Internet Appendix For "Collateral Misreporting in the RMBS Market"

A. New Century-ABSNet matching description

We merge funded first-lien loans associated with single-unit properties in New Century data with those loans in ABSNet whose originator is either New Century Mortgage Corporation or its subsidiary, Home123 Corporation. We keep loans for which the lien position or the number of units in the underlying property are missing. This results in initial samples of 952,289 loans in the New Century data and 577,899 loans in ABSNet. We first match the loans based on their zip code, first payment date, interest rate type (fixed- or adjustable-rate mortgage), and purpose of transaction (purchase or refinance). Second, we require the New Century's status date to be within 30 days from the loan origination date in ABSNet, and loan amounts and credit scores to be within a $1,000 and 10 points, respectively. Third, we only consider the remaining loan pairs a match when it is unique. This procedure results in 363,623 unique matches, which represents 38.2% on the initial New Century data sample. Restricting the sample based on the criteria described in Section 1.1 results in a sample of 70,325 matched loans, which are described in Table IA.**2**.

To confirm the accuracy of our matching procedure, we repeat the matching exercise with all loans in ABSNet regardless of their originator. Using this methodology, we match 468,676 pairs of loans. Of the 363,623 pairs that we obtained through the original matching, 363,434 (99.95%) coincide with those obtained through the less restrictive matching procedure, which provides reassurance about the accuracy of the database merge.

B. Pool selection and pool data calculation description

The unit of observation for our RMBS analysis is the RMBS deal pool, which is a pool of loans that support a specific set of securities within a RMBS deal. For deals with a "Y" structure, we conduct our analysis at the more general loan pool level corresponding to the subordinated securities. From the ABSNet loan data, we calculate pool-level average appraisal difference, percent of refinance loans with even LTV, and control variables, including average FICO score, average CLTV ratio, percentage of loans with low or no documentation, and percentage of loans that are refinance. Like Piskorski, Seru, and Witkin (2015), we restrict our sample to loan pools with at least 25% of loans in our loan sample. In addition, we only consider loan pools for which at least 95% of the underlying loans have both FICO score and CLTV ratio information. Our regressions also control for deal year and fixed effects for the top six underwriters in the sample. The remaining underwriters, which jointly correspond to 247 pools, are grouped together.

We use ABSNet pool and security data to calculate pool-level losses and pricing. Losses are pool-level cumulative realized losses as of September 2014 as a percent of the pool's original balance. Yield spreads are average floating rate interest margins across all of the securities supported by the pool. Because this data is limited to floating rate securities, we limit our analysis to pools in which at least 90% of pool security value comes from floating rate securities with available interest rate margin data. AAA subordination is the fraction of the security balance in the pool that is subordinated to the AAA securities. We calculate this as the minimum subordination of any AAA security in the pool. Security-level credit ratings come from Standard & Poor's, supplemented by Moody's. Because we need credit ratings for this calculation, we limit our analysis to pools in which we have credit ratings for at least 90% of the security value in the pool.

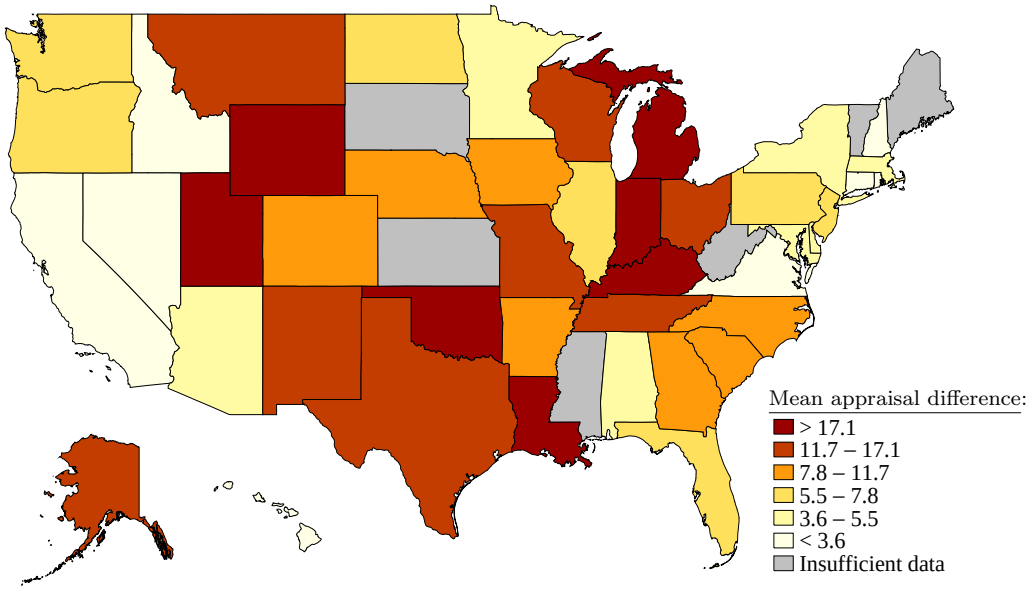
As a control variable, we also collect pool-level overcollateralization. Overcollateralization is based on the difference between subordination and total credit support. We also compute overcollateralization based on reported overcollateralization tranches with similar results.

To eliminate outliers and potential errors in the data, we drop pools with losses, yield spreads, or AAA subordination above the 95^{th} percentile and require pools to have data on all three outcome variables. This results in a sample of 694 loan pools, which come from 681 deals and contain 2.6 million underlying loans.

