Comparison of Intra-day and Daily Stock Price Data

Bradley Okresik
http://bokresik.home.comast.net bokresik@comcast.net

Introduction

The purpose of this analysis is to compare 30 minute data to daily data obtained from the same source. The 30 minute data is complied into daily open, high, low, close, and volume. If the data source is consistent then the 30 minute data compiled to daily will be equal to the raw daily data. I calculated and plotted the difference between these two daily price records.

Body

Apparently the only reliable security price data is time and sales from an exchange where all transactions are processed through one platform (Ex. CME Globex). I already know that the data provided to us by the brokerage houses is not totally accurate, but I wanted to quantify that lack of accuracy. I spent a weekend comparing 30 minute data with daily data from TD Ameritrade's Strategy Desk application and found the data to be fairly inconsistent with itself. TD tells me that they obtain the data from E-Signal. I notice that most of the common websites also use this same data because their price charts have always matched those from TD. I exported daily and 30 minute data for 4 securities (SPY, QQQQ, DELL, AAPL) from 8/16/05 to 9/7/07. I compiled the 30 minute data into a daily timeframe calculating the daily open, high, low, close, and volume from the 30 minute data. I subtracted the 30 minute compiled data from the raw daily data and plotted the difference. Ideally all plots should exhibit a straight line at zero, indicating that the 30 minute data is consistent with the daily data. The code that processes this data was developed using SPY and was tested repeatedly by sampling several data points from the plots presented below. Each sampled data point was verified against the 30 minute and daily bar chart in my TD Ameritrade application. In the plots below, the X axis represents the number of trading days since 8/16/07. The Y axis is the difference (raw daily – 30 minute compiled to daily) for the open, high, low, close, or volume as labeled in the plot title. In the case of volume data, I have also presented the volume discrepancy as a percentage of daily reported volume.

One sampling of 10 minute data was found to be exactly the same as the 30 minute data. 1, 3, and 5 minute data have not been analyzed and I do not currently intend to do so.

Results

The data for the equities (DELL, AAPL) were much more consistent than the index tracking ETFs (SPY, QQQQ). Daily highs are regularly higher than their intra-day counterparts and daily lows are regularly lower than their intra-day counterparts. The daily data typically exhibits more range than the intra-day data. Opening price is fairly consistent except for a few occasions, but closing price is almost never consistent. The daily volume is always higher than the intra-day volume.

SPY comments:

The opening price was consistent until some event occurred on 3/1/07, where the daily open is reported \$1.59 higher than the intra-day open. Since 3/1/07 there have been regular discrepancies less than \$ .20. 1/26/06 is an interesting data point. On the daily candle chart the high is stated as 127.67 and low is \$127.76, thus the open, close and high are all lower than the low! The intra-day low is \$126.67. This is the only time an intra-day low was found lower than a daily low. On 8/17/07 the SPY high has a discrepancy of \$2.45 (daily high – \$148.35, intra-day high – \$145.90). SPY fairly regularly has low prices that have discrepancies of more than \$ .50 and volume discrepancies can reach 20-30% of the daily reported volume.

QQQQ comments:

In general QQQQ appears to be more accurate than SPY, although its price is 1/3 that of SPY. On 2/26/07 QQQQ exhibits a low price discrepancy of \$1.61 (daily low – \$43.15, intra-day low – \$44.76). The closing prices maintained reasonable consistency until the market volatility increased during the last several months.

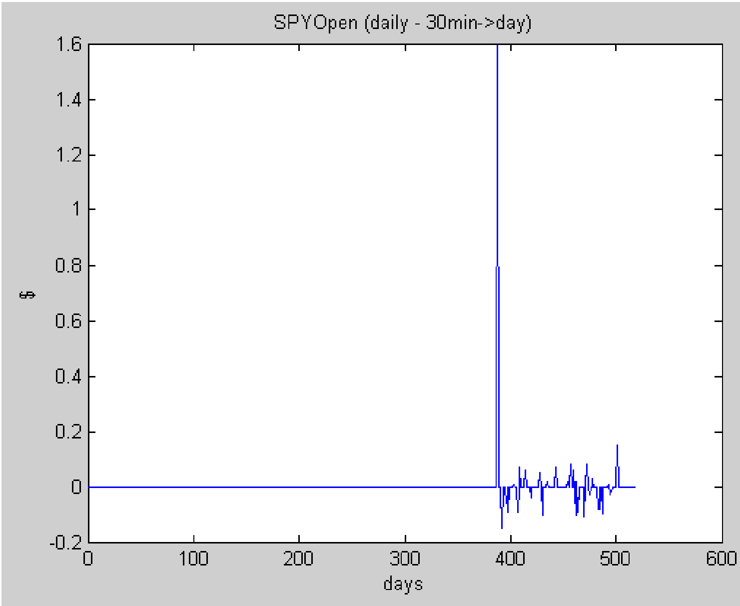
AAPL comments:

