Quantitative Finance and the Critical Role of Statistics and Data Science

Katherine B. Ensor, Ph.D.
Editor, Applied Stochastic Models in Business and Industry

Noah G. Harding Professor of Statistics
Director, Center for Computational Finance and Economic Systems
Rice University, Houston, Texas

2021 ISBIS Webinar Series
Leading Women in Business and Industry
January 21, 2021
Much of today’s global prosperity

• Can be attributed to the intelligent use of modern statistics.
  – In technology / and certainly in social media
  – In manufacturing
  – In product development
  – In food production
  – In drug discovery and patient care
  – In healthier societies

• And most certainly in the global financial industry
From strong foundations....

Statistics has changed the world, with more to come.

Our Professions AMAZING History

Machine, Deep Learning and Artificial Intelligence – at the base is Multivariate Statistics, Time series functional methods...

The interplay between strong statistical paradigms, ubiquitous statistical computing, algorithms, and large and fast computers ensures a world increasingly reliant on data for knowledge, and action.
Global Financial Markets

Then
Brookmire (1913)
The American Economic Review

Now
Data Science is Ubiquitous
• Research and education in quantitative finance at Rice
• Collaboration between engineering (statistics), social sciences (economics) and business (finance)
• Advance the quantitative study of markets AND their impact on society
• Data science MEETS finance
“Annual” Eubank Conference on Real World Markets

• 2009 Modeling Real World Markets
• 2010 Modeling Financial Markets in a World of Fiat Money
• 2011 Critical Issues in Today’s Dynamic and Uncertain Political Climate and Markets
• 2012 Creating Growth, Entrepreneurship and Analytics
• 2015 A Focus on Energy
• 2017 High Frequency Trading: Mitigating its Impact on Trading and Investing
• 2018 Intelligent Advising in the Fintech Era
• **April 26, 2021 Sports Analytics and Gaming Risk Management**
Show me the money...then.

AMERICAN STATISTICAL ASSOCIATION.

NEW SERIES, No. 92. DECEMBER, 1910.

THE CORRELATION OF ECONOMIC STATISTICS.

BY WARREN M. PERSONS, ASSISTANT PROFESSOR OF ECONOMICS, DARTMOUTH COLLEGE.
The cause and effect relation existing between economic events is especially difficult to ascertain because of the presence of innumerable variable elements. In solving his problems the economist cannot, like the physicist or chemist, eliminate all causes except one and then by experiment determine the effect of that one. Causes must be dealt with en masse.

It is rarely, if ever, possible for the economist to state more than “such and such a cause tends to produce such and such an effect.” Events can only be stated to be more or less probable. He is dealing mainly, therefore, with correlation and not with simple causation.
Persons 1910 cont. Correlation not causation...

Just as the biologists cannot predict a man's height or color of eyes or temper or combativeness by knowing those qualities in his ancestors, so economists cannot predict that a definite call rate in Wall Street will go with a given percentage of reserves to deposits in New York banks or that a given supply of wheat will result in a definite price per bushel. But, on the other hand, just as it has been observed that there is a relation existing between a man's stature and the stature of his ancestors, so it has been observed that a relation does exist between bank reserves and call rates and between supply of wheat and its price per bushel.
Corn Production of United States and Farm Prices, 1866-1906.

- Full line = Million bushels of corn
- Dotted line = Price in cents per bushel

The line shows the course that production must take in order to keep the price uniform, according to the equation of regression.
The importance of business forecasting

Methods of Business Forecasting based on Fundamental Statistics

James H. Brookmire

Stable URL: http://www.jstor.org/stable/1828258

METHODS OF BUSINESS FORECASTING BASED ON FUNDAMENTAL STATISTICS

A business man succeeds or fails in proportion to his ability to forecast the future trend of the influences determining the relation of supply and demand in his business. Some of these influences are technical and concern his particular business without affecting other lines; others are of a general nature and affect all lines of business definitely and vitally.

