Estimates of Real-Time Volatility in Interest Rate Futures

Ramnik Arora
Advisor: Robert Almgren

December 21, 2011
# Contents

1 Introduction .................................................. 2

2 Volatility Estimators ....................................... 6
   2.1 Evaluation of Volatility Estimators .................. 6
   2.2 Realized Variance ....................................... 7
      2.2.1 Criticism .......................................... 7
   2.3 Garman-Klass Estimators ................................ 8
      2.3.1 Criticism .......................................... 9
   2.4 Hawkes Process ......................................... 9
      2.4.1 Criticism .......................................... 9
   2.5 Brownian Motion with Invisible Bands ................ 9
      2.5.1 Criticism .......................................... 10

3 AR(1) Model ................................................ 12
   3.1 Characteristics of the Market ......................... 12
   3.2 Daily Estimator ........................................ 13
      3.2.1 Volatility Forecasting and Comparisons ........ 13
      3.2.2 TWAP Schedules .................................. 14
   3.3 Real-time Volatility Estimator ....................... 16
      3.3.1 Constant $\beta$ .................................. 16
      3.3.2 Real-time AR(1) Volatility ..................... 16
      3.3.3 Correlations and Clustering .................... 16
   3.4 Extensions .............................................. 19
      3.4.1 Binary-Tree Model ................................ 21
      3.4.2 Vector Autoregressive Model ..................... 21
   3.5 Contrast with other models ............................ 22
   3.6 Conclusion ............................................. 22

A Appendix .................................................. 24
   A.1 Properties of Interest Rate Futures ................. 24
   A.2 Diffusion with Barriers ................................ 24
   A.3 AR(1) Coefficient Update ............................. 26
Volatility is an estimate of the fluctuations in the price of a underlying contract, and is used to measure the uncertainty or risk posed by the financial instrument. It signifies the risk associated with the price process, and useful in many applications, such as risk-management, option pricing, portfolio allocation and execution. While executing a order we have a trading schedule, which we try to adhere to (often determined by the benchmark, for example *Time Weighted Average Price* or *Volume Weighted Average Price*), and some forward and behind boundaries which will make sure we don’t stray too far away from the trade schedule. However, these bands must be constricted in accordance to our forecast of the risk, in a mean-variance framework. Often, volume is used as a proxy for risk\(^1\), but this relationship is not exact, and frequently large volumes print without significant large price movements.

The question of volatility estimation is closely related to fitting a model to the market prices\(^2\), and in a “random walk plus microstructure effects” framework, we are concerned with the quadratic variation of the true price. However, the market prices don’t follow a model perfectly, and moreover we only have access to the discretized observations\(^3\) of a continuous latent underlying price process, making this a difficult ordeal. To further complicate matters, the price series have irregular temporal spacing, and peculiar diurnal patterns.

It is desirable of any real-time estimator to reproduce the properties and characteristics of high frequency data. Owing to the large amounts of data,

---

\(^1\)The correlation between volume traded and range(maximum price - minimum price within the timeperiod) in prices, in 15-minute bins for Sept’13 Eurodollar on 2011.12.06 is 0.73.

\(^2\)A familiar example is *implied volatility* which is deeply connected with lognormal asset prices, in the Black-Scholes framework.

\(^3\)The minimum price increments are *ridiculously* large in the interest rate future, and the price movements are very discrete, often not changing for large periods in time. See reftab:tick for details on tick size.
Figure 1.1: An illustration of the orderbook for March’14 Eurodollar on 2011.12.06. The microprice is depicted in black, the midprice in the green(dashed) while the trade prints at shown by dark green dots. The bid/ask volume is represented accordingly. It is interesting to note the flickering quotes, and the presence of ghost liquidity. The market is essentially multi-tick, with often people narrowing the spreads.
the need for a fast estimation process is heightened. Besides, it would be very useful if the estimation process, converted volatility to be an observable (much like volume traded), since it would allow greater use cases in the algorithm. Rather than provide volatility per unit time, the algorithm could subscribe to its own measure based on requirements. Another crucial property of real-time estimator is that it must be parameter-free, since it would be difficult to guess appropriate parameters, which maybe dependent on the changing market conditions. As a matter of usage, we will compute volatility in terms of ticks\(^4\) per unit time, in order to standardize over different interest rate futures contracts.

