Autocorrelation in Daily Stock Returns

ANTÓNIO CERQUEIRA

ABSTRACT
This paper examines the roles of spread and gradual incorporation of common information in the explanation of the autocorrelation in daily stock returns. The study develops a theoretical model and provides empirical evidence using an innovative test procedure. The data used in the empirical study comes from the Center for Research in Security Prices (CRSP) daily return files for NYSE stocks. The development of the theoretical model predicts that the autocorrelation in stock returns is the sum of two terms. The spread is the source of a negative term, while the gradual incorporation of common information generates a positive term. In the case of daily stock returns, the model predicts that, taking into account the sign, the autocovariance can be positive, negative or approximately zero depending on the absolute values of the two terms. The results of empirical tests, which use a modified version of the statistic variance ratio, seem to indicate that both the spread and the gradual incorporation of common information are relevant to explain the dynamics of daily prices.

This study examines, both theoretically and empirically, the autocorrelation in daily stock returns using an approach that takes into account the trading mechanism and the specific and common components of information. In the case of daily stock prices, the study proposes that the statistical properties of price changes depend on the interaction between the spread effects and the effects due to the gradual incorporation of common information. An important contribution of this study is providing an innovative test procedure that minimizes potential biases arising from unwanted factors. For each firm in the sample, and over the same period, this procedure compares two variance ratio statistics, as described in section V. Another important feature of the study is the application to individual daily stock returns, which requires a large amount of data, whereas most of the works on serial correlation analyse portfolio returns or longer time horizons. Other
recent studies also examine the autocorrelation in short-term individual stock returns. This is the case of Avramov, Chordia and Goyal (forthcoming). However, they use a different approach and analyze weekly and monthly returns.

Stock price changes reflect new market-level and firm-level information. It is likely that common information, which refers to the market-level information, has a significant impact on the dynamic properties of stock prices, because several empirical works found that stock price changes have informational implications for other assets in the same sector, market or even geographically separated markets, for example, Forbes and Rigobon (2002). Moreover, several studies suggest that the gradual incorporation of marketwide information causes serial correlation in short-term stock returns, for example, Chordia and Swaminathan (2000), Sias and Starks (1997), and Lo and MacKinlay (1990).

In this study, the theoretical model uses the Glosten (1987) market microstructure framework, because microstructure models have been used to analyze stock transactions, O’Hara (1999) and Madhavan (2000). However, in order to incorporate into the model the assumptions described in the sections that follow, several modifications have had to be made to the Glosten (1987) framework. Given that the study focuses on the relevance of the information components and of the gradual incorporation of common information to the interpretation of the statistical properties of stock prices, then the stock price model must explicitly describe the incorporation of common and specific information. Another major modification is related to the composition of the spread, which is based on the empirical findings of Huang and Stoll (1997). Additionally, the study proposes a simple process of diffusion of information in a market operated by a competitive market maker. The model predicts that the autocovariance in stock returns is the sum of two terms. The spread is the source of a negative term, while the gradual incorporation of common information generates a positive term. In the case of daily stock returns, the model predicts that, taking into account the sign, the autocovariance can be positive, negative or approximately zero depending on the absolute values of the two terms.
Several papers on serial correlation investigate the presence of abnormal returns using trading strategies, for example momentum and contrarian, that exploit the predictability of stock returns based on past returns. This study proposes that the statistical patterns in stock returns depend crucially on the time horizon. For example, several studies have found positive autocorrelation in medium-term stock returns, while long-term stock returns exhibit negative autocorrelation. Additionally, this study suggests that we can expect to find significant differences between the dynamic properties of individual stock returns and portfolio returns. The empirical tests in the study, which use daily returns, document a positive and substantial autocorrelation in stock index returns, while individual stock returns exhibit a weak autocorrelation.

The empirical study examines the relationship between the autocorrelation in daily stock returns and the relevance of common information. The empirical tests use CRSP data on daily returns for NYSE stocks. The results of the theoretical model show that the autocorrelation in stock returns has two terms: the term related to the spread and the term due to the gradual incorporation of common information. The term related to the spread is negative while the other is positive, and the second term is expected to be higher when common information is more significant. In the study, the CRSP Index return is the variable used to represent the importance of common information. Therefore, the autocorrelation in daily stock returns is expected to be higher when the absolute value of the CRSP Index return is higher, because of the positive sign of the autocorrelation term associated with the incorporation of common information. As regards the meaning of high autocorrelation it is important to note that, in this study, a zero autocorrelation is higher than a negative autocorrelation.

The choice of an appropriate statistic to test the serial correlation in stock returns is based essentially on the criterion proposed by Richardson and Smith (1994). They analyze a general class of test statistics that are linear combinations of the autocorrelations coefficient estimators, such as the variance ratio (Lo and MacKinlay), J-statistic (Jegadeesh) and multiperiod autocorrelation estimator (Fama and French). They found that, given an alternative distribution for stock
returns, the most powerful test statistic places its relative weights on the autocorrelation estimators in a one-for-one correspondence with the magnitude of the true autocorrelation. The criterion suggests the use of the Lo and MacKinlay (1988) variance ratio, given that our theoretical model predicts that the magnitude of the autocorrelations is decreasing in the order of the autocorrelation. The Lo and MacKinlay variance ratio statistic is sensitive to serial correlation and robust to time-varying volatilities and deviations from normality, and has been used to detect deviations of stock returns from the random walk model.

A main contribution of this study is providing an innovative test procedure that minimizes potential biases arising from unwanted factors. For each firm in the sample, and over the same period, this procedure compares two variance ratio statistics. However, this procedure implies the use consecutive and non-consecutive observations whereas the Lo and MacKinlay (1988) statistic is based on consecutive observations. Therefore, a new expression for the estimator of the multiperiod variance, which uses overlapping observations, is proposed. To confirm the accuracy of this expression it is also shown that we can obtain the Lo and MacKinlay estimator in the case where there is no discontinuity in observations.

The results of empirical tests seem to indicate that both the spread and the gradual incorporation of common information are relevant to explain the dynamics of daily prices. The first test applies the variance ratio statistic to the CRSP Index daily returns. The results of the test show that the autocorrelation in the Index return is positive and significant. When applied to individual stock returns, the tests show a weak autocorrelation. Moreover, the sign of the autocorrelation is not always the same, and frequently the absolute values of the autocorrelations do not lead to the rejection of the null hypothesis that the autocorrelations are different from zero.

Another empirical test uses the median of the absolute values of the Index return to divide the individual stock return observations into two sets and, for each one of the two sets, the variance ratios in the daily individual stock returns are calculated. After that, tests are performed to investigate the null hypothesis that
the mean variance ratio is equal between the two sets. The observed ANOVA F-statistic and the probability $p$ lead to the rejection of the null hypothesis, suggesting that the mean variance ratio is higher when common information is more relevant. This test uses 2-period variance. Similar tests were performed using 3-period variance and 4-period variance, leading to similar test results. Another test, which is based on the two-period variance, uses more extreme values of the CRSP Index return to split the stock returns into two sets. In this test the 66th percentile and the 33rd percentile are used to divide the return observations into two sets. This test leads to similar results.

Assuming the results of the theoretical model that the autocorrelation in stock returns has two terms: the term related to the spread and the term due to the gradual incorporation of common information and that the term related to the spread is negative while the other is positive, then the autocorrelation in daily stock returns is expected to be higher when the absolute values of the CRSP Index return are higher. The results of the empirical tests are consistent with the theoretical model. Moreover, the more stronger evidence is provided in the case where the 2-period variance is used, consistent with the hypothesis that common information is impounded into the stock price during the disclosure day and during the next day, Brennan, Jegadeesh and Swaminathan (1993).

