# Estimating beta

**Donald Robertson**[1]

**Faculty of Economics**

**University of Cambridge**

May 2026

[1] This note is written in a personal capacity. The opinions and conclusions expressed are solely those of the author and should not be attributed to the University of Cambridge.

# Summary

## Objective

The key question is what beta it is intended to measure. For regulatory decisions, the relevant object is typically a forward-looking, medium to long run measure, rather than a purely statistical fit to historical data.

Estimating beta is inherently a modelling exercise where the methodology used cannot be separated from the assumptions one is prepared to make about beta and the forecast time horizon. Simple approaches such as OLS do not avoid this issue  they include their own implicit assumptions.

A robust approach should:

- explicitly consider the time variation of beta.
- apply appropriate techniques for outliers and/or structural change.
- assess sensitivity across methods.

Explicit judgement is required on the appropriate technique to adopt to develop a suitable and internally consistent methodology. Differences in estimation approach (window length, treatment of outliers, modelling of time variation) can lead to materially different beta estimates and therefore materially different allowed returns.

## Overview

The CAAs H8 methodology uses unweighted Ordinary Least Squares (OLS) beta. The standard CAPM beta estimator based on OLS is simple and widely used, but its interpretation relies on strong assumptions:

- Beta is constant over both estimation period and forecast horizon.
- Errors are well behaved (no major structural breaks or influential observations).

When these assumptions are violated  even if unbiased  the OLS estimator may be:

- inefficient (less precision) compared to alternative estimators.
- sensitive to outlier observations in small samples.
- not the best forward-looking measure if beta is actually time-varying (which use of a cut off date would suggest).

This report does not assess whether comparator airport betas in fact violate those assumptions. Rather, it explains how such issues should be treated if they are present. In particular:

- Robust estimators (e.g. Huber, LAD, Theil-Sen) are more appropriate for outliers.
- Weighted least squares (WLS) addresses time-varying volatility, but not outliers.
- Structural change should be modelled explicitly (e.g. via dummies or time-varying models).

It is not apparent that the CAA has explicitly assessed the impact of outliers, time-varying volatility or structural change in their H8 Initial Proposals.

In addition to selecting the chosen model, estimating beta requires a choice of estimation window. Choosing an estimation window involves a fundamental trade-off:

- **Short windows**: capture local structure but can be noisy and sensitive to outliers, implicitly assumes short-run estimates are informative about longer run behaviour.
- **Long windows**: reduce sensitivity to noise and align better with long horizon forecasts, but may mix different structural regimes and hence strain the assumption of a constant beta.

There is no single “correct” window length. The choice must depend on assumptions one is prepared to make about the nature of any time variation of beta. A method that uses several windows without explaining how they should be combined does not solve this problem; it shifts the modelling judgement to a later, less transparent stage.

The CAA uses 2yr, 5yr and 10yr estimation windows. This approach raises several important concerns.

- It assumes beta is broadly stable within each estimation window and over the forecast horizon, with the latest spot estimate representative of the future.
- At the same time, the use of multiple windows appears to recognise instability across time.
- It does not formally model the source, persistence or likely future path of that instability.
- It does not provide justification or explanation (statistical or otherwise) for how estimates from overlapping 2, 5 and 10 year samples should be combined, even though they are not independent and can imply materially different allowed returns.
- It leaves unclear whether unusual periods, such as COVID, are being treated as time volatility (heteroscedasticity), outliers, or structural changes in beta.

These assumptions are difficult to reconcile. A two year window may be more representative of recent dynamics, but it is also more exposed to noise and outliers. A ten year window may provide greater statistical stability, but only by imposing a stronger assumption that beta has been sufficiently stable over a much longer period. The CAA approach therefore risks presenting a range of estimates without an explicit framework for deciding what each estimate means.

## Implications for the CAA

The CAAs purely mechanical application of OLS risks:

- relying on estimates that may not be aligned with the regulatory horizon.
- implicitly imposing strong and inconsistent assumptions about beta dynamics.
- implicitly rejecting a materially better alternative method for beta estimation.

In addition we have the following issues:

- what method will be used to combine estimates across different windows?
- shorter windows may be more representative of recent dynamics but can be more sensitive to noise and outliers so may need robust estimators.
- the approach does not account for any potential dynamic structure in the evolution of beta over time (e.g. mean reversion within the forecast period).

