the appraisal LTV, although it is extremely small (a 5e-03 increase in Somer's D).

We also score customers that fall on LTV notch points with an AVM LTV, with the rest left with an appraised LTV, and we find that risk rank-ordering is inferior compared to simply using an appraised LTV across the entire spectrum. This is in stark contrast to results from Calem et al. (2017), and means that appraisal values at LTV notch points actually have a higher information value than the AVM-derived LTV, although only be a small margin.

9 Swap-Set Analysis

We now relate the differences in rank-ordering ability of the AVM and appraisal LTVs to their impact on the bank's balance sheet. We can accomplish this using a so-called swap-set analysis where we examine which loans would be acquired if using a AVM-based LTV that would otherwise be rejected by the appraisal LTV, or the swap-ins and compare them to the loans that would be rejected using AVM-based LTVs that would otherwise be accepted using an appraisal LTV, or the swap-outs.

Note that this analysis would require the formulation of an actual credit policy. In the prior analysis, we have only estimated risk scores used for rank ordering loans, but have not chosen a concrete thresholds that the sample bank would use to accept or decline the loans. To make our analysis as general as possible, we conduct our analysis of swap-ins and swap-outs for the entire range of possible cut-offs.

There are a variety of ways to compare the swap-ins and swap-outs of a particular credit policy-one being in the performance of the resulting binary classification system through precision and recall. Precision is the percentage of cases that are marked as risky that are truly risky, or the percent of true positives from the predicted positives. Recall is the ability of the binary classifier to

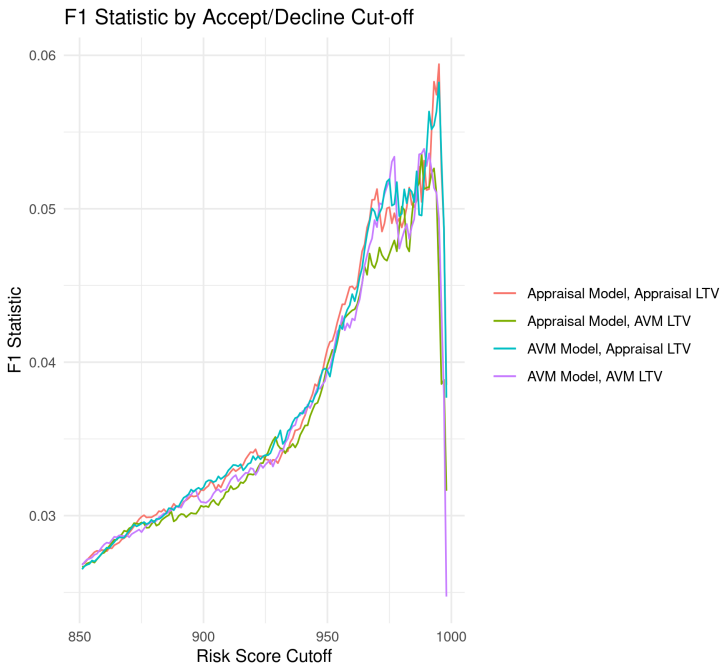

Figure 12: High risk score implies higher probability of bad event. Portfolio with the an x cut-off includes all customers with risk score *at or below* x and leaves out customers with risk score *above* x.

be able to identify the true positives, or the percent of predicted positives out of total true positives. Both are highly relevant for the performance of any binary classification system. For example, it could be that the swap-ins demonstrate fewer bads than the swap-outs, which means that the new credit policy has better recall and fewer false negatives, but also the swap-outs may include a larger proportion of goods, demonstrating poor precision and more false positives.

For any given credit score which rank orders risk, a looser cut-off in the credit spectrum for accept/decline would typically imply a higher precision but a worse recall of bads, while a cut-off that accepts only the highest credit quality individuals would have a higher recall but poor precision of bads. Hence, the setting of any credit policy necessitates weighing the importance of precision and recall. One typical way to do this would be to utilize the F1 statistic, which is the harmonic mean of precision and recall and provides a balance between the two statistics.

We bin risk score into 1,000 buckets, with 0 being the lowest risk individuals, and 1,000 being

