Journal of Strategic Security

Volume 6
Number 5 *Volume 6, No. 3, Fall 2013 Supplement: Ninth Annual IAFIE Conference: Expanding the Frontiers of Intelligence Education*

Article 6

Forecasting the Failed States Index with an Automated Trader in a Combinatorial Market

Anamaria Berea
C4I Center of Excellence

Charles R. Twardy
George Mason University

Daniel T. Maxwell
KaDSci, LLC

pp. 38-51

Recommended Citation
Berea, Anamaria, Charles R. Twardy, and Daniel T. Maxwell. "Forecasting the Failed States Index with an Automated Trader in a Combinatorial Market." *Journal of Strategic Security* 6, no. 3 Suppl. (2013): 38-51.

Forecasting the Failed States Index with an Automated Trader in a Combinatorial Market

Forecasting the Failed States Index with an Automated Trader in a Combinatorial Market

Anamaria Berea, Charles R. Twardy and Daniel T. Maxwell
C4I Center of Excellence, George Mason University and KaDSci, LLC.

Introduction

Forecasting the risk of a failed state in the intermediate future is an important intelligence and social question. Even the ability to anticipate state failure – let alone avoid it – could save thousands of lives and hundreds of millions of dollars annually, just by prepositioning humanitarian relief and security forces. Although state failure is a highly complex event, recent high-profile projects like ICEWS revisit the goal of automated or semi-automated crisis warning.[1] Depending on who you talk to, this goal is either ludicrous or obvious.
It's ludicrous because the world is complex, and most attempts have failed. Indeed, the 2010 Journal of Peace Research article of the year showed that the two most trusted models of civil war were worthless:[2]

Large-n studies of conflict have produced a large number of statistically significant results but little accurate guidance in terms of anticipating the onset of conflict.

It's obvious because it has worked in other fields when models are actually developed for prediction. The most famous discussion began in 1954 with Paul Meehl's book, Clinical Versus Statistical Prediction.[3] In the ensuing seventy years, there have been hundreds of follow-on studies. Two notable meta-analyses found that "mechanical" prediction was more accurate than clinical prediction overall.[4] In fact, much of the evidence suggests that the more complex the situation, the stronger the advantage for models over experts.

In geopolitics, Tetlock's *Expert Political Judgment* notably established that as far as forecasting was concerned, there wasn't much expert judgment to be had, and humans were often beaten by embarrassingly simple statistical models ("no change"), and comprehensively beaten by sophisticated models.[5]

---

[1] Sean P. O'Brien, "Crisis Early Warning and Decision Support: Contemporary Approaches and Thoughts on Future Research," *International Studies Review* 12:1 (2010): 87–104.
[2] "JPR Article of the Year Award, 2010, Goes to Michael D Ward, Brian D Greenhill & Kristin M Bakke," *Journal of Peace Research* 48:2 (March 1, 2011): 143–143; Michael D. Ward, Brian D. Greenhill, and Kristin M. Bakke, "The Perils of Policy by P-value: Predicting Civil Conflicts," *Journal of Peace Research* 47:4 (July 1, 2010): 363–375.
[3] Paul Everett Meehl, Clinical Versus Statistical Prediction: A Theoretical Analysis and a Review of the Evidence (University of Minnesota Press, 1954).
[4] Robyn Dawes, David Faust, and Paul Meehl, "Clinical Versus Actuarial Judgment," *Science* 243:4899 (1989): 1668–74; M.C. Marchese, "Clinical Versus Actuarial Prediction: a Review of the Literature," *Perceptual and Motor Skills* 75:2 (October 1992): 583–94; William M. Grove et al., "Clinical Versus Mechanical Prediction: a Meta-Analysis," *Psychological Assessment* 12:1 (2000): 19–30.
[5] Philip Tetlock, Expert Political Judgment: How Good Is It? How Can We Know? (Princeton University Press, 2005).

Our approach hypothesizes that a hybrid approach that integrates the judgment of experts with "mechanical" prediction models can perform better than either individually. There is significant evidence in other domains that supports this belief. Heckerman's Pathfinder model[6], now twenty years old integrated expert judgment with a Bayesian Network for cancer diagnosis and significantly outperformed experts, especially in the difficult cases.

The Fund for Peace's Failed States Index is an example of an approach that integrates expert judgment with models to assess the relative stability of countries.[7] For our experiments we are using the Failed States index score as a proxy for state stability by attempting to forecast the score a country will receive when the annual results are released. To do this we create a Bayesian network template containing the Failed States index variables, instantiate specific models for several countries, and test it both alone and using estimates from a public prediction market.

This paper first discusses briefly the Failed States Model, and then describes a template Bayes Net that can be used as a foundation for detailed modeling of specific countries and questions. We then discuss a prototype model focused on Sudan for the year between June 2012 and June 2013 and describe an automated agent, called an autotrader that traded in the prediction market alongside human users. The paper closes with a few conclusions and recommendations for future research.

