Academy of Economic Studies Doctoral School of Finance and Banking

Lead – Lag Relationship between the Romanian Cash Market and Futures Market

MSc Student: Streche Lucian

Supervisor: Ph.D. Professor Moisă Altăr

Bucharest, July 2008

Table of Contents

Abstract

The fundamental purpose of this paper is to unravel the way price discovery works in the Romanian markets and at the same time explain its most obvious mechanisms. This is an aid for traders that use both markets (cash and futures) but at the same time it is a relevant input when trying to assess local market and investor maturity. It may also be a relevant piece of information for market regulators, as it gives an inside into the way the whole stock market set works; the indicators analyzed are a result of all the elements interacting in the stock market, not only some.

Price discovery mechanisms in the equity and futures markets yield important data for traders if uncovered. One important factor in price discovery is the exchange of information between the cash market and the futures market, when futures contracts with listed equities as underlying assets are traded. When new information emerges, it is integrated in the two markets with different speeds, depending upon the characteristics of the markets and the investors involved. Hence, a lead-lag relation between the two markets emerges.

We try to discover and explain this relation using two different models, (and two implicitly different approaches: top-down and bottom-up). The data series used are high frequency observations of the instantaneous return rates for two listed market funds (SIF2 and SIF5) along with their futures contracts (DSIF2 and DSIF5); the traded volumes are also inputs.

The results show that, in opposition to US markets results, the Romanian cash market leads the futures market by three to five minutes. The results generally hold strong under different conditions: long data series – short data series (top-down approach with 45.000 observations, bottom-up approach with 500 observations), higher frequency – lower frequency data (one minute – five minutes), high volume – low volume, good news – bad news and bull – bear market.

I. Introduction and short history

Business is the cornerstone of every economy. Another law that governs the economy is that a business must grow or otherwise succumb to the ever changing market conditions. In order for a business to expand, other than a good product or service, a company needs to be able to finance its expansion. Owners generally have two options to overcome this. They can either borrow the money from a bank or venture capitalist, or sell part of the business to investors and use the money to fund growth. Taking out a loan is common, and very useful, however the institutions that lend money have specific requirements to meet in order to accept the applicant's loan request. Banks will not always lend money to companies, and over – eager managers may try to borrow too much initially, disturbing the balance of the balance sheet. Factors such as these often provoke owners of small businesses to issue stock. In exchange for giving up a fraction of control, they are given cash to expand the business. In addition to money that doesn't have to be paid back, "going public", the terms used when a company sells stock in itself for the first time, gives the business managers and owners a new tool: instead of paying cash for an acquisition, they can use their own stock.

A text-book definition of the stock exchange is an institution established for the purpose of assisting, regulating and controlling the business of buying, selling and dealing in securities.

The global network of stock markets is the heart of the global economy, pumping the finances needed in different parts of the world through complicated structures that large investors have set up in order to take advantage of high yield opportunities, no matter the geographical coordinates. This interconnection amongst the markets can prove both beneficial and problematic for investors, since capital can easily flow, but along with it, shocks also spread easily. The main world indexes are some of the best indicators evaluating the global economy's state of health. Globally, the size of the stock markets is estimated at about $51 trillion.

Today, the largest stock markets are the London Stock Exchange and New York Stock Exchange. For the London Stock Exchange there are more than 380 firms worldwide trade as members. Approximately 1,840 companies are listed on the London Stock Exchange with a total market value of 4.3 trillion British Pounds. In 2005, there were roughly 63 million trades executed on this market involving 2.5 trillion British Pounds.

For many years, more securities were exchanged on the New York Stock Exchange than any other trading floor in the world. This made the NYSE not only the busiest exchange in the world, but also the most prestigious. Today, roughly 1.5 billion shares worth roughly $87 billion are exchanged daily on the floor of the NYSE. There are currently around 2,764 companies listed on the exchange (December 2007) with a capitalization of nearly $20 trillion.

Equities are not the only products traded on the stock exchanges. In time, more and more complex derivatives have emerged to fulfill the ever growing in complexity investor needs. The world derivatives market has been estimated at about $480 trillion face or nominal value, 30 times the size of the U.S. economy and 12 times the size of the entire world economy.

Aside from structural changes, some derivative exchanges also changed the way they conducted trading. Old systems of face-to-face trading on trading floors have been replaced with electronic trading, and telephone and computer networks. With the advent of Internet, electronic trading evolved into e-trading.

There is a general consensus that London and New York are the world's primary markets for over-the-counter derivatives. Notably, significant derivatives trading is also happening in Tokyo, Paris, Frankfurt, Chicago, Amsterdam, etc.

In spite of the close connection between worldwide markets, many parameters, like market maturity, liquidity, total size of the market, influence heavily how markets react to information. With such a large set of parameters to consider, markets retain important specific characteristics.

In Romania there are two operational Exchanges. The Bucharest Stock Exchange is the only market for local equities. It is a small market compared to LSE or NYSE: around 20 million EUR are traded daily, while the total market capitalization is around 30 billion EUR. There are two derivative markets, Sibiu Monetary Financial and Commodities Exchange (BMFMS) and a very illiquid Bucharest Stock Exchange. On the BMFMS daily trades account around 18.000 futures contracts, with a total nominal value approaching 15 million EUR.

Even thou the tradition exists, the Bucharest Exchange opened on 1 December 1882, the long period of non-trading that ended in 1995 classifies it as a young market. The Sibiu Monetary Financial and Commodities Exchange began its history in 1994 and is the only relevant derivatives market.

For a stock market system to function at full potential, besides the stock market, there is a need for other institutions. The clearing and settlement is handled by The Bucharest Clearing House. Investor information and protection should be assured by legislation and BVB rules. As with any legislation, the Romanian one has its strong points and its weak ones, being in development. Romania doesn't have any important rating agencies, nor a well known rating system, which means that the small companies don't have to fear a negative public rating.

The education and training of investors as well as company managers, regarding the stock market, is one of the biggest problems that plagues this market. The way investors trade, the way listed companies share relevant information are both considerable problems that influence our markets and obviously any analysis based on them.

II. Literature review

Since the futures contracts have as underlying assets common stock, it is obvious that there is a strong connection between their prices. On the Romanian market less than on the US market, futures traders take coincident positions in the cash market such that a substantial volume of equity transactions is tied to futures activity. On the US markets futures trading influences the underlying equity prices, especially on days when institutional investors implement program trading strategies. While this inter-market effect can appear at any time, it is commonly associated with the final hour of trading on the futures contract expiration days. Proponents of futures argue that these markets provide an important price discovery vehicle and offer an alternative marketplace for adjusting equity exposure. Moreover, they do not view the price swings as a problem.

The lead-lag relation between price movements of stock index futures and the underlying cash market illustrates how fast one market reflects new information relative to the other, and how well the two markets are linked. In a perfectly frictionless world, price movements of the two markets are contemporaneously correlated and not cross-auto correlated. However, if one market reacts faster to information, and the other market is slow to react, a lead-lag relation is observed.

There are a number of papers that try to review the relationship between the cash markets and futures markets, but almost all of them focus on the US stock exchanges. More precisely they use the S&P 500 market index and its futures contract.

One of the earlier papers is "The temporal price relationship between S&P 500 futures and the S&P 500 Index" (Kawaller, Koch & Koch, 1987)[10]. This is focused on the effect of the futures market on the cash market in the last days of the futures contracts.

The primary objective was to determine whether movements in the futures price provide predictive information regarding subsequent movements in the S&P 500 index and/or vice versa. They employed time-series regression analysis to identify the nature of this intraday dynamic relationship and test whether a systematic lead/lag relationship exists. The tests distinguish between the prices relationships on expiration day versus days prior to expiration because market activity on expiration day may differ from that on non-expiration days.

The model used was as follows:

$$i_t = z_1 + \sum_{k=1}^{\infty} a_k i_{t-k} + \sum_{k=0}^{\infty} b_k f_{t-k} + e_{1t},$$

$$f_t = z_2 + \sum_{k=0}^{\infty} c_k i_{t-k} + \sum_{k=1}^{\infty} d_k f_{t-k} + e_{2t},$$

Where z_1 and z_2 are the intercept terms, i_t equals the change in I_t, f_t equals the change in F_t (i.e., $i_t = (1 - L)I_t$ and $ft = (1 - L)F_t$, with L equal to the lag operator ($L_kI_t = I_{t-k}$, $L_kF_t = F_{t-k}$)), and other relevant market information affecting these prices is represented by random noise, e_{lt} and e_{2t}. I_t is the cash market price and F_t is the futures price.

Results suggest that S&P 500 futures prices and the index are simultaneously related on a minute-to-minute basis throughout the trading day. Further, significant lag coefficients suggest that the lead from futures to cash prices extends for between twenty and forty-five minutes, while the lead from cash prices to futures prices, though significant, rarely extends beyond one minute. The length of the lead from futures to the index reflects, in part, inertia in the stock market. Stocks are not traded as frequently as futures contracts.

The lead/lag relationships are remarkably stable across the different days and futures contracts examined in 1984 and 1985. Interestingly, the lead from futures to the index on expiration day is at least as long as other days prior to expiration, suggesting that expiration days do not demonstrate a temporal character substantially different from earlier days.

In December 1990, Hans R. Stoll and Robert E. Whaley publish "The Dynamics of Stock Index and Stock Index Futures Returns"[13]. This paper's purpose was to model, empirically, the temporal relation between the price movements of index futures contracts and stocks. It was distinguished from prior papers because of two aspects: they have used a long time interval, five years, and a fine return grid, five minutes. Second, they have treated the delay in the price change of a stock index, due to infrequent trading of component stocks explicitly.

They have showed that, when the effects of infrequent trading and bid/ask spread are incorporated, observed portfolio returns follow an ARMA (p,q):

$$R^{o}_{S,t} \approx \omega_{S,0}\mu_S + \sum_{k=1}^{\infty}\phi_k R^{o}_{S,t-k} + \epsilon_{S,t} - \sum_{k=1}^{\infty}\theta_k \epsilon_{S,t-k}$$

Where the error term $\varepsilon_{S,t}$ contains three error components:

- the random error from an infrequent trading model

$$R^{o}_{S,t} = \sum_{k=0}^{n-1}\omega_{S,k}R^{*}_{S,t-k} + \nu_{S,t} .$$

- a weighted average error from the individual stock bid/ask spreads
- the true return innovation in the stock portfolio

The regressed model's equation was:

$$Z_{S,t} = \alpha + \sum_{k=-3}^{3}\beta_k R^{o}_{F,t-k} + u_t$$

Where Z_S is the cash market's instantaneous return rate while R_F is the future's one.

They have found that S&P 500 futures index leads the stock index by about five minutes on average, but occasionally as long as ten minutes or more, after the observed stock index returns have been purged of infrequent trading and bid/ask price effects. The futures returns indexes tend to lead even the return of the most actively traded stocks, from the cash index. They have also uncovered that this lead effect in not completely unidirectional. There is a weak positive predictive effect of lag stock index returns on current futures returns; however, this effect has grown smaller, as the markets have matured.

There was evidence that the futures market leads the cash market mainly because not all stocks of the index are traded continuously.

Another important article for this paper is "A Further Analysis of the Lead – Lag Relationship between the Cash Market and Stock Index Futures Market"[3] published in 1992 by professor Kalok Chan.

The paper focuses on two issues concerning the temporal relationship between futures and cash index returns: the first is whether the lead-lag relation is induced by the infrequent trading of component stocks. The second issue to be examined is why, if not because of non-synchronous trading, futures prices lead cash index prices.

He uses two set of data, both of about nine months, one from 1984 and one from 1987. The main model used has the following form:

$$R_{S,t} = a + \sum_{k=-3}^{3} b_k R_{F,t+k} + \epsilon_{s,t}$$

The results confirm previous findings that there is an asymmetric lead-lag relation between the two markets; there is strong evidence that the futures leads the cash index, and weak evidence that the cash index leads the futures. The results are robust even in 1987, when the cash market seems to be faster in processing market wide information. Several sets of results suggest that non-synchronous trading cannot completely explain why futures prices are dominant in leading the cash index. First, an asymmetric lead-lag relation holds between futures and all component stocks, even in 1984-1985 when some stocks are more frequently traded than the futures. Second, even for some stocks that are actively traded and have non-trading probabilities close to zero (e.g., IBM and AT&T), the returns still lag futures returns significantly. Therefore, the lead-lag relation is not well explained by non-synchronous trading.

It also finds that the asymmetric lead-lag relation between cash and futures markets can be attributed to two forces. First, the futures market is faster than all individual stocks in processing information. Second, futures prices seem to be better at reflecting market-wide information than cash index prices. Certainly, the two forces are interrelated. It may be that because the futures market is better at reflecting market-wide information, it leads all component stocks.

In recent studies, the frequency is of data used is increased to one, five or ten seconds resolution. Yiuman Tse, Paramita Bandyopadhyay and Yang-Pin Shen use one second resolution for the analyzed time series in their article "Intraday Price Discovery in the DJIA Index Markets"[14] published in 2006. Their main goal is to assess the relative influence of different traded contracts in price discovery and not necessary the lag between them.

The paper explores the dynamics of price discovery between the Dow Jones Industrial Average (DJIA) index and its three derivative products: the DIAMOND exchange-traded fund (ETF), the floor-traded regular futures, and the electronically traded mini futures. A mini futures contract is an electronically traded futures contract on the Chicago Mercantile Exchange that represents a portion of the normal futures contracts. For example, the E-mini S&P 500 futures contract is one-fifth the size of the standard S&P 500 futures contract. Advantages to trading E-mini contracts include liquidity, greater affordability for individual investors and around-the-clock trading.

They use the concept of information share to analyze the contributions of different markets to this efficient price in terms of the variance of innovations in the common factor. This information share (IS) model does not establish any price as the best price. Rather, it allows them to determine which entity moves first in the price adjustment process.

The basis of the IS model is the vector error correction model:

$$\Delta Y_t = \alpha\beta' Y_{t-1} + \sum_{j=1}^{k} A_j \Delta Y_{t-j} + e_t$$

where $Y_t = \{y_{it}\}$ is an n x 1 vector of co-integrated prices, α is the error correction matrix, β is a matrix of co-integrating vectors, and et is a zero-mean vector of serially uncorrelated innovations with covariance matrix $\Omega = \{\sigma_{ij}\}$.

