Nearby and deferred quotes: What they tell us about linkages and adjustments to information

Mindy Mallory, Phillip Garcia, and Teresa Serra

June 04, 2015

# Introduction

Can we make the citations and biliography work (Hasbrouck 2006; Easley, Lopez de Prado and O’Hara 2011)?

# Background

The literature on how information affects liquidity in securities markets is long and rich.[[1]](#footnote-23) Bagehot (1971) is regarded as the first to demonstrate that the BAS arises even if inventory and transactions costs are assumed to be zero when asymmetric information is present. Copeland and Galai (1983) built upon Bagehot's work by assuming that the proportion of informed traders in the market was . Knowing this the market maker adjusts his quoted bids and offers to maximize expected profit. Copeland and Galai's model, however, did not account for the fact that the trades themselves could reveal information about whether or not traders were informed. Glosten and Milgrom (1985) formalized this concept and developed a model where the market maker adjusts his beliefs based on the trades that occur. The market maker knows that at least some of the traders are informed so sell orders revise the market maker's belief about the value of the security down and buy orders revise his belief upward. They show that the spread is increasing in the proportion of informed traders, and there is a point at which too many informed traders require the market maker to set the spread so wide that trade does not occur and the market halts (an example of the famous "Market for Lemons" described by Akerlof (1970)).

Easley and O’Hara (1987) and Easley and O’Hara (1992) incorporate trade size and its effect to a model similar to Glosten and Milgrom. An informed trader chooses between small and large size trading blocks. The informed trader understands that a large size trading block essentially tips their hand to the market maker and the other uninformed traders, causing the market maker to revise bids and offers in accordance with the revealed information. This results in a trade off for the informed trader. Transact at 'good' price with one large lot, or transact with several small lots at 'good', but worsening, prices as the market maker learns slowly from the informed traders order flow.

Hasbrouck (2006) provides and excellent overview of how Easley, Hvidkjaer and O’Hara (2002) and Easley, Kiefer and O’Hara (1997) use the Easley and O'Hara models of informed trading to develop a measure of the probability of informed trading (PIN). This measure, though, is estimated solely based on the distribution of order arrivals, where a trade is labeled as a buy, or buyer initiated if the trade occurs above the midpoint of the quoted spread and a sell if the trade occurs below the midpoint of the quoted spread. Numerous studies have documented that there may be problems with downward bias in the estimated PIN (Yan and Zhang 2012; Vega 2006; Boehmer, Grammig and Theissen 2007) and estimating information based trading in this way ignores some aspects of futures markets that are not present in securities markets.

For example, in commodities futures markets several contracts with different maturities trade in the marketplace, each reacting in a correlated but non-synchronous way to information and liquidity based shocks. They react differently to information based shocks because there is a cost to store the physical commodity through time. Therefore, an information based shock that is most pertinent to the nearby contract may have large effect on the nearby futures contract, but a dampened effect in more distant maturing contracts. Information about shortages and surpluses in the near-term are mitigated in distant contracts by the assumption that high (low) prices today will ration (encourage) consumption and therefore mitigate price effects. Further, each contract maturity attracts different levels of liquidity. Generally liquidity decreases as more distant contracts are considered except in the case of commodities produced once per year - corn, for example, is harvested in the fall of each year and the 'new crop contract' with a December maturity will routinely attract more volume than the May or July contract even when the May or July contract expire sooner. For these reasons we do not adopt the PIN measure of information based trading in the marketplace

To the best of our knowledge there are no market microstructure models that explicitly take into account the features of commodity futures markets. The closest models come from work on index futures that cover a basket of securities. Most prominent is the work by Kumar and Seppi (1994) who assume *N* different non-dividend paying securities and an index futures contract on a buy-and-hold portfolio of a subset of these stocks. In Kumar and Seppi's model specialists in the cash market observe a signal, and floor traders of the futures index observe a signal about the value of the index but not the individual securities. A key feature they build into the model is a lag in the information transmittal between the cash and futures markets because specialists only observe order flows from their own market, and not the other. This lag in information transmittal allows for arbitrageurs, who possess faster telecommunication technologies, to learn from transactions in both markets and make profitable trades in the cash and futures markets. In common futures market parlance, these 'arbitrageurs' are analogous to spread traders who trade in both nearby and deferred contracts hoping to profit on relative price movements.

There are some important distinctions between the arbitrageurs as imagined in the Kumar and Seppi model and spread traders in a futures market. Namely, the basis between a composite of cash security prices and the price of a futures index of the same basket should behave in very predictable ways (the basis, in theory, should only vary with interest rates and expected changes in dividend yields). In contrast, the spread between the price of two commodity futures contracts with different maturities depends on many more uncertain variables related to domestic and international consumption, exchange rates, production or distribution bottlenecks, weather, among other things. The arbitrageur in Kumar and Seppi's model need only to wait to profit for others in the marketplace to learn the information they already know and placed a trade based upon. The futures market spread trader entertains much more risk in betting on relative price changes between to futures maturities.

However, Kumar and Seppi's model yields two empirical predictions that provide insight. Their Proposition 1 states that it may not be optimal for the arbitrageur to hold a perfectly offsetting position. Because there is a time lag over which the specialists learn about realized market signals in the other market, it may be optimal for the arbitrageur to hold a net long or net short position in the cash or index futures market. As noted above, the commodity futures market is even noisier and so this proposition may hold for commodity futures as well. Next, their proposition two states that the correlation between arbitrageurs' stock and futures orders is negative, meaning if they are long the in the cash basket of securities they are likely to be short the basket of index futures. If arbitrageurs only made perfectly offsetting orders in the cash and futures markets, this would trivially be a perfect negative correlation. What this behavior means for bids, offers, and relative transaction prices in the cash and futures market is unclear because sometimes the arbitrageurs' activities will cause the cash and futures prices to come together and sometimes it will cause them to diverge.

In the next section we draw insights from the sequential trading models described above to generate some empirical predictions about the correlations between the bids and offers of nearby and deferred maturity commodity futures contracts.

# Conceptual Framework

The role of liquidity trades verses information based trades.

# Data

The data used in this analysis comes from the CME Group's Top of Book (BBO) database for corn futures contracts from 1/4/2010-11/4/2011. The data contains the best bid, bid volume, best offer, offer volume, last trade price, and last trade volume of the order book for each active futures contract, time-stamped to the second. For those familiar with the CME Group's Time & Sales database, the BBO contains the Time & Sales data as well as information about the best bids and offers on the top of the limit order book.

The data are time-stamped to the second, but trades and updates to the top of the book routinely occur more rapidly than once per second. This results in several updates to the top of the book displaying the same time stamp. This requires us to either aggregate to the second, or to simulate sub-second time stamps (cite Wang and Easley). Since we calculate correlations between updates to the top-of-the book for several contract maturities, simulation would need to preserve (at least) the order in which updates arrived to each respective contract to be useful. Since preserving the order in which updates to the top of the book is impossible, we aggregate to the second instead.

