WHAT HAPPENS AFTER YOU OVERPAY FOR A HOUSE

Weifeng Wu†

Eric Rosenblatt

Fannie Mae

This version: December 01, 2018

First version: January 31, 2018

Abstract

In this paper we study whether borrowers who overpaid for their house, are more likely to default later on, *ceteris paribus*. The occurrence of financial crisis in 2008 has seriously challenged the practice of using appraisals to safeguard collateral value. Here we instead use several measures of independent model-based valuations, against which we compare the purchase price. We demonstrate that compared to these valuations, borrowers who overpay are indeed monotonically more likely to experience serious default later on. Further we show that even if these loans are performing fine, unsurprisingly the borrowers eventually realize less gain when selling the house later on, controlling for the local house appreciation trend. This implies that these model-based valuations, while by methodological design are not completely removed from the general rising or declining price trend, can still detect these most egregiously inflated bubbles, and thus serve as a useful yardstick in prudential lending, which in the end not only help the lender and GSE, but also benefit the borrowers in the long run; otherwise, they either miserably default more, or if not default, receive less wealth accumulation from owning this overpaid house.

Keywords: mortgage loans performance, collateral valuation, appraisal bias, financial crisis

JEL Classification Codes: G21, G28, K11, L85, R31

---

† Corresponding author: weifeng_wu@fanniemae.com. We thank colleagues at Fannie Mae and FHFA for their suggestions and comments. All remaining errors are our own. The views presented in this paper are those of the authors alone and do not necessarily reflect those of the Fannie Mae or FHFA.

1 Introduction

Most residential home sales in the US are backed by mortgages from the lenders and for them to be sure of the collateral value, an appraisal is required. By policy design, appraisers are licensed or certified by governmental agencies, and expected to have expert knowledge in local markets as compared to any other interested parties, like buyers, sellers, and realtors. They are required to enter the property, inspect the location, condition and quality, and evaluate the unique features/improvements; and based on this on-site observation as well their understanding on recent local demand and supply trends, they develop a professional opinion of the value. This process is human intensive, simply because every real estate property is unique, and these properties transact very infrequently. However, before appraisers do their professional assessment, bank lending rules require that they are to be presented with a sales contract already negotiated by the seller and buyer. This extra instruction does not only lead to a confirmation bias[1] to the process, but also more importantly, it completely changes the incentive of appraisers: appraisers would like and also have to confirm the contract, in order to make sure the transaction can go through smoothly. In this case, the opinion of value provided by the appraiser is heavily influenced by the contract price and thus is of questionable value.

By a similar token, potential buyers on the market, ideally after analyzing their own financial situation and personal preference, know what their target houses are in terms of location, size, style, condition and quality, etc. In today's information age, together with the help of real estate agents, it should not be difficult for them to gather these statistics on what the nearby and similar properties are sold for. Empowered with this knowledge, the potential buyers should have an idea of how much the house they are negotiating is worth, at least by looking at Zillow's valuation. However, buyers may not do the homework as rigorously as one might expect, and without knowing the misguided incentive structure in the appraisal profession[2], would naively think an appraiser will be their best safeguard.

Before the housing crisis, the appraisers usually worked directly for the lender, who had

[1] Per Wikipedia, this is a tendency to search for, interpret, favor, and recall information in a way that confirms one's preexisting beliefs or hypotheses.

[2] Even if they know that the appraisers are not trustworthy, the cost of following through a professional appraiser's workflow may prevent many buyers from doing so.

the financial incentive to close the transaction, which means the appraisers were under the constant pressure of saying the house value is equal to or higher than the contract price. This is why after the financial crisis, Attorney General of New York and GSEs jointly issued the Home Valuation Code of Conduct (HVCC) in May 2009, and one mandate of which is to separate the the appraisal workflow from lending decision, leading to the flourishing of appraisal management companies (AMC)[3]. Nevertheless we continued to see a majority of appraisals confirming the contract price. As pointed in Eriksen, Fout, Palim, and Rosenblatt (2016), for three million purchase money appraisals submitted to a national mortgage guarantor from 2012 to 2015, about 64.6% exceed the contract price and 28.8% exactly hit the contract price (even though among these who hit the contract price, a third of contracts say the sellers make concession). And for these where the appraisals exceed or hit the contract, there is little motivation to renegotiate after the appraisal because of the prospect of no success: in about 96.6% of such cases, the contract prices will become the eventual sales price. What is still obviously lacking is that buyers are not getting the independent valuation check that they paid for.

Given that appraisals does not constitute a valid check for the housing value, it should not surprise anyone that at least some of these house buyers could end up paying too much: if they do, they either need to put more money down, or borrow from a higher LTV loan, both of which are detrimental for their long-term financial health. In this paper we would like to look at several benchmark valuations that do not suffer from this man-made disincentive, and see if using these measures rather than the appraised values, will help identify these overpaying borrowers, and improve the quality of mortgage underwriting.

While it is common knowledge that over-inflating collateral value is bad for loan performance, there is little quantitative evidence for that, because of the difficulty to locate convincing benchmarks[4]. This paper is one of the first to use various empirical valuations in the secondary mortgage market to compare with the selling price. We carefully examined four benchmarks: (a) the predicted collateral value at the time of loan origination by an automatic valuation model (AVM); (b) a mark-to-market value prediction around the time

[3]AMC as a fire wall between the lenders and the appraisers does not seem to solve the problem. As pointed in Shui and Murthy (2018), the empirical comparisons indicate no clear evidence of any systematic quality differences between AMC and non-AMC appraisals.

[4]In individual cases, AVM could suffer from a lot of impreciseness, like lack of human inspection of the house condition, etc.

of loan origination, utilizing the available contemporaneous home price index (HPI); (c) a retrospective value prediction around loan acquisition/delivery to the mortgage guarantor; and (d) an ex post mark-to-market value prediction, using the HPI based on data available at a later date.

While each of the benchmarks may differ in terms of available information incorporated and thus prediction coverage and accuracy, none of them face the pressure from the lenders to let the loan go through, a byproduct of which is we can have a substantial percentage of houses identified as being overpaid. More importantly, compared to these benchmarks, loans with an over-inflated selling price (which in turn was confirmed by an over-inflated appraisal report), defaults at a significantly higher probability: the quantitative difference in six-month delinquency within the first five years[5] could be 5%-10% in the peak of the housing price. We prove this by looking at the contribution of the appraisal bias in a default model, by comparing the loan performance by the appraisal bias groups, and by comparing the defaults in a controlled setting as is done through a propensity score matching exercise. Moreover, perhaps more importantly for the majority of borrowers, we then trace these overpaid borrowers who did not default in their mortgage overtime, until the next time their houses go on the market. It shows that these houses eventually sold with less profit than the average house in the neighborhood, as indicated by a local price index. This implies that, these model-based valuations, while by methodological design are not completely removed from the rising or declining price trend in the underlying data[6], can still detect these most egregiously inflated bubbles, and thus serve as a useful yardstick in prudential lending.

This paper proceeds as follows: the next section describes each of these four benchmarks, section III investigate the role of appraisal bias in a default regression framework and in a propensity matching framework, section IV examines the profits from future sales for the majority of borrowers who did not default, and section V offers the concluding remarks.

[5]So this is more likely a default measure rather than serious delinquency rate which is usually defined as 60 or 90 days past due. We use delinquency and default interchangeably in this paper.

[6]This is because these model valuations are generated using the reported sales prices as input.

2 Data Description

2.1 Benchmark Definition

While it is intuitive to say that an individual buyer has overpaid for his/her house likely in a buyer's remorse setting, it is difficult to precisely define "over-inflation" systematically because this requires a fundamental value to compare with what the buyers paid; however, such fundamental is a latent variable, and is also likely to fluctuate as time goes on. This unobserved nature of true (or intrinsic) collateral value is the key reason why we see a whole set of lending rules are designed to solve this problem, yet the housing market still saw a dramatic up and down, cumulating in the 2007/2009 subprime mortgage crisis.

Here we will explore several benchmark valuations against which the purchase price of an arms-length transaction can be compared. These benchmarks are generated by statistical models that take into account nearby sales information but most importantly, avoids the incentive issue in human appraising: models are neither given the contract price, nor to face the pressure from the loan officers or the AMCs hired by banks.

First we start with the value predictions from a large mortgage guarantor's AVM framework, at the time of mortgage origination. This framework consists of a hybrid of models like mark-to-market, tax assessment, and hedonic regressions. At the time of loan application, lenders will submit information regarding the borrowers and the collateral, into the guarantor's proprietary underwriting system, which will then give a prediction on the value of the collateral at that time. Since this prediction is a number generated mechanically by computer, without the knowledge of the contract price, it could be and is usually higher or lower from the contract price, and sometimes dramatically so. The framework also includes a confidence score which is related to the density of information the computer is using, where 1, 2 and 3 refers to "good", "fair", and "OK", while 4 and 5 are usually considered as "not acceptable" and 0 means "a confidence score cannot be produced". To minimize the effect of imprecise predictions, we limit the sample to those with a confidence score between 1 and 3.