The paper is organized as follows. Section I analyzes de gradual incorporation of common information. Section II focus on the serial correlation in stock returns. Section III presents the theoretical model. Section IV analyzes the statistic used in the empirical tests. The data and test methodology are presented in section V. Section VI presents the results of the empirical tests. Section VII concludes.

I. Gradual Incorporation of Common Information

Several studies of the autocorrelation in short-term stock returns, daily and intradaily, assume that returns reflect the dynamic effects associated with the spread and with changes in market participants’ expectations. Glosten (1987) examines the autocorrelation in stock returns using a model of price formation. More specifically, the Glosten (1987) model investigates the relationship between
the statistical properties of transaction prices and the spread, and uses a market microstructure framework where the stock price is the sum of two terms: the expected value conditional on the market maker information set and the term associated with the spread.

In this study it is proposed that in the case of daily prices, the statistical properties of price changes depend on the changes in market participants’ expectations, and on the interaction between the spread effects and the effects of gradual incorporation of common information. We can accept the relevance of the gradual incorporation of common information if we think that stock price reveals information to the rest of the market, and that at least part of this information has value implications for other firms in the same market or in the same industrial sector. After an informational event, each stock price will reflect part of the information related to the event. This first stage is followed by price adjustments based on information revealed by other securities price changes. A related question is whether there is a linkage between geographically separated markets concerning international transmission of price variations. Some empirical studies report that when volatility is high, the price variations in major markets tend to be correlated, but the correlation appears to be asymmetric with US markets leading other countries’ markets.

A number of studies on gradual incorporation of information into stock prices propose price models with imperfect incorporation of information. For example, the Amihud and Mendelson (1987) work captures not only the gradual incorporation of information, but also the overreaction of prices to information. The most common explanations for the gradual incorporation of information are based on nonsynchronous trading, on transaction costs or on the price adjustments made by the market maker based on the observation of the market reaction to the informational event. However, other alternative approaches have been used to explain the gradual incorporation of information and the consequent cross-serial correlation in stock returns or the autocorrelation in stock returns. This section continues with a description of several alternative interpretations of the imperfect incorporation of information into prices.
The Sias and Starks (1997) study is also consistent with gradual incorporation of information. The study assumes that institutional investors spread block trades over consecutive trading days, and that this trading is the source of a positive autocorrelation in stock returns. They argue that institutional investors use this strategy to hide their information or to minimize transaction costs.

The Daniel, Hirshleifer and Subrahmanyam (1998) behavioral model is based on investors overconfidence and biased self-attribution. They use the investors psychological reactions to explain overreaction and underreaction in financial markets. In the study the overreaction is the source of a positive autocorrelation in stock returns, and the negative autocorrelation is an outcome of the subsequent price correction.

The De Long, Shleifer, Summers and Waldman (1990) work examines the importance of the uninformed trader’s activity to explain that stock prices may diverge from stock values. The article also explains the autocorrelation in returns in the case where uninformed investors use momentum strategies. The model assumes that rational traders react to information in an excessive way. For example, if the case of good news, they drive prices high since they know that in the future momentum traders will accept paying higher prices.

Bagnoli and Watts (1998) analyze private information taking into account the investors expectations about the time of the information disclosure. In the model, informed investors will trade based on private information until the time of the information dissemination. This trading pattern can explain the gradual incorporation of information.

Even in the case of public news, the gradual incorporation of information into stock prices seems to be a reasonable assumption. For example, in the case of earnings announcements, the interpretation of the public information requires some private information about the stochastic process of earnings to calculate the stock value. Moreover, the computational capacities of each agent are not unlimited and the agents have to correct their forecasts analyzing other investors’ actions.
In this study, the price model assumes that common information is impounded into stock prices by the market maker after having analyzed the market reaction to the information. This is consistent with the assumption that the market maker is competitive and takes into account the cross-security information aggregation. Moreover, if we take into account the empirical work of Brennan, Jegadeesh and Swaminathan (1993) this information must be impounded into prices during the disclosure day and next day. Brennan, Jegadeesh and Swaminathan (1993) found that most of the common information is impounded into stock prices during days 0 and 1 relative to the informational event.

II. Serial Correlation in Stock Returns

The serial correlation in stock returns is a central issue in different financial areas. For example, several studies that investigate the process followed by stock returns analyze serial correlation, and the understanding of the origins of the serial correlation, and the consequent predictability of stock returns based on past returns, are essential to examine the stock market efficiency hypothesis. Moreover, the investment strategies, for example momentum and contrarian strategies, are based on the serial correlation. It is important to note that serial correlation refers to the cross-serial correlation of individual stock returns, or to the autocorrelation in individual stock returns, or to the autocorrelation in portfolio returns.

During the last years, the nature of the process of stock prices has been intensively investigated. Several studies examine the statistical properties of stock prices in terms of a dichotomy between two different processes, for example, they investigate whether the process is chaotic or random walk. Most of the works assume a stochastic process and investigate whether prices follow a random walk or a different process.

The strong-form market efficiency implies that, in cases where information and trading costs are zero, the stock prices reflect all available information. Prices fail to incorporate all available information when the costs of information acquisition or the trading costs are taken into account, Grossman and Stiglitz (1980). Several works analyze the relationship between price changes and information arrivals. If
there is no special pattern in information disclosure then stock returns must exhibit no autocorrelation. This implies that investment strategies based on historical data must earn zero abnormal returns.

The majority of the empirical studies on market efficiency examine the random walk hypothesis for stock prices. These studies use different methods, which analyze, for example, the serial correlation in stock returns, or the expected returns, or the conditional heteroscedasticity. However, it is important to note that in cases where the random walk hypothesis is rejected this does not imply the rejection of the efficient market hypothesis, because the procedures that has been used to test the market efficiency are a joint test of market efficiency and of the asset pricing model. To test the market efficiency it is necessary to use a model that incorporates, for example, the investors’ preferences and their information structures. For example, Fama and French (1988) assume that the market is efficient and explain the autocorrelation in long term returns using a price model with a mean reverting component.

The results of the studies on serial correlation must be analyzed taking into account their different assumptions. Analyzing previous studies on serial correlation we found that the statistical properties of stock returns vary according to the length of the time period used to calculate the returns. For example, Sias and Starks (1997) analyze the serial correlation in daily returns; Lo and MacKinlay (1990) and Kadlec and Patterson (1999) analyze the weekly stock returns; Chan, Jegadeesh and Lakonishok (1996) use time periods between three and twelve months; Fama and French (1988) analyze the behavior of long term stock returns. We also found considerable differences in statistical properties between individual stock returns and portfolio returns. Most of the studies of serial correlation analyze portfolio return autocorrelations. The source of the autocorrelation in portfolio returns, however, remains the subject of much debate. Possible explanations include nonsynchronous trading, transaction costs, time-varying expected returns. This study investigates the autocorrelation in the individual stock returns. The serial correlation in individual returns is analyzed, for example, by Sias and Starks (1997).
In the case of short-term returns, empirical tests have documented a positive autocorrelation in portfolio returns. This is the case, for example, of the positive and substantial autocorrelation in stock index returns. The results of the empirical tests in this study, which use daily returns on the equally weighted CRSP stock index, are in agreement with those of previous works, showing that the autocorrelation in the index returns is positive and significant. Moreover, these empirical tests indicate that the first order autocorrelation is larger than higher order autocorrelations.

The studies on short-term individual stock returns investigating the statistical distribution parameters are few in number. Usually, the results of such empirical tests found a weak autocorrelation, they also indicate that the sign of the autocorrelation is not always the same and that, most of the times, the absolute value of the autocorrelation does not lead to the rejection of the null hypothesis of random walk. In the case of daily returns for NYSE stocks, the results of the empirical tests in this study lead to the same conclusions.