Price changes are often unstable, and rapid reversions in the prices are rampant. An example would be that for the market to actually move, one would require either the bid moves down first and the ask follows down, or the ask moves up followed by the bid moving up. Rarely do both the actions occur simultaneously, and the multi-tick model\(^5\) is transient. Our model must somehow minimize the impact of these unstable quotes, or rapidly-reverting prices changes (see figure 1.1).

In order to devise a volatility estimator, various data series maybe used (See figure 1.1 for illustration). The market provides us with trade/order book data, and both may very well be used for a real-time volatility estimator.

1. Trade data: Trade data represents the timestamped series of the trades observed by the market. Since trade prices may print on the bid or the ask, these inherently contain microstructure noise and are unsuitable for direct usage. Plenty of real-time estimators have been proposed using high-frequency trade data, but all these make significant assumption on microstructure noise\(^6\). A recent paper \cite{Gatheral2010} concludes that in a Zero-intelligence model, in terms of sampling, the midprice and microprice are 40-60 times less noisy than trade data (as measured by microstructure noise variance). For these reasons, we choose not to use trade prints from the markets.

2. Microprice: Micro-price is defined as (bid volume) * ask + (ask volume) * bid, and is representative of the current state of the market. However, this metric is gameable, by ghost liquidity, and in \cite{Gatheral2010}, the mid-quote is weakly preferred over micro-price in a zero-intelligence setup.

3. Midprice: Mid-price is defined as the arithmetic mean of the inside quote, or the mean of the bid and the ask. This is known to be more

\(^4\)A tick is used to denote the minimum price increment allowed by the CME for the particular contract. Please see A.1 for contract-wise details.

\(^5\)Multi-tick markets are when the bid-ask spread exceeds the minimum tick, for example when the bid moves down or the ask moves up.
stable (albeit discontinuous) and doesn’t incorporate microstructure noise.

For the study, the interest rate futures tick-by-tick data from the CME has been used. Although we only present the results for volatility estimates from Eurodollar and Treasury future outrights, similar property are observed in Packs, Spreads and Butterfly contracts on these outrights. In section \textsuperscript{2} we briefly discuss a few models for real-time volatility estimators, like sum of squared returns, Garman-Klass estimator, modelling based on Hawkes process, and use of invisible bands around midprices for mean-reverting characteristics of the market. In section \textsuperscript{3} we choose to study the sum of squares of AR(1) corrected residuals in greater details, and it’s suitability in a high-frequency environment.
Chapter 2

Volatility Estimators

In this chapter, we will discuss some of the volatility estimators implemented and the corresponding results obtained. These estimators were eventually discarded and a quick discussion of their suitability is presented.

2.1 Evaluation of Volatility Estimators

A critical problem one faces in volatility estimation is reliable quantitative comparison of various estimators. A test for volatility estimators is to normalize the period returns by the volatility estimates, and one must have a draw from a standard gaussian distribution. The empirical distribution maybe tested against a standard gaussian distribution, and metric suitably defined on the performance of the metric. However, hereby we are assuming that the distribution of returns is normally distributed, and this observation is empirically flawed. This test is strict and would not be met by any estimator perfectly, owing to the large error introduced from discretization, leptokurtosis of return series, microstructure noise amongst other factors. It does however form a suitable benchmark for comparison.

There are some empirical observations that every volatility estimator must satisfy:

1. Eurodollar volatility should mostly be in increasing order of time to maturity, since the uncertainty is greatest for far out contracts.

2. Amongst the treasury future outrights, the five-year are the most volatile, while the two-year behave similar to the Eurodollars.

3. Any metric imposed on the correlations amongst treasuries should cluster them in increasing order of time to maturity.

4. Volatility should spike up around Non-Farm Payroll, FOMC Meetings and other significant number releases.
Figure 2.1: A comparison of the Realized Variance, Garman-Klass and AR(1) corrected daily volatility estimators. The curves represent the daily returns normalized by the volatility estimator, and a histogram is compared with the a histogram from standard gaussian distribution.

5. Volatility should spike around May 6th 2010 (Flash Crash) and 12th August 2011 (US Sovereign Downgrade).

2.2 Realized Variance

Realized Variance is a simple estimator given by the quadratic variation of the return series, which in single tick markets often reduces to counting the total number of price changes. Because of large serial autocorrelation between returns, arising from microstructure noise, we expect realized variance to overestimate the variance in a period.

