Using a “Human + Machine” Approach to Determine Financial Market Regimes

Friday, September 07, 2018

Jennifer Wang, Zaid Shahzadeh – Totem Equities, IHS Markit

Bill Chung, Alvin Lee, Eric Takigawa – Financial Risk Analytics, IHS Markit

Financial markets are volatile and unpredictable, but do they display certain patterns that could be helpful in risk management and portfolio allocation?

We embarked on a 10-week journey and applied a “human + machine” approach to our Totem dataset to try and find some answers.

Totem Equity Implied Volatility Dataset

The journey started with the extraction of stock index implied volatility time series for S&P 500 (SPX), Euro Stoxx 50 (SX5E), Nikkei 225 (NKY) and Hang Seng (HSI) from the Totem Equity Implied Volatility database. The time series covered January 2005 to June 2017, included more than 180,000 volatility data points, and spanned across periods of major market turbulence, including the 2008 global financial crisis (GFC), 2010-2012 European sovereign debt crisis and 2015 CNY devaluation shock, among many others.

The unique feature of the Totem dataset, compared with publicly-available alternatives is the high quality and richness of the implied volatility data. In particular, we explored the Totem consensus data which is sourced from major financial institutions and has extended coverage of long-dated and deep out-of-the-money options.

Figure 1 shows the month-end Totem consensus valuation of SPX options in June 2017 which contained more than 400 individual data points. The Totem consensus pricing is widely used for product control, generating accounting statements and meeting regulatory requirements.

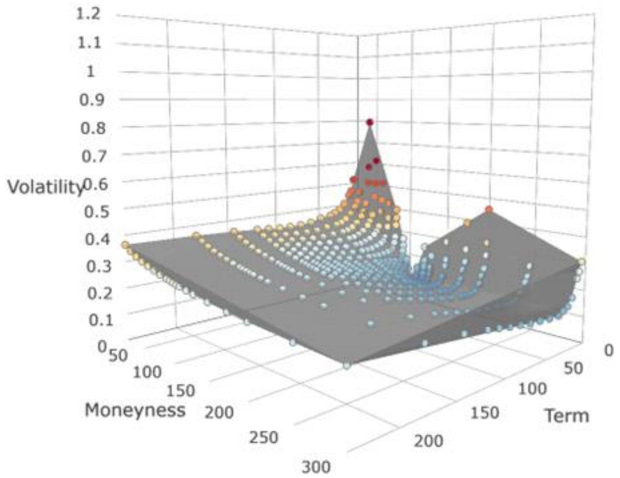

Figure 1: Totem consensus implied volatility data for SPX in June 2017

“Human + Machine” Data Science Process

We used clustering analysis, a machine learning algorithm which groups similar data points into partitioned clusters. In our case, each observation represented a set of features summarizing the state of the stock market, and observations being assigned to the same cluster were considered to be falling into the same market regime.

Specifically, we ran the Gaussian Mixture Model (GMM) algorithm, which can be considered as an extension of the more commonly used K-Means algorithm. Coding was done in Python and executed in Jupyter notebooks, using various open-source data manipulation, computation and visualization libraries such as pandas, numpy, scikit-learn and plotly.

A typical data science process was adopted, with steps such as exploratory data analysis, feature engineering, model building, model assessment and result analysis. In addition to just “letting the data speak for itself” through machine learning algorithms, we added a human overlay of subject matter expertise and judgement into the process.

For example, in the feature engineering step, techniques commonly used in financial markets to extract skewness and curvature using risk reversals and butterfly spreads were adopted instead of naively using all available data points. This relied on our regular conversations with Totem clients on financial market conventions and developments.

One significant challenge was encountered during the model building and assessment phase, where there was no “ground truth” available to quantitatively compare our results against. The usual process of calculating an error metric and using a train/test approach could not be applied to this unsupervised learning problem. Hence, there was no clear-cut way to determine if the specific set of features and number of clusters chosen were correct.

Or, to put it in another way: “Who truly knows how many market regimes there are? And when exactly do they occur?”.

