



Leveraging Data for Buy Side Risk & Analytics

Data centralization and data considerations for superior analytics

Charlie Browne

Contents

1. Introduction.....	3
2. Data Foundations for Buy Side Analytics	3
2.1. The importance of data lineage.....	5
3. Use Cases	7
3.1. Efficient Frontier	7
3.2. The Capital Market Line and Sharpe Ratio	9
3.3. Beta and CAPM.....	11
4. Liquidity Concepts.....	13
4.1. Anatomy of a market.....	13
4.2. Liquid vs illiquid markets	14
4.3. Liquidity indicators.....	15
4.4. Proxy pricing for illiquid instruments.....	16
5. Data Domains for Buy Side EDM	18
6. The Buy-Side Data Ecosystem	19
6.1. The Buy-Side end-to-end	21
6.2. Instruments of interest list.....	22
6.3. An example flow	25
7. Instrument Reference Data.....	26
7.1. Instruments that can be traded	26
7.2. Instruments that are used to build curves	27
8. Centralizing and Managing Market Data	28
8.1. Sample success factors	28
8.2. Sample KPIs	29
8.3. A typical set of problems.....	29
8.4. A comprehensive market data system will comprise	29
8.5. Price Requests	31
8.6. Market Data Collections	32
8.7. Market Data Validations	33
Snapshots & Time Zones.....	34
8.8. Validations and Correction Methods	34
8.9. The Golden Price Level.....	35
8.10. Golden Price Approaches	36

8.11.	Managing Market Data Exceptions.....	37
8.12.	Managing Time-Series data.....	38
9.	The Complete EDM Solution.....	41

1. Introduction

Analytics, portfolio optimization, and portfolio risk are all terms that are used for calculations that are required for a buy-side firm to efficiently manage its portfolios. These calculated values can either be bought from vendors or calculated within the firm. Duration, effective spread, convexity, tracking error, yield curve sensitivity, value-at-risk, beta, portfolio standard deviation, Sharpe ratio are all examples of the type of calculations that are required by portfolio managers. While purchasing analytics directly from vendors has the benefit of being a very simple operating model, the approach offers neither flexibility nor transparency into the underlying calculations. With high portfolio returns becoming increasingly difficult to generate, firms are choosing to refocus their efforts by adopting technology that offers better insights into their product portfolios and profitable investment strategies. A key aspect of this is the ability to leverage Enterprise Data Management (EDM) approaches that will drive the business forward over the long term. Clean validated data is key. By being better able to analyze time-series of prices, risk factors and portfolio static data, firms can carve out an edge in their portfolio optimization approaches and strategies, such as smart beta and factor returns analysis.

2. Data Foundations for Buy Side Analytics

The approach described in the introduction requires an EDM capability that delivers data and analytics using a two-tier approach

- A flexible analytics layer on top of
- A data Foundation

The diagram below illustrates the concept



concepts:

- Data is validated and sometimes derived in the data layer. While those validations and derivations can contain some calculation logic, for the most part there is very little derivation / analytic logic in the data layer.
- The calculations that do exist in the data layer can be said to be “linear” calculations such as the derivations of FX cross rates or the gap filling of time-series or statistical checks on the quality of the data.
- The data layer does not include business specific calculations such as CAPM, Beta, or factor returns analysis nor does it include non-linear quantitative calculations such as the valuation of OTC options. These calculations all sit in the analytics layer
- Analytics and Quantitative development are the core competencies of the team that work in the Analytics layer. They are not necessarily concerned with the data cleansing approaches and algorithms used in the data layer
- Examples of data cleansing in the data layer are:
 - Instrument gold copy creation
 - Daily price movement checks
 - Validation and gap filling of time-series
- Much of the analytics and quantitative development is done using pricing data (enriched with static data where required).
- The analytics and quantitative development required for portfolio risk and optimization primarily uses time-series /historical data. Whereas NAVs and other valuation-related calculations use today's / current valuation-date data.

Typical Elements of Buy Side Analytics

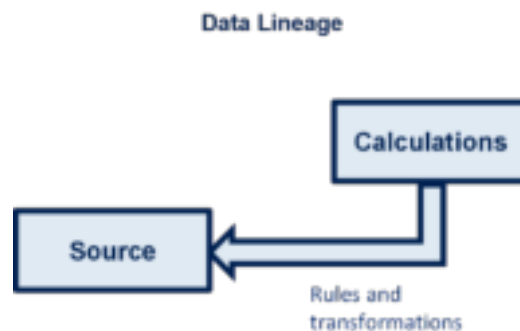


2.1. The importance of data lineage

In the sections following there are examples of investment and portfolio-related calculations and analytics. These calculations include

- Expected Return
- Asset standard deviation
- Asset Skewness
- Asset Kurtosis
- Portfolio standard deviation
- Covariance of two assets
- Correlation of two assets
- Efficient Frontier
- Capital Market Line
- Minimum Variance of a Portfolio
- Risk Premium
- Sharpe Ratio
- Beta of an Asset
- Capital Asset Pricing Model

One observation with respect to these calculations is that the underlying data content for all of these calculations is the time-series of prices or the time-series of returns (price changes) associated with the prices. The ability to examine one of these calculations and have full transparency in terms of both the logic used within them and the inputs to them is essential for audit purposes. Data lineage is the ability to trace the inputs to the calculations all the way back to source.



The tables below show time-series data for instruments representing an Asset A, an Asset B, the Risk-Free asset as well as an instrument representing the market.

Time Series Data for Instruments

Security Master				
			Issuer of Asset A	
			Rating of Asset A	
			Maturity of Asset A	
			Coupon of Asset A	
Time-Series Record			etc	
PriceQuote	Mid	Asset A	22.9	-0.87%

		Asset A		Asset B		RF Asset		Market	
		23.1		4.5		0.40%		123.1	
		22.9	-0.87%	4.7	4.44%	0.40%	0.00%	122.9	-0.16%
		21	-8.30%	4.6	-2.13%	0.40%	0.00%	121	-1.55%
		23	9.52%	4.7	2.17%	0.40%	0.00%	123	1.65%
		24.7	7.39%	4.6	-2.13%	0.40%	0.00%	124.7	1.38%
		22.5	-8.91%	4.4	-4.35%	0.40%	0.00%	122.5	-1.76%
		22.4	-0.44%	4.5	2.27%	0.40%	0.00%	122.4	-0.08%
		25	11.61%	4.6	2.22%	0.40%	0.00%	125	2.12%
		23.9	-4.40%	4.7	2.17%	0.40%	0.00%	123.9	-0.88%
		24.5	2.51%	4.6	-2.13%	0.40%	0.00%	124.5	0.48%
		26.7	8.98%	4.5	-2.17%	0.40%	0.00%	126.72	1.78%
		27.8	4.12%	4.6	2.22%	0.40%	0.00%	127.8	0.85%
		28.9	3.96%	4.7	2.17%	0.40%	0.00%	128.9	0.86%
		29.8	3.11%	4.8	2.13%	0.40%	0.00%	129.8	0.70%
		30.8	3.36%	4.7	-2.08%	0.40%	0.00%	130.8	0.77%
		28.4	-7.79%	4.6	-2.13%	0.40%	0.00%	132.9	1.61%
Expected Return:			1.59%		0.18%		0.00%		0.52%
Stdev:			6.61%		2.66%		0.00%		1.19%
Variance:			0.44%		0.07%		0.00%		0.01%
Skewness:			-29.80%		-13.83%				-69.27%
Kurtosis:			-94.89%		-149.43%				-40.23%
Correlation:			0.31						
Covariance:			0.00052						

Time Series Data

Statistical Calculations

From a data perspective the following comments apply:

- A time-series row represents a historical price or rate
- A price or a rate is a price or a rate of an instrument
- Asset A, Asset B, the Risk Free Asset and the Market are each represented by instruments
- Different price types and price attributes determine the type of price that is being stored, e.g.
 - Bid, Mid, Ask
 - PriceQuote, Percentage, Volatility
- Full data lineage from the instrument identifier back to the underlying security master and source data is required for audit purposes

3. Use Cases

The use cases below are all delivered using the time-series inputs described above.

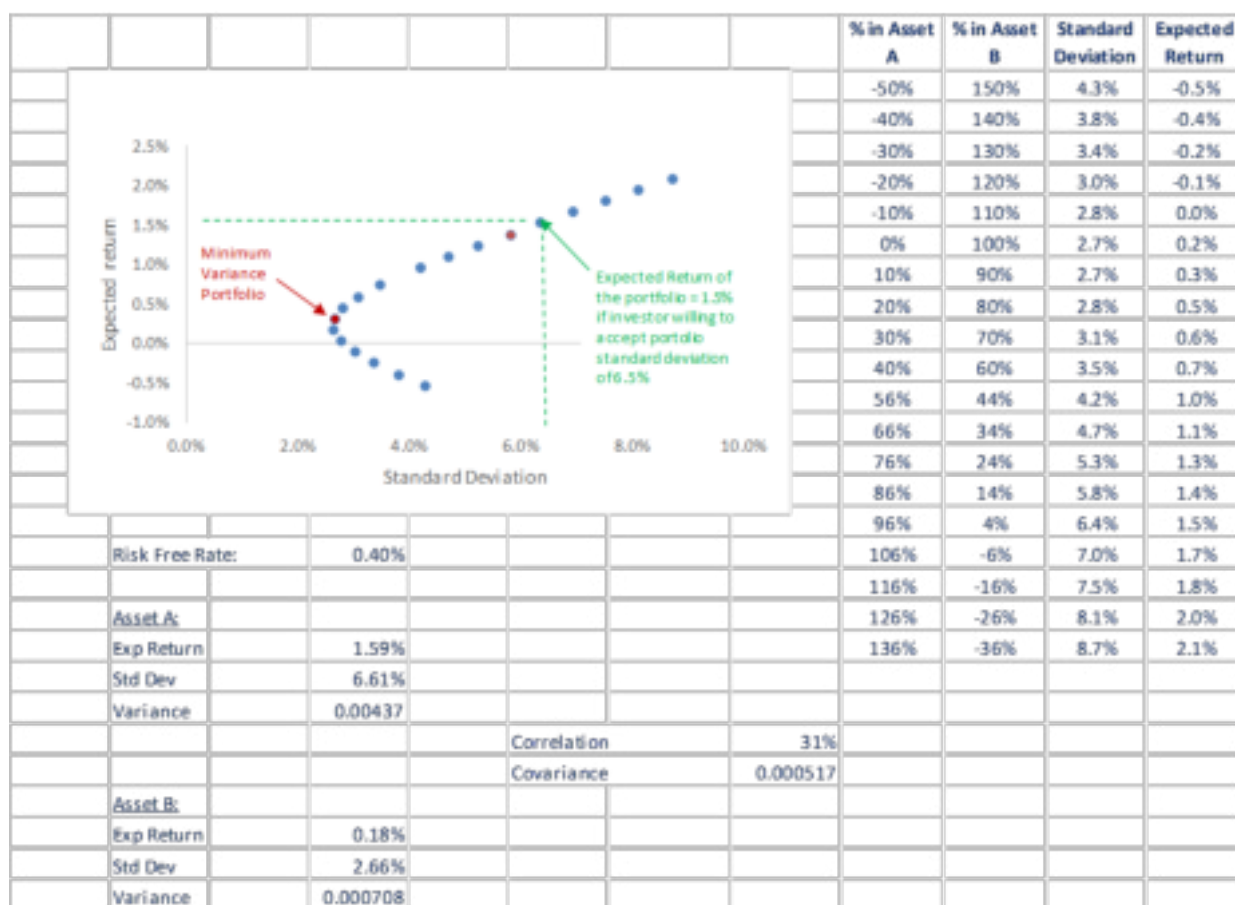
3.1. Efficient Frontier

An *efficient frontier* is a graph that plots the expected return on a portfolio of investments on the y-axis and its risk, as measured by its standard deviation, on the x-axis

In a portfolio, we can have a range of different combinations of the individual assets. The total portfolio return and standard deviation changes in response to any change in the asset allocation.

