Ofgem

Beta Study – RIIO-2

Main Report

Final

December 2018

Indepen Limited, 4th Floor, 1 Fore Street, London, EC2Y 9DT
www.indepen.uk.com

Indepen is a management consultancy working with clients facing the challenges of regulation, deregulation, competition and restructuring. We help investors, boards and senior managers identify and assess political and regulatory risk and to develop and implement internal and external strategies to manage their exposure.

Our clients are the organisations involved in financing, constructing, managing and regulating built and natural infrastructure – water, energy, transport, land and property. We have constructive relationships with relevant government departments and agencies.

Our team combines experience of public policy, regulation, corporate finance, communication and engagement and organisational development. We complement this with input from our associates – CEOs and chairs of FTSE and privately owned companies, regulators, government ministers and academics.

The Indepen Forum provides the opportunity for investors, government and business leaders to debate, under the Chatham House Rule, issues that if mishandled could undermine well-intentioned policy initiatives.

Table of Contents

IMPORTANT NOTICE

Executive Summary

Introduction

For capital intensive infrastructure companies, the weighted cost of capital (WACC) when combined with the regulatory asset base forms the largest building block in the estimation of cost. All the elements of the WACC have come under scrutiny in the recent past. The measure of relative risk (the equity beta – β) is an element where historically regulators have adopted a range of approaches.

Concerns about the evidence on equity βs is that the estimates require multiple assumptions and they appear to be unstable.[1] Critically, the size of the estimate has a material effect on the ability of companies to cover the cost of funding new investment and remunerating previous investments. The observed volatility of the estimates leads to questions about the estimation approaches adopted.

Regulators have tended to take a conservative view of the equity β to use in setting prices which has led to the adoption of values greater than those justified by the available statistical evidence. [2] One reason for this is that the regulator thinks the cost to the economy of using a too low estimate of the WACC is greater than the cost of using an estimate that is too high.

This report investigates the following issues in establishing a defensible estimate of the equity β.

- Whether there are structural breaks in the data used for estimation and the implications for the estimates of the cost of capital.
- Whether the standard approach to estimation, using OLS regression analysis, is appropriate given the assumptions underlying the method.
- The choice of data frequency for the returns (daily, weekly or monthly).
- Whether the estimation window (which could be two-years, five-years, the period since the last structural break) has an impact on the appropriateness of using OLS.
- Whether other estimation approaches are more consistent with the characteristics of the underlying data.
- Whether alternative approaches to estimating risk provide valid options for estimating equity β values, or at least providing supporting information.

In addition, we consider how equity βs have been used by regulators in relation to

[1] Similar concerns have been raised about the estimation of other cost drivers like cost assessment where multiple econometric models are used with strong assumptions around aspects such as economies of scale.

[2] See for example Jenkinson (2006) and Ofgem (2004b).

- the impact of capital structure (gearing) on the β value and whether the approach of de-gearing and re-gearing equity and asset βs is appropriate
- establishing individual business unit values when an observed equity β is a portfolio value for a group of businesses

In conclusion we develop an assessment of

- options for a process to generate defensible equity β values
- an appropriate range for values today

Some of these issues have recently been addressed in a study undertaken for UKRN, *Estimating the cost of capital for implementation of price controls by UK Regulators* by Wright et al (2018) and one for Ofgem, *Estimating β*, Robertson (2018).[3] In this study we address these and additional questions from a broader perspective and seek to understand the interlinkages between the different issues around the estimation.

Direct measurement

Since the privatisations of the 1980s and early 1990s, the number of listed UK infrastructure companies has fallen from a high of over 30 in the mid-1990s to six now. Understanding the data was important even with the larger set of companies and it is even more important given the role that each listed company provides today in the estimation process.

In an ideal world the estimation of equity β would be based upon all available information back to the date of listing. However, given the likelihood of structural breaks due to company specific, regulatory or market wide factors, the data used for estimation may be restricted. If structural breaks affect relative risk, it will be important to know whether an event had a significant effect or not and whether the effect is permanent or transitory. The answers to these questions will affect the time period over which we can derive stable estimates of equity βs.

Our initial analysis was focused on identifying the possible timing of structural breaks. We used a range of tests, described in Section 2 of the report. If we focus on the post 2000 period, multiple structural breaks can be identified for each of the six listed companies. For the majority there was a break in September/ October 2008.

Ideally, analysis would focus on the period since an identified structural break as this should result in a stable estimate of the equity β. In the report we have considered three possibilities: post-2000, post-2008 and post-2013. These periods correspond to what we see to be commonly used and each choice has implications for interpretation of the results of the analysis. For example, if the post-2000 period were used, the impact of the global financial crisis would need to be factored into the interpretation of the results.

[3] This paper provides a good overview of the issues linked to GARCH and technical aspects of the ways in which GARCH βs can be estimated that are not covered in this report.

Having decided on the estimation window, it is necessary to determine the best method of estimation. Given the prevalence and simplicity of the OLS estimator, the preferred approach of UK regulators is to estimate equity β using daily returns over a rolling window of two years (or sometimes more) going back at least five years. Given our conclusions on structural breaks, we considered rolling estimates over various windows for the three time periods since 2000 and using daily, weekly and monthly returns data.

In each case we tested for homoscedasticity in the errors. The existence of heteroscedasticity in errors may be evidence of time variation or an underlying non-linear process for the equity β. Concerns that the underlying assumptions made by OLS generally do not hold are valid. In almost all cases, this assumption did not hold when using daily or weekly returns; the findings were less clear with monthly returns.

In the presence of heteroscedasticity and based on the existence of a time varying β a common approach is to adopt the Generalised Autoregressive Conditional Heteroscedastic (GARCH) model. Various GARCH specifications exist and a process is needed to find the one that best characterises a given company's data. This leads us to company specific solutions. If outliers are a concern, then Least Absolute Deviation (LAD) offers a possible solution (described in Appendix D).

In this report we focused on the GARCH (1,1) model for the period since-2000. In Annexes to Appendix C we examine the sensitivity of our findings to the use of higher order models and the other two time periods.

Indirect measures

Given the paucity of data, and the issues around the stability of the estimates of equity β, we considered other sources of information on β

- accounting β values - estimates of β based on accounting relationships rather than equity market data
- other risk measures – such as elements of the yield to maturity on debt which provide an indication of the market's view of the riskiness of a business

We find that neither approach is a viable alternative to direct estimation. The accounting β approach is relatively untested, often does not provide statistically significant results and does not always provide parameter estimates consistent with prior expectations.

Other risk measures can be informative but they too require significant assumptions. We conclude that, at best, they may provide a possible lower bound estimate for the equity β.

Other issues

Other issues considered are the role of asset βs; and how portfolio βs should be disaggregated.

Regulators tend to use forecasts of equity β based on their view of a notional efficient capital structure. This is often different from the actual gearing of the regulated business and so regulators have sought to adjust observed values of β to allow for this. Several elements of the adjustment process need to be considered, the two most important being debt βs and the measurement of gearing.

If, as we believe is the case, debt βs are not zero as was once assumed, we need a way to generate reliable debt β estimates. Evidence suggests that debt βs vary between companies and over time, so a simple assumption of a single constant value may be unhelpful. More work on estimating debt βs is needed but for short- and medium-term applications, it would seem to be necessary to adopt pragmatic approaches. We note UK regulatory precedent of debt βs in the range 0.05 to 0.22 and observed that academic literature, albeit not without issues, typically supports this range or higher.

Moving between actual and notional gearing requires clarity about what the measure of gearing. Conventionally, gearing is measured on an Enterprise Value basis, assuming the book and market values of net debt are similar. Notional gearing tends to be measured on an regulatory asset base (RAB) basis where the equity is valued at an indexed book value rather than market value. If the market to asset ratio (MAR) is close to unity then a notional RAB estimate can be used alongside an observed Enterprise Value. However, if the MAR is not close to unity and varies over time then this simplification may not be appropriate and mixing the two approaches to gearing will lead to an unreliable estimate of the equity β.

Solutions to this consistency issue could be greater transparency of assumptions or the application of a "normal" MAR to adjust the RAB gearing estimate to make it a closer proxy to an Enterprise Value measure.

The fact that most listed infrastructure companies undertake more than one activity, both in the UK and internationally, means that an observed equity β is effectively a portfolio β. Ideally, the portfolio β should be disaggregated into its constituent parts so that we can focus on the β for the regulated activity. While in principle disaggregation is straight-forward, its application requires numerous assumptions with a material impact on the estimate. As data constraints make disaggregation imperfect, the best solution is to be clear about the assumptions being made so that the implications can be appreciated by those who will use the results.

International comparators are also proposed as a solution to data limitations. While there are more comparators listed in other countries that can be considered we believe that

- any estimate derived from these companies needs to be subjected to the same analysis that we apply to UK companies, so a full investigation of structural breaks, model selection etc

- issues linked to de-gearing and re-gearing become even more pronounced owing to different tax regimes
- clarity about what the results mean when different economies have different underlying risk as well as possible differences arising from regulatory regimes makes drawing definitive conclusions difficult

Taken all together we do not think international comparisons provide an answer to the question of how to estimate a β for setting price controls in the UK.

Findings and approaches

We find that the available data are subject to structural breaks and the approaches that have been used have weaknesses that affect their statistical validity. The weaknesses stem from the characteristics of the data and inconsistencies in the assumptions made (or a lack of transparency in those assumptions). The weaknesses do not mean that the estimates are always wrong, but that testing their reliability is difficult.

The regulatory regime needs a credible way of estimating the equity β for price control periods. This will generally mean forecasting a value for a five-year period but given the duration of the price review processes, regulators will need to forecast further ahead than five years. In the light of the findings, we conclude that there are three approaches for the measurement of equity βs for consideration.

Approach A – establish the range of feasible results and agree a set of principles for judgements to identify the estimate to be used

- Start from a set of definitions, principally about actual and notional gearing, that are internally consistent
- Consult on the factors that should affect a judgement of the equity β for use in setting prices
- Collate an extensive data set – probably for the period from 2000, even though this will include some structural breaks
- Review data for structural breaks and decide how to proceed
- Consider the distribution of results from estimates using different time windows and frequencies of returns (this can include using OLS and other estimation approaches)
- Apply judgements derived from the consultation process to arrive at the preferred estimate of the equity β within the distribution
- Where portfolio βs need to be decomposed, make explicit assumptions including about taxand gearing

Approach B - a more technocratic approach using a decision tree and well-defined criteria

- Start from a set of definitions, principally about actual and notional gearing, that are internally consistent

- Collate an extensive data set – probably for the period from 2000
- Review data for structural breaks and decide how to proceed
- Where portfolio βs need to be decomposed, make explicit assumptions including about tax and gearing

The decision tree, an example of which is given in Figure E1, is for dealing with the issue of autoregressive conditional heteroscedasticity. It assumes we have at least a decade of data since the last structural break. If the structural break is more recent then we might use a shorter time period or place greater weight on the evidence from the period since the break.

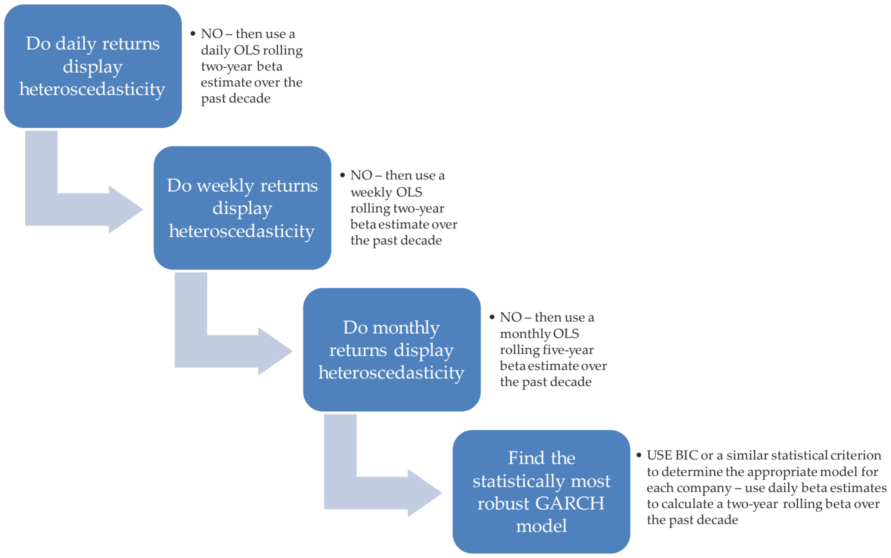


Figure E1: Decision tree for estimating equity β

Approach C - follow the approach used by the Australian Energy Regulator - decide on a number and stick to it unless there is a good reason to the contrary.

Summary of results

The results we have generated for equity βs are summarised in figure E2 below.

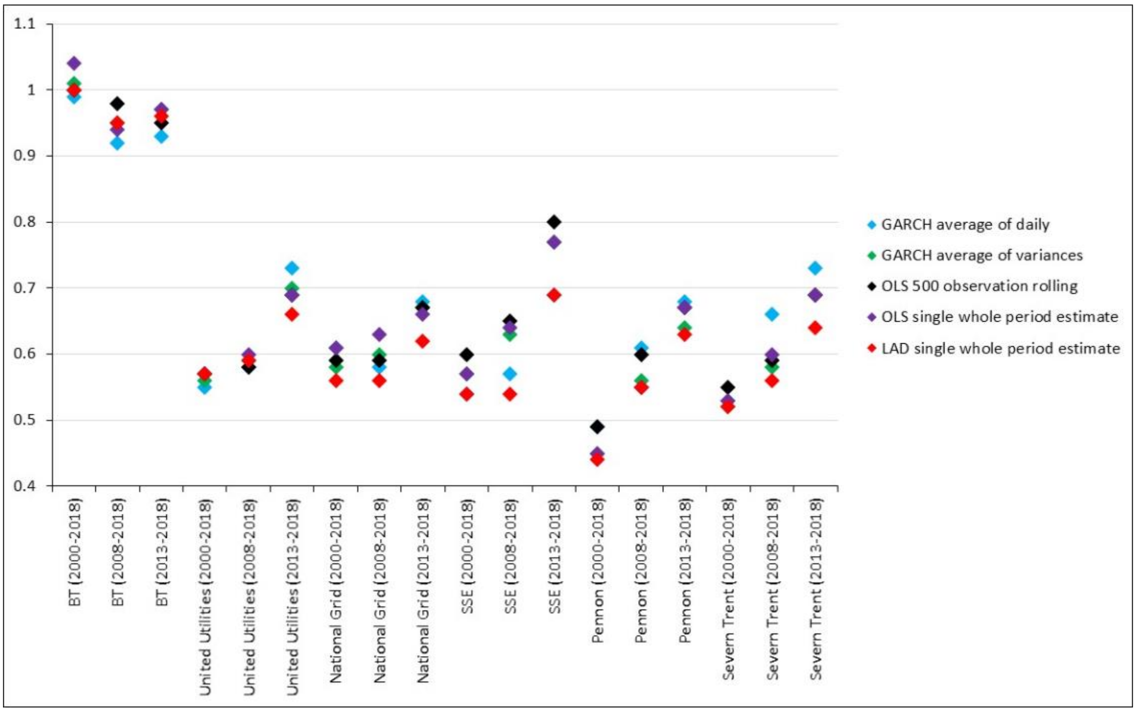


Figure E2: Equity β estimates

Note: in cases where some measures do not appear in the figure this is because the results are at a level where they appear the same as another measure. Full details of the results are given in Section 5 of the report.

The different estimation approaches give numbers that are not widely divergent.

In view of this, and assuming that there is no prior information with which to choose a preferred estimate of the equity β, our position is that OLS can continue to be used as part of the regulator's toolkit – provided that time window (to address the time varying nature of β) and appropriate corrections to standard errors (to adjust for ARCH) are considered. The reason for this recommendation is that OLS is well understood across a broad set of stakeholders, and in the analysis conducted in this report its estimates are not much different from those derived using more sophisticated approaches. This approach is consistent with other areas of regulatory analysis, where results based on the use of the OLS model are often used as a benchmark with which to compare other estimators.[4]

Given that this recommendation follows in part from the specifics of the study, we would advise further analysis to determine whether the range of estimates of equity β across different estimators is repeated using different datasets.

Based on these results, if we were estimating the equity β for energy companies at the present time, subject to further analysis and refinement, we would find a range of **0.55 to 0.70**. This excludes data on BT as we consider it to be significantly different from the utility

[4] This recommendation is consistent with Robertson (2018) who found that rolling OLS estimation can generate patterns very similar to those observed in real data, although it can also substantially overstate the true parameter. So the conditions in which OLS is being employed and the way it is used are important to understand.

companies and is also subject to the application of judgement as would be the case for other building blocks. Our judgement is explained further in Section 5 of the report.

If a narrower range is desired, we would opt for **0.57 to 0.65**, with **0.60** as a point estimate. The rationale for this includes the following.

- The 2008 to 2018 data window is the period after the last general structural break and pushes us towards 0.6.
- Evidence from 2013 to 2018 captures more recent market sentiment, which suggests a higher number, but as explained in section 2 of the main report, this was driven by a spike which is reversed after the election in 2017. This is reinforced by the lower LAD estimates, that place less weight on outliers.

The longest time period considered, 2000 to 2018, gives somewhat lower numbers but it has limited value given at least one major structural break in the relationships. A long-term view has some validity, however, and so our suggested range extends below 0.6.

1 Introduction

This independent study, commissioned by Ofgem while working with other UK regulators (via the UK Regulators Network (UKRN)), investigates issues linked to the measurement and estimation of beta (β) as part of the Capital Asset Pricing Model (CAPM) approach to estimating the cost of equity. It has been undertaken by Indepen in conjunction with a panel of academics, described further below. The work builds on previous work, especially that undertaken recently for UKRN and Ofgem by various academics/consultants. The role of β and the questions that arise with it are summarised below.

1.1 What is β and why is it important?

When estimating the allowed revenue for a regulated entity a key component is the required return on equity. In the UK this is normally estimated using CAPM which relates the return required for a particular stock to the risk-free rate and the relative exposure to market risk. Market risk is represented by the equity β which is multiplied by the market risk premium to give the company's equity risk premium.

The equity β has conventionally been estimated using the ordinary least squares estimator applied to the linear model

$$r_i = \alpha + \beta_i r_m + \varepsilon_i \qquad (1)$$

where β_i denotes the time invariant equity β and its estimate is given by

$$\beta_i = \frac{covar(r_i, r_m)}{var(r_m)} \qquad (2)$$

Note, if β is, in fact, time varying rather than constant – β_{it} instead of β_i – and we use the OLS estimator above then that results in a mis-specified equation and problems of interpretation. In a world where equity β varies over time, conventional estimation methods are not appropriate. For example, applying OLS estimation to (1) assumes a constant equity β estimate which is not appropriate given the time varying nature of the true β. Further, the heteroscedasticity in the residuals from OLS estimation then makes statistical testing of the parameter estimates unreliable.[5]

The importance of using an equity β with good properties follows directly from the capital-intensive nature of regulated utilities where small changes in the required return can have a

[5] The key report to Ofgem that we are reviewing and building on is that of Robertson 2018 – *Estimating β*. This report is being published alongside our report as part of Ofgem's consultation process. Industry was previously provided with a copy and it has been discussed at various meetings. We replicate Robertson's results as part of this study. There are also elements of the recent UKRN study that feed directly into our report. Robertson's paper provides a good overview of the issues linked to GARCH and technical aspects of the ways in which GARCH βs can be estimated that are not covered in this report.

significant impact on required revenues and consequently prices. For example, if the regulatory capital value of an industry was £50 billion, a 0.05 increase in the equity β could generate annual additional required revenues of £62.5 million. In this example, that is just under a 10% increase in the required return on equity.[6] In the low interest rate environment we have now, the importance of the assumed equity β is greater as the market risk premium accounts for a larger proportion of the total market return.[7]

If current βs are time varying but are estimated using an OLS estimator assuming a constant β then we have a mis-specified linear model. In this case, a critical concern is how can a more appropriate measure of β be estimated? If the stability issue is, in part, linked to the fact that equity β varies over timethis will have implications for the econometric modelling approach. One possible way in which we can test for time-varying βs is through whether the errors are heteroscedastic.

There are circumstances where it is necessary to estimate the underlying business risk, referred to as the asset β. The relationship between asset and equity βs is given in (3).

$$\beta_a = (g \times \beta_d) + \{(1 - g) \times \beta_e\} \qquad (3)$$

where g is the level of gearing in the company (proportion of debt in the capital structure) and β_a, β_e, and β_d are the asset, equity and debt βs of the company.

Our focus in this report is both estimating the equity β and examining several related issues, such as the decomposition of a given β estimate into estimates for the component businesses. In addition, we investigate whether there are alternative indirect approaches which do not involve the use of share price data, to estimate proxy β values. Throughout the main part of the report we focus on the validity of different approaches to estimation, returning in the conclusions to the practical implications of the findings of the report.

Regulators outside the UK, including the Australian Energy Regulator (AER), have investigated the issues we consider in this report and we draw on such results on some of the issues in this report.

Our analysis focuses on the six listed UK utility companies as at September 2018. Data were provided by Ofgem and cover the period from 1987 to 2018. For reasons that we explain below, most of the analysis we report covers the period from 2000 to 2018 (Robertson 2018 focused on the period 2000 to 2017).

[6] Assume that the capital base is 50/50 debt equity, a starting equity β of 0.6 and a market risk premium of 5%. Assume the risk-free rate is 0.5.

[7] This is true if movements in the risk-free rate are not fully reflected in movements in the total market return. Evidence suggests that this is the case – see CEPA (2013) or PwC (CAA publication Q6?).

1.2 The academic panel

The analysis and interpretation have been undertaken by a team from Indepen, with advice and guidance from a panel of academics. These are:

Professor Seth Armitage – Seth joined the University of Edinburgh as a lecturer in finance in 1989, having previously worked as a credit analyst and lending officer for two merchant banks. He has since taught a wide range of finance courses on undergraduate, MBA and MSc programmes, and for companies. In 2002 he moved to Heriot-Watt University, where he established an MSc in Finance and became Head of Department of Accounting and Finance. He rejoined Edinburgh University in 2007. His current teaching and research interests are mainly in corporate finance. He holds a BA in Politics, Philosophy and Economics (Oxford), an MPhil in Philosophy of Mind (St Andrews), and a PhD in Finance (Edinburgh). He is an associate editor of European Journal of Finance, and the author of *The Cost of Capital: Intermediate Theory* (Cambridge University Press, 2005). Seth's research and teaching are mainly in corporate finance, including seasoned equity offers, the cost of capital, leverage and dividend policy.

Professor Sudi Sudarsanam – Sudi is emeritus Professor of Finance at the Cranfield School of Management and was previously Professor of Finance and Accounting at Cass Business School, City University, London. He is currently Senior Research Adviser at the Mergers & Acquisitions Research Centre at Cass. Sudi was a member of the UK Competition Commission (CC) from 2005 to 2013 and sat in many inquiries into the cost of capital for regulated companies. He was a member of the CC's Expert Panel on Cost of Capital and the Finance & Regulation Group. Since leaving the CC, Sudi has been an expert advisor to firms and sectoral regulators including Ofgem, Ofcom, CAA and Ofwat through economic consultants. Sudi has published numerous research papers on corporate restructuring, mergers and acquisitions and their shareholder value effects based on cost of capital and asset pricing models. He is the author of the standard works in this area, *The Essence of Mergers and Acquisitions* (Prentice Hall, 1995) and *Creating Value from Mergers and Acquisitions: The Challenges* (FT Prentice Hall, 2003, 2010), translated into Chinese and other Asian and European languages.

Dr Melvyn Weeks – Melvyn is Assistant Professor in Economics, University of Cambridge and Fellow of Clare College. He has been an Indepen Associate for over 15 years. Melvyn's work spans theoretical and applied micro-econometrics including: understanding behaviour in discrete choice; modelling demand systems; revealed and stated preference models; model testing and evaluation; and computationally intensive methods including machine learning for predictive and causal inference, simulation-based inference and the bootstrap. His research interests are the development and application of models of choice: parametric and nonparametric methods to represent flexible demand behaviour; and estimation of willingness-to-pay, with application to utility markets. Melvyn has published in highly respected journals including the Journal of the American Statistical Association, The Economic Journal and the Journal of Applied Econometrics.

The academic panel was involved in the design of the study, the ongoing review of outputs during the project and checking specific questions/preparing aspects of the report.

Numerous meetings with Ofgem staff have helped with the development of the report and Dr Robertson kindly met with the team to discuss his and our findings.

1.3 Selection criteria

To be able to choose between different estimation approaches and data definitions, there is a need for criteria against which alternatives can be assessed. We agreed the following criteria with Ofgem at the beginning of the project.

- Appropriateness – do the characteristics of the data as established through the use of exploratory methods and econometric testing match the assumptions underlying an approach for estimating β?
- Feasibility of implementation – is an approach feasible to implement and can it be replicated by others?
- Regulatory policy implications – does the approach meet Ofgem's requirements or raise broader implications for regulatory determinations?

1.4 Structure of the report

The remainder of this report is structured as follows.

- Section 2 considers the strengths and weaknesses of options for the direct measurement of an equity β.
- Section 3 investigates alternative ways in which β can be estimated.
- Section 4 examines the relationship between equity and asset βs and how asset βs should be estimated for regulated businesses.
- Section 5 concludes and provides recommendations to Ofgem and possible areas for further research.

Appendices and supporting Annexes provide further information.

2 Direct measurement approaches

In this section we present evidence and draw conclusions about the data and modelling issues relevant to the direct estimation of equity βs.

A key issue to note is that there are few listed utility network companies in the UK. At the peak of privatisation in the early 1990s there were some 30 such companies.[8] Currently there are six and few of these focus on single activities. The small number of listed companies and the fact that most of them represent portfolios of activities poses problems for regulators.

In response to the small number problem, options include making do with the small number of equity βs (the AER's solution in Australia) and using international comparators (suggested by several of the UK regulated energy companies). Decomposing equity βs into their portfolio elements might give some more accurate estimates of 'pure-play' βs. Another possibility altogether is to consider approaches to measuring βs that do not rely on share price information. These are discussed in later sections of the report.[9]

2.1 Structural breaks

When considering the length of the estimation window, an important factor is the existence of structural breaks. Reasons why a time series (or a relationship between time series) may include structural breaks include the following

- Changes in the regulatory regime affecting risk
- Political and other external events, such as financial crises, affecting utility companies differently from the effect on other companies
- Business changes whereby divestments, mergers, acquisitions or the organic development of new business lines significantly change the risk of the company
- Changes to the constituents of the market index, such as the inclusion of new technology companies, which change the overall market risk

Over the period since the privatisations of the late 1980s and early 1990s factors that might have introduced structural breaks include the following

- Regulation – e.g. changes in the regimes: creation of Ofgem, RPI-X@20/RIIO, Future Price Limits in the water industry

[8] There were 10 water and sewerage companies, 15 electricity distribution companies (12 in England & Wales, two in Scotland and one in Northern Ireland), British Gas and British Telecom. In addition there were several larger water-only companies listed on the stock market and other network companies in transport like BAA and Railtrack (from 1996). Note, National Grid was initially owned jointly by the English & Welsh electricity distribution companies and only spun out and listed on the stock market in 1995.

[9] We do not consider the option of requiring companies owning regulated network companies to keep them listed. This would reflect a significant shift in policy and would not necessarily avoid other aspects of the data and measurement problems.

- Political and other events – e.g. leaving the ERM, re-evaluation of utility risk post-Enron, the dot.com boom and bust, the East Asian financial crisis, the Global Financial Crisis (GFC), Brexit, renationalisation policy of the Labour Party

If structural breaks change the relative risk relationship between the company and market it will be important to know whether the effect was a significant effect and whether it was permanent or not. The answers to these questions will affect the time period that can be used to derive consistent estimates of equity βs.

For example, consider a change in circumstances that causes a flight from equity. In the short-term there could be an increase in the equity β as the flight causes a higher correlation between market and company returns that results from the common reaction by investors rather than a change in the underlying relative risk relationship. Once the crisis is over, if the underlying relative risk relationship reasserts itself the equity β would return to its original value. In this case we would use observations from before and after the crisis, and investigate whether it is appropriate to include data from the period of the crisis.[10] On the other hand if the shift was permanent then using observations from before the break will bias the result for current application.

To understand our data a series of structural break tests were applied to an equity β estimation model as given by (1). We utilised tests of residual stability (cumulative sums – CUSUM – of overall and recursive OLS residuals) and parameter stability (recursive and moving window estimation of parameters) along with the Chow F test for stability. In the case of the residual and parameter tests, we look for significant shifts away from zero mean and stability while the F test compares the residual sums of squares from a single model with that from two separate models, one pre-break and one post. The details of the tests are provided in Appendix A to this report.

The majority of these tests demonstrate the presence of structural breaks in the full dataset (1987-2018) as well as the post-2000 dataset. Table 2.1 provides estimates of the date when the post-2000 breaks occur.[11] The number and dates of breaks were not common across companies, which suggests that some are business specific. There appears to be a significant break in September/October 2008.

[10] Similar arguments can be raised about political effects. If elections do not happen at regular intervals (as was the case pre the coalition Government in the UK) but need to happen no later than a set time period, then election periods could have an exaggerated effect, especially if the regulated network companies are an issue during the election period. See *Quantifying Political Risk* in *Utility Finance* April 1994 (OXERA).

[11] This is based on a consideration of the results and figures in Appendix A and its annexes.

Table 2.1 - Summary of findings on structural breaks

Stock	*CUSUM residuals*	*Recursive residuals*	*Recursive parameters*	*Moving parameters*	*Chow (1987-2018)*	*Chow (2000+)*	*Years of likely post-2000 breaks*
BT	-	-	Break	Break	Break	Break	01 & 03/4
NG	-	Break	Break	Break	Break	Break	00, 01, 08 & 09
PNN	-	Break	Break	Break	Break	Break	04/5 & 08
SSE	-	-	Break	Break	Break	Break	01, 02, 03, 04/5, 08 & 13/15
SVT	-	-	Break	Break	Break	Break	02/3, 04/5, 08 & 12/13
UU	-	Break	Break	Break	Break	Break	02, 03, 04/6, 08 & 13/15

As discussed in Partington & Satchell (2017), the identification of structural breaks through statistical analysis is a starting point, but it is not the whole story. There are also issues related to whether there is ex ante clarity around the date of a potential break. In addition, priors about the potential causes of any break are important when deciding how to address the implications of the statistical results. As the authors note, the absence of an obvious cause does not imply we can ignore evidence of a break. While we may choose to continue to use a longer time series, including the period before and after the break, we need to do that in the knowledge that there is a break in the data.

Likely causes of structural breaks include regulatory changes and corporate transactions. Significant changes in regulatory regime, like the shift from RPI-X to RIIO in the energy sector or the implementation of the Future Price Limits changes at PR14 in the water sector, suggest that the assumption of a constant equity β is likely to be untenable. National Grid's merger with Lattice Group (the gas transmission and distribution network formerly owned by British Gas) in 2002 and subsequent divestment of its gas distribution businesses in two phases in 2005 and 2017 significantly altered the make-up of the company, as did its acquisitions in the US in 2002, 2002 and 2007. While such transactions did not affect the risk of the individual businesses, they may have altered the observed equity β of National Grid Group and would need to considered when estimating and using National Grid's β.

We conclude that using the period since the most recent structural break is appropriate unless it can be shown that any such break did not significantly alter the underlying relative risk relationship captured by β. This does not mean earlier data should be discarded, but the weight placed on it should be considered carefully.

If the break is recent, occurring within the last five years, further consideration is needed as to how to obtain a sufficiently long time series for estimating equity β. We return to the question of the choice of period at the end of this section.

2.2 Frequency of observations

When thinking about data frequency, it is important to consider the relationship between the length of the estimation window and the frequency of observations. For example, although daily returns provide more information and a larger number of observations, evidence suggests that this comes with the complication that the estimates vary significantly over time. Less frequent measurement of returns, say weekly or monthly, may overcome this problem.

In addition, the estimate of the equity β may depend on the choice of when in thr week or month the returns are measured. The evidence presented in Table 2.2, provided by British Gas in a 1996 submission to Ofgas, illustrates the well-known problem that estimates of equity β depend on the day of the week used to estimate it. As set out in Appendix B, this issue is encountered with the six utility companies in this report. Consequently, if weekly or monthly returns are employed, the issue of variability arising from day of the week/month should be considered.

Table 2.2: Weekly estimates of β by day of the week

	Monday	Tuesday	Wednesday	Thursday	Friday
Equity β	0.97	0.93	0.86	0.88	1.03

Note: based on information provided in table 3.7 of Appendix 2 of British Gas TransCo's *A framework for Efficient and Effective Regulation: TransCo's Submission to the Ofgas Review of Price Controls*, March 1996

The choice between daily, weekly and monthly data frequency involves a balance between

- the need for a sufficient number of observations so that inference on the estimate is possible. This depends on the length of the dataset as well as the frequency of observations. For example, to get the same number of observations for weekly data requires a dataset that is five times longer in calendar time than daily data), and
- whether greater frequency of observations breaches the statistical assumptions underlying OLS calculations, in particular about the homoscedasticity of error terms

With a five-year data window or more, all three frequency options are feasible and it is necessary to test whether returns demonstrate ARCH (autoregressive conditional heteroscedasticity) behaviour or follow some other form of heteroscedasticity. The choice of which test is most appropriate is explained in the text box below.

Three common tests for heteroscedasticity are f Breusch & Pagan (1979), Engle (1982) and White (1980). Each tests for a different form of heteroscedasticity.
The Breusch-Pagan test regresses squared residuals on the levels of the independent variables and so tests for a form of heteroscedasticity that is linear in the variables. The White test expands the Breusch-Pagan to include nonlinearity: squared residuals are regressed on levels, squares and cross-products of the independent variables. The Engle test regresses the squared residuals on lagged values of the squared residuals – it is testing for autoregressive conditional heteroscedasticity (ARCH).
The issue being addressed with respect to financial returns data is the presence of ARCH, the Engle test is the appropriate heteroscedasticity test..

Box: Testing for heteroscedasticity

Table 2.3 - Results of Engle test for heteroscedasticity (2000-2018)

Stock	*Monthly (2 lags, first trading day basis)*	*Weekly: Mon (8 lags)*	*Weekly: Tue (8 lags)*	*Weekly: Wed (8 lags)*	*Weekly: Thu (8 lags)*	*Weekly: Fri (8 lags)*	*Daily (40 lags)*
BT	0.007	0.000	0.000	0.003	0.000	0.000	0.000
NG	0.862	0.000	0.000	0.000	0.000	0.000	0.000
UU	0.831	0.000	0.000	0.000	0.000	0.000	0.000
SSE	0.000	0.000	0.000	0.000	0.000	0.000	0.000
PNN	0.004	0.000	0.000	0.000	0.000	0.000	0.000
SVT	0.166	0.000	0.000	0.000	0.000	0.000	0.000

Note: p-values – H_0 of homoscedasticity

Given that we are dealing with financial returns data, ARCH or GARCH behaviour in the residuals is commonplace. This is best tested for using the Engle test (an auxiliary regression of the squared OLS residuals on a number of their own lagged values). Table 2.3 provides the results the Engle test for daily, weekly and monthly data for the six network companies. For completeness further tests for all data frequencies are reported in Appendix B.

Our results demonstrate strong evidence for heteroscedasticity/ARCH processes. Note that a low probability (p) value indicates that the probability of the event under the null, here homoscedasticity, is low, i.e. heteroscedasticity exists. A higher value implies it is more likely that the null hypothesis, homoscedasticity, is not rejected. Consequently, values below the cut-off point – say 5 or 10% probability (0.05 or 0.10 in the table) – lead to the rejection of the null hypothesis and consequently an expectation that conditional heteroscedasticity exists. The cases where the null hypothesis is not rejected are coloured green in the table.

Heteroscedasticity (ARCH form) is, as expected, universal at the daily level.[12] Even moving to weekly returns – using one model for each trading day, so the Monday model uses Monday-to-Monday returns, we see that ARCH remains universal. At the monthly level the averaging effects of aggregation serve to remove the ARCH effect for three of the companies: NG, UU and SVT. However, for BT, PNN and SSE even the monthly returns exhibit autoregressive conditional heteroscedasticity. This has implications for the approach to estimating equity βs and, in particular, means the use of OLS may not be appropriate.

Based on the evidence presented above, OLS should only be used for three of the companies and then only with monthly returns data. Any estimation of β using daily or weekly returns for any of our sample of companies will breach the homoscedasticity assumption underlying the BLUE-ness of OLS. Even monthly returns for the other three companies breach the homoscedasticity assumption. The implications of violating this assumption and that of time varying βs are discussed further below.

[12] In this section we only report heteroscedasticity across the 2000-2018 period. Tests of shorter periods, depending on the choice of estimation window should also be undertaken. Annex C4 of Appendix C provides some examples of heteroscedasticity tests over shorter time periods. The general existence of heteroscedasticity continues to be a problem.

2.3 Estimation approach

Traditionally estimation of the equity βs in equation (1) has been undertaken using OLS. Issues arise with respect to this approach (including the day of the week/month issue noted above).

First, the existence of autoregressive conditional heteroscedasticity in the returns data occurs along with time varying βs which are best modelled through non-linear models such as GARCH rather than the linear OLS. Further, in a world where the OLS model is used, the existence of heteroscedastic errors biases the estimate of the standard error, making significance testing difficult being problematic for generating a confidence interval.[13] Heteroscedasticity is common with high frequency financial data such as daily returns – see for example Armitage and Brzeszczynski (2011) and Robertson (2018).

Second, OLS is considered by some commentators, such as Henry (2008), to be unduly influenced by outliers as they will attract disproportionate weight.[14] Henry recommended use of the Least Absolute Deviation (LAD) estimator rather the minimisation of the sum of squared residuals. The Australian Energy Regulator (AER) has followed Henry's recommendation in its price determinations.

The choice of estimation approach depends on the characteristics of the data, or more precisely, the data generation process that we are modelling.[15] If hetero-scedasticity/ARCH is not an issue then it is possible to use OLS and there is an argument that LAD should be employed given the problem of outliers.[16] Appendix D provides LAD estimates that can be compared with our standard OLS estimates. If heteroscedasticity/ARCH is an issue – as it is always at weekly and daily frequencies – then OLS significance tests may be biased and consequently an alternative approach is needed.

