Constructing VWAP Curves

Pragma Securities

Abstract—**One of the most popular trading strategies is the Volume-Weighted Average Price (VWAP) trading algorithm. A VWAP algorithm tries to achieve an average execution price which is as close as possible to the realized VWAP in the market. VWAP trading algorithms use a pre-defined trading schedule that tries to approximate the volume pattern realized over the trading day. This schedule is calculated by averaging historical trading patterns. In this note we update our previous studies [1] [2] and consider further refinements.**

I. Preliminaries

A. Introduction

Over the last several years, the financial landscape has changed dramatically, moving from high-touch human traders to black-box systems where people are not involved in the actual trading process. Whether humans or machines are doing the actual trading, their performance is measured relative to some benchmark. Over the years many benchmarks have been proposed, but probably the most popular one is the volume-weighted average price, or VWAP. The VWAP represents the average price per share that was paid in the market during the life of the trade. As such it represents an ideal price that a trader would like to capture for the principal. This simple and intuitive benchmark is very appealing to many market participants.

There are many trading algorithms that try to achieve an average trade price close to the market-wide VWAP for the day. These algorithms are divided into two main groups: static and dynamic. Static algorithms use a pre-defined trading policy which does not change in response to market conditions. Dynamic algorithms use various real-time indicators in order to decide at what rate to trade. In essence, a dynamic strategy makes small deviations from the commands supplied by a static algorithm, based on current market conditions.

The most common static VWAP strategy is to follow the historical intra-day seasonality. While intuitively simple, the task of estimating intra-day patterns is tricky. There are two approaches: estimate the intra-day seasonality on a stock by stock basis [3], or average over many stocks [4]. The first approach requires averaging over a long history in order to achieve a stable estimate for each stock. This long period results in a mis-specification, i.e., errors, in the estimated curves due to changing market conditions. The second approach requires a much shorter history, which results in a much faster adaptation to changing market conditions. However, this faster adaptation does not come for free as it introduces an error due to inter-stock variability.

The published literature on intra-day seasonality is sparse. Moreover, even the papers that do explore this issue more often than not treat it as a nuisance that needs to be addressed in order to solve a different problem. Apart from [3][4], it is worth mentioning [5], which investigated a very close question to the one we consider here, and reached conclusions similar to ours.

B. VWAP, Price Moves, and Performance Criteria

A key issue in the analysis of VWAP strategies is how to measure their performance. In practice, VWAP traders try to minimize the difference, or shortfall, between the average price of their trade and the market-wide VWAP. Considering that traders can take either side of a trade, the *average* shortfall over a large number of trades is approximately zero for any curve. So, what characterizes a good VWAP strategy? The answer is that good VWAP strategies have a low shortfall *dispersion*—ideally, the shortfall is always near zero. In this document, we compare the performance of various VWAP curves using standard deviation of VWAP shortfall as a measure of dispersion. We justify the use of this metric in the appendix.

In what follows we ignore two effects that have an impact on VWAP shortfall. One is the effect of commissions and fees, which are costs that depend on the size of the order but not on the strategy. Therefore we can safely ignore them for the purpose of comparing strategies.

Secondly, we assume that we are able to capture the market's VWAP over short time intervals, e.g. over a few minutes. In practice, our average execution price will be slightly worse than the market's average price, because we need to pay for liquidity, and the average VWAP shortfall will be slightly positive (i.e. unfavorable). Note, however, that this average is dwarfed by the dispersion: the average may be 1-4 basis points, while the standard deviation is typically about 20 basis points, meaning that

95% of the orders have a VWAP shortfall within $\pm$ 40 basis points. The focus of this study is minimizing this dispersion.

We consider hypothetical full-day VWAP orders, and compute the shortfall relative to daily VWAP that would have been incurred by each order if a given volume pattern had been used. We then aggregate over symbols and dates, and compute the shortfall standard deviation.

We divide the trading day in 26 15-minute bars, plus the open and close auction, for a total of 28 trading periods. According to our assumption above, our order captures the market price over each of those intervals. However, the order will incur a non-zero VWAP shortfall because of the mismatch between the order's VWAP schedule and the realized volume profile for that name on that date.

In our study we examine the performance over the period between January 1, 2011 and December 31, 2011, unless otherwise specified. Generally, we compute the standard deviation of VWAP shortfall using a rolling one-month window. This allows us to examine both the absolute performance of the various algorithms and the temporal behavior of the performance.

C. Scope and Main Conclusions

In this document we examine the performance of static VWAP algorithms. Recall that in a static VWAP algorithm we trade based on a pre-defined schedule, one that tries to match the intra-day seasonality as much as possible. As was mentioned, there are various ways of constructing this schedule, e.g., averaging cross-sectionally, averaging temporally, etc. In this document we examine various ways for constructing the trading schedule curve and the effect these schedules have on the performance.

One could also argue that different groups of stocks behave differently. We examine this issue and we demonstrate that the use of curves tailored to narrower groups mostly does not improve the results in a meaningful way. For example, assume the universe of interest is composed of the largest 500 stocks. We can use a curve constructed specifically for that group. However, using a curve constructed from the largest 1000 stocks instead does not have any meaningful influence on the performance. Another possible grouping is economic sectors, i.e., use different curves for different sectors. This study shows that using sector-specific curves can occasionally lead to better performance. However, this improvement is not persistent. Considering the risks in using sector-specific curves (e.g., sensitivity to outliers), one is better off by trading according to one universal curve. Other groups we consider are ADRs and NYSE- vs. NASDAQ-listed stocks. In both cases we find that special-purpose VWAP curves do not improve performance.

One group of equities that does benefit from a specialized curve is Real Estate Investment Trusts (REITs). These securities are more directly influenced by events in non-equity markets, notably interest rates. It turns out that they trade more heavily at the end of the day, and using a matching VWAP curve improves performance.

We also consider adjusting the VWAP curves with idiosyncratic per-stock components, in particular the relative size of the open and close auctions. This adjustment yields marginal improvement.

Finally, we consider using special curves for days that are known in advance to be special. Here we generally find that special curves do improve performance. First, we examine the optimal strategy during days in which there are scheduled Federal Reserve announcements. During these days we do observe lower shortfalls for a special curve that places more weight at 2.15pm, the time of the announcement. We also examine days in which equity options expire and so-called Triple/Quad Witching days, when other equity-linked derivatives also expire. Finally, we consider the days in which Russell indices are reconstituted. In all these cases we find that special curves are beneficial.

II. AVERAGING WINDOW

We have established in the past that the volume pattern drifts over time, so that one should re-compute it periodically [1]. In practice we re-compute every day or every week, and calculate the average volume profile over a window of n trading days. The question, then, is how to choose n. Generally, shorter averages react more quickly to changes but produce noisier results.

Figure 1 shows the results of a simple experiment in which different averaging lengths were used, ranging from 1 day to 3 months (65 trading days). In all cases we use a universe of the top 500 stocks by market cap for averaging. Previously we had determined that a one-month window provided a good trade-off between quick adaptation to changes and robustness to random fluctuations. As seen in Fig. 1, different curves perform very similarly, except for the one-day window, which is clearly worse. The one-month window (22 trading days) continues to perform well over a wide range of market conditions.

Fig. 2 shows the VWAP curves computed with different averaging windows, as of the end of the experiment (December 31, 2011.) The horizontal axis shows the time of day (in hours). The first and last points are the open

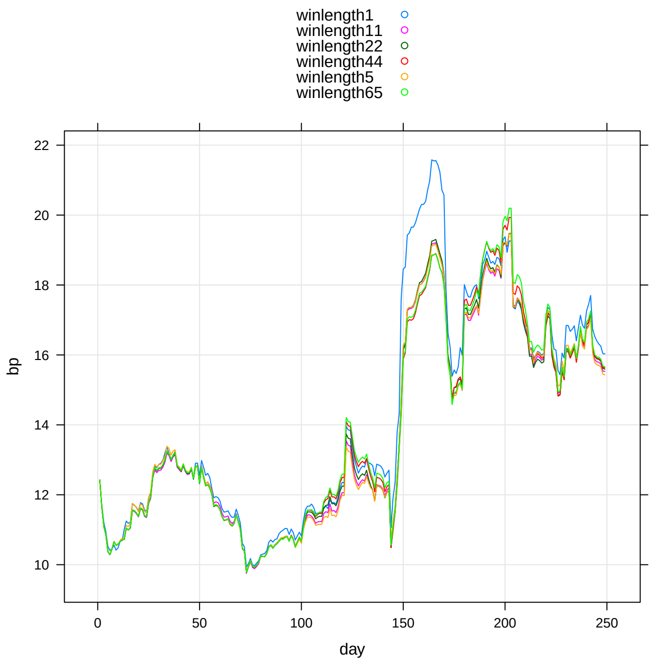

Fig. 1. Standard deviation of VWAP shortfall when the rolling pattern is updated using 1, 5, 11, 22, 44 and 65 trading days.