A risk-averse investor's aim is to identify the mix of investments which yields the highest return for the lowest level of investment risk or expose the firm to the lowest level of risk for a given investment return. The efficient frontier plots portfolios that yield the same risk-return tradeoff.

Efficient Frontier



As you can see from the diagram, two of the key inputs into the calculation of a portfolio's Efficient Frontier are:

- Portfolio Expected Return
- Portfolio Standard Deviation

Portfolio Expected Return

The expected return of the portfolio is the weighted average return expected from the portfolio. It is calculated by multiplying the expected return of each individual asset by its allocation percentage within the portfolio and summing all of the resulting weighted expected returns. The formula for portfolio expected return is:

$$E_r = w_1 \times R_1 + w_2 \times R_2 + \dots + w_n \times R_n$$

Where

E_r = is the portfolio expected return,

w_1 = is the weight of first asset in the portfolio,

R_1 = is the expected return on the first asset,

w_2 = is the weight of second asset and R_2 is the expected return on the second asset and so on

Portfolio Standard Deviation

Portfolio standard deviation is the standard deviation of a portfolio of investments. It is a measure of the total risk of the portfolio. It is also a key input into the calculation of the Sharpe ratio (see below). Investment management finance principles tell us that diversification leads to a reduction in portfolio risk unless there is a perfect correlation between the returns of the portfolio's assets. In statistical terms this diversification benefit means that the standard deviation of a portfolio of assets should be lower than the weighted average of the standard deviations of the individual investments. The formula for the standard deviation of a two-asset portfolio is:

$$\sigma_P = (w_A^2 \sigma_A^2 + w_B^2 \sigma_B^2 + 2w_A w_B \sigma_A \sigma_B \rho_{AB})^{1/2}$$

Where

σ_P = is the portfolio standard deviation;

w_B = weight of asset B in the portfolio;

σ_A = standard deviation of asset A;

σ_B = standard deviation of asset B; and

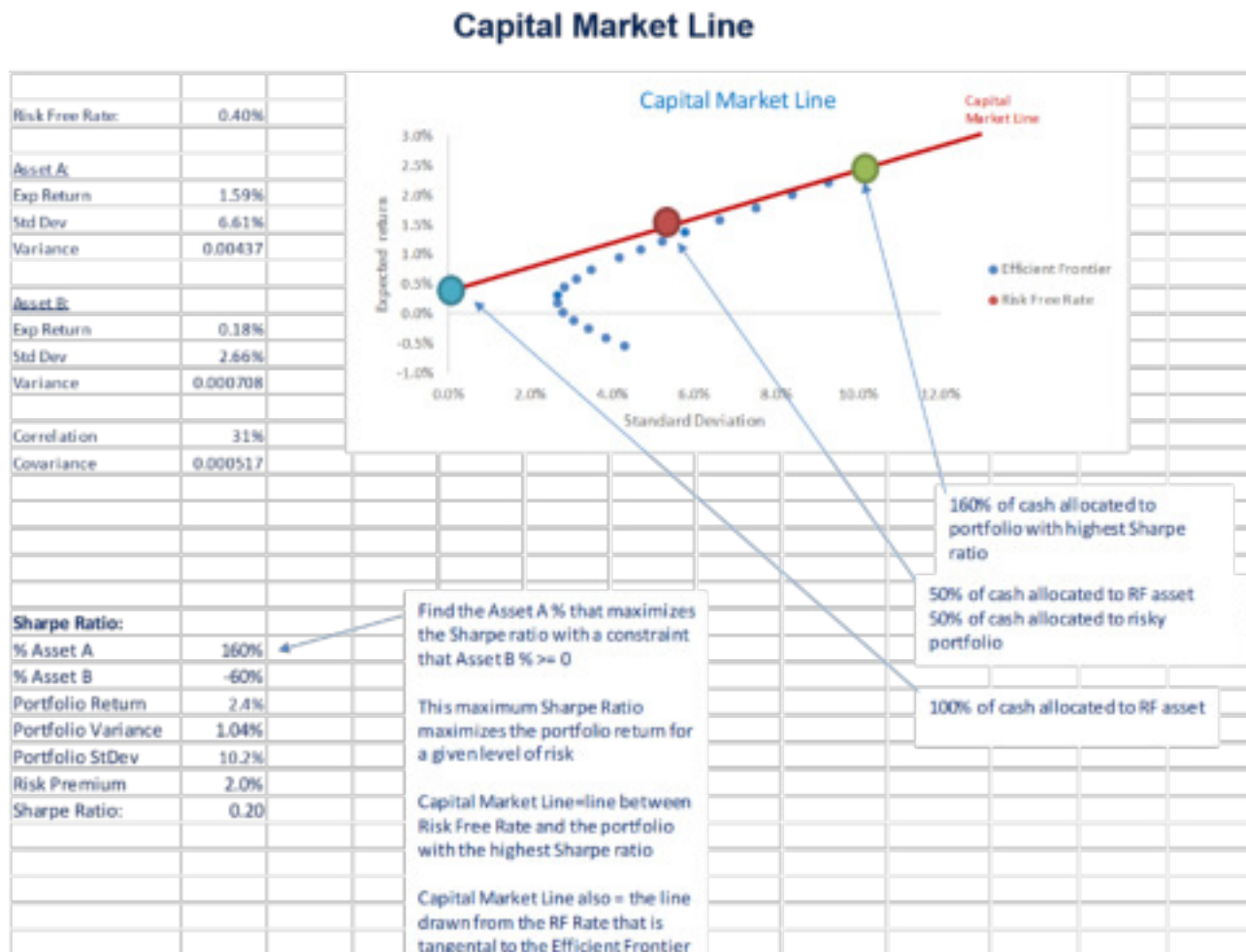
ρ^{AB} = correlation coefficient between returns on asset A and asset B.

3.2. The Capital Market Line and Sharpe Ratio

The Capital Market Line (CML) and the Sharpe Ratio are intrinsically related. The Sharpe Ratio is the expected return of the asset divided by the asset's standard deviation.

Sharpe Ratio: $\frac{\text{Asset's Expected Return}}{\text{Asset's Standard Deviation}}$

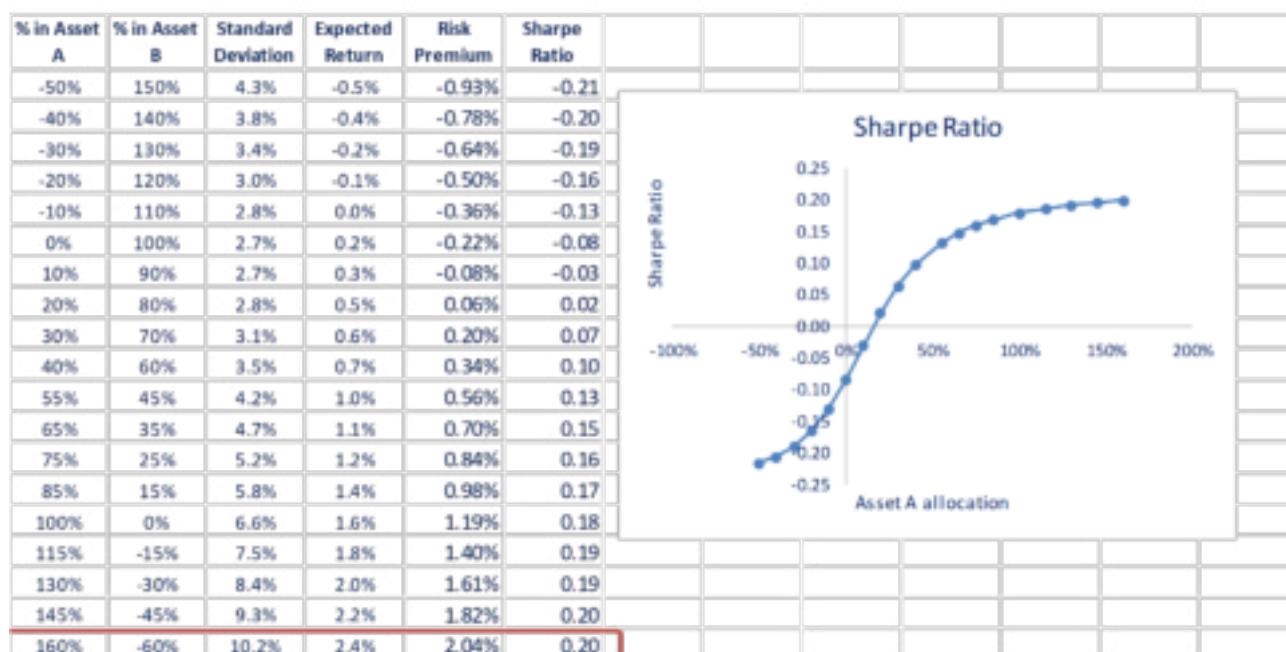
In the two-asset portfolio example below, the CML calculation process involves finding the Asset A percentage allocation that maximizes the Sharpe Ratio with a constraint that Asset B's percentage allocation is greater than or equal to zero. This maximum Sharpe Ratio maximizes the portfolio's return for a given level of risk (where risk is measured by standard deviation). The Capital Market Line is the line between Risk Free Rate and the portfolio with the highest Sharpe ratio. In the diagram below, the CML is also the line drawn from the RF Rate that is tangential to the Efficient Frontier



For our two-asset portfolio, the table below shows a series of portfolios made up of the same two assets, Asset A and Asset B. Each portfolio represents a different allocation of Asset A and Asset B within the portfolio. For each portfolio the table shows

- the portfolio standard deviation
- the portfolio expected return given the portfolio standard deviation
- the Risk Premium (expected return less risk free rate)
- the Sharpe Ratio for the portfolio

Sharpe Ratio



The Sharpe Ratio line chart above plots the Sharpe ratio (Y-axis) against different combinations of Asset A's percentage allocation within the portfolio. As you can see, the portfolio that gives the highest Sharpe Ratio is:

Asset A: 160%
Asset B: -60%

3.3. Beta and CAPM

CAPM stands for Capital Asset Pricing Model and it is a formula that defines the relationship between the expected return of an asset and market risk. The formula is:

$$\bar{R}_a = R_f + \beta(\underbrace{\bar{R}_m - R_f}_{\text{Market Premium}})$$

Where

- R_f: Risk Free Rate, e.g. 10 year Treasury Note
 β: Beta, the relationship between Beta and the Market
 R_m: Expected Return of the Market