Second given that this AVM framework is a sophisticated black box, we decide to test a more transparent alternative, the *contemporaneous* mark-to-market value. Using the property transaction data, whether or not they have mortgages, we can identify the prior arms-length sales, where here "prior" means at least six months[7] before the earlier of the current loan's acquisition date and first payment date. The contemporaneous HPI data is produced by the mortgage guarantor for zips where we have seen sufficient single-family sales and refinance data every quarter since 1997. So for each mortgage in our sample, we can find the appropriate index series whose vintage quarter is equal to the loan origination quarter, and from that particular vintage, we can find the index corresponding to the quarter of the prior sale quarter and that of the current loan. The prominent feature of this mark-to-market valuation is that it is utilizing a contemporaneous index series, so it is only using information already available in the origination quarter; the impact from any future sales/refinance data is nil.

The third benchmark is the guarantor's retrospective property value, which comes from a similar AVM framework. After the loans are originated, within a couple of months, most of them are sold to a national mortgage guarantor, who will then conduct a loan quality review, part of which is to assess the collateral value with updated information as of the loan acquisition and every month after that, up to one year from the loan acquisition. The value we pull is the prediction using all information as of one year after loan acquisition, since by that time, all relevant transactions up to the loan origination should be observed by the guarantor and thus should contain pretty accurate information regarding the local market. We will just call this the acquisition AVM prediction.

The fourth benchmark is an ex post mark-to-market prediction, which is similar to the second one, except that we now use only a single HPI vintage, the one generated in the second quarter of 2016. This 2016Q2 vintage was estimated utilizing all transaction data available to the guarantor up to 2016, and should have almost complete coverage of transactions before 2015. Because all future transactions are incorporated in this index and we have experienced a big boom-and-bust housing cycles, valuations based on this index have a strong forward-looking component, with futurity in existence. It should be interesting to compare this benchmark with the second one, to assess the effect of future data on prior HPIs and on the appraisal biases. For example, to look at the average appreciation between 2006 and 2007, if we estimate the HPI at 2008Q1, the contribution comes from any house that was transacted before 2006 and in 2007 (not any future years);

[7] So if there is another sale within 180 days prior, it is unlikely that this is a typical arms-length sale: maybe a house fix-and-flip, or a complete teardown, where the mark-to-market will be wide of the mark.

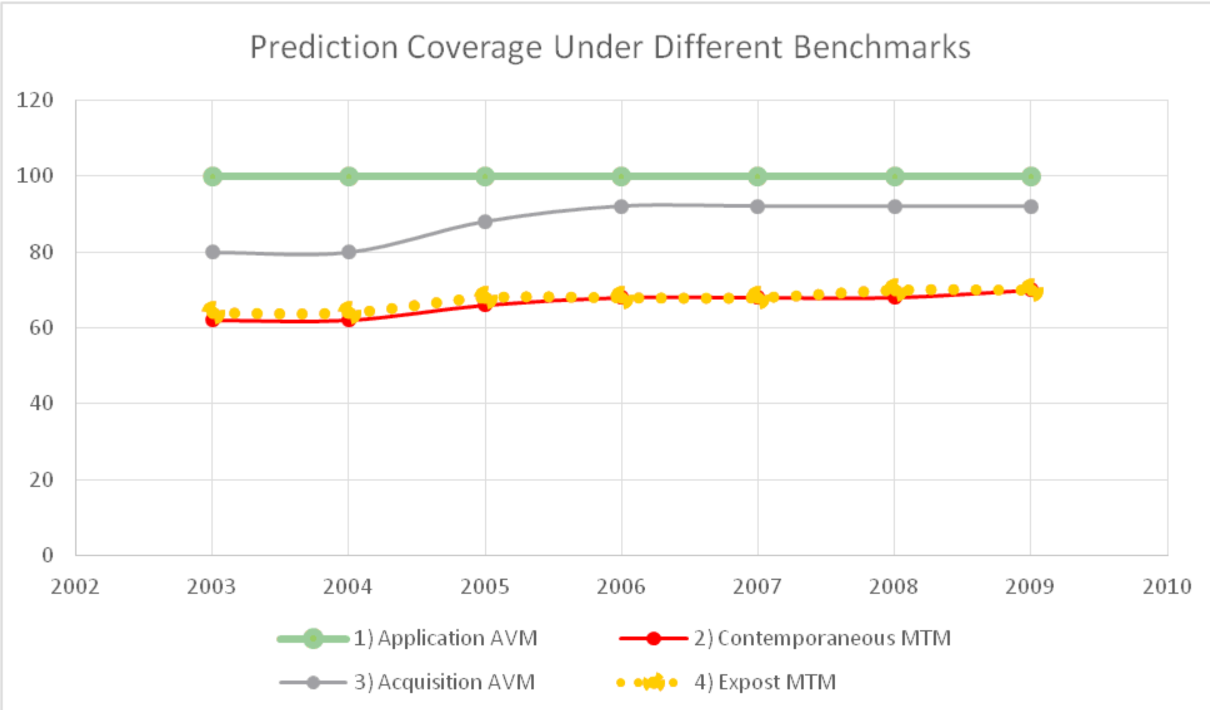
on the other hand, if we estimate the HPI at 2016Q2, a house sold in 2005 and then again in 2015 will also affect the return between 2006 and 2007. We did see a price decline in 2008, which should make the appreciation estimated in 2016Q2 smaller than a contemporaneous estimate.

2.2 Benchmark Coverage

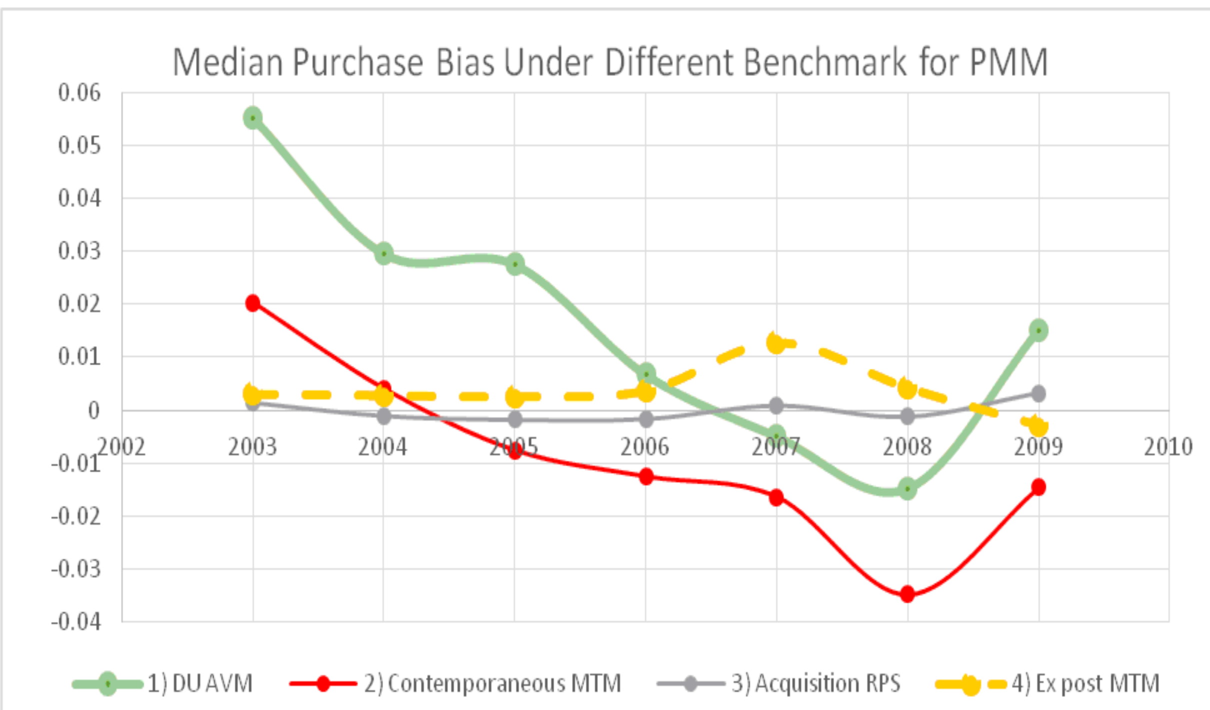
Before we proceed to use these benchmarks, we would like to do some comparisons regarding their coverage. Our sample starts with pulling all conventional (first-lien, owner-occupied) loans where we have an AVM prediction with sufficient confidence at origination. So the first benchmark's coverage is 100%. The next broadest coverage is from the retrospective prediction: because of confidence requirements, the coverage is about 80% before 2005, and rises to 92% since 2006. This should come as no surprise, because these two predictions are from the similar modeling methodology, both created for loan quality control purpose, with timing and thus information being the only difference. The mark-to-market predictions are between 60 and 70%, where the ex post version is slightly better than the contemporaneous one. The remaining 30% are either because we cannot find the prior sale or because the zip- level index is not readily available due to the scarcity of nearby sales. Of course as time goes by, the data lag problem will mitigate so not surprisingly the ex post HPI has slightly more coverage than the contemporaneous one. Overall, coverage is strongly related to the underlying methodology rather than the information set. This is displayed in Figure 1(a)