They find that the Diamond ETF relatively dominates the price discovery process of the ETF shares. The results remain robust when they perform the analysis using the SPY (an S&P 500 ETF), the S&P 500 regular futures, and the E-mini futures. The conclusion is that the quotes of the SPY contribute more than the E-mini futures.

Among the three derivatives of the DJIA, the E-mini futures contribute the most to price discovery, 69.1%, while the ETF quotes of Diamond also contribute significantly, 28.6%. For the derivatives of the S&P 500, the ETF quotes from Diamond contribute about 49%.

III. General considerations and input data

The most important difference between this paper and the ones being mentioned earlier is the fact that I want to investigate the lead-lag relationship in the Romanian markets and not another foreign market.

Let us first mention the elements that are common to virtually all markets.

Both futures prices and cash index prices reflect the aggregate values of underlying stocks. Futures and cash prices will differ, however, because of differences in carrying costs. But if interest rates and dividend yields were non-stochastic, contemporaneous price changes in the two markets would be perfectly correlated and no lead-lag relation would exist between them. Various frictions, however, may cause one market to react faster to information than the other, so that the lead-lag relation is observed.

The use of stock futures typically falls into one of three categories:

- hedging, which involves the purchase or sale of index futures in anticipation of an intended cash market trade, whereby the hedge provides compensation for adverse price moves prior to the cash transaction,
- arbitrage, which involves the simultaneous purchase and sale of stocks and futures in order to capture realignments of relative prices following a perceived mispricing opportunity, and
- trading, which involves the active use of futures to speculatively take advantage of anticipated broad market price movements.

While arbitrage uses both cash and futures contracts, hedging and trading strategies normally incorporate only one type of instrument at any given time.

Futures prices normally vary relative to stock prices without triggering the arbitrage, so that arbitrage opportunities are available infrequently.

If the actual futures price is higher than the predicted value, the futures contract is overvalued, justifying the purchase of the stocks and the simultaneous sale of the futures contract. If the actual futures price is below the predicted value, the futures is undervalued and the reverse trade is initiated. Upon the convergence of the futures price to the actual stock value at the expiration of the contract, the arbitrager is assured of achieving some predetermined fixed-income return or fixed-rate cost of borrowing. It is possible that a liquidation opportunity may arise sooner if the futures price returns to its predicted value prior to expiration.

The futures-to-cash price differential, labeled the basis, normally falls within boundaries determined by financing costs adjusted for dividend uncertainty, transactions costs, and taxes. Because market interest rates have historically exceeded the dividend rate on common stocks, the stock index futures price normally exceeds the stock index value, and the basis is positive.

Two phenomena, market sentiment and arbitrage trading, are the major determinants linking stock index futures and the stock market on mature markets. The conventional wisdom among professional traders, in the US, is that movements in the futures price reflect market expectations of subsequent movements in cash prices. The futures price presumably embodies all available information regarding events that will affect cash prices and responds quickly to new information. Stock price movements may similarly convey information regarding subsequent price variation in the futures contract. It is unlikely, however, that the relation-ships are symmetric.

Consider a trader reacting to new information on the health of the economy. If the information is bullish, a trader has the choice of buying either the futures contract or the underlying stocks. While the futures trade can be executed immediately with little up-front cash, actual stock purchases require a greater initial investment and may take longer to implement as they involve a subsequent stock selection and numerous individual stock transactions. All of the reasons above create investor preference for futures and explain why changes in futures prices lead changes in stock prices in the papers that describe the US market. Futures prices may thus provide a sentiment indicator of forthcoming cash prices, which follow when investors who are unwilling or unable to use futures incorporate the same information into their cash market transactions.

Changes in the stock prices may also lead changes in the futures price as the value of the underlying stock represents part of the information that affects futures prices. Futures traders likely incorporate recent changes in the index in their pricing decisions. Put another way, if the index were to decline or rise for whatever reason, the price change might induce a change in sentiment that would be reflected in subsequent declines or increases in futures prices.

As long as the basis lies within the no-arbitrage trading range, changes in market sentiment would affect both futures prices and the index in the same direction. If the basis varies outside the no-arbitrage range, however, arbitragers would take opposite positions in the two markets so that the basis would ultimately approach its predicted value. This adjustment could arise because both prices move in a common direction, with one price moving more rapidly than the other, or because the two prices move oppositely. Regardless of which occurs, the lead-lag relationship during periods when arbitrage activity is present might reasonably be expected to differ from the lead-lag relationships present when no arbitrage activity occurs.

Most of the studies done on this subject focus on the S&P 500 index, as a representative for the cash market, and S&P 500 futures contract. While the liquidity of the futures contract is very good, the lag induced by the non-simultaneous trading of the component S&P 500 stocks represents a problem.

As we have mentioned before the Bucharest Stock Exchange is the only Romanian market that trades in stocks. It has several indexes: BET, BET-C and BET-FI. Until September 2007, there weren't available any futures contracts having as underlying asset a stock index. At that time, the Bucharest Stock Exchange began trading futures contracts based on the BET index. The liquidity of the futures market, introduced by BVB, has been very low; there have been a lot of days, in the analyzed time interval, when no trades have been made.

The main market for Romanian futures contracts is Sibiu Monetary - Financial and Commodities Exchange (BMFMS, the Romanian abbreviation). Because of the fact that it doesn't have any important leadership ties with the BVB, from a management perspective, BMFMS never introduced a futures contract with a BVB stock index as underlying asset.

Thus, the only chance to make a relevant analysis on the lead-lag relationship lies in using a stock from the BVB and its futures contract listed on the BMFMS.

During the privatization process, World Bank representatives suggested the development of financial entities, resembling investment funds, which would serve as a vehicle to divide the country's assets to the individuals. This is how the SIFs were created. Nowadays, the five SIF entities, numbered from one to five, lead the transaction value charts most of the time, both on the BVB and the BMFMS.

The SIFs were created by law and have a distinct trait: no entity or group of entities that act with concerted actions may hold more than 1% out of any SIF. This is a factor that influences stock perception. Unlike any other stock that has a majority holder these ones don't have one in the traditional sense. This is why the fund's behavior is not influenced by the collateral interests of the main shareholder.

For this analysis I have chosen the SIFs with the most important liquidity on both cash market and futures market: SIF 2 and SIF 5.

Being practically investment funds, the two SIFs have minority or majority stakes in other companies. SIF 5 has stakes in 257 companies, while SIF 2 in 371. Most of these companies are listed therefore, the results obtained on these equities are as close as possible, for the Romanian market, to the ones obtained on the S&P 500. There are listed companies where both the SIFs hold stakes, but the most of the capital is invested in different companies.

The SIF2 or SIF5 futures contract (that will be referred to from now on as DSIF2 and DSIF5) represents the purchase or sale of 1000 underlying stocks.

Before accessing the market, traders must post an initial margin deposit or collateral equal to only a fraction of the stocks' market value. Futures prices change intermittently throughout each trading day. At day's end, there is a marking to market of the contract position, whereby traders must cover any losses when prices move against them or may withdraw any profit in excess of their initial margin requirement when prices move favorably.

Based on previous studies the market that is cheaper to trade and that has more liquidity is the prime candidate to lead the price discovery process.

One very important aspect, that is present in the US market and absent on the Romanian one, is the impossibility of short-selling stocks on the BVB. Short selling involves the sale of securities borrowed from brokers who, in turn, usually borrow them from third party investors. The short seller pays a negotiated fee for the privilege and has to "cover" his position: to re-acquire the securities he had sold and returns them to the lender (again via the broker). This allows her to bet on the decline of stocks he deems overvalued and to benefit if she is proven right: he sells the securities at a high price and re-acquires them once their prices have, indeed, declined. If pension funds and institutional businesses were not generally long term holders of securities, then the arbitrageurs would not have taken advantage of this. In Romania, pension funds are now in the process of acquiring enough money to be able to meet such a role.

Many economists insist that short selling is a mechanism which stabilizes stock markets, reduces volatility, and creates incentives to correctly price securities. Under all other conditions equal, the lack of short-selling should lead to an increased role in the price discovery process of the traded futures contracts. This effect would propagate better into the Romanian market should the local market have a similar investor (both individual and institutional) structure.

The margin required to buy/sell a contract is 350 RON for DSIF2 and 450 RON for DSIF5. This margin needs to be augmented or decreased, according to the price movements of the stock, at the end of the day, as mentioned before.

As it is expected, the commissions required for the futures contracts are much lower compared to the ones required for equity trading. The funds required to make the stock purchase is close to ten times as large as the ones required to enter into a SIF futures contract.

Several factors can influence how fast the cash and futures markets reflect information, and thus affect the lead-lag relation. This paper tries to examine multiple situations and to reflect how these special conditions can influence the lead- lag relationship between the data series.

Ticker	Number of stocks (contracts)	Margin (RON)	Cost (RON)	Commission (RON)	Total funds required (RON)
SIF2	1000	-	3000	9	3009
SIF5	1000	-	4000	12	4012
DSIF2	1 (equivalent 1000 stocks)	350	-	0.6	350.6
DSIF5	1 (equivalent 1000 stocks)	450	-	0.6	450.6

Table 1. Trading costs of considered stocks and futures

One factor that influences this relationship is short-sale constraints. Diamond and Verrecchia (1987)[5] show that prohibiting traders from shorting slows the adjustment of prices to private information, especially with respect to private bad news. Since on the BVB cash market it is not allowed to short-sell, this adjustment of prices should be slower in this market than on the BMFMS.

Because of short-sale constraints in the cash market, there should be noticeable a difference in lead-lag relation under bad news or good news. In a bullish market the lag interval should be smaller, while in a bearish market the same interval should be even larger.

Another factor that influences the lead-lag relationship is the intensity of trading in the two markets. Lower trading activity implies that the securities are less frequently traded, so observed prices lag "true" values more. Also, information dissemination may be related to the intensity of trading activity. Admati and Pfleiderer (1988)[1] show that, in general, trades of both discretionary liquidity traders and informed traders cluster, with each group preferring to trade when the market is thick. The clustering of trades causes more information to be released when trading activity is higher. Therefore, the lead-lag relation is expected to vary with the relative intensity of trading activity in the two markets. Another paper, Stephan and Whaley (1990)[12] study the intraday relation between the stock market and the stock option market. They find that not only do price changes of stocks lead price changes of options, but that trading activity (proxied by the number of transactions and trading volume) in the two markets also bears the same kind of lead-lag relation. This provides evidence that price discovery and trading activity are related.

Due to all previously mentioned factors, this paper chooses to analyze their effects separately. Although the short-selling constraint is very important in cash market evolution and implicitly in the lead-lag relationship, it is almost impossible to quantify its importance because there are no stocks or indexes that can be sold short on any Romanian market.

1. Data

This paper makes use of one data interval, from the 2nd of August 2007 to the 14th of March 2008. This choice has been made for two main reasons:

- Market liquidity is a problem in Romanian markets, and as we will show later, low liquidity influences results considerably. The chosen interval has a high average trading volume both on the cash market and in the futures market. There are available statistics regarding trading volumes and trading probabilities for the stocks analyzed further on.
- In the close past, Romanian market began to be more closely correlated to external markets, mostly New York and London. Thus, the period considered was able to capture a bull market, in the first few months, and a bearish one towards the end; the bearish market was an effect of the globally propagated subprime crisis. This should provide a more reliable result, as it is verified in very different market conditions.

Data was obtained directly from the BVB and the BMFMS servers in the shape of all transactions made for the period. This data contained the date, the time, the volume, the markets, and the settle price from all transactions.

The BVB has several markets:

- REGS - is the regular market, where the most general-type transactions take place. Most of the information gathered in the data series comes from this market. Its only restriction is a certain number of minimum stocks traded on any given transaction. This minimum number is 100 shares for SIF2 and SIF5, in the neighborhood of 100 US$, thus being a very light restriction.
- DEAL- is the market for large transactions. The costs associated with trading on this market are a couple of times larger than REGS costs therefore it is generally avoided. Since the SIFs have the largest liquidity on the BVB, the price should reflect very closely market expectations, hence considerable settle price changes,

from REGS values, are rare. Transactions on this market have been included for a very important reason: although very rare, transactions on the DEAL market are justified only when large quantities are exchanged at a considerable price difference. Because we are talking about very large investors, these trades usually contain strong new information for the market, so the price from the DEAL's transactions is quickly incorporated into the REGS market serving as a target price for future transactions. Since the interval to price incorporation into the REGS market may vary from a few seconds to a few minutes, and since the futures traders are able to see real-time the DEAL transactions, using only the REGS market would not show the real lead-lag relationship that we are studying.

- ODDS – is a market for smaller than 100 stock trades. Its settle price is close to the REGS market. It has also been included in the data series for a completeness reason; because of small volumes and because of the method used for data aggregation it does bear a small role in final price/ instantaneous rate of return results.

The BMFMS has only one market for futures contracts called DEAL. There wasn't any need for data aggregation in this case.

Trading hours are different for the two institutions:

- The BVB opened trading, for the analyzed period, at 10:00 and closed it at 14:15 local time. Only the REGS market has a pre – open period, half an hour from 9:30 to 10:00, and a pre-close period, 15 minutes from 14:15 to 14:30. During these periods investor can enter orders in the market but they will only be executed in the first interval. Since there is no equity being exchanged pre-close and pre-open do not offer relevant information for this paper.
- The BMFMS opened trading at 10:00 and closed at 16:00. Settle price for the futures contracts being traded after BVB closes are influenced mainly by very recent news and pre-open status of the US markets. The NYSE opens at 16:30, local time.

Considering the fact that the two markets do not have the same trading hours, the data series have been made similar. This paper took into consideration only trades in the futures market that took place before the BVB closing, meaning before 14:15 local time.

For the Romanian market, at this point in time, there aren't available price series, for the stocks or futures considered, distributed in time according to a specific period (shorter than a day). To solve this problem, all the individual data transactions have been introduced in a database and aggregated to reflect a minute by minute price. This was accomplished with the help of a weighted average. The price of a stock and the exchanged equity volume, for a specific minute was determined using these two formulas:

$$P_t = \frac{p_1 * v_1 + \cdots + p_n * v_n}{v_1 + \ldots + v_n} \qquad (1)$$

$$V_t = \sum_{i=1}^{n} v_i \qquad (2)$$

Where P_t and V_t are the price and the volume of stocks traded in the t minute. p_i and v_i are the prices and the volumes for the i trades that have been executed in that specific minute.