Further, we exclude days on which there was a limit price move in any of the contracts, since when prices are locked at the limit, calculating correlations is impossible (Dates deleted due to limit price moves are as follows: 1/12/2010, revision to a Crop Production report; 3/31/2011, Prospective Plantings report; 6/30/2011, Planted Acres report; 10/8/2010, World Agricultural Supply and Demand Estimates (WASDE); and 12/9/2010, WASDE). Also, we exclude 4/5/2010, because there was an unusually high number of revisions to the best bid and best offer. Since we were not able to process all of the data for this day in a reasonable amount of computing time, we drop this day from our sample. Additionally, 7/5/2011 was an unusually light trading day after the Fourth of July holiday and resulted in no data for the third to mature contract.

Table 1 shows the first ten entries to our data after manipulating the raw BBO data set from CME Group to display the entire top of the book on one line with the appropriate time stamp. The first column is the time-stamp, the second column is the trade sequence number, which the CME Group gives to individual trades to identify separate orders that arrive on the same second. The third column, SYMBOL, identifies which futures maturity the observation represents. In this case, 1003 stands for March 2010, with the first two characters representing the year and the second two characters representing the month.

For each date in our sample, we define the first to mature (nearby), one, two, and three contracts deferred. To avoid undesirable delivery month effects, we roll these definitions at the end of the month prior to contract maturity (For example, on March 1st the May futures contract is considered the nearby as opposed to the March futures contract, and so forth).

# Analysis

Our analysis simply considers the correlation of revisions to nearby contract bids (offers) with one, two, and three deferred contract bids (offers), respectively. The operating assumption here is that when information arrives to the market, it usually will affect the entire forward curve in the same direction. In other words, information that raises the best bid (offer) in the nearby contract, should usually raise the best bid (offer) in the deferred contracts as well. Linkages between the nearby and deferred contracts can be measured with simple correlations without making the distributional assumptions required by a more intricate econometric model. Given the high frequency and irregularly spaced nature of the data, such distributional assumptions are often problematic (cite).

We have two primary objectives in this study: 1) calculate the strength of correlations between the order books of the nearby and deferred contracts, and 2) measure (or bound) the time it takes for information to be transmitted from nearby to deferred contracts. To measure the first, we calculate contemporaneous correlations between the log changes of the bid (offer) in the nearby and the deferred contracts. Then, to measure the second we calculate the correlation between time lagged log changes of the bid (offer) of the nearby with log changes of the bid of the deferred contracts. We lag the nearby by one second and ten seconds. The time lagged correlations provide a measure of how long it takes for information to be transmitted from nearby to the deferred contracts. The logic is that if we observe contemporaneous correlation between the nearby and deferred contracts, we can search for the time lag at which we observe the correlation disappearing. We conclude that information has been fully transmitted when the time lagged nearby and deferred contract order book revisions become uncorrelated.

Since the corn futures contract experiences non-uniform trading volume throughout the day, there may be time of day effects in the strength and rate at which information is transmitted through the futures market. To measure how the correlations change throughout the trading day, we divide the day into ten minute intervals and calculate correlations between the nearby and the deferred contract bids (offers) for each interval with one day's worth of data. This allows us to detect if there are any discernible patterns to the correlations over the trading day. Further, since we calculate one correlation per day we also recover a whole distribution of the correlations for each time interval.

## USDA Announcement Days

On USDA report announcement days there is often a significant amount of information that market participants receive at the same time, causing large price fluctuations and larger than usual trading volumes. Therefore, in our analysis we also separate out days on which major USDA reports are released and calculate the same correlations described above. During our sample period, the USDA reports we include were released at 8:30 am CST, before the day trading session began.

## Treatment of Zero Changes

We mentioned above that it is common to have multiple revisions to the order book happen in the same second (and consequently receive the same time-stamp in the data). The converse is also true, however. There are also many instances where a number of seconds will transpire before a revision to the top of the order book arrives. This results in our variables, log changes in best bid (offer), containing many zeros. How these zeros are distributed between the contracts is related to the concepts of liquidity-based activity verses information-based activity discussed in the conceptual framework above.

To fix ideas consider the case where we intend to calculate the contemporaneous correlation between the nearby and the deferred contracts. There are four possibilities for the presence of zeros. On any time stamp one of four situations may be operative: 1) neither the nearby nor the deferred has a zero log change in the bid (offer), 2) either the nearby or the deferred has a zero log change in the bid (offer), but not both, or 3) both the nearby and the deferred have a zero log change in the bid (offer).

Based on the definition of liquidity-based activity and information-based activity in the conceptual framework from above, we present a case for interpreting (1) as information-based activity, (2) liquidity-based activity, and (3) neither liquidity based activity nor information-based activity.

The intuition is that if both the nearby and deferred contracts experience a revision in the same direction, they are likely responding to the the arrival of information to the marketplace, and best bids (offers) adjusted accordingly. This in contrast to the case where one of the two contracts experienced a revision and the other contract had no change. If one contract experienced a revision in the best bid (offer) and the other did not, it is likely that the revision resulted from someone making a liquidity based order in an effort to exit their position.

If this intuition is correct, it is informative to consider only time-stamps for which both contracts experienced a revision. In the next section we will present results based on all data in our sample and the sub-sample where only the time-stamps for which both contracts experienced a non-zero revision to the top of the book are included as separate cases.

# Results

In this section we present a series of charts that display the calculations described in the preceding section. First we present all the correlation charts calculated contemporaneously: full sample, liquidity activity only, USDA report days only. Then we present correlation charts calculated at one and ten second time lags of the nearby contract for the full sample, liquidity activity only, and USDA report days only.

For each set of figures panel A is produced with bid-to-bid correlations and panel B is produced with offer-to-offer correlations of the nearby and deferred maturity contracts. Correlations are calculated over ten minute intervals starting at 9:30am Central Standard Time, the beginning of the daytime trading session for CBOT corn futures. One correlation is is calculated per day per ten minute bin, so for every ten minute bin we recover a distribution of correlations. The dots represent the mean of the distribution and the bars represent one standard deviation.

## Contemporaneous Correlation

In figure 1 contemporaneous correlation between the nearby and one, two, and three deferred maturity contracts are calculated. Correlations between both the nearby bid and deferred bids in panel A and nearby offers and deferred offers in panel B seem to have fairly stable means beginning around 0.6 at the start of the trading day and falling slightly to around 0.5 as trading progresses. Both means and one standard deviation error bars seem to exhibit a fairly constant and stable relationship over the trading day; there is no trend up or downward in the mean correlations as trading progresses.

It is interesting that the distribution of correlations between the nearby and deferred contract bids are at such similar levels, hovering between 0.5 and 0.6. Also, the correlation between nearby and two contracts deferred is highest, nearby and three contracts deferred is next highest, and the nearby and one contract deferred is lowest, across all ten minute intervals. This result is puzzling. We expected either a random ranking among correlations with the different deferreds or that the highest correlations would result from the nearest to mature contracts.

