

# OPPORTUNITY IDENTIFICATION AND BACKTESTING

## 1. THE ROLE OF HISTORICAL SIMULATION

A well-designed quantitative trading strategy nearly always undergoes a certain amount of analysis against historical data. There are various reasons to do this

- Find out if the strategy is worth pursuing as a business case
- Make as much money as possible, by optimizing its parameters
- Develop a sense of risks and rewards expected in various live trading scenarios

## 2. MAJOR STRATEGY METRICS

When we judge a strategy, either in historical simulation or in actual performance, there are various metrics we can use. They range from extremely simple ones, such as annualized rate of return with no correction for fees, to more complex ones such as cost-adjusted risk/reward ratios. Here we look at some of the common choices.

**2.1. Return Opportunity.** To compute return opportunity, we simply take a set of actual or hypothetical trades, often of equal size and sometimes without regard to position accumulation costs (bid/ask spreads), figure out the corresponding *PL* (profits or losses), and then divide each by position notional. This gives a rate of return per trade, which we can then average over all trades to get a return opportunity estimate.

Certainly, we can do more sophisticated things such as examining the interquartile range of returns, or weighting simulated trades, or altering presumed trade sizes, but that takes us more into the realm of a full backtest.

**2.2. Strategy Returns.** To compute strategy returns, we construct a somewhat more plausible set of trades, sized perhaps by daily traded notional, sum up the PL, and divide by the sum of notionals to develop strategy returns. At this point, we are likely to begin taking costs into account (hedge fund-style management fees if we are addressing outside investors, internal tech and overhead costs for insiders).

**2.3. Backtested Returns.** Here, we make clear and defensible assumptions about trading costs and trade sizes attainable. We simulate using a reasonable set of risk limits and a reasonable risk model. We assume large positions may require many trades to obtain or liquidate. Preferably, we run any simulations against actual market tick data. We make assumptions about capital requirements and capital held, and report returns on capital rather than on notional.

**2.4. Metrics Derived From Return Series.** Once we have returns, there are a few important metrics we can derive from them

**2.4.1. Sharpe Ratio.** The original *Sharpe ratio* was defined in terms of some “risk-free” rate  $r_f$  as

$$S_{\text{Orig}} = \frac{\mathbb{E}[r - r_f]}{\sqrt{\mathbb{E}[r^2]}}$$

The Sharpe ratio is now universally used in its modified form, sometimes called the *information ratio* where we base on some benchmark return  $r_b$  (which might be 0 for some cases)

$$S_{\text{Modified}} = \frac{\mathbb{E}[r - r_b]}{\sqrt{\mathbb{E}[(r - r_b)^2]}}$$

In the rare case the strategy exactly replicates the  $r_b$ , Sharpe ratio is undefined.

**2.4.2. Sortino Ratio.** The *Sortino ratio* is much like the Sharpe ratio, but it only penalizes cases where return disappoints us versus the benchmark

$$S_{\text{Sortino}} = \frac{\mathbb{E}[r - r_b]}{\sqrt{\mathbb{E}[(r - r_b)^2 | r < r_b]}}$$

Note the Sortino ratio may not be well-defined if our benchmark is not competitive, because then we may have no returns that underperform the benchmark, and thus an empty set in the expectation.

If everything is close to gaussian, Sharpe ratio and Sortino ratio scale (approximately) with the square root of time, so that a strategy with daily Sharpe ratio of 0.1 will have an annual Sharpe ratio of about 1.6.

**2.4.3. Largest Single-Period Drawdown.** Taking some pre-defined period, typically a day, week, or month, we examine the history of the strategy for the worst-performing such period.

2.4.4. *Maximum Drawdown.* We look at maximum drawdown as maximum possible losses incurred by someone who joined the strategy at some peak, and liquidated in some later valley.

2.4.5. *Downside Beta.* We compute at the beta (essentially, correlation) of strategy returns to market returns in cases when the overall market is losing money. To do so, we look at strategy returns  $r_t$  conditional on market returns  $r_t^{(M)}$ , and fit the model

$$r \sim r^{(M)} + \text{Const}$$

on the data set

$$\left\{ r_t, r_t^{(M)} \mid r^{(M)} < 0 \right\}$$

2.4.6. *Stress Scenarios and Stress Beta.* We can choose some scenarios we think would be particularly stressful for the strategy, and examine is average returns or betas to the market in those cases.

