# Modeling Latent Variables in Economics and Finance

**By**


## Keven Bluteau


PhD thesis submitted for the joint degree of

Doctor of Finance (UniNE)
&
Doctor in Business Economics (VUB)

Promotors: Prof. David Ardia (UniNE) and Prof. Kris Boudt (VUB)




Institute of Financial Analysis
University of Neuchâtel
Switzerland

Solvay Business School
Vrije Universiteit Brussel
Belgium


April 17, 2019

# Abstract

The subject of unobservable variables encompasses this thesis. These latent (*i.e.*, unobservable) variables must be inferred using statistical models or observable proxies. The objectives of my doctoral thesis are to develop and test new statistical models to infer these variables and link them to the analysis and improvement of economic and financial decisions.

In my first essay, I tackle the evaluation of volatility models which allow for (latent) structural breaks. It is of utmost importance to capture these breaks in a timely manner, as a precise measure of volatility is crucial for optimal decision–making that requires a trade–off between expected return and risk, as well as for applications in asset pricing and risk management. However, no empirical study has been done to evaluate the overall performance of volatility model considering structural breaks. To that end, I perform a large–scale empirical study to compare the forecasting performance of single–regime and Markov–switching GARCH (MSGARCH) models, from a risk management perspective. I find that, for daily, weekly, and ten–day equity log–returns, MSGARCH models yield more accurate Value–at–Risk, Expected Shortfall, and left–tail distribution forecasts than their single–regime counterpart. Also, my results indicate that accounting for parameter uncertainty improves left–tail predictions, independently of the inclusion of the Markov–switching mechanism.

While my first essay tackles the modeling of latent variables from a statistical point of view, my second and third essay capture a more novel variable, namely the sentiment expressed in written communications.

My second essay addresses the development and testing of new text–based proxies for economic sentiment. More specifically, I introduce a general sentiment engineering framework that optimizes the design for forecasting purposes in a high–dimensional context. I apply the new methodology to the forecasting of the US industrial production, which is usually predicted using available quantitative variables from a large panel of indicators. I find that, compared to the use of high–dimensional forecasting techniques based solely economic and financial indicators, the additional use of optimized news–based sentiment values yield significant forecasting accuracy gains for the nine–month and annual growth rates of the US industrial production.

My third essay focuses on the analysis of the dynamics of abnormal tone or sentiment around the time of events. To do so, I introduce the Cumulative Abnormal Tone (CAT) event study and Generalized Word Power methodologies. I apply these methodologies to media reports

in newswires, newspapers, and web publications about firms' future performance published around the quarterly earnings announcements of non–financial S&P 500 firms over the period 2000–2016. I find that the abnormal tone is more sensitive to negative earnings surprises than positive ones. Additionally, I report that investors overreact to the abnormal tone contribution of web publications at earnings announcement dates, which generates a stock price reversal in the following month. This result is consistent with an overreaction pattern on the abnormal tone and psychological biases such as the representativeness heuristic. Moreover, it highlights that there is heterogeneity in the informational value of different types of media.

**Keywords:** abnormal return, abnormal tone, earnings announcements, elastic net, expected shortfall, forecasting performance, GARCH, generalized word power, large–scale study, MSGARCH, news media, risk management, sentiment analysis, sentometrics, textual tone, time–series aggregation, topic–sentiment, US industrial production, value–at–risk