Boston University Questrom School of Business

MF793 – Fall 2021
Eric Jacquier

Filtering Methods for Time-Varying Volatility

1. Rolling vs. extending window estimators

2. Issues to address
    - Sample size, horizon
    - Outliers and regime shifts
    - Spurious memory

3. **Risk Metrics** – Declining (EWMA) weights estimator

4. From block sampling to **realized volatility**: high frequency estimators

5. Recall the good news: Variance estimated more precisely with high frequency data

1. Rolling vs. extending window estimator

- **Extending window** estimator:
  - Use all available data at all times
  - Say we start at time t with **n available observations** from **t-n+1** to **t**
  - Every period, we get one new observation, and we can **re-estimate**

---

t-n+1				t	t+1

- Estimate at time t mean $m_t = \sum_{i=0}^{n-1} R_{t-i} \,/\, \mathbf{n}$

  variance $s^2_t = \sum_{i=0}^{n-1} (R_{t-i} - m_t)^2 \,/\, \mathbf{n}$

  sample covariance $s_{AB,t}$ for series $R_{A,t-i}$ and $R_{B,t-i}$

- Re-estimate at time t+1 $m_{t+1} = \sum_{i=0}^{n} R_{t+1-i} \,/\, (\boldsymbol{n}+\mathbf{1})$

  s^2_{t+1}

  $s_{AB,\, t+1}$

  with one more observation

- **Rolling window** estimator

  - We decide on an arbitrary number **n** of observation, e.g.,
    5 years of monthly data
    1 or 2 years of weekly or daily data

  - Every period: add a new observation, remove the oldest, re-estimate

  - => We keep the **window length n** constant

Rolling window length n

at t: uses observations [t-n+1, ... , t]
at t+1: uses observations [t-n+2 , ... , t+1]

| | | | | | | | |

t-n+1 . . . . t

s^2_t

t-n+2 . . . . t+1

s^2_{t+1}

- Note the **overlapping observations** from one estimate to the next ones

- <u>Extending</u> window estimator consistent with:

    - Underlying belief: parameter estimated is constant
      => Old observations remain useful
    - more observations => higher precision

- <u>Rolling</u> window estimator consistent with:

    - Underlying belief: parameter varies with time
      => older observations are less relevant
      ``parameter has changed, we remove old data as time goes''

    - Little guidance on window size apart from common practice
      σ : up to 1 year of daily data
      β : 5 years of monthly data 1 or 2 years of weekly or daily data

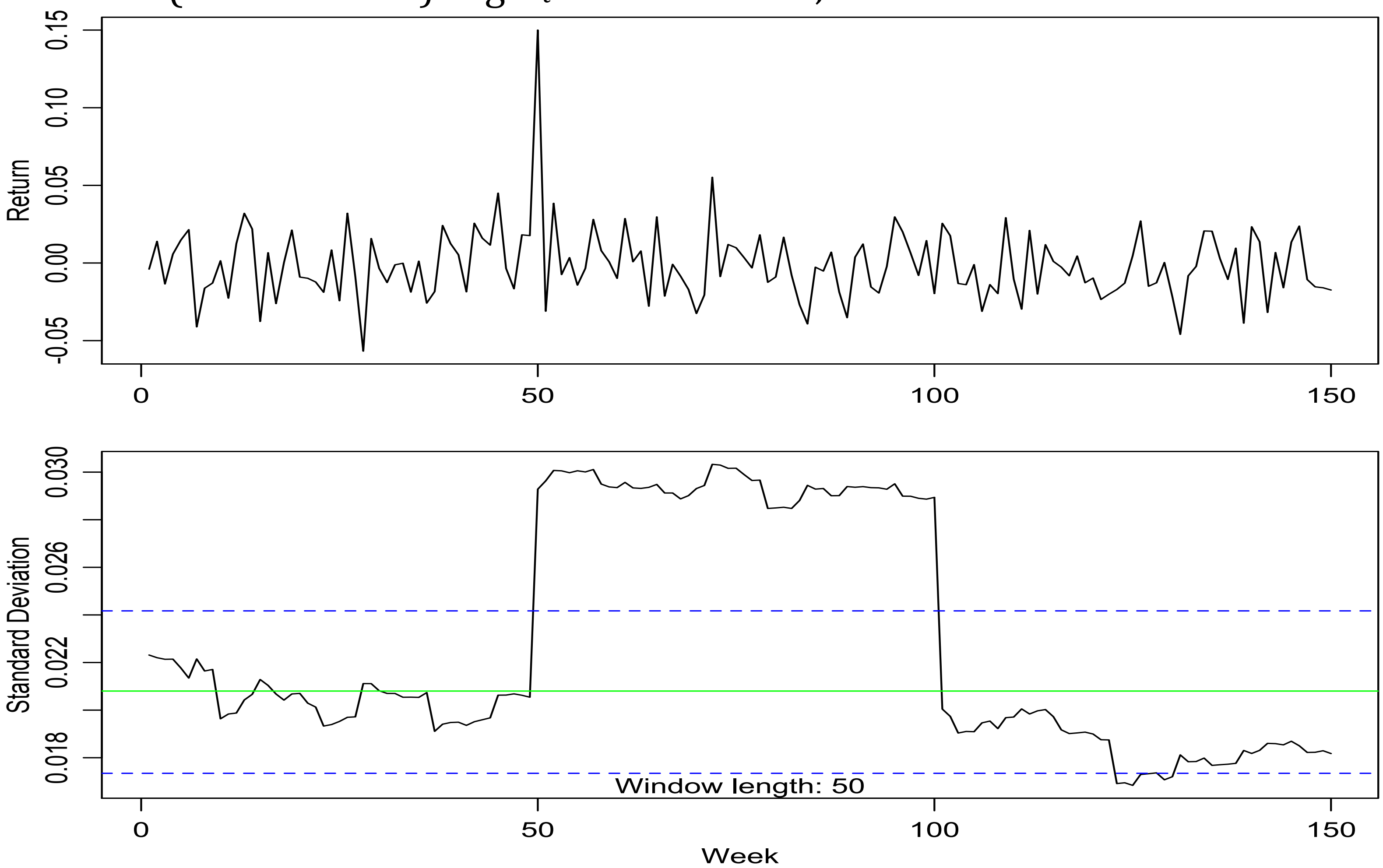
- Rolling or Extending ? The usual dilemma

