INTRA-DAY CO-MOVEMENT AMONG STOCK RETURNS IN US BANKS

by

Shadi Najiazar

B.Sc., Electrical Engineering – Amir Kabir University of Technology, 1996

PROJECT SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF BUSINESS ADMINISTRATION IN GLOBAL ASSET AND WEALTH MANAGEMENT

In the
Faculty of Business Administration

SIMON FRASER UNIVERSITY

Summer 2009

APPROVAL

Name: **Shadi Najiazar**

Degree: **Master of Business Administration in Global Asset and Wealth Management**

Title of Project: **Intra-Day Co-Movement of Stock Returns In US Banks**

Supervisory Committee:

____

Dr. Peter Klein
Senior Supervisor
Professor

____

Dr. Evan Gatev

Second Reader

Assistant Professor

Date Approved: ____

TABLE OF CONTENTS

List of Tables:

Abstract:

This paper evaluates co-movement of asset returns at intra-day bases. We regressed the quote revision of one asset on lagged returns and trades of second asset and also on pseudo-error correction part that was used by Zebedee and Kasch-Haroutounian (2009).

We have used 5 minutes interval for trades of six large financial institutions in U.S. for the period of 27.03.2009 to 22.06.2009 and found that both lagged quote revisions and trade history of asset B provide valuable predictability for asset A returns. In addition, the suggested modified pseudo error correction mechanism of Zebedee and Kasch-Haroutounian (2009) highlights the predictive power of second asset by presenting new variables of spread and depth. Higher levels of spread is an indicator of asymmetric information in the market and the speed of adjustment increases with higher spread and depth; however this result is consistent in only half of the assets.

ACKNOWLEDGEMENTS

I would like to express my gratitude to my supervisor, Dr. Peter Klein, whose invaluable support, comments and encouragement helped me in all of my research and writing of this thesis and especial thanks to Dr. Evan Gatev for his kind support and stimulating suggestions.

Introduction:

There are number of studies supporting the existence of co-movement among returns of financial assets and the role of information in creating co-movement while some other empirical works reject existence of strong co-movement. For example, Deb et al (1996) implemented a GARCH(1,1) model on nine commodities with monthly time series data in two periods, 1960-85 and 1974-92 and found weak evidence of co-movement. To analyze the mechanics behind the co-movement of asset returns, researchers have taken different approaches to this issue and have analyzed co-movement from different perspectives.

Many researchers analyzed the co-movement of returns at international level. Karoly and Stulz (1996) examined the causes of cross-country stock return correlation. They used transaction data from 1988 to 1992 to build the overnight and intraday returns for a portfolio of Japanese stock and a matched-sample portfolio of U.S. stocks. They concluded that U.S. macroeconomic news affects the Yen/Dollar foreign exchange rate and treasury bill returns while industry announcements have insignificant effect on correlation of U.S. and Japanese return. Rober J.Shiller (1989) attempted to justify the co-movement of US and UK stock returns by co-movement of dividends and real interest rates and found a weak dependency.

To find the mechanism of flow of information and co-movement creation, Epps (1979) analyzed the effect of different sampling frequencies on correlation of asset returns and concluded that lags of an hour or more are needed for adjustment of stock prices. He chose four assets, Chrysler, Ford, GM and AMC in automotive industry and analyzed the correlation between those assets for 125 trading days. He tested correlation in a range of ten minutes to three days intervals and concluded that the correlation of asset returns decreases by changing sampling frequency from inter-day to intra-hour level and at least one hour is needed to observe significant correlation. For example, he monitored GM and Ford and observed 0.055 correlation between changes of log price in ten minutes while correlation was 0.645 in three days interval. He believed that the reason is non-stationary nature of asset price changes and in very short periods, correlation exists between short-term price changes of first asset and short-term price changes of all four assets in previous intervals.

To analyze the co-integration at microstructure level in today's organized financial market, researchers have focused on effects of information on different measurable parameters such as bid-ask spread and depth that are indicators of asymmetric information. Copeland and Galai (1983) concluded that bid-ask spread is a positive function of price and information affects the bid-ask price.

To find the effectiveness of flow of information between two assets, Chan (1992) tested the lead-lag relation between price movement of intraday cash market and stock index futures. He concluded that futures leads the cash index and cash index price lag futures price.

Chan (1993) examined the co-movement at transactional level by developing a two-period partial adjustment model. The concept of this model is existence of causality relationship between each market maker price setting and the observed noise which is followed by adjusting the price by observing other market makers prices. In first period, each market maker sets price based on the observed noise and then in second period, adjusts the price by comparing it to the price that is set by other market makers. In other word, each market maker adjusts the price based on the received feedback of other market makers opinion and this adjustment results in positive cross-correlation with lag. Dufour and Engle (2000) also concluded that when the time gap between transactions decreases, the speed of price adjustment increases and therefore presence of high numbers of informed traders makes the market more active and consequently less liquid.

Laura Veldkamp (2006) also evaluated the effect of information market on co-movement of asset prices and concluded that when investors set price for an asset based on available information, news about one asset affects the price of second asset and consequently prices of assets co-move.

Zebedee and Kasch-Haroutounian (2009) utilized a lag model and a pseudo-error correction mechanism to evaluate the relationship between financial assets. In a lead-lag model, one return is regressed against leads and lags of another asset's return. Based on error correction models, the financials assets are co-integrated and arbitrage does not allow the price of securities to diverge.

To aggregate data, Zebedee and Kasch-Haroutounian (2009) used data at each transaction rather than fixed intervals and the proposed model tested a sample of US passenger airline securities. Since the airline industry is almost a homogeneous group of securities, it is expected that new market information will have similar effects on security pricing.

The proposed model was tested on trade and quote data of Trades and Quotes (TAQ) database from January, February and March 2000 that covered 63 trading days.

Based on empirical research by Zebedee and Kasch-Haroutounian (2009), at the microstructure level, correlation among the paired securities is weak and increased asymmetric information affects the speed of adjustment in intra-day returns. These findings are consistent with empirical research done by Chan (1993).

