AUTOMATIC SYSTEM OF DETECTING INFORMED TRADING ACTIVITIES IN EUROPEAN-STYLE OPTIONS

Moshenets M. K. and Kritski O. L.
The Department of Higher Mathematics and Mathematical Physics, Tomsk Polytechnic University, Russia
E-Mail: olegkol@tpu.ru

ABSTRACT
We propose a computer-based system of detecting the informed trader activities in European-style options and their underlying asset. The model (9) with moving average component was written. Being added to it ARMA-process for log-price differences of underlying asset, the generalized model is written as Vector ARMA, stable at $|\rho|<1$. We also proposed a mathematical criterion of informed trader activity presence which is a corner stone of automatic detecting system constructed.

Keywords: computer detecting system, informed traders, high-frequency trading, ARMA, European style options, underlying asset.

INTRODUCTION

Constructing an automatic system of identifying the informed traders is an actual and complex problem. In this work we define informed traders as large institutional investors (such as pension or hedge funds, market makers, dealers and etc.). They play an important role in maintaining the operating performance of the stock market by providing it with a high liquidity and low risk aversion levels [3], by keeping up some level of prices and trading volumes of financial instruments and by smoothing significant price volatility. Besides the special functions, they use privileges which are not accessible to individual investors and speculators [8]: they may obtain information about important macroeconomic indicators faster; they may also possess nonpublic information about firms before it is officially published in the quarterly or annual financial reports. For example, in [8] authors determine the statistical significance of return changes and show an increase in trading activity before the official announcement time of such important economic indicators as Chicago PMI and UMich Sentiment. It appears that subscribers receive information 3 and 5 minutes earlier than others. And, at the time of the announcement published data is already included in the price and significant spikes disappear.

The quantity of papers devoted to the detection of insider and informed traders in option and futures markets is not so large. Thus, in [7] authors verify the put-call parity for American options with two different strikes for different market agents and compute the probability of informed trading (PIN). And, Hu [1] found an order imbalance in trades of the underlying asset and its three options, i.e. in-the-money (ITM), at-the-money (ATM) or out-of-the-money (OTM) options that enables someone to compute the probability of informed traders. Muravyev et al. [5] combine those two approaches.

The possibility of addressing futures prices for detection of informed traders was first noted in [10], where it is shown that the enhancement of the trading activity in the futures market anticipates an increased activity in the spot-market, but in most cases the direction of trade (to sell or to buy) remains unclear, even if using the non-public information.

This work develops the multi-period mixed model [4] of monitoring insider trading activities on a stock market. We generalized detecting model [6] to option trading case and found the dependence between differences of function values computed at some deltas for European-style options, and underlying asset price returns. Equations obtained were written as Vector ARMA-process. The stationarity of this one was proved. At last we formulated, constructed and tested a detection criterion for automatic system to identify the informed traders.

The model
Let’s assume that number of all players on the market, trading an underlying asset and their put and call options, are divided into informed traders and ordinary “noise” traders. Let a macroeconomic announcement affecting the price, be publicly known at a future time $T$, while the informed trader has obtained the data at time $t<T$. We assume that he decides to buy (sell) the underlying asset or options in equal installments at regular intervals, i.e. at time $t, (t+1), \ldots, T$. Then the effect of changes in the underlying asset price could be defined as

$$X_t = v_t + u_t,$$

where $u_t \sim N(0, \sigma_u^2)$ – addition to the price, offered by uninformed traders, $v_t$ – surcharge to the price that an informed trader is willing to pay.

Let $v_t$ obey the relation

$$v_t = \beta \Psi_t,$$

where $\beta$ – coefficient of proportionality, $\Psi_t = \ln \theta$, – nonlinear asset price dynamics, $\theta$ – value of trading transaction of informed trader.