In panel B, we see a very similar panel after calculating the correlations for offers between the nearby and deferred maturities.

## Contemporaneous Correlation - Information-Based Activity Only

In figure 2 we see a very different pattern of correlation when we perform the calculations based only on time-stamps where both the nearby and deferred maturity experience non-zero revisions to the best bid (panel A) or offer (panel B). When we focus on what we are calling information-based activity, the correlations between the nearby and deferred best bids (offers) basically go to one.

This implies that in the event both contracts experience revisions to their respective limit order books, they are revised in lockstep. While some of this correlation is artificially induced due to the tick structure of price changes in this market (prices move in 1/8th of a cent increments.), not all of it can be attributed to that. Additionally, we since our data is only time-stamped to the second, we may be missing nuance that would be captured with data time stamped to the millisecond. However, we contend that the result further supports interpreting this activity as information based, even considering that the market has 1/8th of a cent tick structure and the data is time stamped to the second.

## Contemporaneous Correlation on USDA Report Days

Figure 3, is exactly analogous to figure 1 in that it calculates the contemporaneous correlation between nearby and deferred bids (panel A) and offers (panel B), keeping data for which zeros are present. We see a remarkably similar depiction compared to figure 1, in that the correlations hover between 0.5 and 0.6 throughout the trading day. There is visibly more variation in the means of these distributions, resulting from the small sample of report days we average over, but the one standard deviation error bars are similar to that generated by the full sample.

If there had been a difference in the pattern of correlations on USDA report days, one would have expected the first ten minutes of trading to display the largest effect. Contrary to this, however, there does not seem to be any discernible effect of USDA report releases on the market open, or in any other time bin. Two primary causes are suspected: 1) Since we removed days were the report release corresponded to limit price moves, we systematically removed report days where the most important information was conferred on the market. It is possible that the remaining days corresponded to USDA reports that contained relatively less relevant market information, or they contained information market participants widely expected beforehand. 2) Since USDA reports were released prior to the market open during this time period, the information may have already been fully incorporated by market participants by the time the market opened, resulting in no discernible difference in the pattern of correlations in the first (and subsequent) time bins.

## Contemporaneous Correlation - USDA Report Days and Information-Based Activity Only

Figure 4 is analogous to figure 2 in that is it generated by calculating the contemporaneous correlation between nearby and deferred bids (panel A) and offers (panel B), removing data where zeros are present. Again, this figure is very similar to its full sample counterpart, correlations between the nearby and deferred are nearly one throughout the trading day. Like in figure 3, we observe more variation due to the small sample of USDA report days relative to the full sample.

## Contemporaneous and Time-lagged Correlations

An important contribution of this analysis is to provide a measure of the speed at which information is transmitted from the nearby to the deferred contracts. To measure this, we lag the nearby series of log changes in the bid (offer) and calculate the correlation with the deferred bids (offers). This allows us to determine the length of time it takes for information to be fully transmitted to the deferred contracts. The assumption here is that the length of time it takes for the revisions to the top of the nearby limit order book to become uncorrelated with revisions to the top of the deferred limit order books is the length of time it takes for information to be transmitted between the two markets.

Figure 5 contains the correlations between the nearby and one deferred contracts. The graph shows the contemporaneous correlation from figure 1 as a reference, and correlations generated by lagging the nearby by one second and ten seconds respectively. Here we expected to see a clear pattern of decreased correlation as we increased the length of the lag in the nearby. However, we see that the correlation drops to zero with a lag of one second, which in this data set is the shortest time lag possible.[[2]](#footnote-35)

There are two possible explanations for this. First, it is possible that there is a clear and decreasing correlation between lags of the nearby and the deferreds, but it can only be observed on mili- or micro-second time stamps. Then, when aggregating to the nearest second, we only observe contemporaneous correlation, but zero correlation even at the shortest possible time delay.

This would also occur if linkages between the nearby and deferred contracts were immediately enforced by spread traders. Surely the spread trade is an important component that keeps nearby and deferred contracts linked in economically meaningful ways. Spread trading does not really explain this phenomenon, however, because a spread trade is entered as a bid (offer) in the nearby and a offer (bid) in the deferred contract. Until now, we have been measuring correlations bib-to-bid and offer-to-offer between the nearby and deferred contracts.

In figure 6 we generate a figure analogous to figure 1 in that it calculates the contemporaneous correlations between the nearby and deferred revisions to the top of the book. Figure 6, however, calculates the correlation between revisions to the nearby by bid with the deferred offer in panel A, and nearby offer with deferred bid in panel B. Here, as in figure 1 we see correlations hovering between 0.5 and 0.6. Perhaps the only distingishable characteristic of this figure compared to figure 1 is that the unexpected ordering of correlations highest to lowest as 2, 1, and 3 deferred that is present in figure 1 is no longer evident. The correlations of the nearby with the different deferred maturities is noticeably tighter in the bid-to-offer and offer-to-bid correlations.

# Discussion

Spread traders? High frequency trade?

