Merging Time Series with Specialist Experts

Tim Scarfe   Yuri Kalnishkan

RHUL Computer Learning Research Centre
Colloquium 2013

June 5, 2013

In this presentation:

Algorithm proposed for predicting options implied volatility

- Options/Data Set
- Transformation into time-series
- Aggregating Algorithm for Specialist Experts (AAS)
- Results and Conclusions

What is an option?

An option is a derivative financial instrument linked to an underlying asset, which is usually a share, but can also be a portfolio of shares, a futures on a share etc. There are two popular types of options, European and American, which differ by their execution arrangements.
For our purposes we can think of an option as being an object with parameters; *Strike Price* X, *Put/Call (1/0), Time To Maturity* T, *Volatility* σ. We are trying to predict the *Volatility*.

RTSSE Datasets

The datasets we use were provided by the Russian Trading System Stock Exchange (RTSSE) and record data from mid-2000s, when the Russian stock market was experiencing steady unperturbed growth. The options studied were American rather than European, which means that they could be executed any time before maturity, not simply on maturity.

Datasets Summary

Dataset	Underlying asset	Maturity	Number of transactions
eeru1206	futures on share	December 2006	13152
gaz307	futures on share	March 2007	10985
rts307	futures on index	March 2007	8410

Volatility vs strike, transactions 1000-2000 (gaz307)

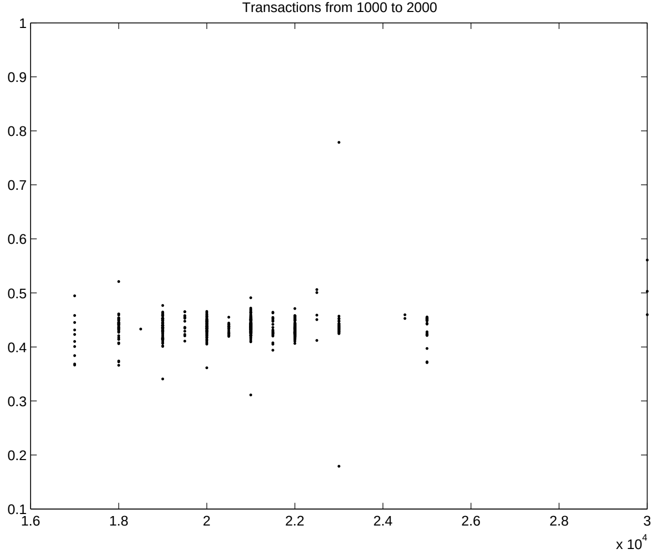


Figure : Volatility vs strike, transactions 1000-2000

Volatility vs strike, transactions 10000-11000 (gaz307)

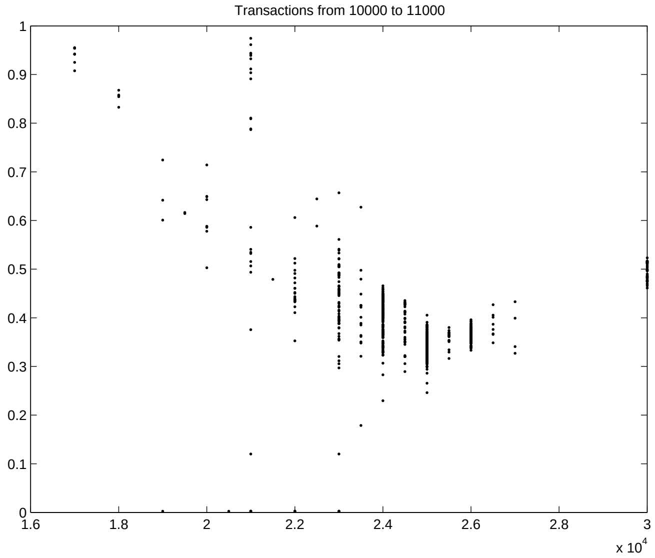


Figure : Volatility vs strike, transactions 10000-11000

Volatility vs number, transactions 1000-2000 (gaz307)

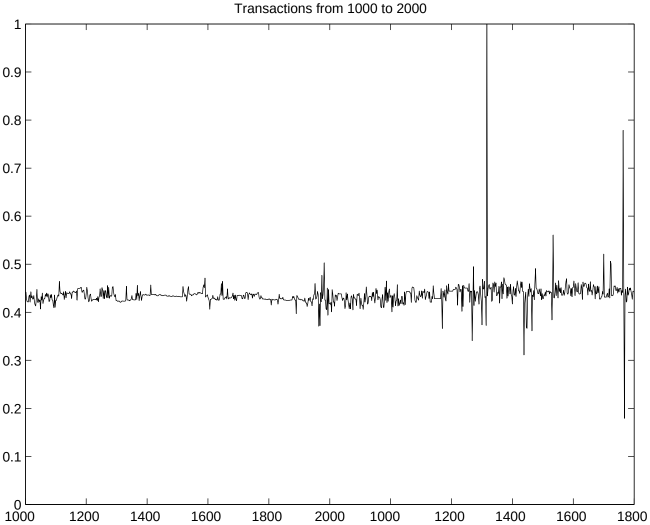


Figure : Volatility vs number, transactions 1000-2000

Volatility vs number, transactions 10000-11000 (gaz307)

