THE VALIDITY OF CONSUMER SENTIMENT IN SMALL-AREA ECONOMIC FORECASTING: A NAÏVE BAYES ANALYSIS

A.E. Rodriguez & Carolyne Cebrian*

Abstract

Obtaining an accurate picture of the current state and direction of the regional economy is particularly important to local decision-makers, including shopkeepers, academic institutions, and state and local government agencies.

Traditional, survey-based sentiment indices have long-existed and are used for this purpose. But current abilities to source online data to map consumer sentiment has kindled interest in their usefulness in regional economic forecasting. The appeal of tailored sentiment indices and other similar online-sourced measures are their seeming immediacy and their ability to capture information in more relevant geographic and product domains.

If decision-makers are to profitably rely with reasonable confidence from the increased availability of sentiment indices they will have to learn to effectively integrate domain knowledge, conventional or tailored online sentiment indices and traditional data. Perhaps more importantly, users will have to be assured of sentiment index validity in enhancing regional economic forecasts. We test sentiment index relevance in this paper reproducing results of a popular local forecast.

Specifically, we appraise whether there are measurable improvements from the presence of a sentiment index to the New Haven Register's Economic Scorecard, a popular regional forecast model. The model is a binary directional prediction model. Succinctly, we find measurable improvements in the model's predictive accuracy of the Economic Scorecard. We speculate as to the generalizability of our results, especially regarding the use of other online-sourced nowcasting metrics.

DRAFT: January 8, 2017

JEL: C43, M21, H70, R15

* Rodriguez, University of New Haven. Email: arodriguez@newhaven.edu. Cebrian, Sacred Heart University, Email: ccebrian@lincolncollegene.edu. Rodriguez is corresponding author.

As do many individuals, business operators and establishments often forge their economic understanding of local business conditions from a combination of sources: local media business analysts, newscasts, specialized web-sites, published government statistics, inter alia.

Obtaining an accurate picture of the current state of the economy is critical to all of us, but it is especially important to regional economic decision-makers. There are countless localized economic forecasting models in service of this effort throughout the country. Their coverage scope varies: they range from metropolitan area ones, statewide ones, to multi-state or regional forecasts.

For the most part, these models rely on conventional statistical models and conventional economic time series as reported by state and federal government agencies. Both conventions are beset with considerable problems.

There are any number of limitations associated with reliance on statistics gathered and provided by the federal government. For instance, much of the data assembled reflect historical, pre-establish zones and areas, jurisdictions which may not necessarily reflect current circumstances, nor the circumstances of all. Another limitation, an especially critical one is the reporting lag – federal supplied data are provided often several months in arrears.

Thus, to improve forecast accuracy, forecast specialists and analysts often incorporate regional or contemporaneous information into their models: consumer sentiment indexes are a popular choice. And historically, the sentiment indexes of choice were ones assembled via traditional survey methods.

But do sentiment indexes contribute to forecast accuracy or improving forecasts generally? This is an especially important question nowadays, because of the current ability for forecasters to draw online-data sourced indices.

It is possible to construct and use tailored sentiment indices given the increased availability of social and online data, what is known generally as "nowcasting" (Choi & Varian, 2009) (Das & Chen, 2007) (Scott & Varian, 2013) (Mago, 2016).[1] These nowcasting indices can be crafted to cover a particular zone or region in a manner that conventionally sourced statistics cannot.

If both professional forecasters and practical decision-makers who are capable of constructing their own forecasts are to profitably rely with reasonable confidence from the increased availability of sentiment indices, they will have to learn to effectively integrate domain knowledge, conventional or tailored online sentiment indices and traditional data. Perhaps more importantly, users will have to be assured of sentiment index validity in enhancing regional, area economic forecasts.