2.4.7. *Return Moments.* In addition to the first moment (mean) of returns, we can learn important information about risk from seconds and third moments, *skew* and *kurtosis*.

2.5. **Choice of Benchmarks.** Good benchmarks for a strategy consist of tradable alternatives that are highly formulaic (often buy-and-hold), work with publicly traded and marked liquid assets, and are expected to be related in some meaningful way. For example, a strategy trading convertible bonds might be judged against a buy-and-hold strategy of simply maintaining a position in some convertible bond ETF (*exchange-traded fund*). Occasionally a reasonable benchmark may be comprised of a competing internal strategy.

It is a common phenomenon that, over time, we observe *beta creep* in quantitative investing, where at first no simple benchmark appears to correlate to a strategy returns. As markets learn and mature, products and methods arise to replicate the strategy (at least partially). Whereas they previously had no “beta” to any other security, they begin to exhibit beta to these new securities. it has “crept up” on them.

We have observed this in the equity markets from original risk-free benchmarks, through CAPM and index benchmarks, Fama/French and now ETFs.

2.5.1. *Fama, French and Carhart.* Long ago, asset managers were judged only by gross returns. Eventually, investors realized only market-beating performance was truly worth rewarding. As managers and investors have grown more sophisticated, the concept of outperformance has become more sophisticated.

The Fama French factor model is rightly famous, because it has a long pedigree of explaining otherwise unusual-seeming outperformance in returns. It contains a market factor, a value minus growth factor, and a small minus big factor. Carhart added a momentum factor, though the validity of a momentum factor remains debatable.

**2.6. Return Decompositions.** Once we have a (historical or simulated) time series of strategy returns, it is a simple matter to regress them against a few benchmarks simultaneously. By examining the coefficients and residuals of the regression, we are able to form a *return decomposition*, which attributes a proportion of PL to each benchmark, and the residual PL, denoted alpha or  $\alpha$ , to strategy novelties. Those novelties can then be presumed to consist of some combination of luck, statistical bias, and skill.

Reasonable benchmarks for this purpose include

- Fama/French/Carhart
- Yield Curve Duration, Credit Risk
- Macroeconomic growth and inflation
- Market state (volatility, market volume and spreads)

One underappreciated property of buy-and-hold benchmarks is that, while they are easy to create for a 100% long or 100% short strategy, it is puzzling to form benchmarks for strategies that are long some securities and short others. Even a 130/30 mutual fund is tricky to handle.

If we simultaneously run against a reasonably complete set of benchmarks, the residual return, which has theoretically been controlled for factor exposures of any important kind, is called *portable alpha*. In principle it will be uncorrelated with any other strategy and so can be advantageously combined with other portfolios. When it really exists, it is highly desirable for those reasons.

By hedging, we can truly construct the portable alpha. At any moment, for each benchmark  $B$ , we take its beta from the regression and enter an “extra” position of size

$$-\beta \cdot B$$

The gross pool of alpha available to all actors is zero. Of course, each of them has overhead so the net alpha is in fact negative. Any actor has some finite supply available to them given their skills, and it is constantly reduced by beta creep.

For any given trade, the trader on the other side is likely an informed investor. Overall, seeking alpha is extremely difficult.

**2.7. Further Metrics.** We have further metrics that are important to examine for a strategy. In particular, we care about trade frequency, turnover, capacity, competition/crowding.

### 3. BACKTEST PREREQUISITES

What do we need to form a good backtest?

**3.1. Asset Prices.** We cannot get very far without some way of judging PL, for which we need asset prices. In public markets such as US equity and futures markets, this is usually fairly simple. For OTC and closed markets, such as exotic options or rates options, the problem becomes more difficult. If we have not already got the data we must find a vendor.