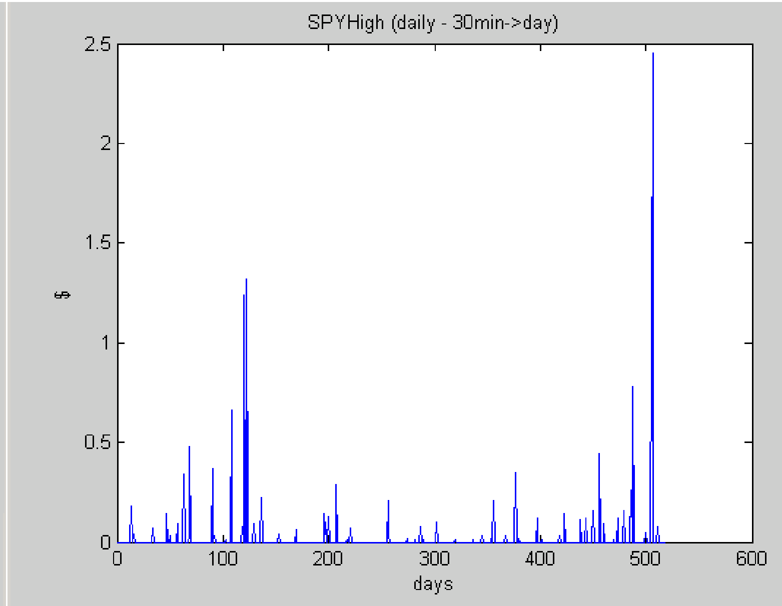
The volume discrepancies appear to exhibit a periodic nature.

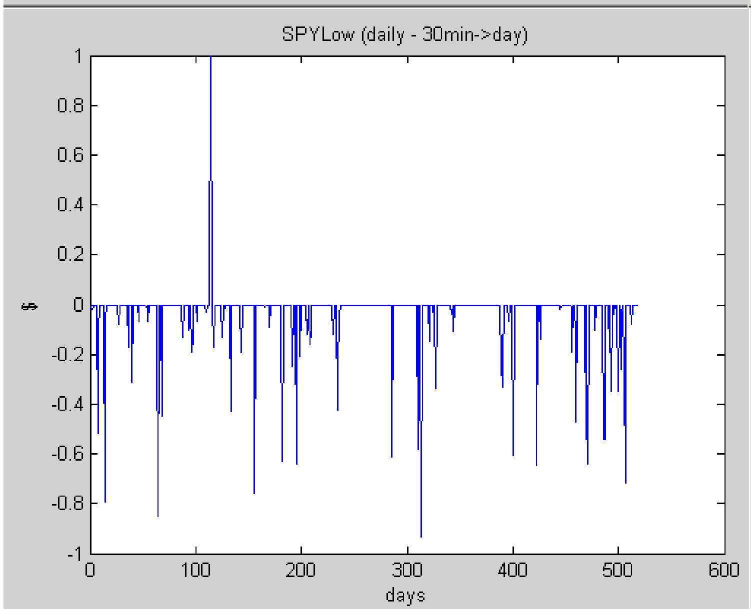
DELL comments:

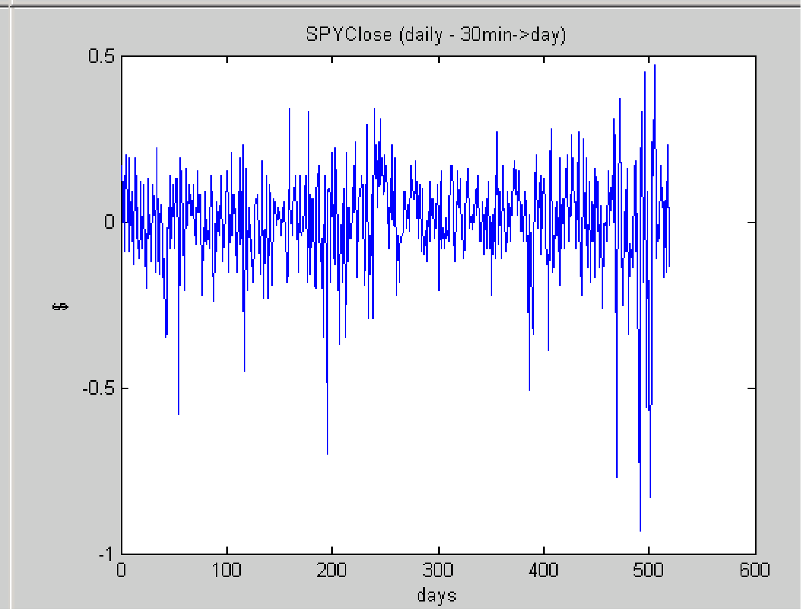
Volume discrepancies have approached 45% of the daily reported volume.