2.3 Overpayment Measure

With four benchmarks in place, we define the overpayment measure as purchase price divided by the benchmark minus 1; so a positive measure means the house buyer has "overpaid" as compared to the benchmark, while a negative one indicates an "underpayment", i.e. the buyer is getting a good deal. Ideally for the purchase money mortgage, if the benchmark prediction is estimated with full transaction data observed, the median or average bias for the whole estimation sample should be close to zero, because if not,



(a) Coverage



(b) Median Bias

Figure 1: Four Benchmark Predictions

then the estimation process can always correct for this bias[8]. This is indeed true for the acquisition AVM, which is generated one year after loan acquisition where the majority of transactions have been observed. The ex post MTM values are slightly under-predicting the purchase price in 2006 through 2008 and over-predicting the price in 2009, but the degree of such deviation is small.

On the other hand, the two contemporaneous benchmarks (origination AVM and the contemporaneous MTM predictions) are far away from the actual purchase price, which happens because of the information disadvantage. It is obvious that there is a delayed learning process: from 2003 to 2006, as the actual home prices rose quickly, while the median contemporaneous predictions were always under-shooting the transaction price, the gap is becoming smaller and smaller; from 2007 to 2009, the prediction turns to over-shoot the actual sales price, as a result of the delayed error-correction mechanism. Comparing the two contemporaneous predictions, because the origination AVM uses information other than home price index like dated tax assessments, so it seems to be more cautious in capturing the recent trend, hence it is even more conservative than the contemporaneous MTM benchmark. This comparison can be seen on Figure 1(b).

3 Mortgage Delinquency Comparison

3.1 Evidence From A Logistic Regression

One straightforward way to check the effect of overpayment on loan performance is to run a conventional default model, where we supplement the usual risk characteristics with the overpayment measure. It is expected that the sign will be positive, means a higher overpayment is correlated with an elevated level of loan default, but we also would like to see the magnitude in terms of final mortgage performance status.

[8]Ultimately given so much data coverage used in the model, the prediction of the house value has to be about correct.

3.1.1 Spread between Sales and AVM

Here we use the origination AVM as an example benchmark. For each loan, we define the spread as sales price divided by the AVM measure then minus one. We then classify the sample by the magnitude of this spread: as seen in Table 1, the spread can be as small as -15% (indicating the borrower has paid a price less than 85% of the benchmark valuation), to as large as 20% (the borrower's price is more than 20% over the benchmark valuation). Not surprisingly, we have positive coefficient estimate for positive spread, meaning if one pays more than the benchmark, then s/he will be more likely to default than otherwise, and vice versa for negative spread. We also have observed the monotonicity of the estimated coefficients, implying that the more one pays over (less than) the benchmark valuation, the more (less) likely s/he will incur default in mortgage payment later on. On the significance, for bigger absolute spreads, most of the coefficients are significant at 1% or 5% confidence level.

Table 1: Estimated Coefficients of Spread Between Sales and AVM

Spread	Loan Acquisition Year						
	2003	2004	2005	2006	2007	2008	2009
<-15%	-0.19^{**}	-0.36^{***}	-0.33^{***}	-0.24^{***}	-0.22^{***}	-0.23^{***}	-0.48^{***}
[-15%, -10%)	-0.10	-0.17^{**}	-0.16^{***}	-0.15^{***}	-0.07^{***}	-0.08^{*}	-0.40^{***}
[-10%, -5%)	-0.13	-0.12^{**}	-0.13^{***}	-0.03	-0.04	-0.04	-0.25^{*}
[-5%, -1%)	-0.05	-0.04	-0.05	0.01	-0.04^{*}	-0.05	-0.17
[-1%, 1%]	0.00	0.00	0.00	0.00	0.00	0.00	0.00
(1%, 4%]	0.03	-0.01	-0.01	0.08^{**}	-0.02	0.00	-0.16
(4%, 8%]	0.06	-0.02	0.14^{***}	0.12^{***}	0.07^{***}	0.06	-0.05
(8%, 12%]	0.15^{*}	0.06	0.20^{***}	0.14^{***}	0.13^{***}	0.15^{***}	0.12
(12%, 20%]	0.23^{***}	0.08	0.27^{***}	0.21^{***}	0.23^{***}	0.30^{***}	0.02
>20%	0.43^{***}	0.32^{***}	0.43^{***}	0.29^{***}	0.35^{***}	0.57^{***}	0.49^{***}

Note: *** represents significant at 1% confidence level, ** at 5%, and * at 10%.

In Table 2, we translate the estimated coefficients into the probability of default, for an otherwise average loan. To be explicit, we translate the average default rate into the estimated coefficient ($X\beta$), ask what if the overpayment spread instead is not zero but rather in any of the categories defined, and finally translate the coefficients back to the probability of default. This exercise shows the discriminatory power of the estimated coefficients: in 2007, while the average default rate is 19.10%, if one pays less than the benchmark by at least 15%, then his or her default rate will drop by like 300 basis point to 15.88%; on the other hand, if one pays more than the benchmark by 20%, the default rate skyrockets to 25%, an increase of 600 basis point. Consider the stressful environment during the crisis, these hundreds basis points of difference could easily mean whether a lender still had some liquidity and could survive the crisis or had to immediately file bankruptcy.

Table 2: Implied Default Probability (in %) of Spread Between Sales and AVM

Spread	Loan Acquisition Year						
	2003	2004	2005	2006	2007	2008	2009
<-15%	1.58	2.22	4.31	9.24	15.88	6.79	0.70
[-15%, -10%)	1.72	2.68	5.10	10.02	18.04	7.76	0.75
[-10%, -5%)	1.68	2.79	5.25	11.15	18.55	8.09	0.87
[-5%, -1%)	1.82	3.04	5.65	11.60	18.42	8.03	0.95
[-1%, 1%]	1.91	3.15	5.92	11.45	19.10	8.37	1.12
(1%, 4%]	1.97	3.12	5.87	12.28	18.74	8.40	0.95
(4%, 8%]	2.03	3.10	6.77	12.72	20.21	8.83	1.07
(8%, 12%]	2.21	3.34	7.13	13.00	21.24	9.55	1.26
(12%, 20%]	2.39	3.42	7.60	13.77	22.90	10.94	1.14
>20%	2.90	4.30	8.78	14.72	25.14	13.93	1.81

3.1.2 Spread between Sales and Appraisal

As a contrast, we also look at the spread between sales and appraisals: here the spread is defined the difference between sales price and appraisal, divided by the sales prices[9]. Not surprisingly, only a smaller minority of loans have appraisals less than sales, so we only focus on the other cases. However, even in this case, the spread is quite limited: the biggest category is that appraisal exceeding the sales by 6%. This is understandable: since the core objective of appraisals is for the lenders to be sure of the collateral value being at least the sales price, there is not much incentive for the appraisers to say more than that. Nevertheless, from the estimation it seems if appraisers say the borrowers are

[9] It makes little difference if the denominator is the appraisal value. Given that the majority of loans have appraisals above sales prices, if we use appraisal value as denominator, then the distribution of spread will be slightly narrower.

getting a really good deal, then the borrowers are less likely to default. In Table 4, the translated probability does show some discriminatory power: in 2007, while the average default rate is 19.10%, if one gets a deal of 6% less than the appraisal, the default rates drop by at least 200 basis point. However, the main problems with this spread definition are (a) per appraisal values, only a few percentage of borrowers are "overpaying", and (b) even for these over-payers, there is little difference in default rates as compared to these non-over-payers if any at all.