Armitage and Brzeszczynski (2011) discuss

- GARCH models – the preferred solution proposed by Robertson (2018) and UKRN (2018)
- Kalman filters – suggested in NERA (2018), OXERA (2018b) and used in previous research such as Buckland (2001)
- other approaches such as Blume and Vasicek adjustments to OLS estimates[17]

---

[13] If OLS is used as a source of parameter estimates then any tests and confidence interval need to be undertaken using the estimates corrected for the appropriate form of heteroscedasticity. We return to the question of continued use of OLS in Section 5 of the report.

[14] This follows given that the OLS estimation criteria is based on a quadratic loss function such that the square of an outlier error will exert a significant effect on the location of the least squares regression line. LAD minimizes the sum of absolute residuals/errors.

[15] Various options exist and choosing between them becomes an issue of statistical methodology.

[16] This approach is employed by AER in Australia as a cross-check for OLS estimates. See for example section G1 of AER 2017

[17] We do not consider these approaches in this report. This type of Bayesian adjustment is one that we do not think appropriate when considering regulated infrastructure pricing determinations. The Blume adjustment assumes that the value of the equity β should be 1. While that is true across the average of all listed companies, it is not appropriate for a single company or sector. The Vasicek adjustment is more defensible inasmuch as the prior value that is weighted against the observed value does not

We focus on GARCH models in this report as they provide the most flexible response to the underlying problem of time varying βs and seek to best capture the characteristics of the underlying data.[18] There are different specifications of the GARCH model, with their use being driven by the characteristics of the data as well as assumptions about the underlying stochastic relationships. Robertson (2018) used one specific version of GARCH(1,1), namely diagonal BEKK (D-BEKK). We have replicated his results and updated them for the latest year's data.

As noted in Robertson's report, one particular GARCH form may not be appropriate for all the network companies.[19,20]

Using the 2000 to 2018 dataset and using daily returns, and using the Bayesian Information criterion (BIC). to choose between models, we find that no one GARCH specification is preferred for all companies. Our results are summarised in Table 2.4.

Table 2.4 - Summary of preferred models for order (1,1) 2000-2018

Model	***Number of lowest BIC cases***	***Companies***
T-BEKK (Triangular BEKK)	2	UU and SSE
Full VECH (Half-vectorisation)	4	BT, NG, SVT and PNN

Using daily data, the moving averages of the GARCH and rolling OLS equity βs, are illustrated in the following graphs. We have focused on a two-year rolling window for this illustration.[21] Annex A2 of Appendix A investigates drivers of the changes in the β values including the variance of the market and the covariance between the stocks and the market. We return to some of that evidence in Section 5.

have to be 1. However, if there is a prior expectation of the value that should be stated transparently, justified and taken as a piece of evidence to be incorporated into any decision.

[18] LAD might also be a but it does not address the finding that β is, fundamentally, time varying and, apparently, non-stationary: it is discussed in Appendix D.

[19] This is explained in detail in Appendices B and C. In addition we considered whether the choice of the order of GARCH mattered. This is explained in Annex C1 of Appendix C. While we found two companies ought to follow order (2,2) concerns about the statistical validity of the preferred specifications means that we have focused only on order (1,1) in the main report.

[20] We have also tested for stability in the choice of GARCH specification across different time windows. Later in this report we consider two sub-periods, 2008-18 and 2013-18. For each of these sub-periods we determined what was the appropriate model specification and the results arising from that choice. As explained in Annex C3 of Appendix C, there was some stability in the model specification choices but some of the differences arise from the data requirements for the different specifications. As discussed elsewhere, the specification choice has, in most cases, an insignificant impact on the estimate.

[21] Two year rolling daily β calculations are commonly used as a basis for estimating β.

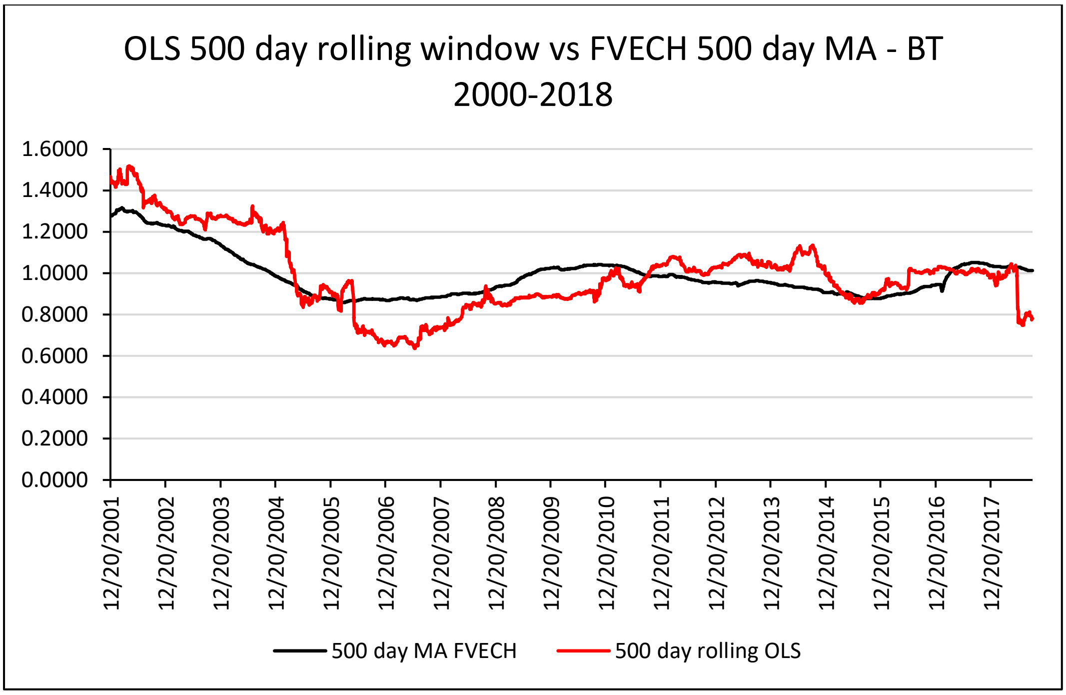


Figure 2.1: Comparison of equity β estimates for BT

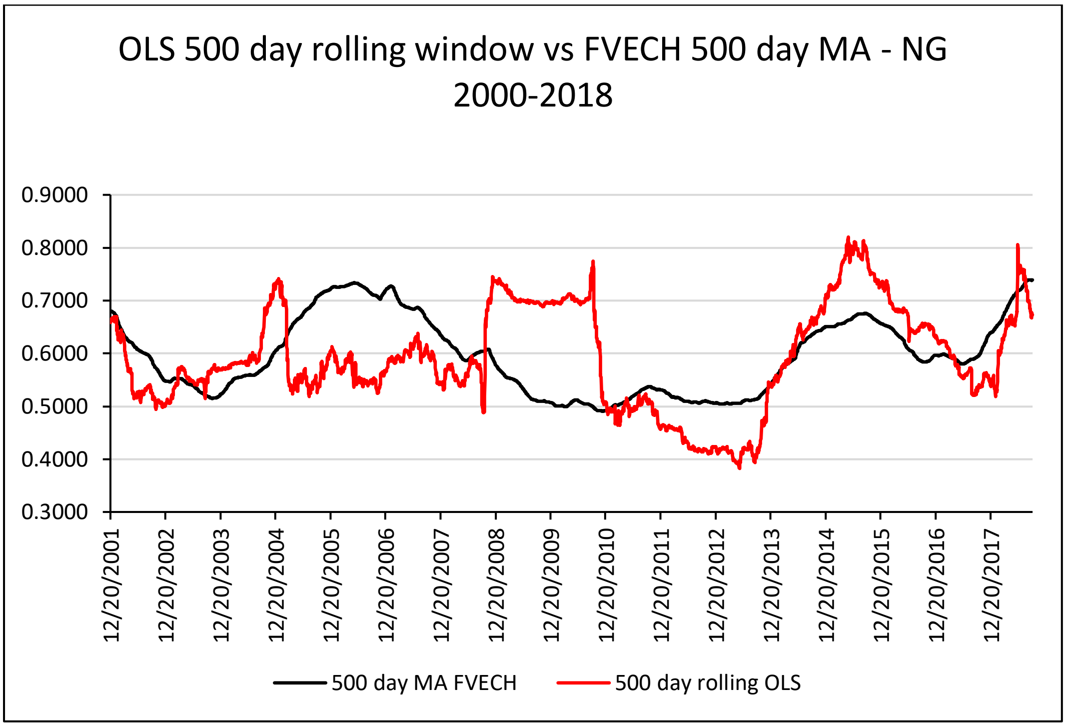


Figure 2.2: Comparison of equity β estimates for National Grid

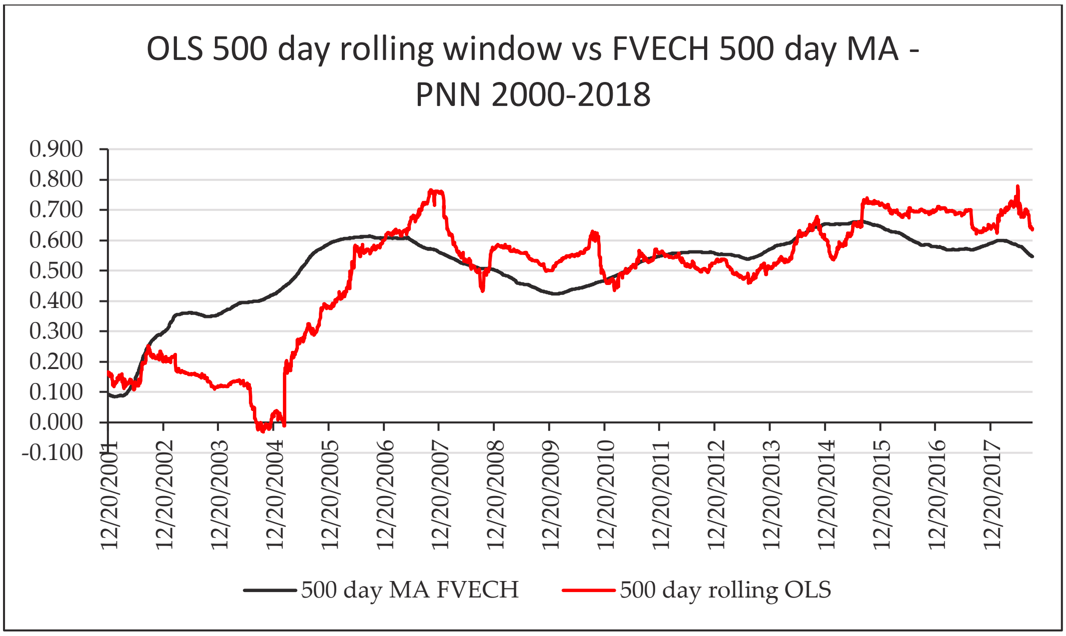


Figure 2.3: Comparison of equity β estimates for Pennon

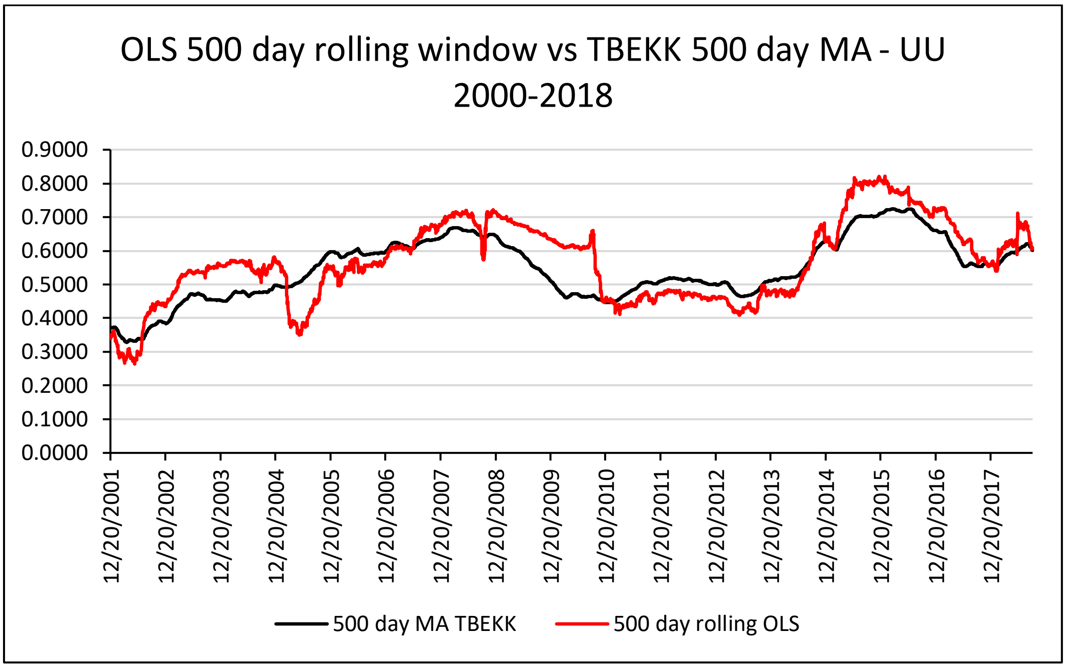


Figure 2.4: Comparison of equity β estimates for United Utilities

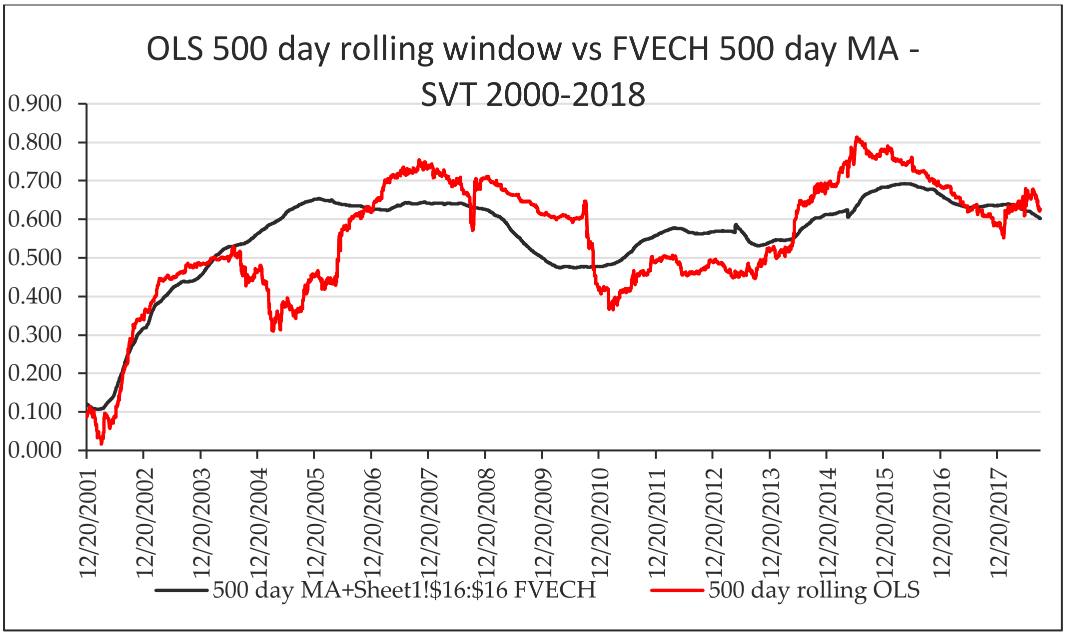


Figure 2.5: Comparison of equity β estimates for Severn Trent

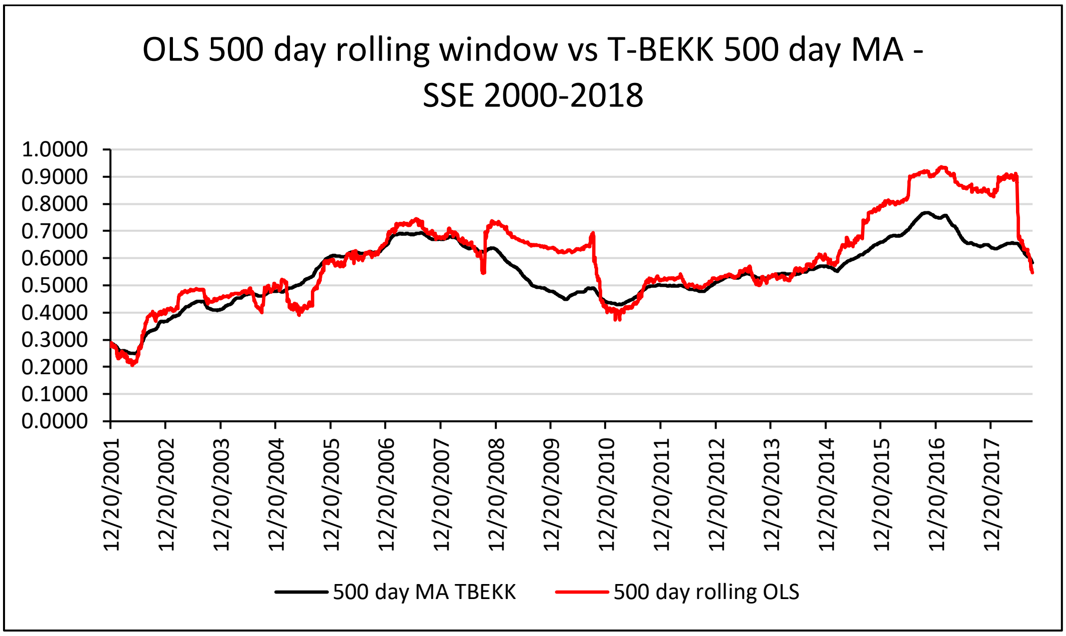


Figure 2.6: Comparison of equity β estimates for SSE

As can be seen, the moving average of the βs from the GARCH model in each case tracks the rolling OLS estimate but with less volatility. These results suggest the following.

- The volatility in the GARCH estimates means that focusing on short-term/recent estimates of the equity β could lead to bias and that estimates using a longer time period are better.

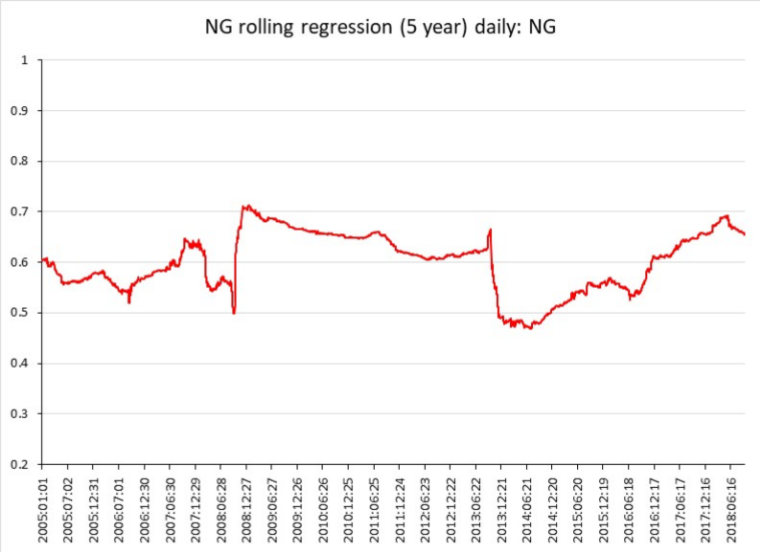
- The long-term pattern of GARCH estimates is such that an investigation of the use of ARIMA may create a more stable estimate for the five- to seven-year period that Ofgem requires (this is considered further in Section 2.4 below).
- Given the volatility in both measures, seeking to understand what drives the changes in the estimates is helpful. For example, was the recent increase due to low volatility in the market and unchanged volatility in the utility stocks? This would allow us to draw conclusions about whether changes are short-term after which mean reversion occurs or deeper structural changes in the relative risk relationship.[22]

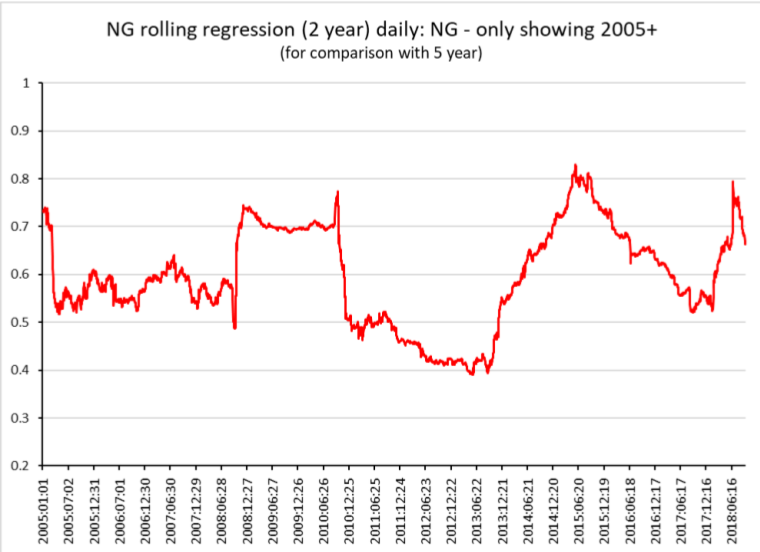
Also it is important to consider the impact of different timeframes in order to understand what is driving the changes in β in OLS. This is illustrated with data for National Grid in figure 2.7 below, which show the estimated equity βby rolling OLS in various ways

- in the top box over a five-year window using daily data since 2000
- in the middle box over a two-year window using daily data since 2000
- in the bottom box over a five-year window using monthly data since 2000

As can be seen from the middle box, the equity β peaked around 2015-2016 and since then has fallen (we discuss this peak in Section 5 and Annex A2 of Appendix A). That fall has not yet been reflected in the five-year rolling monthly β. We consider in Section 5 which of these is the value on which a regulator should base a forward-looking allowed cost of equity.

[22] In part our understanding of this will be driven by the structural change analysis already undertaken.





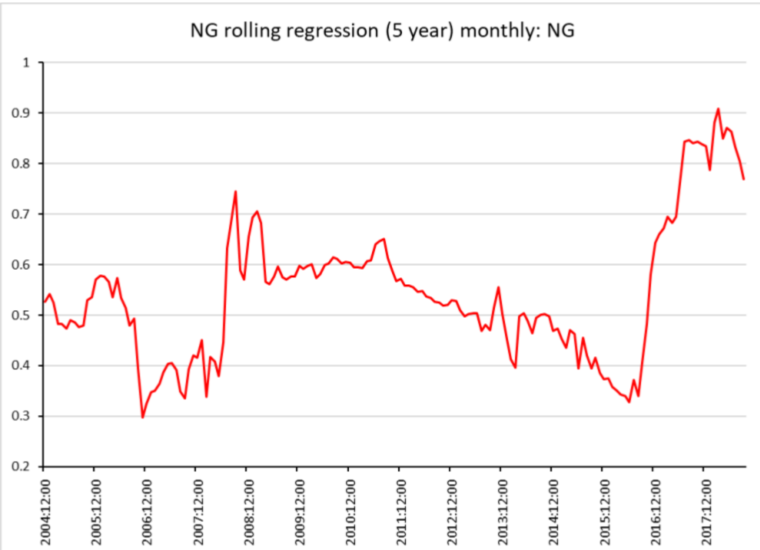


Figure 2.7: Comparison of equity β estimates using different windows for National Grid

2.4 Forecasting β

How might we use our models grounded in historical data to think about future values of β? We considered whether the best forecast of the β at the time of a price determination is given by ARIMA modelling applied to the daily GARCH estimates of β. Appendix H discusses the approach adopted to estimating ARIMA – using the daily β estimates generated by the GARCH(1,1) specifications as set out earlier in this Section.

Table 2.5 shows the two-year estimated mean β value as well as the average for the period from which the ARIMA is generated (2000-2018).

Table 2.5 Mean β sample and forecast

Series	*Pre-forecast mean*	*Forecast mean*
BT Full VECH(1,1)	0.99	0.94
UU Triangular BEKK(1,1)	0.55	0.25
SSE Triangular BEKK(1,1)	0.54	0.17
SVT Full VECH(1,1)	0.53	0.50
NG Full VECH(1,1)	0.61	0.65
PNN Full VECH(1,1)	0.49	0.35

The first column of β values shows the average value for the period from which observations were used to estimate the ARIMA specification (GARCH daily β values from 2000 to 2018). The second column of β values, the forecast mean, provides the average β value over the forecast period – in this case 500 daily β values were forecast, giving an estimate over the next two years. As shown in Appendix H, the forecasts quickly converge on a single value and any longer-term forecast would remain at that value.

In most cases in which the GARCH specification is VECH we find a forecast value close to the pre-forecast mean and this apparent stability leads us to believe that the ARIMA model provides a good estimate. For the T-BEKK specifications the results are significantly different, and this requires further analysis before the reliability of using the ARIMA estimate for regulatory purposes can be assessed.

A useful piece of information for a regulator seeking to establish a forecast value for a future price control period is the ARIMA model estimated using the daily GARCH β estimates.

2.5 Summary of findings and conclusions

Consideration of the evidence and results derived from the six listed UK network companies leads to the following findings.

- There have been structural breaks in the CAPM β relationship for most companies since 2000.

- Autoregressive condition heteroscedasticity makes OLS unsuitable when using daily or weekly data for all the companies and remains a a problem for half of them even when considering monthly data.
- GARCH models can be estimated using daily returns data but there is no one preferred model appropriate for all six companies.[23]
- The averaged GARCH estimates calculated are similar to the rolling OLS daily results for the same time periods, with the OLS estimates showing more volatility.
- For OLS estimates calculated over different time windows, a relatively small number of observations can influence results for a significant period, especially when the window is quite long – say five years. This means it is important to take a longer view, such as consideration of rolling estimates, and to use high frequency data and longer windows, to ensure that underlying changes are captured rather than noise.
- It is possible to generate ARIMA forecasts for the daily β estimates from the GARCH models which should provide a useful piece of information for any regulator having to estimate a β value for a future price control period.[24]

The evidence suggests that the process for estimating statistically stable equity βs is problematic and may mean different models for different companies at the time of a price control. Further, the process would need to be undertaken for each price control and could lead to different models being used each time. From a regulatory policy perspective this may be undesirable.

There appear to be two options available.

- Acknowledge the weakness in the current OLS approach and collect a broad range of information – such as calculations since 2000, or from the end of the GFC, over different time windows and frequencies of returns – and use regulatory judgement to draw a conclusion on the appropriate estimate of the equity β.[25] Or
- Develop a decision tree that uses statistical testing to narrow the set of estimation approaches, requiring clear criteria to generate a narrower range from which the equity β is chosen.

An example of the type of decision tree that a regulator could use is given below.

[23] While this is not a problem from a statistical perspective it is not a desirable outcome from both a replicability and policy perspective.

[24] Other approaches may also exist.

[25] While OLS daily estimates from 2000 (or 2008), either as a single long-term value or in some form of rolling estimate should be included in the information set collected, by themselves they are not sufficient. The range of information collected should include GARCH specifications and possibly LAD as a cross-check against the impact of outliers on the parameter estimates.

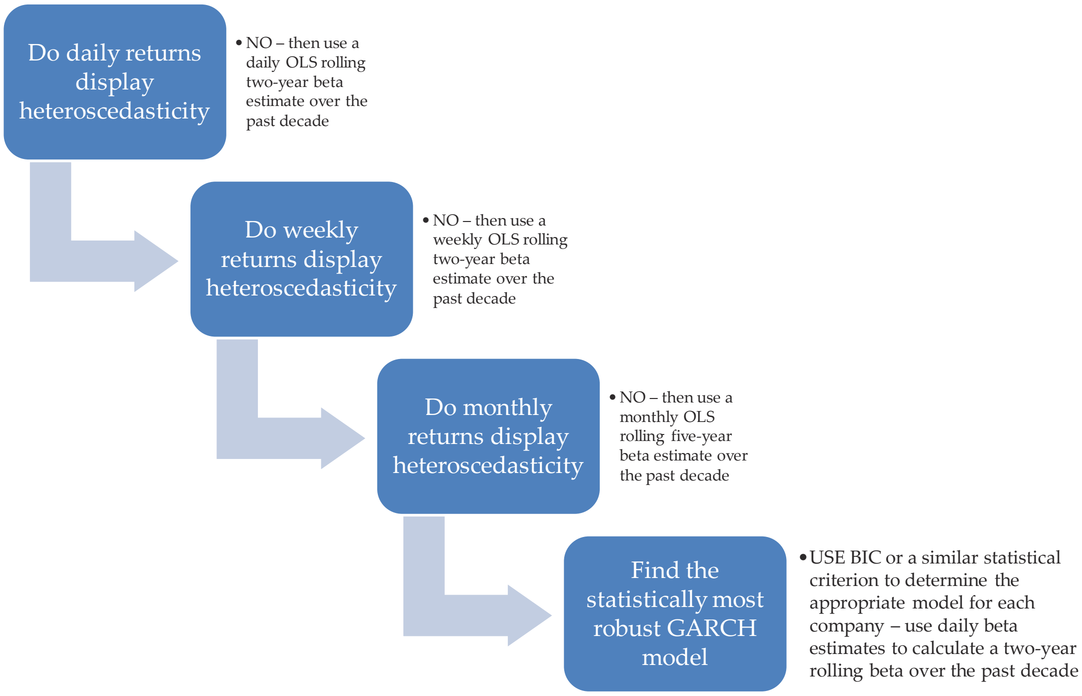


Figure 2.8: Decision tree for estimating equity β

In all cases, a look back over at least five and probably ten years is desirable and any programme should commence with an evaluation of structural breaks. Where there are significant breaks these should be investigated and their implications for the choice of β explored and made clear.

Whether the monthly option should be included depends on the specifics of the company under consideration. Daily and weekly returns data are preferable given the information content and the beneficial effect of the greater number of observations. Combined with the fact that structural breaks seem relatively common, at least for some companies, this means that the use of monthly information is likely to be problematic. We have included the option in the diagram as it may be appropriate in some cases and regulators should consider whether useful information can be generated.

3 Indirect measurement approaches

As seen in Section 2, direct measurement poses problems even when there is data from which estimates can be derived. Are there indirect measurement methods that can be used to derive equity or asset β estimates?

Two possible sources of information are considered in this section

- accounting βs, and
- other risk measures

In both cases the information can be applied more widely than just to listed companies, namely to unlisted companies and individual businesses within larger groups.

While not exactly an indirect approach we also discuss international comparators at the end of this section.

3.1 Accounting βs

When companies are not listed, or are businesses within a larger listed group, there is no stock market share price information on which a β can be estimated. One solution proposed to this is to consider accounting data that captures risk and to use these as explanatory variables for a β estimate.

Work in the 1970s identified four key explanatory accounting determinants of β.[26] They are

- earnings cyclicality – β depends on the relationship between swings in the firm's earnings and swings in the economy generally
- earnings variability – β is strongly related to the volatility of earnings
- financial leverage – β is highly related to financial risk
- growth – β is positively related to growth, given the traditional association between rapid growth and high business risk

It is possible to estimate an equation linking the accounting factors of listed firms to their estimated β values, along the lines of:

$$\beta = a_0 + a_1X_1 + a_2X_2 + \cdots . + a_nX_n + \varepsilon \qquad (7)$$

Where X_i represents the accounting factors. This equation is estimated using a cross-section of companies so that a statistical relationship between the accounting factors and β can be established.

[26] These factors are reported in Morin (2006), drawing on previous academic work.

For non-listed entities it is possible to estimate a β value using the relevant accounting measures in equation (7).

Morin (2006) quotes an example of work in the late 1970s to estimate β values for unlisted oil pipeline businesses. An asset β estimation model was derived using the standard deviation of book rates of return for 18 industries and then applying the standard deviation of book rates of return for oil pipelines to estimate a β for them.

In a similar vein, Morin (2006) discusses an approach where an Earnings β is estimated. This relates the earnings of a business to the aggregate earnings across the economy. Regressing these figures – say quarterly over a decade – generates an earnings β. As with the oil pipeline example, it is possible to consider the relationship of the earnings β with estimated β values and use this relationship to generate estimates of β for unlisted companies. The following example is provided as example 7-3 in Morin's book:

$$Stock\ Market\ \beta = 0.564 + 0.251\ Earnings\ \beta$$

Other academic evidence, such as Laveren (1997) and Toms (2005), is that while some variables, such as operating leverage and financial leverage, can be important determinants of risk, the overall explanatory power of such models is low. Further, explanatory variables do not necessarily have the expected sign calling into question the use of the variable and possibly the model.

To illustrate how this could work with current information we investigated the relationship between the estimated βs for the six companies and a series of accounting measures using a panel dataset over a nine-year period. While not exhaustive and potentially in need of further separate research, we found no strong relationship and unexpected results, such as financial gearing having a negative effect on the observed β value.

Given the academic results, limited apparent practical use and the issues, such as expected signs not being met, we believe that this is an area which requires more research before it could be an element in any β estimation process for a regulatory determination.

3.2 Broader measures of risk

Another possible source of information about risk relates to how other financial instruments are priced.[27] While this may not provide a direct estimate of the risk for a company it can provide indicators of

- the market's overall view of the riskiness of the company
- lower limits for the equity premium (the equity β multiplied by the market risk premium)

This latter point is useful only if we believe it possible to determine an asset β (discussed in more detail in Section 4). It should be the case that, when properly measured, the expected

[27] OXERA (2018a) section 5.1 discusses this in some detail.

return on debt is never larger than the equity premium – given the risks attached to the two types of investment. So, a clear lower bound for the equity premium is the expected return on debt – this means it is possible to infer a lower bound to the equity β. Note, it is the expected return on debt not the yield to maturity (YTM)/debt premium that should be used here. The YTM is the return on debt plus default risk (this is discussed in more detail in Section 4 when we are discussing debt β values). In this example we use the YTM for a bond but it should be remembered that this then will be an over-estimate of the expected return on debt and consequently an over-estimate of the debt β.

The calculation is for an unlevered equity β which is also the asset β. As such it is important to remember that the debt premium should be based on an appropriate bond. Companies may have debt premia above 150 basis points but that is in part because of the high gearing they have – reflected in part in the level of investment grade rating they have. An ungeared company would have a better credit rating and ought to face a lower debt premium. While the rating may not be AAA, where premia can be in the range of 0 to 50 basis points, it should be lower than the observed values. In addition, as noted above, this number includes default risk (which while low, is still positive) and so is a slight over-statement of the expected return on debt.

Of course, given the risk that equity faces relative to debt, there would need to be a premium for the equity β. Exactly what level of premium is required is not clear and is worth further separate investigation.

Taken together this means that the lower bound for the asset β is given by the following equation:

$$\beta_a^{Lower} = \frac{(Debt\ YTM + Premium)}{MRP} \qquad (4)$$

Suppose the YTM is 100 basis points, an assumed premium of between 25 and 50 basis points and the market risk premium is 500 basis points. This would generate a lower bound for the asset β of between 0.25 and 0.3.

This would then need to be re-geared to get an equity β value at the required gearing level – exactly what we mean by gearing in this instance is investigated further in section 4 and all we do here is provide an illustration of the way the calculation works.

If the required gearing level was 50% and the debt β was 0.1, then the equity β would be between 0.4 and 0.5 and if 60% then the range would be 0.475 and 0.6.[28]

Does this approach provide us with anything useful? In principle yes, if we can overcome all the measurement issues linked with deriving the correct expected return on debt and the premium for equity relative to debt as well as solving the methodological issues discussed in Section 4 about de- and re-gearing β values. However, it is unlikely that all those issues can be adequately addressed and so, while it is possible to make assumptions about the various points needed to allow the calculation, the result will be as much a product of the

[28] If a zero debt β was assumed the values would be 0.5 and 0.6 at 50% gearing and 0.625 and 0.75 at 60% gearing.

assumptions as a true value and, as such, will need to be weighted accordingly when a judgement is being taken about the appropriate range for an equity β value.

3.3 International comparators

While not necessarily an indirect approach, the use of international comparators is something that is often proposed, either to supplement existing national data or because there is no national data available (say for airports).

Using international comparators is something that respondents to the earlier Ofgem RIIO-2 consultations have raised. In principle the same questions and processes outlined in Section 2 above could be followed to estimate β values – we do this for a set of European and American utilities in Appendix G. Note we do not search for structural breaks in these datasets. The observed equity β values are set out in the following table – the two GARCH estimates are based on the different ways in which values can be generated: (1) an average over the period; and (2) an average of the daily β values.

Table 3.1 Estimates of β values for a range of European and American utilities

Company	OLS	LAD	GARCH 1	GARCH 2	Sector
Eversource Energy	0.60	0.61	0.61	0.49	Distribution
Consolidated Edison	0.52	0.55	0.56	0.52	Distribution
Unitil Corp	0.35	0.39	0.50	0.35	Distribution
Terne Rete Elettrica Nazionale	0.50	0.51	0.58	0.47	Transmission
ACEA	0.64	0.60	0.65	0.64	Generation
Snam	0.43	0.46	0.53	0.41	Gas distribution
EDP	0.59	0.56	0.61	0.59	Integrated utility
Red Electra	0.59	0.56	0.62	0.59	Transmission
Naturgy Energy	0.77	0.78	0.82	0.74	Gas distribution
Enagas	0.61	0.62	0.66	0.60	Gas transmission & storage

Source: Ofgem, Bloomberg and Indepen analysis

Appendix G provides more detail on the GARCH specifications that were used as well as how the β values vary over time.

How useful are these comparators? If a β decomposition is being undertaken for a company that has operations in one of the countries covered by the comparators, then they may be helpful – Section 4 provides an example that utilises US information.

If the comparators are a key input to the calculation of an equity or asset β value for a price determination then the following issues should be considered.

- How close a match is the company's risk profile with that of the regulated company?

- How different is the financial structure, tax regime and business environment in the country where the comparator is based?
- How different is the financial structure and tax regime faced specifically by the comparator company?
- How much do the β values vary over time and what is the appropriate time frame to be considered?

The answers to these questions are likely to be such that significant care needs to be taken when trying to draw anything more than a broad range from the international data. ACCC (2017) is a good example of a case where some weight is placed on international comparators but much greater weight is placed on Australian evidence even though the number of listed utility companies is low.

Overall, when listed UK examples exist, it is more appropriate to seek to understand the β values for these rather than to research international comparators.

4 The relationship between asset and equity βs

The previous sections have been concerned with estimating equity β for the set of listed business for which data are available. Each business has its own financial structure and most are different to the notional (efficient) structure assumed by the regulator. Also most of the observed equity βs are portfolio βs, inasmuch as the listed entity contains more than one business.

This section considers how to identify the effect of financial structure on β and whether, and if so how, a group equity β can be decomposed into constituent business equity βs.