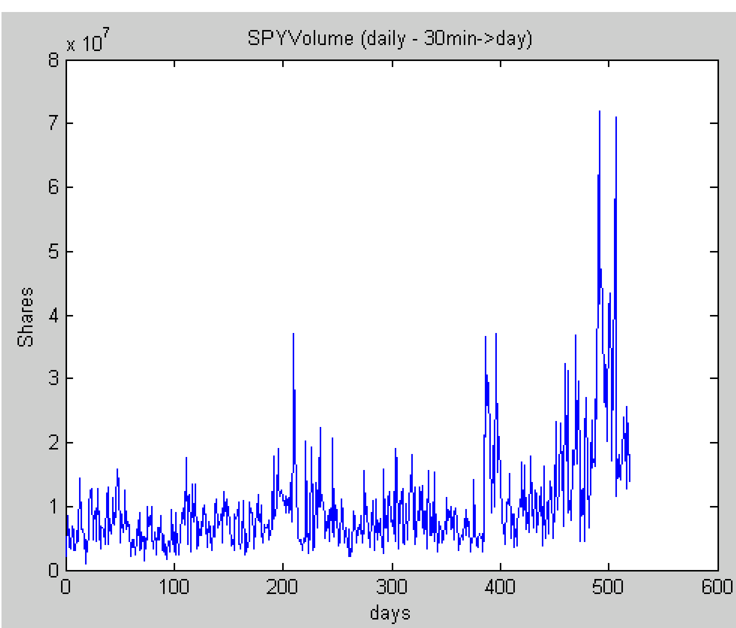
Reference – See Appendix for a statement on premarket and afterhours trade data.
Spreadsheets containing data generated and displayed in this report:
http://bokresik.home.comcast.net/finance/data_integrity_study/data_integrity_study.xls