    - Rolling: less precise but more robust

    - Extending: more precise (if correct) but less robust (if wrong), misspecification risk

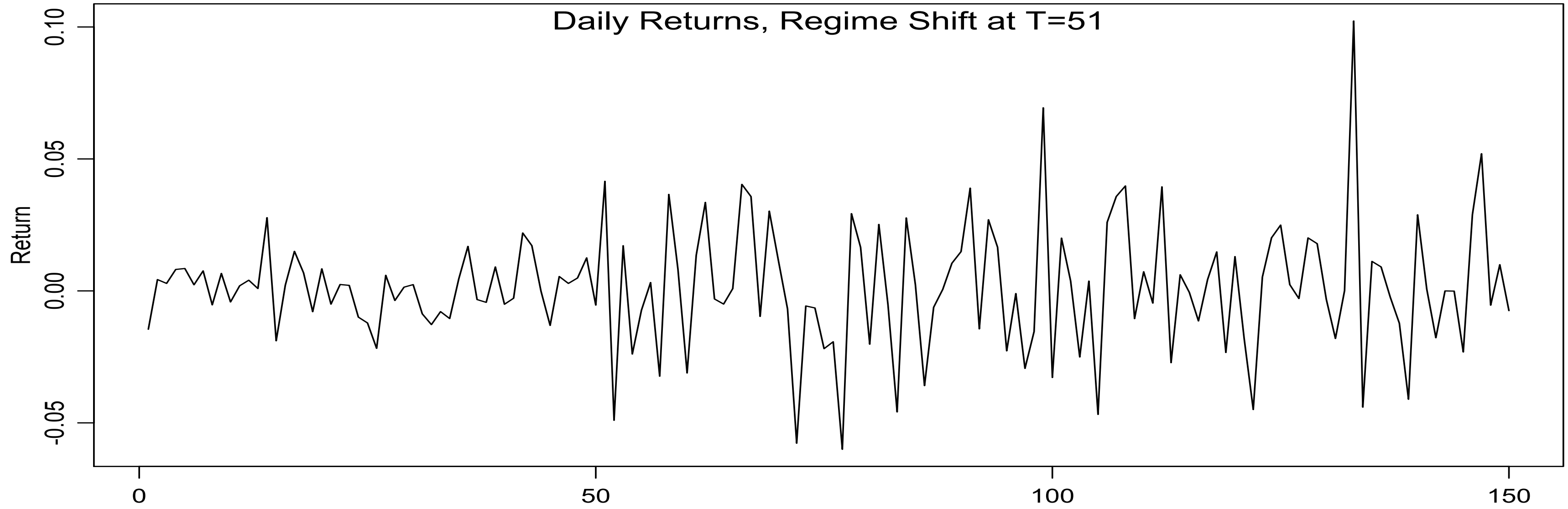
2 Three issues (example with estimating std. dev.)