If there are no trades in a one minute interval, then the price value for that minute is equal to the price calculated in the previous interval.

Chan (1992) chose to use, as price for a time interval, the last price in that interval. I believe that my method is better suited for these data series because of two particularities in the Romanian market:

- On the BVB it is possible to use hidden volume orders, meaning that in the market you appear as having a buy or sell hidden order of unknown volume. Traders very often test these hidden orders with minimum volume trades. Obviously these trades change the price (they are used to test a bid or ask level, different from the one that took part in the last trade). These trades do not reflect new information entering the market, they are only technical tools used by the traders. Using a volume weighted price calculation model the effect of these trades is kept to a minimum.
- On the BMFMS, trades that involve only a small number of contracts are usual (only 1- 4 contracts/trade). Also, since there is a large number of small package contracts in the market, when a higher volume order is executed (market order), it will buy/sell at the desired level (most of the order) but also touch the next bid/ask levels introducing a semi-artificial price change. Using the same weighted average method we are sure to obtain a reliable set of observations.

From the many hundreds of thousands of trades, after the first aggregation we are left with around 45.000 observations. Each day will have 255 observations, in accordance to the 255 minutes of trade, each day on the BVB. From these price series we calculate the instantaneous rate of return using the formula:

$$R_t = \ln(P_t) - \ln(P_{t-1}) \qquad (3)$$

The instantaneous rate of return data series is named “the minute series” and is used directly in the regression model.

Chan (1990) used the five minutes aggregated instantaneous rate of return data series for estimation. Our hope is that using the minute series we will be able to establish a more precise relationship. In order to be able to compare the results this paper also includes an estimation of the model for a five minutes time scale.

The data series for the five minutes intervals has been build differently. We use as starting point the minute data series and not the initial pool of transaction data. The five minutes instantaneous return series is obtained by retaining the last price change in that interval. This way, the new data series has better relevance because five minutes is a long enough period so that enough trades are made to be able to include the new information into the price of the equity. Also, we have already compensated in the minute series for the low trading volumes effects mentioned earlier.

Obviously, the volume for the five minutes data series is obtained by adding the volumes in the minute series for that interval.

2. Preliminary statistics

The most important aspects for reliable price discovery statistics are trading frequency and volume. A high trading frequency means that any piece of information is integrated into the stock price very fast while large trading volumes separate the trades into trades with a solid motive for execution or less thought of sell decisions. Should we be able to attach a relevance coefficient to each trade, in a study about lead-lag relationships, it would definitely be larger for high volume trades than for low volume trades.

In the tables below such statistics are calculated. First are the tables for the minute data series (45.135 entries).

Ticker	Average volume (in stock no or no of contracts, case by case)	Average price (RON)	Trading probability (%)	Non-price change probability (%)
SIF2	6665.05	3.32	51.49	72.26
SIF5	9101.66	4.03	54.8	69.42
DSIF2	15.47	3.37	57.2	53.84
DSIF5	14.21	4.15	53.48	56.77

Table 2. Statistics for minute data series

Statistics about the five minutes series are also presented but, since the aggregation method is different and the minute data series have been used as inputs, table 2 contains the most relevant results. The most traded stock is SIF5, and when taking into consideration the price differential it almost has a double traded volume compared to SIF2. For all series, the trading probability stands at about 50%, meaning that there is a chance in two that a stock/futures contract is traded in a specific minute. The non-price change probability is the probability that a trade (or lack of a trade in that specific second) keeps the quote price unchanged.

One more difference between stock trading and futures trading is the possibility, in the futures case, of setting the ASK/BID price with four decimals. In stock trading, investors are only allowed to use two decimals. For this reason, if a futures contract is sold or bought there is almost a very high chance (almost 90% for both contract types) that there will also be a price change. This type of price changes are usually very small so the "noise" introduced in the instantaneous return series is just as small. For the two BVB quoted stocks, if a trade occurs, there is an almost 50% price change probability.

It is worth mentioning that, just as we predicted earlier, the futures prices are usually above equity prices. This happens not only because of technical reasons but also because of the optimism of the Romanian trader. The data begins to be collected after several years of wide market growth that was in tune with the higher than average growing internal economy.

Ticker	Average volume (in stock no or no of contracts, case by case)	Average price (RON)	Average traded value (RON)	Trading prob. (%)	Non-price change probability (%)
SIF2	33325	3.32	110,804	93.08	41.56
SIF5	45508	4.03	183,375	94.5	39.83
DSIF2	77.36	3.38	261,465	92.26	17.69
DSIF5	71.07	4.15	294,976	90.67	20.04

Table 3. Statistics for five minutes data series

Looking over the results as a whole, it is easier to believe that the SIF5-DSFIF5 applied models will be more relevant than the ones for SIF2-DSIF2, since higher trading volume and higher trading probability should reduce latencies and increase information density in the data series.

The traded value for the futures contracts has been calculated not in cash exchanged but in equivalent stock value “exchanged”. For the five minutes data series we also find that: trading volume SIF2/SIF5 = 73% while DSIF2/DSIF5 = 109%; traded value SIF2/SIF5 = 60% while DSIF2/DSIF5 = 89%. It remains probable the fact that the SIF5/DSIF5 data series should hold more information than the SIF2/DSIF2 ones.

Compared to the minute series there is a doubling in trading probability which comes close to the 100% mark. At the same time the non-price change probability for the series reduces considerably. There are the important facts that need to be considered when analyzing the results.

The data aggregation was done by retaining the last price from the five minute interval. I chose this method to obtain data as close as possible to the one used by Chan (1992) so that a comparison between the results would be relevant. Using a five minutes data interval for the model regression makes hardly noticeable any lead-lags of one to three minutes. The method of aggregation also excludes most of the information available inside the five minutes interval. Considering the trading statistics, the Romanian market specific elements and the more recent studies, it is probable that these data series will yield less reliable results.

<table>
<tr><th>SIF2,DSIF2(-i)</th><th>SIF2,DSIF2(+i)</th><th>i</th><th>lag</th><th>lead</th></tr>
<tr><td>|** |</td><td>|** |</td><td>0</td><td>0.2351</td><td>0.2351</td></tr>
<tr><td>|* |</td><td>|* |</td><td>1</td><td>0.1229</td><td>0.1095</td></tr>
<tr><td>| |</td><td>|* |</td><td>2</td><td>0.0335</td><td>0.0527</td></tr>
<tr><td>| |</td><td>| |</td><td>3</td><td>0.0131</td><td>0.0307</td></tr>
<tr><td>| |</td><td>| |</td><td>4</td><td>0.0176</td><td>-0.0007</td></tr>
<tr><td>| |</td><td>| |</td><td>5</td><td>0.0008</td><td>0.0071</td></tr>
<tr><td>| |</td><td>| |</td><td>6</td><td>0.0053</td><td>0.0057</td></tr>
<tr><td>| |</td><td>| |</td><td>7</td><td>0.0105</td><td>0.0001</td></tr>
<tr><td>| |</td><td>| |</td><td>8</td><td>0.0061</td><td>-0.0127</td></tr>
<tr><td>| |</td><td>| |</td><td>9</td><td>0.0050</td><td>0.0033</td></tr>
<tr><td>| |</td><td>| |</td><td>10</td><td>0.0025</td><td>-0.0064</td></tr>
</table>

Fig1. Cross-correlogram of SIF2 and DSIF2 (minute data series)

Kawaller and Koch (1987) and Chan (1992) show that in the early stages of developing a model, it is useful to compute the cross-correlation function between the data series to identify the empirical dynamic relationships. Such cross-correlation analysis reveals the first hints towards the final results.

<table>
<tr><th>SIF5,DSIF5(-i)</th><th>SIF5,DSIF5(+i)</th><th>i</th><th>lag</th><th>lead</th></tr>
<tr><td>| |</td><td>| |</td><td>0</td><td>0.0126</td><td>0.0126</td></tr>
<tr><td>| |</td><td>| |</td><td>1</td><td>0.0299</td><td>0.0036</td></tr>
<tr><td>| |</td><td>| |</td><td>2</td><td>0.0128</td><td>0.0056</td></tr>
<tr><td>| |</td><td>| |</td><td>3</td><td>0.0000</td><td>-0.0033</td></tr>
<tr><td>| |</td><td>| |</td><td>4</td><td>-0.0045</td><td>0.0074</td></tr>
<tr><td>| |</td><td>| |</td><td>5</td><td>0.0009</td><td>0.0008</td></tr>
<tr><td>| |</td><td>| |</td><td>6</td><td>0.0046</td><td>0.0011</td></tr>
<tr><td>| |</td><td>| |</td><td>7</td><td>-0.0010</td><td>-0.0038</td></tr>
<tr><td>| |</td><td>| |</td><td>8</td><td>-0.0006</td><td>0.0024</td></tr>
<tr><td>| |</td><td>| |</td><td>9</td><td>-0.0018</td><td>-0.0010</td></tr>
<tr><td>| |</td><td>| |</td><td>10</td><td>0.0010</td><td>-0.0025</td></tr>
</table>

Fig2. Cross-correlogram of SIF5 and DSIF5 (minute data series)

For the minute data series, the cross-correlograms do not contain strog information. The noise in the futures series and the considerable non-price change probability differential influences the results. There is a rather unreliable hint, considering the lead-lag probabilities, that the cash market integrates information faster than futures markets.

<table>
<tr><th>SIF2,DSIF2(-i)</th><th>SIF2,DSIF2(+i)</th><th>i</th><th>lag</th><th>lead</th></tr>
<tr><td>|****** |</td><td>|****** |</td><td>0</td><td>0.5909</td><td>0.5909</td></tr>
<tr><td>| |</td><td>|* |</td><td>1</td><td>0.0311</td><td>0.0576</td></tr>
<tr><td>| |</td><td>| |</td><td>2</td><td>0.0358</td><td>-0.0104</td></tr>
<tr><td>| |</td><td>| |</td><td>3</td><td>0.0003</td><td>-0.0196</td></tr>
<tr><td>| |</td><td>| |</td><td>4</td><td>0.0016</td><td>0.0029</td></tr>
<tr><td>| |</td><td>| |</td><td>5</td><td>0.0016</td><td>-0.0046</td></tr>
<tr><td>| |</td><td>| |</td><td>6</td><td>0.0143</td><td>-0.0191</td></tr>
<tr><td>| |</td><td>| |</td><td>7</td><td>0.0096</td><td>0.0065</td></tr>
<tr><td>| |</td><td>| |</td><td>8</td><td>-0.0037</td><td>-0.0068</td></tr>
<tr><td>| |</td><td>| |</td><td>9</td><td>-0.0208</td><td>-0.0131</td></tr>
<tr><td>| |</td><td>| |</td><td>10</td><td>0.0062</td><td>0.0101</td></tr>
</table>

Fig3. Cross-correlogram of SIF2 and DSIF2 (five minutes data series)

<table>
<tr><th>SIF5,DSIF5(-i)</th><th>SIF5,DSIF5(+i)</th><th>i</th><th>lag</th><th>lead</th></tr>
<tr><td>|* |</td><td>|* |</td><td>0</td><td>0.1384</td><td>0.1384</td></tr>
<tr><td>| |</td><td>| |</td><td>1</td><td>0.0073</td><td>0.0195</td></tr>
<tr><td>| |</td><td>| |</td><td>2</td><td>0.0150</td><td>-0.0081</td></tr>
<tr><td>| |</td><td>| |</td><td>3</td><td>0.0044</td><td>-0.0149</td></tr>
<tr><td>| |</td><td>| |</td><td>4</td><td>0.0328</td><td>0.0002</td></tr>
<tr><td>| |</td><td>| |</td><td>5</td><td>-0.0019</td><td>-0.0071</td></tr>
<tr><td>| |</td><td>| |</td><td>6</td><td>0.0125</td><td>0.0064</td></tr>
<tr><td>| |</td><td>| |</td><td>7</td><td>0.0086</td><td>-0.0079</td></tr>
<tr><td>| |</td><td>| |</td><td>8</td><td>0.0003</td><td>-0.0019</td></tr>
<tr><td>| |</td><td>| |</td><td>9</td><td>0.0005</td><td>-0.0130</td></tr>
<tr><td>| |</td><td>| |</td><td>10</td><td>-0.0303</td><td>0.0035</td></tr>
</table>

Fig4. Cross-correlogram of SIF5 and DSIF5 (five minutes data series)

The results for the five minutes series confirm the ones for the minute data series. Apparently, on these data series, the correlogram is not able to accurately separate information from the background noise. In the annex there are also available the cross-correlation tables for minute data series and the correlograms for the data series, but only considering 254 recordings (one trading day). As unreliable as they are, these cross-correlograms show the first main difference between the Romanian and US market: stock prices appear to react faster to new information compared to futures.

IV. Methodology and results

This paper tries to paint a more complete description of the lead-lag relationship and to achieve this purpose it uses two types of regression models. The first one, which coincides with the model used by Chan (1992), tries to use a top-down approach: the data series are very long (minute and five minute data frequency along with eight months of recorded data) so that that the results, if any, should be persistent and consistent. The disadvantage of this model is that the estimates are put under great pressure, from an econometric point of view, because of many and different perturbations they must compensate for in such a large interval.

The second model, proposed by I. Kawaller, P. Koch and T. Koch (1987), is a bottom-up approach. Data series used are as short as one trading day, with the same high frequency. The advantage of this model is the fact that the results should be more obvious; however it is tested only for some trading days and not the whole period. Using random days for the regression along with some specific condition days should yield relevant results for the entire period.

1. Top-down approach

The model used for this approach is a regression model that uses a variable number of lags for the independent variable:

$$S_t = a + \sum_{i=-c}^{+c} b_i * F_{t+k} + \varepsilon_{St} \quad (4)$$

Where S_t is the equity's instantaneous return rates, F_{t+k} is the futures contract's instantaneous return rates and c is the number of lags considered. Depending on data frequency, the number of lags increases or decreases in order to make the model encompass all of the necessary data. The coefficients with negative subscripts (b_{-i}) are lag coefficients, and those with positive subscripts (b_{+i}) are lead coefficients. If the lag coefficients are significant, the cash index lags futures. If the lead coefficients are significant, the cash index leads futures.