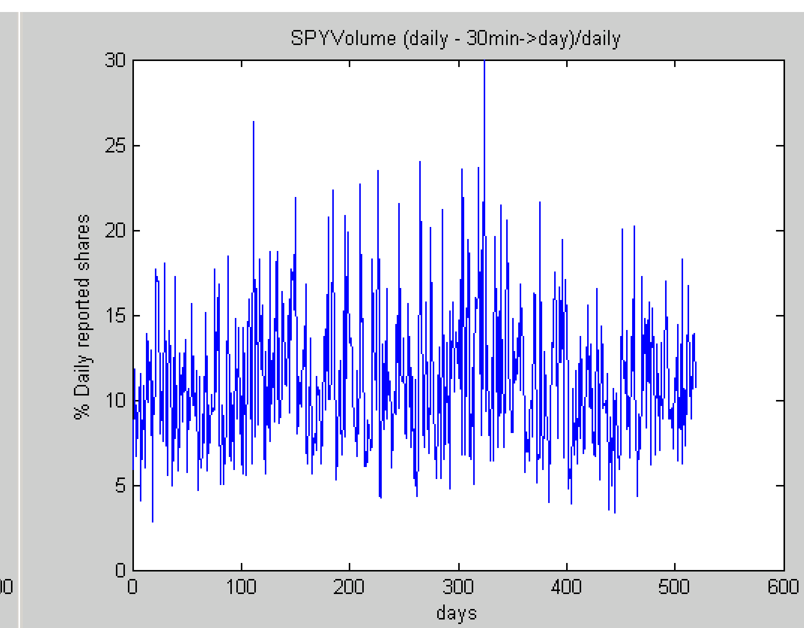

SPYOpen (daily - 30min->day)
$
days

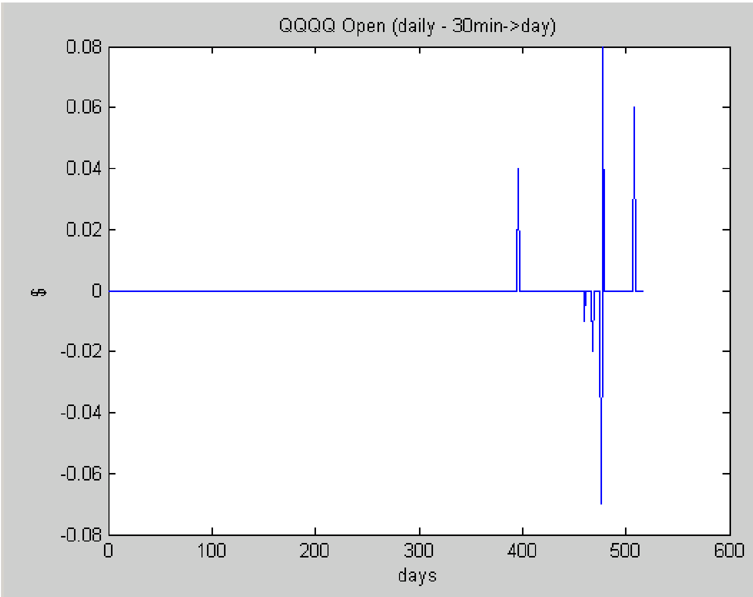

SPYHigh (daily - 30min->day)
$
days

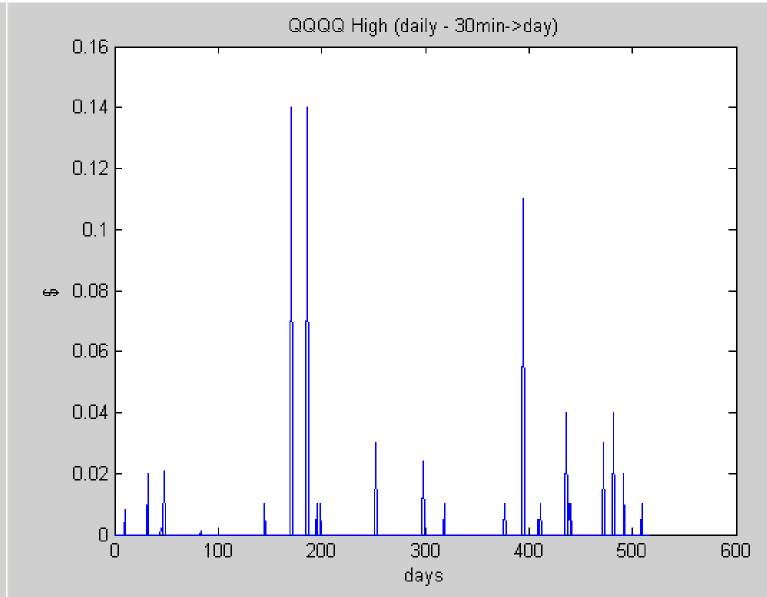

SPYLow (daily - 30min->day)
$
days

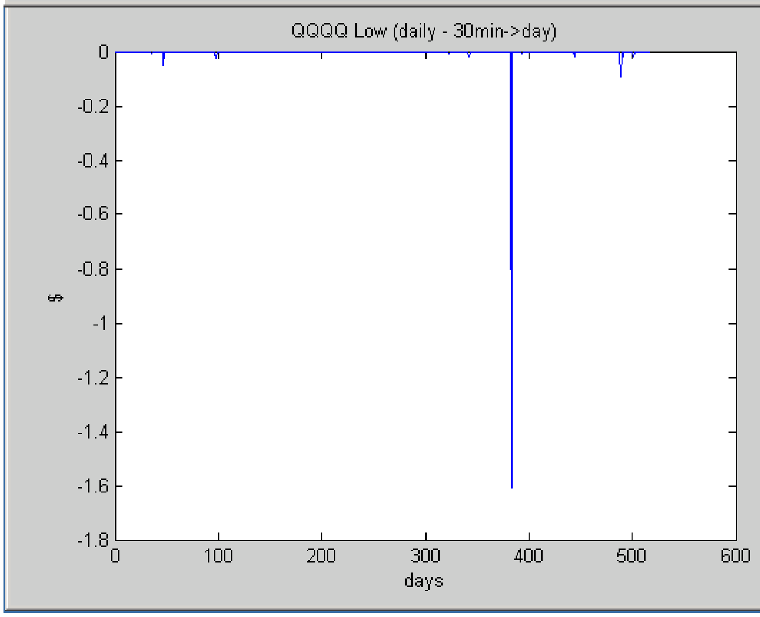

SPYClose (daily - 30min->day)
$
days

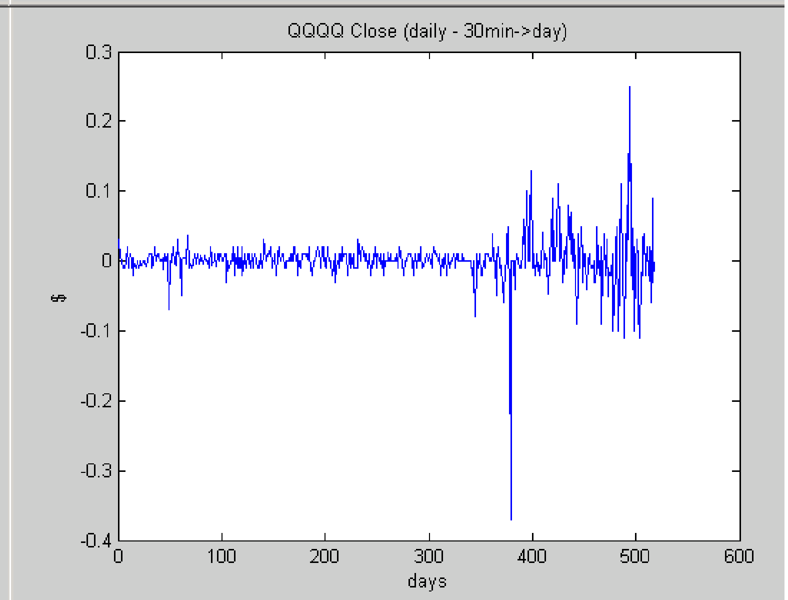

SPYVolume (daily - 30min->day)
x 10^7
Shares
days

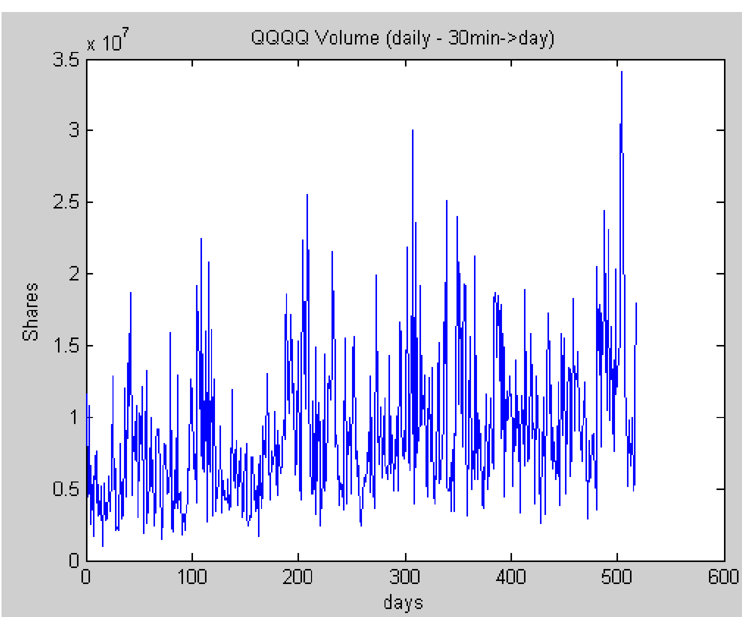

SPYVolume (daily - 30min->day)/daily
% Daily reported shares
