

DYNAMICS AND PREDICTABILITY IN RUSSIAN
FINANCIAL MARKETS

Stanislav Anatolyev, Alexander Gerko, Dmitry Shakin, Olga Kryukovskaya

September 25, 2003

Contents

1	The Trading Approach to Testing for Predictability and Market Efficiency	5
1.1	Introduction	5
1.2	Trading approach and trading strategy	6
1.3	Test for conditional mean independence	8
1.4	Nonparametric prediction	11
1.5	Data and testing results	13
1.6	Evolution of market efficiency	14
1.7	Conclusions	16
2	High frequency transactions at an emerging stock market	21
2.1	Introduction	21
2.2	Data, data adjustment, and preliminaries	22
2.2.1	MICEx electronic trading system	22
2.2.2	MICEx database	22
2.2.3	Interday seasonal adjustment	24
2.2.4	Intraday seasonal adjustment	24
2.2.5	Empirical strategy	25
2.2.6	Market microstructure theories	26
2.3	Modeling durations	27
2.3.1	ACD-type models	27
2.3.2	Log-ARMA model	28
2.3.3	Empirical results	29
2.4	Modeling price dynamics	31
2.4.1	UHF-GARCH and ACD-GARCH models	31
2.4.2	Flexible Time Standardization GARCH model	32
2.4.3	Empirical results	33
2.5	Conclusion	34
2.6	Appendix A: MICEx trading regimes	37
2.7	Appendix B: stocks under analysis	37

1. THE TRADING APPROACH TO TESTING FOR PREDICTABILITY AND MARKET EFFICIENCY

1.1 Introduction

The notion of market efficiency, broadly understood as impossibility of obtaining abnormal returns, has been a subject of intensive discussion since the seminal paper by Fama (1970). Numerous theoretical and empirical studies have been devoted to testing the hypothesis of market efficiency. By neglecting a possibility of time-varying risk premia and ignoring all sorts of operational market inefficiencies such as transactions costs and infrequent trading, one reduces the problem to testing the conditional mean independence property of returns (i.e. that the return has constant expectation conditional on past information). For decades, testing for efficiency was performed via regression-based tests, although it is not clear how the significance of coefficients in estimated regressions can be transformed into profits. For numerous other caveats in market efficiency testing, see Timmermann and Granger (2002) for an excellent survey.

In recent years there has been an increasing attention to measuring profitability instead of coefficient significance in the context of testing for market efficiency. Several studies evaluate various technical trading rules for their ability to generate excess returns in “real time” by simulating traders’ decisions. For example, Brock, Lakonishok and LeBaron (1992) find that such simple rules as moving averages and trading-range breaks have predictive power; Pesaran and Timmermann (1995) examine profitability of a portfolio generated by a recursive search for a model specification that maximizes certain financial criteria among available linear forecasting equations; Gençay (1998) evaluates profitability of a simple trading strategy, the one which is at the core of our analysis in this paper, relative to the “buy-and-hold” strategy.

In this paper, we show that the trading approach may be used to construct a formal test for mean predictability and market efficiency. The trading strategy that we use is a rule which issues a buy signal if the next period return forecast is greater than 0 and a sell signal otherwise, where the forecasts are generated by an arbitrary forecasting model. Using this trading strategy, we construct a new test for conditional mean independence of returns in the spirit of Pesaran and Timmermann’s (1992) directional accuracy (DA) test. Our test is based on a suitably normalized profitability of the portfolio implied by the trading strategy over a certain “coin toss” portfolio, which we call a profitability statistic. The asymptotic distribution of this statistic is standard normal. We study power and size properties of the profitability and DA test statistics via simulations. Owing to its orientation towards detecting profitability, the profitability statistic proves to have superior power and acceptable size properties.

Furthermore, motivated by the idea that the very fact that some strategy is able to obtain an abnormal profit during a certain period of time does not imply inefficiency (Pesaran and Timmermann (1995)), we propose computing the profitability statistic from data in a moving time window of fixed length. In this way, we are able to assess the degree of inefficiency in various time periods, and thus track the evolution of market efficiency through time.

Naturally, the better the forecasting model that underlies the trading strategy is, the higher power we can expect from our test. A forecasting model may be taken from a simple class such as the class of linear autoregressive models, or it may be more sophisticated than that and try to exploit nonlinear predictability. Even though it is generally accepted that various financial time series exhibit nonlinear patterns in their behavior (e.g., Abhyankar, Copeland and Wong (1997)), it is still an open question whether it is possible to build an econometric model capable of capturing this nonlinearity *and* yielding good out-of-sample forecasts. Among most popular parametric methods used in applied literature are threshold and Markov switching models (e.g., Aslanidis, Osborn and Sensier (2002), Turner, Startz and Nelson (1989)). Nonparametric methods are less prone to misspecification biases and are able to detect unknown sources of nonlinearity. Very popular are various local regression and nearest neighbors methods used in numerous studies on nonparametric prediction of stock market indices (e.g., Hsieh (1991), Gençay (1998)) and exchange rates (e.g., Diebold and Nason (1990), Mizrach (1992), Satchell and Timmermann (1995)). In this paper, we explore various nonparametric models to construct the forecasts, including a nearest neighbor local regression with our special bandwidth selection rule.

We apply our tests to two stock market indices. One is a half of a century long weekly S&P500 index. The other is a 7.5-year series of the daily RTS (“Russian Trading System”) stock index. The RTS index has been used in several other studies (Hall and Urga (2000), Fedorov and Sarkissian (2000)), and is a most frequently used indicator of the Russian stock market. The papers by Hall and Urga (2000), Rockinger and Urga (2000) and Urga and Rockinger (2001) analyze the Russian and other Eastern European stock markets for informational market efficiency by looking at evolution of a time-varying autoregressive parameter. Our results on the whole agree with conclusions of this alternative approach giving evidence of improving weak form efficiency in Russian financial markets. For the S&P500 index we observe a non-trivial dynamics of the level of weak form market efficiency throughout the last half of the 20th century. In particular, we provide some evidence that it has substantially decreased during the last decade.

The paper is organized as follows. In Section 2 we discuss the statistical and trading approaches to predictability, introduce a simple trading strategy and show how it may be used for testing purposes. In Section 3 we construct a new test for conditional mean independence, and discuss its power and size properties. In Section 4 we digress to outline the nonparametric techniques that lie in the heart of our trading strategy. In Section 5 we describe the data and report and discuss testing results. In Section 6 we apply our test for tracking the evolution of market efficiency through time. Finally, Section 7 concludes.

1.2 Trading approach and trading strategy

Usually, to make predictions of some economic or financial variable, one first constructs a regression model of this variable on its past. The quality of this model is usually judged by in-sample or, more appropriately, out-of-sample fit summarized in certain statistics. In the latter case these statistics are usually mean squared or absolute prediction error, or other predictability measures such as the directional accuracy statistic of Pesaran and Timmermann (1992). This “statistical” approach to forecasting model building actually aims at minimizing some unrelated to profitability loss functional rather than at getting a significant outcome from the viewpoint

of profit maximization, the ultimate goal of making predictions in the context of forecasting returns in a financial market.

Hellström (1998) argues that the “trading” and “statistical” forecasting problems should be separated and different quality measures should be used depending on what the objective of model construction is. A more formal approach is found in Leitch and Tanner (1991) who ask the following question: “Why do profit maximizing firms buy professional forecasts when statistics such as MSPE or MAPE often indicate that a naive model will forecast about as well?” They analyze the performance of several linear forecasting models and professional services for the three-month Treasury bill rate. It appears that although the “naive” forecast is best in terms of conventional forecast quality measures, it is inferior to professional service forecasts in terms of profit obtained via various trading strategies based on corresponding forecasts. By splitting the dataset into several periods they also find that the correlation between the MSPE and profits is positive in most cases. Ramaswamy (1998) also offers examples where the model with smaller MSPE is less profitable, and uses the total return over an investment period to formulate an optimization problem and to compute optimal weights for the Nadaraya–Watson estimator.

We conclude that it is likely that the formal inference on predictability and efficiency, if based on profitability measures rather than on standard forecasting quality statistics, may turn out to be more successful in identifying predictable patterns. Let the variable y_t represent the return on some financial asset or index. Let \hat{y}_t be a forecast of y_t provided by a meaningful parametric or nonparametric regression of y_t on elements of its past $\{y_{t-1}, y_{t-2}, \dots\}$, or, more generally, on elements of the extended information set $\mathcal{I}_{t-1} \supset \{y_{t-1}, y_{t-2}, \dots\}$ which may include other historical variables. The forecast \hat{y}_t is allowed to depend only on the data at or before time $t - 1$.¹ Consider the following trading rule based on \hat{y}_t :

$$\begin{cases} \text{buy} & \text{if } \hat{y}_t > 0, \\ \text{sell} & \text{otherwise.} \end{cases} \quad (1.2.1)$$

That is, a trader goes long if the prediction of the next period return is positive, and goes short otherwise. For brevity, we will call this rule *the trading strategy*. Equipped with the trading strategy, a trader forms a portfolio and modifies it each trading period. The one period return of this portfolio is

$$r_t = \text{sign}(\hat{y}_t)y_t. \quad (1.2.2)$$

The profitability of the trading strategy (1.2.1) was evaluated by Gençay (1998) to measure whether forecasts have economic value in practice. Using two and a half decades of the DJIA data Gençay (1998) finds that such trading strategy is able to provide perceptible profits relative to the “buy-and-hold” strategy. We will use this trading approach to construct a formal test of conditional mean independence of returns.

Let us first illustrate how the series of returns r_t can be useful for testing purposes. Suppose we want to test the null hypothesis that the return y_t is a martingale difference sequence (MDS) relative to \mathcal{I}_{t-1} , i.e. $E[y_t|\mathcal{I}_{t-1}] = 0$. Under this null, the return r_t is an MDS, too. Indeed,

$$E[r_t|\mathcal{I}_{t-1}] = E[\text{sign}(\hat{y}_t)y_t|\mathcal{I}_{t-1}] = E[\text{sign}(\hat{y}_t) E[y_t|\mathcal{I}_{t-1}]] = 0.$$

Moreover, if we assume that the distribution of \hat{y}_t is continuous, under the null we have

$$\text{var}[r_t] = E[\text{sign}(\hat{y}_t)^2 y_t^2] = E[y_t^2] = \text{var}[y_t].$$

¹For example, estimates of unknown parameters in a parametric predictor should not utilize future observations.

Then under the null the appropriately normalized cumulative return of the corresponding portfolio is asymptotically distributed as standard normal:

$$\frac{\sum_t r_t}{\sqrt{\sum_t y_t^2}} \xrightarrow{d} N(0, 1). \quad (1.2.3)$$

The same null could be tested by using a similar statistic based on the normalized cumulative original return y_t . However, the test (1.2.3) is expected to be more powerful. Indeed, if under the alternative the underlying forecasting model has high predictive power, the absolute value of the test statistic (1.2.3) should be much greater than that of $\sum_t y_t / \sqrt{\sum_t y_t^2}$.

1.3 Test for conditional mean independence

In this section we introduce a test for the null hypothesis of conditional mean independence, i.e. that the return series has constant expectation conditional on the information set \mathcal{I}_{t-1} . This test may be viewed as a generalization of our illustration given in the previous section. Under the assumption of a constant risk premium and no operational market imperfections, such tests are usually referred to as tests of “informational market efficiency”. We elaborate the link to market efficiency more in Section 6.

The test is based on the out-of-sample profitability of the trading strategy (1.2.1). Formally, we are interested in testing the null hypothesis $H_0 : E[y_t | \mathcal{I}_{t-1}] = \text{const}$, where $\mathcal{I}_{t-1} \supset \{y_{t-1}, y_{t-2}, \dots\}$ is some information set relative to which the forecast \hat{y}_t is measurable. Technically, we require that under the null a stronger property hold, that \hat{y}_t be independent from y_t for all lags and leads. See the discussion of this technical issue at the end of this section.

The expected one period return of the trading strategy (1.2.1) is $E[r_t]$, which is consistently estimable under the null by the following two estimators:

$$A_T = \frac{1}{T} \sum_t r_t \quad (1.3.4)$$

and

$$B_T = \left(\frac{1}{T} \sum_t \text{sign}(\hat{y}_t) \right) \left(\frac{1}{T} \sum_t y_t \right). \quad (1.3.5)$$

Indeed, under the null, $A_T \xrightarrow{P} E[r_t]$ and

$$\begin{aligned} B_T \xrightarrow{P} E[\text{sign}(\hat{y}_t)] E[y_t] &= E[\text{sign}(\hat{y}_t) E[y_t]] \stackrel{H_0}{=} E[\text{sign}(\hat{y}_t) E[y_t | \mathcal{I}_{t-1}]] \\ &= E[\text{sign}(\hat{y}_t) y_t] = E[r_t]. \end{aligned}$$

Note that A_T represents the average one period return of the portfolio based on the trading strategy, and B_T can be interpreted as the average return of the portfolio based on the “coin toss” strategy with the same proportion of “buy” and “hold” signals as the trading strategy (1.2.1). We base our test statistic on the difference between A_T and B_T . Under the alternative, if the underlying forecasting model has predictive power, the trading strategy will generate a higher cumulative return than the “coin toss” strategy, and the test will have power. The power of the test will be greater the better the underlying forecasting model is.

To complete the construction of the test, it remains to compute the asymptotic variance of $A_T - B_T$ under the null. Let $p_{\hat{y}} = \Pr\{\text{sign}(\hat{y}_t) = 1\}$, then $E[\text{sign}(\hat{y}_t)] = 2p_{\hat{y}} - 1$ and $\text{var}[\text{sign}(\hat{y}_t)] = 4p_{\hat{y}}(1 - p_{\hat{y}})$. Recall that under H_0 the series \hat{y}_t and y_t are independent for all lags and leads. Then we have

$$\begin{aligned}\text{var}(A_T) &= \frac{1}{T} E[\text{sign}(\hat{y}_t)^2 y_t^2] - \frac{1}{T} E[\text{sign}(\hat{y}_t) y_t]^2 \stackrel{H_0}{=} \frac{1}{T} E[y_t^2] - \frac{1}{T} (2p_{\hat{y}} - 1)^2 E[y_t]^2, \\ \text{var}(B_T) &= E\left[\left(\frac{1}{T} \sum_t \text{sign}(\hat{y}_t)\right)^2 \left(\frac{1}{T} \sum_t y_t\right)^2\right] - E\left[\left(\frac{1}{T} \sum_t \text{sign}(\hat{y}_t)\right) \left(\frac{1}{T} \sum_t y_t\right)\right]^2 \\ &\stackrel{H_0}{=} E\left[\left(\frac{1}{T} \sum_t \text{sign}(\hat{y}_t)\right)^2\right] E\left[\left(\frac{1}{T} \sum_t y_t\right)^2\right] - E\left[\frac{1}{T} \sum_t \text{sign}(\hat{y}_t)\right]^2 E\left[\frac{1}{T} \sum_t y_t\right]^2 \\ &= \frac{1}{T} (2p_{\hat{y}} - 1)^2 \text{var}(y_t) + \frac{4}{T} p_{\hat{y}} (1 - p_{\hat{y}}) \left(\frac{1}{T} E[y_t^2] + \left(1 - \frac{1}{T}\right) E[y_t]^2\right).\end{aligned}$$

$$\begin{aligned}\text{cov}(A_T, B_T) &= E\left[\left(\frac{1}{T} \sum_t \text{sign}(\hat{y}_t) y_t\right) \left(\frac{1}{T} \sum_t \text{sign}(\hat{y}_t)\right) \left(\frac{1}{T} \sum_t y_t\right)\right] \\ &\quad - E\left[\left(\frac{1}{T} \sum_t \text{sign}(\hat{y}_t) y_t\right)\right] E\left[\left(\frac{1}{T} \sum_t \text{sign}(\hat{y}_t)\right) \left(\frac{1}{T} \sum_t y_t\right)\right] \\ &\stackrel{H_0}{=} \frac{1}{T^2} E[\text{sign}(\hat{y}_t)^2 y_t^2] + \frac{T-1}{T^2} E[\text{sign}(\hat{y}_t)^2 y_t] E[y_t] \\ &\quad + \frac{T-1}{T^2} E[\text{sign}(\hat{y}_t) y_t^2] E[\text{sign}(\hat{y}_t)] + \frac{(T-1)^2}{T^2} E[\text{sign}(\hat{y}_t)]^2 E[y_t]^2 \\ &\quad - E[\text{sign}(\hat{y}_t)]^2 E[y_t]^2 \\ &= \frac{1}{T} (2p_{\hat{y}} - 1)^2 \text{var}(y_t) + \frac{4}{T} p_{\hat{y}} (1 - p_{\hat{y}}) \left(\frac{1}{T} E[y_t^2] + \left(1 - \frac{1}{T}\right) E[y_t]^2\right) \\ &= \text{var}(B_T).\end{aligned}$$

Taking things altogether², we obtain

$$\text{var}(A_T - B_T) = 4 \frac{T-1}{T^2} p_{\hat{y}} (1 - p_{\hat{y}}) \text{var}(y_t).$$

We can estimate this variance in a variety of ways. The most straightforward is

$$\hat{V} = \frac{4}{T^2} \hat{p}_{\hat{y}} (1 - \hat{p}_{\hat{y}}) \sum_t (y_t - \bar{y})^2,$$

where we corrected for degrees of freedom when estimating the variance of y_t , and

$$\hat{p}_{\hat{y}} = \frac{1}{2} \left(1 + \frac{1}{T} \sum_t \text{sign}(\hat{y}_t)\right)$$

²That the test is of Hausman type can be seen from the following argument: the estimator B_T is semiparametrically efficient for $E[\text{sign}(\hat{y}_t)] E[y_t]$ as it is a product of independent semiparametrically efficient estimators of $E[\text{sign}(\hat{y}_t)]$ and $E[y_t]$. However, this argument is based on *asymptotic* considerations, while in the text we showed explicitly that $\text{cov}(A_T, B_T)$ equals $\text{var}(B_T)$ *in finite samples*.

is a consistent estimator of $p_{\hat{y}}$. The estimate \hat{V} is positive by construction unless by chance all forecasts are of the same sign. The test statistic and its asymptotic distribution are

$$\frac{A_T - B_T}{\sqrt{\hat{V}}} \xrightarrow{d} N(0, 1). \quad (1.3.6)$$

We will refer to this as *the profitability statistic*.