**3.2. Trading cost assumptions.** We often assume that trades happen only by crossing bid/offer spreads. This raises questions of the quality of historical data we have on spread sizes and quantities available for liquidity takers. For equities, the information is at least public. For corporate fixed income, the spreads can be 2-5%, for government fixed income about 1/8% and for liquid FX pairs about 1/4%.

Transactions taking place on public exchanges also attract fees. The typical structure is that the liquidity taker (the entity crossing the spread) pays a small fee, while the market maker collects a slightly smaller fee, with the difference going to the exchange.

Trading costs are higher for bigger transactions. A rule of thumb is for total transaction sizes of  $N$  unit in some security  $S$ , trading costs will look like

$$c_0^{(S)} + c_{\frac{1}{2}}^{(S)} \sqrt{N}$$

where the constants  $c_0$  and  $c_{\frac{1}{2}}$  vary by security.

Almgren and Chriss (2000) provide a more sophisticated picture combining a GBM-like random walk with *price impact functions* including a permanent impact based on trading rate  $v$

$$g(v) = \gamma v$$

This sums to a quadratic in size and trading rate, varying from

$$\frac{1}{2}\gamma N^2 + \epsilon N + \left(\eta - \frac{1}{2}\gamma\tau\right)\frac{N^2}{T}$$

to

$$\epsilon N + \eta\frac{N^2}{\tau}$$

depending on the scaling plan.

**3.3. Capital Size and Funding Rates.** Bigger capital means more dollars of profit, but probably smaller returns. It is normal to operate with *leverage* where the capital used is rather less than the notional of positions taken. In this case, one has to *fund* ones positions by borrowing, typically from the prime broker. This is done with reference to some standard rate like fed funds (or formerly LIBOR).

Excess capital in a portfolio may receive interest, but never at the same rate paid for borrowing. The overall difference is called the *borrow/lend spread*.

**3.4. Taxes and fees.** Asset trades are frequently subject to taxes, for example SEC Section 31 fees on US equity transactions. These are passed on to investors.

In addition, asset managers nearly always charge fees to outside investors. Therefore, backtest results of interest to investors are based on returns net of these fees, and all other predictable costs.

**3.5. Available Assets.** We call the assets presumed available for trade at a given time the *universe*. A wider universe means the strategy has a broader choice of investments, and can either cherry-pick the very best return opportunities, or deploy more capital in the pursuit of higher absolute PL.

**3.6. Human/Computer System Performance Assumptions.** Lags from decision time to action vary widely by the systems in place as well as what the market makes practical/. For futures, equities and FX, largely traded electronically, it is reasonable to assume 1 second to 5 minutes. For government fixed income, 30 seconds to 60 minutes, and for corporate fixed income 1 to 10 days.

**3.7. Position Size and Scaling Rules.** Part of any trading strategy is a means of deciding target position size and how to scale into a position. One might choose equal-weighting within a universe, or weight by liquidity, expected profitability, a Markowitz-style analysis, or any combination of these.

Most historical databases are limited to point-in-time capture of prices, either of transactions or quotes. It is important not to rely heavily on the assumption that point-in-time prices would have been transactable in significant size. There is a lot of complexity in inferring, from historical point-in-time data, what transactions really *could* have been made. The question has to do with market size and structure, desired position size, and so on.

When a desired position is small compared to the liquidity of a security, scaling into it isn't an issue. However in many cases we desire position sizes that can take hours, days or weeks to accumulate. There is of course risk in missing out on good positions or continuing to hold bad positions which must be balanced against costs and availability. The seminal paper in this field is Almgren and Chriss.

A common rough solution begins with *VWAP*, or volume-weighted average pricing.

The classical perspective on position sizing is the *Kelly criterion* where we have a given size of capital and want to make the most of it without risking ruin. We have just one "thing" we can do, and know its expected return and variability. Kelly finds under simple assumptions that we should invest a proportion  $s$  of trading capital where

$$s = \frac{\Pr\{w > 0\}}{\mathbb{E}[w | w < 0]} - \frac{\Pr\{w < 0\}}{\mathbb{E}[w | w > 0]}$$

is based on profit probability and expected PL sizes.