Table 3: Estimated Coefficients of Spread Between Sales and Appraisals

Spread	Loan Acquisition Year						
	2003	2004	2005	2006	2007	2008	2009
<-6%	-0.38^{***}	-0.41^{***}	-0.31^{***}	-0.17^{***}	-0.16^{***}	-0.21^{***}	-0.28^{***}
[-6%, -3%)	-0.27^{***}	-0.09^{**}	-0.14^{***}	-0.04	-0.09^{***}	-0.19^{***}	-0.16^{*}
[-3%, -2%)	-0.14^{**}	-0.18^{***}	-0.05	-0.04	-0.08^{***}	-0.24^{***}	-0.07
[-2%, -1%)	-0.19^{***}	-0.07	-0.06^{**}	-0.04^{*}	-0.13^{***}	-0.19^{***}	-0.24^{**}
[-1%, -0.5%)	-0.09^{*}	-0.07	-0.04	0.02	-0.17^{***}	-0.20^{***}	-0.12
[-0.5%, 0%)	-0.05	-0.04	0.03	0.00	-0.08^{***}	-0.17^{***}	-0.05
=0%	0.00	0.00	0.00	0.00	0.00	0.00	0.00
>0%	0.03	0.01	0.12	-0.04	0.00	-0.08	-0.03

Note: *** represents significant at 1% confidence level, ** at 5%, and * at 10%.

Table 4: Implied Default Probability (in %) of Spread Between Sales and Appraisals

Spread	Loan Acquisition Year						
	2003	2004	2005	2006	2007	2008	2009
<-6%	1.32	2.11	4.41	9.81	16.72	6.91	0.85
[-6%, -3%)	1.46	2.88	5.17	11.06	17.69	7.01	0.96
[-3%, -2%)	1.67	2.65	5.62	11.02	17.85	6.73	1.05
[-2%, -1%)	1.58	2.95	5.57	11.01	17.23	7.01	0.88
[-1%, -0.5%)	1.74	2.94	5.72	11.63	16.60	6.93	1.00
[-0.5%, 0%)	1.82	3.03	6.11	11.44	17.94	7.19	1.07
=0%	1.91	3.15	5.92	11.45	19.10	8.37	1.12
>0%	1.97	3.17	6.63	11.09	19.09	7.77	1.08

3.1.3 Both Spreads In the Model Together

We also conduct another exercise by putting both spreads in the model together, and see if any change in the discriminatory power of the spread categories. Comparing Table 5 with Table 2, we can not see any major change; on the other hand, comparing Table 6 with Table 4, we notice the difference in predicted probability between the zero spread and these who get a deal of at least 6% is much smaller now we have both spreads in the same model. In 2007, the gap in default rates is 238 basis points in Table 4, and now becomes only 53 basis points in Table 6. This is due to that now the estimated coefficients on the spread between sales prices and appraisal are much smaller in absolute value.

This exercise shows that appraisal value as compared to sales price, while standing alone, can provide some hint on future mortgage performance, its predictive power will be overshadowed much by the AVM benchmark we use. It is possible that because the distribution of the spread between sales and AVM is much wider, such spread can provide more insights on the probability of whether the loan is going to have payment problem down the road.

Table 5: Implied Default Probability (in %) of Spread Between Sales and AVM (in a model with both spreads)

Spread	Loan Acquisition Year						
	2003	2004	2005	2006	2007	2008	2009
<-15%	1.69	2.41	4.51	9.34	15.74	6.67	0.70
[-15%, -10%)	1.79	2.78	5.21	10.62	17.94	7.70	0.75
[-10%, -5%)	1.72	2.84	5.30	11.17	18.50	8.07	0.87
[-5%, -1%)	1.84	3.06	5.67	11.60	18.41	8.02	0.95
[-1%, 1%]	1.91	3.15	5.92	11.45	19.10	8.37	1.12
(1%, 4%]	1.96	3.12	5.86	12.28	18.73	8.38	0.95
(4%, 8%]	2.01	3.09	6.74	12.72	20.19	8.79	1.07
(8%, 12%]	2.18	3.33	7.08	13.00	21.20	9.49	1.26
(12%, 20%]	2.35	3.40	7.54	13.78	22.85	10.84	1.14
>20%	2.86	4.29	8.73	14.74	25.04	13.80	1.81

Table 6: Implied Default Probability (in %) of Spread Between Sales and Appraisals (in a model with both spreads)

Spread	Loan Acquisition Year						
	2003	2004	2005	2006	2007	2008	2009
<-6%	1.45	2.36	5.18	11.02	18.57	8.00	1.03
[-6%, -3%)	1.54	3.00	5.50	11.63	18.54	7.56	1.06
[-3%, -2%)	1.73	2.72	5.85	11.36	18.43	7.09	1.09
[-2%, -1%)	1.61	2.99	5.70	11.18	17.58	7.21	0.90
[-1%, -0.5%)	1.76	2.95	5.76	11.70	16.76	7.00	0.99
[-0.5%, 0%)	1.82	3.01	6.03	11.36	17.88	7.14	1.03
=0%	1.91	3.15	5.92	11.45	19.10	8.37	1.12
>0%	1.91	3.09	6.30	10.64	18.41	7.30	0.98

3.2 Raw Sample Comparison

While regression as in Section 3.1 tell mechanically that "overpayment", i.e. the spread between sales and benchmark value, is correlated with mortgage delinquency, it is still kind of abstract in that it is not intuitive to translate the coefficient from that estimation into actual delinquency difference. Therefore, to make our point more transparent, we would like to do some propensity score matching, to show in actual loan samples, such spread is an helpful metric. As an illustration, we now focus on a subsample where the LTV equals 80: this is a rather conservative choice[10], since by putting 20% down, these loans are in no way connected to the subprime underwriting which is prevalent before the crisis[11]. The borrowers usually are in a better position in terms of credit score, regular cash flow, debt management, etc. Hence we will go directly to the sample summaries. To make the logic flow easier, we will first look at the origination AVM as a benchmark.

Specifically for each year's LTV=80 purchase money mortgages, we can generate a distribution of overpayment for that year, according to which we can divide the loan sample into ten deciles: for example, the first decile contains these loans whose overpayment is in

[10]It should be noted that all the reasoning about overpayment applies for any loan. If the reader is convinced by the argument made on the 80 LTV sample, then the effect of overpayment will be even stronger for loans underwritten under the loose underwriting environment, like the subprime products or high LTV loans.

[11]Calem, Lambie-Hanson, and Nakamura (2017) points that at the loan-to-value boundaries where mortgage insurance premium matters, like LTV=80, appraisers are more likely to inflate the value and lower down the LTV.

the 0-10% percentile of that distribution, i.e. these house transactions who are seriously underpaid, or the least overpaid, per our benchmark used.

First we can do a raw comparison of the default rates across these overpayment deciles. Table 7 is the loan counts per decile for each acquisition year: it is obvious that we divide the sample into ten equal sub-samples. Table 8 is the median overpayment by decile: note that the overall median for each year is declining from 5.5% in 2003 to roughly 0 in 2007, and then -2% in 2008. Overall the spread between the first and the tenth deciles is increasing: in 2008 and 2009, these who appear to underpay the house can do so as much as 30%, i.e. getting a 30% discount compared to the benchmark value.

Table 7: Loan Counts By Overpayment Decile: Raw Sample

Overpayment Deciles	Loan Acquisition Year						
	2003	2004	2005	2006	2007	2008	2009
1st (00%-10%)	8214	8368	9686	9615	8560	7053	6820
2nd (10%-20%)	8214	8368	9686	9615	8560	7052	6819
3rd (20%-30%)	8213	8368	9686	9614	8560	7052	6820
4th (30%-40%)	8214	8368	9686	9615	8559	7053	6819
5th (40%-50%)	8213	8367	9686	9614	8560	7052	6820
6th (50%-60%)	8214	8368	9686	9615	8560	7052	6819
7th (60%-70%)	8214	8368	9686	9615	8559	7053	6820
8th (70%-80%)	8213	8368	9686	9614	8560	7052	6819
9th (80%-90%)	8214	8368	9686	9615	8560	7052	6820
10th (90%-100%)	8213	8367	9685	9614	8559	7052	6819

Table 9 displays the average default rates in basis points, where we use a relatively serious default measure which is 6 months delinquent within 5 years of acquisition[12]. We prefer this measure because for this kind of serious delinquency, it is unlikely that the loan can be cured by itself and thus such delinquency will more likely move to the default stage and have a consequential lingering effect on the borrowers' future financial life like credit reports, next housing choice, etc. Consistent with the overall housing cycles, there is a clear inverse-U shape in default rates; but within a given year, the general trend is that

[12]The distinct between loan application, origination and acquisition is not that important. With GSEs' market share, majority of loans will be delivered to one of the two national mortgage guarantors within a couple of months after origination.