The profitability statistic (1.3.6) is reminiscent of the directional accuracy (DA) test of Pesaran and Timmermann (1992) that is routinely used as a predictive-failure test in constructing forecasting models. Let us have a quick look at the construction of the DA statistic. Under the same null hypothesis as before, the success ratio defined as

$$\tilde{A}_T = \frac{1}{T} \sum_{t=1}^T \mathbf{I}[y_t \hat{y}_t \geq 0],$$

cannot differ much from the expected success ratio that would obtain in case y_t and \hat{y}_t were independent. An efficient estimate of the latter, say, \tilde{B}_T , can be expressed in terms of shares of positive values among y_t and \hat{y}_t in the sample. The variances of \tilde{A}_T and \tilde{B}_T can be straightforwardly estimated, so that the Hausman-type DA test statistic, an appropriately normalized difference between \tilde{A}_T and \tilde{B}_T , is asymptotically distributed as standard normal.

Note that, apart from certain renormalization, a typical summand in \tilde{A}_T equals

$$\text{sign}(\hat{y}_t) \text{sign}(y_t),$$

while that in A_T equals

$$\text{sign}(\hat{y}_t) \text{sign}(y_t) |y_t|.$$

This implies that if the series of returns y_t is predictable, our test takes fuller advantage of this predictability, while the DA test exploits it prudently and ignores the fact that higher returns are associated with better forecasts. Further note that the DA test will successfully detect deviations from the null of no predictability when y_t exhibits sign predictability, while the profitability statistic (1.3.6) will when y_t exhibits conditional mean dependence. Christoffersen and Diebold (2002) show that the sign predictability may be merely a result of volatility dependence in the absence of conditional mean dependence, and that sign dependence does not imply violation of market efficiency. Thus, our profitability statistic is better suited for predictability and efficiency considerations.

We investigate the power of both tests in a small Monte–Carlo experiment. We use two data generating processes (DGPs) that exhibit different types of predictability and whose coefficients are calibrated using the S&P500 index during the period from 1954 to 1973 when the index seems to be most mean predictable. The DGPs are:

$$\text{AR: } y_t = 0.1256 \cdot y_{t-1} + \varepsilon_t, \sigma^2 = 0.000249;$$

$$\text{SETAR: } y_t = \begin{cases} 0.000844 + 0.2453 \cdot y_{t-1}, & |y_{t-1}| \leq 0.01848 \\ 0.002679 + 0.0664 \cdot y_{t-1}, & |y_{t-1}| > 0.01848 \end{cases} + \varepsilon_t, \sigma^2 = 0.000245.$$

The sample length is 1,000. In each of 10,000 simulations, estimation is performed over a rolling sample with 100 observations. The DA and profitability statistics are computed over 900 predictions. The two top panels of Table 1 show actual rejection frequencies corresponding to nominal rejection frequencies of 10%, 5%, and 1% for two- and right one-sided alternatives. We use the least squares linear predictor (labeled OLS) and nearest neighbors local regression (labeled NNA), the latter being detailed in Section 4. The power of our test appreciably exceeds

that of the DA test, sometimes by a factor greater than 1.5. It can be also seen how much the power of either test may vary with the choice of a forecasting model. When the DGP is truly linear, fitting a linear model has a much greater predictive power than nonparametric methods, which, in turn, increase power when the series is nonlinearly predictable.

To conclude the discussion of properties of the tests, it should be noted that in derivations of the asymptotic distribution of the DA³ and our profitability statistics, it is presumed that under H_0 there is independence of y_t and \hat{y}_t at all lags and leads. Under the null of interest when we posit only the contemporaneous independence between y_t and \hat{y}_t , the total independence is questionable. Indeed, even under no predictability it is highly unlikely that the forecast is completely unrelated to the variable being forecast. For example, when a lagged value y_{t-1} is used to make a prediction of y_t , this prediction is by construction not independent from y_t at lag one. Thus, in practical situations when the total independence does not hold both tests may have wrong sizes. To assess size distortions, we perform a small Monte–Carlo experiment with the design outlined above, and the following DGP:

$$\text{GARCH: } y_t = 0.002483 + \varepsilon_t, \sigma_t^2 = 0.0000223 + 0.1773 \cdot \varepsilon_{t-1}^2 + 0.7397 \cdot \sigma_{t-1}^2.$$

Thus, we deliberately invoke strong GARCH effects to introduce volatility predictability and further exacerbate the dependence between returns and their forecasts. The two bottom panels of Table 1 show actual sizes corresponding to nominal sizes of 10%, 5%, and 1%. In spite of possible distortions, the actual sizes of both tests are very close to nominal.

1.4 Nonparametric prediction

In this section we digress to outline the nonparametric predictors we will use in the empirical analysis. We consider the nonparametric time series regression

$$y_t = m(y_{t-1}, y_{t-2}, \dots, y_{t-k}) + \varepsilon_t, \quad (1.4.7)$$

with $E[\varepsilon_t | \mathcal{I}_{t-1}] = 0$. Let us denote $z_t = (y_{t-1} y_{t-2} \dots y_{t-k})'$ and $z_{t,i} = y_{t-i}$. The aim is to construct a forecast $\hat{y}(z)$ of y_t given the history $z = (z_1 z_2 \dots z_k)'$. One class of predictors uses the Nadaraya–Watson nonparametric kernel estimators:

$$\hat{y}(z) = \frac{\sum_{s=k+1}^{t-1} y_s \mathbf{K}(z - z_s, h(z))}{\sum_{s=k+1}^{t-1} \mathbf{K}(z - z_s, h(z))},$$

another uses locally weighted linear estimators

$$\hat{y}(z) = \hat{\alpha} + (z - z_s)' \hat{\beta},$$

where

$$(\hat{\alpha}, \hat{\beta}) = \arg \min_{\alpha, \beta} \sum_{s=k+1}^{t-1} (y_s - \alpha - (z - z_s)' \beta)^2 \mathbf{K}(z - z_s, h(z)).$$

³See the manipulations below equation (4) in Pesaran and Timmermann (1992, page 462).

In these formulas, $\mathbf{K}(\cdot, \cdot)$ is kernel function and $h(z)$ is (possibly non-scalar) bandwidth. Both estimators are special cases of local polynomial kernel estimators. For more on nonparametric kernel regression, see Härdle (1990) and Härdle and Linton (1994). Below, we discuss options for the kernel function, the use of nearest neighbors, and bandwidth selection rules, paying close attention to the specifics of prediction of financial data.

Although there is a theory of optimal kernels, the choice of the shape of the one-dimensional kernel function is not most important, because the loss of efficiency from the use of non-optimal kernels is negligible in practice (Härdle (1990, section 4.5)). In the empirical literature the most commonly used kernels are uniform, triangle, tricube and standard normal. For example, in the context of time series prediction Diebold and Nason (1990) and Meese and Rose (1990, 1991) use the tricube kernel, Mizrach (1992) and Barkoulas and Baum (1996) employ the triangle kernel.

In cases when $k > 1$ a multi-dimensional kernel should be constructed from one-dimensional ones. There are two main approaches. In the first approach, one forms the product kernel:

$$\mathbf{K}(z - z_s, h(z)) = \prod_{i=1}^k K\left(\frac{z_i - z_{s,i}}{h_i(z)}\right). \quad (1.4.8)$$

In the second approach, one transforms a vector argument into its scalar norm:

$$\mathbf{K}(z - z_s, h(z)) = K\left(\left\|\left(\frac{z_1 - z_{s,1}}{h_1(z)}, \frac{z_2 - z_{s,2}}{h_2(z)}, \dots, \frac{z_k - z_{s,k}}{h_k(z)}\right)\right\|\right), \quad (1.4.9)$$

where $\|\cdot\|$ is some norm on \mathbb{R}^k , and $K(\cdot)$ is one-dimensional kernel function. All empirical papers mentioned in the previous paragraph use the Euclidean norm in (1.4.9).

More important is the choice of the way the function $h(z)$ is constructed. In dealing with financial data we have a high degree of non-uniformity of the regressor space. This implies that adjusting the bandwidth at the boundary of the regressor space becomes critical, and suggests that various global bandwidth selection rules, such as cross-validation, are not likely to perform well. The situation becomes exacerbated when we deal with the multi-dimensional locally weighted linear regression as we have to increase the bandwidth at the boundary of the regressor space for each of the regressors.

A family of nearest neighbors algorithms offers a neat solution to this problem. Unlike various global bandwidth selection methods they allow for the automatic bandwidth adjustment at boundaries of the space of regressors. The idea is that a set of n nearest neighbors of z in the sample is used to determine the bandwidth. More precisely, a set of neighbors of the point z is identified, and every component of $h(z)$ is set to be the distance from z to its farthest neighbor or some other measure of the set of neighbors.

Nearest neighbor methods seem to be most popular in the literature. Cleveland and Loader (1996b) note that they provide the control over the number of points smoothed over, which is convenient from the practical point of view. Cleveland and Loader (1996a) find that estimation with a global bandwidth turns out to perform worse in practice than nearest neighbors methods. In empirical work, Diebold and Nason (1990) and Meese and Rose (1990, 1991) use the distance to the farthest neighbor as the bandwidth, while the number of neighbors is selected manually; Mizrach (1992) and Barkoulas and Baum (1996) – the sum of distances to all neighbors as the bandwidth, while the number of neighbors is also selected manually.

We propose and exploit in our empirical investigation a bandwidth selection rule in the framework of the nearest neighbors procedure, which tends to work better in the case of predicting stock indices. The idea of this rule is based on the following observation. Suppose that one of the components of z , say, z_i , is relatively close to the boundary of the corresponding one-dimensional regressor space. It is natural to presume that the corresponding component of the

optimal bandwidth, $h_i(z)$, should be greater than in the case when z_i is in the interior of the regressor space. Let $\mathbb{N}_n(z)$ be the set of n nearest neighbors of the point z in the L^∞ metric. Then set

$$h_i(z) = \sup_{\zeta \in \mathbb{N}_k(z)} (|z_i - \zeta_i|). \quad (1.4.10)$$

1.5 Data and testing results

We apply our analytical results to two stock market indices. One is the daily RTS (“Russian Trading System”) index for the period from September 01, 1995 to April 18, 2003, totaling to 1909 observations. This index is a most frequently used indicator of the Russian stock market, and has been used in several studies (Hall and Urga (2000), Fedorov and Sarkissian (2000)). As of the beginning of the year 2003 it was comprised of dollar prices of 59 shares of 35 leading Russian companies, weighted according to their capitalization. Detailed information on the index’s construction and the data are available from www.rts.ru. The second series we analyze is the weekly S&P500 stock market index for the period from January 03, 1950 to May 05, 2003, totaling to 2783 observations, which we obtained from finance.yahoo.com. We use it to see, on the one hand, how sensitive our results are to the choice of a dataset, and on the other hand, how our methods work with less frequently sampled data on a longer time scale. Before the analysis we take first differences of logarithms of the indices to obtain series of returns.

We employ various nonparametric predictors, and use as benchmarks the “naive” random walk (RW) forecasts implied by the “buy-and-hold” strategy, and forecasts provided by a linear model estimated by OLS. After some experimentation, we chose the number of lags k in the regression to be 5 for the daily RTS index, and 1 for the weekly S&P500 index. The motivation for these numbers is that in the former case we expect to find a lot of predictability at the day-to-day frequency, while in the latter we can hardly expect any dependence for the lags beyond the first. We assess the quality of one step ahead forecasts using rolling regression with blocks of constant length. The block length equals 300 for the RTS data (which corresponds to a little more than one year), and 104 (which corresponds to two years) for the S&P500 data. This scheme results in 1603 ($= 1909 - 1 - 5 - 300$) and 2677 ($= 2783 - 1 - 1 - 104$) predictions, respectively.

We use the standard normal kernel function as its unbounded support allows one to set a smaller global bandwidth, and the product kernel in the multivariate case. We show the results for nonparametric predictors implied by the Nadaraya–Watson (NW) and locally weighted linear regression (LWR) with a global scalar bandwidth, and the nearest neighbors local regression (NNA) augmented by our bandwidth selection rule (1.4.10). The global bandwidth for the NW and LWR estimators is chosen manually as cross-validation produces unstable results which though did not differ much from the eventually chosen bandwidth. After some limited experimentation we selected a couple of bandwidth values $h = 0.03$ and $h = 0.1$ noting that excessive playing with smoothing parameters may result in data-snooping biases (cf., for example, Lo and MacKinlay (1990)). Similarly, for the NNA estimator the number of neighbors selected by the cross-validation tends to be greater than optimal, hence it is manually set to $n = 10$ and $n = 100$.

The testing results are reported in Table 2. As follows from its column 2, no model succeeded in improving upon RW forecasts in terms of mean squared prediction error (MSPE), which is in line with empirical literature. However, the results of the DA test of Pesaran and Timmer-

mann (1992) in column 3 show strong predictability even in the case of the linear model. The results become even more impressive, at least for the RTS index, when we take into account the profitability of the trading strategy based on corresponding predictors. Column 4 reports the values of the profitability statistic which are highly significant in most cases. It is clear that the conditional mean independence hypothesis both for the daily RTS index and the weekly S&P500 index is strongly rejected.

An interesting thing to note is that while the MSPE tends to decrease when a smoothing parameter (h or n) increases, the other forecast quality measures behave in the opposite way. Taking a smaller number of neighbors (or smaller h) weights more heavily similar histories in the data and allows one to capture subtler forms of dependence. In this case outliers can strongly influence the quality of forecasts as measured by the MSPE. However, the DA and profitability statistics are driven by prediction errors only to the extent to which they may change the sign of a prediction. The following explanation for the apparent impossibility of a significant improvement of the MSPE can be suggested. Even if the data does contain predictable nonlinearity, it is clear that the stochastic unpredictable component dominates the data (cf. Abhyankar, Copeland and Wong (1997)), and its variance dominates in the MSPE. In contrast, a properly normalized out-of-sample profitability is free from this drawback, and hence may be suggested as the right measure of forecast quality in the case of financial data.

Column 5 reports cumulative returns that a trader would obtain if equipped with the trading strategy and corresponding forecasting models. Although these figures are quite impressive, especially for the NNA model, they should be taken with a grain of salt, especially in case of RTS data, since no transaction costs and market restrictions are taken into account. Illuminating patterns unfold on Figures 1 and 2 which show the graphs of the log cumulative returns (for $h = 0.03$ and $n = 10$) of portfolios based on our trading strategy and various forecasting models. For the RTS index the cumulative returns rank according to our perception of relative powers of forecasting methods, obtained from Table 1. The best performing one is the nearest neighbors local regression with the bandwidth (1.4.10), then come the locally weighted linear regression with a global bandwidth, the Nadaraya–Watson estimator with a global bandwidth, and, finally, the linear least squares model. The only difference for the S&P500 index is that the Nadaraya–Watson estimator performed worst of all linear and locally linear estimators. This can be explained by the fact that it is a locally constant estimator, thus, if a linear model is a better proxy for the true data generating process, this estimator has a substantial bias at boundaries of the regressor space, which is the most critical region for profitability. The apparent similarity between the graphs of cumulative returns should also be noted. The periods where the gains from the trading strategy are most evident are the same for all predictors. The robustness to a choice of a forecasting model gives us grounds to interpret different trading possibilities as evidence of time-varying level of market efficiency.

1.6 Evolution of market efficiency

As noted before, under the assumption of a constant risk premia and no operational market imperfections, tests of the hypothesis $H_0 : E[y_t | \mathcal{I}_{t-1}] = \text{const}$ are usually referred to as tests of “informational market efficiency”. If by \mathcal{I}_{t-1} we mean information embedded in the historical values of y_t , we obtain a test of the “weak form” market efficiency, which can be interpreted as the impossibility of profitable trading on the basis of only historical price information. If

we add to \mathcal{I}_{t-1} other variables publicly observable at moment $t - 1$, we obtain a test of the “semi-strong form”, the impossibility of obtaining abnormal profits on the basis of the publicly available information.