Motivation and Background:

The early signs of global financial crisis of 2008-2009 showed up as liquidity crisis in July 2007 when investors loss their confidence in the value of securitized mortgages in the United States. The indicator of perceived risk in general economy, TED spread, spiked up in July 2007 and remained volatile until it spiked to 4.65% on October 10, 2008. In September 2008, stock markets worldwide crashed and many insurance companies and financial institutes collapsed. Given the high volatility in 2008-2009, I would like to implement Zebedee and Kasch-Haroutounian (2009) model to evaluate the co-movement of six largest banks in US in the period of March 2009 to June 2009 (Table 1). These events create signals for market makers to adjust the price of securities and by implementing Zebedee and Kasch-Haroutounian (2009) model I can verify the co-movements of financial assets in intra-day level in a volatile period.

Methodology:

There are two approaches in this project for modeling the relationship: A Lead-lag model and an Error correction model. Lead-lag models have been used to model the flow of information between assets for many years. In a lead-lag model, return of asset A is regressed against leads and lags of the other asset's return and the estimated coefficients indicate the direction and timing of information flow. For example, futures and their underlying indices and options and their underlying securities could be used as paired assets in this model. Over all, previous studies by Ng (1987), Kawaller and Koch (1987), Harris (1989), Stoll and Whaley (1990), Chan (1992), Huang and Stoll (1994), and Tse (1999) show that futures returns lead stock returns. Result of empirical research on relationship between options and stocks are more controversial. Researchers have not reach an agreement about the direction of information flow (Zebedee and Kasch-Haroutounian, 2009). For example, research by Manaster and Rendleman (1982) shows that options return lead the asset returns while the result of empirical research done by Chan (1993) shows that stock returns lead option returns. In this paper, a lead-lag model has been deployed to assess the direction of information flow for one asset and its effect on pricing of second asset in fixed intervals of 5 minutes.

Error correction model is based on co-integration of paired assets in response to arbitrage theory in financial market. The concept of arbitrage theory has been in finance world as early as 17th century based on Geoffrey Poitras (2009) paper about history of arbitrage. In an arbitrage trading, an arbitrager takes advantage of existence of different prices for one good and generates positive profit with no risk and no net investment. Koutsougeras and Papadopoulos (2004) define arbitrage as "simultaneous purchase and sale of the same, or essentially similar, security in two different markets for advantageously different prices" . This approach is less clear for paired stocks since firm specific risk exists for different stocks. For this reason, Zebedee and Kasch-Haroutounian (2009) suggested a modified pseudo-error correction mechanism that recognizes intra-day relationship but allows day by day deviation.

To develop a lead-lag model, I have used Zebedee and Kasch-Haroutounian (2009) methodology to regress asset A's quote revision on lags of quote revision and trade history of asset A as well as asset B. The proposed autoregressive system has two parts; the first part evaluates the effect of quote revision process and the second part tests the outcome of trade history on pricing process for asset A.

$$q_t^A = \alpha + \sum_{i=1}^{p} \beta_i^A q_{t-i}^A + \sum_{i=0}^{q} \chi_i^A \chi_{t-i}^A + \sum_{i=0}^{r} \beta_i^B q_{t-i}^B + \sum_{i=0}^{s} \chi_i^B \chi_{t-i}^B + \varepsilon_t^q \quad (1)$$

q_t^A, q_t^B are the quote revisions for asset A and B, respectively at time t and it is calculated as : $q_t = \ln((a_t + b_t)/2) - \ln((a_{t-1} + b_{t-1})/2)$

where a_t and b_t are the quoted ask and bid prices for observation t. χ_i^A and χ_i^B are signed trade indicator variables at time t. This variable takes the value of -1 if the trade is initiated by a seller and +1 if it is initiated by a buyer. Zebedee and Kasch-Haroutounian (2009) used Lee and Ready (1991) mid-quote classification rule which states that a trade is initiated by a buyer if the transaction price is higher than mid-quote price and is considered initiated by a seller if the transaction price is lower than mid-quote price. The trade indicator variable is categorized as zero if the trading price is equal mid-quote price. Unlike Zebedee and Kasch-Haroutounian (2009), we have used buy and sell data reported by Bloomberg terminal. β_i^B and χ_i^B are estimated coefficients and ε_t^q is a normally distributed error. This method addresses how a market maker watches quoted price of other assets and adjusts quoted price for asset A. Since market makers adjust the price based on price of other assets, quote revisions will have positive cross-correlation with lag. The same concept suggests positive sign for lagged trades of the second asset. If there is asymmetric information in the market and asset B traders react to this information faster than asset A traders, asset A traders and market makers could use trade history of asset B. Therefore, there would be positive correlation between quote price of asset A and lagged trade history of asset B; certainly, these lagged parameters depend on the number and time-interval of the lags.

We have used Zebedee and Kasch-Haroutounian (2009) modified pseudo-error correction model. To modify pseudo-error correction model, Z&K added the difference of the intra-day returns for two assets which is $daily_{t-i}^{A} - daily_{t-i}^{B}$ in equation 2. More specifically , $daily_{t-i}^{A} - daily_{t-i}^{B}$ is $\ln((a_t + b_t)/2) - \ln((a_{open} + b_{open})/2)$. a_{open} and b_{open} are the opening ask and bid quoted price of asset A and B for the day that observation t is occurred. This correction assumes that market makers verify quote revisions on daily basis and consequently returns deviated by asymmetric information converge to an equilibrium at the end of the day which results a zero difference between intra-day returns of asset A and B.