Compares
• “Jevons -- correlates commercial crises with fixed tendencies and solar influences”
• “Brenner -- shows that panics were periodic”
• “Babson --- believes in periodicity of plotted areas”

• Brookmire -- “Professor Fisher’s ‘Equation of Exchange,’ however, is used, not as a mechanical forecaster in and of itself, but merely as a picture of past and present conditions. The future is not indicated, but must be judged from the general outlook, using the diagram as a guide.”
Babson Periodicity of Plotted Intervals

Chart C.—Babson’s area plot, showing periods of prosperity and depression.
Bookmire’s 1913 composite chart of fundamentals

CHART E.

Bookmire's composite chart of business, banking, and investment conditions, showing cycles of banking credit.
Global Financial Markets

**Algo(rithmic) Trading**
- Automated
- Statistics, data science and AI have transformed global financial markets. *AND will continue to do so.*
- New risks, new rewards.
- Critically important that we have smart statistical scientists in this industry. *An industry which underlies our future.*

**New modeling needs**
- Market dynamics have changed.
- Adaptive, models and methods.
- Big data, with temporal and strong network structure.
- Understand the interconnected nature of the global system.
- ROBUST, FAST and EFFICIENT
Modern Finance & Statistics

- Complex data types and structures
- Massive data to “little” data
- Difficult dependencies in all dimensions
- Mixtures
- Dynamic and evolving
- Robust

  - Statistical Graphics
  - Multivariate dynamic time series
  - Bayesian methods
  - Nonparametric methods
  - Simulation and Resampling
  - Statistical / machine learning
  - Network models (Neural, Bayes, General)
  - Regression trees
  - Hierarchical models
  - Functional Data Analysis
  - Categorical methods
  - Nonlinear regression
  - Dependent series
  - Stochastic processes
  - Survival Analysis
  - Agent based model
  - Anomaly detection

Ensor, 2013, WIRES

120 Stock Return Series
10 Sectors
Finance - What about today?

- **Double wow – the financial world is REALLY open to the quantitatively savvy individual.**
- Still “insider” info
- A LOT OF data is available; SOME is not
- Portfolio management is possible – but takes time – many of us don’t have this kind of time.
- Should you turn over management of your investments to financial analyst with an undergraduate education (or basic MBA)??? The million dollar question -- LITERALLY
  - maybe yes if they keep educating themselves
  - a definite no if they stop at a basic college CAPM, normal based education
    - It is a full time job
    - Or invest in a highly rated index fund
- **There can be huge value in fundamental analysis with statistical guidance**
- Algorithmic traders represent the competition on the day-to-day real-time trading
Nobels for Nonsense....

J.R. Thompson SIMUGRAM

• Award winning paper
• Markets are not efficient
• Simugram - Novel strategy for choosing portfolio weights
• Subject of multiple successful dissertations
• Patent protected!
• The “poor man’s” version that works well – the max median rule.
• Founding chair of Rice Statistics, and key leader of CoFES
QUESTION: IS THERE AN OPTIMAL PORTFOLIO DESIGN?
IT’S ALL ABOUT THE WEIGHTS
Anomaly Detection

Modeling Correlation

• QUESTION: Is there a stock that whose basic trading characteristics have changed relative to the market?
  • Basic technology – multi-layer multivariate GARCH model with REGIME SWITCHING
  • Capitalize on correlation WITHIN and BETWEEN financial sectors
  • FAST, EFFICIENT and ROBUST anomaly detector
  • Uses daily closing PRICE and VOLUME (publicly available information)
The Fall of Enron Timeline

- 10/4 our methods identify the futility of the situation
- 10/16 announced nonrecurring losses of $1 billion
- Competing state of the art statistical tools, pinpoint the problem on 11/28 after junk status is achieved

An example: Identifying Anomalous Behavior

Koev and Ensor, 2006

11/21 junk status
12/1 bankruptcy
Dynamic Outlier Detection

Stock market example

• Dynamic multivariate advanced time series model coupled with dimension reduction strategies.


