

Incorporating Compustat and CRSP in the BS Finance Curriculum

Leslie Boni, California State University, Monterey Bay
Pat Tinsley McGill, California State University, Monterey Bay
Marylou Shockley, California State University, Monterey Bay
Christina Zhang, California State University, Monterey Bay

ABSTRACT

We describe a series of projects for the undergraduate finance course curriculum that require students to analyze large financial databases (COMPUSTAT and CRSP) to make real-world decisions and recommendations. Students use the databases to obtain financial data about companies of interest to them. They evaluate large amounts of quantitative data on past and current operating performance using company financial statements and Securities and Exchange Commission data (via the COMPUSTAT database) as well as the risk and return associated with investments in the company's securities (via the CRSP database). Analyses include comparison of the company to its competitors and forecasting future performance. Final written reports and oral presentations are suitable for presentation to perspective employers. The projects further students' learning outcomes at multiple levels: students produce business writing that meets professional standards, produce critical analysis of business scenarios, use relevant resources to draw conclusions, and analyze data using quantitative tools to support business analysis. Our projects contribute to our university's commitment to preparing students to be competitive in the today's technology-focused and data-driven workforce as well as serving underrepresented minority (URM) students. We believe our work is also particularly timely given the AACSB's recent addition of a section on "Technology Agility" to accreditation Standard 9.

INTRODUCTION

We describe incorporating two large financial databases – COMPUSTAT and CRSP – in the undergraduate finance course curriculum. These databases, often used by researchers and in PhD and Master of Finance programs, are rarely used in undergraduate programs.

Three motivations inspire this work. First, our university, California State University, Monterey Bay (CSUMB), located just 65 miles south of Silicon Valley, is committed to preparing students to be competitive in the today's technology-focused and data-driven workforce. CSUMB's founding vision statement includes the following: "The university will invest in preparation for the future through integrated and experimental use of technologies as resources to people, catalysts for learning, and providers of increased access and enriched quality learning." The skills, knowledge, and confidence gained from analyzing real-world problems using large databases increases students' chances of obtaining better entry-level positions, succeeding in those

positions, or being accepted by leading graduate programs to advance their educational goals.

Second, in undergraduate business programs, students often struggle in quantitatively-focused courses. Some of them need to repeat the courses multiple times. This is more prevalent among underrepresented minority (URM) students. For example, our university's undergraduate core finance course has one of the largest URM achievement gap among its business courses. Among those enrolled in this course, the average is 3.04 among non-URM students and only 2.68 among URM students, indicating a course GPA gap of 0.36. CSUMB is a designated "Hispanic serving institution," with 51% of the university's student body in the underrepresented minority (URM) category. Between 2010 and 2014, the number of URM freshmen at CSUMB increased by over 19%. In the College of Business, 45% of 2016 incoming students are URM students. It is crucial for us to identify possible ways to embrace this opportunity.

One of the reasons students find the quantitative courses to be challenging is because prior to the incorporation of the Compustat and CRSP database projects, the finance curriculum lacked real-life examples. Unlike marketing courses in which instructors can easily show a commercial to illustrate the underlying theory, it is difficult to connect finance with real life without the use of data. Therefore, students often felt the material in class was somewhat disconnected and useless. With data-driven projects, students are more engaged in the learning process because they are able to see the real data of the companies they know. They can see how finance theories work in the reality. In addition, the projects foster a collaborative learning environment, which aligns best with the learning styles of minority groups that emphasize connected knowing, collaborative problem solving, and socially based knowledge (Cabrera et al., 2002).

Furthermore, to better serve time-impooverished students who juggle the demands of school, work, and family, the hybrid learning format, has become an increasingly appealing solution. The hybrid approach, also identified as blended learning (Bersin, 2004), combines traditional face-to-face learning with online classes, offering opportunities for synchronous and asynchronous learning. One of the biggest issues with hybrid learning is students are often not studying when they are not in class. When they are not interested in the subject, they will only learn when they are forced, which generally happens in class. Each data-driven project is designed to address this problem by making the project fun to work so that students are self-motivated to work on the material outside the classroom, making the hybrid model more effective.