4.1 Asset βs and de-gearing/re-gearing

As noted in Section 1, equation (3) sets out the relationship between the equity β, asset β and financial structure as measured by gearing – the debt β is also a consideration. Gearing, in UK regulatory precedent is defined as per (5).

$$g = \frac{ND}{(ND+E)} \qquad (5)$$

where ND is net debt, ie gross/total debt minus cash and short-term financial instruments (near-cash) and E is the market value of equity.[29]

We are interested in asset βs because companies' capital structures differ, for all sorts of reasons. Regulators want to focus on the underlying business risk, captured by the asset β, to be able to estimate the equity β for the regulated company at the notional capital structure.

Equation (3) is often simplified by assuming the debt β is zero, giving a straight relationship between the asset β and equity β of[30]

$$\beta_a = (1-g)\beta_e \qquad (6)$$

This relationship, while simple, has not proven robust and increasingly UK regulators are applying non-zero debt βs – for example Ofwat on the recommendation of Europe Economics (2017) has proposed a range of 0.1-0.15 for PR19 (further regulatory evidence and precedent is presented later in this section). Other complications also exist, for example, tax ought to be considered as part of the relationship and there is a question of whether a linear relationship is appropriate. Each of these is discussed below.

[29] This definition has been used by the majority of regulators in the UK since the 1993 MMC gas appeal (MMC(1993)).

[30] Ofgem (2004a) is a good example of the acceptance of a zero debt β.

Debt βs[31]

What happens if we relax the assumption of a zero debt β? Our estimate of the asset β, using equation (3), when we de-gear the equity β will be affected and any subsequent equity β calculated by re-gearing the asset β will be different from that derived by assuming a zero debt β. Consider the following example which illustrates the impact of the non-zero debt β.

> An equity β of 0.7 is observed for a company with 50% gearing. The regulator wishes to establish an estimate for a notional 60% geared company.
>
> Assume equation (4) holds – so a zero debt β. Then the asset β is 0.35 – calculated as 0.7 x (1-0.5).
>
> Now, applying the 60% gearing gives an estimate of the equity β of 0.875 – found by rearranging equation (4) and using the resulting equation to give 0.35/(1-0.6).
>
> Now, if a debt β of 0.1 is used in equation (3) the asset β is 0.4 (given by {0.7 x (1-0.5)} + (0.1 x 0.5). When this is re-geared to 60% you get an equity β of 0.85.
>
> If the estimates were re-geared at 40% then equation 4 gives an equity β of 0.583 and equation 3 gives 0.6.

So, the assumption of a non-zero debt β has around 3% effect in these examples. Given the uncertainty around the estimation of the equity β, this could easily be lost in the noise around the estimate. But, the assumption of a non-zero debt β is more appropriate and is increasingly becoming the standard UK approach.

Given the above example, it is important to understand how to estimate the debt β; and whether it should be allowed to vary according to the riskiness of the company or over time. Two approaches to estimating debt βs tend to be discussed[32]

- direct estimation through OLS regression of debt returns against market returns
- decomposition of observed debt premia into different elements including systematic risk which then allows the estimation of a debt β

Few examples of direct estimation exist although Europe Economics (2007) provides an estimate for BAA (0.17) and quotes evidence from Fama and French (2003) which ranges from 0.19 to 0.30 depending on the credit-rating of the debt.[33] Pratt & Grabowski (2014) also provide estimates of directly calculated debt βs by credit-rating for 2010, 2011, 2012 and the

---

[31] Appendix E provides more information on these and related issues.

[32] See for example Europe Economics (2007).

[33] No further citation is provided in Europe Economics (2007) and we have not been able to source the document directly. We will approach Europe Economics to determine which Fama and French paper they actually reported in that document. However, the results are in line with Pratt & Grabowski (2014) although the extremely long period covered by the Fama and French data (it is report as being 1963 to 1991) is likely to cover significant variability if the more recent results are representative.

first part of 2013 – with significant variation in values over time and a trend reduction in the debt β for the most highly rated firms falling to 0 at the end of the period covered.

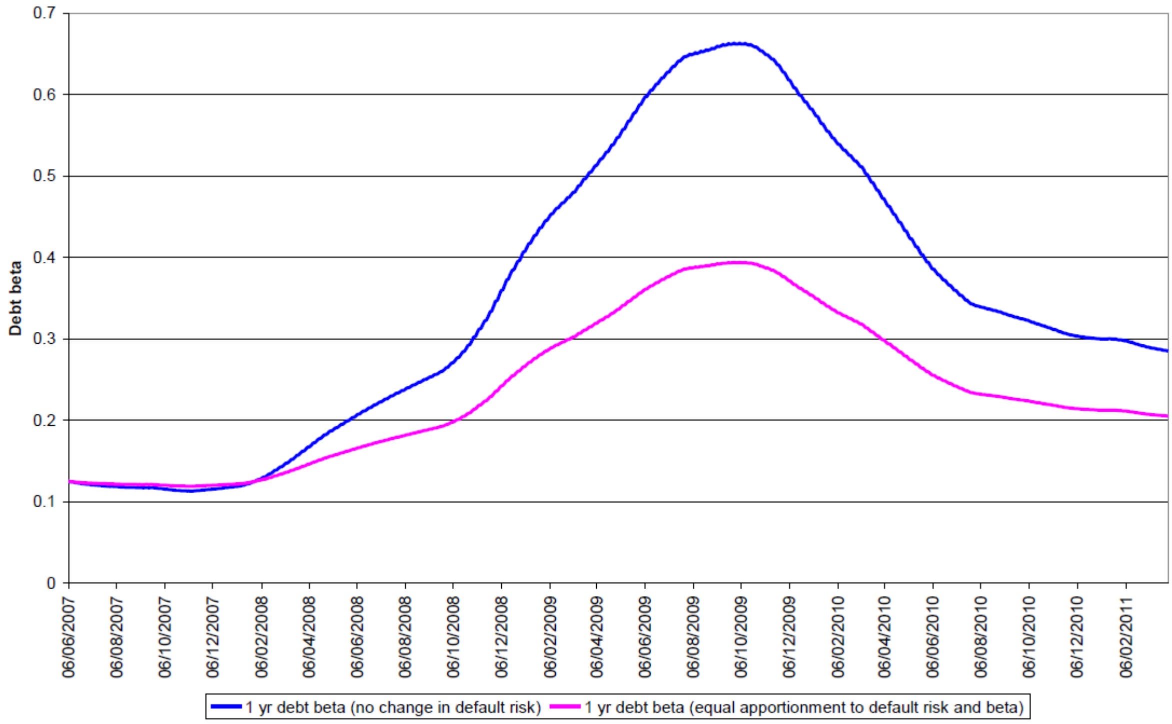
UK regulatory precedent has tended to focus on debt premium decomposition. An observed debt premium captures systematic risk (which drives the debt β value) plus other risk such as default risk (and its associated costs). It is possible to determine a specific value for the risk of default – credit rating agencies tend to produce cumulative default risk tables. Europe Economics (2007) did this for BAA and found a debt β value of 0.21.

The second question about debt βs becomes important when a significant difference exists between the actual gearing of the observed equity β and the notional level of gearing that the de-gearing/re-gearing calculation is used to estimate. If the debt β is assumed to be constant then there is no calculation issue, but the assumption is unlikely to hold. Would a AAA rated borrower have the same debt β as a BBB borrower? If not, then assuming that the debt β is constant will lead to mis-estimation of the re-geared notional equity β. This is supported by the evidence in Pratt & Grabowski (2014). Further, mirroring some of the concerns about equity βs, are debt βs constant over time? If not, again as suggested by Pratt & Grabowski (2014), understanding how they move over time becomes important.

Europe Economics (2011) addressed this question with respect to Ofcom's WBA determination.[34] While not directly seeking to determine how much of the change in debt premium that occurred between 2007 and 2011 was linked to default risk and how much to a changing debt β. Ofcom had argued for a constant debt β over this period. Figure 5.3 of Europe Economics (2011) illustrates what would happen to the debt β estimate under two scenarios: (i) 100% of the change in debt premium being linked to the debt β; and (ii) 50% being default risk and 50% being debt β. The figure is reproduced below.

In scenario i the debt β moves from a little over 0.1 to almost 0.7 before dropping back to about 0.3. Scenario ii sees a smaller impact – a peak of about 0.4 and an end-point close to 0.2. But in both cases significant changes occur over this time. Given the impact that the value of the debt β can have on the asset β and the re-geared equity β, knowing whether the value is constant or subject to variation is clearly important.

[34] See Ofcom (2011) and Europe Economics (2011).



Source: Europe Economics

Figure 4.1 Impact of assumptions on time varying debt βs

As mentioned earlier in this Section, while multiple issues with debt βs can be, and have been, identified, UK regulators are increasingly using them. As such, understanding recent regulatory precedent is important. Table 4.1 below provides a summary of recent evidence.

Table 4.1 Recent UK regulatory precedent on debt βs

Regulator	Date	Price control	Status	Value
CAA	2014	Q6 (Heathrow and Gatwick)	Determination	0.1
CAA	2014	NERL	Determination	0.1
Ofcom	2014	WBA	Determination	0.1
Ofcom	2015	MCT	Determination	0.1
Ofcom	2016	LLCC	Determination	0.1
CAA	2017	H7	Proposal	0.05
Ofwat	2017	PR19	Proposal	0.1-0.15
Ofcom	2018	WLA	Determination	0.1

Source: UKRN (2018), CAA (2017), Europe Economics (2017)

While much of the recent evidence has been from telecoms, both airports and water have used or proposed the use of debt βs for ongoing price determinations. As noted in Appendix E, in its initial investigation of debt βs for airports at Q5 the Competition Commission calculated a range of 0.09 to 0.19 before using a value of 0.1.

This discussion and the evidence presented show that debt βs need further research and then estimation approaches should be refined and agreed. Ensuring that clear and defensible assumptions are made when moving between equity and asset βs is key if the conventional approach is to continue to be used. As any analysis will take time, if Ofgem is minded to use a debt β value in the short-term then regulatory precedent would suggest a value of around 0.1 (bearing in mind that recent proposals cover the range 0.05 to 0.15). If such a value were to be used, transparency is key so that the implications of the choice can be assessed.

Tax

The value of the company is affected by any tax shield arising from the tax deductibility of interest. If companies face different effective tax rates, then this is likely to affect the relationship between asset and equity βs. There are versions of equation (3) that allow for tax rates.[35]

The primary issue that arises here is 'what tax rate should be used'? Is it the statutory or headline rate of tax or the effective tax rate? These can be substantially different owing to the tax treatment of investment etc.[36] Within a country this can be complex and when comparators from other countries are used the situation becomes even more difficult.

Work by Cooper and Nyborg (2004) suggests that use of equation (3) rather than a tax adjusted version could cause errors in the asset β of about 10%.[37] The actual size of the error depends on the corporate and personal tax rates.

Overall, tax is most important when there are significant differences in the effective rates faced by the comparators used for de-gearing and the company being re-geared. Where differences are small, not applying tax is an acceptable assumption. Evidently, clarity at a more detailed level is needed when international comparators are used.

Linear relationship

An assumption underlying equation (3) is that there is a linear relationship between asset βs and equity and debt βs. Work on estimating the cost of debt in the 1990s showed non-linear relationships, either quadratics or elbow points.[38]

If this is so, assuming that the debt or equity β follows a linear relationship is a mis-specification. It is easy to see how investor sentiment could lead to a stable equity β for a range of gearing levels with an increasing equity β if a company went beyond what investors see as the acceptable range. This phenomenon would explain the evidence presented in the CEPA 2010 note for Ofgem better than the traditional assumption.[39]

---

[35] This is an issue increasingly covered in text books. See the two chapters on tax in Armitage 2005.

[36] The standard text on this is King-Fullerton (1984). Jorgenson and Landau (1993) provides good examples of how tax regimes differ across countries and the implications of this, using the King-Fullerton framework.

[37] *Tax adjusted discount rates with investor taxes and risky debt*, 2004, Cooper and Nyborg. Downloadable from here.

[38] Confidential client reports prepared by a UK consultancy.

[3939] Jenkinson (2006) noted a similar apparent mismatch between theory and evidence where gearing levels were increasing but equity βs were unchanged – implying either that asset βs were falling or the assumed relationship was not holding.

If this is the case, then great care is needed when interpreting estimates that have been subject to de-gearing and re-gearing. It may even make this a pointless task without further research on the implications of non-linear relationships.

Other issues

Two further issues should be noted.

First, implicit in the relationship in equations (3) and (5) is the assumption that the absolute change in the level of gearing for the company is what matters. This will only be the case if the national level of gearing underlying the market risk premium is stable. Generally, it is the relative change in gearing that matters – has the company increased or decreased its gearing relative to the change in national level gearing?

If we want to estimate the observed equity β over a period of time, at least two to five years and possibly much longer, then it is possible that market gearing could change and the evidence suggests that such changes could be quite significant. Consider the figure below, reproduced from the 2005 Bank of England document *The determination of UK corporate capital gearing* (Brierley and Bunn). This shows that gearing levels can be relatively stable for periods of time with bouts of change. Whether those periods of change correspond to structural breaks identified earlier is an issue that deserves further attention.

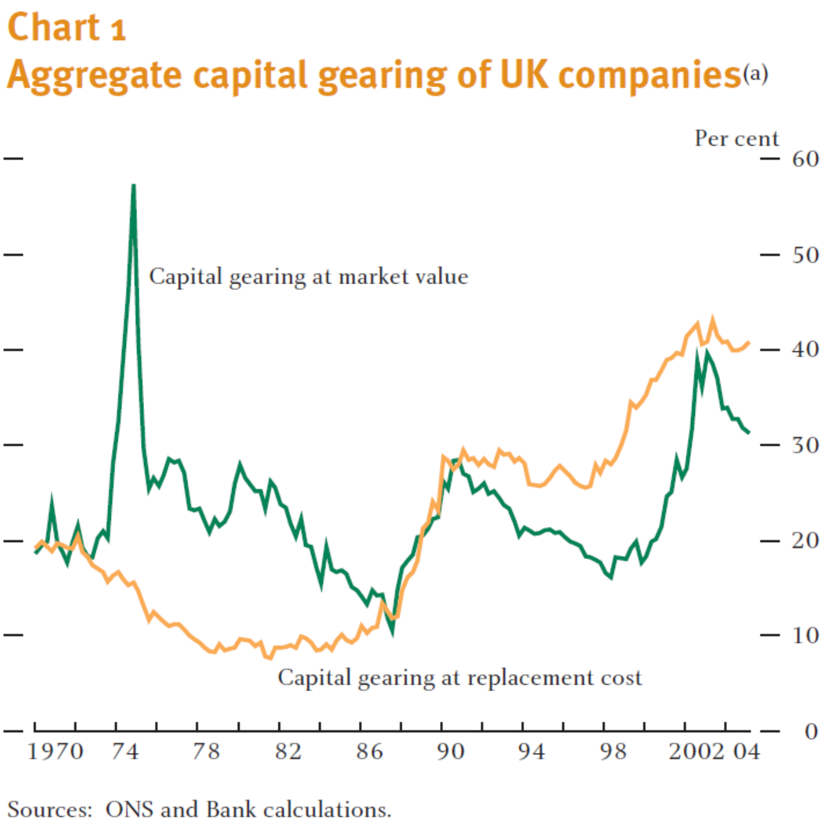


Figure 4.2 Changes in national level gearing from 1970 to 2004

Possible solutions include

- seeking to establish a new relationship and assumptions to allow us to rewrite equations (3) and (5), possibly as relative gearing
- measuring equity βs during periods of stability so that the underlying relationship and assumptions in equations (3) and (5) hold
- acknowledging that the problem exists when interpreting the results if limiting the data period or rewriting the relationships are not desirable

Our recommendation is that a mix of the second and third options should be used. It is quite possible that some of the structural breaks that are identified capture changes in national gearing and consequently seeking periods of stability in national gearing may coincide with periods where no structural breaks occur.

The second issue relates to the measurement of gearing. It was noted earlier that some regulators in other jurisdictions use total debt rather than net debt. It is also possible that book value of debt, used in the UK, may not adequately capture the value of the business and market value should be employed.

Europe Economics' report for Ofwat (2017) provides a taxonomy of gearing definitions that is helpful. They define three measures

- RCV gearing: measured as book value of net debt to regulatory capital value (RCV) or regulatory asset base (RAB) – we will use RAB gearing in this report except when referring to specific calculations undertaken in the Europe Economics report
- Enterprise value gearing: measured as book value of net debt to enterprise value (defined as market value of equity plus book value of net debt)
- Fair value gearing: measured as fair value of net debt to the sum of the fair value of net debt and the market value of equity[40]

Europe Economics calculated the different measures of gearing for two of the water and sewerage companies. The different measures are reported in table 4.2 below. We have generated a series of Enterprise Value gearing estimates for the six utilities we are considering, reported in table 4.3 and RAB gearing estimates for five of the six companies in table 4.4.

Table 4.2 Water company gearing measures (%)

Company	RAB Gearing	Enterprise Value Gearing	Fair Value Gearing
United Utilities	60	51.1	52.1
Severn Trent	60	47.4	49.6

Note: RAB gearing is approx. as read from figure 6.1 of Europe Economics (2017)

Source: Europe Economics (2017) chp 6

[40] Rather than measure the market value of the debt a regulatory value was calculated as the treatment of embedded debt provides protection against some of the market movements in the price of debt.

Table 4.3 Enterprise Value Gearing, 2014-2018 (as at April each year)

	BT	NG	UU	SSE	PNN	SVT
2014	19.2%	41.0%	52.4%	29.0%	44.4%	50.8%
2015	13.7%	40.8%	50.2%	23.6%	40.0%	49.1%
2016	19.8%	39.4%	51.8%	31.2%	42.6%	48.3%
2017	25.2%	30.6%	51.3%	30.7%	42.2%	47.8%
2018	32.2%	44.7%	60.3%	39.1%	50.9%	55.5%

Source: Ofgem, Bloomberg and Indepen analysis

Table 4.4 Comparison of gearing measures, October 2018

	NG	UU	SSE	PNN	SVT
RABGearing	65.3%	62.4%	100.9%	88.2%	60.6%
Enterprise Value Gearing	44.9%	58.5%	41.7%	50.1%	54.6%

Source: Ofgem, Bloomberg and Indepen analysis. Appendix F provides more information on these gearing estimates.

Note that the figures presented in the three tables are similar where there are overlaps but some small differences arise from the timing of the measurement of the market value of equity. We have excluded pension fund deficits in our calculation of company gearing for the purposes of this analysis. Also we have not treated the SSE hybrid security as part of its gross debt.

We have used book value for debt on the basis that this best reflects the expected future cash flows of the business to service debt. An alternative approach to use fair market values for debt was not considered in this analysis but could be used.

The fact that we are using observed equity βs and de-gearing them using the actual gearing (based on Enterprise Value gearing) while potentially re-gearing them based on a RAB gearing notional value creates an inconsistency. Consistency requires that an Enterprise Value notion gearing level be established. When the Enterprise Value and RAB gearing is close, ie the MAR is close to 1, then this is not a problem and the notional RAB gearing can be considered an Enterprise Value gearing value. But how should a regulator react if the MAR is significantly different to 1? One possible solution is set out in the box below.

What this solution shows is that to generate a notional Enterprise Value level of gearing we need to have a view about what an appropriate MAR should be. What is clear, is that assuming an MAR of 1 is a strong assumption and one that leads to a significant de- re-gearing effect on equity βs.

If a notional RAB gearing level of 60% is used by a regulator, what value should that be as Enterprise Value gearing?

One solution is to use a pure-play MAR to revalue the RAB and to generate an implied Enterprise Value (iEV) level of gearing.

Suppose a MAR of 1.1 is appropriate – not an unusual long-term value for UK utilities. Then:

60% RAB gearing is generated by net debt of 60 over an RAB of 100

A MAR of 1.1 means an implied Enterprise Value of 110 (the RAB of 100 multiplied by the MAR of 1.1).

This gives an iEV gearing of 54.5% (net debt of 60 divided by the implied Enterprise Value of 110)

So, in this case RAB gearing of 60% becomes iEV gearing of 54.5%

What happens if a higher MAR value is used? Consider the case where MAR is 1.5.

This would generate an implied Enterprise Value of 150 (RAB of 100 multiplied by 1.5)

iEV gearing is then 40% (net debt of 60 divided by the implied Enterprise Value of 150)

Box: Estimating a notional Enterprise Value gearing level

It appears that regulators currently mix RAB and Enterprise Value gearing. When the market to asset ratio is close to 1 this will not cause much of an inconsistency but if there is significant divergence from 1 it will be more of a problem.

Given that evidence suggests that MAR values tend to be above 1 then some form of adjustment is necessary. Two options exist, regulators can use

- a "normal" MAR to revalue the notional gearing value
- the actual MAR

The latter raises circularity and valuation issues, especially for portfolio companies while the former requires a view be taken about an appropriate normal value. The problems associated with the latter approach would mean that using the former is appropriate. As a starting point, a MAR of 1.1 could be used but further research on this issue should be undertaken.[41]

This value is based on a mix of information. Annex I of the UKRN study (UKRN(2018)) provides a detailed discussion of the possible reasons why a value greater than 1 is normal and notes that an average value in excess of 1.2 over the last 20 years for the energy and

[41] See Annex J of the UKRN study for a fuller description of the arguments for a MAR above 1 and the possible ranges based on efficiency outperformance etc.

water networks.[42] This period includes several acquisitions where the premium was significant and given the impact of the acquisition premia we do not think it would be appropriate to use a value as high as 1.2. However, what is clear is that justification for a number in excess of 1 exists and 1.1 would appear to be defensible – the recent MARs for the water pure-plays is an average of about 1.1.

Options and recommendation

Even if the relationship between equity and asset βs follows the simple linear relationship set out in equation (3) there are measurement issues that could have a significant effect on the estimated asset β. Whether this would have a material impact on the equity β used for the notional cost of equity calculation will depend on

- the degree of consistency in the application of the assumptions
- the quantum of the difference between market and book values
- the quantum of the difference between the actual and notional gearing

It is important to consider whether these errors are multiplicative or additive as this will affect how big the error might be.

Some of these points are not easy to address and their materiality will depend on market circumstances at the time of the determination.

There are two main options.

- Provide clarity about what assumptions have been made when using either equation (3) or (5) (correcting the obvious inconsistencies that currently exist where they can be adequately addressed) and noting that regulatory discretion will play a role in the final choice of the cost of equity. Some discretion will be in judgements about the appropriateness of the assumptions. This is what happens in other building blocks during a price determination, such as estimating efficient totex allowances.
- Using evidence of observed equity βs, gearing levels, risk profiles etc and using this as the basis for judgement about what the notional equity β should be without the potentially spurious application of equation (3) or (5).

While the latter option has some clear advantages with respect to implementation and minimising the arguments about assumptions, UK precedent favours a more formal calculation and might favour the former option.

Our view is that the former option should be used – using regulatory precedent and "normal" values to derive assumptions – but the latter option also employed as a sense check on the numbers arising from the application of some form of equation (3) or (5).

[42] This number is based on work by PwC and is reported in UKRN(2018).

When implementing this, we think the order of priority needs to be addressing the possible gearing inconsistency and then the debt β. Until further research has been undertaken we would recommend

- for the gearing the use of a normal MAR of say 1.1 as a starting point or be explicit about what is being assumed
- regulatory precedent of a debt β of between 0.1 and 0.15

4.2 Portfolio asset decomposition

The second issue when estimating a notional equity β relates to the fact that most listed businesses are not pure-plays but a portfolio of businesses with the consequent observable equity β being a weighted average of the business equity βs of the individual businesses. If unregulated businesses account for a significant proportion of the activities and are likely to have a different risk profile, it is necessary to decompose the β to obtain an estimate for the regulated activity. This is something that Ofcom has done when regulating BT and has been suggested in NERA (2018) as a necessary adjustment to the National Grid observed β because the risk profile of the regulated US businesses (about 40% of the company's assets) differs from that of the UK regulated businesses.

In this section we consider the ways in which βs have been decomposed and the issues that can arise. It should be noted that a similar approach could be adopted for handling an expected change in risk within the regulated business, say owing to the development of new capacity. For example, R3 and T6 during development at Heathrow could be considered as a separate risk that is then blended into the different risk associated with existing operational assets.[43]

Decomposition approach

The approach to decomposition that is normally used is shown in equation 6 below.[44]

$$\beta_G = \beta_1 \frac{Assets_1}{Assets_G} + \beta_2 \frac{Assets_2}{Assets_G} + \cdots + \beta_N \frac{Assets_N}{Assets_G} \quad (7)$$

In this the observed β for the group (G) is a weighted average of the βs of the N separate businesses, weighted by the proportion of assets employed in each business.

[43] Normally new capacity would not be sufficient to warrant this treatment, but as seen with T5, when a significant proportion of assets are associated with new capacity there may be a material impact on the forward looking β estimate. This type of approach could have been adopted for the significant transmission capex at RIIO-T1. Implicitly the choice of a higher percentile in the final range achieved something similar but is less transparent than a weighted β approach.

[44] Implicitly this approach assumes that there is no correlation between the βs of the different businesses.

This is the basis of the approach employed by Ofcom – which initially considered two businesses within BT and now considers three. Similar approaches have been recommended elsewhere.[45]

Issues

Questions arise when employing this sort of decomposition.

- Should it be applied to equity or asset βs?
- If applied to asset βs should a group average, group actual or industry specific gearing be used?
- Are net assets the right way of measuring the weights?

Each question is addressed in turn. A simple numeric example is used throughout this section. The basic information for this example is provided in the table below.

Table 4.4 Information used in β decomposition example

Business	%age of net assets	Observed equity β	Actual gearing (%age)	Comparator equity β	Comparator industry gearing (%age)
Group	100	0.6	40		
Business 1	40		45		50
Business 2	40		50	0.6	60
Business 3	20		10	0.5	0

Asset or equity β

Equation (7) was left ambiguous with respect to whether the decomposition is done for equity or asset βs. If it is undertaken for equity βs then the decomposition is simple. In our example, with a Group equity β of 0.6, comparator equity βs of 0.6 (business 2, 40% of net assets) and 0.5 (business 3, 20% of net assets) then the equity β for business 1 is 0.65.

This assumes that all the businesses have the same financial structure –in our example, a gearing level of 40%. Given this, an asset β decomposition could also be undertaken, the Group asset β is 0.36, business 2's asset β is also 0.36 and business 3's asset β is 0.3. Decomposing those numbers gives an asset β of 0.39 for business 1, which is what you get if the calculated equity β of 0.65 is de-geared at 40% gearing.

Measure of gearing

While the asset β example above is simple, is it correct? Assuming that each of the businesses have the same capital structure is a strong assumption and may be inappropriate. Consequently the question of what gearing should be used arises. The impact of Group average gearing was shown above.

[45] See for example Appendix 4 of *Cost of Capital: The application of financial models to state aid*, Alexander, 1995.

Two other possibilities exist

- actual gearing reported for the businesses with the Group (if regulated businesses account for the majority of the Group it should be possible to use regulatory accounts to determine the capital structure in the regulated business and use this to calculate the capital structure of the remainder of the Group), or
- industry average gearing or whatever measure is consistent with the comparator equity βs

Under the first approach above the total gearing across the businesses will be consistent with the Group gearing while under the second it is possible that the gearing figures will not be consistent.

In the example, the use of actual gearing would reduce the asset β for business 1 to 0.375. This would give an equity β of 0.68. While industry levels of gearing would see business 1 have an asset β of 0.41 and an equity β of 0.82.

In this example the choice of approach to gearing would have a significant impact on the asset β with the range being almost 9% of the average value (range of 0.035 over an average value of 0.3925). Given this, it would seem to be appropriate to check what the actual gearing is within the Group and how this compares to industry levels of gearing and if they differ significantly to investigate further why there is a difference and whether that needs to be adjusted for.[46]

Assets or profitability?

The net assets approach to the weighted average β is an appropriate basis if the majority of the businesses are asset intensive. What happens if some of them are asset light and a focus on net assets would bias the result? A company like SSE is increasingly focused on asset light (retail) or unregulated businesses (generation) – the OXERA 2018a report states that approximately half the profits of SSE come from these activities.

In these circumstances should a decomposition be done using profitability as the weight? Or should implied asset values based on multiples of earnings/EBITDA be generated for the asset light businesses? Again, this is something to be considered if a β decomposition is needed.

Applying decomposition to National Grid

NERA in its reports to industry participants as part of the RIIO-2 consultation has raised the point that National Grid ought to be considered a portfolio of UK and US assets and decomposed accordingly. In principle we agree with NERA, provided that good

[46] Of course, if there are numerous comparators for the main regulated business then this decomposition is more of a sense check rather than primary source of information. However, as noted in Section 2, the paucity of listed network businesses in the UK means that any information that can be gained should be.

comparators can be found for National Grid's US businesses and that any assumptions are explicit so the sensitivity of the results to the assumptions can be assessed.

In the following example we seek to illustrate the impact that different assumptions have on a decomposition rather than seek to establish a single "true" value. For this example, rather than seeking to find specific comparators that we believe are sufficiently similar, we focus on the three companies noted by NERA in one of its reports: Consolidated Edison, Eversource Energy and Unitil Corp.

In the two tables below we set out two possible calculations for the National Grid Group β value to be decomposed. In both cases we keep the Group and US comparator equity β values unchanged. Also the weights between the US and UK businesses ae fixed.[47] Instead we consider two different levels of Enterprise Value gearing – in example 1 we focus on numbers that are consistent with the last couple of years and in example 2 we take a longer perspective. (Note that we have kept the de-gearing and re-gearing simple by excluding debt βs and tax – for this illustration they would be an unnecessary complication.)

Table 4.5 National Grid example 1 (data from 2017 and 2018)

		UK	US	Group
%age assets		51%	49%	
Gearing		35.6%	39.9%	37.7%
Asset B		0.44	0.30	0.37
Equity B		0.69	0.50	0.6

Table 4.6 National Grid example 2 (data from 2000 to 2018)

		UK	US	Group
%age assets		51%	49%	
Gearing		42.1%	46.0%	44.0%
Asset B		0.40	0.27	0.34
Equity B		0.69	0.50	0.6

In both examples we generate an equity β value for the UK regulated business of 0.69 – that is a coincidence rather than by design. How should we interpret this value? The equity β value is at an Enterprise Value gearing level below that of the Group, so by implication the RAB gearing for the business would be lower than the overall RAB gearing (simplifying by assuming that there is a constant MAR between the UK and US businesses). This is likely to bring the RAB gearing closer to the notional (efficient) level assumed by Ofgem.

Table 4.7 below shows a different example, focusing on the most recent 2018 evidence. In this, the Enterprise Value gearing for the UK business is higher than that of the US and Group – so while a higher equity β value is found, that has to be weighed against a potentially even higher level of RAB gearing – above the notional level assumed by Ofgem.

[47] The values for the UK and US businesses are based on regulatory values that will differ from book values. However, for this example these values provide a suitably illustrative basis for the calculations.

Table 4.7 National Grid example 3 (2018 data)

		UK	US	Group
%age assets		51%	49%	
Gearing		48.3%	41.0%	44.7%
Asset B		0.38	0.34	0.36
Equity B		0.73	0.58	0.65

These simple examples, based on illustrative but informative values, show that while it is possible to derive a decomposed set of values, clarity about the assumptions and data used is important as this can have a significant impact on the results. Further, as noted earlier, consistency in the definition of gearing is important to allow a better understanding of the results.

Options and recommendation

β decompositions may be an important part of the analysis given that a strong case can be made for three of the six UK listed network businesses being portfolios of some form (BT, National Grid and SSE). BT is already subject to β decomposition.

What is the appropriate way to do this and what should be the weights? Given the likely differences in risk profiles it then becomes important to focus on asset β decomposition. The sensitivity of the results to different assumptions will help to form a view taken on the best approach to adopt.

We do not believe that the assumptions that have to be made for this calculation currently justify the use of the results of the decomposition. With further work on the evidence and assumptions it ought to be possible to refine this and estimate a range of values that can help inform a regulatory determination.

5 Conclusions and Recommendations

This report investigates a number of issues linked to the estimation and use of βs when setting an allowed cost of equity for a regulatory price determination. This section sets out our conclusions and recommendations for regulators after a brief recap on the evidence that we have found.

Focusing on CAPM as the basis of estimating the allowed cost of equity places a significant emphasis on the β value as it is the only company/sector specific aspect of the calculation. As such, significant time and resources are regularly spent seeking to estimate the appropriate β value.

While evaluating the various approaches to estimating β we have considered three criteria – set out in Section 1. We have given greatest weight to the assessment of the statistical appropriateness of different approaches and the degree to which assumptions have to be made explicitly or implicitly to use an approach. Replicability and broader implications for regulatory policy also need to be considered.

5.1 Findings

Approaches

The approaches of UK regulators have been complex and it has been standard practice to make a new estimate of β for each price review.

The methods have been inconsistently applied with changes in key assumptions.

Regulators have adopted estimates that have been close to 1, which except in the case of telecoms are substantially higher than justified by the evidence.

In Australia the AER has adopted an approach based on a small number of local companies with estimates updated infrequently only when significant changes have occurred[48]. However, the range is broad – from 0.4 to 0.7 for equity βs – with recent decisions focused on the upper end of this empirical range.

Statistical evidence

Academic analyses of returns and the estimation of βs for use in the CAPM have found that the OLS estimator can suffer from heteroscedasticity in the error term – discussed in Section 2 and several of the appendices. Various ways of alleviating the effects have been proposed but they all impose additional constraints or assumptions.

[48] See AER(2017) for an example of this.

Alternative sources of β estimates, such as accounting data (discussed in Section 3) are reasonable in theory, but do not generate estimates that are statistically significant.

As assumptions are relaxed, such as those about taxation, debt risk etc, the standard relationships become complex. This was investigated in detail in Section 4.

Evidence, both that collected for this report and more generally, has shown that

- observed βs vary over time which means that estimates are sensitive to the dataset used, the length of the estimation period and the period over which we measure returns (daily, weekly or monthly)
- heteroscedasticity is at its greatest in the daily data which means that OLS significance tests and confidence intervals should be undertaken with corrected standard error values if OLS is used
- GARCH models can be found which address the heteroscedasticity problem but model selection does not lead to a single preferred specification for all companies which makes the estimation process more onerous and requires clear processes and selection criteria
- there are structural breaks in the data, both in the aggregate and for individual companies, so we need to interpret the breaks and their implications for any dataset
- the simple asset β/equity β relationship does not hold consistently under a reasonable range of assumptions

From our consideration of gearing, we find

- estimates of debt βs are likely to vary according to the riskiness of the company and to change over time, leading to more complex specifications of the relationship which are difficult to estimate, and giving estimates that are not reliable for the purpose
- the definition of gearing has not been applied consistently and the inconsistencies are material – for example, observed βs are evaluated by using Enterprise Value gearing to derive asset βs and then re-gearing to the notional (efficient) level of RAB gearing[49]

In the main body of the report we have sought to show the materiality of these points and their likely impact. While any actual impact will depend on the specific circumstances under consideration, current circumstances mean that, for example, the inconsistency between Enterprise Value and RAB gearing can significantly increase the equity β value generated through the approach currently applied.

Evidence from sources of information, detailed in Section 3, other than time series of returns for the listed companies shows that

- accounting numbers do not provide reliable estimates as the models are not statistically robust
- other sources of information, such as debt premia, provide some information but it is of limited value

[49] What effect this has depends on the relationship between the actual and notion levels of gearing as well as the MAR.

- the variability of the assumptions underlying βs estimated in other jurisdictions makes the interpretation of international evidence unreliable

5.2 Conclusions

We draw the following conclusions.

- There is a substantial body of company and sector information that can be used to inform regulatory judgements about the range of values within which the β lies.
- The characteristics of the data series are such that making a statistical estimate of a company's/sector's β at a point in time entails a process that is complex and sensitive to several assumptions and potentially subject to bias and inaccuracy.

The purpose of the estimation poses a greater challenge than this. Regulators are seeking an estimate of β to be applied for a period of years and usually they are doing this a year or more before the determination is to come into effect. This means they are looking for an estimate that will be appropriate for a period of more than five years. This leads to the conclusion that reliance on any single estimate based on a short run of data is not appropriate.

Our overall conclusions are that

- none of the specific approaches we have considered is without significant failings
- accounting approaches do not help
- international comparisons do not help

5.3 Options

In the light of the conclusions, there are three feasible approaches.

- **Approach A** – establish the range of feasible results and agree a set of principles for the application of judgement to identify the estimate to be used
- Start from a set of definitions, principally about actual and notional gearing, that are internally consistent
- Consult on the factors that should affect a judgement of the equity β for use in setting prices
- Collate an extensive data set – probably for the period since 2000 even though this will include some structural breaks
- Review data for structural breaks and decide how to proceed
- Consider the distribution of results from estimates using different time windows and frequencies of returns (this can include using OLS and other estimation approaches)
- Apply judgements derived from the consultation to arrive at the preferred estimate of the equity β within the distribution

- Where portfolio βs need to be decomposed, make explicit assumptions including tax and gearing

Approach B - a more technocratic approach using a decision tree and well-defined criteria. It has features in common with Approach A, namely

- Start from a set of definitions, principally about actual and notional gearing, that are internally consistent
- Collate an extensive data set – probably for the period since 2000
- Review data for structural breaks and decide how to proceed
- Where portfolio βs need to be decomposed, make explicit assumptions including tax and gearing

The decision tree is primarily to deal with the fundamental issue of heteroscedasticity. It should be noted that in the tree we assume that there is at least a decade of data since the last structural break. If the structural break is more recent then this needs to be allowed for, either by using shorter time periods or by using the whole decade but placing greater emphasis on the evidence from the period since the structural break.