## A Measure of the Proportion of Market Activity Attributable to Information- verses Liquidity-Based Activity

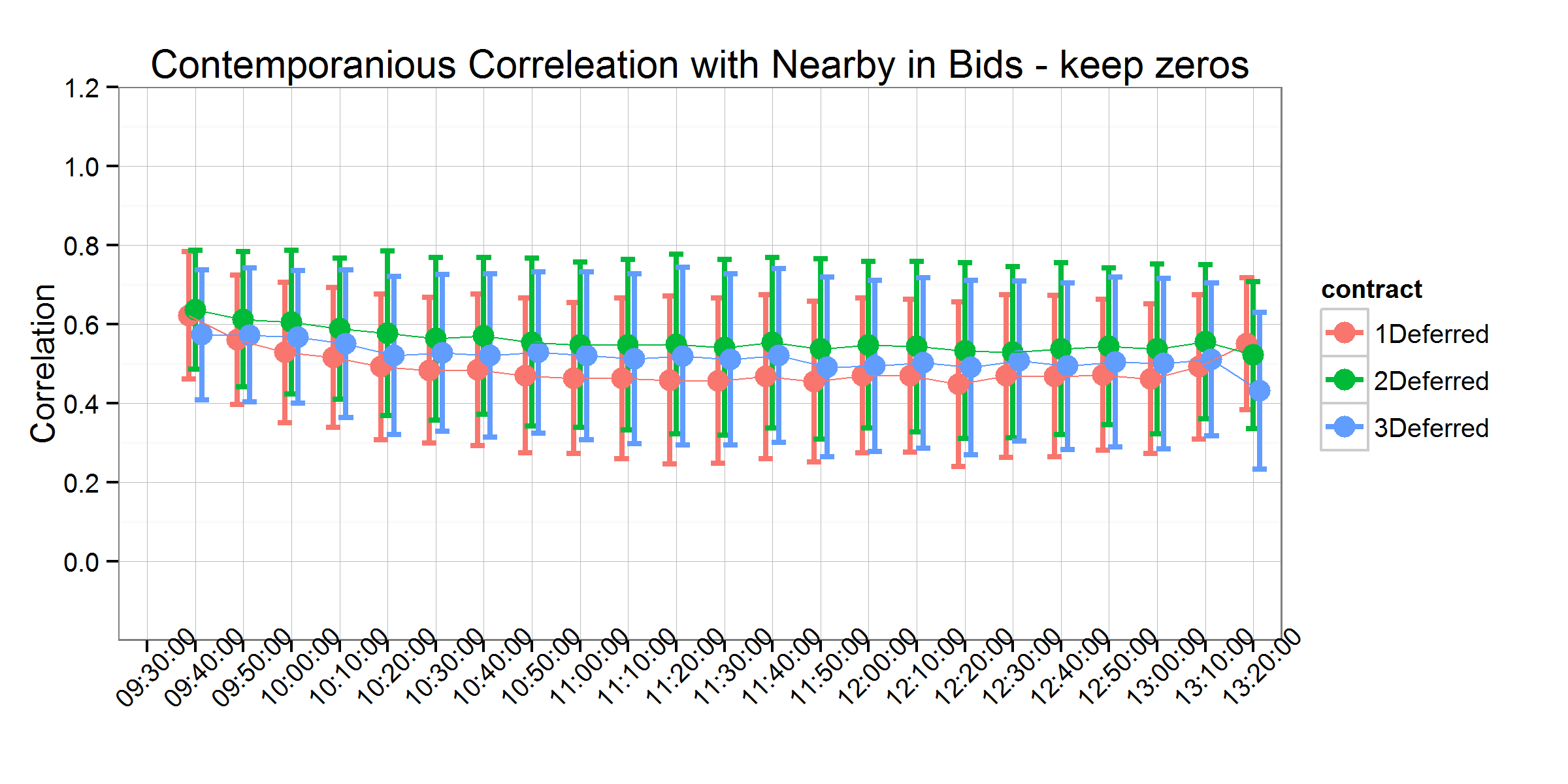
# Conclusions

# Tables

Table 1: First ten entries in our data set.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | EX | SYMBOL | OFRSIZ | OFR | BIDSIZ | BID |
| 2010-01-04 09:30:00 | 98790 | 1003 | 1475 | 423.75 | 1188 | 423.75 |
| 2010-01-04 09:30:00 | 98800 | 1003 | 1483 | 423.75 | 1188 | 423.75 |
| 2010-01-04 09:30:00 | 98810 | 1003 | 1483 | 423.75 | 1197 | 423.75 |
| 2010-01-04 09:30:00 | 98820 | 1003 | 1486 | 423.75 | 1197 | 423.75 |
| 2010-01-04 09:30:00 | 98830 | 1003 | 1486 | 423.75 | 1231 | 423.75 |
| 2010-01-04 09:30:00 | 98840 | 1003 | 1494 | 423.75 | 1231 | 423.75 |
| 2010-01-04 09:30:00 | 98850 | 1003 | 1496 | 423.75 | 1231 | 423.75 |
| 2010-01-04 09:30:00 | 98860 | 1003 | 1510 | 423.75 | 1231 | 423.75 |
| 2010-01-04 09:30:00 | 98870 | 1003 | 1510 | 423.75 | 1233 | 423.75 |
| 2010-01-04 09:30:00 | 98880 | 1003 | 1520 | 423.75 | 1234 | 423.75 |