The decline of Lehmann Brothers

Caution
Modern Statistics is powerful

*Estimating the Term Structure With a Semiparametric Bayesian Hierarchical Model: An Application to Corporate Bonds*

Alejandro CRUZ-MARCELO, Katherine B. ENSOR, and Gary L. ROSNER, JASA 2011

- Estimated Term Structure for each Corporate Bond in the Data Set
- Method “borrows strength” from similar groups.
- Classification is semi-parametric and robust
Modern Statistics used well can make a big difference...

The above estimated yield curves are based on methods proposed as late as 2004; fall in to the class of non-linear regression strategies. Undiscovered challenge was the mixture of bond behaviors below AAA ratings.

<table>
<thead>
<tr>
<th>Method</th>
<th>Dollar</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
</tr>
<tr>
<td>SC-rating</td>
<td></td>
</tr>
<tr>
<td>BHM-rating</td>
<td></td>
</tr>
<tr>
<td>BHM-firm</td>
<td></td>
</tr>
<tr>
<td>SC-rating</td>
<td></td>
</tr>
<tr>
<td>BHM-rating</td>
<td></td>
</tr>
<tr>
<td>BHM-firm</td>
<td></td>
</tr>
<tr>
<td>SC-rating</td>
<td></td>
</tr>
<tr>
<td>BHM-rating</td>
<td></td>
</tr>
<tr>
<td>BHM-firm</td>
<td></td>
</tr>
<tr>
<td>SC-rating</td>
<td></td>
</tr>
<tr>
<td>BHM-rating</td>
<td></td>
</tr>
<tr>
<td>BHM-firm</td>
<td></td>
</tr>
<tr>
<td>SC-rating</td>
<td></td>
</tr>
<tr>
<td>BHM-rating</td>
<td></td>
</tr>
<tr>
<td>BHM-firm</td>
<td></td>
</tr>
</tbody>
</table>

Ensor, Eubank Conference 2011
Dynamic Continuous Time Models
Impact of News on Oil Futures

Crude Oil Futures: Price

NEW QUESTIONS
WE CAN ANSWER
Dynamic jump intensities and news arrival in oil futures markets

Katherine B. Ensor\textsuperscript{1} • Yu Han\textsuperscript{1} • Barbara Ostdiek\textsuperscript{2} • Stuart M. Turnbull\textsuperscript{3,4}

Revised: 18 May 2020 / Published online: 26 June 2020
© The Author(s) 2020

Abstract
We introduce a new class of discrete-time models that explicitly recognize the impact of news arrival. The distribution of returns is governed by three factors: dynamics volatility and two Poisson compound processes, one for negative news and one for positive news. We show in a model-free environment that the arrival of negative and positive news has an asymmetric effect on oil futures returns and volatility. Using the first 12 futures contracts, our empirical results confirm that the effects of negative and positive news are described by different processes, a significant proportion of volatility is explained by news arrival and the impact of negative news is larger than that of positive news.
Background – Oil Futures

• Well-developed derivative markets for oil
• To price and hedge derivative contracts written on oil, it is necessary to understand the price dynamics
• Derivative pricing literature for oil is relatively sparse compared to the existing literature on equity derivatives.
• Earlier authors argue that allowing for jumps is crucial for modeling crude oil futures.
• It is also thought that news impacts crude oil futures.
Modeling the impact of News on Futures Market

• Schwartz (1997) developed a classic 3 factor model for commodity futures.
• We adopted a similar approach and extended to a 5 factor to include a:
  – GARCH stochastic volatility process
  – News process
    • Number of negative news items and the size of the jump
    • Number of positive news items and the size of the jump
Enter the “big data” era

• Collection and analysis of large data banks detailing different types of news articles is growing.
• Machine Learning algorithms are used to classify news, and its sentiment.
• We can now incorporate the arrival of information into our asset pricing model dynamics.
• We explicitly include the news arrival processes in our pricing.
Futures Data