An explicit framework for assessing beta is required to give confidence that the chosen methodology is the optimal framework to address the issues highlighted around time variation, outliers, highly influential data points, and structural breaks. Table 1 contains a discussion of when methods should be considered suitable for beta estimation.

| Method | Description / Suitability | Main risk / drawback |
|---|---|---|
| OLS (least squares) | Standard estimator assuming constant beta. Use long samples with no major outliers or structural change. | Sensitive to outliers and high-leverage points. May give misleading results if beta is time-varying. |
| Weighted Least Squares (WLS) | Downweights observations with high conditional variance. Needs model of heteroscedasticity. | Requires correct specification of weights. Misspecification or data derived weights can cause problems |
| Huber M-estimator | Robust regression that downweights large residuals smoothly. Performs well for moderate number of outliers. | Requires tuning parameter. May still be influenced by high leverage points. |
| LAD/Median regression | Minimizes absolute residuals so estimates conditional median. Robust to moderate outliers. | Less efficient under normal errors. Not estimating mean relation. |
| Theil-Sen | Median of pairwise slopes. Non-parametric and robust in small samples or heavy-tailed data. | Computationally intensive in large samples. Less efficient under well-behaved (normal) errors. |
| Least Median Squares (LMS) | Minimizes median of squared residuals. Very high breakdown point (50%). | But low efficiency under normality. Estimates can be unstable in small samples. |
| Least Trimmed Squares (LTS) | Minimises sum of squares after trimming largest residuals. Robust to outliers. | Requires arbitrary trimming choice. Discards potentially informative data. |
| Winsorising | Caps extreme values at chosen thresholds to limit outlier influence. | Ad hoc choice of thresholds. May bias estimates. |
| Shrinkage | Shrinks beta toward prior (often 1) reducing MSE. | Introduces bias if wrong prior. Assumption of representative asset may not hold. |
| Rolling window OLS | Estimates beta over moving window to capture time variation. | Results can be sensitive to window length. Can miss rapid changes. |
| Kalman filter | Models beta as evolving over time in state space framework | Requires model specification. Results can be sensitive to assumptions and parameter choices. |
| GARCH-based beta | Models time-varying covariance and variance to derive beta. | Model complexity and risk of misspecification. Can be unstable in small samples. |
| Dummy for structural change | Allows discrete shifts in beta for specific periods (e.g. COVID). | Requires correct identification of break periods. Ignores gradual changes in beta. |

Table 1: CAPM Beta Assumptions and Estimation Methods

# 1 Regulation and Beta

## Background

Monopolies extract consumer surplus by raising price above marginal cost, moving up the demand curve and generating extra profits. This is inefficient from a societal point of view where it is generally viewed optimal to have price = marginal cost. To address this, whilst retaining the incentive for innovation and cost control, regulated monopolies should be limited in their profits. This is typically enforced by a return on capital type rule with an assumed model to derive the cost of equity. From the CAA's H8 Initial Proposals

> For the cost of equity, we propose to use beta estimates that we have derived from a comparator group of listed European airport operators using multiple estimation windows to ensure both relevance and statistical robustness.

## The CAPM Approach

The standard model for returns based regulation is to use the CAPM. Under reasonable conditions the CAPM gives in equilibrium the relationship between excess returns:

$$E\left(R_i - R_f\right) = \beta_i E\left(R_M - R_f\right)$$

where $R_i$ is the return on asset $i$, $R_M$ the market return and $R_f$ the risk free rate. $\beta_i$ is $\frac{Cov(R_i,R_M)}{Var(R_M)}$.

Estimating beta correctly is then essential to compute a fair cost of equity for regulated monopolies. Suppose we have a time series of excess returns observations $r_{it}, r_{Mt} \quad t = 1, \ldots, T$ on asset $i$ and market returns. If $\beta_i$ can be assumed constant then a straightforward way to estimate would be to plug in estimates of the numerator and denominator.

$$\widehat{\beta}_i = \frac{\sum_{t=1}^{T}\left(r_{it} - \bar{r}_i\right)\left(r_{Mt} - \bar{r}_M\right)}{\sum_{t=1}^{T}\left(r_{Mt} - \bar{r}_M\right)^2}$$

and this will be a consistent estimator of the true $\beta_i$.[2]

The usual way to obtain this estimate is via least squares regression

$$r_{it} = \alpha + \beta_i r_{Mt} + \epsilon_t$$

where the least squares formula returns an identical expression for $\widehat{\beta}_i$.[3]

[2] As long as the numerator and denominator satisfy some regularity conditions (such as weak law of large numbers).