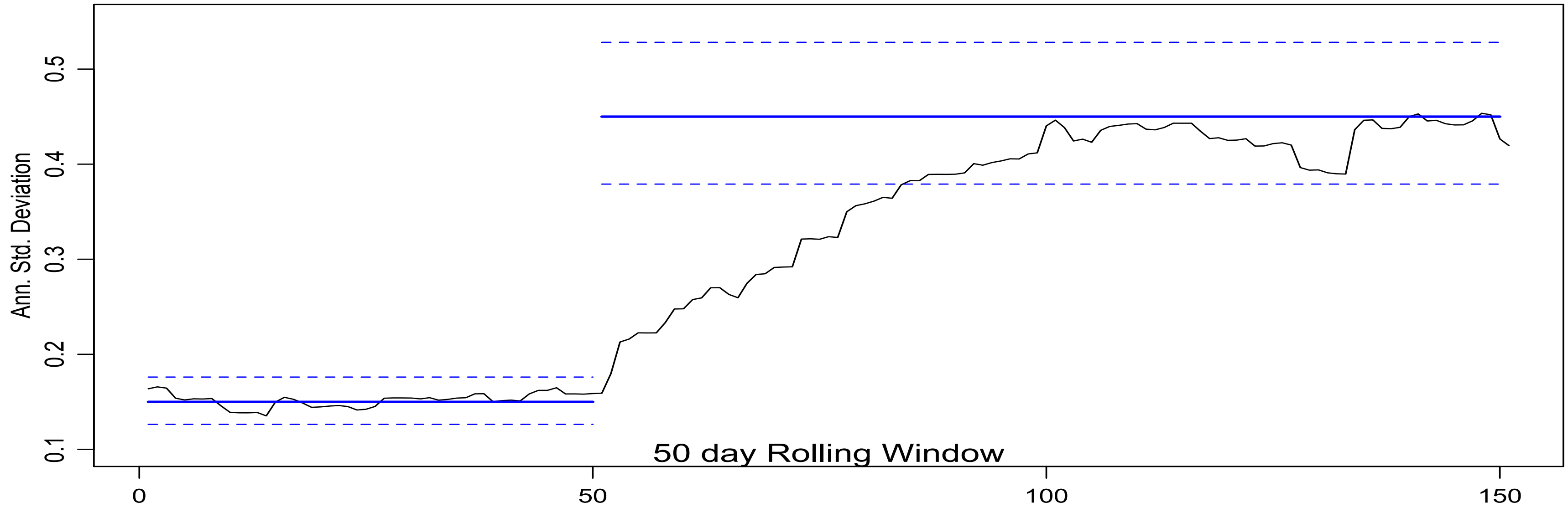
- **Outliers** (extreme data): e.g. s_t around Oct 19, 1987 ?



Outliers come in an out of estimation window abruptly, filter has unrealistic patterns

- **Regime shifts** in true variance: Durable change in true volatility ...





Filter does not incorporate regime shifts quickly, they come in slowly (linearly)

- *Spurious Memory*:

Overlapping moving window creates unwanted memory from one estimate to the next ones

$s^2_t = \sum_{i=0}^{n-1} (R_{t-i} - m)^2 / n$

returns used: [t-n+1, ... , t] ~~t+1~~

$s^2_{t+k} = \Sigma (R_{t-i+k} - m)^2 / n$

returns used: [t+k -n+1, ... t+k] n-k common terms with s^2_t

$s^2_{t+n} = \Sigma (R_{t-i+n} - m)^2 / n$

returns used: [t+n -n+1, ... , t+n]: no common term

=> Causes a spurious linearly declining autocorrelation structure up to lag n even if the true variance is not autocorrelated or is constant.

=> One can't use ACF up to lag n to conclude on the time series process of the true σ_t. .. because of the spurious autocorrelation in s_t.

Spurious memory affects all overlapping estimators

3 Risk Metrics: a filter better adapted to heteroskedasticity

- RM uses declining weights

$$s^2_t = \sum_{i=0}^{n-1} (R_{t-i} - m_{t-i})^2 \; (w_i / \Sigma w_i), \qquad w_i = \lambda^i, \quad 0 < \lambda < 1$$ e.g., 0.9 , 0.81, ...