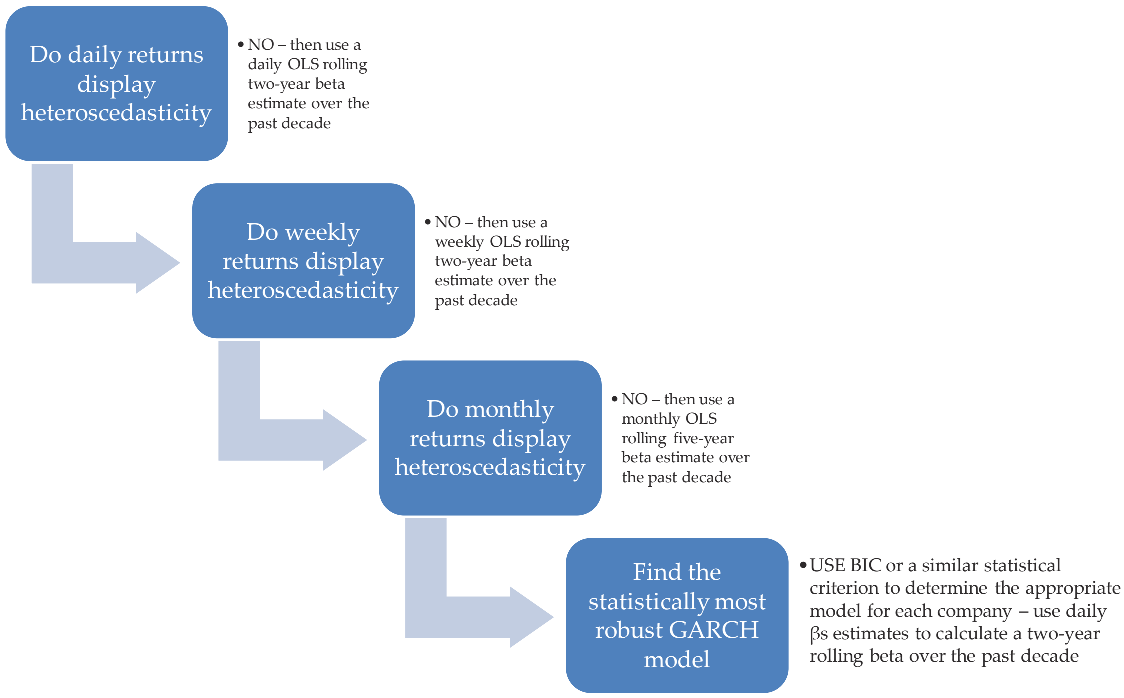


Figure 5.1: Decision tree for estimating equity β

Approach C - follow the approach used by the Australian Energy Regulator, decide on a number and stick to it unless there is a good reason to the contrary.

5.4 Recommendation

What are the pros and cons of each of the options outlined in Section 5.3 above? Table 5.1 below summarises our assessment against the criteria set out in Section 1 of the report.

Table 5.1 Pros and cons of each of the options

Approach	Approach A	Approach B	Approach C
Pros	Is robust inasmuch as a wide range of information is being considered to inform the decision Is not that different to how some of the other building blocks are treated Should be easy to replicate the information set on which the judgement is made	Is robust inasmuch as the approach finds the best statistical approach given the data and constraints faced From a broader policy perspective this approach is good as it is flexible and adapts as needed	Is simple and provides a clear expected value from which deviations only occur if sufficient evidence exists to justify the deviation Implicitly acknowledges the methodological and data issues that exist with the existing approaches
Cons	Depending on how judgement is used the approach could be open to appeal inasmuch as different stakeholders could arrive at different judgments given the same information set	May be difficult to replicate given the process requirements – although this could be overcome through the clarity of the rules in the decision tree The individual nature of the approach may raise some broader policy issues as it highlights differences between the listed companies It may appear unstable as the chosen approach for one or more of the companies could change from one price review to the next	At risk of appeal from companies and other stakeholders given the possible perceived lack of statistical credibility Loses possible up-to-date information depending on the rules for testing to see if change is necessary
Overall	Is closer to the way that some of the other building blocks for prices are addressed. Realistic but may be perceived as placing too much power in the hands of regulatory judgement Answer unlikely to be that different to Approach B	Technically correct and likely to lead to the "best" answer but requires significant work at each price review and may be perceived as unstable	Simple but lacks regulatory precedent in the UK and is a significant change from the current "precise" approach

On balance, UK regulators are unlikely to favour the simplicity of Approach C.

We recommend a combination that starts with Approach B and elements of Approach A to engage investors and customers to seek to establish a degree of consensus.

5.5 Some values

The evidence provided in the Appendices is summarised in the tables below.[50]

Table 5.2 Equity β estimates – 2000 to 2018 based on daily returns

Company	GARCH average of daily	GARCH average of variances	OLS 500 observation rolling	OLS single whole period estimate	LAD single whole period estimate
BT	0.99	1.01	1.00	1.04	1.00
United Utilities	0.55	0.56	0.57	0.57	0.57
National Grid	0.61	0.58	0.59	0.61	0.56
SSE	0.54	0.57	0.60	0.57	0.54
Pennon	0.49	0.45	0.49	0.45	0.44
Severn Trent	0.53	0.52	0.55	0.53	0.52

Table 5.3 Equity β estimates – 2008 to 2018 based on daily returns

Company	GARCH average of daily	GARCH average of variances	OLS 500 observation rolling	OLS single whole period estimate	LAD single whole period estimate
BT	0.92	0.95	0.98	0.94	0.95
United Utilities	0.58	0.59	0.58	0.60	0.59
National Grid	0.58	0.60	0.59	0.63	0.56
SSE	0.57	0.63	0.65	0.64	0.54
Pennon	0.61	0.56	0.60	0.55	0.55
Severn Trent	0.66	0.58	0.59	0.60	0.56

Table 5.4 Equity β estimates – 2013 to 2018 based on daily returns

Company	GARCH average of daily	GARCH average of variances	OLS 500 observation rolling	OLS single whole period estimate	LAD single whole period estimate
BT	0.93	0.97	0.95	0.97	0.96
United Utilities	0.73	0.70	0.69	0.69	0.66
National Grid	0.68	0.67	0.67	0.66	0.62
SSE	0.69	0.77	0.80	0.77	0.69
Pennon	0.68	0.64	0.67	0.67	0.63
Severn Trent	0.73	0.69	0.69	0.69	0.64

Based on the evidence found in this report and summarised above, if we were setting the equity β for the energy companies at the moment, subject to further analysis and refinement, we would recommend a broad range for the equity β of **0.55 to 0.70**. Note, in this range we exclude the information from BT which we consider to be significantly different to the other utility and infrastructure companies – however, we report it as a piece of information. In addition we have used our judgement to exclude some observations owing to factors including peakiness.

It should be noted that the different estimation approaches produces figures that are not significantly different from each other. While in this case it could be used as a justification for only using OLS we think that good regulatory practice involves checking the assumptions and sense checking results through the consideration of multiple approaches

[50] In each sub-period the GARCH specification has been chosen on the BIC – but constrained so that Cholesky specifications are not chosen. This was discussed earlier in footnote 14 and is set out in more detail in Annex C2 of Appendix C.

when they exist. However, what is clear is that OLS should continue to be one of the approaches that regulators consider (given the ease of replication, understanding etc and the fact that the equity β parameter estimates from OLS may be consistent with other approaches) and that corrections to OLS results are possible to correct for heteroscedasticity.[51]

A narrower range that we think appropriate is **0.57 to 0.65**, with **0.60** as the central estimate. The rationale for this is

- the 2008 to 2018 data window is the most important as it captures the period from the last general structural break, that pushes us towards 0.6
- the more recent evidence from 2013 to 2018 is informative and captures more recent market sentiment, this leads to higher numbers. However, the illustrations in section 2 show that this was driven by a spike (caused by both a reduction in the volatility in the market and an increase in the covariance between the stocks and the market which falls away after the election in 2017 – discussed in more detail in Annex A2 of Appendix A) which is reversing, reinforced by the smaller increases in the LAD estimates that place less weight on outliers, and consequently the weight given to this should be low
- the longest time period considered, 2000 to 2018, has only limited value given it includes the major events of the GFC and the likely associated structural break. However, some consideration of these long-term numbers should be given, hence the range extending a little below 0.6

We have not de-geared and re-geared the estimates as we believe that the core comparators have gearing levels sufficiently close to the notional level that the impact would be small and does not justify the numerous assumptions that have to be made to apply de- and re-gearing.

[51] As reported in Robertson (2018), OLS estimates can generate patterns very similar to those observed on real data or can substantially overstate the true parameter. Consequently estimation across other time periods may lead to a divergence in OLS and GARCH estimates.

Bibliography

AER (2017). Electranet transmission determination, 2018-2023 – Attachment 3: Rate of Return

Alexander, I. (1995). Cost of Capital. The Application of Financial Models to State Aid. OXERA.

Armitage, Seth (2005). *The cost of capital: Intermediate theory*. Cambridge University Press.

Armitage, S., & Brzeszczynski, J. (2011). Heteroscedasticity and interval effects in estimating beta: UK evidence. *Applied Financial Economics*, *21*(20), 1525-1538.

Becker, T. E., Robertson, M. M., & Vandenberg, R. J. (2018). Nonlinear Transformations in Organizational Research: Possible Problems and Potential Solutions. *Organizational Research Methods*.

Blume, M. E. (1971). On the assessment of risk. *The Journal of Finance*, *26*(1), 1-10.

Breusch, T. S. & Pagan A. R. (1979). A Simple Test for Heteroscedasticity and Random Coefficient Variation. *Econometrica*, 47, 1287-1294

Brierley, P. (2005). The determination of UK corporate capital gearing.

British Gas TransCo (1996). A framework for Efficient and Effective Regulation: TransCo's submission to the Ofgas Review of Price Controls.

Buckland, R., & Fraser, P. (2001). Political and regulatory risk: beta sensitivity in UK electricity distribution. *Journal of Regulatory Economics*, *19*(1), 5-25.

CAA (2017). CAP 1610 Economic regulation of capacity expansion at Heathrow: Policy update and consultation.

CEPA (2010). Short-term relationship between equity and asset betas, prepared for Ofgem

Cooper, I., & Nyborg, K. (2004). Tax-adjusted discount rates with investor taxes and risky debt. *Unpublished working paper London Business School, London, UK*.

Cooper, Ian (2005)."Comments on the document: Beta analysis of British Telecommunications."

Engle, R. F. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica*, 50, 987-1007

Europe Economics (2007). Supporting paper II: Cost of capital – Analysis of responses to CAA's initial proposals

Europe Economics (2011). Estimating BT's cost of capital: A report from Europe Economics for TalkTalk and BSkyB

Europe Economics (2017). PR19 – Initial assessment of the cost of capital

Henry, O. (2008). Econometric advice and beta estimation.

Jenkinson, T. (2006). Regulation and the Cost of Capital. In Crew and Parker *International Handbook of Economic Regulation*, 146-163, Edward Elgar publishing.

Jorgenson and Landau (1993). Tax reform and the cost of capital: An international comparison, Brookings Institute

King, M. and Fullerton, D. (1984). The taxation of income from capital. University of Chicago Press

Monopolies and Mergers Commission (1993). Gas and British Gas plc: Volume 2 of reports under the Gas and Fair Trading Acts, available to download from the national archives

Morin, R. (2006). New Regulatory Finance. Public Utilities Reports, Inc.

NERA (2018) Cost of Equity for RIIO-2: A report for Western Power Distribution

Ofcom (2011). WBA Charge Control: Charge control framework for WBA market 1 services

Ofgem (2004a). Electricity Distribution Price Control Review: Background information on the cost of capital

Ofgem (2004b). Electricity Distribution Price Control Review: Final proposals

OXERA (1994), Quantifying Political Risk, *Utility Finance,* April 1994

OXERA (2018a). *The Cost of Equity for RIIO-2: A review of the evidence*, a report prepared for the ENA, February 2018

OXERA (2018b). *Estimating Cost of Capital in the Presence of Noise*. OXERA (2018b), www.oxera.com/wp-content/uploads/2018/07/Estimating-cost-of-capital-in-the-presence-of-noise.pdf.pdf.

Partington, G., & Satchell, S. (2017). REPORT TO THE AER: DISCUSSION OF SUBMISSIONS ON THE COST OF EQUITY.

Pratt, S., & Grabowski, R. (2014). Cost of Capital: Applications and Examples, 5th edition, Wiley

Sudarsanam, P. Sudi (1995). *The essence of mergers and acquisitions*. Pearson PTR.

Sudarsanam, Sudi (2003). *Creating value from mergers and acquisitions: The challenges: An integrated and international perspective*. Pearson Education.

UKRN (2018). Cost of Capital – Annual Update Report.

Vasicek, Oldrich A. "A Note on Using Cross-Sectional Information in Bayesian Estimation of Security Betas." Journal of Finance, Vol. 28 (12 1973), pp. 1233–1239

White, H. (1980). A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. *Econometrica*, 48, 817-838

Wright, S., Burns, P., Mason, R., & Pickford, D. (2018). Estimating the cost of capital for implementation of price controls by UK Regulators. Prepared for UKRN

https://www.ofcom.org.uk/__data/assets/pdf_file/0026/53549/disaggregating.pdf. Web. 2005

https://www.ofcom.org.uk/__data/assets/pdf_file/0020/112457/Annex-16-NERA-Report-The-Evidence-for-Differences-in-Risk-for-Fixed-vs-Mobile-Teleco.pdf. Web. 2017

https://www.ofcom.org.uk/__data/assets/pdf_file/0028/97039/annex_31.pdf. Web. 2016

https://www.ofcom.org.uk/__data/assets/pdf_file/0022/112459/MCT-review-statement-annexes-115.pdf

https://www.ofcom.org.uk/__data/assets/pdf_file/0035/99638/Annexes1-19.pdf.

https://www.ofcom.org.uk/__data/assets/pdf_file/0032/54977/final-annexes-29-30.pdf.

https://www.ofcom.org.uk/__data/assets/pdf_file/0012/50520/annexes1-17.pdf.

Ofgem

Beta Study – RIIO-2

Appendices A – C

Final

December 2018

Indepen Limited, 4th Floor, 1 Fore Street, London, EC2Y 9DT
www.indepen.uk.com

Indepen is a management consultancy working with clients facing the challenges of regulation, deregulation, competition and restructuring. We help investors, boards and senior managers identify and assess political and regulatory risk and to develop and implement internal and external strategies to manage their exposure.

Our clients are the organisations involved in financing, constructing, managing and regulating built and natural infrastructure – water, energy, transport, land and property. We have constructive relationships with relevant government departments and agencies.

Our team combines experience of public policy, regulation, corporate finance, communication and engagement and organisational development. We complement this with input from our associates – CEOs and chairs of FTSE and privately owned companies, regulators, government ministers and academics.

The Indepen Forum provides the opportunity for investors, government and business leaders to debate, under the Chatham House Rule, issues that if mishandled could undermine well-intentioned policy initiatives.

Table of Contents

Appendix A Data and structural stability

A.1 Introduction

This appendix describes the data employed and looks at evidence for structural breaks in the CAPM relationship as estimated by ordinary least squares (OLS) regression. Structural breaks will be considered solely from a statistical viewpoint rather than tested for as responses to known real-world events.

A.2 Data

All data were provided by Ofgem from Bloomberg sources: 2 indices and 26 stocks. The requested data window was 1 January 1987 to 24 September 2018. Where data were sparse or missing – common in earlier years – a decision was taken as to whether to fill with the prior value (in the case of gaps of just a few days) or to start the effective data series once major gaps had ceased. Returns were calculated as day-to-day percentage changes (expressed as numbers rather than decimals). Given that, on a daily basis, the risk-free return is negligible the analyses in this paper were undertaken using just individual stock returns and those on the market indices. Data availability is summarised in Table . In this appendix we use only data on the All Share Index and the six utilities (BT, NG, UU, SSE, PNN and SVT).

Table A1 - Data employed

Ticker	*Company*	*Sector*	*Returns Effective Start*	*Returns End Date*
ASX	FTSE All Share Index	Market index	05/01/87	24/09/18
UKX	FTSE 100	Market index	05/01/87	24/09/18
BT	BT Group PLC	Utility	05/01/87	24/09/18
NG	National Grid PLC	Utility	23/11/95	24/09/18
UU	United Utilities Group PLC	Utility	20/07/90	24/09/18
SSE	SSE PLC	Utility	23/09/91	24/09/18
PNN	Pennon Group PLC	Utility	24/07/90	24/09/18
SVT	Severn Trent PLC	Utility	12/07/91	24/09/18
BBY	Balfour Beatty PLC	Other	13/09/88	24/09/18
SMDS	Smith (DS) PLC	Other	15/01/91	24/09/18
WEIR	Weir Group PLC	Other	17/01/91	24/09/18
NEX	National Express Group PLC	Other	27/04/95	24/09/18
SRP	Serco Group PLC	Other	26/03/91	24/09/18

Ticker	Company	Sector	Returns Effective Start	Returns End Date
ATK	WS Atkins PLC	Other	26/07/96	30/06/17[1]
BVS	Bovis Homes Group PLC	Other	10/12/97	24/09/18
BKG	Berkeley Group Holdings PLC	Other	21/06/90	24/09/18
GRG	Greggs PLC	Other	13/05/92	24/09/18
TATE	Tate & Lyle PLC	Other	13/09/88	24/09/18
FAS	Fidelity Asian Values PLC	Other	14/06/96	24/09/18
AAS	Aberdeen Asian Smaller Companies Investment Trust PLC	Other	20/10/95	24/09/18
VCT	Victrex PLC	Other	22/12/95	24/09/18
ANII	Aberdeen New India Investment Trust PLC	Other	05/04/94	24/09/18
CPI	Capita PLC	Other	02/03/92	24/09/18
CNCT	Connect Group PLC	Other	11/08/06	24/09/18
NTG	Northgate PLC	Other	27/02/92	24/09/18
BRLA	Blackrock Latin American Investment Trust PLC	Other	01/03/91	24/09/18
SKY	Sky PLC	Other	09/12/94	24/09/18
CTY	City of London Investment Trust PLC	Other	15/01/91	24/09/18

As is to be expected from financial returns data, the series almost always appear to be mean-stable but with apparent clusterings of volatility – that is, typical GARCH behaviour (Engle, R. F. 1982). Returns are shown in Figure to Figure (only utilities and the All-share index shown – other stocks available upon request).[2]

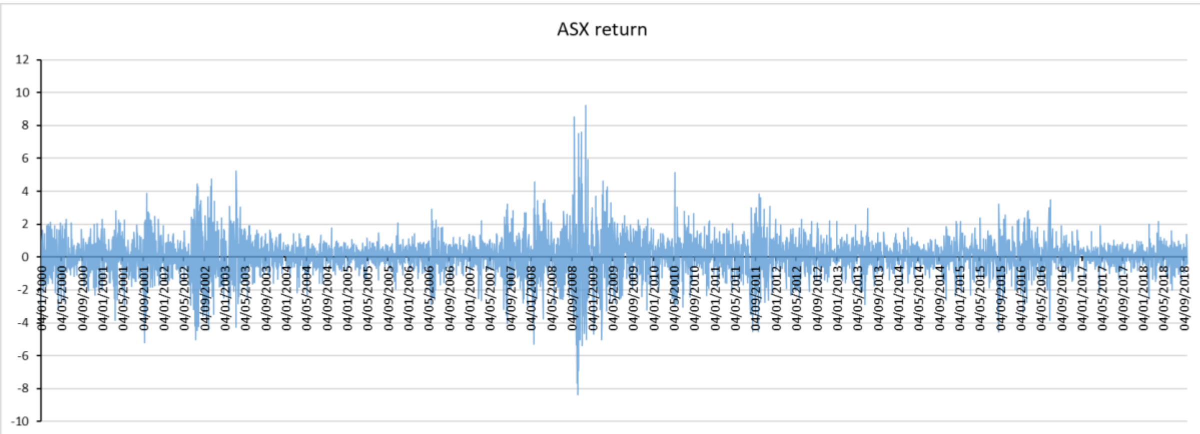


Figure A1 - ASX Daily Returns

[1] Purchased by SNC-Lavalin

[2] Annex 1 to this appendix provides further information on the All-share Index.

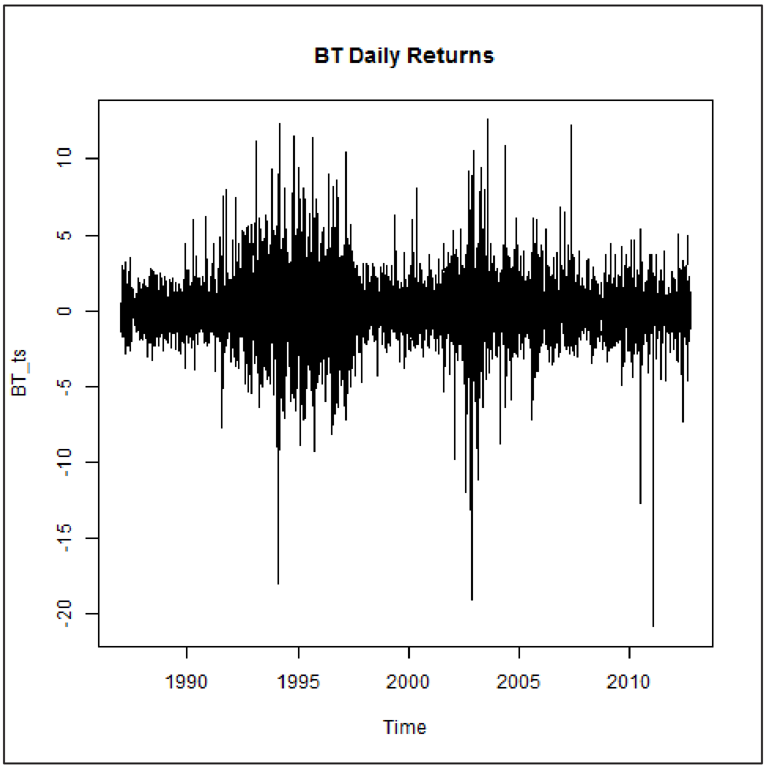


Figure A2 - BT Daily Returns

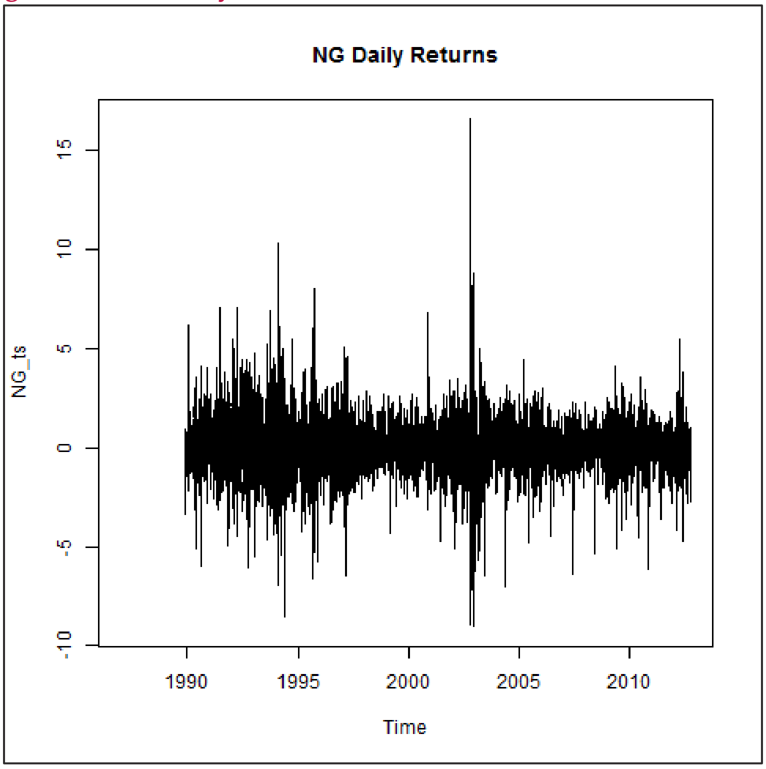


Figure A3 - NG Daily Returns

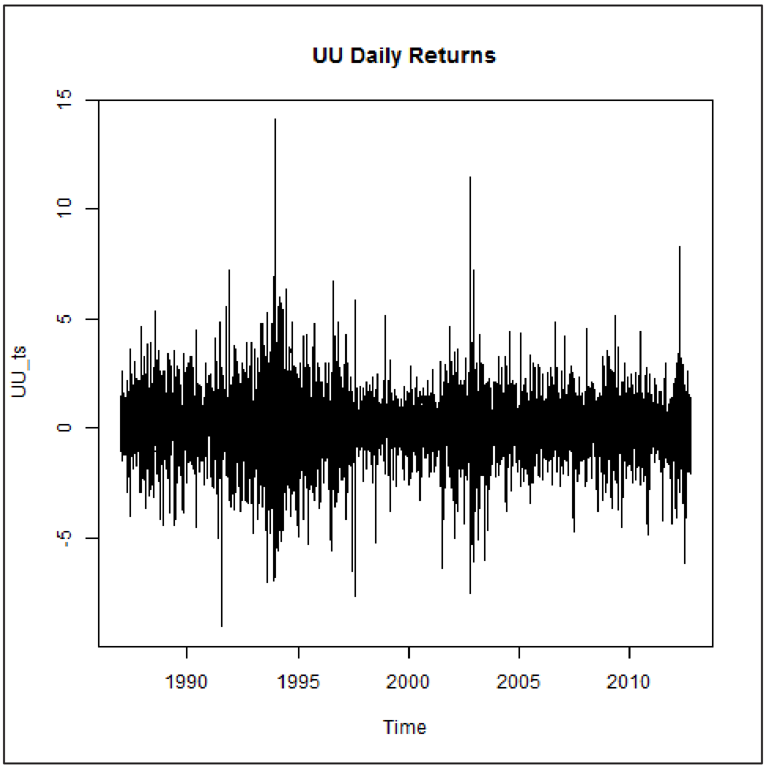


Figure A4 - UU Daily Returns

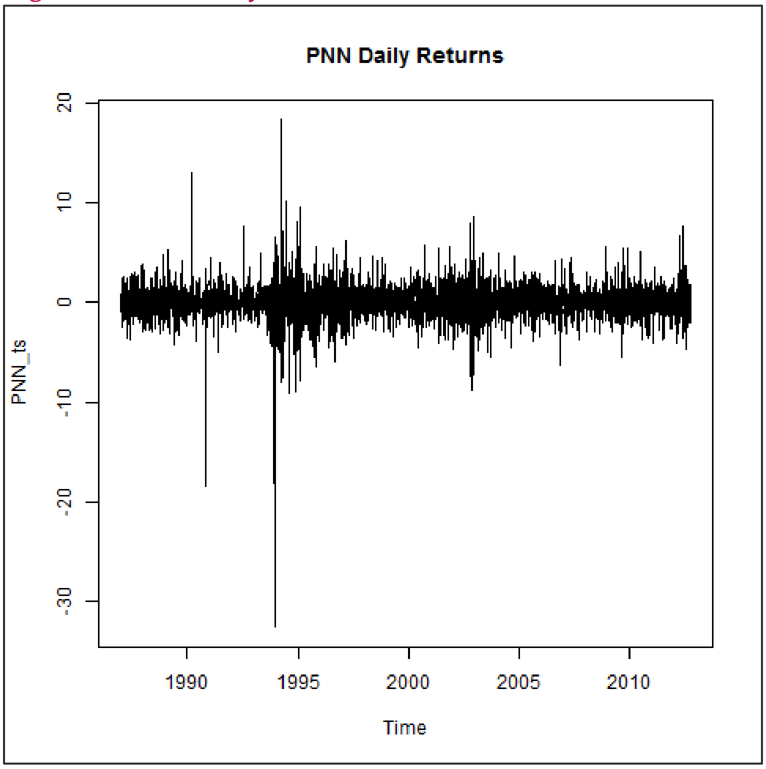


Figure A5 - PNN Daily Returns

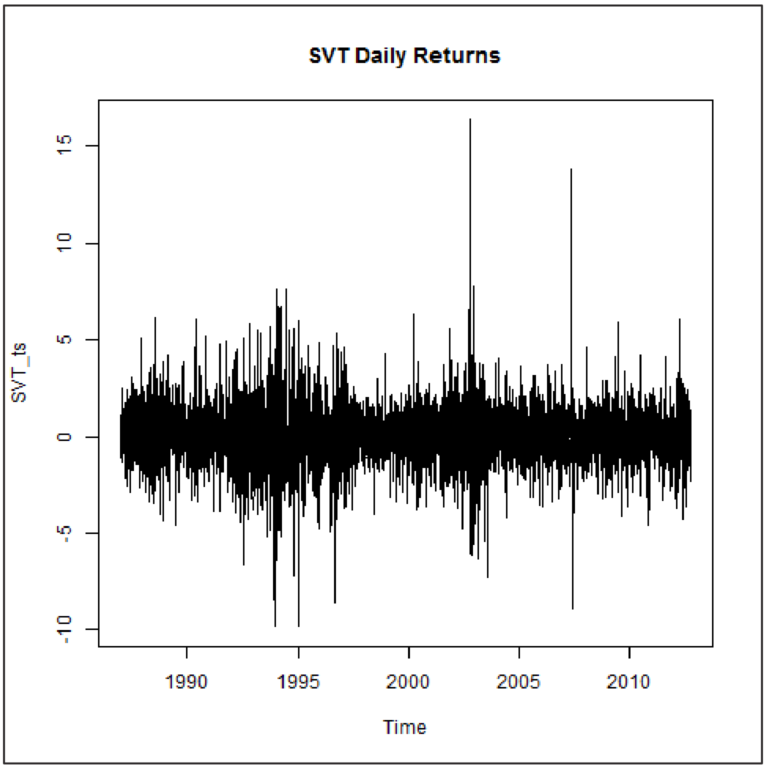


Figure A6 - SVT Daily Returns

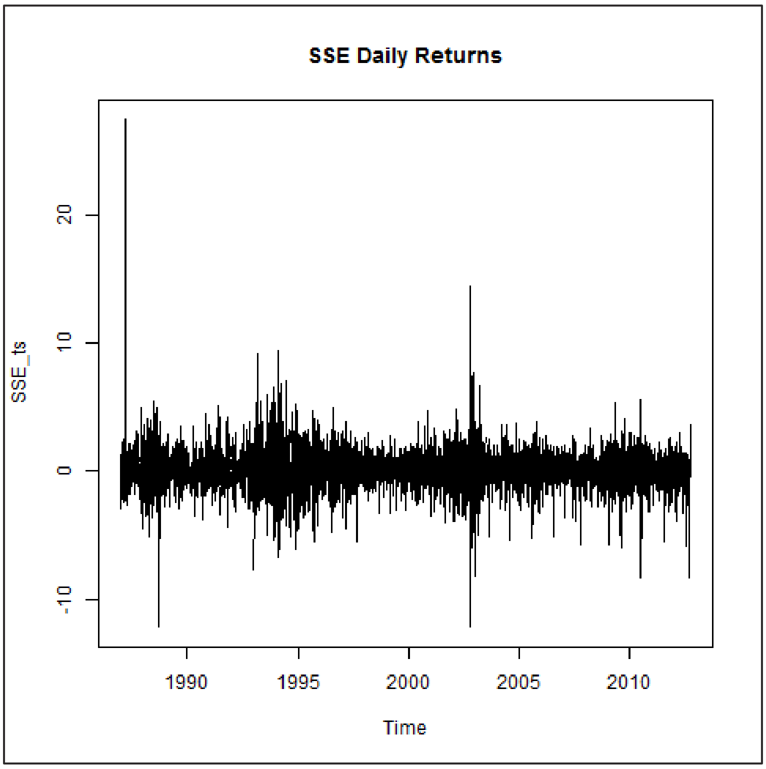


Figure A7 - SSE Daily Returns

A.3 Frequencies

Subsequent analyses – described later – used returns data at three frequencies:

- Daily
- Weekly: data series were created for Monday-to-Monday return, Tuesday-to-Tuesday return and so forth
- Monthly: two types of data series were created – first-trading-day to first-trading-day and then 20 data sets using returns between first and twenty-first trading days, second and twenty-second trading days and so forth

A.4 Structural stability

Methodology

Analysis was conducted in R version 3.4.4 (64-bit) (R Core Team 2018). The primary packages used were *lmtest* (Zeileis, A. and T. Hothorn 2002), *strucchange* (Zeileis, A., F. Leisch, K. Hornik and C. Christian Kleiber 2002) and *zoo* (Zeileis, A. and G. Grothendieck 2005).

Three primary sets of tests were investigated:

- Residuals-based approaches
- Parameter-based approaches
- F-test based approaches

The residuals based approaches look at the cumulative sum of the regression residuals, which should remain close to zero if there is no structural change and diverge dramatically if there is such change. The two variants used employ, respectively, residuals from recursive model estimation (Brown, R. L., J. Durbin and J. M. Evans 1975) and the residuals from a single OLS estimation (Ploberger, W. and W. Krämer 1992). The lines shown in the charts in the Results section below are 5% critical values.

The parameter- based approaches apply a similar logic but consider the estimated parameters of the models and how they change rather than focusing on residuals. Here, again, there are two approaches. The first uses recursive estimation (Ploberger, W., W. Krämer and K. Kontrus 1989) while the second uses a moving window (in this case one fifth of the total series length) to assess stability (Chu, C.-S. J., K. Hornik and C.-M. Kuan 1995).

The F-test approaches – the classic Chow test (Chow, G. C. 1960) applied at all possible breakpoints in a defined range – consider the sums of squared residuals from regressions on the data up to the breakpoint, from the breakpoint onwards and without any breakpoint. For all series the tests were applied to the entire available data range with both market indices as independent variable while, for the six utilities and the All-Share index the F-test was also applied to the period from 2000 onwards. Additionally, each series was examined for breakpoints in its univariate return series – a regression on a constant and so a test of intercept/mean shift.

The three approaches above admit the use of formal significance tests, which have been conducted. Results almost uniformly reflect the impressions given by visual analysis of the graphs – and so are not presented in this summary report. The tests and associated p-values are, of course, available upon request.

Three further analyses were undertaken for the utilities and the All-Share index. The first is rolling regression with a 5-year time window (already implicitly undertaken as part of the parameter variation analysis described above but here the parameter values are examined explicitly). The second involves the selection of "optimal breakpoints" in the series using the Bayesian Information Criterion – a development of the Akaike Criterion that generally places a greater penalty on the number of estimated parameters (Schwarz, G. 1978). (Such criteria are to be preferred to measures like the sum of squared residuals as they explicitly penalise additional parameters – the sum of squared residuals can be driven down and down by adding more sub-models.) The "optimal breakpoint" analysis was also run for the six utilities (with the All-Share index as independent variable) for the period since 2000. Finally, the Breusch-Pagan test (Breusch, T. S. and A. R. Pagan 1979) was used as a check for the (expected) heteroscedasticity of the daily returns data[3].

A.5 Results

Here we present results only for the six utilities: all stock series showed evidence of structural breaks over the entire data period (1987-2018) and analyses for non-utilities are available upon request. They are not dissimilar to those presented here.

We do not expect radically different results from analyses using the two indices ASX and UKX as the two have a correlation of 0.984 across the entire period.

The x-axis shows the proportion of the sample period as a decimal fraction from zero to one. As almost every series has a different effective start date the x-axes need to be interpreted in the light of the start and end dates given above in Table .

Results: Residuals analysis

As can be seen from comparison of Figures A9-A11 below for BT, the results of the residuals analysis using ASX and UKX as independent variable are essentially identical. In the interests of space, we will only present the ASX equation results for the other utilities.