In figure 2.2 we compare the \( \frac{\text{daily returns}}{\text{realized volatility}} \) against a standard gaussian distribution (green curve for standard gaussian, while blue for returns normalized using realized volatility), and in 2.3 the blue curve represent the squared sum of returns for front month Eurodollar from 2011.01.01-2011.12.06.

2.2.1 Criticism

Realized Variance is a naive estimator since it doesn’t discount for unstable quotes or mean reversion of interest rate future returns. Although the estimator can be computed rapidly, and would treat volatility as an observable, it doesn’t replicate the characteristics of high-frequency data. In early
Figure 2.2: A plot of the Realized, AR(1) and Garman-Klass Volatility Estimators. The impact of US Debt downgrade, the unusual non-farm payroll, important FOMC announcements and the Eurozone debt crisis is evident.

2.3 Garman-Klass Estimators

A standard approach to daily realized volatility estimation was provided in Garman-Klass(1980). Their volatility estimators use open, high, low, close prices for the given period to estimate the volatility.

\[
\sigma^2 = \frac{1}{2}(\log_e \text{High} - \log_e \text{Low})^2 - (2 \log_e 2 - 1)^2(\log_e \text{Close} - \log_e \text{Open})^2
\]

A comparison of the Garman-Klass estimator with the Realized variance, and the AR(1)-Volatility estimator can be seen in Figure 2.2. Although Garman-Klass seems to do fairly well as a estimator of daily volatility, it breaks down as we increase the sampling frequency. In figure 2.3, the Garman Klass volatility estimator for daily returns is presented for front Eurodollar contract between 2011.01.01-2011.12.06. Garman-Klass estimator allows subsampling a time period, which was briefly looked into revealing similar results.
2.3.1 Criticism

Garman-Klass estimators only rely on the OHLC data for volatility estimation, and discard the remaining data. Garman Klass estimator were designed for low frequency data and fail to capture the peculiar characteristics of interest rate futures.

2.4 Hawkes Process

In the paper [Bacry(2010)] the authors model microstructure noise using Hawkes process for tick-by-tick variation. Two stochastic counting processes are associated with the price processes, for the positive and negative jumps in prices. These stochastic processes, are assumed to be mutually stimulating, which is helpful to reproduce the high-frequency mean-reversion characteristics observed in the data.

The volatility maybe expressed as the limit of the signature plot of the mid-price series. The paper prescribes the calibration of the limit using a 3 parameter model.

2.4.1 Criticism

The model parameters converge to un-intuitive values, suggesting that the base diffusion rate is close to zero and the processes are simply mutually stimulated, making them unstable. These are not responsive in real-time and are useful for calibrating volatilites using lower frequency data.

2.5 Brownian Motion with Invisible Bands

An interesting approach to high-frequency volatility is proposed in [Robert(2011)]. They suggest that the underlying latent process $X(t)$ is observed on discretized bands and incorporate the market characteristics using uncertainty zones, which are bands around the mid-tick grid where the efficient price is too far from the tick grid to trigger a price change.

Assuming we have two absorbing barriers, at $[0, 1]$, with a particle starting at $\alpha \in [0, 1]$. We can calculate the total probability of the particle being

\[
\hat{C}(\tau) = \frac{1}{T} \sum_{n=0}^{n=\frac{T}{\tau}} (X((n+1)\tau) - X(n\tau))^2
\]

---

1. The signature plot of a time series $X(t)$, corresponds to the realized variance at different sampling frequency. Thus, it’s a plot of $\hat{C}$, over a time period $[0, T]$ at a scale $\tau > 0$, where $\hat{C}$ is defined as:
Figure 2.3: The signature plot for March’12 Eurodollar with sampling frequency varying from 15-1015 seconds. As can be seen the signature plot is very choppy, and the smooth approximation is shown by the red/green curves, whose asymptotic limits represent the volatility.

absorbed at the two barriers analytically as a function of time(t)(see Appendix A.2). The probability of being absorbed at the lower boundary is \(1 - \alpha\), while at the upper boundary is \(\alpha\). This maybe empirically calibrated from the market, where absorption to the lower boundary can be treated as reversals, and absorption by the upper boundary is equivalent to continuations. The scaling in time whereby the two curves are closest (in \(L_2\)) would be representative of volatility, which is equivalent to scaling in time.

