MGT/ISYE/MATH 6785 : Practice of QCF

Georgia Tech Ernest Scheller Jr. College of Business

Class Information

Instructor: **Sudheer Chava**
Office: **402 Scheller College of Business**
Email: **sudheer.chava@scheller.gatech.edu** (best way of reaching me)
Phone: **404-894-4371**
Web site: `http://www.prism.gatech.edu/~schava6`
Office Hours: **4:30-5:30pm on Tue; or by appointment**
Class Room: SCOB Room 101
Class timings: Thursday 6:30-9:15pm

Course overview

This course is a *practicum* or *project based capstone course* for QCF students. As such, the class is a project based seminar course and not a regular lecture based class. The classes would involve student presentations of papers, methodologies, implementation and results. The experiential learning projects in the class or grouped into two categories

1. *Industry mentored projects* - Many a time the industry mentored projects would involve the students accessing proprietary data bases and systems of the industry partners. In this case, typically an NDA is required between the student and the company to get access to the systems and data. These data should be considered **confidential** and should be used only for class project and shouldn't be shared with anyone.

2. *Projects with data that GT has access to / public data* - If any of the students don't want to sign an NDA with the experiential project partner, they can work on projects that involve data that GT has access to or public data. Please note all the data made available for the class - either through WRDS or shared through QCF servers - should be used *only* for the class purposes and shouldn't be shared with anyone outside the group. There are strict

licensing and usage restrictions on data purchased by Georgia Tech. These would mainly deal with the following topics

- Quantitative Trading Strategies
- NLP models applied to finance
- Bayesian and Machine Learning Models applied to Credit Risk
- FinTech, Crypto Currencies and Blockchain
- Other topics are possible, but we need to discuss first

Required Material

There is no *required* text book for the class. For each of the topics, some relevant papers are assigned. All the papers would be available online and all datasets would be available for download through WRDS or on canvas.

Online Information

Course page is maintained on canvas `canvas.gatech.edu` . All relevant information including but not limited to: readings, lectures, presentation schedules and class schedules and any changes to the schedule would be posted on the website. I would strongly recommend that you check the course page on canvas website every week.

Grading

This course is primarily a project based class. As such there are no quizzes or final exams in the class. The grading for the class is determined by

- **Project**: The project would have a cumulative weight of 80%.
- **Presentations**: Each group is expected to present multiple times - to me and to the class. These presentations would carry a weight of 20% in aggregate.

Course Rules

- No cell phone usage in class. Please turn your ringers off.
- Please do **NOT** use laptops or computers in the class room unless I ask you to do so.

- You are expected to be familiar with the Georgia Tech Honor Code `http://honor.gatech.edu`, and to conform to all its requirements.
- To ensure a prompt response to e-mail, please put **course number and class time** in the subject line, e.g. **MGT 6785-6:30-9.15: Question about Project**

Data and Programming

- If you selecting a non-industry project, make sure you apply for a WRDS account (`https://wrds.wharton.upenn.edu`) to access some of the data licensed by Georgia Tech
- Make sure that relevant data is available on WRDS for trading strategies. (check with me in the first meeting)
- Bankruptcy, Default and Covenant Violation data and other related datasets will be provided for Credit Risk estimation.
- Python, R and SAS are suggested for most of the projects. SAS can also be useful for data manipulation and estimation can be done in Python or R. It's also OK to do the entire project in Python or R.

Data Confidentiality

Please note all the data made available for the class - either through WRDS or shared through QCF servers - should be used *only* for the class purposes and shouldn't be shared with anyone outside the group. There are strict licensing and usage restrictions on data purchased by Georgia Tech. Some other data that is made available for the class is confidential and the students may need to sign an *NDA* to access the data.

Selection of the Projects

The project selection for the class is enumerated in the following steps

1. The companies proposing experiential learning projects come and do a project pitch of their proposed projects or the project pitch is posted on canvas. This year tte participating companies include **Invesco**, **Voya Asset Management**, **Wells Fargo** and **Truist**.
2. Students interested in the projects would need to fill a survey and express their top choices among the proposed projects

3. Students would be selected from the submitted applications from interested students. The project mentor will provide further details to the students regarding NDA and other requirements to access the systems and data needed for the project.