Use of the terms "lead" and "lag" does not necessarily mean that price movement in one market causes price movement in the other market. It is more appropriately interpreted as one market reacting faster to information than the other market, which lags and then catches up.

For a number i of lags used in the model, in order to be able to estimate it, the data series is shortened with 2*i entries. Since the data series for these models has many thousand recordings (45.135 for the minute series), this aspect bears no effect on the outcome.

Another difference from Chan's method of processing primary data is the way end-of-day, beginning-of-day data is treated. Chan's data series doesn't include the instantaneous rate of returns calculated from market closing price to market opening price. The main reason I chose to include these recordings in the data series is because of the existence of BVB market pre-open period. In the last few minutes, before the cash market opens, the opening price is obvious to investors. Therefore a futures trader already has this information at the moment when he is able to execute his first futures order on the newly opened market. My choice would be less justified, even detrimental, in the case when the futures market would react faster to information than cash market. At the same time, should the opposite occur, and considering the fact that the price differences are usually larger for the period between market close and open, useful information is added to the data series.

The model used has only one purpose: to estimate the intraday relation between listed equity prices and futures prices and does not investigate the variability of the disturbances. It also does not aim to explain the behavior of a variable completely (meaning that a high value of R^2 in the regression models is not really the goal) but only to correctly estimate the value and signs of the coefficients that tie different lags of futures and equity data series.

Based on previous evidence, the error terms (ε_t) in regression (4) are likely to be time-varying heteroskedastic. The dynamics of the conditional variances are not explicitly modeled in this article. However, since heteroskedasticity generally leads to inconsistent estimates of standard errors and invalidates inference, all of the t-ratios for the coefficients are adjusted using the Newey- West HAC (Heteroskedasticity and Autocorrelation Consistent Covariances). The main problem is that the present heteroskedasticity is not known most of the time.

When the form of heteroskedasticity is not known, it may not be possible to obtain efficient estimates of the parameters using weighted least squares. OLS provides consistent parameter estimates in the presence of heteroskedasticity, but the usual OLS standard errors will be incorrect and should not be used for inference.

Using the Newey-West HAC consistent covariance estimates does not change the point estimates of the parameters, only the estimated standard errors.

The White covariance matrix is not used because it assumes that the residuals of the estimated equation are serially uncorrelated. The tests that have been applied to the input data show that there is serial correlation in the estimation residuals. Newey and West (1987) have proposed a more general covariance estimator that is consistent in the presence of both heteroskedasticity and autocorrelation of unknown form.

The values of the calculated coefficients are not modified but, using HAC, we can now trust their relevance probability. No other methods for serial correlation in the residual series or heteroskedasticity elimination have been used. The main obstacle was the unknown form of heteroskedasticity in data series present.

A) High frequency data

The first set of regressions is done on the minute series. Although Chan (1992) used only the five minutes series, recent studies (for example Joel Hasbrouck 2001)[8] use higher frequencies data. Electronic trading, the internet and the development of mobile communications have brought a faster news response time even in trading.

In figures 5 and 6 the results of the regressions are presented.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-1.57E-06	4.95E-06	-0.317008	0.7512
DSIF2(-10)	0.006501	0.005537	1.174031	0.2404
DSIF2(-9)	0.010414	0.006836	1.523349	0.1277
DSIF2(-8)	0.009483	0.007052	1.344794	0.1787
DSIF2(-7)	0.014294	0.005861	2.438847	0.0147
DSIF2(-6)	0.010287	0.007348	1.399910	0.1615
DSIF2(-5)	0.005146	0.008442	0.609576	0.5421
DSIF2(-4)	0.026680	0.009410	2.835314	0.0046
DSIF2(-3)	0.018238	0.008439	2.161056	0.0307
DSIF2(-2)	0.036748	0.014357	2.559586	0.0105
DSIF2(-1)	0.145417	0.041355	3.516311	0.0004
DSIF2	0.280706	0.035940	7.810422	0.0000
DSIF2(1)	0.128021	0.019344	6.617981	0.0000
DSIF2(2)	0.060112	0.012772	4.706442	0.0000
DSIF2(3)	0.039200	0.008261	4.744956	0.0000
DSIF2(4)	0.003617	0.008218	0.440188	0.6598
DSIF2(5)	0.012955	0.008317	1.557609	0.1193
DSIF2(6)	0.011634	0.006907	1.684204	0.0921
DSIF2(7)	0.001525	0.006793	0.224579	0.8223
DSIF2(8)	-0.012394	0.006385	-1.941002	0.0523
DSIF2(9)	0.008695	0.005988	1.452213	0.1464
DSIF2(10)	-0.004441	0.006262	-0.709202	0.4782

R-squared	0.087494	Mean dependent var	-7.57E-06
Adjusted R-squared	0.087070	S.D. dependent var	0.001845
S.E. of regression	0.001763	Akaike info criterion	-9.842889
Sum squared resid	0.140188	Schwarz criterion	-9.838639
Log likelihood	222053.0	F-statistic	205.8899
Durbin-Watson stat	2.312463	Prob(F-statistic)	0.000000

Fig 5. Regression results for minute series, SIF2 and DSIF2 variables, i = 10

For SIF2/DSIF2 the most important coefficient is the contemporaneous one at 0.28. This suggests that the market responds simultaneously to most of the information available. In this case, the lead coefficients (DSIF (+i)) are the most important. Their values come down from 0.128 to 0.039 for the third lead. Only the first lag coefficient is relevant at a 1% significance level.

The fact that the lag coefficient (meaning only one minute delay) is relevant has two main reasons:

- Usually in the market, when the price increases or decreases it is followed by a correction/rebound of a smaller magnitude. This is caused by the traders that try to take advantage of the new price value, selling or buying at least the quantity in the first level of bid/ask, thus touching the second level.
- The second reason, with a much lighter influence is a feed-back relationship that can occur at times from the futures market to the cash market. Investors that trade in both markets might be tempted to take the opposite action, compared to the momentary trend, based on information from the futures market. This relation can exist because there may be high volume, informed trades, on the futures market while the cash market trades at very low volumes.

The rest of the calculated coefficients are small, so that the standard error is very large compared to their values. The significant lead coefficients tend to decrease by half from one to the next. This shows that the quantity of information that is integrated with delay decreases exponentially.

The Durbin-Watson statistic has low relevance since the model it is applied on is not well suited. Autocorrelation in the residuals has already been taken into consideration and eliminated with the help of Newey-West method.

One more aspect, that is even more important for the SIF5/DSIF5 series, is the value of the R^2 coefficient. They are both small showing that the model doesn't explain very well the behavior of the endogenous variable. We are not interested in explaining the behavior, but only to analyze the lead-lag relation.

The important result is the fact that the cash market seems to react faster to information than the futures market. This is an unexpected result, based on the other studies available on the

US market, but understandable when considering the before mentioned Romanian markets' specific attributes.

Null Hypothesis:	Obs	F-Statistic	Probability
DSIF2 does not Granger Cause SIF2	45132	491.476	0.00000
SIF2 does not Granger Cause DSIF2		302.786	1.E-194

Fig 6. Granger causality test for SIF2/DSIF2, minute series

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-9.56E-06	7.36E-06	-1.300017	0.1936
DSIF5(-10)	0.002306	0.001922	1.200314	0.2300
DSIF5(-9)	0.002160	0.002291	0.942742	0.3458
DSIF5(-8)	0.004263	0.002809	1.517819	0.1291
DSIF5(-7)	0.006042	0.003064	1.971668	0.0487
DSIF5(-6)	0.010024	0.003413	2.937171	0.0033
DSIF5(-5)	0.008850	0.003572	2.477536	0.0132
DSIF5(-4)	0.007396	0.005109	1.447687	0.1477
DSIF5(-3)	0.012005	0.008185	1.466725	0.1425
DSIF5(-2)	0.020716	0.010446	1.983042	0.0474
DSIF5(-1)	0.029999	0.011844	2.532844	0.0113
DSIF5	0.021895	0.008162	2.682423	0.0073
DSIF5(1)	0.015162	0.005299	2.861409	0.0042
DSIF5(2)	0.013516	0.004197	3.220364	0.0013
DSIF5(3)	0.009444	0.003232	2.922424	0.0035
DSIF5(4)	0.013189	0.003630	3.633279	0.0003
DSIF5(5)	0.008539	0.003066	2.784522	0.0054
DSIF5(6)	0.006723	0.002740	2.453653	0.0141
DSIF5(7)	0.003235	0.002753	1.175063	0.2400
DSIF5(8)	0.004352	0.002595	1.676952	0.0936
DSIF5(9)	0.002202	0.002380	0.925151	0.3549
DSIF5(10)	0.000857	0.002062	0.415431	0.6778

R-squared	0.003217	Mean dependent var	-1.09E-05
Adjusted R-squared	0.002752	S.D. dependent var	0.001781
S.E. of regression	0.001778	Akaike info criterion	-9.825931
Sum squared resid	0.142585	Schwarz criterion	-9.821680
Log likelihood	221670.4	F-statistic	6.929250
Durbin-Watson stat	2.139038	Prob(F-statistic)	0.000000

Fig 7. Regression results for minute series, SIF5 and DSIF5 variables, i = 10

Since the lead coefficients are relevant we conclude that SIF2 reacts, on the whole period analyzed and at a one minute frequency, faster to new information than DSIF2. Not only that, but we can establish the lead to 3 minutes on average. The fifth and sixth coefficients are somewhat larger, but close to the standard error, showing that sometimes the lead extends further than 3 minutes but not very often. This may be the case because there are periods with very low to no liquidity even in the futures market.

The Granger causality tests yield an interesting result. For both regressions, the null hypothesis is rejected at an extremely low significance level. At the same time it points out that both cash market representatives Granger cause their futures counterparts, while the reverse still hold true. It is important to note that the statement "x Granger causes y" does not imply that y is the effect or the result of x. Granger causality measures precedence and information content but does not by itself indicate causality in the more common use of the term.

Null Hypothesis:	Obs	F-Statistic	Probability
DSIF5 does not Granger Cause SIF5	45130	14.8078	1.5E-14
SIF5 does not Granger Cause DSIF5		6.11146	1.2E-05

Fig 8. Granger causality test for SIF2/DSIF2, minute series

For SIF5/DSIF5 the results are even more conclusive. Although the R^2 is very low, the correct length of the lead/lags should hold true, and the values of the coefficients can be compared from a relative perspective (we take into account their proportionality).

At the same 1% relevance level, there are now five relevant leads. It appears that for these series the lead of the cash market extends to five minutes. The sixth's coefficient's probability is also close to the 1% level proving that the lead extends at times.

The first order lag coefficient is also barely not relevant showing that the same effects of correction/rebound and feed-back still apply, but to a lesser degree. The main difference from SIF2/DSIF2 is the fact that, as the leads increase, the coefficients decrease but by a smaller margin. Since SIF5 has an almost double traded value compared to SIF2, the differences in cash market prices are lower; more intermediary price levels are touched. The price information is incorporated into the futures market with smaller jumps, hence the relative grouping of the coefficients.

The difference in lead length between the two equities may be cause by the fact that DSIF2 has greater liquidity than DSIF5. Corroborated with the lower relative SIF2 liquidity, the equity price can be faster incorporated in the futures one.

Overall, there is evidence that there is a lead-lag relation between the cash market and the futures market, regardless of whether SIF2 or SIF5 futures are used. Further, the lead-lag relation is asymmetric – the feedback from the cash market into the futures market is higher than the reverse.

B) Medium frequency data

For these regressions we use the five minutes data series. The aggregation method has been the same with the one used by Chan (1992), however the input data was not raw data but rather the minute series. This choice was made because of a low stock trading probability and because of the many minimal volume trades that have the potential to strongly influence the results.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-7.54E-06	2.03E-05	-0.371848	0.7100
DSIF2(-3)	0.002337	0.010259	0.227793	0.8198
DSIF2(-2)	0.053389	0.011717	4.556689	0.0000
DSIF2(-1)	0.057681	0.016815	3.430279	0.0006
DSIF2	0.634183	0.041015	15.46234	0.0000
DSIF2(1)	0.083106	0.014183	5.859335	0.0000
DSIF2(2)	0.004253	0.009551	0.445300	0.6561
DSIF2(3)	-0.019627	0.008977	-2.186389	0.0288

R-squared	0.361073	Mean dependent var	-3.74E-05
Adjusted R-squared	0.360562	S.D. dependent var	0.003699
S.E. of regression	0.002958	Akaike info criterion	-8.807958
Sum squared resid	0.076593	Schwarz criterion	-8.801497
Log likelihood	38604.47	F-statistic	706.8887
Durbin-Watson stat	2.520371	Prob(F-statistic)	0.000000

Fig 9. Regression results for 5 minutes series, SIF2 and DSIF2 variables, i = 3

Even though many measures have been taken in order to compensate for the disturbances that can be induced by the local markets, the results for the five minutes SIF2/DSIF2 present a somewhat different result. From this regression it appears that SIF2 trails DSIF2 in terms of information integration.

The important fact is that the first lead coefficient is not only relevant, but also almost double in value to the lag coefficients. Since the time step is five minutes, from the previous results, a one step lead is expected and confirmed.

For the SIF5/DSIF5 data series the results are closer to what is to be expected. There are only two relevant coefficients, above the 1% level, the simultaneous one and the first lead. It is noticeable that, although in this case, the first two lag coefficients are around the 5% probability level their relative values are very close to the ones found for SIF2. They practically have the same ratio: 0.033/0.021 = 1.57 and 0.083/0.053 = 1.56. The R^2 value for SIF5/DSIF5 is again very small.