Our third motivation is the AACSB's recent addition of a section on "Technology Agility" to accreditation Standard 9. Among the 5 bullet points included in the new section are:

- "Evidence-based decision making that integrates current and emerging technologies, including the application of statistical tools and techniques, data management, data analytics and information technology throughout the curriculum as appropriate"
- "Demonstration of higher-order cognitive skills to analyze an unstructured problem, formulate and develop a solution using appropriate technology, and effectively communicate the results to stakeholders" (AACSB, 2018, page 35).

The inclusion of Compustat and CRSP projects in the finance curriculum provide an ideal setting for managing large datasets, applying statistical analysis, and analyzing unstructured problems. Including written reports and presentations as deliverables allows students to practice and receive constructive feedback on how to communicate effectively.

The paper proceeds as follows. Section 2 describes the projects. Section 3 (and an Appendix) describes how we manage the data for the core finance course. Section 4 concludes.

DESCRIPTION OF PROJECTS

The series of projects we develop incorporate active or experiential learning and include collaborative (“group”) projects as well as individual projects. Johnson and Johnson (2009) and Marzano, Pickering, and Pollock (2005) discuss collaborative learning environments leading to positive outcomes (i.e. greater student achievement, higher-quality reasoning and/or critical thinking strategies, social skills improvement, higher degree of students’ engagement) for learners of all ages. In addition, positive experiences in collaborative learning have been shown to contribute to student learning, retention, and overall college success (Astin, 1993; Tinto, 1987; National Survey of Student Engagement, 2006). Researchers have highlighted the positive effects of collaborative work on minority students relative to other students (Fullilove and Treisman 1990; Levine and Levine, 1991). Aside from the benefits to learning of content material, participation in group projects can also help students develop a host of skills that are important in the business world (Caruso and Woolley, 2008; Mannix and Neale, 2005).

Students are introduced to Compustat and CRSP via projects in the core finance course that all business students are required to take. This course is a prerequisite to finance concentration courses, and these projects provide a foundation for concepts reinforced and extended to more complicated projects in two finance concentration courses (“Advanced Corporate” and “Investments”). The following table provides an overview of the projects and learning objectives for the core finance course projects.

Course	Project	Description	Deliverable
Finance Core	“Compustat: Financial Ratios”	Individual project	Excel spreadsheet
Finance Core	“CRSP: Risk and Return” (Part 1)	In-Class Collaborative Activity	Excel spreadsheet
Finance Core	“CRSP: Risk and Return” (Part 2)	Individual project	Excel spreadsheet & Written report

Additional description for the projects follows.

- Finance Core Course “Compustat: Financial Ratios”: Students create Excel spreadsheets to analyze real-world company financial ratios using data from Compustat. The instructor provides an Excel spreadsheet of Compustat financial data for the most recent calendar year end obtained via WRDS for 40 firms (8 firms from 5 different industry sectors). Recognizing students prefer to pick their own firms, each student can work on any firm they want. The instructor also provides the industry average of each ratio for year-end for each of the 5 industry sectors. The deliverable is an Excel spreadsheet that includes a comparison of whether each financial ratio is better, worse, or about the same for firm’s industry. Learning objectives include: understand financial ratios; understand what Compustat is and that its data are from company filings at the SEC; develop data checking skills and careful analysis techniques; develop a plan for how to interpret the ratios and judgement for whether each ratio is good, bad or about the same as the industry average.

- Finance Core Course “CRSP: Risk and Return” (Part 1): Students work in teams to analyze risk and return. Students create Excel spreadsheets to analyze real-world company stock risk and return using the most recent calendar year’s weekly data from CRSP. Part 1 is a collaborative activity in class and students obtain CRSP data via a WRDS class account. Working collaboratively, students are asked to calculate and interpret sample standard deviations, means, correlation coefficients, covariances, and Capital Asset Pricing Model (CAPM) alpha and beta and understand the risk-free rate as well as standard deviation and its implications for 1 sigma, 2 sigma and 3 sigma versus the mean; understand the difference between stand-alone risk and CAPM beta risk; understand what CRSP is and that its data are per actual U.S. stock market trading; practice data checking skills and careful analysis techniques; and develop a plan for how to interpret the results of the calculations.
- Finance Core Course “CRSP: Risk and Return” (Part 2): Part 2 tests the student's ability to calculate and interpret return and risk measures as an individual. For Part 2, students calculate sample standard deviation and mean of monthly returns, covariance of a company's stock monthly returns with the market (i.e., the CRSP value-weighted index), and CAPM beta, using monthly return data for 5 years (2013 - 2017). Students obtain feedback on their Excel spreadsheets. The calculations provide the elements needed for the Written Report deliverable, which asks the student to communicate an understanding of return and 2 measures of risk (stand-alone and CAPM Beta), an understanding of how standard deviation can be used as a rough estimate (1 sigma, 2 sigma and 3 sigma versus the mean) generally for data analysis if we are willing to assume the data are normally distributed, and the limitations of using historical data for forward-looking risk-for-return forecasts.