The term γ in equation 2 indicates the speed of adjustment between two assets in a trading day. In the cross term $\gamma(daily_{t-i}^{A} - daily_{t-i}^{B})$, γ is expected to have a negative sign. When intra-day return of asset A is higher than that of asset B, a negative γ reduces the quote revision and returns it to an equilibrium; it will function in an opposite direction when Intra-day return of asset A is lower. In the following equations (2 & 3) , $spread_{t-1}^{A}$ is the demeaned quoted spread of asset A for observation t-1; $depth_{t-1}^{A}$ is the demeaned average quote size of asset A for observation t-1; $activity_t^A$ and $activity_t^B$ are the number of trades executed for asset A (B) in a 5 minutes interval; $revisions_t^B$ is the number of intermittent quote revisions for asset B in a 5 minutes interval.

$$q_t^A = \alpha + \sum_{i=1}^{p} \beta_i^A q_{t-i}^A + \sum_{i=0}^{q} \chi_i^A \chi_{t-i}^A + \sum_{i=0}^{r} \beta_i^B q_{t-i}^B + \sum_{i=0}^{s} \chi_i^B \chi_{t-i}^B + \gamma(daily_{t-i}^A - daily_{t-i}^B) + \varepsilon_t^q \quad (2)$$

$$\gamma = \gamma_0 + \gamma_1 spread_{t-1}^A + \gamma_2 depth_{t-1}^A + \gamma_3 activity_t^A + \gamma_4 activity_t^B + \gamma_5 revisions_t^B \quad (3)$$

Variables spread and depth are indicators of asymmetric information in the market. When there is uncertainty for the real value of an asset, spread is wider and market makers try to adjust the price and eventually price movement of the asset occurs. This price movement even in a relatively small trade size, results a lower quoted depth. Variables activity and revisions are included to control non-synchronicity. The concept is that if the paired assets are not balanced in terms of liquidity, the return of less active asset may converge to the return of more active asset. To be consistent with Zebedee and Kasch-Haroutanian (2009) empirical research, in both parts of equation 3, five lags are included and the coefficients are estimated by ordinary least squares.

Data integration:

We have used Bloomberg terminal for data integration purpose and downloaded data for 6 largest banks in US in the period of 27.03.2009 to 22.06.2009 (Table 1). Unlike Zebedee and Kasch-Haroutanian (2009) that used data at every transaction level and implemented a combination of Finucaneu (1999) lead-lag analysis and fixed intervals for data integration, we have used Bloomberg intra-day tools with arbitrary fixed 5 minutes intervals, same time-interval used by Chan (1993). Integrating data with fixed intervals simplifies the analysis of a lead-lag model because data integrated with variable intervals needs more adjustments. Zebedee and Kasch-Haroutounian (2009) utilized a method of variable intervals that was first used by Finucane (1999). Finucane (1999) used variable intervals to analyze the lead-lag relationship between options and stocks. He first integrated data based on clock time and arranged it with respect to event time. Then he omitted the prices that were registered at the same time. He considered option returns as the dependant variable and stock returns as the independent variable. He considered the first stock return recorded prior to the option return as the first lag of stock return. He used the same routine to develop other lagged independent variables. By using this method for data integration, the analysis is more precise since data for all the dependant variables are collected sequentially and no single event is missed and the timing of the dependant variable is significant in the analysis.

Zebedee and Kasch Haroutounian (2009) used slightly different method in data integration. They used same procedure for recording the dependant variables but for the lagged independent variables, fixed intervals were used. More specifically, they defined fixed intervals for independent variables and matched the dependent variable with the last independent variable in the previous fixed interval. Figure 1 shows how fixed and variable intervals are set for dependant and independent variables. For example, at t2, lagged return of second asset which is the independent variable is ln(midB2/midB1) while in t3, return of asset 2 is zero since in the previous fixed interval (t2-t3) there was not any quote change for second asset. If we intend to use variable intervals for second asset as well, in the later case, the return of the second asset would be ln(midB2/midB1).

In fact, although Zebedee and Kasch Haroutounian (2009) used fixed intervals for second asset but since these fixed intervals are matched to irregular intervals for the first asset, they still have retained the variable nature in data integration process.

Although choosing flexible intervals and integrating data at each transaction has many advantages, we could not follow same method because of our limited access to data of all transactions. Choosing an interval with lower time period such as 1 minute also creates large data that it is hard to process with ordinary computers. Therefore, we decided to implement the model on the data integrated at fixed 5 minutes intervals that had already been implemented by Chan (1993), and like Zebedee and Kasch-Haroutanian (2009) we chose a window of 63 trading days. There are some problems however with this method. As it can be seen in figure 2, we have used the last price in each 5 minutes interval to calculate the return. Based on timing of each event, we will have different scenarios that in some of them the integrated data leads incorrect computation.

Interval 2: In this scenario all data is collected correctly and the computed returns are correct.

Computed Return for Asset A:	Correct Return of A:	Computed lagged Return for Asset B:	Correct lagged Return of B:
Ln(MIDA2/MIDA1)	Ln(MIDA2/MIDA1)	Ln(MIDB1/MIDB0)	Ln(MIDB1/MIDB0)

Interval 3: In this scenario, return of asset A is wrong and we miss A3 for computing return of asset A. Also for asset B we miss the most recent quote for asset B, B3 that has occurred before A4.

Computed Return for Asset A:	Correct Return of A:	Computed lagged Return for Asset B:	Correct lagged Return of B:
Ln(MIDA4/MIDA2)	Ln(MIDA4/MIDA3)	Ln(MIDB2/MIDB1)	Ln(MIDB3/MIDB2)

Interval 4:

In this scenario, return of asset A is correct while return of asset B is wrong. Instead of B6 and B5 that are most recent quotes for asset B and happened prior to asset A's quote, we have used older quotes, B4 and B2.

Computed Return for Asset A:	Correct Return of A:	Computed lagged Return for Asset B:	Correct lagged Return of B:
Ln(MIDA5/MIDA4)	Ln(MIDA5/MIDA4)	Ln(MIDB4/MIDB2)	Ln(MIDB6/MIDB5)

Interval 6:

In this scenario, asset A has not been quoted or its quote has not been changed. So, based on our computation in fixed intervals, return of asset A is zero since Bloomberg considers same price when price is unchanged; In variable intervals that are defined based on event time, return of asset A will be same as interval 4 which is ln(MID A5/Mid A4) not zero. The correct lagged return for asset B is ln(MIDB6/B5) while we are using B7 and B6 that happened after A5 and they are not correct lagged quotes.

