Technomics
Better Decisions Faster

Social Media and Submarines

How machine learning and unconventional methods can change cost estimating

Presented at the 2018 ICEAA Professional Development & Training Workshop - www.iceaaonline.com

Today's Presentation

- Social Media and Cost Estimating
- Submarines!
- How accurate are price indexes anyway?
- One example: Twitter
- Can unconventional data sources replace escalation?
- Applying big data and machine learning
- Potential impact for the cost industry

Today's Presenters

Jeff Pincus

Senior Associate

Jeff Pincus is a senior associate with Technomics, focusing on data analysis, design and architecture. He has worked closely developing complex data solutions to improve federal cost estimation, as well as designing tools to make data analysis as seamless as possible. Previously, Jeff worked in big data and economic modeling for economic development project in the Americas. Mr. Pincus holds degrees in economics and government.

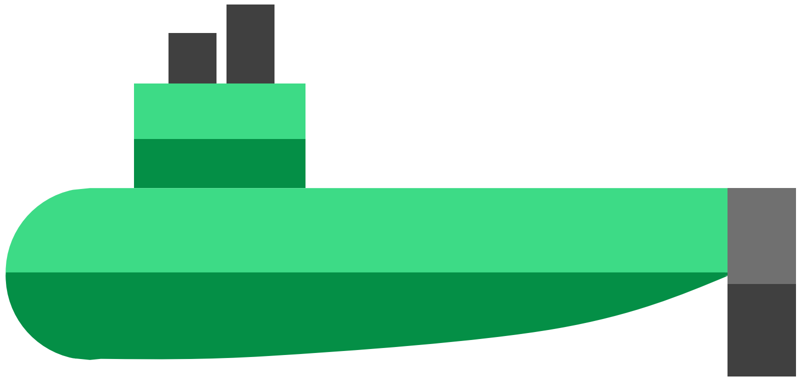
Omar Akbik

Senior Analyst

Omar Akbik is a senior analyst with Technomics specializing in analytics, cost analysis, and strategic advisory. His work focuses on interagency data and technology integration as well as quantitative analysis of federal programs. He has previously supported life cycle cost estimating, alternatives and business case analysis, as well as application development for various clients throughout the federal space. Mr. Akbik is a Certified Cost Estimator/Analyst, an Agile Certified Practitioner (PMI-ACP) and holds degrees in economics and finance

Lets have a discussion…

You're estimating a program…



- Want a submarine? Well, shipbuilding Inflation has grown ~60% faster than OSD inflation estimates
- Something else in defense? It is estimated that 25% of DOD MDAPs had cost growth in 2007. In 2015, it is estimated that **48%** of DOD MDAPs had cost growth. And they are projected to rise to 51% by 2020
- Put it in dollars: Cost growth for DoD acquisitions **exceeded $90B** over the past year
- Why are cost estimates missing the mark?
  - Volatile commodity markets?
  - Exceeding operation and maintenance costs?
  - Unpredictable events like labor strikes, sever weather events?
  - Are conventional indexes not accurate enough?

Social Media and Cost Estimating

- Social media has become a significant topic of interest in academic literature
  - Titanic quantities of data, produced every minute, make “big data” seem miniscule
  - More data than any human could dissect
- If not humans, than computers. Enter **Machine Learning** and **Automated Data Processing and Mapping**
- Behavioral economics, finance, and other industries are researching how consumer sentiment on social media platforms predicts market demand for goods, services, as well as for securities
- Why not **Cost Estimating?**

Researchers have successfully forecasted indexes, such as the Dow Jones Industrial Average, with an 88% accuracy by scanning emotional sentiment on Twitter

If price indexes used in program estimates were derived from Twitter, would they be more accurate?

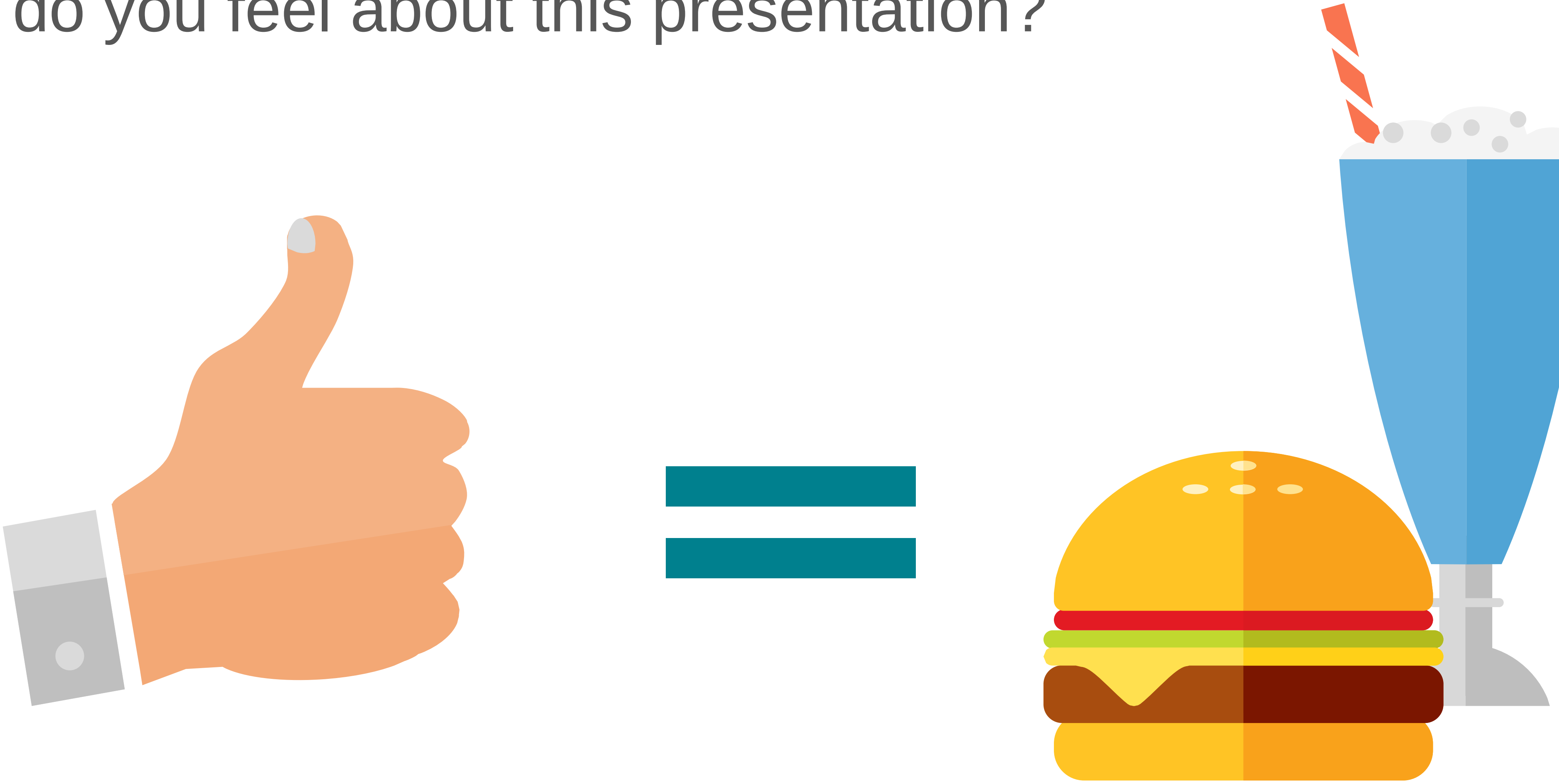
From Twitter to… Escalation?

