Overview of the Steps used for the Analyses in Bushman, Hendricks and Williams (2016) 
Voluntary Compliance with JAR’s New Data Policy

This document details the steps necessary to conduct the analyses performed by Bushman et al. (2016). The data was handled by two of the coauthors (Hendricks and Williams) and one RA (Tianshu Qu). The data sources included Dealscan, Edgar, Bank Compustat, the Bank Regulatory filings, CRSP, Federal Reserve Bank of New York, and the CBOE. SAS was the primary program used to handle the extraction, manipulation, and cleaning of the data. STATA was then used for the regression analyses and to provide descriptive statistics. Perl was used to extract regulatory filings from the SEC Edgar database and run the word counts necessary to construct the text-based measure of competition. Our sample period runs from 1996 (the first year of Edgar filings) to 2010. The data for the paper was extracted summer of 2012. In addition to the above this document contains all the required items under the new data policy at JAR.

**Steps used to compute the BCE measure**

1. We download all 10-k filings from the SEC’s EDGAR database in May 2012.

2. We follow the instructions outlined in Section 3 of Li et al., (2013). Specifically, we use textual analysis to count the number of occurrences of “competition, competitor, competitive, compete, competing,” including those words with an “s” appended, and then remove any case where “not,” “less,” “few,” or “limited” precedes the word by three or fewer words. To control for 10-k length, we scale the number of competition-words by the total number of words in the 10-k.

3. To encourage future use of this textual-analysis measure of competition, we have made the calculation of the measure publicly available for all financial firms during the years of our sample (https://sites.google.com/sites/christopherwilliamsphd/Home/research). We also include GVKEY in that file so that it can be quickly linked with other information about these financial firms.

**Steps used for the contracting results**

1. Obtain the Dealscan database (Accessed in May, 2012). Remove all deals that were not completed, not denominated in US currency, or not the original loan.

2. Determine the number of covenants attached to each syndicated loan package.

   a. Combine the two lists of covenants (net worth & financial) into a single list of covenants.

   b. Create a new list of all package id numbers from the newly created list of covenants.

   c. At the package level, create indicator variables for each of the 24 different loan covenants indicated in the net worth and financial covenant lists. There were a total of 4 (20) covenants included in the net worth (financial) covenant list.

   d. Sum the number of the 24 covenants attached to each package id.

3. Gather both the lender and loan information for each loan included in the syndicated loan package.
a. Obtain the detailed loan information (maturity, secured, loan type, facility ID) from the “Facility” database (included in the Dealscan database). We used Package ID to link the two information sets.

b. Obtain the spread charged on the loans by using the “Currfacpricing” database (included in the Dealscan database). We defined spread as the average basis points over LIBOR charged on all of the loan facilities within a syndicated loan package.

c. Obtain information about each lender and the syndicate (lender id, role in syndicate, allocated percentage of syndicate) by using the “Lendershares” database (included in the Dealscan database). We used the Facility ID number to link the two information sets.

d. Consistent with the lending decision being made by the lead arranger of the syndicate, we deleted all lenders that were not indicated to be a lead arranger in the lending syndicate. Specifically, we retained only those lenders that had been identified as the lead arranger in the leadarrangercredit field.

4. Gather detailed information for both the borrowers and lenders included in our sample.

a. Use the Dealscan (Facility ID) to Compustat (GVKEY) linking table for each of the Facility ID numbers included in Dealscan. Because Dealscan does not provide extensive financial data about borrowers, this linking table is necessary in order to account for the effects of the borrower’s information in our analyses.

   i. We obtained detailed historical financial information from Compustat for each borrower in our sample (accessed in May, 2012). For our analyses, in all instances, we use the most current balance sheet data and income statement data prior to the loan origination date. All income statement data was annualized by considering the trailing twelve months.

   ii. We use the code provided by Bharath & Shumway (2005) to calculate each borrower’s expected default frequency (EDF) at the time of loan origination. This code required us to obtain monthly observations for each firm’s debt (both current and long-term) from Compustat. The calculation also required daily stock prices, shares outstanding, and monthly risk-free rates which were all obtained from CRSP (accessed in May, 2012).

b. Create a Dealscan (Company ID) to Compustat (GVKEY) linking table for each of the Company ID numbers included in Dealscan. Because Dealscan does not provide extensive financial data about borrowers, this linking table is necessary in order to account for the lender’s information in our analyses.

   i. We obtained detailed historical financial information from Compustat for each lender in our sample (accessed in May, 2012). For our analyses, in all instances, we use the most current balance sheet data and income statement data prior to the loan origination date. All income statement data was annualized by considering the trailing twelve months.

c. Use GVKEY to link to the textual-analysis measure of competition (BCE) to each lender. We use the most recent BCE value, calculated from firms’ 10-k filings, at the time the
loan was originated. As noted above, we have made this data publicly available to encourage others to use it in future research.