Studies that have previously examined the question of sentiment index-enhanced economic forecasts have found a connection between sentiment and relevant

[1] There has been an increased interest in the relevance of consumer sentiment and consumer confidence as a result of a renewed interest into Keynesian "animal spirits" explanations for the great recession. See, e.g. (de Bondt & Schiatti, 2015).

economic variables, such as consumption (Doms & Morin, 2004) – albeit modest ones (Garrett, Hernandez-Murillo, & Owyang, 2005).

However, most of these previous studies were conventional statistical econometric models. And most appraisals of sentiment validity relied on traditional consumer sentiment indicators such those constructed by the University of Michigan and the Conference Board's Survey Indices.

We test sentiment index relevance in this paper reproducing results of a popular New Haven, Connecticut regional model. Specifically, we appraise whether there are measurable improvements from the presence of a sentiment index in the input-variable set of the New Haven Register's Economic Scorecard, a popular regional forecast model. The model is a binary directional prediction model and we use a Bayesian net classifier algorithm.

Thus our inquiry as to the validity of the sentiment index combines several literatures. Succinctly, we rely on machine learning algorithms to estimate bidirectional predictive models for purposes of determining the predictive relevance of sentiment indexes.

All aspects of our work is set forth in this paper. It is structured as follows. The next section provides a description of the forecast-making process. It highlights the embedded sources of possible forecast error. The third section discusses the relevance of the chosen model – noting that it is especially well suited to handle the observed data and data-construction limitations. Bayesian models are known to closely resembling decision-making in financial and economic markets. Section four discusses the model, the data and the results. The last section concludes with a discussion of the generalizability of the results.

FORECASTING

Forecasts derived from economic models are often simply reported. At other times, the reporting conduit is via a commentator or analysis, although an explicit or implicit economic model underscores the analysis and commentary provided. A typical sequence of events leading up to a professional analyst forecast goes something like this: the analyst/forecaster waits for applicable data to be served up by the Bureau of Labor Statistics or relevant government agency at pre-announced release dates; the analyst then acknowledges the information provided in the release with some pithy comment; she then combines the insights from the data judiciously with professional experience; and then projects accordingly. Of course, the analyst's seeming success corresponds almost precisely with her ability to avoid both categorical statements and declarative sentences and on her skill in couching the prediction in the finest two-handed language possible.

Here is an example of this routine. It is drawn from a relatively recent report in our local, regional newspaper but it is not unlike those likely to be found in any, and practically every, regional newspaper, radio, TV station, and online media outlet across the country. It is a ritual keenly tuned to the release dates of official economic data.

> *The area economy will have recovered all of the jobs it lost during the last recession. Through November, New Haven and surrounding communities had recovered 94.3 percent of the jobs that were lost…*
>
> *For the second month in a row, the scorecard for November showed six of the eight indicators headed in a positive direction.*
>
> *"New Haven is clearly out-performing the state as a whole," … "This is a continuation of what started happening earlier this fall and it should continue well into 2015."*[2]

And notwithstanding the carefully crafted ambiguity and deliberate imprecision the forecast provided by our local media expert is, in turn, folded into our individual analysis by the rest of us, individuals, shop-keepers, managers, government employees, academic administrators, to organize and plan our next steps accordingly.

