

Notes on An Intro to High-Frequency Finance

2021 – 04 – 17

Junfan Zhu

(junfanz@gatech.edu; junfanzhu@uchicago.edu)

<https://www.linkedin.com/in/junfan-zhu/>

This is a notes composed by Junfan Zhu on the book.

Notes published on <https://github.com/junfanz1/Quant-Books-Notes>.

Book Link

An Introduction to High-Frequency Finance, by Ramazan Gençay, et al.

<https://www.amazon.com/Introduction-High-Frequency-Finance-Ramazan-Gen%C3%A7ay/dp/0122796713>

Table of Contents

- [Notes on An Intro to High-Frequency Finance](#)
- [Junfan Zhu](#)
- [Book Link](#)
- [Table of Contents](#)
- 1. Introduction
- 2. Markets and Data
 - 2.1. Foreign Exchange (FX) Market
 - 2.2. Spot Market
 - 2.3. Futures Market
 - 2.4. Option Markets
 - 2.5. Structure of Foreign Exchange Spot Market
 - 2.6. Contributor Effects
 - 2.7. OTC Interest Rate Markets
 - 2.7.1. Spot Interest Rates
 - 2.7.2. Foreign Exchange Forward Rates

- 2.8. Interest Rate Futures
 - 2.8.1. Implied Forward Interest Rates and Yield Curves
- 2.9. Bond Futures Markets
- 2.10. Commodity Futures
- 2.11. Equity Markets
- 3. Time Series
- 3.1. Variables in Homogeneous Time Series
 - 3.1.1. Interpolatin
 - 3.1.2. Effective Price
 - 3.1.3. Realized Vol
 - 3.1.4. Bid-Ask Spread
 - 3.1.5. Tick Frequency
 - 3.1.6. Other Variables
 - 3.1.7. Overlapping Returns
- 3.2. Covolution Operators
 - 3.2.1. Build-Up Time Interval
 - 3.2.2. Homogeneous Operators and Robustness
 - 3.2.3. Iterated EMA Operator
 - 3.2.4. Differential (Wavelet Transforms)
 - 3.2.5. γ -Derivative
 - 3.2.6. Vol
 - 3.2.7. Windowed Fourier Transform (Complex Moving Average)
- 4. Adaptive Data Cleaning
- 4.1. Using Filter to clean data
- 4.2. Data Errors
 - 4.2.1. Filtering of Single Scalar Quotes: The Level Filter
 - 4.2.2. Pair Filtering: The Credibility of Returns
 - 4.2.3. Expecte Vol
 - 4.2.4. Pair Filtering: Comparing Quote Origins
 - 4.2.5. Full Quote Filtering Window and Quote Splitting
- 4.3. Multivariate Filtering on Sparse Data
- 5. Basic Stylized Facts
- 5.1. Introduction
- 5.2. Price Formation Process
 - 5.2.1. Distributional Properties of Returns
 - 5.2.2. Tail Index of Return Distributions
 - 5.2.3. Extreme Risk
- 5.3. Scaling Laws

- 5.3.1. Market Maker Bias
 - 5.3.2. Limitations of Scaling Laws
- 5.4. Autocorrelation and Seasonality
- 6. Modeling Seasonal Volatility
- 6.1. Modeling Vol Patterns with Alternative Time Scale and Activity Variable
 - 6.1.1. Geographical Seasonality Patterns
- 6.2. ν -Scale
 - 6.2.1. Adjustments of ν -Scale
 - 6.2.2. Filtering Intraday Seasonalities with Wavelets
- 7. Realized Vol Dynamics
 - 7.0.3. Short/Long Memory
- 7.1. Heterogeneous Market
 - 7.1.1. Vol of Different Time Resolutions
 - 7.1.2. Asymmetric Lead-Lag Correlation of Vol
 - 7.1.3. Conditional Predictability
- 8. Vol Processes
- 8.1. Temporal Aggregatin of GARCH
- 8.2. Heterogeneous Vol: HARCH
- 8.3. EMA-HARCH
- 9. Forecasting Returns over multiple time horizons
- 9.1. Intrinsic Time
- 9.2. Linear combination of nonliner indicators
- 9.3. Measuring Forecast Quality
- 10. Correlation & Multivariate Risk
- 10.1. Covolatility Weighting
- 11. Trading Models
- 11.1. Modified Sharing Function for Robust Optimizations
- 11.2. Real-time Testing
- 11.3. Dynamic Hedging with Exposure Constraints

1. Introduction

The utilization of high frequency leads to more robust annualized volatility estimatins by minimizing the influence of the random noise in the market.

We shouldn't seek simplicity at the cost of missing important features of data-generating process. It's useful to explore nonlinear models which contain more parameters, the increasing complexity is strongly penalized when explored with low-frequency data because of the loss of degree of freedom. But for high frequency data, the penalty is relatively small, because the abundance of independently measured observations approximates an asymptotic environment.

High frequency studies can be done for limited sampling periods with large samples. The market properties with such periods are nearly unchanged. The results are less affected by structural breaks or shifts in the economy than low-frequency ones.

1. Aggregation factor of 4 to 5 orders of magnitude.

- Long memory effect in 20-min absolute returns, Dacorogna *et al.* (1993).
- Similar hyperbolic decay of the autocorrelation function was observed on daily returns, Ding *et al.* (1993)
- Difficult to distinguish rigorously in the data between long memory effects and regime shifts. Hyperbolic decay is empirically found at time scales that differ by 2 orders of magnitude in aggregation is a sign that the process must include long range dependence or there are regime shifts at *all time scales*.

2. Scaling properties and scaling laws.

- Scaling laws of different moments of return distributions. Challenge is to develop models that simultaneously characterize the short-term and long-term behaviors of a time series.

2. Markets and Data

2.1. Foreign Exchange (FX) Market

FX market has the highest market volume of all financial markets. A large part of volume is traded OTC between banks, but there is electronic trading through centralized system as well. The coexistence of interbank and centralized trading implies the volume figures will refer to a market segment rather than all transactions of the whole market.

Data are available over long sampling periods in high frequency 24 hours a day. The market is high liquid and symmetric as both exchanged assets are currencies. But daily/weekly data don't hold up in intraday analysis. In intraday data, the homogeneity of market agents disappears. There are interaction of market

agents with heterogeneous objectives due to different geographical locations, the various forms of institutional constraints, and risk profiles.

The heterogeneous structure of intraday data can explain that practitioners like to do technical analysis, which is using the interaction of different components of markets.

ETF FX market volume is much lower than FX spot market, and OTC derivatives (FX forward) market.

2.2. Spot Market

Direct markets for primary assets, like FX, equity, they are traded immediately at the time of transaction.

Cons: The timing is not flexible, traders have to deal with physical delivery of the traded assets (like commodities), and interest rate spot market is affected by counterparty default risk. So, the derivatives markets are more important than spot markets in some cases. FX market is a market when spot trading is strong.

2.3. Futures Market

Higher liquidity and volume than underlying spot markets. More high frequency data.

All clients buying/selling futures contracts have to put money in a collateral account, which covers the counterparty default risk of the exchange. If future prices move to the disfavor of a client, and the money on the collateral account may not cover the risk, and the exchange will tell client to increase it through *call for margin*.

Behavior of a contract changes when approaching expiry, and price volatility systematically grows/shrinks due to the nature of underlying asset. So it's necessary to construct long samples joining several contracts together. Traders will join price histories of several future contracts with successive expiries, this prescription are based on rollover schemes: they replicate the behavior of a trader holding a contract and switching (*roll over*) to the next contract before the expiry of current contract.

The opening hours of futures markets are sometimes modified, for centralized markets. Long samples may extend over periods with different fixed opening hours. This opening hours leads to difficulties in intraday studies.

2.4. Option Markets

Implied vol of at-the-money ooptions (strike price is not far from base spot price) are slowly changing over time by Black Scholes. This is market forecast of vol of underlying asset for the time period until expiry. So the time series of implied vol are interesting especially compared to historical or realized vol computed from time series of underlying.

2.5. Structure of Foreign Exchange Spot Market

FX market = FX Spot + FX Forward + (small) FX derivatives.

The first one of a currency pair is exchanged currency, whose value is expressed in the second currency (numeraire currency).

Collecting tick-by-tick quotes has problems, like transmission delays, breakdowns, or aberrant quotes due to errors, so it's important to do data cleaning filter to eliminate outliers.