- Research has found that the type of information shared on social media has **significant impact** on the topic of interest in the markets.
  - Stocks with increased social media "buzz" have significantly higher idiosyncratic volatility and higher trading volume over the next month.
  - Stocks with no, or little, social media buzz experience less of both.
- The most talked about stocks experience an increase in mean volatility of ~50%, and an increase in trading volume of ~25%.
- Text-scraping and natural language processing can aid in tapping twitter for indicators of these feelings.

Twitter is my Price Index

- Twitter is proving to be a powerful source for data.
- Research finds that specific emotions have statistically significant predictive power over commodity markets.
    - Found correlations between crude oil, gold, and Independent variables for “Optimistic” sentiments are statistically significant at the 1% level.
- There is evidence that **financial risk managers should use social media-based indices** to help account for market fluctuations.
- Utilize social media sentiment to estimate short-term price movements
- Instead of attempting to estimate years in advance, can attempt to capture short term changes in the marketplace utilizing data that is readily available
- More relevant to oversight committees or program office estimators?

How do you feel about this presentation?

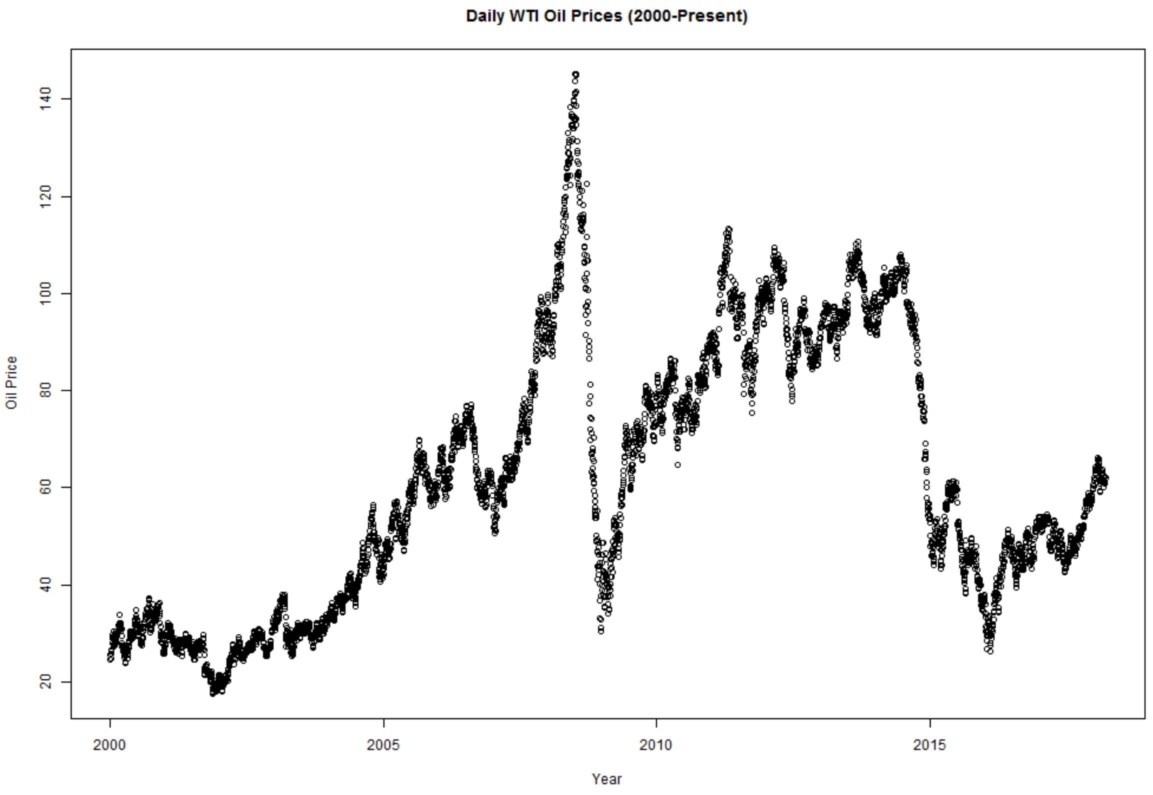


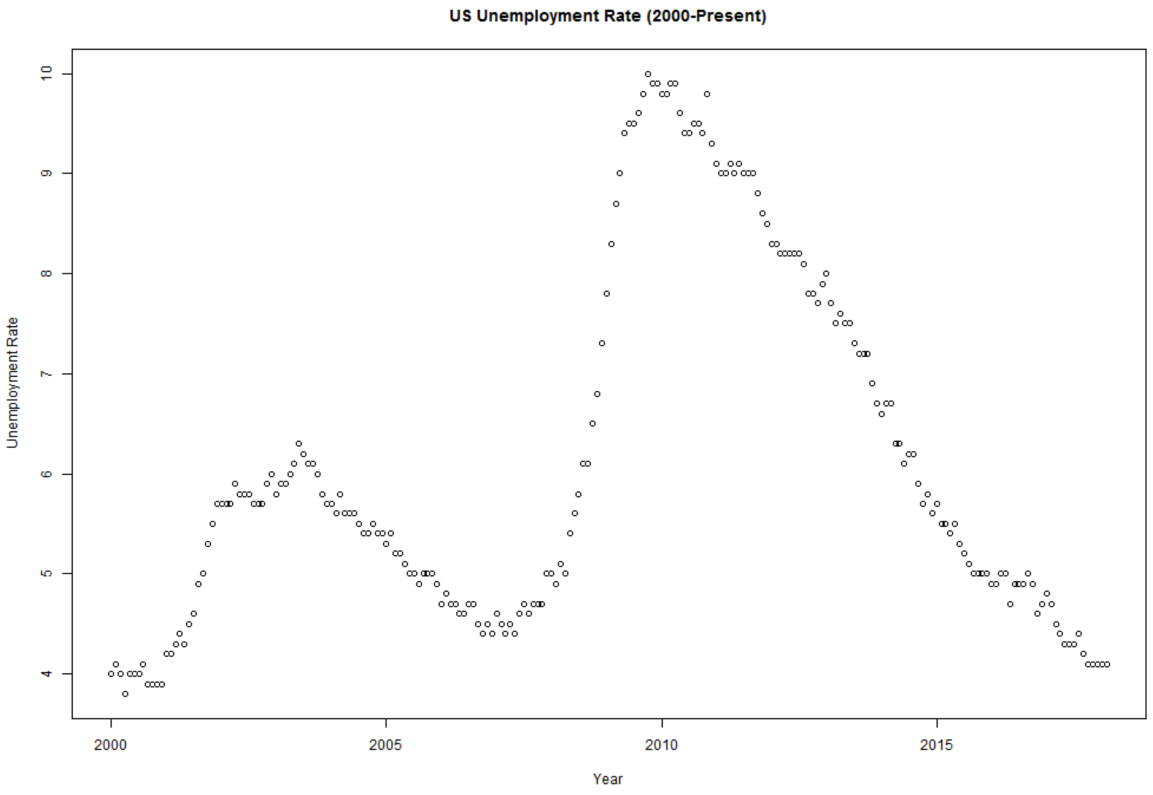
“I love this presentation” = Good feeling make you hungry? = You go buy food

Social Media in Industry Indices

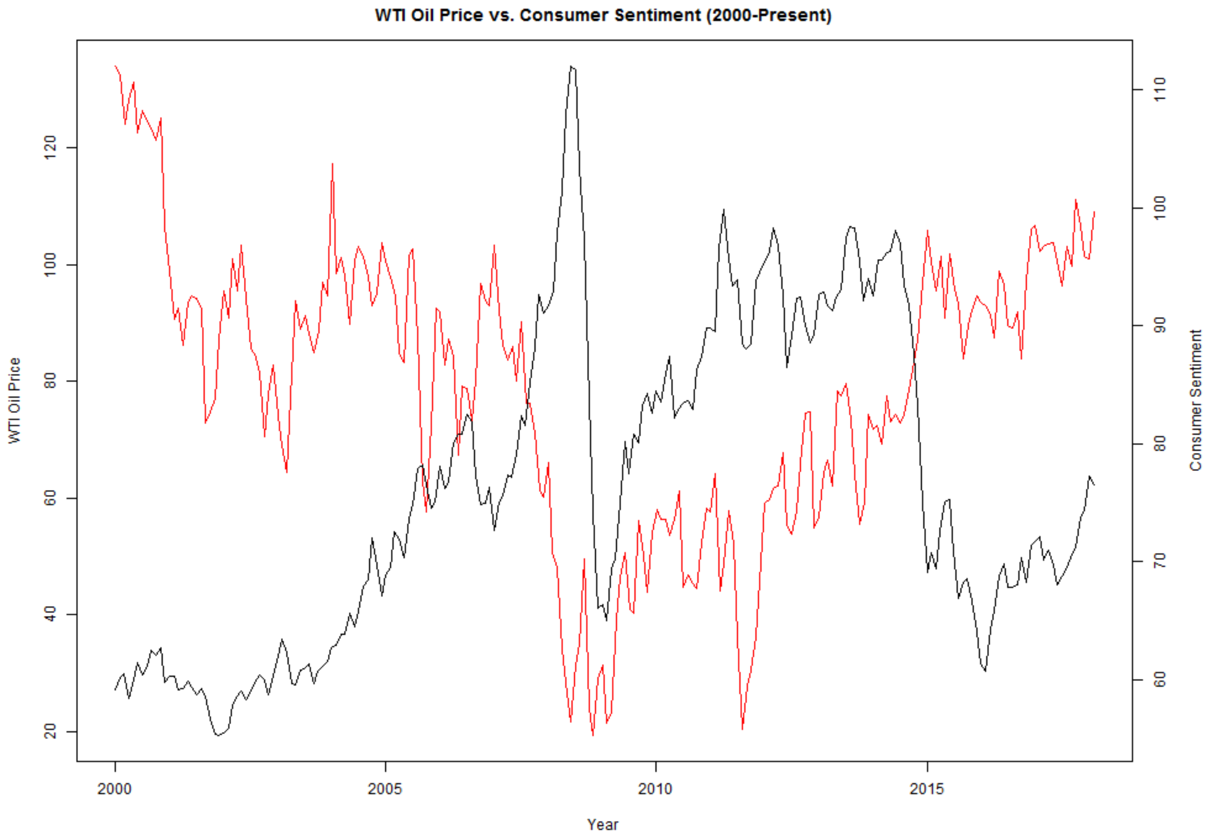
- Utilize social media sentiment to estimate short term price movements.
- Instead of attempting to estimate years in advance, one can attempt to capture short-term changes in the marketplace leveraging existing data.
  - Likely more relevant to oversight committees than program office estimators.
- Thomas Reuters MarketPsych Indices
  - Analyzes news and social media in real time across 2,000 top global news sites and 800 global financial media sites
  - Aimed at driving investor and risk management practitioners decision making
  - Utilizes natural language processing algorithms developed by MarketPscyh Data, LLC
  - Researchers have previously combined data from the Thomas Reuters Core Commodity CRB Index with the Thomas Reuters MarketPsych Index (TRMI) to measure sentiment in social media posts
- Michigan Consumer Sentiment Index