5. We remove all observations that did not have complete information for our analyses. Specifically, we removed all observations that were missing any of the following variables: bank assets, bank tier 1 capital, bank competition (BCE), borrower assets, borrower z-score, borrower edf, loan amount, loan maturity, loan spread, or loan covenants. Because we include borrower and bank fixed-effects in our analyses, we also remove all borrowers and lenders that had only one observation in our sample.

6. We winsorize by setting all extreme values, at the level of the entire pooled sample, of the dependent and independent variables to the 1st and 99th percentiles.

**Steps used in estimating the Accrual Choices results**

1. We start with the bank Compustat universe and restrict the sample to the consolidated financials that have gvkey, permco and permno. We also require the bank’s permno to match to a RSSDID from the Federal Reserve Bank of New York’s permco/RSSDID matching table.

2. From Bank Compustat we collect the following variables: Loan loss provision (pllq), Non-Performing Loans (npatq), Earnings before the loan loss provisions and tax (piq + pllq), Loans (Intalq), total assets (atq), and Tier 1 capital (capr1q). We use date and RSSDID to merge this information with the bank regulatory filings (Y-9c) data. Specifically, we obtain the loan portfolio characteristics: commercial loans (BHCK 1766), Consumer loans (BHCK 1975), and Real Estate loans (BHCK 1410).

3. We use GVKEY and lagged datadate to merge this data with the BCE measure (calculated as described above).

4. We compute each of the variables, as described in Appendix C of our paper, required to estimate Equation (4). We winsorize by setting all extreme values, at the level of the entire pooled sample, of the dependent and independent variables to the 1st and 99th percentiles.

**Steps used in estimating the Revenue Mix and Tier 1 Capital results**

1. We start with the bank Compustat universe and restrict the sample to the consolidated financials that have gvkey, permco and permno. We also require the bank’s permno to match to a RSSDID from the Federal Reserve Bank of New York’s permco/RSSDID matching table.

2. We collect the following variables from Bank Compustat: Total deposits (dptcq), Earning before extraordinary items (ibq), Loans (Intalq), Non interest income (TNIIQ), Interest income (IDITQ), deposit service charges (scdaq), trading income (TAIQ), total assets (atq), non-interest expense (xnitbq), and Tier 1 capital (capr1q). We also use bank RSSDID to merge the following portfolio characteristics from bank regulatory filings (Y-9c): commercial loans (BHCK 1766), Consumer loans (BHCK 1975), Real Estate loans (BHCK 1410). We calculate the maturity mismatch variable as (atq-ceqq-dlttq-cdbtq)/(atq-ceqq).
3. We use GVKEY and lagged datadate to merge this data with the BCE measure (calculated as described above).

4. We compute each of the variables, as described in Appendix C of our paper, required to estimate Equation (5). We winsorize by setting all extreme values, at the level of the entire pooled sample, of the dependent and independent variables to the 1st and 99th percentiles.

**Steps used in estimating the Tier 1 Capital results**

1. We start with the bank Compustat universe and restrict the sample to the consolidated financials that have gvkey, permco and permno. We also require the bank’s permno to match to a RSSDID from the Federal Reserve Bank of New York’s permco/RSSDID matching table.

2. We collect the following variables from Bank Compustat: Trading account assets (tdstq), Earning before extraordinary items (ibq), Loans (Intalq), total assets (atq), book equity (ceqq) and Tier 1 capital (capr1q). We also use RSSDID to merge this data with the following portfolio characteristics from bank regulatory filings (Y-9c): commercial loans (BHCK 1766), Consumer loans (BHCK 1975), Real Estate loans (BHCK 1410). We calculate the maturity mismatch variable as (atq-ceqq-dlttq-cdbtq)/(atq-ceqq).

3. We use CRSP to extract data needed to calculate each bank’s market value of equity. Specifically, we require price (prc) and shares outstanding (shrout). We convert the shares outstanding into millions and multiply the absolute value of price times the number of shares outstanding (in millions).