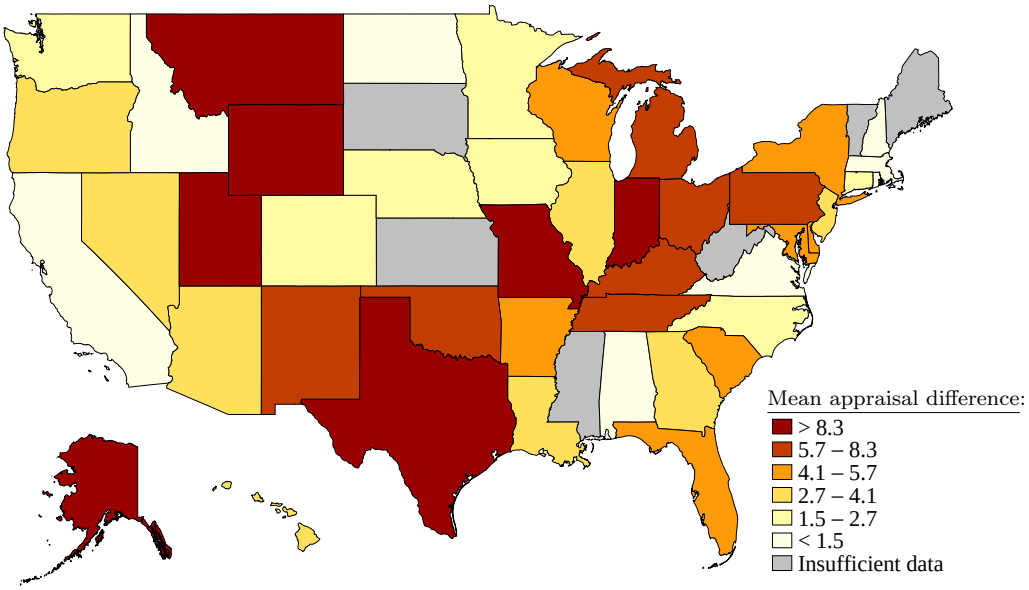
C. Supplemental figures and tables

Figure IA.1. Geographic distribution of appraisal differences

Panel A: Refinance loans

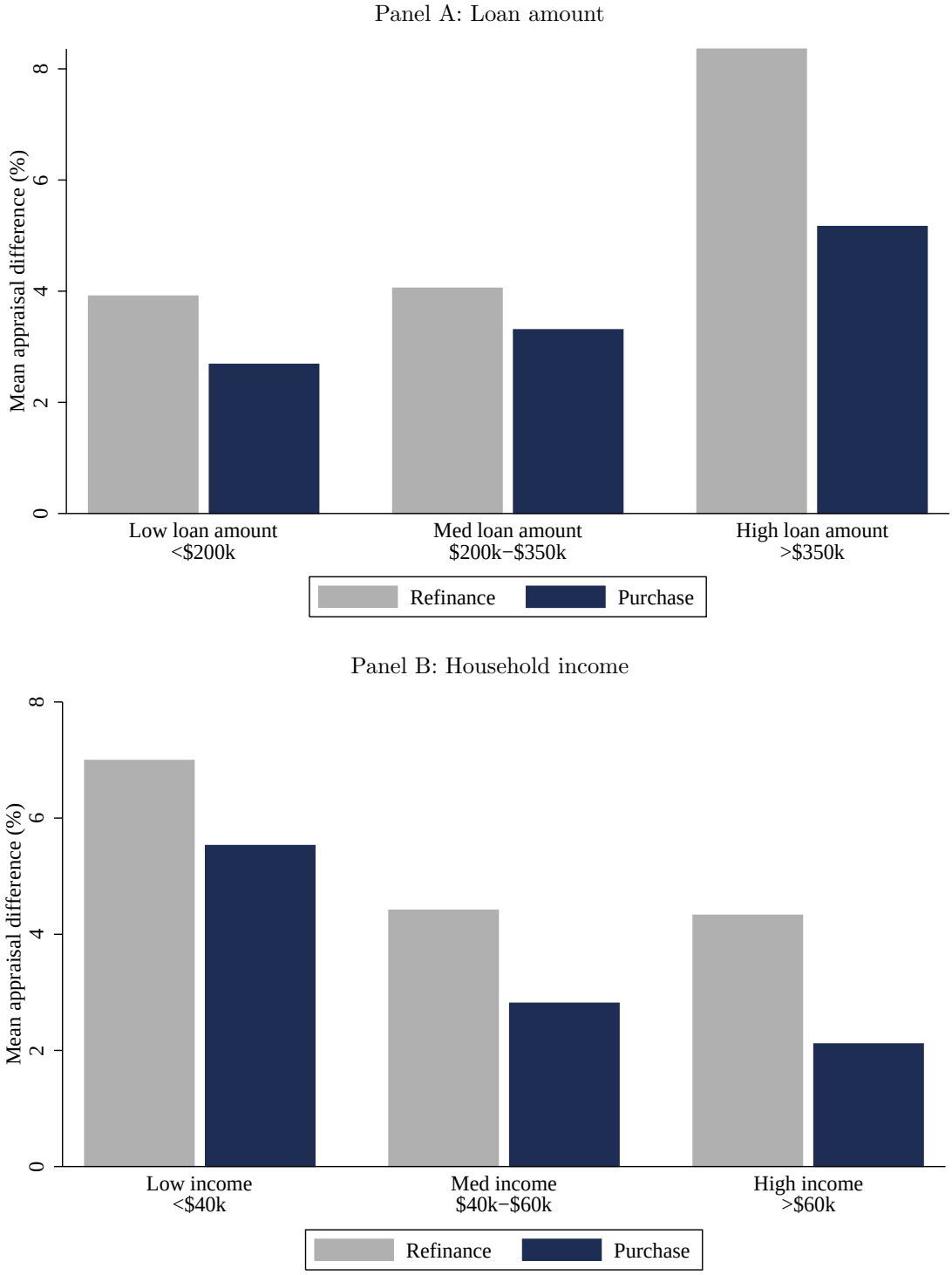


Panel B: Purchase loans



This figure plots average appraisal differences for refinance loans and purchase loans by state. Appraisal difference is defined as the difference between appraised value and AVM value, divided by the average of both values. States with less than one hundred observations are omitted.

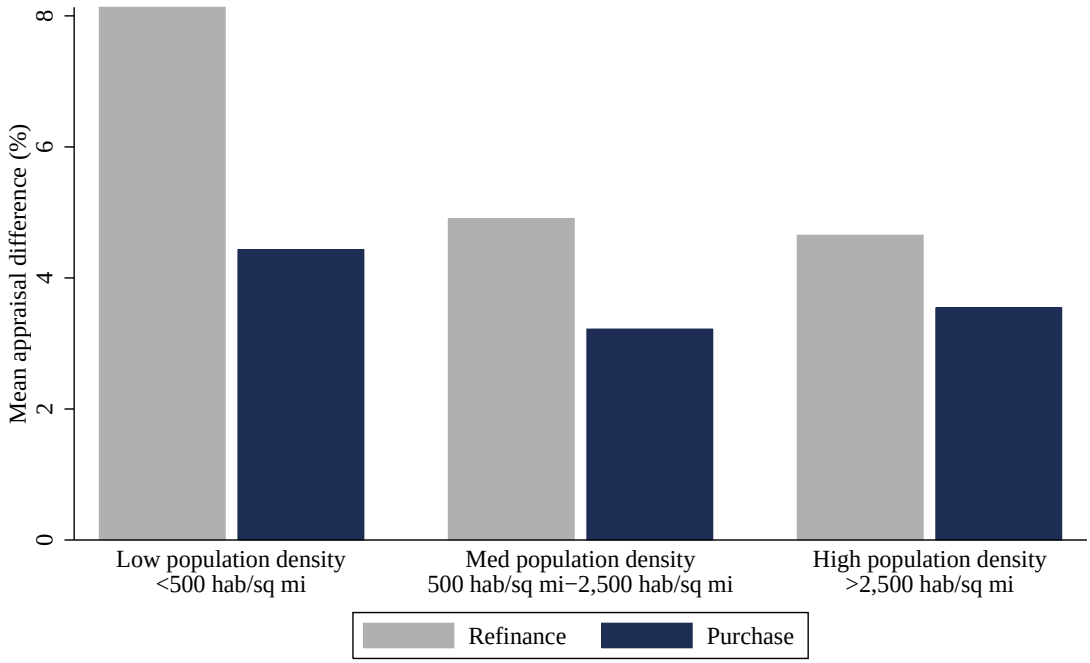
Figure IA.2. Additional cross-sectional description of appraisal differences



This figure plots the average appraisal difference by loan amount, zip code-level income in 2001 (from the SOI IRS database), zip code-level population density (from the 2000 Decennial Census), and zip code-level house market liquidity (measured as the number of purchase transactions reported by DataQuick in the loan's zip code during the 12 months prior to loan origination month). Appraisal difference is the difference between appraised value and AVM value, divided by the average of both values.

Figure IA.2 (continued). Additional cross-sectional description of appraisal difference

