FEN_842 Risk Measurement

Lecture 4
Risk Measure Evaluation

Winfried G. Hallerbach PhD

March 26, 2025

Make an impact

Outline of lecture 4

Risk measure evaluation

- **From building to evaluating risk models**
- Evaluating volatility
- Evaluating VaR
- Risk horizon & time scaling
- Using overlapping observations

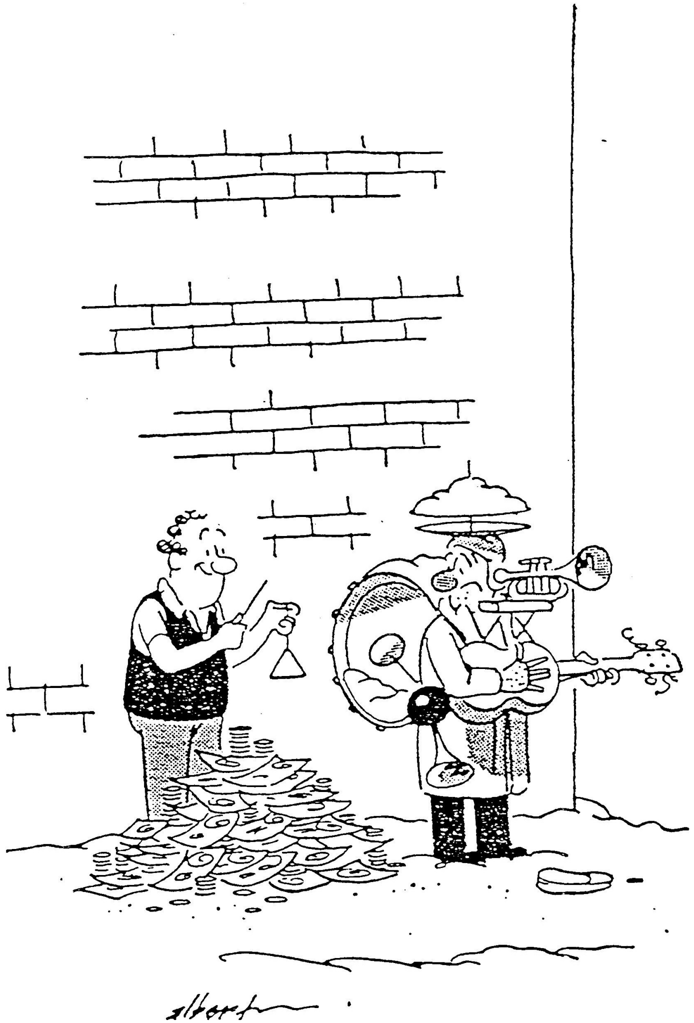
Specifying risk models

Thou shalt not complicate

- strive for **parsimony** :

  "Occam's Razor", "low fat modeling"

- "Make everything as simple as possible, but not simpler"   (Albert Einstein)



Example :

- EWMA is "poor man's GARCH"
- **pragmatic** approach to state-of-the-art volatility modelling
- simple, **robust**, "low fat" (small number of parameters)

Teaser

Parsimony

portfolio

asset :	Equities	Tsies	CorpIG	CorpHY	AbsReturn	Oil
weight :	50%	30%	5%	0%	10%	5%

- portfolio composition
- asset return data

return data

N obs	date	Equities	Tsies	CorpIG	CorpHY	AbsReturn	Oil
1	1/3/1995	-0.2400	-0.2733	-0.2136	0.0572	0.8810	-1.8008
2	1/4/1995	0.3500	0.4110	0.3686	0.1627	0.6150	0.6304
3	1/5/1995	-0.0300	-0.4099	-0.2310	0.2019	-0.0290	1.1390
:	:	:	:	:	:	:	:
7537	12/9/2024	-0.7230	-0.4049	-0.2226	0.0092	0.5980	0.1021
7538	12/10/2024	-0.3530	-0.1618	-0.1142	-0.0369	0.4160	0.2913
7539	12/11/2024	0.8870	-0.3230	-0.1809	0.0351	-0.6760	2.4982

Q : *what type of returns ?*

Q : *how to calculate portfolio vola ?*

Q : *how to generate portfolio risk statistics ?*

→ bottom-up : aggregate covariances
→ portfolio aggregation : first aggregate returns into one portfolio series !

More complex models ?

The case of multivariate GARCH : MGARCH

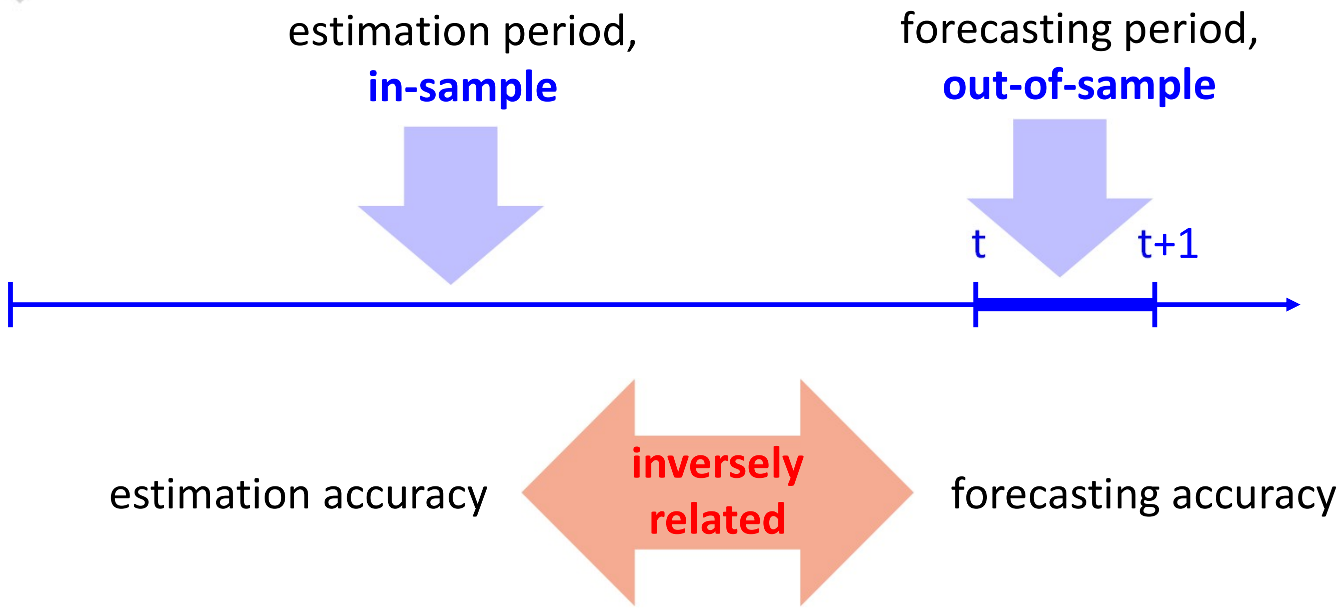
- investment portfolio context
- first use multivariate MGARCH for individual assets & next aggregate to portfolio level

?

first aggregate assets to portfolio level & next use portfolio return vector ?

- Brooks & Persand [2003, p.22] :
  "Given the complexity, estimation difficulties, and computer-intensive nature of MGARCH modelling, we conjecture that unless the conditional covariances are required, **the estimation of multivariate GARCH models is not worth while**. In the context of portfolio volatility, more accurate results can be obtained by **aggregating the portfolio constituents into a single series, and forecasting that**, rather than modelling the individual component volatilities and the correlations between the returns."