To overcome this challenge, we again applied human judgement and considered several criteria to pin down the most appropriate model. The criteria included: (i) visually comparing clustering results against key market events globally and regionally, (ii) assessing cluster separation for key features such as the at-the-money (ATM) term structure and 6-month implied volatility smile, (iii) computing Silhouette Coefficient metrics and (iv) reviewing radar and parallel co-ordinate plots. Figure 2 gives an example of applying human judgement on SPX clustering results using the first two criteria above.

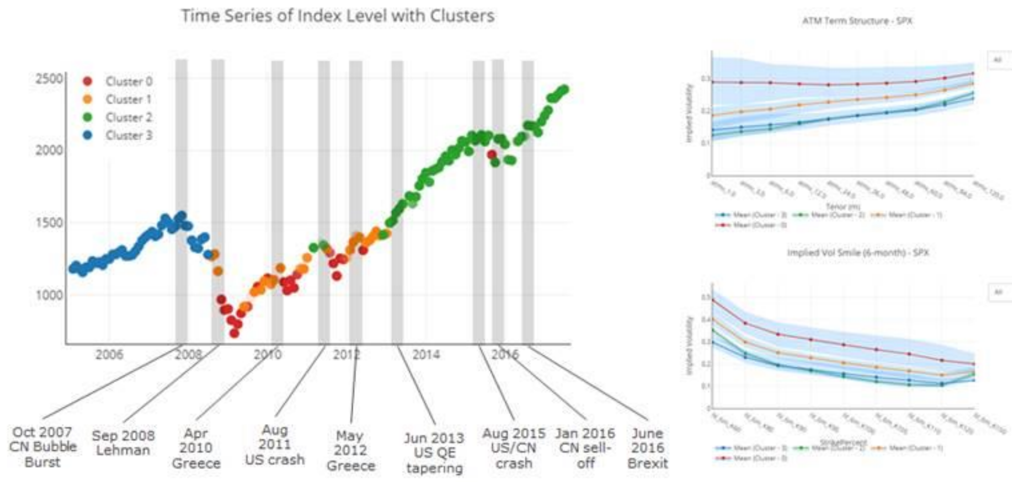

Figure 2: Human judgement on key market events and cluster separation of key features

Clustering Results and Interpretation

Once the most appropriate model was chosen, the next step was to interpret and label each of the four clusters based on their different characteristics. This helps to assign economic intuition to each of the clusters, instead of simply stating that “these are just results from the black-box model” and is another example of a “human + machine” approach. To compare the cluster characteristics along the seven features in the model, we used a radar plot of the normalized centroids (sample means in a cluster), as shown in Figure 3.

This is how we interpreted the results: cluster 0 (red) was labelled as “Crisis period” as it exhibited the highest ATM volatility level and risk reversal, indicating extreme levels of market stress. Another example, comparing cluster 2 (green) and cluster 3 (blue), even though both were during periods of relative calm with upward trending stock markets, the green cluster had significantly higher values for “dif_lowest_6m”, which is the feature for “extreme skew” representing the cost of buying protection against market crashes. Economically, it can be interpreted that in the Post-GFC calm periods, the financial markets have been pricing in the possibility of another global crisis ahead, resulting in higher “extreme skew” values compared to Pre-GFC and hence their respective labels.