Panel C: Population density



Panel D: House market liquidity

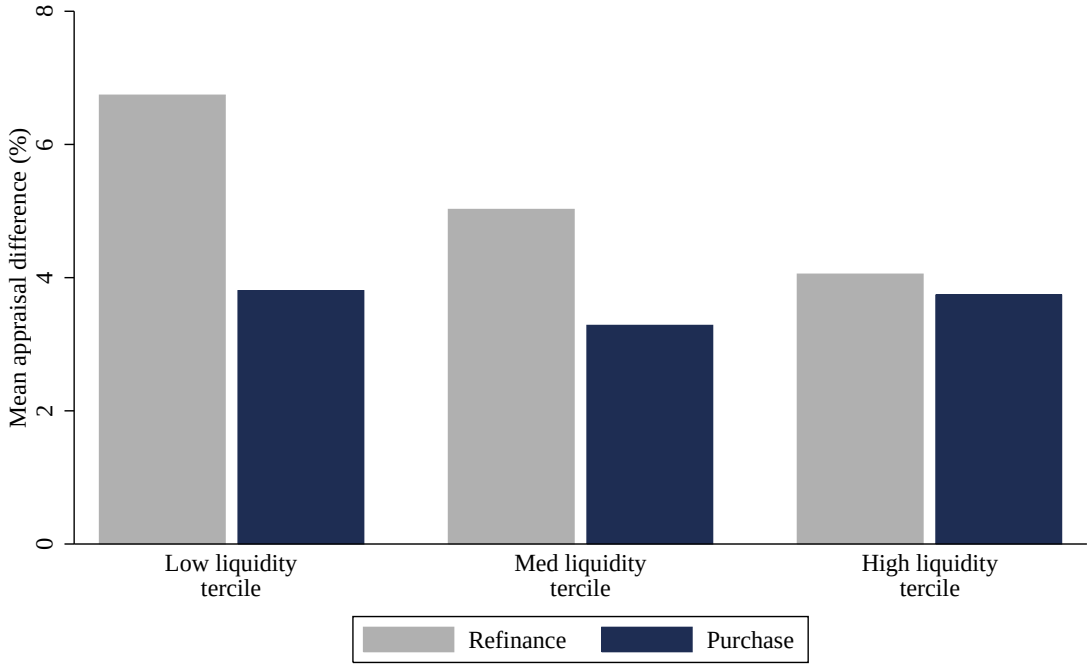
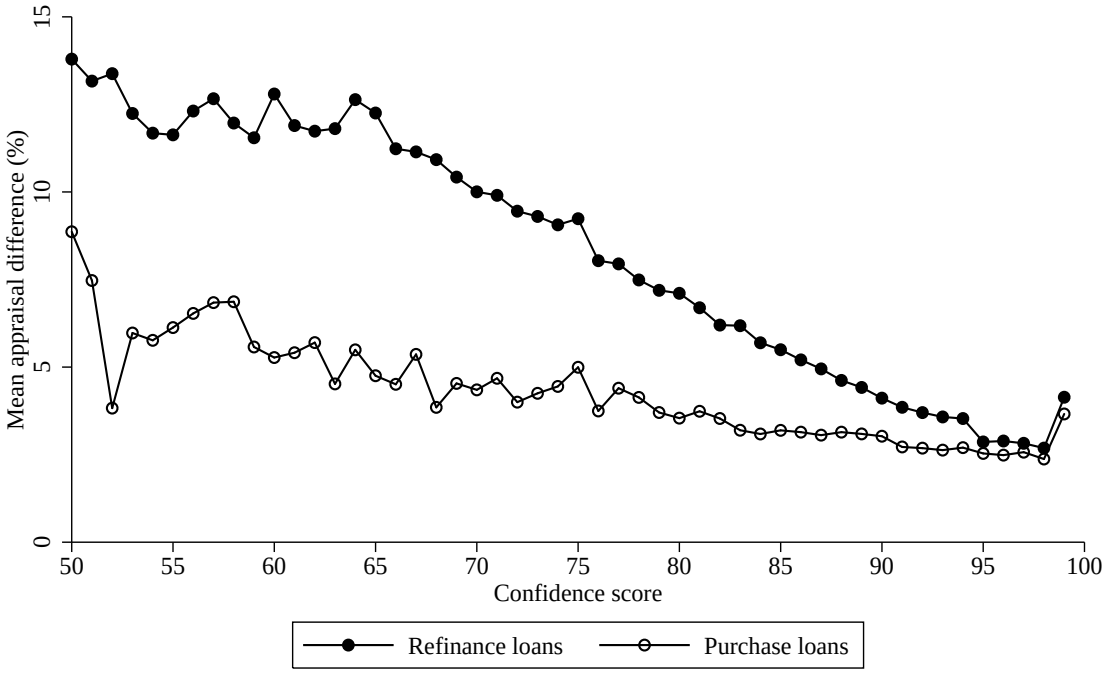


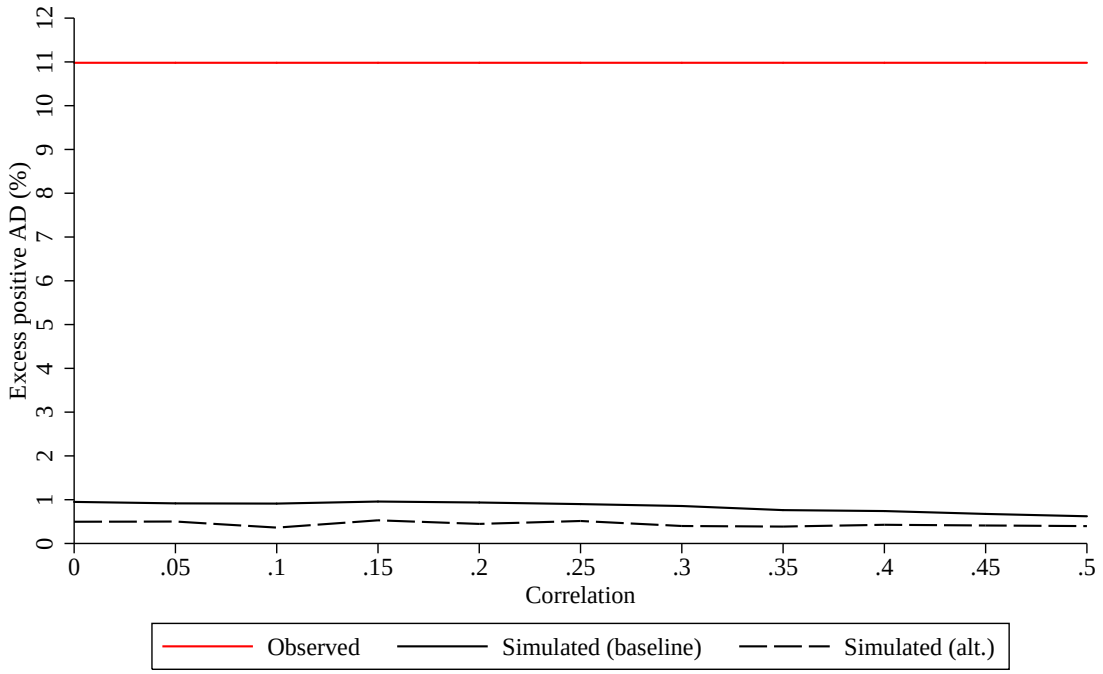
Figure IA.3. Appraisal difference by confidence score