- Light Sweet Crude oil futures (ticker WTI) contracts traded on the Chicago Mercantile Exchange (CME)
- Daily settlement prices, trading volume and open interest from January 5, 2004 through December 28, 2012
- 2261 trading days in our sample
- Contracts mature approximately 3 weeks into the delivery month, and delivery can take place earlier than the first calendar day of the delivery month.
- Consecutive months are listed for the current year and the next five years.
- The first 6 months tend to be the most liquid
- For contracts with maturities greater than 6 months, trading is concentrated in contracts expiring March, June, September and December
- To avoid the delivery period
  - We roll the futures contracts before they have less than one month to maturity
  - Maturity is defined as the first day of the delivery month
Futures Data

- F1 price
- F1 log returns
- F1/F6 ratio
- F1 realized volatility
- **BUBBLE**: Dec 1, 2007 through January of 2009: Doubling of oil prices
- Post January 2009: collapse with sustained period of high volatility
Interest Rate Data

- 3-month U.S. Treasury bill interest rate
- Varied greatly over our study period

Table 2  Interest rate data

<table>
<thead>
<tr>
<th>Year</th>
<th>Mean</th>
<th>Median</th>
<th>SD</th>
<th>Max</th>
<th>Min</th>
<th>ACF(1)</th>
<th>ACF(3)</th>
<th>ACF(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2004</td>
<td>0.0140</td>
<td>0.0134</td>
<td>0.0045</td>
<td>0.0226</td>
<td>0.0087</td>
<td>0.26</td>
<td>0.04</td>
<td>0.15</td>
</tr>
<tr>
<td>2005</td>
<td>0.0322</td>
<td>0.0315</td>
<td>0.0052</td>
<td>0.0408</td>
<td>0.0231</td>
<td>0.23</td>
<td>0.14</td>
<td>0.26</td>
</tr>
<tr>
<td>2006</td>
<td>0.0485</td>
<td>0.0492</td>
<td>0.0024</td>
<td>0.0513</td>
<td>0.0416</td>
<td>0.20</td>
<td>0.10</td>
<td>0.23</td>
</tr>
<tr>
<td>2007</td>
<td>0.0447</td>
<td>0.0484</td>
<td>0.0071</td>
<td>0.0519</td>
<td>0.0287</td>
<td>0.10</td>
<td>0.16</td>
<td>0.22</td>
</tr>
<tr>
<td>2008</td>
<td>0.0139</td>
<td>0.0155</td>
<td>0.0081</td>
<td>0.0327</td>
<td>0.0000</td>
<td>0.26</td>
<td>0.16</td>
<td>0.32</td>
</tr>
<tr>
<td>2009</td>
<td>0.0015</td>
<td>0.0016</td>
<td>0.0007</td>
<td>0.0032</td>
<td>0.0002</td>
<td>0.11</td>
<td>0.14</td>
<td>0.23</td>
</tr>
<tr>
<td>2010</td>
<td>0.0014</td>
<td>0.0015</td>
<td>0.0003</td>
<td>0.0018</td>
<td>0.0004</td>
<td>0.21</td>
<td>0.07</td>
<td>0.05</td>
</tr>
<tr>
<td>2011</td>
<td>0.0005</td>
<td>0.0003</td>
<td>0.0005</td>
<td>0.0016</td>
<td>0.0000</td>
<td>0.31</td>
<td>0.22</td>
<td>0.19</td>
</tr>
<tr>
<td>2012</td>
<td>0.0009</td>
<td>0.0009</td>
<td>0.0002</td>
<td>0.0014</td>
<td>0.0001</td>
<td>0.20</td>
<td>0.08</td>
<td>0.07</td>
</tr>
</tbody>
</table>
### Panel A: F1 Price During Oil Bubble

![Graph of F1 Price During Oil Bubble]

### Panel B: F1 Return During Oil Bubble

![Graph of F1 Return During Oil Bubble]