[3] This is essentially a test for heteroscedasticity that is a linear function of the independent variables. See below for tests specifically focused on ARCH errors

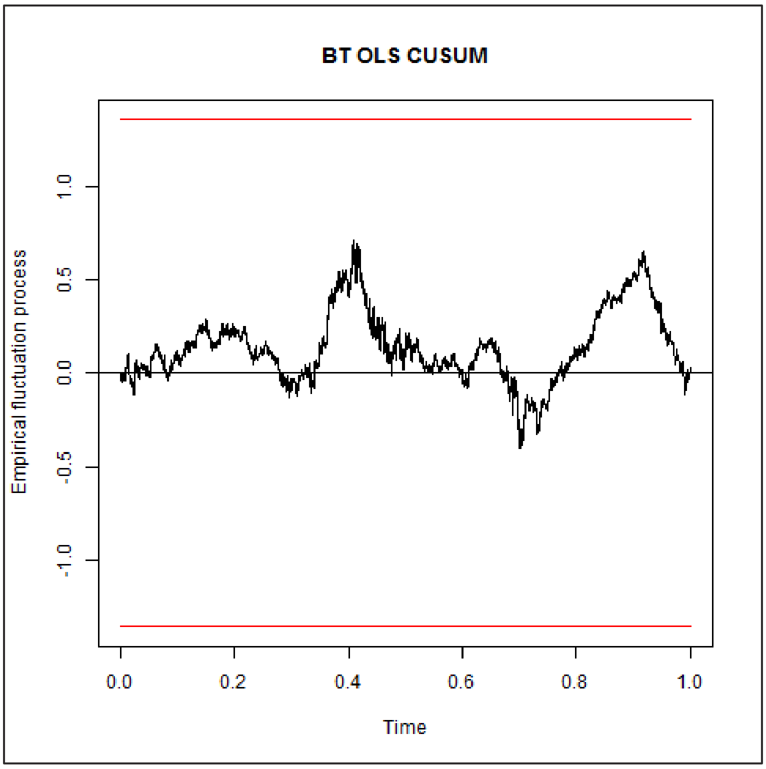


Figure A8 - BT CUSUM Analysis with ASX

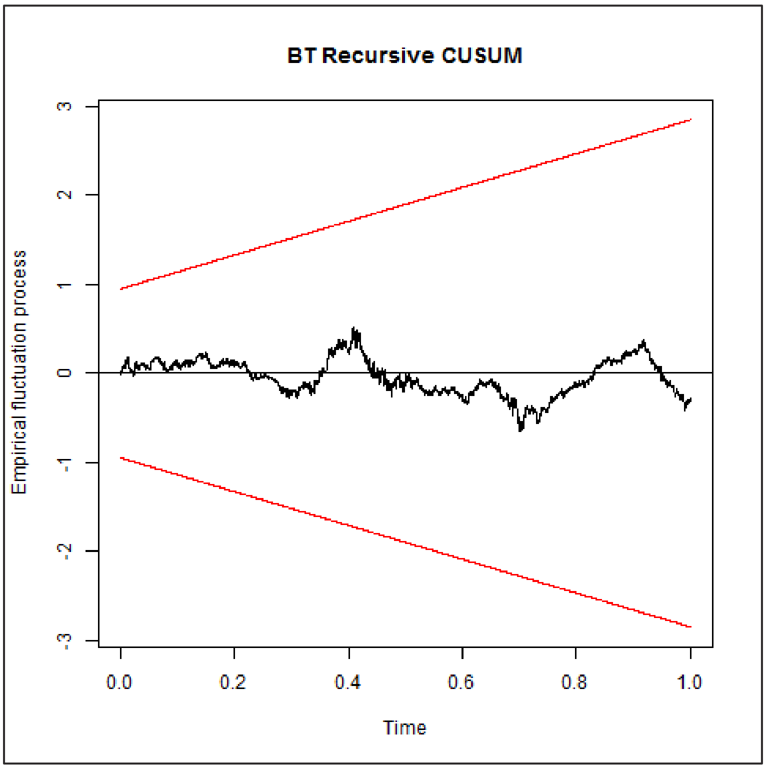


Figure A9 - BT Recursive CUSUM Analysis with ASX

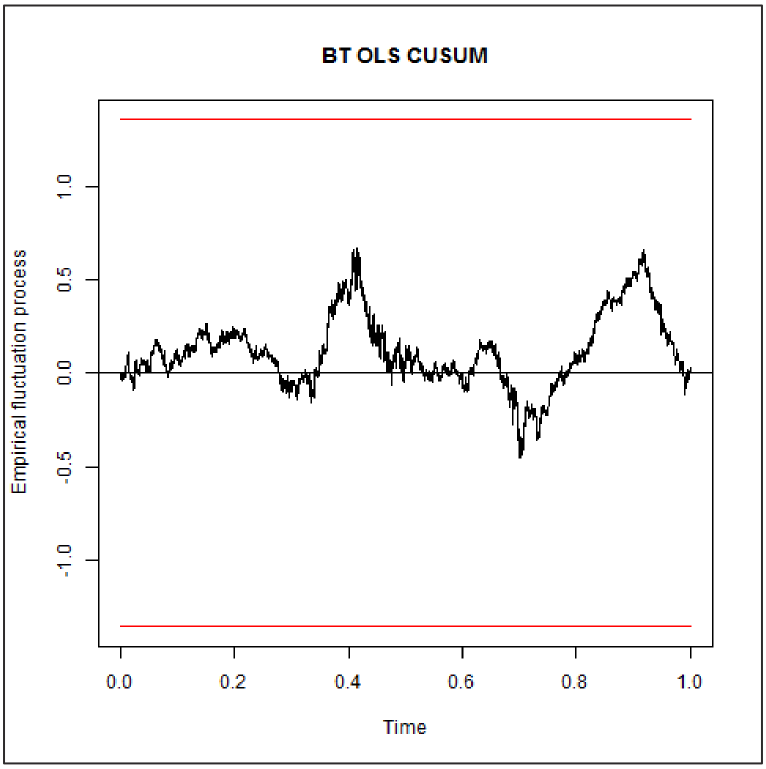


Figure A10 - BT CUSUM Analysis with UKX

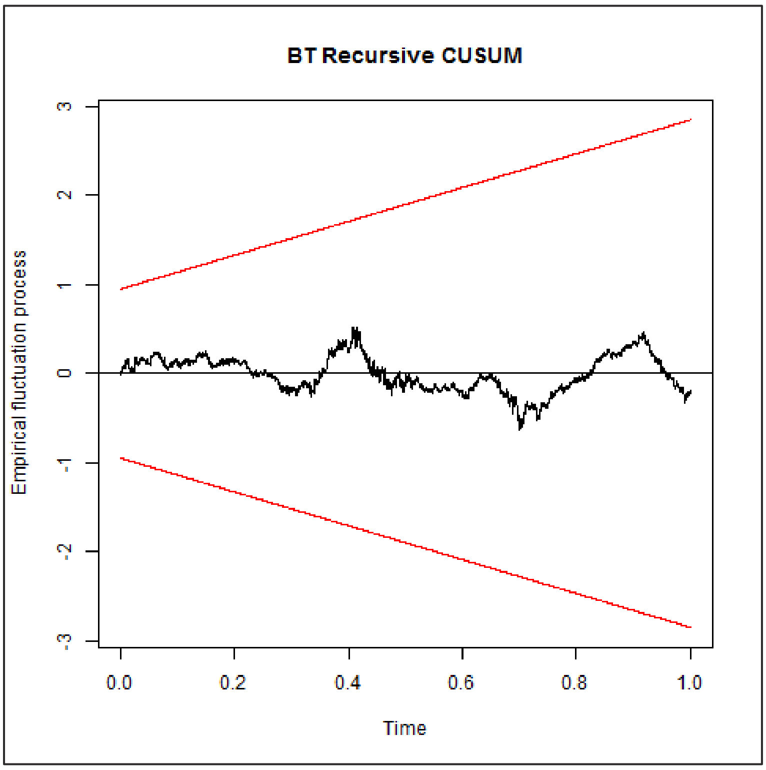


Figure A11 - BT Recursive CUSUM Analysis with UKX

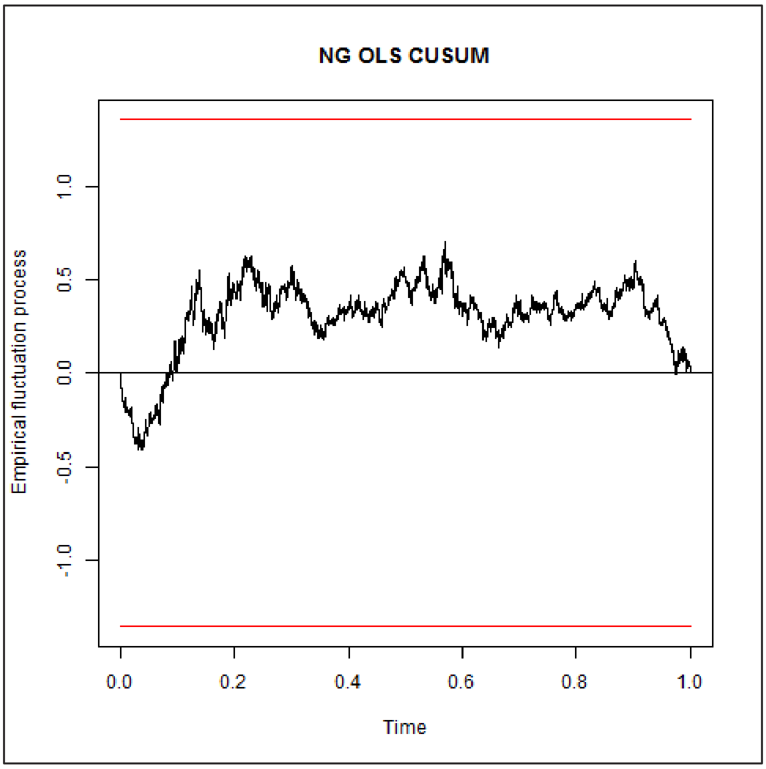


Figure A12 - NG CUSUM Analysis with ASX

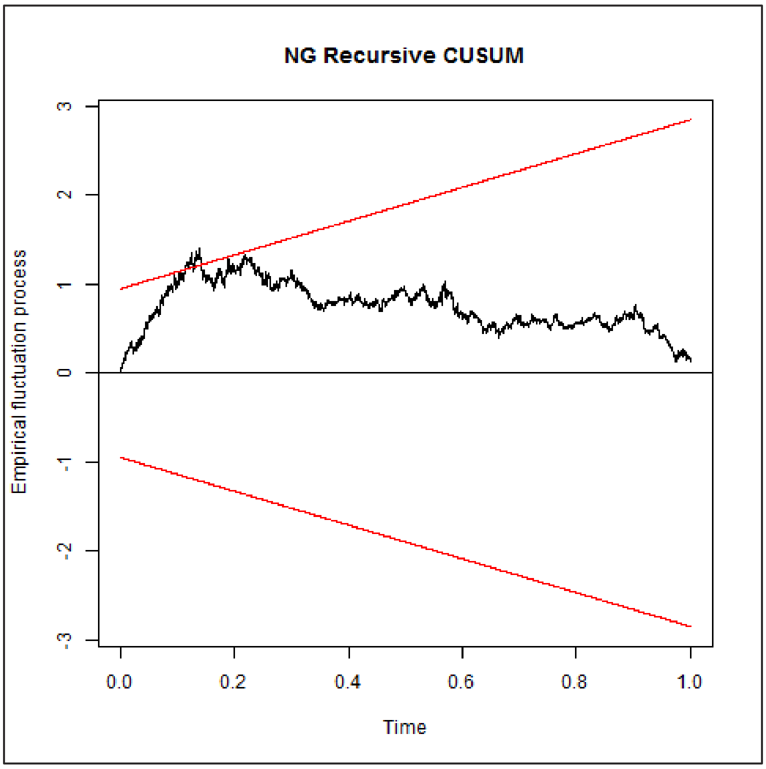


Figure A13 - NG Recursive CUSUM Analysis with ASX

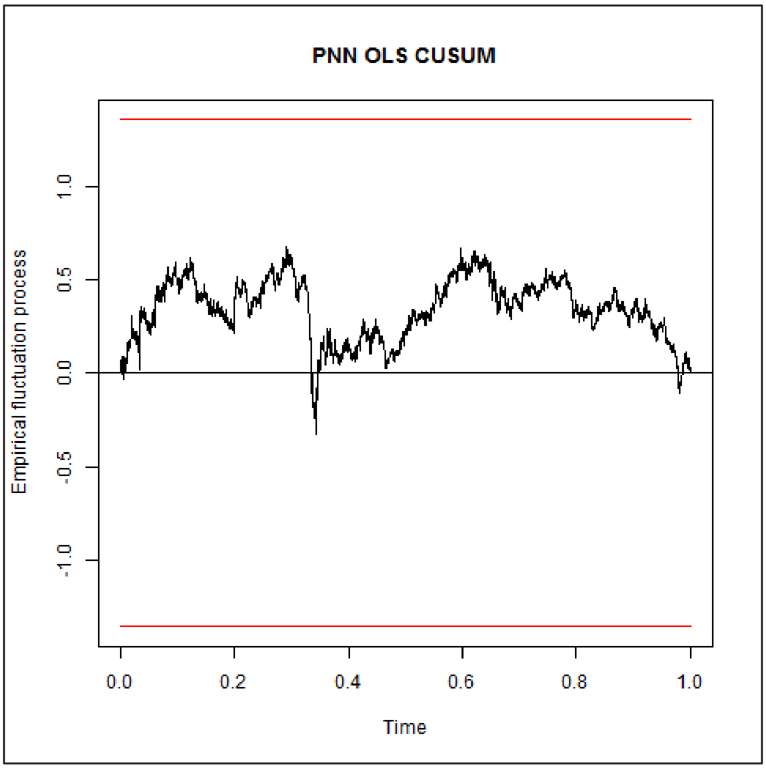


Figure A14 - PNN CUSUM Analysis with ASX

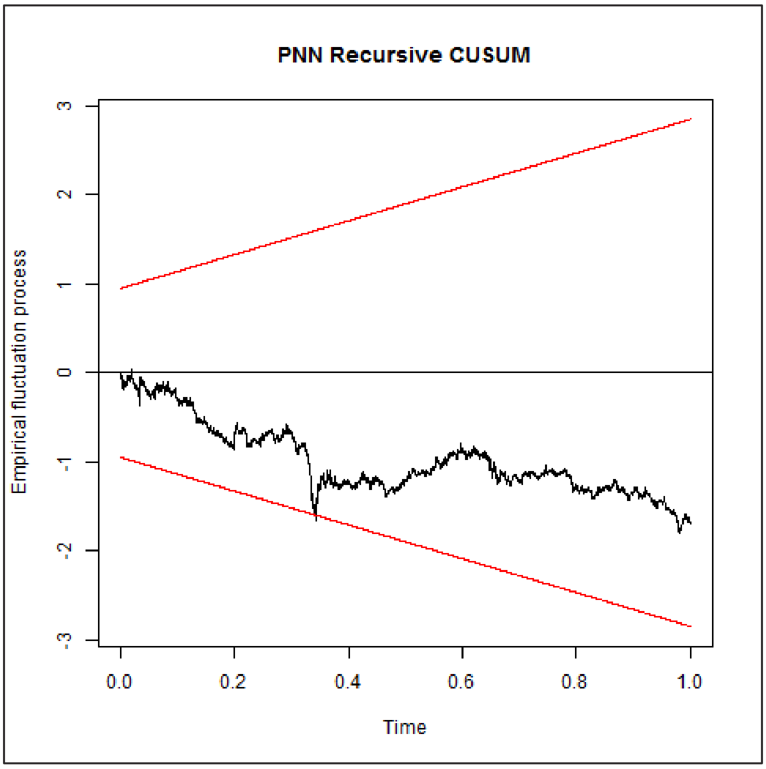


Figure A15 - PNN Recursive CUSUM Analysis with ASX

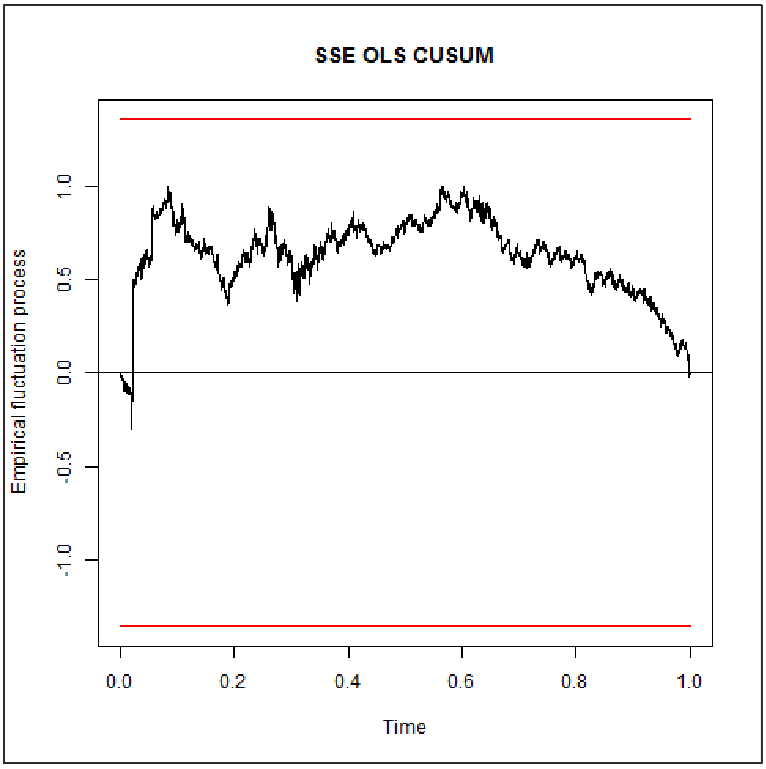


Figure A16 - SSE CUSUM Analysis with ASX

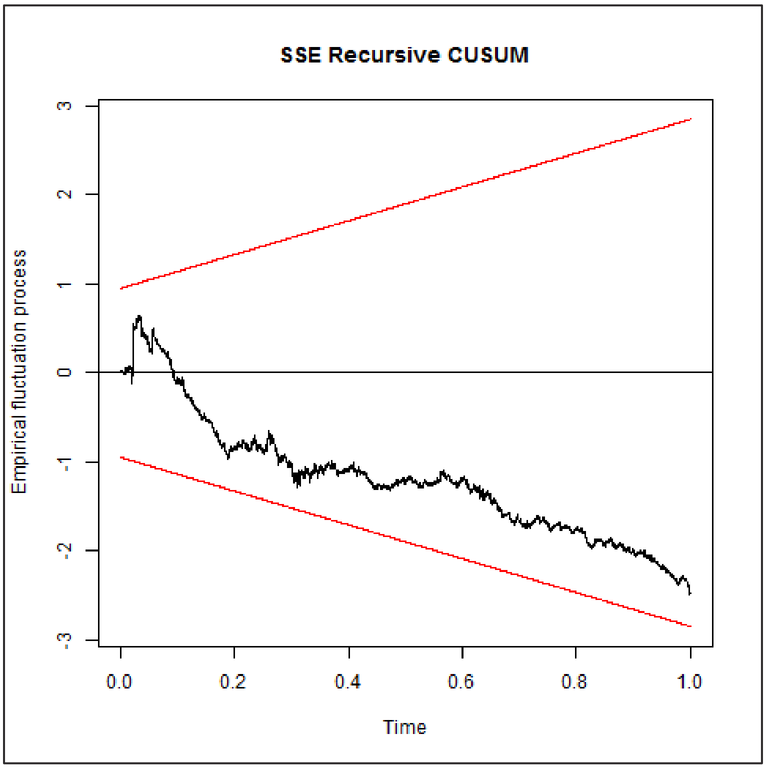


Figure A17 - SSE Recursive CUSUM Analysis

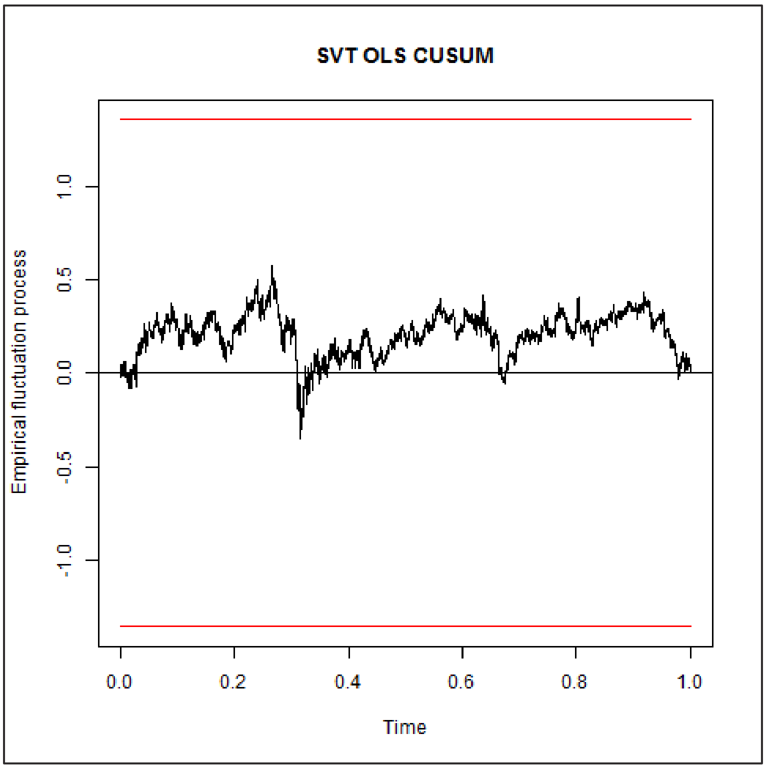


Figure A18 - SVT CUSUM Analysis with ASX

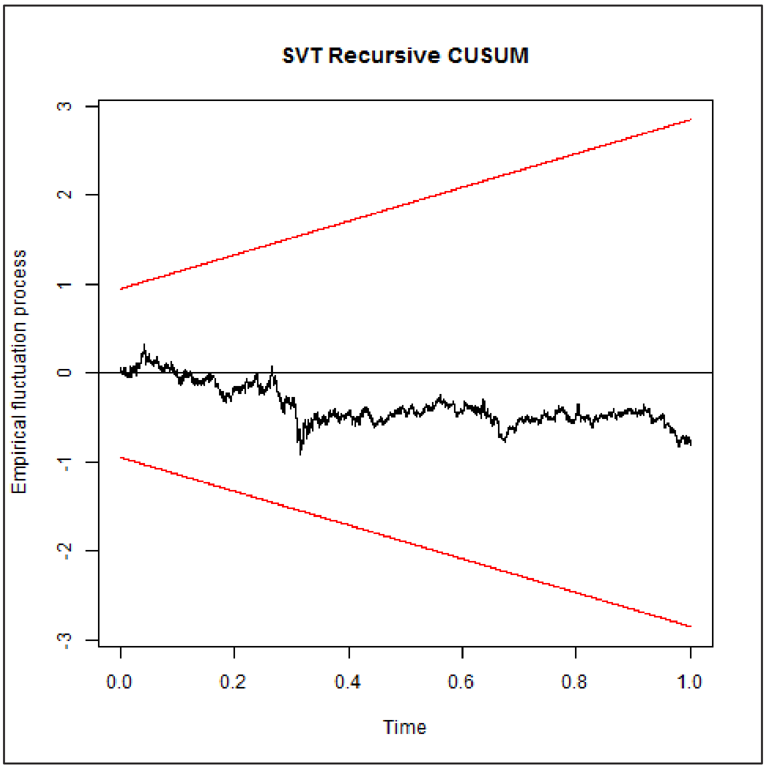


Figure A19 - SVT Recursive CUSUM Analysis

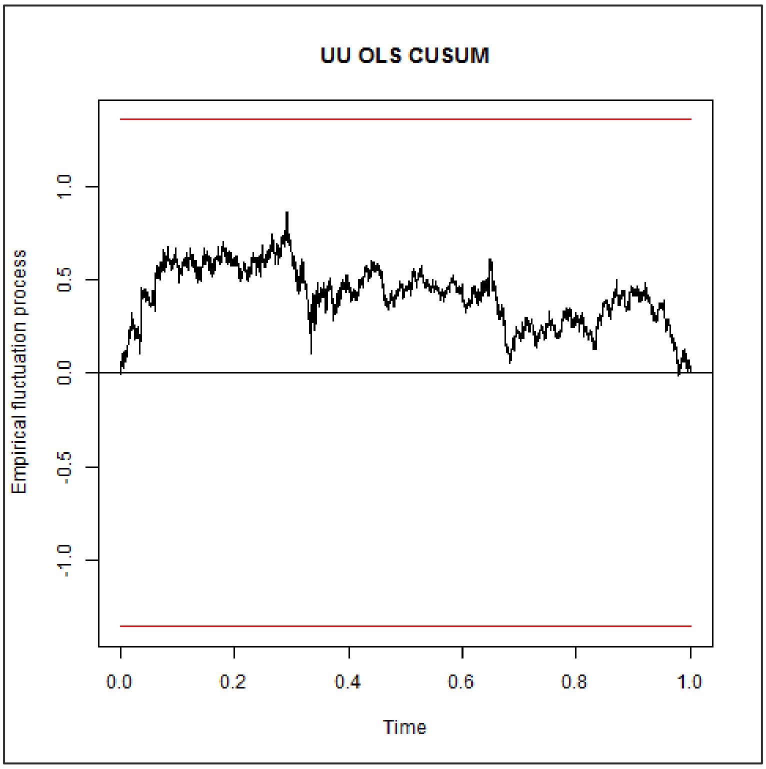


Figure A20 - UU CUSUM Analysis with ASX

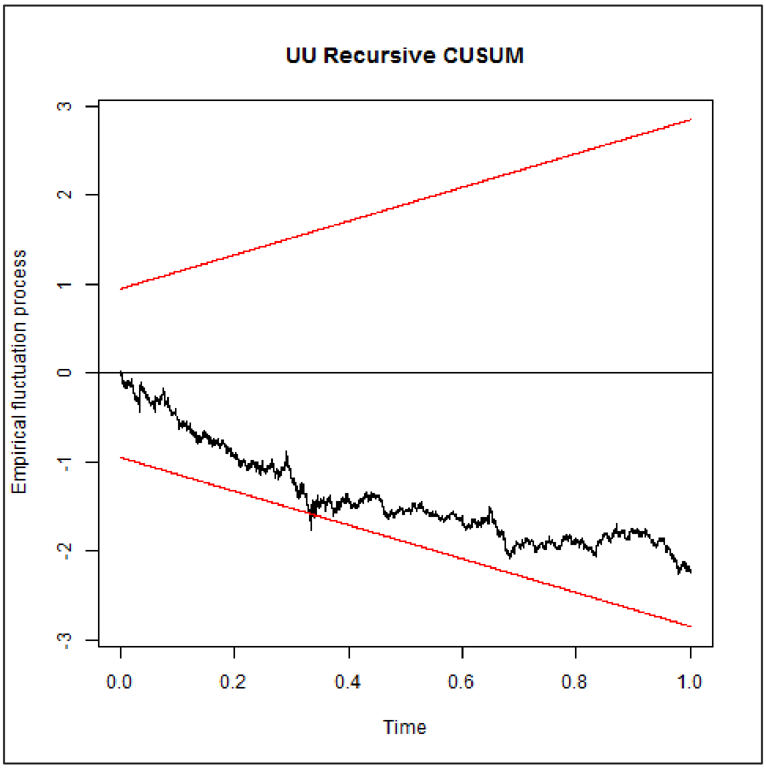


Figure A21 - UU Recursive CUSUM with ASX

One sees immediately that the CUSUM analyses suggest very few structural breaks. There is limited evidence for breaks in the following series:

- NG
- PNN
- UU

The possible breaks are only detected by the recursive method.

Results: Recursive and rolling window estimates

As above, we present only the ASX-based results although UKX-based results are also available.