FX rates between 2 currencies other than USD are *cross rates*. A trader calling market makers to execute an indirect transaction has to pay the bid-ask spreads of 2 markets, both USA-CAD and USD-JPY (to trade USD-JPY), both bid-ask spreads may be higher than ¹ the bid-ask spread of a direct transaction. So we need to compute synthetic cross rates (proxies for unknown direct rates). The 2 ticks used for the cross rate computation should be synchronous so that their time stamps deviate by no more than a few seconds. Otherwise, the synthetic cross rate is distorted by price moves in the time interval between 2 ticks. The bid-ask spreads of missing direct quotes can be lower than synthetic bid-ask spreads.

2.6. Contributor Effects

Data suppliers mix data with quotes of contributors, thus creating a multicontributor data feed.

- Depending on inventory position, market makers have preferences for selling/buying. They publish new quotes to attract traders to make a deal. The other price of bid-ask pair is pushed away to less attractive region by adding/subtracting a large, nominal spread. This leads to high unrealistic quoted bid-ask spreads and a negative autocorrelation of returns at lags of 1 minute.
- FX quotes for some contributors lag behind 1 minute of real market prices. There's *lead-lag* correlation analysis of returns time series.

¹Direct forward quotes exist for most FX rates against USD, but may not be available for some FX cross rates. For these currency pairs, synthetic forward rates are needed.

- Some contributors are laggards, they published prices copied from quotes of other contributors, whereas true prices are negotiated over phone. They use computers to publish fake quotes at high frequency, this data-copying methods lead to low data quality.

2.7. OTC Interest Rate Markets

2.7.1. Spot Interest Rates

A bank with low credit rating has to pay higher IR to attract a lender, IR level is increased by *credit spread*, so IR quote is less universally applicable than FX quot.

Credit spread can lead to series data problems, like 1990s Japanese interest rates market is split to 2, low-rating banks and high-rating banks, due to the daytime Asia time zones issue:

- Too high absolute values of returns (level changes) over time intervals of around 12 hr
- Strong negative autocorrelation of returns at lags of around 12hr

These are simply caused by periodically shifting credit ratings, which can be avoided by eliminating all low-rating (or high-rating) quotes from the sample. JPY interest rates were below 1%, the credit spread led to $IR < 0$ for non-Japanese banks (dominating Euro and US time zones).

Spot interest rates is a deposit of fixed duration, the maturity period.

2.7.2. Foreign Exchange Forward Rates

Due to delayed transaction, the buyer of FX forward contract earns some interest on the base currency of FX rate instead of exchanged currency. If interest rates of 2 currencies deviate, there is net interest payment flow from/to buyer during maturity period. This determines the price of FX *outright* forward rate. To avoid riskfree arbitrage, the outright forward rate deviates from simultaneously quoted FX spot price, by an amount to offset the deviations in interest payment. The price deviation between FX spot prices and outright forward rates reflects the interest rate *differential* between 2 exchanged currencies, rather than absolute level of IRs. (See formula 2.2)

Instead of outright FX forward prices f , the difference (outright forward - simultaneously valid FX spot price) $f - p$ is quoted, which is less volatile than outright forward price. $f - p > 0$: forward premium. $f - p < 0$: forward discount. Both are called forward points. Together are called FX swap rates. FX forward transaction = spot transaction + swap transaction (swapping 2 currencies during maturity)

2.8. Interest Rate Futures

IR futures markets with expiry periods of up to 1 year are more liquid than OTC spot IR market. The transaction costs are lower, a typical bid-ask spread is 10% of the quoted spread of cash IR. So IR futures markets are high-quality intraday data.²

IR futures contract has non-stationary effect, there is a systematic decrease of mean vol when moving closer to expiry. Rollover is not the most suitable method for IR futures, because several contracts with different expiries trade at the same time with comparable liquidity, unlike what happens in bond futures and other future markets, where only 1 or 2 contract are traded at any given time.

IR futures has inherently multivariate problem, such that several contracts with different expiries but simultaneous high open interest levels can't be neglected. We need to infer *implied IR* from prices of futures.

2.8.1. Implied Forward Interest Rates and Yield Curves

Implied IR is better than IR futures prices, as they can be studied in long time series. The behavior of returns is close to stationary, though there may be some local vol peaks before expiry. Both 2 forms need to be computed³:

- Forward IR: the interest rate for a period of 3 months, with starting point shifted to the future by a fixed time interval
- Yield curve: spot interest rates for periods starting now. It's full term structure of interest rates of different maturities.

To use info from IR futures to construct a yield curve / forward rates. (omitted a lot of formula for interpolations here...) A consequence of the polynomial interpolation is the potential overshooting of yield curve. If forward rates implied by a series of IR futures have a maximum around medium-term expiries, an interpolated forward rate may exceed all the implied forward rates corresponding to original future quotes. The modest overshooting of polynomial interpolation isn't undesirable as it leads to strong reduction of 3-month seasonalities obtained for a forward ID series generated by linear interpolation (no overshooting). So, polynomials with smooth but sometimes overshooting behavior represent true yield curves better than piecewise linear interpolation or spline with hard corners at nodes, so that even very distant contracts have influence on local behavior.

²The bid-ask spread on CME Euro contract is small, minimum price movement = half a basis point = 1/200 of 1%.

³Futures prices alone are not enough to construct implied spot rates (points on yield curve), as IR futures market doesn't convey any info about applicable spot rate for the period from current time to next futures expiry. By studying forward IRs, we can avoid the need to use data from other instruments (spot IRs).

Here, overshooting is a local effect, distant contracts can't affect behavior of $\rho(T)$.

There may be undershooting of forward IR, where some parts of $\rho(T)$ may be below 0. The method should include an element to avoid, although interest rates may be slightly below 0 for some circumstances.

The polynomial method relies on regular quarterly sequence of expiries, it can be adapted to include irregular contracts based on serial months. The period from current time to the first expiry can't be covered by IR futures, we need spot IRs to fill the gap⁴. After filling the gap, the method compute implied forward interest rates and can be used to compute implied spot rates ($t=0$) and imply the current time. Yield curve consists of a set of implied spot rates with different $t=end$. There is arbitrage opportunities for us to compare the yield curve derived by IR futures at any time with curve derived from other instruments (deposits, OTC forward rate agreements).

Convexity corrections of yield curve shows difference between futures and forward contracts, due to margining arrangements for futures (collateral accounts), but for this case we can ignore the small value, because futures contracts are never more than 18 months from expiry.

2.9. Bond Futures Markets

Bonds are long-term interest rates. Bond futures are standardized, just like short-term IR futures, they are liquid, high frequency, high quality intraday data.

Rollover scheme can create a continuous time series with different expiries futures contracts, they are good for transaction cost analysis, different optimization goals lead to different rollover schemes. Rollover must follow economic constraint imposing that the value of portfolio changes only when the market prices of individual bond futures move, with no change arising solely from rollover procedure. When rolling over, the number of new contracts to be bought is calculated, so that total amount invested is constant, and number of new contracts is given by number of old contracts being sold multiplied by middle price, divided by middle price of new contracts being bought.

- Scheme 1: use conversion factor to render continuous the transition from one contract to the next one, at a fixed date before the first contract's expiry.

The timing of rollover determines the character of obtained continuous time series. Simple rollover, α factor (see formula 2.10) varies over time, often slower than prices. Rollover time is chosen when both contracts are liquid. So we take

⁴The IR swap rates is another source of high frequency IR data.

a mean α from daytimes where simultaneous quotes of both contracts are found, rather than an α determined at only one fixed daytime. Balocchi and Hopman (1997) says the discontinuities of α factors by asynchronicities in the underlying bonds, are called *cheapest-to-deliver* (or benchmark bonds), which is helpful to analyze continuous series. If underlying bonds of 2 successive contracts don't change exactly at the same time, the α factor is affected by difference.

- Scheme 2: construct a bond futures portfolio with const mean time-to-expiry, though a daily partial rollover, whereas the constituent bond futures have a fixed calendar data expiry.

Based on const-time-to-expiry, to have a portfolio doesn't expire at a fixed date, but keeps const mean time-to-expiry as time moves on, by means of an appropriate daily rollover. This rollover doesn't need to be executed so high transaction costs are avoided. The time-to-expiry h of a portfolio is defined as weighted average of time-to-expiry of constituents. (See formula 2.11) Each day we arrange partial rollover, selling a proportion of holdings in the first expiry and buying the second one to keep h const. We can use volume-dependent timing of rollover (roll over when volume of a new contract overtakes that of an old contract).

2.10. Commodity Futures

Most commodity futures traders offset their contracts (or roll over, which can build long time series from different contracts) before expiry, in some market so early that the second position (contract with the second next expiry) has a higher liquidity than the first position. Purpose of commodity futures trading is hedging and portfolio diversification.

Futures of agricultural commodities have irregular schedule of expiry dates due to seasonality, like New York cocoa futures has expiry dates on Mar, May, Jul, Sept, Dec.