The criterion is slow to converge to optimality, requires *a priori* probabilities, assumes event independence, and ignores the vanishing edge of profitability as  $s$  goes to zero. Some of this can be addressed with the *fractional Kelly* criterion.

Another way to address some of these problems is due to *Browne*, who noted that by altering the investment schedule one can obtain far better short-term behavior at the cost of a small positive risk of ruin. This corresponds more closely to true human interests and behavior.

Browne's mathematical target is to maximize the probability of achieving a given wealth  $b$  before some time  $T$  deadline, given some expected excess profitability  $x$ . It results in

$$s = \frac{1}{\sigma\sqrt{T}} \frac{be^{-rT}}{x} \phi\left(\Phi^{-1}\left(\frac{x}{b}e^{rT}\right)\right)$$

which readers familiar with options will recognize as the hedging strategy of a binary call.

The result is that, given some threshold  $\epsilon$  Browne will outperform Kelly by that threshold with probability  $(1 + \epsilon)^{-1}$ . A Kelly criterion on a strategy with 30% volatility requires about 10,000 years to beat a buy-and-hold strategy with 95% confidence. A Browne criterion that accepts 5% ruin probability reduces the time to 85 years.

**3.8. Hedging Rules.** Most quantitative trading strategies aim to hedge out certain risks. A long-short equity strategy may be run a 0%, 25%, 100% or any other proportion net long. An options strategy will hedge delta and maybe gamma. Credit strategies often hedge using CDX or bond ETFs.

Note that (fully) hedging against some security at least theoretically removes it from the list of reasonable benchmarks.

**3.9. Evaluation Frequencies.** Fully algorithmic trading strategies attempt to handle all incoming information in real time. Others must decide on target time periods in which to let analyses lie fallow. These time periods need not be purely clock-based, since a flood of news in the market should naturally be accompanied by re-evaluations.

**3.10. Position Age Rules.** Sometimes quantitative models just “like” certain positions, and we don’t really believe they should be held for extended periods. To this end, many strategies create aging rules in which positions are liquidated as they become too “stale”.

**3.11. Short Position Assumptions.** Terminology for short positions includes *shorting* a security, which means opening a short position<sup>1</sup>, and *covering a short*, which means closing the position by buying securities in the market and sending them to the lender.

Shorting securities is quite different from the common perception. Except in futures and options, is not easy to do, it costs extra money, and it involves (invisibly) selling an option.

Unless bypassed by a swap agreement, creating a short position involves finding a *locate* (usually through the prime broker but sometime in a *self-locate*, agreeing on *repo rate* which may be *special*, abiding by regulations and contracted limits, and running the risk of forced *buy-in*.

For equities, most shorting is done at *general collateral* or GC rates, but some cases, the “special” ones, provide less repo or even cost money, in effect having a special rate less than zero.

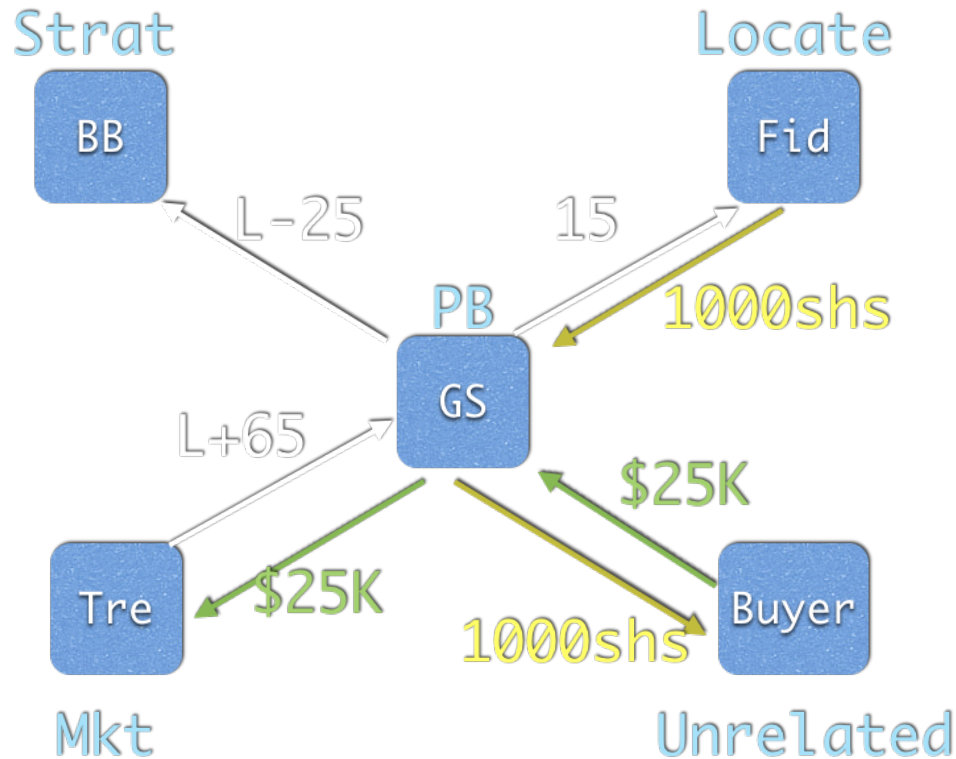
For most other securities, the repo rate is nearly always special.