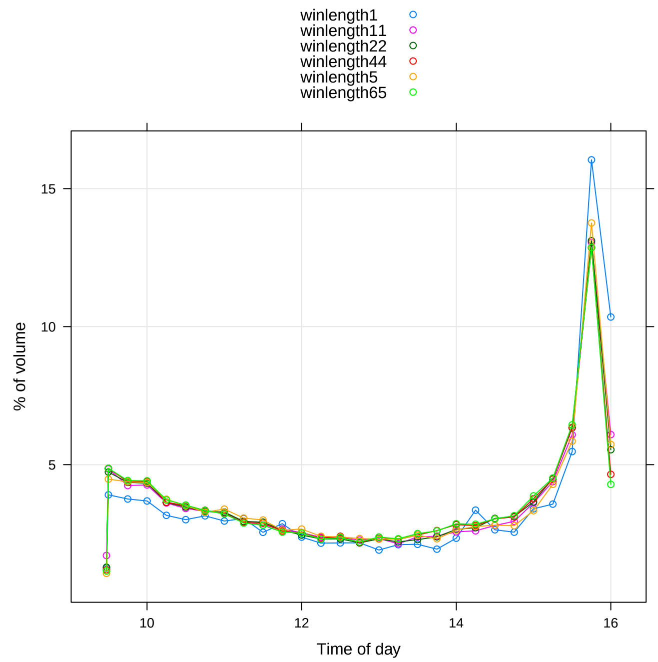

Fig. 2. Volume profile (in %) computed with a rolling pattern updated using 1, 5, 11, 22, 44 and 65 trading days. The horizontal axis is the time of day, in hours; the first and last points are the open and the close.

and close prints, respectively. The vertical axis shows the percentage of the day's volume executed over the corresponding time period. The curves look very similar to each other, which explains why their performance is roughly equivalent. The exception is the one-day curve, which is naturally noisier and performs worse.

We note that the volume profile has shifted over time. As of 2009, around 11% of the day's volume was traded over the last 15 minutes of the day, and the close print was about 4%. At the end of 2011, these figures are closer to 14% and 5%. The beginning of the day has remained stable at 5-6%, so we conclude that some volume has shifted from the middle of the day to the end.

Note that for an n-day window we use equal weights for all n days in the sample. One could consider giving more weight to recent days, for example by using an exponentially-weighted moving average (EWMA). In practice, if we use a 22-day window with more weight on recent days, we will get a curve that is somewhere in between the 22-day curve and 11-day and 5-day curves. As we have seen, all of these curves re very similar to each other, so an EWMA curve would also yield the same performance.

III. Symbol Universes

In general, one can use a single intra-day pattern for all symbols, or use tailor-made curves for specific classes of stocks. Tailor-made curves may yield benefits but they introduce software complexity and they are necessarily "noisier": a 22-day average of the volume profiles of 500 stocks has 500 times more data points than a 22-day average of one stock.

In this section, we investigate the performance of tailor-made curves for various classes of symbols. Specifically we consider:

- Classifying stocks by market cap and using a different curve for each class.
- Classifying stocks by sector and using a different curve for each sector.
- Using a different curve for European ADRs.
- Using a different curve for Real-Estate Investment Trusts.
- Using different curves for NYSE-listed and NASDAQ-listed stocks.

In what follows we define an estimation universe and a simulation universe. The estimation universe is the universe of stocks we use for estimating the average curve. The simulation universe is the universe of stocks we use for testing the performance of the estimated curve. For example, we consider the case when our estimation universe is based on the 100 stocks with the largest market cap, while we examine the performance over a simulation universe constituted by the 3000 stocks with the largest market cap.

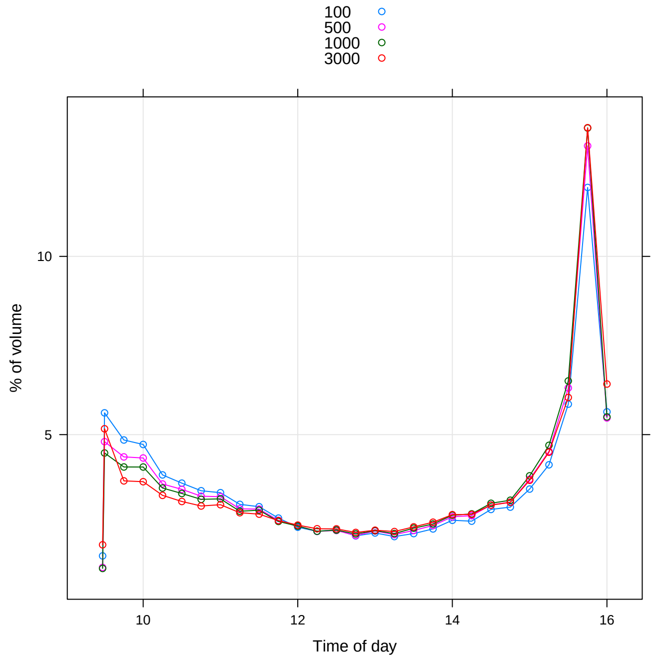

Fig. 3. Volume patterns generated using four overlapping universes of stocks: largest 100, 500, 1000 and 3000 by market cap.

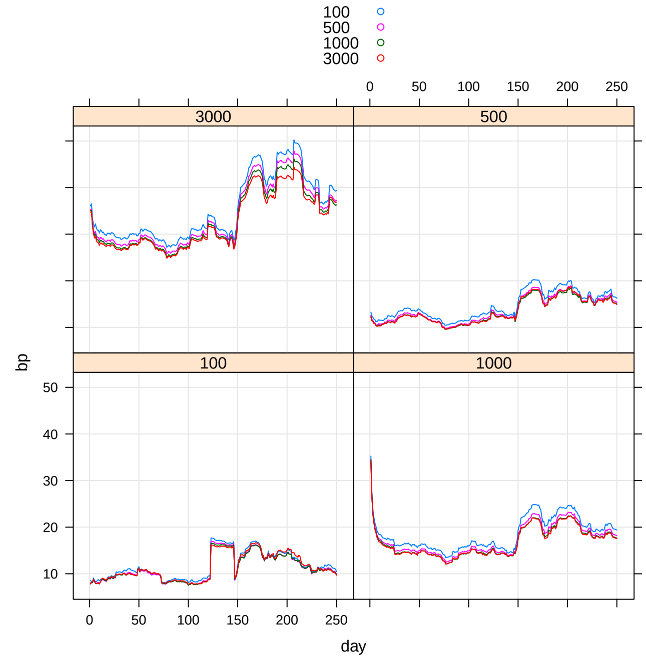

Fig. 4. Standard deviation of VWAP shortfall for the four simulation universes in Fig. 3. For each simulation universe, we show the performance of curves computing each of the four groups.

We assume that we capture the VWAP price within each bar, as explained in Section I-B. Small deviations in the executed price relative to the average market price over time scales of minutes introduce negligible increases in the shortfall standard deviation, which we ignore.

A. Market Cap Effects

In this subsection we explore the effects on performance of dividing the stocks into different market cap groups.

First, we divided the stock population into four *overlapping* classes according to their market cap rank[1]:

Group Name	Market cap ranks	Market cap range
100	1-100	>51 $B
500	1-500	>9.7 $B
1000	1-1000	>3.6 $B
3000	1-3000	>0.5 $B

We used each of these classes as an estimation universe, and then applied the estimated curves to all four classes as simulation universes. In other words, we compute one curve per class but test each curve against all classes. The curves were computed using a 1-month moving-average as in the rolling curve of the previous section. The curves generated are shown in Fig. 3.

We expect the four classes to behave differently: the large stocks are very liquid and have more or less predictable patterns, whereas the smaller stocks are thinly traded and are subject to wide day-to-day variations in volume. Therefore, we expect that the standard deviation of VWAP shortfall will be lowest for the top-100 class and highest for the top-3000 class, regardless of which curve is used.

Fig. 4 shows the performance results. Each panel depicts the shortfall standard deviation for stocks in one of the four classes, and contains results for the four curves under test. Clearly, the differences between the curves are dwarfed by the time variations driven by market volatility. We might expect that each curve would do best when used for stocks in its corresponding class, but this is not the case. For example, the top-100 curve is the worst performer in all cases, including for the top-100 stocks. The top-500 and top-1000 curves perform consistently well.

For our second experiment, we divided the stock population into four *non-overlapping* classes according to their market cap rank[2]:

[1]as of January 1, 2011.

[2]as of January 1, 2011.

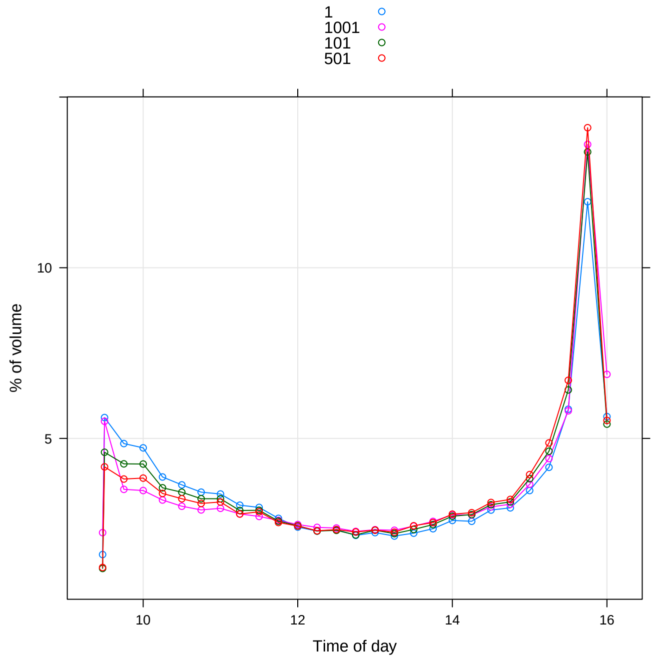

Fig. 5. Volume patterns generated using four non-overlapping universes of stocks by market cap.