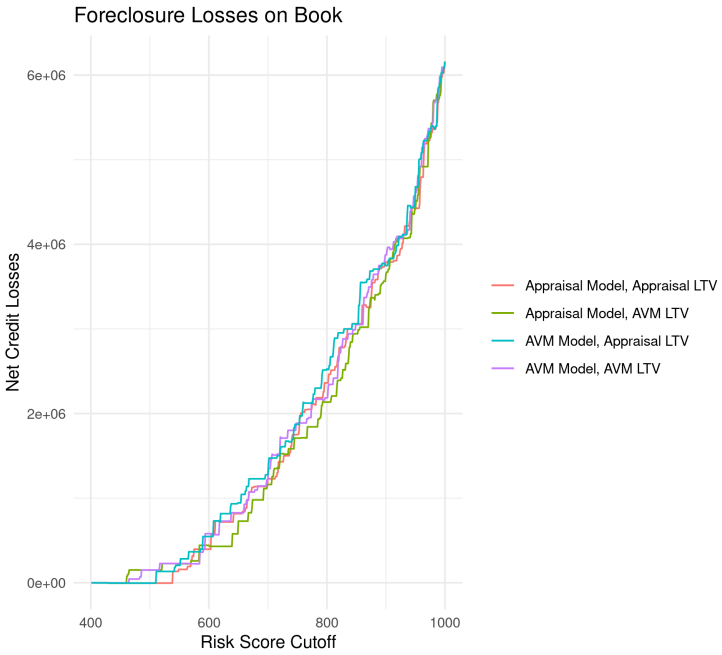

Figure 13: High risk score implies higher probability of bad event. Portfolio with the an x cut-off includes all customers with risk score *at or below* x and leaves out customers with risk score *above* x. Based on total of 123 foreclosures observed within 2 year performance window in this sample.

the highest risk individuals. We compare scoring customers with Appraisal LTV and AVM LTV both in specifications where the model was estimated on an AVM-derived LTV and an appraisal-derived LTV. In figure (12), we examine F1 score across a range of possible cut-offs that maximize F1 score. In general, precision and recall are extremely similar across the range of these models and scores - the bias in the appraisals in this scenario does not seem to make them inferior to using them compared to the AVM value. In fact, using an appraisal-provided LTV both in an AVM and appraisal LTV-estimated model provides the credit policy with the largest F1-score by a small margin. However, again, the differences are minor.

We also examine how foreclosure losses would compare when using these different credit policies. However, the analysis is limited by the small number of foreclosed properties in this particular sample - there are only around 130 properties that foreclosed within the performance window of 2

years in this sample of loans.[11] In figure (13), we compare losses on book from using the different credit policies and once again find that they are broadly similar, with few significant differences. The small size of this foreclosure sample also precludes an analysis of comparative loss reserves using the different valuation methods, since there would be insufficient data to construct a PD-LGD-EAD model. This we leave to future research.

10 Conclusion

We identify anchoring bias in refinance appraisals stemming from appraisers setting home valuations equal to an initial estimate of home value provided by borrowers/loan officers provided in the loan application, a phenomenon we refer to as hitting the mark. This bias occurs significantly less frequently than in the case of purchase money mortgages, with only around 7.5% of appraisals affected.

We find that this amount of hitting the mark cannot arise from rounding or chance, and is highly concentrated among certain appraisers. However, hitting the mark does not seem to be associated with increased loan approval or closing, or appraiser work volume, contrary to popular beliefs on the root causes of appraisal bias. Factors that predict hitting the mark include high LTV loans, and initial estimates of loan value that are significantly over an unbiased AVM value, raising concerns for quality of the appraisal. Hitting the mark is highly concentrated among certain *appraisers* and loans with collateral valued by appraisers that routinely hit the mark are generally riskier controlling for all other observable characteristics, and LTVs derived from their valuations demonstrate worse risk-rank ordering properties. However, when compared to using a credit score that relies on AVM-based LTVs, the appraisal-based LTV still shows slightly favorable risk rank-ordering properties even for loans with biased appraisers, in contrast to findings on the purchase mortgage side. However, this difference is small, meaning that the noise from the AVM and the bias from these appraisals largely cancel out when translated into the risk-rank ordering context.

[11] In robustness analysis we expand the performance window to 4 years, which includes several more foreclosed properties, but restricts the sample to loans originated prior to 2015. This means that the total number of observed foreclosures in the sample remains at around 150, making the sample size roughly equivalent.

The results here show that the presence of bias does not necessarily preclude the use of a particular valuation tool, and appropriately weighting bias against potential noise is essential in understanding the tool's efficacy in the context for which it is used. Indeed, it is conceivable that a valuation tool is biased but is still effective, since it provides satisfactory performance in its use-case of risk rank-ordering.

However, it is worth mentioning that there are other, unexamined ways that the different valuation methods can make an impact on the bank's book. We examined risk rank-ordering for use in underwriting because it is most relevant for the recent policy change on the allowable usage of AVMs for collateral valuation. Collateral valuation is also a key driver in mortgage loss forecasting models. However, due to limited foreclosure sample size in this data set, we reserve this analysis for future research.

References

Agarwal, Sumit, Brent Ambrose, and Vincent Yao (2019), "Can regulation de-bias appraisers?" *Journal of Financial Intermediation*. forthcoming.

Agarwal, Sumit, Itzhak Ben-David, and Vincent Yao (2015), "Collateral valuation and borrower financial constraints: Evidence from the residential real estate market." *Management Science*, 61.

Ben-David, Itzhak (2011), "Financial constraints and inflated home prices during the real estate boom." *American Economic Journal: Applied Economics*, 3, 55–87.

