Intelligent Systems for Price Taking and Market Making

Sid Ghoshal, Steve Roberts
Department of Engineering Science, University of Oxford

OXFORD-MAN INSTITUTE

1. Introduction

Transactions occur in markets whenever one party (the *price taker*) buys or sells at the ask and bid prices of a liquidity provider (the *market maker*). This poses two of the most fundamental questions in finance:

- When is it a good time to trade, i.e. what data is most relevant in deciding when a price taker should act?
- What is the optimal bid and ask price the market maker should show to the price taker, subject to the dual constraint of profit maximisation and risk minimisation?

We assess the efficacy of Bayesian methods in solving both these problems. Using Automatic Relevance Determination (ARD) Gaussian Processes, we fuse time series data, sentiment metrics and analyst recommendations into a high-dimensional mean surface for time series returns, identifying the most salient features behind market fluctuations.

The same framework is used to provide the market maker with a probabilistic representation of each price taker's behaviour. Reflecting the market maker's preference for predictable clients, we measure the entropy of each counterparty's probability distribution as a proxy for adverse selection risk, and include it as an adjustment to the Hamilton-Jacobi-Bellman stochastic control solution to inventory risk.

2. Dataset and Methodology

The price taker attempts to forecast S&P500 stock market returns using information from 4 domains: technical analysis (MA, MACD, Divergence), sentiment (optimism in social media), the curvature of price space (inferred by options market activity) and broker opinions (Buy, Hold, Sell announcements).

The market maker tracks counterparty activity as a function of time of day, 1-minute anticipatory and reactive volatility, returns at the 1-min, 1-hour and 1-day timescale and prevailing bid-offer spreads.

A common thread to both the price-taking and market-making problem is the need to identify relevant features in their datasets. We compare standard student-t testing to the ranking produced by a Matérn 3/2 ARD kernel, chosen for matching the low smoothness and once-differentiability of financial time series.

$$f(\mathbf{x}) \sim \mathcal{GP}(m(\mathbf{x}), k(\mathbf{x},\mathbf{x}'))$$

$$k(\mathbf{x},\mathbf{x}') = \sigma_f^2\left(1+\frac{\sqrt{3}|\mathbf{x}-\mathbf{x}'|}{l}\right) \times \exp\left(-\frac{\sqrt{3}|\mathbf{x}-\mathbf{x}'|}{l}\right)$$

An adjacent concern for the market maker is the need for scalability, addressed by clustering counterparties into a small number of archetypes. We build an adjacency matrix from the Bhattacharyya distances between each pair of counterparties, and compare the results of Bayesian non-Negative Matrix Factorisation (NMF) to traditional k-medoid and hierarchical dendrogram approaches.

3. Price Taking

We compute the Pearson and Spearman correlation levels between each feature and the next-day return on the S&P500. Statistical significance tests provide a benchmark for gauging the ARD framework's ability to identify salience. We find that ARD Gaussian Processes provide a robust means of filtering irrelevant features, and produce interpretable visualisations of stock market return dynamics.

1. Technical Indicators, used by day-traders

Technical Indicators are metrics derived directly from the price history of a financial instrument. We do not believe the arcane definitions of these metrics is inherently meaningful, but rather that they provide measurable thresholds at which chartist market participants will react.

2. Sentiment Analysis, used by retail investors

While factual newsflow matters, it is specifically the polarity of its interpretation by markets – as beats or disappointments – that drives market movement. We capture market sentiment using bullishness metrics derived from both Twitter and Stocktwits, a social media site dedicated to real time discussions of financial markets. Stocktwits called into question the wisdom of crowds: optimism foreshadows declines, and conversely.

3. Price Space Curvature of the Options Market, used by hedge funds

Options markets provide a glimpse into the positioning of the market's informed, leveraged players. As strike-sensitive instruments, they can be used to anticipate price levels where competing forces will constrict returns, as well as areas where consensus will enable prices to gap. This motivates the representation of price space as an inhomogeneous fluid, with regions exhibiting directional or viscous behaviour.

$$\text{Directionality(t)} = \sum_{s\in S}\left(\text{OI(s,t)}_{Call} - \text{OI(s,t)}_{Put}\right) - \sum_{s\in S}\left(\text{OI(s,t-1)}_{Call} - \text{OI(s,t-1)}_{Put}\right)$$

$$\text{Viscosity(t)} = \sum_{s\in S}\left(e^{-\lambda|\text{price(t)-s}|} \times \log\left[\min\left(\text{OI(s,t)}_{Call}, \text{OI(s,t)}_{Put}\right)\right]\right)$$

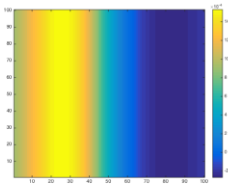

Fig. 1: S&P500 Daily Return Variation as a function of Stocktwits sentiment (x-axis) and Twitter sentiment (y-axis). "Be greedy when others are fearful, fearful when others are greedy."

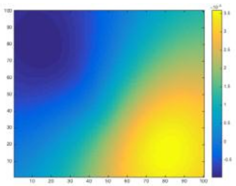

Fig. 2: S&P500 Daily Return Variation as a function of Directionality (x-axis) and Viscosity (y-axis). Option volumes indicate expert prepositioning for directional swings, as well as areas where competing forces will clash.