In the case of intermediate time horizons, for example Grundy and Martin (2001), Hong and Stein (1999), Chan, Jegadeesh and Lakonishok (1996), Jegadeesh and Titman (1993), most of the empirical studies found a positive autocorrelation in stock returns. This autocorrelation is consistent with the profitability of momentum strategies. However, the source of the profitability of momentum strategies remains an open question. A number of different explanations have been proposed, such as the market underreaction to firm-specific information, variations in risk exposure between the formation period and the investment period and lead-lag effect in the component of returns related to common information.

The DeBondt and Thaler (1985) and Fama and French (1988) models, which analyze the negative autocorrelation in long-term returns, differ with respect to the efficient market hypothesis. The DeBondt and Thaler (1985) study is based on the idea that investors are not always rational and that the evolution of their expectations shows an overreaction to news over long-term horizons. The approach proposed by Fama and French (1988), where mean-reversion in stock
prices is the source of the autocorrelation, is consistent with the efficient market hypothesis. Mean-reversion can be explained by business cycles and by changes in investors’ consumption-investment behavior.

III. Model

In this paper, the theoretical model is based on the Glosten (1987) framework. However, the assumptions used in this study require the introduction of several modifications. The major modifications, which are described in the rest of this section, are related to the composition of the spread, the focus on the components of information and on the gradual incorporation of common information, the role played by the market maker in the incorporation of information, the analysis of the process of diffusion of information between market participants and the potential application of the model to different time periods ranging from the high-frequency intraday to daily returns.

As in the model of Glosten (1987), the transaction price consists of two terms: the stock expected value conditional on the market maker information set and the term related to the spread. However, unlike the Glosten (1987) model, which uses the transaction costs and the asymmetric information components of spread, this study incorporates also the inventory costs component.

A main point in this paper is that the process of diffusion of the information plays an important role in the interpretation of the dynamic properties of stock prices, and that the diffusion of the information between market participants depends on the characteristics of the information. Different assumptions are made as regards the incorporation of information into prices. Additionally, the model uses a framework of gradual incorporation of common information with common information being impounded into prices during the information disclosure day and during the next day. In order to incorporate into the model the characteristics of the information diffusion process the study assumes that the market is operated by a competitive market maker that quotes the bid-price and the ask-price based on his expectations about the stock value. The market maker will update his expectations about stock value taking into account the transaction type, which
reveals part of the informed investor private information, and the publicly available information.

As regards common information, this study proposes that common information is incorporated into stock prices by the market maker, which behaves competitively. In setting quotes the competitive market maker incorporates the public information revealed during the current transaction period. The proposed gradual incorporation of common information is modeled by assuming the market maker’s ability to analyze the impact of the whole market price changes on a particular stock price. Then the change in stock expected value between two consecutive transaction periods depends on common information produced during the current transaction period, informed investors’ private information revealed by the transaction type, and common information produced during a previous period.

In the Glosten (1987) paper the stock price is the sum of two terms: the stock expected value and the term related to the spread. The term related to the spread includes two components of spread: the component due to asymmetric information and the component that takes into account the costs associated with the activity of the market maker as well as his normal profits. The first component produces permanent changes in transaction prices and the other component induces changes that affect only the price in the current period, as we can see from the Glosten (1987) formula,

\[
p_{k+n} = E(V|S_{k-1}) + \sum_{i=0}^{n} Z_{k+i} Q_{k+i} + C_{k+n} Q_{k+n}
\]

(1)

where \(p_{k+n}\) is the transaction price in period \(k+n\), \(E(V|S_{k-1})\) the expected value conditional on the information set of the market maker, \(Z\) the asymmetric information component of spread, \(C\) the transaction costs component of spread and \(Q = (1,-1)\) if the investor (buy, sell) one asset.

The specific assumptions described above can now be considered in order to generate the price model that will be used later in the empirical tests. The model proposed in this study assumes that stock price changes are the sum of two terms: an information related permanent term and a transitory term. The permanent term is assumed to include part of common information produced during the current
period, denoted by $\theta M_t$, $0 \leq \theta \leq 1$, part of common information produced during a previous period, denoted by $(1-\alpha)M_{t-i}$, $0 \leq \alpha \leq 1$, and the asymmetric information component of the spread $Z$. The transitory term includes the transaction costs and the inventory costs components of the spread. Given the competitive behavior of the market maker, the transaction price in period $t$ is given by,

$$p_t = E(V|S_{t-1}) + (1-\alpha)M_{t-i} + \theta M_t + \sum_{j=1}^{t-1} Q_j I + Q_t (Z + C + I) \quad (2)$$

where $E(V|S_{t-1})$ is the stock expected value conditional on the information set of the market maker, $Z$ the asymmetric information component of the spread, $C$ the transaction costs component of the spread, $I$ the inventory costs component of the spread, and $Q_t = (1, -1)$ if the investor (buys, sells) one share. The term $E(V|S_{t-1})$ refers to the expected value before the transaction that occurs in period $t$. The term $\sum_{j=1}^{t-1} Q_j I$ represents the accumulated effect associated with inventory holding costs.

The inventory costs component of the spread is used by the market maker to balance supply and demand across time. The market maker changes the quote midpoint depending on his previous inventory position to reduce the imbalance of buy and sell orders. Given that the inventory reversal is not immediate the impact on price lasts for several transaction periods and this explains the term $\sum_{j=1}^{t-1} Q_j I$ in expression (2).

From expression (2) it is easy to prove that the price change between two consecutive transaction periods is given by,

$$\Delta p_t = p_t - p_{t-1} = (1-\alpha)M_{t-i} + \theta M_t + Q_t (Z + C + I) - Q_{t-1} C$$

The above section on the gradual incorporation of information, proposes that common information produced during day $T$ is likely to be incorporated into stock price during day $T$ and day $T+1$. In a working paper, which is available upon request, we extended previous analysis in order to examine longer time
intervals. In what follows, \( T \) denotes the time interval including transaction periods \((T-1) \times n+1, (T-1) \times n+2, \ldots, T \times n\). For example, the interval 1 includes periods \( \{1, \ldots, n\} \). In this case the daily price change, using closing prices, is given by,

\[
\Delta p_T = p_{t+n} - p_t = (1 - \beta) M_{T-1} + \lambda M_T + \sum_{j=1}^{n} Q_{t+j}(Z + 1) + Q_{t+n} C - Q_t C
\]

\[
= (1 - \beta) M_{T-1} + \lambda M_T + \Delta S_T
\]

(3)

where \( t = (T-1) \times n \), the term \((1 - \beta) M_{T-1}, 0 \leq \beta \leq 1\) represents common information related to the time interval \( T - 1 \) and incorporated during \( T \), and \( \lambda M_T, 0 \leq \lambda \leq 1 \) is the common information related to \( T \) and incorporated during \( T \), and \( \Delta S_T \) represents the component of the price change related to the spread.