Calem, Paul S., Lauren Lambie-Hanson, and Leonard I. Nakamura (2017), "Appraising home purchase appraisals." *FRB of Philadelphia Working Paper No. 17-23*.

Ding, Lei and Leonard Nakamura (2016), "The impact of the home valuation code of conduct on appraisal and mortgage outcomes." *Real Estate Economics*, 44, 658–690.

Eriksen, Michael D., Hamilton B. Fout, Mark Palim, and Eric Rosenblatt (2019), "Contract price

confirmation bias: Evidence from repeat appraisals." *Journal of Real Estate Finance and Economics*, 1–22.

Fannie Mae (2010), "Appraiser independence requirements." https://www.fanniemae.com/content/fact_sheet/appraiser-independence-requirements.pdf.

Fannie Mae (2019), "LLPA matrix." https://www.fanniemae.com/content/pricing/llpa-matrix.pdf.

FDIC (2018), "FIL-76-2018: Appraisal threshold for residential real estate loans." https://www.fdic.gov/news/news/financial/2018/fil18076.pdf.

Griffin, John M. and Gonzalo Maturana (2016), "Who facilitated misreporting in securitized loans?" *The Review of Financial Studies*, 29, 384–419.

Housing Wire (2017), "When do AVMs make the most sense?" https://www.housingwire.com/articles/40935-when-do-avms-make-the-most-sense.

Investopedia (2019), "6 reasons to avoid private mortgage insurance." https://www.investopedia.com/mortgage/insurance/avoid-pmi/.

Quicken Loans (2015), "Quicken loans launches revolutionary end-to-end online product Rocket Mortgage transforming how consumers experience the home loan process." https://www.quickenloans.com/press-room/2015/11/24/launches-rocket-mortgage/.

Shi, Lan and Yan Zhang (2015), "Appraisal inflation: Evidence from the 2009 GSE HVCC intervention." *Journal of Housing Economics*, 27, 71–90.

Tzioumis, Konstantinos (2017), "Appraisers and valuation bias: An empirical analysis." *Real Estate Economics*, 45, 679–712.

Yun, Lawrence (2017), "Realtors Confidence Index, report on the March 2017 survey." https://www.nar.realtor/sites/default/files/reports/2017/2017-03-realtors-confidence-index-04-21-2017.pdf.

A Rounding Analysis

One potential explanation behind "hitting the mark" is that it is simply the artifact of rounding. It could be that both borrowers and appraisers independently come up with an estimate of value and these have some chance of coinciding because they would both round to the nearest thousandth or ten thousandth. Indeed, examining table (7) one can see a high rate of rounding in both the first estimate of value and the appraisal amount, with a slightly higher rate of rounding in the first estimate of value compared to the appraisal.

However, if one restricts to appraisals where the initial estimate was not rounded, one can see in table (8) there is still a significant rate of hitting the mark. For example, for a realistic level of rounding to the nearest $\$10,000$, appraisals with rounded initial estimates still have a slightly higher rate of hitting the mark, but the level of hitting the mark is only 2 percentage points lower for the non-rounded values. This suggests that hitting the mark occured through appraisal knowledge of the initial estimate rather than random chance.

One can also see that in table (8 that there are fewer appraisals that hit the mark among the loans where the first estimate was not rounded to the nearest $\$1,000$. However, this is not likely explained by a lack of appraisal foreknowledge. Indeed, it would take an incredible level of appraisal accuracy if hitting the mark was explained by unbiased appraisers rounding to the nearest $\$1,000$. In figure (17) we simulate the level of "random" hitting the mark that would occur if appraisals were just noisy signals of the initial estimate of borrower value, for various levels of noise and rounding. We find that there would only be a significant rate of random hitting the mark when rounding to the nearest $\$1,000$ if appraisals fell incredibly closely to the initial estimate. Indeed, appraisals would have to have a PPE10 of over 95% relative to the initial estimate to get more than a few percentage points of hitting the mark with that level of rounding.

In addition, note that these simulations would be an upper bound on the effect of rounding on hitting the mark since the initial estimate itself is a noisy and potentially biased signal of true value, and an unbiased appraisal may or may not be correlated with it, which would further reduce the likelihood of it randomly coinciding with the initial estimate when rounded.

