ECON 18: Quantitative Equity Analysis

Winter Study 2009

Professor David Kane
dkane at iq.harvard.edu
Teaching Fellow David Phillips '11
David.E.Phillips at williams.edu

Overview

This class will introduce students to applied quantitative equity research. We will briefly review the history and approach of academic research in equity pricing via a reading of selected papers. Students will then learn the best software tools for conducting such research. Students will work as teams to replicate the results of a published academic paper and then extend those results in a non-trivial manner. This course is designed for two types of students: first, those interested in applied economic research, and second, those curious about how that research is used and evaluated by finance professionals. See "Publication, Publication" for useful background reading on my approach. Special thanks to Bill Alpert, Senior Editor at Barron's, Joe Masters '96, Assistant Professor of Mathematics, SUNY Buffalo, George Poulogiannis and Dan Gerlanc '07 for agreeing to participate in the class.

Prerequisites and Background Reading

There are no formal prerequisites for the class. Students who have taken STAT 201 will have an advantage in terms of their exposure to **R**, the programming language that we use. But other students will be able to learn enough **R** both before class starts and in the first week of Winter Study. You can learn more about and download R. See Wikipedia for an overview of R. Read *An Introduction to R*.

You are welcome to use whatever editor you like for writing code, but I strongly recommend Emacs. It's a powerful tool that is useful for creating documents, coding, and running R. Start with this introduction and do the tutorial (which appears in emacs when you press Ctrl-h t). More details (than you really need) can be found at the main documentation page. One of the advantages of using Emacs is that you can run R within your Emacs session. To do so, an XEmacs package called ESS must be installed. Read this overview for an introduction and the ESS manual for more details. Windows users should install this version of Emacs, with ESS already included.

You are welcome to write your paper using whatever software you like, but I recommend Sweave, an R package that lets users execute chunks of R code in a LaTeX document. Read *Sweave, Part I: Mixing R and LATEX*, Sweave, Part II: Package Vignettes, and the *Sweave User Manual*. If you do not already know LaTeX, a powerful typesetting system, start here. Read *The Not so Short Introduction to LaTeX2* for more details, but you only care about LaTeX because it serves as the basis for Sweave. For those who find LaTeX frightening, another option is to use the odfWeave package. References in LaTeX are typically done with BibTeX. See here for a tutorial. Again, students may use whatever software they like to write their papers, but serious professionals use these tools. For this class, even R is optional for any student with a burning desire to use STATA or some other statistical program. The key is not the particular software that you use but the scientific robustness of your results. You need to be able to prove --- quickly and transparently --- the reliability of any empirical claim that you make in your research. Reproducibility is key.

Class Sessions

We will meet three times a week: Monday mornings from 9:00 AM to Noon, Wednesday evenings from 7:00 to 10:00 and Friday afternoons from 1:00 to 4:00. All meetings will be in Greylock D. Meetings will be working sessions. Each team/student should bring her laptop and be prepared to work on data analysis during class. I will do very little lecturing. Think about the class like an apprenticeship. If you wanted to be a blacksmith in 15th century France, there was no book to read or class to take. Your best option was to become the apprentice of the local blacksmith. He taught you what you needed to know by working closely with you. Similarly, there is no (good) book to read about how to conduct applied quantitative equity analysis. The best way to learn is to work closely with someone who already knows how to do it. So, that's what we will do during class. Each student/team will work on their own project as I move around the room, answering questions and providing suggestions. If a particularly important point comes up, I will stop everyone and give a brief overview to the whole class about the topic.

Data

The hardest part of financial research is gathering and cleaning the necessary data. Unfortunately, this is too complex a task for us to learn over Winter Study. So, I will provide you with an R package (ws.data) with all the necessary data. We will be working with 10 years of daily data covering the 1,500 largest US stocks. This data will not exactly match the data that is used in our 6 target articles, but it is close enough for you to (try to) replicate some of their major findings.

Articles

All 6 of these articles are important papers in the quantitative equity literature. Although we do not have the time and data resources to *exactly* replicate *all* of their results, we should be able to use the same basic approaches, albiet over a different universe and time period.

- "Does the stock market overreact?" by DeBondt, Werner, and Richard Thaler, 1985.

Journal of Finance, 40, 793–805.