Table 8: Median Overpayment in Percentages By Overpayment Decile: Raw Sample

Overpayment Deciles	Loan Acquisition Year						
	2003	2004	2005	2006	2007	2008	2009
00%-10%	-15	-16	-16	-17	-20	-30	-30
10%-20%	-6	-8	-8	-10	-12	-18	-16
20%-30%	-2	-4	-4	-6	-8	-12	-10
30%-40%	1	-1	-1	-3	-5	-7	-5
40%-50%	4	2	1	-1	-2	-4	-1
50%-60%	7	4	4	2	1	0	2
60%-70%	11	7	7	4	3	3	5
70%-80%	15	11	10	8	6	6	9
80%-90%	21	17	16	12	11	11	15
90%-100%	34	31	28	23	22	23	26
40%-60%	3.5	3	2.5	0.5	-0.5	-2	0.5

as people overpay more relative to the benchmark, they do default more; the contrast is more obvious between the lowest and the highest deciles.

Table 9: Average Default Rate in Basis Point By Overpayment Decile: Raw Sample

Overpayment Deciles	Loan Acquisition Year						
	2003	2004	2005	2006	2007	2008	2009
00%-10%	84	116	291	703	940	462	104
10%-20%	62	127	310	752	999	450	110
20%-30%	63	125	318	698	910	444	107
30%-40%	54	131	369	751	880	417	91
40%-50%	50	121	401	701	904	384	117
50%-60%	52	129	402	758	806	357	94
60%-70%	60	134	465	759	845	377	70
70%-80%	51	143	518	796	864	366	126
80%-90%	52	140	559	846	894	492	109
90%-100%	85	184	691	866	1000	746	185

It is critical to point that when we categorize the loans according to overpayment measures, we completely ignore other risk attributes that could also affect the defaults. Table 10 and 11 show the average and median predicted default risk, where such prediction is generated from a conventional mortgage default model using all available risk attributes like credit score except the overpayment measure. As is evident in both the average and median statistics[13], these people who appear to underpay the property actually do have loans that are more risky, which is possible if their FICO is lower or the DTI is higher, even though we have already controlled their LTV at 80. On the other hand, these people who appear to overpay the property do have loans that are less risky. So this means if we control for the inherent risk in the loan, the contrast between underpaying and overpaying borrowers in Table 9 should be even sharper.

Table 10: Average ***Predicted*** Default Rate in Basis Point: Raw Sample

Overpayment Deciles	Acquisition Year						
	2003	2004	2005	2006	2007	2008	2009
00%-10%	89	157	406	803	1072	579	156
10%-20%	74	143	406	777	1013	562	130
20%-30%	68	141	409	770	974	490	119
30%-40%	61	131	419	766	929	448	108
40%-50%	64	130	417	761	890	402	106
50%-60%	57	132	420	756	869	401	96
60%-70%	54	125	443	753	860	380	96
70%-80%	52	128	455	755	827	383	97
80%-90%	51	124	476	754	843	423	96
90%-100%	55	130	487	742	846	454	102

3.3 Matched Sample Comparison

As is seen above, a raw comparison of actual mortgage defaults across overpayment deciles is not telling the full story because we do not take into account other risk attributes that are also associated with the default probability. Now we would like to do some controlled experiments, through the propensity score matching. The idea is for each acquisition year, to vigorously select loans from each decile that have similar risk, while throwing away these loans that do not have a counterpart in any of the other deciles; at the end of the matching, each decile of the sample is guaranteed to not only have the same number of loans, but also the similar distribution of predicted mortgage risk, where such risk is predicted by

[13] There is huge difference between the average and the median in predicted risk. This is because for most loans their default risk is low, so the median is as much as half size as the average predicted risk.

Table 11: Median ***Predicted*** Default Rate in Basis Point: Raw Sample

Overpayment Deciles	Acquisition Year						
	2003	2004	2005	2006	2007	2008	2009
00%-10%	40	84	257	506	582	302	67
10%-20%	35	80	264	494	579	280	58
20%-30%	32	78	264	479	531	232	53
30%-40%	29	76	269	482	522	208	47
40%-50%	29	73	261	478	488	186	45
50%-60%	26	73	271	466	487	175	43
60%-70%	24	69	278	473	476	175	44
70%-80%	23	70	282	480	461	175	42
80%-90%	23	69	298	481	470	178	42
90%-100%	24	71	308	467	455	188	41

a standard default model using all risk attributes like LTV and FICO, but excluding the overpayment measure.

After the matching exercise, each year we are capturing roughly about 80% to 90% of the original loans. This percentage is affected by how homogeneous the raw sample is distributed with regard to the predicted risk. In year 2005 and 2006, as seen in Table 10 and 11, the predicted risk is rather uniform so we have not thrown away many records; in 2007 the deciles are pretty different with average risk being 1000 basis point for one decile and 850 for another one, so we have to drop a bit more loans for comparison purposes. But overall, 80% of raw samples are being retained in the matched sample.

Now on the median overpayment measure in Table 13, unsurprisingly few median overpayment measure are changing as compared to that in Table 8. This confirms that while we throw out some loans at the extreme ends, the distribution of overpayment does not change if any at all.

In the matched sample, the effect of overpayment is more transparent in Table 14. As we move from the 1st decile to the 10th decile, we can see a clear steady increase in the default rates: the most over-paying decile defaults materially more often than the decile where borrowers get good deals. And the contrast is more evident than Table 9 simply because during the matching, we throw away a lot of high-risk loans from these deciles that do get a good deal.

Table 12: Loan Counts By Overpayment Decile: Matched Sample

Overpayment Deciles	Acquisition Year						
	2003	2004	2005	2006	2007	2008	2009
00%-10%	6459	7364	8673	8880	7514	5566	5554
10%-20%	6459	7364	8673	8880	7514	5565	5554
20%-30%	6459	7364	8673	8880	7514	5565	5554
30%-40%	6459	7364	8673	8880	7514	5565	5554
40%-50%	6459	7364	8673	8880	7514	5565	5554
50%-60%	6459	7364	8673	8880	7514	5565	5554
60%-70%	6459	7364	8673	8880	7514	5567	5554
70%-80%	6459	7364	8673	8880	7514	5565	5554
80%-90%	6459	7364	8673	8880	7514	5565	5554
90%-100%	6459	7364	8673	8880	7514	5565	5554
% of sample	79	88	90	92	88	79	81

Table 13: Median Overpayment in Percentages By Overpayment Decile: Matched Sample

Overpayment Deciles	Acquisition Year						
	2003	2004	2005	2006	2007	2008	2009
00%-10%	-15	-16	-16	-17	-20	-30	-29
10%-20%	-6	-8	-8	-10	-12	-18	-16
20%-30%	-2	-4	-4	-6	-8	-12	-10
30%-40%	1	-1	-1	-3	-5	-7	-5
40%-50%	4	2	1	-1	-2	-4	-1
50%-60%	7	4	4	2	1	0	2
60%-70%	11	7	7	4	3	3	5
70%-80%	15	11	10	8	6	6	9
80%-90%	21	17	16	12	11	11	15
90%-100%	34	31	28	23	22	23	26

Table 14: Average Default Rate in Basis Point By Overpayment Decile: Matched Sample

Overpayment Deciles	Acquisition Year						
	2003	2004	2005	2006	2007	2008	2009
00%-10%	56	88	291	631	745	336	67
10%-20%	28	106	310	684	838	322	79
20%-30%	46	118	308	662	760	381	85
30%-40%	43	120	352	699	761	374	79
40%-50%	33	117	384	657	825	376	88
50%-60%	48	106	386	707	748	368	92
60%-70%	60	140	415	713	801	401	76
70%-80%	51	128	459	743	860	403	130
80%-90%	59	139	489	798	872	456	110
90%-100%	87	166	597	836	977	654	175

Together with Table 12, Table 15 and 16 confirm our expectation that the matched sample has both equal number of loans in each decile, and more importantly, its distribution of predicted risk is very similar.

Table 15: Average ***Predicted*** Default Rate in Basis Point: Matched Sample

Overpayment Deciles	Acquisition Year						
	2003	2004	2005	2006	2007	2008	2009
00%-10%	52	117	398	693	807	386	90
10%-20%	51	116	403	702	804	385	89
20%-30%	50	117	399	709	806	379	88
30%-40%	50	119	397	705	801	381	90
40%-50%	50	118	399	706	804	392	92
50%-60%	51	119	396	712	813	391	94
60%-70%	54	123	391	705	814	400	95
70%-80%	53	122	392	703	831	405	95
80%-90%	54	121	390	715	821	386	95
90%-100%	52	120	389	707	818	385	95

Figure 2(a) displays the comparison between different deciles and between raw and matched samples. The red thick line represents these loans that the borrowers have overpaid the most; the yellow represents those in the middle, i.e. the 5th and 6th deciles; and the green dotted represents the first decile, i.e. who overpaid least or who got a good deal.

