Description of datasets and programs
Accompanying the article "Collateral booms and information depletion"
Published in Review of Economic Studies
Copyright (2021) Vladimir Asriyan, Luc Laeven and Alberto Martin

This file is called readme.pdf and is contained in the root directory of the zip file
InformationDepletion.zip, available from the REStud website. Once unzipped, the folder
"InformationDepletion" contains all the data and programs necessary to replicate our
results. The readme file contains the information needed to construct the intermediate and
final datasets (section 1) and instructions for replicating the results (section 2).

1. Construction of Intermediate and Final Datasets

This section describes the construction of the intermediate and final datasets used in the
empirical analysis. The final dataset is the same as in Chaney et al. (2012) with two
exceptions: we add measures of information and we expand the sample to 2016 in order to
cover the post-2007 housing bust. For details on the construction of the Chaney et al. (2012)
dataset, we refer to the description of their dataset available from the website of the
American Economic Review at http://doi.org/10.3886/E112551V1.

- COMPUSTAT: We collect information on firm financials from COMPUSTAT. As in
  Chaney et al. (2012), we start the sample in 1993 because the accumulated
  depreciation on buildings is not available in COMPUSTAT after 1993. We include
  firms headquartered in the United States and exclude firms operating in the
  construction, finance, insurance, real estate, and mining sectors. We keep only firms
  that appear at least three consecutive years in the sample. This leaves a sample of
  3,126 firms and 35,346 firm-year observations for the period 1993 to 2016.

  Market value of real estate assets. RE Value is the ratio of the market value of real
  estate assets normalized by the lagged value of Property, Plant and Equipment (PPE)
  (COMPUSTAT item No. 8). Real estate assets include buildings, land and
  improvement, and construction in progress. These assets are valued at historical
  cost. To impute their market value, we follow the procedure in Chaney et al. (2012),
  which calculates the average age of these assets and uses historical prices to
  compute their current market value. The ratio of the accumulated depreciation of
  buildings (COMPUSTAT item No. 253) to the historic cost of buildings (COMPUSTAT
  item No. 263) measures the fraction of the initial value of a building that has been
  depreciated. We impute the average age of real estate assets by assuming that these
  assets depreciate over 40 years, and we infer the market value of these real estate
  assets by inflating their historical cost with state-level residential real estate inflation
  after 1975, and CPI inflation before 1975. We use the headquarter location
  (COMPUSTAT variables STATE and COUNTY) as a proxy for the location of real estate.

  Control variables. We compute cash holdings as the ratio of cash flows (COMPUSTAT
  item 51 No. 18 plus item No. 14) to lagged PPE. Market-to-Book ratio is the total
  market value of equity divided by the book value of assets (COMPUSTAT item No. 6).

The market value of equity is calculated by multiplying the number of common stocks (COMPUSTAT item No. 25) by the year-end closing price of common shares (COMPUSTAT item No. 24) plus the book value of debt and quasi equity, computed as book value of assets minus common equity (item No. 60) minus deferred taxes (COMPUSTAT item No. 74). We use the one-year lagged value of the market-to-book ratio in the regression. Following Chaney et al. (2012), we include initial firm characteristics to control for potential firm heterogeneity. These controls, measured in 1993, are Return on Assets (operating income before depreciation (COMPUSTAT item No. 13) minus depreciation (COMPUSTAT item No. 14) divided by the book value of assets, Size measured as the natural logarithm of the book value of assets, Age measured as number of years since initial public offering (IPO), two-digit SIC codes and state of headquarters' location.

All variables defined in terms of ratios are winsorized at five times the interquartile range from the median.

- LPC DEALSCAN: As main measure of information, we compute the volume-weighted average length of the firm's main bank relationship, expressed in years, at the monthly level using data from LPC Dealscan. Because we need information on the history of loan transactions to construct a measure of lending relationships over our sample period, we use Dealscan data starting in 1985 which is the first year with adequate coverage in the Dealscan dataset. We restrict the sample to US borrowers and syndicated loans issued in US dollars with a defined facility amount and maturity. Following Sufi (2007), we define a lender as lead lender if the variable "Lead Arranger Credit" takes on the value of "Yes," and if the lender is the only bank specified in the loan deal.