In-sample vs out-of-sample



- the stronger a model is fitted to in-sample data → the lower its forecasting ability
  - . more parameters (less dgf)
  - . more complex relations (non-linearities)
- over-complication leads to over-fitting or "**noise fitting**"

More complex models ?

"GARCH Zoo"

- don't more complex GARCH models perform better ?
- capturing vs forecasting cond'l variance
  in-sample ↔ out-of-sample
- "**curse of dimensionality**" :
  more parameters to estimate,
  info or noise ?

Hansen & Lunde [2005] & Stamos [2023] compare various models :

- more lags than in GARCH(1,1) do not help
- GARCH(1,1) $\approx$ more complex specifications
- GARCH(1,1) $\approx$ EWMA
- incorporating vola asymmetry helps in-sample,
  but symmetric model is not inferior out-of-sample

ARCH:	$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2$
GARCH:	$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2$
IGARCH	$\sigma_t^2 = \omega + \varepsilon_{t-1}^2 + \sum_{i=2}^{q} \alpha_i (\varepsilon_{t-i}^2 - \varepsilon_{t-1}^2) + \sum_{j=1}^{p} \beta_j (\sigma_{t-j}^2 - \varepsilon_{t-1}^2)$
Taylor/Schwert:	$\sigma_t = \omega + \sum_{i=1}^{q} \alpha_i \lvert\varepsilon_{t-i}\rvert + \sum_{j=1}^{p} \beta_j \sigma_{t-j}$
A-GARCH:	$\sigma_t^2 = \omega + \sum_{i=1}^{q} [\alpha_i \varepsilon_{t-i}^2 + \gamma_i \varepsilon_{t-i}] + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2$
NA-GARCH:	$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i (\varepsilon_{t-i} + \gamma_i \sigma_{t-i})^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2$
V-GARCH:	$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i (e_{t-i} + \gamma_i)^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2$
Thr.-GARCH:	$\sigma_t = \omega + \sum_{i=1}^{q} \alpha_i [(1-\gamma_i)\varepsilon_{t-i}^{+} - (1+\gamma_i)\varepsilon_{t-i}^{-}] + \sum_{j=1}^{p} \beta_j \sigma_{t-j}$
GJR-GARCH:	$\sigma_t^2 = \omega + \sum_{i=1}^{q} [\alpha_i + \gamma_i I_{(\varepsilon_{t-i}>0)}] \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2$
log-GARCH:	$\log(\sigma_t) = \omega + \sum_{i=1}^{q} \alpha_i \lvert e_{t-i}\rvert + \sum_{j=1}^{p} \beta_j \log(\sigma_{t-j})$
EGARCH:	$\log(\sigma_t^2) = \omega + \sum_{t=1}^{q} [\alpha_i e_{t-i} + \gamma_i (\lvert e_{t-i}\rvert - E\lvert e_{t-i}\rvert)] + \sum_{j=1}^{p} \beta_j \log(\sigma_{t-j}^2)$
NGARCH:[a]	$\sigma_t^\delta = \omega + \sum_{i=1}^{q} \alpha_i \lvert\varepsilon_{t-i}\rvert^\delta + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^\delta$
A-PARCH:	$\sigma^\delta = \omega + \sum_{i=1}^{q} \alpha_i [\lvert\varepsilon_{t-i}\rvert - \gamma_i \varepsilon_{t-i}]^\delta + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^\delta$
GQ-ARCH:	$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i} + \sum_{i=1}^{p} \alpha_{ii} \varepsilon_{t-i}^2 + \sum_{i<j}^{p} \alpha_{ij} \varepsilon_{t-i}\varepsilon_{t-j} + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2$
H-GARCH:	$\sigma_t^\delta = \omega + \sum_{i=1}^{q} \alpha_i \delta \sigma_{t-i}^\delta [\lvert e_t - \kappa\rvert - \tau(e_t - \kappa)]^\nu + \sum_{i=1}^{p} \beta_j \sigma_{t-j}^\delta$

Back testing

Back testing principles

Back testing :

- "retrodiction" : predicting the past
- the process of analyzing and evaluating **ex-ante** risk estimates using **historical** or simulated data
- "how would your risk estimator have performed when used in the past ?"

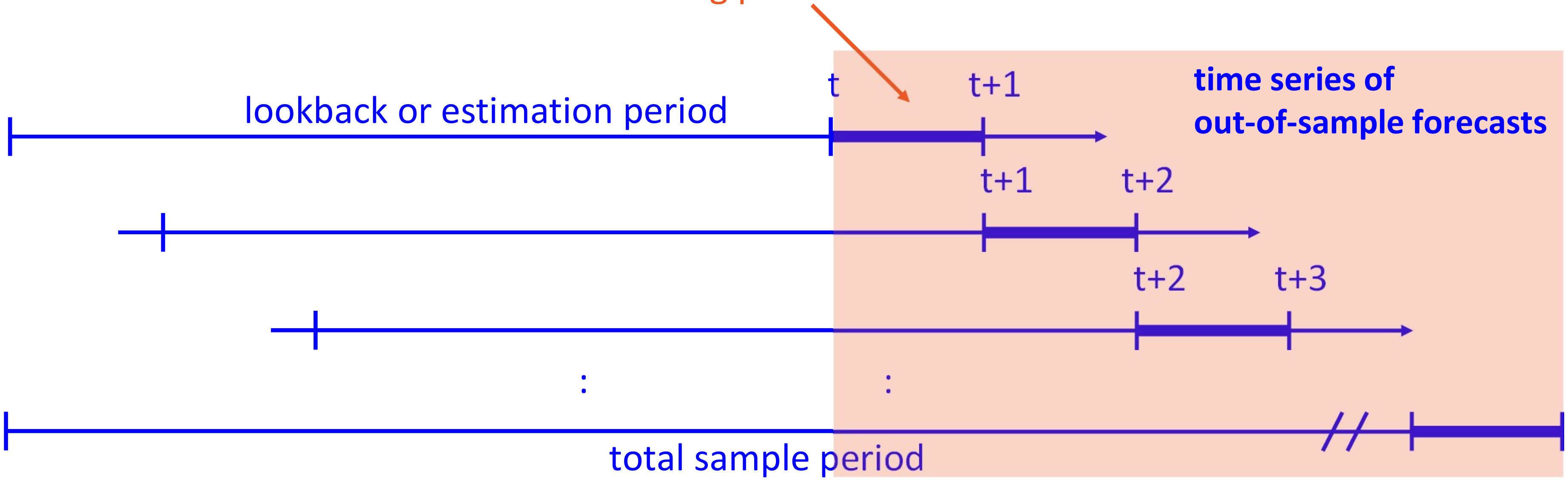
Principles :

- use a **representative sample period** :
  - of sufficient length so as to reflect different market conditions, at least one full "market cycle"
  - assumption that the past is a mirror of the future
- **no cheating :** no forward looking (look ahead) bias, do not use future or full sample information
- use clear **evaluation criteria**
- in principle : use data set only once when optimizing parameters → avoid data snooping

Evaluating risk estimates

Setting up a proper back test

- choosing proper risk measures : regulation
  portfolio type – see lecture 2
  internal preference : 1st, 2nd line of defense
  client preference
- determine relevant risk horizon = forecasting period



Outline of lecture 4

Risk measure evaluation

- From building to evaluating risk models
- **Evaluating volatility : 4 criteria**
- Evaluating VaR
- Risk horizon & time scaling
- Using overlapping observations