We suppose that $\Psi_t$ satisfies AR(1) model, which is explained by the desire of the informed trader to hide his activity and, for instance, to reduce his contribution at low levels of market activity:
\[
\Psi_t = \overline{\Psi} + \rho \Psi_{t-1} + \varepsilon_t, \quad (1)
\]

where \( \overline{\Psi} \) – average log-value of trading transaction price per time unit, \( \varepsilon_t \sim N(0, \sigma^2_t) \) – price noise.

Let \( S_t \) be a quote of the underlying asset at time \( t \). Since a trader buys a large quantity of the underlying asset, we assume that log-price will vary proportionally to the change of prices:

\[
\ln S_t = \ln S_{t-1} + \lambda X_t, \quad (2)
\]

where \( \lambda = \text{cov}(\Psi_t, X_t) / D(X_t | \Psi_{t-1}) = \beta \sigma^2 / (\beta \sigma^2 + \sigma^2_u) \) – market depth variation coefficient, if noise processes \( \varepsilon_t, u_t \) are independent of one another.

Denote \( R_t = \ln S_t \), so, we can rewrite (2) in more convenient way as

\[
\Delta R_t = \lambda X_t, \quad (3)
\]

that describes log-return price dynamics for underlying asset. The expression (3) coincides with model of asset price differences investigated in [6]. So, we can state our main results of this paper for \( \Delta R_t \) implements a mixed strategy (1), (3) with \( \lambda = \frac{\beta \sigma^2}{\beta \sigma^2 + \sigma^2_u} \), the log-price returns \( \Delta R_t \) follows the ARMA(1,1)-process below

\[
\Delta R_t = \gamma + \rho \Delta R_{t-1} + \varepsilon_t + \delta \varepsilon_{t-1}, \quad (4)
\]

where \( \varepsilon_t \sim N(0, \sigma^2_t) \), \( \gamma = \beta(1-\rho) \ln S_T - \ln S_0 / T \),

\[
\sigma^2_t = \lambda^2 \beta^2 \sigma^2_u \left( 1 + \rho^2 \right) \left( \beta^2 \sigma^2 + \sigma^2_u \right) \delta^2 + \delta^2 \left( 1 + \rho^2 \right),
\]

\[
\delta = \left[ \sigma^2_u \left( 1 + \rho^2 \right) + 2 \beta^2 \sigma^2_u \right] \left( 1 - \rho \right) \sigma_u \sqrt{4 \beta^2 \sigma^2 + \sigma^2_u},
\]

\[
\tau_1 = \left( 2 \rho \sigma^2_S - 2 \beta^2 \sigma^2_u \right)^{\frac{1}{2}}.
\]

The proof is listed in the Appendix A.

Theorem 1. In case an informed trader implements a mixed strategy (1), (3) with \( \lambda = \frac{\beta \sigma^2}{\beta \sigma^2 + \sigma^2_u} \), the log-price returns \( \Delta R_t \) follows the ARMA(1,1)-process below

\[
\Delta R_t = \gamma + \rho \Delta R_{t-1} + \varepsilon_t + \delta \varepsilon_{t-1}, \quad (4)
\]

where \( \varepsilon_t \sim N(0, \sigma^2_t) \), \( \gamma = \beta(1-\rho) \ln S_T - \ln S_0 / T \),

\[
\sigma^2_t = \lambda^2 \beta^2 \sigma^2_u \left( 1 + \rho^2 \right) \left( \beta^2 \sigma^2 + \sigma^2_u \right) \delta^2 + \delta^2 \left( 1 + \rho^2 \right),
\]

\[
\delta = \left[ \sigma^2_u \left( 1 + \rho^2 \right) + 2 \beta^2 \sigma^2_u \right] \left( 1 - \rho \right) \sigma_u \sqrt{4 \beta^2 \sigma^2 + \sigma^2_u},
\]

\[
\tau_1 = \left( 2 \rho \sigma^2_S - 2 \beta^2 \sigma^2_u \right)^{\frac{1}{2}}.
\]

The proof is listed in the Appendix A.