Many empirical studies use logarithmic returns. The use of logarithms is expected to linearize the rising of stock prices over time and to stabilize the variance of price changes. If we assume that in the above price change formula \( p_T = \ln p_T^* \) is the logarithm of the transaction price \( p_T^* \), the logarithmic return in day \( T \) is given by,

\[
r_T = \ln \frac{p_T^*}{p_{T-1}} = p_T - p_{T-1} = \Delta p_T
\]

(4)

The autocovariance in daily returns is given by,

\[
\text{Cov} (r_{T-1}, r_{T}) = \text{Cov} (\Delta p_{T-1}, \Delta p_{T})
\]

(5)

substituting the price changes above into (5) and calculating the covariance, we obtain,

\[
\text{Cov} (r_{T-1}, r_{T}) = \beta (1 - \beta) \text{Var} (M_{T-1}) + \text{Cov} (\Delta S_{T-1}, \Delta S_T)
\]

(6)

The calculations assume independence between different pieces of information and between information and spread components. We can show that \( \text{Cov} (\Delta S_{T-1}, \Delta S_T) < 0 \), proof available upon request. The positive term \( \beta (1 - \beta) \text{Var} (M_{T-1}) \) is larger when common information is significant and \( \beta \) is near 0.5. Given that the two terms have opposite signs, the autocovariance
expression is consistent with a positive or negative autocovariance and, in the case of similar absolute values, the autocovariance is expected to be approximately equal to zero. The empirical tests in this study try to investigate if, all other things being equal, the autocovariance in daily stock returns is higher when common information is more significant.

IV. Variance Ratio Statistic

In this study, the Richardson and Smith (1994) criterion has been used to select the appropriate test statistic. Using autocorrelations at different lags, they begin by describing a general class of test statistics that are linear combinations of consistent estimators of the autocorrelations. Richardson and Smith show that many of the statistics that have been used to test the random walk hypothesis fall into this class, as is the case for the variance ratio (Lo and MacKinlay), J-statistic (Jegadeesh) and multiperiod autocorrelation estimator (Fama and French). The underlying idea of all these test statistics is that, under the hypothesis that returns follow a random walk, the variance of the n-period return must be equal to n times the one period return variance. They found that, given an alternative hypothesis, the most powerful test statistic places its relative weights on the autocorrelation estimators in a one-for-one correspondence with the magnitude of the true autocorrelations. In the case of this study, the theoretical model predicts that the magnitude of the autocorrelations is decreasing in the order of the autocorrelation and this suggests the use of the variance ratio test, which places declining weights on the autocorrelation estimators.

The variance ratio statistic has been used to detect deviations of stock returns from the random walk model. Assuming n-period time intervals, the variance ratio is the sample n-period return variance divided by n times the sample single period return variance. Under the null hypothesis that returns follow a random walk, the correspondent population variance ratio is “1” for any value of n. Assuming that the one-period return variances \( \sigma^2(r_i) = \sigma^2(r_k), i \neq k \), and that the one-period return autocovariances \( \text{cov}(r_i, r_{i-\tau}) = \text{cov}(r_k, r_{k-\tau}), i \neq k \), then the n-period variance is defined as:
The variance ratio is given by,
\[
\sigma^2 \left( \sum_{\tau=0}^{n-1} r_{t-\tau} \right) = n\sigma^2(r_t) + 2\sum_{\tau=1}^{n-1} (n-\tau) \text{cov}(r_t, r_{t-\tau})
\]  
(7)

and the expression shows that the variance ratio is a linear combination of the autocorrelations and that for large samples the variance ratio equals one for all \(n\) if the returns are serially uncorrelated. The variance ratio will be less than one if there is negative autocorrelation and above one for positive autocorrelation.

The Lo and MacKinlay (1988) paper proposes a variance ratio statistic that is sensitive to serial correlation and robust to time-varying volatilities and deviations from normality. In the model the log-price is represented by,
\[
X_t = \ln P_t = \mu + X_{t-1} + \epsilon_t
\]  
(9)
where \(\epsilon_t\) is a random term and \(\mu\) an arbitrary drift. First they consider the case where \(\epsilon_t\) is i.i.d., \(N(0, \sigma^2_0)\). The variance estimator for the one-period log-return is given by,
\[
\hat{\sigma}^2_a = \frac{1}{nq} \sum_{k=1}^{nq} \left( X_k - X_{k-1} - \hat{\mu} \right)^2
\]  
(10)
where \(\hat{\mu}\) is a mean return estimator and \(nq+1\) is the number of observations. An alternative estimator can be obtained using the multiperiod return variance,
\[
\hat{\sigma}^2_b = \frac{1}{nq} \sum_{k=1}^{n} \left( X_{qk} - X_{qk-q} - q\hat{\mu} \right)^2
\]  
(11)

They found that the distribution of the statistic \(J_n(q) = \frac{\hat{\sigma}^2_b(q)}{\hat{\sigma}^2_a} - 1\) converges to a normal distribution,
\[
\sqrt{nq}J_n(q) \sim N[0,2(q-1)]
\]  
(12)

The use of overlapping observations allows for more observations and increases the power of the test. They define a new variance estimator,
They also present the following unbiased variance estimators and a new statistic,

\[ \sigma_q^2 = \frac{1}{nq} \sum_{k=q}^{nq} (X_k - X_{k-q} - q\hat{\mu})^2 \]  

They found that the statistic distribution converges to a normal distribution,

\[ \sqrt{nq} \hat{M}_r(q) \left[ 2(q-1)(q-1)/3q \right]^{1/2} \sim N[0,1] \]  

They derive a version of the test that is robust to heteroscedasticity. Denoting by

\[ \hat{\sigma}_c^2 = \frac{1}{m} \sum_{k=q}^{nq} (X_k - X_{k-q} - q\hat{\mu})^2 \] \hspace{1cm} \text{where} \hspace{1cm} m = q(nq + 1) \left( 1 - \frac{q}{nq} \right) \]  

\[ \hat{M}_r(q) = \frac{\hat{\sigma}_c^2(q)}{\hat{\sigma}_a^2} - 1 \]  

In this section a new statistic variance ratio is developed that takes into account the specific assumptions that will be used later in the empirical tests. As described later in the methodology of the empirical tests, the data samples use consecutive and non-consecutive observations whereas the Lo and MacKinlay (1988) statistic is based on consecutive observations. A new expression for the estimator of the
multi-period variance, which uses overlapping observations, is proposed. To confirm the accuracy of this expression, it is also shown that we can obtain the Lo and MacKinlay estimator in the case where there is no discontinuity in observations.

In this section, \( N \) denotes the total number of one-period (daily) return observations and \( T \) the total number of \( q \)-period return observations. In the case of consecutive observations the generation of returns uses overlapping observations. For example, in the sample including the daily log-returns \( \{ r_2, r_3, r_4, r_5, r_9, r_{10}, r_{11} \} \) and assuming that \( q = 3 \), then \( N = 7 \) and \( T = 3 \), being the 3-period log-returns \( R_1 = r_2 + r_3 + r_4 \); \( R_2 = r_3 + r_4 + r_5 \) and \( R_3 = r_9 + r_{10} + r_{11} \).

To obtain an estimator for the multi-period log-return variance we begin by calculating the expected value,

\[
E \left[ \sum_{i=1}^{T} (R_i - q \bar{R})^2 \right], \text{ where } \bar{R} = \frac{1}{N} \sum_{i=1}^{N} R_i \tag{21}
\]

Calculating the sum we obtain,

\[
\sum_{i=1}^{T} (R_i - q \bar{R})^2 = \sum_{i=1}^{T} \left( R_i^2 + q^2 \bar{R}^2 - 2R_i q \bar{R} \right) = \sum_{i=1}^{T} R_i^2 + q^2 \bar{R}^2 - 2q \bar{R} \sum_{i=1}^{T} R_i
\]

\[
= \sum_{i=1}^{T} R_i^2 + q^2 T \bar{R}^2 - 2q^2 T \bar{R}^2 = \sum_{i=1}^{T} R_i^2 - q^2 T \bar{R}^2 \tag{22}
\]