B Additional Figures and Tables

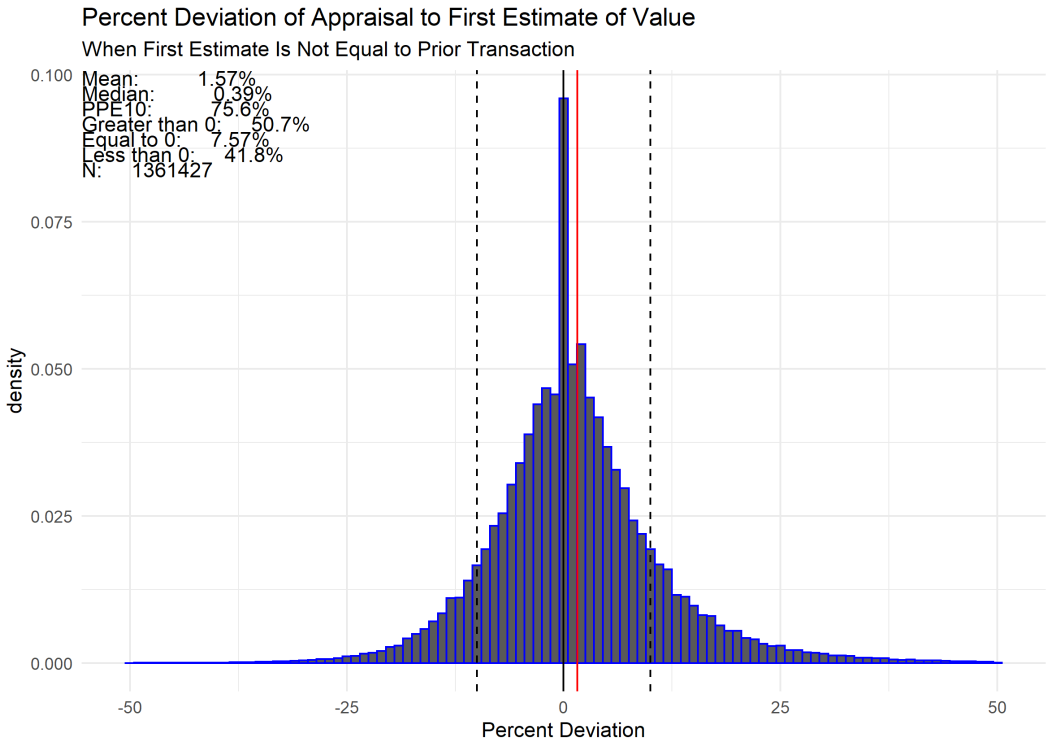

Figure 14: Restricted to Limited Cash-Out Refinance.

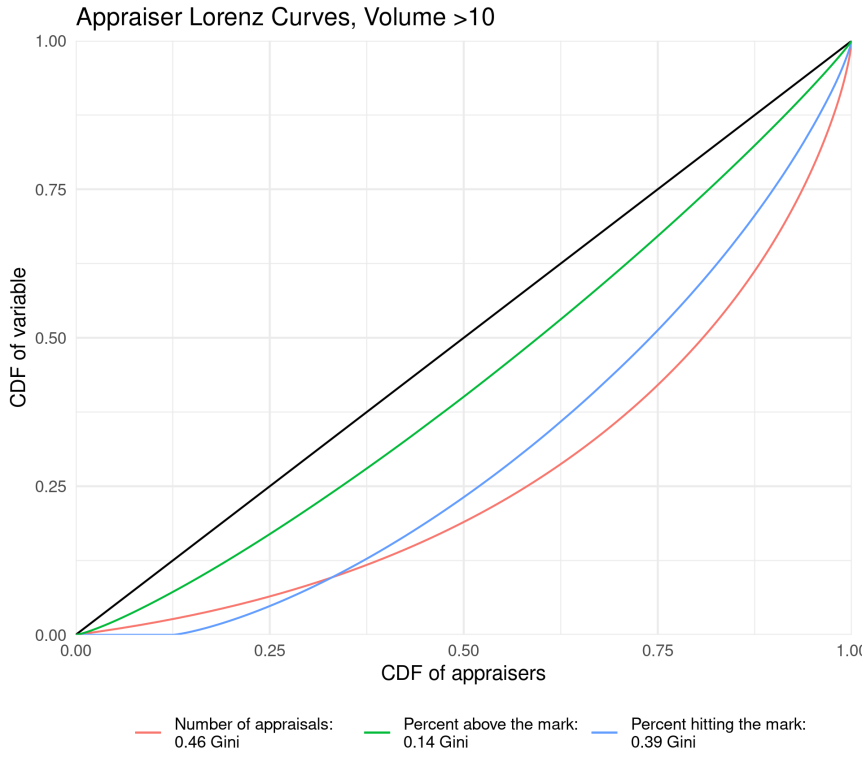

Figure 15: Concentration of hitting the mark among appraisers with at least 10 appraisals.

