

Beware of Low Frequency Data

By Ernie Chan, Ph.D. (www.epchan.com)

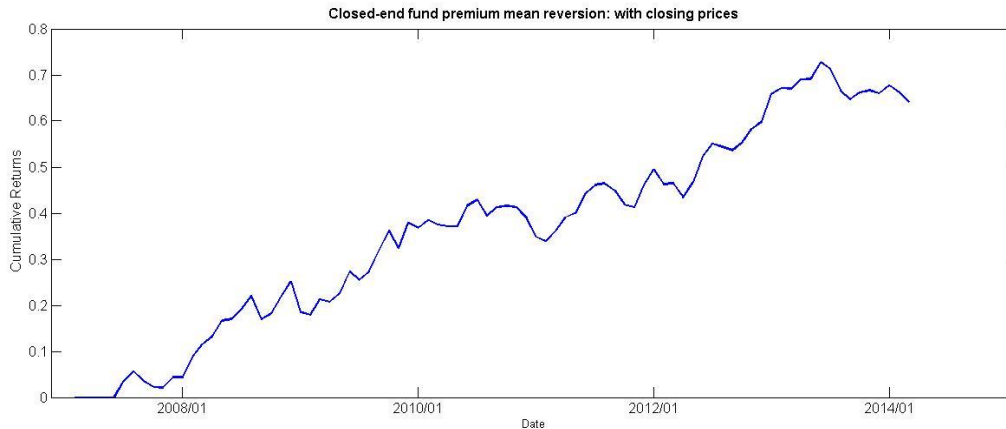
(This post is based on the talk of the same title I gave at Quantopian's NYC [conference](#) which commenced at 3.14.15 9:26:54. Do these numbers remind you of something?)

A correct backtest of a trading strategy requires accurate historical data. This isn't controversial. Historical data that is full of errors will generate fictitious profits for mean-reverting strategies, since noise in prices is mean-reverting. However, what is lesser known is how perfectly accurate capture of historical prices, if done in a sub-optimal way, can still lead to dangerously inflated backtest results. I will illustrate this with three simple strategies.

CEF Premum Reversion

Patro et al published a [paper](#) on trading the mean reversion of closed-end funds' (CEF) premium. Based on rational analysis, the market value of a CEF should be the same as the net asset value (NAV) of its holdings. So the strategy to exploit any differences is both reasonable and simple: rank all the CEF's by their % difference ("premium") between market value and NAV, and short the quintile with the highest premium and buy the quintile with the lowest (maybe negative) premium. Hold them for a month, and repeat. (You can try this on a daily basis too, since Bloomberg provides daily NAV data.) The Sharpe ratio of this strategy from 1998-2011 is 1.5. Transaction costs are ignored, but shouldn't be significant for a monthly rebalance strategy.

The authors are irreproachable for their use of high quality price data provided by [CRSP](#) and monthly fund NAV data from Bloomberg for their backtest. So I was quite confident that I can reproduce their results with the same data from CRSP, and with historical NAV data from Compustat instead. Indeed, here is the cumulative returns chart from my own backtest (click to enlarge):