# Figures



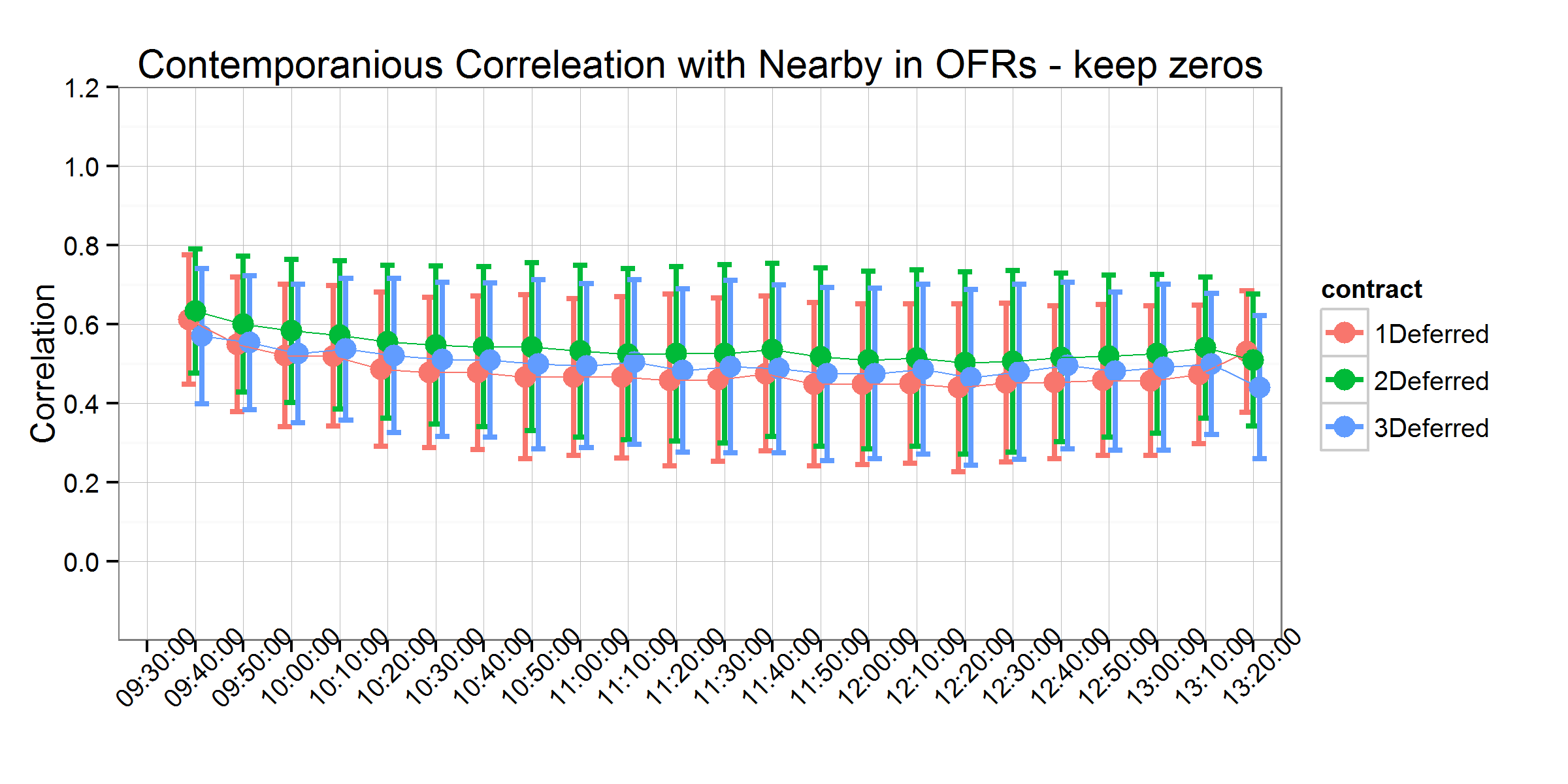
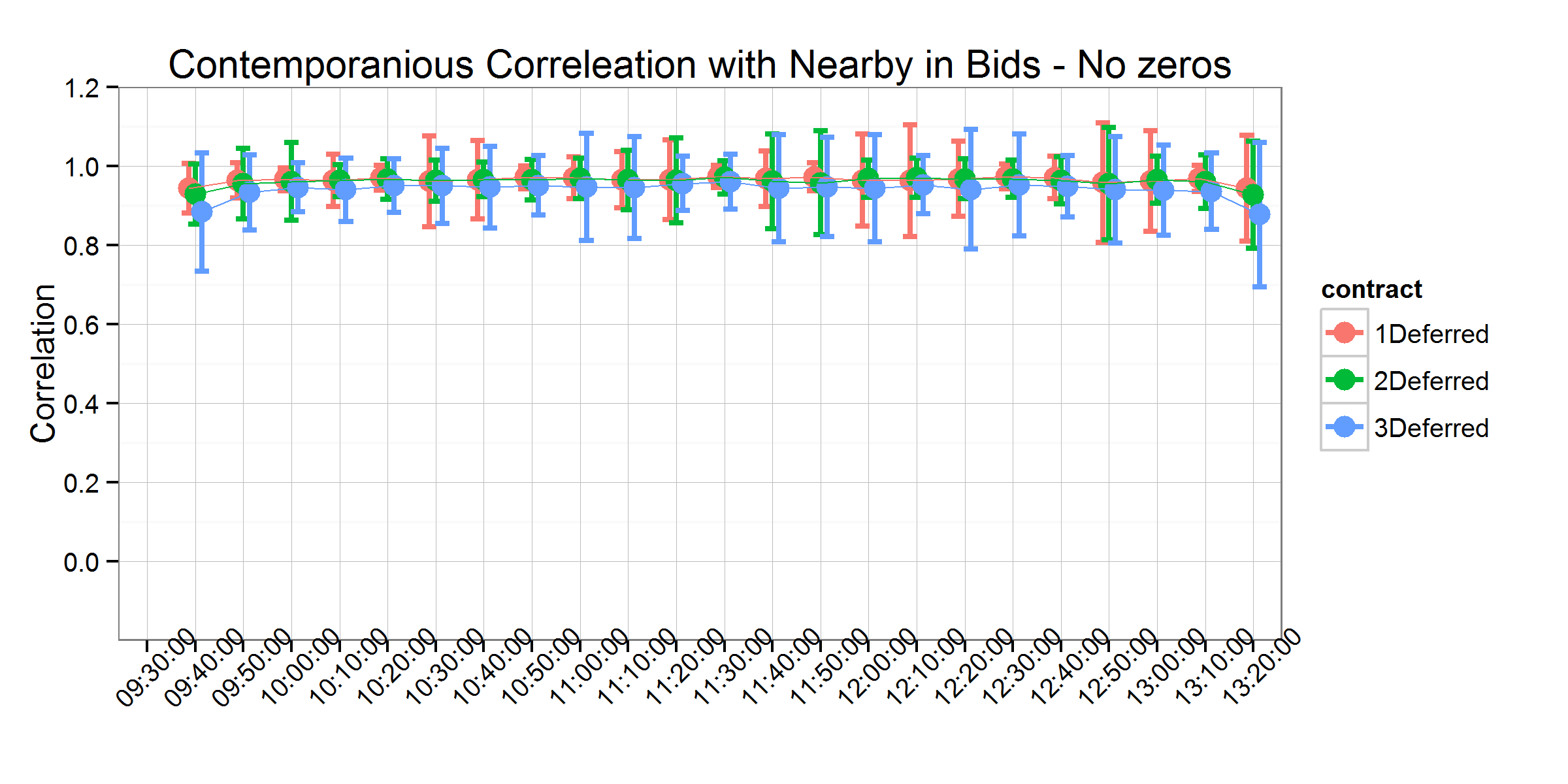