[3] Note this NOT because $r_{it}, r_{Mt}$ are assumed to obey a Classical Linear Regression Model but is simply a convenient way to calculate the plug in estimator given the ubiquity of statistical software. Consequently the output and diagnostics from such a regression should not necessarily be viewed through the lens of the CLRM.

# 2 Estimation Issues

With the assumption of constant $\beta$ there seems little reason not to use as much data as possible, though higher frequency than daily may cause market microstructure issues.

A number of problems can arise with high frequency financial data. They often show changes in volatility (known as heteroscedasticity), there may be anomalous observations or outliers that can distort estimation especially in finite samples, and there can changes in coefficients or structural shifts. All may affect the properties of least squares estimation and need careful treatment. Of course all three problems may coexist, and distinguishing what may be the most severe issue may be challenging simply from looking at the data.

## Heteroscedasticity and weighted least squares

In a regression framework with heteroscedasticity

$$y_t = \alpha + \beta x_t + u_t$$
$$\mathbb{E}\left(u_t|x_t\right) = 0 \quad \mathbb{E}\left(u_t^2|x_t\right) = \sigma_t^2$$

we can transform the model into one with homoscedastic errors by a weighted least squares (WLS) procedure

$$\frac{y_t}{\sigma_t} = \alpha\left(\frac{1}{\sigma_t}\right) + \beta\frac{x_t}{\sigma_t} + \frac{u_t}{\sigma_t}$$

which amounts to downweighting observations with high variance ($\sigma_t$).

Note that such heteroscedasticity does not itself cause biases in least squares estimation as long as $\mathbb{E}\left(u_t|x_t\right) = 0$ is maintained (in which case OLS of CAPM also gives an unbiased estimate of the linear projection of asset excess returns onto market excess returns). Rather this is an efficiency issue with OLS. So if an event such as COVID is assumed temporarily to change $\sigma_t^2$ then the use of WLS could perhaps be justified if efficiency (i.e. maximum precision for unbiased estimators) is required. But a number of issues arise:

1. If the COVID event introduced heteroscedasticity why not treat other episodes showing differing volatility the same?

2. How to choose weights (ie what is $\sigma_t^2$). What is needed is weights $w_t \propto \frac{1}{Var(u_t|x_t)}$. Two approaches:

   i. Specify the weights *a priori*. But if they are just chosen at random then a poor choice of weights may not improve the estimator. If done by running a regression, looking at the residuals and picking "weights" to downweight observations you don't like (e.g. large residuals) and re-estimating then this is essentially data snooping.

ii. Use data to estimate weights i.e. estimate relation of $\hat{u}_t^2$ with some assumed functional form of $x_t$ and plug these in. This can make the problem worse if the weights are chosen with an incorrect model of the relation - essentially you run the risk of over weighting already high variance observations (the estimator would still be consistent but OLS is consistent here anyway). In small samples estimator may have poor properties.

3. In this setting feasible WLS seeks the efficient estimator but at possible cost of introducing extra biases in finite samples. Efficiency may not be a sensible requirement and in attempting to increase precision of the estimates we risk introducing bias. If sample sizes are relatively small robustness may be preferred to efficiency.

4. If second moments vary in a way that affects covariance with the market, it is difficult to sustain the assumption that $\beta$ is constant (unless there is exact proportionality in the covariance and variance movements) so there is an internal inconsistency here. If $\beta$ is time varying then in assuming a constant $\beta$ the CAPM regression is misspecifed with a more complex error structure

$$r_{it} = \alpha + \bar{\beta}_i r_{Mt} + \left(\epsilon_{it} + (\beta_{it} - \bar{\beta}_i) r_{Mt}\right)$$

and using WLS in a misspecified model could be a bad idea.

**When would WLS downweighting COVID be optimal?**

If beta is constant (so either covariances and variances are constant or they move proportionally) then suppose

$$y_t = \alpha + \beta x_t + \epsilon_t$$

$$Var\left(\epsilon_t | x_t\right) = \begin{cases} \sigma_1^2 & \text{non-covid} \\ \sigma_2^2 & \text{covid} \end{cases}$$

then weighted least squares (with weights $\frac{1}{\sigma_{1,2}^2}$) would be optimal (given other Gauss Markov assumptions hold).