The expected value is,

\[
E \left[ \sum_{i=1}^{T} (R_i - q \bar{R})^2 \right] = E \left[ \sum_{i=1}^{T} R_i^2 \right] - q^2 T E \left[ \bar{R}^2 \right] \tag{23}
\]

and taking into account that \( E(x^2) = Var(x) + (E(x))^2 \), the above formula becomes,

\[
E \left[ \sum_{i=1}^{T} (R_i - q \bar{R})^2 \right] = \sum_{i=1}^{T} \left\{ Var(R_i) + \left[ E(R_i) \right]^2 \right\} - q^2 T \left\{ Var(\bar{R}) + \left[ E(\bar{R}) \right]^2 \right\}
\]
Under the random walk hypothesis for stock returns, \( \text{Var}(R_i) = q \text{Var}(r_i) = q N \text{Var}(\bar{r}) \) and,

\[
E\left[\sum_{i=1}^{T} (R_i - q \bar{r})^2\right] = T \text{Var}(R_i) + q^2 T \bar{r}^2 - \frac{q^2 T \text{Var}(R_i)}{q N} - q^2 T \bar{r}^2
\]

\[
= T \left(1 - \frac{q}{N}\right) \text{Var}(R_i)
\]

(24)

The estimator of the multiperiod return variance is given by,

\[
\frac{1}{T} \frac{1}{1 - \frac{q}{N}} \sum_{i=1}^{T} (R_i - q \bar{r})^2
\]

(25)

and the estimator of the one-period return variance is given by,

\[
\hat{\sigma}_c^2 = \frac{1}{q T} \frac{1}{1 - \frac{q}{N}} \sum_{i=1}^{T} (R_i - q \bar{r})^2
\]

(26)

In the case of the Lo and MacKinlay estimator,

\( T = nq - q + 1 \) and \( N = nq \)

and we obtain the parameter \( m = q (nq - q + 1) \left(1 - \frac{q}{nq}\right) \) of the Lo and MacKinlay formula.

**V. Data and Test Methodology**

This study provides empirical evidence to test the validity of the model described in section III. The data come from the Center for Research in Security Prices (CRSP) daily return files for NYSE stocks, over the periods from January 1992 through December 1993 and from January 1996 through December 1997. The tests use data on individual stock daily returns and on the equally weighted CRSP Index daily returns. The use of this sample, instead of more recent data, has the advantage of avoiding the potential bias that could plausibly be attributed to the positive abnormal returns over the period (1998, 2000) and to the subsequent correction over the period (2001, 2003).
The study sample size is comparable to those of previous works on individual stock return autocorrelation, such as Amihud and Mendelson (1987) and Sias and Starks (1997). The Amihud and Mendelson paper examines the stochastic process of daily stock returns, using data from a sample of 30 NYSE stocks over the period 08 Feb 1982 / 18 Feb 1983. They compare the estimators of variance, normality and autocorrelation in stock returns based on the opening price to the same estimators related to the closing price. Sias and Starks (1997) examine the daily stock return autocorrelation using CRSP data for NYSE firms over a 15-year period. However, they split the sample into 30 sets while in our study only two sets are considered. For each one of these 30 sets they calculate the mean autocorrelation and the \( t \) statistic. Using the \( F \) statistic, and for similar capitalization firms, they test the null hypothesis that the mean autocorrelation is equal between sets with different levels of institutional ownership.

This study empirically examines the daily stock returns because we propose that this is an appropriate sampling period to detect the effects analyzed in the theoretical model: the spread effects and the effects related to the gradual incorporation of common information. For example, the intraday data can be used to analyze the spread effects, but this analysis is likely to be biased by the U-shaped curve of these returns. The most significant transactions for the daily returns occur at the beginning and at the end of the trading day. Furthermore, this approach is not appropriate to analyze the effects of the gradual incorporation of common information that we have described, which are expected to last for longer time periods. The use of longer sampling periods is not appropriate due to the overreaction of prices to information and to the effects related to shocks that affect the investors’ preferences. For example, Lo and MacKinlay (1990) and Jegadeesh and Titman (1993) analyze the contrarian strategy and they use the overreaction of prices to information to explain part of the performance of this strategy. The case related to changes in investor’s preferences is analyzed, for example, in the Fama and French (1988) paper.

The stock return in the CRSP files is calculated as follows:

\[
r'(t) = \frac{p(t)f(t) + d(t)}{p(t-1)} - 1
\]  
(27)
where \( p(t) \) denotes the stock price in period \( t \), \( d(t) \) is a cash adjustment for \( t \) that takes into account, for example, dividends, and \( f(t) \) is a price adjustment factor for \( t \) that takes into account, for example, stock splits. Our theoretical model analyzes stock price changes and, if we consider log-price changes, the stock log-return is given by,

\[
    r(t) = \ln \left[ \frac{p(t) f(t) + d(t)}{p(t-1)} \right] = \ln \left[ r'(t) + 1 \right].
\]

(28)

The results of the theoretical model show that the autocorrelation in daily stock returns is the sum of two terms: the term related to the spread and the term due to the gradual incorporation of common information. The term related to the spread is negative while the other is positive, and the second term is expected to be higher when common information is more significant. In this study, the CRSP Index return is the variable that represents the importance of common information. This method is used, for example, in the Chang, Pinegar and Ravichandran (1998) model, where changes in the market value of a portfolio of large firms represent the importance of macroeconomic information. Therefore, the autocorrelation in daily stock returns is expected to be higher when the absolute value of the CRSP Index return is higher, because of the positive sign of the autocorrelation term associated with the incorporation of common information. Note that, in this study, a zero autocorrelation is higher than a negative autocorrelation.

In this study, statistical tests have been developed to analyze the relationship between the autocorrelation in daily stock returns and the importance of common information. However, given that the arrival rate and relevance of information are highly variable over a time period, an innovative test procedure had to be developed. This procedure consists in dividing the return observations for each security according to specific rules, based on the magnitude of common information.

One of the empirical tests uses the median of the absolute values of the Index return to divide the observations into two sets and, for each one of the two sets, the variance ratios in the individual daily stock returns are calculated. After that,
tests are performed to investigate if the mean variance ratio is equal between the two sets. More formally, $H_0$ being the null hypothesis and $H_1$ the alternative hypothesis, we have the following test:

$H_0$: The mean variance ratio is equal between the two sets.

$H_1$: The mean variance ratio is higher in the set related to the higher absolute values of the CRSP Index return.

Table I provides an example where the 2-period returns are used. The stock return observation $r(i)$ is selected for the $H$ sample if the absolute value of the return Index $m(i)$ is higher than the median in the period $i$ or in the $i - 1$ period.

Table I

Selection of observations

The observation “i” is selected for the “H” sample if the absolute value of the return Index is higher than the median in the period “i” or in the “i-1” period.

| $r(t)$ | $m(t)$ | $|m(t)| > median$ | $H$ sample |
|--------|--------|--------------------|------------|
| 0,0134235 | 0,00930259 | Yes | 0,0134235 |
| 0,00652866 | 0,00652866 | Yes | 0 |
| -0,00487255 | -0,00487255 | Yes | 0 |
| 0,0066788 | 0,0066788 | Yes | 0 |
| -0,00668932 | 0,00352117 | Yes | -0,00668932 |
| 0,01333371 | -0,0116837 | Yes | 0,01333371 |
| -0,0133335 | 0,00467902 | Yes | 0 |
| 0,00668858 | 0,0006892 | Yes | 0,00668858 |
| -0,00668932 | -0,00311258 | | |
| -0,00668932 | 0,00250698 | | |
| -0,00668932 | 0,0005169 | | |
| 0,00668858 | 0,00271558 | | |
| 0 | 0,00332903 | | |
| 0 | 0,0034181 | Yes | 0 |
| -0,00668932 | 0,00104543 | Yes | -0,00668932 |
| 0,00668858 | 0,00474922 | Yes | 0,00668858 |
| -0,00668932 | -0,00041375 | Yes | -0,00668932 |