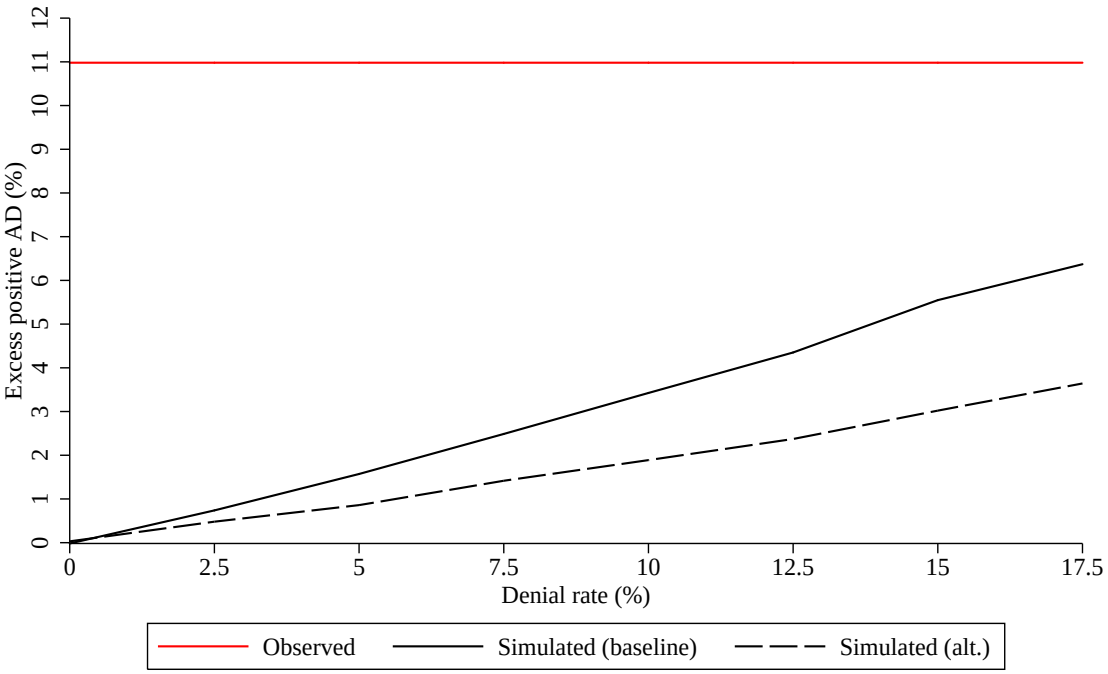
This figure plots the mean appraisal difference for refinance loans and purchase loans by AVM confidence score. Appraisal difference is the difference between appraised value and AVM value, divided by the average of both values.

Figure IA.4. Simulation sensitivity analysis for refinance loans

Panel A: Sensitivity with respect to error correlations



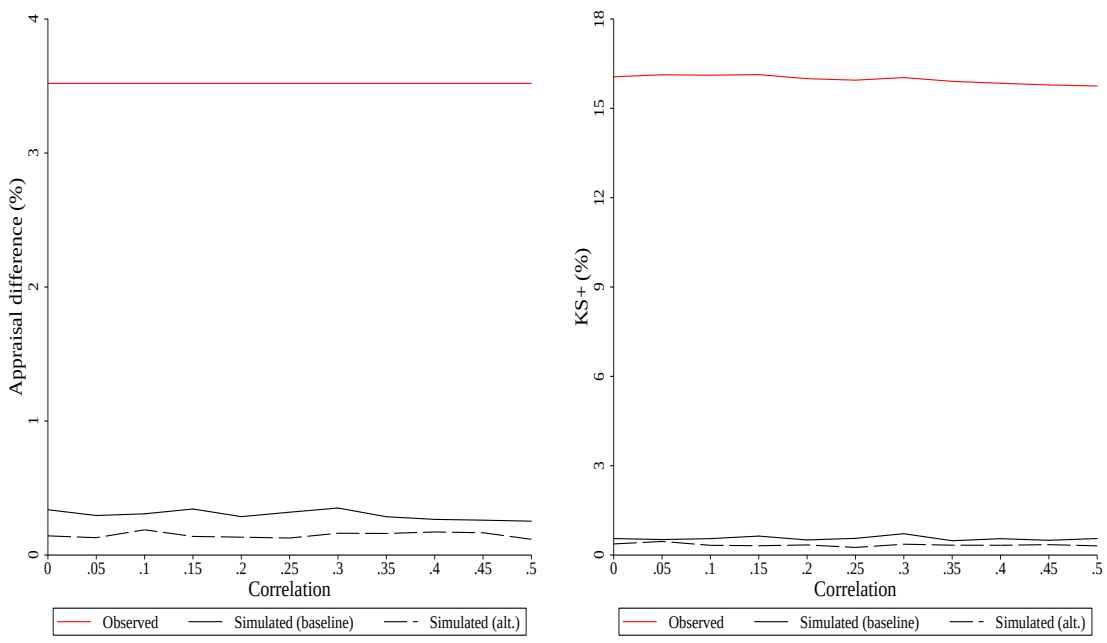
Panel B: Sensitivity with respect to denial rates