Contracts with different expiry dates aren't independent, similar to other futures markets. A distant expiry contract can't be much more expensive than a near contract, similar to forward IR can't be far below 0.

Commodity futures aren't liquid enough for huge transactions, large orders will cause considerable *slippage* with immediate price movements.

2.11. Equity Markets

1. Equity of *individual companies* traded by stock exchanges. Stock splits and dividend payments affect equity prices.

2. Equity *indices* are weighted (adapte from time to time) sums of individual equity prices, a basket including important equities of industry sectors. They often show positive autocorrelation of returns at a lag of 15 minutes. So there is lag structure between leading main equities and less liquid equities of the basket.
3. Equity (index) *futures*, liquid, high quality, high frequency.
4. *Options* implied vol.

3. Time Series

When considering spacing of data in time, we should discuss time scale.

1. Homogeneous: regularly spaced time series. Most time series analysis.
 - Use regularizing operator to do time series of same variable. Sampling an inhomogeneous time series at regular time intervals for time series analysis
 - Microscopic operator depends on actual sampling of inhomogeneous time series. Eliminating random ticks lead to different results.
2. Inhomogeneous: irregularly spaced time series. It's different from missing observations (homogeneous + few gaps). Use Operator Ω , do time series $\Omega[z]$, new variable on same time points.
 - Compute a new variable from initial variable while keeping initial inhomogeneous time points. *Ex*: computing a series of local vol from initial price series.
 - Macroscopic operators extract *average* behavior of time series. They are immune to small variations of individual ticks, including adding/eliminating few ticks. They have well-defined limit when price quotes become infinitely dense. If price quotes are sufficiently dense inside the range of operator, we are close enough to its limit. For inhomogeneous time series, macroscopic operators are robust ⁵. The archetype of macroscopic operator is exponential moving average (EMA), with decaying weight on the past.

3.1. Variables in Homogeneous Time Series

3.1.1. Interpolatin

Construct homogeneous time series with linear interpolatin, see figure 3.2.

⁵For homogeneous ones, macroscopic are unnecessary because sampling frequency is fixed and we don't need to take continuous-time limit or add/remove ticks. Moreover, homogeneous time series rely on backward shift operator \mathcal{B} , which is microscopic. See Zumbach and Muller (2001).

3.1.2. Effective Price

In FX market, we use logarithmic middle price x at time t_j :

$$x(t_j) = \frac{\log p_{bid}(t_j) + \log p_{ask}(t_j)}{2} \quad (3.1)$$

, due to antisymmetric property: if x is USD-JPY price, then JPY-USD is $-x$, and they are the same market. Another advantage of logarithmic is to make returns (difference of x) dimensionless, which is independent of original units in which price is measured.

In FX spot market, bid/ask prices are just indicative quotes produced by market makers who are interested in either bid or ask; the other price are dummy value. This leads to small error for this formula, and often the bid-ask spread is often smaller than the real spread. Due to transmission delays, like market maker B enters a quote after market maker A, but the quote of market maker B is the first to appear on multi-contributor data feed. That's why we use effective price, with best bid/ask in a time window of the size of a quote lifetime (2 mins). And also, we can eliminate negative first-order autocorrelation of returns at high frequencies.

3.1.3. Realized Vol

$$v(t_i) = v(\Delta t, n, p; t_i) = \left[\frac{1}{n} \sum_{j=1}^n |r(\Delta t; t_{i-n+j})|^p \right]^{1/p} \quad (3.2)$$

where n = number of return (price changes) observations. 2 time intervals: return interval Δt (= 10 min) and size of total sample $n\Delta t$. If $p = 2$, v^2 is variance of returns ≈ 0 .

For p choosing, should $<$ tail index of distribution (3.5 for high frequency FX data, 4-th moment of return distribution often diverges.) p can be limited to *half* of tail index.

By Gaussian scaling law, $v^2 \propto \Delta t$, scaled vol = $v_{scaled} = \sqrt{\frac{\Delta t_{scale}}{\Delta t}} v$. If $\Delta t_{scale} = 1$ year, then it's annualized vol. Typical annualized vol for FX rates are around 10%.

So we have 3 time intervals:

- time interval of return observations Δt (if chosen too small, may also be biased, 15min - 2hr is OK)
- sample size $n\Delta t$ (n=250 working days)
- scaling interval Δt_{scale} (1 year: annualized vol)

Three volatilities:

- Realized vol = historical vol: determined by past observations, like formula (3.1). Based on homogeneous series of returns. Can also do interval overlapping for more precise realized vol. Coarse realized vol with large Δt predicts the value of fine vol (small Δt). This lead-lag effect indicates that the dynamics of vol are complex, and realized vol with one choice of Δt isn't a perfect substitute for realized vol with another value of Δt . For low-frequency (daily) data, GARCH has better estimation of vol than realized vol. For high-frequency (intraday), realized vol is better.
- Model vol = virtual variable in theoretical model like GARCH or stochastic vol.
- Implied vol = vol forecast computed from market prices of options, based on e.g. Black Scholes.

3.1.4. Bid-Ask Spread

Bid-ask spread reflects the transaction and inventory costs and risk of institution that quotes the price. For traders, the spread is the only source of costs as intraday credit lines on FX market are free of interest. So, the spread is a measure of friction between market participants and market efficiency.

Relative spread is dimensionless, while nominal spread is units of underlying price. log spread (log of (3.3)) is much less skewed and closer to symmetric than distribution of s .

$$s(t_j) = \log p_{ask}(t_j) - \log p_{bid}(t_j) \quad (3.3)$$

For different market makers, $s(t_j)$ is affected by individual preferences of market makers, so a homogeneous time series of spreads $s(t_j)$ generated by interpolation has high noise. So we need to compute *average* spreads within time windows and build homogeneous time series of average spreads.

3.1.5. Tick Frequency

log tick frequency. Tick frequency is a proxy for transaction volume on the market. But it's not good to equate tick frequency to transaction volume or use it as proxy for both volume and strength of bank. Because:

- Although it takes a few secs to enter price quotation in terminal, if 2 market makers happen to simultaneously enter quotes, only one quote may appear on data collector's screen.
- During high volume trading, some operators are too busy to enter quote into system

- Bank can use automatic system to publish price to advertise itself on market
- Representation of banks depends on coverage of market by data vendors like Reuters. The coverage is changing and doesn't represent the whole market. (Asian market makers aren't as well covered by Reuters as Euro's.)
- Banks have many subsidiaries, they may use one subsidiary to quote prices made by a market maker in another subsidiary on another continent. Quotes from differently reliable and renowned sources have different impacts on the market.

3.1.6. Other Variables

- Realized skewness of the return distribution.
- Vol ratio (2 vols of different time resolutions) can detect trending behavior.
=1: Brownian motion; higher: follows a trend; lower: mean-reverting noise.
- Direction change indicator = # of trend reversals within a time interval.

3.1.7. Overlapping Returns

Panel regression can apply to overlapping returns, where they're grouped in several nonoverlapping series with phase-shifted starting points, and overlapping scheme is phase-shifted intervals on two lower time axes added (See Figure 3.3). Overlapping is a way to reduce stochastic error of realized vol, but can't reduce error of empirical mean. Heavy tailed distribution, serial dependence, heteroskedasticity and choice of p may affect the behavior of stochastic error.

3.2. Covolution Operators

- EMA operator, can compute efficiently with big data. A convolution is an integral, interpolation is needed to define the convolution integral. Our macroscopic operators are composed of iterated moving averages, all EMA operators have noncompact kernels where kernels decay exponentially. But this implies infinite memory, so a build-up must be done over initialization period before the error of operator value become negligible.
- For tick data, values as well as time points of series are both stochastic, so pointwise values are not useful, interval average is more important. We define daily return = difference between average price of last few hours and average price from yesterday. So, we can build smooth variables well suited to stochastic process.
- We want smooth operator with smooth kernels. There is a trade-off between fast reaction with more noise, or a smooth average with slow reaction time.

n -th moment of causal kernel w is $\langle t^n \rangle_w := \int_0^{\inf} dt w(t) t^n$. Also, range R and width w are defined.