days

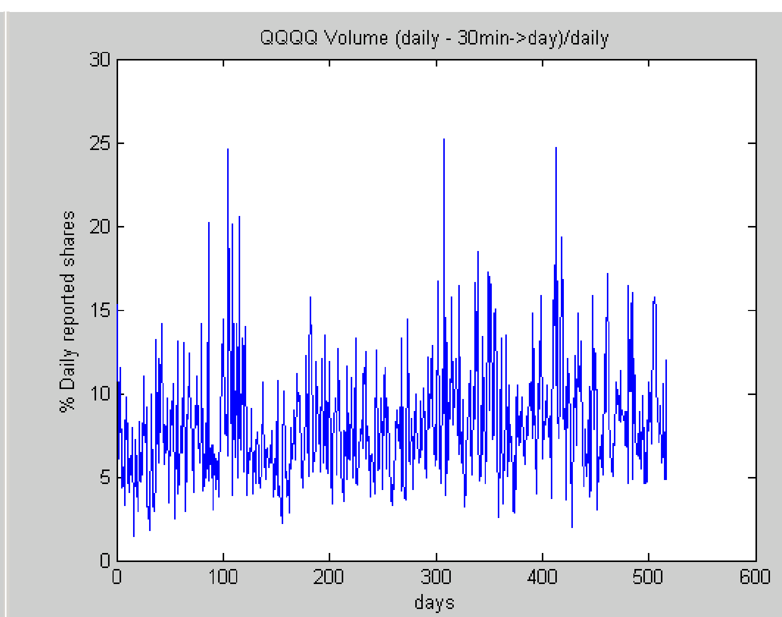

QQQQ Open (daily - 30min->day)
$
days

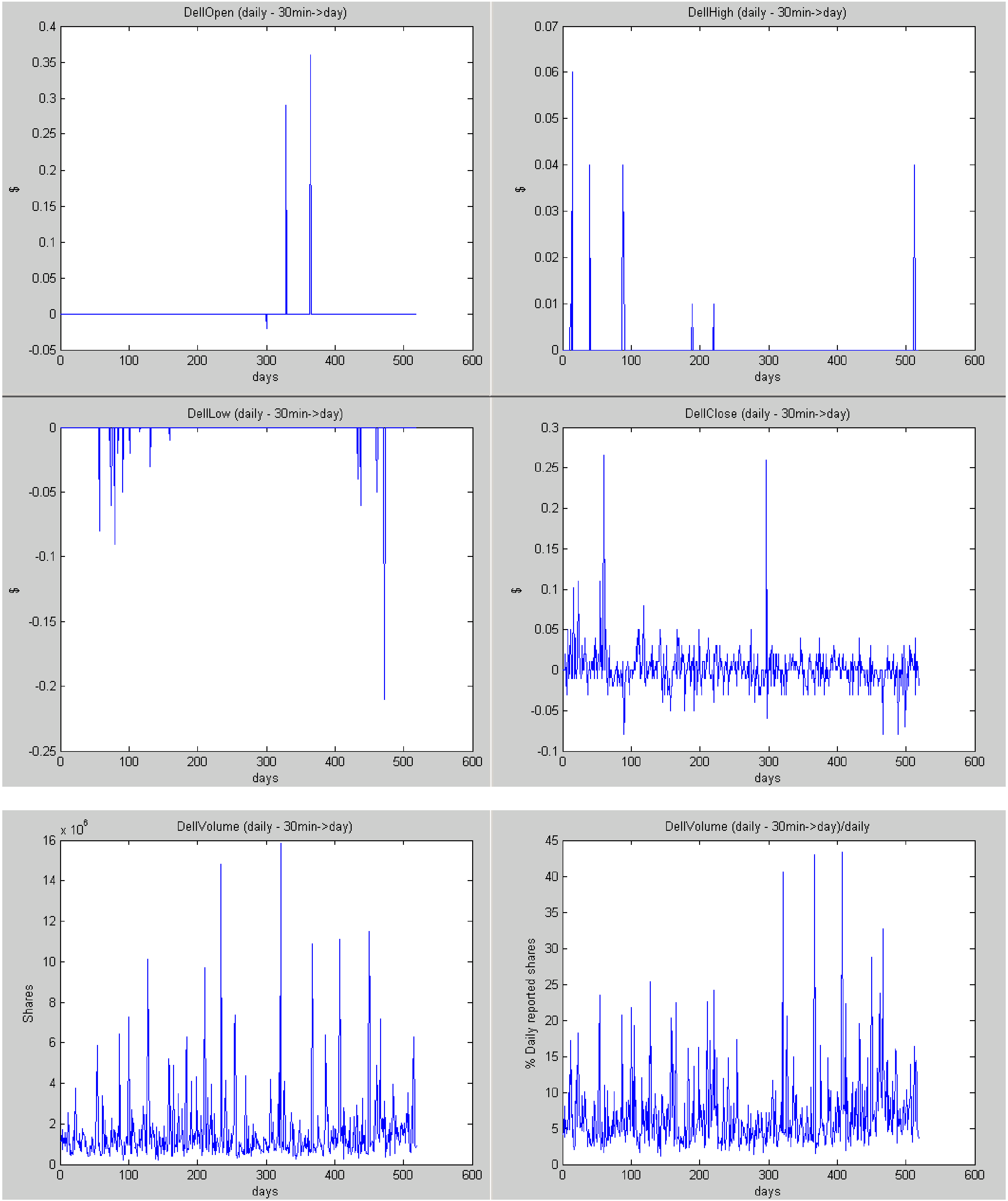

QQQQ High (daily - 30min->day)
$
days


QQQQ Low (daily - 30min->day)
$
days

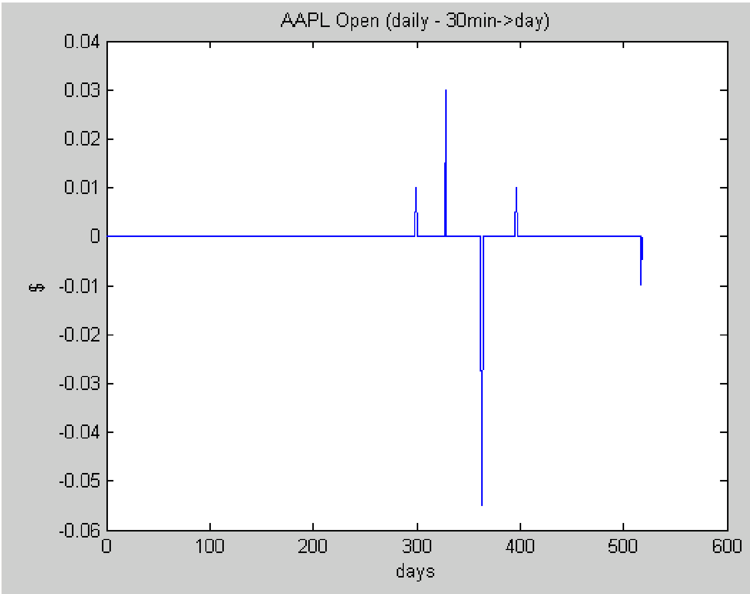

QQQQ Close (daily - 30min->day)
$
days

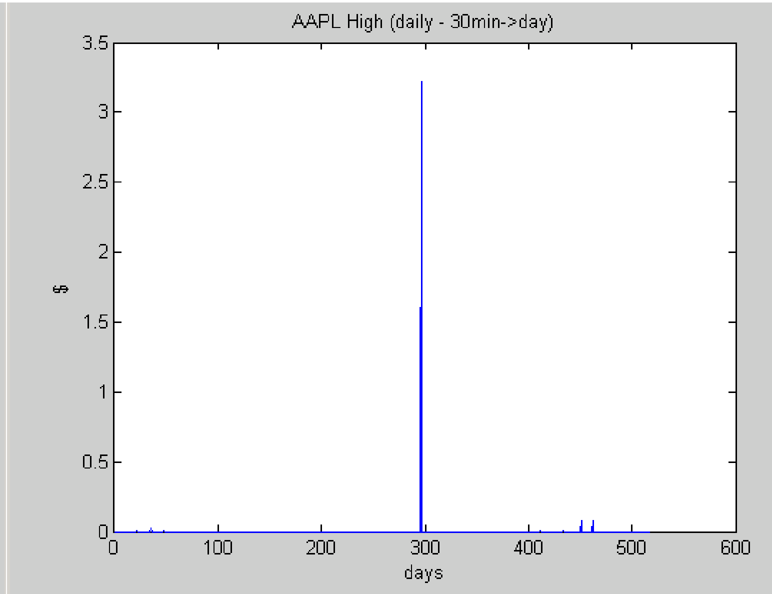

QQQQ Volume (daily - 30min->day)
x 10^7
Shares
days

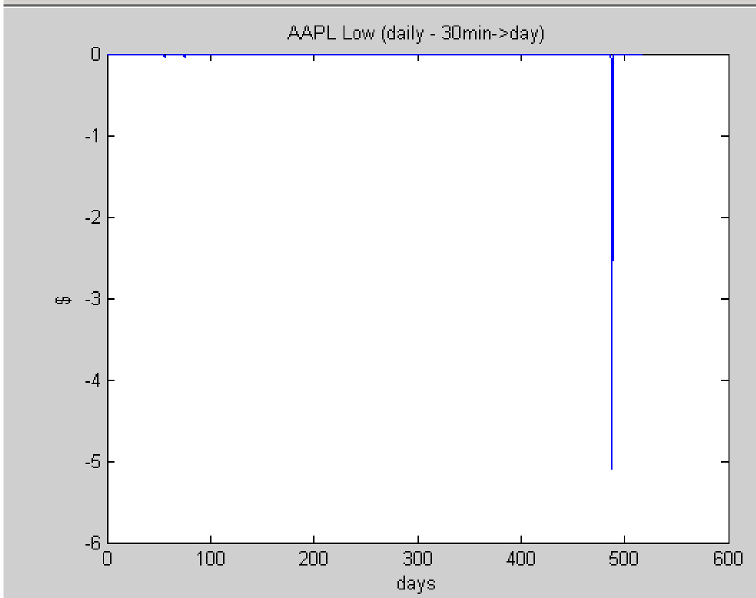

QQQQ Volume (daily - 30min->day)/daily
% Daily reported shares
days