The Failed States Index – A Case Study in Structured Analysis

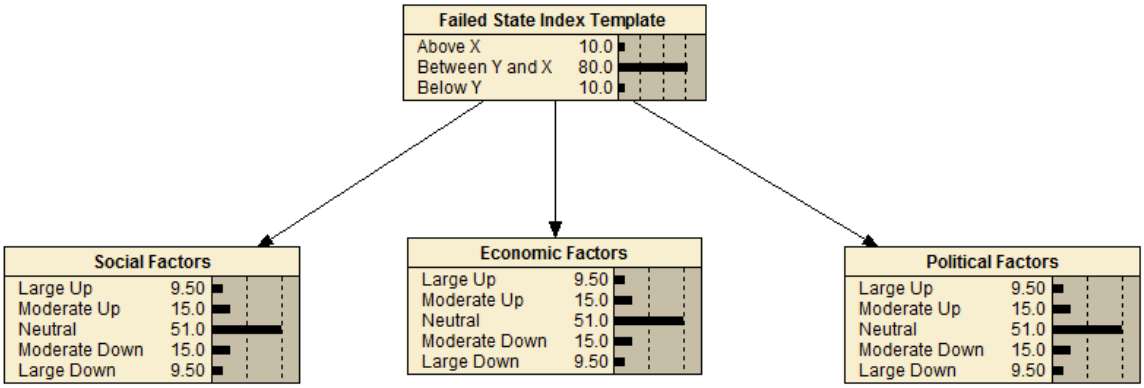
The Fund for Peace has developed a model it calls the "Conflict Assessment Tool" (CAST) that summarizes "twelve conflict risk indicators to measure the condition of a state. The indicators provide a snapshot in time than can be measured against other snapshots in a time series to determine whether conditions are improving or worsening."

The model applies content analysis techniques to thousands of open-source documents to derive a score from 1-10 on each of the indicators, where 1 indicates stable and 10 indicates unstable. The overall score for a country is the linear combination of the twelve scores. At the positive end of the scale are countries like Canada with a 2012 score of 26.9. At the negative end of the scale, are countries like the Congo with a score of 110 out of a possible 130. The score is published annually in June. The score indicates a country's stability relative to other countries, and by extension provides a proxy for its risk of state failure. This approach is the current state of the art.

[6] David Heckerman, *Probabilistic Similarity Networks.* (MIT Press: Cambridge, MA, 1991).
[7] Fund for Peace (2012) The Fund for Peace Country Analysis Indicators and Their Measures, Publication CR-10-97-CA (11-05C), available at: *www.fundforpeace.org*.

https://digitalcommons.usf.edu/jss/vol6/iss5/6
DOI: http://dx.doi.org/10.5038/1944-0472.6.3S.4

Figure 1: Top level model template.



While the CAST approach is one of the most widely accepted and applied indicators of country stability, it has a few key limitations that the use of other supplemental techniques could ameliorate. Specifically, the Failed States Index is published annually. It would be valuable to have earlier indications of instability. The linear combination of factors does not accommodate some of the unique considerations associated with differing state maturity, geography, and culture. Tailoring the component parts of CAST to each country should improve forecasting performance. Finally, CAST is based on large quantities of open source information of varying quality. It may be the case that there is classified, proprietary information, or expert judgment available that is of higher quality one may wish to give additional weight. Therefore, we wish to model explicitly the components of the CAST score.

We use Bayesian networks – Bayes nets or BNs for short. A Bayes net is a specialized probability model that allows for analysts to combine their subjective beliefs about the likelihood of events in the real world with evidence that is observed and collected over time. In addition to the ability to combine subjective judgment with evidence, Bayes nets have at least two other strengths. First, they support rather complicated models with very efficient computational algorithms. This allows analysts to represent explicitly many interacting factors, which is often necessary in complex situations. Second, a Bayes net does not require all of the evidence to be collected to start providing insights about changes in the likelihood of outcomes. As we shall see later, this opportunistic updating makes BNs ideal for application in crowd-sourced forecasting environments like prediction markets. There are several sumarizing publications in Bayes Nets that describe the mechanism and the applications of this method.[8] We patterned a Bayes net template from the Fund for Peace CAST model. **Error! Reference source not found.**Figure 1 shows the top level of the model.[9] It consists of a hypothesis node (about the FSI) associated with the overall index score and three nodes that aggregate indicators into the there major categories of interest; social, economic, and political factors. For our purposes, the hypothesis has three states: Improvement in state stability (Below Y), No significant Change (Between Y and X), and Deterioration in stability (Above X). The specific values for each of these states,

