Liar's loan? –Effects of Origination Channel and Information Falsification on Delinquency

Wei Jiang	Columbia University
Ashlyn Nelson	Indiana University
Edward Vytlacil	Yale University

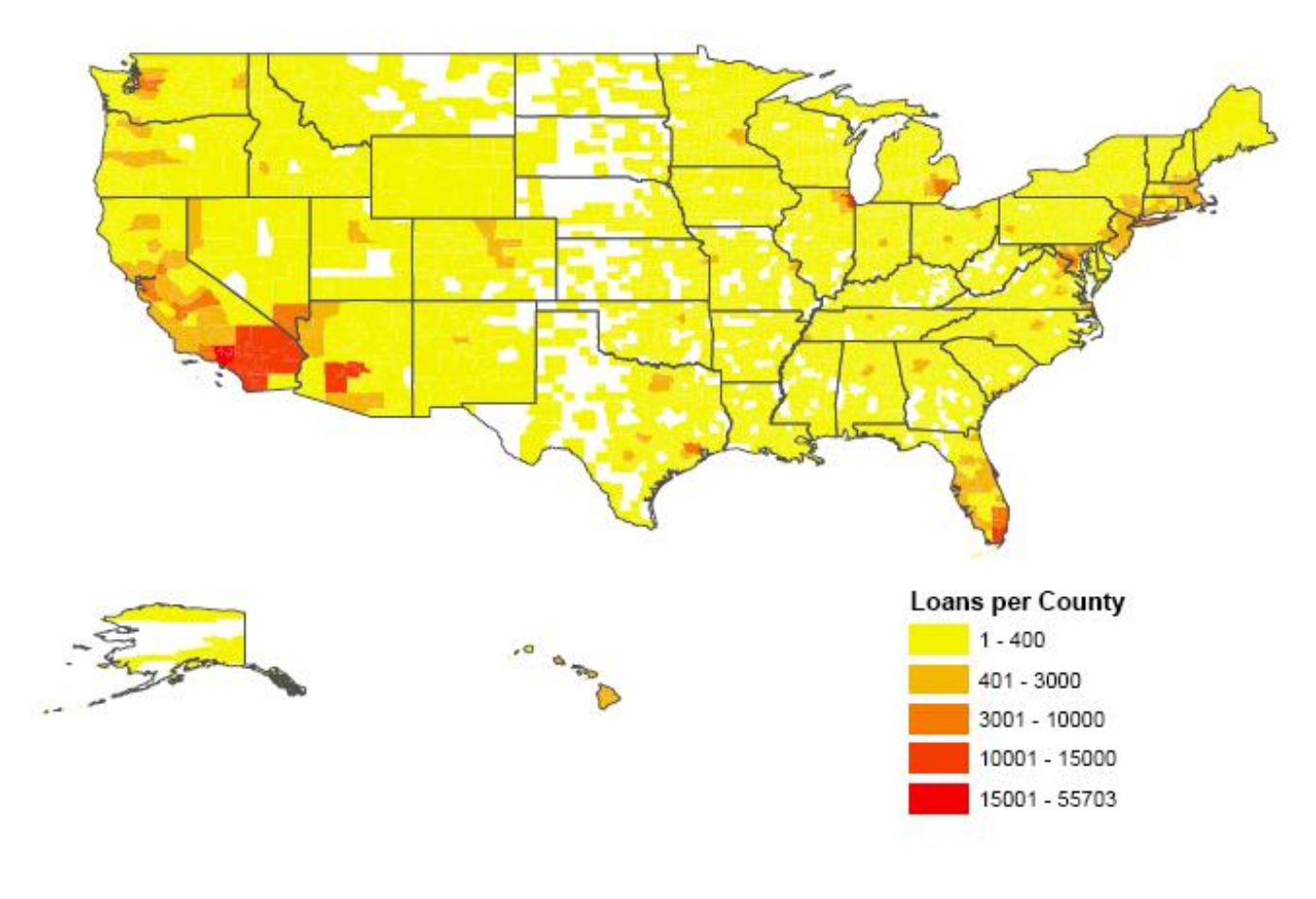
September, 2009

Mission of the paper

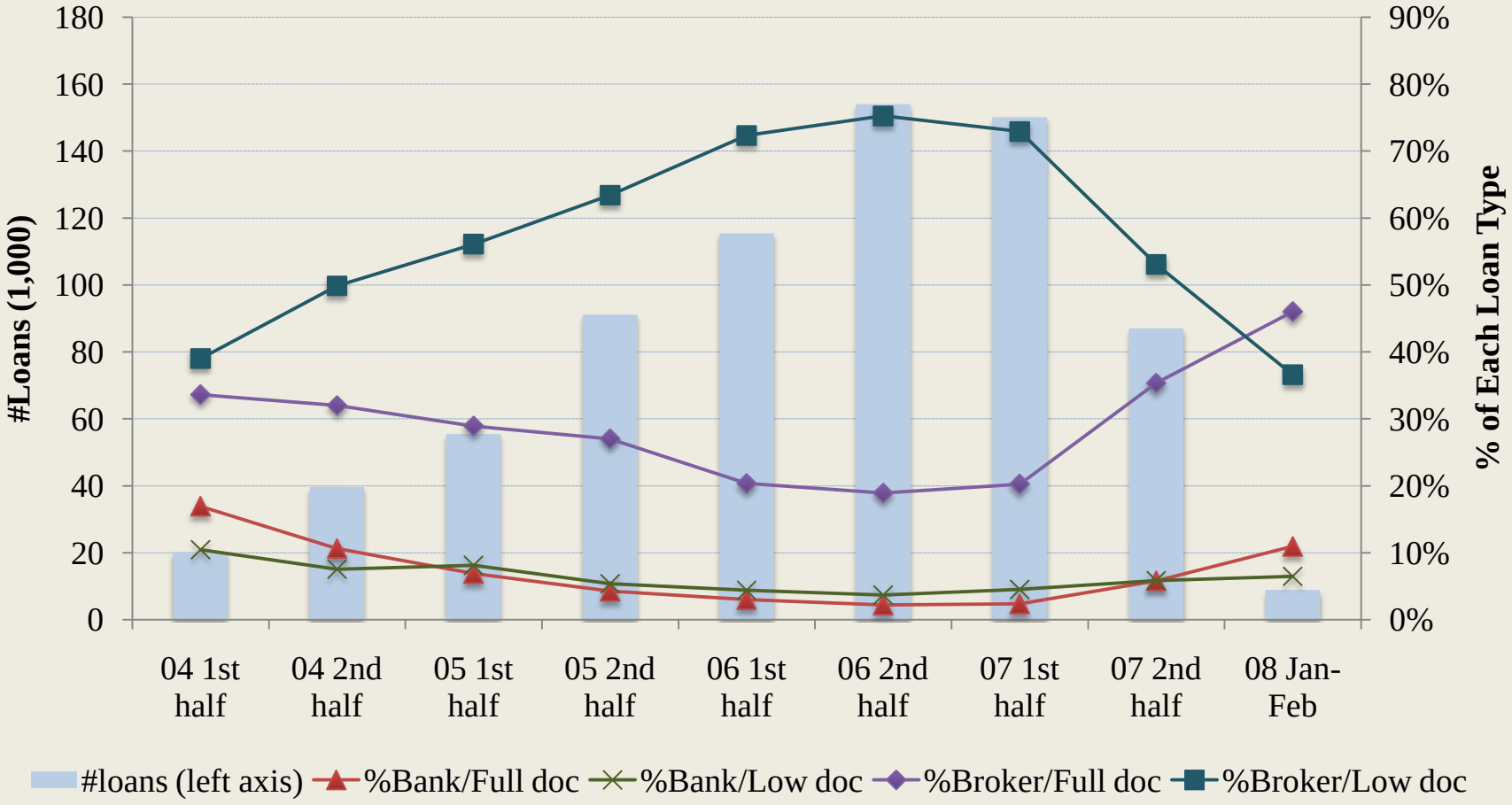
- A comprehensive delinquency predictive analysis.
  - Have detailed loan, property, and borrower information, and origination channels—all information the bank recorded at origination.
  - Updated performance to early 2009.
  - Allows an accurate calibration of the hard information set by the bank.
    - Analyze "soft information."
    - Analyze agency problem/adverse selection.
- Analyze two layers of agency problems:
  - Bank vs. third-party (correspondents and brokers): origination channel
  - Lender vs. borrower: information falsification

