Estimating β

Donald Robertson*

April 19, 2018

1 Introduction

This document describes some issues with the current approach to estimating the CAPM beta as a method to calculate appropriate discount rates to use to value risky investments and suggests a way forward through the use of multivariate modelling. We take the use of the single factor (beta) CAPM as the correct model and explore statistical issues around the estimation of beta. The existing approach typically uses rolling least squares estimation over some window to produce a current estimate of beta which is then treated as the estimate of beta going forward. Whilst this procedure is computationally straightforward it raises a number of statistical and conceptual issues. In particular the importance of developing models that account for time variation in both variances and covariances of asset returns, and the requirement for an estimate of beta that is suitable for an investment horizon that may stretch to several years. We argue that if data is to be used to inform decision making then this has to be done in a way which respects the statistical framework.

2 Background

The standard economic approach to modelling the required expected return on an asset is via the Capital Asset Pricing Model (CAPM) which captures the relationship between risk and expected return in the pricing of risky investments. The CAPM states that for an asset i the expected return $E(R_i)$ over some horizon should satisfy

$$E(R_i) - R_f = \beta_i (E(R_M) - R_f) \tag{1}$$

where R_f is the risk free rate over the same horizon, R_M is the corresponding market return and $\beta_i = \frac{Covariance(R_i, R_M)}{Variance(R_M)}$ is the (market) beta of asset i. R_M is usually taken as the return on a broad index of assets and R_f usually proxied by return on Govt discount bills. One can interpret the CAPM as measuring the

*Faculty of Economics, University of Cambridge. Any opinions expressed in this paper are those of the author alone. Please do not quote without permission.

required return on an asset as depending on the amount of risk that asset holds, given by β_i, times the market price of holding risk, given by the expected market excess return $(E(R_M) - R_f)$. The CAPM captures the idea that expected (excess) returns are lower for assets where their actual returns covary less with the market (excess return). We accept a lower expected return on these assets because when market returns drop the individual asset return does not drop so much.

We can rewrite the CAPM in terms of actual returns as

$$R_i - R_f = \beta_i (R_M - R_f) + \epsilon_i$$

where ϵ captures the differences between expected returns and actual returns. This now looks like a regression equation (with zero intercept). The excess return on any asset has two components. One is a linear function of the excess return on the market, this is usually called the non-diversifiable or systematic risk and holding this risk is what requires compensation in the market. The other is a random component which is usually called non-systematic or diversifiable risk.

$$R_i - R_f = \underbrace{\alpha_i}_{\text{zero}} + \underbrace{\beta_i (R_M - R_f)}_{\text{nondiversifiable}} + \underbrace{\epsilon_i}_{\text{idiosyncratic}}$$

In a time series context we define at time t the returns $R_{it} = E(R_{it}) + \eta_{it}$ and $R_{Mt} = E(R_{Mt}) + \xi_{Mt}$ where η_{it} and ξ_{Mt} are then random variables capturing the difference of outcomes from expectations. Under rational expectations these will be zero mean conditional on the information set available. Then we have the following relation between the observed returnsR_{it} and R_{Mt} (we include a constant which the CAPM implies should be zero).

$$\begin{aligned} R_{it} - R_{ft} &= \alpha_i + \beta_i (R_{Mt} - R_{ft}) + \eta_{it} - \beta_i \xi_{it} && (2) \\ &= \alpha_i + \beta_i (R_{Mt} - R_{ft}) + u_{it} && (3) \end{aligned}$$

In practical terms one wishes to obtain estimates of the unknown parameter β_i. The long established technique is to note that if we have a time series $t = 1, \ldots, T$ of observations on R_{it} and R_{Mt} (or on the excess returns) then we can write a regression equation

$$R_{it} = a_i + b_i R_{Mt} + v_{it} \tag{4}$$

the least squares estimator of b is given mechanically by

$$\hat{b}_i = \frac{\widehat{Cov(R_i, R_M)}}{\widehat{Var(R_M)}} = \frac{\Sigma_{t=1}^{T} \left(R_{it} - \bar{R}_i\right)\left(R_{Mt} - \bar{R}_M\right)}{\Sigma_{t=1}^{T} \left(R_{Mt} - \bar{R}_M\right)^2} \tag{5}$$

where$\widehat{}$ denotes sample estimates and $\bar{R}_i = \frac{1}{T}\Sigma_{t=1}^{T} R_{it}$ and $\bar{R}_M = \frac{1}{T}\Sigma_{t=1}^{T} R_{Mt}$ are the sample averages. Since the least squares formula for $\hat{b}_i$ in (4) coincides with the formula for β_i in (1) we can use $\hat{b}_i$ obtained from a simple OLS regression as an estimator of β_i and this can easily be implemented in any standard statistical software.

If the sample averages $\widehat{Cov(R_i, R_M)}$ and $\widehat{Var(R_M)}$ tend to their true values as the sample size grows then this estimator will have desirable properties. On daily data it is usual to assume $R_{ft} = 0$ and in estimation we usually find essentially no difference between the estimates including or excluding the constant.

This motivates the use of ordinary least squares to estimate beta. Note that although the theoretical relationship (2) looks a little like a regression equation it is actually derived from the CAPM relationship (1) which itself follows from assumptions about utility maximisation, law of one price, no arbitrage and other conditions. The expression for β_i comes out of these assumptions. The regression equation (4) is then simply a mechanism to estimate β_i. This difference is important because there is nothing in (1) that prevents β_i from changing over time whereas an implicit assumption in the use of (4) is that the regression coefficient does not change over time. Most empirical applications of this methodology use a rolling sample period of T observations presumably reflecting an underlying belief that β_i is in fact changing in some way over time. The issue to be addressed is whether least squares estimation of (4) can be expected to deliver reasonable estimates of the parameter of interest in (1), in particular in a situation where there may be time variation in β_i.

2.1 Some modelling issues

In estimating beta there are certain issues we have to consider

- the assumptions we are prepared to make about beta
- the purpose for which we wish to estimate beta (in particular, if we wish to forecast β, the horizon over which we may need estimates)
- the nature of the data available in particular the frequency and time interval over which estimation can take place

It is important to recognise that these issues cannot be separated. Given a set of assumptions the appropriateness of a particular modelling strategy can then be evaluated. For example if we assume that beta is a constant then the best strategy is probably to use OLS estimation on all the data available. Low frequency data reduces the number of observations used and may also require some consideration of the risk free rate that appears in the CAPM (both in estimation and in predicting future returns). Higher frequency data provides more precision in estimation and the usual assumption of a zero risk free rate is also more defensible. But if one moves to even higher frequency such as intra-daily data then market microstructure issues (for example non-synchronous trading times or bid-ask spreads) may become an important modelling concern. Finally if beta is indeed a constant then whether we are interested in short horizon forecasts or forecasting beta over longer horizons in this framework the same estimate should be used.

If beta is time varying then estimation can still be done by OLS but other techniques may be more appropriate depending (partly) on the reasons for which we wish to estimate beta and the way in which beta varies. We may be interested in short horizon (conditional) $E(\beta_{t+h} \mid \Omega_t)$ where h may be a few days or weeks and Ω_t represents the information set available at time t. Or we may be interested in much longer run (unconditional) $E(\beta)$ or indeed something in between. The appropriate modelling strategy in the first case may not be the same as in the second. In particular to forecast over longer horizons it will be usually be necessary to specify some model for the time evolution of beta. It is important to see in forecasting some model of the time evolution of beta is always assumed. For example if we have a current estimate of $\hat{\beta}_t$ and use this to forecast forward β_{t+h} we are implicitly assuming $E(\beta_{t+h} \mid \Omega_t) = \beta_t$ for all h.

If beta varies only slowly (relative to the data sampling frequency) then beta in the immediate future may be well approximated by the current estimate, in which case OLS on the most recent data may still be the most useful strategy and this is very much the standard approach where a recent window (of say five years of monthly data) is used to obtain a current estimate of beta. But this current estimate may be a poor guide if beta reverts to some longer run level from its current levels and we are interested in longer horizons.

For example if we assume that β_t evolves as an AR1 around some long run level β^* we might write a simple AR1 model

$$(\beta_t - \beta^*) = \lambda(\beta_{t-1} - \beta^*) + \eta_t$$

and if we wish an estimate of the average β over the next h periods. Then it is easy to show

$$\tilde{\beta}_{t,t+h} = (1 - \theta_h)\beta^* + \theta_h \hat{\beta}_t$$

where $\tilde{\beta}_{t,t+h}$ is the average expected β over the period $t \to t+h$ (so $\tilde{\beta}_{t,t+h} = \frac{1}{h}\sum_{\tau=1}^{h} E(\beta_{t+\tau} \mid \beta_t)$) and $\theta_h = \frac{\lambda(1-\lambda^h)}{h(1-\lambda)}$. The forecast average over the period is thus a weighted average of current the estimate of β_t and the long run (equilibrium) value β^* (which again we may have to estimate). Note as well one implication is that if we estimate or require beta over quarterly data we have roughly 66 days daily beta. Some sort of mean reverting behaviour in the process for β_t could make the former estimate quite different from the latter.

3 OLS estimation

Given a time series of observations $R_{i,t}$ and $R_{M,t}$ over an interval $t = 1, \ldots, T$ the least squares estimator is given by

$$\hat{b}_i = \frac{\Sigma_{t=1}^{T}\left(R_{it} - \bar{R}_i\right)\left(R_{Mt} - \bar{R}_M\right)}{\Sigma_{t=1}^{T}\left(R_{Mt} - \bar{R}_M\right)^2}$$

If the underlying population parameters (ie $Cov\left(R_i, R_M\right)$and $Var\left(R_M\right)$) are constant then under fairly mild conditions the sample estimates will converge to their population values in probability and so we can expect $\hat{b}_i$ to give consistent estimation of β_i. If this is the belief then the appropriate estimation strategy should be to use all available observations on R_i and R_M to obtain the most precise estimates of $Cov\left(R_i, R_M\right)$and $Var\left(R_M\right)$ and the estimate of β_i simply can be obtained from a regression of R_i on R_M over the full sample. This can then be used whether we require an estimate of beta for short or long horizon forecasts.

There is a very large literature using the simple OLS approach to estimation of beta originating with the work of Fama-Macbeth. In most of this literature there is at least an implicit acknowledgment that beta for individual stocks may vary over time. Much of the discussion focusses on the choice of sample period and observation frequency within that window (and the trade offs between these two), for example for a monthly forecast of beta one might use the previous five years of monthly data in estimation). If beta is believed to vary only slowly over the estimation interval and the requirement is for an estimate of beta over a comparable short future horizon this approach has much to recommend it.

3.1 What if beta varies over time?

It seems almost universally accepted that beta varies over time (and this is the usual justification for truncating the observation window for estimation). The hope presumably is that if $Cov\left(R_i, R_M\right)$and $Var\left(R_M\right)$ are approximately constant over some interval $t = 1, \ldots, T$ (at levels σ_{iM} and σ_M^2, say) then the least squares estimator in a regression of $R_{i,t}$ on $R_{M,t}$ will be approximately $\frac{\sigma_{iM}}{\sigma_M^2}$ over that interval and this may provide a suitable estimate for β_i for subsequent use. That is, the average value of beta over the recent past can be used to estimate the expected value of beta over some comparable future interval.

In this context the following questions arise:

1. Is there evidence that beta changes over time?

2. If so does this invalidate OLS as an estimation method?

3. If so what other estimation procedures might be available and what advantages might they have?

We address each in turn.

3.1.1 Is there evidence that beta changes over time?

Figure 1 below shows estimates of beta from a rolling OLS of window size approx two years using daily data for SVT,NG and UU with ASX as the market proxy - these are 500 days ahead rolling regressions over the period 5 January 2000 to 10 Sep 2015 with the final 500 obs running to 31 Aug 2017.