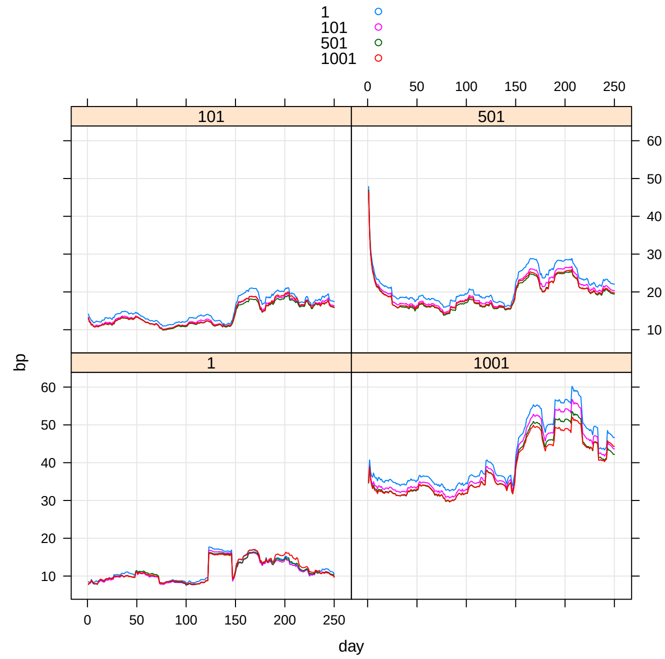

Fig. 6. Standard deviation of VWAP shortfall for the four simulation universes. For each simulation universe, we show the performance of curves computing each of the four groups.

Group Name	Market cap ranks	Market cap range
1	1-100	>51 $B
101	101-500	9.7 - 51 $B
501	501-1000	3.6 - 9.7 $B
1001	1001-3000	0.5 - 3.6 $B

The results are shown in figures 5 and 6. Again we find that the curve computed with fewest stocks (the top 100) performs worst, even for the top-100 group. The 101 and 501 curves perform uniformly well.

In conclusion, there is no advantage in using curves specifically tailored to market cap groups. The common average pattern should be computed using the top 500 or 1000 stocks.

B. Sectors

Another way to divide the universe of stocks in various groups is by economic sector. Here we compute sector-specific VWAP curves by using each of the ten GICS sectors as a separate estimation and simulation universe. We then estimate the overall performance of a sector-specific strategy by pooling the 10 standard deviations obtained with the 10 sector-specific VWAP curves. We compare the pooled standard deviation with the results of using a single moving-average curve for the entire population. In all cases we use a moving average over the last month, and we consider only stocks in the group of the largest 500 by market cap.

The results are depicted in Fig. 7. Sector-specific curves perform very similarly to the common curve, with no persistent advantage or disadvantage. The difference between the two is within $\pm 2\%$. We conclude that using sector-specific curves does not improve performance.

C. European ADRs

In this section we consider European ADRs. Specifically, we consider US-listed ADRs whose corresponding primary security is listed in a European exchange. This universe comprises approximately 90 symbols. Presumably, these ADRs should trade much like the primary security, and therefore should have daily volume patterns strongly influenced by European trading hours. Table I lists the main European exchanges and their trading hours. Most European trading takes place between 3am and 11.30am New York time. Therefore we expect the ADRs to trade mostly between the open and 11.30am.

To build ADR-specific curves we averaged the historic trading pattern of our ADR universe. We constructed three curves, using moving averages over 22, 66 and 130 trading days. For comparison, we also used a standard curve constructed from the largest 500 common stocks, updated daily using a moving-average window of 22 trading days. The 130-day ADR curve and the standard curve are thus constructed using approximately the same number of data points for averaging.

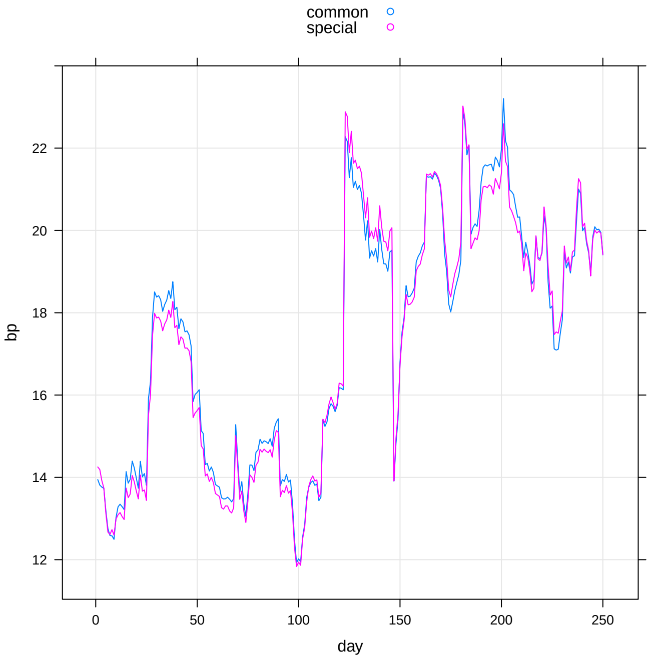

Fig. 7. Standard deviation of shortfall using sector-specific curves and a common rolling curve.

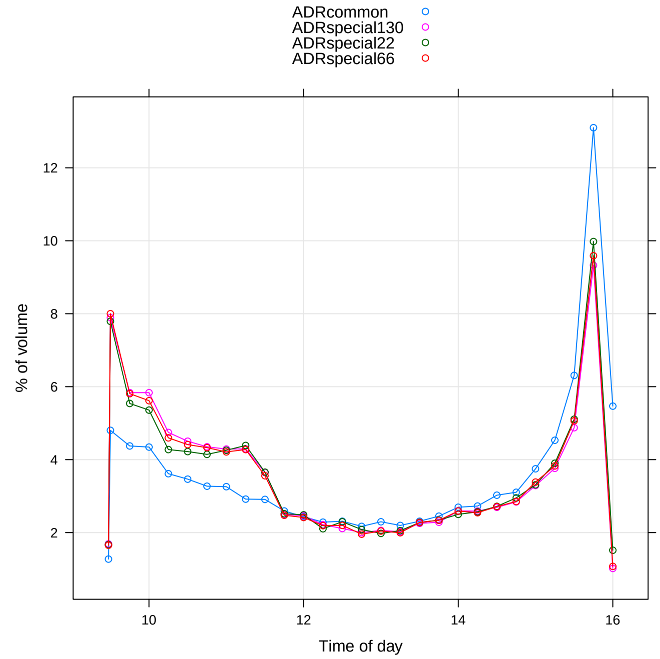

Fig. 8. ADR-specific daily volume patterns (solid lines) with 2σ confidence intervals (dashed lines.)

TABLE I
EUROPEAN EXCHANGE HOURS

Exchange	Local trading hours	Timezone (Europe)	NY hours[a]
LSE	8am - 4.30pm	Western	3 - 11.30am
Euronext (Paris, Brussels, Lisbon, Amsterdam)	9am - 5.30pm	Central	3 - 11.30am
Frankfurt (specialist)	8am - 8pm	Central	3 - 2pm
Xetra	9am - 5.30pm	Central	3 - 11.30am
Chi-X	8am - 4.30pm	Western	3 - 11.30am
BME (Madrid, Barcelona)	9am - 5.30pm	Central	3 - 11.30am

[a]except when daylight-saving time does not coincide

The special ADR curves are depicted in Fig. 8. The ADR curves have relatively higher weights in the mornings and lower in the afternoons and at the close. Specifically, the ADR patterns are higher up to 11.30am - 11.45am. This is consistent with the hypothesis that European ADRs trade like Europe-listed stocks.

We computed the standard deviation of VWAP shortfall for our ADR universe using the four curves described above over 2011. The standard deviation is shown in Fig. 9. The spike near the beginning of the year is due to a single outlier: ARMH had a positive news event and spiked 13% with heavy volume throughout the afternoon.

The special curves perform similarly to each other but differently from the common curve. However, there isn't a persistent advantage or disadvantage. If we ignore the large spike in January, the special curve does have a small advantage, of the order of 5% on average. Yet even this is not persistent: at the end of the year, the common curve outperforms. We conclude that a special-purpose ADR curve is not warranted.

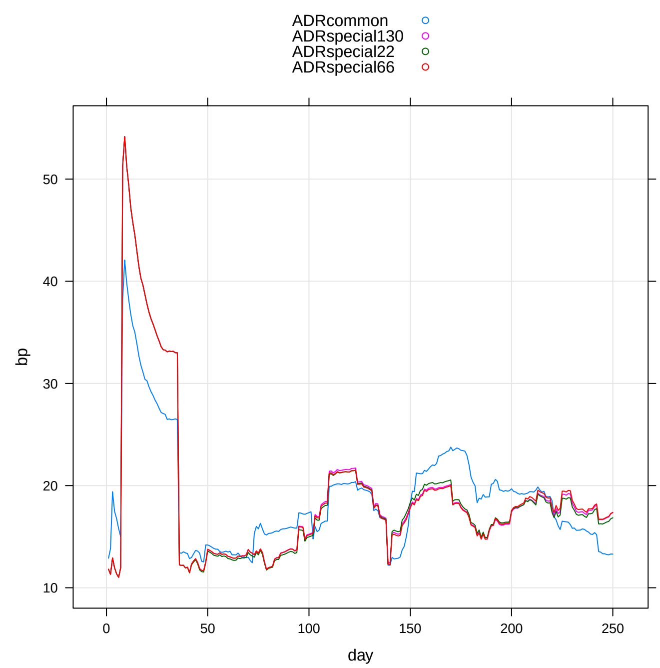

Fig. 9. Standard deviation of VWAP shortfall for European ADRs.