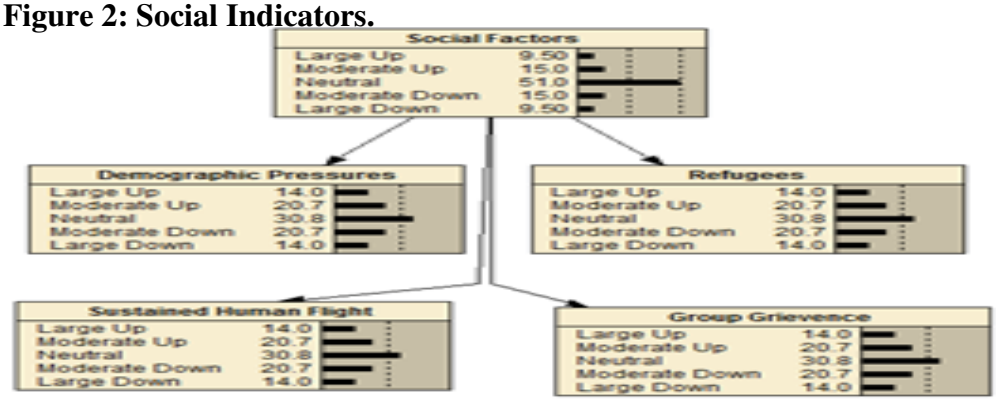
[8] Charniak, E., "Bayesian Networks Without Tears", *AI Magazine*, Vol 12:4 (1991).

[9] Figures are screenshots from a software tool called NETICA™, but many packages are available. See *http://www.cs.ubc.ca/~murphyk/Software/bnsoft.html* or Appendix B of Kevin B. Korb and Ann E. Nicholson, Bayesian Artificial Intelligence (CRC Press, 2003).

depends on the specific country under consideration. The default values in the template model place a higher likelihood on no change (80 percent) than a movement either up or down (10 percent each).

The three category nodes are similar to the hypothesis node, but slightly more detailed, to accommodate evidence that has varying weight on the movement of the hypothesis. The "Up"

Figure 2: Social Indicators.

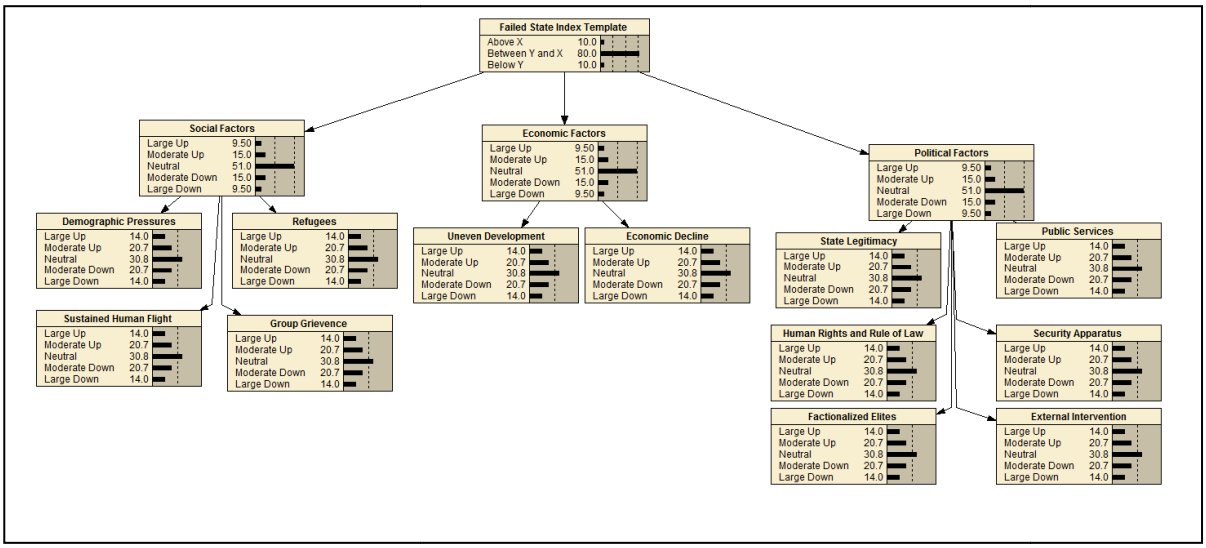


and "Down" states are each split into "Large" and "Moderate". Once again, neutral is the most likely, moving the likelihood of the hypothesis neither up nor down. By default, the moderate states marginally increase the likelihood of movement in the hypothesis, and the large states impact the movement of the hypothesis significantly. These default values are easily adjusted by an analyst if they are inconsistent with the specific situation being modeled. Additionally, if an analyst is more comfortable providing information differently, the structure of the model can be modified to ease the elicitation burden by reversing arcs or modifying the nodes.