Our results of the previous section can be used to answer the question of whether the Russian and American stock markets were “weak form” efficient during the periods under consideration. The answer is firm “no”. However, it is more interesting to have an answer to the question: in which periods of time were these markets inefficient? Indeed, the recent literature points out that the period of inefficiency may alternate with those of efficiency, and that the degree of predictability is different in different periods (Timmermann and Granger (2002)). If we base our profitability statistic on predictors estimated from the data in a moving time window of fixed length, we can assess the degree of inefficiency during the period of this window, and thus track the evolution of market efficiency through time.

The length of the time window should satisfy the following two conditions: it should not be too small in order not to include too much noise, and it should not be too large in order not to smooth out interesting details. For the daily RTS data we find the compromise at the block length equal to 250 (which corresponds to approximately 1 year); for the weekly S&P500 data – to 520 (which corresponds to approximately 10 years). As the nearest neighbors local regression with the bandwidth (1.4.10) showed to advantage before, we report the results with this predictor.

The tracks of the profitability statistics for the RTS and S&P500 indices, along with the upper 5% critical value line, are depicted on Figures 3 and 4, respectively. Recall that the market inefficiency at a point in time is unobservable, so the values of the profitability statistic over the designated period of 1 year (10 years) should be interpreted as the measure of market inefficiency over the whole period. We are interested in periods when a track exceeds the bound of the critical line, as well as global minima and maxima that may be interpreted as an indication of structural breaks in the evolution of market efficiency.

For the RTS data our results are not completely in line with those of Hall and Urga (2000) who claim a substantial increase in the market efficiency of the Russian stock market starting from the first quarter of the year 1999. The first period for which the hypothesis of conditional mean independence is not rejected at the 5% level by our test is 08.1999 – 07.2000. This is more in line with the results of Rockinger and Urga (2000) who notice that the predictability of the Russian market remains high up until the end of the year 1999. From the beginning of 2001 there is no evident trend in the data and the hypothesis of conditional mean independence is not rejected.

For the S&P500 data one can clearly observe two points when the structural breaks in the market efficiency may have occurred. The first corresponds to the global maximum of the profitability statistic over the whole sample, which occurs for the period from 1963 to 1973. Starting from the end of this period and up to the year 1992 one can see an apparent trend towards improving market efficiency. In fact, for the period from 1982 to 1992 the value of the profitability statistic is close to zero. However, after the year 1992 this trend is reversed, and for the period from 1993 to 2003 the hypothesis of conditional mean independence can be only marginally accepted. If this 11-year trend continues, we are likely to see the hypothesis of weak form informational market efficiency rejected for the S&P500 index in the nearest future.

1.7 Conclusions

We have proposed a forecast quality measure which, in case of a series of returns, has a clear interpretation as a properly normalized return of a certain trading strategy that uses as an input forecasts generated by an arbitrary forecasting model. Based on this measure, we have developed a new formal test for conditional mean independence based on a suitably normalized profitability of the portfolio implied by the trading strategy, over a certain “coin toss” portfolio. To make forecasts for the trading strategy, we use rolling regressions and nearest neighbor methods with a new rule of bandwidth selection.

In our empirical study we have applied the constructed test to examine the efficiency of Russian and American stock markets and track the evolution of the degree of efficiency in time. We have found that although for the daily RTS and weekly S&P500 indices the hypothesis of conditional mean independence is rejected, there is remarkable dynamics in the degree of market efficiency throughout the observation period.

References

1. Abhyankar, A., L.S. Copeland, and W. Wong (1997) Uncovering nonlinear structure in real-time stock-market indexes: The S&P 500, the DAX, the Nikkei 225, and the FTSE-100. *Journal of Business & Economic Statistics* 15, 1–14.
2. Aslanidis, N., D.R. Osborn, and M. Sensier (2002) Smooth transition regression models in UK stock returns. SES/CGBCR, University of Manchester.
3. Barkoulas, J. and C.F. Baum (1996) Essential nonparametric prediction of U.S. interest rates. Working paper #313, Boston College.
4. Brock, W.A., J. Lakonishok, and B. LeBaron (1992) Simple technical trading rules and the stochastic properties of stock returns. *Journal of Finance* 47, 1731–1764.
5. Christoffersen, P.F. and F.X. Diebold (2002) Financial asset returns, market timing, and volatility dynamics. Manuscript, University of Pennsylvania.
6. Cleveland, W.S. and C.L. Loader (1996a) Smoothing by local regression: Principles and methods. In: W. Härdle and M.G. Schimek (eds.), *Statistical Theory and Computational Aspects of Smoothing*, 10–49. Springer, New York.
7. Cleveland, W.S. and C.L. Loader (1996b) Rejoinder to discussion of “Smoothing by local regression: Principles and methods”. In: W. Härdle and M.G. Schimek (eds.), *Statistical Theory and Computational Aspects of Smoothing*, 113–120. Springer, New York.
8. Diebold, F.X. and J.A. Nason (1990) Nonparametric exchange rate prediction. *Journal of International Economics* 28, 315–332.
9. Fama E.F. (1970) Efficient capital markets: A review of theory and empirical work. *Journal of Finance* 25, 383–423.

10. Fedorov, P. and S. Sarkissian (2000) Cross-sectional variations in the degree of global integration: the case of Russian equities. *Journal of International Financial Markets, Institutions and Money* 10, 131–150.
11. Gençay, R. (1998) Optimization of technical trading strategies and the profitability in security markets. *Economics Letters* 59, 249–254.
12. Gençay, R. (1999) Linear, non-linear and essential foreign exchange rate prediction with simple technical trading rules. *Journal of International Economics* 47, 91–107.
13. Hall, S. and G. Urga (2000) Testing for time-varying stock market efficiency using Russian stock prices. Manuscript, IRMI, CUBS.
14. Härdle W. (1990) *Applied Nonparametric Regression*. Cambridge University Press, New York.
15. Härdle, W. and O. Linton (1994) Applied Nonparametric Methods. In: R.F. Engle and D.L. McFadden (eds.), *Handbook of Econometrics*, Vol. 4. Elsevier Science, North Holland.
16. Hellström, T. (1998) A random walk through the stock market. Licentiate thesis, Umeå University.
17. Hsieh, D. (1991) Chaos and Nonlinear Dynamics: Application to Financial Markets. *Journal of Finance* 46, 1839–1877.
18. Leitch, G. and J.E. Tanner (1991) Economic forecast evaluation: profits versus the conventional error measures. *American Economic Review* 81, 580–590.
19. Lo, A.W. and A.C. MacKinlay (1990) Data-snooping biases in tests of financial assets pricing models. *Review of Financial Studies* 3, 431–468.
20. Meese, R.A. and A.K. Rose (1990) Nonlinear, nonparametric, nonessential exchange rate estimation. *American Economic Review* 80, 192–196.
21. Meese, R.A. and A.K. Rose (1991) An empirical assessment of non-linearity in models of exchange rate determination. *Review of Economic Studies* 58, 603–619.
22. Mizraç B. (1992) Multivariate nearest-neighbor forecasts of EMS exchange rates. *Journal Of Applied Econometrics* 7, 151–163.
23. Pesaran, M.H. and A. Timmermann (1992) A simple nonparametric test of predictive performance. *Journal of Business and Economic Statistics* 10, 561–565.
24. Pesaran, M.H. and A. Timmermann (1995) Predictability of stock returns: Robustness and economic Significance. *Journal of Finance* 50, 1201–1228.
25. Ramaswamy, S. (1998) One-step prediction of financial time series models. Working paper #57, Bank for International Settlements, Switzerland.
26. Rockinger, M. and G. Urga (2000) Evolution of stock markets in Central and Eastern Europe. *Journal of Comparative Economics*, 28, 456–472.
27. Satchell, S. and A. Timmermann (1995) An assessment of the economic value of non-linear foreign exchange rate forecasts, *Journal of Forecasting* 14, 477–497.

28. Timmermann, A.G. and C. Granger (2002) Efficient market hypothesis and forecasting. Discussion paper #3593, CEPR.
29. Turner, C.M., Startz, R., and Nelson, C.R. (1989) A Markov model of heteroskedasticity, risk, and learning in the stock market. *Journal of Financial Economics* 25, 3–22.
30. Urga, G. and M. Rockinger (2001) A time varying parameter model to test for predictability and integration in stock markets of transition economies. *Journal of Business and Economic Statistics* 19, 73–84.

Table 1												
Model	DA statistic						Profitability statistic					
	10%		5%		1%		10%		5%		1%	
	OLS	NNA	OLS	NNA	OLS	NNA	OLS	NNA	OLS	NNA	OLS	NNA
Power, testing against two-sided alternatives												
AR	40.7	25.0	30.1	16.3	13.5	5.8	51.9	32.5	40.9	22.6	21.5	9.15
SETAR	39.3	39.8	29.3	29.6	14.3	13.9	50.2	52.1	40.1	41.1	22.6	22.1
Power, testing against right one-sided alternatives												
AR	53.3	36.6	40.2	24.2	18.9	8.8	64.0	44.2	51.7	32.0	28.7	13.6
SETAR	51.2	53.2	38.8	39.6	19.4	19.1	61.5	63.9	49.8	51.9	29.0	29.3
Size, testing against two-sided alternatives												
GARCH	9.3	9.9	4.5	4.9	0.7	1.0	14.1	11.3	7.9	6.1	2.5	1.2
Size, testing against right one-sided alternatives												
GARCH	9.9	9.9	4.6	4.8	0.9	1.0	8.7	9.2	4.4	4.7	1.0	1.0

Notes:

Entries in the table are actual rejection frequencies from 10,000 simulations of series with sample size 1,000, corresponding to nominal rejection frequencies of 10%, 5%, and 1%.

For power properties, the following DGPs are used:

$$\text{AR: } y_t = 0.1256 \cdot y_{t-1} + \varepsilon_t, \sigma^2 = 0.000249;$$

$$\text{SETAR: } y_t = \begin{cases} 0.000844 + 0.2453 \cdot y_{t-1}, & |y_{t-1}| \leq 0.01848 \\ 0.002679 + 0.0664 \cdot y_{t-1}, & |y_{t-1}| > 0.01848 \end{cases} + \varepsilon_t, \sigma^2 = 0.000245.$$

For size properties, the following DGP is used:

$$\text{GARCH: } y_t = 0.002483 + \varepsilon_t, \sigma_t^2 = 0.0000223 + 0.1773 \cdot \varepsilon_{t-1}^2 + 0.7397 \cdot \sigma_{t-1}^2.$$

OLS refers to least squares on a linear projection, NNA – to the nearest neighbors local regression using a bandwidth selected according to the rule (1.4.10). One lag of y_t is used in the regression function.

The profitability statistic is defined in (1.3.6). The DA and profitability statistics are asymptotically distributed as $N(0,1)$ under the null of no predictability.

In each simulation, estimation is performed over a rolling sample with 100 observations. The DA and profitability statistics are computed over 900 predictions.

Table 2				
Prediction method	Sum of squared prediction errors	Pesaran–Timmermann DA statistic	Profitability statistic	Cumulative return
Forecasting performance: daily RTS index				
RW	1.62			2.10
OLS	1.65	3.36**	4.56**	345.89
NW, $h = 0.1$	1.61	1.71*	1.33*	6.35
NW, $h = 0.03$	1.69	3.05**	4.66**	362.22
LWR, $h = 0.1$	1.67	4.03**	5.24**	811.80
LWR, $h = 0.03$	2.43	3.67**	5.81**	1680.21
NNA, $n = 100$	1.63	4.31**	5.69**	1462.77
NNA, $n = 10$	1.70	4.15**	5.98**	2090.73
Forecasting performance: weekly S&P500 index				
RW	1.07			38.97
OLS	1.10	2.52**	2.48**	50.94
NW, $h = 0.1$	1.08	-0.05	0.10	10.80
NW, $h = 0.03$	1.09	0.18	0.13	9.82
LWR, $h = 0.1$	1.10	2.92**	2.73**	64.56
LWR, $h = 0.03$	1.16	2.31*	2.97**	77.63
NNA, $n = 100$	1.10	2.63**	2.66**	60.40
NNA, $n = 10$	1.14	4.24**	4.03**	138.50

Notes:

OLS refers to least squares on a linear projection, NW – to the Nadaraya–Watson estimator, LWR – to the locally weighted linear regression using a global bandwidth, NNA – to the nearest neighbors local regression using a bandwidth selected according to the rule (1.4.10). Five lags of y_t are used in the regression function for the RTS index, one lag – for the S&P500 index. Estimation is performed over a rolling sample with 300 (104) observations for the RTS (S&P500) index. The DA and profitability statistics are computed over 1603 (2677) predictions.

The cumulative return is defined as $\exp(\sum_t r_t)$ except in the case of RW when it equals $\exp(\sum_t y_t)$. DA is the directional accuracy statistic of Pesaran and Timmermann (1992).

The profitability statistic is defined in (1.3.6). The DA and profitability statistics are asymptotically distributed as $N(0,1)$ under the null of no predictability.

* (**) indicates that the corresponding statistic is significant at 5% (1%) level.