We have seen that the statistic $\hat{M}_r(q)$ converges to a normal distribution,

$$\sqrt{nq} \hat{M}_r(q)/\sqrt{\hat{\theta}} = \sqrt{\frac{\hat{M}_r(q)}{\sum_{j=1}^{q-1} \left( \frac{2(q-j)}{q} \right)^2}} \sim \mathcal{N}[0,1]$$ (29)
where,

\[
\hat{M}_r(q) = \frac{\sigma_c^2(q)}{\sigma_a^2} - 1
\]

(30)

We use the corresponding data, for example de “H” sample data, to obtain the unbiased estimator of the one-period return variance \( \sigma_a^2 \) and the estimator of the q-period return variance,

\[
\sigma_c^2 = \frac{1}{qT} \left( 1 - \frac{q}{N} \right) \sum_{i=1}^{T} (R_i - q\bar{r})^2
\]

(31)

VI. Empirical Results

The theoretical model examines the autocovariance in stock returns, while the variance ratio statistic, which is used in the empirical tests, is a linear combination of the estimators of the autocorrelations. The variance of the individual stock return is expected to change according to the volatility of the Index return, and it is necessary to take this into account when analyzing the results of the empirical tests. A simple way of showing the dependence of the two variances is to use the median of the absolute values of the Index return to divide the individual stock return observations into two sets, as described in section V. After that, for each one of the two sets, the variance in the daily individual stock returns is calculated. In the total sample, including 1625 (2016) stocks for the periods 1992/93 (1996/97), we found that the individual stock return variance is higher for the subsample related to high Index returns in 1238 (1674) cases. The dependence between these variances represents an additional difficulty in testing the hypothesis that the autocovariance in stock returns is higher when the absolute values of the Index return are high. In fact, given a constant autocovariance, the autocorrelation is a decreasing function of the variance. In spite of this adverse effect, the study supports the hypothesis that the positive component of the autocovariance is higher when the market Index returns reveal that common information is more relevant. However, this adverse effect influences the results of the tests that use more extreme values of the Index return, for example, when the
33rd percentile and the 66th percentile are used to divide the observations into two sets.

The first test uses the variance ratio statistic to examine the equally weighted CRSP Index daily returns. Three variance ratio estimators are calculated using either two, three or four period variances. The normalized estimator $Z$ is calculated by dividing the variance ratio estimator by the respective variance estimator. The test results are shown in table II and table III. All the $Z$’s are statistically significant at 0.01 significance level (1992/93) and 0.2 significance level (1996/97). However, the $Z$’s are inversely related to the order of the statistic.

<table>
<thead>
<tr>
<th>Table II</th>
<th>CRSP Index Daily Returns (92/93)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Three variance ratio estimators are calculated using either two, three or four period variances. The normalized estimator $Z$ is calculated by dividing the variance ratio estimator by the respective variance estimator. All the $Z$’s are statistically significant at 0.01 significance level. However, the $Z$’s are inversely related to the order of the statistic.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CRSP Index variance ratios (92/93)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Var. rat. Ord.2</td>
</tr>
<tr>
<td>$Z$</td>
</tr>
<tr>
<td>Var. rat. Ord.3</td>
</tr>
<tr>
<td>$Z$</td>
</tr>
<tr>
<td>Var. rat. Ord.4</td>
</tr>
<tr>
<td>$Z$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table III</th>
<th>CRSP Index Daily Returns (96/97)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Three variance ratio estimators are calculated using either two, three or four period variances. The normalized estimator $Z$ is calculated by dividing the variance ratio estimator by the respective variance estimator. All the $Z$’s are statistically significant at 0.2 significance level. However, the $Z$’s are inversely related to the order of the statistic.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CRSP Index variance ratios (96/97)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Var. rat. Ord.2</td>
</tr>
<tr>
<td>$Z$</td>
</tr>
<tr>
<td>Var. rat. Ord.3</td>
</tr>
<tr>
<td>$Z$</td>
</tr>
<tr>
<td>Var. rat. Ord.4</td>
</tr>
<tr>
<td>$Z$</td>
</tr>
</tbody>
</table>

The statistic variance ratio that uses the two-period variance is approximately equal to the first-order autocorrelation in returns. Therefore, the variance ratios in
tables II and III show that the first-order autocorrelation in the CRSP Index return is the most relevant and that this autocorrelation is positive and significant. These results are consistent with those of other empirical studies on portfolio return autocorrelation, such as Conrad and Kaul (1988). The results are also consistent with the findings of Brennan, Jegadeesh and Swaminathan (1993). Using CRSP data for NYSE, AMEX and NASDAQ stocks, they found evidence that common information is impounded into prices over the disclosure day and the next day. Therefore, assuming that the gradual incorporation of common information can explain part of the serial correlation in individual stock returns and part of the autocorrelation in portfolio returns, then the first-order autocorrelation in the CRSP Index return is likely to be the most relevant.

The study also empirically examines the daily stock return distributions over the sample periods 1992/93 (1996/97), by analyzing the mean, standard deviation, skewness and kurtosis. The skewness coefficient is positive for 979 (1147) stocks and negative for 646 (869) stocks. The kurtosis coefficient is higher than three, which is the kurtosis of the normal distribution, for 1586 (1969) stocks and lower than three for 39 (47) stocks. The kurtosis of the sample stock returns is almost always higher than the kurtosis of the normal distribution.

A. Autocorrelation in daily individual stock returns: two-period variance

As an example, the calculations that follow, in the case of 2-period time intervals, are performed for just one of the companies for the periods 1992/93 and 1996/97. The company returns are reported in appendix B, available upon request, and the formulas are presented in the section on the Variance Ratio statistic.

The stock return observations associated with periods of more significant common information are selected taking into account the position of the absolute value of the CRSP Index return relative to the median. Assuming that \( r(t) \) is the individual stock daily return on day \( t \), then the 2-period stock return \( R(t) = r(t - 1) + r(t) \) is selected if the absolute value of the Index return on day \( t - 1 \) is higher than the median, \( \text{ABS}(m(t - 1)) > \text{median} \).
A similar method is used to select the stock return observations associated with periods of less significant common information. The 2-period stock return \( R(t) = r(t-1) + r(t) \) is selected if the absolute value of the Index return on day \( t-1 \) is lower than the median, \( \text{ABS}(m(t-1)) < \text{median} \).

**Table IV**

**Variance ratio calculations**

The variance ratio calculations shown in this table, in the case of 2-period time intervals, are performed for just one of the sample companies. The company returns are reported in appendix B, available upon request, and the formulas are presented in the section on the Variance Ratio statistic.

**Company:** McDonalds Corp

**Period:** 1992 / 93

(a) Stock log-return subsample for \( \text{ABS} \) (Index return) > median

(b) Stock log-return subsample for \( \text{ABS} \) (Index return) < median

<table>
<thead>
<tr>
<th></th>
<th>Stock return Subsample (a)</th>
<th>Stock return Subsample (b)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( q )</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>( N )</td>
<td>378</td>
<td>379</td>
</tr>
<tr>
<td>( T )</td>
<td>252</td>
<td>253</td>
</tr>
<tr>
<td>( \overline{r} )</td>
<td>1.0648055 \times 10^{-3}</td>
<td>9.3027712 \times 10^{-4}</td>
</tr>
<tr>
<td>( \sigma_2^2 )</td>
<td>2.2280776 \times 10^{-4}</td>
<td>1.9172624 \times 10^{-4}</td>
</tr>
<tr>
<td>( \sigma_2^2 (2) )</td>
<td>2.4513469 \times 10^{-4}</td>
<td>1.5441600 \times 10^{-4}</td>
</tr>
</tbody>
</table>

**Similar calculations** were performed for all the global sample companies, comprising a total of 1625 + 2016 NYSE stocks. Furthermore, a third variance
ratio estimator that uses the total return observations for each stock, has been calculated. The results of the tests show a weak autocorrelation in individual stock returns, show that the sign of the autocorrelation is not always the same and that the absolute values of the autocorrelations do not lead to the rejection of the null hypothesis that the autocorrelations are different from zero. Similar empirical results on the sign and the magnitude of the autocorrelation in individual stock returns were found, for example, by Amihud and Mendelson (1987).