<table>
<thead>
<tr>
<th>Label</th>
<th>Date</th>
<th>$R_{t-3}$</th>
<th>$R_{t-2}$</th>
<th>$R_{t-1}$</th>
<th>$R_t$</th>
<th>News</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>2008-11-04</td>
<td>-0.02</td>
<td>0.03</td>
<td>-0.05</td>
<td>0.10</td>
<td>OPEC countries announce major cut; Federal Reserve loans $133 billion to credit facilities</td>
</tr>
<tr>
<td>P2</td>
<td>2008-12-11</td>
<td>0.08</td>
<td>-0.04</td>
<td>0.03</td>
<td>0.10</td>
<td>Saudi propose big production cut plan; IEA report shows higher growth for oil and gas</td>
</tr>
<tr>
<td>P3</td>
<td>2008-12-31</td>
<td>0.06</td>
<td>0.06</td>
<td>-0.03</td>
<td>0.13</td>
<td>US oil inventory rises half as much as expected</td>
</tr>
<tr>
<td>P4</td>
<td>2009-01-02</td>
<td>0.06</td>
<td>-0.03</td>
<td>0.13</td>
<td>0.12</td>
<td>US oil inventory rises half as much as expected</td>
</tr>
<tr>
<td>P5</td>
<td>2009-04-02</td>
<td>-0.08</td>
<td>0.03</td>
<td>-0.03</td>
<td>0.12</td>
<td>G20 summit shows optimism on economy; FASB announces new accounting fair market value accounting standard</td>
</tr>
<tr>
<td>N1</td>
<td>2008-09-29</td>
<td>-0.01</td>
<td>0.02</td>
<td>-0.01</td>
<td>-0.10</td>
<td>US House rejects bailout act</td>
</tr>
<tr>
<td>N2</td>
<td>2008-10-10</td>
<td>0.02</td>
<td>-0.00</td>
<td>-0.02</td>
<td>-0.10</td>
<td>Global equity market crashes (Dow down 697 points)</td>
</tr>
<tr>
<td>N3</td>
<td>2008-12-01</td>
<td>-0.07</td>
<td>0.07</td>
<td>-0.00</td>
<td>-0.10</td>
<td>OPEC defers production cuts</td>
</tr>
<tr>
<td>N4</td>
<td>2009-03-02</td>
<td>0.06</td>
<td>0.06</td>
<td>-0.01</td>
<td>-0.11</td>
<td>US reports disappointing GDP data</td>
</tr>
</tbody>
</table>
Panel A: F1 Realized Variance During Oil Bubble

Panel B: F1 Realized Variance Daily Change During Oil Bubble

<table>
<thead>
<tr>
<th>Label</th>
<th>Date</th>
<th>$\Delta \sigma_i^2$</th>
<th>$\Delta \sigma_{i-1}^2$</th>
<th>$\sigma_{i-2}^2$</th>
<th>$\sigma_{i-1}^2$</th>
<th>$\sigma_i^2$</th>
<th>News</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>2008-07-17</td>
<td>-0.23</td>
<td>0.70</td>
<td>0.36</td>
<td>0.13</td>
<td>0.83</td>
<td>Bernanke sounded bleak note on the growth of the U.S. economy (Dow down 697 points)</td>
</tr>
<tr>
<td>P2</td>
<td>2008-10-16</td>
<td>-0.14</td>
<td>0.39</td>
<td>0.28</td>
<td>0.15</td>
<td>0.54</td>
<td>EIA reports crude stocks up 5.6 million barrels vs forecast of 1.9 million</td>
</tr>
<tr>
<td>P3</td>
<td>2008-11-13</td>
<td>-0.05</td>
<td>0.27</td>
<td>0.25</td>
<td>0.20</td>
<td>0.47</td>
<td>IEA cuts world oil demand forecast on weak economy</td>
</tr>
<tr>
<td>P4</td>
<td>2008-12-10</td>
<td>-0.05</td>
<td>0.47</td>
<td>0.32</td>
<td>0.27</td>
<td>0.74</td>
<td>EIA predicts world oil demand to fall for first time in decades</td>
</tr>
<tr>
<td>P5</td>
<td>2008-12-31</td>
<td>-0.14</td>
<td>0.47</td>
<td>0.37</td>
<td>0.23</td>
<td>0.70</td>
<td>EIA reports crude stock up 500,000 barrels vs forecast of 1.5 million</td>
</tr>
</tbody>
</table>
News Data

