Discussion of “Liar’s Loan?” by Jiang, Nelson and Vytlacil

Recent Developments in Consumer Credit and Payments
September 25, 2009

Ronel Elul
Federal Reserve Bank of Philadelphia

These comments are those of the author and do not necessarily represent policies or positions of the Federal Reserve Bank of Philadelphia nor the Federal Reserve System.

Introduction

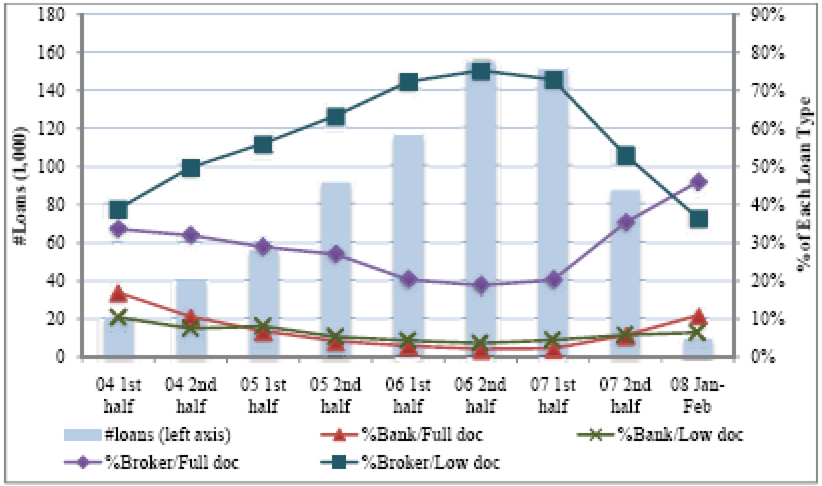
- Looks at the performance of broker-originated and lowdoc mortgages originated during the housing bubble
- An interesting and important question, since there was an explosion in the origination of these loans during the bubble, and they also defaulted at much higher rates

- Previous work
  - Many papers have taken the simple step of adding a dummy variable for lowdoc loans or broker originations to a delinquency model (e.g. Elul, 2009). Generally associated with increased risk.
  - Also, Garmaise (2009) has found that the relationship between broker and lender gets worse over time
  - This paper uses a nice dataset that allows it to examine these loans in detail

Data

- 700,000+ loans originated by a single lender between January 2004-Feb 2008.
  - Represents almost all of this lender's loans
- This lender specialized in lowdoc and broker-originated loans

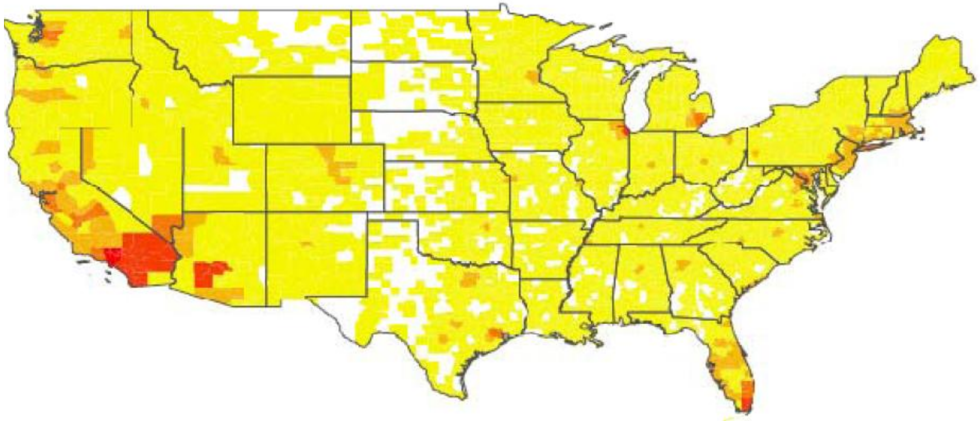
Figure 2. Number of Loans and Composition by Semi-Year: 2004-2008



- Advantages of looking at a single lender:
  - More uniform: less need to worry that differences in performance are due to differences in lenders' underwriting. Less geographically diverse
    - E.g. Elul (2009) finds an effect on delinquency rates of broker and lowdoc using McDash – to what extent is this due to the fact that the worst lenders originated these loans?
  - Potentially have access to additional variables
    - Income, age, cash reserves, first-time homebuyer, job tenure, race

- Drawbacks
  - How representative is this sample?
    - Less geographically diverse (mainly Southern CA.)
    - Almost all loans are lowdoc/broker-originated.
      - So maybe the ones that were not are special in some way
    - By contrast in McDash: 12% lowdoc, 43% broker
  - Some key variables missing/not used: type of mortgage (ARM, optionarm, current balance), interest rate, prime/subprime
    - This can be important: lowdoc and broker loans are 4 times as likely to be Option-ARMs and 40% more likely to be ARMs (McDash). This may be driving some of the observed difference in delinquency rates.