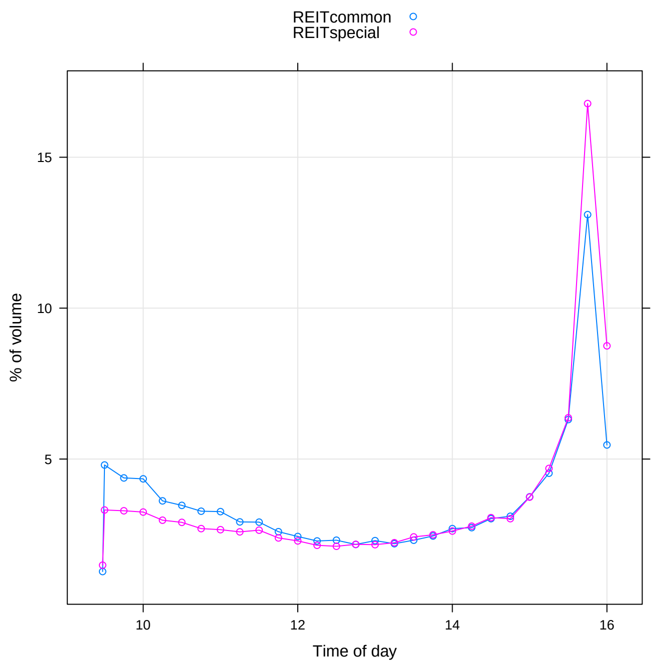

Fig. 10. REIT-specific and common patterns.

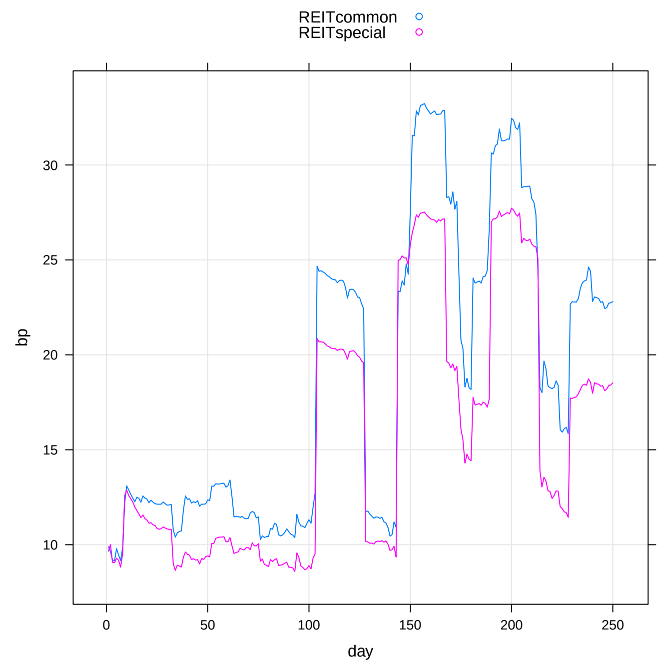

Fig. 11. VWAP shortfall standard deviation for REITs.

D. REITs

Real-Estate Investment Trusts (REITs) are investment vehicles that give investors exposure to real estate. In this sense they could behave like a sector, and perhaps have an idiosyncratic volume pattern.

To build REIT-specific curves, we considered an estimation universe of the largest 100 REITs by market cap. For comparison we used a common curve with the largest 500 common stocks. In both cases we used a 22-day moving average window, updated daily. The simulation universe is the largest 100 REITs in both cases.

Fig. 10 depicts the REIT pattern as of the last day of the simulation, as well as the common pattern. The REIT curve places relatively more weight in the last 15 minutes of the trading day, and less weight in the morning.

Fig. 11 depicts the standard deviation of VWAP shortfall for simulated full-day VWAP tasks. The REIT curve has a small but persistent advantage. On average, using a REIT-specific curve decreases the standard deviation by 10%. Moreover, this improvement is persistent: not only is it visible throughout 2011, but our 2009 study shows the same pattern. We conclude that a special-purpose REIT curve is warranted.

E. Listing Exchange

Another way to divide the stock population is according to primary exchange. In this section we consider NYSE-listed and NASDAQ-listed stocks and test exchange-specific curves.

We consider two universes: the largest 500 NYSE-listed stocks and the largest 500 NASDAQ-listed stocks, by market cap. We use these universes for estimation and simulation. We also consider a common curve built from the largest 500 stocks, in general.

Fig. 12 depicts the 22-day rolling standard deviation for full-day VWAP tasks. The left and right panels show NASDAQ- and NYSE-listed stocks respectively. The blue lines correspond to the common curve, and the red lines to the special-purpose curve.

It is obvious the NYSE stocks have lower standard deviations than NASDAQ stocks, indicating more-stable trading patterns. However in both cases the special curves make virtually no difference. The explanation can be seen in Fig. 13, which depicts the volume profiles. There is very little difference between the NYSE-specific curve, the NASDAQ-specific curve, and the common curve.

IV. Special Dates

On some dates, it is known ahead of time that an event will occur that will affect the trading patterns of market participants. In this section we investigate whether one should construct special-purpose VWAP curves for these specific dates.

A. Scheduled Fed days

Days on which it is known that the Federal Reserve will make an interest rate announcement are of particular

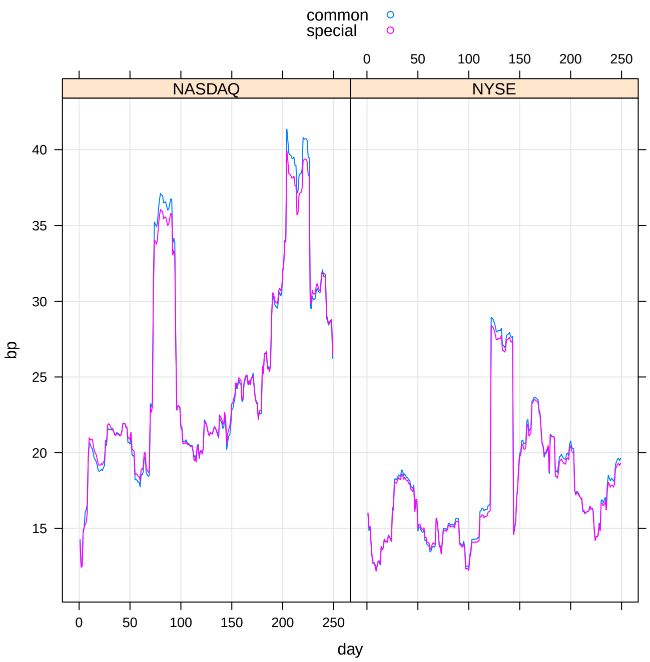

Fig. 12. Rolling shortfall standard deviation from NYSE-listed and NASDAQ-listed stocks, using special-purpose and common curves. The two universes are clearly different, but special curves provide no benefit.

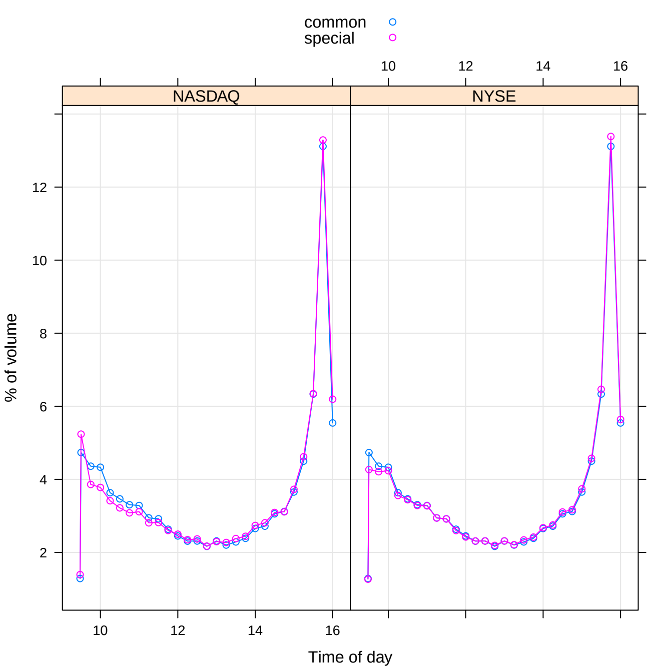

Fig. 13. NASDAQ-specific, NYSE-specific and common VWAP curves.

interest, since much volume is driven by the announcement at 2.15pm. Thus one might consider using a special curve for those days.

A major drawback of this idea is the relative scarcity of data: the FOMC meets regularly only 8 times a year. In order to obtain more-stable estimates, we used data for all *scheduled* Fed days between January 2007 and May 2012. We exclude days in which there was a scheduled press conference by the Fed chairman at 12.30, namely 2011-04-27, 2011-06-22, 2011-11-02, 2012-01-25 and 2012-04-25. We further eliminate 3 dates in 2009 and 2010 because of data issues. This leaves us with 35 sample dates.

We simulated the effect of using a specially-designed Fed-day VWAP curve for those days, versus using a standard VWAP curve. For each of the 35 scheduled Fed days in our sample, we computed the shortfalls obtained by trading the 1000 largest stocks with those curves. The standard deviation of the computed shortfalls is shown in Fig. 14.

For each date in the sample, the Fed-specific curve was fit using all the *other* Fed dates. This prevents in-sample bias, but it also means that a different curve was used for each of the dates shown. The common curve was updated using the top 1000 stocks over a moving 22-day window.