  As syndicated loan contracts often consist of multiple tranches, each with at least one lead lender, it is common for multiple banks to be registered as lead banks on the same deal. In such cases, we select the "main" lead bank in two steps. First, we filter for the lead banks whose contracts offer the longest loan maturity. Second, we choose among these banks the ones with the largest amount pledged. In those cases where this algorithm leads to multiple "lead bank-borrower pairs", we treat those as distinct syndicated loans.

  As "lead bank-borrower pairs" interact repeatedly with each other, it is necessary to evaluate information production over the pairs' entire relationship history. We compute the duration of the lending relationship as the difference between the pairs' latest loan contract expiration date and the earliest loan contract signing date, expressed in years. However, when borrowers switch lead banks, we reset this variable to zero for all bank-borrower pairs without active relationships (i.e., no credit outstanding). Moreover, this variable drops to zero whenever the last loan contract in our sample expires and there are no new lending relationships. To smooth this transition, in such cases we set this variable equal to its last positive observation for up to three more years. To aggregate this relationship variable at the bank-firm level into an information measure at the firm level, the relationship

measure is volume-weighted by the respective amount pledged for each "lead bank-borrower pair" relative to the total loan amount received by each borrower.

Loan amounts are expressed in real terms using the US GDP deflator obtained from the US Bureau of Economic Analysis. We follow the linking procedure in Chava and Roberts (2008) to merge our relationship measure based on Dealscan data to COMPUSTAT. The linking file is available on Wharton Research Data Services (WRDS) and can be found at: Thomson Reuters --- WRDS-Reuters Dealscan --- Compustat Linking Table.

- I/B/E/S: We construct the analyst coverage variable using data on the number of analysts who make annual earnings forecasts for a firm in a given month using data from the I/B/E/S Historical Summary Files. We define the Analyst variable as the maximum number of analysts who make annual earnings forecasts for a given firm in any month during the year. The Analyst variable is then matched to the main dataset by linking the I/B/E/S CUSIP code via the CRSP database, both available in WRDS, to COMPUSTAT.

- OFHEO and FRED: We use data on residential real estate prices, both at the state and at the MSA level. Residential real estate prices come from the Office of Federal Housing Enterprise Oversight (OFHEO). The OFHEO Home Price Index (HPI) is a broad measure of single-family home prices in the United States. We match the state level HPI to our main dataset using the state identifier from COMPUSTAT. To match the MSA level HPI, we link Federal Information Processing Standards codes from COMPUSTAT to MSA identifiers using the correspondence table obtained from OFHEO.

Following Chaney et al. (2012), we instrument local real estate prices using the interaction of long-term interest rates and local housing supply elasticity. Local housing supply elasticities for a total of 95 MSAs are obtained from Saiz (2010). These elasticities capture the amount of local land that can be developed and are estimated using satellite-generated images of the terrain. We measure long-term interest rates using the 30-year conventional mortgage rate from the Federal Reserve's FRED database. Mortgages rates are adjusted for inflation using the CPI inflation rate from the Federal Reserve's FRED database.

2. Instructions for replicating results:

In order to replicate our results in the regression tables, follow these steps in Stata:

- APPENDIX TABLE 2: Run the Stata do file entitled "ALM_FirstStage.do" together with the data file "ALM_firststage.dta"

- TABLE 1 and APPENDIX TABLES 3, 4, 5, and 6: Run the Stata do file entitled "ALM.do" together with the data file "ALM.dta".

References:

Chaney, Thomas, David Sraer, and David Thesmar, 2012. "The Collateral Channel: How Real Estate Shocks Affect Corporate Investment," American Economic Review, 102(6), 2381–2409, https://doi.org/10.1257/aer.102.6.2381.

Chava, Sudheer, and Michael R. Roberts, 2008. "How Does Financing Impact Investment? The Role of Debt Covenants," Journal of Finance, 63, 2085–2121, https://doi.org/10.1111/j.1540-6261.2008.01391.x.

Saiz, Albert, 2010. "The Geographic Determinants of Housing Supply," The Quarterly Journal of Economics, 125(3), 1253–1296, https://doi.org/10.1162/qjec.2010.125.3.1253.

Sufi, Amir, 2007. "Information Asymmetry and Financing Arrangements: Evidence from Syndicated Loans," The Journal of Finance, 62(2), 629–668, https://doi.org/10.1111/j.1540-6261.2007.01219.x.