4. **Students have a choice to not sign the NDA**. In that case, they can't work on the industry project and I will instead assign them projects involving data that GT has access to or using public data.

5. I would encourage any student considering a Ph.D. to select a non-industry project as there is more flexibility in the choice of topics that are more suitable for future Ph.D. work

6. I would also suggest that students working part-time in the Fall consider non-industry projects so that there is no conflict between the NDAs and their part-time employment.

7. There would be a large selection of non-industry projects available for any student not interested in considering industry mentored projects.

Quantitative Trading Strategies

Factors

- Ang, Andrew, Jun Liu, and Krista Schwarz. 2018. "Using Individual Stocks or Portfolios in Tests of Factor Models." SSRN Working Paper 1106463.
- Arnott, Robert D, Mark Clements, Vitali Kalesnik, and Juhani T Linnainmaa. 2020. "Factor Momentum." Journal of the American Statistical Association 3116974.
- Asness, Clifford, Swati Chandra, Antti Ilmanen, and Ronen Israel. 2017. "Contrarian Factor Timing Is Deceptively Difficult." Journal of Portfolio Management 43 (5): 72–87.
- Asness, Clifford, and Andrea Frazzini. 2013. "The Devil in Hml's Details." Journal of Portfolio Management 39 (4): 49–68.
- Asness, Clifford, Andrea Frazzini, Ronen Israel, Tobias J Moskowitz, and Lasse H Pedersen. 2018. "Size Matters, If You Control Your Junk." Journal of Financial Economics 129 (3): 479–509.
- Chinco, Alexander, Adam D Clark-Joseph, and Mao Ye. 2019. "Sparse Signals in the Cross-Section of Returns." Journal of Finance 74 (1): 449–92.
- Chinco, Alexander, Andreas Neuhierl, and Michael Weber. 2020. "Estimating the Anomaly Baserate." Journal of Financial Economics Forthcoming.
- Chinco, Alex, Samuel M Hartzmark, and Abigail B Sussman. 2019. "Necessary Evidence for a Risk Factor's Relevance." SSRN Working Paper 3487624.
- Chordia, Tarun, Amit Goyal, and Alessio Saretto. 2020. "Anomalies and False Rejections." Review of Financial Studies 33 (5): 2134–79.
- Daniel, Kent, and Sheridan Titman. 2012. "Testing Factor-Model Explanations of Market Anomalies." Critical Finance Review 1 (1): 103–39.
- Fama, Eugene F, and Kenneth R French. 1993. "Common Risk Factors in the Returns on Stocks and Bonds." Journal of Financial Economics 33 (1): 3–56.
- Fama, Eugene F, and Kenneth R French. 2015. "A Five-Factor Asset Pricing Model." Journal of Financial Economics 116 (1): 1–22.

- Fama, Eugene F, and Kenneth R French. 2018. "Choosing Factors." Journal of Financial Economics 128 (2): 234–52.
- Gu, Shihao, Bryan T Kelly, and Dacheng Xiu. 2020. "Empirical Asset Pricing via Machine Learning." Review of Financial Studies 33 (5): 2223–73.
- Harvey, Campbell R, Yan Liu, and Heqing Zhu. 2016. "... And the Cross-Section of Expected Returns." Review of Financial Studies 29 (1): 5–68.
- Harvey, Campbell R. 2017. "Presidential Address: The Scientific Outlook in Financial Economics." Journal of Finance 72 (4): 1399–1440.
- Harvey, Campbell, and Yan Liu. 2019. "Lucky Factors." SSRN Working Paper 2528780.
- Harvey, Campbell R, and Yan Liu. 2019. "A Census of the Factor Zoo." SSRN Working Paper 3341728.
- Harvey, Campbell R. 2020. "Replication in Financial Economics." Critical Finance Review, 1–9.
- Harvey, Campbell R, and Yan Liu. 2020. "False (and Missed) Discoveries in Financial Economics." Journal of Finance Forthcoming.
- Harvey, Campbell R, Yan Liu, and Alessio Saretto. 2020. "An Evaluation of Alternative Multiple Testing Methods for Finance Applications." Review of Asset Pricing Studies 10 (2): 199–248.
- Hou, Kewei, Chen Xue, and Lu Zhang. 2015. "Digesting Anomalies: An Investment Approach." Review of Financial Studies 28 (3): 650–705.
- Hou, Kewei, Chen Xue, and Lu Zhang. 2020. "Replicating Anomalies." Review of Financial Studies 33 (5): 2019–2133.
- see papers at `https://teams.microsoft.com/_#/school/files/General?threadId=19%3A6c118ed860324aad8b66e8d0b7105c8c%40thread.tacv2&ctx=channel&context=Literature&rootfolder=%252Fsites%252FTradingStrategies-Summer2020%252FShared%2520Documents%252FGeneral%252FLiterature`