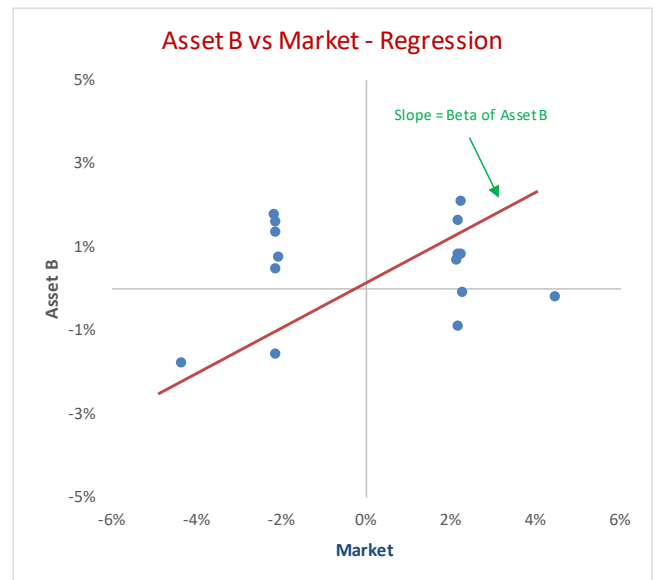
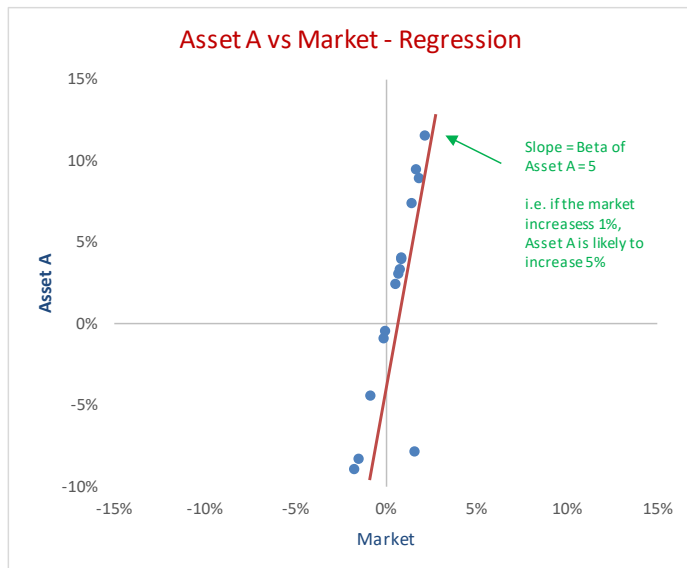
CAPM states that the expected return of an asset equals the return of the risk-free asset plus the market premium multiplied by the Beta of the asset.

Beta is the covariance of the return of an asset with the return of the benchmark divided by the variance of the return of the benchmark. Intuitively, the asset's beta informs users whether, on average, the asset is more or less volatile than the market as a whole.

The table below shows the prices and returns of Asset A, Asset B, the Risk-Free asset and the Market (e.g. an equity index such as the S&P 500).

Calculating the Beta of an Asset

	Asset A		Asset B		RF Asset		Market	
	23.1		4.5		0.40%		123.1	
	22.9	-0.87%	4.7	4.44%	0.40%	0.00%	122.9	-0.16%
	21	-8.30%	4.6	-2.13%	0.40%	0.00%	121	-1.55%
	23	9.52%	4.7	2.17%	0.40%	0.00%	123	1.65%
	24.7	7.39%	4.6	-2.13%	0.40%	0.00%	124.7	1.38%
	22.5	-8.91%	4.4	-4.35%	0.40%	0.00%	122.5	-1.76%
	22.4	-0.44%	4.5	2.27%	0.40%	0.00%	122.4	-0.08%
	25	11.61%	4.6	2.22%	0.40%	0.00%	125	2.12%
	23.9	-4.40%	4.7	2.17%	0.40%	0.00%	123.9	-0.88%
	24.5	2.51%	4.6	-2.13%	0.40%	0.00%	124.5	0.48%
	26.7	8.98%	4.5	-2.17%	0.40%	0.00%	126.72	1.78%
	27.8	4.12%	4.6	2.22%	0.40%	0.00%	127.8	0.85%
	28.9	3.96%	4.7	2.17%	0.40%	0.00%	128.9	0.86%
	29.8	3.11%	4.8	2.13%	0.40%	0.00%	129.8	0.70%
	30.8	3.36%	4.7	-2.08%	0.40%	0.00%	130.8	0.77%
	28.4	-7.79%	4.6	-2.13%	0.40%	0.00%	132.9	1.61%
Risk Free Rate					0.40%			
Expected Return:		1.59%		0.18%		0.00%		0.52%
CAPM Return:		0.88%		0.44%				
Stdev:		6.61%		2.66%		0.00%		1.19%
Variance:		0.44%		0.07%		0.00%		0.01%
Covariance with Market:		0.058%		0.005%				
Beta		4.08		0.35				



In mathematical terms the Beta of an asset equals the slope of regression line drawn through the plot of the asset's returns plotted against the market's returns. E.g. in the diagram above for Asset A, the slope of the regression line = 5, implying that the Beta of Asset A equals 5.

4. Liquidity Concepts

Market liquidity is a market's ability to purchase or sell an asset without causing a material change in the price of the asset. It describes the asset's ability to sell quickly without having to reduce its price. Liquidity is about how big the trade-off is between the speed of the sale and the price it can be sold for. In a liquid market, the trade-off is mild: selling quickly will not reduce the price much. In a relatively illiquid market, selling it quickly will require cutting its price by some amount.

4.1. Anatomy of a market

Markets in their purest form are relatively simple constructs. There are buyers, sellers, ask prices and bid prices. There isn't much more to them than that. Buyers always have to buy at the price the seller is willing to sell at (the ask price). And sellers always have to sell at the price the buyer is willing to buy at (the bid price). These basic concepts apply to anything that is being bought and sold in any market in the world.

Financial markets are like any other market. Exactly the same principles apply as apply for the exchange of goods and services in the rest of the economy. The complexity that many people feel when thinking about financial markets comes from the fact that in financial markets the products that are being traded are less familiar to people. Exchanging money for widgets in a manufacturing environment is a relatively straight-forward concept. Indeed, exchanging money for a bond or an equity is also a relatively straight-forward concept, regardless as to whether you understand what a bond or equity is. For them to be bought and sold they must be perceived by the buyer and seller as having a value.

Conceptualizing payment for an over-the-counter (OTC) derivative that transfers interest rate risk from a fund to an investment bank is more difficult to achieve for most people. Or when a pension fund manager transfers 50 year inflation rate risk to an investment bank via an inflation swap, the concepts become that bit harder to grasp. They become even harder to grasp when non-market practitioners are told that the inflation swap is exposed not just to inflation rate risk but also to interest rate risk, because the future cash flows generated by the inflation swap need to be discounted with an interest rate curve.

Explaining that the interest rate curve that is used to discount the future cash flows of the swap depends on the collateral posted against the swap, can lead to further confusion¹. And telling someone who doesn't have experience with derivatives or financial markets that the asset manager who needs to post collateral has an option to post dollars, sterling or US Treasuries as the collateral amounts, will likely lead to that final bit of confusion. And if it doesn't, it certainly will when you tell them that the optionality the asset manager holds when posting cash or bonds to collateralize to the swap position, means the firm has exposure to volatility rates that need to be derived from the market.

Luckily, anyone tasked with managing financial market data need not fully understand financial market concepts if their data management platform is run on a data model that natively understands all types of financial instruments and the relationships between all types of financial

¹ A swap that is fully collateralized with margin cash posted on a daily basis should be discounted at the rate of return that the cash would give an investor if he invested it for one day (the overnight rate)

data. All a practitioner needs to know is the fundamentals of how a market works, i.e. there are buyers, sellers, ask prices, bid prices and (conceptually) a mid-price. These concepts are illustrated using the diagram below.



4.2. Liquid vs illiquid markets

Markets can be liquid or illiquid. If they are liquid, there is lots of trading going on and their bid-offer² continuums as illustrated in the diagrams below are crowded, leading to a small gap between the highest bid and the lowest offer. This gap, whether big or small, is called the bid-offer spread. As you can also see from the diagram below, in an illiquid market the bid-offer continuum is typically sparsely populated, leading to a wide bid-offer spread. In illiquid markets, sellers find it hard to sell and buyers find it hard to buy. When bid-offer spreads are wide enough, mid-price valuations become meaningless, and large reserves for bid-offer spreads need to be taken on to the firm's balance sheet.

² The terms "offer" and "ask" are analogous in the context of financial markets

The Liquidity of Markets



4.3. Liquidity indicators

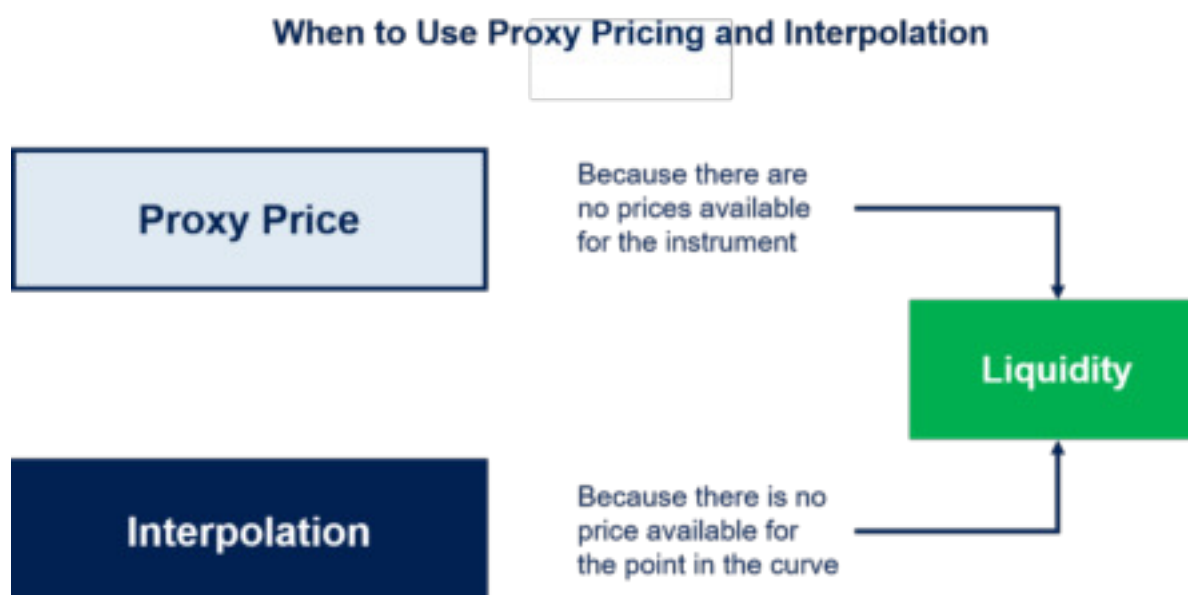
There are different ways to measure liquidity. The size of the bid-offer spread, the number of trades or quotes in a given time-period, the length of time it takes to exit a position without moving its price. The table below provides an overview of the different types of liquidity indicators there are and the regulations that they apply to

Liquidity Indicators Across Buy and Sell Sides

Liquidity Indicator	Related Regulatory Concept
Size of Bid-Offer spread	NIFOR1 Bid-Offer Provisional PRUVAL Close-out Costs AWA
Number of ticks in a time period	NIFOR1 IFRS13 Level 1/2/3 Classifications
Market Depth Distribution of price sources	PRUVAL Market Price Uncertainty AWA
No of trades in a time period	NIFOR1 IFRS13 IFRS13 Level 1/2/3 Classifications
Exit Time for a position	PRUVAL Concentration AWA FRIB ES Liquidity Horizon
Liquidity breakdown of fair value balance sheet	NIFOR1 IFRS13 Level 1/2/3 Classifications

4.4. Proxy pricing for illiquid instruments

Proxy pricing, model pricing and interpolation are methods for deriving prices when there are insufficient bids and asks available in the market to come up with a price that can be used for valuation purposes.



When securities or curve points are illiquid, there are a lot of different ways that proxy / model prices can be derived for them. The use cases described below cover proxy pricing, interpolation and calibration. But they are all doing the same thing; deriving prices where there is insufficient liquidity in the market to get them from contributors. Using combinations of native market data derivation functionality and application programming interfaces (APIs) to 3rd party quant libraries, a market data system needs to ensure that there is a framework in place that can handle each of these use cases.