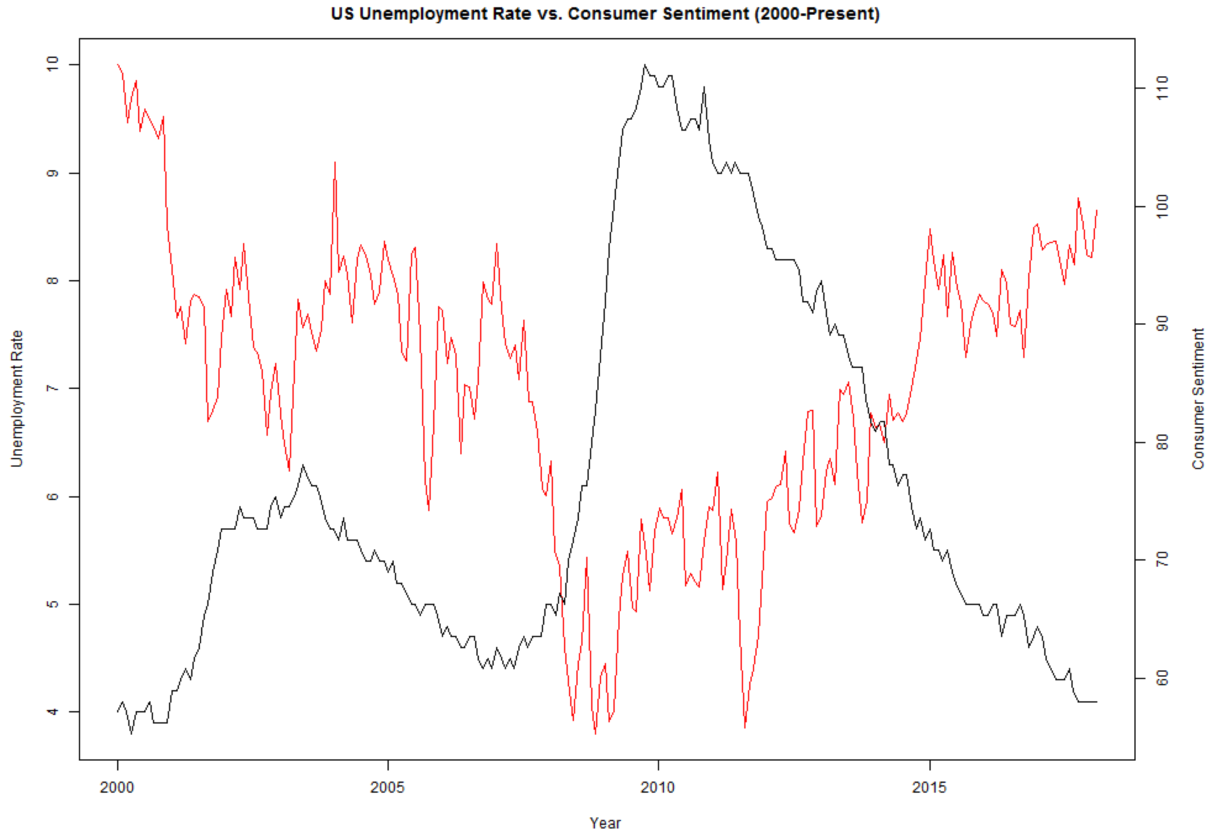
What might be the cause of such cost growth?