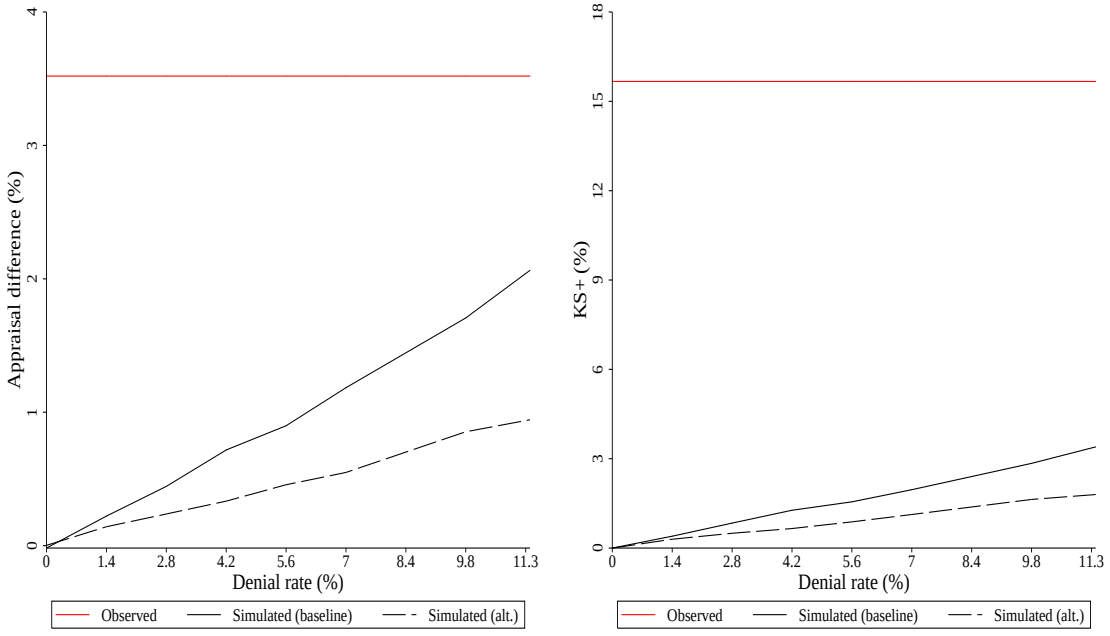
This figure plots refinance simulation results for excess positive AD under different assumptions regarding error correlations and denial rates. In the alternative simulations, we change the threshold for 100% origination probability from $A \geq V$ to $A \geq 1.25V$ while keeping the same linear structure for loan completion probability when appraisals are below the $1.25V$ threshold. Excess positive appraisal difference measures the amount of appraisals that are higher than the AVM in excess of 50%.

Figure IA.5. Simulation sensitivity analysis for purchase loans

Panel A: Sensitivity with respect to error correlations



Panel B: Sensitivity with respect to denial rates



This figure plots purchase simulation results under different assumptions regarding error correlations and denial rates. In the alternative simulations, we change the threshold for 100% origination probability from $A \geq V$ to $A \geq 1.25V$ while keeping the same linear structure for loan completion probability when appraisals are below the $1.25V$ threshold. Appraisal difference is defined as the difference between appraised value and AVM value, divided by the average of both values. KS^+ measures maximum difference from the bias-free simulated distributions. Because KS^+ is computed relative to the bias-free simulation, observed KS^+ changes slightly across the correlation scenarios in Panel A.

Table IA.1. New Century unfunded loan application summary statistics

	Refinance loans N = 976,737		Purchase loans N = 300,223	
Variables	Mean	SD	Mean	SD
Appraisal bias measures				
Appraisal-price difference (%)	-	-	2.0	20.4
Appraisal-price difference≥0 (d,%)	-	-	98.0	-
Loan/borrower characteristics				
Loan amount ($000)	172.8	124.8	201.5	143.4
LTV (%)	77.3	14.57	85.6	9.8
ARM (d,%)	65.5	-	78.5	-
Prepayment penalty (d,%)	71	-	72.1	-
Owner occupied (d,%)	95.8	-	79.7	-
Interest rate (%)	7.5	2.1	7.6	2.0

This table reports summary statistics for the sample of unfunded loan applications from New Century internal records. The sample consists of first-lien loan applications submitted between 2001 and 2007 for purchase or refinancing with original loan balances between $30k and $1 million. Loans with original LTV ratios over 103% or with CLTV ratios below 25%, as well as loans reported as being for homes of over one unit are excluded. FHA and VA loans are also dropped. Appraisal-price difference is the difference between appraisal and the property's purchase price divided by the purchase price.