Figure 1. Contemporaneous Correlation with Nearby - Keep Zeros



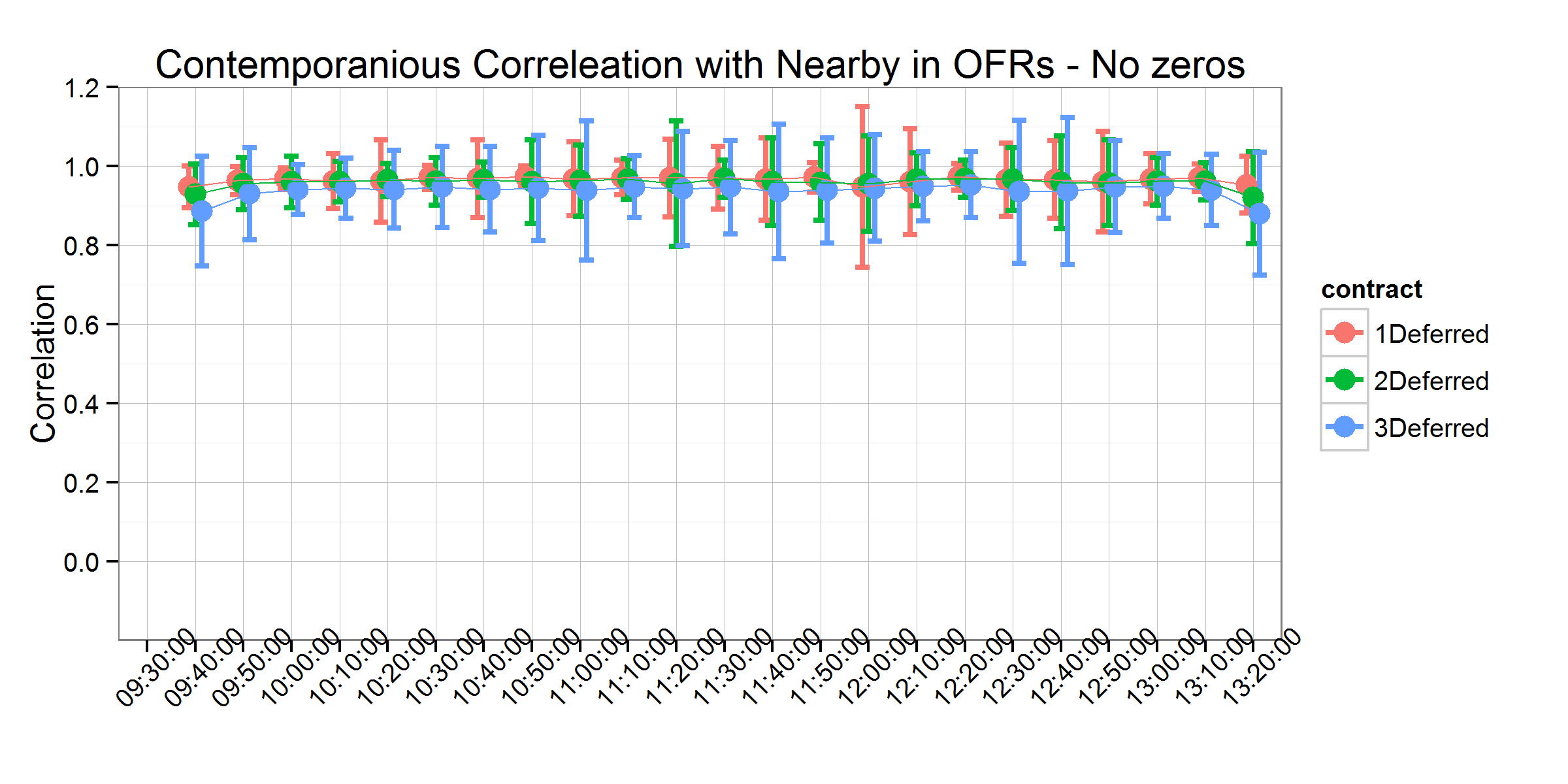


Figure 2. Contemporaneous Correlation with Nearby - No Zeros



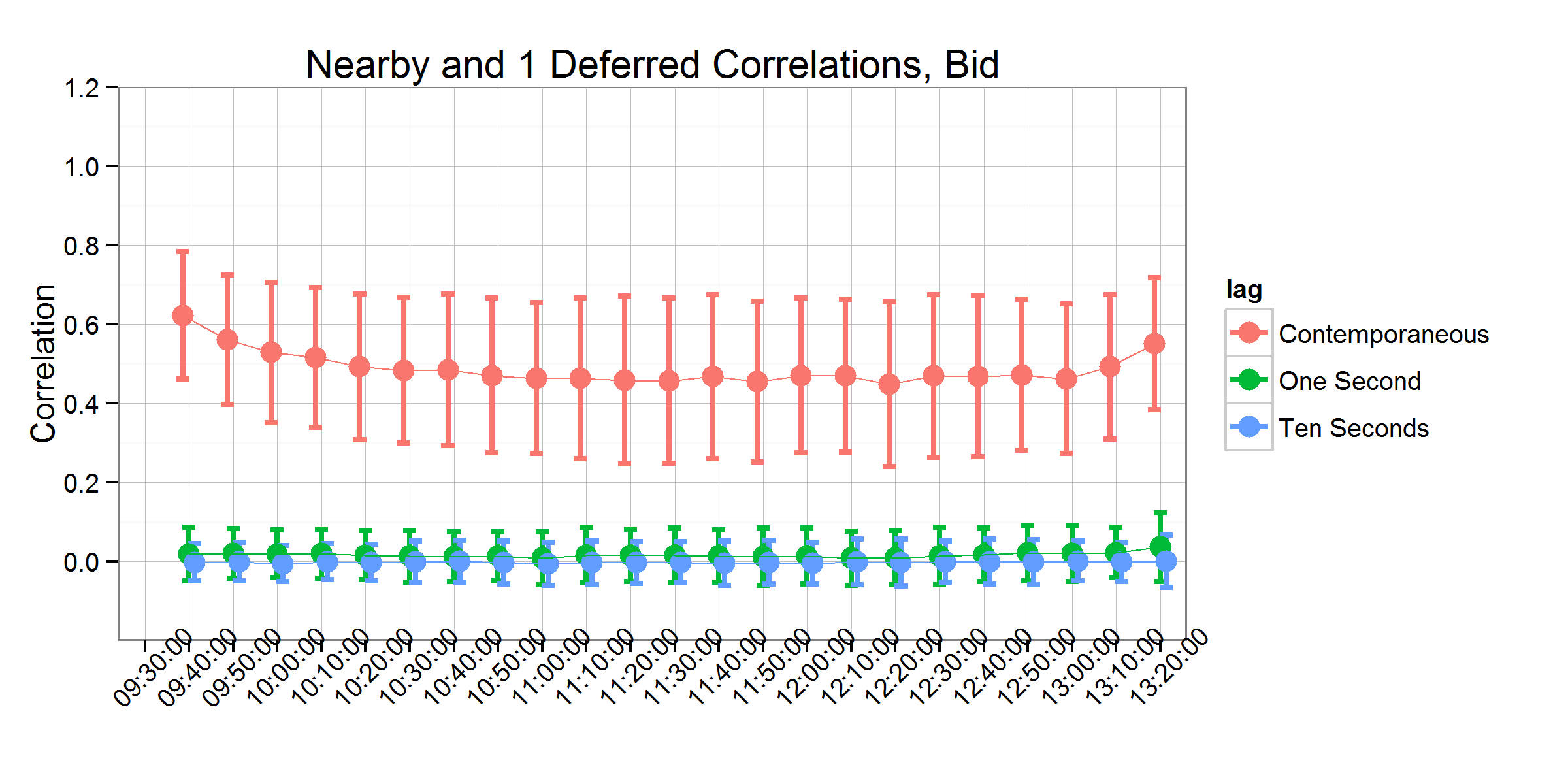


Figure 3. Contemporaneous Correlation with Nearby on USDA Report Days - Keep Zeros