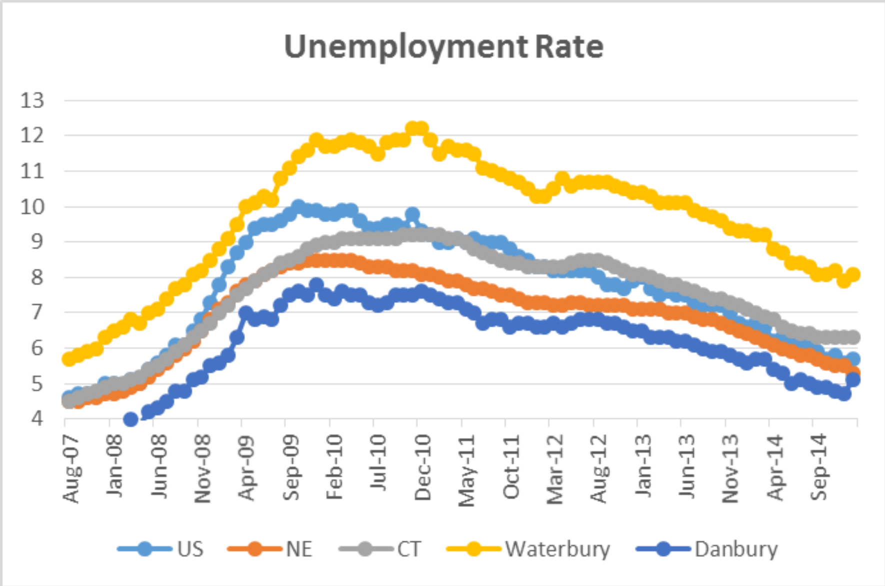
The problems with this exercise are numerous but invariably include the following two. The first is that the analyst cited in the article is relying on dated data. November figures are used to inform a late January forecast! This is a well-known shortcoming of our official statistical reporting system. And it is a handicap especially unhelpful for businesses organizations burdened with long-lead times in production and distribution such as the fashion (retail) industry and the toy industry.

A second problem follows from the fact that the data reflects a pre-selected frame or category of analysis that may resemble a user's situation only partially or indirectly. For example, Figure 1 provides a visual representation of the reported unemployment rates that contain applicable information for our local economy. Waterbury and Danbury are small cities approximately 34 miles apart. As can be seen, the variation among the reported rates is enormous. Within Connecticut alone,

[2] (Turmelle, 2015) The underlining of the last phrase is our doing.

the difference in the unemployment rates between Danbury and Waterbury averaged 2.7 standard deviations and reached a maximum difference of 3.5 standard deviations in November of 2010.[3]

Figure 1



Clearly, the unemployment figure for the region or sub-region used by the analysts, necessarily an aggregate, may disguise important variation in the underlying data. Thus, the particular rate relied upon by the original analysts to form his or her opinion is critically import. In fact, as can be seen in the example cited above, the particular instance of the unemployment rate (or applicable statistic) is often left unreported.

By appealing to an unemployment rate calculated over a particular town or area solves what is known generally as the "reference-class" problem or as the "partition" problem (Colyvan & Regan, 2007). The problem arises because there is no a priori way to privilege one classification over another and therefore any proposed outcome is infinitely malleable. For economic series in general this might appear to be nitpicking because many current series reflect reasonable commonly established and accepted boundaries that define the partition set at hand. In fact, to a large extent these boundaries have been defined by common practice. The problem arises when the users fall outside the original intended audience. Thus, returning to the graph above, if you are a resident of Southbury, a town halfway between Danbury and Waterbury – which unemployment rate applies to you?

[3] These statistics are calculated using Bureau of Labor Statistic for Seasonally Adjusted Unemployment Rates for Connecticut, Waterbury and Danbury for the following period August 2007-February 2015; the Waterbury and Danbury series are New England Cities and Town Areas (NECTA). The US and New England series are also BLS monthly, seasonally adjusted series. The standard deviation is the difference between the reported rates for Waterbury and Danbury divided by the standard deviation of the reported SA-Unemployment series for Connecticut.

There is a derivative problem. There are simply too much data. Indeed, FRED, the Federal Reserve Economic Database alone boasts of 247,000 time series from 77 sources.[4] In perhaps a variant of Fredkin's paradox we are paralyzed, incapable of deciding when confronted with choices among and between similar series. The massive amounts of data are effectively useless when combined with our inability to successfully separate the signal from the overwhelming noise. And perhaps this is what drives us to rely on more idiosyncratic, contemporaneous data. We turn to examine whether this heuristic is a valid one.