Computed Return for Asset A:	Correct Return of A:	Computed lagged Return for Asset B:	Correct lagged Return of B:
Ln(MIDA5/MIDA5)	Ln(MIDA5/MIDA4)	Ln(MIDB7/MIDB6)	Ln(MIDB6/MIDB5)

Interval 7:

While in this scenario both returns are correct , the inference might be wrong since it is possible that B8 has occurred just few seconds before A6 but we interpret it as 5 minutes lagged quote.

Computed Return for Asset A:	Correct Return of A:	Computed lagged Return for Asset B:	Correct lagged Return of B:
Ln(MIDA6/MIDA5)	Ln(MIDA6/MIDA5)	Ln(MIDB8/MIDB7)	Ln(MIDB8/MIDB7)

As it is shown in figure two and different scenarios discussed, incorrect inferences might happen when the length of the intervals are long compare to duration between two events. This incorrect inferences are also likely when too many events happen in each interval that we miss them since we extract older data from previous interval. In addition, as it was shown in interval 7, in some cases that events are so close in time sequence, we misinterpret the time interval. In this case, there is a possibility for two kinds of errors. First, two events are very close in time space and in reality asset A trader did not monitor asset B quotes and decided independently, but we consider two actions to be related. Second, two events are very close in time space and in reality asset A trader monitored asset B and made decision based on asset B's quotes and trade history, but we consider the lag to be 5 minutes while it is much shorter. Finally, the other problem with fixed intervals is missing data like interval 3 that we discussed before and missing observations might lead to incorrect inferences and analysis.

Descriptive Statistics:

Table 2 presents the correlation coefficients between paired assets in daily, hourly, 30 minutes and 5 minutes bases. All coefficients are significant at 5% level of confidence. In daily returns, correlation coefficients are ranged from 0.58 between STI and WFC and 0.91 between JPM and HBC as well as JPM and PNC. Moving from daily bases to hourly, it could be seen that correlation decreases between paired assets. This result is aligned with Zebedee and Kasch-Haroutanian (2009) result. In 30 minutes intervals, except two cases, correlation coefficients are either decreased or have not changed. In 5 minutes intervals, compare to 30 minutes interval, in 4 out of 15 cases correlation has slightly increased and in 11 other cases there is either no change or coefficients are decreased. The range of coefficients in 5 minutes interval is between 0.46 to 0.77 while based on Zebedee and Kasch-Haroutanian (2009) empirical research on samples of airline industry in the period of January to March of 2000, the maximum of estimated correlation coefficient is 0.18. Thus, overall the level of correlation in different intervals in this project compare to Zebedee and Kasch-Haroutanian (2009) result has increased while the general concept of decrease of correlation in intra-day bases is supported. The higher degree of correlation in short intervals such as 5 minutes might be the result of advanced technologies used in quoting process in the market or presence of more informed traders that react faster in data adjustment process compare to last decade, however, this needs more research and tests to be verified.

Table 3 presents the summary statistics for a 5 minutes interval on 27.03.2009. Quote revision data is presented on the left side of the table, column 2 to 5 and trades data is presented on the right hand side , column 6 to 9. Column one presents the number of quote revisions; quote revisions and average quote revisions are presented on column 2 and 3. Quote revision is the change in the quoted ask or bid price. As it is shown in Table 3 column 3, the large number of quote revisions indicate that market makers actively watch and adjust ask and bid price. The range of spread is 0.02 to 0.19 and the range of quote revision is 0.17% to 0.92%. The maximum quote revision belongs to STI and spread of this asset is 0.15 and the minimum quote revision belongs to WFC with 0.19 spread, however, to find the relation between quote revisions and spread we need to compare the quote revisions with lagged spread.

Trades are reported in right hand side of the table 3. Total volume is the aggregate volume, negative for sell, positive for buy and zero for otherwise and as a result total number of buy and sales volume does not match with total volume of trades. Since we do not have the details of all transactions for each 5 minutes interval, we could not report the standard error of data.

Asset A :

The first series of tests evaluates the effect of trade and quote history of asset A on pricing process of itself. There are two parts in this test. First, we examine the quote revision process and more specifically regress asset A quote on lags of quote revisions.

Second, we test the effect of trade history on quote process. For this reason we implement equation 4 on the sample data.

$$q_t^A = \alpha + \sum_{i=1}^{p} \beta_i^A q_{t-i}^A + \sum_{i=0}^{q} \chi_i^A \chi_{t-i}^A \qquad (4)$$

The result is different from one asset to other and it is not consistent with Zebedee and Kasch-Haroutanian (2009) result. Based on their empirical research, estimated coefficients are highly significant in autoregressive model while in this research, assets have significant coefficients ,mostly on first lag, and in few of them on other lags. Panel A and B of table 4 present the estimated coefficients for lags of quote revisions and trade history of asset A respectively. For example, as it is presented in table 4 Panel A, the estimated coefficients for Citibank with symbol C are significant at 5% level in first and fourth lag while JP Morgan with symbol JPM only has significant coefficient for quoted revision at second lag with 10% level of significance. Result for Wells Fargo with symbol WFC is different. Coefficients for all the lags are significant but only third lag is significant at 1% level of significance. On the other hand for PNC, the coefficients are all insignificant except fifth lag. Zebedee and Kasch Haroutounian's (2009) paper does not have detailed report but they have stated that the result is highly autoregressive. The different result of this project might be the outcome of choosing fixed intervals for data integration. By missing many quotes in a fixed interval and using older quotes instead of recent quotes in computation, we are trying to match unrelated quotes. For this reason and to verify the code used in this project, I have executed the code on the same data that Z&K used. I integrated data for two securities, United Airline and US Airways Group from Trade and Quotes (TAQ) database for January, February and March (TAQ industry code: 202A) and I used the same methodology of Z&K (2009) for data integration as well as Lee and Ready (1991) method for data sequencing. Like Z&K (2009) the result shows the autoregressive nature of data and coefficients are strongly significant for all five lags.