This Paper



McDash

Are Lowdoc/Broker Loans Worse?

- Split data into four subsamples:
  - Bank/Broker×FullDoc/LowDoc
- Observed Delinquency Rates:
  - Bank/Full: 13%; Bank/Low: 18%; Broker/Full: 24%; Broker/Low: 32%
- Run separate delinquency models for each subsample and then decompose the difference in these delinquency rates into “Endowment “ and Coefficient” effects (next slide)
  - Maybe should instead have started with a big regression that includes all loans, with dummies for lowdoc, broker, and maybe with further interactions (w/income, vintage, time)

- Decomposition
  - How much of the difference in delinquency rates is due to differences in observables [like lower FICO] ("endowment") vs. unobservables ("coefficient")?
  - Results: Lowdoc: almost all "coefficient". Broker: ¾ endowment, ¼ coefficient.
  - Methodology (for a linear model):
    - Run separate regressions for each subsample
    - Endowment: apply fulldoc coefficients to difference in average observables between fulldoc and lowdoc. $(X^{full}-X^{low})\beta^{full}$
    - Coefficient: what's left over. $X^{low}(\beta^{full} - \beta^{low})$.
    - Note: β's include constant term ("lowdoc dummy")

- Comments:
  - Omitted variables can bias this decomposition in favor of the coefficient effect (e.g. attribute optionarm to the coefficient effect)
  - I would have personally started with the simpler strategy outlined above, and then used the decomposition to flesh out those results
    - This would give us a benchmark for how much worse these lowdoc/broker loans are, and whether they get worse over time…
  - Interpretation of the decomposition?
    - In some sense we *expect* there to be differences in the coefficients, esp. between full-doc and low-doc loans (e.g. income, FICO coefficients). These are different products with different "production functions".
    - So even if there was no difference in default rates, there might well still be a nontrivial coefficient effect

- Bigger Picture: to what extent do we *care* about these differences?
  - Maybe these were just riskier products, with higher returns for the bank
    - Rates are 15 bp higher for lowdoc, 8bp for broker (McDash-ARMs)
    - Dropping interest rate may obscure this
    - Could try comparing delinquency rates on these mortgages to those on other loans with similar interest rates (i.e. similar risk)

Information Falsification

- How much information falsification was occurring with lowdoc loans?
  - One approach: examine coefficients in estimations for each subsample.
    - Fulldoc: higher reported income → less default
    - Lowdoc: higher income → more default
      - caution: may reflect noisier income
  - Second approach:
    - Estimate "true income" for fulldoc sample w/other observables.
    - Apply these coefficients to lowdoc borrowers (except self-employed).
      - Lowdoc borrowers exaggerate "true" income by 20%
      - Degree of income exaggeration predicts default
    - Need to flesh this out more (regression results not in paper?)

- Third approach: out-of-sample prediction
  - Fix some semi-annual time period. Estimate delinquency models for each subsample using data from past vintages
  - Use this to estimate default probability for each loan originated in this time period. Classify the loan as "predicted to be delinquent" if predicted prob>average in this period.
  - Calculate fraction of correct predictions using model.
  - Results:
    - Models better at predicting fulldoc loan delinquency than lowdoc
    - Predictive power gets worse over time, esp. 2006
    - No significant difference between broker and bank-originated

- Comments:
  - This is not necessarily a surprise. Again, maybe these are just “noisier” borrowers.
  - But it is nice to quantify this
  - Question: does the probit model include variable related to loan seasoning?

Conclusion

- Nice paper, novel dataset
- Authors should consider starting with a simple estimation with all the samples together, and also with vintage interactions.
- Try to exploit interest rates and also see if can get additional variables