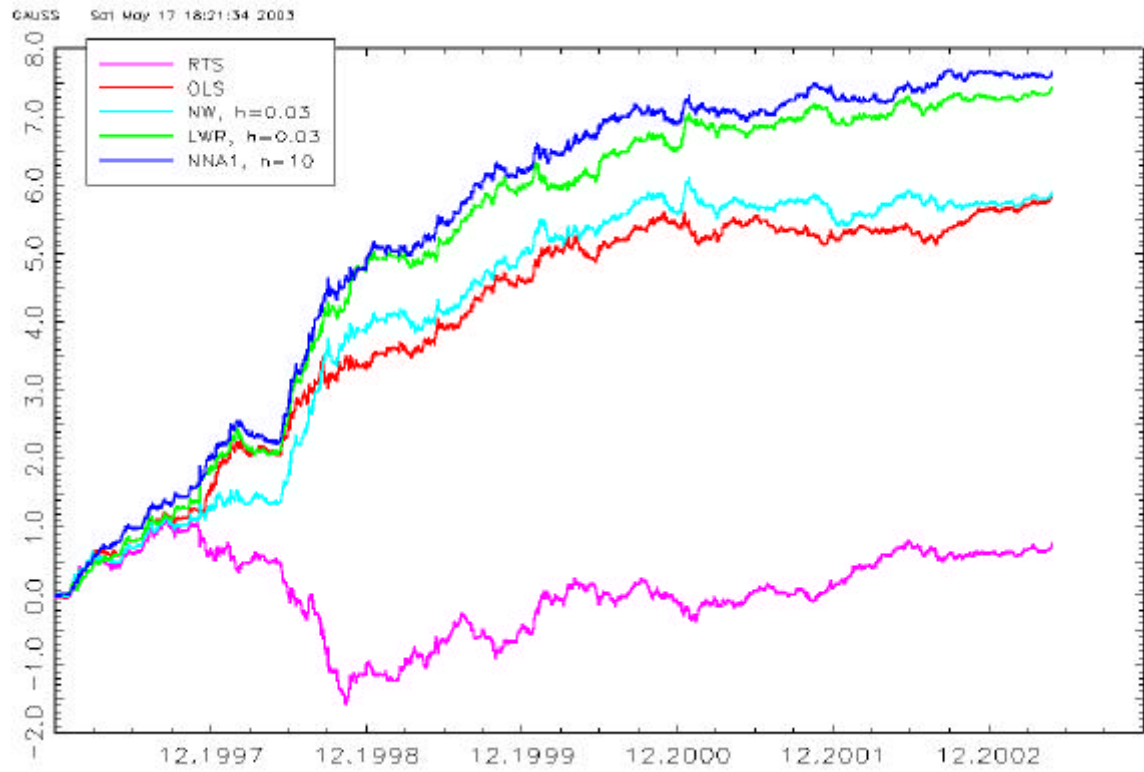


Figure 1. Cumulative returns, daily RTS index

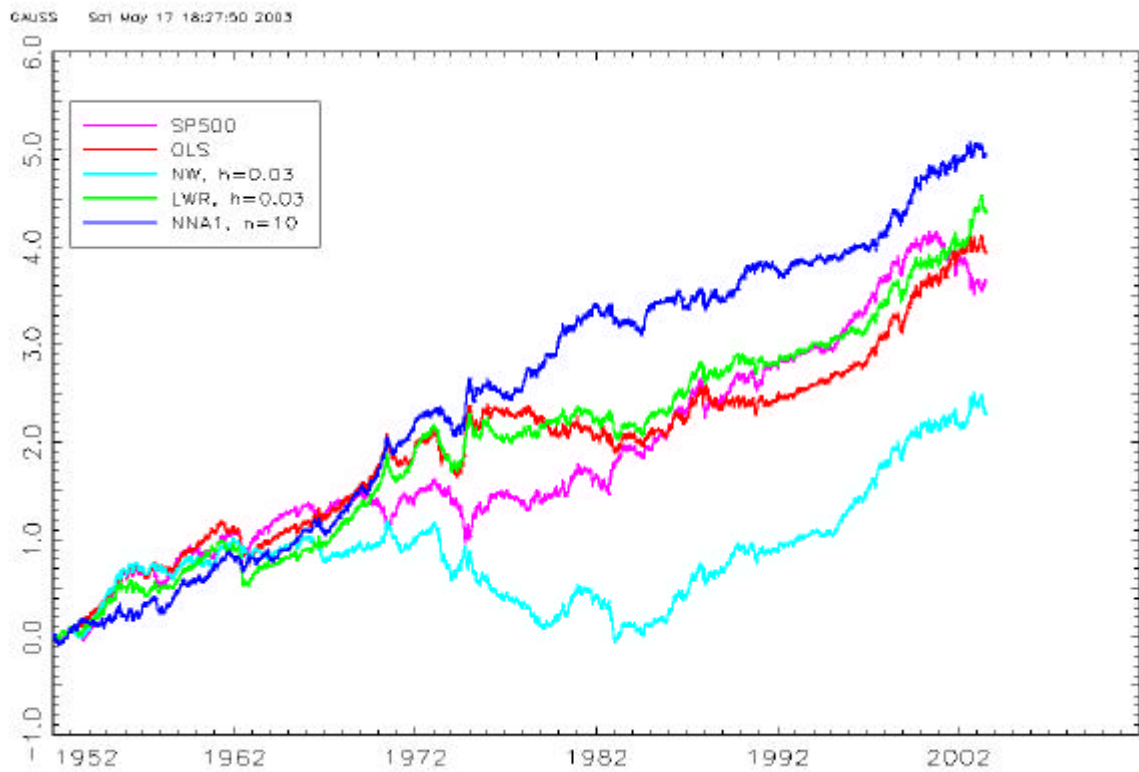


Figure 2. Cumulative returns, weekly S&P500 index

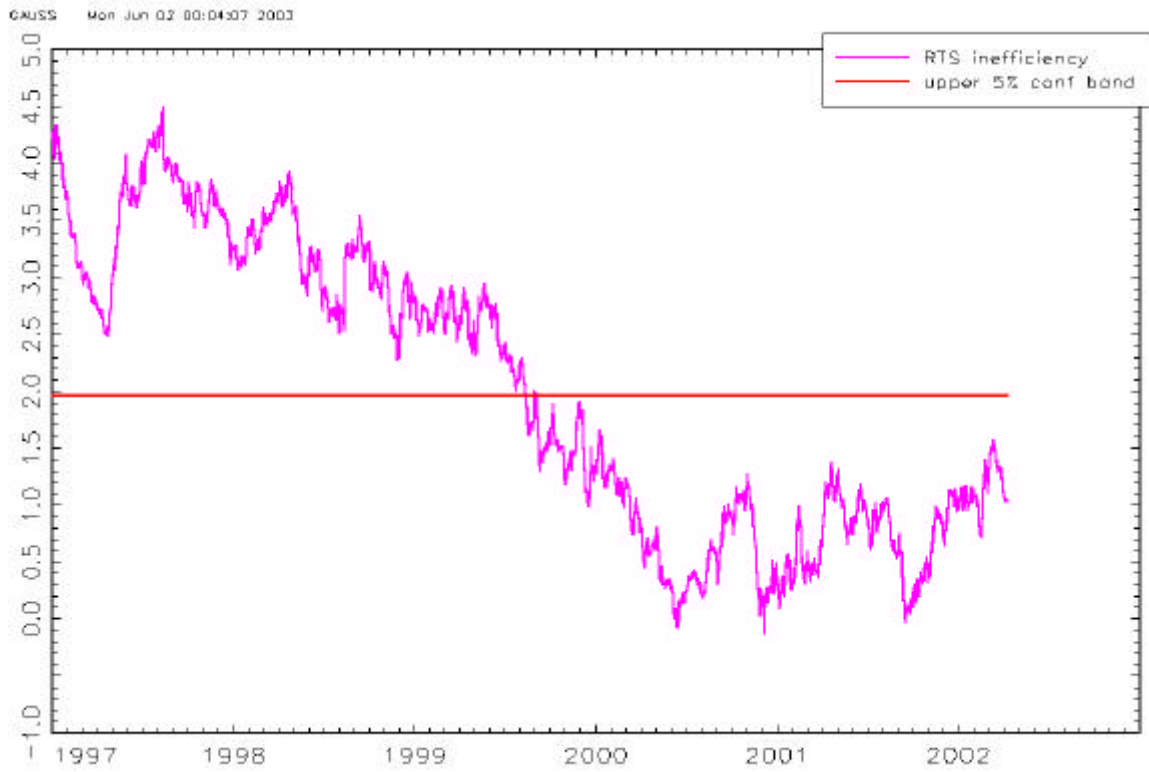


Figure 3. Evolution of market efficiency, daily RTS index



Figure 4. Evolution of market efficiency, weekly S&P500 index

2. HIGH FREQUENCY TRANSACTIONS AT AN EMERGING STOCK MARKET

2.1 Introduction

Modeling high frequency transaction data has been an intensive topic of research during recent decades. The seminal papers of Engle and Russell (1998) and Engle (2000) generated a burst of new interest to modeling the dynamics of duration and price processes and gaining understanding of market microstructure. In this paper, we contribute to both theoretical and applied sides of this literature by analyzing tick-by-tick data from one of emerging stock markets.

We use the data on trades in seven most frequently traded common stocks at the Moscow Interbank Currency Exchange (MICEX). The total value of trades in these stocks takes around 97% of the overall value of trades. Thus, in contrast to other studies which analyze one stock or a tiny sample from a large population of stocks, we essentially investigate the whole population. In the existing literature, most frequent is the analysis of IBM stocks at the NYSE (Engle and Russell 1998, Ghysels and Jasiak 1998, Jasiak 1999, Engle 2000, McCulloh and Tsay 2000, Hafner 1999, Zhang, Russell and Tsay 2001, Tsay 2002), different studies often using the same database. Some exploit data for other stocks at the NYSE (Engle and Lunde 1998, Dufour and Engle 1999, Bauwens and Giot 2000). Exploration of other stock exchanges is much more rare and is not left unnoticed. Several papers (e.g., Jasiak 1999 and Ghysels, Gouriéroux and Jasiak 2002) use data from the Paris Bourse, Grammig and Wellner (2002) – from the German Security Exchange and its electronic trading system IBIS, Tyurin (2002) – from the Reuters D2000-2 foreign exchange electronic brokerage system. To our knowledge, there does not exist any study that analyzes high frequency data from an emerging stock market in a transitional country.

There do exist a number of papers devoted to the microstructure of Russian financial markets. Medvedev and Kolodyazhny (2001) investigate the behavior of non-residents in the state bond market, and find evidence of herding behavior. Barinov, Pervozvansky and Pervozvanskaya (1999) study the behavior of traders at the Russian bond market and test some implications of the optimal portfolio theory. They also find that market microstructure theories work well for Russian financial markets. Goldberg and Tenorio (1997) apply the auction theory to the investigation of the Russian market for foreign exchange. None of these and other researchers though makes an attempt to investigate the microstructure of the Russian stock markets from the viewpoint of high frequency data.

In this paper, we use our own modifications of the classic autoregressive conditional duration (ACD) model of Engle and Russell (1998) and ultra high frequency (UHF) GARCH model of Engle (2000). We arrive at more flexible specifications that fit the data under study better, as indicated by the Ljung–Box and Nyblom tests. Our Log–ARMA–GARCH specification for durations accounts better for some of intrinsic features of their dynamics than the ACD and closely connected Log–ACD models, and at the same time is computationally simpler than the stochastic volatility duration (SVD) model of Ghysels, Gouriéroux and Jasiak (2002). We model the process for returns by a new FTS–ARMA–GARCH specification which in turn is

a generalization of the UHF-GARCH model and admits endogenous determination of time standardization for measuring volatility of high frequency returns.

An inspection of the data reveals that along with intraday deterministic patterns there are also interday variations in durations and return volatilities. We remove both sorts of nonstationarity from the data prior to the analysis. Subsequent estimation of both Log-ARMA-GARCH and FTS-ARMA-GARCH models on the adjusted data show that the behavior of Russian traders is quite similar to that of traders at developed financial markets. We find that the Russian stock market data support some of the implications of Admati and Pfleiderer (1988) and Easley and O'Hara (1992) models. At the same time, no evidence is found in favor of the Diamond and Verrecchia (1987) model which assumes constraints on short selling.

The plan of the paper is following. The description of our database to some detail is given in Section 2 and Appendices A and B. Section 2 also describes the data adjustment, characterizes our empirical strategy, and reviews market microstructure theories we intend to test. Section 3 is devoted to the development of a new model for evolution of intertrade durations, together with its empirical implementation, Section 4 – to the analysis of high frequency returns. Section 5 concludes.

2.2 Data, data adjustment, and preliminaries

2.2.1 MICEx electronic trading system

The electronic system at the MICEx is organized as an order-driven market. An excellent explanation of an order-driven system can be found in Tyurin (2002); here we provide only a brief description. At each moment of time the system keeps track of all buy and sell offers for each particular security in a so called *order book*; these offers, in particular, include prices and amounts. As soon as there is a buy offer with a price which is at least as great as the price of some sell offer for the same security, the bargain is immediately executed, and these two offers in the order book are modified according to the bargain's amount. However, along with this "normal" trading regime with continuous matching of orders, there are special regimes of trades in the same security described in Appendix A. Undoubtedly, the trades in these regimes are observed by traders and their results may be taken into account when making decisions. However, these data are very noisy (especially for the "negotiated deals" regime when trades are not executed immediately at the time they are recorded, but may be delayed for days or weeks). We remove such trades from the sample and concentrate on trades in the "normal" trading regime which opens at 10:30AM and closes at 6:45PM.

2.2.2 MICEx database

For the analysis we take the period from August 12, 2002 to October 27, 2002, composed of 50 trading days. We consider common stocks of Lukoil, Yukos, SurgutNefteGaz, Rostelecom, Sberbank, MosEnergo, and UES. These are seven most frequently traded stocks during the period under consideration. For each of the seven corporations we identify the most heavily traded release of common stocks and ignore all other stock releases. The total value of trades

in these seven stocks takes nearly 97% of the overall value of trades. A description of the stocks under study can be found in Appendix B. In what follows, we will report ultimate estimation results for all stocks, and intermediate and illustrative results – for Lukoil only.

Each record in our database corresponds to one trade transaction for which the following three characteristics are available:

1. Time T_t (with the precision of one second) when the transaction occurred;
2. Price P_t at which the transaction occurred;
3. Number of shares V_t bought or sold.

The database does not contain bid or ask data. Using the available characteristics we generate the following three variables of primary interest:

1. Duration between consecutive trades: $\hat{d}_t = T_t - T_{t-1}$;
2. Log return: $\hat{r}_t = \ln P_t - \ln P_{t-1}$;
3. Log volume: $\hat{v}_t = \ln V_t$.

It is worth noting that the price step for all stocks is tiny in contrast to that at other exchanges such as the NYSE. This can be clearly seen from Figure 1 which shows the histogram of nonzero price changes for Lukoil; about one third of transactions having no price change are excluded (cf., for example, Russell and Engle 1998, Figure 1). Even then, the discreteness of price movements, together with the associated microstructure questions such as the analysis of possible front-running behavior on the part of small traders, is an additional interesting issue that deserves separate investigation (e.g., along the lines of Russell and Engle 1998, Rydberg and Shephard 1998, Darolles, Gouriéroux and Le Fol 2000).

When generating the variables we pay special attention to simultaneous transactions. Such transactions arrive from two sources. First, simultaneous transactions are recorded when a new buy (sell) offer matches several sell (buy) offers and is in sufficient amount to execute all these matching offers. It is natural to treat such trades as one big transaction. The characteristic feature of simultaneous transactions of this type is that their prices are either non-increasing or non-decreasing. The second possibility is that a duration between consecutive trades is so small that the precision of one second is not sufficient for their discrimination and they are marked with the same time stamp. Since we cannot distinguish between the two sources, we use the following mnemonic procedure. If there are several transactions recorded at the same time stamp, we aggregate transactions with non-increasing or non-decreasing prices, using for these aggregated transactions a weighted average price and aggregate volume, and remove all remaining transactions with this time stamp.¹

Table 1, upper panel, shows descriptive statistics of transaction variables for Lukoil. Some specific characteristics of high-frequency data can be noted. First, both durations and returns exhibit very strong autocorrelation, which is reported by the Ljung–Box statistics with fifteen lags $Q(15)$ whose 1% critical value is 30.58. Another feature of the duration process is overdispersion (when the standard deviation is higher than the mean), which is often observed in high-frequency duration data. Transaction returns have a much higher kurtosis than that implied by the normal distribution.

¹This deletion is relatively innocuous because the number of such situations is small for all stocks (except very heavily traded UES).

2.2.3 Interday seasonal adjustment

Most of previous literature studying high frequency transactions data does not reveal significant interday variation in variables under investigation. In contrast, we find that in different days the transaction activity and price volatility at the MICEx fluctuate noticeably. In particular, running a Wald test for equality of all coefficients in a regression of average durations on five day-of-the-week dummies using the Newey–West variance estimator yields rejections at the 1% significance level for all stocks except Sberbank. The interday variations may be easily detected in Figures 2 and 3, where the plots of average durations and return volatilities throughout trading days are depicted. For the volume variable \hat{v}_t , the effect of interday variations is also present, although not as clearly as for durations and returns.

Several explanations can be suggested for these interday movements. First, there are day-of-the-week deterministic patterns: on average the durations are larger on Mondays and Fridays than on Tuesdays, Wednesdays and Thursdays. Second, Russian stocks are strongly dependent on external factors. For example, the prices of oil extractor stocks (Lukoil, Yukos, Surgut-NefteGaz) are closely tied to world oil prices, hence the volatility of their returns may vary significantly from one day to another. Finally, different intensities of trades may be caused simply by the highly varying number of players in the market.

One way to deal with the interday variations is to create a dummy variable for each day, akin to fixed time effects in the panel data analysis, or to apply the model to each day separately. Another approach was taken in the hierarchical model of McCulloch and Tsay (2001) where the parameters are modeled as random variables which are drawn from some distribution and take new values every day, akin to random time effects in the panel data analysis. We instead remove interday variations from the data from the outset in the following way:

$$\begin{aligned}\tilde{d}_t &= \frac{\hat{d}_t}{\bar{\hat{d}}_s}, \\ \tilde{r}_t &= \frac{\hat{r}_t}{\sqrt{\bar{\hat{r}}_s^2}}, \\ \tilde{v}_t &= \hat{v}_t - \bar{\hat{v}}_s,\end{aligned}$$

where $\bar{\hat{d}}_s$ is the average duration for day s if observation t belongs to day s , $\bar{\hat{r}}_s^2$ is the average squared return for day s , and $\bar{\hat{v}}_s$ is the average log volume for day s . For durations and returns we use a multiplicative form of adjustment as durations and squared returns are naturally positive. However, for log volumes we use additive adjustment as this variable may take any sign; this adjustment implies the multiplicative form for volumes in levels as well.

2.2.4 Intraday seasonal adjustment

Beside interday variations in the data, there are specific intraday patterns of deterministic nature in the duration, return, and volume series. It is conventional in the literature to remove such patterns prior to estimation by using various nonparametric methods. Most of papers (e.g., Ghysels and Jasiak 1998, Engle and Lunde 1998, Engle 2000, Grammig and Wellner 2002) use cubic or piecewise linear splines, while some authors estimate regressions based on kernels methods (e.g. Hafner 1999, Tsay 2002, Zhang, Russell and Tsay 2001). Yet more rarely

they include diurnal dummy variables (as in Dufour and Engle 1999), or remove the seasonal component by a linear regression on time (as in Ghysels, Gouriéroux and Jasiak 2002) or by simple averaging over a moving window and linear interpolation (as in Bauwens and Giot 2000).

Here we estimate the equations

$$\begin{aligned} E[\tilde{d}_t|T_{t-1}] &= f_1(T_{t-1}), \\ E[\tilde{r}_t^2|T_{t-1}] &= f_2(T_{t-1}), \\ E[\tilde{v}_t|T_{t-1}] &= f_3(T_{t-1}), \end{aligned}$$

by using the local linear regression with the gaussian kernel and manually selected bandwidth. Moreover, we do not assume that the intraday patterns are the same across different days of the week, and do estimation for each of them separately. This is also different from previous studies some of which allow but do not detect weakly seasonal components, e.g., Ghysels and Jasiak (1998), Bauwens and Giot (2000), although Grammig and Wellner (2002) also fit separate splines for different weekdays and a few other special dates. The estimates of intraday patterns for durations and squared returns of the Lukoil stocks are presented in Figures 4 and 5. The pictures are consistent with previous studies: the durations are longer in the mid-day and shorter in the morning and in the evening; the return volatility is high in the morning and pretty stable during the rest of the day. Note that there is a noticeable day-of-the-week effect, especially Monday and Friday patterns differ from those of the other days of the week. It is interesting that there is no increase in volatility in evenings, which can be explained by the fact that although at 6:45PM the “normal” trading regime is over, there is a possibility to trade in other regimes (see Appendix A).

