

# DISSERTATION PROPOSAL

Vitaly Merso

“Essays on Novel Methods in Empirical Finance”

Monday, December 3, 2018

10:00 am

Tepper Quad 5219

My dissertation consists of three chapters exploring applications of novel statistical methods to questions in Empirical Finance. The first two chapters adapt models from the Natural Language Processing literature, and the third one -- a firm-consumer shock decomposition from applied economics literature.

In the first chapter of my dissertation, I use statistical models of language to quantify the informativeness of earnings announcements language for investors and identify the textual features most associated with the market activity. I find evidence that language and numbers are substitutes for investors to a large extent, but contain a complementary component. On aggregate, the announcement language explains 11% of the variation in absolute returns around the release date out-of-sample, relative to 15% explained by contemporary values of accounting and market-based numerical variables. The best explanatory power is achieved by combining textual and numerical data. Mentions of earnings components are associated with repricing, suggesting that investors pay attention to how the firm arrived at its earnings value.

In the second chapter, joint with Bryan Routledge, we analyze a corpus of approximately 30,000 semi-annual letters from investment fund leadership to their shareholders. The letters are not structured and cover a range of topics about the stock market, the fund's performance, and the macro-economy. For these economic quantities, the letters touch both on the past and the outlook. We use the Hierarchical Attention Network (HAN) to analyze the narrative text. Specifically, we use the text to predict measured economic quantities such as aggregate stock market performance, aggregate market volatility (VIX), and measures of the fund performance. The main advantage of the HAN in this setting is that it produces a more interpretable mapping despite the complex non-linearities of the neural architecture. The model allows us to see the sentences and phrases driving the prediction. This interpretability is helpful in addressing interesting economic issues such as the difference between investment funds with a different structure: mutual and closed-ended funds.

In the third chapter, I provide evidence that bank and non-bank mortgages can be either complements or substitutes depending on local demand conditions. Normally, shadow lending grows faster in counties exposed to increases of bank credit supply. However, non-bank mortgages can become substitutes for bank mortgages in counties that experience an increase in demand. At the same time, non-bank lenders' own supply changes explain more variance of shadow lending growth than bank supply changes. My findings suggest that decreases in bank supply, including the ones due to increased regulation, are unlikely to be the main driver of shadow lending growth. Other factors, like operational differences or technology, can have a stronger impact.