2.5.1 Criticism

From the figure it can be seen that the model doesn’t calibrate very well to the market, especially for continuations. In the market, a large number of times (and especially during events), the market has a strong directional trend, which is not captured by this model. Assuming that the latent underlying process follows a Simple Brownian Motion seems like an oversimplification. It maybe benificial to extend the model for generic Levy processes to account for rapid continuations.
Figure 2.4: The shaded region represents the cumulative flux for a brownian motion, starting at point $\alpha$ (calibrated using market data as the mean-reversion probability), and the dotted line is the scaled empirical cumulative flux observed from market data. Notice the significant difference in the cumulative flux of continuations with respect to the analytical results.
Chapter 3

AR(1) Model

3.1 Characteristics of the Market

As has been alluded to often in the report, the returns have a serial autocorrelation. As can be seen in figure 3.1, the midpoint changes in single tick market have a negative AR(1) coefficient, or are mean-reverting. This is quite unlike daily returns, which have been empirically observed to have zero serial autocorrelation [Engle(1995)]. This is a critical observation, and a defining characteristic of the market.

Let $r_t$ be the midpoint changes in single tick markets. We hypothesize that:

$$r_t = \beta r_{t-1} + \epsilon_t$$

whereby $\epsilon_t$ is a mean zero process. The quadratic variation of the residuals $\epsilon_t$ would be the volatility of the return series.

This metric is consistent, since volatility is the conditional variance of asset returns\(^\text{1}\) or mathematically:

$$\sigma_t^2 = \mathbb{E}[(r_t - \mu_t)^2 | \mathbb{F}_{t-1}]$$

where $\mathbb{F}_{t-1}$ is the information available till $t-1$, and $\mu_t = \mathbb{E}[r_t | \mathbb{F}_{t-1}]$, which displays a definite AR(1) structure.

A lucid illustration of this measure is provided in 3.2, where we plot the actual prices, and the scaled (to same range) the AR(1) corrected price series, assuming a constant $\beta$ for the day\(^\text{2}\). In our measure of AR(1) corrected

---

\(^1\)To quote Paul Wilmott from his blog: “The definition of volatility is subtly different when there is SAC[serial autocorrelation]. The sequence +1, -1, +1, -1, +1, has perfect negative SAC and a volatility of zero!”.

\(^2\)Let $y_t$ be the AR(1) corrected price series. This is computed as $y_t = x_0 + \sum_{k=1}^{t} \epsilon_k$, where $x_0$ is the starting price of the underlier. Scaling is done such that the series underlier price series($x_t$), and AR(1) corrected price series($y_t$) have the same range.
volatility, we are effectively (loosely speaking, because of scaling) taking the quadratic variation of returns from the AR(1) corrected price series (shown in blue), which seems to dampen the mean reverting quotes.

3.2 Daily Estimator

Using the above formulation, we initially come up with daily estimators of volatility. Given a day’s data, we compute the first-order serial autocorrelation of the returns in single-tick wide markets, and the quadratic variation of the residual represents the AR(1) variance for the given contract on the given day. Figure 2.2 plots the normalized AR(1) volatility realized returns against equal draws from a standard normal for the first(rolling) eight eurodollar contracts between 2011.01.01-2011.12.06. Figure 2.3 is a plot of the AR(1) volatility of the near expiry contract between the same time period.

3.2.1 Volatility Forecasting and Comparisons

The quadratic variation of the residuals may also be computed in non-intersecting 15-minute bins, spanning the entire day using historical data. An exponential weighted average of previous periods, super-imposing the
Figure 3.2: The black line is the price for March’13 Eurodollar on 6th December 2011, and the blue line is the scaled AR(1) corrected price series. As is evident, the AR(1) corrected price series dampen the contribution to quadratic variation from mean-reverting quotes.

event model\(^3\) yields the volatility forecast for the various Interest Rate Futures complex in 15-minute bins. Figure 3.3 is a forecast of the Eurodollar complex volatility forecast generated on 2011.12.05 for the trading day 2011.12.06, while the comparisons with realized AR(1) volatility are presented to the right.

3.2.2 TWAP Schedules

Given our forecast of volatility, we adjust out forward and backward boundary in accordance with our volatility forecast\(^3\). Figure 3.4 is an illustration of this behaviour, where boundaries are adjusted on the basis of volume and volatility respectively. The volume and volatility profile can easily be contrasted here, and the difference in the ahead and behind boundaries is noticeable. The contrast between volume and volatility is extreme on 2011.09.21, an important FOMC meeting announcement, and the constriction of the forward and behind boundaries is obvious.