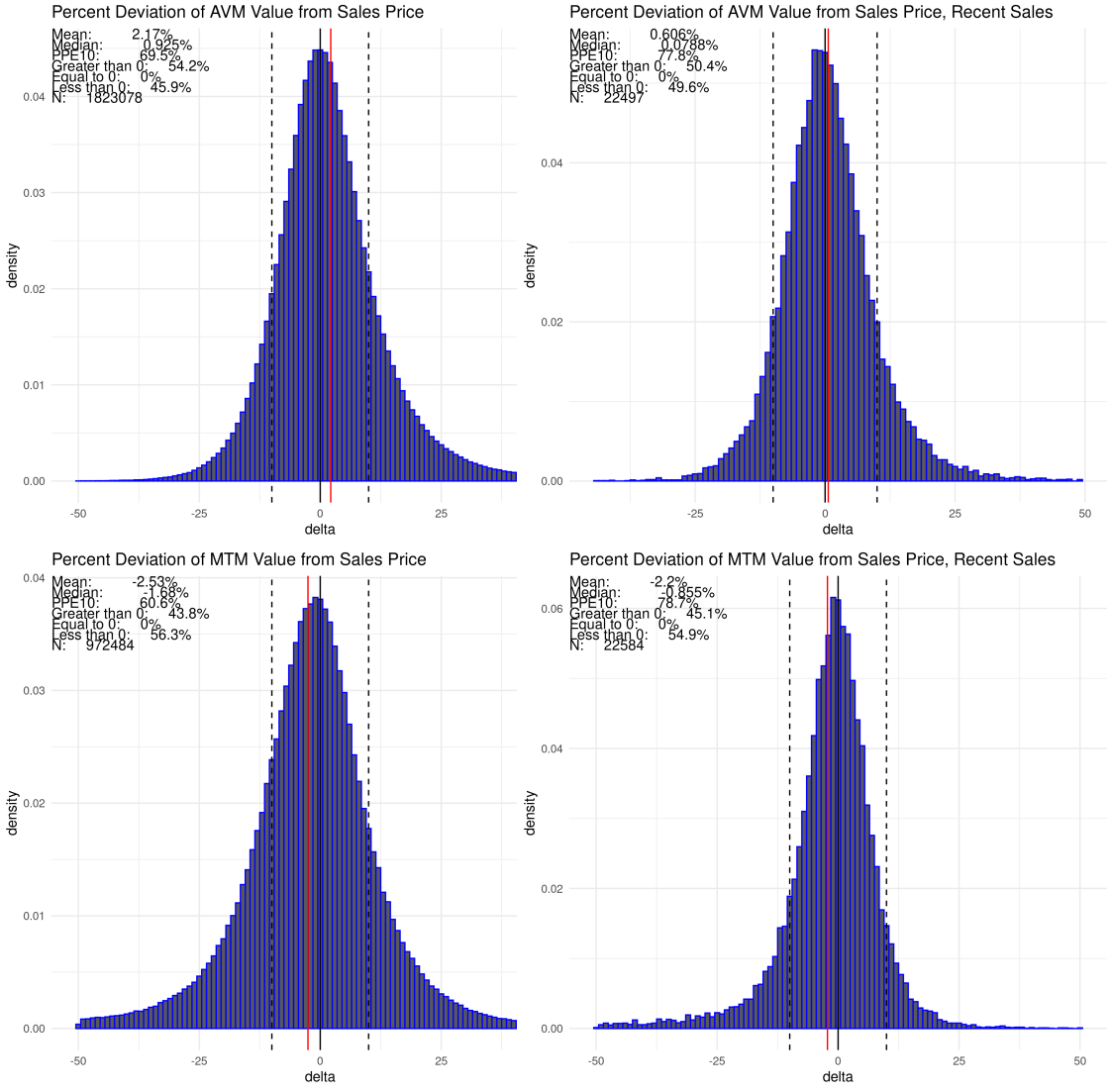

Figure 16: Estimated on set of purchase mortgages from 2013-2017. "Recent sales" are those purchases that are repeat transactions with prior sales between 1-2 years ago.