Lead-Lag in Stock Returns

- Lo and MacKinlay (RFS, 1990): Document positive cross-autocorrelation, which explains: (1) positive index autocorrelation (2) profits from contrarian portfolio strategies. Display lead-lag effects in size-sorted portfolios. What are the economic sources of positive cross-autocorrelation across securities?

- Brennan, Jegadeesh, and Swaminathan (RFS, 1993): Many analyst firms lead fewer analysts firms. Many analyst firms respond more rapidly to market returns. (Arbitrage portfolios (long many analyst, short few analyst) are negatively related to lagged market index.)

- Badrinath, Kale, and Noe (RFS, 1995): Portfolios of stocks with the highest level of institutional ownership lead the returns on portfolios of stocks with lower institutional ownership.

- Chordia and Swaminathan (JF, 2000): High volume portfolios lead returns on low volume returns, because returns on low volume portfolios respond more slowly to information in market returns.

- Sias and Starks (JFE, 1997): Autocorrelations are an increasing function of the level of institutional ownership, implying that institutional trading reflects information and increases the speed of price adjustment.

- McQueen, Pinegar, and Thorley (JF, 1996): When returns on large stocks are negative, the concurrent beta for small stocks is high, but the lagged beta is insignificant. When returns on large stocks are positive, small stocks have small concurrent beta and very significant lagged betas, suggesting all stocks react quickly to negative macroeconomic news.

- Bessembinder, Chan, and Seguin (JFE, 1996): No relationship between market-wide information flows and small-firm volume.

- Nofsinger (JBF, 2001): Investors buy large firms after good economic news and sell large firms after bad economic news. The trading of small firms does not appear to be motivated by macro-news.

- Adams, McQueen, and Wood (JB, 2004): Find weaker evidence of a delayed small stock response to inflation news.

- Hou and Moskowitz (RFS, 2005): The most delayed firms (speed of response to common information) command a large return premium.

- Hou (RFS, 2007): Lead-lag effect is predominantly an intra-industry phenomenon. The effect is driven by small firms' sluggish adjustment to negative information on big firms.
- Hameed, Kang, and Viswanathan (JF, 2010): Contrarian profits are higher following market declines.
- Chordia, Sarkar, and Subrahmanyam (JFQA, 2011): In equilibrium, large stock illiquidity induced by informed trading portends stronger cross-autocorrelations. Find that the lead-lag relation increases with lagged large stock illiquidity, which is stronger before macro announcements, and weaker afterward.
- Du et al (2011): Size factor is positively related to lagged market return.

Networks, Industries, Return Predictability, and the Cross-section of Stock Returns

- Do Industries Lead Stock Markets by Harrison Hong, Walter Torous, Rossen Valkanov (JFE, 2007)
- Market Segmentation and Cross-predictability of Returns, by Lior Menzly and Oguzhan Ozbas (JF, 2010)
- Network Centrality and the Cross-section of Stock Returns, 2015 by Kenneth R.Ahern
- Industry Networks and the Speed of Information Flow, 2015 by Joonki Noh
- Rapach, Strauss, Tu and, Zhou, 2015, Industry Interdependencies and Cross-Industry Return Predictability
- Chava, Hsu and Zeng, Does History Repeat Itself? Business Cycle and Industry Returns

Information Content of 8-Ks

- Xiaofei Zhao, 2013, Information Intensity and the Cross-Section of Stock Returns.
- Sudheer Chava and Nikhil Paradkar, 2016, "December Doldrums, Investor Distraction, and Stock Market Reaction to Unscheduled News Events"

Impact of Monetary Policy

- Bernanke and Kuttner, What Explains the Stock Market?s Reaction to Federal Reserve Policy?
- Lucca and Moench, The Pre-FOMC Announcement Drift
- Chava and Hsu, 2014, Monetary Policy Shocks, Financial Constraints and the Cross-Section of Equity Returns
- Chava, Gallmayer and Park, Credit Conditions and Stock Return Predictability, Journal of Monetary Economics

Event Studies in Stock and CDS Markets

- Kothari and Warner Review
- Chava, Ganduri and Ornthanlai, 2016, Are Credit Ratings Still Relevant?