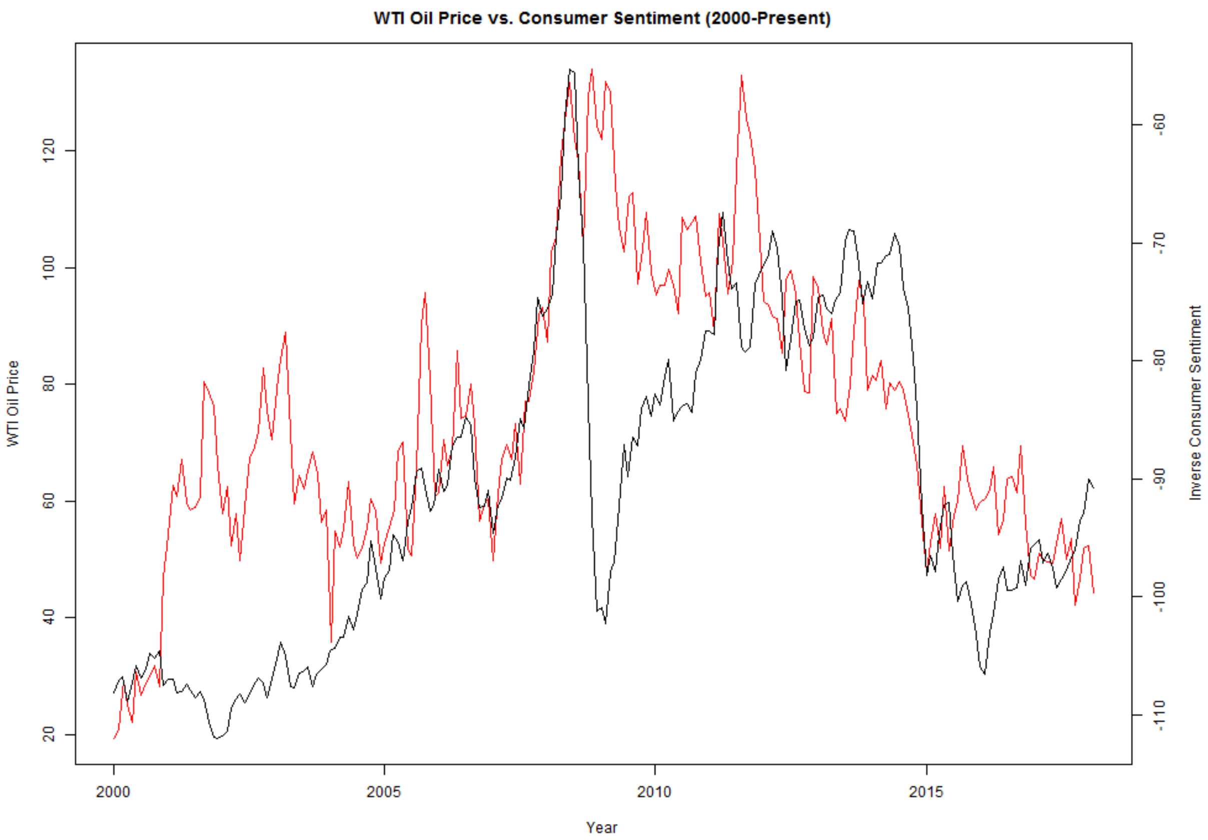


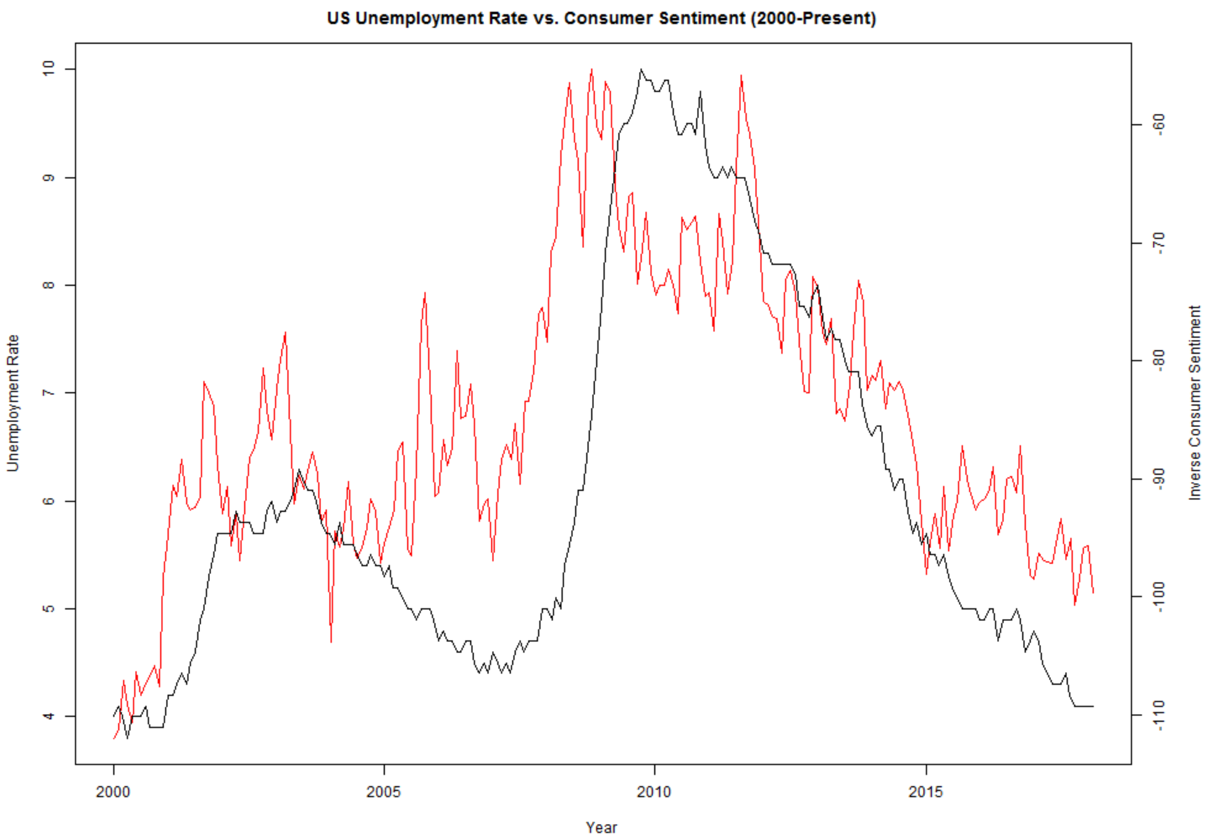
Can Consumer Sentiment Help?



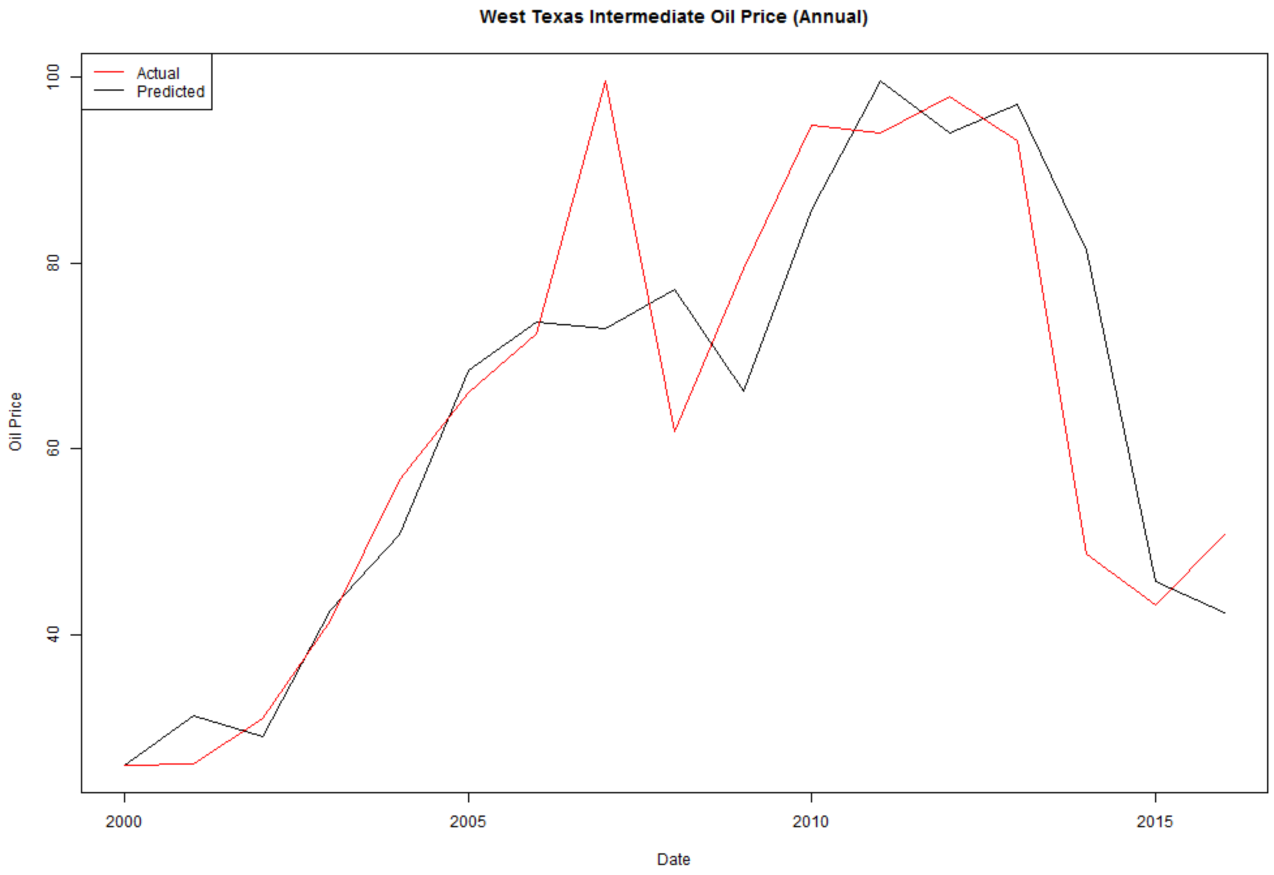


Can Consumer Sentiment Help?