Volatility evaluation criteria

Testing models & parameter settings

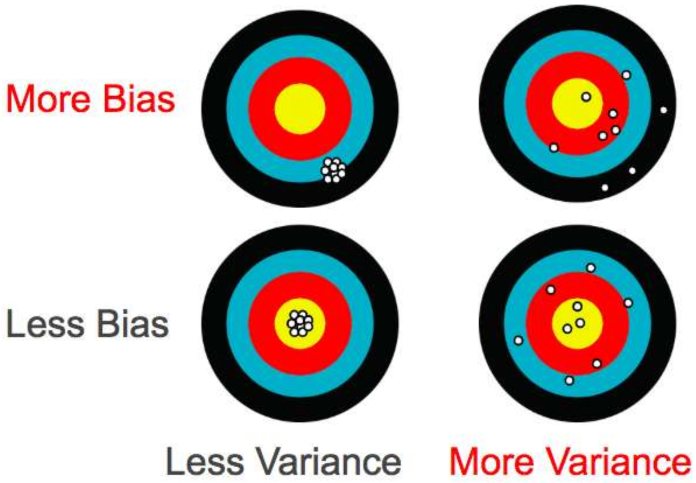
Given a risk measure & risk predictions over the back test sample period :

1. what is the performance **on average** ?
   → bias : long term over-/under-prediction

2. what is the performance **over time** ?
   → variability : sub-periods of over-/under-prediction

3. what is the improvement in **distribution** ?
   → do z-scores have a better behaved distribution ?
   → less fatter tails

4. what is the **relative** performance ?
   → best single criterion to rank & test across models



Evaluation criterion 1 : avge bias

Long term precision : performance on average

Recipe bias test :

- normalize returns with conditional volatility : $z_{t+1} = \dfrac{r_{t+1}}{\sigma_{t+1|t}}$ **= "z-scores"** $\sim N(0,1)$

- **criterion :** if volatility estimates are unbiased, then ***stdev(z)* = 1** over the back test period *M*

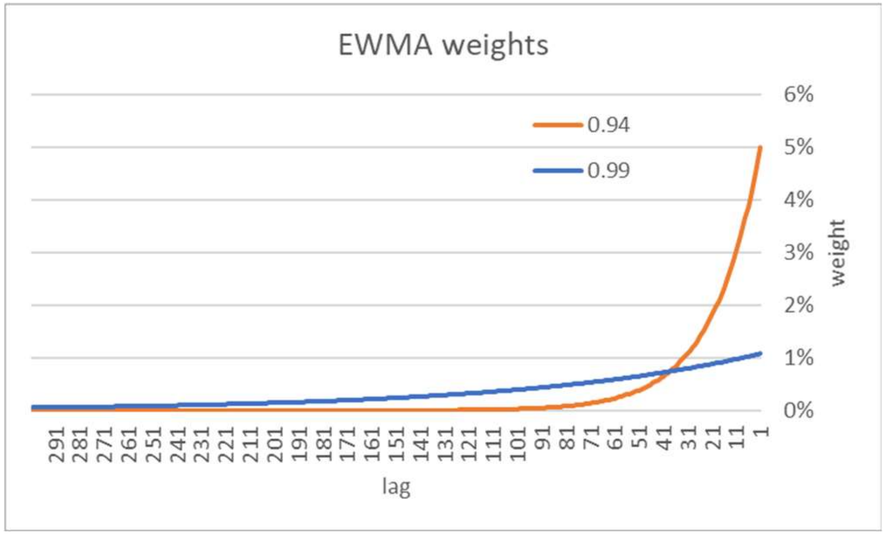
- the standard error of the standard deviation is approximately : $\boxed{s.e.(\hat{\sigma}) \approx \dfrac{1}{\sqrt{2M}} \cdot \sigma}$

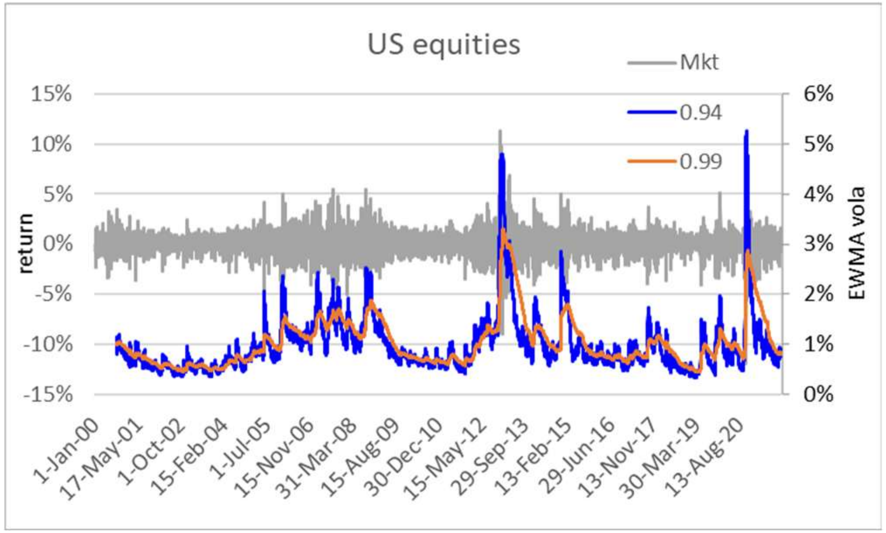
- hence, a 95% confidence interval is : $\left[1 - \sqrt{\dfrac{2}{M}}\right] \cdot stdev(z) \;\leq\; stdev(z) \;\leq\; \left[1 + \sqrt{\dfrac{2}{M}}\right] \cdot stdev(z)$

Evaluation criterion 1 : avge bias

Long term precision : US Equity example revisited

- Kenneth French data library, US equity market factor
- daily total returns, 2-Jan-1990 – 29-Oct-2021 (8,021 observations)
- consider two EWMAs with persistence of 0.94 and 0.99
- fixed T=260 days estimation window → re-scale EWMA weights to sum to unity





Evaluation criterion 1 : avge bias

Long term precision : US Equity example revisited

- normalize returns with conditional volatility forecasts
- calculate standard deviation of these z-scores :

z-scores	**0.94**	**0.99**
stdev	1.05	1.02
95% conf	1.033:1.067	1.004:1.037

- **bias :** on average, both EWMAs slightly **under**-estimate volatility
- *note :* stdev(z) > 1 indicates under-estimation of volatility,
  stdev(z) < 1 indicates over-estimation of volatility

- number of predictions = 8,021 – 260 = 7,761
- hence, a 95% confidence interval is : $0.984 \cdot stdev(z) \leq stdev(z) \leq 1.016 \cdot stdev(z)$
- so both EWMAs have a small but **significant negative bias**

Evaluation criterion 2 : bias over time

Short term precision : performance over sub-periods

- **bias** test indicates the **average** over-or under-estimation of volatility
- still, over sub-periods there can be considerable bias
- "with my head in the oven & my feet in the freezer, I'm comfortable on average"

solution :

- consider *stdev*(*z*) over **rolling windows**
- how long should this rolling window be ? "not too long, not too short"

  shorter : better insight in performance to capture volatility dynamics
  longer : higher precision of *stdev*(*z*)

- choose between 1 - 3 years for daily / weekly data
- **criterion :** the smaller the time variation in the rolling window *stdev*(*z*) the better