Use the nonlinear price dynamics (3) to discover an informed trading activity in European-style option trades.

It is well-known that fair price of European-style option with expiry date \( T \), implied volatility \( \sigma \), underlying asset price \( S_0 \), expiration price \( E \) and risk-free interest rate \( r \) follows the well-known Black-Scholes formula [2] at any time \( t \):

\[
V_t = S_t \Phi(d_1) - E e^{-r(T-t)} \Phi(d_2),
\]

where

\[
d_1 = \frac{\ln S_t - \ln E + (r + \sigma^2 / 2)(T-t)}{\sqrt{(T-t)}}, \quad d_2 = d_1 - \sigma \sqrt{T-t},
\]

\[
\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2} t^2} dt, \quad 0 \leq t \leq T.
\]

Because variables \( V_t \) and \( d_i, i = 1,2 \), depend on \( S_t \) in a non-linear way, it is impossible to write model for differences \( \Delta V_t \) like it was made earlier in expressions (3) and (4). So, we will use classical hedging coefficients \( \Delta_t = \frac{\partial V_t}{\partial S_t} \) to obtain all equations of our model.

**European style call option**

It is well-known that delta for European-style call option is equal to

\[
\Delta_t = \Phi(d_i),
\]

where

\[
d_i = \frac{\ln S_t - \ln E + (r + \sigma^2 / 2)(T-t)}{\sqrt{(T-t)}}, \quad 0 \leq t \leq T.
\]

Therefore if we take a reversed to normal p.d.f. function \( \Phi^{-1}(y) \) from all parts of expression above, we get

\[
d_i = \Phi^{-1}(\Delta_i), \quad d_{i+1} = \Phi^{-1}(\Delta_{i+1}),
\]

i.e., denoting \( Q_i = \Phi^{-1}(\Delta_i) \), we obtain

\[
\Delta t_i = \Delta Q_i, \quad (5)
\]

We take advantage of Taylor series expansion of the function \( \sqrt{1 + \tau} \) near the point \( \tau_0 = \tau - 1 \):

\[
\sqrt{1 + \tau} = \sqrt{\tau} + \frac{1}{2\sqrt{\tau}} + O\left( \frac{1}{\sqrt{\tau}^3} \right).
\]

Then

\[
\sqrt{1 + T-t} = \sqrt{T-t} + O\left( \frac{1}{\sqrt{T-t}} \right). \quad (6)
\]

Applying (6) to computing an approximation of \( d_{i+1} \), we find, that

\[
d_{i+1} = \frac{R_{i+1} - \ln E + (r + \sigma^2 / 2)(T-t+1) + O\left( (T-t)^{-1} \right)}{\sqrt{(T-t)}} \quad \cdot \frac{1}{\sqrt{(T-t)}} \sigma,
\]

i.e.

\[
\Delta d_{i+1} = \frac{\Delta R_{i+1} - (r + \sigma^2 / 2) + O\left( (T-t)^{-1} \right)}{\sqrt{(T-t)}} \cdot \frac{1}{\sqrt{(T-t)}} \sigma. \quad (7)
\]
So, the final form of expression (5) is as follows

\[
\Delta Q_t = \Delta R_t - \left( r + \sigma^2 / 2 \right) + O(T-t)^{-1} / \sigma_\sqrt{T-t}. \tag{8}
\]

Substitute (8) to (4) for taking into account the underlying asset log-return dynamics:

\[
\Delta Q_t = \gamma \left( r + \sigma^2 / 2 \right) + O(T-t)^{-1} + \rho \Delta R_{t-1} + \frac{1}{\sigma_\sqrt{T-t}} \epsilon_t + \frac{\delta}{\sigma_\sqrt{T-t}} \epsilon_{t-1}. \tag{9}
\]

