READ ME FILE

Title: The Role of Collateral in Borrowing

Authors: Nicholas Garvin, David W Hughes and José-Luis Peydró

Description

This 'read me' file provides information about the regression code and data construction for RDP 2021-01. Plotting data for all figures except Figure 4 are publically available and can be found in the spreadsheet 'rdp-2021-01-graph-data.xlsx'.

Regressions

The results reported in this RDP were generated using StataSE 13 and R version 3.6.1.

Each of the regression output tables, except for Table B3, is generated by one of the three accompanying Stata '.do' scripts.

These three scripts each read in a csv data file that contains the confidential data.

The R code that produces the coefficients in Table B3, from the data object 'bank_data' that contains bank-level variables, is:

```
summary(lm(Bnpl ~ Bsz + Blev + Bdom + Bw1btr, data=bank_data))
```

Data sources

See below for a description of the process for formatting the raw data sources into the regression dataset read by Stata.

The TED spread data for the two sample periods can be found in the 'TED spread 2008' and 'TED spread 2006' csv files and are sourced from the Federal Reserve Bank of St. Louis. The relevant RBA cash rate data were manually input into the code from the RBA website (https://www.rba.gov.au/statistics/cash-rate/). All other data are not publically available due to either confidentiality or third-party provider provisions.

For any questions, including more detail about the data formatting code, please contact the authors.

Process for transforming raw data into datasets read by Stata

Raw datasets

Each of the following datasets have a 2008 and a 2006 version:

1. Interbank loan data. This includes interbank repos and unsecured loans of maturities 8 days or less, initiated on 1 September or later and closed out no later than 31 October, excluding any loans involving accounts associated with RBA, ASX or International Central Securities Depositories. These datasets are produced by algorithms run on transaction-level data, and are described in detail in:

   • Brassil, Hughson and McManus (2016): RBA RDP 2016-11 Identifying Interbank Loans from Payments Data, and

   • Garvin (2018): RBA RDP 2018-09 Identifying Repo Market Microstructure from Securities Transactions Data.

   The Austraclear data require manually aligning the account names used by Austraclear with organisation names of the associated parent companies.

2. Collateral holdings data. These include securities (by ISIN) held each night in the Austraclear accounts of the entities (i.e. banks) in dataset (1).

3. Daily TED spread data from the Federal Reserve Bank of St. Louis.

4. Bank characteristics. For each bank: non-performing loans (as a proportion of total loans); total asset value; leverage; and domicile. These are sourced from a combination of publicly released reports and SNL Financial (now called S&P Global Market Intelligence). This dataset intentionally excludes BankWest, which had corporate activity during the sample.

5. Banks' open market operations (OMO). This includes banks' daily changes in OMO positions with RBA, at the level of bank*collateral type (AGS, semis or other), as implied by transactions between that bank and the RBA's OMO Austraclear account in Austraclear transactions data.

The process for creating the regression data is as follows. It is the same for the 2006 and 2008 datasets.

I. Format dataset (2) into a single measure of AGS holdings and a single measure of semis holdings for each bank in dataset (1):

   a. Identify any account names containing the character strings 'NOM', 'nom', 'Nom', 'CUST', 'cust', 'Cust', and/or 'client', and remove them. These strings are chosen based on what appears in the account names.

   b. At the bank*ISIN level, calculate the total face value of securities held on the night of 1 September, and repeat for the night of 8 September.

   c. At the bank*ISIN level, discard the higher of the face values on these two days.

   d. Identify ISINs issued by the Australian Government (AGS) and ISINs issued by organisations tied to the Australian state governments.

   e. At the bank level, sum all AGS ISINs from step (c), and all semis ISINs from step (c), so that each bank has one measure of AGS holdings and one measure of semis holdings.

II. Transform dataset (1), which is at the loan level, to a borrower*lender*day*market level dataset with loans outstanding between banks each night, and combine with other datasets.

   a. In the repo data in dataset (1), remove any loans outside the interest rates described in footnote 4 in the paper. There are no interest rate outliers in the unsecured data in dataset (1).

   b. Remove any loans for which the borrower and lender are the same. (Intrabank repos can occur because banks can have multiple Austraclear accounts operated by different desks within the Bank.)

   c. Create the lender*borrower*day*market level dataset (the 'regression dataset'). To do this, at the lender*borrower*market level, loop through each day in the sample, summing the value of all loans initiated that day and subtracting the value of all loans closed out that day. In many cases loans outstanding are zero.

   d. At the lender*borrower*market level, create a dummy variable for whether there are any positive loans outstanding on the days in the regression sample (i.e. four weeks from the first Monday in September). Add this dummy variable into the regression dataset. It is used to exclude any borrower*lender*market units that had no positive loans outstanding from regressions.

   e. Add in the lagged TED spread from dataset (3) into the regression dataset.

f.   For each observation, add variables for the borrower and lender characteristics, from dataset (4) and step (Ie), into the regression dataset. For borrowers or lenders without characteristics (e.g. client 'entities' and BankWest), these characteristics are left as missing values.

g.   Truncate the regression dataset by removing observations for which the day is outside the sample period (i.e. four weeks starting from the first Monday in September). It is important that the loans-outstanding dataset starts after and finishes before the loan-level dataset, to capture, for example, loans that are initiated before the loans-outstanding dataset sample starts, but continue to be open within some of the loans-outstanding sample range.

h.   From dataset (5), sum the changes in OMO positions to the bank*day*collateral level, where collateral is either government (AGS and semis) or 'other'. Then, for each bank*day*collateral observation, take the cumulative sum of all past observations, to get a measure of the outstanding OMO position for that day. Shift all of these observations up by adding an amount that makes the minimum observation for each bank equal to $1 million. This ensures outstanding OMO positions are never negative. Add these into the regression dataset.

III. Repeat most of step (II) to create a borrower*lender*day*collateral level regression dataset.

a.   Repeat the whole process in step (II) but replacing the references to the borrower*lender*market level with references to the borrower*lender*collateral level, where collateral is either AGS, semis, other or none (i.e. unsecured).

b.   Add in the weighted average interest rate for repos initiated that day. That is, for each borrower*lender*day*collateral level observation, identify the repos initiated that day, and take their weighted average spread to the cash rate, weighting the spreads by the value of the loan. Leave as a missing value if no repos were issued that day.

20 January 2021