Evaluation criterion 2 : bias over time

Short term precision : performance over sub-periods

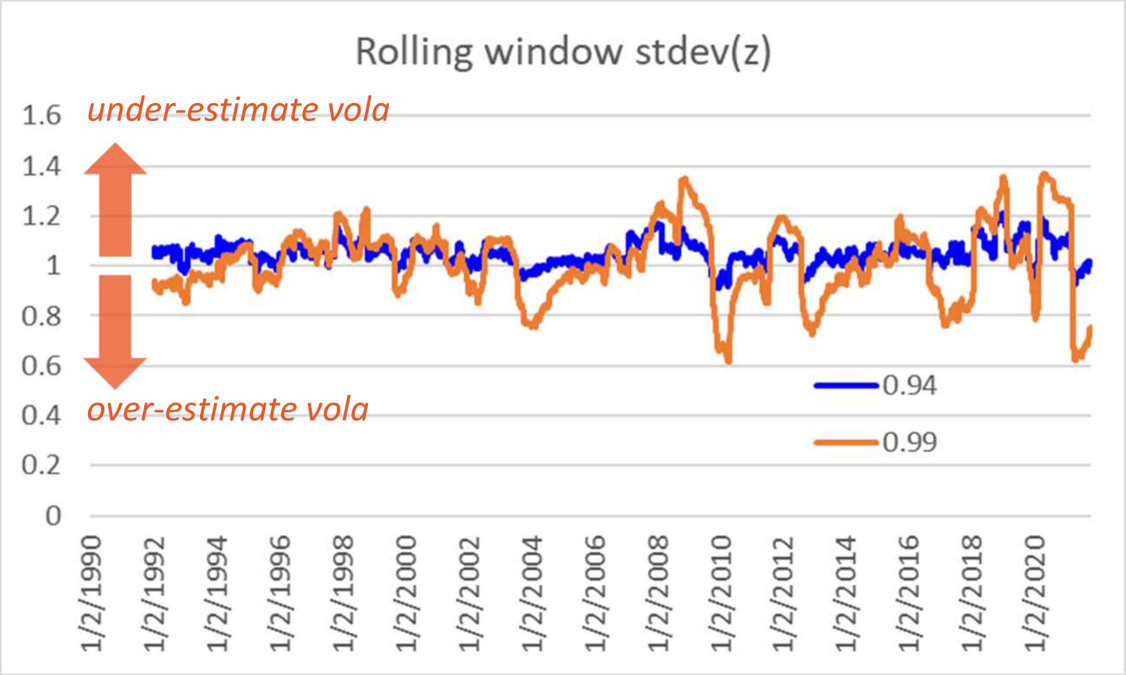
- here, we choose rolling windows of 260 days to calculate a times series of *stdev*(*z*) :

- measure variation by **Mean Rolling Absolute Deviation from unity** :

$$MRAD = \frac{1}{M}\sum_{i=1}^{M}\left|stdev(z_i)-1\right|$$

- EWMA 0.94 better captures vol dynamics

260d ***stdev*(*z*)**	**0.94**	**0.99**
avge abs deviation from 1	0.06	0.12



Evaluation criterion 3 : distribution

Reduction in heteroskedasticity

- time-varying volatility → mixture distribution → fat tails → excess kurtosis
- normalize returns with conditional volatility predictions → z-scores
- **criterion :** the smaller the kurtosis of the z-scores, the better the volatility estimate
- also relevant for volatility-weighted strategies : returns are scaled *z*'s

- calculate excess kurtosis :

	returns	z-scores	
	Mkt	**0.94**	**0.99**
kurtosis	10.51	2.72	3.68

- **improvement in distribution** :
  large reduction in **kurtosis**, both EWMAs capture time-varying volatility
- 0.94 is slightly better

Q : *why would 0.94 be better than 0.99 ?*

Evaluation criterion 3 : distribution

What about skewness ?

- the conventional skewness statistic :

$$skewness = E\left[(r-\overline{r})^3\right] / \sigma^3$$

"positive skewness → right tail"
"negative skewness → left tail"

- however : sensitive to extreme returns, large sampling error, may give wrong information

- alternative : **robust skewness** measure :

$$robust\ skewness \;=\; \frac{avge(r) - median(r)}{stdev(r)}$$

US Equity example	**returns**	**z-scores**	
	Mkt	**0.94**	**0.99**
skewness	-0.27	-0.62	-0.59
robust skewness	-0.028	-0.041	-0.036

- surprising : slight increase in negative skewness,
  usually, volatility-targeting mitigates kurtosis ànd skewness

Evaluation criterion 4 : relative

Evaluating relative performance

- look for the best **single** criterion to **rank** models, allows for **testing** of significant differences
- it is robust for using a variance proxy r^2 instead of the true unobservable variance

relative performance : Quasi Likelihood, Patton [2011]

- defined as the logarithm of the joint "likelihood" function of the data given the predicted volatilities
- start again with the *z*-scores
- calculate the time series of the Quasi Likelihood statistic : $QL_t = \ln\left(z_t^2\right) - z_t^2$

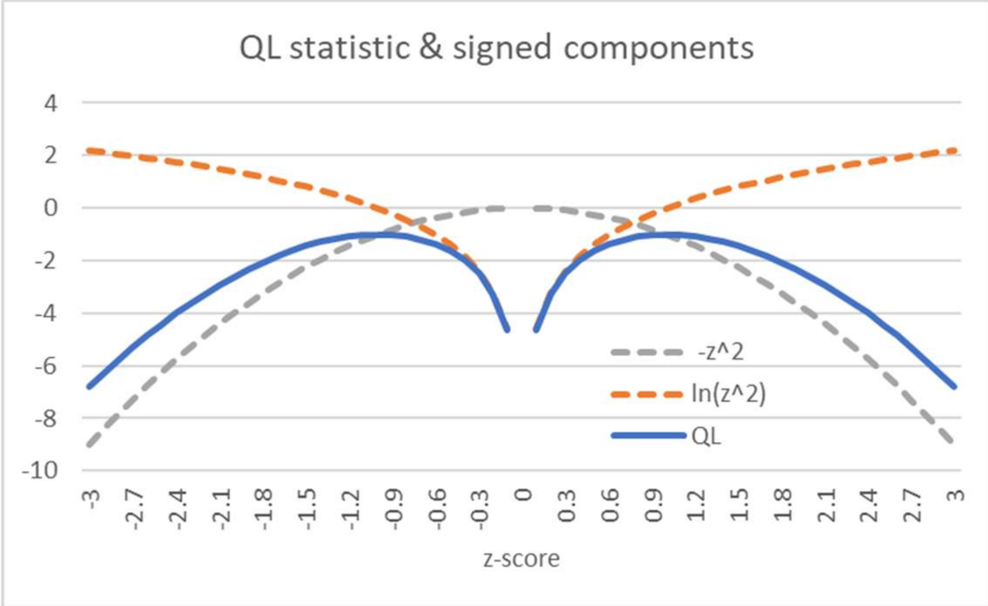
you don't need to know this formula for the exam

- **criterion :** select the model that has the highest likelihood, i.e. **largest sum** of QL_t's : $\max \sum_{t \in T} QL_t$
- **test for comparing** models A and B : calculate difference statistic : $d_t^{A-B} = QL_t^A - QL_t^B$
- calculate average over time series : $\overline{d}^{A-B} = \sum_{t \in T} d_t^{A-B}$ and perform *t*-test

Evaluation criterion 4 : relative

Looking at the two components

- Quasi Likelihood statistic : $QL_t = \ln\left(z_t^2\right) - z_t^2$



- **maximum** *QL* for |z| = 1
- two components :

  logarithm penalizes for smaller |z|
  → prediction too high

  quadratic penalizes for larger |z|
  → prediction too low

  larger penalty for **under**-estimation of volatility

- test for comparing models A and B,

  calculate **difference statistic** : $d_t^{A-B} = QL_t^A - QL_t^B$

Evaluation criterion 4 : relative

US Equity example cont'd

select the model that maximizes likelihood :

- sum of QL_t's is max for 0.94 EWMA

EWMA	0.94	0.99	0.94-0.99
sum *QL*	-19681	-20385	
avge(*d*)			0.0909
t(*d*)			6.14

is the **difference** in performance significant ?