The first lag coefficient is relatively small compared to the simultaneous one proving that most of the information is integrated simultaneously in the two markets. The results from these two regressions indicate that the use of high frequency data was necessary to establish an accurate lead-lag difference. Chan (1992) found fifteen minutes lead for the 1985 data series, so the five minutes data was sufficient for a conclusion. For the second data series, from 1987, the lead comes down to only five minutes.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-2.83E-05	3.45E-05	-0.820382	0.4120
DSIF5(-3)	0.008749	0.010952	0.798785	0.4244
DSIF5(-2)	0.021864	0.010998	1.987944	0.0468
DSIF5(-1)	0.022855	0.011003	2.077199	0.0378
DSIF5	0.147769	0.011004	13.42838	0.0000
DSIF5(1)	0.033003	0.011003	2.999461	0.0027
DSIF5(2)	-0.003195	0.010998	-0.290535	0.7714
DSIF5(3)	-0.013085	0.010953	-1.194588	0.2323

R-squared	0.021151	Mean dependent var	-3.52E-05
Adjusted R-squared	0.020368	S.D. dependent var	0.003264
S.E. of regression	0.003231	Akaike info criterion	-8.631305
Sum squared resid	0.091392	Schwarz criterion	-8.624844
Log likelihood	37830.38	F-statistic	27.02835
Durbin-Watson stat	2.180426	Prob(F-statistic)	0.000000

Fig 10. Regression results for 5 minutes series, SIF5 and DSIF5 variables, i = 3

Except for the difference in lead length between the model applied on US data and the one applied on Romanian one, the only other important difference is the absolute values of the coefficients. For US data, they are considerably larger proving that there is a stronger conventional causal relation between the cash market and the futures one.

The leads underlined by the model are the same for the 2007 Romanian market and the 1987 US market, in accordance to the size differential and the experience one. The fundamental difference is that in the US, the futures market moves faster than the cash one, the reverse being true for the Romanian one. This is caused directly by the way the local investor trades. While in the US market many futures contracts are traded with the purpose of hedging, in the Romanian market these kinds of trades are at a very low level. Arbitrage opportunities are more frequent in the local market, but again a low number of trades seem to be executed for this purpose. The average Romanian investor uses simpler trading strategies and focuses more on entry and exit levels. When such a strategy is used, it is normal that the non-price change probability remains very low, although the trade probability is not overwhelming.

The majority of trades and the majority of investors, use the futures market in order to speculate. Since important mass-media transmitted news can't come very often, information from the cash market is used for futures trading.

C) Behavior under good news and bad news

The BVB cash market in Romania doesn't have the option of selling stocks short. This is why the behavior under good news should be different than the one under bad news.

In order to be able to estimate this effect, the observations have been sorted by their absolute value and by their sign. The input data is the five minutes series. Trading hours are partitioned into 85-minute intervals (i.e., each interval contains seventeen observations), and cash index returns are calculated for each interval. The length of the interval has been chosen to be short enough to avoid many different bits of information, and long enough to allow the information effect to have an impact on the lead-lag relation of some observations. It also was important to choose an interval length that was a divisor of the 255 minute BVB trading day.

The 85-minute intervals are ranked according to five quintiles based on cash index returns, and observations are allocated into the five quintiles according to the ranking of the interval. There have been kept only the first and the last quintile. Each observation is actually represents a 17 numbers long data series. This data was the input for the good-news series and the bad-news series.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.000377	7.51E-05	-5.011143	0.0000
DSIF2(-3)	-0.003591	0.015281	-0.234990	0.8142
DSIF2(-2)	0.026520	0.017269	1.535719	0.1248
DSIF2(-1)	0.033644	0.016524	2.036050	0.0419
DSIF2	0.756528	0.050645	14.93792	0.0000
DSIF2(1)	0.071442	0.017923	3.986063	0.0001
DSIF2(2)	-0.006885	0.017499	-0.393466	0.6940
DSIF2(3)	-0.041461	0.015240	-2.720605	0.0066

R-squared	0.542930	Mean dependent var	-0.001156
Adjusted R-squared	0.541032	S.D. dependent var	0.004639
S.E. of regression	0.003143	Akaike info criterion	-8.682806
Sum squared resid	0.016651	Schwarz criterion	-8.657139
Log likelihood	7362.336	F-statistic	286.1018
Durbin-Watson stat	2.359909	Prob(F-statistic)	0.000000

Fig 11. Regression results for SIF2/DSIF2 bad news group

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.001006	9.78E-05	-10.28898	0.0000
DSIF5(-3)	0.051345	0.028071	1.829127	0.0676
DSIF5(-2)	0.000474	0.028086	0.016880	0.9865
DSIF5(-1)	-0.027835	0.028088	-0.990999	0.3218
DSIF5	0.198767	0.028068	7.081678	0.0000
DSIF5(1)	-0.019145	0.028090	-0.681541	0.4956
DSIF5(2)	-0.023248	0.028094	-0.827493	0.4081
DSIF5(3)	0.001762	0.028081	0.062764	0.9500

R-squared	0.031731	Mean dependent var	-0.001028
Adjusted R-squared	0.027711	S.D. dependent var	0.004067
S.E. of regression	0.004010	Akaike info criterion	-8.195433
Sum squared resid	0.027109	Schwarz criterion	-8.169767
Log likelihood	6949.532	F-statistic	7.893074
Durbin-Watson stat	2.055725	Prob(F-statistic)	0.000000

Fig 12. Regression results for SIF5/DSIF5 bad news group

The coefficients presented in figure 11 for SIF2 are in accordance with the other results obtained so far. There is only a 5 minutes lead of the cash market over the futures, and the lag coefficient is not relevant at 1% (its absolute value is also low).

The difference from previous tests is in the fact that the instantaneous coefficients are much larger. This means that, in the presence of important news, be it bad or good, markets react very fast and incorporate much of the information just as fast.

For the SIF2 data series there is no real separation in the results for the good news and the bad news. The lead-lag relation seems to work the same no matter the news content, news intensity being the only real influence.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.000340	7.01E-05	4.853872	0.0000
DSIF2(-3)	0.016756	0.018366	0.912367	0.3617
DSIF2(-2)	0.046527	0.021303	2.183991	0.0291
DSIF2(-1)	0.025363	0.024748	1.024851	0.3056
DSIF2	0.774133	0.046137	16.77909	0.0000
DSIF2(1)	0.067741	0.026772	2.530269	0.0115
DSIF2(2)	-0.034842	0.019008	-1.833015	0.0670
DSIF2(3)	-0.005590	0.016051	-0.348248	0.7277

R-squared	0.482937	Mean dependent var	0.001130
Adjusted R-squared	0.480791	S.D. dependent var	0.004651
S.E. of regression	0.003351	Akaike info criterion	-8.554278
Sum squared resid	0.018935	Schwarz criterion	-8.528611
Log likelihood	7253.473	F-statistic	224.9609
Durbin-Watson stat	2.512119	Prob(F-statistic)	0.000000

Fig 13. Regression results for SIF2/DSIF2 good news group

We can also notice that larger leads (two or three) are close to the 5% relevance level, and have a negative sign. The correction/rebound effect is proportionally large to the price jump, which is also directly influenced by the intensity of news.

In regards to SIF5/DSIF5 results, they are quite different from what was expected. Lead and lag coefficients, of the first order, are very small, sometimes even having a different sign. The most probable explanation is the fact that it reacts faster than the five minutes interval chosen, thus the initial increase/decrease in price overlaps the correction/rebound.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.000996	0.000104	9.545842	0.0000
DSIF5(-3)	-0.010431	0.025032	-0.416719	0.6769
DSIF5(-2)	0.030533	0.024991	1.221751	0.2220
DSIF5(-1)	0.027506	0.024987	1.100845	0.2711
DSIF5	0.201210	0.024988	8.052177	0.0000
DSIF5(1)	0.016905	0.024985	0.676628	0.4987
DSIF5(2)	-0.029220	0.024989	-1.169302	0.2424
DSIF5(3)	-0.023309	0.024884	-0.936674	0.3491

R-squared	0.039194	Mean dependent var	0.001024
Adjusted R-squared	0.035205	S.D. dependent var	0.004354
S.E. of regression	0.004276	Akaike info criterion	-8.066658
Sum squared resid	0.030834	Schwarz criterion	-8.040992
Log likelihood	6840.459	F-statistic	9.825342
Durbin-Watson stat	2.068748	Prob(F-statistic)	0.000000

Fig 14. Regression results for SIF5/DSIF5 good news group

Other than the reversed lead-lag relation, the results are in accordance with Chan (1992). He finds that, for the US markets, it does not seem that the futures leads the cash index only under bad news. Neither is there a stronger tendency for the futures to lead the cash index under bad news than under good news.

The lead of the futures market over the cash market, for the US, stands at the usual five minutes interval.

D) Lead-Lag relation under different intensities of trading

The initial data series from BVB and BMFMS contained, along with the temporal and pricing information, the volume of each trade. Data aggregation for the volume series is the most straight-forward. For the minute series, as well as the five minute series have been build simply by adding the traded volume in that time interval.

The dataset used by Chan (1992) doesn’t include all the information available to us because his data series had entries only when a price change was observed (for futures transactions).

For the purpose of observing the lead-lag relation under heavy or light trading we have once again used the 85 minutes series. The volumes have been added for every 85 minutes interval, and these sums of volumes have been sorted according to their value. Once sorted, they are reverted to the instantaneous return rates, calculated for five minutes intervals. The new full length return series is divided in three equal parts: the high, medium and low return series.

The same regression model is used, with three lead-lags considered.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-3.86E-06	4.15E-05	-0.092939	0.9260
DSIF2H(-3)	0.012445	0.017775	0.700131	0.4839
DSIF2H(-2)	0.055051	0.024833	2.216858	0.0267
DSIF2H(-1)	0.049369	0.020271	2.435472	0.0149
DSIF2H	0.685531	0.036759	18.64921	0.0000
DSIF2H(1)	0.108749	0.022852	4.758827	0.0000
DSIF2H(2)	-0.003597	0.014393	-0.249911	0.8027
DSIF2H(3)	-0.011691	0.015363	-0.760975	0.4467
R-squared	0.411441	Mean dependent var		7.06E-05
Adjusted R-squared	0.410025	S.D. dependent var		0.004208
S.E. of regression	0.003232	Akaike info criterion		-8.628390
Sum squared resid	0.030407	Schwarz criterion		-8.611999
Log likelihood	12596.82	F-statistic		290.6109
Durbin-Watson stat	2.454341	Prob(F-statistic)		0.000000

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-6.13E-06	5.96E-05	-0.102778	0.9181
DSIF5H(-3)	0.004528	0.015300	0.295931	0.7673
DSIF5H(-2)	0.016188	0.017628	0.918310	0.3585
DSIF5H(-1)	0.042544	0.028567	1.489304	0.1365
DSIF5H	0.175046	0.058727	2.980669	0.0029
DSIF5H(1)	0.074652	0.026263	2.842433	0.0045
DSIF5H(2)	0.017417	0.018166	0.958772	0.3378
DSIF5H(3)	-0.017960	0.014422	-1.245254	0.2131
R-squared	0.033744	Mean dependent var		-1.76E-05
Adjusted R-squared	0.031420	S.D. dependent var		0.003447
S.E. of regression	0.003392	Akaike info criterion		-8.532093
Sum squared resid	0.033480	Schwarz criterion		-8.515702
Log likelihood	12456.32	F-statistic		14.51785
Durbin-Watson stat	2.143381	Prob(F-statistic)		0.000000

Fig 15. Regression results for SIF2/DSIF2 and SIF5/DSIF5, high volumes

The results remain in the same general lines designated by all the other five minutes series regressions.

For SIF2/DSIF2, no matter the volume traded the cash market remains ahead of the futures market by one lead, meaning five minutes. The instantaneous coefficient remains rather large being surpassed only by the one from the good/bad news regressions. It seems plausible that the high importance information is very fast integrated into the prices. Higher trading volumes means that even low importance news get integrated faster, however there simply is not a lot of information to be integrated, resulting in lower absolute value coefficients. At the same time, high volume is associated with high volatility and corroborated with the lack of important information it is to be expected that the zero lead coefficient is smaller.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-2.77E-05	4.17E-05	-0.663182	0.5073
DSIF2M(-3)	0.004738	0.014125	0.335463	0.7373
DSIF2M(-2)	0.019918	0.016071	1.239351	0.2153
DSIF2M(-1)	0.041726	0.022645	1.842628	0.0655
DSIF2M	0.631013	0.105443	5.984391	0.0000
DSIF2M(1)	0.044068	0.017057	2.583509	0.0098
DSIF2M(2)	-0.002145	0.017850	-0.120155	0.9044
DSIF2M(3)	-0.008191	0.013906	-0.588972	0.5559

R-squared	0.344104	Mean dependent var	-7.62E-05
Adjusted R-squared	0.342526	S.D. dependent var	0.003836
S.E. of regression	0.003110	Akaike info criterion	-8.705400
Sum squared resid	0.028153	Schwarz criterion	-8.689009
Log likelihood	12709.18	F-statistic	218.0971
Durbin-Watson stat	2.486898	Prob(F-statistic)	0.000000

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-3.92E-05	6.70E-05	-0.585816	0.5580
DSIF5M(-3)	-0.004412	0.019005	-0.232154	0.8164
DSIF5M(-2)	0.027303	0.019034	1.434462	0.1515
DSIF5M(-1)	0.033999	0.019039	1.785749	0.0742
DSIF5M	0.177252	0.019042	9.308221	0.0000
DSIF5M(1)	0.026472	0.019042	1.390221	0.1646
DSIF5M(2)	-0.012140	0.019038	-0.637640	0.5238
DSIF5M(3)	-0.006203	0.018999	-0.326487	0.7441

R-squared	0.030561	Mean dependent var	-5.86E-05
Adjusted R-squared	0.028229	S.D. dependent var	0.003664
S.E. of regression	0.003612	Akaike info criterion	-8.406553
Sum squared resid	0.037959	Schwarz criterion	-8.390162
Log likelihood	12273.16	F-statistic	13.10514
Durbin-Watson stat	2.167631	Prob(F-statistic)	0.000000

Fig 15. Regression results for SIF2/DSIF2 and SIF5/DSIF5, medium volumes

The first lead coefficient remains relevant, with a large difference between the high volume one and the rest. The feed-back effect is also noticeable at high volume, the first lag coefficient being close to the 1% relevance level.