3.2.1. Build-Up Time Interval

Most kernels have exponential tail for large t , which means when starting evaluation of operator at time T , a build-up time interval must be elapsed before result of evaluation is accurate enough (influence of initial error at T has sufficiently faded). Assume process $z(t)$ is known, operator Ω :

$$\Omega[-T; z](t) = \int_{-T}^t dt' w(t-t') z(t') \quad (3.4)$$

For $-T < 0$, average build-up error ϵ at $t = 0$ is defined as (omit a lot of steps):

$$\epsilon^2 = E[(\Omega[-T; z](0) - \Omega[-\infty; z](0))^2] = 2\sigma \frac{\tau}{1y} \int_{\bar{T}}^{\infty} dt \tilde{w}(t) \int_{\bar{T}}^t dt' \tilde{w}(t')(t' - \bar{T}) \quad (3.5)$$

The heavier the tail of kernel, the longer the required build-up. Build-up interval should be computed numerically, and can be solved analytically for simple EMA kernel. The build-up time interval is large for small error tolerance and for process with high vol.

To measure tail, we can use first 2 moments off kernel, defined as aspect ratio $AR[\Omega] = \langle t^2 \rangle_w / \langle t \rangle_w = 2/\sqrt{3}$ for rectangular kernel and $\sqrt{2}$ for simple EMA. Low aspect ratio means that the kernel of operator has short tail, so a short build-up time interval is good. For nonnegative causal kernels, the aspect ratio is less useful for choosing build-up interval of causal kernels with more complicated, partially negative shapes.

3.2.2. Homogeneous Operators and Robustness

Most nonlinear operators are homogeneous of degree p , $\Omega[ax] = |a|^p \Omega[x]$, nonlinear operators can build robust estimators.

Translation-invariant homogeneous operators of degree pq take form of convolution.

$$\Omega[z](t) = \left[\int_{-\infty}^t dt' w(t-t') |z(t')|^p \right]^q \quad (3.6)$$

For moving norm, $w = \text{average}$, $q = 1/p$.

Nonlinear operators can build robust estimators. Data errors (outliers) should be eliminated by data filter prior to any computation. Prior data cleaning, robust estimators can reduce dependency of results on outliers or the choice of data cleaning algorithm. This problem is acute mainly when working with returns, because the difference operator needed to compute returns from prices is sensitive to outliers. So the modified operator achieves robustness by giving a higher weight to the center of the distribution of returns than to the tails.

Robust operator mapping function $sign(x)|x|^\gamma = x|x|^{\gamma-1}$ has exponent $0 \leq \gamma < 1$. For vol estimates, usual L^2 vol operator based on squared returns can be made more robust by using mapping function: the signed square root $f = sign(x)\sqrt{|x|}$, which transforms L^2 vol into L^{2p} vol. This power law transformation includes in the definition of moving norm, moving variance or vol operators.

3.2.3. Iterated EMA Operator

EMA Decaying kernel $ema(t) = \frac{e^{-t/\tau}}{\tau}$.

Technical analysis is simple iterated EMA operators to homogeneous time series for long-term.

Kernel of $EMA[\tau, n](t) = \frac{1}{(n-1)!} \left(\frac{t}{\tau}\right)^{n-1} \frac{e^{-t/\tau}}{\tau}$. This is Laguerre polynomials, which are orthogonal w.r.t. measure e^{-t} for $\tau = 1$. Any kernel can be expressed as sum of iterated EMA kernels. So, the convolution with an arbitrary kernel can be evaluated by iterated exponential moving averages. But the convergence is slow. For different τ, n choosing, see *Figure 3.5*.

3.2.4. Differential (Wavelet Transforms)

Low-noise differential operator suitable to stochastic process should compute an *average differential*, the difference between an average around time now over a time interval τ_1 and an average around time $now - \tau$ on a time interval τ_2 . Kernel of similar kind are used for wavelet transforms.

Our point of view is different from continuous-time stochastic analysis. In continuous time, $\tau \rightarrow 0$, leading to Ito derivative with subtleties. Here, we keep the range τ finite to analyze process at different time scales (different orders of magnitudes of τ). Moreover, the $\tau \rightarrow 0$ is not true in financial data, because a process is known only on a discrete set of time points and doesn't exist in continuous time.

Differential operator $\Delta[\tau] = \gamma(EMA[\alpha\tau, 1] + EMA[\alpha\tau, 2] - 2EMA[\alpha\beta\tau, 4])$, see figure 3.8.

3.2.5. γ -Derivative

γ -derivative $D[\tau, \gamma] = \frac{\Delta[\tau]}{(\tau/1y)^\gamma}$. When $\gamma = 0$, differential, $\gamma = 0.5$, stochastic diffusion process, $\gamma = 1$, derivative.

3.2.6. Vol

Vol is a measure to quantify the size and intensity of stochastic process movements, it's the width of probability distribution of the process increment.

Realized vol based on regularized data has drawbacks:

- Inhomogeneous time series, a synthetic regular time series must be created, which involves interpolation scheme.
- Difference is computed with pointwise difference, which implies noise in stochastic data.
- Only some values at regular time points are used. Info from their points of series between regular sampling points is thrown away. Due to information loss, the estimator is less accurate.
- It's based on rectangular weighting kernel, that all points have const weight of $1/n$ or 0, as soon as they're excluded from sample. Continuous kernel with declining weights is better with less noisy.
- By squaring returns, the realized vol puts a large weight on large changes, so it increase the impact of outliers and tails. As 4-th moment of prob distribution of returns might not be finite, the vol-of-vol might not be finite. So this estimator is not robust. So we prefer realized vol defined in L^1 norm.

A soft kernel will lead to lower mean value of vol than hard kernel (positive/negative parts are close to delta functions). In risk, we are interested in conditional daily vol, given today prices, we want to estimate/forecast the size of price move to tomorrow (vol within a small sample of only one day). The actual value of vol can be measured one day later.

3.2.7. Windowed Fourier Transform (Complex Moving Average)

For time series and vol at different time scales, we want a double representation in time and frequency, but we don't need an invertible transformation, because we need to analyze rather than further process the signal. This gives us flexibility in choosing transformations. We want a couple of oscillations in window 2τ , large k increase frequency resolution at the cost of time resolution. We compute EMA with complex τ , which is equivalent to include sine and cosine in the kernel. Because this has nice computational iterative property of moving average. Kernel of complex ema: $ema[\zeta](t) = \frac{e^{-\zeta t}}{\tau}$, where $\zeta = \frac{1}{\tau}(1 + ik)$ is complex. By using

convolution formula, we can have iterative kernel of complex $EMA[\zeta, n; c] = \frac{c}{(1+ik)^n}$.

Windowed Fourier transforms can be computed for a set of different τ values to obtain a full spectrum. But there's an upper limit in the range of computable frequencies. Results are reliable if $\tau >$ average time interval between ticks; but for $\tau <$ average tick interval, results are biased and noisy, which also applies to other time series operators.

4. Adaptive Data Cleaning

4.1. Using Filter to clean data

There's a less favorable alternative to prior data filtering: robust statistics, where all data and outliers are included. Cleaning high frequency data has problems in:

- Variety of possible errors and causes
- Variety of statistical properties of filtered variables (distribution functions, conditional behavior, structural breaks)
- Variety of data sources and contributors of different reliability
- Irregularity of time intervals (sparse/dense data, data gaps of long duration)
- Complexity of quoted data: transaction prices, indicative prices, FX forward points (negative value allowed), IR, figure from derivatives markets, transaction volumes, bid-ask quotes vs. single-valued quotes
- Real-time filtering

Data Cleaning Algorithm (Adaptive) and Filtering methodology:

- Cause of data errors is rarely known, so validity of a quote is judged according to plausibility given statistical properties of series.
- Filtering window, to judge credibility of quote, can grow and shrink according to data quality.
- Quotes with complex structure (bid-ask) are split into scalar variables to be filtered separately. Filtered variables are derived from raw variables (logarithm of bid-ask). Special error occurs for full quotes before data splitting.
- Avoid numerical methods with convergence problems (model estimation, nonlinear minimization)
- Computationally fast, so the algo is sequential and iterative, use existing filter info base when a new tick arrives, with a minimum amount of updating.

4.2. Data Errors

The filter needs a build-up period to learn from the data, so it's adaptive filter. Raw data \rightarrow filter \rightarrow computational block \rightarrow filter \rightarrow application. Some computational blocks like cross rate, yield curve computations require filtered input and produce an output that user want to filter again.

4.2.1. Filtering of Single Scalar Quotes: The Level Filter

Comparison between quotes. This applies to volatile but mean-reverting time series. The timing of mean reversion is fast. For IR, IR futures prices, level filtering isn't suitable, because they are mean-reverting only after time intervals of years, and they're freely floating within smaller intervals.