Fed-specific curves do better on 32 out of 35 days in the sample. The average improvement is about 10%.

Figure 15 shows the average Fed-specific pattern computed from our sample of 34 Fed days. The volume spike at 2.15pm is clearly visible.

We conclude that a special-purpose Fed day curve improves performance.

Beginning in 2011, the Fed conducts scheduled press conferences on some, but not all, of their meeting days. The press conference takes place at 12.30pm, and the rate decision is announced then. On these days, the volume spike does in fact happen at 12.30 and not at 2.15pm, so we use a special "Fed press conference" curve.

B. *Option Expiration and Triple/Quad Witching Days*

The third Friday of each month is the last trading day before the expiration of equity options. Therefore we may expect unusual trading activity and perhaps a special volume pattern. To test this theory, we compared the performance of a common rolling curve and special-purpose curves for these days.

Additionally, we consider Triple (or Quadruple) Witching Days, when equity options, equity futures, index futures and options on index futures expire concurrently. This happens every three months, in March,

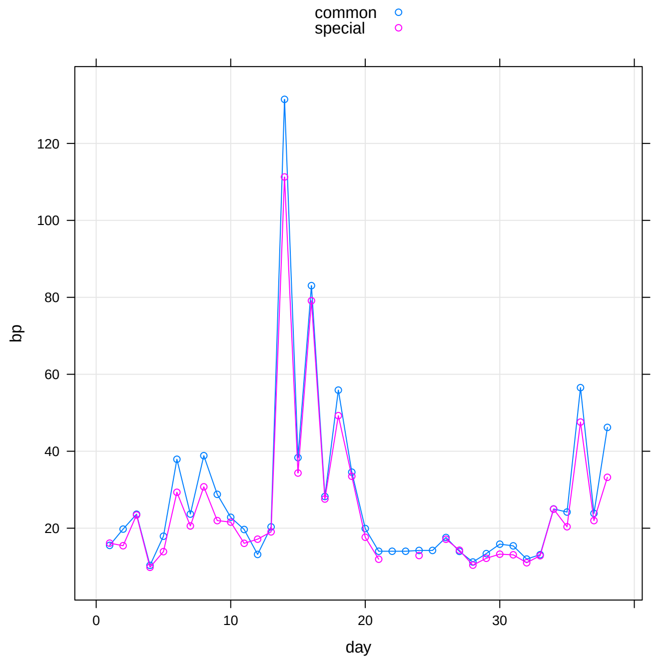

Fig. 14. Standard deviation of shortfall for scheduled Fed days, 2007-12. The red line was obtained by averaging the pattern for Fed days only. The blue line results from using a moving-average pattern.

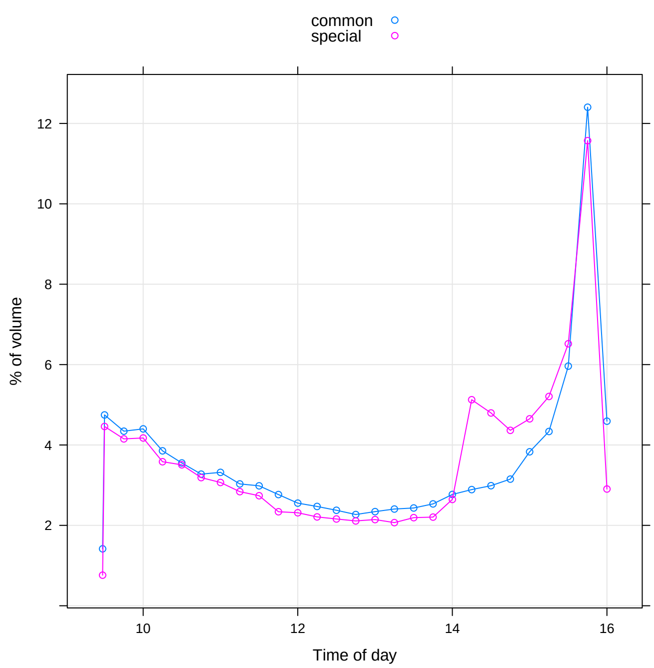

Fig. 15. Average realized volume pattern for scheduled Fed days, 2007-2012. The Fed-day curve does not exhibit the typical U pattern of normal days.

June, September and December. Also on these days, at the close of business, certain accumulated changes to S&P equity indices go into effect. Volume tends to be very high on triple-witching days.

In our experiment, we consider the dates corresponding to the third Friday of each month between January 1, 2007 and March 31, 2012[3]. For each date, we built a VWAP curve by averaging volume patterns for all *other* dates, to prevent in-sample bias. Then we computed VWAP shortfalls for the top 1,000 stocks, using this special-purpose curve as well as a common rolling curve.

The results are plotted in Fig. 16. The standard deviation is shown in red for the special-purpose curve and in blue for the common curve. Fig. 17 shows the corresponding volume patterns. Note that considerably more volume than normal is traded at the open, in the first 15 minutes of trading, and at the close. The performance of the special curve is *better* than that of the common curves. This is a new phenomenon: previous studies were inconclusive. Fig. 18 confirms this result. The X-axis shows the improvement of the special curve over the common curve (positive means improvement), and Y-axis shows the percentage of days with such an improvement (or less). It can be seen that 90% of the time the special curve does better. The average improvement is about 15%. We conclude that a special option expiration curve improves performance.

Fig. 19 compares the performance of a common and a special curve for triple/quad witching days. The special purpose curve is better in 20 out of 21 days in our sample. The average improvement in standard deviation is about 15%.

Figure 20 depicts the common rolling curve as of December 31, 2011 (in blue) and the special triple-witch curve (in red). The special curve has the same qualitative shape as the option expiration curve, but with even more weight at the open and close auctions. We conclude that a special triple-witch curve does improve VWAP performance.

C. *Russell Reconstitution Days*

The Russell indices are market capitalization weighted broad-based equity indices. The list of stocks that make up the indices is updated once a year in June. The changes go into effect after the close on reconstitution date. Many mutual funds and ETFs track these indices, so they update their holdings accordingly. Typically on reconstitution day trading volumes are larger than usual, especially towards the close. Therefore we may expect

[3]except March 21, 2008, which was a holiday.

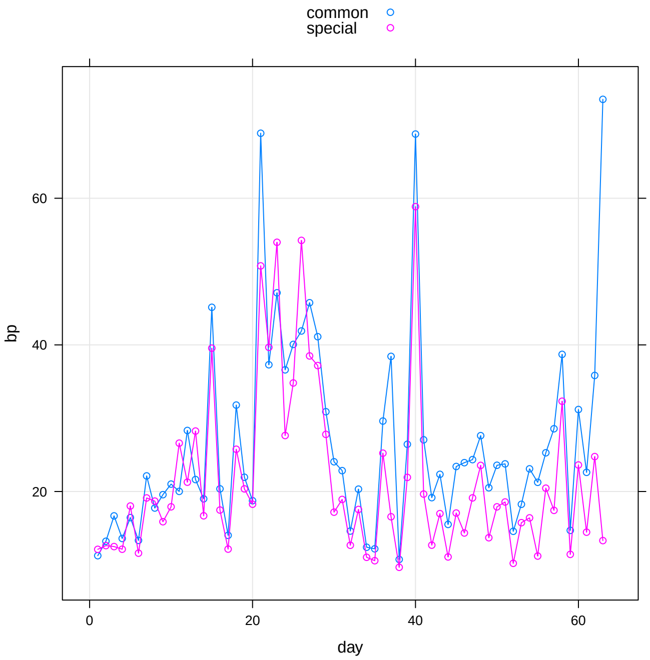

Fig. 16. Standard deviation of VWAP shortfall for option expiration days. The blue points correspond to a common rolling curve. The red points were obtained with an average pattern for all other option expiration days.

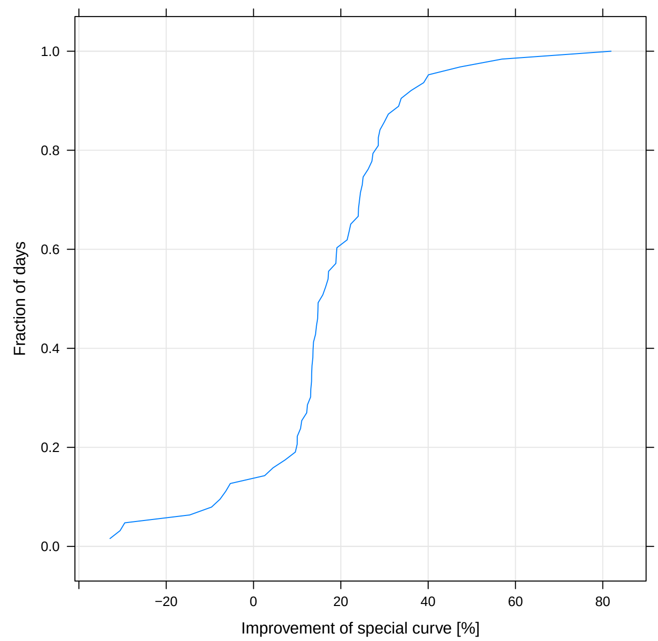

Fig. 18. Improvement of the special curve over the common curve for option expiration days. The X-axis shows the magnitude of the improvement, in percent. The Y-axis shows the fraction of days with such an improvement, or less.