It also is easily noticeable a degradation of all the statistics calculated when volume decreases. Not only are all the coefficients smaller but even the R^2 statistic decreases in value. For the low volume regression even the two lag coefficient becomes relevant, a sign of abnormality.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-2.47E-05	3.33E-05	-0.742609	0.4578
DSIF2L(-3)	-0.016157	0.013149	-1.228755	0.2193
DSIF2L(-2)	0.050294	0.017460	2.880530	0.0040
DSIF2L(-1)	0.044350	0.022453	1.975227	0.0483
DSIF2L	0.534609	0.049251	10.85480	0.0000
DSIF2L(1)	0.058225	0.020754	2.805485	0.0051
DSIF2L(2)	0.006829	0.017084	0.399712	0.6894
DSIF2L(3)	-0.021412	0.016135	-1.327066	0.1846

R-squared	0.290858	Mean dependent var	-0.000105
Adjusted R-squared	0.289152	S.D. dependent var	0.002927
S.E. of regression	0.002468	Akaike info criterion	-9.168159
Sum squared resid	0.017723	Schwarz criterion	-9.151768
Log likelihood	13384.34	F-statistic	170.5069
Durbin-Watson stat	2.467319	Prob(F-statistic)	0.000000

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-2.68E-05	4.76E-05	-0.563126	0.5734
DSIF5L(-3)	0.033069	0.020137	1.642232	0.1007
DSIF5L(-2)	0.042308	0.020154	2.099262	0.0359
DSIF5L(-1)	-0.019193	0.020161	-0.951973	0.3412
DSIF5L	0.048947	0.020153	2.428726	0.0152
DSIF5L(1)	-0.006048	0.020146	-0.300196	0.7640
DSIF5L(2)	0.016737	0.020129	0.831482	0.4058
DSIF5L(3)	-0.020675	0.020116	-1.027773	0.3041

R-squared	0.005365	Mean dependent var	-2.42E-05
Adjusted R-squared	0.002973	S.D. dependent var	0.002575
S.E. of regression	0.002572	Akaike info criterion	-9.085853
Sum squared resid	0.019244	Schwarz criterion	-9.069462
Log likelihood	13264.26	F-statistic	2.242410
Durbin-Watson stat	2.205662	Prob(F-statistic)	0.028328

Fig 15. Regression results for SIF2/DSIF2 and SIF5/DSIF5, low volumes

In accordance with previous results the SIF5/DSIF5 data series shows a clear lead-lag relationship only for the high volume series. Apparently, the high volume, allows the trend to resume within the five minutes interval, so that the last price in the interval shows the larger trend. The other two series seem affected by the same events as the good/bad news SIF5 series.

In conclusion, the lead doesn't generally extend or contract, however there are noticeable effects when the volume of trading is varied. For the US markets, Chan (1992) finds that there is no compelling evidence to suggest that the lead-lag relation may be affected by the relative intensity of trading activity in cash and futures markets. Variation in the infrequent trading component may slightly affect the relation between the two markets. However, if the futures contracts are traded actively enough, changes in the trading intensity will not have a significant impact on the lead-lag pattern. The coefficient values retain their growing value in accordance with growing trading volume.

2. Bottom-up approach

The bottom-up approach has the purpose of identifying a lead-lag relation in high frequency data, on short trading intervals (one or two days) that is relevant enough to be extended for the whole analyzed period.

I have chosen a model that was first used by Kawaller, Koch & Koch (1987)[1] to determine the relation between the cash and futures US markets, employing the help of the S&P 500 index.

As we have discussed in the first part of the paper, movements in the prices of the cash market and futures market representatives can each transmit information regarding subsequent price variations in both markets. It is reasonable to assume that the SIFx (x = 2 or 5) price at t+1 is given by the past prices of SIFx, the past prices of the DSIFx (x = 2 or 5) and the new information available at the moment of price formation. The price for the DSIFx futures contract is identified the same way.

In order to estimate these temporal price relations it is necessary to estimate the distributed lags between the first differences of the index and nearby futures price, in a model like the following:

$$S_t = c_1 + \sum_{i=1}^{jS} a_i * S_{t-i} + \sum_{i=0}^{kS} b_i * F_{t-i} + e_{1t} \quad (5)$$

$$F_t = c_2 + \sum_{i=0}^{jF} d_i * S_{t-i} + \sum_{i=1}^{kF} e_i * F_{t-i} + e_{2t} \quad (6)$$

Where c_1 and c_2 are intercept terms, S_t and F_t are the instantaneous returns for the cash market representative respectively the futures one, while other relevant market information that is affecting the prices is represented by random noise e_{1t} and e_{2t}.

Equations (5) and (6) represent a simultaneous-equations model because futures and cash prices may affect each other contemporaneously through b_0 and d_0. If such is the case, ordinary least-squares estimation of the equations would yield biased and inconsistent estimates. The model can be rewritten using matrixes, keeping in mind that we have a number of lags to consider. To be able to estimate the system we need to choose a number of lags to introduce into the model. The tradeoff is the fact that the longer the lag lengths are the chance of misspecification decreases but we also lose more degrees of freedom. Hence, it is desirable to choose the minimum lag length that specifies the relationship accurately. In an article from 1978 Geweke[6] argues that the lags of the dependent variable should be kept large, to be able to minimize the chance of serially correlated errors, while the number of lags on the other variable should be set lower to retain power in the hypothesis tests. In these paper's estimations the lag of the dependent variable has been chosen at 10, while the lag of the independent one has been chosen at 8. In choosing the lag length, I have used as start-up information the results found using the top-down approach and the correlogram/ cross-correlogram of the data series used. From there (eight lags) I have chosen a larger dependent variable length (the ten lags).

The estimation of the model is done using a three-stage least squares method. There are two potential advantages over an estimation using ordinary least squares that are applied to a single equation. The first one is that because of potential simultaneity amongst the variables, an instrumental-variables estimator is required to produce consistent estimates. Second, the "other relevant information" embodied in each error term, e_{1t} and e_{2t} may affect both prices. This would imply contemporaneous correlation between the error terms, and, even in the absence of simultaneity (if b_0 and d_0 are simultaneously 0), ordinary least squares would yield inefficient estimates.

A) High volume trading, bear market

The first data series that is used as input for this model is the instantaneous return rates for the 20th and the 21st of November 2007. These two days have been the last days of a bear market that stretched over a month. Since it is exactly the turning point, the volumes are very large; in fact they are the largest volume days for 2007. The 23rd of November was actually the first day with green indexes.

From the minute series, there have been extracted the 510 observations recorded on the two mentioned days. The method of aggregation has remained the same with the minute series.

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	-9.89E-05	7.81E-05	-1.266857	0.2055
C(2)	-0.218107	0.044539	-4.896969	0.0000
C(3)	-0.108976	0.045515	-2.394300	0.0168
C(4)	-0.016900	0.045780	-0.369149	0.7121
C(5)	-0.042247	0.045801	-0.922393	0.3566
C(6)	-0.009642	0.045493	-0.211951	0.8322
C(7)	-0.009365	0.045761	-0.204643	0.8379
C(8)	-0.060130	0.045610	-1.318375	0.1877
C(9)	0.064134	0.045673	1.404208	0.1606
C(10)	0.076539	0.045197	1.693443	0.0907
C(11)	0.086010	0.044572	1.929694	0.0539
C(12)	0.249319	0.054985	4.534285	0.0000
C(13)	0.038456	0.055903	0.687914	0.4917
C(14)	0.141107	0.053754	2.625044	0.0088
C(15)	0.044601	0.054097	0.824468	0.4099
C(16)	0.168985	0.054078	3.124807	0.0018
C(17)	0.073760	0.053894	1.368600	0.1714
C(18)	-0.003506	0.053714	-0.065276	0.9480
C(19)	0.106223	0.053477	1.986330	0.0473
C(20)	0.083811	0.053552	1.565026	0.1179
C(30)	-3.42E-05	6.27E-05	-0.546239	0.5850
C(31)	-0.125985	0.044617	-2.823688	0.0048
C(32)	-0.039958	0.043461	-0.919394	0.3581
C(33)	0.067838	0.043335	1.565434	0.1178
C(34)	-0.114636	0.043609	-2.628719	0.0087
C(35)	-0.039878	0.043818	-0.910078	0.3630
C(36)	-0.041424	0.043607	-0.949944	0.3424
C(37)	-0.077665	0.042886	-1.810952	0.0705
C(38)	0.012104	0.043382	0.279016	0.7803
C(39)	-0.013752	0.043008	-0.319752	0.7492
C(40)	-0.044771	0.042414	-1.055569	0.2914
C(41)	0.164016	0.035572	4.610751	0.0000
C(42)	0.080119	0.036687	2.183858	0.0292
C(43)	0.119560	0.036791	3.249703	0.0012
C(44)	0.067762	0.036930	1.834856	0.0668
C(45)	0.049638	0.036903	1.345101	0.1789
C(46)	-0.002535	0.036998	-0.068505	0.9454
C(47)	0.084280	0.037046	2.275000	0.0231
C(48)	0.084678	0.036853	2.297716	0.0218
C(49)	-0.040039	0.036204	-1.105936	0.2690
Determinant residual covariance	4.52E-12			

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	-0.000122	6.56E-05	-1.852646	0.0642
C(2)	-0.268904	0.045164	-5.953886	0.0000
C(3)	-0.208654	0.046541	-4.483257	0.0000
C(4)	-0.166685	0.047647	-3.498305	0.0005
C(5)	-0.018245	0.048002	-0.380091	0.7040
C(6)	0.027638	0.047963	0.576226	0.5646
C(7)	-0.045618	0.048138	-0.947647	0.3435
C(8)	0.013206	0.047397	0.278623	0.7806
C(9)	0.034737	0.047085	0.737758	0.4608
C(10)	0.056841	0.045897	1.238437	0.2159
C(11)	0.087861	0.044781	1.962043	0.0500
C(12)	0.424806	0.055639	7.635087	0.0000
C(13)	-0.014889	0.057303	-0.259835	0.7950
C(14)	0.200744	0.057357	3.499888	0.0005
C(15)	-0.045909	0.057712	-0.795495	0.4265
C(16)	0.060063	0.057580	1.043126	0.2972
C(17)	0.038112	0.057427	0.663666	0.5071
C(18)	0.087148	0.057469	1.516439	0.1297
C(19)	0.108648	0.057154	1.900970	0.0576
C(20)	0.031243	0.059211	0.527651	0.5979
C(30)	-1.43E-05	5.15E-05	-0.278255	0.7809
C(31)	-0.085305	0.044772	-1.905342	0.0570
C(32)	-0.154430	0.045279	-3.410609	0.0007
C(33)	0.013127	0.045304	0.289762	0.7721
C(34)	-0.018195	0.045299	-0.401653	0.6880
C(35)	-0.052667	0.045175	-1.165845	0.2440
C(36)	0.006155	0.045205	0.136152	0.8917
C(37)	-0.048833	0.045320	-1.077503	0.2815
C(38)	-0.092088	0.046612	-1.975628	0.0485
C(39)	-0.000489	0.045589	-0.010737	0.9914
C(40)	0.020389	0.045376	0.449334	0.6533
C(41)	0.261299	0.034297	7.618713	0.0000
C(42)	0.118517	0.036652	3.233534	0.0013
C(43)	0.118467	0.037369	3.170202	0.0016
C(44)	0.097320	0.038150	2.550966	0.0109
C(45)	0.033535	0.037881	0.885273	0.3762
C(46)	0.042769	0.037779	1.132068	0.2579
C(47)	0.074907	0.037526	1.996155	0.0462
C(48)	0.033370	0.036747	0.908097	0.3641
C(49)	-0.031526	0.035964	-0.876606	0.3809
Determinant residual covariance	1.76E-12			

Fig 16. Regression results for SIF2/DSIF2 (left) and SIF5/DSIF5 (right), high volumes, bearish

The equations used are (the same for SIF5/DSIF5, with different data series):

SIF2= c(1)+c(2)*SIF2(-1)+c(3)*sif2(-2)+c(4)*sif2(-3)+c(5)*sif2(-4)+c(6)*sif2(-5)+c(7)*sif2(-6)+c(8)*sif2(-7)+c(9)*sif2(-8)+c(10)*sif2(-9)+c(11)*sif2(-10)+C(12)*DSIF2+c(13)*dsif2(-1)+c(14)*dsif2(-2)+c(15)*dsif2(-3)+c(16)*dsif2(-4)+c(17)*dsif2(-5)+c(18)*dsif2(-6)+c(19)*dsif2(-7)+c(20)*dsif2(-8)

DSIF2=c(30)+c(31)*DSIF2(-1)+c(32)*dsif2(-2)+c(33)*dsif2(-3)+c(34)*dsif2(-4)+c(35)*dsif2(-5)+c(36)*dsif2(-6)+c(37)*dsif2(-7)+c(38)*dsif2(-8)+c(39)*dsif2(-9)+c(40)*dsif2(-10)+c(41)*SIF2+c(42)*sif2(-1)+c(43)*sif2(-2)+c(44)*sif2(-3)+c(45)*sif2(-4)+c(46)*sif2(-5)+c(47)*sif2(-6)+c(48)*sif2(-7)+c(49)*sif2(-8).

We find that both intercept are 0, and that both series are connected very weakly to their past values (SIFx(t) and DSIFx(t) determined by SIFx(t-i) and DSIFx(t-j)). If any, the relevant correlation coefficient is the first or second degree one (meaning that price in a series is influenced only by the last price and possibly second-to-last one). At the same time, the equity price is influenced by the second lag futures price, a possible result of the feed-back effect. Just as before, the contemporaneous coefficients remain the largest.

The most important result is that we are able to identify a sufficiently clear lead-lag relation. The cash market continues to lead the futures market with a time difference of two (SIF2) to three minutes (SIF3).

B) High volume trading, bullish market

The second data series used along with the model is the instantaneous return rates from the 20th to 21st of December 2007. It represents the peak of a bullish move, again with very high volumes characteristic in market mood reversal. The same minute series represents the starting data series used. The estimation equations also remain the same.

The results show that price discovery, for both cash and futures markets, involves the same first or second lag dependent variable. DSIFx influences SIFx with only one lag.

The most important result is the one linked to the lead of cash markets over futures markets. We find that SIF2 leads DSIF2 by three minutes while SIF5 leads DSIF5 by almost five minutes in complete agreement with the previous results from the top-down approach, minute series, and also in line with the rest of the results.

The coefficients of these lagged variables are also close together, in absolute values, when compared (SIF2 ones compared to SIF5 ones).