Figure 4. Contemporaneous Correlation with Nearby on USDA Report Days - No Zeros



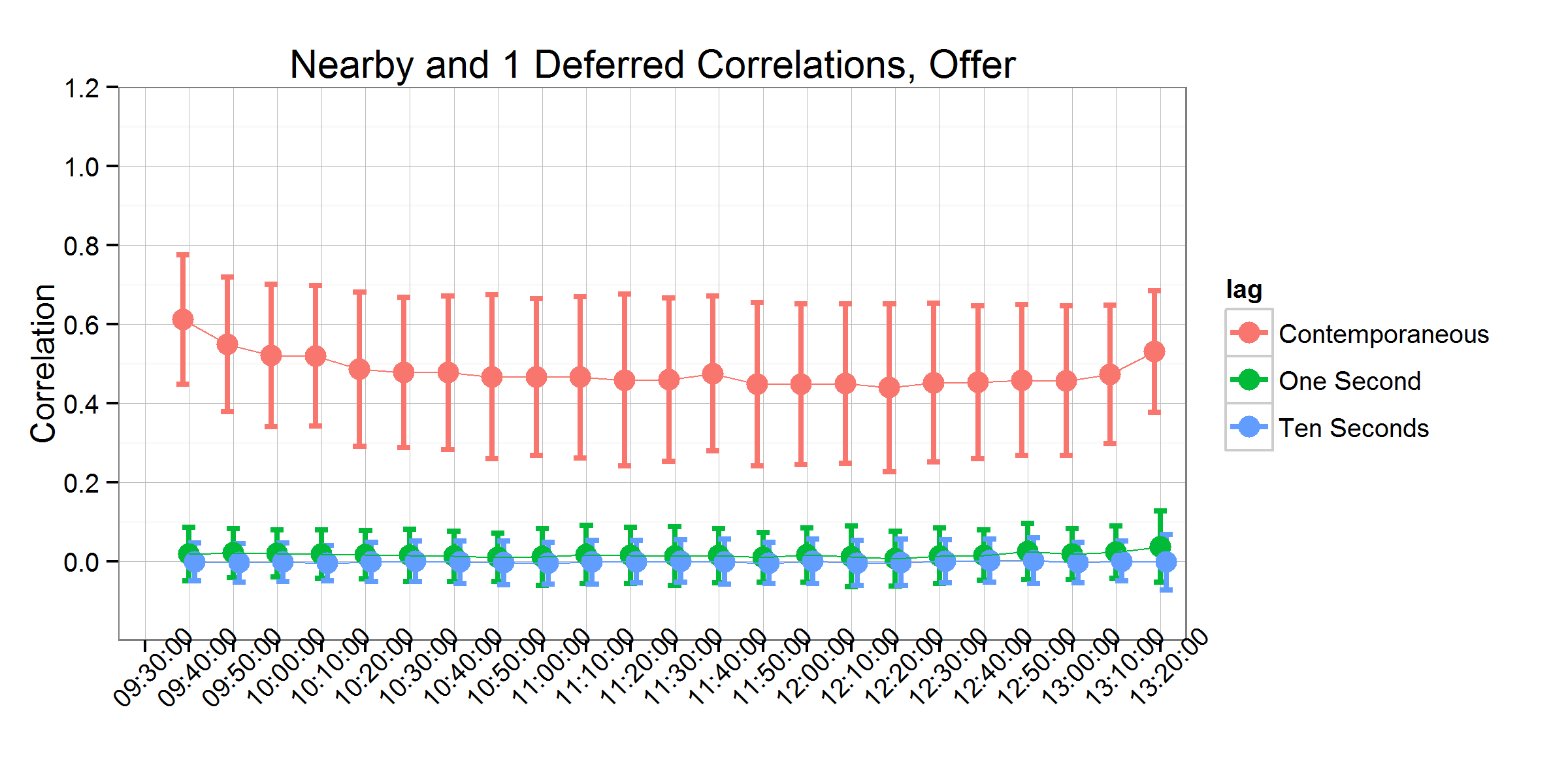
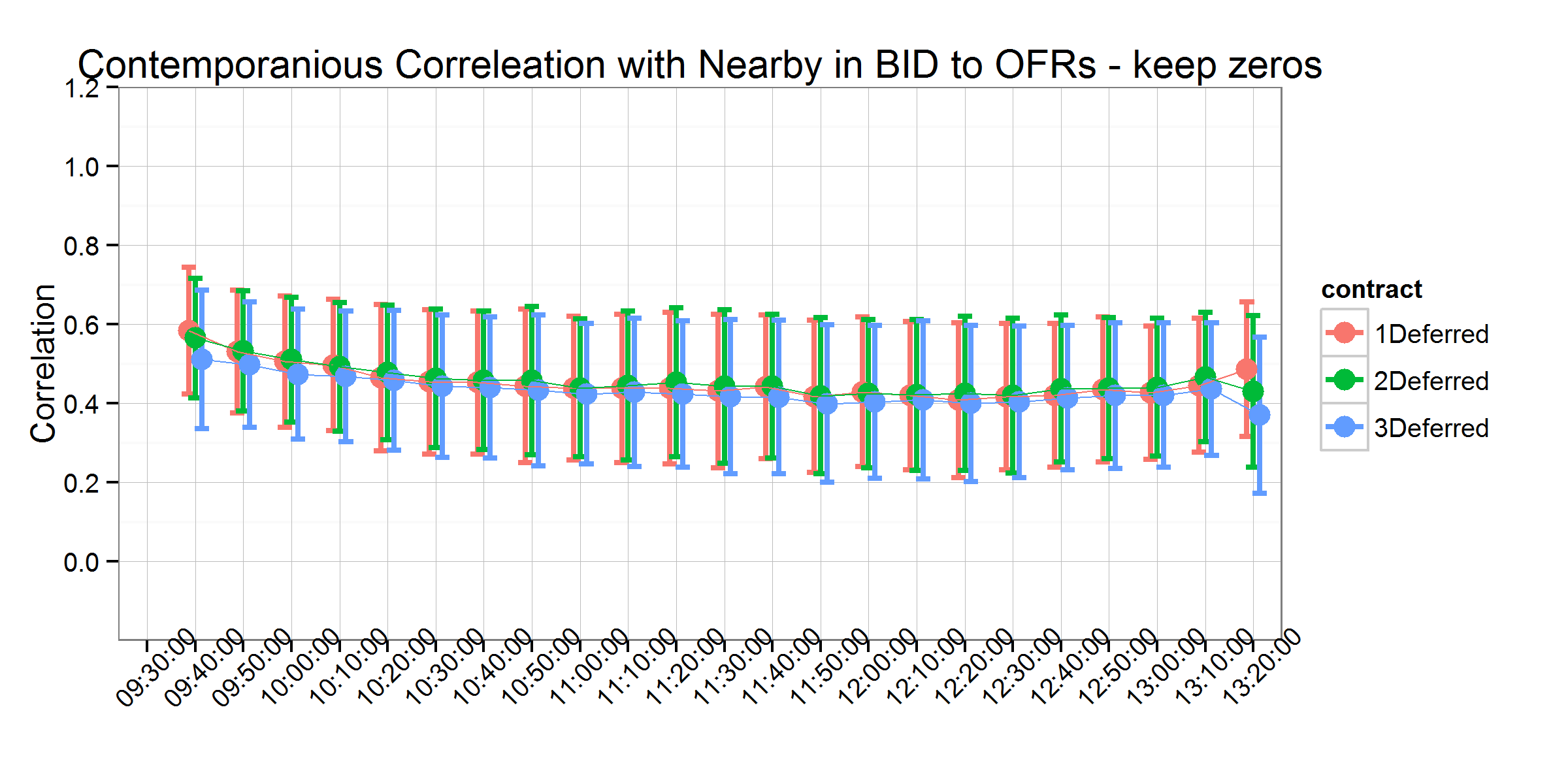


Figure 5. Contemporaneous and Time-lagged Correlations between the Nearby and One-deferred Contract





Figure 6. Contemporaneous and Time-lagged Correlations between the Nearby and One-deferred Contract on USDA Report Days