Data: 700,000 + loans

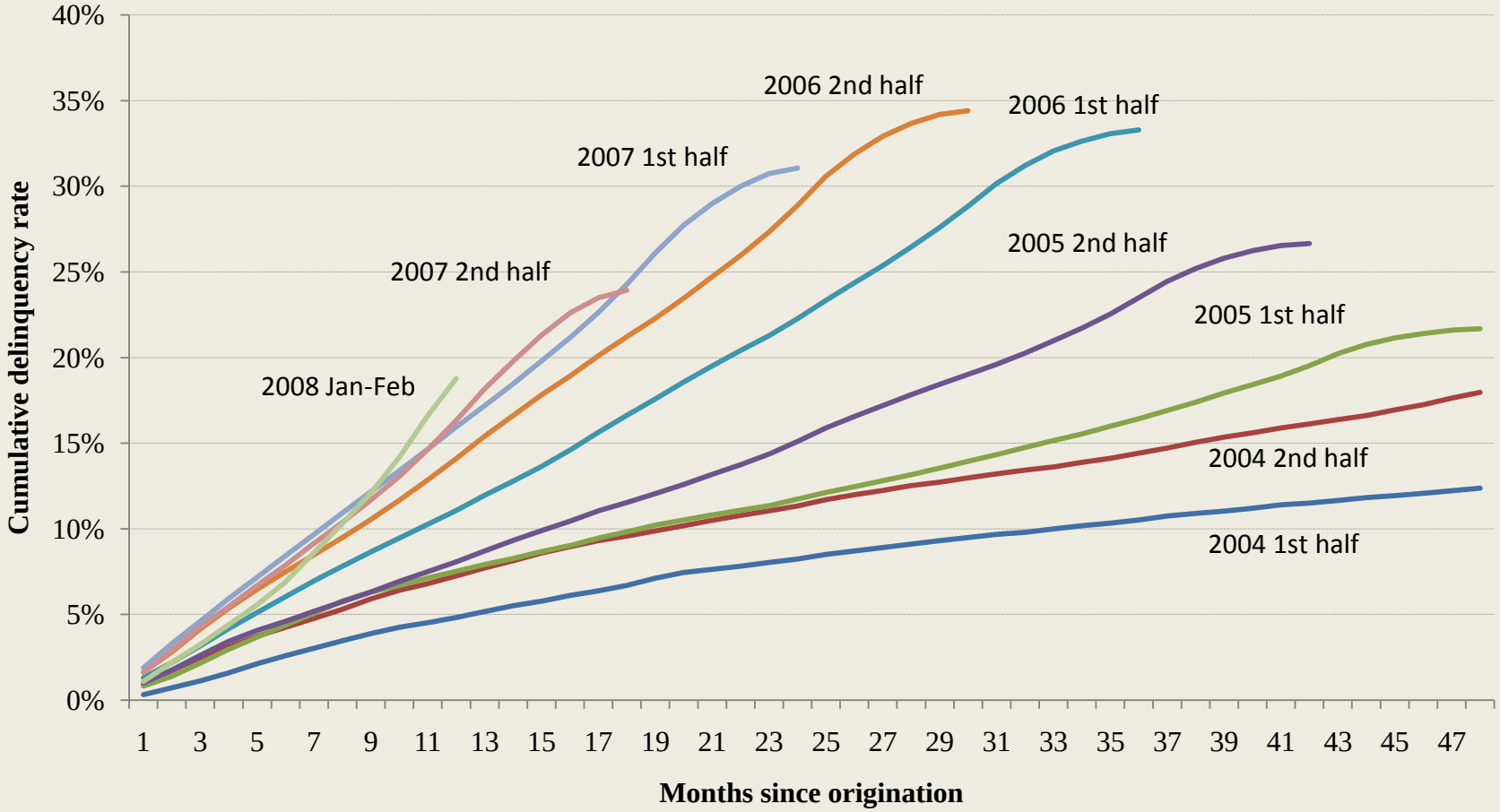
issued in Jan. 2004 – Feb. 2008 by a top national mortgage bank.