## Outliers and Robust estimation

An alternative interpretation is that these observations are simply data outliers that might distort inference especially in small samples. The sensitivity of an estimator to anomalous or outlier observations is often discussed in terms of "breakdown value". Roughly speaking, it indicates the smallest fraction of contaminants in a sample that cause the estimator to break down, that is, to take on values that are arbitrarily bad. For example, suppose we wish to estimate the mean from a sample of $n$ observations. We can use $\overline{x}_n = \frac{x_1 + \ldots + x_n}{n}$ which will have a breakdown value of zero because contaminating even a single observation can make $\overline{x}_n$ arbitrarily large. The median has a much higher breakdown value (of 0.5 which is in fact the maximal value for an estimator of location since if half the observations are contaminated the notion of central location itself becomes lost).

It is well known that estimators such as least squares have a very low breakdown value and are not robust to even a single anomalous observation. In this situation weighted least squares is usually not appropriate. It addresses heteroscedasticity, not outliers. To deal with outliers one can minimise a loss function less sensitive to large values, for example estimation via least absolute deviation penalises absolute rather than quadratic loss. More generally one can choose weights as a function of residual size, $w_t = f(|e_t|)$, in a systematic way giving rise to robust M-estimators such as Huber.

Alternatively, if the concern about outliers is severe, very high breakdown robustness can be achieved using least median squares (LMS), where one minimises the median of squared residuals. This estimator has a breakdown point of $\frac{1}{2}$, though at the cost of lower efficiency under Gaussian errors (but we may care more about bias than efficiency here).

M-estimators can be implemented via an iteratively reweighted least squares routine, where each step solves a weighted least squares problem with residual dependent weights.

For Huber estimation with loss function $\rho(e)$ (usually a symmetric function with $\rho(0) = 0$) the estimator solves

$$\min \sum_t \rho(y_t - \alpha - \beta x_t)$$

where $e_t = y_t - \alpha - \beta x_t$. Let $\rho' = \psi$ then the first order conditions for minimising are

$$\sum_t \psi(e_t) = 0 \quad \sum_t \psi(e_t) x_t = 0$$

or writing with $w_t(e_t) = \frac{\psi(e_t)}{e_t}$ for $e_t \neq 0$ the second is

$$\sum_t w_t(e_t) x_t e_t = 0$$

so has the form of a weighted least squares normal equation where the weights depend on the residuals (and therefore on the estimated parameters), making the estimating equations nonlinear.

Alternatives such as winsorizing the data or least trimmed squares can also be considered though these require choices of thresholds or trimming fractions that can be seen as *ad hoc*.

## Structural Change

Lastly we could see COVID primarily as a one off event giving a sequence of data with a discrete change in beta solely during that period. In this case a better approach would be to use dummies to account for the change (assuming there is sufficient data)

$$r_{it} = \alpha + \beta_{i,pre} r_{Mt} + \beta_{i,COVID} D_{COVID} r_{Mt} + \epsilon_t$$

or, one could allow for differences pre and post,

$$r_{it} = \alpha + \beta_{i,pre} r_{Mt} + \beta_{i,COVID} D_{COVID} r_{Mt} + \beta_{i,post} D_{post} r_{Mt} + \epsilon_t$$

which gives a clean estimate of $\beta_{pre}$ against $\beta_{COVID}$ (and indeed could be used to test for structural change by e.g. $H : \beta_{i,COVID} = 0$) though if beta changes say continuously then Chow type tests would be misspecified. It is also worth noting the challenge of robustly identifying the dates of any structural change(s) which can introduce the risk of p-hacking.

However this again highlights the broader issue, allowing for this sort of structural change is inconsistent with a constant beta framework.

## Estimation for constant vs non-constant beta

With structural change, or with heteroscedasticity that affects the covariance with the market, it may be better to try and model the time variation directly.