RELEVANT LITERATURE

Once a regional forecast is advanced there is evidence to suggest that it is routinely interpreted, by laymen and individuals, in a simple binary manner: whether "things" will get better, or not. This, simplification suggests a wariness of the false precision imparted in level forecasts. In fact, as a general principle it is probably wise to view most economic and financial forecasts with some modicum of caution, especially given economic and financial forecasting's poor track record recently.

The New Haven Register, the influential, regional newspaper in the broader New Haven, Connecticut area, publishes a popular forecast which they call the "Economic Scorecard." It is an instance of a binary directional prediction model which relies on a conventional set of Economic time series. The model returns an outcome that is described as either "thumbs-up," or "thumbs-down."

There are considerable parallels between the data generation model underscoring local area and regional models and those underscoring price movements. Thus, we can apply the binary price directional prediction models used for forecasting movements in market indices, individual prices, inter alia, to the Economic Scorecard model and determine whether the presence of a sentiment index improves its forecast accuracy.

Machine Learning Applications of Binary Directional Prediction Models

The literature from which to draw from is considerable. Over the last decade or so, market directionality appraisals featuring machine learning algorithms have gained popularity in both academic and professional research and practice. The binary time

---

[4] This is how a recent a recent announcement issued by the National Science Foundation characterizes the explosion of data, "Data may originate from many disparate sources, including scientific instruments, medical devices, telescopes, microscopes, satellites; digitally-authored media, including text, images, audio, and emails; streaming data from weblogs, videos, financial/commercial transactions; from ubiquitous sensing and control applications in engineered and natural systems, through multitudes of heterogeneous sensors and controllers instrumenting these systems; social interactional data from social networking sites, twitter feeds and click streams; administrative data; or scientific data from large-scale surveys, brain research, large-scale simulations, continuous simulation models, and computational analyses of observational data. The data can be temporal, spatial, or dynamic; structured or unstructured; and the information and knowledge derived from data can differ in representation, complexity, granularity, context, quality, provenance, reliability, trustworthiness, and scope." http://www.nsf.gov/pubs/2015/nsf15544/nsf15544.txt

series problem in these studies is typically modeled as a two-class supervised learning classification problem where the analyst is interested in the direction of stock market indices, individual shares, exchange rates, inter alia. The algorithmic task is to predict classes by examining historical instances of classification given a set of attributes.

In studies examining direction of change, changes are classified as 0 or 1. Changes in the level prices are typically, although not exclusively, examined on a day-to-day basis. Accordingly, a class value of 0 means that the present day's price is less than the previous day; a fall in the price of the stock. Similarly, a class value of 1 means that the present day's price is greater than the previous day; a rise in the stock price.

The specific literature on machine learning-based bidirectional prediction *alone* is extensive.

Kumar and Thenmozhi examine the predictability of the direction of stock index movement by means of machine learning methods (Kumar & Thenmozhi, 2006). They deploy classification models, such as Linear Discriminant Analysis, logit, artificial neural network, Random Forests, and Support Vector Machines.

Choudhry and Garg deploy a support vector machine; they select the set of attributes with a genetic algorithm (Choudhry & Garg, 2008). Juan et all, also resort to support vector machine algorithms and other machine learning methods for predicting stock market direction. Researchers and commentators have relied on other modeling algorithms to examine binary direction models; there include autoregressive models, the Generalized Linear Model or a Hidden Markov Model. Startz, for instance, applies binary autoregressive models and markov processes to US recession data (Startz, 2012). Bicego and co-authors, resort to Hidden Markov Modeling to identify and predict the sign in short financial trends (Bicego, Grosso, & Otranto, 2008).

As a general point, classification models such as linear discriminant analysis and logit used for the direction of stock index movement outperform the level estimation methods such as exponential smoothing or multilayered feed-forward neural networks (Leung, Daouk, & Chen, 2000).