---

<sup>1</sup> Note the apparent antonym “longing” is *not* a permissible term in finance.



One big problem is that historical information on repo rates and short availability is generally hard to obtain, expensive and of poor quality, unless you are at a large institution with good record-keeping.



#### 4. BACKTESTING CONSIDERATIONS BY ASSET CLASS

##### 4.1. Equities.

4.1.1. *As-of data*. Equity models often take into account SEC filings, which are subject to correction.

4.1.2. *Mergers and spinoffs*. Mergers and spinoffs modify the available universe while also (typically) invalidating the economic assumptions behind our models.

4.1.3. *Dividends and splits*. Equity analyses are often performed with *split and dividend adjusted* data. It is important to keep in mind that this in itself is effectively a backtesting assumption of instant costless reinvestment.

We sometimes see *special dividends* in equities, which saddle a portfolio with unusually large cashflows while also modifying effective position sizes. These can cause problems for casual data aggregators, such as Yahoo, as seen in Figure 4.1.3 . In this case, XLF spun off \$4.44356

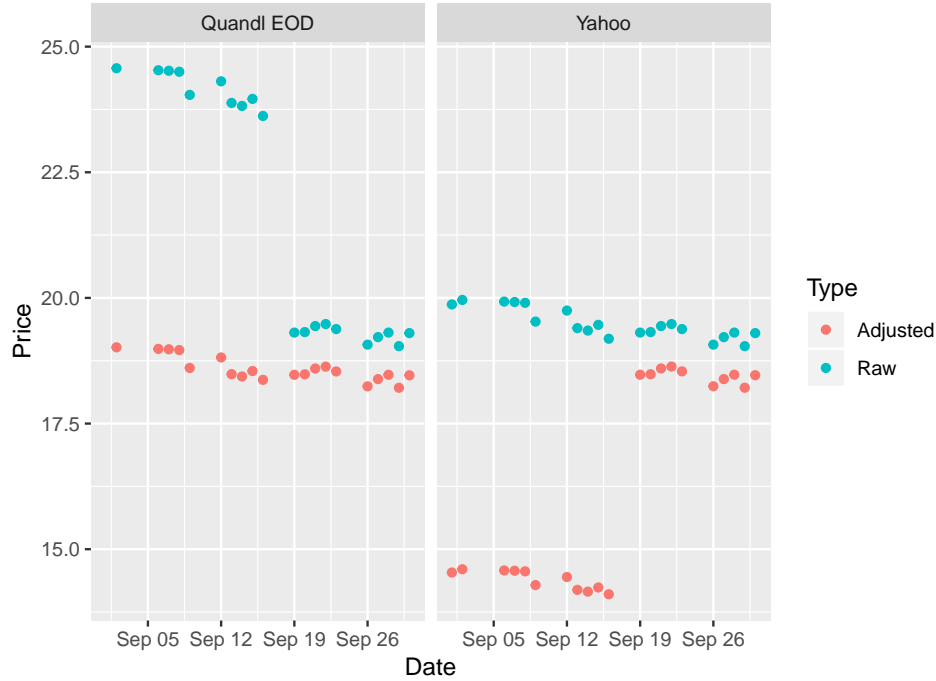


FIGURE 1. XLF dividend adjustment problems for Yahoo: a special dividend and its effect on reported prices on Sep 16, 2016 (as of May 2019) are reflected.