- Thomson Reuters News Analytics (TRNA) database
- Each news item is time stamped with day and time of arrival
- News articles are classified as alerts, articles, append or overwrite
- TRNA publishes ML algorithm that
  - provides probabilities of positive, negative or neutral news
  - Relevance score between 0 and 1
    - Looks at the number of occurrences of asset relative to the number of occurrences of other organizations and commodities
    - If asset is mentioned in the headline, the relevance is set to 1
- Our filtering
  - Relevance of 1
  - Exclude if “PRICE” is mentioned in the title – want to focus on news about “OIL” not “OIL PRICES”
  - Strong positive and negative sentiment cutoff of 0.75
News Summary

• Years 2004 through 2012
• Average daily arrivals
  – 358 total : 165 positive : 59 neutral : 134 negative
• Filtering on relevance 1
  – 280 total : 96 positive : 43 neutral : 91 negative
• Counts elevated in 2007, 2008 and 2011
F1 Returns Volatility and Daily Net Major News Counts

• Median of returns increases with increase in net news
• Negative news is associated with low returns
• Positive news is associated with high returns
• The return volatility is highest for when there is less net news during the bubble period

\[ \text{Net News}_t \equiv \sum_j 1(p_j^+ \geq \mu) - 1(p_j^- \geq \mu) \]

\[ \equiv \text{Positive News}_t - \text{Negative News}_t \]
# News Data with Relevance Filter

## Table 3: News data with relevance filter

<table>
<thead>
<tr>
<th>Year</th>
<th>All news</th>
<th>News with relevance score of 1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total</td>
<td>$n^+$</td>
</tr>
<tr>
<td>2004</td>
<td>277</td>
<td>127 (46%)</td>
</tr>
<tr>
<td>2005</td>
<td>319</td>
<td>159 (50%)</td>
</tr>
<tr>
<td>2006</td>
<td>416</td>
<td>216 (52%)</td>
</tr>
<tr>
<td>2007</td>
<td>409</td>
<td>214 (53%)</td>
</tr>
<tr>
<td>2008</td>
<td>430</td>
<td>213 (49%)</td>
</tr>
<tr>
<td>2009</td>
<td>363</td>
<td>174 (47%)</td>
</tr>
<tr>
<td>2010</td>
<td>316</td>
<td>125 (40%)</td>
</tr>
<tr>
<td>2011</td>
<td>373</td>
<td>139 (37%)</td>
</tr>
<tr>
<td>2012</td>
<td>315</td>
<td>110 (35%)</td>
</tr>
<tr>
<td>Average</td>
<td>358</td>
<td>164 (45%)</td>
</tr>
</tbody>
</table>

### Panel A: Average daily news count

### Panel B: Explanatory power of news for returns and changes in realized variance

<table>
<thead>
<tr>
<th></th>
<th>All news</th>
<th>News with relevance score of 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R^2_{A}$ (%)</td>
<td>0.14</td>
<td>0.14</td>
</tr>
<tr>
<td>$R^2_{AN}$ (%)</td>
<td>6.56</td>
<td>9.34</td>
</tr>
<tr>
<td>$\Delta_{A}$</td>
<td>6.42***</td>
<td>9.20***</td>
</tr>
<tr>
<td>$\Delta_{AN}$</td>
<td>79.04</td>
<td>116.17</td>
</tr>
<tr>
<td>$W_{AN}$</td>
<td>0.02</td>
<td>0.06</td>
</tr>
<tr>
<td>$R^2_{A}$ (%)</td>
<td>-0.12</td>
<td>-0.08</td>
</tr>
<tr>
<td>$\Delta_{A}$</td>
<td>0.18</td>
<td>0.54</td>
</tr>
</tbody>
</table>