- If beta is assumed constant, errors are homoscedastic, no outliers or structural change then use OLS on longest run of data available.
- If beta is assumed constant but there are a few outlier observations then use a robust estimator eg LAD on as much data as possible to minimise impact of outliers.
- With beta constant in general heteroscedasticity is less of a problem than outliers unless one is interested in efficient estimation or inference. Under standard exogeneity conditions, OLS remains unbiased for the population regression (linear projection) coefficient even in the presence of heteroscedasticity.
- If the period chosen for estimation is quite short (because for example beta is felt to be slowly time varying and can be assumed roughly constant over the chosen interval) then the impact of outliers will be exacerbated and a robust estimator should be considered. These are designed to reduce the impact of outlying observations in a finite sample (and include Huber M, median (LAD) regression, Theil-Sen (non-parametric)).
- To deal with outliers winsorising, trimmed OLS or removing observations based on some measure e.g. Cook's distance can also be effective but may be sensitive to the threshold chosen (so sensitivity to any such threshold should be checked).
- Shrinkage type estimators can also improve robustness but standard approach of shrinking towards 1 assumes asset is randomly drawn from market universe which may not be appropriate.

- Otherwise, if beta is time varying it is most appropriate to model the time variation directly
  - If beta shows discrete change only for a specific period (e.g. COVID) then use dummy for this period.
  - Use a rolling window OLS. This imposes a piecewise constant approximation so assumes beta evolves slowly. This may be a poor assumption if beta evolves continuously and reacts quickly in periods of volatility. This is often the approach chosen.
  - With rolling windows, WLS or some form of robust estimation can still be used conditionally within each given period if heteroscedasticity or outliers are present (but these do not of course address the structural problem of time variation).
  - Use Kalman filter and state-space model for dynamic beta estimates
  - Use realised beta i.e. estimate beta using window of high frequency data and then analyse the resulting lower frequency estimates.
  - Model conditional covariance with multivariate GARCH and hence calculate beta.

Time varying methods all require we specify some sort of model for the time variation in beta. Whilst this may seem like a disadvantage compared to the simple OLS or WLS methods, remember that these simply solve this problem by effectively (and often incorrectly) assuming it away.

Unfortunately it is not possible to test for constant beta without a model of what that time variation would be (so in any such test there is a joint hypothesis problem, of constant beta and correct model specification). And even tests that estimate under the null (so LM type tests) need a specification (of model instability) under the alternative to implement the test.[4]

**Time varying coefficients in least squares estimation**

If beta is truly time varying so the model is

$$r_{it} = \beta_{it} r_{Mt} + \epsilon_{it}$$

but this is ignored in estimation, then (see Appendix for details):

- OLS estimates a weighted average of time-varying betas with weights proportional to $r_{Mt}^2$, rather than a simple time average.
- Periods with large $|r_{Mt}|$ have more weight so contribute disproportionally

[4] General stability tests such as the Cusum and Cusumsq (Brown Durbin and Evans) type do not have any specific alternative but do require a number of assumptions e.g. independent errors, constant variance, correct mean specification and violations of these assumptions can create false positives/negatives.

- The OLS estimator is consistent for the (population) mean of beta if the time variation in $\beta_{it}$ is uncorrelated with squared market returns (volatility). Stronger conditional moment restrictions are required for finite sample unbiasedness. If $\beta_i$ rises (or falls) during volatile markets this would be violated.

**Leverage, outliers and Theil-Sen estimator**

We can distinguish two sorts of anomalous points depending on whether we look in the $x$ direction or the $y$. High leverage happens if $x_i$ is far from $\bar{x}$ (unusual in the $x$-direction). An outlier typically has a large residual $y_i - \hat{y}_i$ (unusual in the $y$-direction).

In a simple linear regression with an intercept

$$y_i = \alpha + \beta x_i + \epsilon_i$$

the leverage of any regression point is defined as

$$p_i = \frac{(x_i - \overline{x})^2}{\sum_{j=1}^{N}(x_j - \overline{x})^2}$$

Note that the sum of leverages is 1.

The OLS regression coefficient is then

$$\widehat{\beta}_{OLS} = \sum_{i=1}^{N} p_i \frac{(y_i - \overline{y})}{(x_i - \overline{x})}$$

Now noting that the line that passes through $(x_i, y_i)$and $(\overline{x}, \overline{y})$ has slope $\frac{(y_i - \overline{y})}{(x_i - \overline{x})}$ we have that the slope of the regression line is a weighted sum of the slopes of the lines between each point and the mean point $(\overline{x}, \overline{y})$ where the weights are the leverages. So for points that have large leverage the slope of the regression line is weighted more towards the slope from that point to the mean ie points that have large leverage are the most important in determining the slope of the regression line. Points with small leverage do not count so much, we could remove them without changing the regression line very much. Rule of thumb is leverage over $\frac{2k}{n}$ (or sometimes simply 0.2) should be investigated.