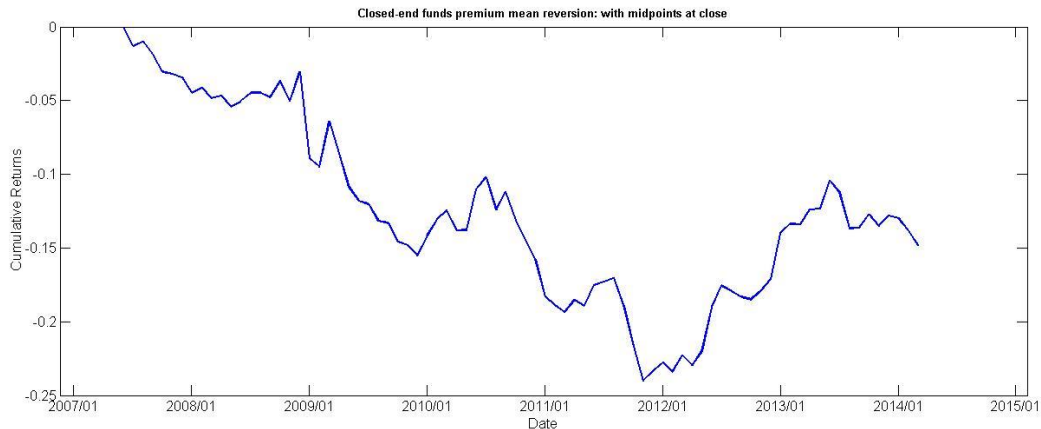


However, I also know that there is one detail that many traders and academic researchers neglect when they backtest daily strategies for stocks, ETFs, or CEFs. They often use the "consolidated" closing price as the execution price, instead of the "official" (also called "auction" or "primary") closing price. To understand the difference, one has to remember that the US stock market is a network of over 60 "market centers" (see the [teaching notes](#) of Prof. Joel Hasbrouck for an excellent review of the US stock market structure). The exact price at which one's order will be executed is highly dependent on the exact market center to which it has been routed. A natural way to execute this CEF strategy is to send a market-on-close (MOC) or limit-on-close (LOC) order near the close, since this is the way we can participate in the closing auction and avoid paying the bid-ask spread. Such orders will be routed to the primary exchange for each stock, ETF, or CEF, and the price it is filled at will be the official/auction/primary price at that exchange. On the other hand, the price that most free data service (such as Yahoo Finance) provides is the consolidated price, which is merely that of the last transaction received by the Securities Information Processor (SIP) from any one of these market centers on or before 4pm ET. There is no reason to believe that one's order will be routed to that particular market center and was executed at that price at all. Unfortunately, the CEF strategy was tested on this consolidated price. So I decide to backtest it again with the official closing price.

Where can we find historical official closing price? Bloomberg provides that, but it is an expensive subscription. CRSP data has conveniently included the last bid and ask that can be used to compute the mid price at 4pm which is a good estimate of the official closing price. This mid price is what I used for a revised backtest. But the CRSP data also doesn't come cheap - I only used it because my academic affiliation allowed me free access. There is, however, an unexpected source that does provide the official closing price at a reasonable rate: [QuantGo.com](#) will rent us tick data that has a Cross flag for the closing auction trade. How ironic: the cheapest way to properly

backtest a strategy that trades only once a month requires tick data time-stamped at 1 millisecond, with special tags for each trade.

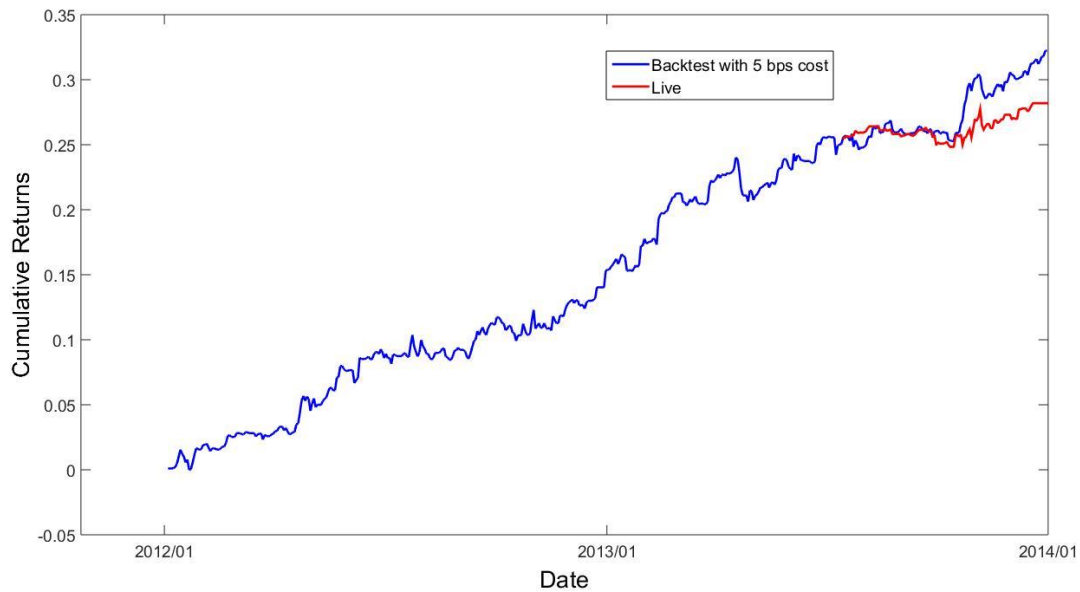
So what is the cumulative returns using the mid price for our backtest?



Opening Gap Reversion

Readers of my [book](#) will be familiar with this strategy (Example 4.1): start with the SPX universe, buy the 10 stocks that gapped down most at the open, and short the 10 that gapped up most. Liquidate everything at the close. We can apply various technical or fundamental filters to make this strategy more robust, but the essential driver of the returns is mean-reversion of the overnight gap (i.e. reversion of the return from the previous close to today's open).

