Does VIX give us more information? – Professor Qin Lu

Outline: VIX is a volatility index for the S&P 500 index return produced by the Chicago Board Option Exchange. The anticipated volatility of the stock market is not directly observable. However, VIX is an index that attempts to indicate the invisible anticipated volatilities underlying option prices. If the VIX rises, the anticipated market volatility is increasing and vice versa. Thus, VIX contains information about anticipated volatility. In 2007, Duan and Yeh published a paper "Jump and Volatility Risk Premiums Implied by VIX", which lays the foundation of the research in this direction. They link the latent volatility to the VIX index via a new theoretical relationship under the risk-neutral measure. In the estimation of parameters, the maximum likelihood method and Bayesian method are the two most commonly used methods. Duah and Yeh use the maximum likelihood method, while we will be using the Bayesian estimation method.

Research objectives: Our project's theoretical framework, hypothesis and results are parallel to Duan and Yeh's. We use a different method to deal with the problem and we can analyze the advantage and disadvantage of the respective techniques. The comparison of the two methods using a variety of models has generated numerous papers in finance.

Outcomes: From mathematical and statistical tool side, the Bayesian method is becoming more and more popular due to the increase of computational capacity. Students can apply the method in different context in their future research. From the finance side, VIX index has a forecast power. We could understand the stock market and its movement better.

Prerequisites: Calculus based probability and statistics, some programming experience

Statistics and Breakpoint Analysis – Professor Jeffrey Liebner

Outline: Many statistical models exist to describe the data and observations that occur in the world around us. These include the familiar ordinary least squares regression model and more complex models such as auto-regression and moving average models to address time series data such as stock prices or climate changes. One challenge with these types of models is that the structure that describes the behavior of the data may suddenly change. These actions cause a break in the model, often called breakpoints or structural changes. Bai and Perron (1998) developed a commonly used frequentist technique to identify these breaks. However, from a Bayesian perspective, this technique is limited in that it does not treat the number of breaks or their location as random variables. There have been various approaches to address this problem from a Bayesian perspective, including Barry and Hartigan (1993) and others. This project will extend work started in an earlier REU in 2018 which used an MCMC routine to identify the number and location of the breaks for the AR(1) models for which the routine was shown to be effective. The technique will be generalized to other classes of models.

Research objectives: To develop a Bayesian statistical tool to detect changes or breakpoints in underlying models, including regression models and time series models. The students will create an algorithm that proposes locations for these breakpoints adjusts the locations in an adaptive manner. Simulations will be performed to assess the quality of the final fit and to compare the technique to existing techniques. The created tool will also be applied to real data to explore the possibility of changes in underlying behavior.

Outcomes: Students will modify and develop code to implement the breakpoint detection algorithm. This code will be applied to data (e.g. stock returns, global temperature data) to explore the nature of the underlying behavior.

Prerequisite: Calculus-based probability and statistics, some programming experience

A Monte Carlo Investigation of the Sampling Variability of Information Criteria for Model Selection and Post-Selection Inference in Mixed Effects Models – Professor Trent Gaugler

Outline: It is relatively common to model variance-covariance matrices when analyzing data via mixed effects models. Indeed, one of the major strengths of such models is their ability to handle data that are not iid, and often in complex ways. For example, the MIXED procedure in SAS offers users 37 different covariance structures to choose from within the 'repeated' statement; see https://documentation.sas.com/doc/en/pgmsascdc/9.4_3.4/statug/statug_mixed_syntax14.htm. As there are so many competing options, a practitioner will want to use some sound objective criterion to choose the best one. Several competing versions of information criteria (e.g. BIC, AIC, AICc) have been proposed as tools for deciding between competing mixed effects models for a sample of data. Some benchmarks have been suggested to decide when observed differences in an information criterion are meaningful and should prompt a practitioner to choose one model over another (see Kass and Raftery, 1995). Though the benchmarks are given as ranges, they may not fully account for the variability inherent in the information criteria statistics. This investigation will explore the characteristics of the data that may be associated with the need for different benchmarks. Further, the investigation will explore the properties of the post-selection inference (type I error rates and power) for parameters.

Research objectives: To evaluate the use of information criteria as a tool for model selection in mixed effects modeling. The students will perform extensive simulations of models under varied dataset characteristics to understand how information criteria work to select the correct model as a function of the simulated data, and how the post-selection inference about model parameters behaves.

Outcomes: The students will write extensive code to perform simulations, and gather the resulting data from these simulations to analyze for patterns in performance. We hope to gain insight into the nuanced behavior of information criteria, and use this to refine approaches to model selection in mixed effects modeling.

Prerequisites: Basic statistics, programming experience

References:

J. Bai and P. Perron. Estimating and testing linear models with multiple structural changes. *Econometrica*. **66** (1998) 47-78.

D. Barry and J.A. Hartigan. A Bayesian analysis for change point problems. *Journal of the Amer. Stat. Assoc.* **35(3)** (1993) 309-319.

J. Duan and C. Yeh,. Jump and Volatility Risk Premiums Implied by VIX (March 30, 2008). Available at SSRN: https://ssrn.com/abstract=966682 or http://dx.doi.org/10.2139/ssrn.966682

R.E. Kass and A.E. Raftery. Bayes Factors. *Journal of the American Statistical Association*, **90:430** (1995) 773-795. DOI: 10.1080/01621459.1995.10476572

G.H. Lubke, I. Campbell, D. McArtor, P. Miller, J. Luningham, and S.M. van den Berg. Assessing Model Selection Uncertainty Using a Bootstrap Approach: An update. *Struct Equ Modeling*. 24(2):230-245. (2017). DOI:10.1080/10705511.2016.1252265