Level filter can be used on bid-ask spread, which are volatile tick by tick but tends to stay in a range. Need to consider properties of bid-ask spread when doing so:

- Quoted bid-ask spread tends to cluster at *even* values (e.g., 10 basis points), but real spread may be odd value oscillating in a range below the quoted value. A series of formal const spreads can hide some substantial vol that's not covered by the statistically determined denominator (in formula 4.3). We need an offset to account for the typical hidden vl in the denominator.
- High values of bid-ask spreads are worse in usability than low spreads. So the quote deviations from the mean are judged with bias. Deviation to high side are penalized by a factor, but no such penalty is applied against low spreads.
- For minor financial instruments, many quotes are posted with 0 spreads, but we penalize them.

4.2.2. Pair Filtering: The Credibility of Returns

Pair filtering's change filter judges the credibility of a variable change. Distributions of high-frequency price changes are fat-tailed with outlier prob $\sim \xi^{-\alpha}$, α is tail index. Determining the distribution in moving sample is not our filter's job here, so we make rough assumption on α that's good across many rates and filtered variable types.

Notes:

1. Filtering should stay a local concept on time axis. But a quote has few close neighbors and distant neighbors, so when the additive trust capital of a quote is determined by pairwise comparisons to other quotes, the results from distant quotes must not dominate those from close neighbors, the interaction range should be limited. So we define trust capital $\propto (\Delta\nu)^{-3}$ for asymptotically large quote intervals $\Delta\nu$.

2. For large $\Delta\nu$, even moderately aberrant quotes are easily accepted. So, the decline of trust capital with growing $\Delta\nu$ is important. But for negative trust capital, should stay negative even if $\Delta\nu$ is large. So there is a selective decline of trust capital with increasing $\Delta\nu$: fast for small ξ (positive trust capital), slow for large ξ (negative trust capital). It's important for data holes or gaps, where there're no close neighbor quotes.

Trust capital from change filter.

$$T_{ij} = T(\xi_{ij}^2, \Delta\nu_{ij}, I_{ij}) = I_{ij}^* \frac{1 - \xi_{ij}^4}{1 + \xi_{ij}^2 + (\frac{d\Delta\nu_{ij}}{\nu})^3} \quad (4.1)$$

where

$$I_{ij}^* = \begin{cases} I_{ij} & \text{if } \xi_{ij}^2 < 1 \\ 1 & \text{if } \xi_{ij}^2 \geq 1 \end{cases}$$

4.2.3. Expecte Vol

An annualized squared micro-vol is defined as variance in the form of moving average. Quotes with coarse granularity have a *hidden* vl such that a series of identical quotes may hide a movement of a size smaller than the typical granule.

4.2.4. Pair Filtering: Comparing Quote Origins

Pair filtering can add credibility to 2 quotes only if these are independent. 2 identical quotes from the same contributr don't add confidence to the quoted level, but 2 nonidentical quotes from the same contributor may imply that the second quote has been produced to correct a bad first one.

4.2.5. Full Quote Filtering Window and Quote Splitting

The full quotes may enter a full-quote filtering window in a form already corrected by filtering hypothesis, but the algorithm of full-quote window doesn't care about quote corrections (it's higher level). So the most important task of full-quote filtering window is quote splitting. Bid-ask quotes are split into 3 scalar quotes: bid quote, ask quote, and bid-ask spread. This is not trivial, because quote splitting is coupled with 2 other operations: basic validity testing and math transformations.

4.3. Multivariate Filtering on Sparse Data

In risk, there is covariance matrix to keep track of correlations between financial instruments, and they can be applied to data cleaning in sparse quotes. Although univariate filtering works well for dense quotes, when new quote of sparse series comes in, there are only few quotes to compare, and these quotes are old, so they're not ideal for filtering. So we need more info from cov matrix, which is artificial quote method. If sparse rate (middle price) is included in cov matrix, and also covers denser rates, we can generate artificial quotes of sparse series by exploiting the most recent quotes from denser series and cov matrix. Expectation Maximization (EM) algorithm can produce such artificial quotes.

Uncertainties in artificial quotes:

- Stochastic error in value, because they are estimated
- Uncertainty in time, due to asynchronicities in the quotes of different financial instruments
- Only a part of full quote is estimated from cov matrix (e.g., middle price, but the bid-ask spread has to be coarsely estimated as an average of past values).

We can use arbitrage conditions to construct artificial quote, such as triangular arbitrage of FX cross rates. Steps:

- Define a basket of high frequency time series, which are well correlated or anticorrelated to sparse series
- Generate artificial quotes from correlation matrix, mix them with normal quotes of sparse series, thus reinforcing the power of univariate filtering algorithm
- Eliminate artificial quotes from final output of filter (because a filter isn't a gap-filter)

5. Basic Stylized Facts

5.1. Introduction

1. At high frequency, middle price is subject to microstructure effects (bouncing of prices between bid and ask levels). The price formation process overshadows some properties at lower frequencies.
2. Distribution of returns are increasingly fat-tailed as frequency increases. Second moment of distribution exists, but fourth moment tend to diverge.
3. Scaling laws describe mean absolute/squared returns of their time intervals. They're proportional to a power of the interval size.

4. Seasonal heteroskedasticity in daily/weekly cluster of vol. This explains fat-tail of returns.
5. Daily/weekly pattern also exist in quote frequency. They're negatively correlated to vol. Trading activity in terms of price quoting frequency has positive correlation to vol, negative to spread. This implies the trading volume is positively correlated to vol. The daily pattern can explain behavior of 3 markets: US, Euro and Asia, whose active periods partially overlap, there're systematic variations of vol.

5.2. Price Formation Process

FX-rate transaction prices has negative autocorrelation. S&P 500 has positive autocorrelation returns, because they're constructed from equities that have different liquidity (lagged adjustment model ⁶). But stock return themselves or future contracts in indices is not the case, negative first-order autocorrelation is unwanted noise. An **effective price** is defined to eliminate negative autocorrelation. Autocorrelations is related to microstructure effects.

Market makers who want to attract buyers more than sellers, tend to publish skewed quote where only either bid or ask price is competitive, and the other is pushed away by spread of conventional size of 5 or 7 basis points. When they're uncertain about direction of price, they may quote larger spreads with conventional values like 10 or 15. Different banks have different conventions and market situation changes, so the distribution of spreads has 4 or 5 peaks. So we need to *model* the *real* spreads. For Eurofutures (IR futures), there is no well-defined spread because the bid and ask quotes aren't synchronized, and depend on market state, there may be only bid or only ask quotes for a while, but a spread can be computed from bid and ask that are a few seconds apart.

5.2.1. Distributional Properties of Returns

At frequencies > 10 min, the fat tails start to decrease. (See figure 5.6) If data generating process is random walk with increments from stable distribution, which is defined by *scaled* returns $r/(\Delta t)^\gamma$, we can obtain uniform distribution with identical moments within significance limits.

5.2.2. Tail Index of Return Distributions

3 kinds of tail of distributions:

⁶Lagged Adjustment model: one group of stocks reacts more slowly to aggregate info than another group of stocks. Because auto-covariance of a diversified portfolio is the average cross-covariance of stocks that make up the portfolio, which results in positive autocorrelations.

1. Thin-tailed: all moments are finite and cumulative distribution declines exponentially in the tails. Tail index $\alpha = \infty$
2. Fat-tailed: cumulative distribution declines with a power in the tails. $\alpha > 0$
3. Bounded distribution with no tails. $\alpha < 0$

Extreme value theory: extreme value distribution of ordered data must belong to one of above 3 general families, regardless of original distribution function.

Hill (1975) Estimator: $\gamma_{n,m}^H = \frac{1}{m-1} \sum_{i=1}^{m-1} \ln X_{(i)} - \ln X_{(m)}$, $m > 1$ is consistent of $\gamma = 1/\alpha$ for fat-tailed distribution. $(\gamma_{n,m} - \gamma)m^{1/2}$ is asymptotically normally distributed with mean 0 variance γ^2 .

Tail index values of FX rates can be estimated by a subsample bootstrap method (see Table 5.3) on Hill estimator, confidence ranges can be obtained by jackknife method.

Tail index reflects the institutional setup and the way different agents on the market interact, so it's another (apart from drift exponent) empirical measure of market regulation and efficiency. Large tail index \rightarrow free interactions between agents in different time zones, with low degree of regulation and smooth adjustment to external shocks.

5.2.3. Extreme Risk

Have we already seen the largest movements or are we going to experience even larger ones? Once we know tail index, we can apply extreme value theory *outside* our sample to see extreme movements that have not yet been observed historically, by computation of quantiles with exceedance probabilities (Dacorogna et al. 2001).

Practically, hedging against extreme risk must be based on *unconditional* distribution. In large portfolio, it's impossible to find counterparties to hedge in turbulent states of market, so we should plan far in advance.