We have backtested this strategy using the closing mid price as I recommended above, and including a further 5 bps transaction cost each for the entry and exit trade. The backtest looked wonderful, so we traded it live. Here is the comparison of the backtest vs live cumulative P&L:



Yes, it is still mildly profitable, but nowhere near the profitability of the backtest, or more precisely, walk-forward test. What went wrong? Two things:

- Just like the closing price, we should have used the official/auction/primary open price. Unfortunately CRSP does not provide the opening bid-ask, so we couldn't have estimated the open price from the mid price. QuantGo, though, does provide a Cross flag for the opening auction trade as well.
- To generate the limit on open (LOO) or market on open (MOO) orders suitable for executing this strategy, we need to submit the order using the pre-market quotes before 9:28am ET, based on Nasdaq's rules.

Once again, a strategy that is seemingly low frequency, with just an entry at the open and an exit at the close, actually requires TAQ (ticks and quotes) data to backtest properly.

Futures Momentum

Lest you think that this requirement for TAQ data for backtesting only applies to mean reversion strategies, we can consider the following futures momentum strategy that can be applied to the gasoline (RB), gold (GC), or various other contracts trading on the NYMEX.

At the end of a trading session (defined as the previous day's open outcry close to today's open outcry close), rank all the trades or quotes in that session. We buy a

contract in the next session if the last price is above the 95th percentile, sell it if it drops below the 60th (this serves as a stop loss). Similarly, we short a contract if the last price is below the 5th percentile, and buy cover if it goes above the 40th.

Despite being an intraday strategy, it typically trades only 1 roundtrip a day - a low frequency strategy. We backtested it two ways: with 1-min trade bars (prices are from back-adjusted continuous contracts provided by eSignal), and with best bid-offer (BBO) quotes with 1 ms time stamps (from QuantGo's actual contract prices, not backadjusted).

For all the contracts that we have tested, the 1-ms data produced much worse returns than the 1-min data. The reason is interesting: 1-ms data shows that the strategy exhibits high frequency flip-flops. These are sudden change in the order book (in particular, BBO quotes) that quickly reverts. Some observers have called these flip-flops "mini flash crashes", and they happen as frequently in the futures as in the stock market, and occasionally in the spot Forex market as well. Some people have blamed it on high frequency traders. But I think flip-flops describe the situation better than flash crash, since flash crash implies the sudden disappearance of quotes or liquidity from the order book, while in a flip-flopping situation, new quotes/liquidity above the BBO can suddenly appear and disappear in a few milliseconds, simultaneous with the disappearance and re-appearance of quotes on the opposite side of the order book. Since ours is a momentum strategy, such reversals of course create losses. These losses are very real, and we experienced it in live trading. But these losses are also undetectable if we backtest using 1-min bar data.

Some readers may object: if the 1-min bar backtest shows good profits, why not just trade this live with 1-min bar data and preserve its profit? Let's consider why this doesn't actually allow us to avoid using TAQ data. Note that we were able to avoid the flip-flops using 1-min data only because we were lucky in our backtest - it wasn't because we had some trading rule that prevented our entering or exiting a position when the flip-flops occurred. How then are we to ensure that our luck will continue with live market data? At the very least, we have to test this strategy with many sets of 1-min bar data, and choose the set that shows the worst returns as part of our stress testing. For example, one set may be [9:00:00, 9:01:00, 9:02:00, ...] and the second set may be [9:00:00.001, 9:01:00.001, 9:02:00.001, ...], etc. This backtest, however, still requires TAQ data, since no historical data vendor I know of provides such multiple sets of time-shifted bars!

As I mentioned above, this type of flip-flops are omnipresent in the stock market as well. This shouldn't be surprising considering that 50% of the stock transaction volume is due to high frequency trading. It is particularly damaging when we are

trading spreads, such as the ETF pair EWA vs EWC. A small change in the BBO of a leg may represent a big percentage change in the spread, which itself may be just a few ticks wide. So such flip-flops can frequently trigger orders which are filled at much worse prices than expected.

Conclusion

The three example strategies above illustrates that even when a strategy trades a low frequency, maybe as low as once a month, we often still require high frequency TAQ data to backtest it properly, or even economically. If the strategy trades intraday, even if just once a day, then this requirement becomes all the more important due to the flip-flopping of the order book in the millisecond time frame.

===

Ernie is the managing member of QTS Capital Management, LLC., a commodity pool operator and trading advisor. Find out more about him at epchan.com