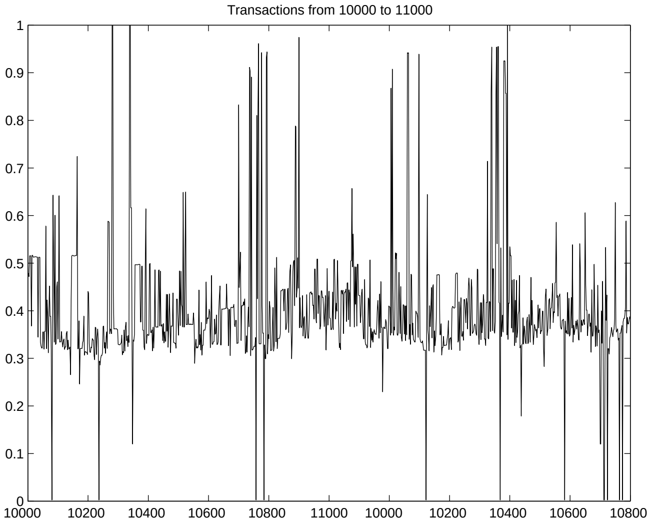


Figure : Volatility vs number, transactions 10000-11000 (looks like a time series)

Splitting into elementary time series

Suppose that we want to apply a time series prediction method. The simplest way of doing this is to treat the outcomes $\omega_1, \omega_2, \ldots$ as a time series ignoring the signals x_t altogether. Obviously this can lead to a loss of potentially useful information.

Splitting into elementary time series

- Interestingly, time series methods work well to predict volatility
- The number of possible strikes is limited. While theoretically the strike can have any real value, stock exchanges usually restrict strikes to some round numbers in order to improve liquidity. Thus one can consider splitting the time series into separate time series.
- Strike filtered time series work even better.

Splitting into elementary time series (2)

Consider an arbitrary finite subset $\{U_1, U_2, \ldots, U_K\} \subseteq 2^X$ such that $\cup_{k=1}^K U_k = X$. We will call sets U_k vicinities because it is natural to choose them in such a way that elements of U_i are in some respect akin to each other. Each vicinity U_i generates a specialist expert that predicts as follows. The expert is awake only on steps t where $x_t \in U_i$. It maintains the series $\omega_{t_1}, \omega_{t_2}, \ldots$ of outcomes for such steps and uses the series to make predictions for steps t where $x_t \in U_i$. For t such that $x_t \notin U_i$ the expert makes no predictions. The experts are then merged using the aggregating algorithm for specialist experts.

Splitting into elementary time series (3)

Let S consisting of $s_1 < s_2 < \ldots < s_L$ be the strikes for a dataset. A simple vicinity of diameter d is a set of d consecutive strikes; there are $L - d + 1$ vicinities of diameter d.
A compound vicinity is a subset of $S \times \{0, 1\}$, where the $0/1$ bit denotes whether the option in transaction is a put or call. A compound vicinity of diameter d is a product of a vicinity of diameter d by either 0 or 1; there are $2(L - d + 1)$ compound vicinities of diameter d. Note that some of them may give rise to empty time series if, say, there were no transactions on put options with particular strikes. However every transaction belongs to at least one compound vicinity.
In the experiments below we took all simple and compound vicinities of diameters from 1 to d with $d = 5$.

Prediction with Expert Advice

```
for t = 1, 2, ...
  experts θ ∈ Θ announce predictions γ_t(θ) ∈ Γ
  learner outputs γ_t ∈ Γ
  nature announces ω_t ∈ Ω
  each expert θ ∈ Θ suffers loss λ(γ_t(θ), ω_t)
  learner suffers loss λ(γ_t, ω_t)
endfor
```

Prediction with Expert Advice (2)

Over T trials each expert θ suffers the cumulative loss

$$\mathrm{Loss}_T(\theta) = \sum_{t=1}^{T} \lambda(\gamma_t(\theta), \omega_t)$$

and the learner suffers the cumulative loss

$$\mathrm{Loss}_T = \sum_{t=1}^{T} \lambda(\gamma_t, \omega_t) \ ;$$

one wants the inequality $\mathrm{Loss}_T \lesssim \mathrm{Loss}_T(\theta)$ to hold for all $T = 1, 2, \ldots$ and $\theta \in \Theta$.