Table 16: Median ***Predicted*** Default Rate in Basis Point: Matched Sample

Overpayment Deciles	Acquisition Year						
	2003	2004	2005	2006	2007	2008	2009
00%-10%	30	76	278	479	507	223	51
10%-20%	30	76	276	479	514	221	51
20%-30%	30	76	277	479	508	222	51
30%-40%	30	76	278	479	510	221	51
40%-50%	29	76	277	478	509	220	51
50%-60%	30	76	278	480	510	223	51
60%-70%	29	76	276	479	514	220	52
70%-80%	29	76	277	480	509	218	51
80%-90%	30	76	278	479	508	219	51
90%-100%	29	76	279	479	516	226	52

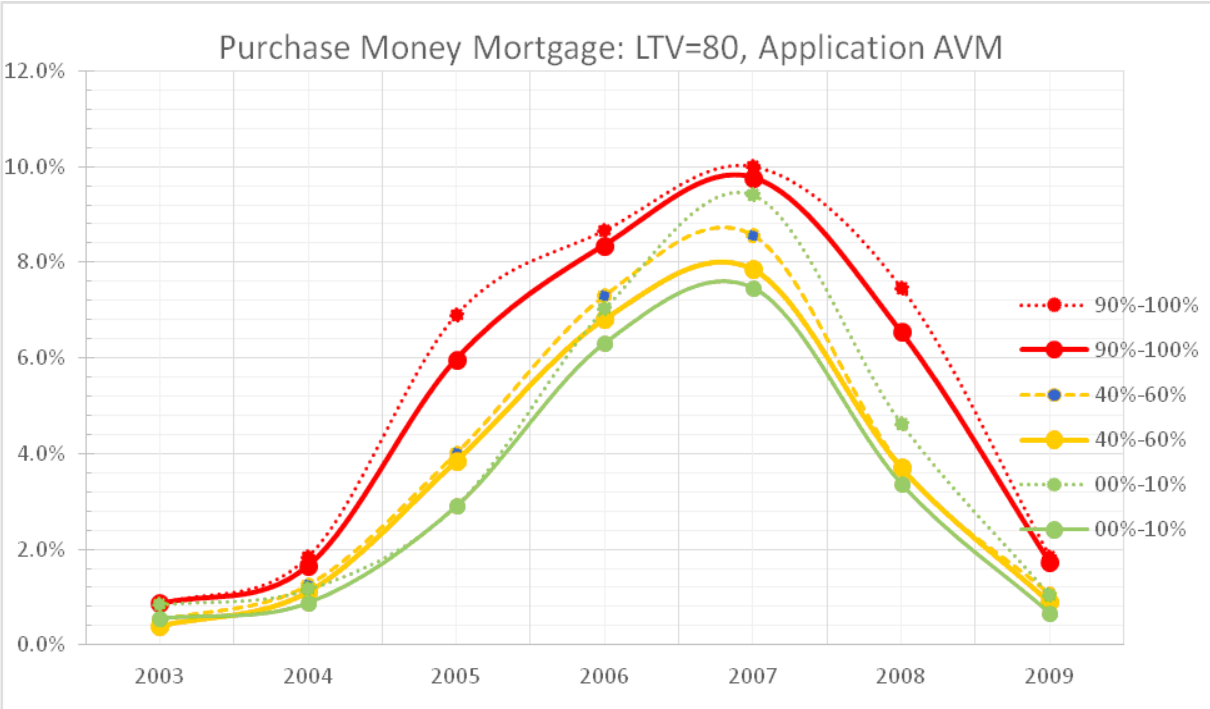
The dashed lines represent the statistics from the raw sample while the solid lines represent that from the matched sample. Not surprisingly, these who overpaid always default more, and by a large magnitude. However in the raw sample, the monotonicity between the middle deciles and the first decile is not always observed: sometimes the yellow line is below the green line like in 2007, as we observed from Table 9; after the match, the contrast is very sharp: the less you overpay, or the more you underpay, the smaller the probability of experiencing a serious default.

3.4 Using Different Benchmarks

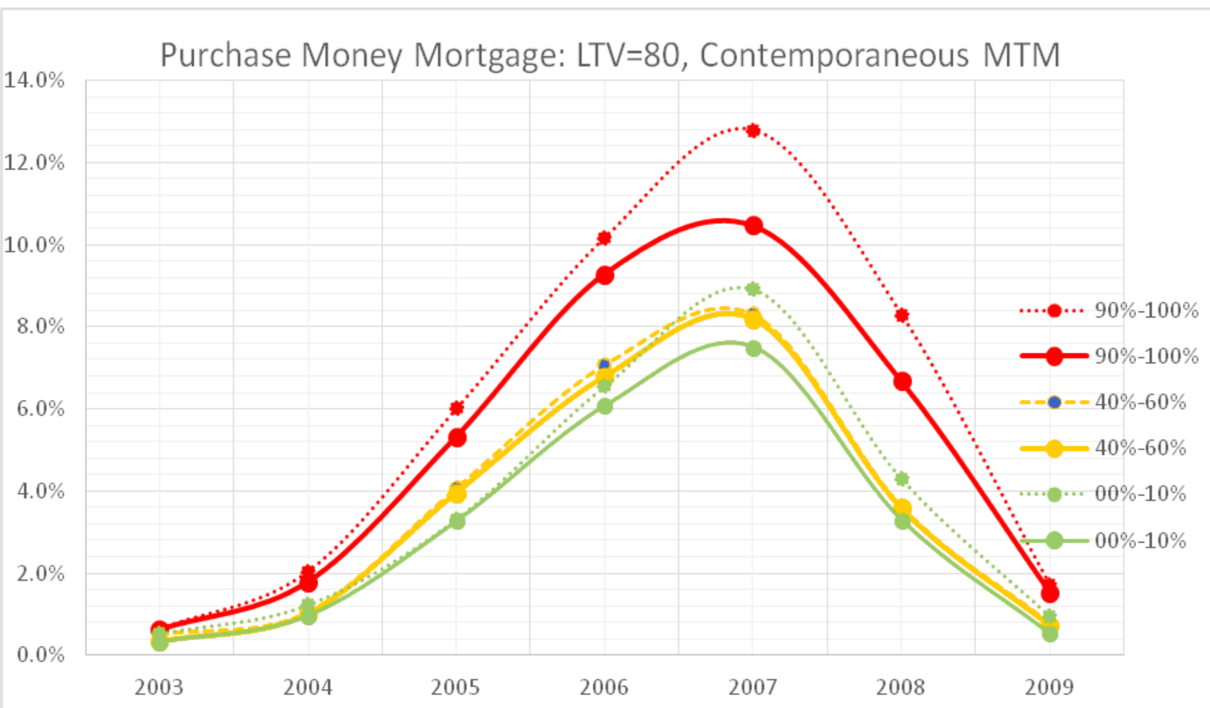
Here we look at other three potentially useful benchmarks and do similar matching exercises: while they differ in terms of data used, real-time availability or coverage, each of the benchmark yields a consistent message: overpaying for a collateral is closely related to later a larger chance of default; and vice versa.

3.5 Expanding beyond 80 LTVs

It is worthwhile to emphasize that the sample we use to illustrate our point consists of only 80 LTV loans which are safe loans as judged by even the strictest underwriting regimes. Using each of the four benchmarks, for loans acquired in 2007, the difference between those