The New Haven Economic Scorecard

The Economic Scorecard is an implicit multi-attribute scoring model. It is a linear unweighted sum of seven conventional economic time series plus the Michigan consumer sentiment index.

The Register's Economic Scorecard popularity is derived from its simplicity. The model relies on a simple appraisal of the change in the various economic series. A series, total employment for example, is a "thumbs-up" if it has increased from the same month in the previous year; obviously, a "thumbs-down" if it has experienced a decrease over the period.[5] The model's eight constituent series are transformed in

[5] The sole exception if for the unemployment series. It is ascribed a "thumbs-up" if it *declines* for the month relative to the same month the previous year, and viceversa.

this manner into a binary score. The individual series' binary score is then summed; thus, the score is at its maximum of 8 if all the attributes result in a "1."

$$Index\ Score = IS = \sum v_i$$

$$where,\ i = 1,8,\ and$$

and, for all *i:*

$$v_i = \begin{cases} 0 \text{ iff } \Delta v_i < 0 \\ 1 \text{ iff } \Delta v_i > 0 \end{cases}$$

The last step is taken by the analyst and it entails an ad hoc determination of the cutoff threshold. Based on the associated analysis by the Scorecard authors of the March Economic Scorecard we establish that any score greater than or equal to a score of five is considered an aggregate thumbs-up.

$$IS = \begin{cases} 1 & \text{if } \sum v_i >= 5 \\ 0 & \text{if } \sum v_i < 5 \end{cases}$$

This result is then announced as the predicted monthly outlook for the region.

Application of a Naïve Bayes Model to Estimate Directionality

To model the binary directional predictive process and appraise the relevance of sentiment indices in a realistic setting we use the simplest form of a Bayes network, the Naive Bayes classifier. Naïve Bayes has been used across a wide variety of fields including RNA sequencing, disease diagnosis, image classification and spam filtering (Raschka, 2014).

The Naive Bayes algorithm is a simple classifier that calculates a set of probabilities. It counts the frequency and combinations of values in a given set of attributes. The algorithm uses Bayes theorem and assumes all attributes to be independent given the value of the class variable.

The most likely class given an attribute, v, and a class, C, is:

$$Class = argmax\ P(C/v)$$

and

$$Class = argmax\ \frac{P(v/C)P(C)}{P(v)}$$

The conditional independence assumption rarely holds true in real world applications and leads to the appellate "naïve" assigned to the algorithm. Yet Naive Bayes tends to perform well and learn rapidly in various supervised classification problems (Domingos & Pezzani, 1997) (Zhang, 2004). The performance of the algorithm is determined by the accuracy of the classification. Classification accuracy is calculated by determining the percentage of tuples placed in a correct class.

To appraise the relevance of the sentiment index, the model is first estimated with the actual sentiment index. This outcome is then compared to the outcome of the same model with the sentiment index replace by a randomly generated one.

Results

The confusion matrix is generated for class Outlook having two possible values i.e. positive (1) and negative (0). The result for the model containing the published sentiment index:

	actual	
predicted	positive	negative
positive	43	4
negative	5	47

Reported accuracy: 0.909

With the randomized sentiment Index

	actual	
predicted	positive	negative
positive	40	8
negative	8	43

Reported Accuracy = 0.838

The reported accuracy is substantive. The results suggest that the sentiment index conveys useful information for purposes of enhancing predictive accuracy.

Concluding Comments

Our results suggest that the presence of the sentiment index in the Economic Scorecard model measurably improves forecast accuracy. Importantly, our result does convey some weight to decision-makers' reliance on idiosyncratic indicators and on constructed sentiment indexes – such as those assembled from online-sources.