Hedge Funds

- Getmansky, Lee and Lo, Review, Hedge Funds: A Dynamic Industry in Transition
- Brav, Jiang, Partnoy and Thomas, 2008, Hedge Fund Activism, Corporate Governance, and Firm Performance
- I may be able to make a widely used hedge fund database available to the students to conduct analysis

Trading Strategies

- Andrea Frazzini, David Kabiller and Lasse H. Pedersen, 2014, Buffett's Alpha
- Cohen and Lou, Complicated Firms
- Cohen, Malloy and Pomorski, Decoding Inside Information
- Clifford S. Asness, Andrea Frazzini, and Lasse H. Pedersen, Quality Minus Junk

Default Risk and Equity Returns

- Chava and Purnanandam, 2010, Is Default Risk Negatively Related to Stock Returns?
- Campbell, Hilscher and Szilagyi, In Search of Distress Risk

Misbehaving

- DAVID HIRSHLEIFER, SONYA SEONGYEON LIM, and SIEW HONG TEOH, Driven to Distraction: Extraneous Events and Underreaction to Earnings News
- Goetzmann et al, Weather-Induced Mood, Institutional Investors, and Stock Returns
- Sudheer Chava and Nikhil Paradkar, 2016, "December Doldrums, Investor Distraction, and Stock Market Reaction to Unscheduled News Events"

Bayesian Model Combination

- Bayesian Model Averaging and other similar techqniques applied to trading strategies
- Other papers if there is interest

Other Papers

- papers and reviews in the literature folder on tsquare
- other papers referenced in these papers
- papers that are of interest to you and not listed here(bring to my attention in the next meeting)

These are steps that students are expected to follow in the class

1. Form a group of three students
2. browse through the topic list, reference material (in the syllabus) and own research. Discuss within the group and come up with a particular trading strategy
3. Read the relevant that is in the literature directory on tsquare or listed in the syllabus carefully.

4. Make a group presentation to me about the trading strategy. The idea is to make a case for the trading strategy and suggested improvements and convince me. Note that I don't expect you to understand the trading strategy completely at this stage. But you should think about the following questions

    - What is the economic idea behind the paper?
    - What is the empirical methodology applied in the paper?
    - What data sources are used in the paper?
    - What are the main results of the paper?
    - What are the implications of the results?
    - Is it a feasible and implementable trading strategy? What about transaction costs? What is the maximum draw down?
    - How can the trading strategy be improved?

5. At a later date, make a presentation to the entire class in more depth about the trading strategy. I will provide more details on the presentation later.

6. The next step is to implement the trading strategy and improvements if any and check the performance

7. All groups have to submit a succinct report on the trading strategy, implementation and results. Detailed instructions on the report structure would be given later

Text as Data and NLP

- For background information, see
  - Diyi's course - https://www.cc.gatech.edu/classes/AY2020/cs7650_spring/
  - Jacob's book - https://mitpress.mit.edu/books/introduction-natural-language-processi
- Spitzley, L. (2014). Predicting fraud from quarterly conference calls: A small-sample study of scripted language.
- Qin, Y., & Yang, Y. (2019). What You Say and How You Say It Matters: Predicting Financial Risk Using Verbal and Vocal Cues. In 57th Annual Meeting of the Association for Computational Linguistics (ACL 2019) (p. 390).
- Keith, K. A., & Stent, A. (2019). Modeling financial analysts' decision making via the pragmatics and semantics of earnings calls. arXiv preprint arXiv:1906.02868.
- Loughran and Mcdonald (2015) for a survey of textual analysis in finance
- Loughran and Mcdonald, When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks
- Also see http://www3.nd.edu/~mcdonald/Word_Lists.html
- Bochkay, Khrystyna, Jeffrey Hales, and Sudheer Chava. "Hyperbole or reality? Investor response to extreme language in earnings conference calls." The Accounting Review 95.2 (2020): 31-60.
- Hassan, Tarek A., et al. "Firm-level political risk: Measurement and effects." The Quarterly Journal of Economics 134.4 (2019): 2135-2202.
- Hassan, Tarek Alexander, et al. Firm-level Exposure to Epidemic Diseases: Covid-19, SARS, and H1N1. No. w26971. National Bureau of Economic Research, 2020.
- Chen, Andrew Y., and Jie Yang. "Financing Concerns in April 2020 Appear Worse Than in 2008 Based on Earnings Calls." Available at SSRN 3592929 (2020).