Projects in two finance concentration courses (“Advanced Finance” and “Investments”) extend students exposure to the WRDS portal for accessing the Compustat and CRSP databases. Rather than the instructor downloading data as in the core course, students access the data and obtain what they need directly using WRDS class accounts. Projects include the use of CRSP data to analyze risk and return analysis using additional risk models (e.g., Fama and French 3- and 4-factor models, asymmetric risk measures, and Value at Risk) and Compustat data to perform analyses of company fundamentals and time trends. Projects are scaffolded, are a mix of in-class and outside-of-class assignments, and include individual and collaborative projects. Deliverables include Excel spreadsheets, written reports, and oral presentations that are suitable for inclusion in students’ interview materials for employers.

MANAGING THE CORE FINANCE COURSE DATA EACH TERM

Because all undergraduate business students take the core finance course, we offer many sections each term. The following were considerations in our decision for selecting companies for the students’ use for the core finance course.

- Use only publicly-traded U.S.-listed companies because we want the students to be able to perform analyses that require the use of 10-k data.
- Make the companies interesting for the students; give the students a long enough list of companies that at least one company will probably be of interest personally.
- Limit the number of companies so that grading is manageable.
- Make the data timely.

- Generate the data elements from WRDS Compustat and CRSP university subscriptions formatted as Excel spreadsheets for the students' use.

We use the following steps.

Step 1: Determine the Compustat and CRSP data elements.

Step 2: Use the WRDS interface to manually set up the Compustat and CRSP data elements and save for future semesters.

Step 3: Use WRDS to manually request data for all U.S.-listed firms at the beginning of each semester or quarter. Save as permanent SAS datasets for the semester or quarter.

Step 4: Use SAS to obtain the S&P 500 current constituent list from the Select Sector SPDR ETFs website for each of the 10 sectors that constitute the S&P 500 index. As of Sept. 15, 2018, the links are:

<http://www.sectorspdr.com/sectorspdr/IDCO.Client.Spdrs.Holdings/Export/ExportCsv?symbol=xlk>

for Technology (Ticker = xlk), and substituting the Ticker for the other 9 sectors:

Sector	Ticker
Consumer Discretionary	xly
Consumer Staples	xlp
Energy	xle
Financials	xlk
Healthcare	xlw
Industrials	xli
Materials	xlb
Real Estate	xlre
Utilities	xlu

Step 5: Use SAS to narrow the list down from 500 firms, obtain the data elements from the manual WRDS downloads,

For each sector ETF, the Select Sector SPDR ETFs website link downloads a csv file of all the companies in the S&P 500 with that sector classification. Ticker symbol and Company name are listed in the first 2 columns, starting in row 3. The third column reports the weight percent of the stock in the sector, with the column total equaling 100%. We use that information to retain the largest market-cap firms.

For example, for the first data project in the core undergraduate finance course, we ask the students to analyze a company's financial ratios using the Compustat data. We have about 40 students in the class. We let students choose a company from the list of companies we generate. We also want to have the students compare the financial ratios to the industry average. We define sector as industry. We have about 40 students per section. For fall semester 2018, we decided to use 5 of the 10 sectors (specifically by ticker: xly, xlp, xle, xlk, and xli). The SAS code provided in the Appendix downloads the tickers, company names, and index weights for those 5 sectors from the Select Sector SPDR ETFs website. It then obtains the Compustat data elements for each company from the manually created Compustat dataset (i.e., Step 3 above). Companies with missing data elements are eliminated. The SAS code retains the largest (by market-cap) 8 companies (with non-missing data elements) for each of the 5 sectors. Finally, the SAS code outputs the Excel dataset with desired data elements in rows with one column for each of the 40

firms. The company's sector name is included so that students know the appropriate mapping of companies to sectors for industry average calculation and comparison. The same approach is used to generate CRSP data elements for risk and return analyses used later in the course.