We note that expression (9) is the desired regression model with moving average, that takes into account quantile differences of normal p.d.f. computed in regression model with moving average, which allows us to calculate the initial estimates of volatility of price return noise then coefficient before log-differences \( \Delta R_{t-1} \) in (9) is much greater than coefficient before \( \epsilon_{t-1} \). Hence, we consider that informed tradings are found if the absolute value of \( \frac{\rho}{\sigma_\sqrt{T-t}} \) is greater than \( \frac{\delta}{\sigma_\sqrt{T-t}} \) at any time \( t \), but their signs should be opposite, i.e. 1) if \( \rho<0 \), then \( 0<\delta<\rho; \) 2) if \( \rho>0 \), then \( -1<\delta<-\rho \). This is the criterion of informed trading to be formulated.

For empirical high frequency values of stock and option prices we should use some statistical estimates of \( \hat{\rho} \) and \( \hat{\delta} \). Let \( R_t, t = 0, 1, \ldots, (T-1) \), be a data set that is available for analysis. Let \( m << T \) be a time window length, which allows us to calculate the initial estimates of coefficients \( \gamma_t = \frac{1}{m} \sum_{s=0}^{m-1} \ln S_{t-s} - \ln S_t \), \( \rho_t \) and \( \delta_t \) with values \( R_{0s}, R_{1s}, \ldots, R_{ms} \) in model equations (4) and (9). Moving the time window to the right per unit until we reach time \( (T-1) \), by known \( R_{s}, R_{s+1}, \ldots, R_{ms} \) we estimate \( \gamma_s, \rho_s \) and \( \delta_s, s = 0, 1, \ldots (T-m-1) \).

Further, for put option the expressions (7) and (4) hold, so, the last equality can be rewritten as follows:

\[
\Delta Q_t = \gamma \left( r + \sigma^2 / 2 \right) + O(T-t)^{-1} + \rho \Delta R_{t-1} + \frac{1}{\sigma_\sqrt{T-t}} \epsilon_t + \frac{\delta}{\sigma_\sqrt{T-t}} \epsilon_{t-1},
\]

and it looks very similar to (9). And so, of course, Theorems 2 and 3 will also be true for European-style put option with \( X_t = (\Delta Q; \Delta R) \).

Therefore if we denote \( \hat{Q}_t = \Phi^{-1}(\Delta_t + 1) \) and take a reversed to normal p.d.f. function \( \Phi^{-1}(y) \) from all parts of expression above, we get

\[
\Delta d_{t, r} = \Delta \hat{Q}_t.
\]

**CRITERION OF PRESENCE OF INFORMED TRADING ACTIVITY**

If process \( X_t \) is stationary, the expression (9) allows us to formulate criteria and make a detecting system, helping to find informed trader activities on stock and option markets. For instance, if their influence is greater than price return noise then coefficient before log-differences \( \Delta R_{t-1} \) in (9) is much greater than coefficient before \( \epsilon_{t-1} \). Therefore if we denote \( \hat{Q}_t = \Phi^{-1}(\Delta_t + 1) \) and take a reversed to normal p.d.f. function \( \Phi^{-1}(y) \) from all parts of expression above, we get

\[
\Delta d_{t, r} = \Delta \hat{Q}_t.
\]

We consider a European-style put option.

**EUROPEAN STYLE PUT OPTION**

It is well-known that delta for European-style put option is equal to \( \Delta_t = \Phi(\sigma \Delta_t) - 1 \).

**DETECTING SYSTEM**

Let choose non-zero price movements \( \hat{\gamma}_k \neq 0 \), \( k = 0, 1, \ldots, (T-m-1) \) and satisfy the inequality \( |\hat{\gamma}_k| < 1 \) if it
is possible. We assume that the informed transaction is detected if one of the following inequalities holds:

\[
\begin{align*}
a) \sum_{k} \beta_k \sum_{k} \delta_k &< 0, ~ \text{or} \sum_{k} \beta_k > 0, \quad \sum_{k} \beta_k > \sum_{k} \delta_k ; \quad \beta, \delta \in \mathbb{R} \\
\sum_{k} \beta_k \sum_{k} \delta_k &< 0, \quad \sum_{k} \beta_k > 0, \quad \sum_{k} \beta_k > \sum_{k} \delta_k .
\end{align*}
\]