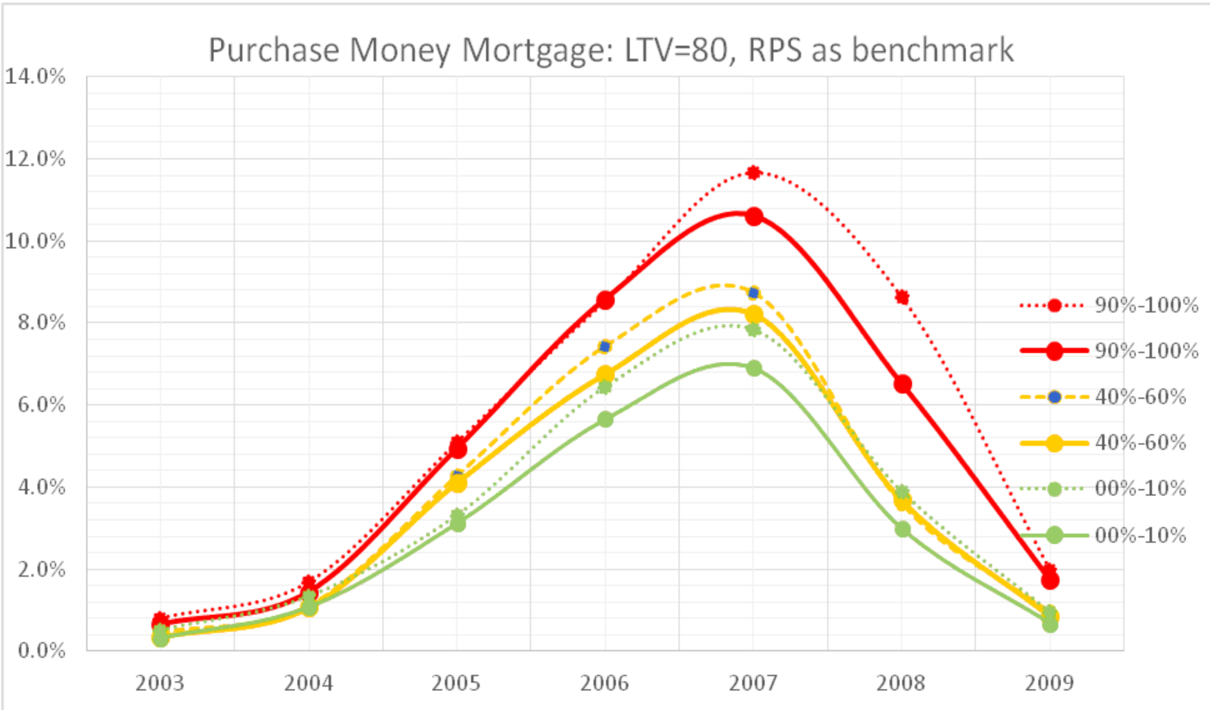


(a) Origination AVM as benchmark)

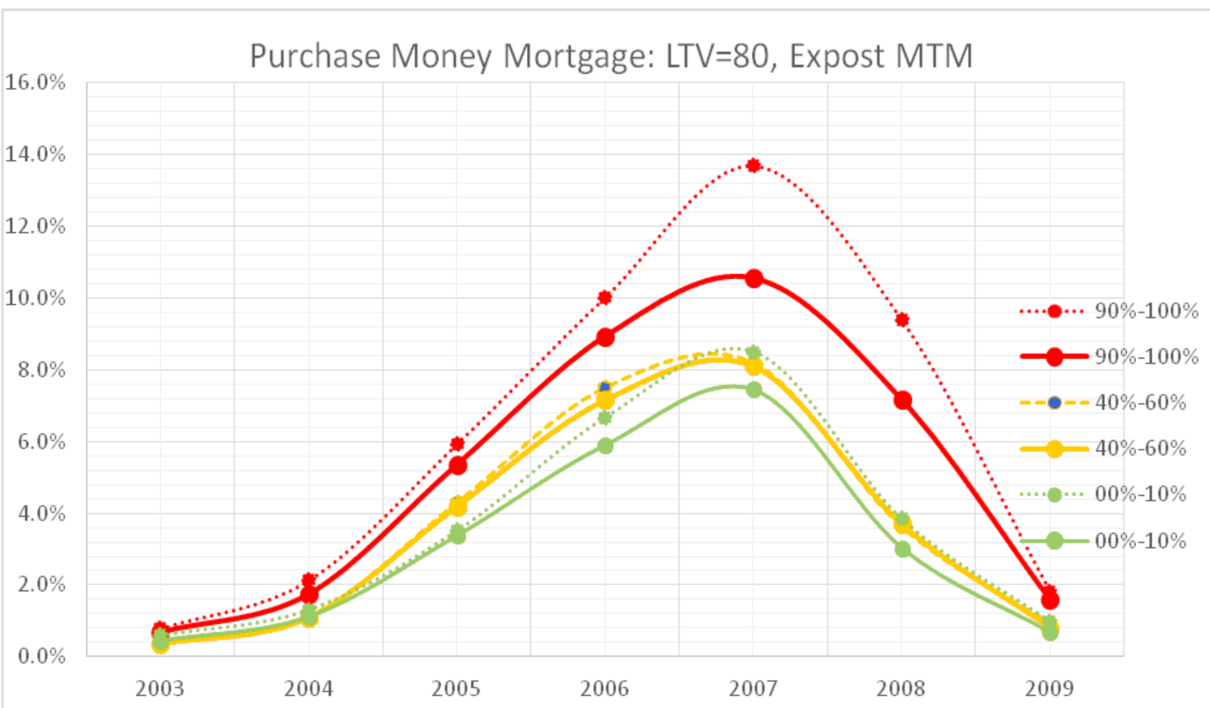


(b) Contemporaneous MTM as benchmark

Figure 2: Performance Comparison Between Deciles: Raw and Matched Samples



(a) Acquisition AVM as benchmark



(b) Ex post MTM as benchmark

Figure 3: Performance Comparison Between Deciles: Raw and Matched Samples

who overpay the most and those who overpay the least, is in the range of 2 percentage points, which is pretty large for such prime mortgages with 20% down payment. However, if we expand to the whole sample including other LTVs[14], then the effect of overpayment on future serious defaults is much larger. This is displayed below in Figure 4 for all the conventional purchase money mortgages between 2003 and 2009: for loans acquired in 2007, the fact that the default rates is 17.6% among these who overpay the least is surprising enough; but on top of that, the default rates rises to 23.9% among these who overpay the most, is even more intriguing, a gap as big as 6.3%.

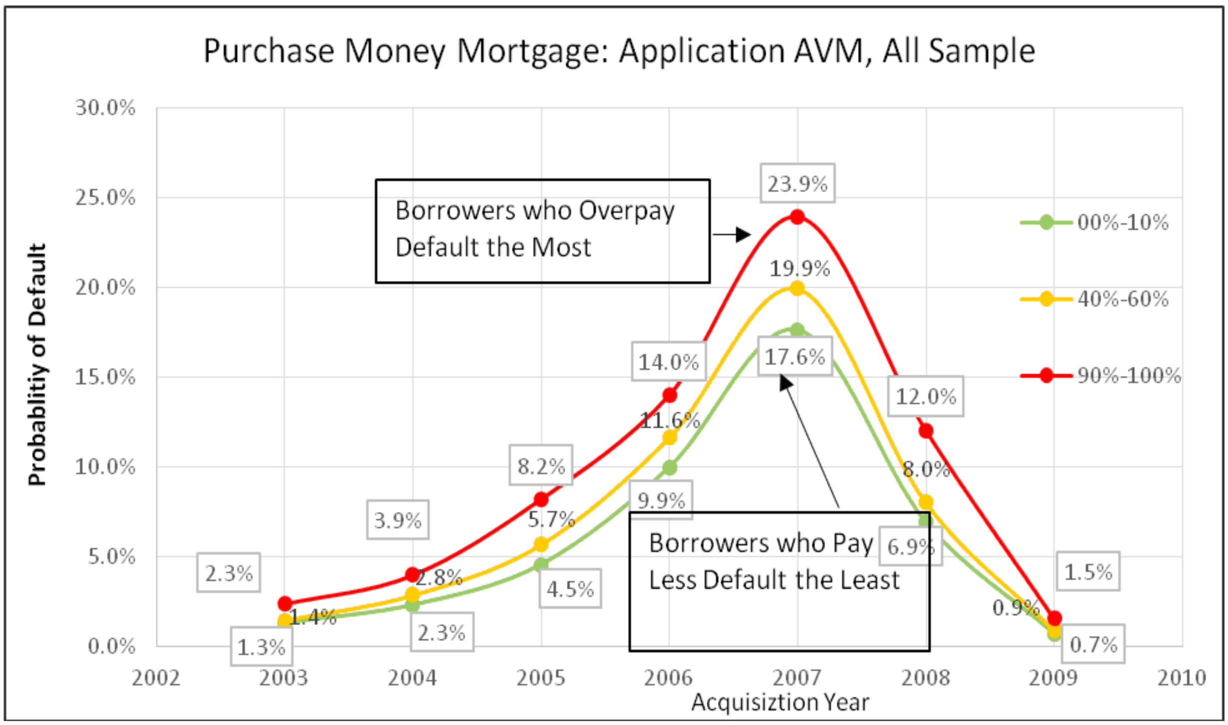


Figure 4: Performance Comparison Between Deciles: Raw and Matched Samples (Application AVM as benchmark; All LTV loans)

[14]Even these loans in the expanded samples are pretty standard: they are conventional and conforming loans. In other words, the subprime mortgages are not part of our sample.