Expansion of asymptotic cumulative distribution function from X_i observations are:

$$F(x) = 1 - ax^{-\alpha}[1 + bx^{-\beta}]$$

x_p, x_t : quantiles w.r.t. exceedance prob p and t . n : sample size, x_t inside sample ($p < 1/n < t$). By def, and based on same logic as Hill estimator, we have $\hat{x}_p = X_{m(\frac{m}{np})}^{\hat{\gamma}}$, where $m = \hat{m}$ from tail estimation corresponding to $\hat{\gamma}$. For m small w.r.t. n , the tail of $F(x)$ is well approximated by Pareto law, and \hat{x}_p is also good.

$\frac{\sqrt{m}}{\ln \frac{m/n}{p}} \left(\frac{\hat{x}_p}{x_p} - 1 \right)$ has the same limiting normal distribution as Hill estimator, which is an estimator of our quantile computation.

To compromise between accuracy of tail estimation and length of interval needed by risk managers, the time interval is shorter than overnight position practically. First, by Monte Carlo simulations of synthetic data, where process was fitted to 30-min return. Second, quantile estimation on FX rates as function of prob of event happening once a year. (See Table 5.7). HARCH model slightly overestimates extreme risk due to its long memory, but GARCH model underestimates the risk a lot. ARCH-type process capture the tail behavior of FX rates, which can provide early warning in case of turbulent situations. However, the Gaussian-based models will overestimate risk, this model is focusing on tails, for the cases far from tails, normal distribution produce higher quantiles than actual data.

5.3. Scaling Laws

Time interval Δt and average vol measured as a power p of absolute returns observed over intervals, has relationship:

$$E[|r|^p]^{1/p} = c(p)\Delta t^{D(p)}.$$

D : drift exponent. Taking logarithm, estimation of c and D can be derived by OLS regression. The lower the weight of tails in statistics, the more empirically determined drift exponent deviates from Gaussian random walk value 0.5, due to the changing form of distribution under aggregation and a sign of multifractality.

Scaling law applies to different currencies and commodities (gold, silver), when return distributions are unstable.

5.3.1. Market Maker Bias

We not only should study statistical error by uncertainty of price, but also the market makers bias like bouncing effect that reflect in negative autocorrelation of returns in very short term. True market price is not middle price, if spreads large (minor FX rates), the uncertainty implies great measurement error.

$\text{Var}(\log |\Delta x|) \approx \frac{\eta^4}{4|\Delta x|^4} + \frac{1}{2n}$. For long time intervals, $|\Delta x| \gg \eta$, and $1/2n$ is essential part of errors. In case of short time intervals, n is big, $|\Delta x|$ has same order as η , so the first term plays the central role. This explains that errors are large for high frequency points, then diminishing and eventually increasing again when number of observations becomes small. (See figure 5.8)

5.3.2. Limitations of Scaling Laws

Gencay et al. (2001): scaling behavior breaks for returns measured at higher intervals than 1 day. See figure 5.9, decomposition of variance on scale-by-scale basis through application of nondecimated discrete wavelet transform. This method doesn't assume distributional form to returns. e.g., first scale is

associated with 20-min changes, second scale is 40-min, etc. Each increasing scale represents lower frequencies. Oscillations with a period length of 40min to 2560min. Each day is 1440 min, and we see the first 6 scales are related with intraday dynamics. But in the 7-th scale, there's an apparent break in the variance.

5.4. Autocorrelation and Seasonality

Extreme events are less correlated with each other than average or small absolute returns. So, the heteroskedasticity is due to average behavior, not extreme events.

Banks issue price quotes have constraints:

- Granularity: FX prices are quoted with 5 digits like 1.6755 or 105.21. The lowest digit sets granularity and thus the unit *basis points*.
- Quoted spreads are wider than traded spreads, as they include safety margins on both sides of real spread negotiated in simultaneous real transactions. These margins allow FX dealers, when called by customer during the lifetime of quote, to make fine adjustment of bid and ask prices within the range given by the wide quoted spread.
- FX dealers have biased intentions: one of the prices, bid or ask, is carefully chosen to attract a deal in desired direction, the other is made unattractive by increasing the spread.

The cumulative distribution functions have properties:

- Not Gaussian, but convex (s strongly, $\ln s$ slightly), indicating a positive skewness and leptokurticity of the tail on positive side.
- Look like staircase with smooth corners. For nominal spread in basis point s_{nom} , we expect a staircase with sharp corners, the vertical parts of staircase function indicating the preferred *even* values like 10 basis points.

6. Modeling Seasonal Volatility

6.1. Modeling Vol Patterns with Alternative Time Scale and Activity Variable

First, model the price process for seasonal vol patterns. Due to nonstationarity, we introduce a new time scale such that the transformed data of the new time scale don't possess intraday seasonalities. Two components:

- Directing process $\nu(t)$, is a mapping from physical time to another pre-determined time scale, to contain intraday seasonal variations.
- Subordinated price process generated from directing process $x(t) = x^*[\nu(t)]$ is tick data with intraday seasonalities, which leads to x^* process (with no intraday seasonalities).

Transaction clock is defined to cumulate transaction volume to obtain new time scale. Since ν -scale considers the seasonality of x , x^* has nonseasonal vol patterns, it may be conditionally heteroskedastic.

Market Activity and Scaling Law

$$\Delta\nu_i = \left(\frac{E[|r_i|]}{c^*}\right)^{1/D}$$

6.1.1. Geographical Seasonality Patterns

See Figure 6.2, panel on left is the shape of geographical seasonality in Euro market, which has 2 peaks with second peak higher than the former. The relative minimum between 2 peaks is the lunch break effect. For North American, there is no lunch break effect.

To sum over the intraweekly sample, the minimization problem is nonlinear for some params (Eqn 6.15), can be solved by Levenberg-Marquardt method (1986).

6.2. ν -Scale

ν can model the seasonal, intradaily and intraweekly aspect of heteroskedasticity. Activity variable can define the speed of ν against time t .

Def

$$\nu(t) = \int_{t_0}^t a(t')dt' = a_0(t - t_0) + \sum_{k=1}^3 \nu_k(t)$$

ν_k represents business time scale of k^{th} market: $\nu_k(t) = \int_{t_0}^t a_{1,k}(t')dt'$, which can model intramarket behavior. Due to normalization, ν -time can be measured in the same units as physical time (hours).

Relative weight W_k of each market component can be defined as the share of the k^{th} market in ν interval of 1 week.

ν -scale contracts periods of low activity and expands period of high activity.

6.2.1. Adjustments of ν -Scale

Recalibrate factor c^* over whole sample. See Figure 6.6 time mapping function between physical time and ν time. It's difficult to consider different holidays of each markets accurately (like half-day holiday, will need to split daily activity functions into halves).

6.2.2. Filtering Intraday Seasonalities with Wavelets

For intraday studies, we need to extract seasonal vol by wavelet multiscaling approach, to decompose data into low and high frequency components by non-decimated discrete wavelet transform. See figure 6.8, wavelet transformation uncovers the nonseasonal dynamics without imposing any spurious persistence into filtered series.

7. Realized Vol Dynamics

Vol dynamics have conditional heteroskedasticity (vol clustering) and long memory of autocorrelation of vol, and asymmetry of info flow between vol computed from returns measured at different frequencies.

Bias factor is the ratio of 2 mean realized vols over the same sample (See Eqn 7.3, figure 7.2). There is also bias correction method. In figure 7.4, autocorrelation function data sampling in 20-min frequency for lags up to 10 days, we can see the peaks, indicating daily seasonality. There's a final structure with small but visible peaks at every 8 hrs, corresponding to 3 continental markets.

Autocorrelation of squared returns is meaningful only if kurtosis of return process is finite, but it's not guaranteed for currency returns. The autocorrelation of vol is significantly positive and declines at an **hyperbolic** rate (See Figure 7.5) rather than exponential rate. This behavior indicates the presence of long memory process in the underlying data-generating process of returns.

7.0.3. Short/Long Memory

Hyperbolic $f_h(\tau) = k\tau^{-h}$, exponential $f_e(\tau) = ke^{-\tau/h}$, where τ : lag order of autocorrelation function. Exponential function can't simultaneously capture the short and long-term persistence, but hyperbolic function can capture both.

This is similar to fractional noise process $a = \frac{|l+1|^{2H} - 2l^{2H} + |l-1|^{2H}}{2}$, where l : lag parameter, H : Hurst exponent $\in [0.5, 1]$ for persistent fractional noise. For larger lags l , autocorrelation converges to $a \approx H(2H - 1)l^{2(H-1)}$ which has a hyperbolic decay. H value tend to be lower like 0.25, indicating vol doesn't follow pure fractional noise process, but vol is positive definite and has skewed

and fat-tailed distribution, whereas distribution function of pure fractional noise is Gaussian. (Return process doesn't have fractional noise since it's unlike vol having autocorrelation)

7.1. Heterogeneous Market

Contrary to traditional beliefs, vol is positively correlated to market presence, activity and volume. So vol is also indicator for trends.