Number of Loans and Composition
(by Semi-Years: 2004-2008)



Cumulative delinquency: origination - Jan 2009



Sample representativeness

	Our sample	General market
% loans originated by third party	90%	60%-70%
% loans securitized	85%	60%-80% for all; 75-91% for subprime
% low-doc	70%	25%
% subprime	15%	18-21%
LTV	About the same	
Loan amount	Our sample is about 15% higher	
Credit score	Our sample is about 5-8 points lower	
Demographics	Our sample has higher % of Hispanic borrowers	
Annual growth 2004-2006	> 50%	30-40%
% Delinquency (early 2009)	26%	11% for all, 39% for subprime

- "Outsource origination to distribution" model.
- A representative yet amplified version of the boom-bust cycle.

Main issue #1: Delinquency prediction and origination channels

- Four subsamples: Bank/Full-Doc; Bank/Low-Doc; Broker/Full-Doc; Broker/Low-Doc.
  - Brokered loans could be divided into "correspondents" and "pure brokers."
- Dependent variable:
  - Delinquency status at the end of the sample (probit).
  - Time to delinquency (duration with censoring).
- Covariates: A fuller set of predictive variables than previous studies.
  - About the loan: LTV (first and second lien); loan amount; refinance; prepay penalty; owner occupancy; first time borrower.
  - About borrower economic condition: income; cash reserve, credit score; tenure; self-employment.
  - About borrower demographics: gender; race/ethnicity; age.
  - Origination year dummies.
- Cluster level: MSA.

Delinquency across origination channels

- Cumulative rate & survival rate after five years:
  - Bank/Full-Doc: 13.2% & 86.3%
  - Bank/Low-Doc: 18.0% & 68.9%
  - Broker/Full-Doc: 23.6% & 64.7%
  - Broker/Low-Doc: 31.6% & 45.9%
  - Correspondent Brokers are between Bank and Non-Correspondents, and closer to the former.
- Two possibilities:
  - Broker and low-doc channels approach observably worse-quality borrower pools.
  - Broker and low-doc channels attract worse types (unobservable).
  - All loans, once originated, are "treated" the same—all serviced by the bank.

Choice of Broker and Low-Doc

- Broker: Observably lower credit quality
  - Less experienced borrower belonging to groups that have lower credit quality on average: first-timer; low credit score; low income; female; minority; young; short tenure.
  - Young neighborhoods with low minority representation.
  - Main issue: aggressive lending to the less-informed.
- Low-Doc: "good on paper"
  - Low LTV; high credit score; high income; non-first-timer. And self-employed.
  - Booming young minority neighborhoods.
  - Hide information unrelated to delinquency (e.g., taxes).
  - Hide information related to delinquency: withholdings on income; other major expenditure.