	Dependent Variable: Hitting the Mark						
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**
Intercept	**-3.14**	**-3.21**	**-2.21**	**-2.47**	**-2.32**	**-2.5**	**-2.31**
	(1.07)**	(1.07)**	(1.18e-01)***	(1.20e-01)***	(1.19e-01)***	(1.20e-02)***	(1.20e-01) ***
Change in Loan Life Cost	**3.30e-06**	**4.68e-06**	**4.40e-06**				
from Going Up 1 LTV bucket	(1.34e-07)***	(1.31e-07)***	(1.32e-07)***				
Dummy for being		**2.63e-01**	**2.65e-01**				
in 76-80 LTV bucket		(5.95e-03)***	(5.97 e-03)***				
Change in LLPA Cost				**4.00e-06**			
from Going Up 1 LTV Bucket				(1.63 e-07)***			
Change in PMI Cost from Going Up 1 LTV Bucket				**2.34e-04**			
				(7.21e-05)***			
Pc. Dev. of First Estimate to AVM			**1.26e-02**	**1.24e-02**	**1.40e-02**	**1.38e-02**	**1.40e-02**
			(1.88e-04)***	(1.89e-04)***	(1.93e-04)***	(1.93e-04)***	(1.93e-04)***
ARM Dummy				**-3.66e-02**	**5.45e-02**	**1.51e-02**	
				(5.30e-03)**	(2.79e-05)***	(1.32e-02)	
Loan Term				**8.97e-04**		**6.46e-04**	**5.31e-02**
				(3.45e-05)***		(3.47e-05)***	(1.30e-02)***
Appraiser Work Volume							5.93e-05
							(2.39e-05)
640 $\leq$ FICO <660				3.62e-02	-2.58e-02	-2.38e-02	-2.31e-02
				(2.39e-02)	(2.37e-02)	(2.37e-02)	(2.38e-02)
660 $\leq$ FICO <680				**8.03e-02**	-5.63e-02	-5.39e-02	-5.50e-02
				(2.26e-02)***	(2.37e-02)*	(2.21e-02)*	(2.38e-02)*
680 $\leq$ FICO <700				**1.32e-01**	**-9.84e-02**	**-9.79e-02**	**-9.83e-02**
				(2.17e-02)***	(2.08e-02)***	(2.08e-02)***	(2.08e-02)***
700 $\leq$ FICO <720				**1.18e-01**	**-1.16e-01**	**-1.14e-01**	**-1.15e-01**
				(2.13e-02)***	(2.04e-02)***	(2.04e-02)***	(2.05e-02)***
720 $\leq$ FICO <740				**8.69e-02**	**-1.67e-01**	**-1.66e-01**	**-1.66e-01**
				(2.12e-02)***	(2.03e-02)***	(2.03e-02)***	(2.04e-02)***
740 $\leq$ FICO				**1.90e-02**	**-2.22e-01**	**-2.15e-01**	**-2.21e-01**
				(2.02e-02)	(1.91e-02)***	(1.91e-02)***	(1.91e-02)***
Dummies for LTV at first submission.					*See figure (10)*	...	...
Somer's D	0.081	0.107	0.144	0.155	0.171	0.173	0.171

Table 3: Regression on "hitting the mark."

	Dependent Variable: Going Above the Mark					
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Intercept	5.22	-3.2	**-1.33e-01**	**-3.91e-01**	**-5.36e-01**	**-3.01e-01**
	(1.39)	*(1.07)***	*(7.75e-02)****	*(7.87e-02)****	*(7.91e-02)****	*(7.93e-02)****
Change in Loan Life Cost	**-2.35e-06**	**-1.89e-06**	**5.59e-07**			
from Going Up 1 LTV Bucket	*(7.27e-08)****	*(8.04e-08)****	*(9.11e-08)****			
Dummy for being		**6.81e-02**	**1.39e-01**			
in 76-80 LTV bucket		*(3.30e-03)****	*(3.80e-03)****			
Change in LLPA Cost				**7.07e-06**		
from Going Up 1 LTV Bucket				*(1.17e-07)****		
Change in PMI Cost				**2.08e-04**		
from Going Up 1 LTV Bucket				*(4.84e-06)****		
Pc. Dev. of First estimate			**-1.15e-01**	**-1.15e-01**	**-1.54e-01**	**-1.15e-01**
to AVM			*(1.92e-04)****	*(1.93e-04)****	*(1.94e-04)****	*(1.94e-04)****
ARM Dummy				5.64e-03	-2.89e-03	**5.02e-02**
				(8.10e-03)	*(8.02e-03)*	*(8.13e-03)****
Loan Term				**-6.35e-04**		**-8.24e-04**
				*(2.07e-05)****		*(2.09e-05)****
640 $\leq$ FICO <660				8.29e-02	2.88e-02	2.62e-02
				(1.55e-02)	*(1.54e-02)*	*(1.54e-02)*
660 $\leq$ FICO <680				**1.73e-01**	3.96e-02	3.61e-02
				*(1.45e-02)****	*(1.43e-02)**	*(1.43e-02)**
680 $\leq$ FICO <700				**2.94e-01**	**4.75e-02**	**4.60e-02**
				*(1.38e-02)****	*(1.35e-02)****	*(1.35e-02)****
700 $\leq$ FICO <720				**3.43e-01**	**9.60e-02**	**9.33e-02**
				*(1.35e-02)****	*(1.32e-02)****	*(1.32e-02)****
720 $\leq$ FICO <740				**3.75e-01**	**1.09e-01**	**1.06e-01**
				*(1.34e-02)****	*(1.31e-02)****	*(1.31e-02)****
740 $\leq$ FICO				**5.21e-01**	**2.75e-01**	**2.65e-01**
				*(1.28e-02)****	*(1.24e-02)****	*(1.24e-02)****
Dummies for LTV					...	...
Somer's D	-0.032	0.123	0.608	0.614	0.614	0.619

Table 4: Regression on going "above the mark."

