DESIGN OF AUTONOMOUS ALGORITHMIC MODELS FOR TIME SERIES PREDICTION

Eva Zámečníková
Doctoral Degree Programme (5), FIT BUT
E-mail: xzamec10@stud.fit.vutbr.cz

Supervised by: Jitka Kreslíková
E-mail: kreslika@fit.vutbr.cz

Abstract: This paper is focused on basic concepts used for processing of high frequency data. The idea of designing of systems for predicting of these time series and its parallel to business proces rules modeling will be mentioned. Designed system will use principles of statististical arbitrage, time series correlation, the use of multivariate variables and characteristics of the distribution of interim data. Newly designed system must meet the condition of econometrics for high frequency data.

Keywords: statistical arbitrage, fair value, high frequency data, time series, econometrics, business rule

1 INTRODUCTION

One of the main source of high frequency data are financial markets. For processing of these data we need to put together statistic, mathematic, economic and also informatic method and algorithms. Statistical methods can predict time series well but the results are not so well when there are noises in the time series such as inaccurate or incomplete data. Market data is very variable as each time interval (known as tick) generates new logical unit of data.. Main focus of todays research is not only good descriptive systems but also the ability to produce predictions of future movements of data. Information for this section were mainly taken and updated from [2].

The goal of this work is not to describe all basic terms as there is no space for it, for explanation of basic notions and terms the author recommends literature [1], [2] and [5]. In the following sections will be given overview of the basic concepts for processing of high freuqency data which are useful for the future work.

2 STATISTICAL ARBITRAGE

Statistical arbitrage belongs to one of the business strategy, more specifically it is an investment strategy. This strategy involves pairs trading. It uses statistical measures to detect mispricing between two assets based on the expected value of these assets. We assume that if the course of these two titles developed historically the same way, then their mutual deviation will only be temporary. The higher the data frequency, the more arbitrage opportunities appear.

Mathematically, the steps involved in the development of statistical arbitrage (or simply stat-arb) trading signals are based on a relationship between price levels or other variables characterizing any two securities. A relationship based on price levels $S_{i,t}$ and $S_{j,t}$ for any two securities i and j can be arrived at through the following procedure:

1. Identify the universe of liquid securities – that is, securities that trade at least once within the desired trading frequency unit. For example, for hourly trading frequency choose securities that trade at least once every hour.

2. Measure the difference between prices of every two securities, i and j, identified in step (1) across time t:

$$\Delta S_{ij,t} = S_{i,t} - S_{j,t}, t \in [1, T] \tag{1}$$

where T is a sufficiently large number of daily observations. According to the central limit theorem (CLT) of statistics, 30 observations at selected trading frequency constitute the bare minimum. The intra-day data have high seasonality – that means that persistent relationships can be observed at specific hours of the day. Thus, a larger T of at least 30 daily observations is recommended. For robust inferences, a T of 500 daily observations (about two years) is desirable.

3. For each pair of securities, select the ones with the most stable relationship – security pairs that move together. To do this perform a simple minimization of the historical differences in returns between every two liquid securities (defined by , Gatev, Goetzmann, and Rouwenhorst (2006) [1]):

$$\min_{i,j} \sum_{t-1}^{T} (\Delta S_{ij,t})^2 \tag{2}$$

The stability of the relationship can also be assessed using cointegration and other statistical techniques. Next, for each security i, select the security j with the minimum sum of squares obtained in equation 2.

4. Estimate basic distributional properties of the difference as follows. Mean or average of the difference:

$$E[\Delta S_t] = \frac{1}{T} \sum_{t-1}^{T} \Delta S_t \tag{3}$$

Standard deviation:

$$\sigma[\Delta S_t] = \frac{1}{T-1} \sum_{t=1}^{T} (\Delta S_t - E[\Delta S_t])^2 \tag{4}$$

5. Monitor and act upon differences in security prices: At a particular time τ, if

$$\Delta S_\tau = S_{i,\tau} - S_{j,\tau} > E[S_\tau] + 2\sigma[\Delta S_\tau] \tag{5}$$

sell security i and buy security j. On the other hand, if

$$\Delta S_\tau = S_{i,\tau} - S_{j,\tau} < E[S_\tau] - 2\sigma[\Delta S_\tau] \tag{6}$$

buy security i and sell security j.