Table IA.2. New Century-ABSNet merged data summary statistics

	All loans N = 70,325		Refinance loans N = 53,330		Purchase loans N = 16,995	
Variables	Mean	SD	Mean	SD	Mean	SD
Appraisal bias measures						
Appraisal difference (AD) (%)	4.63	22.3	4.85	22.5	3.96	21.7
AD>0 (d,%)	60.7	-	62.0	-	56.7	-
Loan/borrower characteristics						
Purchase loan (d,%)	24.2	-	-	-	-	-
Loan amount ($000s)	223.3	130.8	217.7	125.9	240.8	143.6
FICO score	608.8	59	598.5	56.3	641.3	56.6
LTV (%)	78.8	11.9	77.5	12.8	82.5	7.2
ARM (d,%)	74.5	-	70.5	-	87	-
Full documentation (d,%)	58.7	-	63.3	-	44.1	-
Prepayment penalty (d,%)	58	-	56.3	-	63.5	-
Owner occupied (d,%)	92.7	-	94.4	-	87.6	-
Complex (d,%)	0.003	-	0.000	-	0.012	-
Interest rate (%)	7.8	1.2	7.8	1.2	7.9	1.2
Loan performance						
Delinquent 90+ before Sep. 2012 (d,%)	48.9	-	44.6	-	62.5	-

This table reports summary statistics for the sample of New Century-ABSnet matched loans. We match the loans in the two datasets based on their zip code, loan size, first payment date, purpose, type of interest rate (fixed or floating), and credit score, and we require matches to be unique. A more detailed description is available in Internet Appendix A

Table IA.3. Cash-out vs. non-cash-out refinance loans

	Appraisal Difference			Even LTV
	(1)	(2)	(3)	(4)
Mean (%)	5.4	5.4	5.4	45.2
Even LTV	1.518***		1.435***	
	(0.103)		(0.098)	
Cashout		1.319***	1.208***	7.689***
		(0.126)	(0.124)	(0.430)
Controls	yes	yes	yes	yes
CBSA×Quarter FE	yes	yes	yes	yes
N	3,662,156	3,662,156	3,662,156	3,662,156
R^2	0.11	0.11	0.11	0.25

Columns (1) to (3) report results from OLS regressions where the dependent variable is the loan's appraisal difference. The explanatory variables of interest are indicator for even LTV and an indicator for cash-out refinance. Control variables include indicators for full-doc loans, the presence of a prepayment penalty, owner occupied properties, complex loans, adjustable-rate loans, as well as credit score, loan amount, LTV, interest rate at origination, and an interaction term between interest rate and the adjustable rate indicator. Column (4) reports the result from an OLS regression where the dependent variable the indicator for even LTV. Reported t-statistics in parentheses are heteroscedasticity-robust and clustered by CBSA. ***p <0.01, **p <0.05, *p <0.1. *p <0.1.

Table IA.4. Appraisal bias and loan performance and pricing of purchase loans

	Delinquent			Interest rate		
	(1)	(2)	(3)	(4)	(5)	(6)
Mean (%)	48.9	48.9	48.9	7.8	7.8	7.8
AD	7.569***		6.224***	-0.051**		-0.060***
	(0.975)		(1.055)	(0.023)		(0.023)
A=Price		15.341***	13.980***		0.219***	0.220***
		(0.743)	(0.782)		(0.015)	(0.017)
AD×A=Price			8.109***			0.012
			(3.084)			(0.053)
Controls	yes	yes	yes	yes	yes	yes
CBSA×Quarter FE	yes	yes	yes	yes	yes	yes
N	70,325	70,325	70,325	70,325	70,325	70,325
R^2	0.28	0.27	0.29	0.60	0.60	0.60

This table reports results analogous to Table 5, for purchase loans instead of refinance loans. Columns (1) to (3) report results from OLS regressions where the dependent variable is a dummy variable that takes the value of one if the loan became more than 90 days delinquent at any point in time between origination and September 2012, and zero otherwise. The explanatory variables of interest are the loan's appraisal difference and an indicator for appraisal being equal to purchase price. Control variables include indicators for full-doc loans, the presence of a prepayment penalty, owner occupied properties, complex loans, adjustable-rate loans, as well as credit score, loan amount, LTV, interest rate at origination, and an interaction term between interest rate and the adjustable rate indicator. Columns (4) to (6) report results from OLS regressions where the dependent variable is the loan interest rates at origination. The regression specifications are the same as in columns (1) to (3) except that interest rate is not a control variable (because it is the dependent variable) and an additional control variable indicator for LTV ratios above 80 is included. Reported t-statistics in parentheses are heteroscedasticity-robust and clustered by CBSA. ***p <0.01, **p <0.05, *p <0.1. *p <0.1.

Table IA.5. Baseline simulation calibration parameters and appraisal bias moments

Panel A: Refinance loans

	Data	Bias free simulation	Selection bias simulation
Calibration parameters			
σ_A	-	19.10	19.30
σ_{AVM}	-	19.10	19.30
β	-	0	0.33
Appraisal bias moments			
σ_{AD}	24.26	24.28	24.35
d	2.50	0	2.54
Mean AD	5.36	-0.04	0.57
AD>0−0.5	10.98	-0.05	0.79
KS^+	15.59	-	0.87
Mean (A-AVM)/AVM	9.28	3.07	3.74
(A-AVM)/AVM>0.20	21.28	21.96	22.70
(A-AVM)/AVM<-0.20	8.02	17.37	16.77

Panel B: Purchase loans

	Data	Bias free simulation	Selection bias simulation
Calibration parameters			
σ_A	-	20.30	20.30
σ_{AVM}	-	20.30	20.30
β	-	0	0.21
Appraisal bias moments			
σ_{AD}	21.27	21.39	21.29
d	1.70	0.00	1.70
Mean AD	3.62	0.01	0.27
AD>0−0.5	7.56	0.08	0.47
KS^+	15.67	-	0.74
Mean (A-AVM)/AVM	6.69	2.41	2.65
(A-AVM)/AVM>0.20	14.95	18.80	19.06
(A-AVM)/AVM<-0.20	6.41	14.06	13.72

