Discussion of "The Effects of Big Data on Commercial Banks" by Xiao Yin

Enrico Sette

Bank of Italy and CEPR

4th Bank of Italy, Bocconi University and CEPR Conference on Financial Stability and Regulation

The analysis and conclusions expressed are those of the author and should not be interpreted as those of the Bank of Italy

Introduction - This Paper

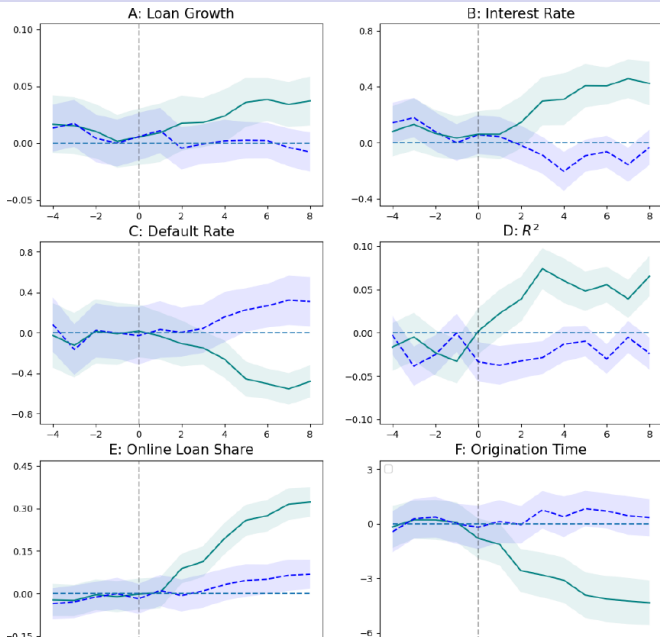What is the effect of providing banks with a large amount of data on borrowers?

Look at credit volumes, interest rates, default rates

Explain the channels through a structural model of lending under asymmetric information (Crawford et al. 2018). Convenience versus Screening.

Chinese government promoted program of information sharing

- Benefit for banks is the easier access to the information (no change in the available information)
- Benefit for borrowers is the wider availability of online applications associated with faster approval time

A: Loan Growth
B: Interest Rate
C: Default Rate
D: R^2
E: Online Loan Share
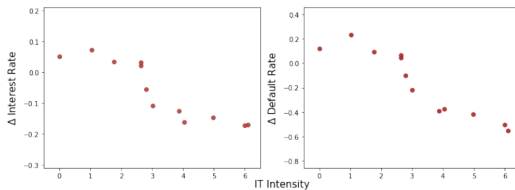F: Origination Time

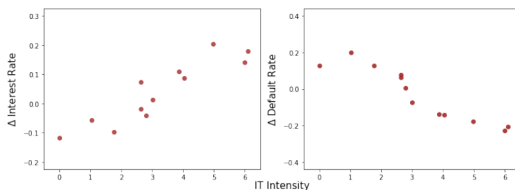# Introduction - Main Findings

A: Only Convenience Channel
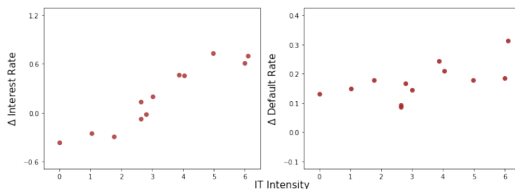
B: Only Screening Channel

C: Both Channels

Comments

- Very important topic

- Paper well executed, rich set of very interesting results

- Blend of reduced form and structural gives valuable insights on mechanism

My comments focus on 3 areas

1. Better understand the experiment

2. Some more thoughts on the results, especially online applications in the demand equation

3. More on heterogeneous effects

Comments I - The experiment

There are two key steps

- Firms need to give consent to share the data
- Banks need to apply fast enough to be among those served by the provider

The paper provides convincing evidence on the lack of correlation between bank observables and the time to reply to the data provider.

Yet:

- How about firms? Any selection? Only firms that had some contract with the government?
- Can you observe firms that give consent and firms that don't? Could create stigma effects, lead to a reallocation of credit towards firms that share the data.

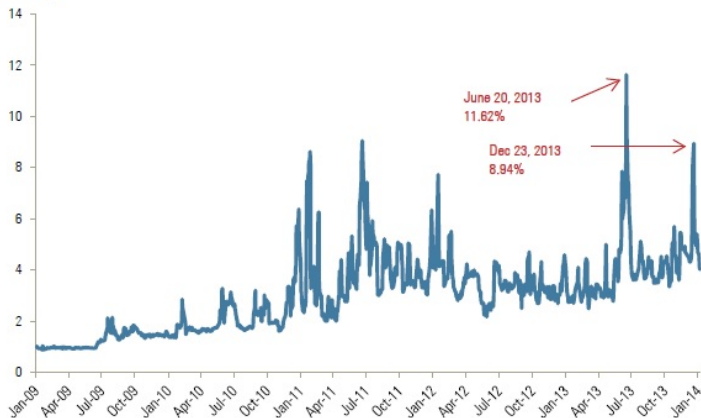At a minimum, more discussion of the institutional setting.

Comments I - The experiment

Say something more about the timing of the program.

Was the gov't initiative a way to cushion the real effects of the Chinese Banking Liquidity Crisis of 2013?

Figure I. Weighted Average Seven-Day Interbank Bond Collateral Repo Rate
Unit: %pa

June 20, 2013
11.62%

Dec 23, 2013
8.94%

Comments II - Results, online applications

	(1) Demand
Interest Rate	-0.38
	(0.15)
Interest Rate × Relationship	-0.72
	(0.24)
Online	-1.34
	(0.18)
Online × Relationship	-0.71
	(0.12)
log(Distance)	-0.20
	(0.04)
log(AT)	-0.06
	(0.14)
log(Volume)	3.43
	(0.10)
Age	0.02
	(0.44)
Profitability	0.00
	(0.41)
Leverage	0.00
	(0.01)

Borrower is **less** likely to choose a bank if it makes the online application? Why is that? This is the demand equation (paper says: "data-sharing allows banks to decrease the origination time for all types of borrowers. This leads to a higher demand, and thus higher interest rate from the average borrower.")

Comments III - Heterogeneous Effects

Standard result in theory: more precise info should benefit better borrowers and penalize worse borrowers. This is found in the paper (Table 5), could be emphasized more.

Explore further the interaction between risk and preference for speed.

- How are they correlated (if any)?

- Explore theory of multidimensional screening for predictions (Rochet and Stole 2010), here the two "types" would be preferences for speed and risk. Could help to further rationalize why the convenience channel dominates in equilibrium leading to higher interest rates.

Speed of internet connection can be a relevant dimension of heterogeneity here, as it influences the ease to access data for banks and online applications for borrowers.
Was there differential access to fast internet across locations? See D'Andrea et al. 2020 for results on how fast internet affects banks access to credit register information.

Other comments

- I see the experiment as related to Credit Register expansions, see Hertzberg et al. 2010, Choudary and Jain 2014, for example

- Some more discussion on the instruments used in the structural estimation

- Cluster std. errors: why also at year-quarter level? Bank level seems the correct dimension

Conclusion

- Very interesting paper on extremely important topic. I really enjoyed reading it.

- Can make important contribution to the literature.

- Thank you for your attention!