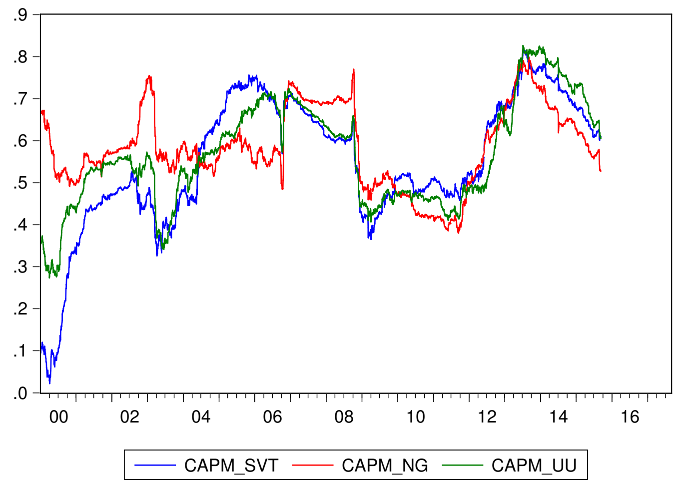


Figure 1: Rolling (daily) beta estimates for utilities

For comparison here's the monthly (using 60 month ie 5yr ahead rolling window again ending Aug 2017)

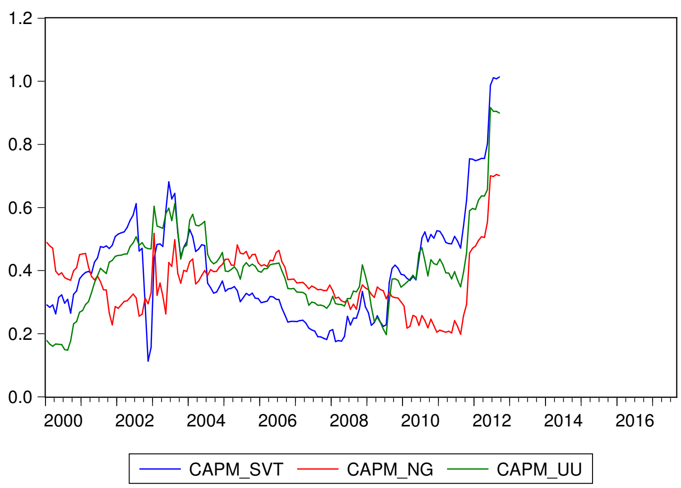


Figure 2: Rolling (monthly) β estimates for utilities

Were beta strictly constant then we should obtain quantitatively similar estimates from daily data and from monthly data. The estimation interval would not make a substantive difference and we should observe little beyond random variation in the rolling estimates around the true (constant) value of β_i for each company, whatever the estimation frequency. This is difficult to reconcile with the Figures. Indeed we see large short term fluctuations where the addition of a single day or a month can generate substantial movement in the estimated β, and possibly a more longer term lower frequency variation.

Since $\beta_i = \frac{Cov(R_i, R_M)}{Var(R_M)}$ time variation in β can only occur if $Cov\left(R_i, R_M\right)$ and/or $Var\left(R_M\right)$ vary over time. We can get a proxy for $Var\left(R_M\right)$ at daily frequency by graphing the squares of the daily return as below. Under homoscedasticity each of these would be an unbiased estimate of the (constant) variance of returns (ie we should see a more or less horizontal line).

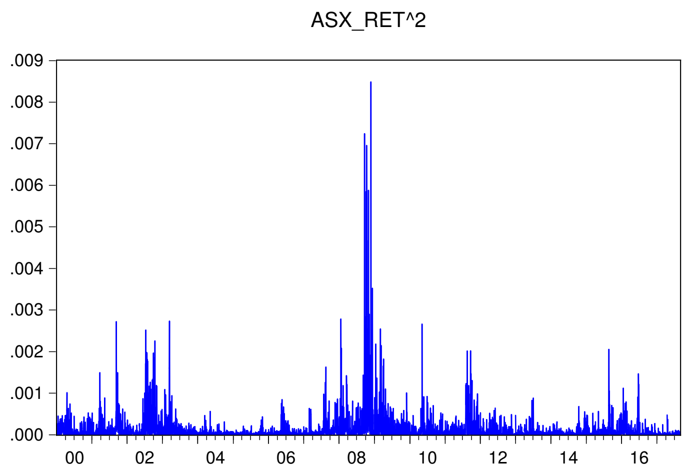


Figure 3: Squared daily stock market returns

The visual impression of time variation is confirmed by a statistical test (a test for no relation between variance at time t and the previous 20 values has $\chi^2_{20} = 1100.8$ which has a p-value around 1.8×10^{-220}). With such overwhelming evidence of time variation in $Var\left(R_M\right)$ the only way to obtain constant β_i would be if $Cov\left(R_i, R_M\right)$ had time variation that precisely mimics that of $Var\left(R_M\right)$. This seems unlikely.

Squared returns for Severn Trent, National Grid and United Utilities are graphed below. Again test for homoscedasticity is overwhelmingly rejected in each case ($\chi^2_{20} = 410, 712$ and 940 respectively)

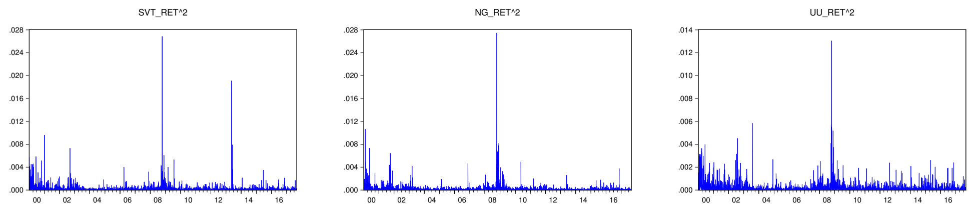


Figure 4: Squared daily returns for utilities

With such evidence of persistent time variation in the variances of these series it is extremely difficult to argue that β should be treated as a constant, except perhaps in the very short run. To make this argument would require that beta is allowed to be time varying but in such a way that it varies only very slowly over the estimation window, so that forecasting using this value over a similar window is a reasonable approximation - that is if beta is approximately constant at monthly intervals then one might try to use a history of monthly observations to forecast one or two months ahead. The problem of course is to obtain a sufficient history of data for estimation would then require using several years of monthly data where the assumption of only slow variation might be difficult to defend. The alternative approach of estimating on a relatively shorter interval of, say, daily data and using this estimate to project beta forward over say a one month horizon is likely to be preferable and this seems to be the main justification for the current approach in financial econometrics. However if one wishes to produce beta estimates for horizons further than days or even months the issue of time variation in the future as well as the past has to be acknowledged. One interesting feature of Figures 1 and 2 is that there appears to be high frequency movements (in beta) that are averaged out in the monthly regressions. This is at least suggestive of time variation in beta driven by a combination of temporary and more persistent shocks. The component GARCH model of Engle and Lee (1999) is one approach that tries to capture this short and longer horizon decay of shocks and provides more flexible specification than the GARCH(1,1) models described below though at the cost of some computational complexity (the component GARCH is a restricted GARCH(2,2)).

3.1.2 Does this invalidate OLS as an estimation method?

If we accept that β_i is time varying the immediate question is what precisely the OLS estimator in the CAPM regression then measures (note that there is nothing to stop us performing such a regression). Let $R_{i,t+1}$ be the one period conditional return on asset i from t to $t+1$ and $R_{M,t+1}$ the corresponding excess market return.

Express the time varying CAPM as (this follows Bollerslev and Zhang, 2003)

$$R_{i,t+1} = \alpha_{t+1|t} + \beta_{i,t+1|t} R_{M,t+1} + \eta_{i,t+1} \quad t = 0, 1, \ldots, T-1 \tag{6}$$

where $\alpha_{t+1|t}$ is the conditional expected return on a zero beta portfolio and $\eta_{i,t+1}$ is zero mean but may be conditionally heteroscedastic.

$$\beta_{i,t+1|t} = \frac{Cov_t\left(R_{i,t+1}, R_{M,t+1}\right)}{Var_t\left(R_{M,t+1}\right)} \tag{7}$$

where now variances and covariances are all conditional on information at time t (and written with subscript t). These conditional betas are, of course, not observed and since we only have a single realisation of the joint time series $(R_{i,t}, R_{M,t})'$ estimation of such time varying moments requires further assumptions.

OLS estimation

If we just run OLS on the T observations at this sampling interval we obtain the estimate

$$\hat{b}_i = \frac{\Sigma_{t=0}^{T-1}\left(R_{i,t+1} - \bar{R}_i\right)\left(R_{M,t+1} - \bar{R}_M\right)}{\Sigma_{t=0}^{T-1}\left(R_{M,t+1} - \bar{R}_M\right)^2} \tag{8}$$

and a simple procedure is to use this as the estimate of beta over the next time interval. For example a standard approach might be to use 3 or 5 years of monthly data to estimate (8) and use this value as the one month ahead β estimate (and then roll the estimation forward). This is the original Fama-MacBeth approach which treats $\beta_{i,t+1|t}$ as a constant in the least squares objective function and chooses an estimation window that is hopefully consistent with that assumption. If this estimate of beta is then used to calibrate future returns then this locally constant assumption has to hold over the future period as well. This is difficult to maintain if the forecast horizon runs to years. If the locally constant assumption is violated then the time variation becomes part of the error structure leading to heteroscedasticity and omitted variable type problems (though in order to assess any bias one also has to specify what the true beta is and consequently decide whether it is actually fixed or time varying).

The exact properties of these rolling OLS regressions when the slope coefficient (and intercept) may be time varying are difficult to calculate and depend on the nature of the joint stochastic process determining the evolution of the conditional market betas and the conditional market risk premium. If the conditional beta on a particular day is equal to some underlying mean value plus a purely random variation for each day then we would have effectively a random coefficient model at daily frequency and OLS will provide unbiased estimates of the mean beta over the sample interval (though this estimate would not be efficient if there is heteroscedasticity). But if the covariance between the conditional β and conditional market return were not zero OLS would not even be unbiased (for the mean beta) in such a model.

Since one can in principle sample at even higher frequency than daily the questions arises over what time interval it might be possible to regard beta as locally constant. In the limit one would have to consider a model where beta evolves continuously in time. This then gives rise to the realised beta approach.

Realised beta

We can make the assumption that returns actually evolve in continuous time but can only be observed at discrete intervals (such as daily, weekly, monthly or quarterly). The process generating the returns will have underlying variance and covariances that are continuous (and possibly time varying) processes. By implication there is also a beta that evolves continuously in time. The realisation of the returns, variances and covariances at the observed frequency will be given by cumulating the returns variances and covariances from the underlying continuous time processes.

In these circumstances an alternative approach is to use the realised beta methodology. To estimate say the variance of a return in a particular quarter we recognise that this can be approximated by the cumulated variances of minute-by-minute or even second-by second returns with the approximation improving the higher the frequency we can sample the data within that quarter. As the sampling frequency increases in the limit we obtain the realised variance measure of Barndorff-Neilsen and Shephard (2002) (see Anderson Bollerslev Diebold and Wu (2006) for discussion). This can also be done to construct realised covariance and hence we can obtain realised beta at say quarterly frequency using higher frequency such as daily (or even intra-daily though remember the caveats above) data.