If we consider the global sample with the three variance ratio estimators for each stock we obtain three sets of variance ratio estimators. For each set, the mean of the variance ratio statistic was near zero mainly due to the presence of positive and negative variance ratios. The data related to the absolute values of the CRSP Index return above the median showed positive variance ratios for 894 (888) stocks and negative for 731 (1128) stocks. The data related to the absolute values of the CRSP Index return below the median showed positive variance ratios for 579 (602) stocks and negative for 1046 (1414) stocks. In the case of the total return observations the statistic was positive for 772 (776) stocks and was negative for 853 (1240) stocks.

The study tests the null hypothesis that the mean variance ratio is equal between the subsamples related to the high absolute values of the CRSP Index return and to the low absolute values of the CRSP Index return. As we will see later, the results of the tests do not support the null hypothesis, suggesting that the mean variance ratio is higher when common information is more relevant. These findings are consistent with the theoretical model. Additionally, if we take into account the total sample with the 1625 (2016) stocks, the estimated z’s lead to the rejection of the null hypothesis that the individual stock return autocorrelations are zero for 513 (664) stocks at 0.05 significance level and for 345 (469) stocks at 0.01 significance level.
Figure 1. Variance ratio distribution in the subsample related to the absolute values of the CRSP Index return above the median, using the two-period variance (1992/93).

Figure 2. Variance ratio distribution in the subsample related to the absolute values of the CRSP Index return above the median, using the two-period variance (1996/97).

More detailed results of the main test are presented in figures 1, 2, 3 and 4. Figures 1 and 2 show the variance ratio distribution in the subsample of daily stock returns related to the absolute values of the CRSP Index return above the median. Figures 3 and 4 show the variance ratio distribution in the subsample of daily stock returns related to the absolute values of the CRSP Index return below the median. The analysis of these distributions suggests that the mean variance ratio is higher for the first subsample. Next, we provide stronger support for this finding.
Figure 3. Variance ratio distribution in the subsample related to the absolute values of the CRSP Index return below the median, using the two-period variance (1992/93).

Figure 4. Variance ratio distribution in the subsample related to the absolute values of the CRSP Index return below the median, using the two-period variance (1996/97).

The mean variance ratio for the period 1992/93(1996/97) is -0.001738(-0.040664) in the subsample related to the high absolute values of the Index return and -0.056295(-0.083306) in the subsample related the low absolute values of the Index return. Moreover, the analysis of the variance ratios shows that the variance ratios related to the first subsample are higher for 1077(1286) stocks in a total of 1625(2016) stocks. The mean autocorrelation is not significant, which is in part due to the existence of positive and negative autocorrelations and this effect is consistent with the results of the theoretical model.

Table V
Test of the null hypothesis

The F statistic and the p value lead to the rejection of the null hypothesis that the mean variance ratio is equal between the data set related to the absolute values of the Index return higher than the median and the data set related to the absolute values of the Index return below the median. Two-period variance (1992/93).
Anova: Single Factor (92/93)

**SUMMARY**

<table>
<thead>
<tr>
<th>Groups</th>
<th>Count</th>
<th>Sum</th>
<th>Average</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rend. Inf.</td>
<td>1625</td>
<td>-91,4796</td>
<td>-0,0562951</td>
<td>0,02070512</td>
</tr>
<tr>
<td>Rend. Sup.</td>
<td>1625</td>
<td>-2,8238</td>
<td>-0,0017377</td>
<td>0,02560174</td>
</tr>
</tbody>
</table>

**ANOVA**

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>SS</th>
<th>Df</th>
<th>MS</th>
<th>F</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between Groups</td>
<td>2,418414752</td>
<td>1</td>
<td>2,41841475</td>
<td>104,451681</td>
<td>3,72E-24</td>
</tr>
<tr>
<td>Within Groups</td>
<td>75,20234246</td>
<td>3248</td>
<td>0,02315343</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>77,62075721</td>
<td>3249</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The test of the null hypothesis that the mean variance ratio is equal between the two subsamples uses the analysis of variance. The analysis of variance allows one to determine if there are significant differences between the two sets. For example, this technique is used to test the equality of two means, comparing the variation between the two data groups to the variation within each particular group. The ratio of the two variations is known to follow an $F$ distribution. The observed value of $F$ and the probability $p$ lead to the rejection of the null hypothesis.

Tables V and VI report the results of this test.

**Table VI**

**Test of the null hypothesis**

The F statistic and the $p$ value lead to the rejection of the null hypothesis that the mean variance ratio is equal between the data set related to the absolute values of the Index return higher than the median and the data set related to the absolute values of the Index return below the median. Two-period variance (1996/97).

Anova: Single Factor (96/97)

**SUMMARY**

<table>
<thead>
<tr>
<th>Groups</th>
<th>Count</th>
<th>Sum</th>
<th>Average</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rend. Inf.</td>
<td>2016</td>
<td>-167,94488</td>
<td>-0,083306</td>
<td>0,0227365</td>
</tr>
<tr>
<td>Rend. Sup.</td>
<td>2016</td>
<td>-81,978787</td>
<td>-0,0406641</td>
<td>0,0274768</td>
</tr>
</tbody>
</table>

**ANOVA**

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>SS</th>
<th>Df</th>
<th>MS</th>
<th>F</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between Groups</td>
<td>1,832879168</td>
<td>1</td>
<td>1,83287917</td>
<td>73,003861</td>
<td>1,81E-17</td>
</tr>
<tr>
<td>Within Groups</td>
<td>101,1796222</td>
<td>4030</td>
<td>0,02510661</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>103,0125014</td>
<td>4031</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The results of this test are consistent with the theoretical model where the positive term of the autocorrelation in daily returns is directly related to the relevance of common information. A similar procedure is used in the Sias and Starks (1997) paper where they show that the autocorrelation in daily returns for NYSE stocks is an increasing function of the level of institutional ownership. They use the \( F \) statistic to test the null hypothesis that the mean autocorrelation is equal between the data set related to the higher level of institutional ownership and the data set related to the lower level of institutional ownership. The average of the autocorrelations was 0.0468 for the first set of data and -0.0439 for the second set of data.

**B. Autocorrelation in daily individual stock returns: three-period variance**

In this section the stock return observations are divided into two sets according to the absolute values of the CRSP Index return. The median of the absolute values of the CRSP Index return is calculated. For each stock in the sample, the daily returns \( r(t) \), \( r(t-1) \) and \( r(t-2) \) are selected for inclusion in the first set if the absolute value of the CRSP Index return \( m(t-2) \) is higher than the median and the remaining observations belong to the second set. A third set composed of all stock return observations is also used.

In the set that includes all the return observations, the variance ratio statistic was positive for 696(709) stocks and was negative for 929(1307) stocks over the periods 1992/93(1996/97). In the set related to the absolute values of the CRSP Index return above the median the statistic was positive for 773(832) stocks and was negative for 852(1184) stocks. In the set related to the absolute values of the CRSP Index return below the median the statistic was positive for 564(500) stocks and was negative for 1061(1516) stocks.