**Theil-Sen estimator**

Least squares is:

$$\widehat{\beta}_{OLS} = \sum_{i=1}^{N} p_i \frac{(y_i - \overline{y})}{(x_i - \overline{x})}$$

The Theil Sen estimator computes:

$$\widehat{\beta}_{TS} = \text{median}_{i<j} \left( \frac{y_j - y_i}{x_j - x_i} \right)$$

So Theil-Sen calculates the median of all pairwise slopes between observations. Both estimators aggregate slope information from the data. OLS uses a leverage-weighted mean so is dominated by extreme $x$-values. Theil-Sen uses a median over pairwise slopes, and will be robust to such extremes.

### Influence and Cook's distance

It is important to note that outliers (high residuals) needn't be high leverage, and high leverage needn't be outliers. Troublesome points (from an estimation point of view) will be those with high leverage *and* a large residual (because the model does not fit well and the point contributes substantially to the estimated slope). Such points are called influential, usually captured by Cook's distance.

Cook's distance is defined as

$$D_i = \frac{e_i^2}{k\hat{\sigma}^2}\left(\frac{p_i}{(1-p_i)^2}\right)$$

where $k$ is number of regressors incl constant and $p_i$ is leverage. Loosely it measures how much the fitted regression would change if observation $i$ were removed. Common rule of thumb suggest investigating where Cook's distance >1 (or, in larger samples $\frac{4}{n}$).

When there are multiple potential outliers there can be a masking effect hindering detection (when these outliers may 'hide' each other, so that none of them appears unusually influential or extreme when assessed individually). So a cautious approach is needed - use multiple diagnostics (residuals, leverage, Cooks distance) and check for clusters of unusual observations, not just single points.

# 3 Relevance to H8

First question is "What is the object we are trying to estimate?"

- Constant (long run) $\beta$?
- Time average of $\beta_t$?
- Forward looking (conditional on data) $\beta$?

A sensible approach would be to investigate several or all of these estimation methods and obtain a range of possible values for beta to assess robustness and allow for model uncertainty, rather than relying on a single potentially misspecified estimate. Table 1 sets out a range of estimation techniques, the situation where they might be suitable and potential drawbacks.

So if view is constant beta but there is heteroscedasticity

- OLS still unbiased (as long as $E(u_t|x_t) = 0$ is maintained)
- WLS corrects for heteroscedasticity if the conditional variance model is correct.
- Potentially try CAPM with ARCH/GARCH errors but empirical experience is not good.

These can correct for time varying heteroscedasticity but are not designed to deal with outliers which require a different treatment.

If COVID period just a sequence of outliers (but concern is that they are influential enough to distort estimate of beta) then need a systematic and model consistent way to detect and robustify the estimator. Ad hoc deletion or selective downweighting can introduce bias and poor small sample problems which undermine the credibility of the estimation procedure. This is especially important if we have restricted the data to a shorter estimation interval because of concerns about longer horizon instabilities in beta.

If COVID changed the true structural beta then constant beta models may not be sensible. In which case we need to model the change and decide if it is permanent or temporary.

## The H8 Approach

Paragraphs 9.74-9.75 of the CAA/CAP3232 document explains

> We propose to rely on unweighted estimates of listed airport company equity betas estimated over different estimation windows . . . We, therefore, estimate equity betas for listed airport companies based on two- year, five-year and ten-year estimation windows based on daily share price data

This is the "Rolling Window OLS" approach in Table 1, with a focus on the final spot value. So assumes beta is constant or very slowly moving within the specified estimation window though they also seem to entertain the possibility of discrete structural shifts. The tension then is the simultaneous assumption of stability within both the estimation window and the forecast horizon, together with across window instability, with no formal modelling of this variation. A two year window makes the assumption of a constant beta more palatable, but exposes estimation to more noise and assumes the (final) spot estimate holds for the forecast horizon. A ten year window may be more appropriate for the forecast horizon (and weakens the impact of outliers etc), but strains the constant beta assumption. There is no free lunch.