6. Once the gap in security prices reverses to achieve a desirable gain, close out the positions. If the prices move against the predicted direction, activate stop loss.

This section was based on [1], [3], and [8].

3 FAIR VALUE

According to [5] the fair value is *"The calculated price of a given security, typically an option, such that neither counterparty to a trade at that price would experience an economic gain or loss. It is also known as fair market value."*. The impact of changes in fair value is recognised as profit or loss in the period they occur.

According to International Financial Reporting Standards (IFRS) [9] fair value is a market-based measurement and the entity's intention to hold an asset or to settle or otherwise fulfil a liability is not relevant when measuring fair value. When measuring fair value we use assumptions that market participants would use when pricing the asset or liability under current market conditions. Characteristics of a particular asset or liability that a market participant would take into account when pricing the item at the measurment date, include: age, condition and location of the asset restrictions on the sale or use, risk characteristics, cost of and return on capital or individually perceived utility.

Fair value is measured by using the price in the principal market for the asset or liability or, in the absence of a principal market, the most advantageous market for the asset or liability.

4 TIME SERIES CORRELATION

In this chapter will be introduced high frequency time series (or high frequency data - HFD) and its correlation. Financial markets are the source of discrete high frequency data. The original form of market prices is *tick-by-tick* data. Each tick is one logical unit of information. According to spacing in time we distinguish two types of data - homogeous (regularly spaced in time) and inhomogeneous (irregularly spaced in time) [2].

Data typically arrive as a random sequence of time points – the more activity on the market the denser the data. We study and make research on this data to understand the markets and to predict the behaviour of data. Tick-by-tick data allow to study the market microstructure and to decide what type of rules are the most appropriate for the markets to function efficiently. There is need to find new ways of defining the analysis of the data because of its volume: interpolation methods, data cleaning, etc. It is important to develop statistical methods with minimal assumptions of the underlying process.

4.1 CORRELATION OF HIGH FREQUENCY DATA

Correlation of the relative measure of mutual dependence in the development of two time series x_t, y_t is given by following relation:

$$s_{xy} = \frac{\sum_{t=1}^{n}(x_t - \overline{x}).(y_t - \overline{y})}{s_x.s_y} \in < -1; 1 > . \qquad (7)$$

Correlation values approaching the limit value -1 means that the two time series have a completely opposite direstions in their time development. Values close to 1 reveals that the time series x and y evolves almost identically in terms of the same direction and have the same relative pace in the mutual development.

Correlation between returns of different financial assets play an important role in fields such as risk management [1]. One of the known problem concerning corelation is that correlation between financial time series data vary over time.

5 MULTIVARIATE RANDOM VALUES

In mathematics, probability, and statistics, a multivariate random variable or random vector is a list of variables with unknown values, either because the value has not yet occurred or because there is not good knowledge of its value. The individual variables in a random vector are grouped together because there may be correlations among them – often they represent different properties of an individual statistical unit (e.g. a particular person, event, etc.). Normally each element of a random vector is a real number [3].

Random vectors are often used as the underlying implementation of various types of aggregate random variables, e.g. a random matrix, random tree, random sequence, random proces, etc.

More formally, a multivariate random variable is a column vector $X = (X_1, ..., X_n)^T$ (or its transpose, which is a row vector) whose components are scalar-valued random variables on the same probability space $(\Omega, \mathcal{F}, P)$, where Ω is the sample space, $\mathcal{F}$ is the sigma-algebra (the collection of all events), and P is the probability measure (a function returning each event's probability) [4].

6 BUSINESS RULES

This section is based on information from [6]. Business rules approach manages the flow of bussines process by using constrains or decision blocks. For use in processsing of high frequency data could be helpful to build the process using business rules at runtime. The aim for my future work will be to use business rules on real historical high frequency data and to do experiments and measurements and compare those results with real results and to control them to ensure econometrics requirements.