Asset B quote and trade history effect:

The first category of tests evaluates the effect of trade and quote history of second asset on the price process of first asset. Two series of F tests have been done to examine the significance of feedback of the second asset's price. In the first series of tests, the null hypothesis is that lagged quote revisions of asset B is ineffective in pricing process of asset A or the estimated coefficients $\beta_0^B, \beta_1^B, ..., \beta_5^B$ are jointly zero. The result of this test is presented in Panel A of Table 5. For instance, the value in first row second column, 189.03 is the F statistics for the F test assessing the significance of C quote revisions on JPM's quote revisions process. High value of F tests is the result of large difference between R^2 value of restricted and unrestricted equation. In all 30 cases, at 1% significance level, the F statistics reject the null hypothesis that estimated coefficients, $\beta_0^B, \beta_1^B, ..., \beta_5^B$ are jointly zero. Therefore, it can be concluded that lagged quote revisions of asset B provide valuable predictability for asset A returns.

To evaluate the effect of lagged quote revisions of asset B in quote revision process, Z&K (2009) performed Wald test and based on their test, the null hypothesis were rejected for 30 of 72 cases at 10% significance level. Z&K (2009) concluded that lagged quote revisions in asset B provide some predictability in asset A returns.

To clarify the difference between the result of Z&K (2009) research and my project, I implement the model for equation (1) and I used the method that Z&K (2009) used for data integration which is using all transaction data rather than fixed 5 minutes interval. The result of F-tests for UAL-U paired assets is close to those in Z&K (2009) paper (Table 4 part A in Z&K). For example, the chi-squared test statistics result for UAL-U paired asset in Table 4 Panel A of Z&K (2009) is 30.46 while the F-test result generated by my code is 13.91.

In the second series of tests, the F test examines the significance of asset B's trade history on the quote revision process of asset A. In this case, the null hypothesis is that estimated coefficients $\chi_0^B, \chi_1^B, \chi_2^B, ..., \chi_5^B$ are jointly zero. Overall, in 17 out of 30 cases, the F statistics rejects the null hypothesis. Thus lagged trade histories of asset B offer some information in predicting price for asset A. Again, to verify the code, I executed the code for UAL-U paired asset while I used transaction data. The results are close; the chi-squared test statistics for UAL-U paired asset in Z&K (2009) report (Table 4 Panel B) is 11.96 while the F test statistics computed by my code is 17.84.

Pseudo-error correction mechanism

Second category of tests evaluates the effect of number of trade and quote characteristics , γ ,on the quote revision process. γ is defined as :

$$\gamma = \gamma_0 + \gamma_1 spread_{t-1}^A + \gamma_2 depth_{t-1}^A + \gamma_3 activity_t^A + \gamma_4 activity_t^B + \gamma_5 revisions_t^B$$

The F test examines significance of Pseudo-error correction mechanism for quote revision process of asset A. For this purpose, the restricted model tests the joint significance of the estimated coefficients $\gamma_0, \gamma_1, \gamma_2, \gamma_3, \gamma_4$ and γ_5. The results are presented in table 6. Overall, at 1% significance level, 27 out of 30 cases reject the null hypothesis. Thus we can conclude that additional parameters specified as spread and depth provide significant predictability.

To evaluate the individual coefficients, two series of t statistic tests have been done on the estimated coefficients for spread and depth. In Panel A of table 7, the estimated coefficients for spread are presented. In 14 out of 30 cases, the result is significant at 1% significance level and 2 out of 30 cases are significant at 5% level. Among the significant coefficients, 14 out of 16 cases are negative. The negative sign of spread is consistent with the concept of paired asset return convergence, since the holder of asset A, quotes higher spread to adjust the price and lowers his or her loss.

The large spread implies greater asymmetric information which encourages the market maker to follow and monitor other assets. Therefore, the large spread results in increasing the speed of return convergence of paired assets.

In Panel B, the estimated coefficients for depth are presented. In 16 out of 30 cases, the estimated coefficients are significant. From 16 cases, 11 cases are positive which is 36% of the total coefficients. The positive sign is consistent with paired asset return co-movement, since price movement of asset A is relatively large compare to the size of the trade , it lowers depth to compensate the loss for holder of asset A. For example, in the scenario that asset A has lower return and asset A holder has loss, the price of asset A should be adjusted upward to compensate the loss. On the right side of equation, depth of asset A diminishes and "daily return A- daily return B" is negative thus quoted price of asset A on the other side of the equation will be adjusted upward. Therefore, the estimated coefficient for depth should be positive to result a higher value for asset A's quote with the new diminished depth. The positive sign of estimated coefficient of depth indicates higher co-movement in presence of asymmetric information.

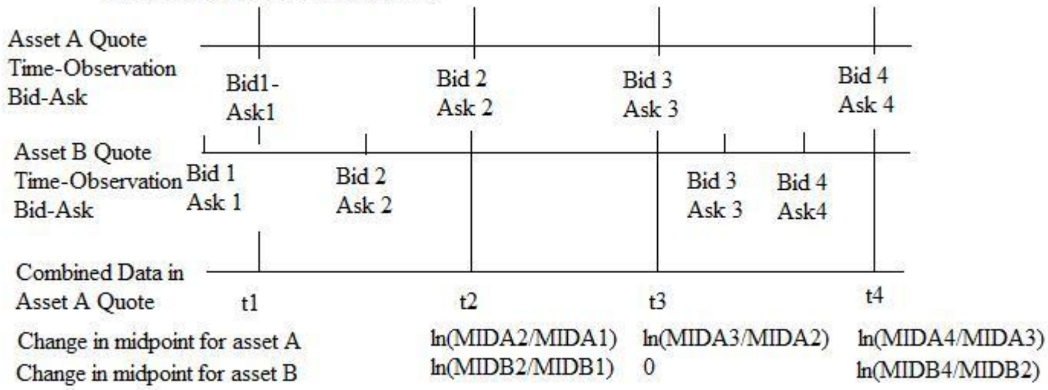
Conclusion:

This paper evaluates the co-movement of asset returns on intra-day bases. The implemented model is regression of quoted revision of one asset on lagged returns and trades of second asset and also pseudo-error correction model that was used by Zebedee and Kasch-Haroutounian (2009).