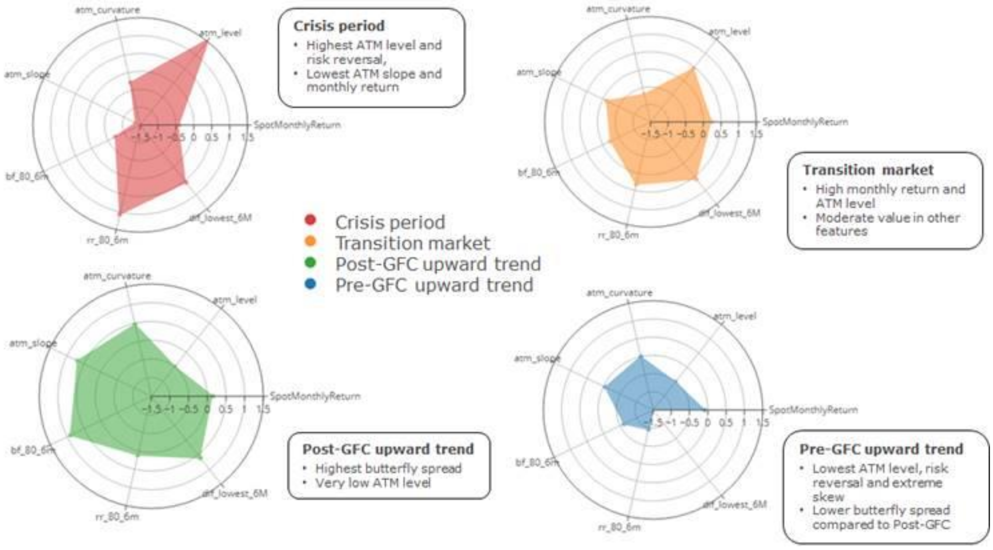

Figure 3: SPX cluster interpretation by comparing radar plots of different cluster centroids

So, going back to the question originally asked, it does appear that information contained in the Totem dataset displays certain patterns in detecting market regime. This observation is in itself interesting, but perhaps what is more interesting is the second part of the question: “… could [the patterns] be helpful in risk management and portfolio allocation?”.

Answering this second part was the focus of the final leg of our journey.

Stress Scenario Generation

One possible application of the clustering results is illustrated in Figure 4, where the red “Crisis period” observations were used to generate stress scenarios that consider different time points. This is different from the commonly used approach of defining stress periods as specific but consecutive time points. For example, the new Basel regulations for market risk (FRTB) require the identification of a specific 12-month stress period.

Arguably, the stress scenarios generated using this clustering approach would be more conservative as they would pick up the worst moments across the entire historical period, whereas a specific 12-month period would include some recovery moments. However, this very conservatism could be a useful gauge of a truly worst-case scenario.

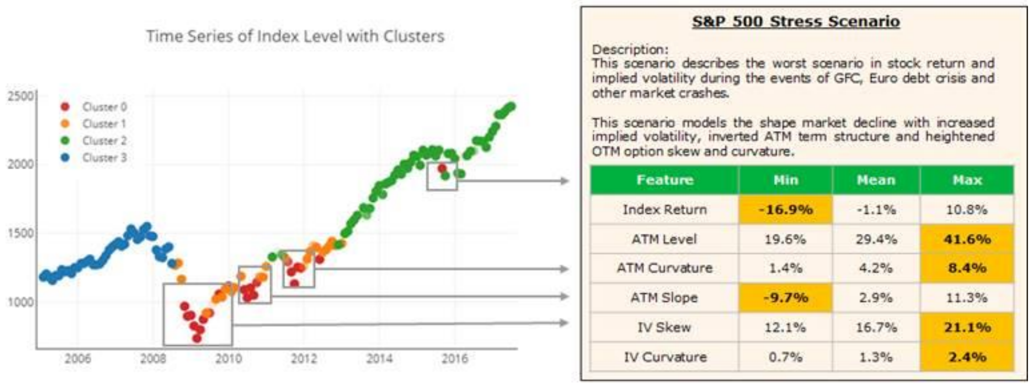

Figure 4: Applying clustering results to stress scenario generation

Counterparty Credit Risk Management

Another possible application is using the clustering results to dynamically manage counterparty credit risk exposure, as illustrated in Figure 5. Typically, counterparty credit risk managers would use a risk metric such as Maximum Potential Future Exposure (Max PFE), and compare it against policy limits determined by their risk appetite. They would also typically have a stress testing framework to compute more extreme risk metrics like stress losses.

However, the practical challenge is to figure out when to use the normal metric and when to use the extreme metric, or some blend of both. Using the extreme metric all the time could result in frequent breaches of policy limits, and result in very few trades being done and drastically reducing profits. Using the “normal” metric all the time could underestimate actual risk during periods of high market volatility and result in large credit losses, e.g. during the August 2015 shock devaluation of the CNH/USD exchange rate. Applying the clustering results to dynamically adjust the risk metric could be one systematic and balanced way to manage credit exposures.

