1 Volatility Calculation

The purpose of this report was to estimate the volatility historical stock prices based on four sample datasets spanning a 170 day period. The data included biases and errors which required cleaning and the resulting code was designed to run automatically and unsupervised as part of a scheduled set of tasks.

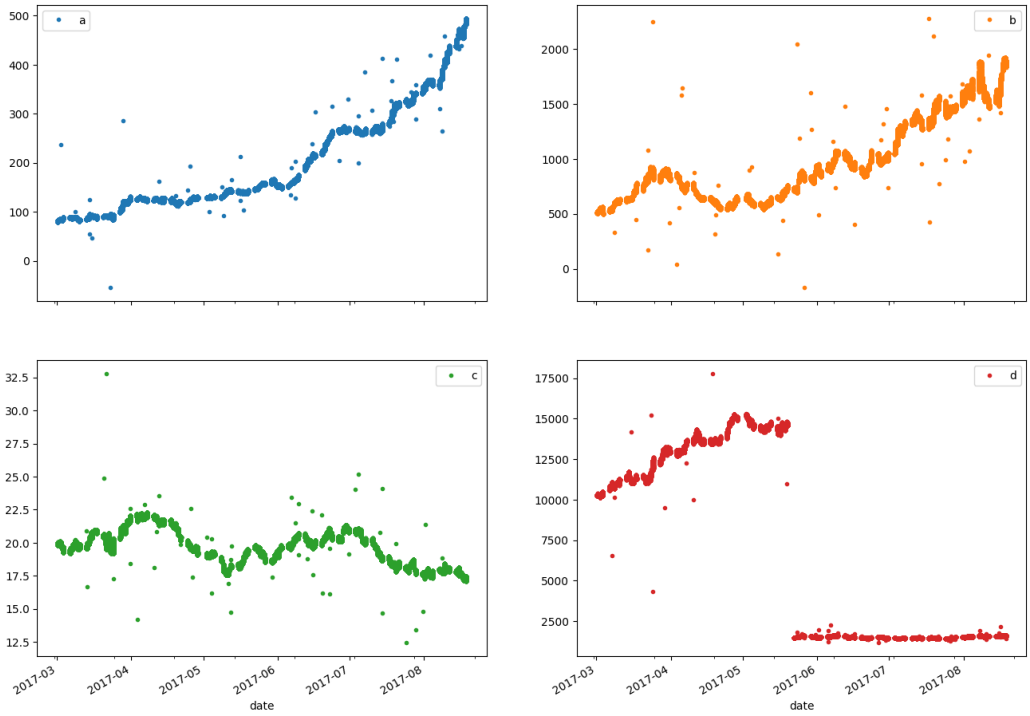


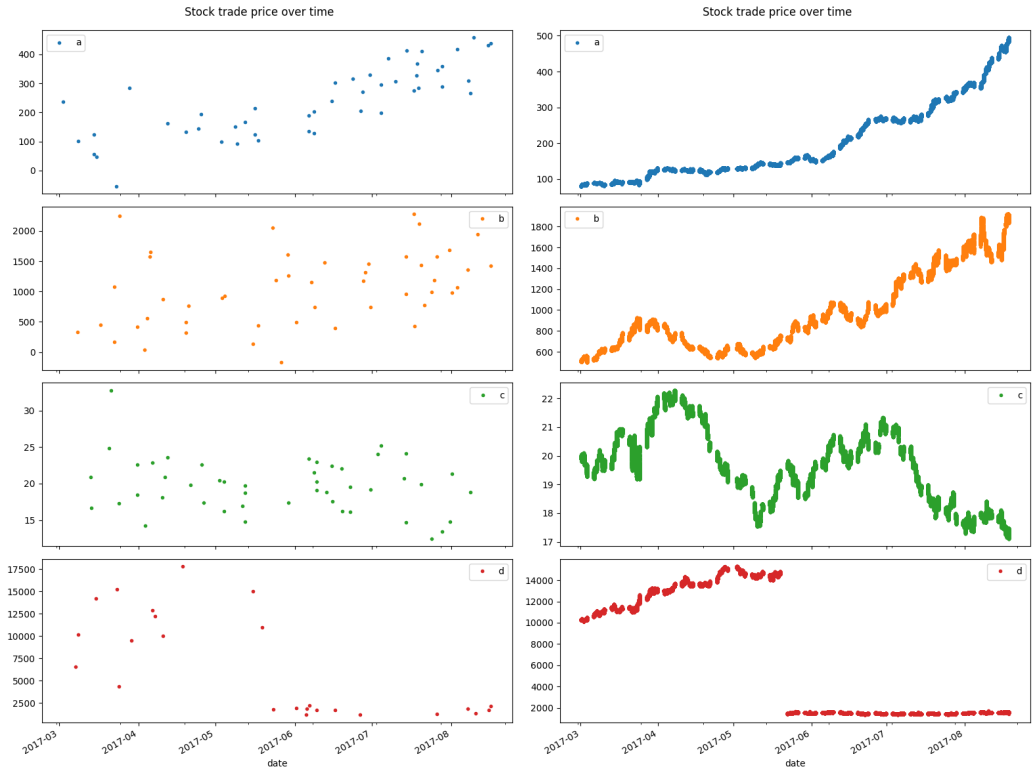
Figure 1: Raw data for four stock trade time series.