Nonlinear Blinder-Oaxaca decomposition

- Full-Doc (D = 0) vs. Low-Doc (D = 1):

	Bank			Broker		
	Difference	t-stat	Percentage	Difference	t-stat	Percentage
Endowment Effect	-0.06%	-0.10	-1.20%	-0.89%	-1.62	-11.10%
Coefficient Effect	4.87%	9.13	101.20%	8.91%	12.84	111.10%
Total	4.81%	5.37	100%	8.02%	8.05	100%

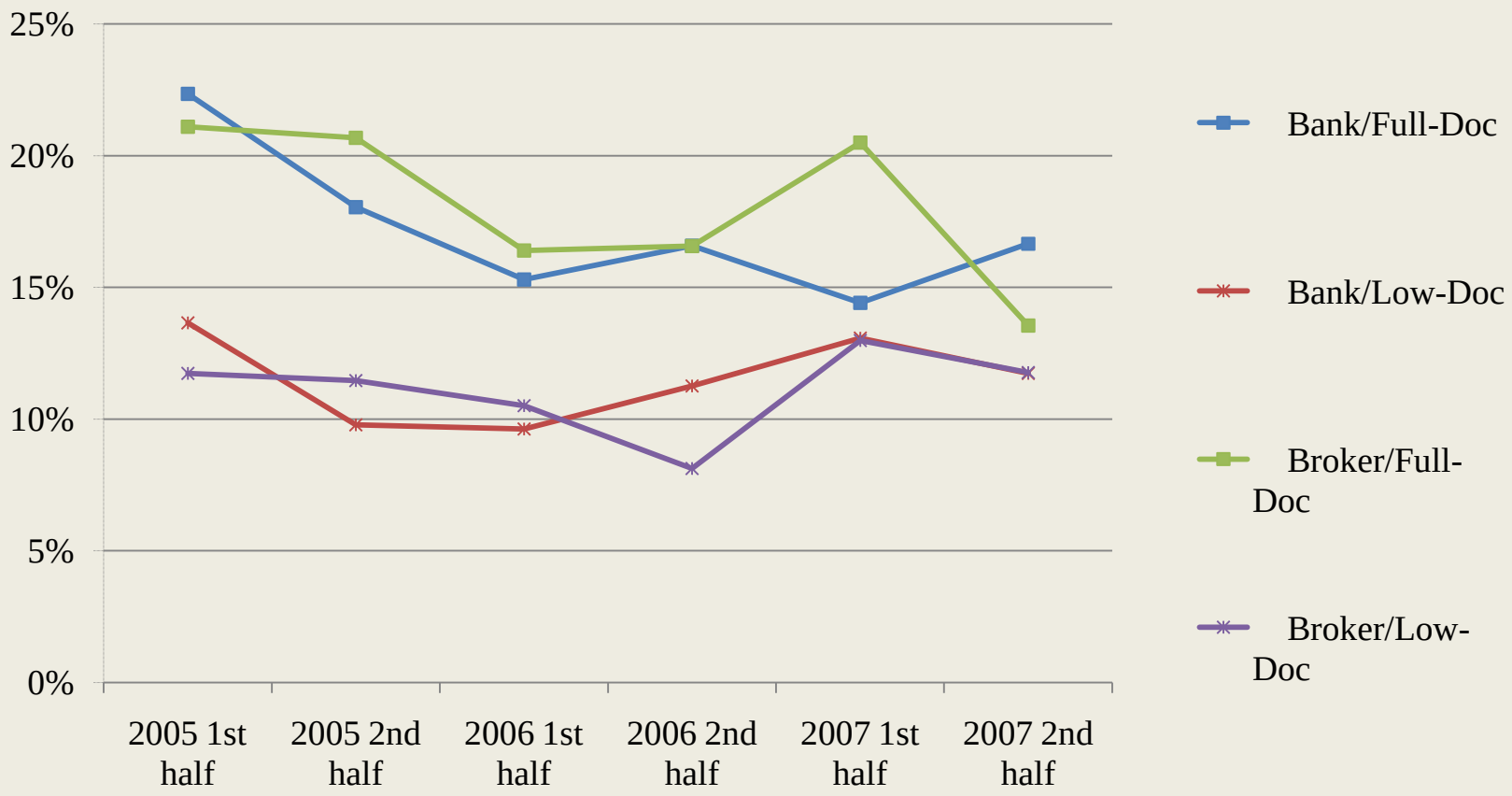
- Bank (D = 0) vs. Broker (D = 1):