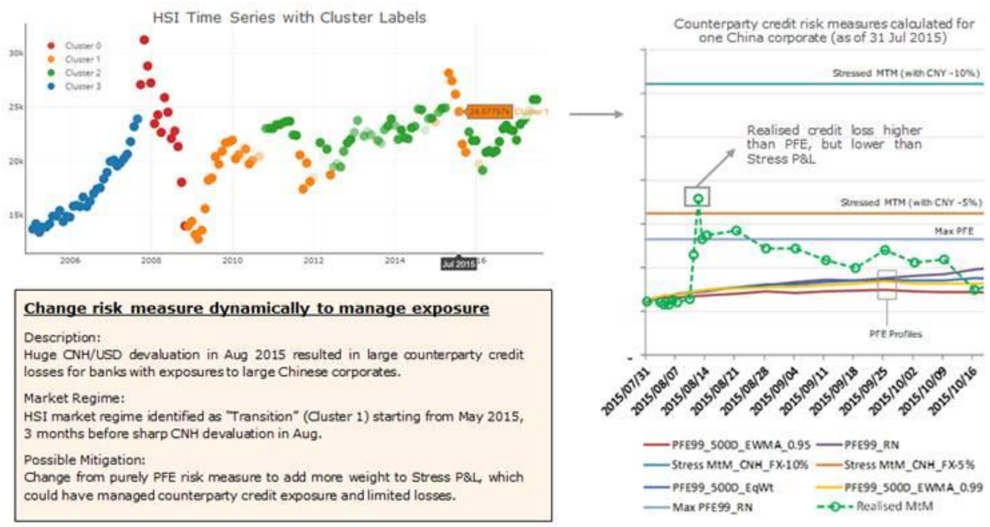

Figure 5: Applying clustering results to counterparty credit risk management

Nowcasting Monitor

By stacking the identified regimes of the four stock indices together as shown in Figure 6, we could generate a Nowcasting monitor ready for a cross-sectional comparison. A more interesting view would be to horizontally extend coverage of stock markets across geographical regions, and vertically drill down into underlying stocks that comprise the index.

The Nowcasting monitor would allow for a bird’s eye view of what is happening in the global financial markets, plus allow for granular stock level analysis to identify potential regional or sectoral spill-over during crisis periods. Preliminary discussions with banking regulators indicate that this would be of value from a macro-prudential risk management perspective, and hence would be a natural extension to the existing results.

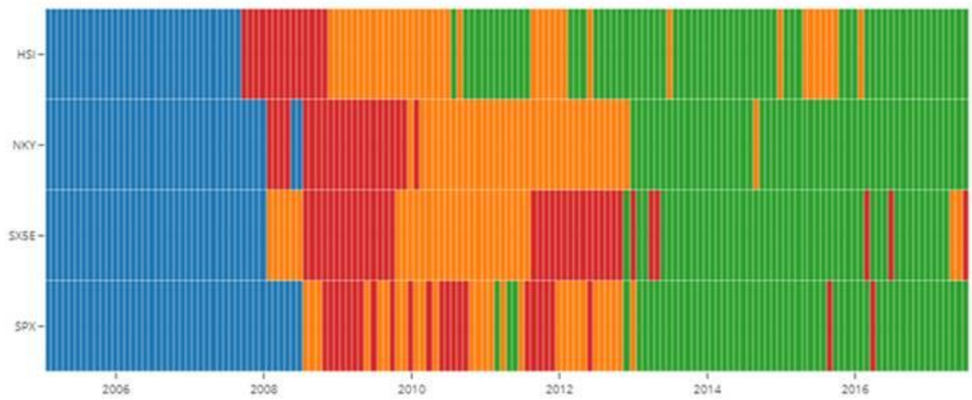

Figure 6: Applying clustering results to stock market nowcasting and cross-sectional comparison

Building on the idea of the nowcasting monitor, it would be useful if regime turning points could be detected. Admittedly, this is a challenging task, but by looking at the distribution of features for each cluster, perhaps we can determine a soft boundary to distinguish between different market regimes. Let us look at Figure 7 and consider the box plots for “ATM Level”. The line in the box represents the median, the top and bottom edges of the box represent the 25th and 75th percentile and the two ends represent the upper and lower bound.

Comparing the red “Crisis period” and green “Post-GFC upward trend” clusters, we can draw a line at the implied volatility level of 23% as a soft boundary separating the two clusters. Therefore, if we are currently in a green cluster, but if the implied volatility level starts creeping up towards 23%, perhaps we need to be prepared for the possibility of a regime change.