per share, in a combination of cash and XLRE positions. Note the data sources agree on raw and adjusted data *after* the dividend, but that Yahoo is incorrect before it.

Naively, one might think it reasonable to use the Yahoo “Raw” series. However, then standard dividends would remain unadjusted for the remainder of the price series. In addition, returns inferred from the Yahoo “Raw” price, though computed from a time series without discontinuity, would be slightly different.

4.1.4. *High sensitivity to common benchmarks.* If we take overall market return as a benchmark, we find most equities have correlations over 60%. If we allow further benchmarks to include Fama French or a set of ETFs, it can be surprising how much the rest resembles low-volatility noise.

## 4.2. Credit Instruments.

4.2.1. *Finding hedging instruments.* There are not many credit ETFs, and credit indexes tend to be untradable, so the only reasonable hedging instruments in corporate credit are CDX (in very mature markets) or highly liquid corporate bonds in related companies, which take care of market risk without addressing default risk.

4.2.2. *Scaling and inventory.* The credit markets are OTC and are traded in large lot sizes, making minimum position sizes quite large. At the same time, scaling strategies need to take into account the fact that these markets are quite slow relative to electronic markets.

nextframe

4.2.3. *Historical data.* Databases of historical prices are relatively easy to obtain, though many are only approximate. Historical quote databases are more expensive but considerably better, but it is important to keep in mind that quotes in credit markets are almost never “hard”.

Historical trade data for non-144A US bonds have TRACE data, though it does not give accurate sizes.

Though credit instruments ultimately trade on probability of default,  $h$ , there is no historical data for  $h$ .

F 5 05/15/18 \$ C 105.000 +1.000 Corp QRD											
As of 3/23 DELAYED Vol 1,395 Op 104.700 Hi 105.000 Lo 102.500 YLD 4.069											
Definitions GR/QR Options Multi-Day Quote Recap Page 1											
Time	07:00:00	To	16:30:00	Min	Vol(M)	1000	Source	TRAC	USD		
Date	3/15	To	3/23	Price	Range		To		Sprrd	To	Benchmark
FORD MOTOR CRED F 5 05/15/18 105.000/105.000 (4.07/4.07) TRAC											
Date	Time	Act	Price	Ind	Yield	RPS	Sprrd	Benchmark	Size(M)	CC	Trd Time
03/23	08:35:52		102.760		4.479	B	284.2 T1 3/8	02/28/19	1000+	e	08:35:08 3/23/12
03/22	11:56:09		103.000		4.435	D	276.9 T1 3/8	02/28/19	1000+	e	11:56:08 3/22/12
03/22	11:32:47		102.938		4.447	D	278.5 T1 3/8	02/28/19	1000+	e	11:32:45 3/22/12
03/22	11:16:37		103.313		4.377	D	271.6 T1 3/8	02/28/19	1000+	e	11:15:14 3/22/12
03/22	11:15:42		103.250		4.389	D	272.8 T1 3/8	02/28/19	1000+	e	11:15:00 3/22/12
03/22	10:28:39		103.313		4.377	D	271.1 T1 3/8	02/28/19	1000		10:25:34 3/22/12
03/22	09:52:33		103.500		4.343	S	269.6 T1 3/8	02/28/19	1000+	e	09:52:22 3/22/12
03/21	16:11:17		104.250		4.206	S	251.1 T1 3/8	02/28/19	1000+	e	16:11:16 3/21/12
03/21	08:39:59		103.625		4.321	B	259.9 T1 3/8	02/28/19	1000+	e	08:27:47 3/21/12
03/19	07:57:21	N/W	104.438		4.173	D	247.7 T1 3/8	02/28/19	1000+	e *C	11:43:37 3/16/12
03/16	15:28:36		104.250		4.207	B	252.5 T1 3/8	02/28/19	1000+	e	15:28:25 3/16/12
03/16	13:36:58		104.250		4.207	D	252.5 T1 3/8	02/28/19	1000+	e	13:33:23 3/16/12
03/16	13:35:44		104.250		4.207	D	252.5 T1 3/8	02/28/19	1000+	e	13:33:26 3/16/12
03/16	13:31:00		104.188		4.219	D	253.9 T1 3/8	02/28/19	1000+	e	13:30:36 3/16/12
03/16	13:26:59		104.188		4.219	D	253.4 T1 3/8	02/28/19	1000+	e	13:26:38 3/16/12
03/16	11:50:11		104.500		4.162	D	246.5 T1 3/8	02/28/19	1000+	e	11:49:38 3/16/12
03/16	11:44:27		104.375		4.185	D	248.8 T1 3/8	02/28/19	1000+	e	11:43:37 3/16/12
03/15	08:30:06		105.063		4.061	D	237.9 T1 3/8	02/28/19	1000+	e	08:28:17 3/15/12
Australia 61 2 9277 8600 Brazil 5511 3048 4500 Europe 44 20 7330 7500 Germany 49 69 9204 1210 Hong Kong 852 2977 6000											
Japan 81 3 3201 8900 Singapore 65 6212 1000 U.S. 1 212 318 2000 Copyright 2012 Bloomberg Finance L.P.											
SN 772916 CDT GMT-5:00 H192-140-0 26-Mar-2012 11:57:13											