	Full-Doc			Low-Doc		
	Difference	t-stat	Percentage	Difference	t-stat	Percentage
Endowment Effect	7.84%	8.09	75.69%	10.40%	12.16	76.67%
Coefficient Effect	2.52%	9.46	24.31%	3.16%	8.76	23.33%
Total	10.35%	10.51	100%	13.56%	13.99	100%

Main issue #2: Liar's loan

- Borrower information falsification, possible encouraged by the brokers.
- Should appear primarily among low- and no-doc loans.
- Information most susceptible to falsification: income; assets; other major expenditure; primary residence.
- Two-level approach:
  - In the aggregate, information falsification should compromise model predictive power. Pseudo R-squared confirms the order between full- and low-doc subsamples.
  - At the covariate level, falsification should distort the relation to delinquency.

Out-of-sample predictive power: Full-doc vs. Low-doc



Falsification of individual variables: income

	Bank/Full-Doc			Bank/Low-Doc			Broker/Full-Doc			Broker/Low-Doc		
	Coef	t-stat	APE	Coef	t-stat	APE	Ceof	t-stat	APE	Coef	t-stat	APE
LTV	1.693	14.61	36.15%	2.48	19.41	56.35%	2.028	17.81	50.99%	3.021	19.15	91.45%
AddLTV	1.467	7.24	31.32%	1.566	7.44	35.57%	1.665	15.98	41.84%	2.975	24.28	90.05%
Loan (log)	0.113	4.08	2.42%	0.178	7.23	4.04%	0.214	8.83	5.38%	0.252	8.78	7.64%
SecondLien	0.245	1.78	5.22%	0.729	6.23	16.56%	0.498	8.07	12.52%	0.297	3.79	9.00%
Refinance	-0.046	-1.08	-0.97%	-0.038	-1.32	-0.86%	-0.05	-2.15	-1.25%	0.097	5.49	2.94%
PrepayPenalty	0.111	2.1	2.37%	0.028	0.7	0.63%	0.005	0.26	0.12%	0.082	6.38	2.49%
FirstTimeOwner	-0.186	-4.2	-3.97%	-0.072	-1.17	-1.63%	-0.01	-0.61	-0.24%	-0.054	-3.81	-1.62%
OwnerOccupied	-0.259	-5.31	-5.53%	-0.275	-8.18	-6.24%	-0.35	-13.75	-8.79%	-0.281	-10.31	-8.51%
OneBorrower	0.267	12.81	5.70%	0.346	15.34	7.87%	0.292	19.32	7.34%	0.298	17.07	9.03%
Income (log)	-0.108	-6.91	-2.30%	0.023	1.32	0.53%	-0.064	-4.33	-1.61%	0.041	4.75	1.26%
IncomeMiss	-0.033	-0.28	-0.71%	-0.006	-0.13	-0.14%	-0.16	-2.97	-4.02%	0.155	6.98	4.71%
CashResv	-0.047	-5.61	-1.01%	-0.027	-3.61	-0.60%	-0.09	-17.94	-2.27%	-0.069	-16.12	-2.10%
CreditScore	-0.009	-53.89	-0.18%	-0.008	-31.84	-0.17%	-0.008	-49.91	-0.21%	-0.007	-71.41	-0.21%
Female	-0.043	-1.71	-0.93%	-0.014	-0.75	-0.32%	-0.003	-0.2	-0.07%	0.003	0.34	0.08%
Hispanic	0.276	5.5	5.89%	0.219	3.78	4.98%	0.391	7.75	9.83%	0.275	10.55	8.33%
Black	0.129	2.74	2.76%	0.156	2.75	3.55%	0.167	5.16	4.21%	0.12	4.53	3.64%
Asian	-0.053	-0.52	-1.13%	-0.052	-1.05	-1.18%	0.022	0.69	0.55%	0.037	1.25	1.12%
Age (log year)	-0.089	-3.65	-1.90%	0.02	1.04	0.45%	-0.02	-1.64	-0.50%	0.005	0.57	0.16%
Tenure(log month)	-0.018	-2.01	-0.38%	-0.045	-5.25	-1.02%	-0.012	-1.87	-0.30%	-0.035	-6.95	-1.06%
TenureMiss	-0.072	-1.16	-1.54%	-0.174	-4.01	-3.95%	-0.251	-7.56	-6.32%	-0.266	-11.52	-8.07%
SelfEmploy	-0.001	-0.03	-0.03%	0.053	2.82	1.20%	0.051	2.44	1.29%	0	-0.01	0.00%
# obs and # clusters		31,408	807		35,553	778		166,402	963		425,181	949

