Some alternative methods:

- Moving window estimates
- Use of high-frequency financial data
- Use of daily open, high, low and closing prices

**Moving window**
A simple approach to capture time-varying feature of the volatility. There is no simple answer to the choice of window size.

**Demonstration:** Consider the daily log returns of the S&P 500 index from 1950 to 2008. A R script, `r-mvwindow.txt`, is available on the course web. [In class demonstration.]

**Use of High-Frequency Data**
Purpose: monthly volatility
Data: Daily returns
Let $r_t^m$ be the $t$-th month log return.
Let $\{r_{t,i}\}_{i=1}^n$ be the daily log returns within the $t$-th month.
Using properties of log returns, we have

$$r_t^m = \sum_{i=1}^n r_{t,i}.$$ 

Assuming that the conditional variance and covariance exist, we have

$$\text{Var}(r_t^m|F_{t-1}) = \sum_{i=1}^n \text{Var}(r_{t,i}|F_{t-1}) + 2 \sum_{i<j} \text{Cov}[(r_{t,i}, r_{t,j})|F_{t-1}],$$

where $F_{t-1} = \text{the information available at month } t-1 \text{ (inclusive).}$

Further simplification possible under additional assumptions.
If \( \{r_{t,i}\} \) is a white noise series, then

\[
\text{Var}(r_t^m | F_{t-1}) = n \text{Var}(r_{t,1}),
\]

where \( \text{Var}(r_{t,1}) \) can be estimated from the daily returns \( \{r_{t,i}\}_{i=1}^n \) by

\[
\hat{\sigma}^2 = \frac{\sum_{i=1}^n (r_{t,i} - \bar{r}_t)^n}{n - 1},
\]

where \( \bar{r}_t \) is the sample mean of the daily log returns in month \( t \) (i.e.,

\[
\bar{r}_t = \frac{\sum_{i=1}^n r_{t,i}}{n}.
\]

The estimated monthly volatility is then

\[
\hat{\sigma}_m^2 = \frac{n}{n - 1} \sum_{i=1}^n (r_{t,i} - \bar{r}_t)^2.
\]

If \( \{r_{t,i}\} \) follows an MA(1) model, then

\[
\text{Var}(r_t^m | F_{t-1}) = n \text{Var}(r_{t,1}) + 2(n - 1)\text{Cov}(r_{t,1}, r_{t,2}),
\]

which can be estimated by

\[
\hat{\sigma}_m^2 = \frac{n}{n - 1} \sum_{i=1}^n (r_{t,i} - \bar{r}_t)^2 + 2 \sum_{i=1}^{n-1} (r_{t,i} - \bar{r}_t)(r_{t,i+1} - \bar{r}_t).
\]

Advantage: Simple

Weaknesses:

- Model for daily returns \( \{r_{t,i}\} \) is unknown.
- Typically, 21 trading days in a month, resulting in a small sample size.

See Figure 1 for an illustration; Ex 3.6 of the text.

**Realized integrated volatility**

If the sample mean \( \bar{r}_t \) is zero, then \( \hat{\sigma}_m^2 \approx \sum_{i=1}^n r_{t,i}^2. \)

\( \Rightarrow \) Use cumulative sum of squares of daily log returns within a month as an estimate of monthly volatility.
Figure 1: Time plots of estimated monthly volatility for the log returns of S&P 500 index from January 1980 to December 1999: (a) assumes that the daily log returns form a white noise series, (b) assumes that the daily log returns follow an MA(1) model, and (c) uses monthly returns from January 1962 to December 1999 and a GARCH(1,1) model.
Apply the idea to *intradaily log returns* and obtain realized integrated volatility.

Assume daily log return $r_t = \sum_{i=1}^{n} r_{t,i}$. The quantity

$$RV_t = \sum_{i=1}^{n} r_{t,i}^2,$$

is called the *realized* volatility of $r_t$.

**Advantages:** simplicity and using intraday information

**Weaknesses:**
- Effects of market microstructure (noises)
- Overlook overnight return

**Further discussion**

1. In-filled asymptotic argument. Let $\Delta$ be the sampling interval, as $\Delta \to 0$, the sample size goes to infinity.

   Under the assumption that the $\Delta$-interval log returns, e.g. 5-minute returns, are independent and identically distributed, then $\sum_{j=1}^{n} r_{t,j}^2$ converges to the variance of the daily log return $r_t$.

2. In practice, however, there are microstructure noises that affect the estimate such as the bid-ask bounce. In fact, it can be shown that as $\Delta$ goes to zero, the observed sum of squares of $\Delta$-interval returns goes to infinity.

**What next?** Two approaches have been proposed:

(a) Optimal sampling interval: Bandi and Russell (2006). Find an optimal $\Delta$. Or equivalently, the optimal sample size $n^*$
= 6.5 hours/\Delta \text{ can be chosen as }
\[ n^* \approx \left[ \frac{Q}{(\hat{\sigma}_{\text{noise}}^2)^2} \right]^{1/3}, \]
where \( Q = \frac{M}{3} \sum_{j=1}^{M} r_{t,j}^4 \) and \( \hat{\sigma}_{\text{noise}}^2 = \frac{1}{M} \sum_{j=1}^{M} r_{t,j}^2 \), where \( M \) is the number of daily quotes available for the underlying stock and the returns \( r_{t,j} \) are computed from the mid-point of the bid and ask quotes.