- No need for an arbitrary finite window

$s_t^2 = \sum (R_{t-i} - m_{t-i})^2 \lambda^i (1-\lambda)$

  - $\Sigma w_i = 1 + \lambda + \lambda^2 + \dots + \lambda^{n-1} = (1-\lambda^n)/(1-\lambda)$
    Weights must sum to 1 for correct scaling
  - $n \to \infty$: $\Sigma\, w_i = 1 / (1-\lambda)$ since $\lambda<1$

- Precision:
  - Equal weights **most** precise if data is homoskedastic (has constant variance)
  - Equal weights **less** precise if volatility changes with time.

- Easy updating rule:

$$s^2_t = \lambda\, s^2_{t-1} + (1-\lambda)\ (R_t - m_t)^2$$

Let $r_t = R_t - m_t$

$s^2_t = (r^2_t + \lambda\, r^2_{t-1} + \lambda^2\, r^2_{t-2} + \ldots)\ (1-\lambda)$

$s^2_{t-1} = (\quad r^2_{t-1} + \lambda\, r^2_{t-2} + \ldots)\ (1-\lambda)$ rewritten at t-1

- **Outliers:** Most effect right away $\lambda\ (1-\lambda)\ (r_{t-1})^2$
  Then weight decays: $\lambda^2\ (1-\lambda)\ (r_{t-2})^2$

- **Regime shift:** new regime taken into account faster since most recent observations have highest weights

4 Realized Volatility (Block Sampling)

1. Say we need estimates at a given frequency. Call it the *low* frequency.

2. Collect data at a frequency **higher** than the frequency needed

   daily if monthly is needed

   5 minutes if daily is needed

$T_r = \int_{t_1}^{t_2} \sigma_t \, dt$

3. Estimate σ_t for each low frequency period using the higher frequency data.

t	t+1	t+2	Low frequency
1 N	N+1 2N	2N+1 3N	
|||||||||||||||||	||||||||||||||||	|||||||||||||||||||||	<- higher frequency data
s_t	s_{t+1}	s_{t+2}	<- low frequency estimate

- Pros:
  - Outliers absorbed right away
  - Regime shifts incorporated right away
  - Subsequent estimates s_1, s_2, do not share common observations:

Block Sampling estimators have no spurious autocorrelation

- Added benefit valid for any estimator based on high frequency returns.

  Recall: $V(x) = E(x^2) - [E(x)]^2$ The mean "shrinks" at the same rate as the variance.

For very high frequencies
we can ignore the estimation of the mean

$$\mathbf{s^2_t} = \mathbf{n} \sum_{i=0}^{n-1} (R_{t-i/n} - \hat{\mu}_t)^2 \,/\, n \approx \sum_{i=0}^{n-1} \mathbf{R^2_{t-i/n}}$$

Why multiply by n ?

Careful to understand the notation, we use the higher frequency returns

Effectively, we compute the sample mean of x^2.

- Cons:

  At ultra-high frequency, the financial data process becomes very complicated, …

  . .. Measurement errors, non-synchronous trading, prices not in equilibrium

  Rule of thumb: Intervals no shorter than 5 minute even for liquid instruments

5 Recall volatility (2^{nd} moment) estimation: precision increases with data frequency

Recall Lecture Note 7, Aggregation and Precision, we did a t-test approach.

- Confidence interval from low frequency data:

$$\left[s_M - 1.96\,\frac{s_M}{\sqrt{2T}}\ ,\ s_M + 1.96\,\frac{s_M}{\sqrt{2T}}\right] \qquad [1]$$

- Confidence interval from high frequency data (N points per low frequency interval):

$$\left[s_D - 1.96\,\frac{s_D}{\sqrt{2TN}}\ ,\ s_D + 1.96\,\frac{s_D}{\sqrt{2TN}}\right] \qquad [2]$$

- Convert the high frequency interval [2] into low frequency to allow for comparison with [1].
  Both lower and upper bounds in [2] are high frequency standard deviations: to be multiplied by $\sqrt{N}$ for annualization into low frequency. [2] becomes

$$\left[s_M - 1.96\,\frac{s_M}{\sqrt{2T\boldsymbol{N}}}\ ,\ s_M + 1.96\,\frac{s_M}{\sqrt{2T\boldsymbol{N}}}\right] \qquad [3]$$

Precision increases by $\sqrt{N}$ when the sampling frequency increases by N