Each of the three category variables has a set of indicators underneath it that further align the BN model with the CAST model. The CAST manual decomposes the primary indicators into a set of measures that are associated with samples of relevant questions that would inform that indicator. For example, the Social Factors category contains four indicators consisting of Demographic Pressures, Sustained Human Flight, Refugees, and Group Grievance (**Error! Reference source not found.**Figure 2) the logic of the conditional probabilities is similar to that described in **Error! Reference source not found.**Figure 1 above. The other two category variables – Economic Factors and Political Factors – are organized similar to Social Factors. The Economic Factors include Uneven Development and Economic Decline, which in turn would have measures like government debt, consumer confidence, and unemployment. Indicators for Political Factors are State Legitimacy, Human Rights and Rule of Law, Factionalized Elites, Public Service, Security Apparatus, and External Intervention. Measures include considerations like corruption of government officials, level of violence, perception of elections, and makeup of the government. Across the three categories there are a total of twelve indicators, with a much broader collection of measures underneath them, not all of which are relevant to the situation being considered by the analyst.

https://digitalcommons.usf.edu/jss/vol6/iss5/6
DOI: http://dx.doi.org/10.5038/1944-0472.6.3S.4

Figure 3: Complete failed-state Bayes net template.

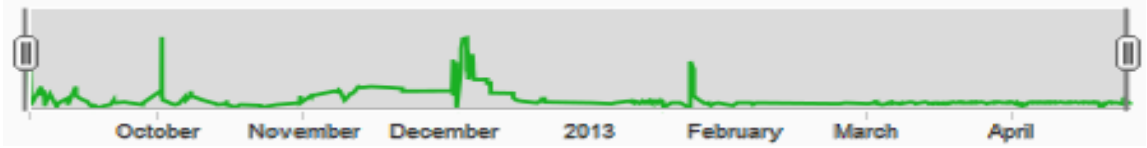


Error! Reference source not found. Figure 3 depicts the entire model, with all twelve indicators integrated into one model. This complete model provides a template the analyst can tailor to the specific country under consideration.

Country Case Study: Sudan

Our first case study is Sudan. For our prediction market, we chose to focus on a psychologically appealing threshold: "Will Sudan score less than 100 in the 2013 Failed States Index?". We launched the question in mid-2012, after the 2012 score was released. At first the target seems hopeless: Sudan scored a dismal 109.4 in 2012, and a nine point move is very rare: CAST scores usually only move a couple of points per year, especially in the positive, or down, direction. However, Sudan split in 2012, with South Sudan getting the poorer regions. Sudan retained the relatively wealthy north, albeit without access to the southern oil reserves. Will that enable it to cross the threshold? As of April 2013, the answer is unknown, but our model currently gives it only an 8 percent chance. As noted in Figure 4, our prediction market has basically let that value stand. How does the model arrive at the estimate?

Figure 4: DAGGRE Prediction Market Estimates.

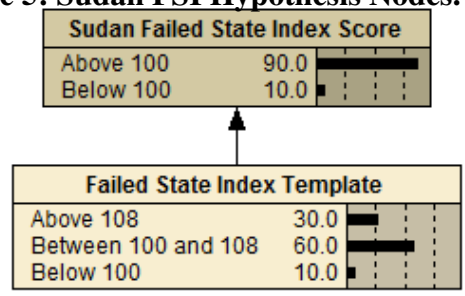


The DAGGRE prediction market estimates of the chance that Sudan will score less than 100 on the 2013 FSI (a substantial improvement). Now at 8% in early May, the chances have rarely exceeded 10%.

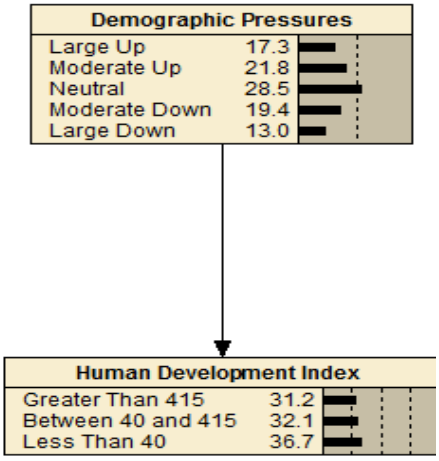
The first step in formulating the model is to adapt the hypothesis in the template to the specific situation under consideration. In this case the template has three states, but the market question only has two. **Error! Reference source not found.**Figure 5 shows how the hypothesis template

Figure 6: Human Development Index.

Figure 5: Sudan FSI Hypothesis Nodes.



was instantiated and then mapped into a two-state hypothesis. First, we set our template states to "Above 108", "Between 100 and 108", and "Below 100". (A score of exactly 100 would resolve in this case as above 100 or false.) Our initial judgment was that "Above 100" was nine times



more likely than "Below 100". Breaking this down, we judged "Above 108" as 30 percent, "Between 100 and 108" as 60 percent, and "Below 100" as 10 percent. We expected improvement at roughly 7:3 odds, but thought nine points was unlikely. But our intuitive starting judgments are of only passing interest. The real task was to identify drivers and indicators that are relevant to the hypothesis and would be known before June 2013.

We need a small set of relevant, non-redundant indicators. In addition to parsimonious models, we must be frugal with our limited forecaster time and points. Research on a question requires time. Forecasting itself requires time, and conditional forecasting multiplies the number of questions a forecaster must monitor.