CONCLUDING REMARKS

In this paper, we describe how we incorporate Compustat and CRSP databases in the undergraduate finance curriculum. Our projects contribute to our university's commitment to preparing students to be competitive in the today's technology-focused and data-driven workforce as well as serving underrepresented minority (URM) students. We believe our work is also particularly timely given the AACSB's recent addition of a section on "Technology Agility" to accreditation Standard 9.

It is worth noting that the commitment of the CSUMB College of Business to incorporating real-world data analysis in the students' undergraduate curriculum is so strong that students also use Compustat data in their Principles of Management core course. The students in Principles of Management complete a group project to provide students with the opportunity to thoroughly research a publicly-traded company by employing many of the concepts introduced in the course. In addition, this project introduces students to many of the business analysis tools, including financial databases (Compustat via WRDS and Hoovers), required for their Capstone project in their senior semester. Aspects of financial data are used throughout the report including the company overview, industry analysis and competitive environment. One of the sections of the report deals specifically with the financial performance including at least 5 key financial ratios and what they mean to the company (2 paragraphs max). They are also required to include graphics as needed to demonstrate the most significant numbers.¹

ENDNOTES

¹Thank you to Dr. Jill Hosmer, CSUMB Lecturer and coordinator of the College of Business Principles of Management course curriculum.

REFERENCES

AACSB International (The Association to Advance Collegiate Schools of Business). (2018). 2013 Eligibility Procedures and Accreditation Standards for Business Accreditation, updated July 1, 2018, available at <https://www.aacsb.edu/-/media/aacsb/docs/accreditation/business/standards-and-tables/2018-business-standards.ashx?la=en>

Astin, A. (1993). *What matters in college? Four critical years revisited*. San Francisco: Jossey-Bass.

Bersin, J. (2004). *The blended learning book: Best practices, proven methodologies, and lessons learned*. John Wiley & Sons.

Cabrera, A. F., Nora, A., Crissman, J. L., & Terenzini, P. T. (2002). *Collaborative learning: Its*

impact on college students' development and diversity. *Journal of College Student Development*, 43(1), 20.

Caruso, H.M., & Wooley, A.W. (2008). Harnessing the power of emergent interdependence to promote diverse team collaboration. *Diversity and Groups*, 11, 245-266.

Fullilove, R. E., & Treisman, P. U. (1990). Mathematics achievement among African American undergraduates at the University of California, Berkeley: An evaluation of the mathematics workshop program. *The Journal of Negro Education*, 59(3), 463-478.

Johnson, D. W., & Johnson, R. T. (2009). An educational psychology success story: Social interdependence theory and cooperative learning. *Educational researcher*, 38(5), 365-379.

Levin, M. E., & Levin, J. R. (1991). A critical examination of academic retention programs for at-risk minority college students. *Journal of College Student Development*.

Mannix, E., & Neale, M.A. (2005). What differences make a difference? The promise and reality of diverse teams in organizations. *Psychological Science in the Public Interest*, 6(2), 31-55.

Marzano, R. J., Pickering, D. J., & Pollock, J. E. (2005). *Classroom instruction that works: Research-based strategies for increasing student achievement*. Upper Saddle River, NJ: Pearson/Merrill/Prentice Hall.

National Survey of Student Engagement Report. (2006).

http://nsse.iub.edu/NSSE_2006_Annual_Report/docs/NSSE_2006_Annual_Report.pdf.

Tinto, V. (1987). *Leaving college: Rethinking the causes and cures of student attrition*. Chicago: University of Chicago Press.