(b) Subsampling: Zhang et al. (2006). Choose \( \Delta \) between 10 to 20 minutes. Compute integrated volatility for each of the possible \( \Delta \)-interval return series. Then, compute the average.

**Use of Daily Open, High, Low and Close Prices**

Figure 2 shows a time plot of price versus time for the \( t \)th trading day. Define

- \( C_t \) = the closing price of the \( t \)th trading day;
- \( O_t \) = the opening price of the \( t \)th trading day;
- \( f \) = fraction of the day (in interval [0,1]) that trading is closed;
- \( H_t \) = the highest price of the \( t \)th trading period;
- \( L_t \) = the lowest price of the \( t \)th trading period;
- \( F_{t-1} \) = public information available at time \( t - 1 \).

The conventional variance (or volatility) is \( \sigma_t^2 = E[(C_t - C_{t-1})^2|F_{t-1}] \).

Some alternatives:

- \( \hat{\sigma}_{0,t}^2 = (C_t - C_{t-1})^2 ; \)
- \( \hat{\sigma}_{1,t}^2 = \frac{(O_t - C_{t-1})^2}{2f} + \frac{(C_t - O_t)^2}{2(1-f)}, \ 0 < f < 1 ; \)
Figure 2: Time plot of price over time: scale for price is arbitrary.
\[
\begin{align*}
\bullet \hat{\sigma}_{2,t}^2 &= \frac{(H_t - L_t)^2}{4 \ln(2)} \approx 0.3607(H_t - L_t)^2; \\
\bullet \hat{\sigma}_{3,t}^2 &= 0.17 \frac{(O_{t-1} - C_t)^2}{f} + 0.83 \frac{(H_t - L_t)^2}{(1-f)(4\ln(2))}, \quad 0 < f < 1; \\
\bullet \hat{\sigma}_{5,t}^2 &= 0.5(H_t - L_t)^2 - [2 \ln(2) - 1](C_t - O_t)^2, \\
&\quad \text{which is } \approx 0.5(H_t - L_t)^2 - 0.386(C_t - O_t)^2; \\
\bullet \hat{\sigma}_{6,t}^2 &= 0.12 \frac{(O_t - C_{t-1})^2}{f} + 0.88 \frac{\hat{\sigma}_{5,t}^2}{1-f}, \quad 0 < f < 1.
\end{align*}
\]

A more precise, but complicated, estimator \(\hat{\sigma}_{4,t}^2\) was also considered. But it is close to \(\hat{\sigma}_{5,t}^2\).

Defining the efficiency factor of a volatility estimator as

\[
\text{Eff}(\hat{\sigma}_{i,t}^2) = \frac{\text{Var}(\hat{\sigma}_{0,t}^2)}{\text{Var}(\hat{\sigma}_{i,t}^2)},
\]

Garman and Klass (1980) found that \(\text{Eff}(\hat{\sigma}_{i,t}^2)\) is approximately 2, 5.2, 6.2, 7.4 and 8.4 for \(i = 1, 2, 3, 5\) and 6, respectively, for the simple diffusion model entertained.

Define

\[
\begin{align*}
\bullet \ o_t &= \ln(O_t) - \ln(C_{t-1}) \text{ be the normalized open;} \\
\bullet \ u_t &= \ln(H_t) - \ln(O_t) \text{ be the normalized high;} \\
\bullet \ d_t &= \ln(L_t) - \ln(O_t) \text{ be the normalized low;} \\
\bullet \ c_t &= \ln(C_t) - \ln(O_t) \text{ be the normalized close.}
\end{align*}
\]

Suppose that there are \(n\) days of data available and the volatility is constant over the period. Yang and Zhang (2000) recommend the estimate

\[
\hat{\sigma}_{yz}^2 = \hat{\sigma}_o^2 + k \hat{\sigma}_c^2 + (1-k) \hat{\sigma}_{rs}^2
\]
as a robust estimator of the volatility, where

\[
\hat{\sigma}_o^2 = \frac{1}{n-1} \sum_{t=1}^{n} (o_t - \bar{o})^2 \quad \text{with} \quad \bar{o} = \frac{1}{n} \sum_{t=1}^{n} o_t,
\]

\[
\hat{\sigma}_c^2 = \frac{1}{n-1} \sum_{t=1}^{n} (c_t - \bar{c})^2 \quad \text{with} \quad \bar{c} = \frac{1}{n} \sum_{t=1}^{n} c_t,
\]

\[
\hat{\sigma}_{rs}^2 = \frac{1}{n} \sum_{t=1}^{n} \left[ u_t(u_t - c_t) + d_t(d_t - c_t) \right],
\]

\[
k = \frac{0.34}{1.34 + (n + 1)/(n - 1)}.
\]

This estimate seems to perform well.

**Takeaway**

Some alternative approaches to volatility estimation is currently under intensive study. It is rather early to assess the impact of these methods. It is a good idea in general to use more information. However, regulations and institutional effects need to be considered.