We used intra-day data with 5 minutes interval of six large financial institutes in U.S. for the period of 27.03.2009 to 22.06.2009. Based on the results of this paper, both lagged quote revisions and trade history of asset B, provide valuable predictability for asset A's returns. In addition, the suggested modified pseudo error correction mechanism of Zebedee and Kasch-Haroutounian (2009) highlights the predictive power of second asset by presenting new variables, spread and depth. Higher levels of spread is an indicator of asymmetric information in the market and the speed of adjustment increases with higher spread and depth; however this result is consistent in only half of the assets. Using 5-minutes fixed interval technique instead of integrating data at transaction level that was used by Zebedee and Kasch-Haroutounian (2009), reduces the power of autoregressive model in particular for asset A, however the final result which is the predictability power of both lagged quote revisions and trade history of asset B as well as the sign of depth and spread variables in pseudo error correction mechanism is consistent with Zebedee and Kasch-Haroutounian (2009).

There are research opportunities such as implementing the model with shorter than 5 minutes interval or with variable intervals at transactions and implementing the pseudo error correction model with different variables that could improve the analysis of trade dynamics and co-integration.

Figure 1 : Example of data integration process with variable intervals used by Zebedee & Haroutounain (2009)



Source: Zebedee and Haroutounian (2009),Journal of Economics and Business 61, Fig 1, page 285

Figure 2
Each interval is 5 minutes. A0 to A5 define trades or quotes for asset A and B0 to B7 define trades or quotes for asset B.

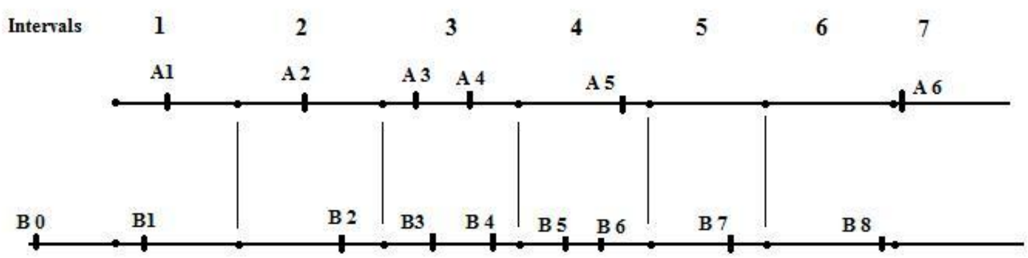


Table 1

Sample identification.

Item	Company Name	Symbol
1	Citigroup	C
2	J. P. Morgan Chase & Company	JPM
3	Wells Fargo & Company	WFC
4	HSBC North America Inc.	HBC
5	SunTrust Banks, Inc.	STI
6	PNC Financial Services Group	PNC

Table 2

This table presents the correlation coefficients between paired assets in daily, hourly, 30 minutes and 5 minutes bases. Returns are computed as $\ln(close_t) - \ln(close_{t-1})$ where $close_t$ is the closing prices for day t for daily returns, and closing price for one hour, 30 minutes and 5 minutes interval for Hourly, 30 minutes and 5 minutes returns respectively.

A) Daily returns						**B) 60 minutes**				
	C	JPM	HBC	WFC	STI	C	JPM	HBC	WFC	STI
JPM	0.67					0.53				
HBC	0.65	0.91				0.50	0.82			
WFC	0.61	0.62	0.68			0.60	0.48	0.51		
STI	0.67	0.79	0.84	0.58		0.56	0.65	0.71	0.51	
PNC	0.63	0.91	0.91	0.67	0.81	0.48	0.72	0.73	0.55	0.66
C) 30 minutes						**D) 5 minutes**				
	C	JPM	HBC	WFC	STI	C	JPM	HBC	WFC	STI
JPM	0.53					0.55				
HBC	0.50	0.80				0.53	0.77			
WFC	0.62	0.51	0.54			0.57	0.52	0.53		
STI	0.54	0.65	0.70	0.54		0.48	0.67	0.70	0.47	
PNC	0.47	0.72	0.71	0.53	0.65	0.46	0.70	0.67	0.49	0.64

Table 3

Summary statistics on 27.03.2009.

Quote revision data is presented on the left side of the table, column 2 to 5 and trades data is presented on the right hand side , column 6 to 9. Column one presents the number of quote revisions; qt is quote revision and is computed as:

$$q_t = \ln((a_t + b_t)/2) - \ln((a_{t-1} + b_{t-1})/2)$$

Mid is average quote revisions; Spread is demeaned quoted spread ; depth is demeaned average quote size; n-trade is total number of trades ;n-quote is total number of quotes. Volume is total number of shares in thousands and Buy Vol and Sell Vol are total number of bought and sold shares respectively.

	Quotes					**Trades**			
	n-quote	**qt**	**Mid**	**Spread**	**Depth**	**n-trade**	**Volume('000)**	**Buy Vol.('000)**	**Sell Vol.('000)**
C	25,641	0.00382	2.61	0.02	36.81	19,536	1329	1582	1343
JPM	22,386	0.00659	28.73	0.18	78.44	17,016	5284	5526	6475
HBC	9,110	-0.00305	16.39	0.14	57.53	6,187	6442	4663	4859
WFC	10,340	-0.00173	29.00	0.19	81.18	9,020	257	634	576
STI	30,228	0.00926	12.89	0.15	35.42	16,214	664	780	744
PNC	31,801	0.00651	32.17	0.17	94.19	17,381	604	1939	971

Table 4

Summary statistics for estimated coefficients of asset A for equation: $q_t^A = \alpha + \sum_{i=1}^{p} \beta_i^A q_{t-i}^A + \sum_{i=0}^{q} \chi_i^A \chi_{t-i}^A$

q_t^A, q_t^B are the quote revisions in asset A and B, respectively at time t. χ_{t-i}^A, χ_{t-i}^B are cumulative signed trades for asset A and B (+1 for a buy;-1 for a sell; and, zero otherwise) between t and t-1 which is a 5 minutes interval; * , ** and *** indicate significance at 10%, 5% and 1% level, respectively.