We use the obtained estimates to generate final adjusted versions of our variables with the forms of adjustment from the previous subsection:

$$\begin{aligned} d_t &= \frac{\tilde{d}_t}{\hat{f}_1(T_{t-1})}, \\ r_t &= \frac{\tilde{r}_t}{\sqrt{\hat{f}_2(T_{t-1})}}, \\ v_t &= \tilde{v}_t - \hat{f}_3(T_{t-1}). \end{aligned}$$

Sample statistics for the deseasonalized variables are shown in Table 1, lower panel. Again, all characteristic features of high frequency data, such as the overdispersion in durations, excess kurtosis in returns and high autocorrelation in all series, are present, although to a lesser degree compared to the original data (see Table 1, upper panel).

2.2.5 Empirical strategy

In the next two sections we will present new models for durations and returns. These models are motivated, on the one hand, by econometric considerations, and on the other hand, by the degree of fit they provide to our data. We follow the prevailing tendency in the high frequency literature that the process for durations is modeled conditionally on the history, and the process for returns is modeled conditionally on the history and current durations (e.g., Engle 2000). Gouriéroux and Jasiak (2001, p. 404) argue that different sets of non-causality restrictions may also be interesting (e.g., Darolles, Gouriéroux and Le Fol 2000).

For making judgments about the quality of fit, we use the Ljung–Box tests for autocorrelation of order 15 of (standardized) residuals and their squares, and the Nyblom test (Nyblom 1989) for structural stability of single coefficients as well as of a whole submodel. The use of the Ljung–Box test is pretty standard in empirical implementation of high frequency data models, despite the literature points out the inappropriateness of usual asymptotic critical values (Li and Mak 1994, Tse 2002). It is sometimes reported that parameters in high-frequency data models are highly unstable. For example, Zhang, Russell and Tsay (2001) used structural break tests of Andrews and Ploberger (1994) and found six structural breaks in their three-month sample. The Nyblom test that we use as a portmanteau check for structural stability has been popularized by Bruce Hansen in a number of articles (e.g., Hansen 1994), and is natural to use when the estimation framework is the method of maximum likelihood. It is based on the average normalized inner square of cumulative sums of estimated score vectors, its critical values depending on the number of coefficients in the model (see Nyblom 1989 and Hansen 1994 for more details).

It is well known that portmanteau test statistics often reject models with a handful of parameters estimated with a large number of observations. The high frequency literature often considers as a success a significant reduction in values of these statistics relative to those for raw data. We are not trying to improve upon the existing models by relaxing the degree of parsimony, so we treat diagnostic testing results liberally and display reasonable tolerance to rejections by portmanteau statistics. It is worth saying, however, that values of the Ljung–Box statistic in our models are never as high as, for example, in Engle (2000), and our minimal P -values are much greater than typical P -values in, for example, Hafner (1999) and Bauwens and Giot (2000).

2.2.6 Market microstructure theories

In this subsection we briefly review some theories of market microstructure whose implications we want to test using our data. More extended discussions of these theories are contained in Engle and Russell (1998, section 7), Engle (2000, section 6A), and the three source papers: Diamond and Verrecchia (1987), Easley and O’Hara (1992), and Admati and Pfleiderer (1988).

We concentrate on three models that describe the behavior of traders on the basis of asymmetric information. All models assume that there are two types of traders: informed and uninformed. The first model is Diamond and Verrecchia (1987) where constraints on short selling imply that informed traders who receive bad news and want to short sell have no such possibility. Consequently, if long intertrade intervals are observed, market participants are inclined to think that there are bad news. Thus, if durations are long, the returns decline and volatility increases because new information is revealed by the fact of slow trading. The second model, Easley and O’Hara (1992), has completely different implications. Here there are no short selling constraints, hence long durations are associated with the absence of new information. In this case the volatility is lower for longer durations. In the third model, Admati and Pfleiderer (1988), short durations are a result of bunching of liquidity traders, who trade for the reasons that are not directly connected with profits (e.g, liquidity, portfolio diversification, etc.). Thus, since short durations accompanied with large volume are not informative, the volatility declines. Also, this model predicts that liquidity traders tend to trade together, hence trade events are clustered and this clustering should strengthen when volumes are high.

2.3 Modeling durations

2.3.1 ACD-type models

The class of autoregressive conditional duration (ACD) models was proposed by Engle and Russell (1998). Their idea was to capture the persistence in durations, which is often found in high frequency data, by means of an autoregressive model for conditional means. Let d_t be the duration between consecutive trades occurred at times T_{t-1} and T_t , which is assumed to be strictly greater than zero. This duration is decomposed in the following way:

$$d_t = E[d_t|I_{t-1}]\varepsilon_t \equiv \pi_t\varepsilon_t, \quad (2.3.1)$$

where I_{t-1} is the information set embedding all previous durations and possibly other variables, and ε_t is a shock independent of variables in I_{t-1} , with a mean of unity. Engle and Russell (1998) use the following functional form of evolution for the conditional expectation π_t :

$$\pi_t = \omega + \sum_{i=1}^p \phi_i d_{t-i} + \sum_{i=1}^q \psi_i \pi_{t-i}, \quad (2.3.2)$$

Such specification is called ACD(p, q). To complete this parametric model one specifies the distribution for ε_t , which can be done in a variety of ways. The most popular choices are exponential, Weibull, and generalized gamma distributions². Given the specified density, one can proceed with the maximum likelihood (ML) estimation. When the above distributions do not fit the data well, one may treat the ACD model as semiparametric, and run quasi-maximum likelihood (QML) estimation presuming the exponential density for ε_t , with robust computation of standard errors. The ACD model also allows a researcher to investigate the influence of external factors by introducing additional regressors in equation (2.3.2).

An arguably more suitable parameterization for dynamics of π_t may be the Log-ACD(p, q) model (Bauwens and Giot 2000) where

$$\pi_t = \exp \left(\omega + \sum_{i=1}^p \phi_i \ln d_{t-i} + \sum_{i=1}^q \psi_i \ln \pi_{t-i} \right) \quad (2.3.3)$$

or

$$\pi_t = \exp \left(\omega + \sum_{i=1}^p \phi_i \varepsilon_{t-i} + \sum_{i=1}^q \psi_i \ln \pi_{t-i} \right). \quad (2.3.4)$$

As Bauwens and Giot (2000) argue, these specifications have an advantage that the conditional mean π_t is always positive, so there is no need to impose non-negativity constraints neither on the coefficients ϕ_i and ψ_i , nor on the coefficients belonging to additional structural variables on the right side of (2.3.3) or (2.3.4). In addition, the authors find that the specification (2.3.4) fits their data better than the ACD specification (2.3.2).

The ACD model may be extended in many other ways using the analogy with the GARCH class. For instance, it is often reported that the sum of autoregressive coefficients in Equation (2.3.2) is very close to unity, which may imply the presence of long memory. This observation leads to the fractionally integrated ACD (FIACD) model proposed by Jasiak (1999).

²There exist more exotic specifications, e.g. the Burr distribution used by Grammig and Maurer (2000).

Introducing nonlinearities leads to, for example, the threshold ACD (TACD) model of Zhang, Russell and Tsay (2001).

In spite of the ability of ACD models to capture high persistence in the duration process well, the major shortcoming³ of models of this type is that the conditional mean and variance are tied to each other:

$$V[d_t|I_{t-1}] = \pi_t^2 V[\varepsilon_t],$$

where $V[\varepsilon_t]$ is a constant depending on the error distribution. To separate out the persistence in the mean and that in variance, Ghysels, Gouriéroux and Jasiak (2002) propose a stochastic volatility duration (SVD) model. They argue that their model is quite successful in capturing these two different sources of persistence. However, the likelihood function has a complicated form due to the presence of latent factors, and thus simulation-based techniques have to be applied for estimation and inference.

In the next subsection we propose a model that combines the simplicity of ACD models and the flexibility of SVD models using an ARMA specification for durations in logs. Such model can be easily estimated with the use of standard econometric packages.

2.3.2 Log-ARMA model

Consider the following ARMA(p, q) representation for log durations:

$$\ln d_t = \alpha + \sum_{i=1}^p \beta_i \ln d_{t-i} + \sum_{i=1}^q \gamma_i \zeta_{t-i} + \zeta_t, \quad (2.3.5)$$

where $\zeta_t = \sigma_t \eta_t$ follows GARCH(r, s):

$$\eta_t \sim IID(0, 1), \quad (2.3.6)$$

$$\sigma_t^2 = \mu + \sum_{i=1}^r \lambda_i \zeta_{t-i}^2 + \sum_{i=1}^s \nu_i \sigma_{t-i}^2. \quad (2.3.7)$$

In the model (2.3.5)–(2.3.7), $\alpha, \beta_i, \gamma_i, \mu, \lambda_i$ and ν_i are unknown parameters. We call this specification the *Log-ARMA*(p, q)–*GARCH*(r, s) model, or *Log-ARMA* for brevity. As before, we have two options to complete the model. One is to specify the distribution of η_t , natural choices for which are the normal and possibly Student t distributions, and proceed with ML estimation. The second option is to view the model as semiparametric and run QML using the normal density.

The Log-ARMA specification (2.3.5)–(2.3.7) has been overlooked in the previous literature. We explain this by the fact that it hides the nature of d_t as a duration variable. At the same time, this specification is natural to consider, it is flexible, and its properties are familiar. In addition, the Log-ARMA model fits our data much better than the ACD and Log-ACD models, as we will see in the next subsection, while the degree of parsimony is comparable. The main advantage of the Log-ARMA model lies in its greater flexibility which originates from the separation of the persistence in conditional mean and that in conditional variance. Thus, we may think of this specification as a simple, in terms of complicatedness of estimation, alternative to the SVD model, and a practitioner may easily use standard statistical software to estimate the parameters in (2.3.5)–(2.3.7).

³See the discussion on this issue in Gouriéroux and Jasiak (2000, p. 462).

Let us see how the Log-ARMA model compares with the Log-ACD(p, q) model (2.3.3). From equations (2.3.3) and (2.3.1),

$$\begin{aligned}\ln d_t &= \omega + \sum_{i=1}^p \phi_i \ln d_{t-i} + \sum_{i=1}^q \psi_i \ln \pi_{t-i} + \ln \varepsilon_t \\ &= \omega + \left(1 - \sum_{i=1}^q \psi_i\right) m + \sum_{i=1}^{\max(p,q)} (\phi_i + \psi_i) \ln d_{t-i} + \zeta_t - \sum_{i=1}^q \psi_i \zeta_{t-i},\end{aligned}$$

where $m = E[\ln \varepsilon_t]$ and $\zeta_t = \ln \varepsilon_t - m$, so that $\ln d_t$ follows an ARMA($\max(p, q), q$) process. If ε_t is distributed lognormally⁴ with parameters⁵ $(-\sigma^2/2, \sigma^2)$, then ζ_t is distributed normally with parameters $(0, \sigma^2)$. Thus, the model under consideration is equivalent to the conditionally normal homoscedastic Log-ARMA model. Other distributional specifications for ε_t would imply conditionally non-normal homoscedastic Log-ARMA models. A conditionally heteroskedastic specification (2.3.7) relaxes the serial independence of ζ_t and thus that of ε_t . If we start from the Log-ACD(p, q) model (2.3.4), however, the nesting of models breaks as the old innovations ζ_{t-i} then enter the ARMA equation nonlinearly.

Another interesting feature of the Log-ARMA model is the type of the implied conditional hazard function for the duration process. For a conditionally normal homoscedastic Log-ARMA model, the conditional distribution of durations is lognormal and thus the hazard function is non-monotonic, while the exponential, Weibull and generalized gamma ACD models imply a monotone hazard. Moreover, in a heteroscedastic Log-ARMA model, the hazard function is time-varying.

In addition, the time-varying conditional variance in the Log-ARMA model can take account of the dynamics of overdispersion in transaction data. Gouriéroux and Jasiak (2000, p. 462) argue that the overdispersion in data is time varying. In the ACD model the conditional overdispersion is constant (recall that $V[d_t|I_{t-1}]/\pi_t^2 = V[\varepsilon_t]$), while in the Log-ARMA it is path dependent. Moreover, the coefficients in the variance equation (2.3.7) provide an idea about the clustering of observations with over- or underdispersion.

2.3.3 Empirical results

To begin with we estimate⁶ the semiparametric ACD(2, 2) model

$$\begin{aligned}d_t &= \pi_t \varepsilon_t, \\ \pi_t &= \omega + \phi_1 d_{t-1} + \phi_2 d_{t-2} + \psi_1 \pi_{t-1} + \psi_2 \pi_{t-2} + \delta v_{t-1}.\end{aligned}$$

Here the lagged volume variable v_{t-1} is introduced to capture the effect of large trades. For the Lukoil stocks, the duration process shows high persistence, with the sum of ϕ - and ψ -coefficients greater than 0.95. The value of the Ljung-Box test statistic for the standardized

⁴Such a choice may be viewed as a limiting case of the generalized gamma distribution.

⁵Recall that if x is distributed lognormally with parameters (μ, σ^2) , then $\ln x$ is distributed normally with parameters (μ, σ^2) , and $E[x] = \exp(\mu + \sigma^2/2)$. Due to the constraint $E[\varepsilon_t] = 1$, μ and σ must be related by $\mu = -\sigma^2/2$.

⁶We estimate all models using Gauss v. 4.0.23 with the maximum likelihood library `maxlik` v. 4.0.24. For numerical optimization, the BFGS and BHHH algorithms are utilized, standard errors are computed using numerical derivatives. All autoregressive processes are reinitialized at the beginning of each day, as in, for example, Engle and Russell (1998) and Zhang, Russell and Tsay (2001).

residuals ($Q(15) = 32.67$) is significant at 1% level (the critical value is 30.58). Moreover, there is significant autocorrelation in the squared standardized residuals, with the Ljung–Box statistic 44.90, and the excess dispersion test shows that there is still overdispersion in the standardized residuals (the value of an asymptotically standard normal statistic is 13.50). The parametric ACD(2, 2) models with the Weibull and generalized gamma distributions also fail to capture the autocorrelation in the squared standardized residuals. The semiparametric Log–ACD(2, 2) model with specification (2.3.3) does not pass both Ljung–Box tests, with the corresponding statistics 108.57 and 137.92 for standardized residuals and their squares.

Dissatisfied with these results, we estimate the following Log–ARMA(1, 2):

$$\begin{aligned}\ln d_t &= \alpha + \beta \ln d_{t-1} + \gamma_1 \zeta_{t-1} + \gamma_2 \zeta_{t-2} + \delta v_{t-1} + \zeta_t, \\ \zeta_t &= \sigma_t \eta_t, \\ \sigma_t^2 &= \mu + \lambda \zeta_{t-1}^2 + \nu \sigma_{t-1}^2.\end{aligned}\tag{2.3.8}$$

The results for seven stocks are presented in Table 4. It can be seen from the table that all models pass the Ljung–Box tests both for standardized residuals and squared standardized residuals at the 1% level. It can be also seen that Lukoil, UES, Sberbank, Yukos, SurgutNefteGaz, and Rostelecom pass Nyblom tests both for individual coefficients and for a whole model at the 1% level. The instability of the model for MosEnergO can be explained by an unstable coefficient for the lagged volume.

It is interesting to look at the distribution of standardized residuals. A nonparametric density estimate for the Lukoil standardized residuals is shown in Figure 7⁷. The residual distribution is very close to the normal, although the Kolmogorov–Smirnov, Jarque–Bera, Cramer–von Mises, and Anderson–Darling tests all reject the null hypothesis of normality. In fact, the distributions for all stocks except UES are slightly skewed to the left and insufficient kurtosis (for Lukoil, the skewness is -0.174 , the kurtosis is 2.678). We thus treat our estimation procedure as quasi-maximum likelihood.

We can compare the corresponding (quasi-)loglikelihood functions of the semiparametric ACD, Log–ARMA and the two versions of the Log–ACD models⁸. For all stocks the mean values of loglikelihood functions for the EACD(2, 2) and both Log–ACD(2, 2) models are appreciably below that for the Log–ARMA(1, 2) specification. For example, if we take the Lukoil stocks, the mean loglikelihood values are -0.9229 for the ACD(2, 2), -0.9226 for the Log–ACD(2, 2) with specification (2.3.3), -0.9243 for the Log–ACD(2, 2) with specification (2.3.4), and -0.8398 for the Log–ARMA(1, 2) model. The big gap between the Log–ARMA and the other models is so large that it also justifies the use of a larger number of parameters: the Akaike and Schwartz information criteria favor the Log–ARMA model. Exactly the same ranking happens with the other stocks, except that for UES, the specification (2.3.4) works better than (2.3.3). We can conclude that the movements in the conditional mean and conditional variance of duration process should not be tied as in the ACD-type models.