APPENDIX: SAS Code for Managing Course Data Each Term (Compustat Example)

```
**** Name permanent dataset location on laptop flash D;;

libname sasDataD "D:";
options linesize = 180 nocenter nonumber nodate;

*** Obtain Compustat data elements from permanent Compustat dataset;

data getCPST;
set sasDataD.alltickers;
run;

data getvars;
set getCPST;
Current_assets = ACT;
Current_liabilities = LCT;
Inventory = INVT;
Cost_of_goods_sold = COGS;
Accounts_receivable = RECT;
Annual_sales = SALE;
Net_fixed_assets = PPENT;
Total_assets = AT;
Total_liabilities = LT;
EBIT = OIBDP - DP;
Interest_charges = XINT;
Net_income = NI;
Common_equity = CEQ;
Market_price_per_share = MKVALT / (CEQ / BKVLPS);
Earnings_per_share = EPSFI;
Book_value_per_share = BKVLPS;

if Current_assets = . or
   Current_liabilities = . or
   Inventory = . or
   Cost_of_goods_sold = . or
   Accounts_receivable = . or
   Annual_sales = . or
   Net_fixed_assets = . or
   Total_assets = . or
   Total_liabilities = . or
   EBIT = . or
   Interest_charges = . or
   Net_income = . or
   Common_equity = . or
   Market_price_per_share = . or
   Earnings_per_share = . or
   Book_value_per_share = .
   then delete;

if Book_value_per_share < 0 then delete;

Year = year(datadate); if year = 2017; run;
```



```

*** Read in 5 sector spreadsheets;

filename fool url
'http://www.sectorspdr.com/sectorspdr/IDCO.Client.Spdrs.Holdings/Export/ExportCsv?symbol=xly';

data getdat1;
length iticker $8 iname $36 iweight1 $8 iweight2 $8 sector $12;
infile fool delimiter=', ' firstobs=3;
input iticker iname iweight1 iweight2;
Sector = 'Cons Discr';
run;

filename fool url
'http://www.sectorspdr.com/sectorspdr/IDCO.Client.Spdrs.Holdings/Export/ExportCsv?symbol=xlp';

data getdat2;
length iticker $8 iname $36 iweight1 $8 iweight2 $8 sector $12;
infile fool delimiter=', ' firstobs=3;
input iticker iname iweight1 iweight2;
Sector = 'Cons Staples';
run;

filename fool url
'http://www.sectorspdr.com/sectorspdr/IDCO.Client.Spdrs.Holdings/Export/ExportCsv?symbol=xle';

data getdat3;
length iticker $8 iname $36 iweight1 $8 iweight2 $8 sector $12;
infile fool delimiter=', ' firstobs=3;
input iticker iname iweight1 iweight2;
Sector = 'Energy';
run;

filename fool url
'http://www.sectorspdr.com/sectorspdr/IDCO.Client.Spdrs.Holdings/Export/ExportCsv?symbol=xli';

data getdat4;
length iticker $8 iname $36 iweight1 $8 iweight2 $8 sector $12;
infile fool delimiter=', ' firstobs=3;
input iticker iname iweight1 iweight2;
Sector = 'Industrials';
run;

filename fool url
'http://www.sectorspdr.com/sectorspdr/IDCO.Client.Spdrs.Holdings/Export/ExportCsv?symbol=xlk';

data getdat5;
length iticker $8 iname $36 iweight1 $8 iweight2 $8 sector $12;
infile fool delimiter=', ' firstobs=3;

```

```

input iticker iname iweight1 iweight2;
Sector = 'Technology';
run;

*** Create a single list of companies and trim quotation marks and % signs;

data getspdrs;
set getdat1 getdat2 getdat3 getdat4 getdat5;
tic = compress(iticker, '');
company_name = compress(iname, '');
weight1 = compress(iweight1, '%') +0;
weight2 = compress(iweight2, '%') +0;
if weight1 = . then weight1 = weight2;
drop iticker iname iweight1 iweight2 weight2;
run;

*** Merge with Compustat data elements;

proc sort data = getspdrs;
by tic;
run;

proc sort data = getvars;
by tic;
run;

data mergel;
merge getspdrs (in = frodo) getvars;
by tic;
if frodo;
if datadate = . then delete;
run;
proc sort data = mergel;
by sector descending Weight1;
run;

*** Retain 8 largest firms by market-cap in each sector and save as Excel
file dataset;
data F18CpstData2017;
set mergel;
by sector descending Weight1;
if first.sector then i = 0;
i+1;
if i > 8 then delete;
drop i;
run;
PROC EXPORT DATA= WORK.F18cpstdata2017
            OUTFILE= "D:\Cpst Fall 2018 Class Data.xlsx"
            DBMS=EXCEL REPLACE;
            SHEET="Data";
RUN;

```