- *t*-test for comparing 0.94 against 0.99
- calculate difference statistic : $d_t^{0.94-0.99} = QL_t^{0.94} - QL_t^{0.99}$
- calculate average & stdev of time series of d_t
- calculate stdev of avge(d_t) : $stdev(\bar{d}) = \dfrac{stdev(d_t)}{\sqrt{T}} = \dfrac{1.3023}{87.99} = 0.0148$
- calculate *t*-statatistic : $t(\bar{d}) = \dfrac{\bar{d}}{stdev(\bar{d})} = \dfrac{0.0909}{0.0148} = 6.14$ → **0.94 is better**

Recap volatility evaluation criteria

Evaluating volatility estimates

1. **long-term precision :** full-sample bias
   . use ex-ante volatilities to normalize the subsequent return → z-scores
   . full-sample standard deviation of z's should be as close as possible to unity

2. **short-term precision :** time series bias
   . over rolling windows, calculate the standard deviation of z's
   . the mean rolling absolute deviation from unity should be as small as possible
   alternative : calculate
   #windows in which volatility is over-estimated : stdev(z) < 1
   #windows in which volatility is under-estimated : stdev(z) > 1

3. **improvement in distribution :**
   skewness, notably reduction in kurtosis

4. **relative :** maximize sum of : $QL_t = \ln\left(z_t^2\right) - z_t^2$, test on differences in QL

Outline of lecture 4

Risk measure evaluation

- From building to evaluating risk models
- Evaluating volatility
- **Evaluating VaR :   Type I & II errors**
- Risk horizon & time scaling
- Using overlapping observations

VaR evaluation criterion

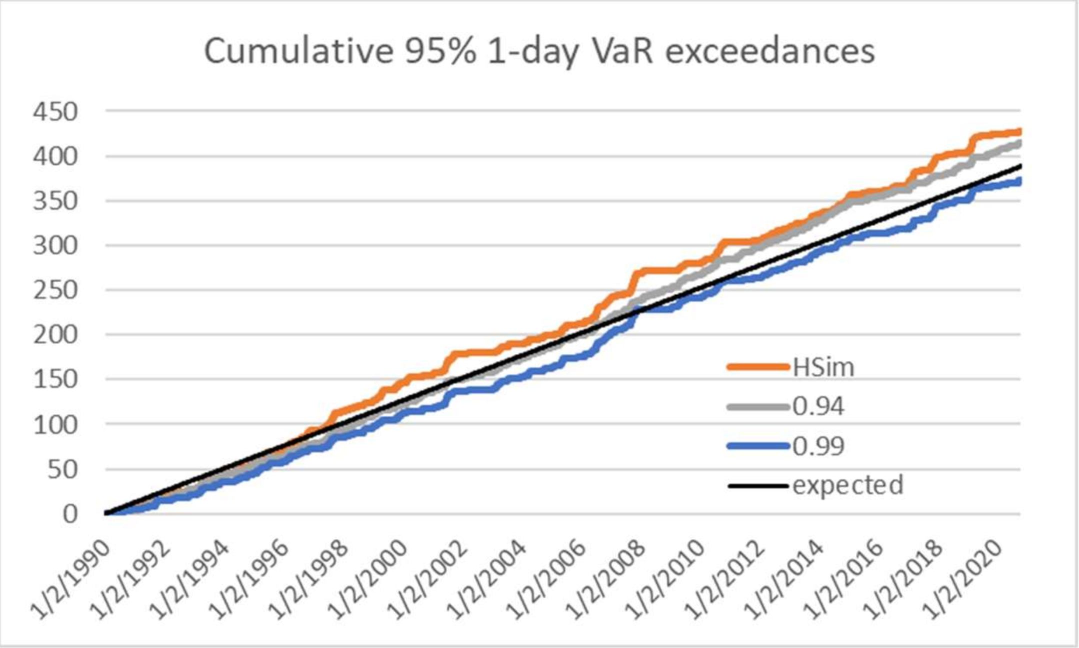
Parametric (0.94, 0.99) against HSim

- the c% VaR is defines as the loss that will not be exceeded in c% of the cases
- evaluate VaR realizations by looking at the number of **exceedances** : $r_t < -VaR_t$

US Equity example 1990-2021 :

- 260-day lookback, 95% 1-day VaR :

exceedances p.a.	0.94	0.99	HSim
expected	13	13	13
on average	13.9	12.5	14.3



- HSim is worst
- EWMA 0.99 VaR looks best, overall
- but sharp increases in # exceedances in crisis periods : 2001, 2008, 2020 ...

VaR evaluation criterion

Zooming in on GFC 2007-2008

- EWMA 0.94 now performed best :

exceedances p.a.	**0.94**	**0.99**	**HSim**
expected	13	13	13
on average	13.9	12.5	14.3
2007-2008			
expected	13	13	13
on average	20.1	27.3	28.9

- conjecture : 0.99 peforms well because it maintains longer a higher vola
  0.94 adapts quicker to higher vola regime

- why did HSim perform so badly ? higher imprecision of estimate (only quantile)
  basically 1-year **equally-weighted** observations (ghost)

VaR evaluation criterion

A balancing act

- test the hypothesis H_0 : "our VaR model is correct"
- reject H_0 → positive test result
  accept (fail to reject) H_0 → negative test result
- you can make two types of errors :

	model	
decision	correct = H_0	incorrect
accept	√	Type II error
reject	Type I error	√

- **Type I error :** "false positive" : when you are too strict
- **Type II error :** "false negative" : when you are too lenient
- low Type I ànd Type II error → powerful test

VaR evaluation criterion

Kupiec' Proportion of Failures test

95% (p = 0.05) non-rejection test confidence regions :

- the higher the VaR confidence level c, the more difficult it is to detect incorrect VaR model
- for c = 99%, there is no way to reject a model that is too conservative

252 day lookback period **VaR confidence level**	***p* = 0.05** **non-rejection region**
99.0%	N < 7
97.5%	2 < N < 12
95.0%	6 < N < 20
92.5%	11 < N < 28
90.0%	16 < N < 36

In our US Equity example, over the last year :

- volatility has decreased
- slow estimators 0.99 & HSim overshoot VaR
- HSim is rejected with 95% confidence
- the more adaptive 0.94 is best

exceedances	**0.94**	**0.99**	**HSim**
expected	13	13	13
actual	12	7	**4**

Outline of lecture 4

Risk measure evaluation

- From building to evaluating risk models
- Evaluating volatility
- Evaluating VaR
- **Risk horizon & time scaling**
- Using overlapping observations

Risk horizon

Depends on type portfolio

Trading portfolios :