Panel A	q_{t-1}^A	q_{t-2}^A	q_{t-3}^A	q_{t-4}^A	q_{t-5}^A
C	-0.039236**	0.009001	-0.032301*	0.048817**	0.032934*
JPM	-0.030442	-0.03259*	-0.009086	0.005371	0.018767
WFC	-0.034955*	-0.049187**	-0.05668***	0.039007**	0.040973**
HBC	-0.057403***	-0.048939**	-0.016322	-0.005718	0.030688
STI	-0.017699	0.007247	-0.008269	0.067522***	0.002247
PNC	0.00089	0.012237	0.027113	0.019371	0.050373***

Panel B	χ_t^A	χ_{t-1}^A	χ_{t-2}^A	χ_{t-3}^A	χ_{t-4}^A	χ_{t-5}^A
C	1e-006***	0	0	0	0***	0
JPM	0***	0	0	0	0	0
WFC	0**	0	0	0	0	0
HBC	2.00E-06	-1.00E-06	-1.00E-06	1.00E-06	-1.00E-06	0
STI	1e-006**	-1e-006***	1e-006***	0	0	0
PNC	-1e-006***	0	3e-006***	0	-1e-006**	0

Table 5

This table presents the F test statistics on the parameters of the model:

$$q_t^A = \alpha + \sum_{i=1}^{5} \beta_i^A q_{t-i}^A + \sum_{i=0}^{5} \chi_i^A x_{t-i}^A + \sum_{i=0}^{5} \beta_i^B q_{t-i}^B + \sum_{i=0}^{5} \chi_i^B \chi_{t-i}^B + (\gamma_0 + \gamma_1 spread_{t-1}^A + \gamma_2 depth_{t-1}^A$$
$$+\gamma_3 activity_t^A + \gamma_4 activity_t^B + \gamma_5 revisions_t^B)(daily_{t-1}^A - daily_{t-1}^B) + \varepsilon_t^q$$

$$q_t = \ln((a_t + b_t)/2) - \ln((a_{t-1} + b_{t-1})/2)$$

q_t^A, q_t^B are the quote revisions in asset A and B, respectively at time t. χ_{t-i}^A, χ_{t-i}^B are cumulative signed trades for asset A and B(+1 for a buy;-1 for a sell; and, zero otherwise) between t and t-1 which is a 5 minutes interval; and $daily_{t-1}^A$, $daily_{t-1}^B$ are the intra-day returns of asset A and B since opening. Asset A is listed on the left-hand side and asset B is listed across the top of the table. There are two sets of restrictions: Panel A: $\beta_0^B = \beta_1^B = \beta_2^B = \beta_3^B = \beta_4^B = \beta_5^B = 0$ and Panel B : $\chi_0^B = \chi_1^B = \chi_2^B = \chi_3^B = \chi_4^B = \chi_5^B = 0$;* , ** and *** indicate significance at 10%, 5% and 1% level, respectively.

Panel A, quote revisions, F test on $\beta_0^B = \beta_1^B = \beta_2^B = \beta_3^B = \beta_4^B = \beta_5^B = 0$

Panel B, trades, F test on $\chi_0^B = \chi_1^B = \chi_2^B = \chi_3^B = \chi_4^B = \chi_5^B = 0$

Panel A	**C**	**JPM**	**WFC**	**HBC**	**STI**	**PNC**
C		189.01***	178.29***	211.71***	135.27***	120.57***
JPM	189.03***		588.41***	156.1***	331.83***	373.25***
WFC	186.56***	584.34***		166.23***	380.47***	322.14***
HBC	200.06***	169.81***	165.9***		115.43***	147.75***
STI	136.39***	345.62***	379.67***	111.47***		289.42***
PNC	119.58***	376.64***	324.08***	141.7***	286.01***	

Panel B	**C**	**JPM**	**WFC**	**HBC**	**STI**	**PNC**
C		2.88***	1.34	1.02	1.04	2.31**
JPM	5.92***		2.02*	3.62***	1.22	0.99
WFC	2*	1.95*		2.75**	1.21	2.76**
HBC	3.27***	0.73	0.61		0.41	2.43**
STI	5.2***	5.11***	1.76	0.57		5.99***
PNC	2.09*	5.91***	8.6***	1.17	0.24	

Table 6

Significance of pseudo-error correction model for asset A's price revision process. This table presents F statistics with restrictions on γ($\gamma_0 = \gamma_1 = ... = \gamma_5 = 0$)in the model:

$$q_t^A = \alpha + \sum_{i=1}^{5} \beta_i^A q_{t-i}^A + \sum_{i=0}^{5} \chi_i^A x_{t-i}^A + \sum_{i=0}^{5} \beta_i^B q_{t-i}^B + \sum_{i=0}^{5} \chi_i^B \chi_{t-i}^B + (\gamma_0 + \gamma_1 spread_{t-1}^A + \gamma_2 depth_{t-1}^A$$
$$+\gamma_3 activity_t^A + \gamma_4 activity_t^B + \gamma_5 revisions_t^B)(daily_{t-1}^A - daily_{t-1}^B) + \varepsilon_t^q$$

q_t^A , q_t^B are the quote revisions in asset A and B, respectively at time t. χ_{t-i}^A , χ_{t-i}^B are cumulative signed trades for asset A and B(+1 for a buy;-1 for a sell; and, zero otherwise) between t and t-1 which is a 5 minutes interval; and $daily_{t-1}^A$, $daily_{t-1}^B$ are the intra-day returns of asset A and B since opening. $spread_{t-1}^A$ is the demeaned quoted spread in asset A for observation t-1; $depth_{t-1}^A$ is the demeaned average quote size for asset A for observation t-1; $activity_t^A$ and $activity_t^B$ are the number of trades executed for asset A (B) in a 5 minutes interval; $revisions_t^B$ is the number of intermittent quote revisions in asset B in a 5 minutes interval. Asset A is listed on the left-hand side and asset B is listed across the top of the table. ;* , ** and *** indicate significance at 10%, 5% and 1% level, respectively.