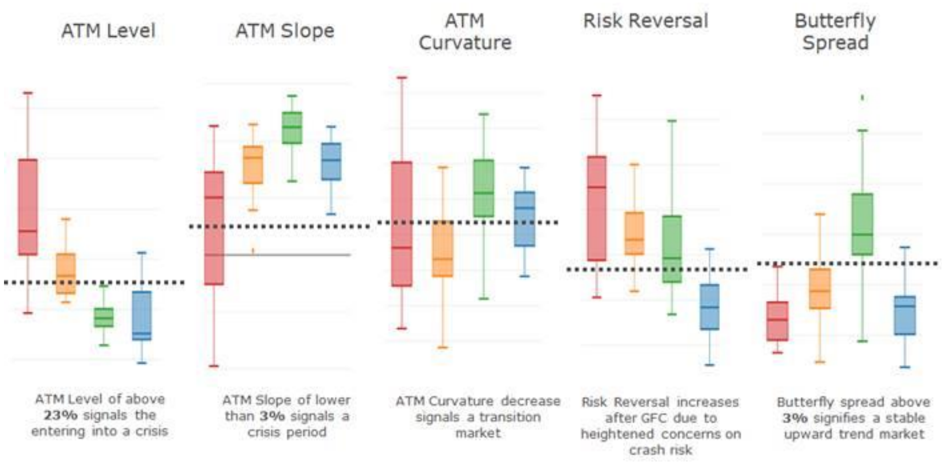

Figure 7: Applying clustering results to detection of market regime turning points

Tactical Portfolio Allocation

The final, and probably the most challenging application to consider is portfolio allocation. Figure 8 illustrates how we used the clustering results to guide dynamic allocation of an investment portfolio into the four stock markets.

The exercise was very simple: we over weighted or underweighted each stock market based on the identified market regime. For example, any market that falls into the red “Crisis period” cluster will be underweighted. The results were compared against an equally-weighted benchmark portfolio with 25% allocated to each stock market. An interesting note is that the GMM algorithm used for clustering comes with a probability measure that indicates the confidence of each clustering result. The tactical allocation incorporated this confidence measure, by rebalancing only when it exceeded a threshold of 70%.

This portfolio allocation strategy was seen to outperform the equally-weighted benchmark by 20% over a historical 12-year period. Indeed, the strategy over weighted HSI during the aftermath of the 2008 GFC and successfully captured the recovery in Hong Kong stock market. It also managed to underweight SX5E during the European sovereign debt crisis period and achieved relative outperformance against the benchmark with a static allocation to SX5E.

We also tried different settings and observed that a more aggressive weighting multiplier allowed us to achieve more than 40% relative outperformance. Another observation was that excluding input features unique to Totem (such as long-dated tenors and deep skew measures) generated strategies that did not pick up recovery periods well, and hence resulted in significantly lower performance.

It should be emphasized that this was just a simple exercise, and by no means a robustly tested investment strategy, hence the results should only be considered as illustrative. Having said that, it does give an idea of how clustering results could be incorporated as a mechanical overlay for tactical portfolio allocation.

Illustration of Tactical Allocation

	SPX	SX5E	NKY	HSI
Equal weighting	**25%**	**25%**	**25%**	**25%**
Cluster	0	1	1	2
Probability	90%	85%	88%	99%
Tilt	X 0.5	X 2	X 2	X 2
Tactical Allocation	**8%**	**31%**	**31%**	**31%**

Probability Threshold	70%

Cluster	Weight Multiplier
0	0.5
1	2
2	2
3	2

Figure 8: Applying clustering results to tactical portfolio allocation

Conclusion

At the end of our 10-week journey, applying a “human + machine” data science approach to the Totem dataset yielded interesting results, and even more interesting applications. We have been sharing these results and applications with various external parties, including banks, fund managers, regulators and university researchers across Asia Pacific and their responses have been uniformly positive.

Many ideas have surfaced during these discussions, including the addition of more IHS Markit datasets such as CDS, Bonds, Economic & Country Risk, Maritime & Trade among others.

We will continue to build on this initial foundation and move forward on our data science journey.