That is to obtain a direct estimate of $\beta_{i,t+1|t}$where $t \to t+1$ is say one quarter we use the higher frequency (such as daily) data within the sampling period $t \to t+1$. If we label higher frequency returns R_{i,t_j} and R_{M,t_j} for $j = 1, \ldots, N$ during a period t (so j represents for example days within a month) then an estimate of the time t conditional beta is obtained as the ratio of the realised covariance and realised volatility

$$\hat{\beta}_{i,t+1|t} = \frac{\sum_{j=1}^{N} R_{i,t_j} R_{M,t_j}}{\sum_{j=1}^{N} R_{M,t_j}^2}$$

which is simply obtained by a regression at the higher frequency within period t of asset returns on market returns. This formulation allows for continuous evolution of the underlying parameter and produces a consistent estimate of the underlying ratio between the integrated stock and market return covariance and the integrated stock market variance over the sampling period.

Having obtained these realised beta estimates Bollerslev and Zhang suggest two ways of using them to estimate future beta, firstly by just holding the current estimate forward, and secondly by producing a time series of estimates of conditional betas at the lower frequency and then using a rolling window to produce (say) AR1 forecasts of the future beta by simply estimating an autoregression for the time series of realised betas.

In principle this realised β method could be used to produce estimates at much greater horizons. If one specifies a sampling period of say 500 observations (two years) then the OLS estimates within that interval provide an estimate of the current 500 day conditional β. This could be projected forward at its current level (imposing a random walk model on future beta) as the forecast beta. Or one could specify some model such as an AR1 for the conditional beta. However even with a total of 16 years of data there are only 8 non-overlapping two year blocks of daily data on which to calculate conditional betas. To estimate an AR1 model on so few observations and project forward for five or even ten years is likely to introduce substantial noise given known problems with small sample autoregressive estimation. A strategy of reducing the sampling frequency for the CAPM to say quarterly and then use the daily data to produce conditional beta estimates at this frequency would allow a longer time series (of estimated conditional betas) for estimating a forecasting model at the cost of fewer daily observations being used to estimate each conditional beta. Bollerslev and Zhang report some evidence in favour of this high frequency plus AR approach to forecasting β at least at short forecast horizons. This methodology blends a non-parametric approach to estimating the conditional betas with a parametric forecasting model. Note also that a simple OLS over the full sample can be interpreted as a measure of the realised beta over that interval. Section 9 below gives some estimates.

3.1.3 What alternative estimation procedures might be appropriate?

If beta is time varying then OLS might still be used as discussed above. An alternative approach is to make parametric assumptions about the time evolution of either $\beta_{i,t}$ or of the underlying variance and covariances at say daily frequency. There is a long established literature in econometrics for modelling time variation of second moments of financial time series through models of conditional heteroscedasticity. For estimating beta the implication would be that we specify (and estimate) some model that allows the (conditional) distribution of the random vector (R_i, R_M) to evolve over time. As time evolves the joint distribution then implies a new value for $\beta_{i,t}$. We discuss these models of autoregressive conditional heteroscedasticity (ARCH) below in Section 4.

3.2 Summary

Least squares estimation of the CAPM model raises some issues

- The CAPM is not a Classical Linear Regression so there is no presumption that OLS has efficiency properties
- If beta is time varying then a linear regression assuming constant coefficient is misspecified and the model will display heteroscedasticity
- If beta is time varying then LS will attempt to estimate some average beta over the estimation window. Whilst this might be appropriate for portfolio analysis over short horizons, especially if beta is relatively slowly varying, if we are interested in longer run estimates of beta this requires some model of how β_t evolves.

We start by revisiting univariate models with conditional heteroscedasticity.

4 Time series models incorporating heteroscedasticity: ARCH Models

For a long time econometricians working particularly with financial data have been aware that the data displays time varying heteroscedasticity and that this should be taken into account in developing models to describe this data. One feature of the heteroscedasticity is that there appear to be clusters of volatility evident in the data, that is an increase in volatility in period t seems to be accompanied by an increase in volatility in period $t+1, t+2, \ldots$ with a gradual reversion towards a base level before a further shock starts this process again. This observation led researchers to propose models where the volatility in any period t might be linked to the shock in period $t-1$. A type of model that captures such variation is the autoregressive conditional heteroscedasticity ($ARCH$) model first proposed by Engle (1982) and its descendants

Consider a univariate time series (such as an asset return)

$$y_t = \alpha + \varepsilon_t$$

where $V(\varepsilon_t \mid \Omega_{t-1}) = h_t$ is the (conditional) variance of the process and Ω_{t-1} the information available at time $t-1$.

Then the autoregressive conditional heteroscedastic model of order 1 ($ARCH(1)$) is defined by

$$h_t = a_0 + a_1 \varepsilon_{t-1}^2$$

Here a high shock in period $t-1$ (ie a large realisation of ε_{t-1}) raises the variance of the shock process in period t. Note that if we define the surprise in the squared shocks as

$$v_t = \varepsilon_t^2 - E(\varepsilon_t^2 \mid \Omega_{t-1}) = \varepsilon_t^2 - \left(a_0 + a_1 \varepsilon_{t-1}^2\right)$$

so we can write

$$\varepsilon_t^2 = a_0 + a_1 \varepsilon_{t-1}^2 + v_t$$

so the variance of the innovation follows an AR1 process (hence the name ARCH).

Now the unconditional mean of y_t is α, and its unconditional variance is $a_0/(1-a_1) > 0$ provided $a_1 < 1$. The unconditional distribution is not normal even if $\varepsilon_t \sim N(0, h_t)$. Engle shows that the kurtosis is $3(1-a_1^2)/(1-3a_1^2)$ which exceeds three for positive a_1, so that the distribution of y_t has fatter tails than the normal (a feature often found for high frequency financial data).

Higher order $ARCH$ models are defined by

$$V(\varepsilon_t \mid \Omega_{t-1}) = h_t = a_0 + a_1 \varepsilon_{t-1}^2 + a_2 \varepsilon_{t-2}^2 + \ldots + a_p \varepsilon_{t-p}^2$$

giving the p^{th} order model denoted $ARCH(p)$.

4.1 Testing for ARCH

It is relatively straightforward to test whether the residuals from a regression display time varying heteroscedasticity without actually having to estimate the $ARCH$ specification. The squared residuals from an OLS regression are regressed on p of their own lagged values ie

$$e_t^2 = a_0 + a_1 e_{t-1}^2 + ... + a_p e_{t-p}^2 + v_t$$

and T times the R^2 from this regression is a χ^2 with p degrees of freedom under the null that ε_t is i.i.d. $N(0, \sigma^2)$. The test is routinely implemented by most time series regression packages. This is the test reported above.

4.2 Estimation of ARCH models

If $ARCH$ effects are detected then estimation proceeds by maximum likelihood. If we assume the ε_t are normally distributed then the conditional distribution of y_t is normal with mean α and variance h_t. The log-likelihood can easily be derived and maximised numerically. One usually needs quite high frequency data (e.g. daily) to be able to identify $ARCH$ effects well.

Of course there is nothing to restrict us to Gaussian distributions. In fact even though $ARCH$ models with Gaussian errors have fatter tails than the normal, many (particularly financial) series that one might wish to consider seem to have even fatter tails. One common assumption is that ε_t has a t-distribution with ν degrees of freedom with ν a parameter that then enters the likelihood function and is estimated along with everything else - this will tend to generate even more kurtosis in the distribution of the $ARCH$ variable and (the hope would be) a better fit to the data. One could easily make further distributional assumptions.

4.3 Extensions

A large number of extensions to the basic $ARCH$ model have been proposed, and many can easily be implemented in standard regression packages.

4.3.1 GARCH

Perhaps the most widely used extension is to the generalised form of $ARCH$ given by the following equation

$$\begin{aligned} h_t &= a_0 + a_1 \varepsilon_{t-1}^2 + a_2 \varepsilon_{t-2}^2 + ... + a_p \varepsilon_{t-p}^2 + \\ &\quad \beta_1 h_{t-1} + \beta_2 h_{t-2} + ... + \beta_r h_{t-r} \end{aligned}$$

giving a $GARCH(r,p)$ model. This allows for greater serial dependence in the variance term. Estimation is by maximum likelihood, though it is somewhat complicated by the need to deal with presample values of the h_t. Note $GARCH(0,p)$ is $ARCH(p)$. In many financial applications the standard model overwhelmingly used by practitioners is the GARCH(1,1) model which is seen to capture many of the features of high frequency returns. Much research has been devoted to models that (it is argued) better capture the heteroscedasticity in the data, for example the class of exponential-GARCH models capture asymmetries in the effect of shocks on the volatility. But for the purposes of modelling joint distributions the degree of tractability offered by plain GARCH models is attractive.

4.4 Multivariate GARCH

In our setting we wish to model both the return on asset i, R_i, and the market return R_M. Both series display time varying heteroscedasticity so we specify a joint process as

$$\begin{aligned} R_{M,t} &= \alpha_M + u_{M,t} \\ R_{i,t} &= \alpha_i + u_{i,t} \end{aligned}$$

so

$$E\left[\begin{pmatrix} R_{M,t} \\ R_{i,t} \end{pmatrix} \mid \Omega_{t-1}\right] = \begin{pmatrix} \alpha_M \\ \alpha_i \end{pmatrix}$$

$$Var\left[\begin{pmatrix} R_{M,t} \\ R_{i,t} \end{pmatrix} \mid \Omega_{t-1}\right] = Var\left[\begin{pmatrix} u_{M,t} \\ u_{i,t} \end{pmatrix} \mid \Omega_{t-1}\right] = \begin{pmatrix} \sigma_{11,t}^2 & \sigma_{12,t} \\ \sigma_{21,t} & \sigma_{22,t}^2 \end{pmatrix} \quad (9)$$

with a short run (conditional)β then given as

$$\beta_{i,t} = \frac{\sigma_{12,t}}{\sigma_{11,t}^2}$$

We then model the joint distribution in (9) directly by specifying an ARCH or GARCH model for the second moments. Once we have the parameters of that distribution we can infer $\beta_{i,t}$ directly. This avoids the problems discussed above of estimating β from a regression where the time varying nature of β means both that there is unmodelled heteroscedasticity and that the least squares assumption of constant coefficients is violated. Instead we model the heteroscedasticity in the variances and covariances directly and calculate β from the implied estimates. By specifying GARCH models for $\sigma_{11,t}^2$, $\sigma_{22,t}^2$ and $\sigma_{12,t}$ we capture a structure where there can be short run (transitory) movements in $\beta_{i,t}$ around some longer run equilibrium value, that is a large shock to returns in the current period changes the conditional variance and covariance of returns in the next period. This causes β to change in the short run and also affects the probability distribution of shocks next period, and the realisation then transmits this heteroscedasticity forward. We observe clusters of increased volatility reverting to some longer run value before another shock sets off the process again. The sequence of β's, implied by this structure will display possibly persistent movements about some longer run average value.

A variety of different specifications have been proposed for the conditional variance process. Two practical difficulties with these models are first to ensure positive definiteness of the conditional covariance matrix and also the number of parameters can easily grow very large indeed causing computational issues. The BEKK (Baba, Engle, Kraft and Kroner (1990) published as Engle and Kroner (1995)) provides a simple tractable model that ensures positive definiteness of the covariance matrix. Alternatives would be the constant conditional correlation model of Bollerslev (1990). Here the conditional correlation is assumed to be constant while the conditional variances are varying. Obviously, this assumption is impractical for real financial time series. Finally there is the Dynamic Conditional Correlation (DCC) model proposed by Engle (2002) which reduces the number of parameters relative to a BEKK model via further restrictions. If we are modelling only the joint distribution of two returns the number of parameters problem is eased relative to say modelling a large cross section of assets in a portfolio.