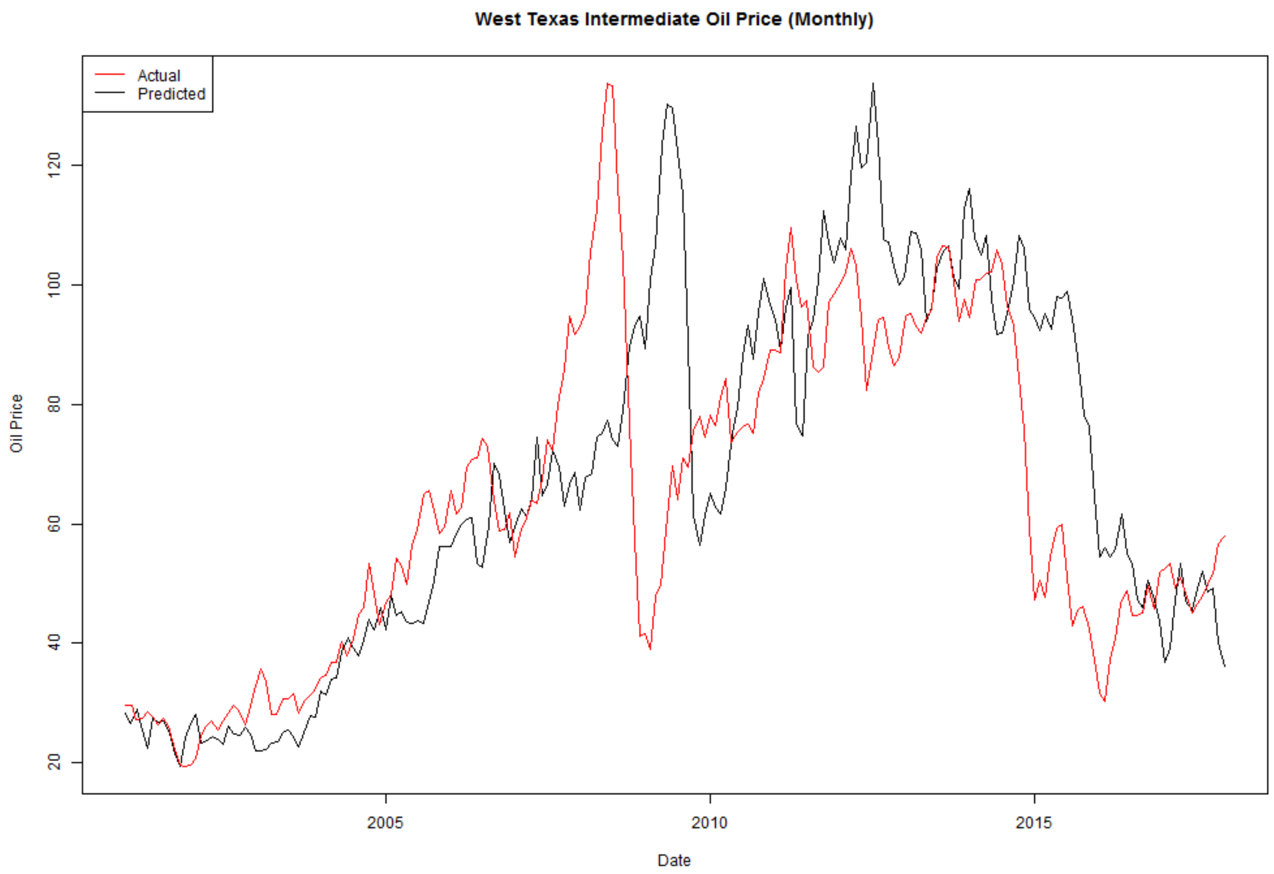
Predicting Oil Prices (A Year in Advance)



Year	Actual Price	Predicted Price	Delta
2001	25.98312	27.71702888	-1.733908877
2002	26.18496	32.97009182	-6.785131824
2003	31.07524	30.91469183	0.160548166
2004	41.5060241	44.69637466	-3.190350564
2005	56.637251	52.8799092	3.7573418
2006	66.05465863	70.86139609	-4.806737456
2007	72.34059524	76.17944893	-3.838853689
2008	99.67150198	75.76799176	23.90351021
2009	61.95043651	80.09937194	-18.14893543
2010	79.47571429	69.32876715	10.14694713
2011	94.88087302	89.47935644	5.401516576
2012	94.05333333	103.8721629	-9.818829526
2013	97.98253968	98.56955795	-0.58701827
2014	93.17222222	102.0435928	-8.871370564
2015	48.65670635	86.11179935	-37.455093
2016	43.29365079	49.73255403	-6.438903238
2017	50.80032	46.53016208	4.270157923

Why were we off in the highlighted years?