- Different agents of heterogeneous market have different time zones and dealing frequencies. Market is heterogeneous with fractal structure of the participants' time zones as it consists of short-term and long-term parts. Each part has its own reaction time to news. If we assume memory of vol of one part to be exponentially declining with time const, like GARCH(1,1), the memory of whole market is composed of many such exponential declines with different time consts. The superposition of many exponential declines with widely differing time consts comes close to hyperbolic decline.
- Homogeneous market, the more agents are, the faster the price converges to real market value with rational expectations. So vol should be negatively correlated with market. But in heterogeneous market, different market actors tend to settle for different prices and decide to execute transactions in different market situations, and this creates vol, thus positive correlation of vol.

7.1.1. Vol of Different Time Resolutions

Long term and short term traders have different fine/coarse time grid. So we define 2 types of vol, the coarse vol and fine vol.

7.1.2. Asymmetric Lead-Lag Correlation of Vol

To analyze correlation of 2 time series, like fine and coarse vols, *lagged* correlation is more powerful to see relation between 2 time series. The lagged correlation function considers simultaneously and has a time shift, and reveals Granger causality relations and info flow structures. If 2 time series are generated by a synchronous info flow, they would have symmetric lagged correlation function. The correlation between fine and coarse vol is a function of number of lags. See Figure 7.8, the negative lags indicate that the coarse vol was lagged compared to fine vol, the thin curve indicates asymmetry.

Heat wave effect: traders have better memory of events about 1 day ago than a broken number of days ago. ν -scale is well able to treat ordinary seasonality as indicated by lack of analogous peak around positive lags. Heat wave effect is not only seasonality, but also can't be eliminated by time scale transformation. That's because vol modeling needs to consider selective memory.

7.1.3. Conditional Predictability

See Figure 7.9, lead-lag correlation of fine and coarse vols for 4 different implied forward rates, 3-hr grid in ν -time. The asymmetry of lead-lag correlation is apparent around diagonal, see figure 7.10. So, in intraday, influence of heat wave effect is important.

8. Vol Processes

3 types:

- ARCH. Autoregressive conditional heteroskedastic models define variance σ_t^2 of return as a function of *past returns*. In GARCH, σ_t depends on its own past values and past returns.
- Stochastic Vol models. σ_t doesn't depend on past returns, and depends on its own past values. Since σ_t is not observable and not computable from past returns, so it's difficult to estimate params of this model.
- Realized vol models. Rather than modeling σ_t , we define σ_t as realized vol computed at $t - 1$. With high frequency data, return interval of 30-min keeps stochastic errors low. Time interval of main model is higher, as using realized vl at $t - 1$ is predictor of vol between $t - 1$ and t relying on the vol clustering. Its advantage is that we use empirical data instead of model assumptions that might be wrong. Note that realized vol is biased if computed at high frequency (need bias correction methd). And realized vol at fine vol lags behind coarse vol in lead-lag analysis, which leads to suboptimal forecat quality when predicting the next step vol. In the end, realized vol at $t - 1$ may not be the best predictor of vol between $t - 1$ and t , which should be replaced by a more sophisticated forecast of realized vol at t .

8.1. Temporal Aggregatin of GARCH

The maximum of likelihood function of GARCH can be found by iterative procedure with genetic algorithm and Berndt-Hall-Hall-Hausman (BHHH) algorithm (a variant of gradient descent, but will trap in local maxima).

GARCH can be viewed as a jump process or diffusion process, both approaches lead to similar results, the autoregressive param $\beta_1 \rightarrow 1$ while moving average param $\alpha_1 \rightarrow 0$ as frequency increases, which means in higher frequency, the vol clustering is longer as measured in time intervals. For disaggregation, we go from low to high-frequency, we need to determine params of aggregated GARCH processes. GARCH doesn't capture the heterogeneity of traders acting at different time zones.

8.2. Heterogeneous Vol: HARCH

HARCH process can be seen as a Markov chain, it takes intervals of different sizes.

$$r_t = \sigma_t \epsilon_t, \sigma_t^2 = c_0 + \sum_{j=1}^n c_j \left(\sum_{i=1}^j r_{t-i} \right)^2$$

HARCH (Heterogeneous Autoregressive Condition Heteroskedasticity, 1997, Muller) is based on squared returns, with good analytical tractability. It has variance equation based on returns over intervals of *different sizes*. The empirical behavior of lagged correlation can be reproduced with this process, indicating HARCH can reproduce long memory of vol⁷. Moreover, the conditional variance of HARCH reflects component structure of market.

Memory of vol is long, so we need a high order of HARCH with large n , to model the time series, which implies a high number of coefficients c_j , and we need a parsimonious parametrization to explore component structure of market.

8.3. EMA-HARCH

It has advantage of including a memory from past intervals. *partial* vol σ_j^2 is the contribution of j^{th} component to the total market vol σ^2 . $\sigma_{j,t}^2 = \mu_j \sigma_{j,t-1}^2 + (1 - \mu_j) \left(\sum_{i=1}^{j_j} r_{t-i} \right)^2$. The depth of vol memory decay is determined by const $\mu_j = e^{-\frac{1}{M(k_j)}}$.

9. Forecasting Returns over multiple time horizons

9.1. Intrinsic Time

FX returns has conditional heteroskedasticity, and can be treated with a change of time scale. Based on scaling law and price vol: $\tau(t_c) = \tau(t_{c-1}) + k \frac{\nu(t_c) - \nu(t_{c-1})}{\Delta \nu} \frac{|\Delta x|^E}{c}$, where t_c is current time, price difference on same interval is $\Delta \nu$, E, c are scaling law inverse exponent and factor. k is calibration factor on time series, to keep τ in line with physical time in the long run. So, it's the reverse of scaling law for a return on const ν -time interval size.

τ -scale is intrinsic time, which expands periods of high vol and contract those of low vol, thus capturing the relative importance of events to market. Any

⁷FIGARCH (1996) can model long memory but can't reproduce lead-lag correlation.

moving average based on intrinsic time τ dynamically adapts its range to market events. Forecasting model based on τ -scale has a *dynamic memory* of price history.

But, intrinsic time τ only know the past, business time scale ν also know future, because it's based on average behavior. So, a forecasting model for price need to combine 2 forecasting models, one for intrinsic time, one for the price.

9.2. Linear combination of nonlinear indicators

An indicator is a predictor of price change for forecasting. For fixed forecasting horizon $\Delta\nu_f$, price forecast $\tilde{x}_f = x_c + \sum_{j=1}^m c_{x,j}(\Delta\nu_f)z_{x,j}(\Delta\tilde{\tau}_f, \tau_c)$, where x_c : current price, m : number of intrinsic-time τ -scale indicators in model, coefficients $c_{x,j}(\Delta\nu_f)$ are estimated with multiple linear regression model. $\Delta\tilde{\tau}_f = \sum_{j=1}^m c_{\tau,j}(\Delta\nu_f)z_{\tau,j}(\Delta\nu_f, \nu_c)$ is intrinsic time forecast.

This model doesn't rely on a fixed basic time interval, but is designed with a concept of continuous time. The time when a price is recorded in the database is unequally spaced in time, and τ -scale implies that our forecasting models must be computed simultaneously over several fixed time horizons $\Delta\nu_f$.

9.3. Measuring Forecast Quality

We use a mapping function of returns: the forecast should fit the mapped returns \hat{Y}_i rather than real returns Y_i , the map returns should have less leptokurtic than original returns.

- Small returns should be amplified when considered in regression, to establish a sufficient penalty against forecasts of wrong direction.
- Large returns should be reduced when considered in regression, so the distribution function of mapped returns is no longer leptokurtic.
- Mapping effects should decrease with increasing time horizon size.

$\hat{Y}_i = M(Y_i) = \frac{AY_i}{[Y_i^2 + B]^\alpha}$, where $\alpha \in [0, 0.5]$ since the mapping function must be underproportional bijection.

10. Correlation & Multivariate Risk

- Correlation of intraday, equally spaced time series derived from unevenly spaced tick data, need careful treatment if a bias resulting from classical missing value problem. We can treat heteroskedasticity by expanding periods of high vol while contracting periods of low vol.

- Correlations between financial time series vary over time
- There's dramatic decrease in correlation as data frequency enters intrahour level (Epps effect: Nonzero correlations of returns are dramatically attenuated when interval decreases and enters the intrahour region).