\(^3\)This is proprietary to Quantitative Brokers, and not a part of this project.
Figure 3.3: Volatility Forecasts and realized AR(1) Volatility produced on 2011.12.05 for the trade day 2011.12.06, in 15-minute bars.

Figure 3.4: The forward and behind boundaries for trading the Eurodollar complex on 2011.12.06 and 2011.09.21. High volume and volatility was expected around FOMC announcement on 2011.09.21, and we force ourselves to return to our trading schedule (in this case TWAP) before the announcement by contricting the bands.
3.3 Real-time Volatility Estimator

3.3.1 Constant $\beta$

For daily volatility numbers, we assume that $\beta$ is constant for the entire day, and is calibrated using the entire days data. In support of this assumption, we fit a time varying AR-1 model using Kalman filter. Essentially, we assume:

\[
\beta_t = \beta_{t-1} + z_t \sim \text{IIDN}(0, \sigma_1^2) \\
r_t = \beta_t r_{t-1} + v_t \sim \text{IIDN}(0, \sigma_2^2)
\]

In figure 3.5, we can see the assumption of constant $\beta$ versus filtered $\beta$ versus real-time $\beta$ are pretty close to each other, especially during market hours. For real-time estimation, we’ll continue with the assumption of constant $\beta$ for the entire day, but our estimator for $\beta$ at time $t$ will use data until time $t$, as discussed below. Graphically, this corresponds to the yellow line on the figure 3.5 which seems to track the green line really well.

3.3.2 Real-time AR(1) Volatility

Given the assumption of constant $\beta$, the algorithm for online update of the AR(1) coefficient is $O(1)$, and can be computed rapidly in a real-time system (See appendix A.3 for details). In figure 3.6 we implement the algorithm for the March’13 Eurodollar contract on 2011.12.06, between 07:30-07:46 a.m. CST. The volatility estimated in the period is 3.67 ticks, while the residuals $\epsilon_t$ can also be seen in the plots. The AR(1) coefficient is -0.64. It is crucial to note that the mean-reverting flickering quotes contributions to the volatility is nominal in comparison to continuations. The behaviour close to 07:44 a.m. is particularly characteristic of the interest rate futures markets whereby the market is essentially in a multi-tick mode, and frequently participants are closing the gap. Any metric such as realized variance, Garman-Klass estimator would be unable to cancel out this noise.

3.3.3 Correlations and Clustering

One would expect a strong correlation within the volatility of neighbouring contracts in the eurodollar complex since segments of the yield curve often move together in unison. Table 3.1 plots the correlation matrix between the front end of the yield curve, and the US Treasury complex for 2011.12.06. An agglomerative clustering of the volatility series, based on 15-minute AR(1)
Figure 3.5: A plot of the returns (in black, where returns vary between \{-1, 1\}), the AR-1 coefficient for the day (in red, which is used for daily estimator), AR-1 coefficient in realtime using data till time $t$ (in yellow, used for real-time volatility estimator), and AR-1 coefficient learned through Kalman filtering model (in green). Note that during market hours, or generally in high volume periods, we tend to do reasonably well with the given assumption of constant AR-1 coefficient.
Figure 3.6: Real-time estimate of AR(1) volatility. The AR(1) residual is shown on each price change along with the AR(1) coefficient and the period volatility. Contribution from mean-reverting quotes is removed, and only continuations contribute to the volatility number.
Dendrogram for the various Interest Rate Futures

2011.11.29

Figure 3.7: An Agglomerative Hierarchical Clustering of the Interest Rate Futures based on the real-time volatility estimate. Clusters of the white (First 4 Eurodollars), reds (1-2 Year out) and the others are clearly seen. The treasury futures are differ than the Eurodollars, and the Two-Year Futures display a suprisingly different volatility characteristics.

Volatility, using $L_1$ norm for 2011.11.29 reveal a very intuitive structure. The whites (first four Eurodollars), reds(4-8 Eurodollars) are very highly correlated, while the greens and blues behave similarly. Amongst the treasury futures, the two-year behave distinctly\footnote{Their trading is patchy, and in alignment with empirical observations.} while the 5, 10 and the T-Bond are closely related.