The tests described in the above section A were also carried out and are available upon request. The results of these tests are similar to those in section A that use the two-period variance and this is consistent with the findings of Brennan, Jegadeesh and Swaminathan (1993) that the first two days are the most important for the incorporation of common information. The study also tests the
null hypothesis that the mean variance ratio is equal between the sets related to the high CRSP Index variations and to the low CRSP Index variations. The results of the test do not support this null hypothesis, suggesting that the mean variance ratio is higher when common information is more relevant. The results of this test are consistent with the theoretical model where the positive term of the autocorrelation in daily returns is directly related to the significance of common information.

C. Autocorrelation in daily individual stock returns: four-period variance

In this section the stock return observations are divided into two sets according to the absolute values of the CRSP Index return. For each stock in the sample, the daily returns \( r(t) \), \( r(t-1) \), \( r(t-2) \) and \( r(t-3) \) are selected for inclusion in the first set of observations if the absolute value of the CRSP index return \( m(t-3) \) is higher than the median and the remaining observations are included in the second set.

When the four-period variance is used the results of the test are similar to the results we found when using the two-period variance and the three-period variance.

D. Autocorrelation in daily individual stock returns: using the 33rd percentile and the 66th percentile

This test, which is based on the two-period variance, uses more extreme values of the CRSP Index return to separate the stock returns into two sets. For each stock in the sample, the daily returns \( r(t) \), \( r(t-1) \) are selected for inclusion in the first set of observations if the absolute value of the CRSP Index return \( m(t-1) \) is higher than the 66th percentile and the daily returns \( r(t) \), \( r(t-1) \) are included in the second set of observations if the absolute value of the CRSP Index return \( m(t-1) \) is lower than the 33rd percentile.

| Table VII |
| Test of the null hypothesis |
| The F statistic and the p value lead to the rejection of the null hypothesis that the mean variance ratio is equal between the set related to the absolute values of the Index return higher than the 66th... |
percentile and the set related to the absolute values of the Index return lower than the 33rd percentile, two-period variance (1992/93).

Anova: Single Factor (92/93)

<table>
<thead>
<tr>
<th>Groups</th>
<th>Count</th>
<th>Sum</th>
<th>Average</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rend.Inf.</td>
<td>1625</td>
<td>-93,1490192</td>
<td>-0,0573225</td>
<td>0,0243529</td>
</tr>
<tr>
<td>Rend. Sup.</td>
<td>1625</td>
<td>2,784376442</td>
<td>0,0017135</td>
<td>0,0294477</td>
</tr>
</tbody>
</table>

ANOVA

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>SS</th>
<th>Df</th>
<th>MS</th>
<th>F</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between Groups</td>
<td>2,83175889</td>
<td>1</td>
<td>2,8317589</td>
<td>105,26874</td>
<td>2,5E-24</td>
</tr>
<tr>
<td>Within Groups</td>
<td>87,372116</td>
<td>3248</td>
<td>0,0269003</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>90,2038749</td>
<td>3249</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The test of the null hypothesis that the mean variance ratio is equal between the two sets uses the analysis of variance. The observed value of $F$ and the probability $p$ lead to the rejection of the null hypothesis. Tables VII and VIII report the results of this test.

Table VIII

<table>
<thead>
<tr>
<th>Test of the null hypothesis</th>
</tr>
</thead>
</table>
The $F$ statistic and the $p$ value lead to the rejection of the null hypothesis that the mean variance ratio is equal between the set related to the absolute values of the Index return higher than the 66th percentile and the set related to the absolute values of the Index return lower than the 33rd percentile. Two-period variance (1996/97).

Anova: Single Factor (96/97)

<table>
<thead>
<tr>
<th>Groups</th>
<th>Count</th>
<th>Sum</th>
<th>Average</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rend. Inf.</td>
<td>1625</td>
<td>-131,2537052</td>
<td>-0,0807715</td>
<td>0,02460529</td>
</tr>
<tr>
<td>Rend. Sup.</td>
<td>1625</td>
<td>-85,08217212</td>
<td>-0,05235826</td>
<td>0,02996528</td>
</tr>
</tbody>
</table>

ANOVA

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>SS</th>
<th>Df</th>
<th>MS</th>
<th>F</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between Groups</td>
<td>0,65594168</td>
<td>1</td>
<td>0,65594168</td>
<td>24,0401254</td>
<td>9,9E-07</td>
</tr>
<tr>
<td>Within Groups</td>
<td>88,6226067</td>
<td>3248</td>
<td>0,02728529</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>89,2785484</td>
<td>3249</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

VII. Conclusions

This study analyzes both theoretically and empirically the serial correlation in short-term stock returns. Serial correlation studies are likely to have important implications for market regulation, design of trading mechanisms, trading
strategies that exploit the predictability of stock returns, market efficiency, spread, volatility, trading volume and liquidity.

A main point of the study is that the selection of the model to examine the statistical patterns in stock returns depends crucially on the time horizon. In the case of daily prices, the study proposes that the serial correlation in individual stock returns depends on the interaction between the spread effects and the effects due to the gradual incorporation of common information.

The statistical tests were developed to analyze the relationship between the autocorrelation in daily stock returns and the relevance of common information. However, given that the arrival rate and relevance of information are highly variable over a time period, an innovative test procedure had to be developed. This procedure consists in dividing the return observations for each security according to specific rules, based on the magnitude of common information. The tests use CRSP data on daily returns for NYSE stocks. The results of the tests are consistent with the conclusion of the model that the positive term of the autocorrelation in daily stock returns is higher when common information is more significant. The results of empirical tests seem to indicate that both the spread and the gradual incorporation of common information are relevant to explain the dynamics of daily prices. Moreover, the stronger evidence was observed in the case where the 2-period variance is used, and this is consistent with the hypothesis that common information is impounded into NYSE stock prices on the disclosure day and on the next day.

It would be interesting to perform additional tests of this model within the McQueen, Pinegar and Thorley (1996) context. They document that prices of small stocks react quickly to the announcement of unfavourable news while reacting slowly to good news. Furthermore, our model predicts that the autocorrelation in daily stock returns has two terms: the spread term and the term related to gradual incorporation of common information. Therefore, the second term is expected to be larger after the announcement of favourable news.

Appendix A
Taking an equally weighted portfolio of $n$ stocks, the portfolio return autocovariance is given by,

$$\text{Cov}\left( \frac{1}{n} \sum_i r_{i,t}, \frac{1}{n} \sum_k r_{k,t+\tau} \right) = \frac{1}{n^2} \sum_i \sum_k \text{Cov}\left( r_{i,t}, r_{k,t+\tau} \right)$$

$$= \frac{1}{n^2} \left( n^2 - n \right) \text{Cov}\left( r_{m,t}, r_{n,t+\tau} \right) + \frac{1}{n} \text{A cov}\left( r_{m,t} \right)$$

$$= \text{Cov}\left( r_{m,t}, r_{n,t+\tau} \right) + \frac{1}{n} \text{A cov}\left( r_{m,t} \right) - \frac{1}{n} \text{Cov}\left( r_{m,t}, r_{n,t+\tau} \right)$$

(A1)

where $m \neq n$, $\text{Cov}$ represents the mean cross-serial covariance and $\text{A cov}$ the mean autocovariance. Taking limits as $n$ approaches infinity we see that the portfolio return autocovariance converges to the average cross-serial covariance of individual returns.

REFERENCES


Avramov, Doron, Tarun Chordia, and Amit Goyal, Forthcoming, Liquidity and autocorrelations in individual stock returns, *Journal of Finance*.


McQueen, Grant, Michael Pinegar, and Steven Thorley, 1996, Delayed reaction to good news and the cross-autocorrelation of portfolio returns, *Journal of Finance* 51, 889 – 919.