Securities

- Use one security price as a straight proxy for the illiquid security³
- Use % change in a security price as a proxy for the % change in price for the illiquid security
- Use % change in a benchmark index as a proxy for the % change in price for the illiquid security
- Use yield on a bond with similar credit rating, sector and maturity to derive (e.g. using a Quantlib function) the price of the illiquid bond

Interest Rate Curves

- Proxy curve = Fixed spread over an existing curve
- Proxy curve (e.g. 6m Libor) = Tenor basis spreads (e.g. 3s6s) added to a base curve (3m Libor)
- Proxy curve (e.g. CHF Libor) = Cross currency basis spreads (e.g. USDCHF xccy basis swaps) added to USD Libor curve to derive a CHF Libor curve
- Proxy curve points where missing curve points are derived using linear interpolation / extrapolation
- Proxy curve points where missing curve points are calculated using a whole curve spline / smoothing (e.g. Cubic spline) interpolation routine

³ The “illiquid” security or curve is the security or curve that the market data solution needs to calculate a proxy price / rate for

5. Data Domains for Buy Side EDM

The purpose of this section is to provide an overview of the data management requirements of a buy-side firm. In this context buy-side covers asset managers, fund managers and investment firms generally. The data requirements of a buy-side firm can be categorized as follows:

- Instrument reference data requirements
- Market data and Risk data requirements
- Funds data requirements

Reference data is a general term used for non-transactional or economic data. Reference data generally does not change that often. Categories of reference data are

- Instrument reference data
- Counterparty reference data
- Entity reference data
- Book reference data
- Location reference data
- Organization hierarchy reference data

Market data and risk is a subset of instrument reference data. It describes the prices of instruments that are traded or used to value instruments that are traded. The term market data covers

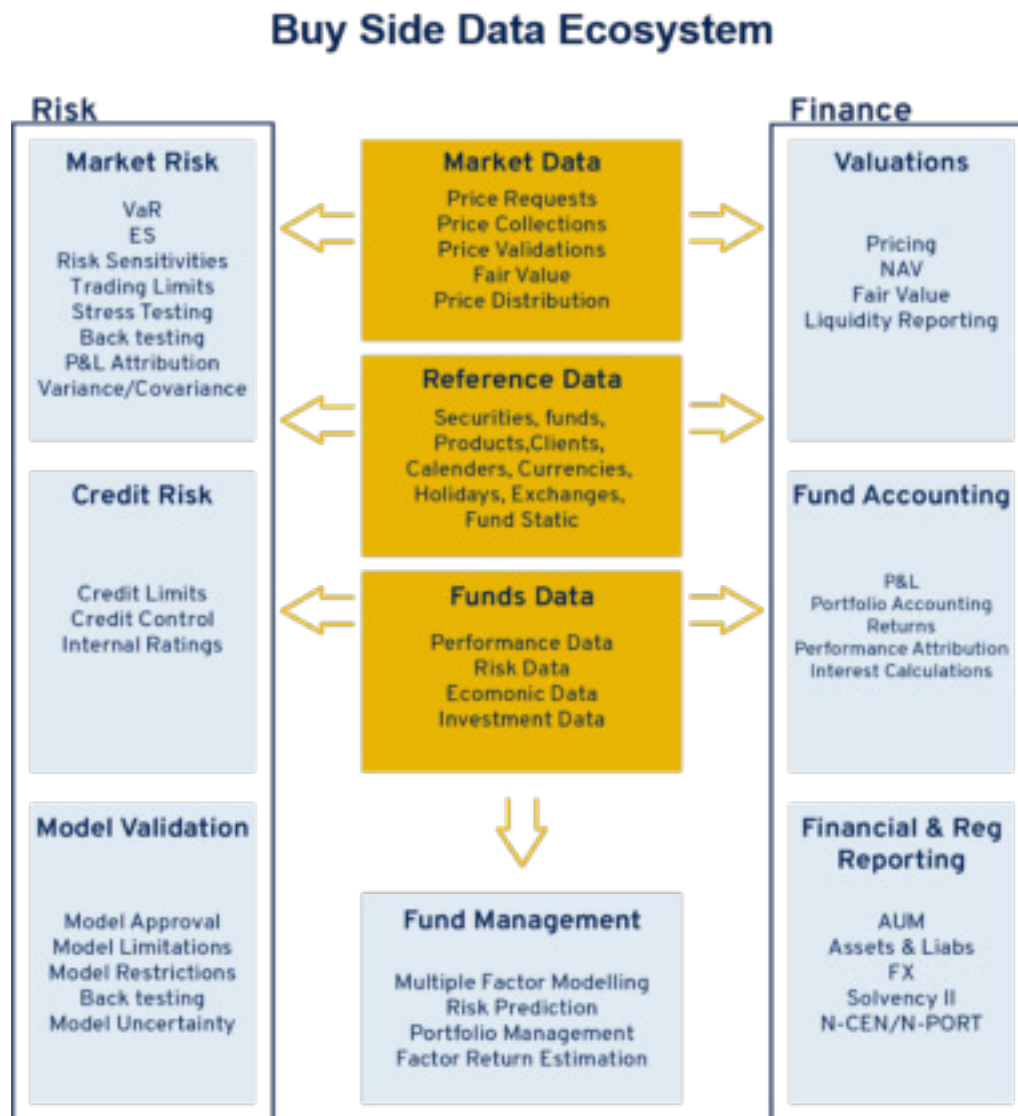
- Prices for securities, typically equities and bonds
- FX rates, either spot rates or forward rates
- Interest rates that apply for a set of maturities that extend out the yield curve
- Credit spreads, typically either derived from prices of Credit Default Swaps or from the prices of traded bonds
- Volatilities, where the volatilities of a given underlying security or rate can be represented on a grid whose dimensions that relate to option expiries, option strike/moneyness, or perhaps maturity of the underlying instrument or derivative
- Time-Series of all of the instrument prices above
- Calculations used for portfolio risk analysis that use that time-series data, including
 - Standard Deviations
 - Correlation
 - Covariance
 - GARCH

Funds data requirements covers

- Performance data, including changes in both prices and Net Asset Values (NAVs) that are used to measure the values of funds
- Funds data, including fund name, notional amounts held, instruments that the funds has positions in, clients that have bought shares in the fund, Net Asset Values,
- Investment data including share turnover,
- Risk data, including Risk Factors, Standard Deviations, Asset Co-variance matrices, Beta values, High-Low ranges, relative strength measures, historical alpha values

6. The Buy-Side Data Ecosystem

The blue boxes in the diagram describe the functions or teams that typically exist within a buy-side firm. The gold boxes in the center show the three categories of data (Market Data, Reference Data, Fund Data) being at the heart of the various calculations that a buy side firm would need to do. The diagram describes the “gold copy” approach. Such an approach assumes a centralized golden source of cleansed and normalized data being distributed to systems that require it for their calculations.



Our mantra at GoldenSource is that 80% of the complexity in Risk and Finance calculations is in the data. Every instrument has a price. Every point on every curve is an instrument. Each position that a firm holds has an exposure to one of those market data instruments. Cash positions have exposures to prices in cash instruments identifiers (e.g. ISINs, SEDOLs). Derivative positions (aka risk sensitivities) have exposures to prices in instruments on curves. Cash instruments typically have associated reference data, e.g.

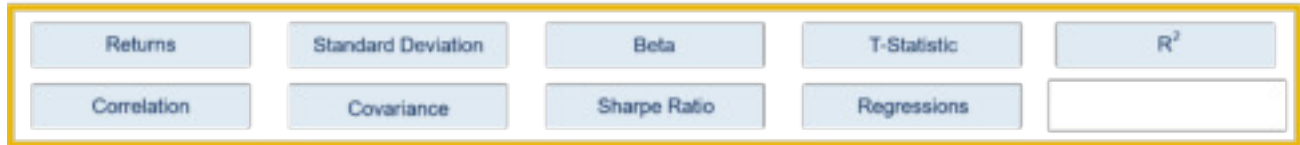
- Issuer
- Rating
- Asset class
- Currency
- Coupon
- Geographic units
- Industry classifications

Each issuer can also be a trading counterparty. Each trading counterparty can also be a financial institution. These reference data relationships can go on and on. Getting them right is key to success in any operations or analytic endeavour.

The diagram below shows that valuations are the common economic data unit for buy-side risk and finance calculations. Performance data and market risk data are derivations of a valuation and the key data inputs into a valuation calculation are

- Position data
- Market data

Buy Side Finance and Risk Calculations



Data



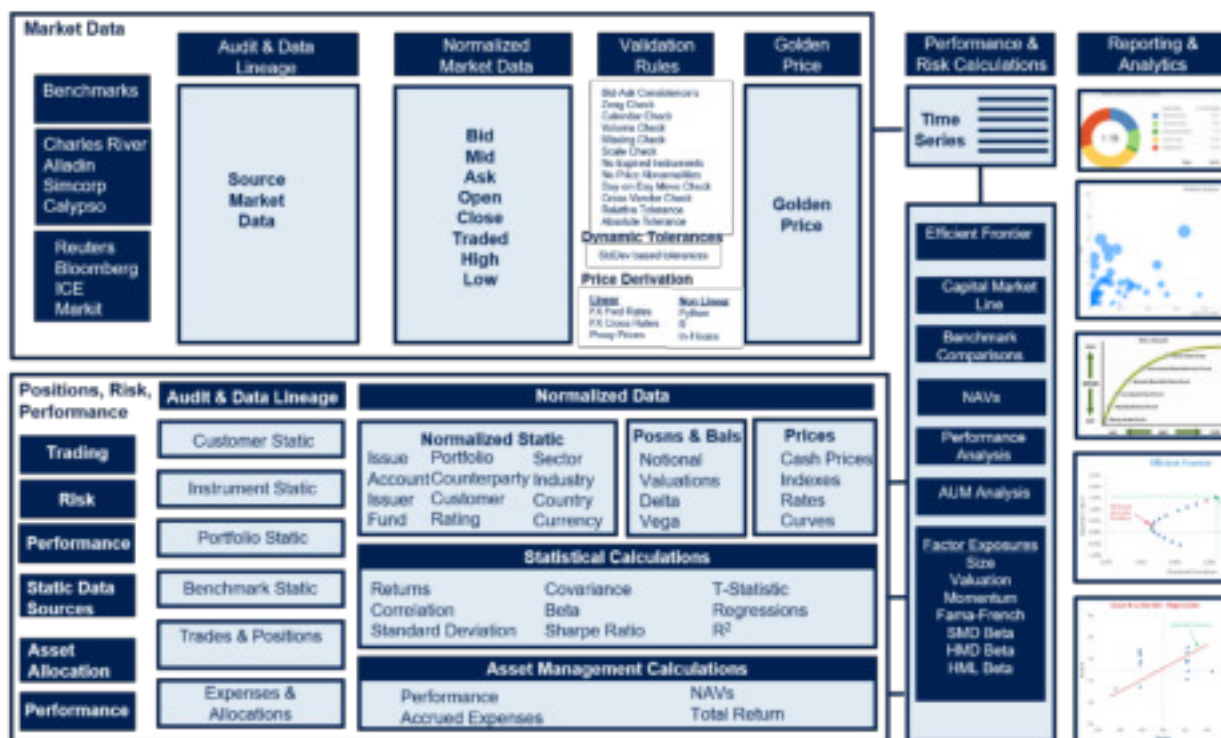
6.1. The Buy-Side end-to-end

The diagram below describes an end-to-end data flow for a typical buy-side firms' data ecosystem. At the heart of the ecosystem are

- A Market data system – see 'Centralized Market Data' below
- Systems that store positions, risk and performance data

All buy-side calculations require these two key categories of data inputs. Market data and instrument reference data can be thought of as external data. It is either freely available data or can be purchased. Position, risk and performance data, on the other hand is internal data. Only the firm itself should know about its positions, its exposures and the performance of those positions as they are grouped into funds.