We stick with a relatively simple formulation that can already capture persistence in the variances and comovements between returns. The first order diagonal BEKK model runs as follows.

$$\begin{aligned}\begin{pmatrix} \sigma^2_{M,t} & \sigma_{iM,t} \\ \sigma_{iM,t} & \sigma^2_{i,t} \end{pmatrix} = \begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{pmatrix} + \\ + \begin{pmatrix} a_{11} & 0 \\ 0 & a_{22} \end{pmatrix} \begin{pmatrix} u_{Mt-1} \\ u_{it-1} \end{pmatrix} \begin{pmatrix} u_{Mt-1} & u_{it-1} \end{pmatrix} \begin{pmatrix} a_{11} & 0 \\ 0 & a_{22} \end{pmatrix} \\ + \begin{pmatrix} b_{11} & 0 \\ 0 & b_{22} \end{pmatrix} \begin{pmatrix} \sigma^2_{M,t-1} & \sigma_{iM,t-1} \\ \sigma_{iM,t-1} & \sigma^2_{ii,t-1} \end{pmatrix} \begin{pmatrix} b_{11} & 0 \\ 0 & b_{22} \end{pmatrix}\end{aligned} \tag{10}$$

where the returns have constant (conditional) means and time varying conditional variance and covariances.

In detail for the BEKK model we have the following equations for the conditional evolution of $Var\left(R_M\right)$ and $Cov\left(R_i, R_M\right)$

$$Var\left(R_{M,t}\right) = \sigma^2_{M,t} = m_{11} + a^2_{11} u^2_{Mt-1} + b^2_{11} \sigma^2_{M,t-1} \tag{11}$$

$$Cov\left(R_{i,t}, R_{M,t}\right) = \sigma_{iM,t} = m_{21} + a_{11} a_{22} u_{it-1} u_{Mt-1} + b_{11} b_{22} \sigma_{iM,t-1} \tag{12}$$

Note that this essentially specifies a GARCH(1,1) process for each return, and a restricted GARCH(1,1) process for the (conditional) covariance (where the restriction ensures that the variance matrix is positive definite). The implied long run (unconditional) variance and covariance are then (noting that the unconditional expectations $E\left(u^2_{Mt-1}\right) = Var\left(R_M\right)$ and $E\left(u_{Mt-1} u_{it-1}\right) = Cov\left(R_i, R_M\right)$)

$$\begin{aligned} Var\left(R_M\right) &= m_{11} / \left(1 - a^2_{11} - b^2_{11}\right) \\ Var\left(R_i\right) &= m_{22} / \left(1 - a^2_{22} - b^2_{22}\right) \\ Cov\left(R_i, R_M\right) &= m_{21} / \left(1 - a_{11} a_{22} - b_{11} b_{22}\right) \end{aligned} \tag{13}$$

and the model can be used to calculate estimates of $\beta_{i,t+h}$ at future horizons. Maximum likelihood estimation returns estimates of the m's, the a's and the b's together with estimates of $\sigma^2_{M,t}$,$\sigma^2_{i,t}$ and $\sigma_{iM,t}$.

4.5 Estimation of long run beta

If we are purely interested in the long run value a number of possible estimators suggest themselves. One is to use the estimated coefficients directly

$$\hat{\beta}_{LR} = \frac{\hat{m}_{21} / \left(1 - \hat{a}_{11}\hat{a}_{22} - \hat{b}_{11}\hat{b}_{22}\right)}{\hat{m}_{11} / \left(1 - \hat{a}_{11}^2 - \hat{b}_{11}^2\right)} \tag{14}$$

An alternative is to use the estimated values $\hat{\sigma}_{M,t}^2$ and $\hat{\sigma}_{iM,t}$. The time series average of each of these should converge to their long run values ie $Var\left(R_M\right)$ and $Cov\left(R_i, R_M\right)$ respectively. So one can also estimate beta as

$$\hat{\beta}_{avs} = \frac{\frac{1}{T}\sum \hat{\sigma}_{iM,t}}{\frac{1}{T}\sum \hat{\sigma}_{M,t}^2}$$

Finally one can calculate short run conditional betas as

$$\hat{\beta}_{SR,t} = \frac{\hat{\sigma}_{iM,t}}{\hat{\sigma}_{M,t}^2}$$

One could then estimate the long run beta as a simple average of these short runs ie

$$\hat{\beta}_{SR} = \frac{1}{T}\sum_t \hat{\beta}_{SR,t} = \frac{1}{T}\sum_t \left(\hat{\sigma}_{iM,t}/\hat{\sigma}_{M,t}^2\right)$$

This gives three methods to estimate beta from the GARCH -BEKK specification.

Firstly note that for $\hat{\beta}_{LR}$ even if we have unbiased estimates of the coefficients $(m_{11}, m_{12}, m_{22}, a_{11}, a_{22}, b_{11}, b_{22})$ we do not get unbiased estimates of $Cov\left(R_i, R_M\right)$ and $Var\left(R_M\right)$ by plugging in these estimates.

To see this note

$$\begin{aligned} E\left[\hat{m}_{21} / \left(1 - \hat{a}_{11}\hat{a}_{22} - \hat{b}_{11}\hat{b}_{22}\right)\right] &\neq E\left(\hat{m}_{21}\right) / \left(1 - E\left(\hat{a}_{11}\right) E\left(\hat{a}_{22}\right) - E\left(\hat{b}_{11}\right) E\left(\hat{b}_{22}\right)\right) \\ &= m_{21} / \left(1 - a_{11}a_{22} - b_{11}b_{22}\right) = Cov\left(R_i, R_M\right) \end{aligned}$$

and similarly for the denominator. But we have

$$plim\left(\hat{\beta}_{LR}\right) = \beta$$

as long as we have consistent estimates of $(m_{11}, m_{12}, m_{22}, a_{11}, a_{22}, b_{11}, b_{22})$ and $a_{11}^2 + b_{11}^2 \neq 1$ and $a_{11}a_{22} - b_{11}b_{22} \neq 1$.

If the GARCH specification implies stationary processes for the conditional moments $\sigma_{iM,t}$ and $\sigma_{M,t}^2$ then time series averages of the estimated processes will converge in probability to the unconditional values. So again we will have

$$plim\left(\hat{\beta}_{avs}\right) = \beta$$

though if we are fairly close to (integrated) I-GARCH this convergence may be quite slow

Finally note also that for $\hat{\beta}_{SR,t}$ since

$$E\left(\frac{\hat{\sigma}_{iM,t}}{\hat{\sigma}^2_{M,t}}\right) \neq \frac{E\left(\hat{\sigma}_{iM,t}\right)}{E\left(\hat{\sigma}^2_{M,t}\right)} = \frac{\sigma_{iM}}{\sigma^2_M} = \beta$$

an average of the short run $\hat{\beta}_{SR}$ will typically not be unbiased for β.

5 Simulations

5.1 GARCH and ARCH processes

To illustrate the effect of ARCH or GARCH models compared to a situation of IID shocks we simulate three models for returns. Firstly we make 3000 observations as

$$RET1_t = u_t$$

where $u_t$is IID $N\left(0, \sigma^2\right)$ and σ^2 is chosen to match the daily volatility of ASX

Secondly we specify

$$RET2_t = u_t$$

where $u_t \sim \sqrt{h_t}\epsilon_t$ with $\epsilon_t \sim IN\left(0,1\right)$ and $h_t = 0.00008 + 0.4u^2_{t-1}$ is an ARCH(1) model with coefficients chosen to match the estimated ARCH(1) values from ASX

Finally we specify a GARCH(1,1) model

$$RET3_t = u_t$$

where $u_t \sim \sqrt{h_t}\epsilon_t$ with $\epsilon_t \sim IN\left(0,1\right)$ and $h_t = 0.0000017 + 0.87h_{t-1} + 0.11u^2_{t-1}$ is a GARCH(1,1) model with coefficients chosen to match the estimated GARCH(1,1) values from ASX.

The squared returns from each of these simulations are graphed below.

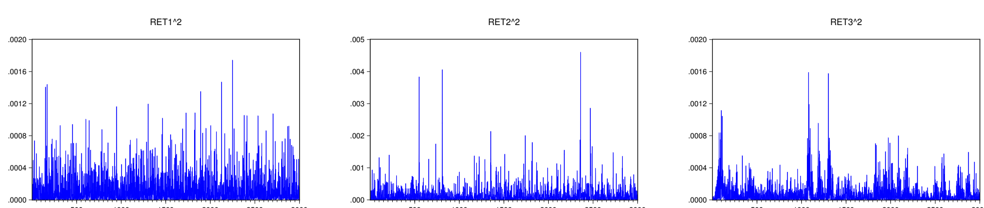


Figure 5: Squared returns for IID, ARCH(1) and GARCH(1,1) processes

We see that the IID return simulation generates squared returns that have a relatively stable pattern over time whereas the ARCH and GARCH models generate clusters of high and low volatility that mimic much more closely the pattern we see in the actual data (note these are simulation so aren't attempting to match the actual pattern observed above).

5.2 Multivariate GARCH

To start we consider what returns series would look like if the world were truly a BEKK-GARCH model, and we investigate the properties of using rolling OLS estimation in this situation. We generate 4000 observations on returns using a BEKK-GARCH(1,1) model. The parameters of the simulation are chosen to match the estimated coefficients from a joint ASX.SVT estimation - these coefficients are set out in the Table below (estimation by maximum likelihood using daily data 9/03/2007-8/31/2017 (2527 obs, roughly ten years).

Coeff	Value used
m_{11}	0.00000139
m_{12}	0.0000015
m_{22}	0.00000532
a_{11}	0.276366
a_{22}	0.184781
b_{11}	0.954948
b_{22}	0.967509

Table 1 Estimated coefficients for simulation exercise

The implied long run β using these parameters using the formula $\beta = \frac{m_{12}/(1-a_{11}a_{22}-b_{11}b_{22})}{m_{11}/\left(1-a_{11}^2-b_{11}^2\right)}$ can be calculated here as 0.50 (to 2dp).

The model is initialised at $u_{M,0} = u_{i,0} = 0$ and $\sigma^2_{M,0}$,$\sigma^2_{i,0}$ and $\sigma_{iM,0}$ set to their estimated unconditional values and then updated using the BEKK formula above. $u_{M,t}$ and $u_{i,t}$ are then generated as random drawings from a joint (conditionally normal) distribution with covariance matrix $\begin{pmatrix} \sigma^2_{M,t} & \sigma_{iM,t} \\ \sigma_{iM,t} & \sigma^2_{i,t} \end{pmatrix}$.

First we graph squared market returns in this simulated world (which should be compared with Figure 3 above)

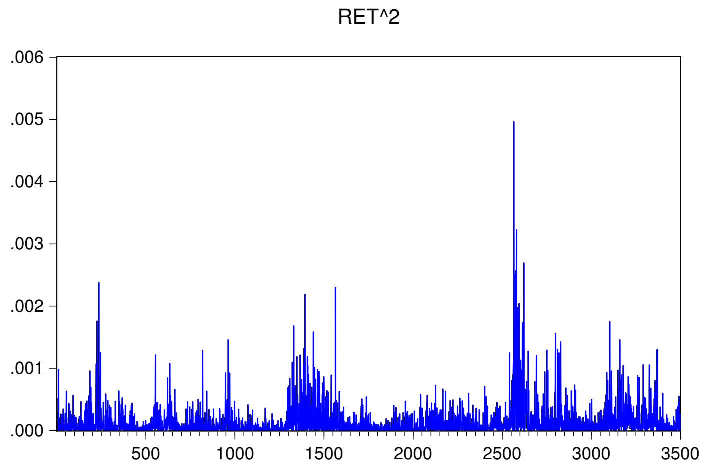


Figure 6: Squared returns for simulated BEKK-GARCH process

Again we see that the BEKK-GARCH model is capable of generating the clusters of volatility seen in the actual data.