### 3.4 Extensions

Hereby, we suggest some improvements, and extensions to the AR(1) model discussed before.
Table 3.1: The correlation matrix of the AR(1) volatility of the first twelve Eurodollars, and the most liquid 2, 5, 10, Long US Treasury Futures.
3.4.1 Binary-Tree Model

In the previous AR(1) formulation, we have assumed \( \epsilon_t \) to be normally distributed. However, since we are taking midpoint changes in tick, we know that \( r_t \) takes mostly \{+1, -1\} values. Since returns have a mean-zero, \( \epsilon_t \) must also be a zero-mean process, with distribution:

\[
\epsilon_t := \begin{cases} 
1 - \beta r_{t-1} & \text{with probability } \frac{1 + \beta r_{t-1}}{2} \\
-1 - \beta r_{t-1} & \text{with probability } \frac{1 - \beta r_{t-1}}{2}
\end{cases}
\] (3.1)

Now, we can construct a markovian binary tree such that each node retains the knowledge of the last path (up/down), that led to the node. In particular, we can define:

\[
u^n_j = \Pr(\text{At node } j, \text{time } n, \text{last step was up})
\]
\[
v^n_j = \Pr(\text{At node } j, \text{time } n, \text{last step was up})
\]

Now, given \( \beta \) as the AR(1) coefficient, we can write a recursive relation for the probabilities \( u^n_j \) and \( v^n_j \) as:

\[
u^n_j = u^{n-1}_{j-1} \frac{1 + \beta}{2} + v^{n-1}_{j-1} \frac{1 - \beta}{2}
\]
and similarly

\[
v^n_j = u^{n-1}_{j-1} \frac{1 - \beta}{2} + v^{n-1}_{j-1} \frac{1 + \beta}{2}
\]

Also, since the sum of the probabilities should sum to 1 at each time step, we must have:

\[\sum_{j=0}^{n} (u^n_j + v^n_j) = 1, \quad \forall \ n \geq 0\]

This presents us with an interesting formulation, since we may calculate the distribution any number of time steps out. The number of time steps is an observable which maybe obtained from the market using the number of moves in the last period, or historically over the given time period. However, in the markets, often the average time taken for continuations are vastly greater than the average time for reversal, and somehow this must be incorporated in the tree structure for practical usage.

3.4.2 Vector Autoregressive Model

An obvious extension to a univariate AR(1) model would be an Vector Autoregressive Model whereby segments of the yield curve maybe modelled directly. It is easy to imagine two price series being modelled on a grid, but an added complexity arises from moving to the return space. Owing
to irregular temporal spacing, the returns vector could not only include the
states whereby the midpoint changed, but the states whereby, for a particu-
lar contract, there was no change in the in the midpoint (while the midpoint
may have changed on another contract). This problem maybe mitigated by
sampling the price series at fixed time interval, and using this to construct
the return series.

3.5 Contrast with other models

In literature (see [Gatheral(2010)]), a number of models for real-time volatil-
ity estimation have been discussed, and we contrast the AR(1)-volatility
model with the some of the popular estimators.

- **ARCH/GARCH estimators** [Engle(1995)]: In ARCH we assume
  that the returns are draws from a gaussian distribution, with the
  variance following an ARMA model. The fundamental observation
  leading to GARCH was the positive autocorrelation in $r^2_t$, while near-
  zero autocorrelation in $r_t$. These observation are invalid for the high-
  frequency returns in the interest-rate futures markets, and not a rep-
  resentative characteristic of the market.

- **Zhou’s Model** [Zhou(1992)]: The author comes up with volatility es-
  timate using tick-by-tick data for foreign exchange rates. Zhou ob-
  serve a very high negative first-order autocorrelation in the returns,
  close to -0.45 for DEM/USD rates, owing to the microstructure noise
  attributed to the bid-ask bounce in trade print. This is again not a
  defining characteristic of the market, and would not incorporate the
  large negative serial autocorrelation in the midprice owing to flickering
  quotes.

3.6 Conclusion

The quadratic variation of the residuals in the presence of an AR-1 model is a
pretty nice estimate for volatility. It seems to be a reliable estimator, suitable
for practical application. In particular, some of the appealing properties of
the estimator are:

- It reproduces the characteristics and properties of the mean-reverting
  high frequency midprice changes.

- We are able to treat volatility as a observable. The process of update
  of AR-1 coefficient is very fast, and this is suitable for high frequency
  strategies.
• The parameters in the model are calibrated from the data directly, and don’t require us to tweak any parameters.