- characterized by short holding periods – (fraction of) days
- risk horizon should correspond to the longest period
  needed for an orderly portfolio liquidation

Investment portfolios :

- **investment horizon :**
  period of time that an investor is willing to hold the portfolio, typically 3-5 years,
  but the horizon for evaluating performance and risk is much shorter
- **risk horizon :**
  linked to portfolio dynamics : hedge fund (intra-day) vs enhanced indexing (monthly)
  benchmark indices are typically rebalanced quarterly (but some semi-annually or annually)
  regulators use 1-month / 20 day horizon (UCITS)

Risk horizon

Investment portfolios

Risk horizon vs investment horizon :

- limiting risk on a short horizon helps restricting the risk level in both the short ànd the long run
- within the investment horizon, risk limits can be used to build a confidence interval of monthly returns

Portfolio management vs risk oversight : 1st vs 2nd line of defense :

- fund reporting is typically monthly, so internal limits can have monthly horizon
- portfolio managers can have shorter risk horizon depending on their decision / trading horizon, or the expected holding period of a position
- **fundamental managers** can be required to unwind a position → risk limits are **restrictive**
- for **quant strategies**, formal risk limits should be **accommodative** :
  it should be possible to run the strategy within the posed risk limits

Match risk horizon & return frequency

M = monthly, W = weekly, D = daily

- Observation interval of returns = length of each time period over which returns are computed
  → if you use M data, the observation interval is M
- **If the risk horizon is D/W/M, it makes sense to use D/W/M returns**
- All statistics you compute from M data automatically apply to a M horizon.
  These statistics include the mean, volatility, VaR, ES, LPMs etc.
  → if you use M data, the volatility is by definition a M volatility
- Next, you can annualize the statistic, express it on an annual basis. This **annualization** is purely **cosmetic**, the statistic remains M
- You can also time-aggregate the statistics to an annual horizon. This **time aggregation** is more complex, it involves compounding statistics over longer time horizons so that it effectively lengthens the observation interval → requires **parametric assumptions**
- Note the subtle difference between an **annualized volatility** and an **annual volatility**

Time-aggregating volatility

More complex than you thought ...

- log returns are **additive** over time → $``\sqrt{T} - rule"$ for volatility
- discrete returns aggregate **multiplicatively** over time : $1 + r_{annual} = \prod_{t=1}^{260} \left(1 + r_{daily,t}\right)$
- assume returns are ***i.i.d.* distributed**
- we start from a horizon of 1 period, with mean return m_1 and variance var_1
- aggregating from 1 to T periods, the variance over T periods becomes :

$$\mathrm{var}_T = \left[\mathrm{var}_1 + (1 + m_1)^2\right]^T - (1 + m_1)^{2 \cdot T}$$

you don’t have to know this formula for the exam

In our US Equity example :

	daily	annualized	annual
mean	0.0476%	12.37%	13.16%
stdev	1.1389%	18.36%	20.95%

Outline of lecture 4

Risk measure evaluation

- From building to evaluating risk models
- Evaluating volatility
- Evaluating VaR
- Risk horizon & time scaling
- **Using overlapping observations**

Problems with longer return horizons

D = daily, W = weekly, M = monthly

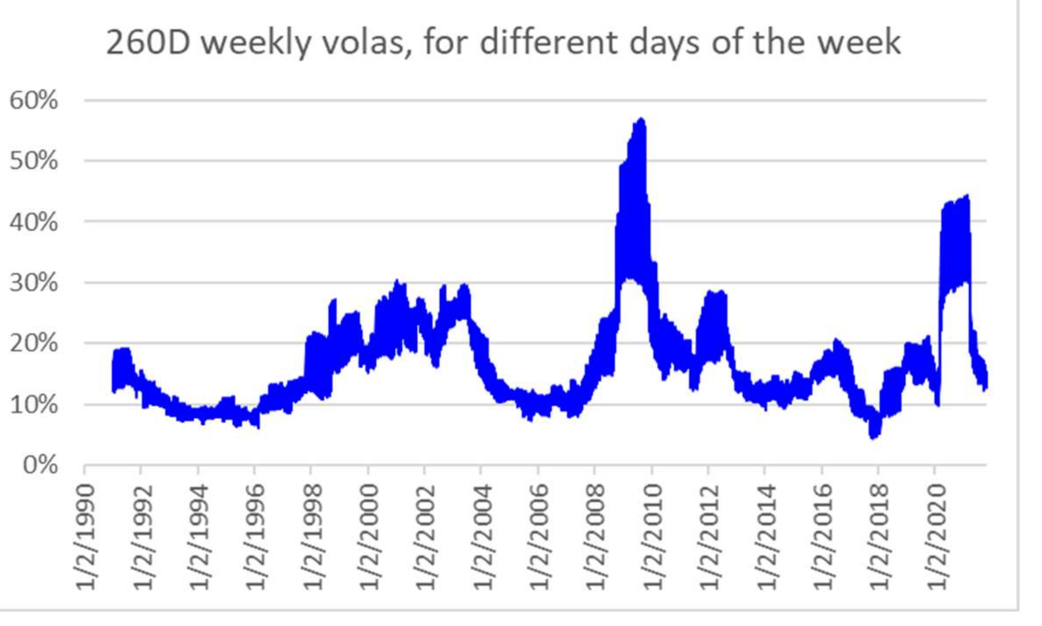
For a M (or W) risk horizon, we should ideally use M (or W) returns

- **problem 1 :** risk monitoring on daily basis is problematic

  . what if we shift window of W/M returns on a D basis ? → **large instability** of estimates

- for **each day** in the sample, estimate the **non-overlapping weekly volatility**, using a 260-day EW window

  shown are weekly volatilities
  based on : Mon - Mon
  Tue - Tue
  Wed - Wed
  Thu - Thu
  Fri - Fri returns
  of US Equities



Problems with longer return horizons

D = daily, W = weekly, M = monthly

For a M (or W) risk horizon, we should ideally use M (or W) returns

- **problem 2 :** for a sufficient number of observations, M or W data require a **long look-back period**
  . what if we use higher frequency (D) data to increase forecasting accuracy ?
- **problem 3 :** this requires **time-aggregating** the risk statistic from D to W / M horizon

why is time scaling of risk metrics so difficult ?

- the underlying distributions change over time → **no i.i.d.**
- cond'l volas change over time → GARCH, EWMA → the $\sqrt{T}$ rule **magnifies** vola fluctuations
- D returns can suffer from spurious serial correlation caused by **non-synchroneities** :
  . a global portfolio that extends over different time zones
  . infrequent trading of some assets in the portfolio,
  this further complicates time aggregation