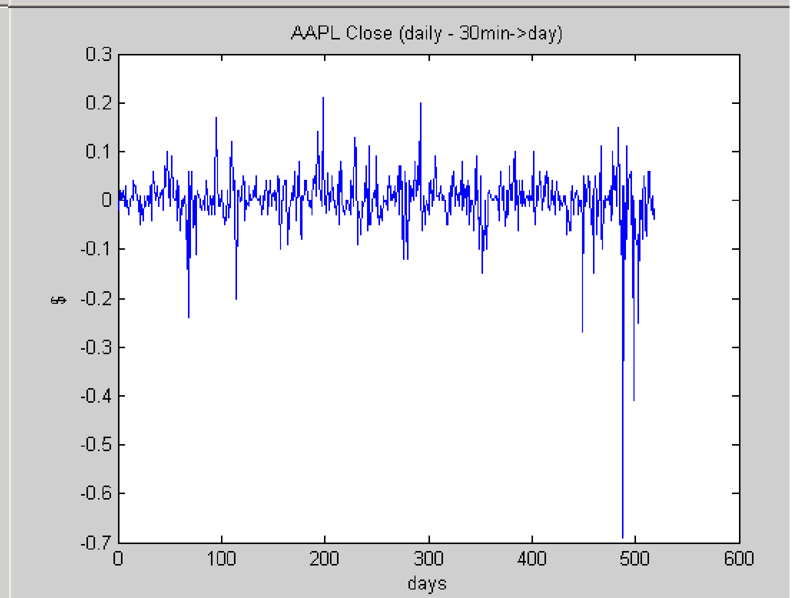


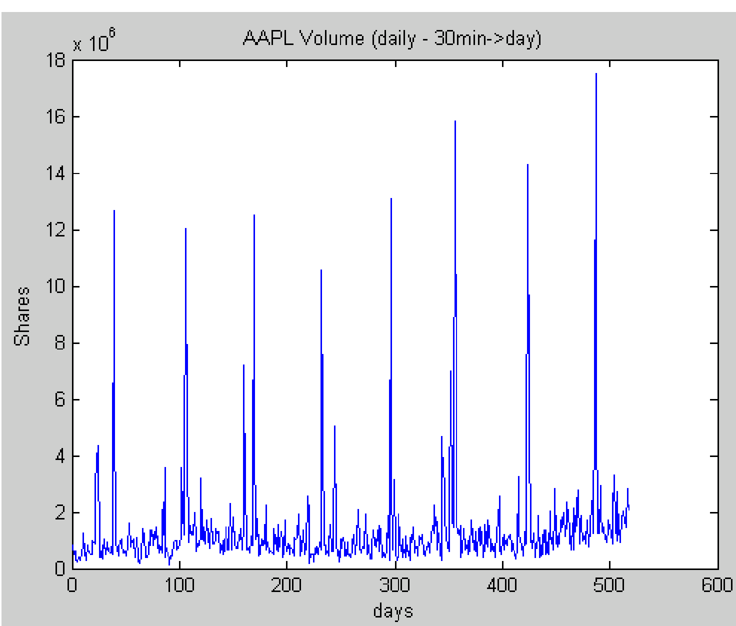
DellOpen (daily - 30min->day)
$
days
DellHigh (daily - 30min->day)
$
days
DellLow (daily - 30min->day)
$
days
DellClose (daily - 30min->day)
$
days
DellVolume (daily - 30min->day)
x 10^6
Shares
days
DellVolume (daily - 30min->day)/daily
% Daily reported shares
days

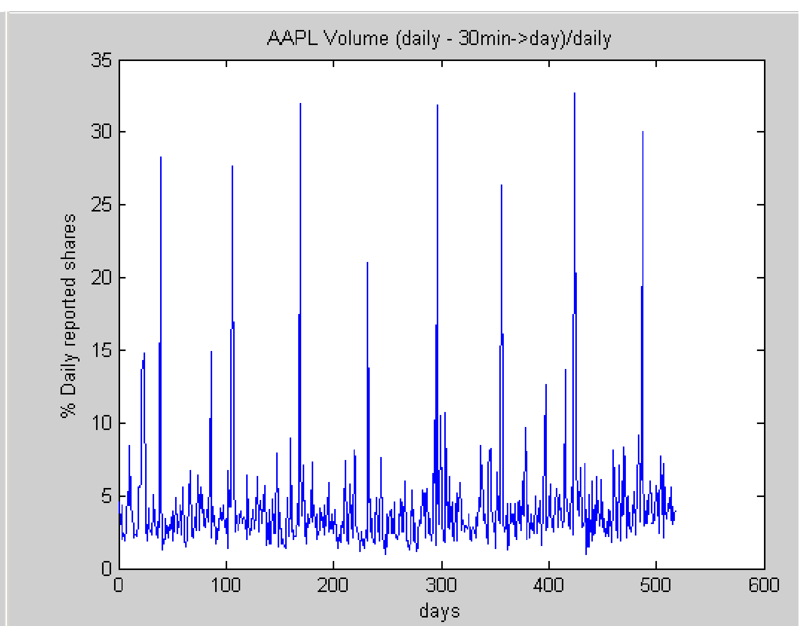

AAPL Open (daily - 30min->day)
$
days


AAPL High (daily - 30min->day)
$
days


AAPL Low (daily - 30min->day)
$
days


AAPL Close (daily - 30min->day)
$
days


AAPL Volume (daily - 30min->day)
x 10^6
Shares
days


AAPL Volume (daily - 30min->day)/daily
% Daily reported shares
days

Appendix

I also considered the premarket and afterhours data available from TD Ameritrade, but this certainly did not account for the discrepancies. I manually compared the 5 days of pre and post market data that are available. The opens are taken from the 9:30 reading (not premarket). The closes are either on the 4pm close or they bounce around aimlessly (same as before). I found several incidents where including the pre and post market trading would have made the intra-day range much larger than the daily range (the opposite of what I currently calculated). I found no evidence that the pre and post market data improved the price comparison. In fact, I found a few occurrences out of a dozen that would have made the price comparison look worse. The intra-day volume however is much more accurate, coming within 1% of daily volume for a all 5 days considered.