4. We perform the following steps to compute the market beta:
   a. From CRSP, we extract daily stock returns (ret) over the quarter for each bank in our sample that has all of the required accounting variables (listed above in Step 2).
   b. For each day, we pull the CRSP value weighted stock return (vwretd).
   c. We require a bank to have a minimum of 40 trading days with non-missing stock returns to be included in the regression.
   d. We winsorize each bank’s quarterly time series of returns by at the 1st and 99th percentiles.
   e. We then regress the bank’s daily return on the market’s value-weighted return over the quarter to obtain an estimate of the bank’s beta.

5. We use GVKEY and lagged datadate to merge this data with the BCE measure (calculated as described above).

6. We compute each of the variables, as described in Appendix C of our paper, required to estimate Equation (6). We winsorize by setting all extreme values, at the level of the entire pooled sample, of the dependent and independent variables to the 1st and 99th percentiles.
**Steps used to compute the Loan Charge-off results**

1. We start with the bank Compustat universe and restrict the sample to the consolidated financials that have gvkey, permco and permno. We also require the bank’s permno to match to a RSSDID from the Federal Reserve Bank of New York’s permco/RSSDID matching table.

2. We collect the following variables from Bank Compustat: Earning before extraordinary items (ibq), Loans (Intalq), total assets (atq), book equity (ceqq), Non-Performing Loans (npaq) and Tier 1 capital (capr1q). We also use RSSDID to merge this data with the following portfolio characteristics from bank regulatory filings (Y-9c): commercial loans (BHCK 1766), Consumer loans (BHCK 1975), Real Estate loans (BHCK 1410), and loan charge-offs (BHCK4635).

3. We sum the quarterly loan charge-off over the subsequent 12 and 24 months to compute the 12-month and 24-month loan charge offs.

4. We use GVKEY and lagged datadate to merge this data with the BCE measure (calculated as described above).

5. We compute each of the variables, as described in Appendix C of our paper, required to estimate Equation (7). We winsorize by setting all extreme values, at the level of the entire pooled sample, of the dependent and independent variables to the 1st and 99th percentiles.

**Steps used to compute the VaR^4 measure**

1. Beginning in 1985, we use the intersection of banks from both Compustat and CRSP that have non-missing permnos and gvkeys. We also require the stock trading to be ordinary common shares (shrcd = 10, 11).

2. We merge in banks’ daily market value of equity using the price (prc) and shares outstanding (shrout) from CRSP and compute the weekend market value of equity for each calendar week.

3. We merge in the value weighted market return from CRSP (vwretd), 10-year BAA bond rate, 10-year T-bond rate, 3-month T-bill rate, the 3-month Libor rate, the VIX and the repo rate for each calendar week.

4. We merge in quarterly data from bank Compustat on total liabilities (ltq), total assets (atq) and book equity (ceqq). We require that the ratio of total assets to book equity be greater than 0 and less than 100.

5. Using data from both SCD and the Federal Reserve Bank of Chicago, we tag weeks during the quarter where there were either mergers or spin-offs. For the quarters where there is either a spin-off or merger, the quarter is dropped from the estimation.

6. We use quarterly financial accounting data for each bank to linearly interpolate the leverage ratios for the each of the weeks within the fiscal quarter.

7. We take the weekly accounting variables and market variables and compute a weekly asset return. To compute the weekly asset return we multiply the interpolated book value of assets to book value of equity ratio by the market value of equity. We then divide that number by the
book value of assets divided by the book value of equity multiplied by the market value of equity at the beginning of the calendar week.

8. We then compute the change in the 3 month T-bill (federal Reserve bank of New York), the yield slope (the difference between the 10-year and the 3 month rate), the credit spread (the difference between the BAA bond yields and the 10-year T-Bill), the weekly value weighted market return from CRSP, the weekly value weighted real estate (sic code 65-66) sector return in excess of the market value weighted return from CRSP, the VIX from the CBOE, and the liquidity spread (the difference between the 3-month general collateral repo rate (Bloomberg) and the 3-month T-Bill rate (federal Reserve bank of New York)).

9. We delete week observations with no assets returns. We then estimate the quartile regressions by each bank over the entire time series. We require a minimum of 260 bank week observations in order to be included in the sample.

10. We estimate three quantile regressions at the 1%, 50% and 99% range. Specifically, we place the asset return as the dependent variable and the macro state variables as the independent variable. We store time-series of estimated coefficients for each bank.

11. Taking the bank’s estimated coefficients for weekly observations, we compute the predicted asset return for each set of coefficients from the 1%, 50% and 99 percentiles. We then sum up the weekly-predicted values for each quantile of interest over the quarter of interest. This gives us the VaR^A metric for the 1%, 50% and 99%.