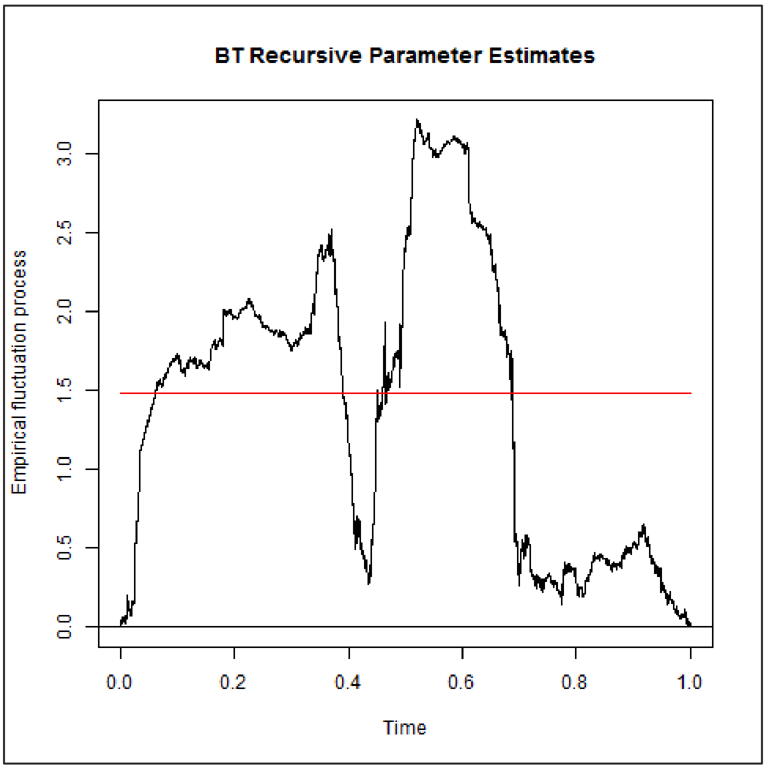


Figure A22 - BT Recursive Parameters Estimates

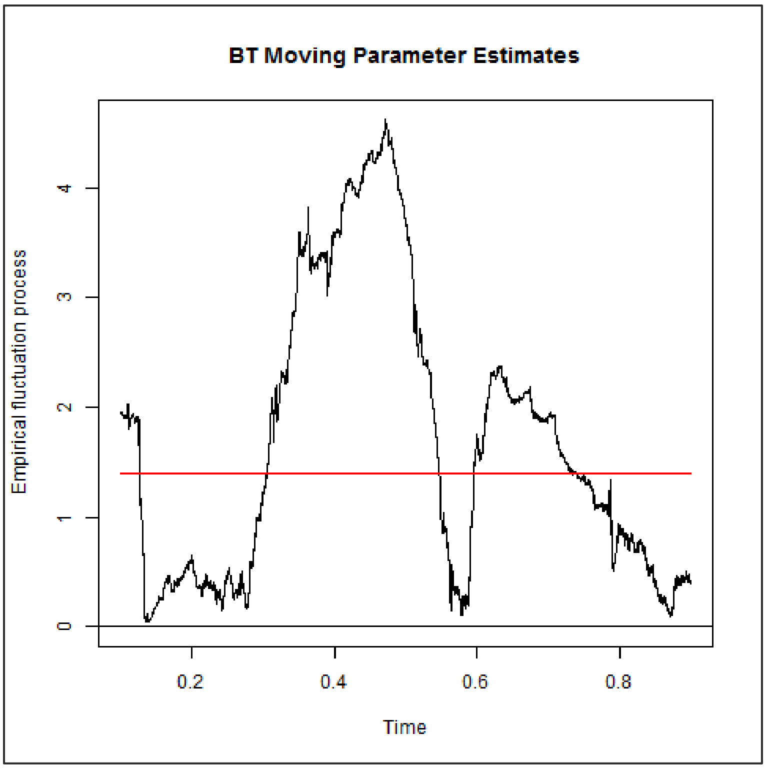


Figure A23 - BT Moving Parameters Estimates

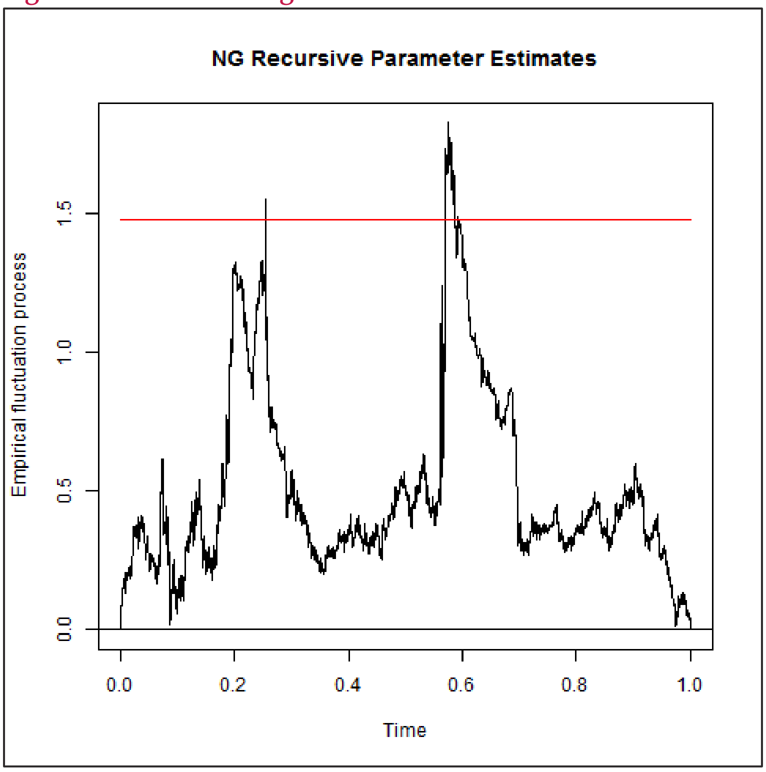


Figure A24 - NG Recursive Parameter Estimates

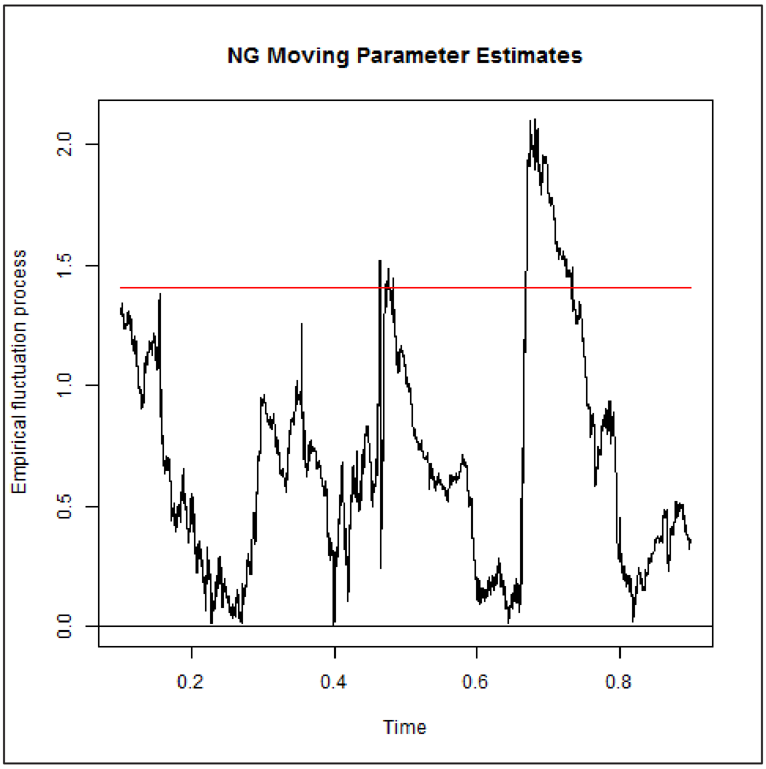


Figure 1 - NG Moving Parameters Estimates

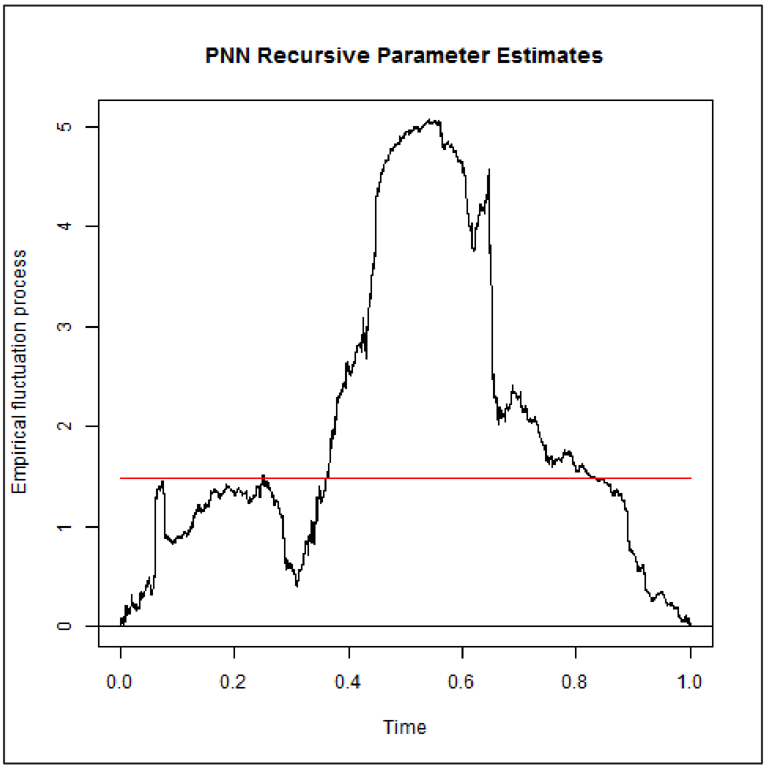


Figure A26 - PNN Recursive Parameters Estimates

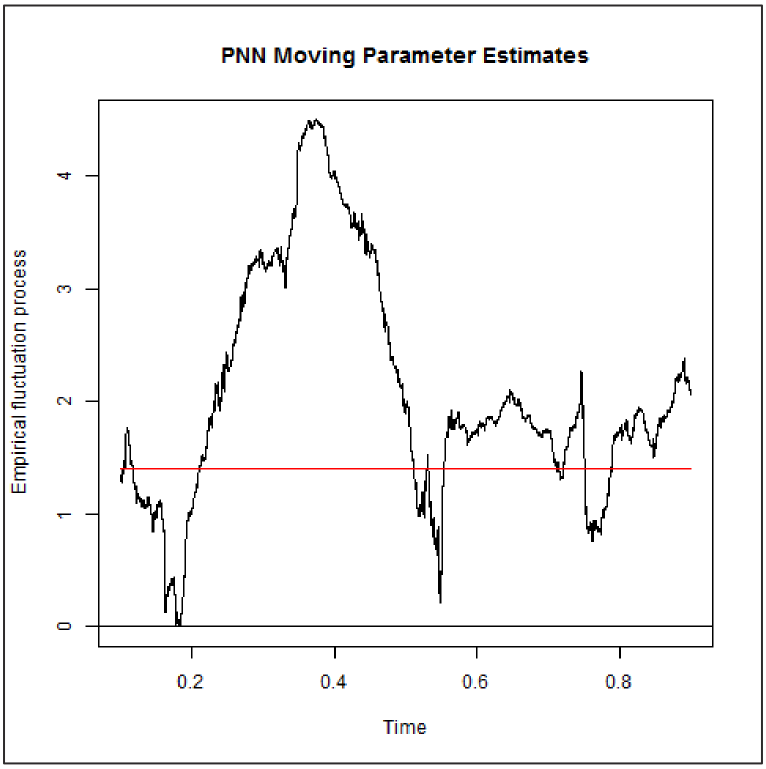


Figure A27 - PNN Moving Parameters Estimates

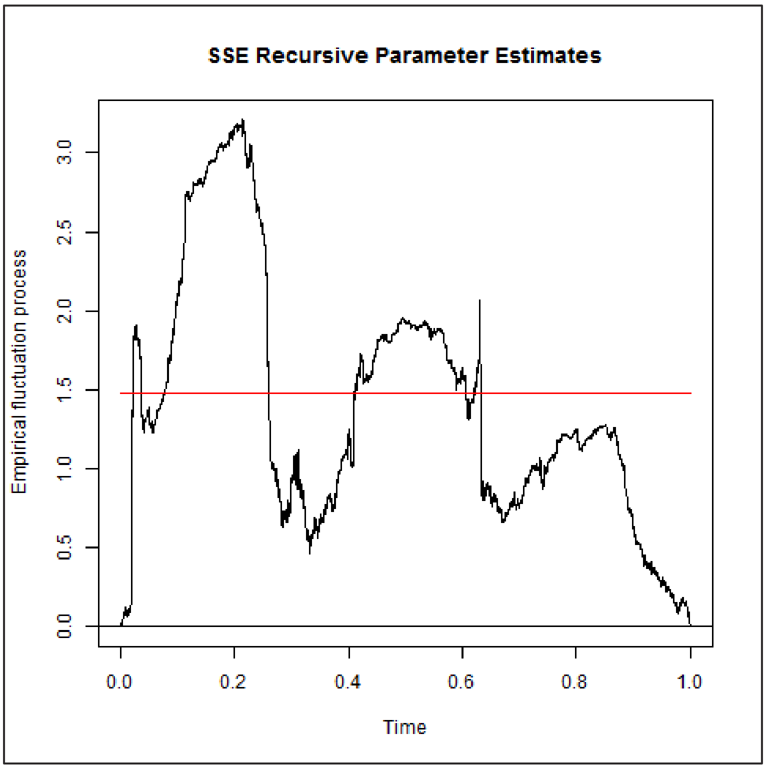


Figure A28 - SSE Recursive Parameters Estimates

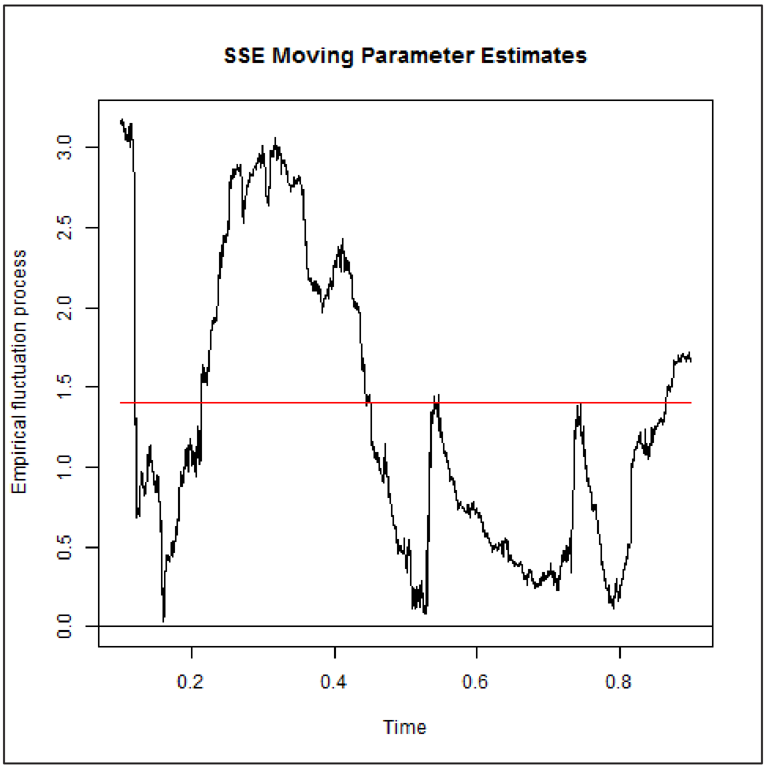


Figure A29 - SSE Moving Parameters Estimates

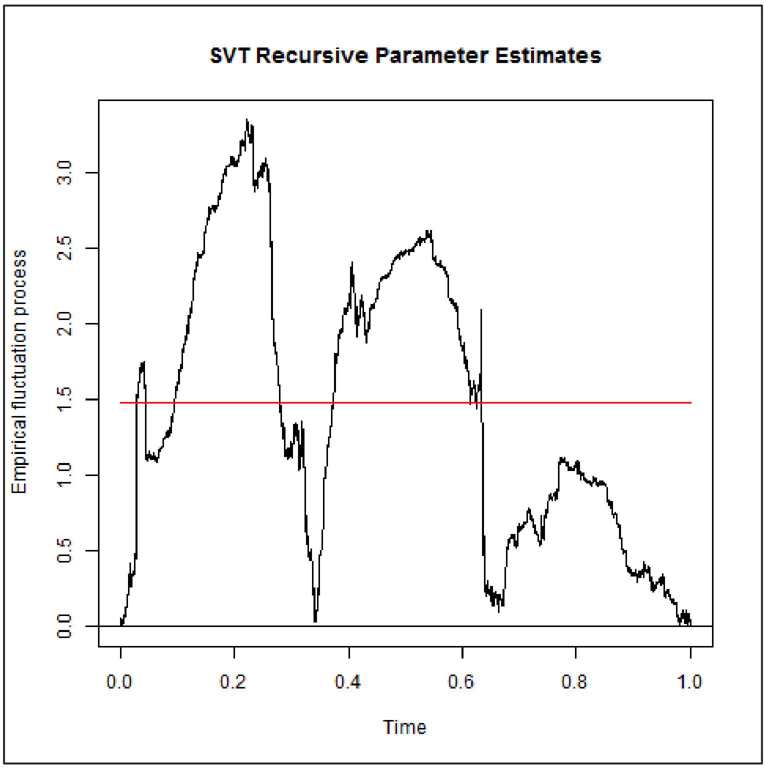


Figure A30 - SVT Recursive Parameters Estimates

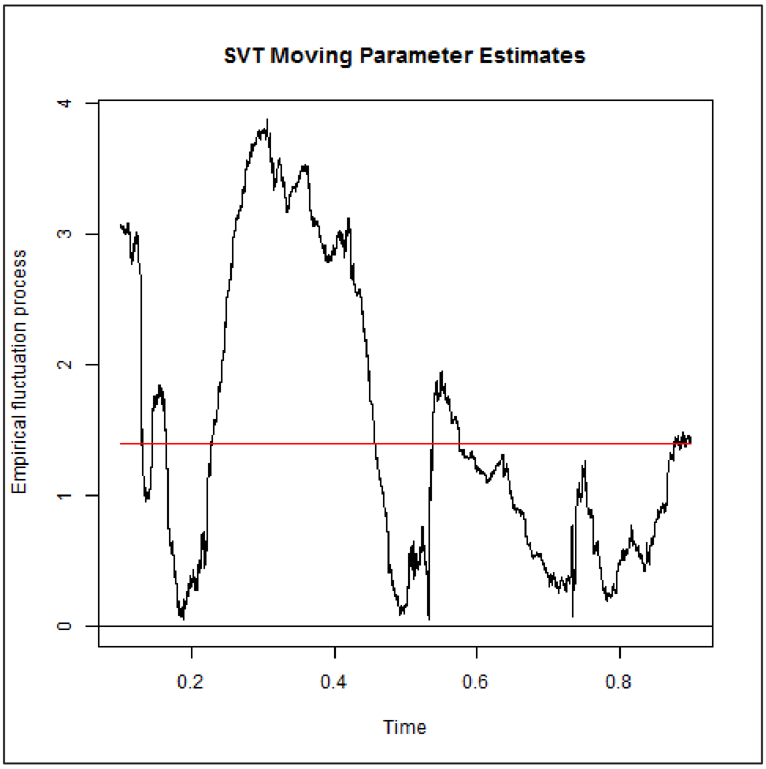


Figure A31 - SVT Moving Parameters Estimates

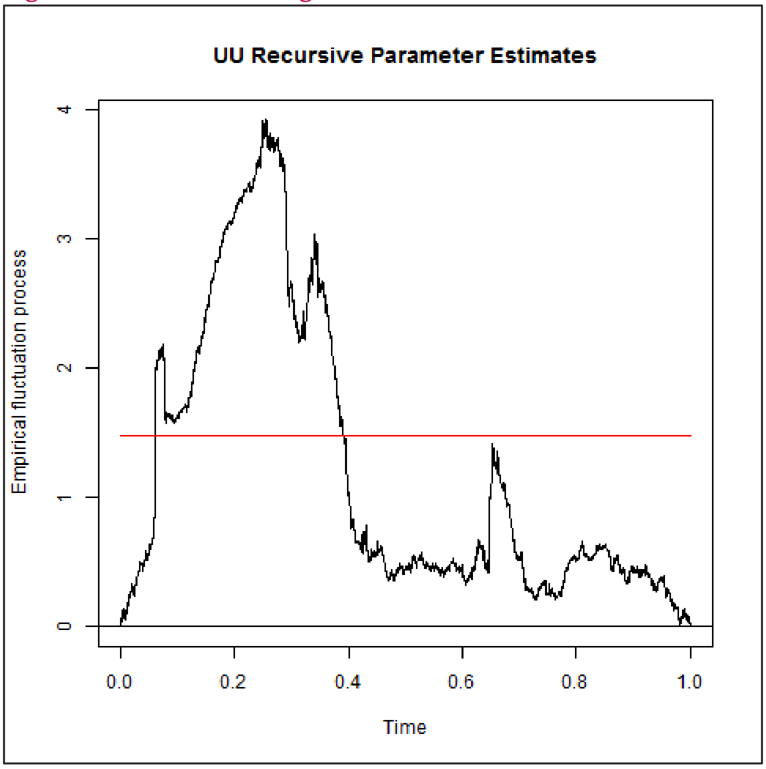


Figure A32 - UU Recursive Parameters Estimates

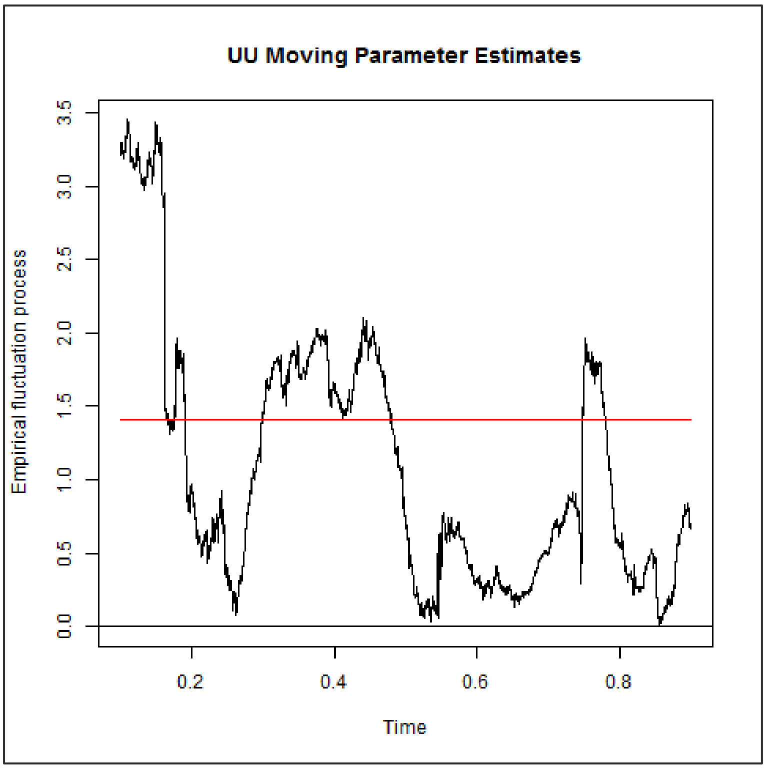


Figure A33 - UU Moving Parameters Estimates

Review of the parameters-based analyses above suggests structural breaks for all six utility stocks.

Results: F-tests

None of the univariate Chow tests (regression on constant) produced significant results for the six utilities. The only stock analysed that showed signs of a structural break in mean return was AAS. Univariate graphs will not be presented here.

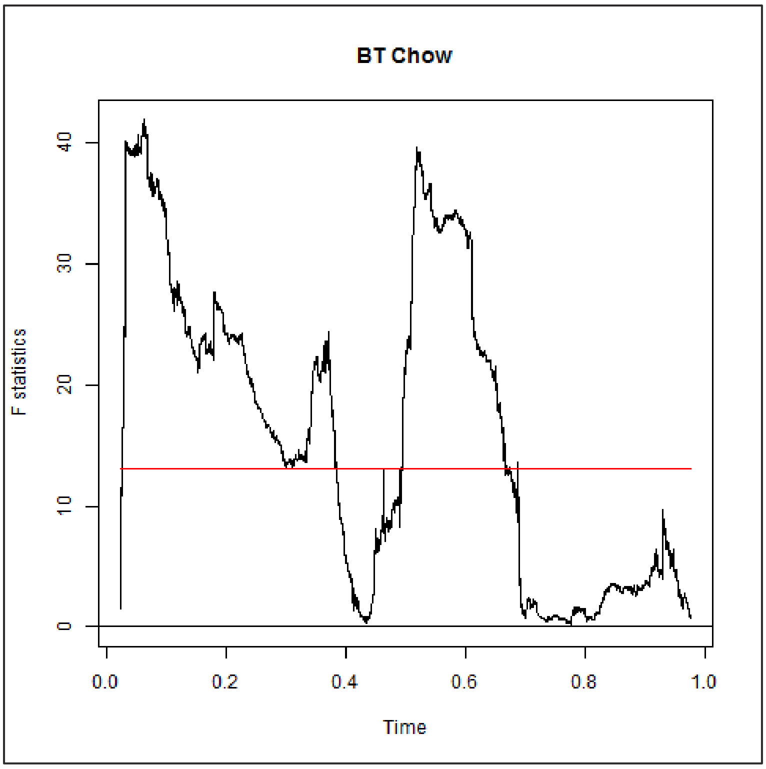


Figure A34 - BT Chow tests (whole period)

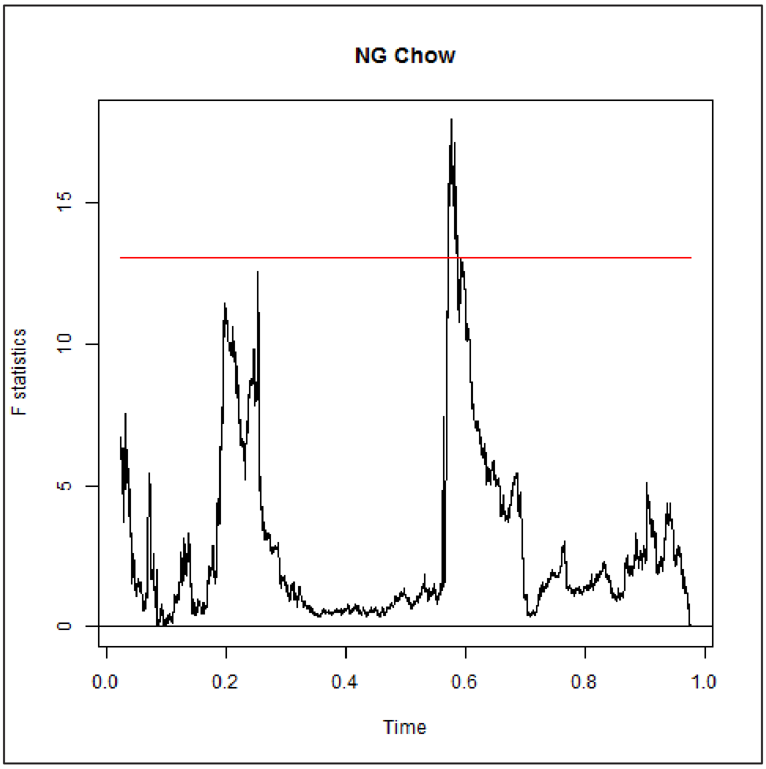


Figure A35 - NG Chow tests (whole period)

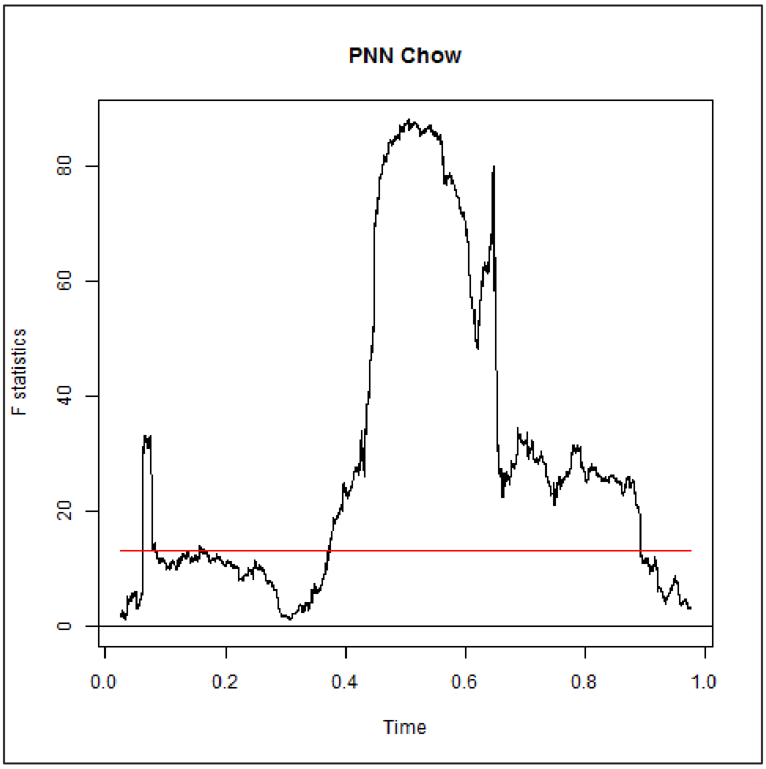


Figure A36 - PNN Chow tests (whole period)

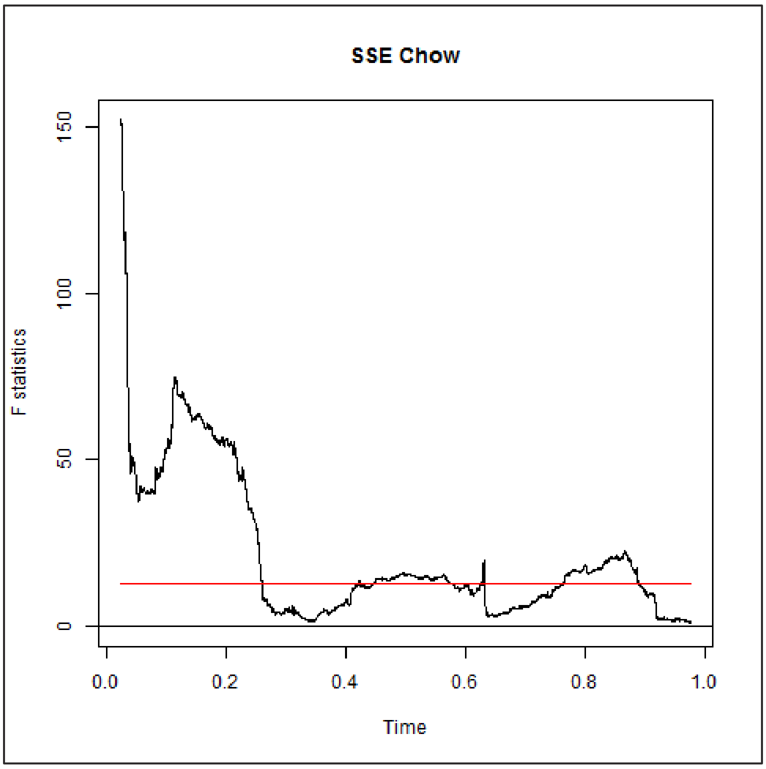


Figure A37 - SSE Chow tests (whole period)

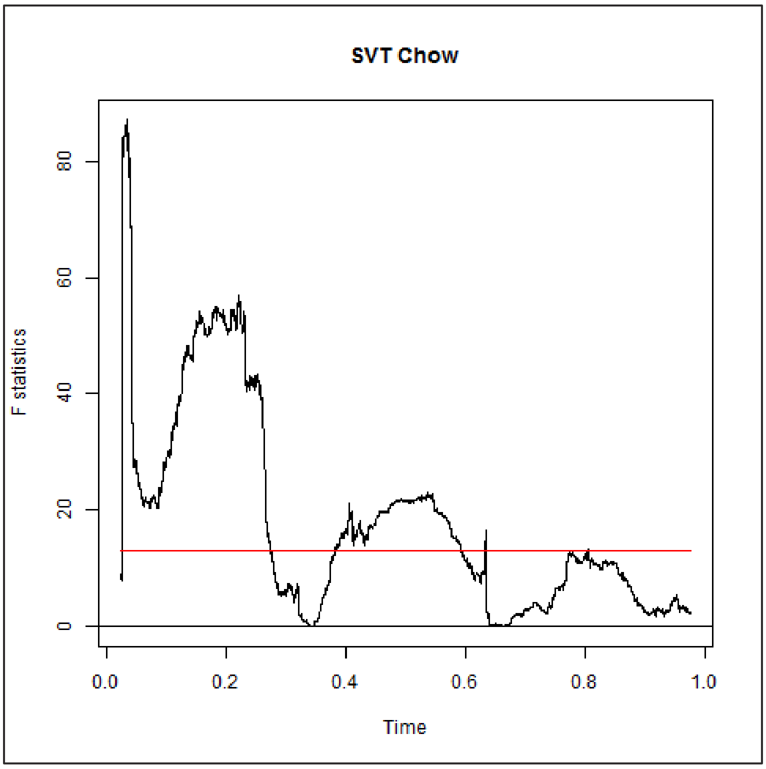


Figure A38 - SVT Chow tests (whole period)

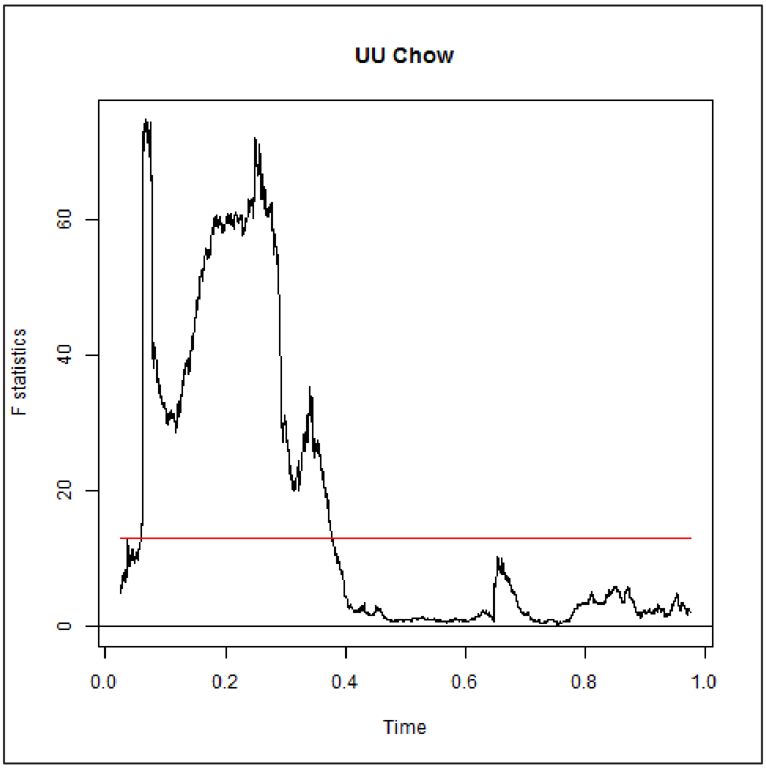


Figure A39 - UU Chow tests (whole period)

It is immediately obvious that the Chow tests suggest structural breaks – often more than one – in each utility series. NG seems the least prone to shifts in the CAPM relationship across the entire data period.

Further Chow tests were carried out on data from 2000 onwards.

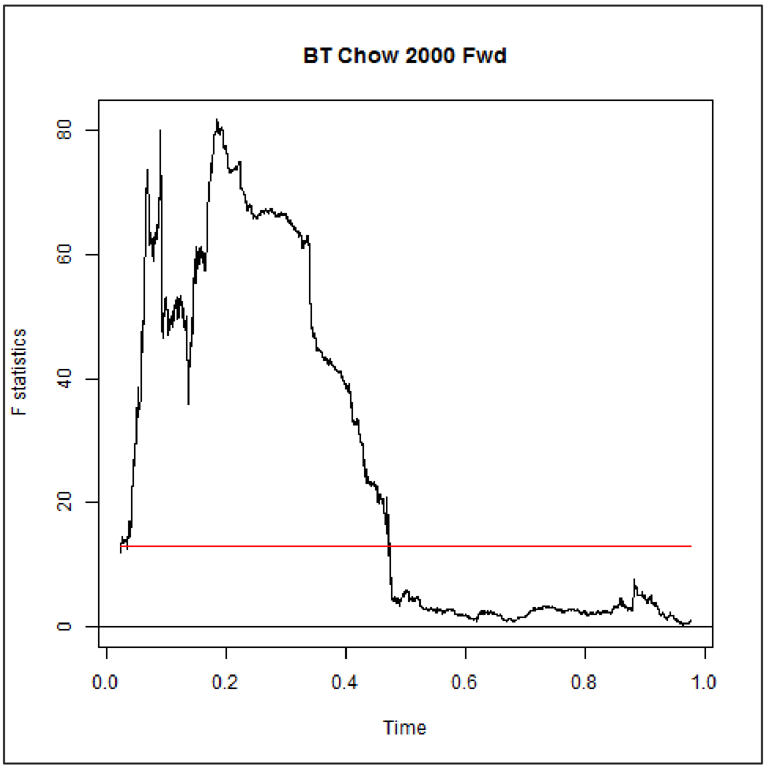


Figure A40 - BT Chow test (2000+)

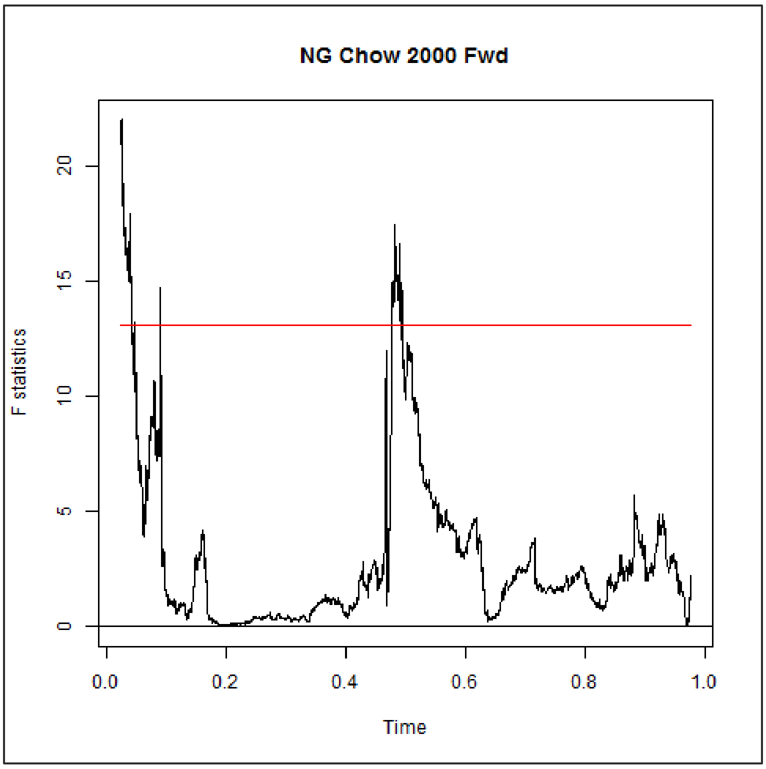


Figure A41 - NG Chow test (2000+)

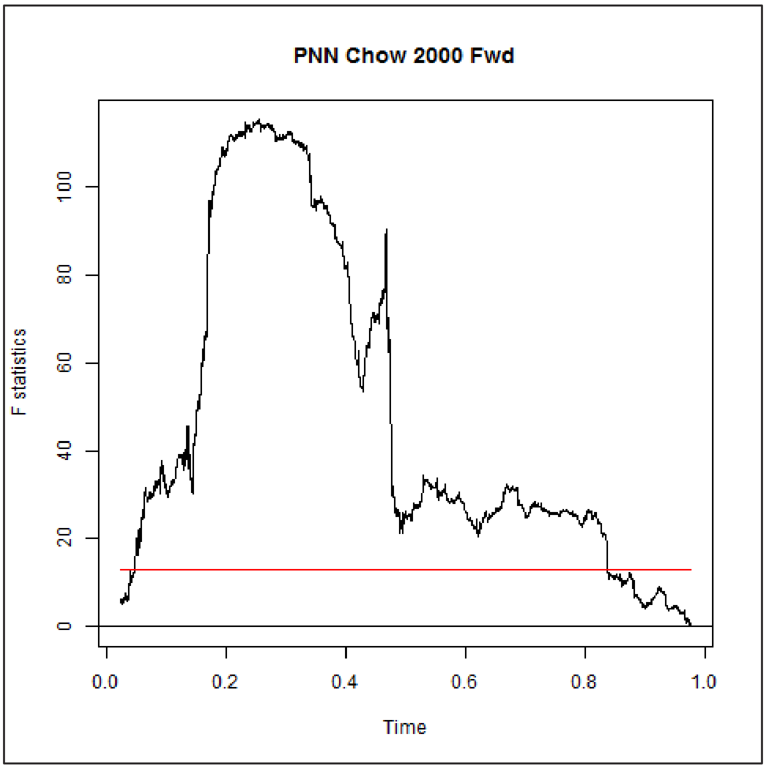


Figure A42 - PNN Chow test (2000+)

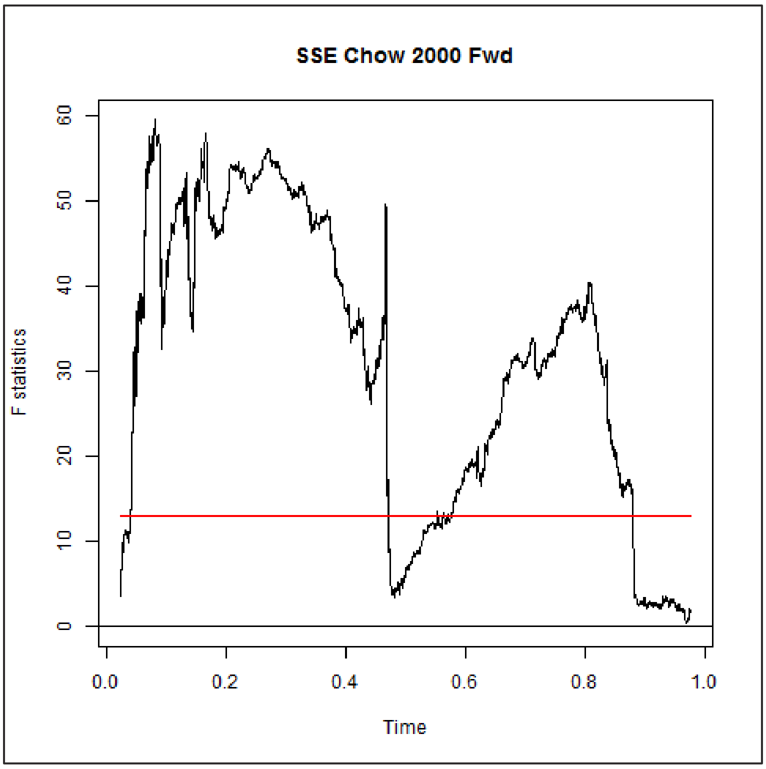


Figure A43 - SSE Chow test (2000+)

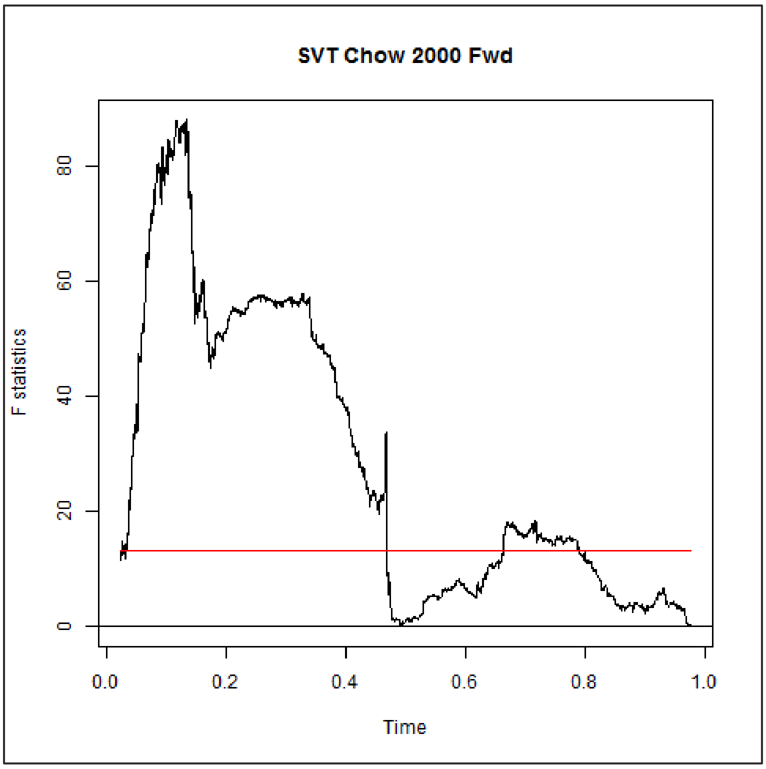


Figure A44 - SVT Chow test (2000+)

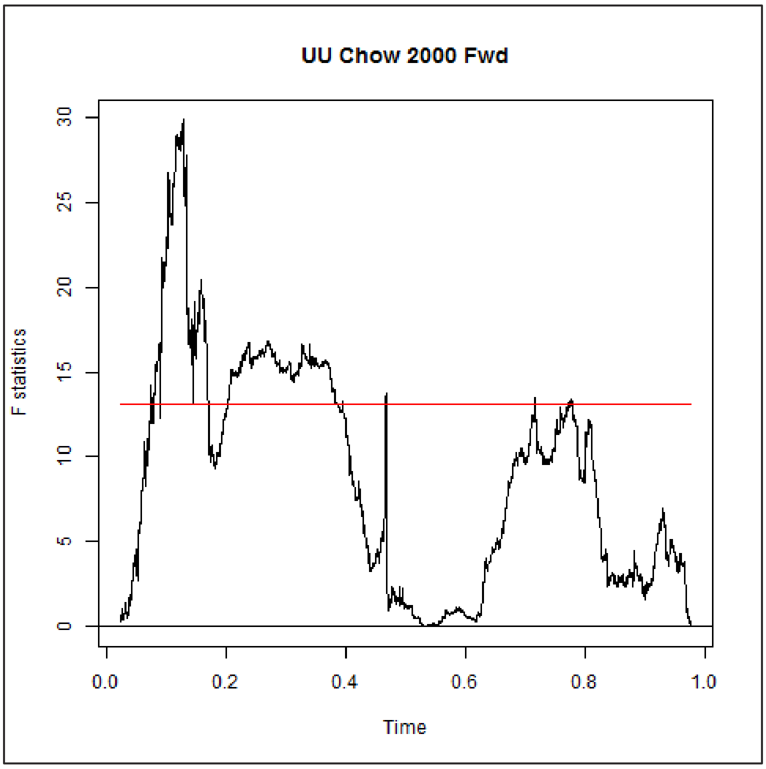


Figure A45 - UU Chow test (2000+)

Again, we see that the F-test provides evidence for structural breaks in the CAPM relationship even considering only the post-2000 timeframe. The evidence is weakest in the case of NG.