To address the Sudan FSI question, we identified four factors in three categories and formulated them into questions for the prediction market. Specifically:

1. Will the Human Development Index for the Sudan for 2012 be: a) Less than 40, b) Equal to 40 but less than 41.5 or c) Equal to or Greater than 41.5? (This score is issued in November of each year by the UN)

https://digitalcommons.usf.edu/jss/vol6/iss5/6
DOI: http://dx.doi.org/10.5038/1944-0472.6.3S.4

2. Will the UN High Commissioner on Refugees report that more than 51,000 refugees were repatriated to the Sudan in 2012? (This is issued each January by the UNHCR)

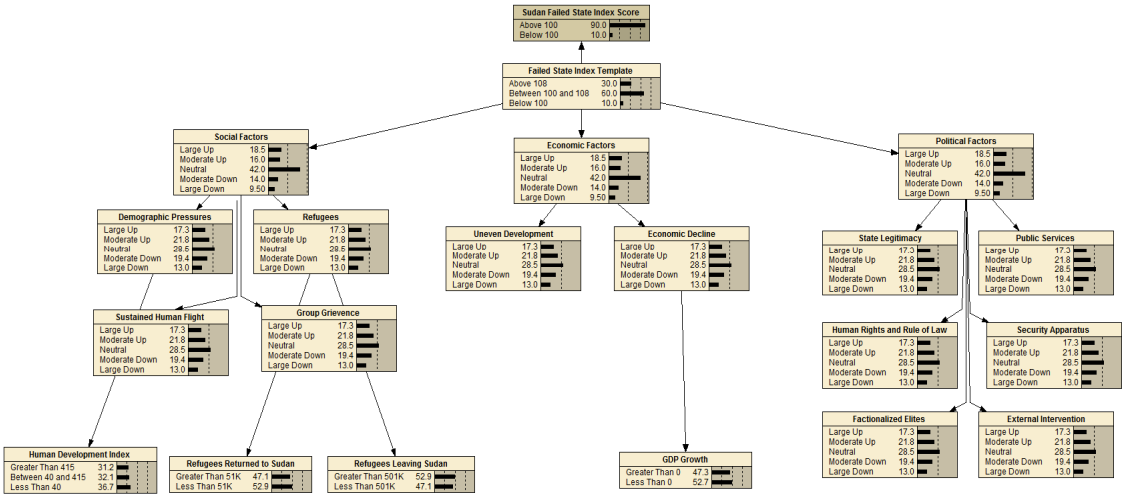
3. Will the UN High Commissioner on Refugees report that more than 501,000 refugees originated from the Sudan in 2012? (This is issued each January by the UNHCR)

Figure 7: Fully Specified Sudan FSI Model

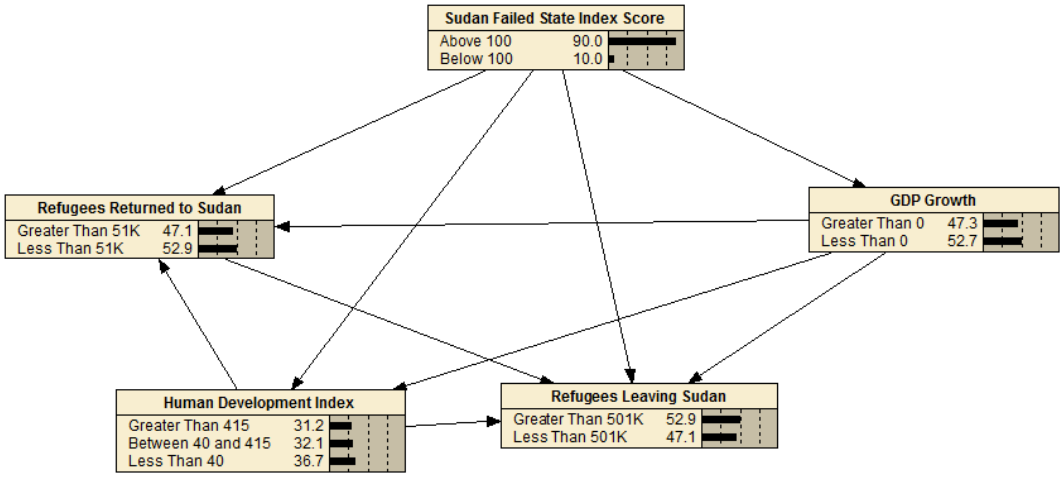
4. Will GDP Growth of Sudan exceed 0 percent in 2012 as published by the World Bank? (World Band Reports on this each February)

These questions are then integrated into the Bayes net model by associating them with a relevant indicator. For example, the Human Development Index is evidence of Demographic pressure,

Figure 8: Finished Sudan FSI Model



either positive or negative depending on the direction it moves. **Error! Reference source not found.**Figure 6 shows how the two nodes are related in the model and that the initial marginal probability of the question is relatively evenly distributed among the outcomes. The "marginal" probability is the unconditional probability. This is a consequence of constructing a conditional probability table similar to the ones described for **Error! Reference source not found.**Figure 2 and **Error! Reference source not found.**Figure 5, but associating improvements in the Human Development Index with downward movement in the indicator node. (Recall that increasing FSI indicates a deteriorating condition.) This process is repeated with the other three questions. The two refugee questions are associated with Refugees and the GDP question is associated with economic decline.