So we construct computer decision system whose power we can check by using data of Taiwan Futures Exchange TAIFEX, freely downloadable and therefore convenient for collecting. TAIFEX also provides delta values for all kinds of derivatives traded. Just for illustrating system abilities we took week calls and puts on index TSEC, the contracts expire at 26/02/14. Strike values were chosen in such a way as to implied volatility index TSEC, the contracts expire at 26/02/14. Strike values for all kinds of derivatives traded. Just for illustration we depicted the system abilities on (IV) possessing the minimum value while a quantity of trade deals reaches a maximum. We state such parameters signs.

**CONCLUSIONS**

We suggested the mathematical procedure and constructed a detecting system for finding informed trader activities because we assumed that the informed transaction is detected if one of the following inequalities holds:

\[
\begin{align*}
a) & \sum_{k} \beta_k \sum_{k} \delta_k < 0, \quad \sum_{k} \beta_k < 0, \quad \sum_{k} \beta_k > \sum_{k} \delta_k ; \\
& \sum_{k} \beta_k \sum_{k} \delta_k < 0, \quad \sum_{k} \beta_k > 0, \quad \sum_{k} \beta_k > \sum_{k} \delta_k .
\end{align*}
\]

We compute auto covariance of some auxiliar expression

\[
\begin{align*}
\Delta R_{t+1} &= \gamma + \rho \Delta R_t + \lambda \beta z_{t+1} + \lambda \beta z_{t-1} + \lambda \mu_t , \\
\Delta R_{t-1} &= \gamma (1 + \rho) + \rho^2 \beta \gamma + \lambda \beta (1 + \rho) z_t + \lambda \beta z_{t-1} .
\end{align*}
\]

We modify (A.2) as

\[
\begin{align*}
\Delta R_{t+1} &= \gamma + \rho \Delta R_t + \lambda \beta z_{t+1} + \lambda \beta z_{t-1} + \lambda \mu_t , \\
\Delta R_{t-1} &= \gamma (1 + \rho) + \rho^2 \beta \gamma + \lambda \beta (1 + \rho) z_t + \lambda \beta z_{t-1} .
\end{align*}
\]

\[
\begin{align*}
\Delta R_t &= \gamma + \rho \Delta R_{t-1} + \lambda \beta z_t + \lambda \beta z_{t-1} + \lambda \mu_t , \\
\Delta R_{t-1} &= \gamma (1 + \rho) + \rho^2 \beta \gamma + \lambda \beta (1 + \rho) z_t + \lambda \beta z_{t-1} .
\end{align*}
\]

**APPENDIX A.**

**Proofs**

**Proof of Theorem 1.** After double substitution of equation (1) into (3) with different \( t \) we obtain:

\[
\begin{align*}
\Delta R_{t+1} &= \gamma + \rho \Delta R_t + \lambda \beta z_{t+1} + \lambda \beta z_{t-1} + \lambda \mu_t , \\
\Delta R_{t-1} &= \gamma (1 + \rho) + \rho^2 \beta \gamma + \lambda \beta (1 + \rho) z_t + \lambda \beta z_{t-1} + \lambda \beta z_{t-1} + \lambda \mu_t .
\end{align*}
\]

We modify (A.2) as

\[
\begin{align*}
\Delta R_{t+1} &= \gamma + \rho \Delta R_t + \lambda \beta z_{t+1} + \lambda \beta z_{t-1} + \lambda \mu_t , \\
\Delta R_{t-1} &= \gamma (1 + \rho) + \rho^2 \beta \gamma + \lambda \beta (1 + \rho) z_t + \lambda \beta z_{t-1} + \lambda \beta z_{t-1} + \lambda \mu_t .
\end{align*}
\]