Blockchain and Distributed Ledgers

Another choice for the students to focus on the emerging areas of FinTech, Crypto Currencies and Block Chain. Some examples of possible projects are below, but you can come and discuss with me if you are interested in related areas

Analyzing the Bitcoin Blockchain

The Bitcoin blockchain — currently 170 GB and growing — contains a massive amount of data that can give us insights into the Bitcoin ecosystem, including how users, businesses, and miners operate. BlockSci enables fast and expressive analysis of Bitcoin's and many other blockchains. The accompanying working paper explains its design and applications:

1. https://arxiv.org/pdf/1709.02489.pdf
2. `https://github.com/citp/BlockSci`

Crypto Currencies

See the papers in the Crypto folder on Canvas

Bayesian and Machine Learning Models applied to Credit Risk

Another choice for the students to focus on Credit Risk. Specifically, the class would deal with some bayesian methods and machine learning techniques as applied to credit risk. The focus would be on applications of these tools. The students are expected to read the references and present the techniques in the class. The main goal of the class is implementation of some recent bayesian and machine learning tools on standardized datasets.

Implementation of Models

The models studied in the class would be implemented on the following datasets (some other datasets may be allowed depending on the student interest and data availability)

- HMDA data
- Fannie Mae's Loan Performance Data
- FDIC's bank default datasets
- other datasets
  - Corporate Bankruptcies
  - Covenant Violations
  - Credit Rating Watches and Upgrades / Downgrades

Topics Covered in the Class: Credit Risk

The students are expected to choose some topics from the references provided later in the syllabus. Some examples of relevant topics are (note that some of the following models are suitable for credit risk)

- Bayesian Models
- Bayesian Survival models
- Frailty Models
- Cure Rate models
- Bayesian Model Averaging (BMA)

- Bayesian Hierarchical Mixtures of Experts
- Support Vector Machines (SVM)
- Random Forests
- Neural Networks
- Principal Component Analysis
- Classification and Regression Trees (CART)

Relevant Material

Some of the relevant material can be found in the following:

- Books or Articles available for downloading at no charge
  - Varian `http://pubs.aeaweb.org/doi/pdfplus/10.1257/jep.28.2.3`
  - Hastie et al `http://statweb.stanford.edu/~tibs/ElemStatLearn/index.html`
  - Barber `http://web4.cs.ucl.ac.uk/staff/D.Barber/pmwiki/pmwiki.php?n=Brml.Online`
- Books that can be found in the library
  - Bishop `http://research.microsoft.com/en-us/um/people/cmbishop/prml/`
  - Murphy `http://www.cs.ubc.ca/~murphyk/MLbook/`

These are steps that students are expected to follow in the class

1. Read Varian's paper listed above carefully. Pay special attention to *causality* and *correlation* in the context of prediction.
2. Read Harvey's paper carefully
3. Form a group of three students
4. browse through the topic list, reference material (in the syllabus) and own research. Discuss within the group and come up with a particular modeling technique that is applicable to credit risk
5. Make a group presentation to me about the modeling approach. The idea is to make a case for the model and convince me. Note that I don't expect you to understand the model completely at this stage. But you should think about the following questions

- what are the main features of the model
- which type of problem is the model most suitable for
- which type of data is the model suitable for
- why is the model applicable to credit risk
- what are the advantages and disadvantages of the model
- can it be implemented in SAS? Python? R?
- why do you think the model will beat a simple hazard or logistic model (benchmark that will be used)

6. At a later date, make a presentation to the entire class in more depth about the model. I will provide more details on the presentation later.

7. The next step is to implement the model on the standard datasets provided to the class. The performance would be measured against a bechmark model (a simple hazard model with the same covariates)

8. All groups have to submit a succinct report on the model, estimation and results. Detailed instructions on the report structure would be given later