Results: Rolling Regression

A 5-year rolling window was assumed applied via an OLS estimator. The estimation period is 1993+ to provide consistent results as the programme seems to handle missing data differently to standard regression analysis. Parameter stability (or lack thereof) can be seen in the following graphs:

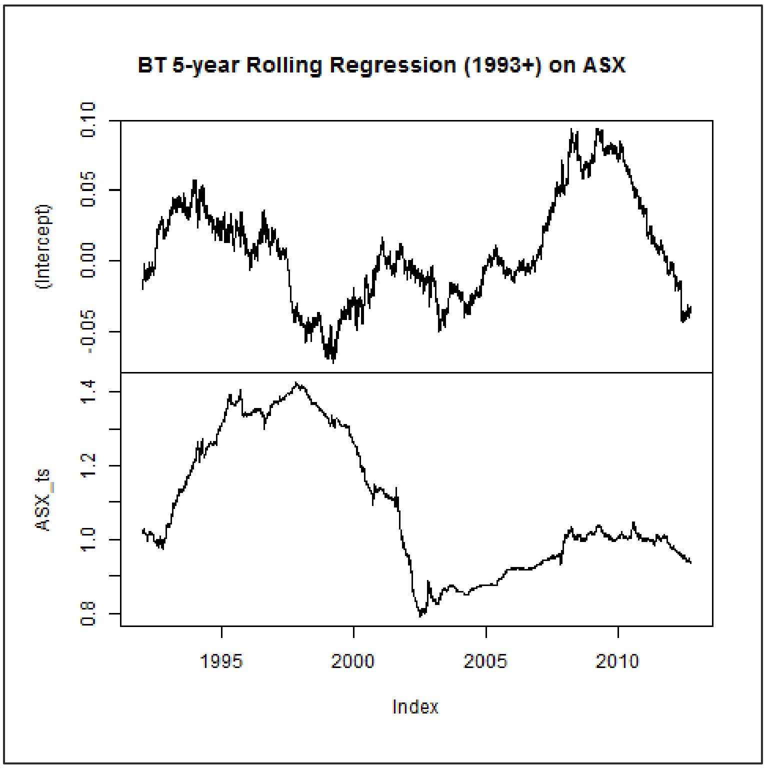


Figure A46 - BT Rolling Regression Analysis

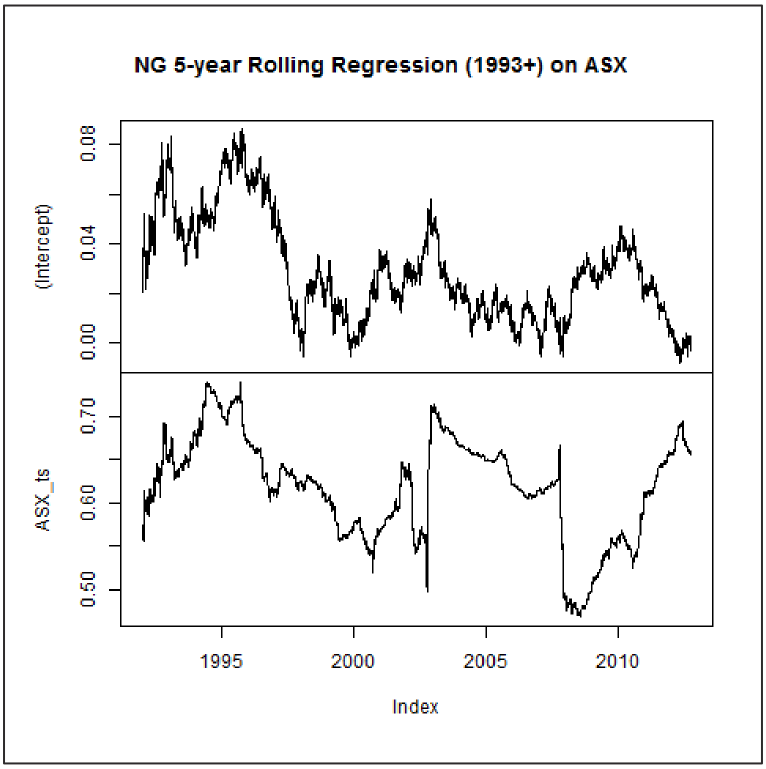


Figure A47 - NG Rolling Regression Analysis

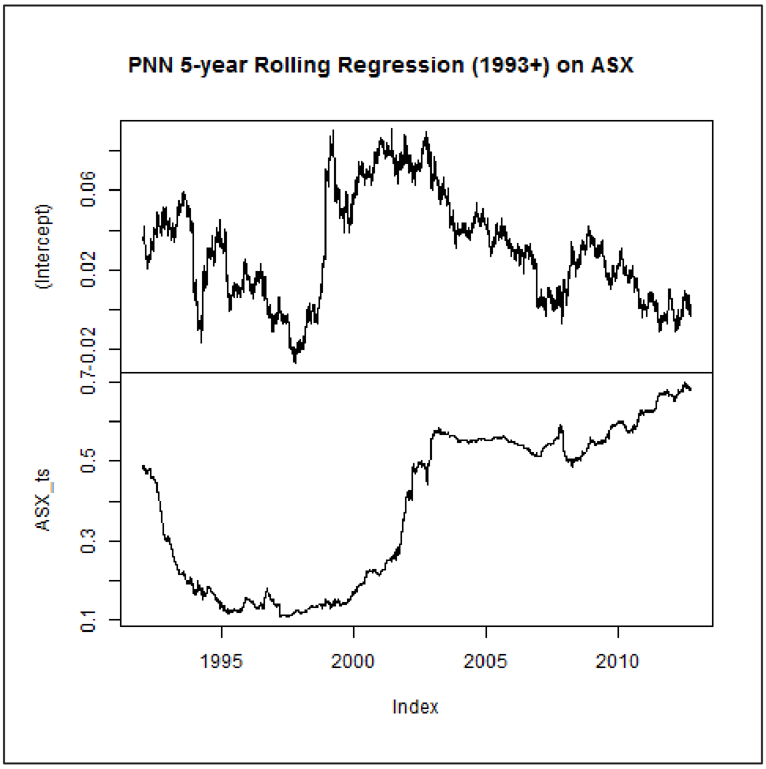


Figure A48 - PNN Rolling Regression Analysis

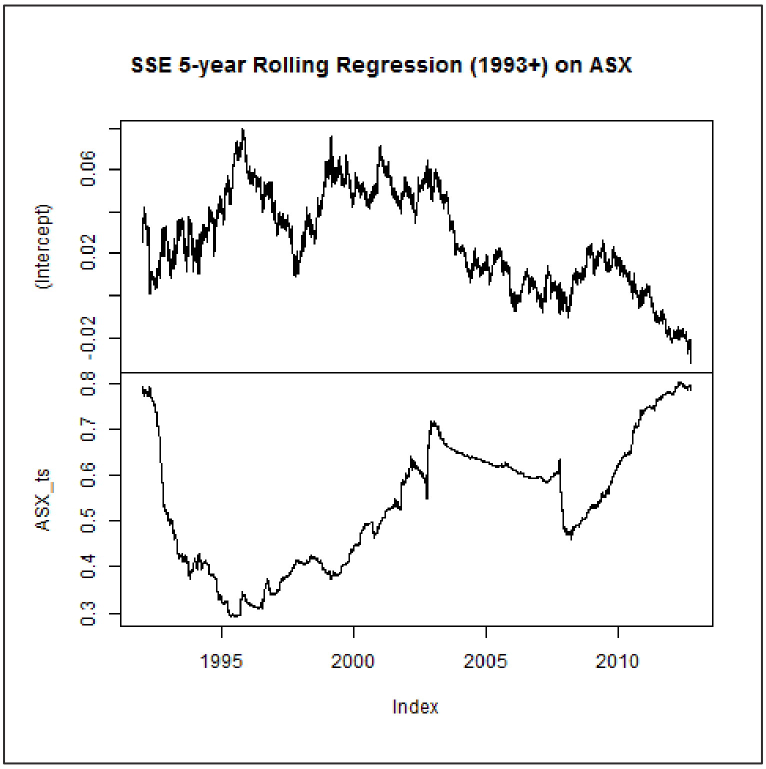


Figure A49 - SSE Rolling Regression Analysis

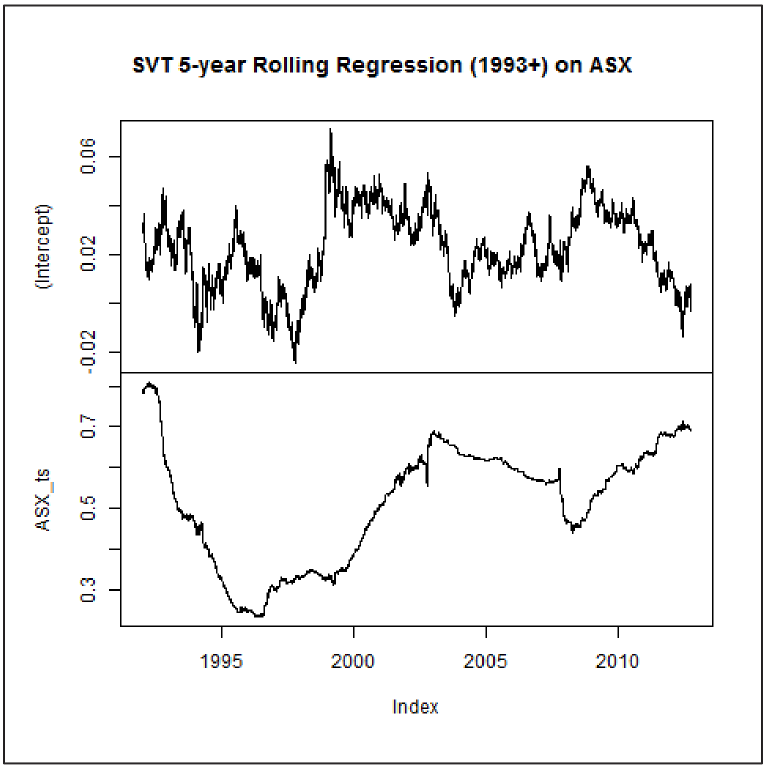


Figure A50 - SVT Rolling Regression Analysis

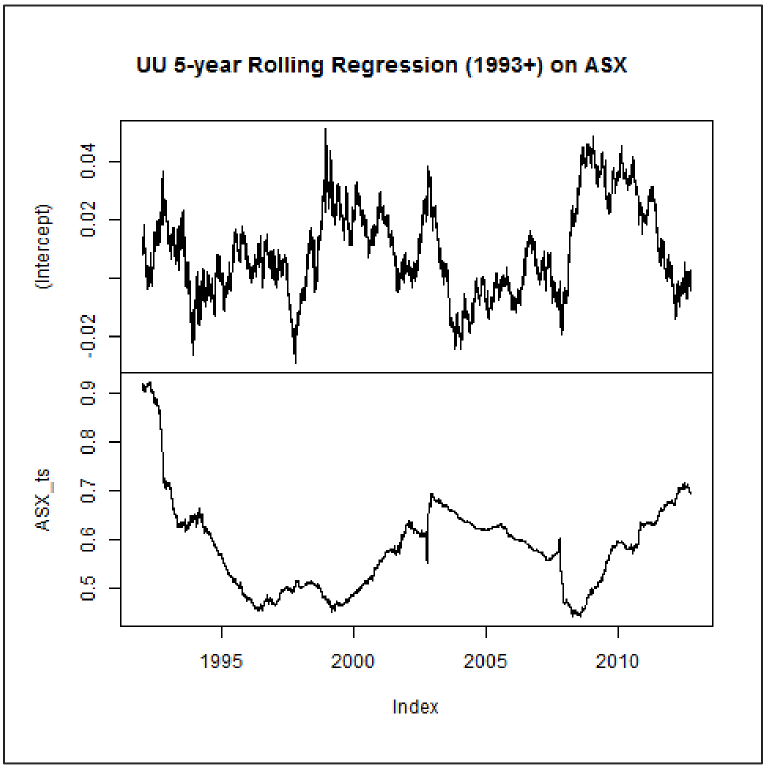


Figure A51 - UU Rolling Regression Analysis

The rolling regressions are interesting – each run of the model drops the oldest observation and adds a new one from those in the future. In the cases of all but BT and NG, we see quite distinct periods for the beta: up to 2002-03, from then until the Great Recession and post-Great Recession. NG and, more strikingly, BT do not follow this pattern.

Results: Optimal breakpoints

The BIC-minimising numbers of sub-periods in the OLS CAPM relationship for each utility are reported in Table A2.

Table A2 - Number of Periods that Minimise BIC

Stock	*Entire period*	*Post-2000*
BT	3	2
NG	1	1
PNN	3	2
SSE	3	4
SVT	3	2
UU	2	1

Only NG seems to be well-characterised by a single CAPM relationship across the entire period (1987-2018). Even if we consider only the post-2000 period there is significant evidence for structural shifts in all cases except those of NG and UU.

Results: Summary

Over the longer time horizon, all CAPM-type relationships considered show structural shifts according to one or more of the measures. Even considering just the post-2000 period, most series show evidence of structural shifts. The various graphical analyses are summarised in Table A3 below.

Table A3 - Summary of Findings

Stock	*CUSUM residuals*	*Recursive residuals*	*Recursive parameters*	*Moving parameters*	*Chow (whole period)*	*Chow (2000+)*
BT	-	-	Break	Break	Break	Break
NG	-	Break	Break	Break	Break	Break
PNN	-	Break	Break	Break	Break	Break
SSE	-	-	Break	Break	Break	Break
SVT	-	-	Break	Break	Break	Break
UU	-	Break	Break	Break	Break	Break

Table A3 above further shows that the CAPM relationship is subject to (multiple) structural breaks in almost all cases if one chooses sub-periods so as to minimise the BIC. Exceptions are NG (at all times) and UU (post-2000). Annex A3 provides further information on the possible dates for the structural breaks.

Bibliography

BREUSCH, T. S., AND A. R. PAGAN (1979): "A Simple Test for Heteroscedasticity and Random Coefficient Variation," *Econometrica*, 47, 1287-1294.

BROWN, R. L., J. DURBIN, AND J. M. EVANS (1975): "Techniques for Testing the Constancy of Regression Relationships over Time," *Journal of the Royal Statistical Society. Series B (Methodological)*, 37, 149-192.

CHOW, G. C. (1960): "Tests of Equality between Sets of Coefficients in Two Linear Regressions," *Econometrica*, 28, 591-605.

CHU, C.-S. J., K. HORNIK, AND C.-M. KUAN (1995): "The Moving-Estimates Test for Parameter Stability," *Econometric Theory*, 11, 699-720.

ENGLE, R. F. (1982): "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation," *Econometrica*, 50, 987-1007.

PLOBERGER, W., AND W. KRÄMER (1992): "The Cusum Test with Ols Residuals," *Econometrica*, 60, 271-285.

PLOBERGER, W., W. KRÄMER, AND K. KONTRUS (1989): "A New Test for Structural Stability in the Linear Regression Model," *Journal of Econometrics*, 40, 307-318.

R CORE TEAM (2018): "R: A Language and Environment for Statistical Computing," Austria, Vienna: R Foundation for Statistical Computing.

SCHWARZ, G. (1978): "Estimating the Dimension of a Model," *Ann. Statist.*, 6, 461-464.

ZEILEIS, A., AND G. GROTHENDIECK (2005): "Zoo: S3 Infrastructure for Regular and Irregular Time Series," *Journal of Statistical Software*, 14, 1-27.

ZEILEIS, A., AND T. HOTHORN (2002): "Diagnostic Checking in Regression Relationships," *R News*, 2, 7-10.

ZEILEIS, A., F. LEISCH, K. HORNIK, AND C. CHRISTIAN KLEIBER (2002): "Strucchange: An R Package for Testing for Structural Change in Linear Regression Models," *Journal of Statistical Software*, 7, 1-38.

Annex A1: Behaviour of All-share Index Returns

AA1.1 Aim

To characterise the univariate time series properties of the ASX.

AA1.2 Data

Ofgem provided ASX and other share price data from Bloomberg for the period from Jan 2000 to Sep 2018. Returns were calculated as 100 times the trading day-to-trading day percentage price change.

AA1.3 Analysis

Initial graphing showed that the ASX returns series appears to exhibit classic ARCH behaviour. Unit root tests (KPSS (Kwiatkowski, D., P. C. B. Phillips, P. Schmidt and Y. Shin 1992) with H_0 of stationarity, Augmented Dickey-Fuller (Said, S. E. and D. A. Dickey 1984) and Phillips-Perron (Phillips, P. C. B. and P. Perron 1988) with H_0 of a unit root) support the view that the series is stationary (p-values of >0.10, <0.01 and <0.01 respectively – stationarity is not rejected while a unit root is rejected).

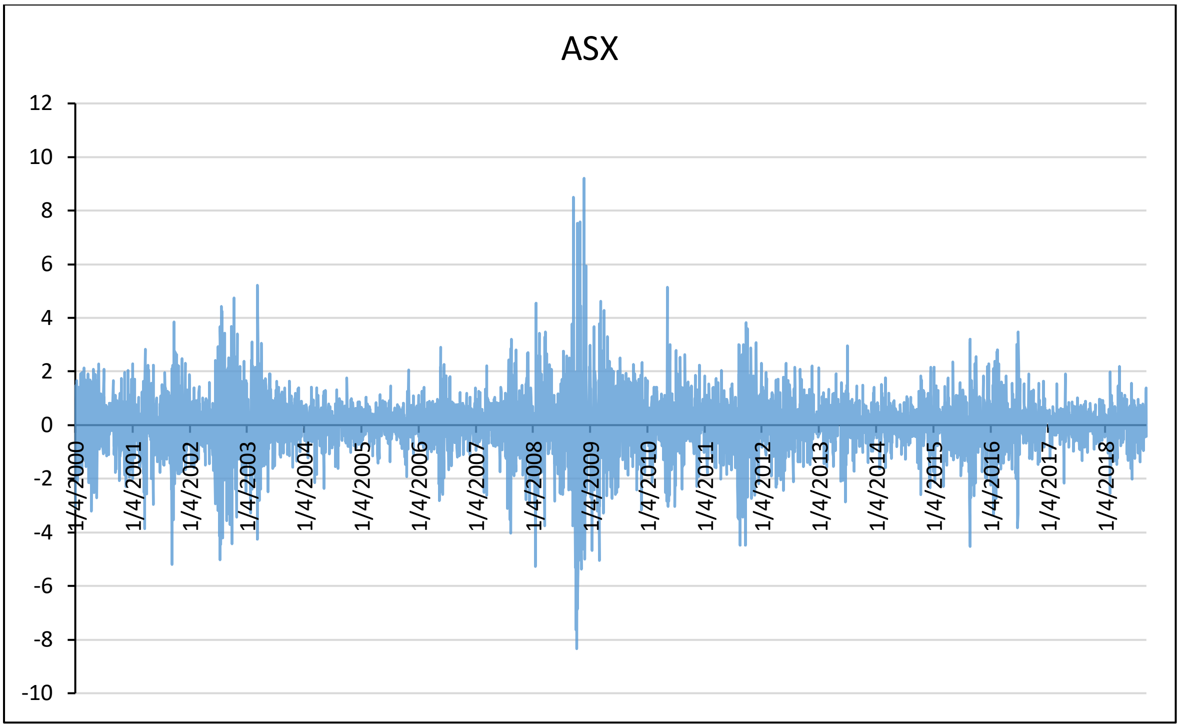


Figure AA1.1 - ASX daily returns 2000-2018

A series of univariate (G)ARCH models were run with orders (1,1), (1,2), (2,1) and (2,2). The Bayesian Information Criterion, or BIC, (Schwarz, G. 1978) was used as the basis for model choice.

Residuals were standardised by the estimated standard errors and then tested at 5 and 10 lags using the Engle test (@archtest) (Engle, R. F. 1982).

All modelling was undertaken in RATS v10.

AA1.4 Results

The table below shows the log-likelihood and BIC calculation: the BIC generally rewards model parsimony and we see that here the (1,1) order form is chosen.

Table AA1.1 - Choice of model order according to BIC

	ARCH(1,1)	*ARCH(1,2)*	*ARCH(2,1)*	*ARCH(2,2)*
Log-Likelihood	-6266.08	-6265.25	-6265.46	-6265.23
N	4733	4733	4733	4733
Parameters	4	5	5	6
BIC	12566.01	12572.81	12573.22	12581.24
Difference from minimum	0.00	6.80	7.21	15.23

Graphs will only be presented here for the chosen order (1,1) model – graphs for the higher order models are available upon request.

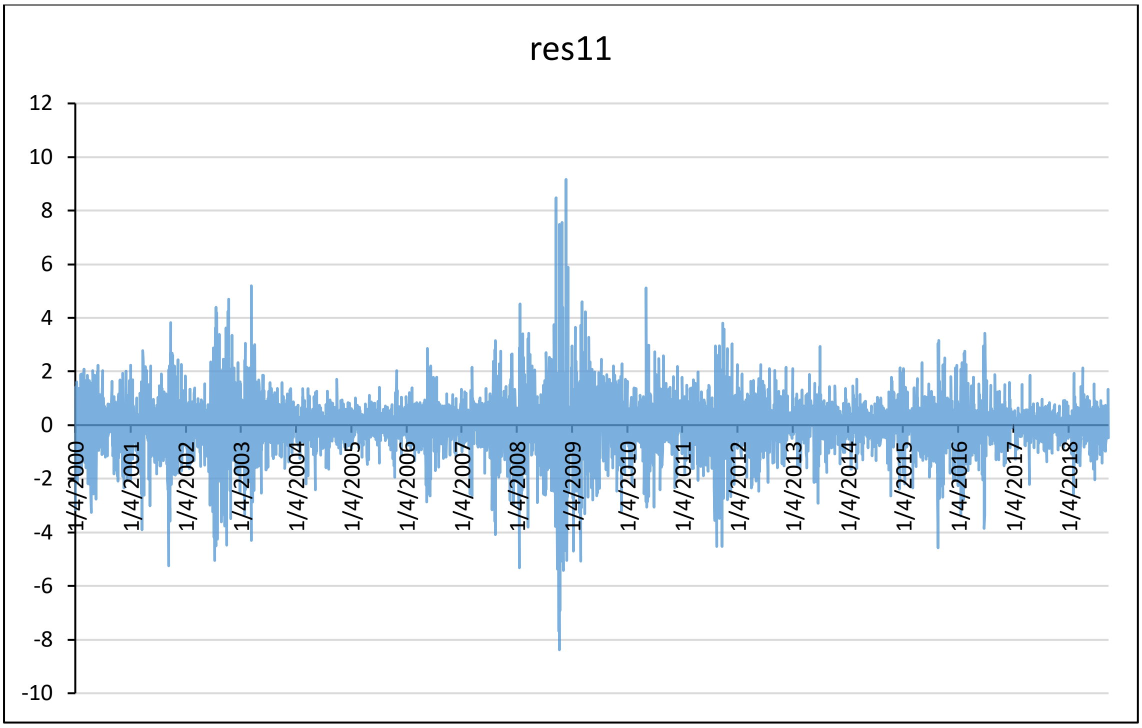


Figure AA1.2- Residuals from order (1,1) ARCH model for ASX

As the model specifies simply a mean for the series, the residuals are essentially just a vertically shifted version of the original data series.

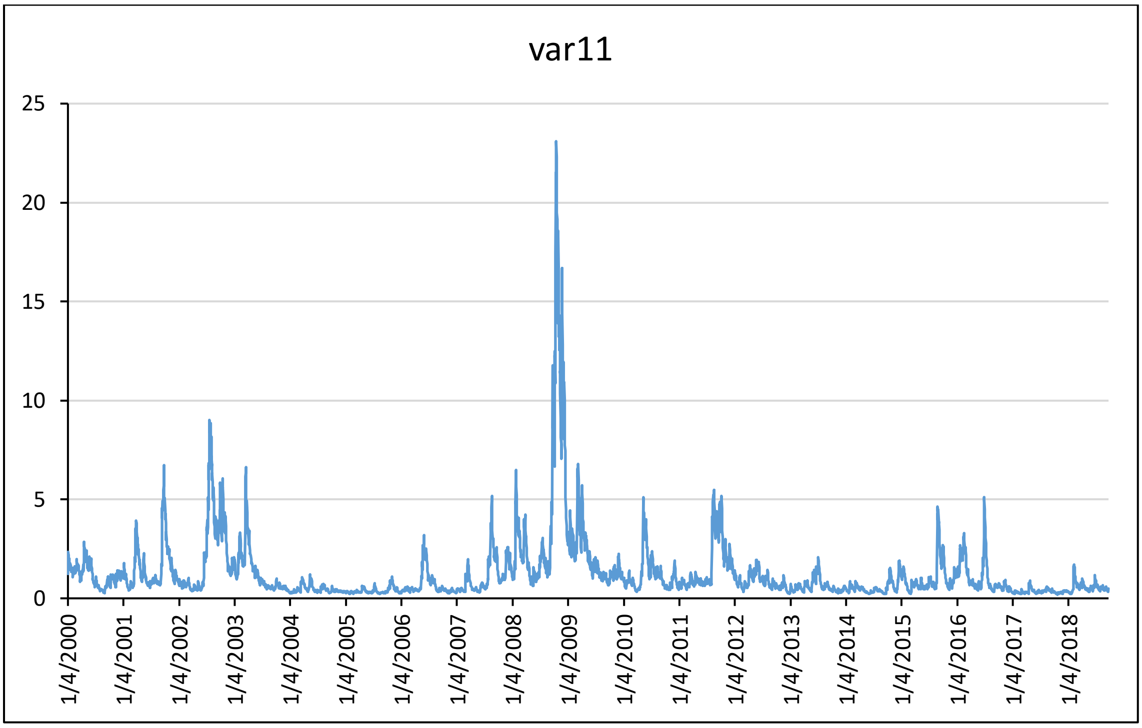


Figure AA1.3 - Estimated error variances from order (1,1) ARCH model for ASX

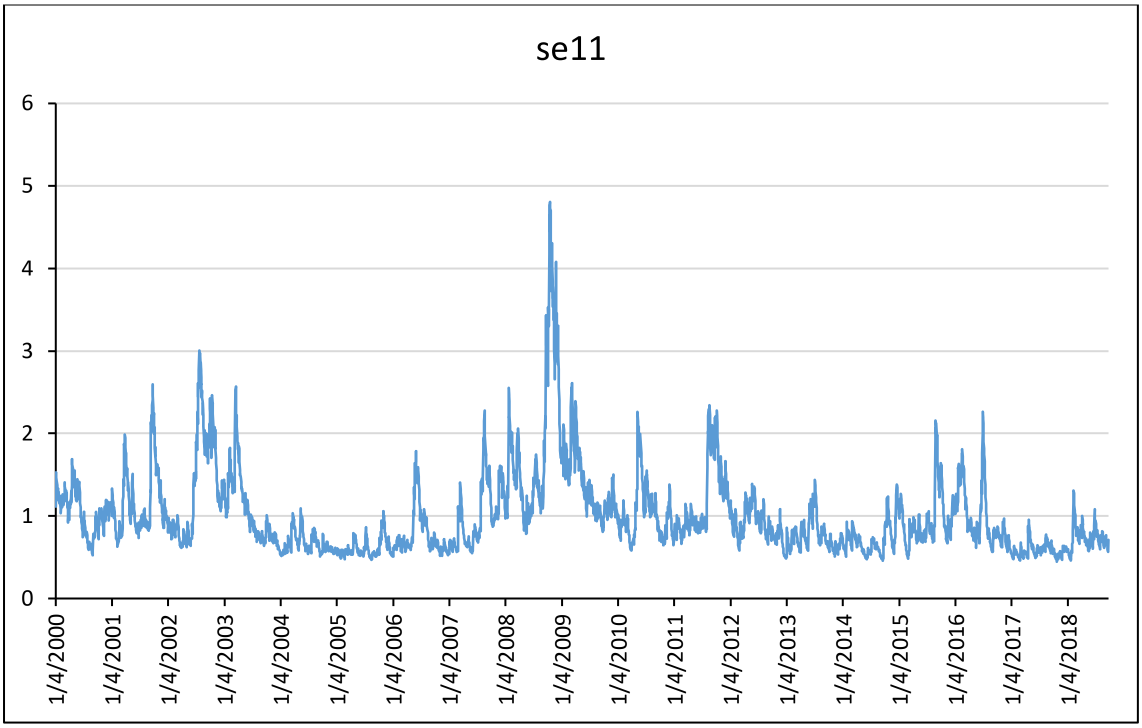


Figure AA1.4 - Estimated standard errors for error term in order (1,1) ARCH model for ASX

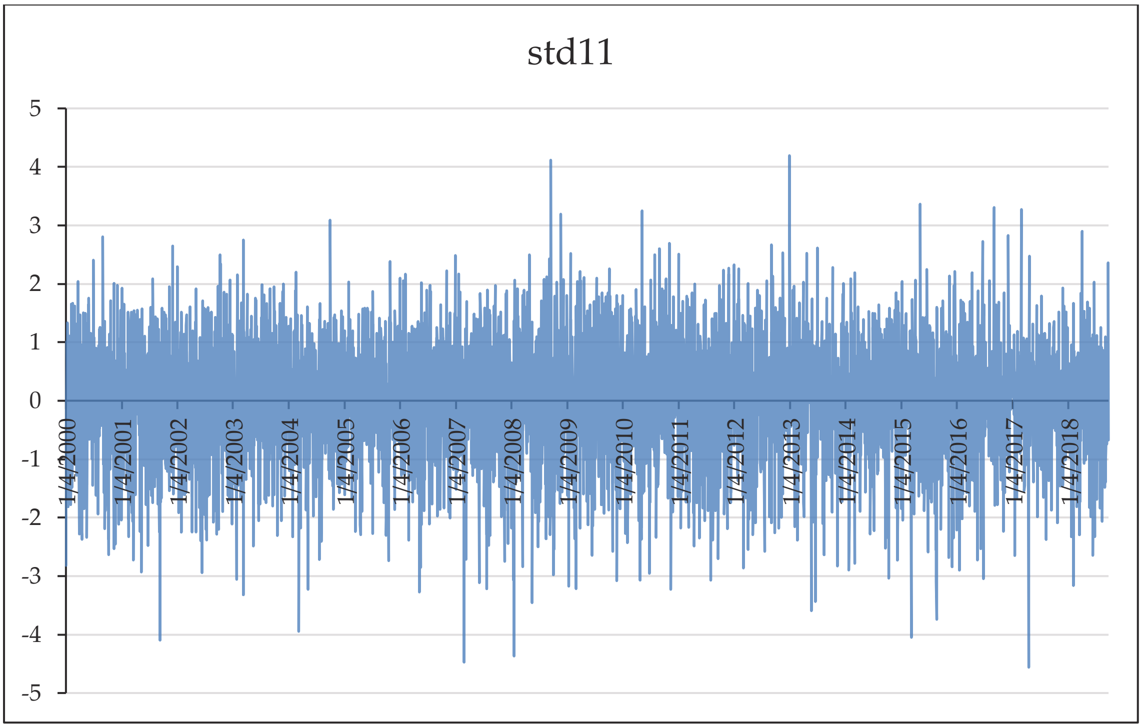


Figure AA1.5 - Standardised residuals from order (1,1) ARCH model for ASX

Comparing Figures AA1.3 and AA1.4, it is obvious that the estimated variances/standard errors serve to render the series non-(G)ARCH. Tests – values are also shown for the other three models estimated – serve to confirm the impression that the ARCH(1,1) model adequately describes the behaviour of ASX returns from 2000 to 2018.

Table AA1.2 - p-values for Engle test for residual ARCH behaviour

	ARCH(1,1)	*ARCH(1,2)*	*ARCH(2,1)*	*ARCH(2,2)*
5 lags	0.60	0.86	0.88	0.87
10 lags	0.48	0.64	0.66	0.65

AA1.5 Discussion

The returns to the ASX are adequately described by an ARCH(1,1) model with there being no evidence for further ARCH behaviour among the standardised residuals from the model.

Bibliography

ENGLE, R. F. (1982): "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation," *Econometrica*, 50, 987-1007.

KWIATKOWSKI, D., P. C. B. PHILLIPS, P. SCHMIDT, AND Y. SHIN (1992): "Testing the Null Hypothesis of Stationarity against the Alternative of a Unit Root: How Sure Are We That Economic Time Series Have a Unit Root?," *Journal of Econometrics*, 54, 159-178.

PHILLIPS, P. C. B., AND P. PERRON (1988): "Testing for a Unit Root in Time Series Regression," *Biometrika*, 75, 335-346.

SAID, S. E., AND D. A. DICKEY (1984): "Testing for Unit Roots in Autoregressive-Moving Average Models of Unknown Order," *Biometrika*, 71, 599-607.

SCHWARZ, G. (1978): "Estimating the Dimension of a Model," *Ann. Statist.*, 6, 461-464.

Annex A2: Volatility of returns

One of the issues that has been identified is a spike in beta values. Understanding whether the spike was caused by an event that should be incorporated into longer term views of the riskiness of companies requires an investigation of the spike.

Figure AA2.1 below provides the rolling variance of returns from 2010. Unexpectedly the variance on the returns of the market are low compared to individual companies – a portfolio effect. But what is interesting is that the market variance has been low while company variances are relatively high. Given the role of market returns in the beta estimation formula, a low value will, *ceterus paribus*, lead to a higher equity beta.

Figure AA2.1 stock variance information 2010-2018

A low volatility of market returns is only one part of the story. The covariance of company and market returns is also a part of the estimation formula. Figure AA2.2 provides the covariance information over the same period as the first figure. As can be seen, the covariance was high until the end of June 2017. So, combined with the low volatility of market returns this led to a spike.

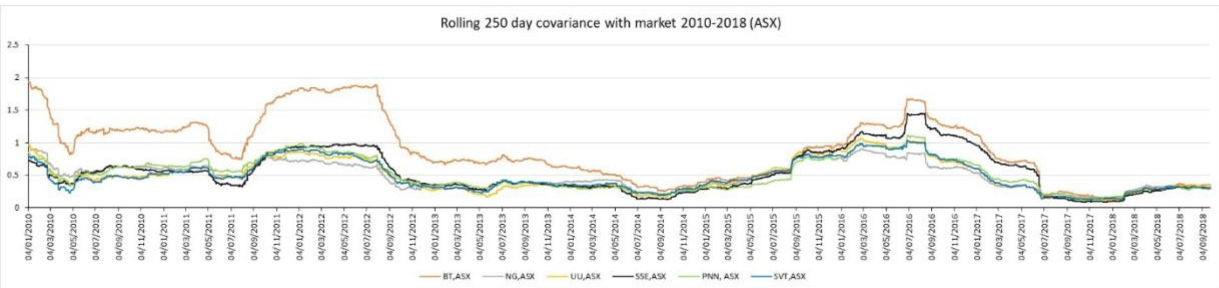


Figure AA2.2 stock variance information 2010-2018

So, the spike was caused by two effects. Understanding those effects is key to knowing whether the resulting beta value is normal. The covariance peak seems to mirror the period from the 2015 election to the 2017 one. Suggesting that political uncertainty was driving some of the spike. The low volatility of market returns was a function of wider financial and economic factors and attracted comment at the time.

Annex A3: Dating the structural breaks

AA3.1 Aim

To compare the Chow F-test approach to identifying structural breaks to that using the Bayesian Information Criterion (Schwarz, G. 1978).

AA3.2 Method

All analyses were conducted in R (R Core Team 2018) using the *strucchange* package (Zeileis, A., F. Leisch, K. Hornik and C. Kleiber 2002).

Data on returns from 2000-2018 were used to estimate the basic CAPM relationship, the independent variable being returns to the ASX.

AA3.3 Results

Visual inspection of the 2000-2018 F-tests suggests the following possibilities:

Table AA3.1 - Visual breaks identified from F-tests

Stock	***Break 1***	***Break 2***	***Break 3***	***Break 4***	***Break 5***	***Break 6***
BT	2001	2003/04				
NG	2000	2001	Oct 2008	2009		
UU	2002	2003	2004/06	Sep 2008	2013/15	
SSE	2001	2002	2003	2004/05	Sep 2008	2013/15
PNN	2004/05	Oct 2008				
SVT	2002/03	2004/05	Oct 2008	2012/13		

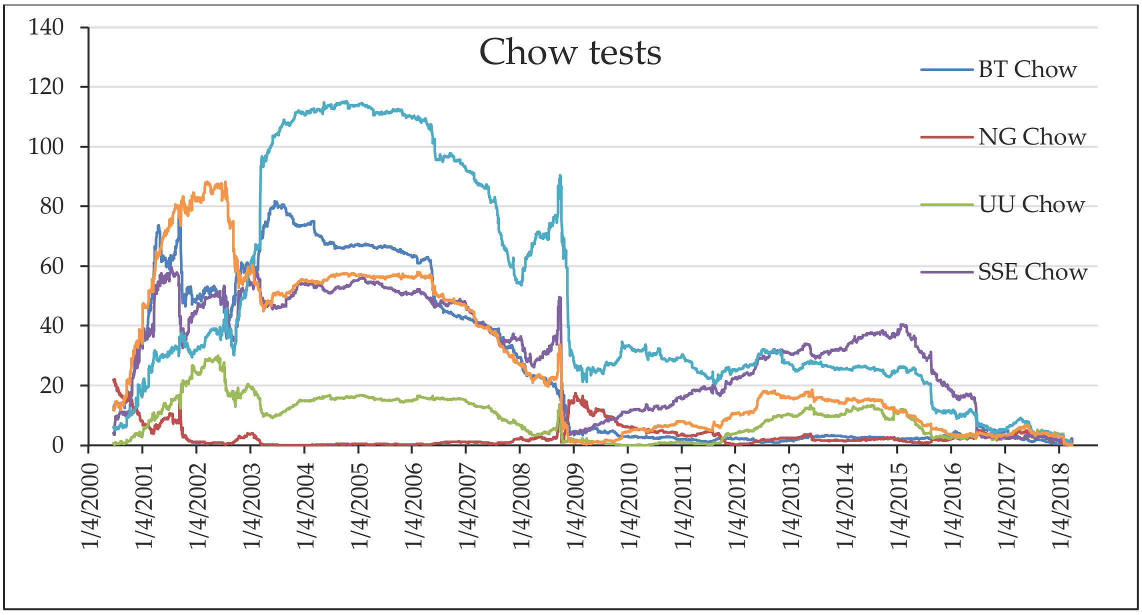


Figure AA3.1 - Chow test values 2000-2018

The BIC-driven "optimal breaks"[4] give the following results:

Table AA3.2 - Breaks identified by R strucchange programme

Stock[5]	***Break 1***	***Break 2***	***Break 3***
BT	20/06/2003		
SSE	28/01/2005	07/01/2009	13/02/2015
PNN	15/10/2004		
SVT	17/01/2003		

The shading of the cells permits cross-referencing between Table and Table : for example, the pink shading highlights the visually-identified break of 2003/04 for BT and the algorithmically-identified break of 20/06/2003.

AA3.4 Discussion

In many cases there is some agreement – indicated by highlight colour – between the visually-identified breakpoints (necessarily a little temporally vague) and those identified algorithmically.

It is also notable that there is a break in autumn 2008 in all models but that for BT. This coincides with the major market crash that formed part of the Great Recession of 2007-2009.

[4] The programme chooses up to 5 breaks (the number possible is user-controlled) using a dynamic programming approach so as to minimise the overall Bayesian Information Criterion

[5] NG and UU have no BIC-minimising "optimal breaks" in the 2000-2018 timeframe.

Bibliography

R CORE TEAM (2018): "R: A Language and Environment for Statistical Computing," Austria, Vienna: R Foundation for Statistical Computing.

SCHWARZ, G. (1978): "Estimating the Dimension of a Model," *Ann. Statist.*, 6, 461-464.

ZEILEIS, A., F. LEISCH, K. HORNIK, AND C. KLEIBER (2002): "Strucchange: An R Package for Testing for Structural Change in Linear Regression Models," *2002*, 7, 38.

Appendix B Ordinary Least Squares analysis and its limitations

B.1 Introduction

This appendix provides OLS results for beta at various frequencies. In the interests of space, only results calculated for the period Jan 2000 to Sep 2018 are included. Results for earlier periods are available upon request but sensible analysis of beta is unlikely to consider data from so far back.

B.2 OLS results daily

Simple OLS produces the results below:

Table B1 - Daily OLS beta estimates 2000-2018

Stock	*Beta (s.e. in parentheses)*
BT	1.04 (0.02)
NG	0.61 (0.02)
UU	0.57 (0.02)
SSE	0.57 (0.02)
PNN	0.45 (0.02)
SVT	0.53 (0.02)

Most of the betas are in the 0.5-0.6 range, as one might expect.

Rolling 500 day (roughly 2 year) rolling regressions were also calculated.