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	-2.72E-05	8.91E-05	-0.305547	0.7600
C(2)	-0.260455	0.046591	-5.590219	0.0000
C(3)	-0.283383	0.049210	-5.758630	0.0000
C(4)	-0.174015	0.051447	-3.382397	0.0007
C(5)	-0.118911	0.051764	-2.297197	0.0218
C(6)	-0.098582	0.052171	-1.889582	0.0591
C(7)	-0.016632	0.052401	-0.317404	0.7510
C(8)	-0.102825	0.052160	-1.971361	0.0490
C(9)	-0.056976	0.051474	-1.106891	0.2686
C(10)	0.040065	0.048307	0.829394	0.4071
C(11)	-0.038007	0.045994	-0.826347	0.4088
C(12)	0.462666	0.078224	5.914620	0.0000
C(13)	0.197191	0.081651	2.415055	0.0159
C(14)	0.172222	0.081970	2.101039	0.0359
C(15)	0.068259	0.081530	0.837231	0.4027
C(16)	0.152902	0.081129	1.884683	0.0598
C(17)	0.147385	0.081484	1.808752	0.0708
C(18)	0.096033	0.082430	1.165034	0.2443
C(19)	0.019832	0.081934	0.242048	0.8088
C(20)	0.132096	0.080521	1.640514	0.1012
C(30)	-3.48E-06	5.03E-05	-0.069150	0.9449
C(31)	-0.036277	0.046356	-0.782577	0.4341
C(32)	-0.090267	0.046014	-1.961713	0.0501
C(33)	-0.065590	0.045934	-1.427925	0.1536
C(34)	-0.092610	0.045986	-2.013863	0.0443
C(35)	-0.069354	0.046119	-1.503803	0.1330
C(36)	-0.151376	0.045208	-3.348408	0.0008
C(37)	0.039180	0.044361	0.883214	0.3773
C(38)	0.034993	0.043627	0.802104	0.4227
C(39)	0.027064	0.042291	0.639951	0.5224
C(40)	-0.024998	0.042168	-0.592833	0.5534
C(41)	0.147975	0.024874	5.948958	0.0000
C(42)	0.146092	0.026333	5.547821	0.0000
C(43)	0.168336	0.027619	6.094894	0.0000
C(44)	0.122296	0.028772	4.250565	0.0000
C(45)	0.049900	0.029243	1.706371	0.0883
C(46)	0.052330	0.029369	1.781829	0.0751
C(47)	0.072608	0.029053	2.499158	0.0126
C(48)	0.068967	0.028155	2.449577	0.0145
C(49)	-0.017450	0.027124	-0.643325	0.5202
Determinant residual covariance		4.29E-12		

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	-1.64E-05	6.66E-05	-0.245574	0.8061
C(2)	-0.211360	0.044862	-4.711355	0.0000
C(3)	-0.088043	0.046270	-1.902825	0.0574
C(4)	-0.070457	0.046953	-1.500573	0.1338
C(5)	-0.097538	0.047750	-2.042676	0.0414
C(6)	-0.090614	0.047963	-1.889250	0.0592
C(7)	-0.082042	0.048031	-1.708114	0.0879
C(8)	0.122641	0.047988	2.555653	0.0108
C(9)	0.011385	0.047698	0.238694	0.8114
C(10)	-0.042057	0.044652	-0.941877	0.3465
C(11)	-0.025870	0.043519	-0.594448	0.5524
C(12)	0.777276	0.066176	11.74562	0.0000
C(13)	0.213262	0.070921	3.007024	0.0027
C(14)	0.080250	0.072220	1.111195	0.2668
C(15)	0.103367	0.072526	1.425246	0.1544
C(16)	0.066609	0.072713	0.916059	0.3599
C(17)	0.059613	0.072498	0.822274	0.4111
C(18)	0.046770	0.071881	0.650660	0.5154
C(19)	-0.007869	0.069669	-0.112953	0.9101
C(20)	0.026647	0.069055	0.385879	0.6997
C(30)	-3.90E-07	4.22E-05	-0.009242	0.9926
C(31)	-0.101006	0.044958	-2.246679	0.0249
C(32)	-0.166087	0.045119	-3.681055	0.0002
C(33)	-0.124017	0.045566	-2.721677	0.0066
C(34)	-0.096183	0.045547	-2.111751	0.0350
C(35)	-0.054648	0.044994	-1.214564	0.2248
C(36)	-0.064445	0.044425	-1.450660	0.1472
C(37)	0.007878	0.043438	0.181357	0.8561
C(38)	-0.019485	0.043136	-0.451704	0.6516
C(39)	0.007861	0.039408	0.199487	0.8419
C(40)	-0.010535	0.039236	-0.268516	0.7884
C(41)	0.309942	0.026480	11.70461	0.0000
C(42)	0.103843	0.028707	3.617371	0.0003
C(43)	0.123176	0.028715	4.289647	0.0000
C(44)	0.116705	0.029059	4.016127	0.0001
C(45)	0.131086	0.029551	4.435925	0.0000
C(46)	0.082249	0.030090	2.733404	0.0064
C(47)	0.073322	0.030228	2.425606	0.0155
C(48)	-0.030746	0.030396	-1.011505	0.3120
C(49)	0.011398	0.029456	0.386937	0.6989
Determinant residual covariance		1.21E-12		

Fig 17. Regression results for SIF2/DSIF2 (left) and SIF5/DSIF5 (right), high volumes, bullish

C) Low volume trading

For the third set of model regressions there have been chosen two days with the smallest liquidity. The 2nd and 3rd of November 2007 counted the least volume traded on both equities considered, SIF2 and SIF5. The assumption was that under low trading volume the lag of the futures market would extend further beyond the 3-5 minutes. Since we are working with minute data, a small change would be observable.

The results obtained for this data series are inconclusive. Even the contemporaneous coefficients are affected; they are smaller than the ones found in previous regressions. As a direct consequence of the small trading volume, autocorrelation in the data series seems to skip some lags. This is because the larger volume trades are mixed with minimal volume trades that, an effect presented earlier, yield no relevant information for the wide market.

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	-6.58E-06	4.98E-05	-0.132073	0.8950
C(2)	-0.124500	0.044064	-2.825414	0.0048
C(3)	-0.130565	0.043679	-2.989204	0.0029
C(4)	-0.029694	0.044057	-0.673983	0.5005
C(5)	-0.054597	0.043938	-1.242610	0.2143
C(6)	0.065887	0.043655	1.509285	0.1316
C(7)	-0.055950	0.043334	-1.291142	0.1970
C(8)	-0.015131	0.043395	-0.348690	0.7274
C(9)	-0.012687	0.043421	-0.292178	0.7702
C(10)	-0.008736	0.042052	-0.207742	0.8355
C(11)	-0.024630	0.041682	-0.590905	0.5547
C(12)	0.160696	0.042635	3.769117	0.0002
C(13)	0.091593	0.042416	2.159427	0.0311
C(14)	-0.009685	0.042636	-0.227154	0.8204
C(15)	0.118384	0.043701	2.708930	0.0069
C(16)	0.124584	0.043785	2.845336	0.0045
C(17)	0.055122	0.044204	1.246993	0.2127
C(18)	0.057548	0.042886	1.341905	0.1799
C(19)	0.103121	0.042759	2.411671	0.0161
C(20)	-0.006032	0.042862	-0.140732	0.8881
C(30)	-4.16E-05	5.19E-05	-0.801635	0.4230
C(31)	0.044662	0.044322	1.007677	0.3139
C(32)	0.028357	0.044334	0.639625	0.5226
C(33)	-0.254434	0.044475	-5.720861	0.0000
C(34)	-0.003671	0.046083	-0.079668	0.9365
C(35)	-0.019055	0.046055	-0.413746	0.6792
C(36)	-0.085986	0.045891	-1.873699	0.0613
C(37)	0.008687	0.046058	0.188613	0.8504
C(38)	0.026411	0.044493	0.593600	0.5529
C(39)	-0.030473	0.044311	-0.687717	0.4918
C(40)	-0.040670	0.044192	-0.920312	0.3576
C(41)	0.174772	0.046287	3.775808	0.0002
C(42)	0.072683	0.046070	1.577644	0.1150
C(43)	0.066447	0.045968	1.445506	0.1486
C(44)	0.084377	0.045981	1.835066	0.0668
C(45)	-0.053465	0.045914	-1.164459	0.2445
C(46)	-0.021899	0.045933	-0.476762	0.6336
C(47)	0.047424	0.045792	1.035655	0.3006
C(48)	0.050251	0.045198	1.111786	0.2665
C(49)	-0.007508	0.044799	-0.167584	0.8669
Determinant residual covariance		1.55E-12		

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	-3.01E-05	3.96E-05	-0.759915	0.4475
C(2)	-0.125780	0.045148	-2.785933	0.0054
C(3)	-0.018301	0.044764	-0.408826	0.6828
C(4)	-0.058942	0.044639	-1.320407	0.1870
C(5)	-0.111282	0.044496	-2.500949	0.0126
C(6)	0.096747	0.044123	2.192649	0.0286
C(7)	0.063121	0.042263	1.493516	0.1356
C(8)	-0.035595	0.042392	-0.839667	0.4013
C(9)	0.084285	0.042312	1.991997	0.0467
C(10)	-0.033558	0.042818	-0.783751	0.4334
C(11)	-0.105062	0.042608	-2.465775	0.0138
C(12)	0.141180	0.047374	2.980100	0.0030
C(13)	-0.065226	0.047491	-1.373449	0.1699
C(14)	0.045022	0.047382	0.950183	0.3423
C(15)	0.023707	0.047373	0.500425	0.6169
C(16)	0.008237	0.047492	0.173440	0.8623
C(17)	0.108963	0.047587	2.289749	0.0223
C(18)	0.087015	0.047545	1.830155	0.0675
C(19)	-0.037540	0.046446	-0.808247	0.4191
C(20)	-0.043004	0.046417	-0.926471	0.3544
C(30)	-3.08E-05	3.72E-05	-0.827516	0.4082
C(31)	-0.010005	0.044489	-0.224886	0.8221
C(32)	0.035632	0.044339	0.803623	0.4218
C(33)	-0.026649	0.044285	-0.601761	0.5475
C(34)	-0.058974	0.044196	-1.334382	0.1824
C(35)	-0.012178	0.044417	-0.274182	0.7840
C(36)	-0.016697	0.044389	-0.376143	0.7069
C(37)	0.084971	0.043471	1.954643	0.0509
C(38)	0.033179	0.043077	0.770219	0.4414
C(39)	-0.049775	0.041880	-1.188508	0.2349
C(40)	-0.103732	0.040665	-2.550891	0.0109
C(41)	0.129780	0.041378	3.136488	0.0018
C(42)	0.161025	0.041812	3.851201	0.0001
C(43)	-0.024093	0.041922	-0.574716	0.5656
C(44)	0.015954	0.041821	0.381477	0.7029
C(45)	0.103681	0.041384	2.505329	0.0124
C(46)	0.056094	0.041048	1.366546	0.1721
C(47)	0.041690	0.039620	1.052248	0.2930
C(48)	0.008771	0.039775	0.220517	0.8255
C(49)	0.029157	0.040016	0.728638	0.4664
Determinant residual covariance		4.99E-13		

Fig 18. Regression results for SIF2/DSIF2 (left) and SIF5/DSIF5 (right), lowest volumes

A lead-lag relation of a certain magnitude no longer holds on these data series. Even the consecutive lead coefficients, between the two markets, which should hold some relation, only display this characteristic amongst non-consecutive ones. For example, in the SIF5/DSIF5 data series, the lead cash market coefficients are C(41), C(42)… While in this particular series, the first two coefficients are relevant (at 1% level), there is no relation to the next two C(43), C(44), which are not relevant, while the fifth lead, C(45), becomes relevant again. In conclusion, the lowest volumes recorded are too low to show a relevant relation amongst the two markets.

Conclusions

This paper has used two different models and just as different methodologies in order to show and explain a lead-lag relation between the Romanian cash and futures markets. The data series have been aggregated, through a weighted average for the price and a sum for the volume, from all transactions that took place, involving two equities (SIF2 and SIF5) and their futures contracts, in a period of eight months. There have been used a minute frequency data set and a five minutes one. The two tickers (SIF2 and SIF5) belong to two funds that hold a highly diversified portfolio of other listed and unlisted equities, which brings them as close as possible to a market index.

In mature markets two phenomena, market sentiment and arbitrage trading, are the major determinants linking stock index futures and the stock market. One particular aspect of the Romanian cash market is the fact that short-selling stocks is not allowed. At the same time, few futures traders use this instrument for hedging or for arbitrage. Thus, most of futures trading is done using a simple strategy of speculating right moments to enter and exit the market.

The top – down approach employs a model also used by Chan (1992) that tries to establish a lead-lag relationship in the data series, involving an equal number of leads and lags in the regression. It is estimated on very long data series (months), at high frequency (minute and five minutes instantaneous return rates). The results prove a lead of three to five (maximum six) minutes depending upon input, of the cash market compared to the futures. At the minute level, a one minute lead of futures over cash prices was observed at the same time. This behavior doesn't seem to change in periods with high importance bad news or good news. Although the lead remains around the five minutes mark, the first lead coefficients seem to decrease along with their relevance, when a high volume trading data series is compared to a medium or a low volume one.

Because of considerable differences in market volume and investor education between the US market and the Romanian one, Chan (1992) finds the opposite, that there is no compelling evidence to suggest the lead-lag relation may be affected by the relative intensity of trading activity in cash and futures markets.

The bottom-up approach uses a system of equations estimated with a three stage least squares method. The data series used are very short, day or two days minute series. Scenarios including high volume bear and bull markets along with low volume series are analyzed. The lead-lag relation is the same as the one found using the first model. Even the time intervals remain in the three to five minutes interval. Again, on the low volume data series it is hard to identify relevant coefficients, along with a clear relation between them. The reasons include a growing number of minimum volume trades (100 equities or 1 futures contract), that incorporate a negligible informational value, along with an extended period for the anti-trend transactions (the temporary rebound/correction trades that take place immediately after a high price change, in the opposite direction)

Consistent results suggest that futures prices and the equities prices are simultaneously related on a minute-to-minute basis throughout the trading day. Further, significant lead coefficients suggest that the lead from cash prices to futures extends for between three and five minutes, while the lead from futures to cash prices, though sometimes significant, rarely extends beyond one minute. The length of the lead from cash prices to futures reflects, in part, the type of investors that trade in the futures market. Furthermore, futures tend to change price between consecutive transactions (partly because four decimals can be used in its price compared to only two for equities) varying around the trend set by the equities. It is also noticeable the fact that most of the new information is integrated in the same time interval (a minute) in both markets: the simultaneous coefficients are always the largest ones.