Predicting Oil Prices (A Year in Advance)



Consumer Sentiment

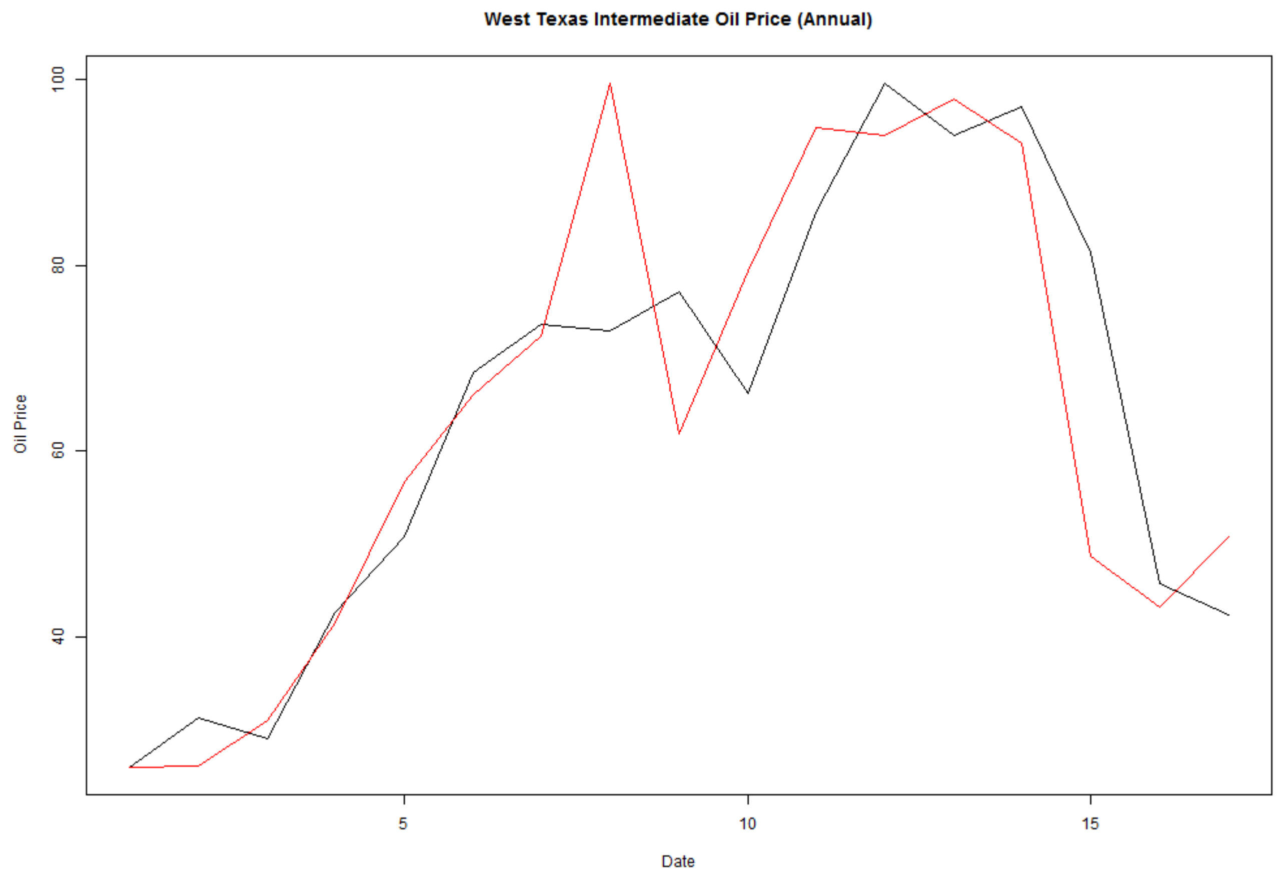
- If there is a long term negative trend of consumer sentiment on social media, estimators can assume a cheaper cost of labor, and vise versa

Using unconventional data sources and machine learning, can we produce more accurate price indices than used in current estimates?

Applications for Cost Estimating

- Commodity-related indicators, such as raw material and fuel, are developed by forecasting historical trends. Some include market speculation:
  - Research indicates that investors and market speculators overestimate their ability to read and interpret market movements.
- Machine learning algorithms tied to unconventional data sources could produce more accurate commodity forecasts.
  - Automated algorithms have the ability to capture “buzz” or trending stories circulating around the web and social media sites.
  - Combining market sentiment analysis with pre-defined relationships developed through economic research can improve accuracy of price indices.
- A better understanding of the price of commodities and the state of the economy can vastly improve budget planning and allocation.
  - There are applications outside of program office estimates, to include improving the accuracy of congressional budget requests

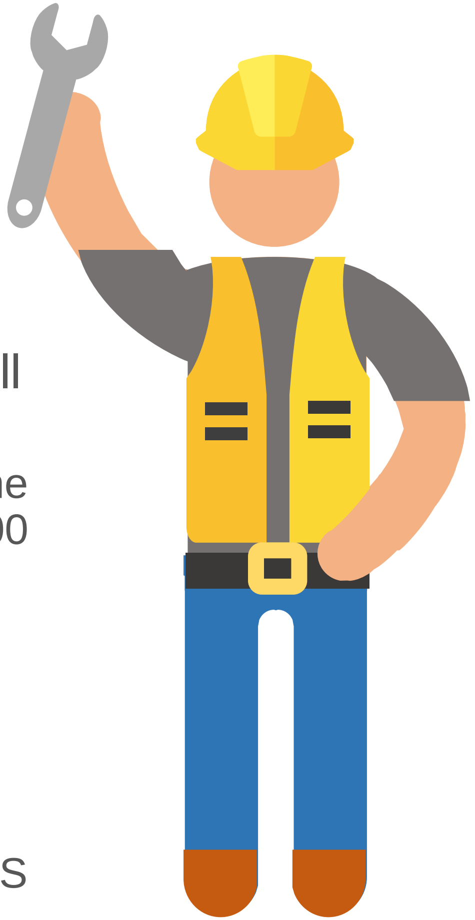
Remember this?



Year	Actual Price	Predicted Price	Delta
2001	25.98312	27.71702888	-1.733908877
2002	26.18496	32.97009182	-6.785131824
2003	31.07524	30.91469183	0.160548166
2004	41.5060241	44.69637466	-3.190350564
2005	56.637251	52.8799092	3.7573418
2006	66.05465863	70.86139609	-4.806737456
2007	72.34059524	76.17944893	-3.838853689
2008	99.67150198	75.76799176	23.90351021
2009	61.95043651	80.09937194	-18.14893543
2010	79.47571429	69.32876715	10.14694713
2011	94.88087302	89.47935644	5.401516576
2012	94.05333333	103.8721629	-9.818829526
2013	97.98253968	98.56955795	-0.58701827
2014	93.17222222	102.0435928	-8.871370564
2015	48.65670635	86.11179935	-37.455093
2016	43.29365079	49.73255403	-6.438903238
2017	50.80032	46.53016208	4.270157923

Examples

- Could “unpredictable” factors such as “union agreements” be predicted via social media indicators?
  - Bootstrapping actuals comparative to long-range weather forecasting/Farmers Almanac to discount for weather impacts?
  - Emotional sentiment to predict if/when labor agreements may occur/when wages will rise.
    - 2008 Boeing Labor Strike – 27,000 machinists went on strike for two months, strike cost the company $100 million per day in revenue and delivery-delay penalties because of the 3,700 jets the company had on backorder.
    - United Space Alliance Strike in 2007. 600 (16% of USA employees) USA claimed that the strike would not effect Space Shuttle launch, however, diverting labor to cover for striking employees represents an opportunity costs and therefore impacts the costs of other programs.
    - 2015 United Steel Workers Oil Refinery Strike: 5,200 workers at 11 refiners across the US went on strike against labor practices. First oil strike since 1982. Strike effects 1/5 of the US refining capacity.