Using our 4000 simulated returns we mimic the current procedure by estimating a rolling regression using a window of 500 observations of the simulated stock return on the simulated market return. This gives 3500 estimates of (daily) β. These are graphed below

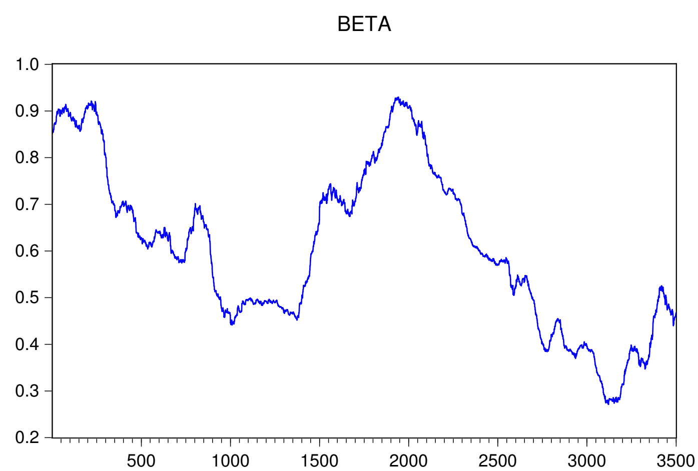


Figure 7: Rolling OLS β estimates from joint BEKK-GARCH process (500 obs window)

The average of the rolling β estimates is 0.61 so in this simulation this average is about 20% too high. It is important to see what this means. In this simulation the true long run value of β is 0.5. In a least squares regression on a randomly chosen window of 500 observations one can expect to see an estimated β of about 0.6. The estimated βs from the rolling regressions are also distributed quite widely around the true value with values as large as 0.9 and as small as 0.3. If one were interested in the short run conditional β (say for portfolio allocation decisions) then the rolling regression may provide a reasonable estimate of the recent average value, but (except by chance) the estimate from any particular regression (or even the average over all regressions) is not a good guide to the true long run value of β (in this case 0.5).

To get a better picture of the effect of the rolling regression approach we repeat this simulation exercise 2000 times. Figure 8 (below) then shows the histogram of the averaged rolling βs from these repetitions (ie in each repetition we estimate 3500 rolling OLS regressions and take the average beta and graph the histogram of the resulting 2000 averages). Relative to the true long run parameter the entire distribution is shifted to the right so both mean and median overstate the true value.

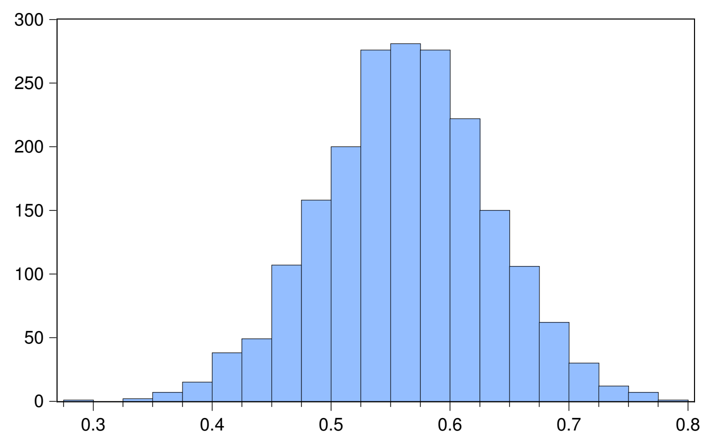


Figure 8: Histogram of average rolling β estimates (2000 repetitions)

Each repetition of 3500 rolling OLS with window 500 from simulated multi-GARCH model. True long run $\beta = 0.5$.

The results presented are for a single simulation model. Unfortunately there are too many dimensions to do an encompassing simulation (that is to vary the original estimation length and parameter values, the simulated data series length, the rolling OLS window length). We have however repeated this exercise using various different sample periods to estimate the coefficients in Table 1, and these are then used to simulate different lengths of artificial returns data and rolling OLS estimation using various window length performed. Throughout we find that the dispersion of the estimated βs typically covers values from 0.2/0.3 up to 0.8/0.9 and the average of these rolling OLS β's tends to overstate the true (calculated) β by a factor of 10-30% with biases seemingly worse when the rolling window is shorter. Further simulations are discussed below.

5.3 OLS vs GARCH approaches

We simulate a world generated as BEKK GARCH(1,1) (parameters are taken from full sample BEKK-GARCH(1,1) estimation for SVT). Error distributions assumed normal. True $\beta_{LR} = 0.4445$

We are interested in the behaviour of four estimators:

1. Rolling OLS using 500 observation window.
2. Full sample OLS which is also the realised beta over this sample.
3. BEKK-GARCH - long run parameter estimated from estimated coefficients.
4. BEKK-GARCH(1,1) long run parameter estimated by ratio of average covaraiance divided by average variance.

For the simulations we consider a sample size of 4000 observations (so roughly 16 years of daily data), 2000, 1000 and 500 (two years daily data).

The Table below summarises the results

Estimation	Mean estimated beta	Standard deviation
Full sample OLS	0.460	0.087
Rolling OLS (500 window)	0.499	0.078
BEKK-GARCH (coefficients)	0.445	0.087
BEKK-GARCH (averages)	0.458	0.084
True beta	0.4445	
Sample size	4000obs	

Estimation	Mean estimated beta	Standard deviation
Full sample OLS	0.468	0.115
Rolling OLS (500 window)	0.499	0.113
BEKK-GARCH (coefficients)	0.448	0.118
BEKK-GARCH (averages)	0.465	0.110
True beta	0.4445	
Sample size	2000obs	

Estimation	Mean estimated beta	Standard deviation
Full sample OLS	0.483	0.148
Rolling OLS (500 window)	0.502	0.167
BEKK-GARCH (coefficients)	0.461	0.181
BEKK-GARCH (averages)	0.478	0.143
True beta	0.4445	
Sample size	1000obs	

Estimation	Mean estimated beta	Standard deviation
Full sample OLS	0.501	0.182
Rolling OLS (500 window)	0.501	0.182
BEKK-GARCH (coefficients)	0.492	0.368
BEKK-GARCH (averages)	0.493	0.184
True beta	0.4445	
Sample size	500obs	

Table 2 Estimating beta on simulated data of various lengths

Comments

- Full sample OLS is pretty good (realised beta). But needs assumptions about forecasting forward. OLS on 500 observations would give 8 non-overlapping betas to forecast forward if one wanted more than two year beta. Likely to have small sample AR problems. OLS on 4000 observations could be forecast forward as is - assumes beta constant in future

- GARCH needs large sample to estimate accurately. Using the estimated coefficients looks risky if we only have say 500 observations. Average covariance divided by average variance looks more well behaved. Some convergence issues in the simulations depending on numerical optimisation routine used.

- Advantage of GARCH model is it can be used to forecast beta forwards directly so obtain transition path to long run as well as long run (if one wants this path - perhaps the pure long run is all we need).

- Need quite long runs of data for the full sample OLS or the BEKK-GARCH models to home in on the true long run parameter. Persistence in the variance and covariance can take beta away for prolonged periods

5.4 Summary

- If beta is time varying it follows that the underlying variances and covariances are time varying. Multivariate GARCH models allows estimation of these time varying second moments directly.

- Simulation of such a model using typical parameters from a utility shows substantial movements in the short term (conditional) beta even when the long run structural beta is constant.
- The simulations suggest rolling OLS estimation of such a model can generate patterns very similar to those observed on real data
- The individual rolling OLS estimates can be far from the true long run beta and rather unstable. Even the average of the rolling OLS coefficients substantially overstates the true parameter.
- OLS using the full sample gets closer to the long run coefficient.
- GARCH estimation provides a good estimate of the long run parameter and also models the short run dynamics of beta

6 Estimation

The following graphs display estimates using daily data for 15 FTSE companies.The companies used and mnemonics are

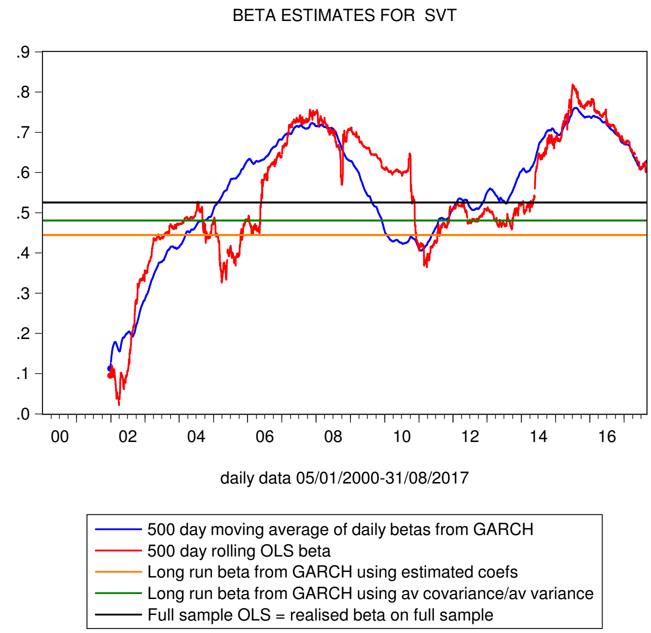
	Company
ATK	Atkins
BBY	Balfour Beatty
BKG	Berkely Group
BVS	Bovis Homes
GRG	Greggs
NEX	Nex Group (ICAP)
NG	National Gridd
PNN	Pennon
SMDS	DSSmith Packaging
SRP	Serco
SSE	Scottish Energy
SVT	Severn Trent
TATE	Tate and Lyle
UU	United Utilities
WEIR	Weir Engineering Group
ASX	FTSE all share index

Each graph plots 5 series - the first two are the time varying, the final three give the horizontal lines.

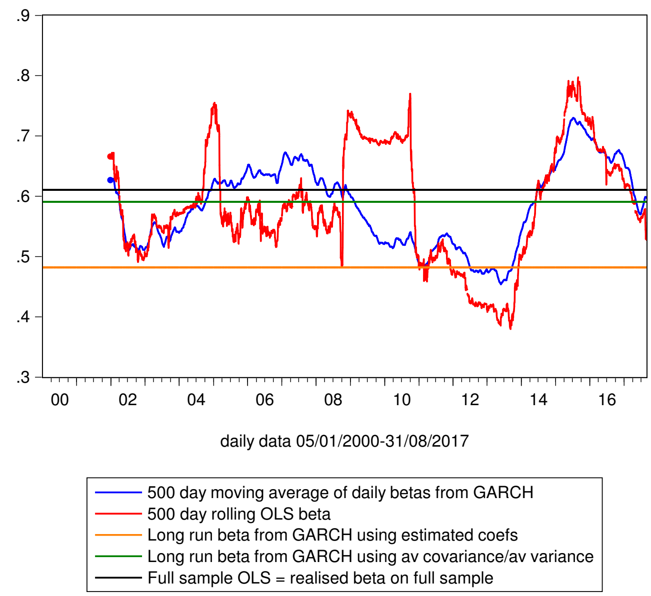
1. Rolling OLS using a 500 observation window [CAPM_name]
2. Short run β_t from a BEKK-GARCH(1,1) model estimated over the full sample and then averaged over 500 obs [@MOVAV(BETAS_name)]
3. Long run β from a BEKK-GARCH(1,1) model estimated over the full sample and calculated using the estimated parameters [BETALR_name]
4. Long run β from a BEKK-GARCH(1,1) model estimated over the full sample using averaged fitted covariance divided by averaged fitted variance [BETALRAV_name]
5. Full sample OLS estimate (ie the realised beta measure over the full sample) [CAPMFULL_name]

We see in each case the moving average of the short run (conditional) estimated β's from the GARCH tracks very closely the 500 observation rolling window OLS. This strongly suggest that much of the time variation in β is being driven by temporary autocorrelated changes in the variances and covariances of the returns series. We also see that full sample OLS returns an estimate not too dissimilar to the long run β derived from the multivariate GARCH estimation. This is at least consistent with a view that there is an underlying long run value of β and short run deviations are driven by time variation of covariances and variances about their long run values. These deviations can be quite persistent. OLS over the longest sample will give an approximation to the unconditional covariance divided by the unconditional variance, though here a sample corresponding to about 16 years of daily data is being used.