This table reports the parameter values and appraisal bias moments from the baseline simulations. Appraisal and AVM values are modeled as bivariate normal random variables with means of zero, equal error standard deviations, and correlations of 0.25 and 0.5 respectively for refinance and purchase loans. We calibrate error standard deviations for Appraisal and AVM such that the simulated appraisal difference (AD) standard deviations for refinance and purchase loans match their empirical counterparts. To model selection, we assume that loan completion probability is one if an appraisal is above the property's true value and is otherwise $max(0, 1 - \beta(V - max(0, A))/V)$, where V represents the property's true value and can be normalized to one. The parameter β is calibrated such that the simulation generates targeted denial rates of 2.5% for refinance loans and 1.7% for purchase loans, which are based on observed HMDA collateral denial rates. Excess positive appraisal difference measures the amount of appraisals that are higher than the AVM in excess of 50% and KS^+ measures the maximum differences from the bias-free simulated distributions.

Table IA.6. Simulation sensitivity analysis

Panel A: A≥V threshold for 100% loan completion probability

		Refinances				Purchases		
		Mean AD	Excess positive AD	KS^+		Mean AD	Excess positive AD	KS^+
$\rho = 0$	d = 0	-0.01	0.03	0	d = 0	0.01	-0.01	0
	d = 2.5	0.57	0.89	0.89	d = 1.7	0.34	0.62	0.64
	d = 7.5	1.93	2.87	2.85	d = 4.9	1.04	1.83	1.86
	d = 12.5	3.50	5.08	5.05	d = 8.1	1.88	3.19	3.23
	d = 17.5	5.31	7.47	7.53	d = 11.3	2.72	4.56	4.58
$\rho = 0.25$	d = 0	-0.04	-0.05	0	d = 0	0.01	0.03	0
	d = 2.5	0.57	0.79	0.87	d = 1.7	0.32	0.53	0.52
	d = 7.5	1.67	2.48	2.54	d = 4.9	0.93	1.58	1.55
	d = 12.5	3.09	4.39	4.47	d = 8.1	1.62	2.71	2.71
	d = 17.5	4.67	6.47	6.58	d = 11.3	2.40	3.98	3.96
$\rho = 0.5$	d = 0	-0.06	-0.12	0	d = 0	0.01	0.08	0
	d = 2.5	0.44	0.60	0.74	d = 1.7	0.27	0.47	0.74
	d = 7.5	1.42	2.00	2.14	d = 4.9	0.78	1.29	2.14
	d = 12.5	2.62	3.57	3.70	d = 8.1	1.36	2.21	3.70
	d = 17.5	3.94	5.27	5.51	d = 11.3	2.03	3.17	5.51

Panel B: A≥1.25V threshold for 100% loan completion probability

		Refinances				Purchases		
		Mean AD	Excess positive AD	KS^+		Mean AD	Excess positive AD	KS^+
$\rho = 0$	d = 0	0.02	0.05	0	d = 0	-0.04	-0.07	0
	d = 2.5	0.26	0.47	0.46	d = 1.7	0.12	0.18	0.42
	d = 7.5	0.88	1.50	1.50	d = 4.9	0.44	0.84	0.95
	d = 12.5	1.60	2.67	2.66	d = 8.1	0.80	1.47	1.59
	d = 17.5	2.38	3.91	3.90	d = 11.3	1.11	2.07	2.16
$\rho = 0.25$	d = 0	0.00	0.01	0	d = 0	-0.02	-0.04	0
	d = 2.5	0.26	0.43	0.49	d = 1.7	0.15	0.28	0.40
	d = 7.5	0.80	1.31	1.34	d = 4.9	0.41	0.77	0.83
	d = 12.5	1.47	2.40	2.41	d = 8.1	0.77	1.47	1.51
	d = 17.5	2.28	3.69	3.68	d = 11.3	1.08	2.02	2.07
$\rho = 0.5$	d = 0	0.02	0.00	0	d = 0	0.04	0.06	0
	d = 2.5	0.29	0.52	0.53	d = 1.7	0.14	0.27	0.53
	d = 7.5	0.79	1.23	1.24	d = 4.9	0.36	0.69	1.24
	d = 12.5	1.30	2.10	2.10	d = 8.1	0.67	1.24	2.10
	d = 17.5	2.12	3.40	3.40	d = 11.3	0.98	1.77	3.40

This table reports sensitivity analysis for the correlation, denial rate, and appraisal thresholds assumptions discussed in Section 3.1. In total, we consider 15 permutations under both baseline and alternative appraisal thresholds. Appraisal and AVM values are modelled as bivariate normal random variables with means of zero and equal error standard deviations. We calibrate error standard deviations for Appraisal and AVM such that the simulated appraisal difference (AD) standard deviations for refinance and purchase loans match their empirical counterparts. To model selection, we assume that loan completion probability is one if an appraisal is above the property's true value and is otherwise $max(0, 1 - \beta(V - max(0, A))/V)$, where V represents the property's true value and can be normalized to one. The parameter β is calibrated such that the simulation generates a targeted denial rate. Excess positive appraisal difference measures the amount of appraisals that are higher than the AVM in excess of 50% and KS^+ measures the maximum differences from the bias-free simulated distributions.