	C	JPM	WFC	HBC	STI	PNC
C		30.46***	29.08***	17.59***	7.62***	31.01***
JPM	7.73***		5.96***	9.72***	13.3***	2.74**
WFC	19.25***	5.56***		12.61***	19.56***	1.35
HBC	7.35***	78.27***	25.07***		17.06***	19.98***
STI	9.29***	30.57***	16.21***	4.04***		12.1***
PNC	1.33	5.14***	0.68	2.84***	4.94***	

Table 7

This table presents the estimated coefficients of spread and depth .

$$q_t^A = \alpha + \sum_{i=1}^{5} \beta_i^A q_{t-i}^A + \sum_{i=0}^{5} \chi_i^A x_{t-i}^A + \sum_{i=0}^{5} \beta_i^B q_{t-i}^B + \sum_{i=0}^{5} \chi_i^B \chi_{t-i}^B + (\gamma_0 + \gamma_1 spread_{t-1}^A + \gamma_2 depth_{t-1}^A$$
$$+\gamma_3 activity_t^A + \gamma_4 activity_t^B + \gamma_5 revisions_t^B)(daily_{t-1}^A - daily_{t-1}^B) + \varepsilon_t^q$$

q_t^A , q_t^B are the quote revisions in asset A and B, respectively at time t. χ_{t-i}^A , χ_{t-i}^B are cumulative signed trades for asset A and B(+1 for a buy;-1 for a sell; and, zero otherwise) between t and t-1 which is a 5 minutes interval; and $daily_{t-1}^A$, $daily_{t-1}^B$ are the intra-day returns of asset A and B since opening. $spread_{t-1}^A$ is the demeaned quoted spread in asset A for observation t-1; $depth_{t-1}^A$ is the demeaned average quote size for asset A for observation t-1; $activity_t^A$ and $activity_t^B$ are the number of trades executed for asset A (B) in a 5 minutes interval; $revisions_t^B$ is the number of intermittent quote revisions in asset B in a 5 minutes interval. Asset A is listed on the left-hand side and asset B is listed across the top of the table. ;* , ** and *** indicate significance at 10%, 5% and 1% level, respectively.

$spread_{t-1}^A$	C	JPM	WFC	HBC	STI	PNC
C		-0.17362***	-0.17598***	-0.085303***	-0.13984***	-0.49377***
JPM	-0.066951***		0.040627**	0.078108**	0.01734	0.01448
WFC	-0.074014***	0.03744		0.0024663	-0.031142	-0.016441
HBC	-0.038798***	-0.27677***	-0.069614***		-0.06657***	-0.082037***
STI	-0.10755***	-0.015752	-0.014216	-0.059145***		-0.0060924
PNC	-0.002877	0.00020852	0.00022535	-0.00049925	0.00068534	

$depth_{t-1}^A$	C	JPM	WFC	HBC	STI	PNC
C		0.00016689***	0.00017477***	6.87E-05	4.61E-05	8.727e-005*
JPM	-1.60E-05		1.36E-05	0.00018276***	1.40E-05	-1.91E-05
WFC	9.76e-006***	-5.50E-06		-2.178e-005***	8.56e-006***	1.039e-005**
HBC	-9.151e-005***	-0.00023528***	-4.431e-005*		-1.44E-05	-4.728e-005**
STI	1.32E-06	4.8e-006***	3.55e-006***	5.9e-007*		3.84e-006***
PNC	2.93E-05	4.32E-05	-9.37E-06	3.77E-05	4.18E-06	

References:

Alfonso Dufour and Robert F. Engle, Dec 2000, Time and the Price Impact of a Trade, The Journal of Finance, VOL. LV, NO. 6

Allan A.Zebedee, Maria Kasch-Haroutounian, November 2008, A closer look at co-movements among stock returns, Journal of Economics and Business, VOL. 61, Issue 4, pages 279-294

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, 5, 123-152

Chan, K. (1993). Imperfect information and cross-autocorrelation among stock prices. The Journal of Finance, 1211–1230.

G.Andrew Karolyi and Rene M.Stulz, 1996, Why Do Markets Move Together? An Investigation of U.S.-Japan Stock Return Co-movements, Journal of finance, Vol. LI, No.3

Geoffrey Poitras, (2009), "Arbitrage: Historical Perspective" in R. Cont. (ed.), The Encyclopaedia of Quantitative Finance, New York: Wiley (2009 forthcoming), Page 6

Harris, L, (1989), the October 1987 S&P 500 stock futures basis, The Journal of Finance, 44, 77-99

Huang, R.D., & Stoll, H.R., (1994), Market microstructure and stock returns. The review of Financial Studies, 7, 179-213

Kawaller, I. G., Koch, P.D., &Koch, T. W., (1987). The temporal relationship between S&P 500 futures and S&P 500 index, The journal of Finance , 42, 1309-1329.

Laura Veldkamp, July 2006, Information Markets and the Comovements of Asset Prices, Review of economic Studies, Vol. 73, No. 3, pp. 823-845

Lee, C. M. C., & Ready, M. J. (1991). Inferring trade direction from intraday data. The Journal of Finance, 46, 722–735.

Leonidas C. Koutsougeras & Konstantantinos G. Papadopoulos, (April 2004), Arbitrage and equilibrium in Strategic Security Markets, Economic Theory, Vol. 23, No. 3, PP. 553-5681

Manaster, S., Rendleman, R. J., Jr (1982), Option prices as pre
dictors of equilibrium stock prices, The Journal of Finance, 37, 1043-1057

Ng, N., 1987, Detecting spot price forecasts in futures prices using causality tests, Review of Futures Markets, 16, 250-267

Partha Deb, Pravin K. Trivedi and Panayotis Varangis, *Journal of Applied Econometrics*, Vol. 11, No. 3 (May - Jun., 1996), pp. 275-291

Robert J. Shiller, July 1989, Comovemnets in Stock Prices and Comovements in Dividends, The Journal of Finance, VOL. XLIV, NO. 3

Stoll, H.R., & Whaley R. E., (1990), The dynamics stock index and stock index futures returns. Journal of Fianance and Quantitative Analysis, 25, 441-468

Thomas W.EPPS, 1979, Co-movements in Stock Prices in the Very Short Run, Journal of the American Statistical Association, Vol. 74, No. 366, pp.291-298

Tse, Y., (1999), Price discovery and volatility spillovers in the DJIA index and futures markets. Journal of Futures Markets, 19, 911-930.