- "Evidence of predictable behavior of security returns" by Narasimhan Jegadeesh, 1990. *Journal of Finance*, 45, 881-898.
- "Returns to buying winners and selling losers: Implications for market efficiency" by Narasimhan Jegadeesh and Sheridan Titman, 1993. *Journal of Finance*, 48, 65–91.
- "Do industries explain momentum?" by Tobias Moskowitz and Mark Grinblatt, 1999. *Journal of Finance*, 54, 1249–1290.
- "Price momentum and trading volume" by Charles M.C. Lee and Bhaskaran Swaminathan, 2000. *Journal of Finance*, 55, 2017–2069.
- "The 52-Week High and Momentum Investing" by Thomas J. George and Chuan-Yang Hwang, 2004. *Journal of Finance*, 59, 2145-2176

If there is a different article that you would prefer to replicate, please let me know. Only one team of students will be allowed to choose each article. First come, first served.

Schedule

Week 1: January 5

Review software tools and academic articles. Assign students to teams. Select target article. Examine the data in the ws.data package. Sign up for r-sig-finance mailing list. Readings:

- "Publication, Publication" by Gary King, 2006. *PS*, 39, 119-125.
- "R for Beginners" by Emmanuel Paradis. Also skim "An Introduction to R".
- Six target articles. As a first pass, you should read the abstract, introduction, conclusion and figure legends for all six. Do so in chronological order, as they are listed above. Once you have selected your target article, you will (obviously) be reading it over and over again.

By the end of the week, you should be comfortable working with R, examining your data, evaluating commands, using the help system and writing functions.

Week 2: January 12

Students create a skeleton of their final paper involving all the basic tools of the class: an Sweave document with embedded R code and bibliography. Skeletons will include at least one figure and one table. Collect and read other articles to include in your bibliography. Sign up for R-help mailing list. (I recommend the digest option.) Explore the finance "task view" at CRAN.Readings:

- "Sweave, Part I: Mixing R and LATEX", "Sweave, Part II: Package Vignettes","An Sweave Demo" and the "Sweave User Manual".
- Skim "R Data Import/Export". You will probably not have occasion to use these tools much, but you should be aware of the power of R. Some teams may benefit from placing the data in a relational database in order to speed and/or organize their calculations.
- "Backtests" by K. Campbell, J. Enos, D. Gerlanc, and D. Kane. *R News*, 7(1):36-41, April 2007.

By the end of the week, you should be comfortable incorporating R code and analysis within an Sweave document. You should be using all the tools that you will need for the class. You

should be able to e-mail your work to another team and have them replicate your results.

Week 3: January 19

Students continue work on their papers, creating an R package that contains the paper (written as a vignette) and the supporting code. After replicating (or at least trying to) one of the main results of their chosen paper, students should extend the analysis in some interesting and non-trivial manner. Much of class discussion will be devoted to the topic of what it means to be "interesting" in the context of quantitative equity research. Readings:

- Read chapters 1 -- 4 in "Writing R Extensions". This provides the details on how to create an R package with its associated data, functions and vignette.
- Other 5 target articles. This will help you understand the style and substance of academic articles.
- "Let's Practice What We Preach: Turning Tables into Graphs" by Andrew Gelman, Cristian Pasarica, and Rahul Dodhia, 2002. *American Statistician*, 56, 121-130.

By the end of the week, you work should be essentially complete.

Week 4: January 26

Finish papers and, depending on student preferences, present results in class. Final papers will be posted on the Internet. You will have worked very hard on your paper and you should be proud of the results.

Commands and Functions

Here is a collection of some of the functions and commands that we will be learning during the course:

- R functions: library, vignette, ls, help, search, head, tail, summary, dim, order, sort, sample, table, subset, function, unique, duplicated, any, all, plot, xyplot, merge, Date, date.mdy, formatC, apply, tapply, lapply, aggregate, by, list.files, cut, data, attach/detach, setwd, getwd, package.skeleton, prompt, edit, if, for, split, install.packages, save and load.
- R packages: backtest, lattice, Rmetrics, and plyr.
- Emacs:C-x C-f, C-x C-s, C-x C-c, C-x b, C-x k, C-x o, C-x 2, C-x u, C-g, C-a, C-e, C-k, C-y, C-SPC, and C-w.
- ESS: M-x R, C-c C-l, C-c C-r, C-c C-n, C-c C-t, M-q, C-c C-z, C-c `, C-c C-o, M-RET, C-c C-s, and C-c C-q.

This may seem a bit overwhelming at first, but regular use makes most of them automatic.

Conclusion

If you had tried to conduct a similar piece of financial research before taking this class, you would have done X well. Now that you have taken the class, you will do Y well, both with your actual paper and with any future financial research you choose to undertake. The success (or failure) of the class can be measured by comparing Y with X.