	Dependent variable: 90+ day delinquency within 2 years of origination.							
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**
Intercept	**3.03**	**3.32**	**3.03**	**2.95**	**3.02**	**3.03**	**3.04**	**3.03**
	*(5.09e-01)***	*(5.06e-01)****	*(5.09e-01)****	*(5.12e-01)****	*(5.11e-01)****	*(5.11e-01)****	*(5.10e-01)****	*(5.10e-01)****
Original Interest Rate	**4.61e-01**	**4.94e-01**	**4.61e-01**	**4.63e-01**	**4.65e-01**	**4.65e-01**	**4.60e-01**	**4.61e-01**
	*(6.29e-02)****	*(6.25e-02)****	*(6.23e-02)****	*(6.31e-02)****	*(6.31e-02) ****	*(6.31e-02)****	*(6.28e-02)****	*(6.28e-02)****
Original Loan Amount	**-1.38e-06**	**-1.37e-06**	**-1.38e-06**	**-1.43e-06**	**-1.39e-06**	**-1.38e-06**	**-1.39e-06**	**-1.38e-06**
	*(2.55e-07)****	*(2.54e-07)****	*(2.55e-07)****	*(2.56e-07)****	*(2.56e-07)****	*(2.56e-07)****	*(2.55e-07)****	*(2.55e-07)****
Appraisal-based LTV	**2.55e-02**		**2.55e-02**	**2.55e-02**	**2.54e-02**	**2.54e-02**	**2.54e-02**	**2.55e-02**
	*(2.15e-03)****		*(2.16e-03)****	*(2.16e-03)****	*(2.16e-03)****	*(2.16e-03)****	*(2.15e-03)****	*(2.23e-03)****
AVM-based LTV		**1.78e-02**						
		*(1.67e-03)****						
DTI	**3.08e-02**	**3.11e-02**	**3.08e-02**	**3.05e-02**	**3.05e-02**	**3.05e-02**	**3.08e-02**	**3.08e-02**
	*(3.35e-03)****	*(3.33e-03)****	*(3.35e-03)****	*(3.35e-03)****	*(3.35e-03)****	*(3.35e-03)****	*(3.35e-03)****	*(3.35e-03)****
FICO Score	**-1.86e-02**	**-1.80e-02**	**-1.83e-02**	**-1.83e-02**	**-1.83e-02**	**-1.83e-02**	**-1.83e-02**	**-1.83e-02**
	*(5.58e-04)****	*(5.53e-04)****	*(5.58e-04)****	*(5.59e-04)****	*(5.59e-04)****	*(5.59e-04)****	*(5.58e-04)****	*(5.58e-04)****
Loan Term	1.19e-03	**1.17e-03**	**1.19e-03**	**1.21e-03**	**1.19e-03**	**1.20e-03**	**1.19e-03**	**1.19e-03**
	*(4.49e-04)**	*(4.48e-04)***	*(4.49e-04)***	*(4.51e-04)***	*(4.51e-04)***	*(4.51e-04)***	*(4.49e-04)****	*(4.49e-04)****
More than 1 Borrower	**-4.94e-01**	**-4.97e-01**	**-4.94e-01**	**-4.90e-01**	**-4.91e-01**	**-4.91e-01**	**-4.94e-01**	**-4.94e-01**
	*(5.47e-02)****	*(5.47e-02)****	*(5.47e-02)****	*(5.48e-02)****	*(5.48e-02)****	*(5.48e-02)****	*(5.47e-02)****	*(5.47e-02)****
Appraisal Hit the Mark			-5.89e-02					
			(9.78e-02)					
Appraiser Hits the Mark >5% of the time				**1.54e-01**				
				*(5.55e-02)***				
Appraiser Hits the Mark >10% of the time					8.38e-02			
					(5.76e-02)			
Appraiser Hits the Mark >20% of the time						3.04e-01		
						(1.64e-01)		
Appraisal Deviation from AVM							2.00e-02	
							(2.34e-02)	
First Valuation Deviation from AVM								2.52e-05
								(1.84e-03)
Holdout Somer's D	**0.679**	**0.673**	**0.667**	**0.667**	**0.667**	**0.667**	**0.679**	**0.679**

Table 5: Dependent variable is 90+ day delinquency within 2 years of origination. 20% holdout set.