Panel A provides the average daily count for total, positive ($n^+$), neutral ($n^0$) and negative ($n^-$) news items for each year in the sample period for all news items and for news items with relevance score of 1. The percentage of news items in each category is provided in parentheses. Panel B provides diagnostic statistics for an AR(1) regression model of daily returns and daily log change in realized variance for the nearby futures contract (F1) for all news and for news with relevance score of 1. $R^2_{A}$ is the adjusted $R^2$ for the basic AR(1) model: $\hat{z}_t = a_0 + a_1 z_{t-1} + \epsilon_t$. $R^2_{AN}$ is the adjusted $R^2$ for an AR(1) model that includes contemporaneous positive and negative news counts: $\hat{z}_t = a_0 + a_1 z_{t-1} + b^+ n^+_t + b^- n^-_t + \epsilon_t$. $\Delta_{AN}$ ($\Delta_{A}$) is the percentage change in the adjusted $R^2$ when contemporaneous (lagged) news variables are included in the model. $W_{AN}$ ($W_{A}$) is the Wald test F statistic for adding the contemporaneous (lagged) news variables, with degrees of freedom (2, 2257). *** (**) indicates significance at 0.001 (0.01) level. The sample period is January 5, 2004, to December 28, 2012 (2661 trading days).
Pricing dynamics in a DISCRETE Time Economy

THE MODEL

- \( F(t,T) \) denotes the futures price at time \( t \) of a contract that matures at time \( T \)
- Over a horizon \([0,T]\) there are \( m \) intervals of equal length \( \Delta \)
- \( S(t) \) is the commodity spot price
- \( X(t) = \ln(S(t)) \)
- The dynamics of \( X \) under the \( P \) measure are given by

\[
X_{t+1} - X_t = (r_t - \delta_t) \Delta - l_t + e^{X_{t+1}}
\]
\[ e_t^X = e_t^V + e_t^J \]

- **GARCH(1,1) -**

\[ e_t^V = \sqrt{h_{t-1}} \Delta e_t^V \]

- Impact of jumps caused by the random arrival of news

\[ e_t^J = \sum_{k=1}^{N^+} \theta_k^+ + \sum_{k=1}^{N^-} \theta_k^- \]
Other model features

• Incorporate dynamics of convenience yield and interest rate
• Covariances between
  – Spot price
  – convenience yield
  – spot interest rate
• The pricing model is derived through a change to the pricing measure, so that

\[
F(t, T) = E_t^Q[S_t]
\]
• Variances are also derived.
Model Specification

We can now derive the model under the pricing measure $Q$. For a futures contract that matures at time $T$, for $t \leq T$ with $n$ intervals to the maturity date the futures price is modeled by:

$$F(t, T) = E_t^Q[S_T] = \exp\left(B_n + D_n^{(X)} X_t + D_n^{(r)} r_t - D_n^{(\delta)} \delta_t + G_n^{(h)} h_t + \sum_{j \in A} D_n^{(j)} n_t^{(j)}\right)$$

with

- $X_t = \ln(S_t)$ is the log spot price.
- $r_t$ is the interest rate process.
- $\delta_t$ is the convenience yield process.
- $h_t$ is the volatility process.
- $D_n$ is the news process.
Estimation

• Separately estimate Interest rate process (MLE)
• Separately estimate News arrival process (MLE of intensity process)
• Simplify convenience yield process by fixing 2 parameters
• The remaining parameters were estimated using Kalman Filtering under the Q measure
What did we learn

- News helps explain prices and volatility
  - 19% of the variation in prices is DUE TO NEWS
  - 24% during the bubble period
- The impact of news is ASYMMETRIC
  - 10% of variation is DUE to NEGATIVE NEWS
  - 9% of variation is DUE to POSTIVIE NEWS
  - Jump size is positive (but not significant) for positive news
  - Jump size is negative and highly significant for negative news
- News arrival affects the price dynamics in a nontrivial way
- News affects all moments of the pricing distribution and hence HEDGING RATIOS
Explained Variation in Futures Market

Contribution From NEWS
Panel A: Paired Plot of Log Conditional Variance From GARCH Process for Model With and Without News

Panel B: Paired Plot of Log Conditional Variance for Model With and Without News
Risk Intertwined: Cautionary notes
Fundamentals of Markets Change
Convergence of Hedge Fund Strategies

First Period

4/04 – 12/00

Second Period

1/01 – 6/07

Third Period

July 2007 – April 2011
Red – correlation >.5
Black – negative correlation

Ensor, Marfin, Seidensticker, Miller 2011
V-lab provides an up to date, time series based volatility and risk analysis of countries and companies.