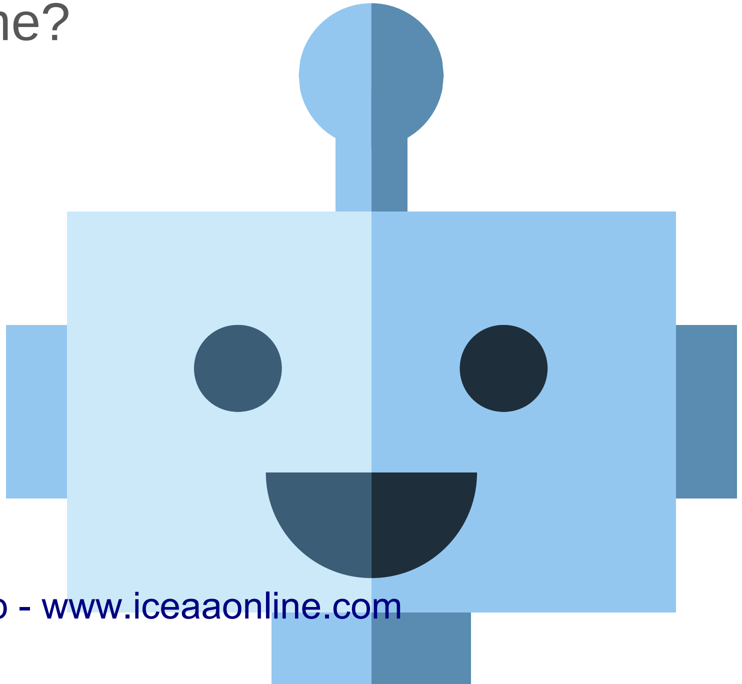


Using **Machine Learning**, deriving price indexes from unconventional data sources, such as social media, may be more accurate than traditional forecasting.

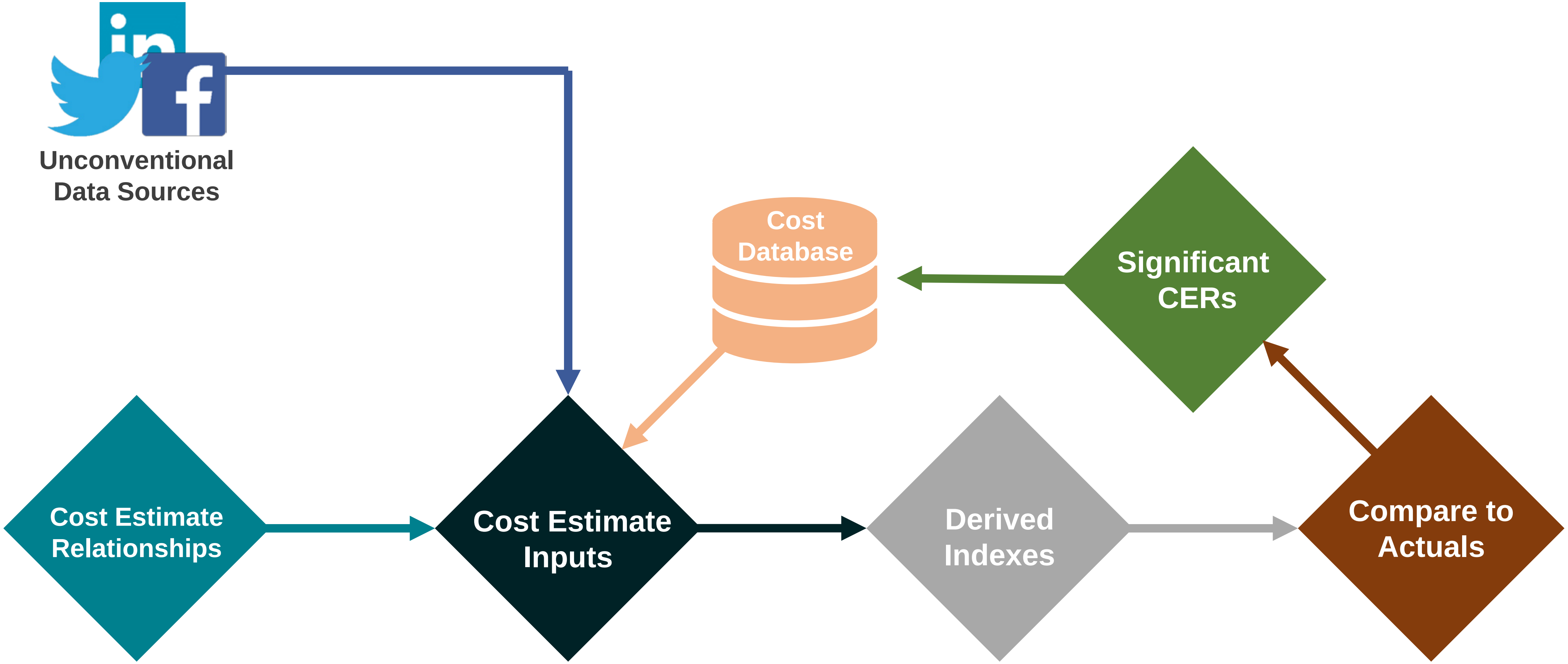
Let’s chat about **Big Data** and **Machine Learning…**

Beyond Social Media

- How can an **Machine Learning** revolution impact Cost Estimating?
- If a Machine Learning is designed in a way to check against hundreds (or more!) of conventional *and* unconventional data sources, what trends will it find and what CERs could it build?
- If an algorithm knows cost estimates and cost actuals, can it independently discover data sources that correctly predict the actual outcome?
- Can an algorithm build a better estimate?



Bootstrap an algorithm with actuals



We know you have questions