The use of sentiment and opinion information is already incorporated into local-area decision-making. Analysts routinely drawing insights into a region's outlook by identifying and tracking non-traditional indicators and data sources capable of providing a more idiosyncratic understanding of a region's performance. Regional and city planners are now using localized urban information such as subway passenger traffic, Broadway ticket sales, parking garage usage counts, and the average price of local apartments on Craig's list to assist in the discerning of the future of their localities.[6]

Organizations, businesses and individuals routinely use tracking services such as customized RSS feeds, IceRocket, and Google's Me on the Web among others to access "social-intelligence", online social-media consumer sentiment and opinion.[7] And with the ascendance and influence of Yelp, Facebook, Twitter, Rotten Tomatoes, and related online fora – where customers provide feedback and opinion, managers monitor their establishment's reputation and performance by tracking web commentary. The popularity and widespread use of sentiment data suggests that users feel intuitively that opinion and sentiment indices convey a sense of the direction of consumer expectations and associated spending (Ludvingson, 2004) (Garrett, Hernandez-Murillo, & Owyang, 2005) (Mago, 2016).

Heuristically, local data – replete with intuitive and familiar features - may prove to be more comforting and natural to use in gauging a future local event. Localized, relatable data may facilitate the formation of more relevant interpretive mental

---

[6] National Public Radio featured a report on the city of San Francisco's use of regional statistics for planning purposes. "A Fresh Approach to Measuring the Economy," April 11, 2010. http://www.npr.org/templates/story/story.php?storyId=125837367 {viewed on 3/14/15}. The New Haven region is one of 45 urban areas across the nation that boasts an online data initiative sponsored by the Urban Institute, "...to further the development and use of neighborhood information systems in local policymaking and community building." Data Haven claims as their mission, "...to improve quality of life by collecting, sharing, and interpreting public data for effective decision making." http://www.ctdatahaven.org/ {viewed on 3/1/4/15}

[7] IceRocket is a free service – among many available providing social media and online tracking capabilities. Meltwater, the enterprise which provides IceRocket services, for a fee, offers more tracking sophisticated capabilities. There are many other similar enterprises available. By referencing them in this paper, we are neither endorsing nor recommending Meltwater; we are merely alluding to them as a representative provider. http://www.meltwater.com/products/?utm_source=IceRocket&utm_medium=banner&utm_campaign=mBuzz_Social_Software For another example, see, also, e.g., socialmention; http://socialmention.com/

models, narratives or algorithms required to process the assembled data – primarily because we have considerable more touchpoints necessary to inform experience.

With the framework that we advance here the information advantage of any particular constructed index can be appraised. It would unlikely convey end-user confidence to be assured that their favored idiosyncratic index is valid.

References

Bicego, M., Grosso, E., & Otranto, E. (2008). Recognizing and forecasing the sign of financial local trends using hidden Markov models. *CRENo.S Working Paper 2008, 03*. Retrieved from http://eprints.uniss.it/22/

Boesler, M., & Perlsberg, S. (2013, October 11). The 41 Most Unusual Economic Indicators. *Business Insider*. Retrieved March 14, 2015, from http://www.businessinsider.com/unusual-economic-indicators-2013-10

Choi, H., & Varian, H. (2009). *Predicting the Present with Google Trends*. Google. Retrieved from http://cs.wellesley.edu/~cs315/315_PPTs/CS315-L19-CS315-L19-PredictingWithSocialMedia/Predicting-the-present-Choi-Varian.pdf

Choudhry, R., & Garg, K. (2008). A Hybrid Machine Learning System for Stock Market Forecasting. *World Academy of Science, Engineering and Technology, 15*, 315-318.

Colyvan, M., & Regan, H. M. (2007). Legal Decisions and the Reference Class Problem. *International Journal of Reference and Proof*, 274-285.

Das, S. R., & Chen, M. Y. (2007, September). Yahoo! for Amazon: Sentiment Extraction from Small Talk on the Web. *Management Science, 53*(9), 1375-1388.

de Bondt, G. J., & Schiatti, S. (2015, December). Confidence Matters for Current Economic Growth: Empirical Evidence for the Euro Area and the United States. *Social Science Quarterly, 96*(4), pp. 1027-1040.