Let us now look at the obtained estimates. Very significant AR and MA coefficients indicate high persistence in the mean equation which is consistent with previous results (e.g., Engle and Russell 1998, Engle 2000, Zhang, Russell, and Tsay 2001). The lagged volume has negative sign and is highly significant for all stocks, so we can conclude that if large volumes are traded, then durations are smaller. This favours the model by Admati and Pfleiderer (1988) and may

⁷We use the Nadaraya–Watson estimator with the gaussian kernel and Silverman rule for selecting the bandwidth (Silverman 1986, eq. 3.31).

⁸Since the Log–ARMA model is written for logarithms of durations, we multiply the likelihood function by the corresponding Jacobian to obtain the likelihood function for durations.

be considered as an indication of the effect of bunching of liquidity traders. Clustering of transactions revealed by high AR and MA coefficients also supports this view.

The variance equation for durations has further implications. The persistence in variance is even greater than in the mean: the sum of ARCH and GARCH coefficients is about 0.99 for almost all stocks. We may interpret this using the hazard functions for the duration process, implied by the normal distribution for errors. If the values of ζ_{t-1}^2 and σ_{t-1}^2 are high so that a previous duration is far from its conditional mean, it is more likely that σ_t^2 will also be high, and the hazard function will be steep, indicating that this duration will be either very short or very long. On the other hand, if previous values of ζ_{t-1}^2 and σ_{t-1}^2 are low, the hazard function will be flatter, implying that the distribution of the next duration is more uniform.

2.4 Modeling price dynamics

2.4.1 UHF-GARCH and ACD-GARCH models

The irregular spacing of stock returns implies that conventional measuring of volatility by the “traditional” variance, i.e. the conditional mean squared return, is no longer appropriate. Engle (2000) argued that it is natural to measure it by the return variance per time unit:

$$\pi_t^2 = \frac{V[r_t | I_{t-1}, d_t]}{d_t} = V \left[\frac{r_t}{\sqrt{d_t}} | I_{t-1}, d_t \right].$$

In the UHF-ARMA(p, q)-GARCH(r, s) model, where “UHF” stands for “ultra high frequency”, or UHF-GARCH for brevity, the weighted return follows ARMA(p, q) process

$$\frac{r_t}{\sqrt{d_t}} = \sum_{i=1}^p \varsigma_i \frac{r_{t-i}}{\sqrt{d_{t-i}}} + \sum_{i=1}^q \tau_i \varepsilon_{t-i} + \varepsilon_t,$$

where $\varepsilon_t = \pi_t \eta_t$ follows GARCH(r, s):

$$\begin{aligned} \eta_t &\sim IID(0, 1), \\ \pi_t^2 &= \omega + \sum_{i=1}^r \phi_i \varepsilon_{t-i}^2 + \sum_{i=1}^s \psi_i \pi_{t-i}^2. \end{aligned}$$

Engle (2000) assumed that the durations are not “Granger caused” by prices (see Engle 2000, p. 11), so additional variables that belong to the information set $I_{t-1} \cup \{d_t\}$ may be introduced into the mean and variance equations. Engle (2000) himself added the variable d_t^{-1} to capture the direct positive effect of data frequency on the volatility. However, although it does seem natural to measure volatility per time units, the choice of the weighting is quite arbitrary.

An alternative model for the return volatility was proposed by Ghysels and Jasiak (1998) who used the idea of temporal aggregation for GARCH processes (Drost and Nijman 1993, Drost and Werker 1996) to account for the irregularity of transaction data. This resulted in the following equation for return variances:

$$\sigma_t^2 = \omega_t + \phi_t r_{t-1}^2 + \psi_t \sigma_{t-1}^2,$$

where the time-varying coefficients ω_t, ϕ_t and ψ_t are certain known functions of deeper parameters and of variables from I_{t-1} . Appended by a duration equation of an ACD-type, this model is called ACD-GARCH. In contrast to the UHF-GARCH model, it does not rely on a specific choice of time units for returns. The suggested flexible parameterization leads to a sophisticated GMM estimation procedure where standard errors for some coefficients are hard to compute because of their complexity (see Ghysels and Jasiak 1998, p. 144). Another problem with this specification is that the current duration does not affect the current return variance, although empirically this influence is very significant (see the discussion in Meddahi, Renault and Werker 2003).

In the next subsection we propose a GARCH-like model that tries to remedy the drawbacks of the UHF-GARCH and ACD-GARCH models.

2.4.2 Flexible Time Standardization GARCH model

In specifying the system of equations which describe the process for returns, we assume an ARMA(p, q) process for returns

$$r_t = \alpha + \sum_{i=1}^p \beta_i r_{t-i} + \sum_{i=1}^q \gamma_i \zeta_{t-i} + \zeta_t, \quad (2.4.9)$$

where $\zeta_t = \sigma_t \eta_t$, and a GARCH(r, s) process for the variance of normalized innovations

$$\eta_t \sim IID(0, 1), \quad (2.4.10)$$

$$\frac{\sigma_t^2}{d_t^\rho} = \mu + \sum_{i=1}^r \lambda_i \frac{\zeta_{t-i}^2}{d_{t-i}^\rho} + \sum_{i=1}^s \nu_i \frac{\sigma_{t-i}^2}{d_{t-i}^\rho}. \quad (2.4.11)$$

In the model (2.4.9)–(2.4.11), $\alpha, \beta_i, \gamma_i, \mu, \lambda_i, \nu_i$ and ρ are unknown parameters. One could also specify the distribution for η , but we prefer to treat the problem as semiparametric⁹. We will refer to such specification as the *FTS-ARMA(p, q)-GARCH(r, s)* model, where “FTS” stands for “flexible time standardization”, or *FTS-GARCH* for brevity. As usual, additional variables, such as lagged volumes and durations, may be included in the mean and variance equations to test implications of microstructure theories.

The FTS-GARCH specification assumes that there is a high frequency measure of volatility σ_t^2/d_t^ρ , where the parameter $\rho \in [0, 1]$ is determined endogenously. This parameter indicates the “equilibrium” weighting for the “traditional” volatility for a particular stock. Putting $\rho = 1$ makes the FTS-GARCH model similar (though not equivalent) to the UHF-GARCH model. Let us rewrite the variance equation (2.4.11) with $r = s = 1$ in the form

$$\frac{\sigma_t^2}{d_t} = d_t^{\rho-1} \left[\mu + \lambda d_{t-1}^{1-\rho} \frac{\zeta_{t-1}^2}{d_{t-1}} + \nu d_{t-1}^{1-\rho} \frac{\sigma_{t-1}^2}{d_{t-1}} \right].$$

When $\rho = 1$, given the past information, the variance per time unit does not depend on the current duration; when $0 \leq \rho < 1$, given the past information, the variance per time unit depends

⁹Imposing normality or Studentcy on this distribution is unreasonable because it may inherit a point mass from the cluster of price changes at zero (recall the discussion in subsection 2.2). On the other hand, the distribution is nearly symmetric, hence the QML estimation must work sufficiently well.

on it negatively. Thus, our specification is richer than the one implied by the UHF–GARCH model. As presented in the next subsection, the estimates of the coefficient ρ are far below unity. To analyze the differences between the UHF–GARCH(1, 1) and FTS–GARCH(1, 1) equations, suppose that the previous duration d_{t-1} equals 1, the sample mean of adjusted durations. Then we have $\sigma_t^2 = d_t s_{t-1}^2$ for the UHF–GARCH and $\sigma_t^2 = d_t^\rho s_{t-1}^2$ with small ρ for the FTS–GARCH, where $s_{t-1}^2 = \mu + \lambda \zeta_{t-1}^2 + \nu \sigma_{t-1}^2$. Thus, the “traditional” volatility σ_t^2 in the FTS–GARCH model is a severely concave function of the current duration and depends on it in higher degree for small values of d_t and in lower degree for large values of d_t than in the UHF–GARCH model.

Note that the variance equation (2.4.11) can be viewed as a GARCH-type specification with respect to the “traditional” variance σ_t^2 with time-varying coefficients that are functions of current and previous durations. Thus, we may think of the FTS–GARCH model as a generalization of the UHF–GARCH model in the direction of the ACD–GARCH. Note however, that in the ACD–GARCH model current durations do not enter the time-varying coefficients. The latter feature of the ACD–GARCH model is heavily criticized by Meddahi, Renault and Werker (2003). When comparing the Engle (2000) and Ghysels and Jasiak (1998) specifications, these authors give a preference to the UHF–GARCH model for that reason, and in addition because their exact discretization method leads to a formulation in terms of the “natural” rather than “traditional” volatility.

On the other hand, our FTS–GARCH specification does not support the implications of the model of Meddahi, Renault and Werker (2003) who propose a continuous time specification for the process for returns. This specification implies that the variance per time unit converges to a non-zero limit as d_t goes to infinity, hence an indefinite increase in d_t implies an indefinite increase in the “traditional” return variance σ_t^2 . This does not seem plausible. As d_t increases, the period without new information about the stock value becomes longer, and it is natural to think that the “traditional” variance should stabilize, i.e. the variance per time unit should go to zero, which corresponds to the case $0 \leq \rho < 1$ in our specification. Some partial evidence in support of the view that ρ is less than 1 is provided by the UHF–GARCH model with the additional variable d_t^{-1} in the variance equation by Engle (2000) who reports that it has a highly significant positive coefficient.

2.4.3 Empirical results

The UHF–ARMA(1, 1)–GARCH(1, 1) specification with additional d_t in the mean equation and additional d_t^{-1} in the variance equation turns out to be inadequate as follows from very high values of the Ljung–Box statistics for pure and squared standardized residuals (e.g., for Lukoil $Q^2(15) = 1109.13$, while the 1% critical value is 30.58). Next, we estimate the following FTS–ARMA–GARCH specification:

$$\begin{aligned} r_t &= \alpha + \beta r_{t-1} + \gamma \zeta_{t-1} + \zeta_t + \delta_1 d_t + \delta_2 i_{t-1}, \\ \frac{\sigma_t^2}{d_t^\rho} &= \mu + \lambda_1 \frac{\zeta_{t-1}^2}{d_{t-1}^\rho} + \lambda_2 \frac{\zeta_{t-2}^2}{d_{t-2}^\rho} + \nu_1 \frac{\sigma_{t-1}^2}{d_{t-1}^\rho} + \nu_2 \frac{\sigma_{t-2}^2}{d_{t-2}^\rho} + \delta_3 i_{t-1}. \end{aligned} \tag{2.4.12}$$

Here the variable i_t is generated by $i_t = v_t - \ln d_t$ and characterizes the intensity of the flow of shares¹⁰. The results of estimation are presented in Table 5. The model passes the Ljung–Box test at the 1% level for all stocks except UES. Moreover, for Sberbank and Yukos additional

¹⁰Recall that v_t is the adjusted value of $\ln V_t$, logarithm of the number of shares bought at time T_t .

constraints $\lambda_2 = \nu_2 = 0$ can be imposed. The model and all coefficients are stable at the 1% level (except the coefficient ρ for Lukoil) as reported by the overall and individual Nyblom statistics.

We start from the analysis of the mean equation. As in UHF-ARMA-GARCH models in Engle (2000) for the IBM data, the MA(1) coefficient is negative and highly significant. The negative autocorrelation in the returns process is often found in high frequency return data and may be explained by the bid-ask bounce (Campbell, Lo and MacKinlay 1996, chapter 3) or structure of electronic trading systems. As Gouriéroux and Jasiak (2001, section 14.2) note, this predictability of returns based on previous information does not contradict market efficiency. The insignificance of the AR(1) coefficient for five stocks out of seven, which is found significant in Engle (2000), can be explained by the fact that the MA(1) term explains most of autocorrelation; in the model without an MA term the first two AR coefficients are negative and significant at the 5% level. However, the persistence of the returns process is not very high. The coefficient of the current duration is positive and significant for three stocks and insignificant for others. We may regard this fact as a contradiction of the Diamond and Verrecchia (1987) model where the duration should enter with a negative sign. This differs from the results of Engle (2000) who found a support for this model in the IBM data. It should be noted, however, that Engle used returns over the square root of duration in place of pure returns, and we may only infer from a negative signs in his UHF-GARCH model that the returns grow not faster than the square root of duration. The positive and significant coefficients for the intensity variable for most stocks can be interpreted as follows. Large trades are likely to be seller initiated, and, consequently, yield negative returns in most cases (because then several best buy offers will be executed, hence the average price for this transaction will go down). The negative autocorrelation in the returns process will most likely lead to a positive return of the next trade. As a result, we observe a positive impact of lagged volume intensity on returns. This effect is not observed only for the UES stocks because of distortions introduced by omitting simultaneous transactions (recall that UES is traded about ten times more frequently than other stocks).

Now let us turn to the analysis of the variance equation. First, there is volatility clustering for all stocks, and the process is sufficiently persistent as can be inferred by summing up the ARCH and GARCH coefficients. The coefficient for the intensity variable tends to be positive and significant, which can be considered as an evidence against the Admati and Pfleiderer (1988) model where large volumes are associated with liquidity traders and lead to a decline in volatility as these trades do not carry new information. Finally, we obtain a highly significant coefficient ρ which ranges from 0.1 for UES to 0.28 for Sberbank, far below unity. The “natural” volatility π_t^2 declines very rapidly in the current duration, which agrees with the result by Engle (2000) who found a significant coefficient on d_t^{-1} . Negative dependence of the “natural” volatility on the current duration can be interpreted in favour of the Easley and O’Hara (1992) model where long durations imply that there are no news. This type of dependence contradicts the model by Diamond and Verrecchia (1987) where long durations imply that the news are bad, and, consequently, volatility should be higher.

2.5 Conclusion

We have proposed a new flexible specification for the analysis of intertrade durations, which allows for richer dynamics compared to the ACD model and is computationally more attractive than the SVD, and a new flexible specification for return volatilities that endogenizes the defi-

nition of time units in the UHF–GARCH model. Both models very well fit the data on trades in seven most frequently traded common stocks at the Moscow Interbank Currency Exchange. We have found that the Russian stock market data support some of the implications of Admati and Pfleiderer (1988) and Easley and O’Hara (1992) models. At the same time, no evidence is found in favor of the Diamond and Verrecchia (1987) model which assumes constraints on short selling. Indeed, the short-selling constraints are not normally an issue for the participants of electronic trades during tranquil times at financial markets.

Future research includes testing the empirical relevance of the developed econometric models by comparing their out-of-sample predictive power with the accuracy of predictions obtained from more traditional models. Another direction is modeling price movements in a framework that acknowledges discreteness of their changes, along the lines of Russell and Engle (1998), Rydberg and Shephard (1998), and Darolles, Gouriéroux and Le Fol (2000). Finally, as the analyzed stocks essentially represent the whole bulk of trades at MICEx, the modeling of joint dynamics of durations and mark processes, including the analysis of effects of market-wide, industry-wide and stock-specific news, is of great interest.

References

1. Admati, A.R. and P. Pfleiderer (1988) *A theory of intraday patterns: volume and price variability*. The Review of Financial Studies 1, 3–40.
2. Andrews, D.W.K. and W. Ploberger (1994) *Optimal tests when a nuisance parameter is present only under the alternative*. Econometrica 62, 1383–1414.
3. Barinov, V., A. Pervozvansky, and T. Pervozvanskaya (1999) *State debt policy and bonds market behaviour*. Working paper 99/05E, EERC–Russia. Available at www.eerc.ru/publications/wp/WorkingPapers.htm.
4. Bauwens, L. and P. Giot (2000) *The logarithmic ACD model: an application to the bid/ask quote process of three NYSE stocks*. Annales d’Economie et de Statistique 60, 117–149.
5. Campbell, J.Y., A.W. Lo and A.C. MacKinlay (1996) *The econometrics of financial markets*. Princeton University Press.
6. Darolles, S., C. Gouriéroux, and G. Le Fol (2000) *Intraday transaction price dynamics*. Annales d’Economie et de Statistique 60, 239–270.
7. Diamond, D.W. and R.E. Verrecchia (1987) *Constraints on short-selling and asset price adjustments to private information*. Journal of Financial Economics 18, 277–311.
8. Drost, F.C. and T.E. Nijman (1993) *Temporal aggregation of GARCH processes*. Econometrica 61, 907–927.
9. Drost, F.C. and B.J.M. Werker (1996) *Closing the GARCH gap: continuous time GARCH modeling*. Journal of Econometrics 74, 31–57.
10. Dufour, A. and R.F. Engle (1999) *Time and the price impact of a trade*. Discussion paper 99–15, University of California–San Diego.