	Dependent Variable: Hitting the Mark		
	(1)	**(2)**	**(3)**
Intercept	4.87	4.96	4.73
	*(1.99)**	*(1.98e)**	*(2.00)**
Original Interest Rate	4.58e-01	4.67e-01	5.15e-01
	(2.48e-01)	*(2.46e-01)*	*(2.46e-01)*
Original Loan Amount	-1.45e-06	-1.51e-06	-1.31e-06
	(8.60e-07)	*(8.61e-07)*	*(8.49e-07)*
Appraisal-based LTV	2.11e-02		
	*(8.99e-03)**		
AVM-based LTV		**1.94e-02**	
		*(6.86e-03)***	
MTM-based LTV			1.20e-02
			*(5.72e-03)**
DTI	**3.53e-02**	**3.53e-02**	**3.49e-02**
	*(1.26e-02)***	*(1.25e-03)***	*(1.25e-02)***
FICO Score	**-2.10e-02**	**-2.09e-02**	**-2.01e-02**
	*(2.16e-03)****	*(2.14e-0 3)****	*(2.10e-03)****
Loan Term	2.00e-03	1.98e-03	1.72e-03
	(1.97e-03)	*(1.97e-03)*	*(1.96e-03)*
More than 1 Borrower	-2.82e-01	-2.82e-01	-2.75e-01
	(5.51e-02)	*(1.85e-01)*	*(1.85e-01)*

Table 6: Robustness Analysis: Hitting the mark with a MTM LTV. Restricted to limited cash-out refinances with prior transaction between 1 and 2 years prior.

	First Estimate	Last Appraisal
Multiple of 100	97.9%	99.8%
Multiple of 1,000	95.4%	96.5%
Multiple of 5,000	81.3%	64.9%
Multiple of 10,000	55.6%	37.4%
Multiple of 50,000	26.2%	10.6%

Table 7: **Level of Rounding in Appraisals.**
Note: Restricted to limited cash-out refinances

Level of Rounding (First Estimate)	Hitting the Mark	
	Rounded	Non-rounded
Multiple of 100	7.57%	1.58%
Multiple of 1,000	7.72%	1.70%
Multiple of 5,000	8.36%	3.48%
Multiple of 10,000	8.43%	6.21%
Multiple of 50,000	8.30%	7.15%

Table 8: **Hitting the Mark Between Appraisals with Rounded and Non-rounded Initial Estimates.**

Note: Restricted to limited cash-out refinances. Rates of hitting the mark were calculated for appraisals where the initial estimate was rounded to various common factors, and contrasted with the rate of hitting the mark with non-rounded initial estimates.

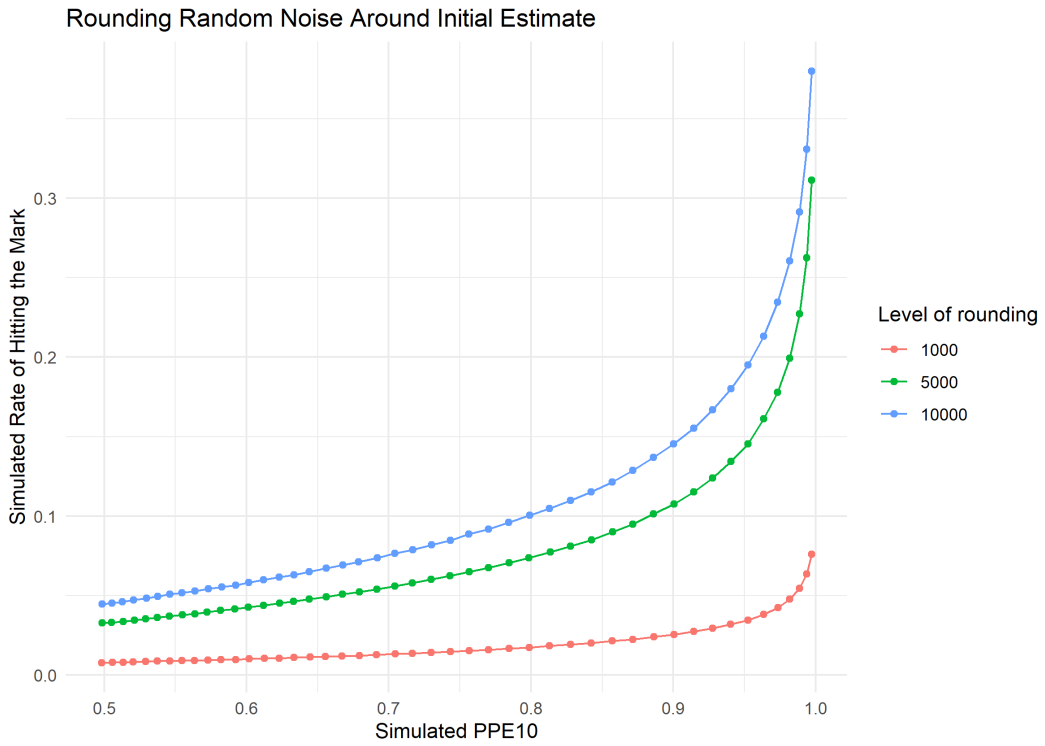

Figure 17: We vary the level of random noise around the initial estimate, with "simulated PPE10" relative to the initial estimate denoted on the X axis. On the Y axis, we denote the rate of concurrence between the rounded noisy estimate with the initial estimate itself. This provides an upper bound to the level of hitting the mark that could be explained by rounding for those values whose initial estimate was rounded.