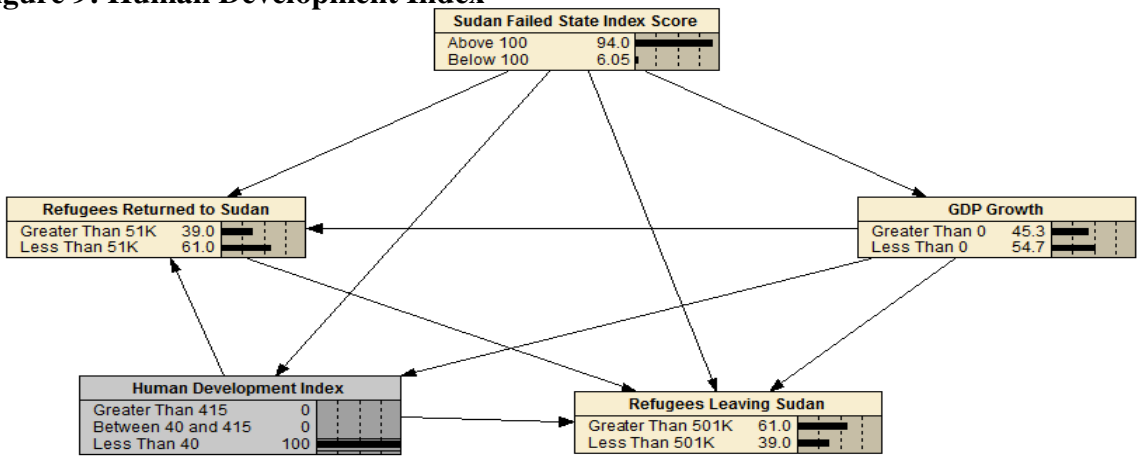
Once the process of customizing the template is complete, we have constructed a twenty-one node Bayesian network with almost 400 conditional probabilities, mostly derived from relationships in the template and only five of which are directly relevant to collecting market judgments or providing a forecast relative to the hypothesis (See **Error! Reference source not found.**Figure 7.) Fortunately, there are algorithmic operations that can reduce the model down to a simpler model that is probabilistically equivalent that contains fewer nodes and is more suitable for use by forecasters and analysts alike.

Error! Reference source not found. Figure 8 depicts the equivalent model for the purpose of forecasting the failed state index of the Sudan using the four identified indicators. Two types of nodes were "absorbed" to simplify the model. The first is nodes that will not inform the hypothesis. For example, we are not using the "political factors" branch at all. The second type is intermediate nodes between the question and the hypothesis. These nodes served their purpose in helping to structure the model and arrive at reasonable initial conditional probabilities. Now they only add complexity to the final model both computationally, and visually. Notice (compare **Error! Reference source not found.**Figure 8 to **Error! Reference source not found.**Figure 7) that the absorption process keeps marginal probabilities the same, and introduces new arrows.

We now have a standalone model which can revise the probability of our target node given beliefs about four key factors. So we could wait for those factors to become known, and update our estimate four times, as in **Error! Reference source not found.**Figure 9. But it would be better if the model could update the FSI continually, as our estimates of the surrounding factors changed. We can do that by putting all five questions on a live prediction market, and embedding our model as an agent in that market. The model can then update the FSI forecast to match changing beliefs about the interim factors, even before they resolve.

https://digitalcommons.usf.edu/jss/vol6/iss5/6
DOI: http://dx.doi.org/10.5038/1944-0472.6.3S.4

Figure 9: Human Development Index



Human Development Index resolves as "Less Than 40", increasing the chance that FSI resolves "Above 100".

Figure 10: DAGGRE Participant Interface

Prediction Markets in Intelligence Analysis

Prediction markets are an increasingly well-known approach for arriving at probability estimates for forecasting uncertain events.[10] From 2011 to 2013, we ran a prediction market called DAGGRE market as part of a geopolitical forecasting research project sponsored by IARPA, the Intelligence Advanced Research Projects Activity. DAGGRE stands for Decomposition Based

[10] Yiling Chen and David M Pennock, "Designing Markets for Prediction," *AI Magazine* 31:4 (January 13, 2011): 42–52; M. A. Chinn and L. A. Huffman, Prediction Markets: A Review with an Experimentally Based Recommendation for Navy Force-shaping Application (DTIC Document, 2009), available at: *http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADA514204*; Bill Gates et al., Prediction Markets for Defense Acquisition: The Devil Is in the Details, May 2010; R. Hanson, "Logarithmic Market Scoring Rules for Modular Combinatorial Information Aggregation," *The Journal of Prediction Markets* 1:1 (2007): 3–15; Justin Wolfers, Eric Zitzewitz, and National Bureau of Economic Research, Prediction Markets in Theory and Practice (National Bureau of Economic Research, 2006).