Given the evident instability of beta in the rolling estimates presented in Figure 9.6, shorter windows may better reflect local dynamics than longer windows. However, they are also more exposed to noise and outliers, so their reliability depends critically on the treatment of extreme observations. So a robust estimator within this window may also be informative.

However this rolling OLS approach ignores any dynamic structure in a time varying beta (for instance beta may be mean reverting around some long run average value) and this could also be checked rather than simply using the final spot value as estimate and would give a natural way to incorporate longer run information.

The document does not really discuss *how* the two, five, and ten year estimates may be combined (whether averaging or some judgemental adjustment). This matters because different window choices can lead to different beta estimates, and therefore materially different allowed returns. The appropriate way to combine these estimates depends on the underlying process governing beta, and therefore requires taking a position on its time variation.

Even under the constant beta assumption, the two five and ten year OLS estimates are not independent given the overlap in their underlying data samples. Although an optimal linear combination exists in principle, the weights would depend on the induced covariance structure across these overlapping windows. The optimal choice of weights is therefore itself model dependent, and cannot be determined without additional assumptions about the data-generating process.

More generally, any method of combining the estimates imposes some - typically unarticulated - assumption about beta dynamics. This does not eliminate modelling assumptions; it simply leaves them implicit rather than explicit. One cannot avoid specifying a model.

## Appendix

### Time varying coefficients and least squares estimation

Suppose (assume no intercept for simplicity)

$$r_{it} = \beta_{it} r_{Mt} + \epsilon_{it}$$

OLS treating beta as constant calculates

$$\begin{aligned}
\hat{\beta}_i &= \frac{\sum r_{Mt} r_{it}}{\sum r_{Mt}^2} \\
&= \frac{\sum r_{Mt}(\beta_{it} r_{Mt} + \epsilon_{it})}{\sum r_{Mt}^2} \\
&= \frac{\sum r_{Mt}(\bar{\beta}_i r_{Mt} + (\beta_{it} - \bar{\beta}_i) r_{Mt} + \epsilon_{it})}{\sum r_{Mt}^2} \\
&= \bar{\beta}_i + \frac{\sum (\beta_{it} - \bar{\beta}_i) r_{Mt}^2 + r_{Mt}\epsilon_{it}}{\sum r_{Mt}^2}
\end{aligned}$$

where $\bar{\beta}_i = \frac{1}{T}\sum_{t=1}^{T} \beta_{it}$.

So $\hat{\beta}_i$ may not reveal the long run average beta unless the second term has expectation/plim zero.

Write $w_t = \frac{r_{Mt}^2}{\sum r_{Mt}^2}$ then

$$\hat{\beta}_i = \bar{\beta}_i + \sum_t w_t(\beta_{it} - \bar{\beta}_i) + \sum_t \frac{r_{Mt}\epsilon_{it}}{\sum r_{Mt}^2}$$

So:

- OLS estimates a weighted average of time-varying betas with weights proportional to $r_{Mt}^2$ not the simple mean.
- Periods with large $|r_{Mt}|$ have more weight so contribute disproportionally
- The second term has expectation zero only if the time variation in beta satisfies conditional mean independence (with respect to market volatility)

$$\mathbb{E}\left[(\beta_{it} - \bar{\beta}_i) | r_{Mt}^2\right] = 0$$

If $\beta_{it}$ is systematically related to market volatility this would be violated.

**Biography** Professor Donald Robertson is Professor of Econometrics at the University of Cambridge and a Fellow of Pembroke College. His academic work sits at the intersection of applied econometrics, finance, macroeconomics and statistics, giving him a strong technical base for advising on regulatory finance questions where empirical estimation, uncertainty and judgement are central. He has published widely in econometrics and applied economics. Professor Robertson drafted a 2018 paper for Ofgem on estimating beta, which examined the limitations of standard OLS approaches, the use of GARCH models, and the implications of unstable or mean-reverting beta estimates for regulatory determinations. Professor Robertson also co-authored an investigation of Beta Estimation for CAPM-WACC at Long Horizons on behalf of the UKRN for their influential 2018 cost of capital study. He has continued to advise across UK economic regulation and recently has worked for Ofwat on their PR24 process, co-authoring multiple papers (with Robin Mason and Stephen Wright). These papers supported Ofwats original PR24 determinations and responding to disputing companies arguments on allowed return and beta estimation as part of the CMA PR24 redetermination.