Except for the reversed relationship, equities leading futures, the largest differences compared to the studies done on the US markets with data from the late 80's, is the considerably shorter lag. Compared to 15 – 20 minutes leads found on the US markets, the three to five minutes encountered on the local market is a very short lead; electronic trading, the internet and the development of mobile communications have brought a faster news response time especially in trading.

The lead/lag relationships are stable across the different time intervals, models, market moods and pairs of equities/futures contracts examined in the 2007-2008 period on the Romanian markets.

Because of differential transaction costs and expected profits, because of the fact that "short-selling" is possible only on this market and the lower non-trading probability, the futures market is expected to lead the cash market. However, because of the very high number of speculative trades, the futures market consistently lags the cash market with 3 to 5 minutes.

Bibliography

1. Admati, A. R., and P. Pfleiderer, 1988, "A Theory of Intraday Patterns: Volume and Price Variability,"Review of Financial Studies, 1, 3-40.
2. Akgiray, V., 1989, “Conditional Heteroscedasticity in Time Series of Stock Returns: Evidence and Forecasts”, The Journal of Business, Vol. 62, No. 1, 55-80.
3. Chan K, 1992, “A Further Analysis of the Lead--Lag Relationship Between the Cash Market and Stock Index Futures Market” The Review of Financial Studies, Vol. 5, No. 1., 123-152
4. Chung, Y. P., 1991, "A Transactions Data Test of Stock Index Futures Market Efficiency and Index Arbitrage Profitability," forthcoming in Journal of Finance.
5. Diamond, D. W., and R. E. Verrecchia, 1987, "Constraints on Short-Selling and Asset Price Adjustment to Private Information," Journal of Financial Economics, 18, 277-311.
6. Geweke J, 1978 "Testing the Exogeneity Specification in the Complete Dynamic Simultaneous Equations Model." Journal of Econometrics 6 , 163-85.
7. Harris, L., 1989, "The october 1987 S&P 500 Stock-Futures Basis," Journal of Finance, 44, 77-99.
8. Hasbrouck Joel, 2001, “Intraday Price Formation in US Equity Index Markets”, Stern School of Business New York University, working paper
9. Karpoff, J. M., 1987,” The Relation Between Price Changes and Trading Volume: A Survey”, The Journal of Financial and Quantitative Analysis, Vol. 22, No. 1, 109-126.
10. Kawaller, I. G., P. D. Koch, and T. W. Koch, 1987, "The Temporal Price Relationship between S&P 500 Futures and S&P 500 Index," Journal of Finance, 42, 1309-1329.
11. MacKinlay, A. C. and Ramaswamy, K.,” Index-Futures Arbitrage and the Behavior of Stock Index Futures Prices”, The Review of Financial Studies, Vol. 1, No. 2, 137-158.
12. Stephan, J. A., and R. E. Whaley, 1990, "Intraday Price Changes and Trading Volume Relations in the Stock and Stock Option Markets," Journal of Finance, 45, 191-220.
13. Stoll, H. R., and R. E. Whaley, 1990, "The Dynamics of Stock Index and Stock Index Futures Returns," Journal of Financial and Quantitative Analysis, 25, 441-468.
14. Tse, Y, P. Bandyopadhyay and Y.-P. Shen, 2006, “Intraday Price Discovery in the DJIA Index Markets”, California State University working paper

Annex

Cross correlogram of SIF2/DSIF2, minute series

SIF2,DSIF2(-i)	SIF2,DSIF2(+i)	i	lag	lead
**	**	0	0.2351	0.2351
*	*	1	0.1229	0.1095
	*	2	0.0335	0.0527
		3	0.0131	0.0307
		4	0.0176	-0.0007
		5	0.0008	0.0071
		6	0.0053	0.0057
		7	0.0105	0.0001
		8	0.0061	-0.0127
		9	0.0050	0.0033
		10	0.0025	-0.0064

Cross correlogram of SIF5/DSIF5, minute series

SIF5,DSIF5(-i)	SIF5,DSIF5(+i)	i	lag	lead
		0	0.0126	0.0126
		1	0.0299	0.0036
		2	0.0128	0.0056
		3	0.0000	-0.0033
		4	-0.0045	0.0074
		5	0.0009	0.0008
		6	0.0046	0.0011
		7	-0.0010	-0.0038
		8	-0.0006	0.0024
		9	-0.0018	-0.0010
		10	0.0010	-0.0025

Cross correlogram of SIF2/DSIF2, five minutes series

SIF2,DSIF2(-i)	SIF2,DSIF2(+i)	i	lag	lead
******	******	0	0.5909	0.5909
	*	1	0.0311	0.0576
		2	0.0358	-0.0104
		3	0.0003	-0.0196
		4	0.0016	0.0029
		5	0.0016	-0.0046
		6	0.0143	-0.0191
		7	0.0096	0.0065
		8	-0.0037	-0.0068
		9	-0.0208	-0.0131
		10	0.0062	0.0101

Cross correlogram of SIF5/DSIF5, five minutes series

<table>
<tr><th>SIF5,DSIF5(-i)</th><th>SIF5,DSIF5(+i)</th><th>i</th><th>lag</th><th>lead</th></tr>
<tr><td>|* |</td><td>|* |</td><td>0</td><td>0.1384</td><td>0.1384</td></tr>
<tr><td>| |</td><td>| |</td><td>1</td><td>0.0073</td><td>0.0195</td></tr>
<tr><td>| |</td><td>| |</td><td>2</td><td>0.0150</td><td>-0.0081</td></tr>
<tr><td>| |</td><td>| |</td><td>3</td><td>0.0044</td><td>-0.0149</td></tr>
<tr><td>| |</td><td>| |</td><td>4</td><td>0.0328</td><td>0.0002</td></tr>
<tr><td>| |</td><td>| |</td><td>5</td><td>-0.0019</td><td>-0.0071</td></tr>
<tr><td>| |</td><td>| |</td><td>6</td><td>0.0125</td><td>0.0064</td></tr>
<tr><td>| |</td><td>| |</td><td>7</td><td>0.0086</td><td>-0.0079</td></tr>
<tr><td>| |</td><td>| |</td><td>8</td><td>0.0003</td><td>-0.0019</td></tr>
<tr><td>| |</td><td>| |</td><td>9</td><td>0.0005</td><td>-0.0130</td></tr>
<tr><td>| |</td><td>| |</td><td>10</td><td>-0.0303</td><td>0.0035</td></tr>
</table>

Cross correlogram of SIF2/DSIF2, two days bear market series

<table>
<tr><th>SIF2,DSIF2(-i)</th><th>SIF2,DSIF2(+i)</th><th>i</th><th>lag</th><th>lead</th></tr>
<tr><td>.|* |</td><td>.|* |</td><td>0</td><td>0.0635</td><td>0.0635</td></tr>
<tr><td>.|. |</td><td>.|. |</td><td>1</td><td>-0.0273</td><td>0.0437</td></tr>
<tr><td>.|* |</td><td>.|* |</td><td>2</td><td>0.1163</td><td>0.0678</td></tr>
<tr><td>.|. |</td><td>.|. |</td><td>3</td><td>0.0023</td><td>0.0352</td></tr>
<tr><td>.|* |</td><td>.|. |</td><td>4</td><td>0.0959</td><td>0.0332</td></tr>
<tr><td>.|. |</td><td>*|. |</td><td>5</td><td>0.0389</td><td>-0.0474</td></tr>
<tr><td>*|. |</td><td>.|* |</td><td>6</td><td>-0.0487</td><td>0.0860</td></tr>
<tr><td>.|* |</td><td>.|* |</td><td>7</td><td>0.0913</td><td>0.0645</td></tr>
<tr><td>.|* |</td><td>*|. |</td><td>8</td><td>0.0572</td><td>-0.0819</td></tr>
<tr><td>.|. |</td><td>.|. |</td><td>9</td><td>0.0135</td><td>0.0254</td></tr>
<tr><td>.|. |</td><td>.|. |</td><td>10</td><td>0.0422</td><td>0.0390</td></tr>
</table>

Cross correlogram of SIF5/DSIF5, two days bear market series

<table>
<tr><th>SIF5,DSIF5(-i)</th><th>SIF5,DSIF5(+i)</th><th>i</th><th>lag</th><th>lead</th></tr>
<tr><td>.|* |</td><td>.|* |</td><td>0</td><td>0.1340</td><td>0.1340</td></tr>
<tr><td>*|. |</td><td>.|. |</td><td>1</td><td>-0.0938</td><td>0.0377</td></tr>
<tr><td>.|* |</td><td>.|* |</td><td>2</td><td>0.1021</td><td>0.0541</td></tr>
<tr><td>*|. |</td><td>.|. |</td><td>3</td><td>-0.0575</td><td>0.0144</td></tr>
<tr><td>.|. |</td><td>.|. |</td><td>4</td><td>0.0448</td><td>-0.0220</td></tr>
<tr><td>.|. |</td><td>.|. |</td><td>5</td><td>-0.0022</td><td>0.0293</td></tr>
<tr><td>.|* |</td><td>.|* |</td><td>6</td><td>0.0617</td><td>0.0626</td></tr>
<tr><td>.|* |</td><td>.|. |</td><td>7</td><td>0.0576</td><td>0.0270</td></tr>
<tr><td>.|. |</td><td>*|. |</td><td>8</td><td>-0.0385</td><td>-0.0729</td></tr>
<tr><td>.|* |</td><td>*|. |</td><td>9</td><td>0.0895</td><td>-0.0856</td></tr>
<tr><td>.|. |</td><td>.|* |</td><td>10</td><td>0.0125</td><td>0.1025</td></tr>
<tr><td>.|. |</td><td>.|. |</td><td>11</td><td>0.0192</td><td>0.0407</td></tr>
</table>

Granger test results for SIF2/DSIF2, five minute series

Null Hypothesis:	Obs	F-Statistic	Probability
DSIF2 does not Granger Cause SIF2	8768	89.5606	3.1E-39
SIF2 does not Granger Cause DSIF2		46.1070	1.2E-20

Granger test results for SIF5/DSIF5, five minute series

Null Hypothesis:	Obs	F-Statistic	Probability
DSIF5 does not Granger Cause SIF5	8768	2.66889	0.06939
SIF5 does not Granger Cause DSIF5		4.61516	0.00992

Granger test results for SIF2/DSIF2, five minute series, bad news

Null Hypothesis:	Obs	F-Statistic	Probability
DSIF2 does not Granger Cause SIF2	1699	4.41239	0.03583
SIF2 does not Granger Cause DSIF2		18.6435	1.7E-05

Granger test results for SIF5/DSIF5, five minute series, bad news

Null Hypothesis:	Obs	F-Statistic	Probability
DSIF5 does not Granger Cause SIF5	1699	0.32495	0.56872
SIF5 does not Granger Cause DSIF5		0.38126	0.53701

Granger test results for SIF2/DSIF2, minute series, two days bull market

Null Hypothesis:	Obs	F-Statistic	Probability
DSIF2 does not Granger Cause SIF2	505	1.69949	0.14882
SIF2 does not Granger Cause DSIF2		8.75710	7.8E-07

Granger test results for SIF5/DSIF5, minute series, two days bull market

Null Hypothesis:	Obs	F-Statistic	Probability
DSIF5 does not Granger Cause SIF5	504	1.30362	0.26098
SIF5 does not Granger Cause DSIF5		7.63679	6.3E-07

Correlogram SIF 2, minute series

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
*| |	*| |	1	-0.076	-0.076	263.40	0.000
| |	| |	2	-0.044	-0.050	351.20	0.000
| |	| |	3	-0.024	-0.031	376.85	0.000
| |	| |	4	-0.011	-0.018	382.79	0.000
| |	| |	5	-0.009	-0.014	386.26	0.000
| |	| |	6	0.003	-0.001	386.58	0.000
| |	| |	7	-0.011	-0.013	391.99	0.000
| |	| |	8	-0.009	-0.012	395.52	0.000
| |	| |	9	0.008	0.005	398.37	0.000
| |	| |	10	-0.008	-0.008	400.99	0.000

Correlogram SIF 5, minute series

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
*| |	*| |	1	-0.069	-0.069	213.16	0.000
| |	| |	2	-0.027	-0.032	246.47	0.000
| |	| |	3	-0.003	-0.007	246.80	0.000
| |	| |	4	0.002	0.001	247.01	0.000
| |	| |	5	-0.005	-0.005	248.30	0.000
| |	| |	6	-0.001	-0.001	248.32	0.000
| |	| |	7	-0.005	-0.006	249.66	0.000
| |	| |	8	-0.002	-0.003	249.93	0.000
| |	| |	9	-0.005	-0.006	251.09	0.000
| |	| |	10	0.006	0.005	252.52	0.000

Correlogram DSIF 2, minute series

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
| |	| |	1	0.004	0.004	0.6057	0.436
| |	| |	2	0.017	0.017	13.632	0.001
| |	| |	3	-0.008	-0.008	16.757	0.001
| |	| |	4	-0.014	-0.014	25.543	0.000
| |	| |	5	-0.003	-0.002	25.886	0.000
| |	| |	6	-0.013	-0.012	33.157	0.000
| |	| |	7	0.006	0.006	34.936	0.000
| |	| |	8	-0.002	-0.002	35.144	0.000
| |	| |	9	-0.013	-0.014	43.175	0.000
| |	| |	10	0.000	-0.000	43.176	0.000

Correlogram DSIF 5, minute series

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
**| |	**| |	1	-0.227	-0.227	2324.2	0.000
*| |	*| |	2	-0.095	-0.154	2728.8	0.000
| |	*| |	3	0.003	-0.061	2729.3	0.000
| |	| |	4	-0.012	-0.045	2735.9	0.000
*| |	*| |	5	-0.095	-0.124	3140.6	0.000
| |	*| |	6	-0.004	-0.075	3141.2	0.000
| |	| |	7	0.001	-0.055	3141.3	0.000
| |	| |	8	-0.001	-0.038	3141.4	0.000
| |	| |	9	-0.006	-0.036	3142.9	0.000
| |	| |	10	-0.004	-0.040	3143.6	0.000