Aggregation. All prediction markets aggregate opinions on individual questions; DAGGRE allows questions to be related. This approach matches nicely with the indicators and measures approach used in the CAST model.

Error! Reference source not found. Figure 10 shows a screenshot of the participant interface for the DAGGRE market. The controls inside the ellipse display the current probability of an outcome and allow the forecaster to adjust the probability up or down. When a participant moves the probability they are wagering points against the outcome of the event. The size of the wager and the payout are determined using a logarithmic scoring rule, which has several nice properties, especially encouraging participants to provide forecasts that are consistent with their beliefs (it is a "proper" scoring rule).[11] This is similar to the approach used to evaluate the performance of weather forecasters. Additionally, the approach provides successful forecasters with supplemental resources, in the form of points won, to increase their participation in the market: over time, the best forecasters get the most influence.

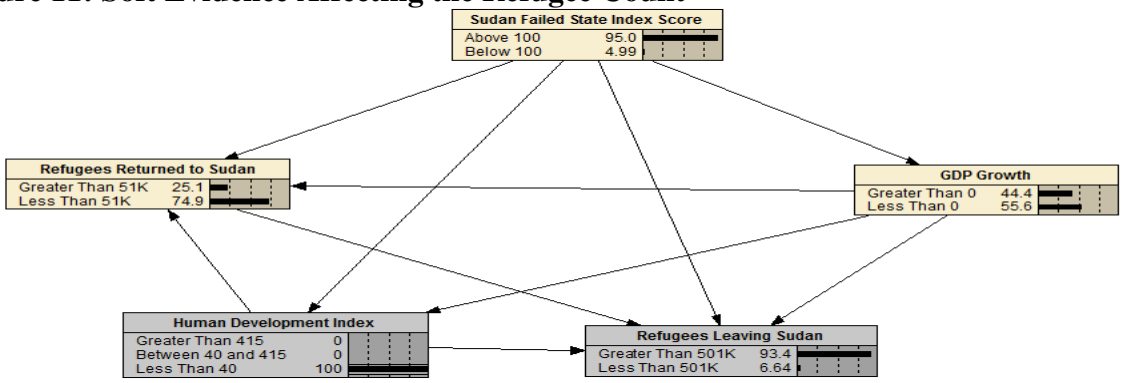
Well-formed questions in prediction markets and variables in Bayes Nets have some similar characteristics that allow them to complement each other nicely. In both cases the questions should pass a clarity test.[12] That is, the outcome of an event is unequivocally observable as having occurred or not. This requirement for precision and accuracy separates this approach from most geopolitical forecasting methods. For example, common forecasts like "Refugees will be a continued issue in Country X" do not pass the clarity test. We need questions like, "Will the January Refugee Report issued by the UN High Commissioner on Refugees indicate that the number of refugees leaving Country X exceeds Y people?" We may care more about whether refugees are "an issue" than about the number on one report, but if we are to evaluate and improve our forecasts, we have to cash out "an issue" in terms of measurable indicators. Only then can we integrate multiple variables into a broader model or to arrive at an unambiguous description of a complex forecasting situation.

The two approaches can be combined by making specific, measurable questions serve as indicators for the more complex hypothesis. This is especially powerful if the indicators will resolve sooner than the hypothesis. In this paper, we use the FSI as the focus hypothesis so that it too can be on the prediction market, but the technique could apply more generally to provide early indicators and warnings for a fuzzier but more interesting core hypothesis.

[11] Hanson, "Logarithmic Market Scoring Rules for Modular Combinatorial Information Aggregation."
[12] Howard, R.A. and Matheson, J., The Principles and Application of Decision Analysis, Strategic Decisions Group, Menlo Park, (1984).

https://digitalcommons.usf.edu/jss/vol6/iss5/6
DOI: http://dx.doi.org/10.5038/1944-0472.6.3S.4

Figure 11: Soft Evidence Affecting the Refugee Count



It remains to show how to update model probabilities from the market *before* the intermediate questions resolve, using the market probabilities as "soft" evidence. For example, let's assume that the Human Development Index question is resolved as in **Error! Reference source not found.**Figure 11. This stimulates a forecaster to do some additional thinking and research concerning refugees leaving Sudan and she concludes that it is 90 percent likely that this question will resolve as more than 501,000 refugees departing the country. She modifies the DAGGRE probabilities as shown in **Error! Reference source not found.**Figure 10 and the system in turn reports those probabilities to the Bayes Net. **Error! Reference source not found.**Figure 11demonstrates that using Bayes Rule to update the network we see that the calculated probability for that node is now over 93 percent and the probability that the Failed State Index will be over 100 is increased to 95 percent. It is important to note that the impact on the probabilities is not linear and is sometimes counterintuitive. These unexpected results are among the most powerful these types of models can provide. That is because the behavior of the model is largely a function of the local judgments the analyst provided at the time the model was constructed. The combination of "local judgments" in the computational model is very often more reliable than a holistic judgment made by an analyst or forecaster.