11. Easley, D. and M. O'Hara (1992) *Time and the process of security price adjustment*. Journal of Finance 47, 577–606.
12. Engle, R.F. (2000) *The econometrics of ultra-high frequency data*. Econometrica 68, 1–22.
13. Engle, R.F. and A. Lunde (1998) *Trades and quotes: a bivariate point process*. Journal of Financial Econometrics, forthcoming in 2003.
14. Engle, R.F. and J.R. Russell (1998) *Autoregressive conditional duration: a new model for irregularly spaced transaction data*. Econometrica 66, 1127–1162.
15. Ghysels, E., C. Gouriéroux, and J. Jasiak (2002) *Stochastic volatility duration models*. Journal of Econometrics, to appear.
16. Ghysels, E. and J. Jasiak (1998) *GARCH for irregularly spaced financial data: the ACD–GARCH model*. Studies in Nonlinear Dynamics and Econometrics 2(4), article 4.
17. Goldberg, L. and R. Tenorio (1997) *Strategic trading in a two-sided foreign exchange auction*. Journal of International Economics 42, 299–326.
18. Gouriéroux, C. and J. Jasiak (2000) *Durations*. In: “A Companion to Theoretical Econometrics”, B. Baltagi (ed.), Blackwell Publishers.
19. Gouriéroux, C. and J. Jasiak (2001) *Financial Econometrics: Problems, Models, and Methods*. Princeton University Press.
20. Grammig, J. and K.-O. Maurer (2000) *Non-monotonic hazard functions and the autoregressive conditional duration model*. Econometrics Journal 3, 16–38.
21. Grammig, J. and M. Wellner (2002) *Modeling the interdependence of volatility and inter-transaction duration processes*. Journal of Econometrics 106, 369–400.
22. Hafner, C.M. (1999) *Durations, volume and the prediction of financial returns in transaction time*. Econometric Society World Congress 2000 Contributed Papers.
23. Hansen, B. (1994) *Autoregressive conditional density estimation*. International Economic Review 35, 705–730.
24. Jasiak, J. (1999) *Persistence in intertrade durations*. Finance 19, 166–195.
25. Li, W.K. and T.K. Mak (1994) *On the squared residual autocorrelations in non-linear time series with conditional heteroscedasticity*. Journal of Time Series Analysis 15, 627–636.
26. McCulloch, R.E. and R.S. Tsay (2001) *Nonlinearity in high frequency financial data and hierarchical models*. Studies in Nonlinear Dynamics and Econometrics 5(1), article 1.
27. Meddahi, N., Renault, E., Werker, B. (2003) *GARCH and irregularly spaced data*. Working paper 2003–27, CentER, Tilburg University.
28. Medvedev, A.N. and G.V. Kolodyazhny (2001) *Financial crisis in Russia: the behavior of non-residents*. Working paper 2K/12E, EERC–Russia. Available at www.eerc.ru/publications/wp/WorkingPapers.htm.
29. Nyblom, J. (1989) *Testing the constancy of parameters over time*. Journal of the American Statistical Association 84, 223–230.

30. Russell, J.R. and R.F. Engle (1998) Econometric analysis of discrete-valued, irregularly-spaced financial transactions data using a new autoregressive conditional multinomial models. Manuscript, GSB, University of Chicago.
31. Rydberg, T.H. and N. Shephard (1998) *Dynamics of trade-by-trade price movements: decomposition and models*. Journal of Financial Econometrics, forthcoming in 2003.
32. Silverman, B.W. (1986) *Density estimation for statistics and data analysis*. Chapman & Hall, London.
33. Tsay, R. (2002) *Analysis of Financial Time Series*, chapter 5. John Wiley & Sons.
34. Tse, Y.K. (2002) *Residual-based diagnostics for conditional heteroscedasticity models*. Econometrics Journal 5, 358–373.
35. Tyurin, K. (2002) *Modeling high-frequency dynamics of financial markets in continuous time: an event-based approach*. Manuscript, Indiana University–Bloomington.
36. Zhang, M.Y., J.R. Russell, and R.S. Tsay (2001) *A nonlinear autoregressive conditional duration model with applications to financial transaction data*. Journal of Econometrics 104, 179–207.

2.6 Appendix A: MICEX trading regimes

There are several trading regimes at the MICEX. Regimes and their brief descriptions are presented in Table 2.

We concentrate on the “trading session”, or “normal” trading regime. All other observations are excluded from the sample, i.e. the following trades were removed:

1. Trades that are recorded out of trading hours for the “normal” regime (10:30AM–6:45PM);
2. Trades that are marked with the string “IN:” (they belong to the “negotiated deals” regime);
3. Trades that are marked with the string “REPO:” (they belong to the “REPO” regime);
4. Trades that are marked with the string “incomplete lots” (they belong to the “incomplete lots” regime).

2.7 Appendix B: stocks under analysis

We use data on common stocks of seven Russian corporations. A brief description of these securities is presented in Table 3. Among the corporations that we consider, three companies are oil extractors (Lukoil, SurgutNefteGaz, and Yukos), UES (short for the Unified Energy System of Russia) and MosEnergo are large electricity producers, Rostelecom is a leading Russian telecommunication company, and Sberbank is the largest Russian commercial bank. For some corporations there are several releases of common stocks, as well as preferred stocks and bonds. However, we focus on only one security for each corporation, the release that is traded much more frequently than other releases. A more detailed description of the companies, their

stocks and trades at the MICEx are available in English and Russian from www.micex.com and www.micex.ru, respectively.

Out of seven stocks, UES is most heavily traded (on average 6000 transactions a day not counting simultaneous transactions); Lukoil, SurgutNefteGaz, Rostelecom and MosEnergO are traded at the frequency of about 600–750 transactions a day; for Yukos we have 400 transactions a day, and only 250 trades for Sberbank. The observation period is 08/12/02–10/27/02, i.e. 50 trading days maximum. Since for UES we have many more transactions than for other companies, we use only one week for model estimation and four weeks for seasonal adjustment. Also, for SurgutNefteGaz the data for August 14, 2002 are incomplete, and we omit them.

The column “stock share” of Table 3 shows the share of each stock value in the value of all trades in stocks at the MICEx during the period 08/19/02–08/23/02. It can be seen that the total value of trades in these seven stocks take almost 97% of the overall value of trades.

Variable	Mean	Standard deviation	Skewness	Kurtosis	$Q(15)$
Raw data					
d_t	43.19	75.43	5.07	44.63	7455.9
r_t	$1.99 \cdot 10^{-6}$	$8.99 \cdot 10^{-4}$	0.09	8.68	2537.2
v_t	4.79	2.19	-0.29	2.62	1439.2
Deseasonalized data					
\hat{d}_t	1.00	1.56	4.04	28.51	3397.6
\hat{r}_t	0.00	0.99	0.06	6.10	2576.5
\hat{v}_t	0.00	2.18	-0.28	2.60	918.46

Trading regime	Timing	Description
Pre-trading period	10:00AM–10:30AM	During this period the opening price is determined.
Trading session	10:30AM–6:45PM	The usual trading regime.
Post-trading period	6:45PM–7:00PM	In this regime trades can be executed at a weighted average price determined during the last 30 minutes of the trading session.
Negotiated deals regime	10:00AM–7:05PM	This regime is designed for large trades, which are not executed immediately, but rather their execution may be delayed up to 30 days.
Special negotiated deals regime	7:05PM–10:00PM	The same as negotiated deals regime, but only specific securities can be traded.
REPO regime	10:15AM–7:05PM	Repurchase agreement trading regime.
Incomplete lots regime	5:00PM–7:05PM	The period when nonstandard lots can be traded.

Company	Security code	Price step	Lot	Stock share	Observation period	Observations in sample
Sberbank	RU0009029540	1	1	0.75%	08/12/02–10/27/02	13156
Yukos	RU0009054449	1	1	1.99%	08/12/02–10/27/02	20125
Lukoil	RU0009024277	1	1	15.71%	08/12/02–10/27/02	34316
SurgutNefteGaz	RU0008926258	0.1	100	5.27%	08/12/02–10/27/02	38275
Rostelecom	RU0008943394	1	1	2.75%	08/12/02–10/27/02	38246
MosEnergo	RU14MSNG3008	0.1	100	1.88%	08/12/02–10/27/02	32452
UES	RU0008926621	0.1	100	68.75%	08/19/02–08/23/02	30537

Note: price step is in kopecks (1/100 of the ruble), lots are in shares.

Table 4. Estimation of Log-ARMA-GARCH models

	Sberbank		Yukos		Lukoil		SurgutNefteGaz		Rostelecom		MosEnerg		UES	
	Coef	Nyb	Coef	Nyb	Coef	Nyb	Coef	Nyb	Coef	Nyb	Coef	Nyb	Coef	Nyb
α	-0.186 (-8.70)	0.15	-0.170 (-8.31)	0.11	-0.109 (-10.54)	0.06	-0.101 (-12.87)	0.13	-0.078 (-11.36)	0.16	-0.085 (-12.01)	0.13	-0.025 (-11.81)	0.14
β	0.812 (39.19)	0.10	0.808 (35.43)	0.20	0.875 (74.67)	0.22	0.876 (93.27)	0.20	0.906 (111.68)	0.20	0.897 (107.73)	0.15	0.950 (237.09)	0.10
γ_1	-0.605 (-26.35)	0.11	-0.639 (-26.24)	0.53	-0.691 (-52.06)	0.07	-0.701 (-63.74)	0.11	-0.736 (-73.92)	0.10	-0.732 (-71.51)	0.17	-0.819 (-115.58)	0.19
γ_2	-0.029 (-2.57)	0.13	-0.023 (-2.35)	0.61	-0.049 (-6.71)	0.19	-0.039 (-6.18)	0.12	-0.048 (-7.66)	0.15	-0.033 (-4.97)	0.50	-0.028 (-4.38)	0.24
δ	-0.061 (-9.28)	0.12	-0.032 (-7.85)	0.35	-0.028 (-11.34)	0.31	-0.028 (-11.78)	0.12	-0.017 (-9.66)	0.17	-0.014 (-7.85)	1.40	-0.013 (-11.37)	0.27
μ	0.019 (2.86)	0.25	0.028 (2.28)	0.35	0.046 (4.84)	0.05	0.022 (4.52)	0.25	0.015 (4.76)	0.19	0.012 (4.47)	0.43	0.011 (6.64)	0.37
λ	0.020 (4.85)	0.12	0.010 (3.21)	0.39	0.020 (8.93)	0.05	0.014 (8.76)	0.25	0.012 (9.06)	0.16	0.013 (9.19)	0.37	0.044 (13.38)	0.38
ν	0.970 (140.07)	0.16	0.975 (106.87)	0.38	0.954 (133.11)	0.05	0.972 (227.44)	0.21	0.978 (334.83)	0.20	0.979 (363.89)	0.39	0.941 (185.68)	0.43
Diagnostic Statistics														
$Q(15)$	27.46		29.88		22.15		24.13		19.94		18.04		18.47	
$Q^2(15)$	19.24		20.81		15.12		27.60		20.96		11.34		12.14	
Nyblom	2.04		1.93		1.52		2.15		1.37		3.86		1.96	

Notes: The table presents estimates of equation (8). “Coef” denotes a point estimate with a robust t -statistic in parentheses. “Nyb” denotes an individual Nyblom statistic whose 5% and 1% critical values are 0.46 and 0.74, respectively. “Nyblom” denotes a Nyblom statistic for the whole model, whose 5% and 1% critical values are 2.12 and 2.58, respectively. “ $Q(15)$ ” and “ $Q^2(15)$ ” denote Ljung-Box statistics of order 15 for pure and squared standardized residuals, whose 5% and 1% critical values are 25.00 and 30.58, respectively.

Table 5. Estimation of FTS-ARMA-GARCH models

	Sberbank		Yukos		Lukoil		SurgutNefteGaz		Rostelecom		MosEnergO		UES	
	Coef	Nyb	Coef	Nyb	Coef	Nyb	Coef	Nyb	Coef	Nyb	Coef	Nyb	Coef	Nyb
α	0.014 (1.69)	0.07 (-2.95)	-0.018 (-3.29)	0.08 (-0.71)	-0.016 (-3.29)	0.40 (-0.71)	-0.004 (-0.71)	0.12 (-0.72)	-0.003 (-0.72)	0.09 (-2.11)	-0.010 (-2.11)	0.08 (-1.15)	-0.006 (-1.15)	0.24 (-0.33)
β	-0.012 (-0.29)	0.11 (0.60)	0.018 (0.60)	0.04 (-0.07)	-0.002 (-0.07)	0.37 (-0.270)	0.005 (0.16)	0.07 (-0.268)	0.026 (0.84)	0.06 (1.99)	0.044 (1.99)	0.31 (-2.25)	-0.033 (-2.25)	0.17 (-0.342)
γ	-0.226 (-5.44)	0.10 (-10.13)	-0.302 (-10.13)	0.04 (-12.55)	-0.270 (-12.55)	0.23 (-12.55)	-0.268 (-8.28)	0.13 (-8.50)	-0.260 (-8.50)	0.09 (-14.59)	-0.311 (-14.59)	0.33 (-24.50)	-0.342 (-24.50)	0.15 (-0.005)
δ_1	-0.008 (-1.26)	0.13 (2.78)	0.013 (2.78)	0.04 (3.05)	0.012 (3.05)	0.13 (3.05)	0.0005 (0.13)	0.08 (0.28)	0.001 (0.28)	0.06 (2.27)	0.009 (2.27)	0.16 (1.32)	0.005 (1.32)	0.27 (-0.004)
δ_2	0.010 (3.05)	0.09 (1.74)	0.004 (1.74)	0.18 (2.56)	0.005 (2.56)	0.40 (2.56)	0.005 (2.70)	0.33 (2.89)	0.004 (2.89)	0.62 (1.69)	0.003 (1.69)	0.25 (0.010)	-0.004 (0.010)	0.17 (6.28)
μ	0.139 (5.72)	0.54 (8.62)	0.224 (8.62)	0.15 (4.54)	0.060 (4.54)	0.22 (4.54)	0.125 (3.50)	0.06 (3.53)	0.048 (3.53)	0.09 (3.60)	0.099 (3.60)	0.20 (0.191)	0.010 (0.191)	0.18 (24.61)
λ_1	0.184 (11.31)	0.33 (14.55)	0.189 (14.55)	0.04 (17.29)	0.212 (17.29)	0.02 (17.29)	0.204 (3.77)	0.03 (3.77)	0.201 (13.86)	0.03 (13.86)	0.193 (10.01)	0.03 (10.01)	0.191 (73.75)	0.07 (1.447)
ν_1	0.676 (19.38)	0.54 (17.29)	0.586 (17.29)	0.15 (-1.69)	1.006 (-1.69)	0.10 (-1.69)	0.683 (-1.69)	0.05 (-1.69)	1.126 (-4.13)	0.05 (-4.13)	0.977 (-5.55)	0.11 (-5.55)	1.447 (-19.88)	0.10 (-0.170)
λ_2					-0.143 (-10.98)	0.02 (-10.98)	-0.064 (-1.69)	0.03 (-1.69)	-0.140 (-8.77)	0.03 (-8.77)	-0.112 (-5.55)	0.04 (-5.55)	-0.170 (-19.88)	0.07 (-0.481)
ν_2					-0.135 (-5.35)	0.12 (-5.35)	0.051 (0.44)	0.06 (0.44)	-0.236 (-4.13)	0.05 (-4.13)	-0.156 (-2.71)	0.11 (-2.71)	-0.481 (-23.84)	0.10 (0.001)
δ_3	0.013 (3.04)	0.21 (3.74)	0.013 (3.74)	0.30 (3.83)	0.005 (3.83)	0.24 (3.83)	0.008 (3.29)	0.07 (3.06)	0.004 (3.06)	0.18 (1.42)	0.002 (1.42)	0.08 (2.86)	0.001 (2.86)	0.10 (0.102)
ρ	0.277 (20.16)	0.07 (16.92)	0.203 (16.92)	0.17 (22.83)	0.223 (22.83)	1.07 (22.83)	0.216 (22.83)	0.46 (24.90)	0.238 (24.90)	0.03 (24.22)	0.202 (24.22)	0.42 (7.60)	0.102 (7.60)	0.03
Diagnostic Statistics														
$Q(15)$	19.95	19.80	30.99	24.93	15.86	15.10	62.74							
$Q^2(15)$	26.26	23.07	25.69	29.32	13.89	27.13	26.38							
Nyblom	1.85	1.63	3.14	2.22	2.10	2.18	2.00							

Notes: The table presents estimates of equation (12). “Coef” denotes a point estimate with a robust t -statistic in parentheses. “Nyb” denotes an individual Nyblom statistic whose 5% and 1% critical values are 0.46 and 0.74, respectively. “Nyblom” denotes a Nyblom statistic for the whole model, whose 5% and 1% critical values are 2.53 and 3.04, respectively, for models with 10 parameters, and approximately 0.20–0.40 bigger for models with 12 parameters. “ $Q(15)$ ” and “ $Q^2(15)$ ” denote Ljung-Box statistics of order 15 for pure and squared standardized residuals, whose 5% and 1% critical values are 25.00 and 30.58, respectively.