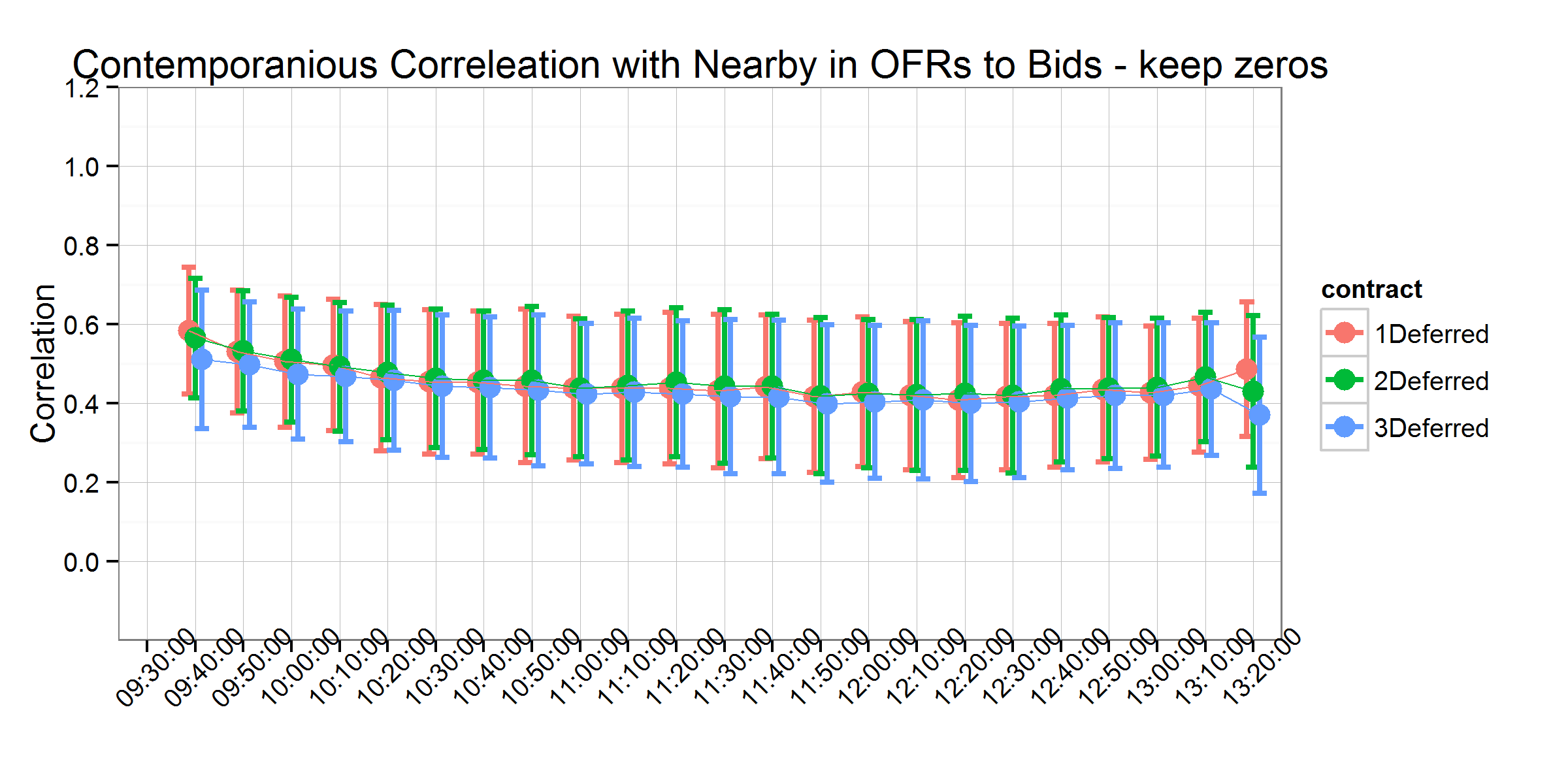


Figure 7. Contemporaneous Correlations with Nearby Bid-to-Ask and Ask-to-Bid - Keep Zeros

# References

Akerlof, G.A. 1970. “The market for‘ lemons’: Quality uncertainty and the market mechanism.” *The Quarterly Journal of Economics* 84(3):488–500.

Bagehot, W. 1971. “The only game in town.” *Financial Analysts Journal* 27(2):12–14.

Boehmer, E., J. Grammig, and E. Theissen. 2007. “Estimating the probability of informed trading-does trade misclassification matter?” *Journal of Financial Markets* 10(1):26–47. Available at: <http://www.scopus.com/inward/record.url?eid=2-s2.0-33846279422&partnerID=40&md5=864042ba646f1d94703dccc526ad2da8>.

Copeland, T.E., and D. Galai. 1983. “Information effects on the bid-ask spread.” *The Journal of Finance* 38(5):1457–1469.

Easley, D., and M. O’Hara. 1987. “Price, trade size, and information in securities markets.” *Journal of Financial economics* 19(1):69–90.

Easley, D., and M. O’Hara. 1992. “Time and the process of security price adjustment.” *The Journal of Finance* 47(2):577–605.

Easley, D., S. Hvidkjaer, and M. O’Hara. 2002. “Is information risk a determinant of asset returns?” *The Journal of Finance* 57(5):2185–2221.

Easley, D., N.M. Kiefer, and M. O’Hara. 1997. “One day in the life of a very common stock.” *Review of Financial Studies* 10(3):805–835.

Easley, D., M. Lopez de Prado, and M. O’Hara. 2011. “The microstructure of the ‘flash crash’: Flow toxicity, liquidity crashes and the probability of informed trading.” *The Journal of Portfolio Management* 37(2):118–128.

Glosten, L.R., and P.R. Milgrom. 1985. “Bid, ask and transaction prices in a specialist market with heterogeneously informed traders.” *Journal of Financial Economics* 14(1):71–100.

Hasbrouck, J. 2006. *Empirical market microstructure: The institutions, economics, and econometrics of securities trading*. Oxford University Press.

Kumar, P., and D.J. Seppi. 1994. “Information and index arbitrage.” *Journal of Business* 67(4):481–509.

O’Hara, M. 1995. *Market microstructure theory*. Blackwell Cambridge, MA.

Vega, C. 2006. “Stock price reaction to public and private information.” *Journal of Financial Economics* 82(1):103–133. Available at: <http://www.scopus.com/inward/record.url?eid=2-s2.0-33748041619&partnerID=40&md5=1eb71f6d3df8978c80f959991cfc4348>.

Yan, Y., and S. Zhang. 2012. “An improved estimation method and empirical properties of the probability of informed trading.” *Journal of Banking & Finance* 36(2):454–467. Available at: <http://www.sciencedirect.com/science/article/pii/S0378426611002433>.

1. The interested reader can refer to O’Hara (1995) for an excellent and detailed overview of the evolution of this literature. [↑](#footnote-ref-23)
2. We generated this figure for USDA report days only. Similar to our result before, it is basically the same as figure 5 with more variation due to the small sample. [↑](#footnote-ref-35)