Estimate the average exaggeration of income

- Identifying assumption:

$$E(Income^* \mid X = x, \text{Low-Doc}) \leq E(Income^* \mid X = x, \text{Full-Doc})$$

- *Income** and *Income* indicate true and reported income.
- The assumption implies *Pr(Full-Doc|X, Income*)* is non-decreasing in *Income*.*
- The assumption may not hold for the self-employed—excluded from the estimation.
- Setting the assumption to *equality* provide a lower-bound estimation of income exaggeration.

Simple estimation: benchmark against neighborhood average income

- Zip code level per capital income from the IRS, 2004-2006.
- Neighborhood size: 2,326 households, 3.3 people each.
- Average ratio of borrower household income to zip-code income:
  - Bank/Full-doc and Broker/Full-doc: 3.6 and 3.3.
  - Bank/Low-doc and Broker/Low-doc: 4.3 and 3.8.
- The average exaggeration is 16-19%.

Refined estimation: benchmark against a linear function of personal & neighborhood attributes

- Projecting income using full-doc observations only:

$$
\begin{aligned}
Income = {} & 0.014 * CreditScore - 0.846 * Female + 0.651 * \ln(Age) - 0.416 * Hispanic \\
& [18.01] \qquad [-16.49] \qquad [13.31] \qquad [-1.92] \\
& - 0.430 * Black + 0.575 * Asian + 0.051 * AvgIncome - 0.030 * Unemprate \\
& [-4.31] \qquad [5.04] \qquad [4.40] \qquad [-2.15] \\
& + 0.131 * Y2005 + 0.373 * Y2006 + 0.299 * Y2007 + 0.010 * Y2008 \\
& [2.58] \qquad [5.40] \qquad [4.76] \qquad [0.096]
\end{aligned}
$$

R-squared: 6.9%; number of observations: 138,514.

- Apply the coefficients onto the low-doc subsample.
- The average (median) exaggeration is $1,830 ($753), or 29% (20%).
- Recover "true" relationship between income and delinquency:
  - Correlation of estimated true income and exaggeration: -7.9%
  - Correlation of estimated true income and delinquency: -23.5%.
  - Correlation of estimated income exaggeration and delinquency: 8.2%

“The color of credit”

- Huge race effect (controlling for observables):
  - Hispanics-white difference: 5.9 to 8.3 percentage points higher.
  - Black-white difference: 2.8 to 5.2 percentage points higher.
  - Asian-white difference: -1.2 to 1.1 percentage points (not significant).
- All coefficients on race/ethnicity/gender variables are essentially zero in the pricing equation for the Bank subsample.
- Broker subsample: Hispanics do not pay higher rates; Black rates are higher by 10-16 bps, amounting to an additional pay of $15-$25 per month on the average loan.
- Dilemma of “discrimination” vs. “irresponsible lending.”

Summary

- Brokerage channel produces low-quality loans, mostly due to looser standards on observable.
- Low-doc loans attract worse borrowers, mostly due to unobserved heterogeneity.
- Evidence of falsification in low-doc loans, most notably in income exaggeration.
- Significant race effects that are not explained by observables and loan pricing.