4. Broker Recommendations, used by institutional investors

Analysts regularly issue Buy, Hold and Sell recommendations on stocks, along with price targets. These have virtually no predictive power, and are discarded from further analysis.

4. Market Making

Optimal market making must account for the two main risks they face: inventory risk and adverse selection risk.

1. Stochastic Control solution to dynamic Inventory Risk Management

We model the intensity of bid and ask order flow as Poisson Processes $\lambda^b(\delta^b)$ and $\lambda^a(\delta^a)$, decreasing functions of the price gaps δ^b and δ^a between the mid price and the market maker's bid and ask prices. By adjusting these price gaps, market makers can control the rate at which their orders are filled, and encourage the execution of trades that neutralise their inventory.

2. Gaussian Process representation of Counterparty Behaviour

We construct probabilistic representations of the behaviour of each client using their entire trade history. Using a combination of Kernel Density Estimation and ARD Gaussian Process Regression, we produce a predictive map of their activity, providing the market maker with insights into which counterparties exhibit desirable, predictable behaviour and which ones are hardest to anticipate.

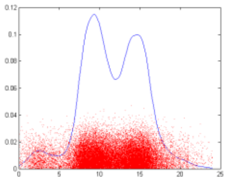

Fig. 3: Kernel density estimator of BNP Paribas's January 2013 trades in EURUSD as a function of time of day. Aggregate client activity peaks twice in the day, at 10:00 GMT and 15:00 GMT, reflecting high points in trade flow for London and New York, respectively.

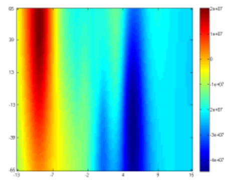

Fig. 4: Client deal volume as a function of 1-minute Returns (x-axis, in basis points) and 1-hour Returns (y-axis, in basis points). Near-term returns are systematically negative, reflecting the price taker paying half the bid-ask spread. Violations of this pattern would suggest adverse selection.

The information advantage of adverse selectors will also manifest as additional entropy in their behaviour, offering us a method for ranking the client base from least desirable to most.

3. Community Detection within the Counterparty Set

Clustering counterparties into separate groupings simplifies the market maker's task of customising bid-ask spreads. We borrow techniques from network analysis, applying Bayesian non-Negative Matrix Factorisation to an adjacency matrix built on the inverse Bhattacharyya distance between each counterparty pair. The method yields very similar results to k-medoids or a hierarchical clustering dendrogram, with the added benefit of highlighting nodes at the intersection between several classes via probabilistic membership scores.

4. A data-adjusted Control solution to jointly solve for Inventory Risk and Adverse Selection

Having devised entropy measures for each client, we adjust the HJB solution by a function f of the entropy of individual counterparties, whose form is a topic for future work.

5. Results

- The fusion of heterogeneous datastreams used by different segments of the market provides a forecasting model that outperforms benchmark Auto-Regressive and Kalman Filter models. Information is gained from each domain except broker recommendations. Most interesting of all: the price space curvature metrics provide the greatest improvement to the model's forecasts, and motivate the pursuit of further research into the price space representation of financial time series.

	Relevance		Pearson	
Feature	Score	Ratio	Correlation	p-value
Directionality	0.0920	2.8×10^3	+0.1460	0.0027
Viscosity	0.0877	2.6×10^3	+0.0560	0.2103
Stocktwits	0.0848	2.5×10^3	−0.1103	0.0133
MACD	0.0630	1.8×10^3	−0.1337	0.0027
Broker Change	0.0001	3.1	+0.0024	0.9564
Noise	< 0.0001	1	−0.0216	0.6288

Table 1: ARD-GP feature relevance alongside Pearson correlation and p-values. Price space metrics were particularly salient, and the correlation analysis broadly corroborated the ranking of the ARD methodology.

- Punitive adjustments to the bid-ask spread for entropic counterparties provides a means of controlling for inventory risk while tracking adverse selectors. Our community detection work provides a tool for the market maker to segment their client base and apply price adjustments to clusters instead of individual counterparties.

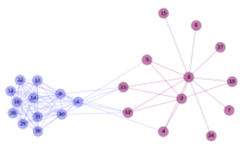

Fig. 5: Colour-coded visualisation of the adjacency structure constructed from Bhattacharyya distances between each pair of counterparties. Bayesian NMF placed nodes 4, 5, 11 and 12 in the overlap between the two clusters.

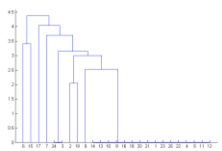

Fig. 6: Dendrogram of the connection IDs (x-axis) branched by Bhattacharyya distance (y-axis). The hierarchy closely matches the results of Bayesian NMF.

6. Conclusions

We have identified two closely related areas where Bayesian non-parametric methods can help rationalise the behaviour of market participants. The salience of price space calls into question the use of time series returns in financial modelling. The transformation to return series loses information about important price loci, an omission that in part explains the poor predictive performance of existing forecasting models.

AIMS Autonomous Intelligent Machines & Systems
EPSRC Pioneering research and skills