Buy Side Analytics End-to-End



6.2. Instruments of interest list

While position, performance and risk data are key “internal data” inputs for buy-side calculations, a buy-side firm may not necessarily be interested in the loading of this internal position related data into a centralized data management platform. Where this is the case, i.e. where a firm is only primarily interested in implementing a centralized market data system, then a process will be required to identify an “instruments-of-interest” list. An instruments of interest list is a list of the instruments that the firm needs prices for. These instruments of interest can either be

- instruments that are traded
- instruments that are used purely for pricing purposes

Instruments that are traded are usually called securities. Instruments that are used purely for pricing purposes are usually used to price derivatives and typically for part of market data structures such as curves. See below for more detail on these two categories of instruments. For securities, the instruments-of-interest list will be the list of instrument identifiers that the firm holds “cash” positions in either for itself or on behalf of clients. These positions will usually be equity or bond/note positions. For derivatives, the instrument-of-interest list will be different. Positions in derivatives are called “risk sensitivities”⁴. Risk sensitivities are exposures to a piece of market data that sits on a curve. An example of a risk sensitivity is a PV01. A PV01 is a currency amount that represents a change in the value of a position if a rate on an interest rate

⁴ Other terms for risk sensitivities are “the greeks” or “exposures”

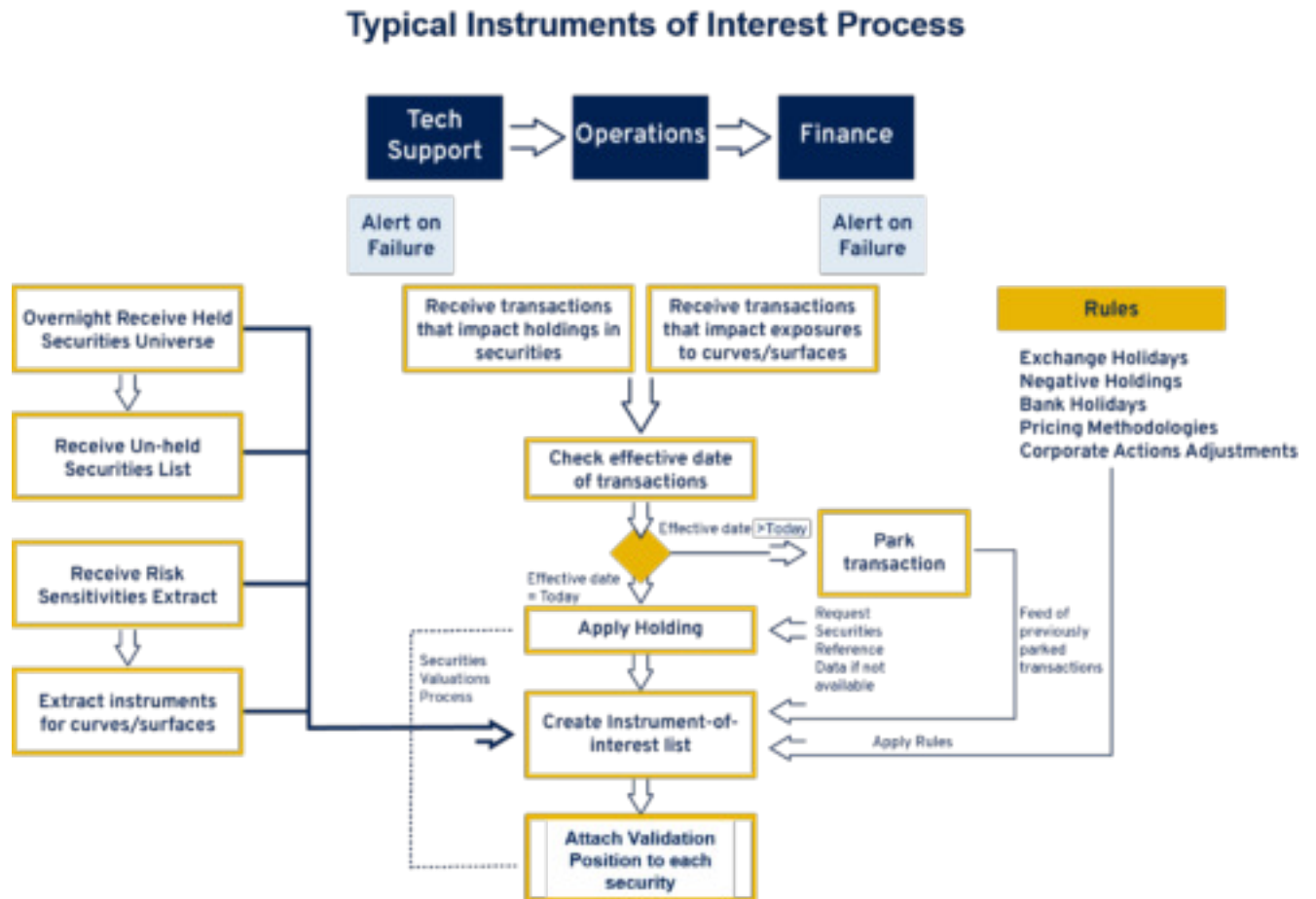
curve changes by 1 basis point. E.g. the 2 year point on an interest rate curve might have a PV01 of \$252k. This means that a 1 basis point⁵ increase in the rate attached to the 2y point on an interest rate curve will lead to a \$252k increase in the value of the derivative that has exposure to that 2y point on the curve. Note: that the rate on the 2y point on the curve is not a currency amount. It is an interest rate. It is the PV01 that is a currency amount and it is the PV01 that represents the derivative position (or one of the positions) that the firm holds. When booking derivatives the firm will likely create multiple risk sensitivities at the same time. For example, if the firm books a Swaption position it will likely create exposures to multiple points on an interest rate curve. The Swaption could have exposures to the 1m, 3m, 6m, 1y, 2y, 3y and 5y points on the USD interest rate curve.

It is important to explain the difference between securities positions and derivatives exposures to points on curves, because the process for extracting an instruments-of-interest list will be different for each. An instruments-of-interest list for securities positions will simply be the list of securities identifiers, e.g. ISINs or SEDOLs, in which the firm has equity or bond positions. The process for extracting an instruments-of-interest list for derivative positions, on the other hand, will require the firm to first produce an extract of risk sensitivities from its risk system or trading system – depending on where the firm stores its risk sensitivities. Once a full risk sensitivities extract has been produced, the firm will be able to derive the actual market data points (the points on curves) that the firm has exposure to. It is these market data points that form the content of the instrument-of-interest list for derivative positions. The instrument identifiers on this derivative instruments-of-interest list cannot be traded, i.e. the firm will not “own” them in the way that it owns a holding in an ISIN or SEDOL for a securities position. But it does have exposures to these instrument identifiers. Hence they are instruments that are used purely for pricing purposes.

Processes which generate instruments-of-interest lists should be scalable so that they can generate instrument identifiers that can price positions in the un-held universe if required. These processes should also be able to identify instruments that are part of positions created by new trades and positions that have been or are about to be impacted by corporate actions.

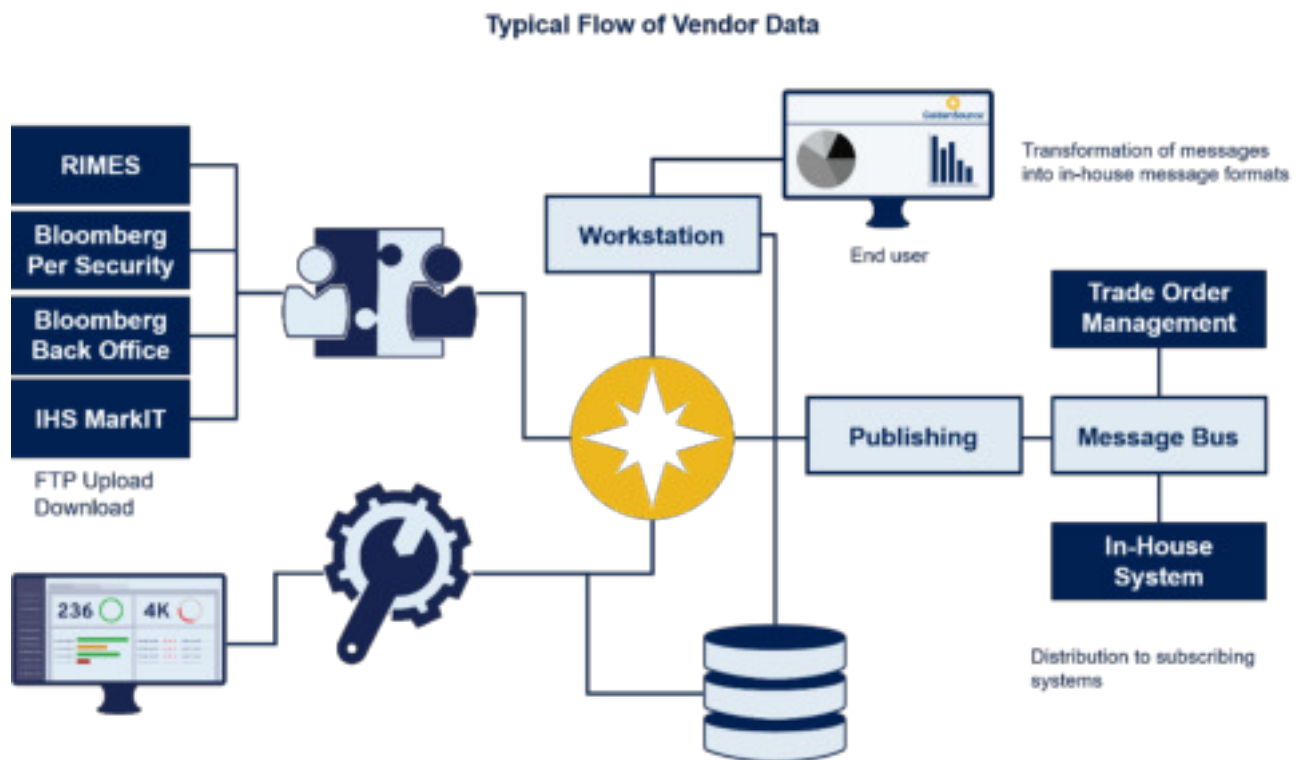
⁵ A basis point equals one hundredth of 1%

The diagram below provides an overview of a typical instrument-of-interest process at a buy-side firm:



6.3. An example flow

The diagram below shows a typical flow of data from data vendors into an Enterprise Data Management (EDM) system. The EDM system validates and creates gold copies of the data and then uses its publishing framework to distribute to the downstream systems that need it for calculations and reporting



A typical EDM platform will comprise:

- **Managed Connections** – these are the connections to data vendors that an EDM system should automatically maintain
- **Internal Data** – these are the trades, fund data, risk and position data feeds that flow from the firms booking and operational systems into the EDM platform
- **Data Governance Structure** – this layer includes validation rules, tools for data lineage, data standards, default values and customizable workflows that ensure that data that flows into the gold copy section is normalized and conforms to the firms standards
- **Gold Copy Data Store** – this is the data mastering layer. Every instrument, customer, entity, price, fund and all related instrument and fund static is stored only once with gold copy identifiers relating the data back to vendor identifiers
- **Business Information and Publishing layer** – this layer requires flexible dashboard and reporting functionality as well as a robust publishing layer that allows both gold copy and source data to be published to downstream systems in multiple formats