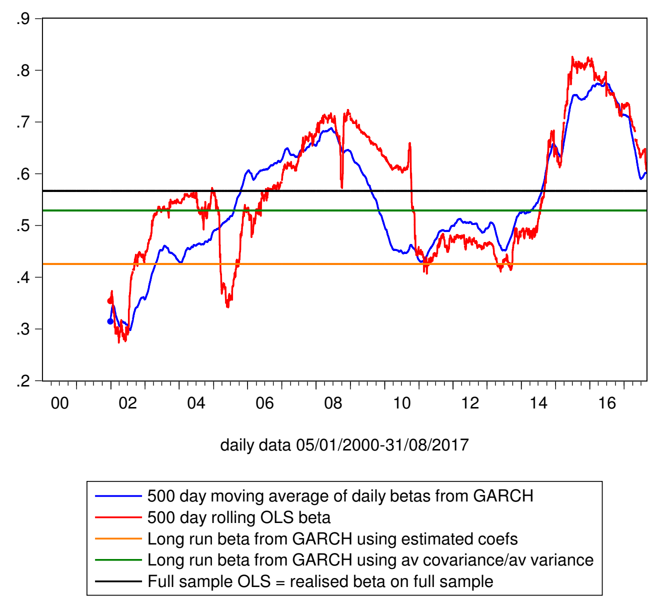
Figure 9 β estimation for various companies, OLS and GARCH approaches. Sample: 1/06/2000 6/30/2017 daily observations. See text for explanation of series.

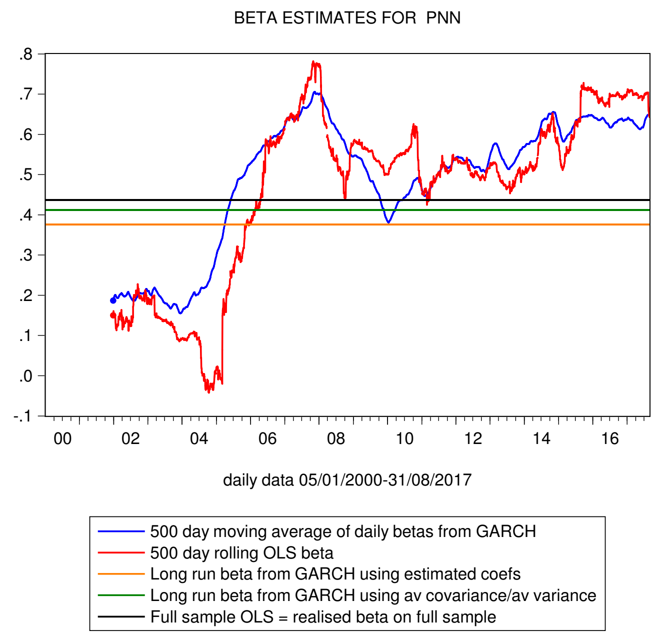


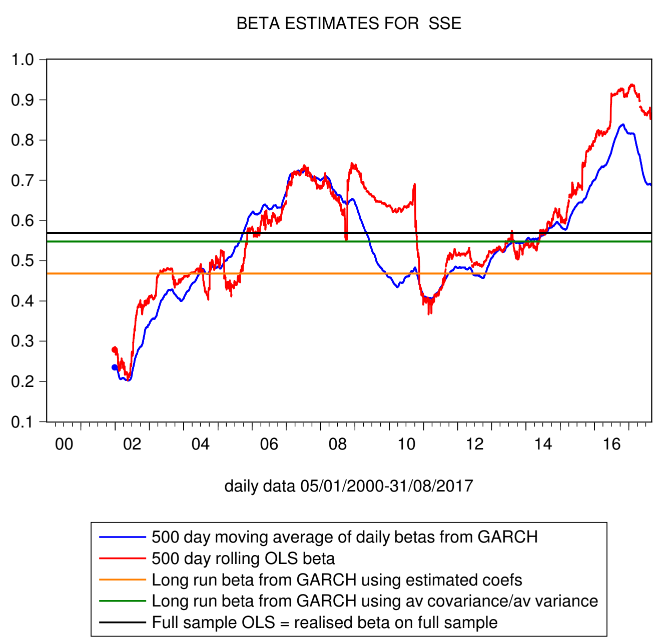
BETA ESTIMATES FOR NG

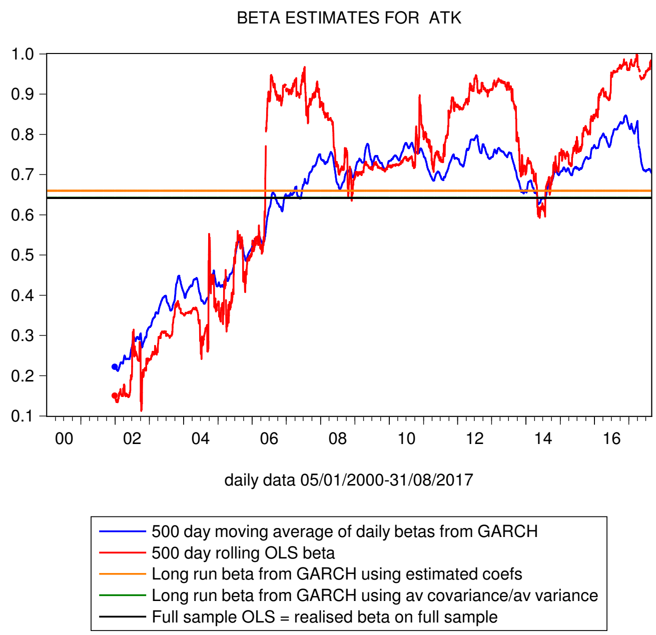


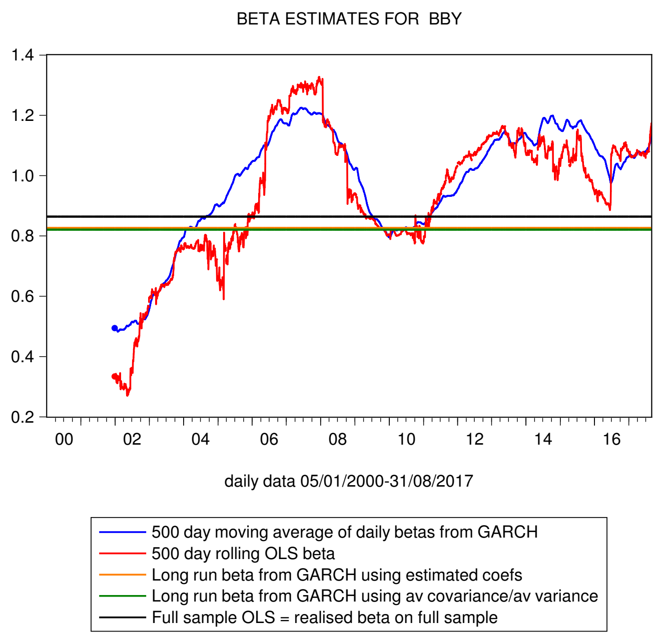
BETA ESTIMATES FOR UU

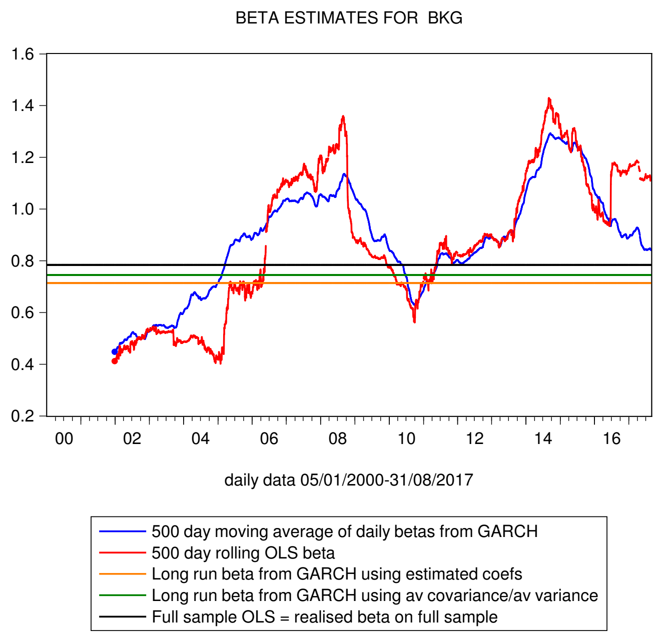


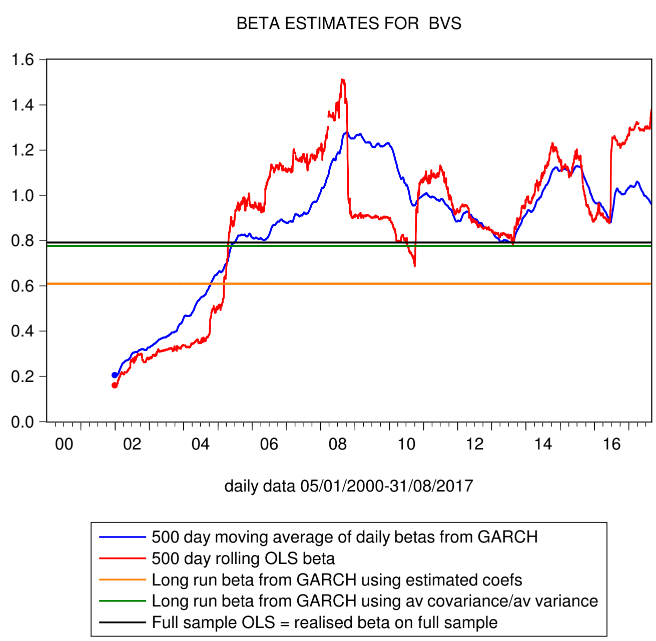

BETA ESTIMATES FOR PNN
.8
.7
.6
.5
.4
.3
.2
.1
.0
-.1
00
02
04
06
08
10
12
14
16
daily data 05/01/2000-31/08/2017
500 day moving average of daily betas from GARCH
500 day rolling OLS beta
Long run beta from GARCH using estimated coefs
Long run beta from GARCH using av covariance/av variance
Full sample OLS = realised beta on full sample

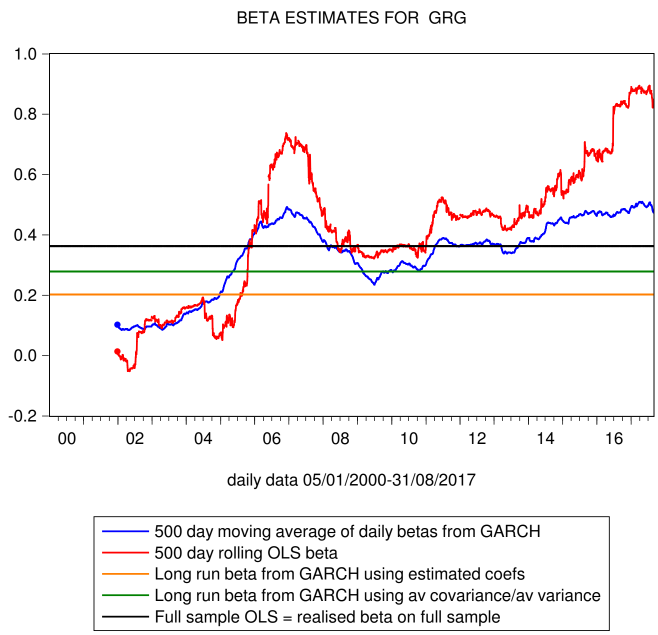

BETA ESTIMATES FOR SSE
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
00
02
04
06
08
10
12
14
16
daily data 05/01/2000-31/08/2017
500 day moving average of daily betas from GARCH
500 day rolling OLS beta
Long run beta from GARCH using estimated coefs
Long run beta from GARCH using av covariance/av variance
Full sample OLS = realised beta on full sample