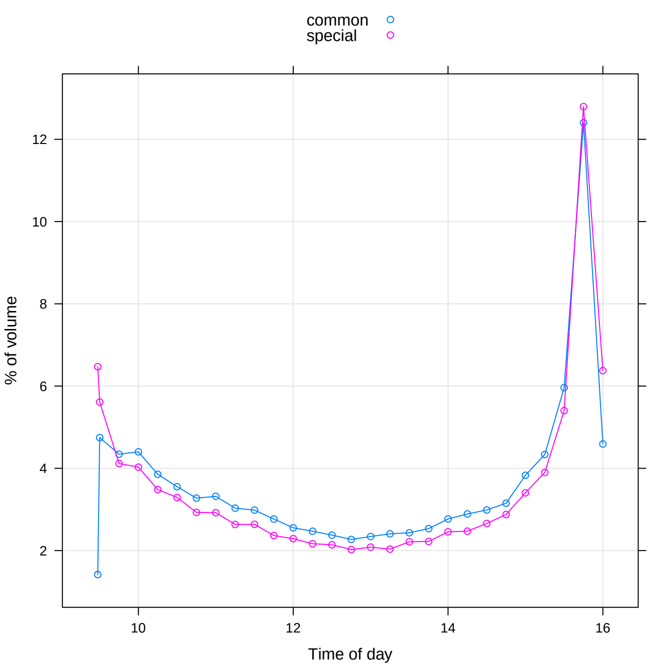

Fig. 17. Common pattern and special-purpose option expiration pattern.

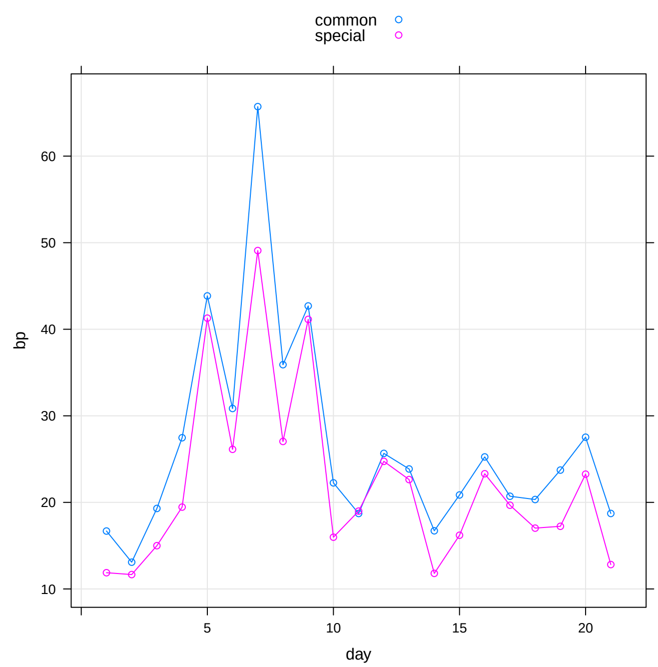

Fig. 19. Standard deviation of VWAP shortfall for triple-witching days. The special-purpose curve achieves a 15% improvement.

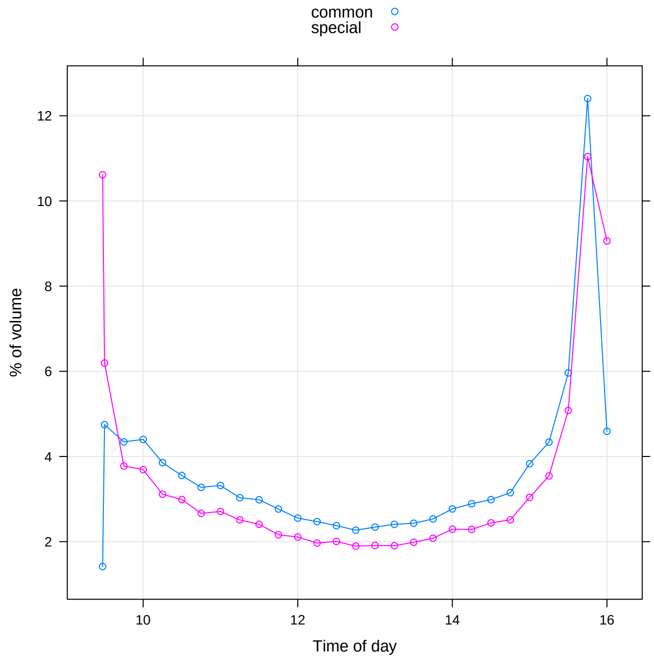

Fig. 20. Common pattern and special-purpose triple-witching-day pattern.

a special volume profile, and a special-purpose VWAP curve may be useful.

Fig. 21 depicts a special-purpose curve, computed using the volume profiles of the top 1000 stocks by market cap, on the 2007-2011 Russell reconstitution days. Note that, since this event occurs only once a year, we have only 5 sample days, possibly not enough to obtain a stable pattern. Nevertheless, the special curve looks as expected, with relatively low volume during the day and a very large close print.

For simulation we proceeded as in previous sections. For each Russell day in our sample, we computed an average curve using the other four Russell days, to avoid in-sample bias. We compared the performance of the special curve to that of a common moving-average curve computed over a 22-day rolling window using the largest 1000 stocks. The results are shown in Fig. 22. The special curve performs better on every day of our sample, with an average improvement of about 6%. We conclude that a special curve is beneficial.

V. Stock-Specific MOO and MOC

A possible scenario is that some parts of the daily volume profile are common to all stocks, while other parts are more idiosyncratic. For example, it has been noted that the open and close auctions are very different for different names. A hybrid curve-building approach is to use a common curve for the continuous-trading

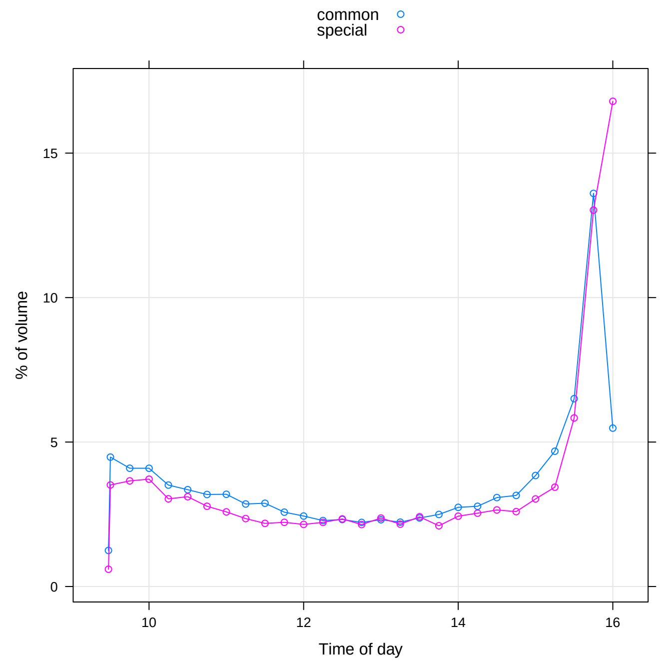

Fig. 21. Common pattern and special-purpose Russell pattern.

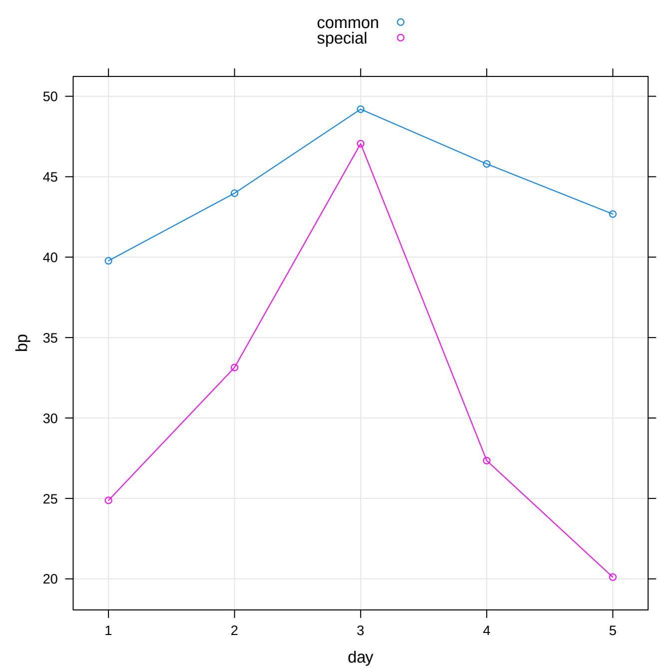

Fig. 22. Standard deviation of VWAP shortfall for Russell days. The special-purpose curve achieves a 6% improvement.

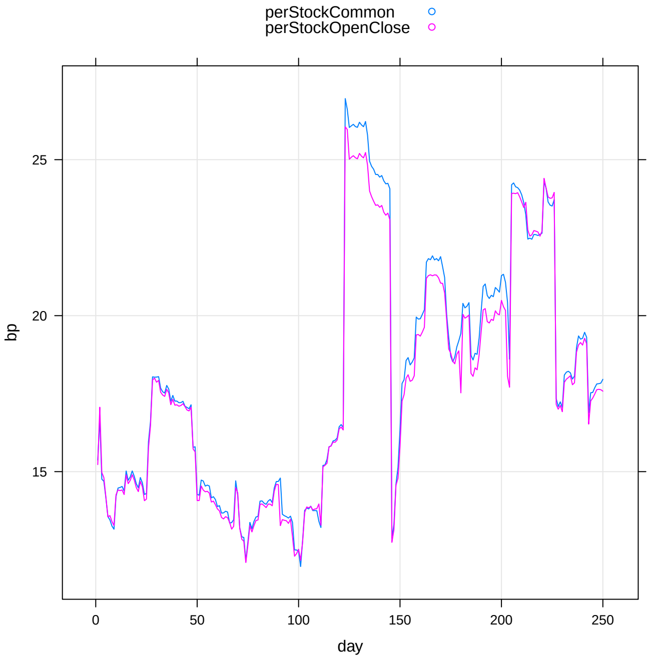

Fig. 23. One-month rolling standard deviation for stock-specific open and close and a common curve for the rest of the day.

portion of the day and specific percentages for the open and close auctions.