• The model inherently dampens contribution from flickering/unstable quotes.
Appendix A

Appendix

A.1 Properties of Interest Rate Futures

We list some of the properties of Interest Rate futures which have a significant impact on the microstructure noise.

The interest rate futures have very large minimum price increments (tick size), as can be seen in table A.1. This forces high percentage of the liquidity to accumulate on the inside levels (See figure A.1) and crossing the spread is more expensive.

Owing to the large minimum tick size, the interest rate futures markets are generally single tick wide (See A.2). The multi-tick mode is often a transient state necessary for price changes, and inherently unstable.

A.2 Diffusion with Barriers

Suppose a brownian motion has absorbing barriers at $x = 0$ and $x = 1$, and the starting point is $z \in (0, 1)$. Let $u(x, t)$ be the probability density

<table>
<thead>
<tr>
<th>Contract</th>
<th>Value of Tick</th>
<th>Minimum Tick</th>
<th>Notional $ Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eurodollar</td>
<td>12.5$</td>
<td>0.5</td>
<td>1M$</td>
</tr>
<tr>
<td>2-Year Futures</td>
<td>15.625$</td>
<td>1/128</td>
<td>200K$</td>
</tr>
<tr>
<td>5-Year Futures</td>
<td>7.8125$</td>
<td>1/128</td>
<td>100K$</td>
</tr>
<tr>
<td>10-Year Futures</td>
<td>15.625$</td>
<td>1/64</td>
<td>100K$</td>
</tr>
<tr>
<td>Treasury-Bond Futures</td>
<td>31.25$</td>
<td>1/32</td>
<td>100K$</td>
</tr>
</tbody>
</table>

Table A.1: A synopsis of the minimum price increments and the notional value of interest rate futures outright contracts.
Figure A.1: Figure shows the percentage of total posted liquidity at the various levels on the bid and the offer. It’s important to notice that Eurodollars and Treasuries have very thin liquidity at far away levels. [Courtesy: Haider Ali, Quantitative Brokers]
Table A.2: The amount of time and the quote changes spent in multitick mode. It is evident from the numbers that the markets (even more so during the floor hours) are primarily one minimum tick wide.

function associated with the particle satisfies:

\[ u_t = \frac{1}{2}u_{xx} \]

with the boundary conditions that \( u(0, t) = 0 \) and \( u(1, t) = 0 \). The solution for the probability density function maybe found using infinite number of positive and negative sources, which yields the solution:

\[ u(x, t) = \sum_{k=-\infty}^{\infty} \left[ e^{-\frac{(x-z-2k)^2}{2t}} - e^{-\frac{(x+z-2k)^2}{2t}} \right] \]

The flux \( \phi = -\frac{1}{2}u_x \),

\[ \phi(x, t) = \frac{1}{2t\sqrt{2\pi t}} \sum_{k=-\infty}^{\infty} \left[ (x-z-2k)e^{-\frac{(x-z-2k)^2}{2t}} - (x+z-2k)e^{-\frac{(x+z-2k)^2}{2t}} \right] \]

A.3 AR(1) Coefficient Update

Given the regression coefficient until time \( t \), on receiving an additional observation, we may dynamically update the regression coefficient. This property is very useful in maintaining a real-time system, since on each inside quote update, we must not retrieve the entire day’s data again. Suppose we have the regression coefficient \( \beta_{t-1} \) and an additional data set \((y_t, x_t)\) arrives,
where $y_t$ is the independent variable, and $x_t$, the dependent variable. Then,

$$\beta_t = (X_t'X_t)^{-1}X_t'Y_t$$

$$= ((X_{t-1}x_t)'(X_{t-1}x_t))^{-1}(X_{t-1}x_t)'(Y_{t-1}y_t)$$

$$= (X_{t-1}'X_{t-1} + x_t'x_t)^{-1}(X_{t-1}'Y_{t-1} + x_t'y_t)$$

where $X_t = (x_1, x_2, \ldots x_t)'$, and similarly $Y_t = (y_1, y_2, \ldots y_t)'$. Thus, if we just store the running sum of $X_t'X_t$ and $X_t'Y_t$, then we $\beta_{t+1}$ in $O(1)$, without accessing previous data. This makes our algorithm very fast, and we can also calculate $\epsilon_t$ as $y_t - \beta_t x_t$. 

27
Bibliography