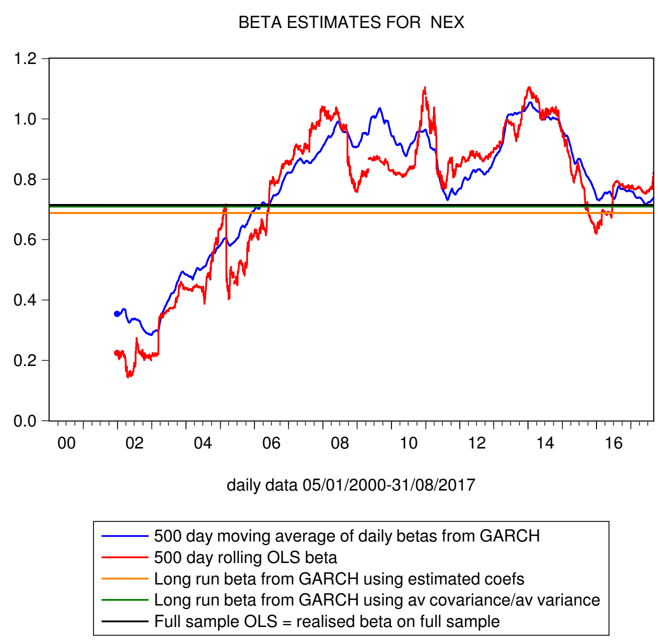

BETA ESTIMATES FOR ATK
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
00
02
04
06
08
10
12
14
16
daily data 05/01/2000-31/08/2017
500 day moving average of daily betas from GARCH
500 day rolling OLS beta
Long run beta from GARCH using estimated coefs
Long run beta from GARCH using av covariance/av variance
Full sample OLS = realised beta on full sample

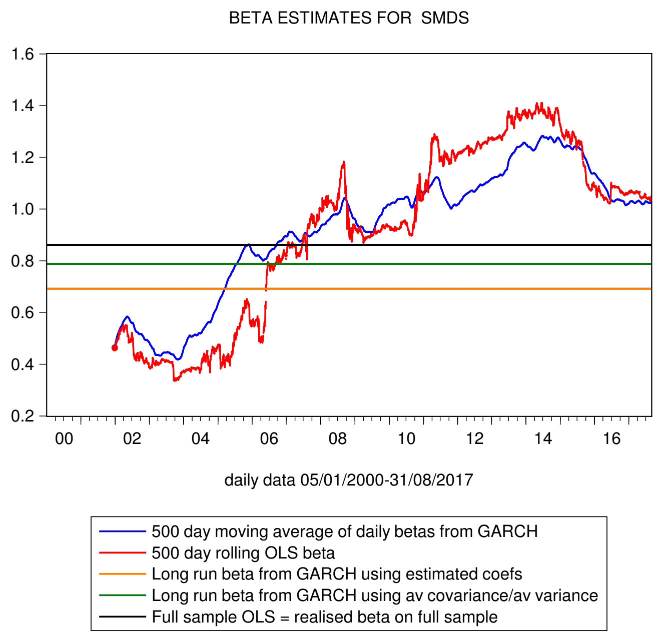

BETA ESTIMATES FOR BBY
1.4
1.2
1.0
0.8
0.6
0.4
0.2
00
02
04
06
08
10
12
14
16
daily data 05/01/2000-31/08/2017
500 day moving average of daily betas from GARCH
500 day rolling OLS beta
Long run beta from GARCH using estimated coefs
Long run beta from GARCH using av covariance/av variance
Full sample OLS = realised beta on full sample

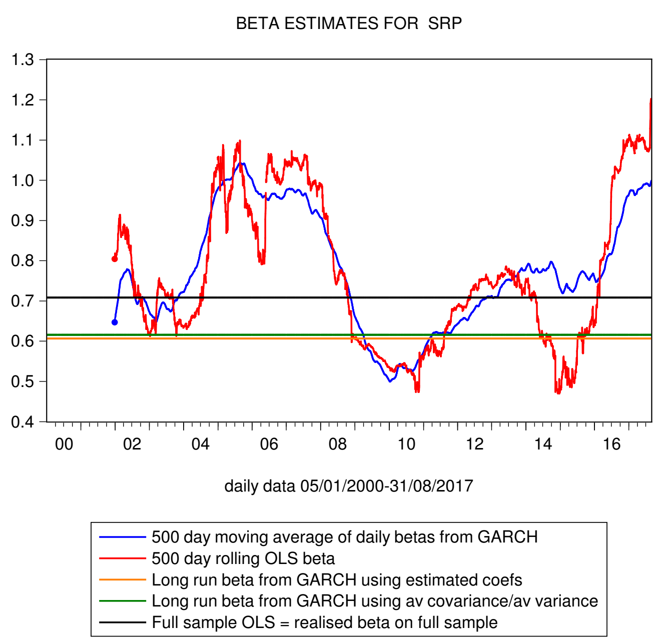

BETA ESTIMATES FOR BKG
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
00
02
04
06
08
10
12
14
16
daily data 05/01/2000-31/08/2017
500 day moving average of daily betas from GARCH
500 day rolling OLS beta
Long run beta from GARCH using estimated coefs
Long run beta from GARCH using av covariance/av variance
Full sample OLS = realised beta on full sample

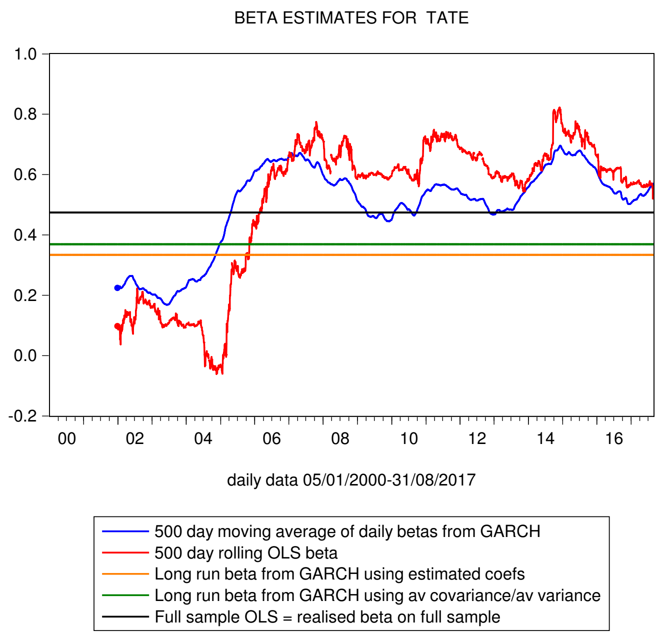

BETA ESTIMATES FOR BVS
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
00
02
04
06
08
10
12
14
16
daily data 05/01/2000-31/08/2017
500 day moving average of daily betas from GARCH
500 day rolling OLS beta
Long run beta from GARCH using estimated coefs
Long run beta from GARCH using av covariance/av variance
Full sample OLS = realised beta on full sample

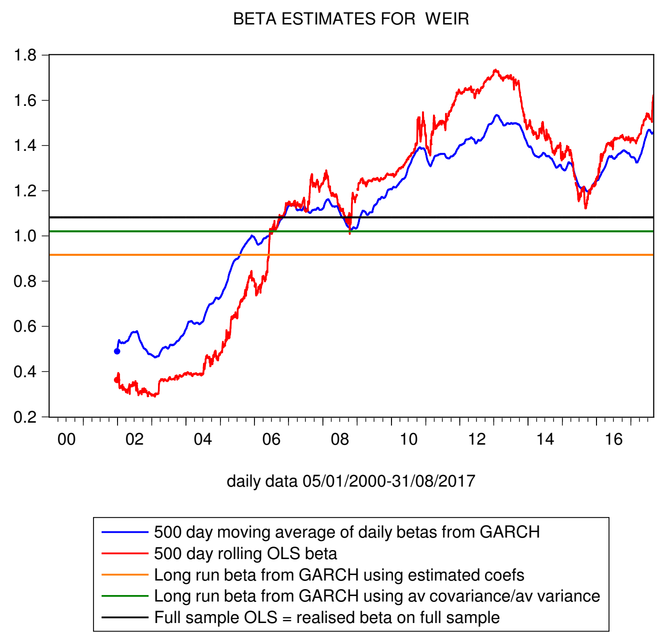

BETA ESTIMATES FOR GRG
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
00
02
04
06
08
10
12
14
16
daily data 05/01/2000-31/08/2017
500 day moving average of daily betas from GARCH
500 day rolling OLS beta
Long run beta from GARCH using estimated coefs
Long run beta from GARCH using av covariance/av variance
Full sample OLS = realised beta on full sample


BETA ESTIMATES FOR NEX
1.2
1.0
0.8
0.6
0.4
0.2
0.0
00
02
04
06
08
10
12
14
16
daily data 05/01/2000-31/08/2017
500 day moving average of daily betas from GARCH
500 day rolling OLS beta
Long run beta from GARCH using estimated coefs
Long run beta from GARCH using av covariance/av variance
Full sample OLS = realised beta on full sample


BETA ESTIMATES FOR SMDS
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
00
02
04
06
08
10
12
14
16
daily data 05/01/2000-31/08/2017
500 day moving average of daily betas from GARCH
500 day rolling OLS beta
Long run beta from GARCH using estimated coefs
Long run beta from GARCH using av covariance/av variance
Full sample OLS = realised beta on full sample


BETA ESTIMATES FOR SRP
1.3
1.2
1.1
1.0
0.9
0.8
0.7
0.6
0.5
0.4
00
02
04
06
08
10
12
14
16
daily data 05/01/2000-31/08/2017
500 day moving average of daily betas from GARCH
500 day rolling OLS beta
Long run beta from GARCH using estimated coefs
Long run beta from GARCH using av covariance/av variance
Full sample OLS = realised beta on full sample


BETA ESTIMATES FOR TATE
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
00
02
04
06
08
10
12
14
16
daily data 05/01/2000-31/08/2017
500 day moving average of daily betas from GARCH
500 day rolling OLS beta
Long run beta from GARCH using estimated coefs
Long run beta from GARCH using av covariance/av variance
Full sample OLS = realised beta on full sample


BETA ESTIMATES FOR WEIR
1.8
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
00
02
04
06
08
10
12
14
16
daily data 05/01/2000-31/08/2017
500 day moving average of daily betas from GARCH
500 day rolling OLS beta
Long run beta from GARCH using estimated coefs
Long run beta from GARCH using av covariance/av variance
Full sample OLS = realised beta on full sample

7 Other Frequencies

In general moving to lower frequency estimation (weekly, monthly or quarterly) would be unwise as it amounts to simply discarding what could be informative data. However the averaging implicit in the lower frequency observations can be expected to reduce heteroscedasticity in the data (as long as the variance structure obeys some limit theorem) relative to what one observes on daily data. We see that for ASX this is indeed the case. The Table below shows the ARCH test discussed above for ASX returns (over the full 2000-2017 sample) at the daily, weekly monthly quarterly frequencies.

Frequency	#lags	test statistic	p-value
Daily	20	1100.8	0.00
Weekly	4	65.2	0.00
Monthly	2	13	0.002
Quarterly	1	1.05	.31

Table 3 ARCH tests for ASX returns at various frequencies.

Sample period is 01/05/2000-2017 - 08/31/2017.