In our first experiment, we considered a simulation universe of the largest 500 stocks by market cap. This universe was determined at the beginning of the simulation (January 2011), and kept fixed throughout the simulation. This is in contrast to other experiments, in which the universe was updated daily according to market cap changes. In this section we consider per-stock curves, so we need to keep a fixed universe (in practice the universe shrinks from 500 to 491 by the end of the year because of corporate actions.)

The curves were constructed as follows. For the continuous-trading bars we used a rolling common curve, constructed from all 500 stocks. We also computed a separate moving-average open and close relative volume for each of the 500 simulated stocks. Thus for each stock we used an idiosyncratic open and close, and the common curve for the rest of the day (scaled to account for the idiosyncratic components). We compare the performance of these curves with that of a full-day rolling common curve constructed from the same 500 largest stocks. Fig. 23 shows the resulting VWAP shortfall standard deviation. The curve with per-stock MOO and MOC performs better 90% of the days (especially in high-volatility periods) but the average improvement is very small (1-2% on average).

In our second experiment, we refined the simulation by splitting our universe into market cap brackets. This is motivated by the fact that small-cap stocks have higher average volumes executed at the open and close prints. We considered four *non-overlapping* classes according to their market cap rank[4]:

Group Name	Market cap ranks	Market cap range
0	1 - 500	> 9.7 $B
500	501 - 1000	3.6 - 9.7 $B
1000	1001-2000	1.1 - 3.6 $B
2000	2001-3000	0.5 - 1.1 $B

In addition, we considered *partial-day VWAP orders*, i.e. VWAP orders that span only a fraction of the day. We divide the day into 28 trading periods, or *bars*: the open auction, 26 15-minute intervals of continuous trading, and the close auction. We considered VWAP orders that include the open auction plus the first few continuous-trading bars, and orders that start near the end of the day and include the close auction. Partial-day VWAP orders may better highlight the difference between using different values for the open and close percentages.

Selected results are shown in figures 24-27. We compute the improvement of using a curve with stock-specific open and close volumes, relative to using a common curve. In all these figures, the X-axis is the improvement in percentage points, and the Y-axis is the fraction of days having up to that improvement. The open auction is bar #1, the 9.30-9.45 am interval is bar #2, the 3.45-4pm interval is bar #27, and the close auction is bar #28.

Fig. 24 refers to bracket 1000, and shows the improvement for partial-day orders that begin with the open auction and run through bars 2, 3, 4 and 28. The latter corresponds to a full-day VWAP order. The blue curve, corresponding to an order that ends at 9.45am, shows that most of the time there is an improvement of 3-7% if one uses a special market-on-open volume percentage for each stock. The green and red lines correspond to orders that end at 10am and 10.15am, and show that using a special open percentage actually hurts performance by a few percentage points. The full-day simulation shows a small improvement of 1-6 %.

Fig. 25 shows corresponding results for bracket 2000. The improvement for orders that end on bar #2 is larger, typically between 0 and 13%. However, orders that end in bars #3 and #4 do worse. Full day-orders show virtually no improvement.

Fig. 26 again refers to bracket 1000. Here we consider orders that end at the close and start at the open (bar #1), 3.15pm (bar #25), 3.30pm (bar #26) and 3.45pm (bar #27). Full-day orders and orders that start at 3.15pm

[4]as of January 1, 2011.

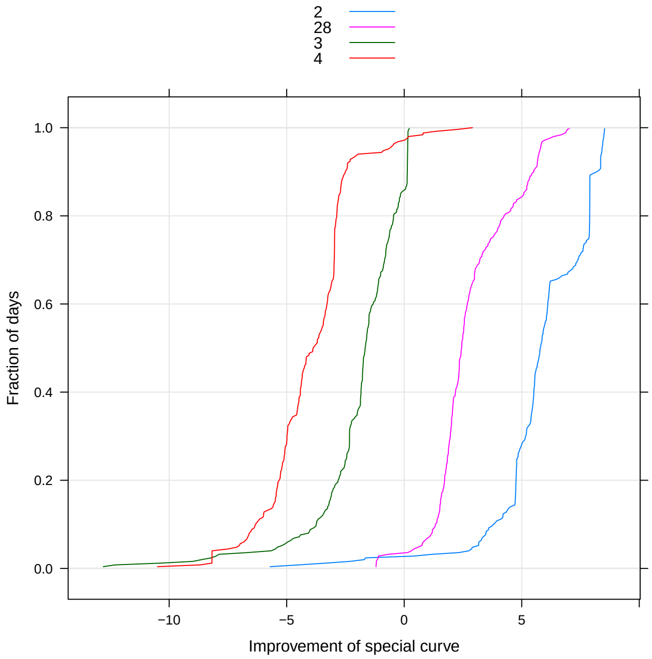

Fig. 24. Improvement of using a curve with stock-specific open and close volumes, relative to using a common curve, for bracket 1000 and orders that start at the open and end on bars #2, #3, #4 and #28 (full day). The X-axis is the improvement in percentage, and the Y-axis is the fraction of days having up to that improvement.

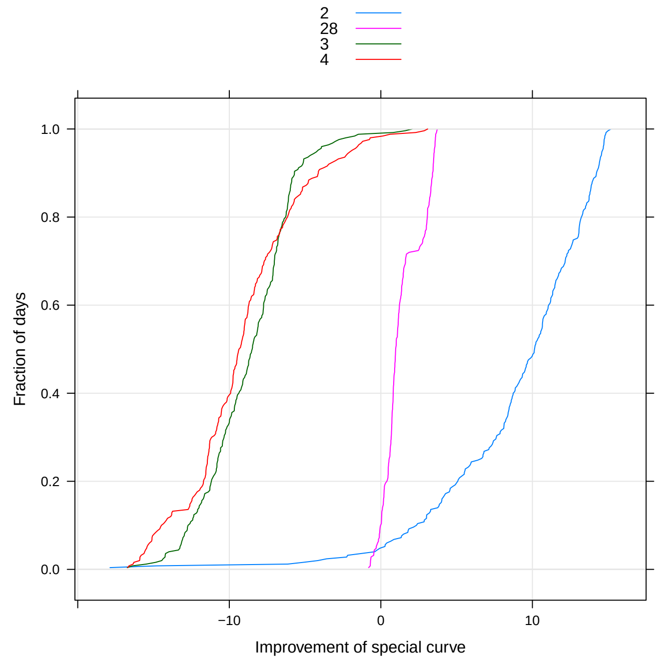

Fig. 25. Improvement of using a curve with stock-specific open and close volumes, relative to using a common curve, for bracket 2000 and orders that start at the open and end on bars #2, #3, #4 and #28 (full day). The X-axis is the improvement in percentage, and the Y-axis is the fraction of days having up to that improvement.

and 3.30pm show some improvement, typically between 2% worse to 7% better. Orders that start at 3.45pm do slightly worse on average.

Fig. 27 shows corresponding results for bracket 2000. Here we see a more marked improvement (2% worse to 13% better) for orders that start near the end of the day. However, Fig. 28 shows the same results for the first half of 2011 only. It can be seen that the improvement is gone.

In conclusion, using stock-specific open and close volumes seems to bring some improvement, but it is small, it is visible only for small caps and partial-day orders, and is not persistent in time. We do not believe that this warrants the software complexity of keeping track of stock-specific open and close volumes.

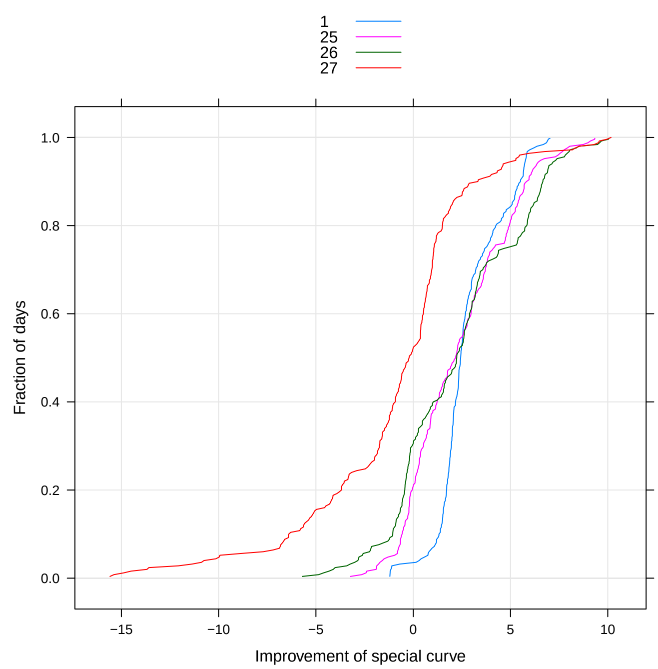

Fig. 26. Improvement of using a curve with stock-specific open and close volumes, relative to using a common curve, for bracket 1000 and orders that end at the close and start on bars #1 (full day), #25, #26 and #27. The X-axis is the improvement in percentage, and the Y-axis is the fraction of days having up to that improvement.