Domingos, P., & Pezzani, M. (1997). On the Optimality of the Simple Bayesian Classifier Under Zero-One Loss. *Machine Learning, 29*, 103-130.

Doms, M., & Morin, N. (2004). *Consumer Sentiment, the Economy, and the News Media*. Board of Governors of the Federal Reserve System. Retrieved from file:///C:/Users/arodriguez/Desktop/Later/sentiment%20paper/Doms%20&%20Morin%20wp04-09bk.pdf

Garrett, T. A., Hernandez-Murillo, R., & Owyang, M. T. (2005, March/April). Does Consumer Sentiment Predict Regional Consumption? *Federal Reserve Bank of St. Louis Review*, pp. 123-135.

Gigerenzer, G., & Hoffrage, U. (1995). How to Improve Bayesian Reasoning Without Instruction: Frequency Formats. *Psychological Review, 102*(4), 684-704.

Goldstein, D. G., & Gigerenzer, G. (1996). Reasoning the Fast and Frugal Way: Models of Bounded Rationality. *Psychological Review, 103*, 650-66.

Griffiths, T. L., & Tenenbaum, J. B. (2011). Predicting the Future as Bayesian Inference: People Combine Prior Knowledge with Observations when Estimating Duration and Extent. *JOurnal of Experimental Psychology: General, 140*(4), 725-743.

Griffiths, T. L., & Tenenbaum, J. B. (Undated). *Teacakes, Trains, Taxicabs and Toxins: A Bayesian Account of Predicting the Future*. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.33.4073&rep=rep1&type=pdf

Healy, J. (2009, September 26). Men's Underwear as an Economic Indictator. *New York Times*. Retrieved March 14, 2015, from http://www.nytimes.com/2009/09/27/weekinreview/27healy.html?_r=0

Hirshleifer, D. (2007). Psychological Bias as a Driver of Financial Regulation. *MPRA Paper No. 5129*, 1-35.

Hubbard, D. W. (2014). *How to Measure Anything*. Hoboken, New Jersey: John Wiley & Sons, Inc. .

Jones, C. L. (2009, June 12). 10 Quirky Economic Indicators. *Kiplingers*. Retrieved March 14, 2015, from http://finance.yahoo.com/news/pf_article_107186.html

Kumar, M., & Thenmozhi, M. (2006, January 24). Retrieved from SSRN Papers: https://ssrn.com/abstract=876544

Leung, M. T., Daouk, H., & Chen, A. S. (2000). Forecasting Stock Indices: a Comparison of Classification and Level Estimation Models. *International Journal of Forecasting, 16*(2), 173-190.

Ludvingson, S. C. (2004, Spring). Consumer Confidence and Consumer Spending. *Journal of Economic Perspectives, 18*(2), pp. 29-50.

Mago, N. (2016, April). Opinion Mining: Applications, Techniques, Tools, Challenges and Future Trends of Sentiment Analysis. *International Journal of Computer Engineering and Applications, X*(IV), 73-82.

Mueller, S. T. (2009). A Bayesian Recognitional Decision Model. *Journal of Cognitive Engineering and Decision Making, 3*(2), 111-139.

Raschka, S. (2014, October). Naive Bayes and Text Classification I: Introduction and Theory. Retrieved from http://arxiv.org/abs/1410.5329

Scott, S. L., & Varian, H. R. (2013). Bayesian Variable Selection for Nowcasting Time Series. *NBER Working Papers 19567.*

Silver, N. (2012). *The Signal and the Noise.* New York: The Penguin Press.

Startz, R. (2012). Binomial Autoregressive Moving Average Models with an Application to U.S. Recessions. *Journal of Business and Economic Statistics, 26*(1), 1-8.

Tetlock, P. E. (2006). *Expert Political Judgment.* Princeton, NJ: Princeton University Press.