#lags is the number of lagged squared residual terms used in the ARCH test

Test statistics are $\chi^2_{\#lags}$ distributed

As we move to the lower frequencies the evidence for heteroscedasticity diminishes. At quarterly frequency there is no longer any statistical evidence of serial correlation in the (squared) residuals. The pattern is similar for the other returns series.

To the extent that heteroscedasticity ceases to be an issue at lower frequencies we might expect OLS to return "better" estimates of beta in that misspecifications due to the heteroscedasticity are reduced. The problem of course is that at lower frequencies we have many fewer non-overlapping blocks of data. Using even longer intervals of data is of course possible but then raises the problem that over very long periods there may be structural changes in beta due to the changing nature of the underlying business (ie beyond the time variation in beta driven by changing covariances and variances). Table 4 estimates beta by OLS over different sample periods ranging from two years to 17 years using various observation frequencies.

	Full	Sample	00-17
	SVT	NG	UU
Daily	0.53(0.02)	0.61(0.02)	0.57(0.02)
Weekly	0.46(0.04)	0.60(0.03)	0.50(0.03)
Monthly	0.36(0.09)	0.43(0.08)	0.36(0.08)
Quarterly	0.29(0.13)	0.48(0.12)	0.30(0.13)

	Ten	Years	07-17
	SVT	NG	UU
Daily	0.60(0.02)	0.61(0.02)	0.60(0.02)
Weekly	0.55(0.04)	0.59(0.04)	0.56(0.04)
Monthly	0.37(0.11)	0.42(0.10)	0.43(0.11)
Quarterly	0.17(0.17)	0.40(0.16)	0.22(0.19)

	Five	Years	12-17
	SVT	NG	UU
Daily	0.68(0.03)	0.64(0.03)	0.67(0.03)
Weekly	0.74(0.08)	0.68(0.06)	0.76(0.08)
Monthly	1.01(0.20)	0.70(0.20)	0.90(0.22)
Quarterly	0.20(0.37)	-0.14(0.40)	-0.00(0.53)

	Two	Years	15-17
	SVT	NG	UU
Daily	0.61(0.05)	0.53(0.04)	0.62(0.05)
Weekly	0.69(0.11)	0.55(0.11)	0.62(0.12)
Monthly	0.96(0.37)	0.88(0.48)	0.96(0.45)
Quarterly	1.01(1.22)	0.04(1.29)	0.75(1.47)

Table 4 Beta estimates from OLS CAPM model over different sample periods at various frequencies (standard errors in parentheses)

Consistent with the discussion above the OLS estimates at lower frequencies are in much closer agreement with the long run estimates of beta obtained from the multivariate GARCH approach. As we move to lower frequencies we generally see a decline in the estimated betas. But for the shorter estimation windows using lower frequency data the estimates essentially become uninformative with very high standard errors.

The conclusion would seem to be that a sensible estimation strategy would be to use high frequency data and also to use the longest estimation window possible unless there clear evidence of changing structure of the business. Variations in short run estimates of beta are to be expected in a heteroscedastic world and a simple constant long run beta multi-GARCH structure is able to mimic the observed time variation in beta pretty well.

7.1 Structural changes and sample period

Nokia was a paper processing business that became a telecoms company. It is difficult to believe this would not affect its stock market beta. So we need to consider the possibility of structural shifts in which all the underlying parameters (of the joint distribution) could change. This argues against using very long runs of data which may include such shifts. If we maintain the idea that forecast of beta should be some weighted average of current and longer run estimates then this would suggest maintaining a rolling window for estimation.

The Table below shows BEKK-GARCH(1,1) estimates of the long run beta for 15 companies using various windows of daily data (computed from formula)

	01/05/2000-08/31/2017	09/01/2007-08/31/2017	09/03/2012-08/31/2017
ATK	.66	.61	1.05
BBY	.83	.88	.97
BKG	.71	.80	1.05
BVS	.61	.89	.96
GRG	.20	.39	.57
NEX	.69	.58	.72
NG	.48	.46	.65
PNN	.38	.42	.63
SMDS	.69	.88	1.1
SRP	.61	.59	.77
SSE	.47	.49	.69
SVT	.44	.50	.67
TATE	.33	.42	.66
UU	.43	.45	.64
WEIR	.92	1.21	1.27
#obs	4463	2527	1264

Table 5 BEKK-GARCH estimates over various sample periods, daily data

We see a pretty common pattern in which the 5yr estimate>10 yr estimate > full sample estimate. This suggests a strategy of estimating short run and long run beta from a window of say ten years of data and updating the estimates every two years or so. Current forecasts of beta would then be some weighted average of these two values. The weight depends both on the degree of persistence of beta and the horizon for which forecast is required. If the horizon is not 0 or ∞ then some model for the time evolution of beta is needed to obtain the weights.

8 Future beta

Estimation can tell us something about past beta. What is more relevant for thinking about returns on assets is beta in the future. So we need some method to link these two. If beta is a constant or, at the other extreme, a random walk[1], then the current estimate should be used as the estimate of future values. However if beta reverts towards some longer run value then this may not be a sensible approach. It is difficult to do this without specifying some model for the time evolution of beta. For example if one were interested in (average) beta over the next 500 days then one could use a rolling 500 observation window and take the final value as the current estimate. One could then augment with an autoregressive model for the evolution of β_t estimated over the history of the rolling beta and project forward using this model.

If we take the rolling beta estimate for SVT shown in Figure 1 the estimated first order autocorrelation is 0.9985 (3964 obs). Such a high degree of autocorrelation is obviously driven by taking 500 day averages of an underlying changing beta (we get slightly different values if we calculate the autocorrelation directly, if we specify an OLS regression on the lagged value or if we use eviews maximum likelihood AR(1) routine which is perhaps not unexpected given the probable presence of MA terms in a regression of such a β_t on its lag and that we are estimating so close to the unit circle, but the argument for calculating beta forward remains the same). Projecting this value forward 500 days implies a coefficient of $(0.9985)^{500} = 0.472$ on the current value so in this setup the two year forward β would have reverted roughly half way to its long run value.

In the GARCH framework estimation provides not only an estimate of current β_t but also estimates the equations of motion for $Cov_t\left(R_i, R_M\right)$ and $Var_t\left(R_M\right)$ given in (11) and (12). These can be iterated forwards to give estimated beta at future horizons.

Notice that the strategy of tailoring the estimation window to the desired forecast window (so for example if one is interested in the one month ahead beta one could estimate by OLS on monthly data and this would provide a measure of the (average) beta over a one month interval that could be used becomes infeasible if the forecast horizon goes much beyond a quarter, there just isn't a sufficient run of returns data to estimate accurately such models.

9 Realised beta estimation

For comparison we also construct realised beta at quarterly intervals and fit an AR1 model to the resulting time series. We estimate realised beta over the 2000-2017 period using 66 days (one quarter) non-overlapping intervals for each realised value. Realised beta is then the ratio of realised covariance given by $\sum_{j=1}^{66} R_{it_j} R_{Mt_j}$ in each quarter t to realised variance $\sum_{j=1}^{66} R_{Mt_j}^2$. An AR1 model is fitted to the resulting time series. Results are reported in Table below (the long run is estimated as $\frac{\alpha}{1-\lambda}$ in the regression $y_t = \alpha + \lambda y_{t-1} + v_t$).

[1] Strictly speaking a martingale

		SVT	NG	UU
Realised	const	0.23	0.46	0.39
beta	AR1	0.55	0.22	0.27
regression	DW	2.01	1.92	2.03
long run		0.51	0.59	0.53

Table 6 Realised beta regressions.

Realised quarterly betas estimated from 66 daily obs.

The AR1 regression is over 68 observations

The long run estimates are simlar to those obtained by the GARCH model. The persistence of the realised beta (at quarterly frequency) is quite low for NG and UU, and even the SVT autoregressive parameter implies relatively swift reversion to the mean. If we use only the last ten years of realised quarterly betas the implied long run values are SVT=0.59, NG=0.58 and UU=0.58.

10 Conclusion

We argue

1. If beta is believed to be a constant one should use the full sample of data, at the highest frequency where accurate measurement is possible, unless there is evidence of a clear structural break in the nature of the underlying business.

2. However there is overwhelming evidence that beta is time varying and these variations can be quite persistent.

3. If beta is time varying forecasts over different horizons really neeed some model of how beta evolves. If beta is stationary then forecast will be some weighted average of current and long run levels. The weight depends on forecast horizon and the persistence of beta.

4. Simulations show much of the pattern of time variation seen in beta would arise from time varying covariance and variances around longer run level

5. Simulations show that if the world is BEKK-GARCH estimation over short run by OLS are overestimates of true beta. GARCH short run estimates are similar suggesting the OLS is actually estimating averages of the conditional betas.

6. Over longer run of data both OLS and GARCH can return good estimates of long run beta. GARCH also implicitly models time variation of beta (ie persistence).

7. Long run estimates from GARCH and OLS are quite similar. The (averaged) conditional beta from the GARCH and the rolling 500 obs window OLS estimates track each other closely.

8. Realised beta assuming an AR1 model for the conditional beta provide similar long run estimates.

9. Using lower frequencies eliminates a lot of the heteroscedasticity and gives estimates closer to the long run betas. But this requires a much longer sample of data for estimation.

10. There is still the possibility of stuctural change. This suggests that using a rolling window may still be sensible.

11. This suggests a strategy of estimating beta over say ten years of daily data and constructing forecast using weighted average of conditional (short run) and unconditional (long run) estimates. Weight depends on forecasts horizon and persistence of beta.

Estimates of long run beta from various estimation techniques all based on daily data are reported below.

	SVT	NG	UU
Full Sample OLS	0.53	0.61	0.57
GARCH (coefs)	0.44	0.48	0.43
GARCH (avs)	0.48	0.59	0.53
Realised beta LR	0.51	0.59	0.53

Table 7 Estimates of long run beta various methods for utilities

Notes:

1. OLS is full sample daily data 2000-2017
2. GARCH (coefs) uses estimated params from BEKK-GARCH(1,1)
3. GARCH (avs) uses averaged covariance and variances from BEKK-GARCH(1,1)
4. Realised beta uses 66 daily observations for each realised variance and covariance and then fits an AR1 to resulting series

Bibliography

Anderson, T.G., T Bollerslev, F Diebold and G Wu (2006) "Realised beta: persistence and predictability" *Advances in Econometrics: Econometric Analysis of Economic and Financial Time Series Vol B* Elsevier.

Baba, Y., R F Engle, D F Kraft, & K F Kroner (1990). "Multivariate simultaneous generalized ARCH", mimeo. Department of Economics: University of California, San Diego.

Barndorff-Neilsen, O E and N Shephard (2002) "Estimating Quadratic Variation Using Realised Variance," *Journal of Applied Econometrics*, 17, 457-477.

Bollerslev, T. (1990) "Modelling the Coherence in Short-Run Nominal Exchange Rates: A Multivariate Generalized ARCH Model" *Review of Economics and Statistics* , 72, 498–505.

Bollerslev, T. and B Y B Zhang (2003) "Measuring and modeling systematic risk in factor pricing models using high-frequency data" *Journal of Empirical Finance* 10 533-558.

Engle, R F (1982) "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation". *Econometrica*, 50, 987–1007.

Engle, R.F. (2002) "Dynamic Conditional Correlation – A Simple Class of Multivariate GARCH Models" *Journal of Business and Economic Statistics* 20(3), 339–350.

Engle, R.F and K F Kroner, (1995) "Multivariate Simultaneous Generalized ARCH" *Econometric Theory*, Volume 11, Issue 1 February 1995 , pp. 122-150.

Engle, R.F and G Lee (1999) "A Permanent and Transitory Component Model of Stock Return Volatility" in *Cointegration Causality and Forecasting* OUP.