There exists many definition of notion of business rules. According to [7] business rule is a statement that defines or constrains some aspect of the business process which is intended to assert business structure, or to control or influence the behavior of the business. A business rule cannot be broken down or decomposed further into more detailed business rules.

Main characteristic of business rules are:

- Business rule classify, compute, compare and control data to direct the flow in a business process.
- Business rules can certify the data in the business forms.
- Business rules are not processes.
- Rules provide criteria for making decisions.

Formalization of business rules lies in identifying of the atomic statement as the definition of a term, fact, constraint, or derivation. Terms, facts, and some of the constraints can be represented as a graphical models. Constraints and derivations must be translated into other formalism.

6.1 CATEGORIES OF BUSINESS RULES

We distinguish several groups of business rules, these groups are:

- Definitions of business terms

  The most basic element of a business rule is the language used to express it. The very definition of a term is itself a business rule that describes how people think and talk about things. Thus, defining a term is establishing a category of business rule. Terms have traditionally been documented in a glossaries or as entities in a conceptual model.

- Facts relating terms to each other

  The nature or operating structure of an organization can be described in terms of the facts that relate terms to each other. To say that a customer can place an order is a business rule, but a fact. Facts can be documented as natural language sentences or as relationships, attributes, and generalization structures in a graphical model.

- Constraints (also called “action assertions”)

  Every enterprise constrains behavior in some way, and this is closely related to constraints on what data may or may not be updated. In many cases, the motive for preventing the creation of a new data record is to prevent the ensuing action from taking place.

- Derivations

  Business rules (including laws of nature) define how knowledge in one form may be transformed into other knowledge, possibly in a different form.

7 CONCLUSION AND FUTURE WORK

The main effort of models for processing and prediction high frequency data is to find model that should have predictive power. In combination with an intuitive understanding of the market, it should be possible to derive models that give more insights into general market behavior. An important goal of model making should be to reproduce also statistical properties found in the data.

The next step of my work will be to design an approach to predict time series based on use of business rules. The main part will be to execute experiments on real historical data and to compare the experimental and real results. New model should adapt the rules and should be dynamic. Model should also recognize noisy data or outlier values in data. Econometricts conditions should be also fullfilled.

ACKNOWLEDGEMENT

This work was partially supported by the BUT FIT grant FIT-S-14-2299, “Research and application of advanced methods in ICT”.

REFERENCES

[1] Franke, J., Hardle, W.K., Hafner, C. M.: Statistics of Financial Markets, Berlin, Springer, 2008, ISBN 978-3-642-16521-4

[2] Gencay, R., Dacorogna, M., Muller, U.A., Pictet, O., Olsen, R.: An Introduction to High-Frequency Finance, London, Academic Press, 2001, ISBN 978-0-12-279671-5

[3] Lai, T. L., Xing, H.: Statistical Models and Methods for Financial Markets, New York, Springer, 2008, ISBN 978-0-3-877-7826-6

[4] Bauwens, L., Pohlmeier, W., Veredas, D. (Eds.).: High Frequency Financial Econometrics, Recent Developments, Heidelberg, Physica Verlag, 2006, ISBN 978-3-7908-1992-2

[5] Durbin, M.: All About High-Frequency Trading, McGraw-Hill, 2010, ISBN 978-0-0-7174344-0

[6] Debevoise, T.: Business Process Management with a Business Rules Approach: Implementing The Service Oriented Architecture, Canada, Business Knowledge Architects, 2005, ISBN 978-1419673689

[7] Herbst, H.: Business Rule-Oriented Conceptual Modeling, Heidelberg, Physica Verlag, 1998, ISBN 978-3-7908-1004-2

[8] Aldridge, I.: High-Frequency Trading: A Practical Guide to Algorithmic Strategies and Trading Systems, New Jersey, John Wiley & Sons, Inc., 2010, ISBN 978-0-470-56376-2

[9] IFRS Foundation [online, visited March 2014] <http://www.ifrs.org/Pages/default.aspx>