4.2.4. *Tradable instruments.* There are extremely many tradable bonds in the universe because most issuers have several bonds outstanding at

any given time. This is in contrast to equity, where it is extremely rare for a corporation to have more than one type of publicly traded equity.

4.2.5. *Optionality adjustments.* Corporate bonds often come with optionality clauses, most commonly embedded calls, but also puts, green shoes, conditionals and more.

4.2.6. *Long time horizons.* Many players in credit markets have very long time horizons, making liquidity very different from outstanding notional, and lengthening the time horizons quantitative traders must be willing to consider.

4.2.7. *Market sentiment regimes.* Credit markets undergo long periods of over- and underestimates of default risk. Clever strategies therefore should generally hedge against overall sentiment.

4.3. **Foreign Exchange.** Foreign exchange, or *FX* trading involves currency transactions. A certain number of euros are swapped for yen, or a contract is made to engage in such a transaction sometime in the future. The markets here are split into widely traded futures and options, plus a cash market that operates on crossing networks mainly controlled by large banks. This latter point has important consequences for capital usage, because offsetting trades in different markets may get poor treatment.

4.3.1. *Interventions.* FX rates follow general market principles most of the time, interspersed with sizable government interventions of varying predictability. The majority of the time, quantitative models are useful, but the interventions are poor fodder for modeling.

4.3.2. *Pegs.* A common form of ongoing intervention is a peg (or sometimes a pair of lower and upper bounds), maintained by frequent interventions. Strategies relying on pegs may go suddenly bad, and strategies expecting pegs to fail can take a very long time to come to fruition.

4.3.3. *Sizing.*

4.3.4. *Similarity to common strategies.* FX rates are clearly linked to many assets, such as foreign equities and bonds. Crises therefore tend to come with high covariance to other asset classes. In addition, these links mean that one should consider which elements of foreign economies may be driving FX rates.

Unlike equities, a net long portfolio of foreign currencies has no theoretical positive excess return expectation. In each respective country, the currency is, of course, just cash, so the risk premium is zero.

#### 4.4. Options.

4.4.1. *Probability of ruin at high leverage.* Options are often very highly levered investments, so their use comes with enhanced probability of ruin. Hedging can control this to an extent, but markets can begin crashing just when liquidity disappears.

4.4.2. *Volatility hedging.* Options strategies almost always involve delta-hedging. It is extremely common to demand control of gamma/vega (volatility exposure) as well.