Aggregating Algorithm

Given a learning rate $\eta \in (0, +\infty)$ and an initial distribution over the set of static experts θ; a distribution can be represented by an array of initial weights $p_0(\theta), \theta \in \Theta$.
The algorithm maintains an array of weights $w_t(\theta), \theta \in \Theta$. Their initial values are $w_0(\theta) = p_0(\theta), \theta \in \Theta$, and they are updated according to the rule

$$w_t(\theta) = w_{t-1}(\theta) e^{-\eta\lambda(\gamma_t(\theta),\omega_t)} = p_0(\theta) e^{-\eta\,\mathrm{Loss}_t(\theta)} \ .$$

Aggregating Algorithm (2)

$$\mathrm{Loss}_T(AA) = \sum_{t=1}^{T} \lambda(\gamma_t, \omega_t) \leq c(\eta)\,\mathrm{Loss}_T(\theta) + \frac{c(\eta)}{\eta} \ln 1/p_0(\theta) \ .$$

The aggregating algorithm performs nearly as well as the best expert loss wise (assume $c(\eta) = 1$ if η is selected optimally).

Specialist Experts

Suppose that an expert in the prediction with expert advice framework can abstain from making a prediction on step t; if it does so, we say that it sleeps on step t.
If an expert θ sleeps on step t, let us assume that it suffers notional loss $\lambda(\gamma_t(\theta), \omega_t)$ – in other words, it goes 'with the crowd'.

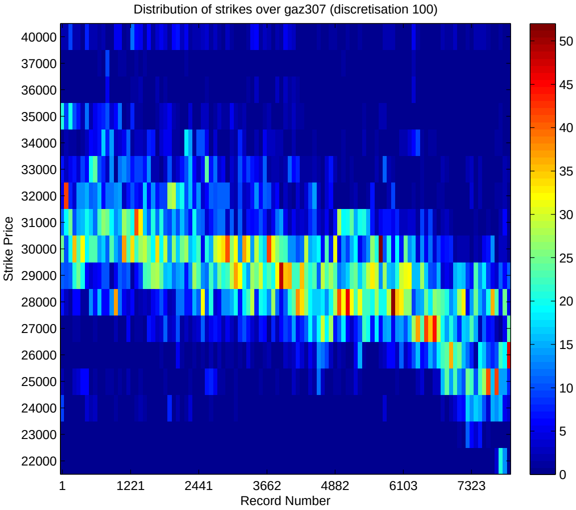
Specialist Experts (2)

Arguing as in the case of the AA, we get a similar bound; by dropping equal terms in the losses on the left- and right-hand side we obtain

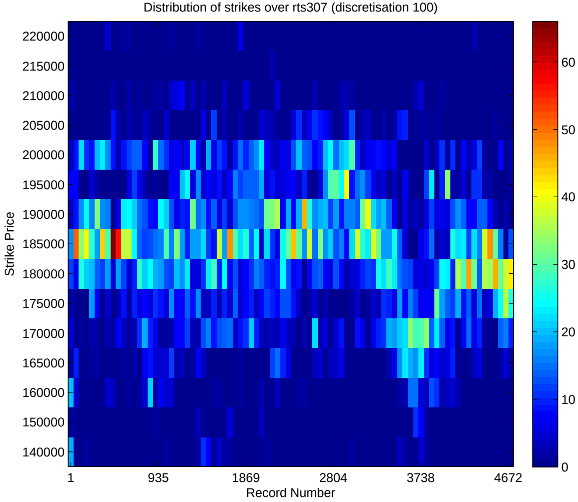
$$\mathrm{Loss}_T^{(\theta)}(AAS) \leq c(\eta)\,\mathrm{Loss}_T^{(\theta)}(\theta) + \frac{c(\eta)}{\eta}\ln 1/p_0(\theta) \ ,$$

where the sum in $\mathrm{Loss}^{(\theta)}$ is taken only over steps when expert θ was not sleeping (again for simplicity assume $c(\eta) = 1$).

