Integration Among US Banks: Data and Methodology Guide

Abhinav Anand and John Cotter

2018/09/03

Data

The starting period for our sample is January 1 1993 and the ending period is December 31 2016.

Our sample consists of all US depository credit institutions and bank holding companies for which data are available in both CRSP and Compustat. We use the WRDS interface to collect data from both sources.

Sample Construction

Finance and insurance related entries are distributed between SIC codes 6000–6799. The following table illustrates this classification in more detail.[1]

Table 1: SIC codes for finance and related industry groups

SIC Code	Industry
6012	Pay Day Lenders
6021	National Commercial Banks
6022	State Commercial Banks
6029	Commercial Banks, NEC
6035	Savings Institution, Federally Chartered
6036	Savings Institutions, Not Federally Chartered
6099	Functions Related To Depository Banking, NEC
6111	Federal & Federally Sponsored Credit Agencies

[1]The full list can be found on the link: http://www.ehso.com/siccodes.php

SIC Code	Industry
6141	Personal Credit Institutions
6153	Short-Term Business Credit Institutions
6159	Miscellaneous Business Credit Institution
6162	Mortgage Bankers & Loan Correspondents
6163	Loan Brokers
6172	Finance Lessors
6189	Asset-Backed Securities
6199	Finance Services
6200	Security & Commodity Brokers, Dealers, Exchanges & Services
6211	Security Brokers, Dealers & Flotation Companies
6221	Commodity Contracts Brokers & Dealers
6282	Investment Advice
6311	Life Insurance
6321	Accident & Health Insurance
6324	Hospital & Medical Service Plans
6331	Fire, Marine & Casualty Insurance
6351	Surety Insurance
6361	Title Insurance
6399	Insurance Carriers, NEC
6411	Insurance Agents, Brokers & Service
6500	Real Estate
6510	Real Estate Operators (No Developers) & Lessors
6512	Operators of Nonresidential Buildings
6513	Operators of Apartment Buildings
6519	Lessors of Real Property, NEC
6531	Real Estate Agents & Managers (For Others)
6532	Real Estate Dealers (For Their Own Account)
6552	Land Subdividers & Developers (No Cemeteries)
6770	Blank Checks
6792	Oil Royalty Traders
6794	Patent Owners & Lessors
6795	Mineral Royalty Traders

SIC Code	Industry
6798	Real Estate Investment Trusts
6799	Investors, NEC

We collect daily price and return data from the CRSP database for all entities whose SIC codes lie between 6020–6079 or from 6710–6712 between the above mentioned dates. Commercial banks lie between SIC codes 6020–6029, saving institutions between 6030–6039, credit unions between 6060–6069; and bank holding companies between 6710–6712. The SIC code ranges $\{6020, \ldots, 6079\} \cup \{6710, 6711, 6712\}$ are referred to henceforth as 'admissible' SICs.

We include only common stocks corresponding to codes 10 and 11 and remove all American Depository Receipts (ADRs) and firms incorporated in non-US countries. We further delete all entities with nominal stock prices less than \$1. For firms whose SIC classification changes from admissible to inadmissible or vice versa, we include them only for the time duration corresponding to their status as admissible banks. Since we include all such banks irrespective of whether they are alive or not, our study is free from survivorship bias.

Finally we include only those US banks whose total assets in 2016 are at least \$1 billion according to data collected from Compustat. This leaves us with a final sample of 349 unique banks.

Our attention on public banks with primary listings in the US excludes several multinational banking corporations which might have secondary listings in the US but primary listings elsewhere. For example, the British bank HSBC has a secondary listing on the New York Stock Exchange but under our definition, we do not include it in the list of US banks. In the same way, financial service providers such as mutual funds, insurance companies etc. are not included in our definition of banks. Since the focus of our paper is to isolate and study integration dynamics of US banks, inclusion of European or Asian banks with secondary listings in the US may bias our estimates.

Methodology

Frequency of Estimation of Bank Integration

Our sample stretches from January 1 1993 to December 31 2016. We partition the duration of the study into quarters and compute bank integration each quarter from daily bank returns. Hence the integration estimates start from 1993 Quarter 1 to 2016 Quarter 4—a total of 96 quarters.

Under this setup, there are between 62–66 daily observations for each bank's return each quarter. We compute the covariance matrix of the 338 US banks each quarter and extract as many principal components each quarter as are necessary to explain 90% of bank returns. For banks which do not contain data for the entire sample period, we start estimating their integration levels from the time their data begin appearing in CRSP.

Construction of Principal Components

Principal components each quarter are constructed from the covariance matrices of all banks with available returns in a particular quarter. For example, at the beginning of the sample there are only 43 banks with available returns and hence the size of the covariance matrix from which principal components are extracted is 43×43. As time progresses, the coverage of banks increases steadily so that by the end of the sample (quarter 96) there are 234 admissible banks and hence the corresponding covariance matrix has dimensions 234×234.

In constructing covariance matrices, we first remove any bank which has no available returns for the entire quarter. Further we also remove banks for which stale returns and missing values exceed a threshold of 30% quarterly observations.[2] This takes care of most of the banks in our sample. For the remnant few banks with leftover missing values, we replace them with their banks' respective quarterly medians and then compute the covariance matrix.

[2]Since there are 62–66 observations each quarter, this means that if a bank in question has greater than 22 missing and/or stale entries, we remove it from the construction of the covariance matrix.

Out-of-Sample Principal Component Construction

Once eigenvectors of the covariance matrices are computed in order of largest to smallest eigenvalue, out-of-sample principal components are estimated by applying them to observed returns for the subsequent quarter. For example, eigenvectors from the covariance matrix in 1993Q1 are applied to the covariance matrix in 1993Q2 to obtain principal components in 1993Q2. Hence there are overall 95 such quarterly principal component computations—from 1993Q2 to 2016Q4.

Each quarter we collect as many principal components as necessary for explaining 90% of returns. Hence the number of principal components used differs from quarter to quarter.