4 What If Loans Do Not Default

In the presence of a secondary national mortgage guarantor, when a loan defaults, the majority of loss has to be absorbed by the guarantor, though the borrowers also lose the equity as accumulated through upfront downpayment and principal payment over the years. Little attention has been paid to loans that are performing well, or loans that are already fully paid due to refinancing or the house being sold. After all in the latter case, neither the banks nor the guarantor will suffer a penny of loss. However as we will see soon, the borrowers still could. The intuition is not very far from a stock market analogy: if you buy a stock when its price is high, your portfolio may not necessarily have a negative return provided you stay in the market long enough, but still you are worse-off as compared to a peer who bought another stock when its price was at its low points. This is not hard in theory, but not many studies have any empirical evidence to prove this.

For some loans we can find their next arms-length transaction from the public records, and we can compare the realized profit to the borrowers (as defined as prices of next sales divided by that of this sales and then minus one) among those who overpay and underpay in the first place. Obviously, not every loan's collateral has been sold in an arms-length way yet[15]. Based on the whole sample, at least half have not been sold among these loans acquired between 2003 and 2009: the success rate for loans acquired in 2003 is about 50%, and smaller for later years.

We will illustrate our point using LTV=80 subsample again. Here the effect we focus is the realized future profitability, and for that purpose, we will proxy the expected profitability using the HPI appreciation in the zip level[16]. So we can do a similar matching exercise: pull all loans with LTV=80 and that we have found the next arms-length transaction, divide them into ten overpayment deciles, and then select loans from each decile where the HPI appreciation is matched across deciles. We can also limit our attention to these loans that did not experience a six-month delinquency within fives years of acquisition. Figure 5 shows the comparison.

[15]Some borrowers are still living in that house and the loan is either performing or just get refinanced, while some have defaulted and the records of sales we found are for REOs or short-sales. Of course, for cases where we find multiple arms-length sales, we will pick up the next immediate one. We accessed the public record database in December 2016.

[16]This avoids the necessity of finding a model that explains the expected house appreciation.

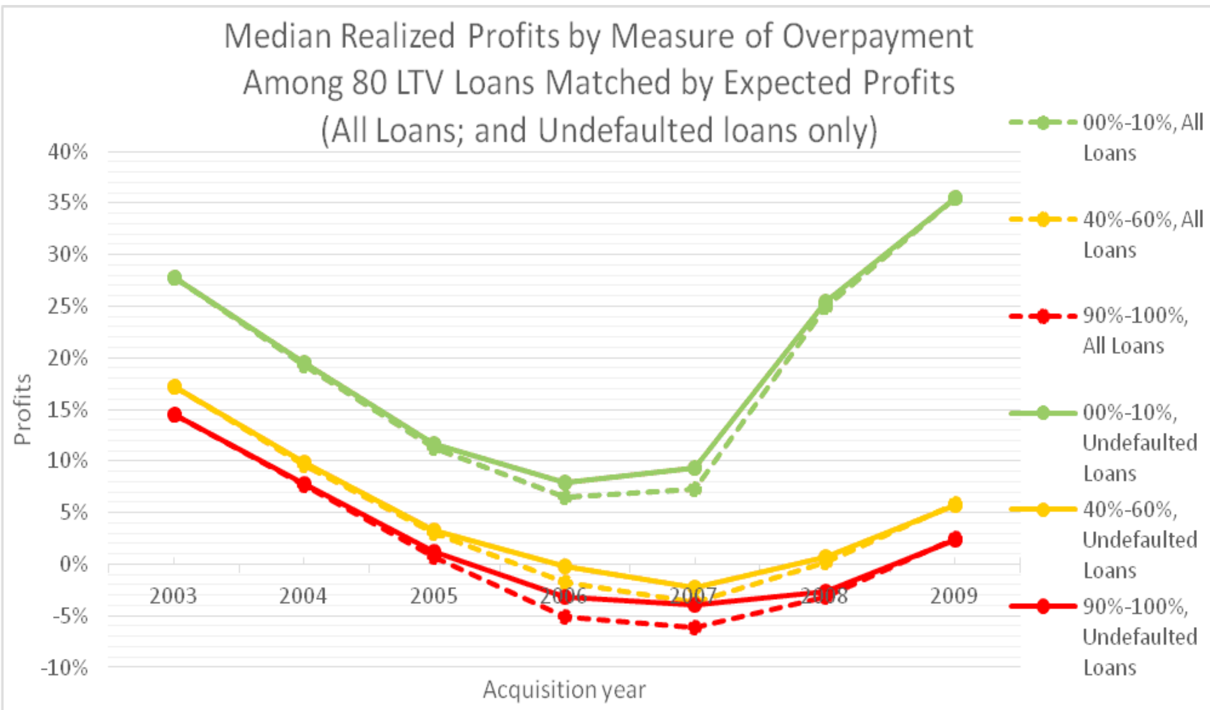


Figure 5: Profitability Comparison Between Deciles: Post-Matching (Application AVM as benchmark)

This plot shows that for loans acquired within a given year, these who overpay the most (least) at the beginning are reaping the least (most) profit from home ownership: if you over-pay in the first place and fortunately the loan survives until the next time you sell the house, then compared to your cohorts who did not overpay, your realized profit is much smaller. For example, for loans acquired in 2006 and 2007 which is at the peak of housing price, the return is near zero or slightly negative if the borrowers overpay at that time; but could be nearly 10% if they did not do so in the beginning. On the other hand, in 2009 immediately after the crash, if you overpay the most, the realized profit is likely less than 5%, while if you overpay the least or underpay the most, the profits can be as high as 35%. Of course if we include the defaulted loans in the sample, overall profits will drop a little bit; but the monotonicity effect of over- or under-paying on future capital gains is still very much pronounced.

We would like to point out that, the above analysis only covers these houses that eventually got sold. Whether overpaying or not will probably have effects on the reservation prices of the owners, and thus how likely one decides to put the house on the market, how long will it stay on the market, and how often you need adjust down the listing price, controlling for other variables affecting these decisions. While these issues require data like listing history that reveals the dynamics of seller decisions, at this moment, we are sure about the seller outcomes: once the house gets sold, an overpaying borrower (the red line) is unlikely to realize a bigger profit than his/her under-pay neighbors (the green line). Considering how often one will move house from house during one's life time, a primary residence's role as a foundation for building wealth could be compromised, unless the borrowers don't overpay in the first place.

5 Conclusion

There are vast studies documenting how appraisers' incentives are affected by lenders' pressures, and how that contributes to the inflating housing bubbles. And as an alternative to appraisers' reports, these model-based valuations are frequently utilized by lenders and GSE for quality control. Arguably no one would like to deny that these valuations are free from any human intervention, and thus are less subjective.

On the other hand, when it comes to their particular house buying, even the buyers most of the time would not like to see their hard negotiated deal to be killed by a low-ball valuation, especially when such valuation is from a computer where no one (real estate agent, buyers, appraisers or lenders) can intervene. Hence this creates a dilemma for home buyers: they want an objective and fair opinion, and yet they also fear of negotiation or losing the negotiated contract and thus would like some one deemed to be professional and objective, to safeguard their interest, at least on the surface. It is this contradiction that leads to the fact that AVM is a reference tool at the back end of loan quality control rather than the forefront of buyer negotiation, even today.

Comparison made in this paper reveals that AVMs or a simple Mark-to-market valuation, can be a useful tool for the home buyers as well. If they overpay in the first place, they are at a bigger risk of becoming serious delinquent; even if they do not run into any mortgage trouble, compared to their peers, next time they sell this house, they will get less profit from the transaction. In theory these are plain simple intuitive arguments, but it is more stronger once we see their empirical support.

Theses stunning comparisons should be viewed by both the policy makers and the mortgage industry as an evidence for the need for more work on evaluating the collateral value, above and beyond the appraisal regulations. It is true that AVMs suffer from their methodological imprecisions, but even within a single AVM for predictions where we have strong confidence, we can still separate these who overpay from these who underpay. The reliance on model-based prediction, could at least act as a secondary check for prudential lending, which is importantly for the well-functioning of the overall mortgage ecosystem. And more importantly in the long run, such sanity check will help the consumers' lifelong financial health.

References

CALEM, PAUL, LAUREN LAMBIE-HANSON, AND LEONARD NAKAMURA (2017): "Appraising Home Purchase Appraisals," *Federal Reserve Bank of Philadelphia Working Paper*, Link here.

ERIKSEN, MICHAEL, HAMILTON FOUT, MARK PALIM, AND ERIC ROSENBLATT (2016): "Contract Price Confirmation Bias: Evidence From Repeat Appraisals," *Fannie Mae Working Paper*, Link here.

SHUI, JESSICA, AND SHRIYA MURTHY (2018): "Are Appraisal Management Companies Value-Adding: Stylized Facts from AMC and Non-AMC Appraisals," *FHFA Working Paper*, Link here.