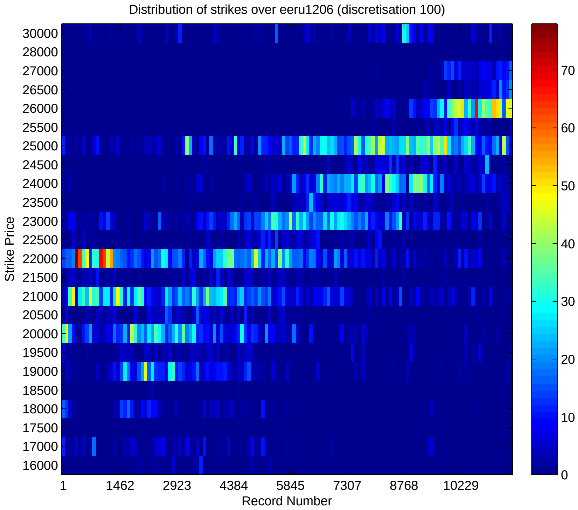
Distribution of Strikes (gaz307)



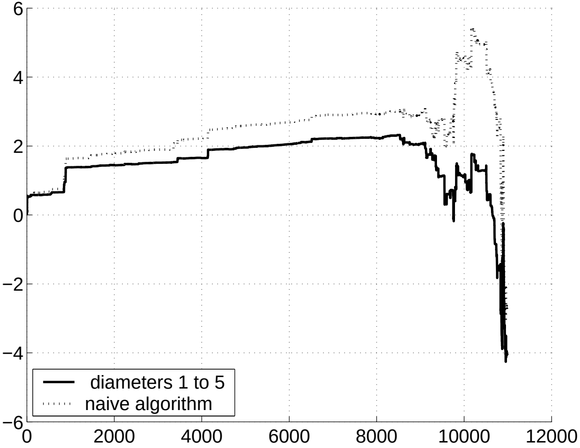
Distribution of Strikes (rts307)



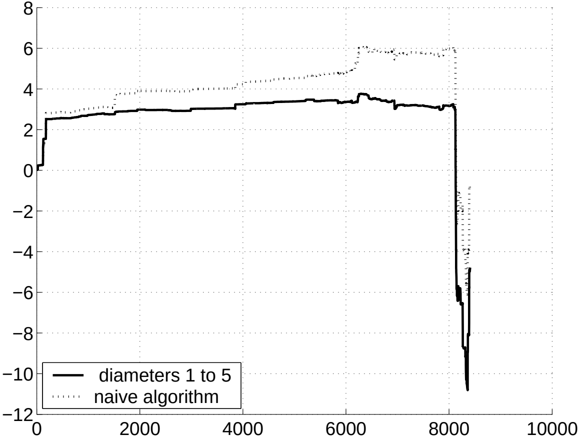
Distribution of Strikes (eeru1206)



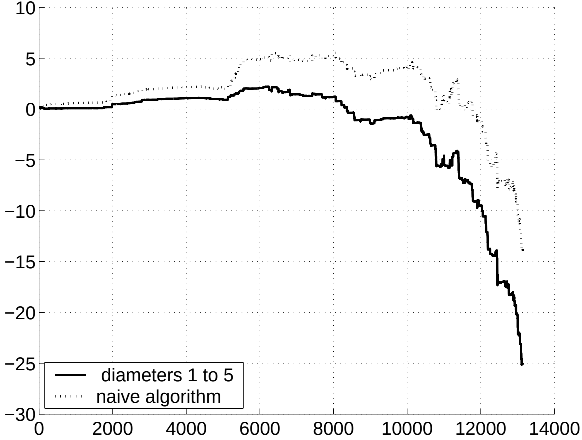
Predict last element on gaz307



Predict last element on rts307



Predict last element on eeru1206



Improvements for different ranges of vicinities

Maximum size	eeru1206	gaz307	rts307
1	20.05	2.55	1.36
2	23.18	4.69	3.57
3	24.60	4.91	4.54
4	25.21	4.91	4.95
5	25.64	4.91	5.14
6	25.84	4.84	5.19
7	25.96	4.76	5.18
8	26.04	4.68	5.15
9	26.08	4.57	5.12
10	26.09	4.49	5.11
11	26.11	4.39	5.08
12	26.10	4.34	5.06
13	26.09	4.31	5.03
14	26.08	4.28	5.01
15	26.07	4.26	5.00

Observations

1. Simple time series methods applied strike-wise perform comparably to the RTSSE proprietary technique (better at end).
2. The figures show that the competitor outperforms our methods at the beginning. A plausible explanation is that the competitor incorporates some prior knowledge about the behaviour of volatility, while our methods need to learn from scratch.
3. Vicinities of different sizes help

Concluding Remarks

1. On some strikes the most recent transaction happened long ago. We may want to collate them.
2. How long ago is long ago? We have a trade-off recent in time vs close in space.
3. We resolve it in the spirit of prediction with expert advice merging all options and letting the weights sort out the trade-off automatically.
4. Making use of sleeping experts, a new and exciting algorithm!
5. Algorithm is fast
6. Parsimonious methods working as well as algorithms like Kernel Ridge Regression!
7. We are looking at ARIMA and other methods to work with the elementary time series

Thank you.