VI. Summary and Conclusions

In this study we examined the performance of various VWAP strategies. The performance of a VWAP trade is measured by the shortfall between the executed average price and the market-wide realized VWAP. We showed that the average shortfall over many trades is zero, and the relevant performance metric is the shortfall dispersion. In this study we measured the dispersion as the shortfalls' standard deviation.

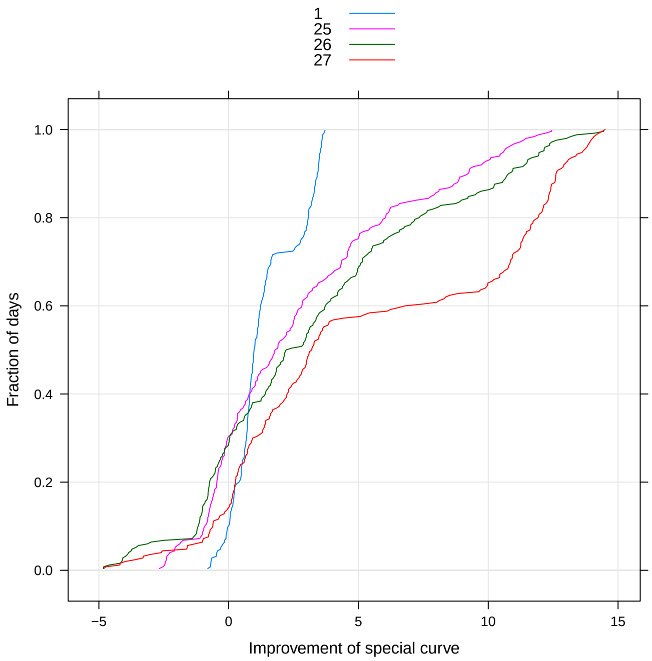

Fig. 27. Improvement of using a curve with stock-specific open and close volumes, relative to using a common curve, for bracket 2000 and orders that end at the close and start on bars #1 (full day), #25, #26 and #27. The X-axis is the improvement in percentage, and the Y-axis is the fraction of days having up to that improvement.

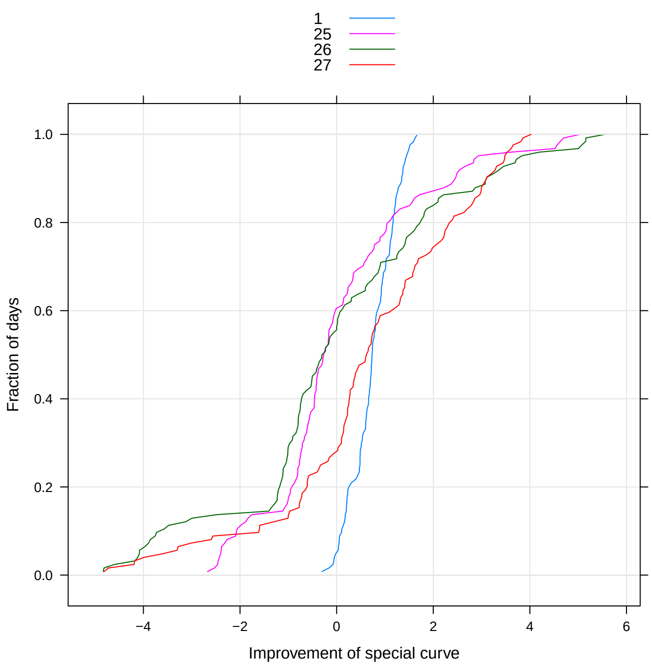

Fig. 28. Improvement of using a curve with stock-specific open and close volumes, relative to using a common curve, for bracket 2000 and orders that end at the close and start on bars #1 (full day), #25, #26 and #27. This includes simulation only the first half of 2011.

We estimated the various algorithms' performance by simulation using historical 2011 volumes and prices, and compared VWAP curves constructed using several combinations of estimation and simulation stock universes. In particular we examined the following cases:

- Curves constructed by stock market cap, sector and exchange.
- Special curves for REITs and ADRs.
- Curves constructed specifically for Fed announcement days, option expiration days, triple-witching days and Russell reconstitution days.
- Per-stock open and close auction percentages, with a common curve for continuous trading.

In general, customizing curves for shorter time spans or smaller stock universes introduces a fundamental trade-off. Curves for specific dates/stocks *decrease* dispersion because they apply to a more-homogeneous group. But since less data is available, there is less opportunity for averaging out any idiosyncratic errors and this may *increase* dispersion.

From the simulation results we can conclude that

- *A window of about 1 month (22 trading days) should be used for averaging.*
- *Slicing stocks by market cap provides no benefit* beyond a minimum size necessary to ensure stable estimation. Any future curves should be fit with an estimation universe of 500-1000 stocks for best performance.
- *Slicing stocks by sector provides no benefit*. Sector-specific curves may perform better or worse than a common curve.
- *Slicing stocks by exchange provides no benefit*. However, NASDAQ-listed stocks can be expected to have more volatile volume profiles than NYSE-listed stocks.
- *Special-purpose curves for Fed dates generally reduce standard deviation* by 10%. If a special curve is used, care must be taken to include as many Fed dates as possible.
- *Special-purpose curves for option expiration days reduce standard deviation* by 15%. The special curve places more weight at the open, the first 15 minutes, and the close, and less weight during the rest of the day.
- *Special-purpose curves for triple-witching days reduce standard deviation* by 15%. The special curve places more even more weight at the open.
- *Special-purpose curves for REITs decrease standard deviation by 10%*. The special curve places more weight on the last 15 minutes of trading and the close, and less weight in the morning.

- *Special-purpose curves for European ADRs provide no benefit*, even though they do place more weight in the morning, as one would expect from European hours trading.
- *Special per-stock open and close auction volumes provide negligible benefit*, not enough to justify the additional software complexity.

For reference,the standard deviation of VWAP shortfall is typically about 20 basis points, so that 95% of the time an order has a VWAP shortfall within ±40 basis points. A 10% improvement will reduce this range to ±36 basis points. This gives an estimate of the level of improvement that we can expect from using special curves.

References

[1] "Static VWAP: A Comparative Analysis". Pragma Securities, August 2009.
[2] "Static VWAP: An Update". Pragma Securities, November 2009.
[3] R. Engle and J. Russel, "Autoregressive Conditional Duration: A New Model for Irregularly Spaced Transaction Data," Econometrica 66, 1127-1162, 1998.
[4] J. McCulloch, "Relative Volume as a Doubly Stochastic Binomial Point Process," Research Paper 146, University of Technology Sidney, October 2005.
[5] D. Hudson, "VWAP and Volume Profiles," Journal of Trading, 1(1) 38-42, 2006.

Appendix A
Why we focus on standard deviation

Below we give a mathematical argument of why standard deviation of shortfall is the most relevant metric to quantify the performance of a VWAP strategy.

First, we note that the performance of a VWAP strategy is relatively independent of market moves. Define by $\bar{p}$ the VWAP prevailing in the market during the life of a particular trade. The shortfall between the executed price and the market's VWAP can be approximated using the following formula,

$$\mathrm{sf} = \frac{10,000}{\bar{p}} \sum_{i=0}^{N-1} (w_i - v_i) p_i \qquad \mathrm{bp}, \tag{1}$$

where N is the number of time periods, T is the life of the trade, w_i is the fraction of the daily market volume traded during the period i-th period, i.e. $[t_0 + i\frac{T}{N}, t_0 + (i+1)\frac{T}{N}]$, v_i is the fraction of the traded shares executed during the i-th period, and p_i is the prevailing market price during the i-th period. This formula tells us that if the fraction of the order traded during a certain period equals the fraction of the total volume traded during that period, the shortfall is going to be zero *independently* of the price moves. If one knew ex-ante how the volume pattern is going to look like at the end of the day, one could achieve a near-zero VWAP shortfall by trading according to that pattern, regardless of price volatility. It is the inevitable deviations between w_i and v_i that result in the VWAP shortfall.

Second, the average shortfall over a large number of trades is zero. Assume that one trades all the shares at the beginning of the trade, i.e. $v_0 = 1$ while $v_1 = \ldots = v_{N-1} = 0$. In this case, for a sell (buy) trade the shortfall is[5] $\frac{\bar{p}-p_0}{\bar{p}}$ ($\frac{p_0-\bar{p}}{\bar{p}}$). Note that the buy and sell shortfalls offset each other. Therefore, if we assume that the side of our trade is random, the average shortfall is zero. This argument can be extended to show that the average shortfall is zero, for any curve $v_0, v_1, \ldots v_{N-1}$ used for trading.

Although the VWAP shortfall is zero on average, the dispersion of shortfall values does have an impact on the profitability of a trading strategy. Consider a trading strategy that has a theoretical Sharpe ratio of r if executed precisely at market VWAP. In reality, some practical VWAP algorithm is used, and a non-zero shortfall is incurred. Since the average shortfall is zero, the average return of the strategy remains the same. But the standard deviation of the shortfall increases the denominator of the Sharpe ratio. For example, if the standard deviation of the shortfall is approximately equal to the volatility of the strategy per bet, the strategy's actual Sharpe ratio is going be about 70% of r. This example and the fact that no VWAP algorithm can systematically outperform the market demonstrates that the trader/algorithm objective should be to minimize the deviation from the market VWAP independent of their side. The most convenient measure of dispersion is the standard deviation of the algorithm's shortfall.

—

For questions or comments, please email Dr. Eran Fishler, Director of Research (technotes@pragmatrading.com).

[5] We use the sign convention that positive shortfalls are unfavorable.