7. Instrument Reference Data

Market Data systems (MDS) should sit on top of, and be fully integrated with a Security Master (SM). Instruments can be set up using either the basic SM functionality that comes with the market data system or using full SM module capabilities. A full SM module should support the storage and maintenance of securities terms and conditions for all asset classes, including

- FX
- Rates
- Credit
- Bonds
- Equities
- Commodities

For a truly integrated solution, underlying all SM and MDS functionality should be a single, integrated data model. The GoldenSource data model, for example, has been built using IP crowd-sourced from the industry over 25 years. All types of financial information (e.g. instrument, counterparty, customer, book, account, entity, product) is defined in the data model. The GoldenSource Data Model R&D team undertakes the research required to ensure the data model stays current – and delivers a governance framework around the data model that ensures it remains consistent

There are two different concepts for instruments types that need to be handled within a data model

- Instruments that can be traded (that the firm has a holding in):
- Instruments that are used to build curves

7.1. Instruments that can be traded

These are either securities (e.g. bonds, equities, notes), FX spot rates, exchange traded derivatives (futures, options, i.e. ETDs) or OTC issues that exist only between the two counterparties of the OTC trade. For securities all industry standard identifiers for instruments need to be supported (e.g. SEDOL, Cusip, ISIN, Valoren) with each of these linking to a single, internally generated, instrument identifier within the system. The 1-N relationship between the internally generated instrument identifier and the various market identifiers ensures instrument referential integrity (critical for instrument usage in all applications that use instrument static – including market data systems). In the cases of securities, ETDs and FX spot rates, there are market prices available for the instruments that are being traded. For OTC trades, which involve multiple unknown future cash flows, single prices cannot be used to value the instruments. Data structures, modelled as Curves, are required

See below for a list of the different types of functionality users should expect to be delivered with an effective SM module

- Multiple instrument identifiers
- Out-of-the-box, auto-maintained, mappings to vendor feeds (See Connections above)

- Customisable mappings using flexible mapping tool
- User friendly, fully configurable set-up and maintenance screens
- 100+ out-of-the-box instrument validation rules
 - Identifier check digits
 - Core Rules (e.g. Issuer validation, Issue Type validation, Market validation)
 - Optional rules (e.g. Date derivations, Maturity schedules)
 - On Demand Rules
- Instrument to issuer referential integrity is a core part of SM with screens and validations for issuers fully integrated with instruments

7.2. Instruments that are used to build curves

Curves are typically built from instrument identifiers that are not traded. E.g. if the firm enters into a 5 year Sterling Fixed-for-Floating Swap with a counterparty, then, in addition to the OTC instrument used when the position is booked, a series of additional instruments are required to be set up so that the position can be valued. These instruments might, for example, be deposit rates, FRA rates, Futures, Swap rates which are used to build the forward curve. A set of discount factors (zero coupon rates) will be generated using the forward curve. Each of the instruments in the forward curve, such as the deposits and the swap rates, should have their own identifiers (e.g. RICs, Bloomberg Tickers). E.g. the Bloomberg ticker for the GBP 5 year swap rate is BPSW5V3. This instrument will not be used to record a position in Finance or Risk systems because, even if the firm originally entered into a swap with a 5 year maturity, in a year's time, the swap will only have 4 years remaining to maturity. Instead, BPSW5V3, will be used as one of the instruments to build the Sterling Libor curve in the firm's market data system – which in turn is used to value the swap position

See Centralised Market Data below for more details of curve validation

8. Centralizing and Managing Market Data

A centralized market data system is a critical component of any firm that needs market prices to run its business. Your firm needs a centralized market data system if one or more of the following applies:

- you trade bonds and shares or derivatives
- you need to mark your positions to market
- you need to calculate Net Asset Values
- you need market data for your time-series based calculations (VaR, ES)
- your market data is currently stored in multiple disparate locations
- you spend a lot of time manually scrubbing market data
- you want an automated approach to market data
- you want a single source for your market data
- you want to reduce operational risk associated with handling market data
- your auditors have told you your market processes are ad-hoc
- you want to reduce the cost of market data in your firm
- you want to reduce the number of market data terminals your firm pays for
- you would like to view market exceptions using user friendly curves
- you want configurable, out-of-the-box validation rules/tolerances
- you want to have a hierarchy of pricing sources

The key requirements in a centralized market data system are:

- It provides a central repository for market data in a firm
- It validates market data in a fully audited system
- It handles EOD and time-series market data
- It generates a “golden price” for each price point
- It allows users to drill from “golden price” through rules to source prices
- It contains analytic tools required to generate proxy and model prices

8.1. Sample success factors

- All instrument reference data should be the gold copy for the firm
- All instrument reference data should be up to date and accurate
- Any changes to the data should be auditable, with information on when the data was created or changed accessible in the database
- Manual intervention between gold copy reference data and target publishing systems should be minimal
- Users should be allowed to re-request data if it does not meet the minimum required quality criteria

8.2. Sample KPIs

- STP Rate > 95%
- Instruments reference data used in Pricing systems should have an error rate that is less than that agreed by the firm's data governance policy
- Instrument reference data used to build curves should have an error rate that is less than that agreed by the firm's data governance policy
- Instrument reference data used on instruments that trades are used to book trades should have an error rate that is less than that agreed by the firm's data governance policy

8.3. A typical set of problems

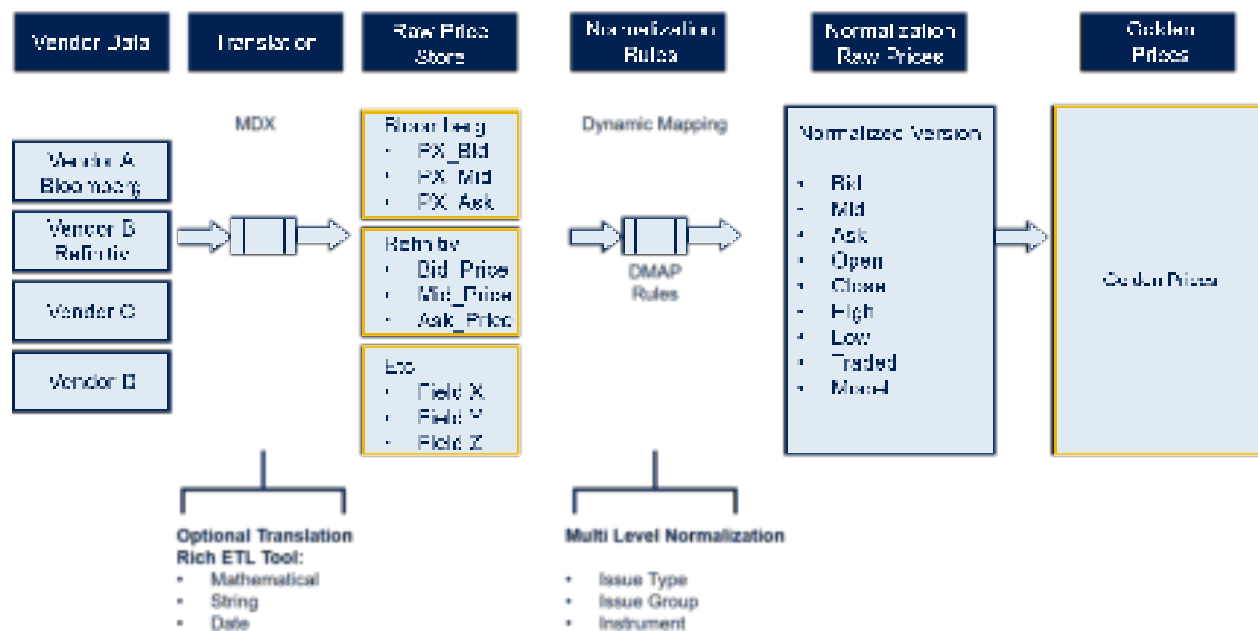
- "My market data is stored in multiple data stores and I am struggling to see how these data stores are populated"
- "I have no transparency on my total market data costs"
- "I need to know which vendors I request market data from and how often I request it"
- "I need to be able to reconcile my market data costs to my vendor market data requests"
- "I need to know which teams in my firm use the various prices that are requested"
- "I need to know where market data is being requested but not downloaded"

8.4. A comprehensive market data system will comprise

- **An extensive list of Pricing/Market Data Connections.** The market data solution will typically come with out-of-the-box connections. These are feeds from market data vendors such as Reuters, Bloomberg, ICE Data Services, MarkIT, WM, etc. Market data solution vendors will sometimes offer "connections" services on top of the APIs. These are connections to feeds that the data vendor provides, which the market data system vendor will maintain on behalf of the firm. Using a market data system vendor to maintain connections reduces the need for the firm to have teams that update their feeds when data vendors update their interfaces.
- **Price Request rules.** Configurable rules that allows market data to be requested from data vendors, reducing both market data costs and cost of market data ownership. These rules should optimise the efficiency of market data usage by reducing the amount of calls for prices made to the data vendors. Tracking when requests for prices were made and the ability to analyse these requests is key to any solution
- **Price Normalisation rules.** Rules that support transformation of source market data attributes into standardised fields for use in reports and in application logic. Market data vendors have bespoke field names on the feeds they use in their market data feeds. MDS rules allow these to be normalised into standard fields (e.g. Bid, Mid, Ask, High, Low, Traded, Open, Close)

- **Vendor Source Hierarchy.** The Gold copy price generation process should support the use of a vendor source hierarchy (VSH) of price sources at many levels (instrument, instrument type, instrument group). The VSH grid allows firms to create a golden price based on its preferences for market data vendor (e.g. Reuters, Bloomberg, IDC) or price contributor (e.g. ICAP, Tullett, BGC, JPM, HSBC) for each data attribute.
- **Daily Market Data Validation rules.** A market data solution should come with out-of-the-box rules for market data validation. Examples are Bid-Ask Consistency, Zero Check, Calendar Check, Volume Check, Missing Check, Stale Check, No Expired instruments, No Price abnormalities, Day-on-Day Move check, Cross Vendor Check, Relative Tolerances, Absolute Tolerances and Tolerances based on Standard Deviation of Historical Price Moves
- **Time-Series Market Data Validation Rules.** A complete market data system needs to be able to import historical time-series data and validate the entire time-series. Validation rules should include checks for zeroes, gaps and statistical
- **Market-aligned exception management screens.** The screens used for managing market data exceptions should allow for the viewing and exception management of market data using screens that can be configured to align with typical market views of curves.
- **Risk-based exception management.** For derivatives market data management, risk-based views of market data should be available. Teams responsible for market data analysis or addressing exceptions should be able to quickly see at which point along a curve the firm has exposures – making for a more efficient market data management process
- **Proxy and Model price derivation rules.** Where market data needs to be derived, price derivation rules should be available to derive proxy prices or calculate model prices where required

The diagram below gives an overview of a typical market data end-to-end solution. The solution should be modularised and fully customisable, with a design that increases user empowerment and reduces cost of ownership. Different firms have different requirements. Most buy-side firms, for example, will not need Price Filtration (Golden Price generation using tick data sources).