In asset allocation and risk, return measurement intervals should be chosen longer than stabilization interval.

10.1. Covolatility Weighting

We need to introduce a time scale that compresses physical time if there is no info and expand t when it exists (similar to ν -time in model vol patterns), and we have multivariate problem of 2 time series for a common time scale.

Large data gaps (varying/nonmatching data arrival frequencies) occur due to failures in data acquisition chain, we can only guess.

Monte Carlo and Empirical Tests

Use synthetic Monte Carlo data to see effectiveness of covolatility adjusted correlation method. (See Figure 10.1)

11. Trading Models

Real time trading strategies:

- Give warning a few minutes in advance of a deal
- Not change recommendations too rapidly
- Not give recommendations outside business hours
- Consider market holidays
- Support stop-loss (around the clock)

Three parts of strategies:

- Generation of trading model recommendations
- Receipt of simulated positions by simulated trader
- Generation of model statistics by performance calculator.

Current Return Calculations. Traders seldom take full exposure at once, they like to build positions in gearing steps, so it's useful to introduce an auxiliary variable, the average price \bar{p} paid for achieving the current exposure (gearing). This simplifies the computation of return of a position built in steps. After a new deal with index i , the average price depends on type of transactions:

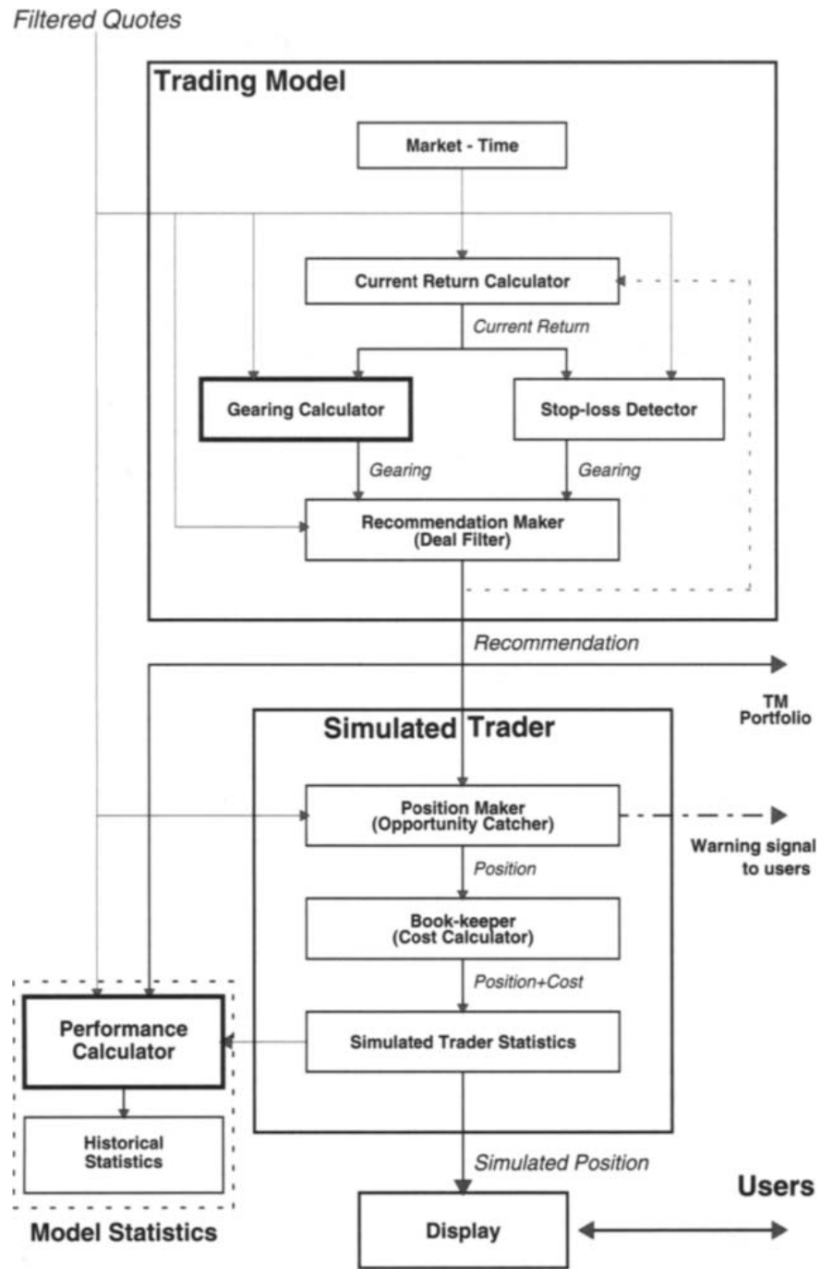


FIGURE 11.1 Data flow of prices and deal recommendations within a real-time trading model.

Figure 1: Data flow of prices and deal recommendations within a real-time trading model

$$\bar{p}_i \equiv \begin{cases} \bar{p}_{i-1} & \text{if } |g_i| < |g_{i-1}| \text{ and } g_i g_{i-1} > 0 \\ g_i \left[\frac{g_i - g_{i-1}}{p_i} + \frac{g_{i-1}}{\bar{p}_{i-1}} \right]^{-1} & \text{if } |g_i| > |g_{i-1}| \text{ and } g_i g_{i-1} > 0 \\ p_i & \text{if } g_i g_{i-1} < 0 \text{ or } g_{i-1} = 0 \\ \text{undefined} & \text{if } g_i = 0 \end{cases} \quad (11.1)$$

where g_{i-1} and g_i are previous and current gearings. The return of a deal $r_i = (g_{i-1} - g'_i) \left(\frac{p_i}{\bar{p}_{i-1}} - 1 \right)$.

Gearing Calculation is the heart of a trading model, it provides trading model with intelligence and ability to capitalize on movements in markets. It's the real model. It consists of indicators (produced from input price data) and trading rules (functions of past dealing history).

Recommendation Maker. Stop-Loss Detection. Stop-Profit Control.

Algorithms

- Trend following indicators, detect and follow market trends
- Overbought and oversold indicators, detect important market tuning points
- Cycle indicators, emphasize periodic market fluctuations
- Timing indicators, provided optimum exit conditions

To minimize overfitting:

- Indicator evaluation for different time series
- Robust optimization technique to find best slution in parameter space
- Strict testing procedures and large data samples

11.1. Modified Sharing Function for Robust Optimizations

Find new genetic algorithm that avoids concentration of many individuals around sharp peaks of high fitness but detects broad regions of param space containing a group of individuals with high average fitness level and small variance of individual fitness values.

So we need a new sharing function that penalizes clusters with: large variance of individual fitness values and too many solutions concentrated inside too small a region. All individuals in a given cluster c share the same fitness value.

$$s_f(i) = \bar{f}_c - \left(\frac{N_c}{N_{av}} + \frac{1 - r_d}{r_d} \right) \sigma(f_c) \quad \forall x_i \in C_c \quad (11.2)$$

where N_c is number of genes in cluster, \bar{f}_c is average fitness value.

$$\overline{f_c} = \frac{1}{N_c} \sum_{i=1}^{N_c} f(i) \quad \text{and} \quad \sigma(f_c) = \sqrt{\frac{1}{N_c - 1} \sum_{i=1}^{N_c} (f(i) - \overline{f_c})^2}$$

If $N_c < \text{expected average number of genes inside each cluster } N_{av}$, the correction is reduced, otherwise increased. The second term $(1 - r_d)/r_d$ is used to penalize clusters with too high a concentration of genes around centroid.

11.2. Real-time Testing

Historical data should be split into at least 2 sample periods, one *in-sample* for optimization, another *out-of-sample* for performance tests. The split must not be modified during optimization or testing, otherwise the risk of overfitting the historical data will be large and test is unreliable.

The heterogeneous market hypothesis attributes the profitability of trading models to simultaneous presence of heterogeneous agents, whereas the classical efficient market hypothesis relates this profitability to inefficiencies, which implies illiquid periods of market are most favorable for excess returns. So the optimal daily trading time and maxima of performance are clustered around opening hours when main markets are active. The best active times are shifted for certain currencies to accommodate their main markets. If the model are only allowed to trade for 1 hour, the best choice is around the peaks in the daily activity of the market.

11.3. Dynamic Hedging with Exposure Constraints

Optimal static hedge can reduce risk, but there may be some distance between static point and efficient frontier of feasible set, which is defined by dynamic allocation models, so we need dynamic hedging.

Constraints (Quadratic programming issues):

- Trading model size must not be negative (which means doing the opposite of the trading recommendations, will lead to new transaction costs)
- Static hedging ratios are limited to an allowed range (risk regulations)

End - Of - The - Book - Notes