2 Outlier Detection

We are interested primarily in *additive outliers* which are defined as a type of outlier which does not affect the trend within the data and typically caused by data entry or human error.

We use the two-sided median method of outlier detection: for a given point in time, y_t, the median value $m_t^{(k)}$ is computed within a $2k$ neighbour-

hood of point y_t. Where the absolute difference between y_t and the median exceeds a threshold τ, y_t is identified as an outlier.[1, 2]



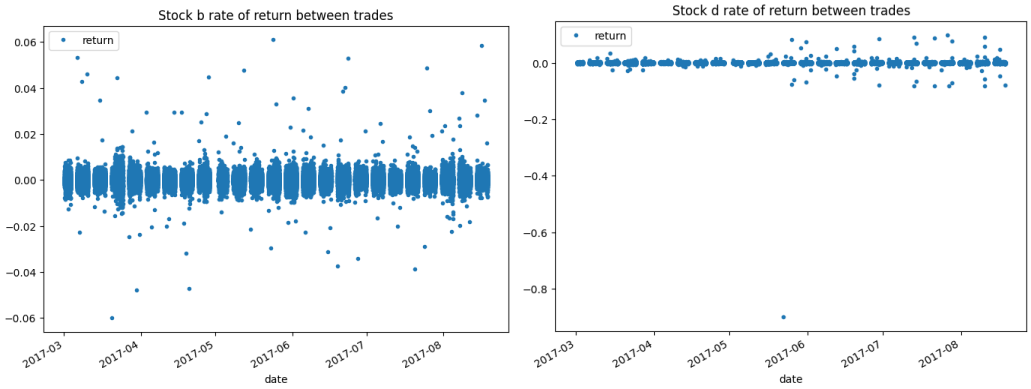
(a) Outliers detected in data. (b) Cleaned data.

Figure 2: Outlier detection and removal.

3 Volatility Calculation

The volatility calculation was based on the percentage annualised volatility of returns, given by the formula: $volatility = std \times \sqrt{252}$ where std is the standard deviation of the daily rate of return. A key consideration for further work is to evaluate changes in volatility over time, as more recent trends may be more relevant to the application. Such changes are seen in the volatility plot for stock d (fig. 3), which becomes more volatile after the mid point of the series. Note that the outlier here represents the level shift and can be

ignored for the purposes of this comparison.



(a) No change in volatility over time. (b) Increased volatility after mid point.

Figure 3: Comparison of volatility over time.

Volatility is then calculated using the latest observed price for each day as an estimation of the closing price of each stock (table 1).

Stock	Volatility
A	145.9%
B	221.1%
C	31.61%
D	149.3%

Table 1: Percentage annualised volatility.

4 Further work

If this process is to be performed every day (and small gains in efficiency are relevant), outlier detection can be performed once per time window and stored to the source data files rather than recalculating each day. It also assumes that data will be supplied in a relatively similar format – further conformity checks are recommended. Finally, the level shift change exhibited in stock d is not handled as doing so was not within the specified requirements, not are changes in volatility over time, but the significance of this change should be discussed with domain experts.

References

[1] Basu, S., and Meckesheimer, M. Automatic outlier detection for time series: an application to sensor data. *Knowledge and Information Systems 11* (2007), 137–154.

[2] Pearson, R. K. Data mining in face of contaminated and incomplete records. In *Proc. of SIAM Intl. Conf. Data Mining* (2002).