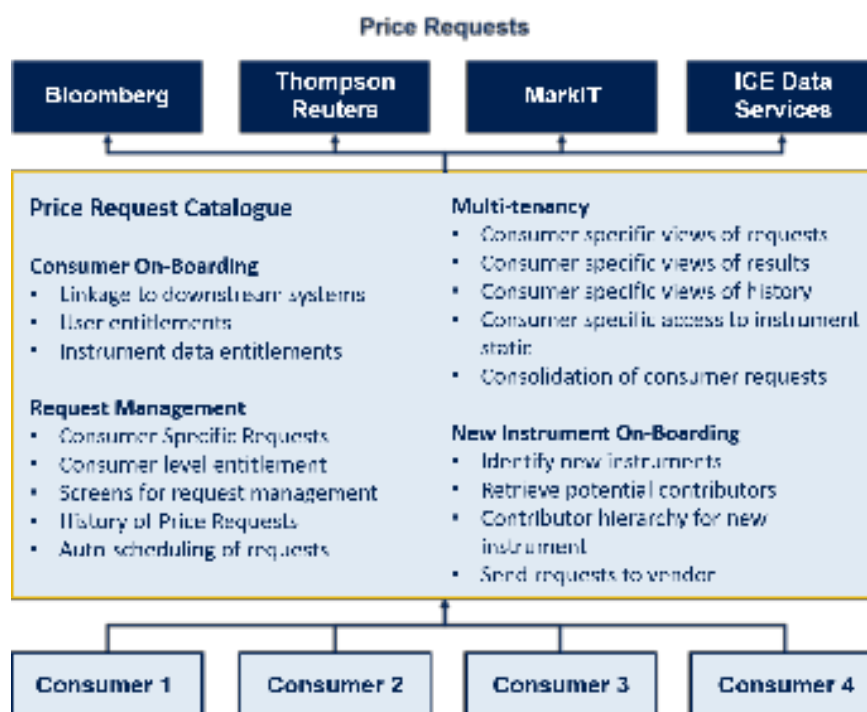
8.5. Price Requests

The key requirements for a robust price request process are

- The optimisation of market data costs
- The ability to on-board new price sources and new instrument types flexibly and quickly
- The automated Price Request hierarchy. The request will move to the next price source in the hierarchy in case of exceptions.
- The integration of the price request process with the late trades process
- If a multi-tenancy solution is required, then
 - Consuming teams need to pay for the market data they use
 - Consumer Price requests should be consolidated so that shared requests are requested only once from market data vendors

The diagram below gives an overview of the basic flow.

- Consumers (teams within the firm) make price requests
- The requests are loaded, consolidated and optimised for vendor market data calls in the Price Request Catalogue
- Where requests for prices for new instruments are made, a new instrument on-boarding is required
- Consumers' data vendor preferences are maintained within the catalogue



8.6. Market Data Collections

The key requirements for a robust market data collection process are

- The accuracy and quality of market data
- The timeliness of loading and issue resolution
- The flexibility of the exception management screens
- The ability to calculate proxy and model prices
- The ability to normalize the data as part of the loading process
- The standardization of price types, e.g. vendor specific price types might be mapped to a standard set such as
 - Bid
 - Mid
 - Ask
 - Open
 - Close
 - High
 - Low
 - Traded
 - Model

The table below contains a typical list of vendor sources that might be loaded as part of the market data collection process

Vendor	Connection	Products
Bloomberg	BO End of Day Pricing	All
	BO Global Commodities	Commodities
	BO Global Commodity Options	Commodity Options
	BO Global Index	
	BO Global Currencies	FX
	BO Equity – MIFID	Equities
	EXT End of Day Pricing data	All
ICE Data Services	APEX - EOD Listed Pricing	Fixed Income
	APEX - EOD Evaluations (Basic Bond Evaluations)	Fixed Income
Thomson Reuters	TR DSE Daily Pricing & Connection, History Service	
	TR DSEI - Bond Pricing Information	Fixed Income
	TR DSS EOD Pricing	All
	TR DSS Intraday Pricing	All
	TR DSS Premium EOD Pricing	All
MarkIT	MarkIT CDS - EOD Composites	CDS
	MarkIT LCDS - EOD Composites	LCDS
Six	SIX VOT E - Valuation Price data	Fixed Income

8.7. Market Data Validations

A market data system needs to be able to validate market data that is loaded into it. The validations need to be instrument level or curve-level depending on the type of validation

A typical set of instrument level validations might be

- Bid-Ask Consistency
- Zero Check
- Calendar Check
- Volume Check
- Missing Check
- Stale Check
- No Expired instruments
- No Price abnormalities
- Day-on-Day Move check
- Weekly or periodic move check
- Dynamic Tolerances (Standard Deviation Based)
- Cross Vendor Difference Checks
- Relative Tolerances (e.g. percentage moves)
- Absolute Tolerances (e.g. basis point)

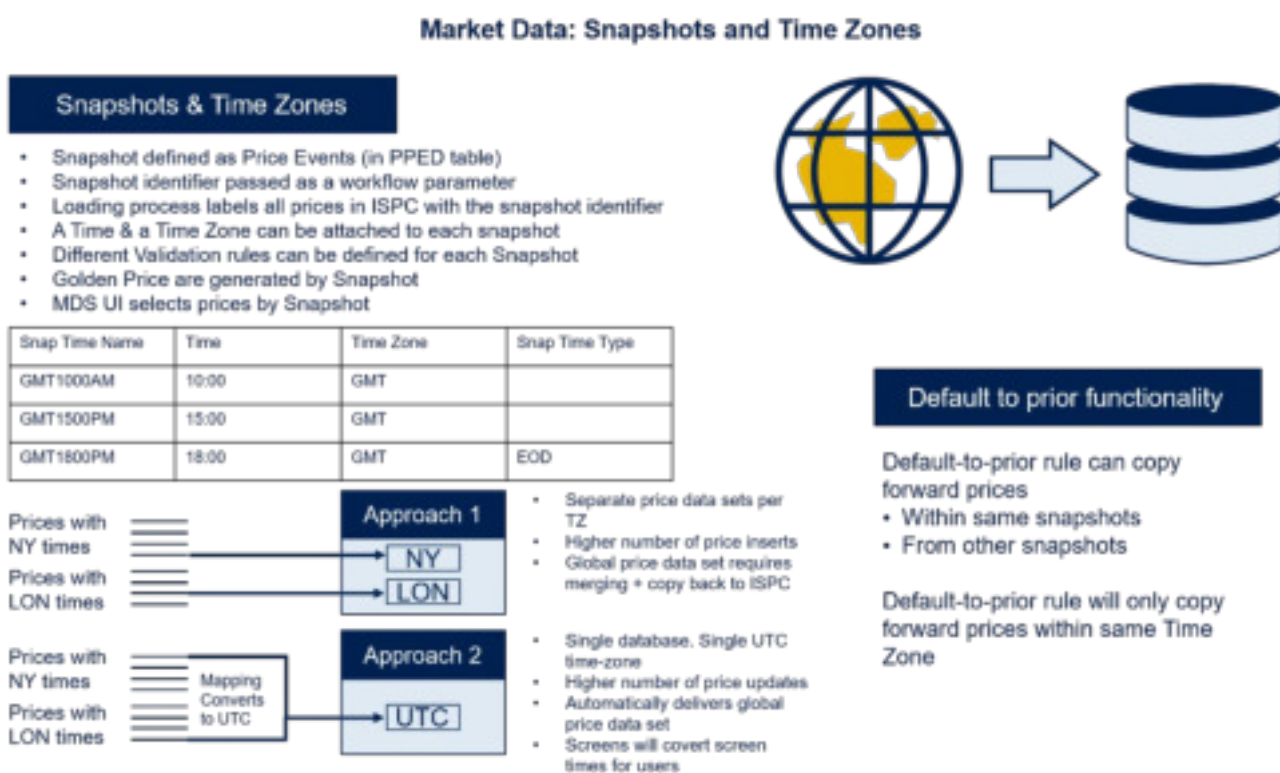
Examples of curve level validations, on the other hand, are

- Curve crossing checks
- Monotonic increasing checks
- Monotonic decreasing checks

It should also be possible to apply the instrument level validations to all instruments on a curve.

Snapshots & Time Zones

Market data needs to be grouped together in specific narrow time-blocks that are often referred to as snapshots. The illustration below highlights some of the concepts



8.8. Validations and Correction Methods

The types of functions that you would expect Market data validation & correction methods to be able to cover are

- Instrument Level validations
- Curve level Validations
- Automatic Correction Methods
- Bulk Correction

Some of the detailed calculations that are possible are:

Parameterizable instrument level validations

- o DoD movement checks (percentage, absolute, relative, standard deviation)
- o Multi-vendor checks
- o Zero checks
- o Missing checks
- o Calendar checks
- o Time-range checks
- o Stale checks
- o Bid-ask checks
- o Black-list check list
- o Price Age
- o Time-series statistical checks
- o Price Level

Parameterizable curve level validations

- o All points valid
- o All points from same contributor

Automatic Correction methods

- o Linear interpolation
- o Copy forward
- o Gap Filling time-series
 - Returns Methods
 - Proxy Approaches

Bulk Correction Methods

- o Suspect-to-Valid
- o Prior Day Price

Manual Correction Methods

- Interpolation
- Copy forward
- Proxy methods
- Choose vendor/contributor

8.9. The Golden Price Level

When implementing a market data system a fundamental design concept is the level at which golden prices can be created. Some sample levels that golden prices can be created are:

- Instrument
- Instrument and currency
- Instrument, currency and market
- Instrument and price type
- Instrument, price
- More granular levels

The most common level that golden prices are created is instrument, currency and market. The diagram below shows tables for sample levels that golden prices can be created at.

Sample Golden Price Levels

GoldenSource Price by Instrument

Golden Records

Instrument	Type	UNIT_CPRC	PRC_TYP	Bid	Mid	Ask	Settle	Open	NAV Price
EUR-DEP-EURIBOR3M-1M	Deposit Rate	-0.34	Bid	-0.34	-0.344	-0.348			
EUR-EURIBOR3M-Future 1	Futures Rate	-0.24	Settle				-0.24		
EUR-EURIBOR3M-2Y	Swap Rate	-0.14	Mid	-0.14	-0.144	-0.148			