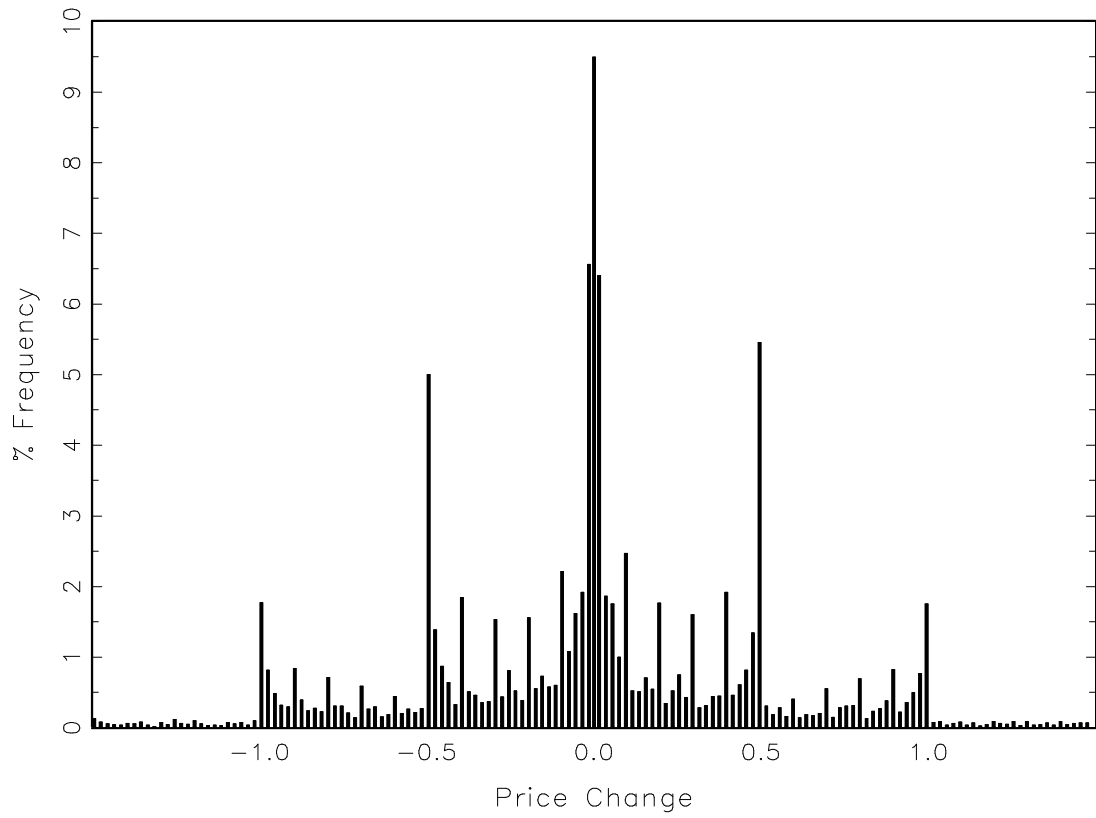


Figure 1. Histogram of non-zero price changes for Lukoil stocks.

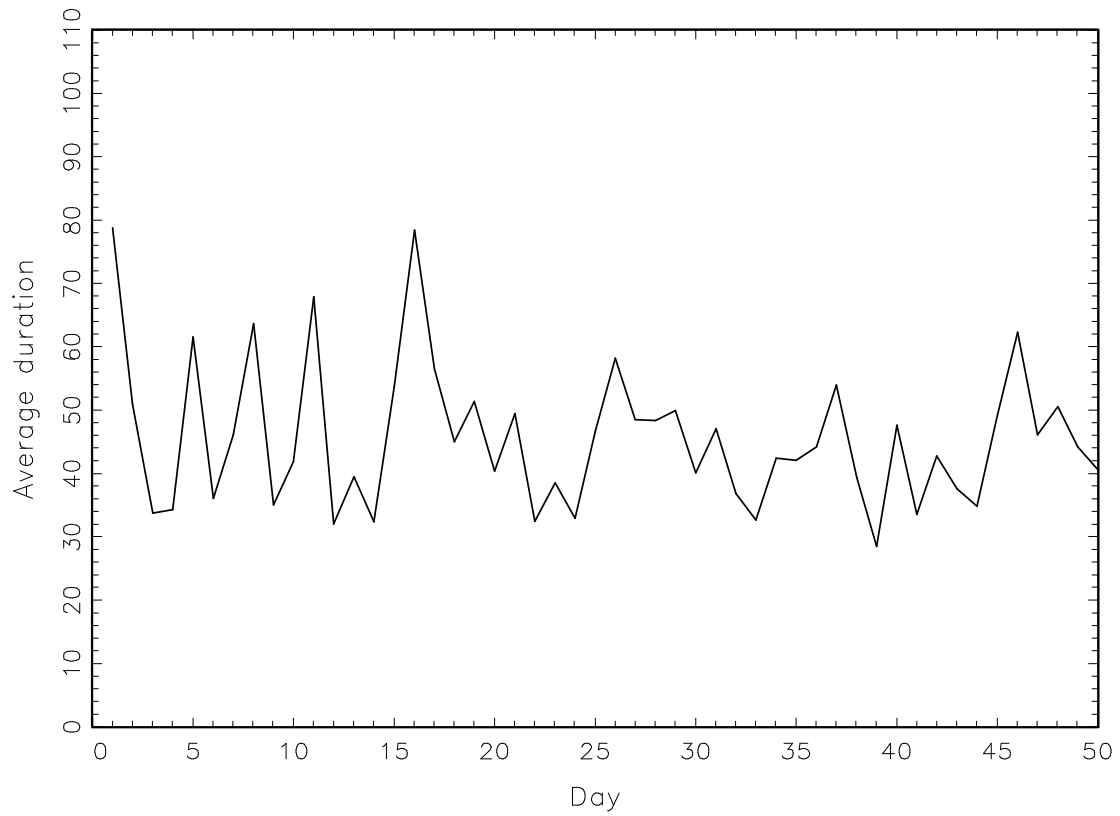


Figure 2. Average durations for different days for Lukoil stocks.

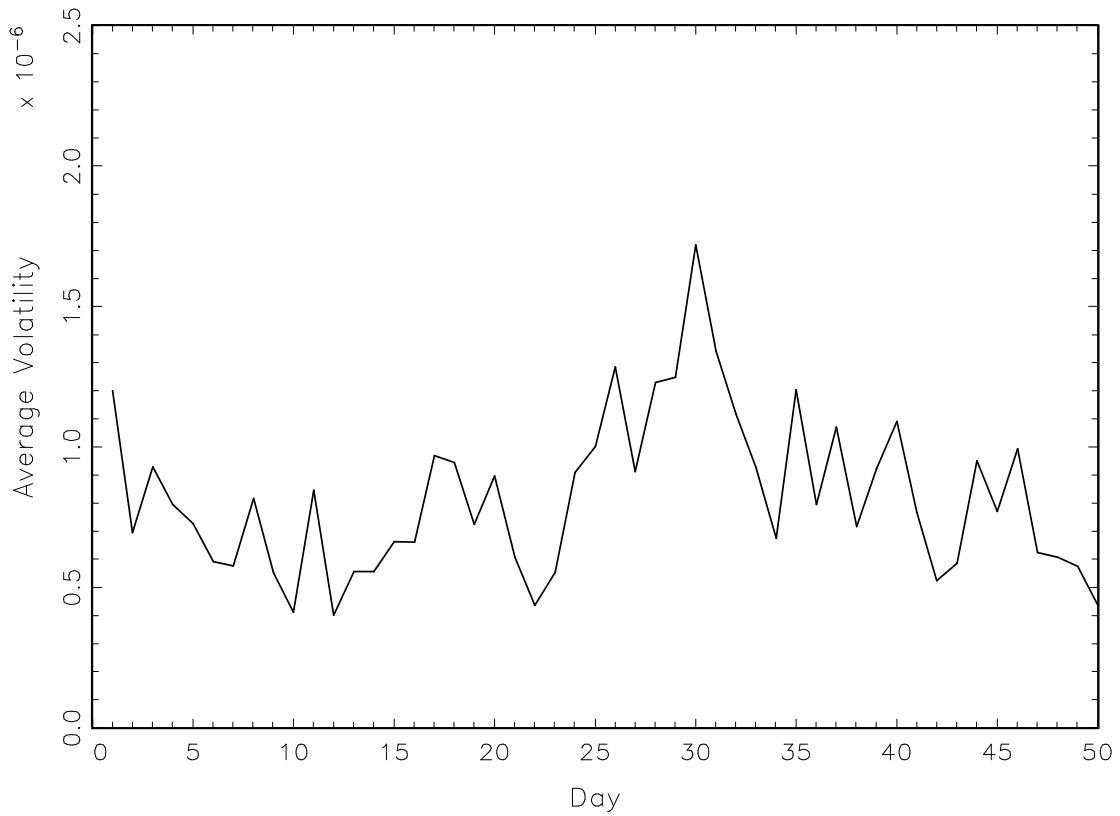


Figure 3. Average squared returns for different days for Lukoil stocks.

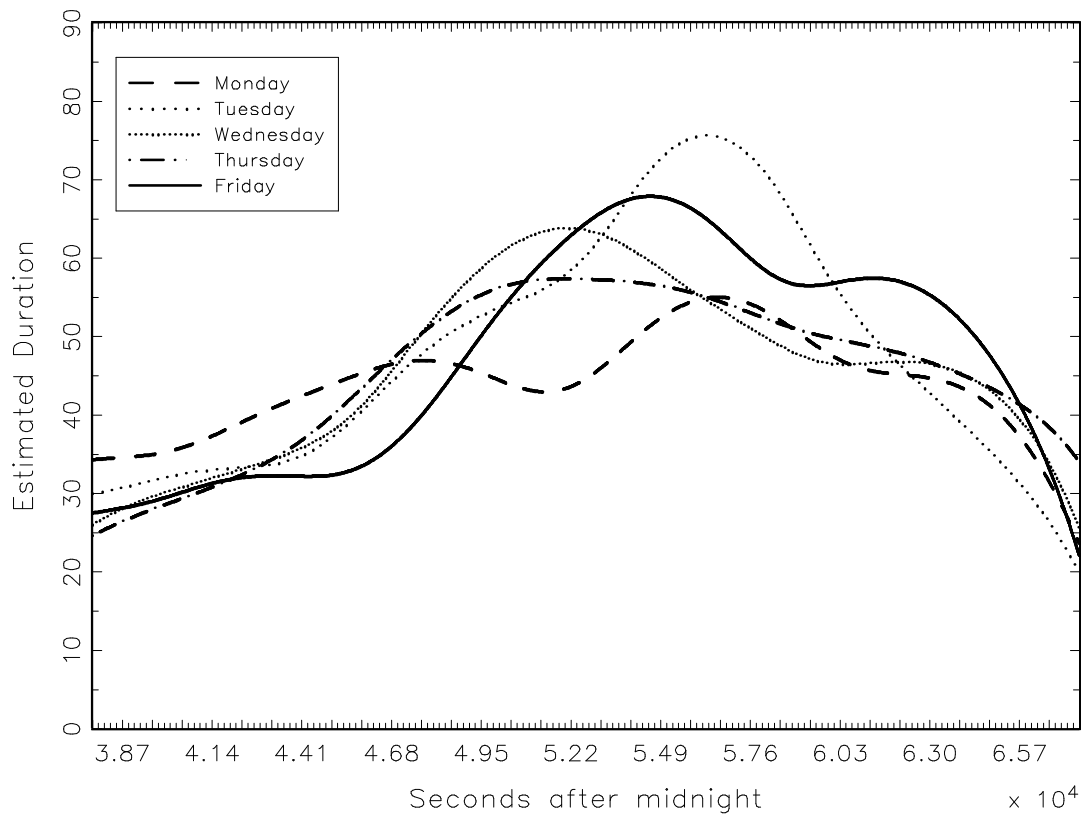


Figure 4. Estimated intraday patterns for durations for Lukoil stocks.

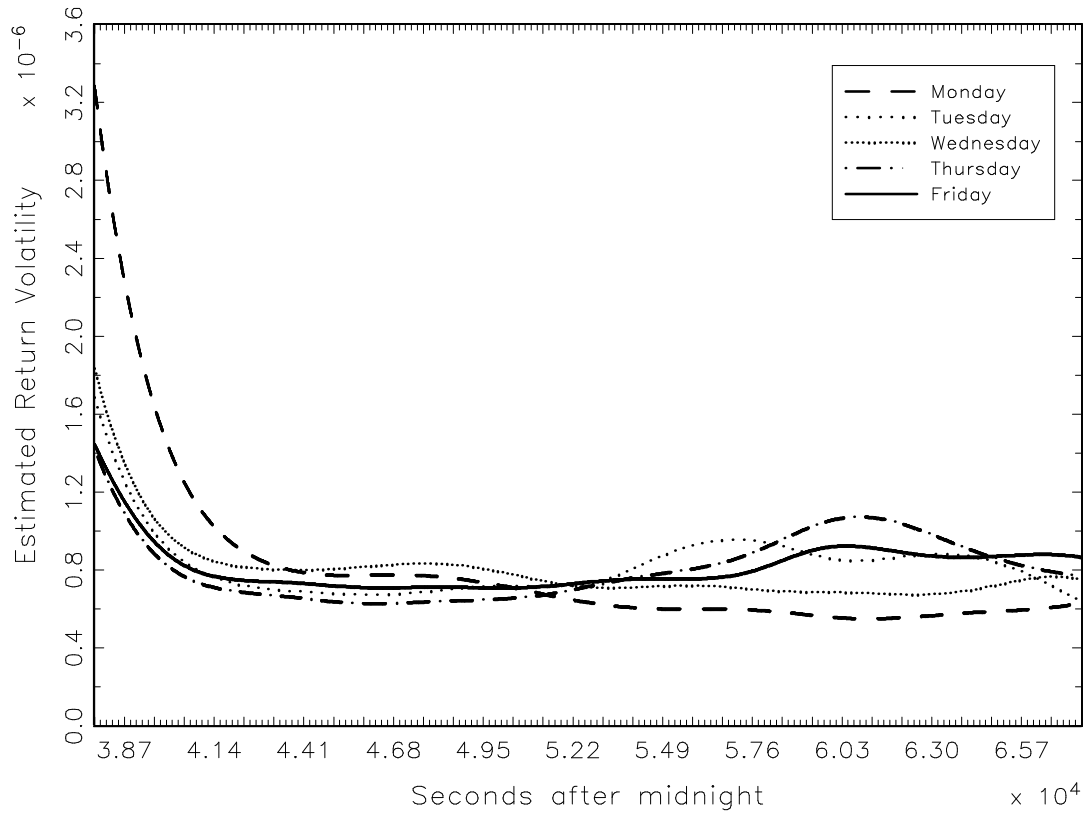


Figure 5. Estimated intraday patterns for squared returns for Lukoil stocks.

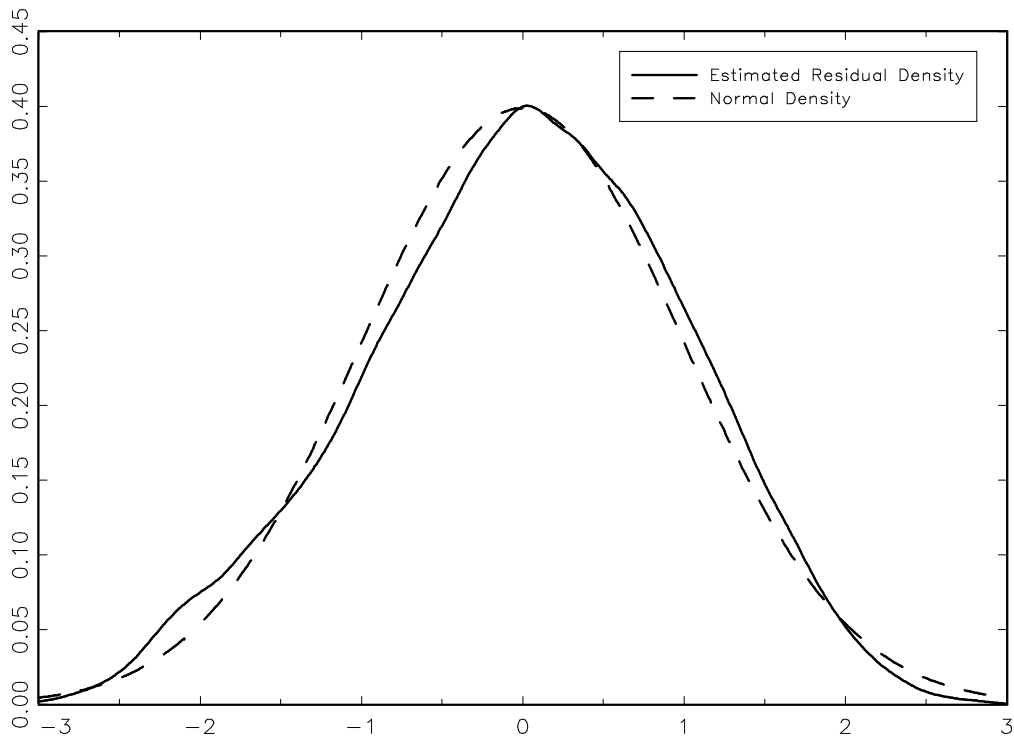


Figure 6. Estimated residual density in Log-ARMA model for Lukoil stocks.