Another possible use of the Bayes Net, not demonstrated in this paper, is to integrate evidence from other sources into the Bayes Net as a complement to the values provided by the market. Using these techniques an updated probability perhaps based on reliable classified information could be entered into the model without compromising sources and methods. The model could then interact with either the market or the forecasters as previously described.

Autotraders – A Method for Improving Forecasting

On the combinatorial prediction market, we also introduced an autotrader that used the Bayes net values over time and traded in the market alongside human participants. The purpose of this was to improve the forecasting accuracy based on the new evidence that comes with respect to the questions in the model as well as the probabilistic coherence the Bayes Net provides.

The Bayes Net autotrader is an algorithm that trades on the targeted question ("Will Sudan score less than 100 in the Failed States Index?") by reading into the Bayes net the information from the

other Sudan questions on the market. The Bayes Net autotrader updates the model with the market's best estimate, and then updates the market to keep it consistent with the model. In the case of Sudan, the offline Bayes Net model originally predicted only a 10 percent chance for the FSI index to drop below 100, and it became even more confident (5 percent chance) as some of the supporting nodes/questions resolved.

At the cut-off time of the analysis, the DAGGRE market predicted 8 percent for the Sudan FSI to drop below 100, and the Bayes Net autotrader (the "online model") predicted 7 percent. The BN autotrader trades once per day for a maximum change of 3 percentage points, so the market substantially agrees with the model.

The users edited this question 321 times and the autotrader edited it ninety-nine times. This means that the autotrader is responsible for 30 percent of the forecasting activity on this question, the rest remaining to the human users. On another hand, the human users used the combinatorial features/ assumptions only twenty-six times (in only 12 percent of the human edits and only 8 percent of the time for the entire activity on this question). The current version of the autotrader does not directly edit the conditional probabilities on the market, but as described above, it uses its own conditional probabilities to keep the related market questions consistent.

Besides the market activity and the frequency of edits, we are also looking at the forecasting accuracy, particularly the Brier score, in order to assess the performance of the market, the offline model and the autotrader.[13] Since the question is still live on the market, we look at the Brier score in two cases – that the question resolves as False (final probability 0 percent) and that the question resolves as True (final probability 100 percent). The Brier score is a distance ranging from zero to two. The closer the Brier score is to zero, the better the forecasting accuracy.

DAGGRE is evaluated on the average Brier score over time for the life of the question, so effectively the average Brier score for **Error! Reference source not found.**Figure 4. The average Brier score of the combinatorial prediction market for this question (2013 Sudan FSI) would be 0.035 if the question closes with "No" and 0.774 if it closes with "Yes". The autotrader would score 0.035 for the "No" option and 0.807 for the "Yes" option. This means that the autotrader only slightly worsens the forecasts if the question closes with "Yes" (basically with a surprise outcome). This is expected, since the online model and the market have been moving in similar directions and the auatotrader is constrained in the size trade it is allowed to execute.

On another hand, the offline model, in the absence of any information from the market, would close with a Brier score of 0.0058 for the "No" option and a score of 0.86 for the "Yes" option. This means that the offline model would perform really well in one case and really bad in the other.

We can conclude though that over the lifetime of the question, the human traders have largely agreed with the model, but tempered its forecasts. If we repeated this trial on dozens of questions, the market (and especially the humans in it) would perform better than the model (and autotrader) for surprises, and earn points at the autotrader's expense. But it is also known that in

[13] Brier, G., "Verification of forecasts expressed in terms of probability". *Monthly weather review* 78 (1950): 1–3

https://digitalcommons.usf.edu/jss/vol6/iss5/6
DOI: http://dx.doi.org/10.5038/1944-0472.6.3S.4

many cases humans update their beliefs too slowly. If the conditional probabilities in the model are correct, then it is merely pushing the humans to be consistent, and over many questions should gain relative to them. Over time, the influence in the market will reach equilibrium between the automated and human traders, with each having the appropriate influence to create the most accurate forecasts.

Conclusion

The model demonstrated here and the associated research to date reinforces our belief in the potential of the DAGGRE, decomposition based approach to forecasting, especially when the market is coupled with Bayesian Networks for medium to long term forecasts of complex situations. We have demonstrated that the technology is computationally efficient, that Bayesian Network models can be constructed and integrated into the market with reasonable amounts of effort. Moreover the evidence to date on forecasting performance, while not conclusive, is very promising.

Future research is required to confirm or refute our beliefs about improved forecasting performance to be gained by integrating Bayesian networks and conditional prediction markets. Additionally, research on how other sources of information can be integrated into the market as well as how to efficiently update the conditional probability distributions in the model appears warranted to assess their potential.