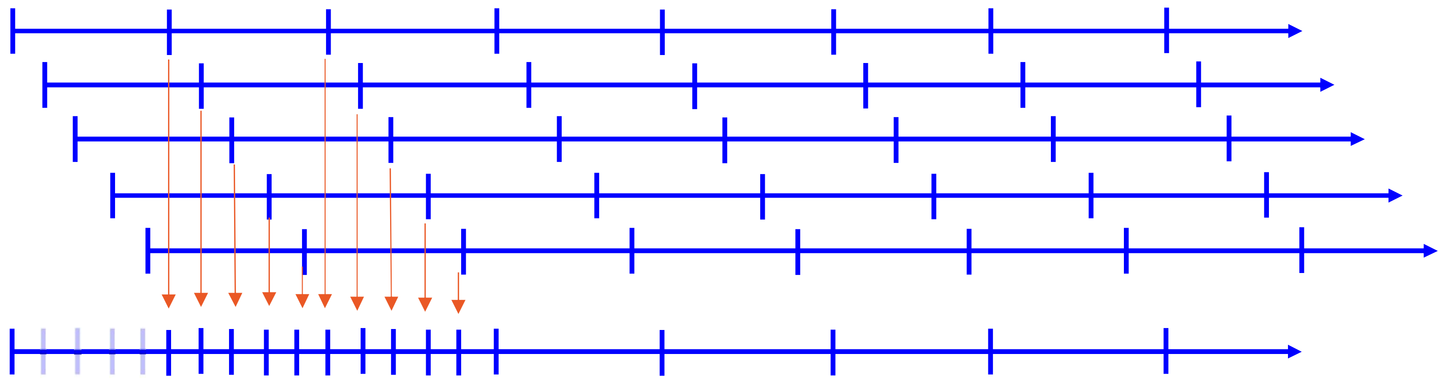
Problems with longer return horizons

D = daily, W = weekly, M = monthly

For a M (or W) risk horizon, we should ideally use M (or W) returns

a way to **solve problems 1-3** is to use **daily overlapping** W or M data :

- increases the "effective number of observations" → increases estimation accuracy
- these W / M statistics can be estimated & monitored daily, but
- do not depend on the day of the W or M (i.e. are stable)
- example daily overlapping weekly (DOW) data :



Using daily overlapping returns

Example daily overlapping weekly returns for volatility

- use daily total returns to make a cumulative total return index series *RI*
- each day *t*, calculate the weekly return : $r_t^w = \dfrac{RI_t}{RI_{t-5}}$
- calculate the volatility using these daily overlapping weekly returns

- *note 1 :* ***auto-correlation ?***
  daily overlapping returns are auto-correlated, but we do **not** have to adjust for this.
  After all, we're using weekly returns for weekly volatility.
  Auto-correlation is only a problem if we aggregate these weekly returns to a longer horizon

- *note 2 :* ***bias !***
  overlapping creates a negative bias : estimated volatility is **smaller** than true volatility
  → inflate estimated vola with **bias factor** → later

What EWMA persistence to use ?

Daily overlapping returns imply using daily persistence

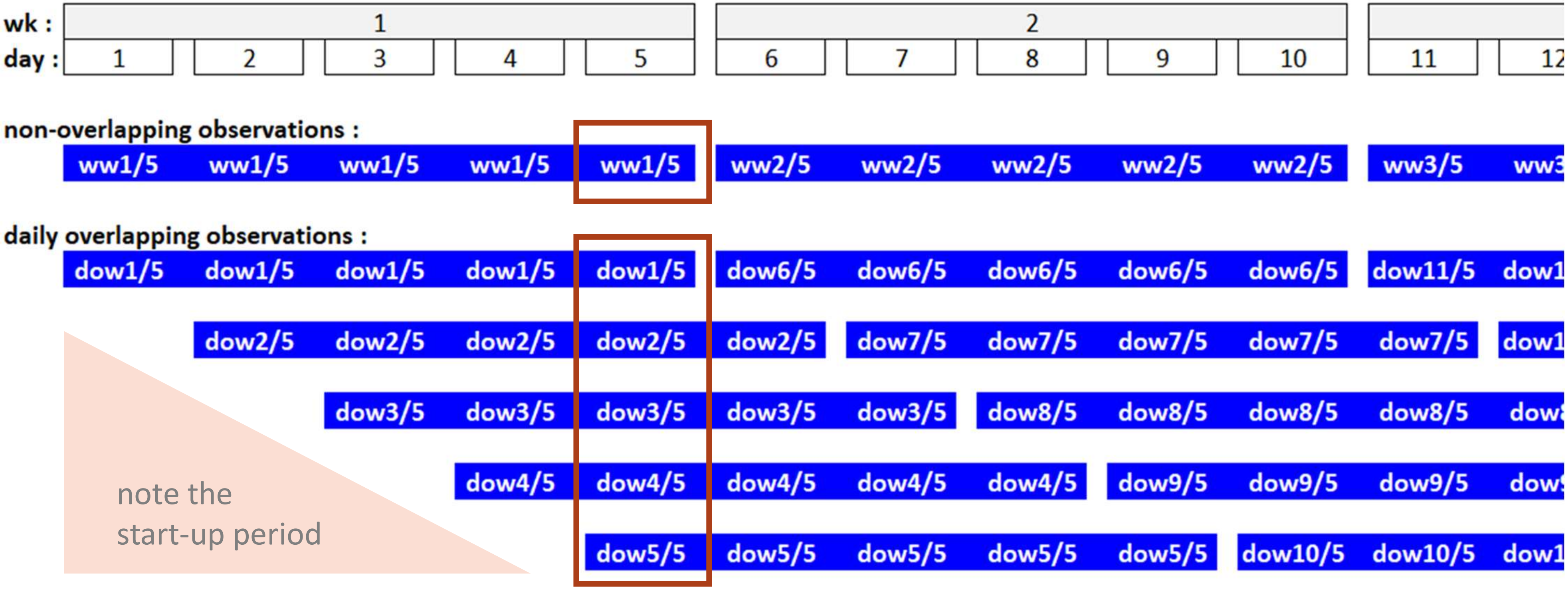
- Changing the return observation interval & using overlapping affects the persistence parameter
- The use of daily overlapping data implies that the persistence parameter should be **re-calculated to a daily basis** → **why ?**

- **Intuitively, the weight for a given day should be the same whether overlapping or non-overlapping data are used**
- So at any day, the sum of the implied weights of the 5 overlapping weekly returns should be the same as the implied weight for the non-overlapping weekly return

→ see next slide

- This implies that **to obtain daily-overlapping weekly weights** we have to :
  scale **down** the weekly non-overlap **weight** with a factor ~5 , and
  scale **up** the weekly non-overlap **persistence** parameter → **how ?**

Look at weight per day



- non-overlapping weekly weight **ww** → = weight of ww/5 per day
- daily overlapping weekly weight **dow** → = weight of 5 times a dow/5 per day
- so daily overlapping weight must be five times as small as non-overlapping weekly weight

Always use daily persistence !

… if daily overlapping !

- daily overlapping weight is non-overlapping weekly weight / 5 = weight for daily observations !

0. intuitively : we should use daily persistence for daily overlapping data !

1. via half-time *H* : exact in λ

- suppose weekly H_w = 9.01 weeks → $\lambda_w^{H_w} = ½$
- convert to daily $H_d = 9.01 \cdot 5 = 45.05$ days → $\lambda_d^{H_d} = ½ = \lambda_d^{5 \cdot H_w}$

$$\left.\begin{array}{l} \lambda_w^{H_w} = ½ \\ \lambda_d^{H_d} = ½ = \lambda_d^{5 \cdot H_w} \end{array}\right\} \Rightarrow \lambda_d = \lambda_w^{1/5}$$

in this example : $\lambda_w = 0.9259$ hence : $\lambda_d = \lambda_w^{1/5} = 0.9847$

QED

2. via total weight in given lookback window of T_w weeks or T_d days ($T_d = 5 \cdot T_w$) :

$$1 - \lambda_w^{T_w} = 1 - \lambda_d^{T_d} \quad \Rightarrow \quad \lambda_d = \lambda_w^{\frac{T_w}{T_d}} = \lambda_w^{1/5}$$

exact in λ

Always use daily persistence !

… if daily overlapping !

- daily overlapping weight is non-overlapping weekly weight / 5 = weight for daily observations !

