Voluntary Compliance to Journal of Accounting Research Data Policy

Biqin Xie

1. A description of which author(s) handled the data and conducted the analyses.

Data handling and programming were performed solely by the author.

2. A detailed description of how the raw data were obtained or generated, including data sources, the date(s) on which data were downloaded or obtained, and the instrument used to generate the data (e.g., for surveys or experiments). We recommend that more than one author is able to vouch for the stated source of the raw data.

The data sources and the years in which data were downloaded or obtained are as follows:

1) Data on residential mortgage applications and approval/denial decisions are obtained from the Home Mortgage Disclosure Act (HMDA) database. The HMDA data for the 2007-2012 period are downloaded from https://www.ffiec.gov/hmda/hmdaraw.htm; the HMDA data for the 1997-2006 period are downloaded from the National Archives website. The HMDA data were obtained in 2013.


3) Data on the identifier of commercial banks and that of their parent bank holding companies are obtained from Call Report downloaded from https://cdr.ffiec.gov/public/PWS/DownloadBulkData.aspx. I downloaded the data in 2013.

4) Risk-free rate and CPI data are from CRSP on WRDS. I first obtained data for the 1997-2010 period in 2010 and 2011, and then added data for the 2011-2012 period in 2013.

5) Data on term spread, default spread, and real GDP growth are obtained from the St. Louis Fed website. I first downloaded data for the 1997-2010 period in 2010 and 2011, and then added data for the 2011-2012 period in 2013.

I vouch for these data sources.

3. If the data are obtained from an organization on a proprietary basis, the authors should privately provide the editors with contact information for a representative of the organization who can confirm data were obtained by the authors. The editors would not make this information publicly available. The authors should also provide information to the editors about the data sharing agreement with the organization (e.g., non-disclosure agreement, any restrictions imposed by the organization on the authors with respect to publishing certain results).
Except for the other-than-impairment data for 2008 and 2009 that I hand collected from 10-Ks, all the other data are obtained from the public databases as described above.

4. A complete description of the steps necessary to collect and process the data used in the final analyses reported in the paper. For experimental papers, we require information about subject eligibility and/or selection, as well as any exclusion criteria.

Part I. I follow the steps detailed below to obtain the loan-level annual sample employed to test the relation between fair value accounting and bank lending.

Step 1: working with bank financial data.

I hand collect other-than-impairment data for 2008 and 2009 from 10-Ks of all publicly listed banks, and merge them with other bank financial data from FR Y-9C reports. The main identifier for bank holding companies in the bank financial data is RSSD9001 from FR Y-9C reports.

Step 2: working with mortgage application data from the HMDA database and merging them with bank financial data.

1) I use two datasets from the HMDA database: the Loan Application Register (LAR) dataset for raw data on residential mortgage applications and banks’ approval/denial decisions, and the Reporter Panel dataset for banks’ identification information.

2) The matching between commercial-bank lenders in the HMDA database and their parent bank holding companies in FR Y-9C reports is achieved by using bank identifier information in Reporter Panel and that in Call Report. The matching process is as follows:

   Step (a). I obtain RSSD9348 (i.e., RSSD ID of the regulatory high holder) and RSSD9364 (i.e., ID of the financial high holder) from Call Report. I define RSSDHH as being equal to RSSD9348 (RSSD9364) when RSSD9348 is nonmissing (is missing). For the 2010-2012 period, I merge Call Report and Reporter Panel by using the fact that RSSDHH corresponds to “Top Holder RSSD ID” in Reporter Panel. For the 2004-2009 period when “Top Holder RSSD ID” is not available in Reporter Panel, I merge Call Report and Reporter Panel by exploiting the fact that RSSD9001 in Call Report corresponds to “Respondent RSSD ID” in Reporter Panel. For periods prior to 2004, Reporter Panel contains neither “Top Holder RSSD ID” nor “Respondent RSSD ID”; thus, I take the RSSD9001- Respondent RSSD ID matching in 2004 and apply them to the 1997-2003 sample period.

   Step (b). I merge the resulting dataset from step (a) with the LAR dataset by requiring observations from the two datasets to have the same “Respondent ID” and the same “Agency Code”.
Step (c). I merge the resulting dataset from step (b) with the bank financial data obtained from step 1 above by exploiting the fact that the RSSDHH in the resulting dataset from step (b) corresponds to RSSD9001 from FR Y-9C reports.

3) I limit mortgage applications to those with clear actions of approval (i.e., “Actions” =1 or 2) or denial (i.e., “Action” = 3). I define the bank-year level total number of applications as the sum of the number of applications approved and the number of applications denied, and define the bank-year level denial rate as the number of applications denied divided by the bank-year level total number of applications. I plug the bank-year level total number of applications and the bank-year level denial rate into the Cochran (1977) formula to determine the optimal size of the stratified random sampling. I then perform stratified random sampling using each bank-year as a stratum, and thereby obtains a stratified random sample of mortgage applications for each bank year.

4) I create a mapping between the “old” MSAs (i.e., MSAs during 1997-2003) and the “new” MSAs (i.e., MSAs during 2004-2012), which is necessitated by a revision of MSAs that occurred in 2003. I use the mapping data from the NBER website (http://www.nber.org/data/cbsa-msa-fips-ssa-county-crosswalk.html) to map between the “old” MSAs and the “new” MSAs, retaining the “old” MSAs for which I cannot find a “new” MSA match.

Step 3: imposing sample selection criteria.

In addition to require sample bank holding companies to have nonmissing values for variables needed in the analyses, I also require them to satisfy the following criteria: (1) home loans represent at least 25% of the bank’s loan portfolio; (2) the bank has at least 200 mortgage applications in the stratified random sample in a given sample year; (3) the bank has total assets of at least $500 million in 2006 dollars; and (4) the bank’s quarterly asset growth does not exceed 10%.

Step 4: winsorization.

All continuous variables are winsorized at the top and bottom 99% by year.

Part II. I follow the steps detailed below to obtain the bank-level quarterly sample used to analyze whether unrealized AFS gains and losses are procyclical.

Step 1: I take the universe of bank holding companies that file FR Y-9C reports over the 1994Q2-2013Q4 period.

Step 2: I keep all bank holding companies with nonmissing variables needed for the analysis.

5. Prior to final acceptance of the paper, the computer program used to convert the raw data into the dataset used in the analysis plus a brief description that enables other researchers to use this program. Instead of the program, researchers can provide a detailed step-by-step description that enables other researchers to arrive at the same dataset used in the analysis. The purpose of this
requirement is to facilitate replication and to help other researchers understand in detail how the sample was formed, including the treatment of outliers, Winsorization, truncation, etc. This programming is in most circumstances not proprietary. However, we recognize that some parts of the data generation process may indeed be proprietary or otherwise cannot be made publicly available. In such cases, the authors should inform the editors upon submission, so that the editors can consider an exemption from this requirement.

I use SAS to convert the raw data into the datasets used in the analyses and use Stata to run the analyses. The detailed step-by-step description on how I arrive at the final datasets is included in part 4 above.

6. **Data and programs should be maintained by at least one author (usually the corresponding author) for at least six years, consistent with National Science Foundation guidelines.**

I will maintain the data and programs for at least six years.