**Steps used to compute the VaR^E measure**

1. To compute the VaR^E metric, we follow the same steps as the VaR^A computation substituting stock returns over the week instead of asset returns over the week.

**Steps used to compute the ΔCoVaR^A measure**

1. To compute the ΔCoVaR^A measure we begin by following steps 1-7 from the computation of VaR^A.

2. For each calendar week, we create a value-weighted asset return for all the banks in the system excluding the bank of interest.

3. We estimate a quantile regression at the 1st percentile of the market asset returns computed in step 2, on the same macro state variable described in step 8 from the computation of VaR^A. We also include the bank of interest’s asset return as an additional independent variable.

4. We compute weekly ΔCoVaR^A as the predicted value from the estimated parameters using the difference between the VaR^A at the 1st percentile and 50th percentile for the bank of interest.

5. We sum the weekly ΔCoVaR^A to arrive at the quarterly value for the bank of interest.
Steps used to compute the $\Delta \text{CoVaR}^E$ measure

1. To compute the $\Delta \text{CoVaR}^E$, we follow the same steps as in the computations for $\Delta \text{CoVaR}^A$ substituting stock returns and the $\text{VaR}^E$ at the 1st percentile and 50th percentile as the input variables.

Steps used to compute the MES measure

1. We begin by taking the universe of bank stocks that are in the two-digit SIC codes 60-62.

2. For each stock and year, we extract the daily stock returns (ret) from CRSP. We also pull each bank’s stock the price (prc) and shares outstanding (shrout) to compute the bank’s total market value of equity each trading day.

3. We compute a daily value-weighted return for the banking industry using all the banks in our sample excluding our bank of interest.

4. For each year, we mark the days in which the banking sector’s return is in the bottom 5% for the year.

5. We compute the average return for each individual bank on the days in which the rest of the market’s value weighted return is in the bottom 5% for the year.

Steps used to compute control variables for VaR, CoVaR and MES Results

1. We use a common set of control variables for our VaR, CoVaR and MES analyses.

2. We begin by merging in from Bank Compustat the following variables: Trading account assets (tdstq), Total deposits (dptcq), Earning before extraordinary items (ibq), total assets (atq), book equity (ceqq) and the maturity mismatch variable (calculated as: $\frac{(\text{atq}-\text{ceqq}-\text{dlttq}-\text{cdbtq})}{(\text{atq}-\text{ceqq})}$).

3. We use CRSP to extract data needed to calculate each bank’s market value of equity. Specifically, we require price (prc) and shares outstanding (shrout). We convert the shares outstanding into millions and multiply the absolute value of price times the number of shares outstanding (in millions).

4. We use RSSDID and date to merge in data on banks’ loan portfolio characteristics from the bank regulatory reports. Specifically we merge in: commercial loans (BHCK 1766), Consumer loans (BHCK 1975), and Real Estate loans (BHCK 1410).

5. We next include a measure of the bank’s beta. We follow these steps to compute the variable:
   a. For each of the banks in the sample that we have accounting data for we pull from CRSP their daily stock returns (ret) over the quarter.
   b. For each day we pull the CRSP value weighted stock return (vwretd).
   c. Before estimating the regression we require a bank to have a minimum of 40 trading days with non-missing stock returns.
   d. We winsorize each bank’s quarterly time series of returns by at the 1 and 99 percentiles.
e. We then regress the bank’s daily return on the market’s value weighted return over the quarter.

f. The estimated coefficient on the market’s value weighted return is used as the estimate of the bank’s beta.

6. We compute a measure of return volatility by performing the following:
   a. We extract from CRSP daily stock returns (ret) during the quarter.
   b. We require a bank to have a minimum of 40 trading days with non-missing stock returns.
   c. We then take the standard deviation of each bank’s daily returns over the fiscal quarter.

7. We compute a measure of illiquidity as follows:
   a. We extract from CRSP each bank’s daily returns(ret), stock price (prc) and the traded volume (vol) for during the quarter.
   b. We require a bank to have a minimum of 40 trading days with non-missing stock returns.
   c. For each day, we compute the ratio of the absolute value of the bank’s return divided by the product of the bank’s price multiplied by traded volume. We use the average value of this ratio over the bank quarter as the measure of liquidity.

8. We use GVKEY and lagged datadate to merge this data with the BCE measure (calculated as described above).

9. We winsorize by setting all extreme values, at the level of the entire pooled sample, of the dependent and independent variables to the 1st and 99th percentiles for Equations (9), (11), and (12).