Pictured is their climate risk metric, by country.

It is a product of the Volatility and Risk Institute at NYU, lead by Robert Engle and Richard Berner

https://vlab.stern.nyu.edu
Modern Methods Used Poorly....

• Recipe for Disaster: The Formula that Killed Wall Street, Wired Magazine, 2/23/2009

\[ P[T_A < 1, T_b < 1] = \Phi_2(\Phi^{-1}(F_A(1)), \Phi^{-1}(F_B(1)), \gamma) \]

• Allowed “hugely complex risks to modeled with more ease and accuracy than ever before”.

• “But it's a very inexact science. Just measuring those initial 5 percent probabilities involves collecting lots of disparate data points and subjecting the... all manners of statistical and risk analysis.”

• Failed to capture the true joint probability of default ... strong assumptions that were not met when market dynamics changed.

• Copula is also used at the regulator level and extensively by Moody’s for ratings at the time (NISS and OCC Explorations Workshop: Financial Risk Modeling and Banking Regulations, Feb. 2009)

Extremely difficult to “fit” in practice.

Missing: A healthy dose of skepticism
FINANCIAL CRISIS
DID WE LEARN?
MAYBE, SORT OF...
MAYBE NOT
What is the quantitative investor to do?

• Ignore the market micro-structure???
• This is one of the struggles we all face
  – Use robust strategies that are not impacted by this structure
  – Properly incorporate the essence of the dynamics into factor based models
  – Rely on the law of large numbers and the central limit theorem and continue business as usual with modern portfolio theory
  – Target “stable” time epochs to place longer term positions
  – Develop model free strategies that capture the dependence structure between stocks, e.g. SIMUGRAM, Thompson et al
  – Develop dynamic time series strategies that fully capture the changing structure – if you can
• Most certainly rely on MODERN STATISTICAL methods, while fully understanding the limitations of the array of methods used, an maintain a healthy dose of skepticism.
And will continue to exist...

- We cannot regulate away systemic risk; we can only change its form.
- If market structures must change I am reminded of Pareto’s chart used so well in quality control across the world – *highlighting the most common sources of defects*
- Understanding the degree of global market dependence and systemic risk is key for all investors regardless of their investment style -- diversification is a moving target

*A common observation is convergence of many financial instruments.*
The “new breed” of statisticians/quants

• Like the “old breed”

• Understand the limitations and complexities of the methodologies and algorithms they are using

• Understand the limitations of the data itself
  – How was it collected? How should it be collected? What does the data truly measure? What statistical methodologies work for the specific type of data?
  – Sadly an often overlooked issue

• Discovering false patterns HELPS NO ONE -- and can have disastrous consequences
Our Answer
Focused Solution Based
Quantitative Financial Research
Statistical Science AND Engineering

• There is much to be gained by sophisticated problem solving related to statistics, data science and AI.
• Overarching issues
  – Data collection modalities
  – Data types, quality and representativeness
  – Complex nature of the observed variables (e.g. extreme values can really throw off answers, mixtures, categorical)
  – Dependence structure (e.g. space – time – dimension – mega dimension)
  – Robust, reproducible and replicable inferences.
  – Automation
    – Asking and answering the right questions, and in the timeframe required.
• Incorporating statistical inference features that account for data and problem complexity has high payoff. This is a lasting principal.
Decisions are based on ``data’’...good use of statistics can guide those decisions

- Decisions are based on some form of data, whether measured or subjective.
- Understanding the quality and source of data is imperative
- The S, R and Python revolution puts modern statistics in the hands of anyone.
- Appropriate use of modern statistics helps immensely toward improving all decisions – but modern statistics is complex and not well understood by the novice practitioner.
- And as in everything, a healthy dose of skepticism is essential.
- And with that I ...
Thank you
ENSOR@RICE.EDU

And join Rice+CoFES for the FREE VIRTUAL Eubank 9th Conference on Real World Markets: SPORTS ANALYTICS
APRIL 26, 2021
Cofes.rice.edu