0. intuitively : we should use daily persistence for daily overlapping data !

1. via half-time *H* : $\lambda_d = \lambda_w^{1/5}$

2. via total weight in given lookback window of T_w weeks or T_d days : $\lambda_d = \lambda_w^{1/5}$

3. via number of periods *N* (= via *WAL*) : exact in *N*

- suppose weekly N_w = 26 weeks → $\lambda_w = 0.9259$
- convert to daily N_d = 26 * 5 = 130 days → $\lambda_d = 0.9847$ check : 0.9259^(1/5) = 0.9847

$$\lambda_d \approx \lambda_w^{1/5}$$

Using daily overlapping returns : bias

Bias adjustment factors for daily overlapping weekly & monthly volas

- we look at **weekly** (5 days) or **monthly** (21 days) returns
- we use **daily overlapping** : the overlap is 4 resp. 20 days
- academic literature provides bias & adjustment factors for **EW** volatility
  → translate to **EWMA**

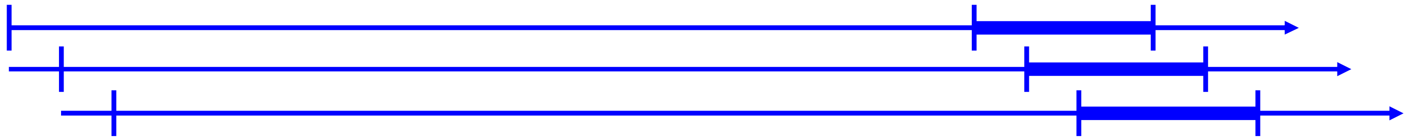
- start with weekly or monthly non-overlapping data
- express λ_w or λ_m in terms of N_w or N_m via $N = \dfrac{1+\lambda}{1-\lambda}$
- determine the implied number of overlapping returns :
  $N_{ow} = N_w \cdot 5 - 4$, $N_{om} = N_m \cdot 21 - 20$
- the table shows the **vola adjustment factors** given N_o
- use these factors to inflate estimated EWMA vola
  (interpolate for other values of N)

	adjustment factor vola	
No	**weekly**	**monthly**
399	1.005	1.026
260	1.008	1.040
199	1.010	1.053
99	1.021	1.110
66	1.031	1.173
49	1.043	1.244
39	1.054	1.321
32	1.067	1.410

Evaluating overlapping cond'l volas

What about the 4 evaluation criteria ?

- cond'l volas from overlapping returns → forecasted cond'l volas & *z*-scores overlap



- hence, the forecasted volas and *z*-scores are serially correlated

implications for vola evaluation criteria :

1. **long-term precision :** full-sample bias
   in calculating the s.e.(stdev(z)), set M to # **non**-overlapping observations
2. **short-term precision :** time series bias : still compare models by MRAD
3. **improvement in distribution :**
   kurtosis of *z*-scores can be compared; only if testing, use # **non**-overlapping observations
4. **max QL :** still compare sums of QL ; only when testing for differences, use # **non**-overlapping observations for T in $stdev(\bar{d})$ when calculating *t*-statistic

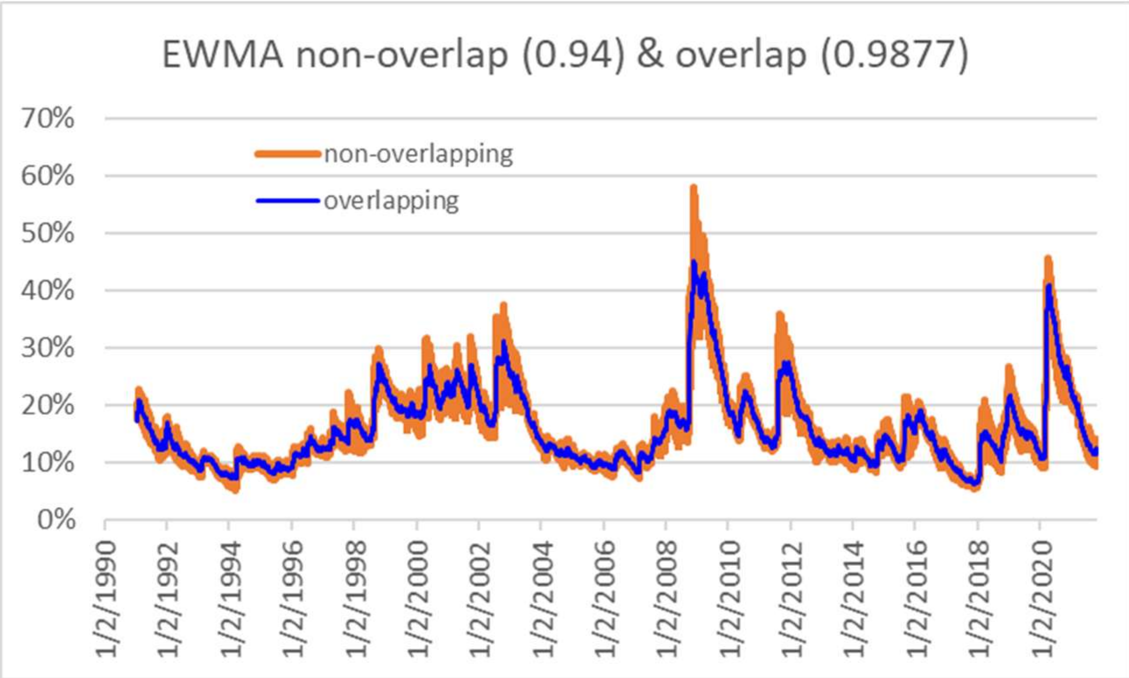
Overlapping vs non-overlapping

US Equity example

- 52-week lookback period, 0.94 EWMA for non-overlapping weekly returns
- 260-day lookback period, 0.94^(1/5) = 0.9877 EWMA for daily overlapping weekly returns

- **non-overlapping :**
  clear day-of-the-week effect

- **overlapping :**
  nice in the ~middle of the
  non-overlapping estimates



Some final words on EWMAs

EWMAs in an investment portfolio context

- EWMA is a **practical**, **parsimonious** and **robust** alternative to GARCH(1,1)

- EWMA is **adaptive** to increases in volatility and allows for **vola clustering**

- a fast EWMA : adapts quickly to changes in vola

  but : . has less precision than a slower EWMA, gives unstable forecasts
  . tends to overshoot after extreme returns
  . displays "**lullaby effect**" during tranquil periods
  because of absence of mean-reversion

- **overlapping** may help to : . align observation interval and forecast horizon,
  . increase estimation accuracy & stability

  note : the overlapping frequency determines how to re-scale the persistence
  (daily overlapping implies daily persistence, e.g.)

Enhanced EWMAs

- adding a **mean-reversion** level :
  allocate a small fraction Υ of the weight to a long-term variance

$$\sigma_{t+1}^2 = \gamma \cdot \sigma_{LT}^2 + (1-\gamma) \cdot \left[(1-\lambda) \cdot r_t^2 + \lambda \cdot \sigma_t^2 \right]$$

→ but how to choose Υ and LT-variance ?

- coping with the **lullaby effect** (also from absence of mean-reversion) :
  - specify a fast and a slow EWMA, for example λ = 0.94 (*N* = 32) and λ = 0.99 (*N* = 199)
  - construct a **"dual metric"** vola estimator :

$$EWMA^* = \max[EWMA_{fast}, EWMA_{slow}]$$