Turmelle, L. (2015, January 15). New Haven Area Outperforming the Rest of the State in Job Recovery. *New Haven Register.* Retrieved March 14, 2015, from http://www.nhregister.com/business/20150117/new-haven-area-outperforming-rest-of-the-state-in-job-recovery

von Helversen, B., & Rieskamp, J. (2008). The Mapping Model: a Cognitive Theory of Quantitative Estimation. *Journal of Experimental Psychology: General, 137*(1), 73-96.

Zhang, H. (2004). The Optimality of Naive Bayes. *Proceedings: 17th International FLAIRS Conference.* AAAI Press.

DATA APPENDIX

Data Treatment

To reproduce the Scorecard results we used the All Transactions House Price Index instead of the reported Warren Group home prices. We used monthly estimates of Weekly Earnings instead of Real Disposable Income. And we used the Michigan Sentiment Index instead of the Conference Board Consumer Confidence Index. All data was obtained from the Federal Reserve Economic Database via Quandl feeds. Housing Starts were for all of Connecticut instead of from the Connecticut Department of Economic & Community Development (DECD. According to the Scorecard authors they obtain their data from 12 area towns from CT DECD. The DECD data is very sparse and the aggregation of towns is probably meant to overcome a possible small numbers problem.

Summary Statistics

housing transactions index		*cpi*		*total labor force*	
Mean	142.0	Mean	224.4	Mean	321031.1
Median	136.5	Median	226.0	Median	322066.0
Standard Deviation	12.2	Standard Deviation	10.4	Standard Deviation	4916.5
Minimum	129.1	Minimum	202.4	Minimum	310364.0
Maximum	170.5	Maximum	238.7	Maximum	328167.0

unemployment		*total employees*		*housing starts*	
Mean	7.3	Mean	274.0	Mean	413.4
Median	7.5	Median	275.6	Median	368.0
Standard Deviation	1.6	Standard Deviation	5.6	Standard Deviation	175.4
Minimum	4.5	Minimum	263.9	Minimum	157.0
Maximum	9.6	Maximum	282.0	Maximum	889.0

sentiment		*weekly earnings*	
Mean	76.8	Mean	875.7
Median	76.0	Median	879.5

Standard Deviation	10.7	Standard Deviation	43.6
Minimum	55.3	Minimum	775.5
Maximum	98.1	Maximum	954.1

Model Economic Series Source and ID Codes

Variable	Source	Identifier
Housing	FMAC	HPI_NEWCT
cpi	FRED	CPIAUCNS
laborforce	FRED	NEWH709LF
unemployment	FRED	NEWH709UR
total_employees	FRED	NEWH709NA
housingstarts	FRED	CTBPPRIVSA
sentiment	UMICH	SOC1
earnings	FRED	SMU09757000500000011SA

A-priori probabilities

```
medvcat_f
 positive  negative
0.4848485 0.5151515
```

Conditional probabilities

```
          housing_f
medvcat_f          0         1
  positive 0.8541667 0.1458333
  negative 0.6470588 0.3529412

          laborforce_f
medvcat_f          0         1
  positive 0.5208333 0.4791667
  negative 0.2352941 0.7647059

          cpi_US_f
medvcat_f           0          1
  positive 0.22916667 0.77083333
  negative 0.05882353 0.94117647

          unemployment_f
medvcat_f          0         1
  positive 0.6041667 0.3958333
  negative 0.5686275 0.4313725

          total_employees_f
medvcat_f          0         1
  positive 0.4166667 0.5833333
  negative 0.1568627 0.8431373

          starts_f
medvcat_f          0         1
  positive 0.6666667 0.3333333
  negative 0.4901961 0.5098039

          sentiment_f
medvcat_f          0         1
  positive 0.5208333 0.4791667
  negative 0.2941176 0.7058824

          earnings_f
```

medvcat_f	0	1
positive	0.5000000	0.5000000
negative	0.1568627	0.8431373