GoldenSource Price by Instrument and Price Type

Golden Records

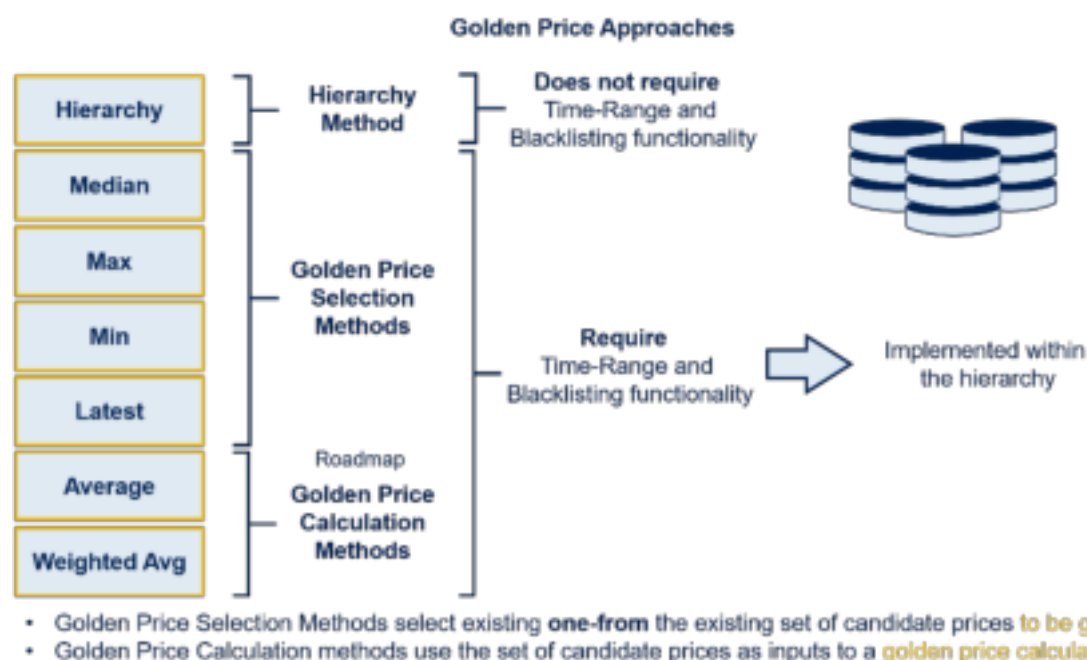
Instrument	Type	UNIT_CPRC	PRC_TYP	Bid	Mid	Ask	Settle	Open	NAV Price
EUR-DEP-EURIBOR3M-1M	Deposit Rate	-0.34	Bid	-0.34	-0.344	-0.348			
EUR-DEP-EURIBOR3M-1M	Deposit Rate	-0.344	Mid	-0.34	-0.344	-0.348			
EUR-DEP-EURIBOR3M-1M	Deposit Rate	-0.348	Ask	-0.34	-0.344	-0.348			
EUR-EURIBOR3M-Future 1	Futures Rate	-0.24	Settle				-0.24		
EUR-EURIBOR3M-2Y	Swap Rate	-0.14	Bid	-0.34	-0.344	-0.348			
EUR-EURIBOR3M-2Y	Swap Rate	-0.144	Mid	-0.34	-0.344	-0.348			
EUR-EURIBOR3M-2Y	Swap Rate	-0.148	Ask	-0.34	-0.344	-0.348			

8.10. Golden Price Approaches

Golden prices can be created using different calculation methods. Typical sample methods are:

- Hierarchy-based methods
- Golden Price Selection methods
- Golden Price Calculation methods

The diagram below illustrates the concept.



Hierarchy and Selection methods always end up using a source price (e.g. a vendor price) as the golden price whereas golden price calculation methods such as average or weighted average use one or more vendor prices as inputs but golden price that's created is not a vendor or source price. It is a calculated price.

8.11. Managing Market Data Exceptions

When market data is loaded into the market data system and validations have been applied, exceptions will be generated. These exceptions need to be managed in a user-friendly exception management and resolution user interface (UI). The UI should

- have the ability to drill from gold copy prices back to the source (vendor) prices that were used to generate them
- have single-click issue resolution
- have the concept of suspect, missing or invalid prices
- have 4-eyes review principles built-in for user edit of prices
- allow users to choose different contributor prices as alternative pricing sources
- allow users to see price histories
- allow users to view exceptions using curves
- allow users to view and resolve exceptions using either bulk exception screens or more visually oriented exception management screens

Example: Approving / Rejecting Market Data Exceptions / Amendments in the Bulk Screen

Bond Name	Face	Index	Maturity	Opt Expiry	Status	Price	Residual
BOND 1	1000000	100	10/1/2025	10/1/2025	100	100	100
BOND 2	1000000	100	10/1/2025	10/1/2025	100	100	100
BOND 3	1000000	100	10/1/2025	10/1/2025	100	100	100
BOND 4	1000000	100	10/1/2025	10/1/2025	100	100	100
BOND 5	1000000	100	10/1/2025	10/1/2025	100	100	100
BOND 6	1000000	100	10/1/2025	10/1/2025	100	100	100
BOND 7	1000000	100	10/1/2025	10/1/2025	100	100	100
BOND 8	1000000	100	10/1/2025	10/1/2025	100	100	100
BOND 9	1000000	100	10/1/2025	10/1/2025	100	100	100
BOND 10	1000000	100	10/1/2025	10/1/2025	100	100	100
BOND 11	1000000	100	10/1/2025	10/1/2025	100	100	100
BOND 12	1000000	100	10/1/2025	10/1/2025	100	100	100
BOND 13	1000000	100	10/1/2025	10/1/2025	100	100	100
BOND 14	1000000	100	10/1/2025	10/1/2025	100	100	100
BOND 15	1000000	100	10/1/2025	10/1/2025	100	100	100
BOND 16	1000000	100	10/1/2025	10/1/2025	100	100	100
BOND 17	1000000	100	10/1/2025	10/1/2025	100	100	100
BOND 18	1000000	100	10/1/2025	10/1/2025	100	100	100
BOND 19	1000000	100	10/1/2025	10/1/2025	100	100	100
BOND 20	1000000	100	10/1/2025	10/1/2025	100	100	100

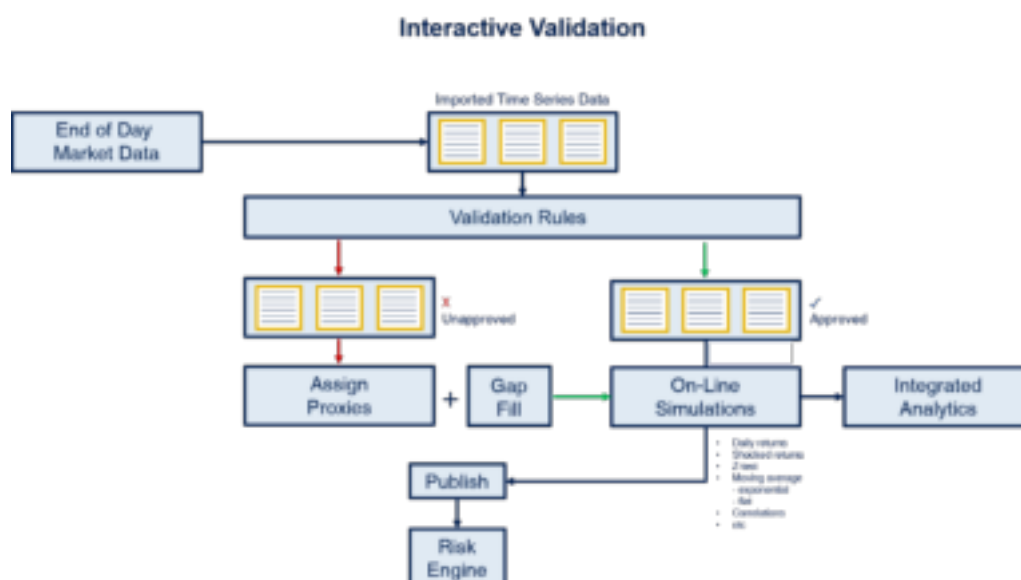
Status	Date	Contributor	Validation Name	Diff To	Diff Amount	Latest	Diff	Spread	BAC	Base Price	Reference Price	Status Flag	Status Defined	Comments
1	10-Dec-2024	Reuters	Calendar Check	10/1/2025	-1.7	4.26284E+024	7.40281E+024	100.0000	100.0000	100.0000	100.0000	0	0	Error: The result is not in the allowable threshold.
2	10-Dec-2024	Reuters	Diff 1/2025	10/1/2025	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0	0	Info: No Calendar for Market found - used default.
3	10-Dec-2024	Reuters	Calendar Check	10/1/2025	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0	0	Info: Default calendar used as base price has no next.

8.12. Managing Time-Series data

Time Series data (historical prices and rates) can be validated in two modes:

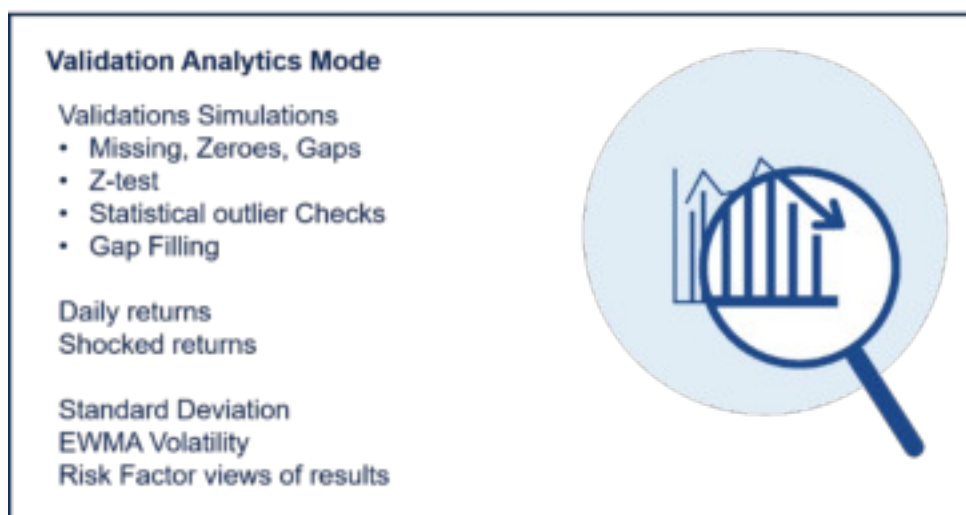
- Automated mode
- Interactive Mode

Automated mode allows for the time-series data to be validated and checked every day.



Whereas running validations in interactive mode assumes that validations are run once (or possibly periodically on an ad-hoc basis) on the historical data using an interactive UI and do not need to be run again after that. Only recently added data needs to be validated after and this can be validated via the (potentially separate) daily market data validation process.

Automated Validation



In both cases (automated and interactive) the types of validation that are run are the same but when validations are run in automated mode specific data sets can be targeted for validation on a pre-determined schedule, e.g. stress period data sets, or periodically to capture statistical outliers, etc.

Typical checks that are required in a time-series validation process:

Basic Checks:

- Zeroes
- Nulls
- Long Stale - the max no. of days a price is unchanged before it is considered stale
- Short Stale - several short stale days within a long period

Statistical Checks:

- Outliers - distance from mean in terms of standard deviation
- Rolling Z-Score - rolling distance from mean in terms of standard deviation
- Auto-Correlation - how correlated is recent data with prior data?
- EWMA Correlation - same as above but more weight given to recent data

If the time-series validations identify errors or gaps in the series, then these errors need to be corrected. Typical gap filling approaches include:

- Copy forward from prior-day
- Linear interpolation
- Gap filling using the returns (price movements) of related (proxied) time-series

9. The Complete EDM Solution

At GoldenSource we believe that we provide the complete Enterprise Data Management solution for buy side firms' data management. Our Security Master (SM) and Market Data Solution (MDS) are best of breed, allowing firms to create gold copy solutions that significantly reduce both the costs of and operational risks associated with the data processing within the firm



Fulfilment of Buy-Side strategic project priorities. GoldenSource will accelerate Buy-Side firms' Market Data Centralization processes through our product ready capabilities



A single platform with flexibility to cater for future changes easily. GoldenSource EDM is a single platform encompassing a modular, multi domain and regulatory ready capability that fulfils all buy side data requirements. Evolved over the last 25 years it is rich in functionality with the flexibility for the user to add new changes easily through configuration or as extensions to the data model.



Commitment to innovation. GoldenSource is solely focused on enterprise data management and at the forefront of the data industry driving best practice. Buy-Side firms will gain from continuous innovations and R&D to our platform that are driven by market trends, technology and our large community of clients.



Accelerate change whilst lowering the cost of such changes. Value for money is demonstrated through cost-effective change, substantial reduction in the cost of regulatory compliance, ability to rationalise and optimise the IT landscape, increased STP and the eradication of bad data and associated cost across the business



Minimise operational impact and delivery risk. By partnering with an organisation who has the most extensive and knowledgeable delivery team in our industry along with the deepest experience in delivering EDM solutions. We have the scale to provide delivery assurance when implementing and supporting such a mission critical solution.