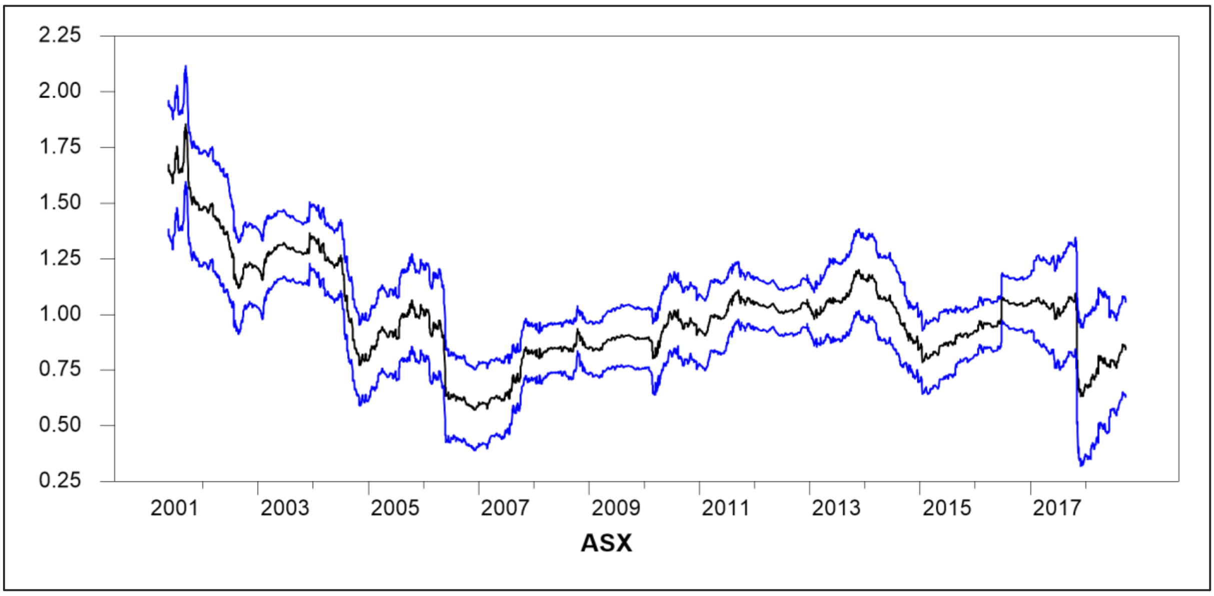


Figure B1 - Rolling beta estimate (500 days) for BT with confidence interval

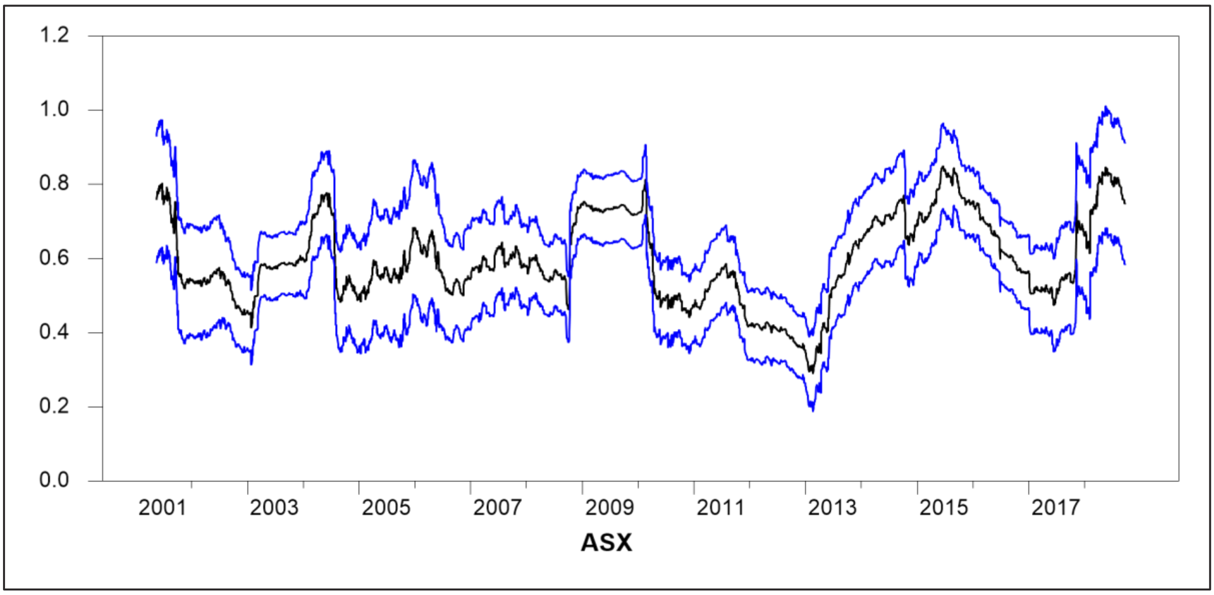


Figure B2 - Rolling beta estimate (500 days) for NG with confidence interval

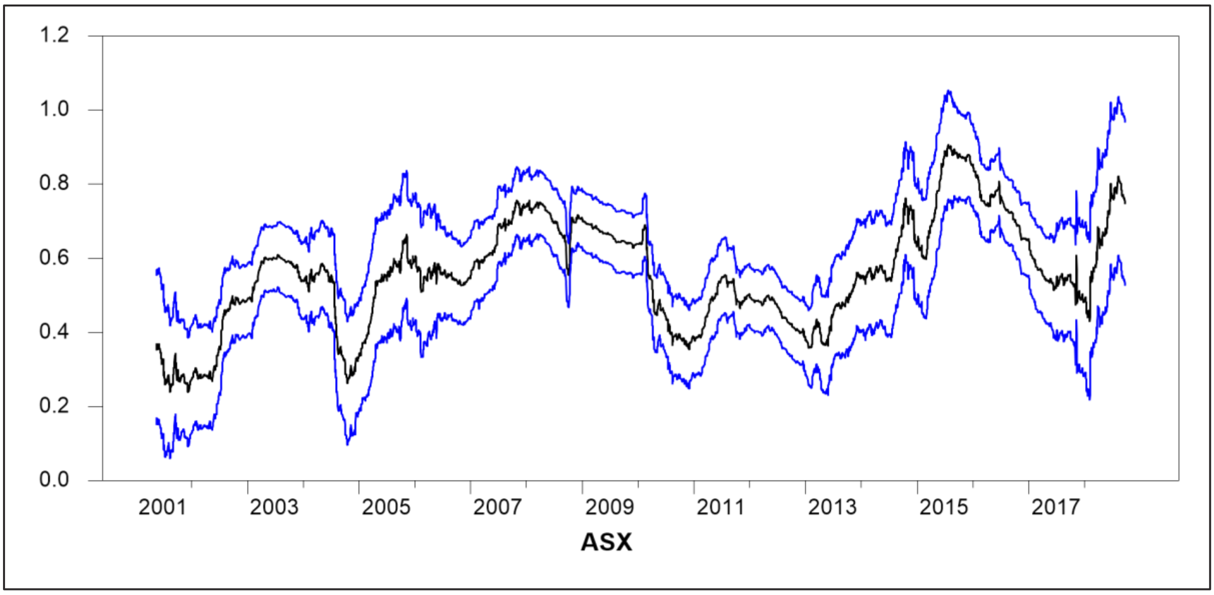


Figure B3 - Rolling beta estimate (500 days) for UU with confidence interval

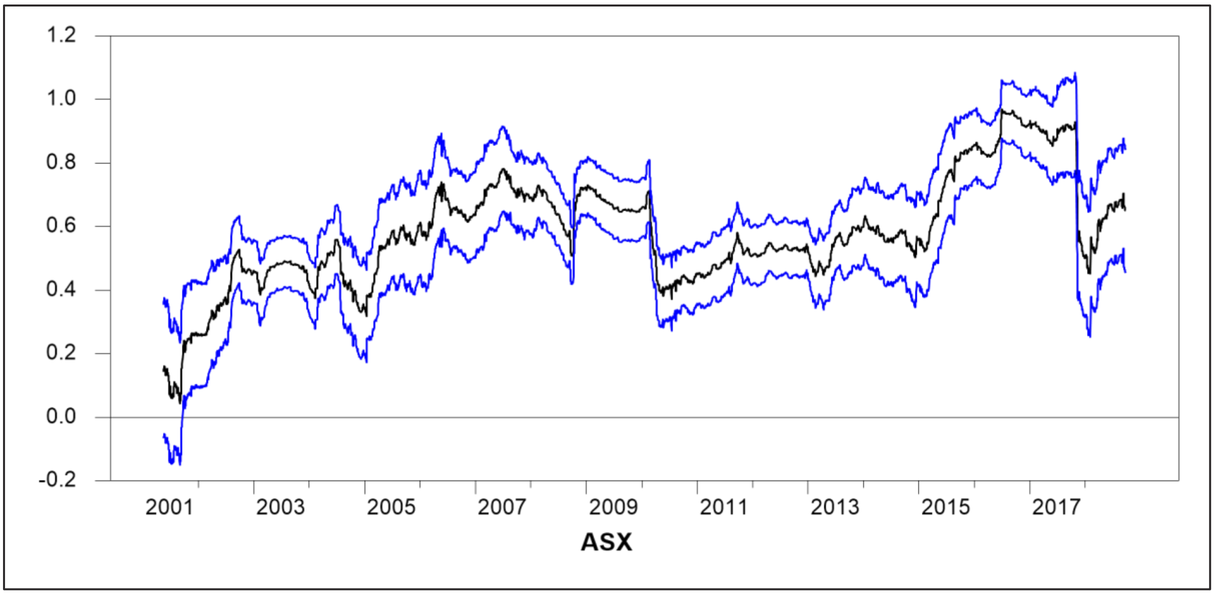


Figure B4 - Rolling beta estimate (500 days) for SSE with confidence interval

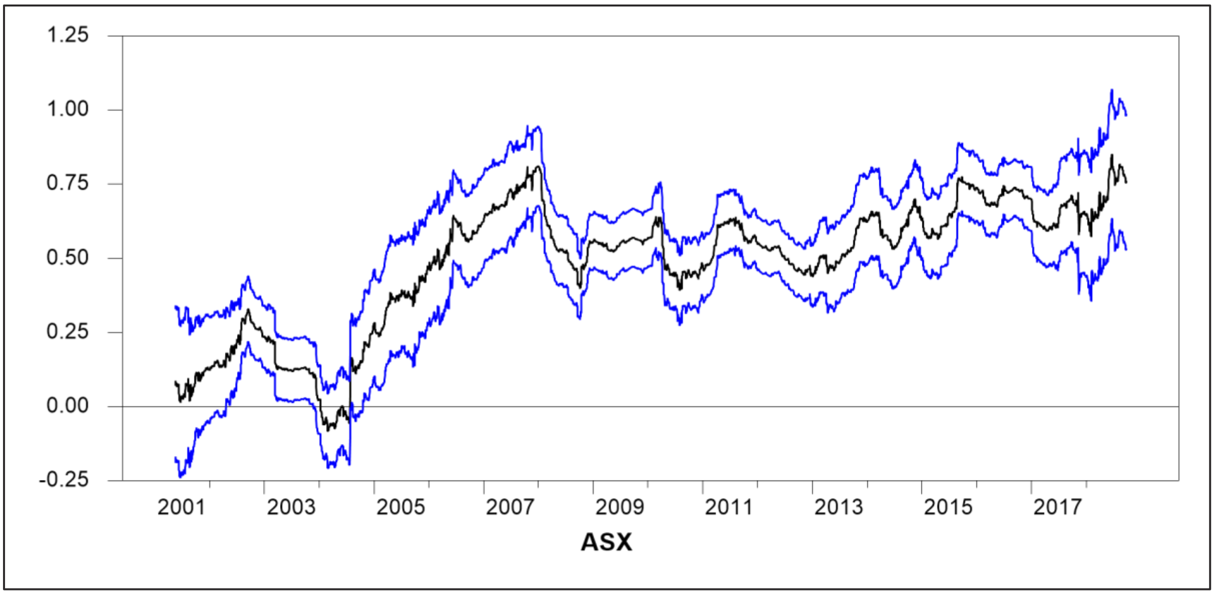


Figure B5 - Rolling beta estimate (500 days) for PNN with confidence interval

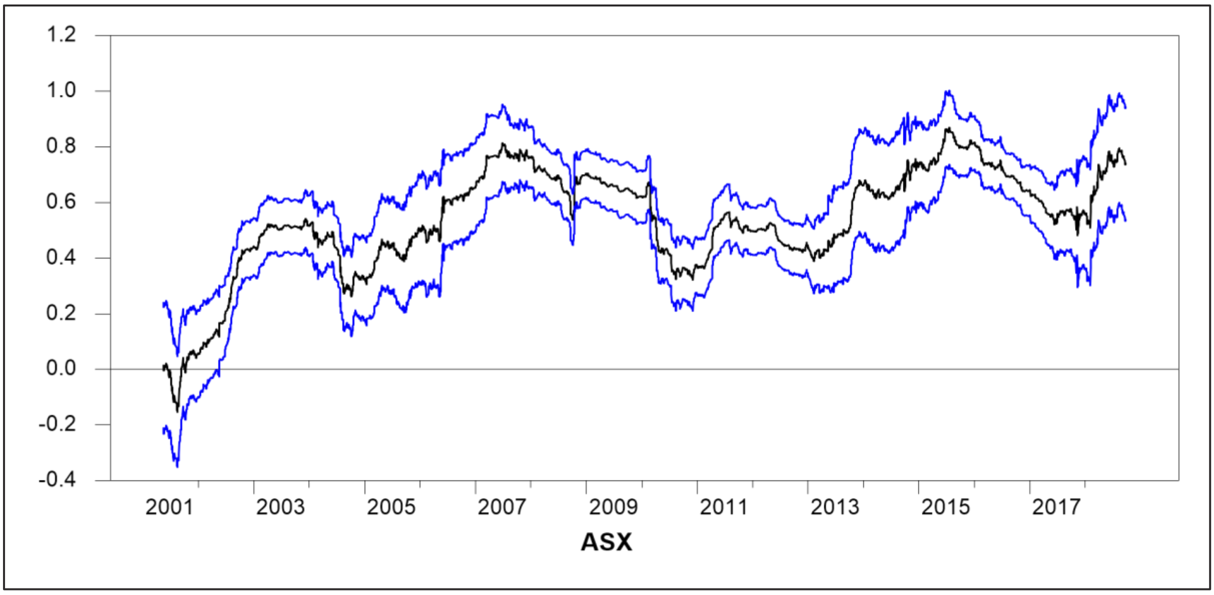


Figure B6 - Rolling beta estimate (500 days) for SVT with confidence interval

It is obvious from the above figures that beta varies over time.

B.3 OLS results weekly

As described in the data appendix, weekly data sets were created for Monday-to-Monday, Tuesday-to-Tuesday etc. Betas were calculated for each trading day and then averaged to give a "weekly beta".

Rolling regressions were then calculated using a window of 100 observations (approximately two years). The results for the five days were averaged and then a 200 element moving average was created for graphing.

All analyses were conducted in RATS v10 using the *linreg*, *rollreg* and *archtest* commands.

OLS results are presented in the table below.

Table B2 - Weekly OLS betas by day of week

Stock	*Monday beta*	*Tuesday beta*	*Wednesday beta*	*Thursday beta*	*Friday beta*	*Average beta*
BT	0.903	1.006	0.928	0.867	0.858	0.912
NG	0.570	0.606	0.555	0.501	0.599	0.566
UU	0.573	0.507	0.482	0.466	0.584	0.522
SSE	0.570	0.563	0.497	0.443	0.526	0.520
PNN	0.505	0.524	0.451	0.439	0.504	0.485
SVT	0.515	0.457	0.453	0.413	0.529	0.473

Table B3 - Ranks of weekly OLS betas

Stock	*Monday rank*	*Tuesday rank*	*Wednesday rank*	*Thursday rank*	*Friday rank*
BT	3	1	2	4	5
NG	3	1	4	5	2
UU	2	3	4	5	1
SSE	1	2	4	5	3
PNN	2	1	4	5	3
SVT	2	3	4	5	1

It is interesting that, for 5 of the 6 stocks, the Wednesday and Thursday betas are the second lowest and lowest respectively. It is, however, unclear whether this observation has real-world meaning.

We now present both a comparison of the average of the 100 week window rolling regression betas and the overall OLS average beta (see Table) as well as the graphs of the rolling regression (Figure Figure to Figure).

Table B4 - Comparison of overall and rolling betas

Stock	*Average weekly beta 2000-2018*	*Average of rolling weekly betas*
BT	0.912	0.871
NG	0.566	0.549
UU	0.522	0.538
SSE	0.520	0.552
PNN	0.485	0.515
SVT	0.473	0.514

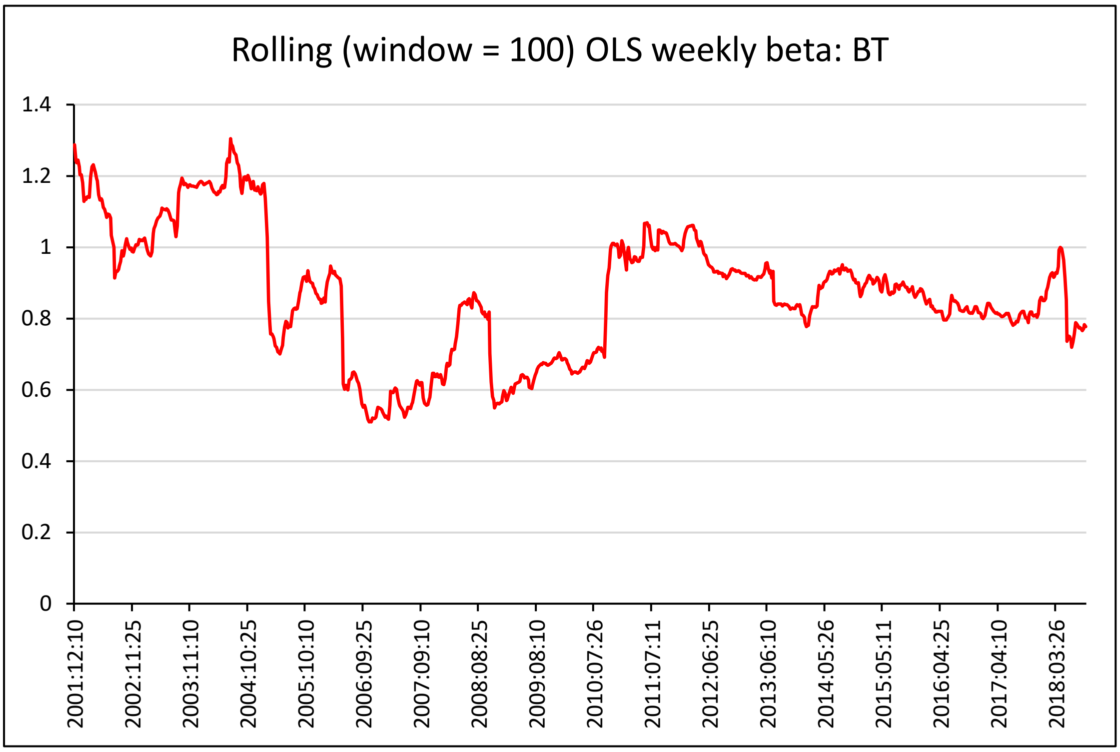


Figure B7 - BT rolling beta 2000-2018 weekly

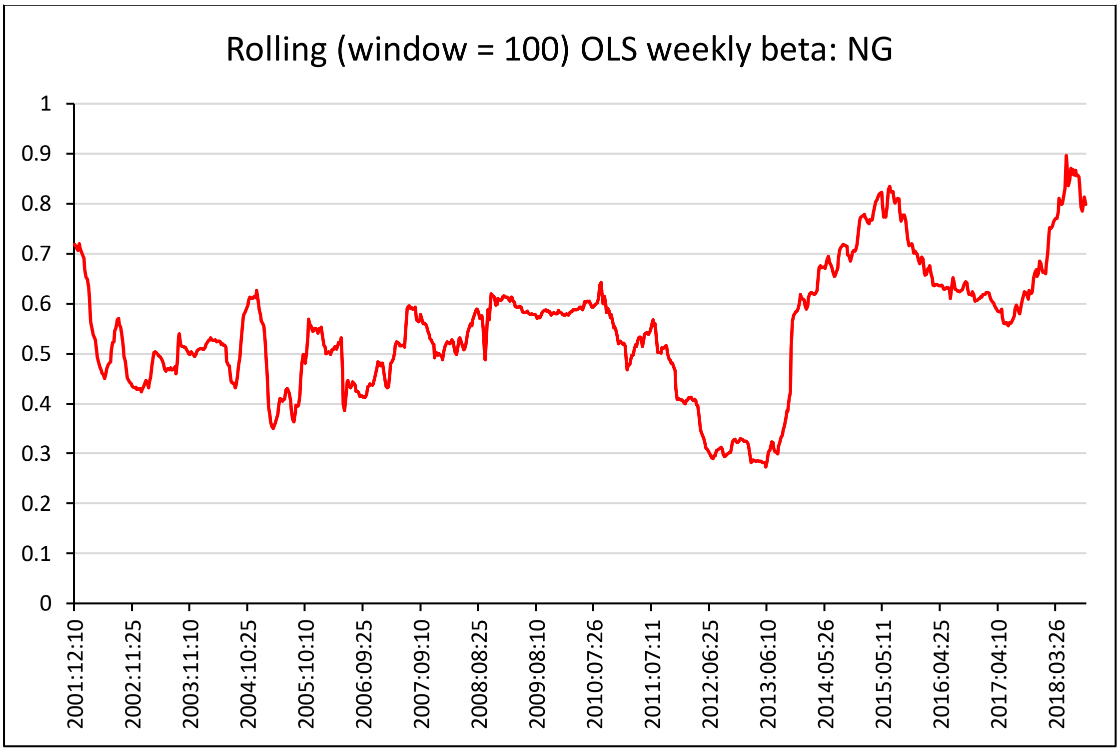


Figure B8 - NG rolling beta 2000-2018 weekly

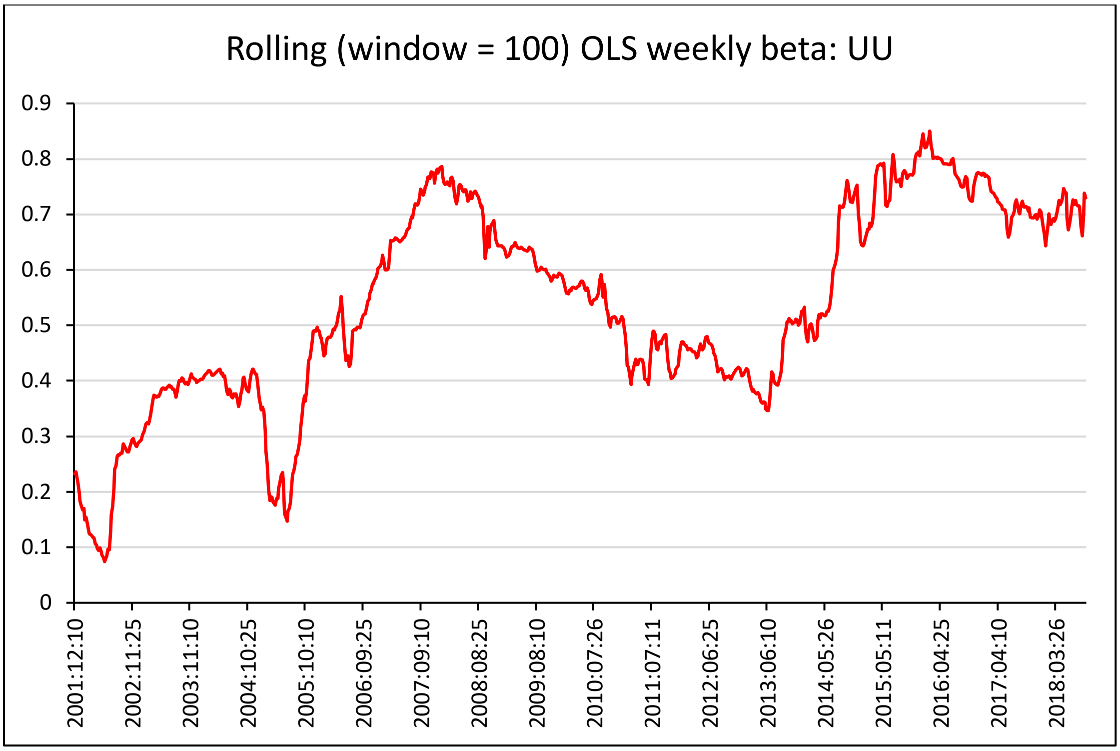


Figure B9 - UU rolling beta 2000-2018 weekly

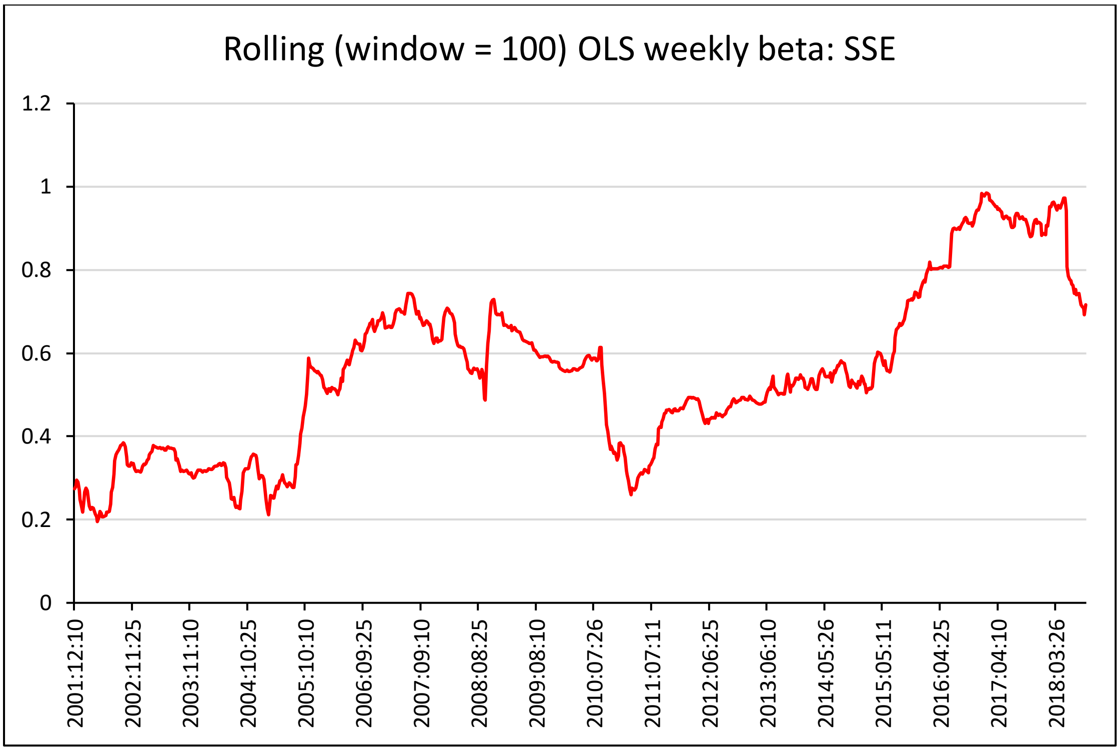


Figure B10 - SSE rolling beta 2000-2018 weekly

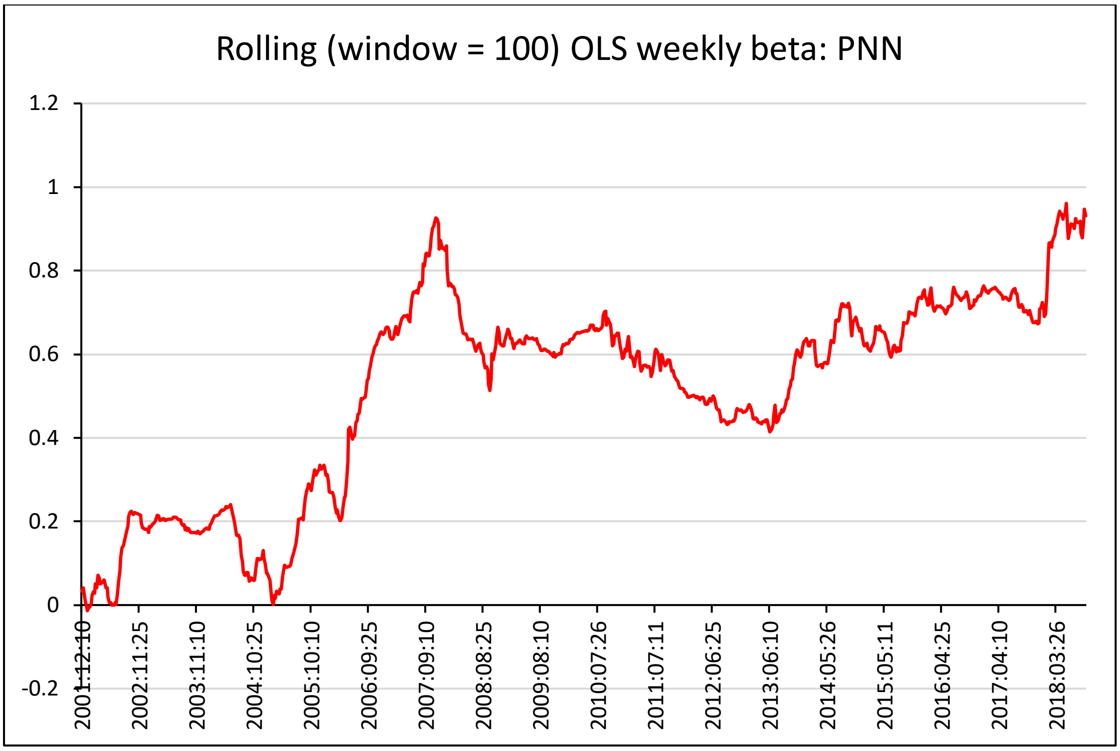


Figure B11 - PNN rolling beta 2000-2018 weekly

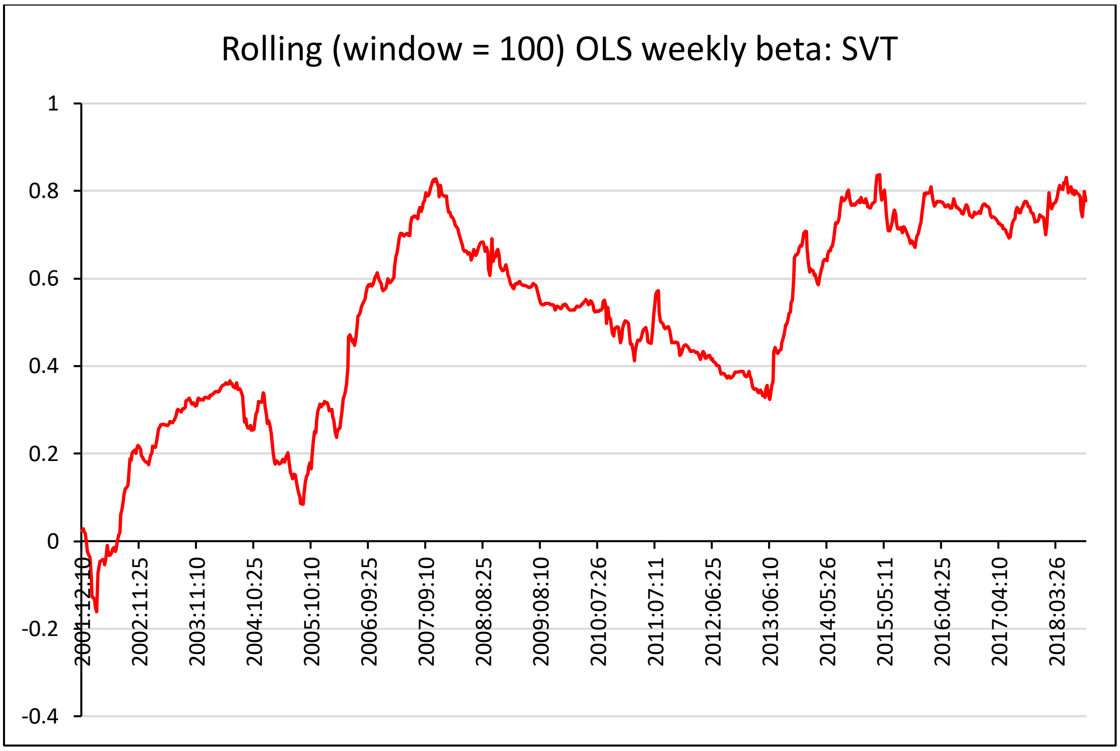


Figure B12 - SVT rolling beta 2000-2018 weekly

B.4 OLS results monthly

Monthly OLS models were run two ways – with first-trading-day to first-trading-day returns and with returns calculated every 20th trading day. Results are summarised (for the utilities) below:

Table B5 - Monthly beta estimates for utilities

Stock	First-trading-day model (s.e. in parentheses)	20th day model (s.e. in parentheses) [6]
SVT	0.43 (0.09)	0.42 (0.08)
NG	0.56 (0.08)	0.58 (0.07)
UU	0.50 (0.08)	0.42 (0.08)
PNN	0.47 (0.10)	0.38 (0.10)
SSE	0.42 (0.08)	0.40 (0.06)
BT	1.11 (0.11)	0.82 (0.10)

In the cases of SVT, NG and SSE the switch from "true" (first-trading-day to first-trading-day) returns to "every 20th trading day" makes essentially no difference while the estimates for UU, PNN and BT are higher – in the case of BT, substantially so.

An average result across the 20 different 20-day months (first to twenty-first, second to twenty-second etc.) was calculated:

Table B6 - Average beta across 20 alternative 20-day month

Stock	*Mean*	*S.D.*	*C.V.*
BT	0.986	0.080284	8.1%
NG	0.571	0.074727	13.1%
UU	0.481	0.072525	15.1%
SSE	0.462	0.059944	13.0%
PNN	0.437	0.065903	15.1%
SVT	0.442	0.07736	17.5%

Variability is quite marked when one considers different trading days of the month – the coefficient of variation for SVT is 17.5%.

Monthly data cannot be regarded as particularly suited to calculation of beta given the large amount of variation in data that is masked during aggregation.

B.5 Limitations of OLS: heteroscedasticity and (G)ARCH

There are various tests for heteroscedasticity – figures in the preceding appendix on data showed that returns appear to exhibit temporal clusters of higher and lower volatility –

[6] Estimated Jan 2000 to Aug 2017 as part of efforts to replicate results in Robertson (2018)

amongst which those of Breusch & Pagan (1979) and White (1980) are classics for general heteroscedasticity. We here present tests for heteroscedasticity in the various models presented above.

The results of the Breusch-Pagan test for daily data are presented in Table B7.

Table B7 - Results of Breusch-Pagan test for heteroscedasticity (p-values – H_0 of homoscedasticity) – daily data, 2000+

Stock	*p-value*
BT	0.11
NG	0.02
PNN	0.67
SSE	0.06
SVT	0.00
UU	0.04

Heteroscedasticity is, as expected, relatively common (although far from dominant) – only PNN and SSE seem to have relatively homoscedastic returns at the daily level across the entire period For BT the picture seems to vary depending on the timeframe considered.

The Breusch-Pagan test (and the White test) are, however, not especially powerful against the specific alternative of (G)ARCH (**G**eneralised **A**uto**R**egressive **C**onditional **H**eteroscedasticity) error processes (Engle, R. F. 1982). The 1982 paper introduces a test – a LM test (actually a generalisation of the Breusch-Godfrey autocorrelation test (Breusch, T. S. 1978, Godfrey, L. G. 1978)) that is now widely known as the Engle test – specifically for ARCH errors. It involves regressing squared residuals on a number of their own lags. Engle test results at various frequencies are presented below.

Table B8 - Results of Engle test for ARCH (p-values – H_0 of homoscedasticity) – 2000+

Stock	***Monthly (2 lags, first trading day basis)***	***Weekly: Mon (8 lags)***	***Weekly: Tue (8 lags)***	***Weekly: Wed (8 lags)***	***Weekly: Thu (8 lags)***	***Weekly: Fri (8 lags)***	***Daily (40 lags)***
BT	0.007	0.000	0.000	0.003	0.000	0.000	0.000
NG	0.862	0.000	0.000	0.000	0.000	0.000	0.000
UU	0.831	0.000	0.000	0.000	0.000	0.000	0.000
SSE	0.000	0.000	0.000	0.000	0.000	0.000	0.000
PNN	0.004	0.000	0.000	0.000	0.000	0.000	0.000
SVT	0.166	0.000	0.000	0.000	0.000	0.000	0.000

Aggregation, especially to a monthly level, might be expected to reduce the appearance of (G)ARCH error processes through a form of averaging and, indeed, for NG, UU and SVT we see that homoscedasticity is not rejected at the monthly level. However, for BT, SSE and PNN the null is rejected even at the monthly level – and the null is always rejected at the weekly or daily level. Further tests for ARCH at differing time horizons are presented in Annex C4 of Appendix C below.

We thus conclude that as weekly, and even more so, daily data contain much more information than monthly data we are better off modelling at higher data frequencies and explicitly accounting for (G)ARCH error processes. An alternative approach is the Newey-West estimator (Newey, W. K. and K. D. West 1987), which subsumes White's (1980) heteroscedasticity-consistent variance-covariance matrix to include autocorrelation. The so-called HAC (**h**eteroscedasticity and **a**utocorrelation **c**onsistent) errors are consistent under quite general forms of error autocorrelation and heteroscedasticity. However, they do not specifically address the (G)ARCH form which is very well-defined: it thus seems appropriate to employ models that explicitly account for the (G)ARCH nature of the process.

Bibliography

BREUSCH, T. S. (1978): "Testing for Autocorrelation in Dynamic Linear Models*," *Australian Economic Papers*, 17, 334-355.

BREUSCH, T. S., AND A. R. PAGAN (1979): "A Simple Test for Heteroscedasticity and Random Coefficient Variation," *Econometrica*, 47, 1287-1294.

ENGLE, R. F. (1982): "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation," *Econometrica*, 50, 987-1007.

GODFREY, L. G. (1978): "Testing against General Autoregressive and Moving Average Error Models When the Regressors Include Lagged Dependent Variables," *Econometrica*, 46, 1293-1301.

NEWEY, W. K., AND K. D. WEST (1987): "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703-708.

WHITE, H. (1980): "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica*, 48, 817-838.

Appendix C GARCH(1,1) modelling

C.1 Introduction

In this appendix we build on Robertson's (2018) paper presented to Ofgem on 19 April, 2018. In particular, we use the Bayesian Information Criterion to choose a statistically "best" GARCH form and estimate rolling betas.

C.2 Donald Robertson's analysis

Robertson (2018) produced a report for Ofgem on the estimation of beta in which he presented OLS results (both static and rolling) and Diagonal BEKK(1,1) GARCH results. The current project replicated Robertson's daily and rolling daily OLS results and his GARCH results. We were unable to replicate his monthly results using the Bloomberg data provided by OFGEM and now understand that this was because Robertson only used the Bloomberg data for his daily analyses, using an older data set for the weekly and monthly analyses (Robertson, *pers. comm.*, 2018).

C.3 GARCH models

GARCH models (Engle, R. F. 1982, Engle, R. F. and K. F. Kroner 1995) explicitly incorporate an autoregressive model of the conditional variance of the error term. There is a wide range of variants, including:

- VECH (half vectorised)
- DCC (Dynamic Conditional Correlation)
- BEKK
- Triangular BEKK
- Diagonal BEKK
- Cholesky
- Diagonal

GARCH models are said to be of order (m,n) where m is the order of the autoregressive component and n is that of the moving average component. Most financial models tend to be of order (1,1) although higher orders are certainly possible.

A generic 2-series (the form we are interested in here, with the series being the returns of a utility and the returns on the All Share Index) BEKK(1,1) model has the following form:

Equation 1 - BEKK VCV Structure

$$\begin{pmatrix} \sigma_{M,t}^2 & \sigma_{iM,t} \\ \sigma_{iM,t} & \sigma_{i,t}^2 \end{pmatrix} = \begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{pmatrix} + \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} u_{M,t-1}u_{M,t-1} & u_{M,t-1}u_{i,t-1} \\ u_{M,t-1}u_{i,t-1} & u_{i,t-1}u_{i,t-1} \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$$
$$+ \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} \begin{pmatrix} \sigma_{M,t-1}^2 & \sigma_{iM,t-1} \\ \sigma_{iM,t-1} & \sigma_{i,t-1}^2 \end{pmatrix} \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix}$$

The form estimated by Robertson is a Diagonal BEKK(1,1) – the elements a_{12}, a_{21}, b_{12} and b_{21} are set to zero. This results in a readily tractable model that can provide 3 estimates of beta, as Robertson noted:

- A long-run beta based on the m_{ij}, a_{ij} and b_{ij} estimates
- An average based on the averages of the daily covariances between the asset and the market and the variance of the market
- Daily estimates based on the daily covariances between the asset and the market and the market variance

As noted above, we successfully replicated Robertson's D-BEKK(1,1) models. We desired, however, to consider a broader range of GARCH models. Using RATS v10 software, we estimated 10 different GARCH(1,1) specifications for the 6 utilities. How, then, to choose between these models? We opted for the Schwarz or Bayesian Information Criterion (BIC) (Schwarz, G. 1978). The BIC penalises models with large numbers of coefficients that do "little" to improve fit. The formula is:

Equation 2 - Bayesian Information Criterion

$$BIC = ln(n)k - 2ln(\hat{L})$$

where n is the number of observations, k the number of estimated parameters and $\hat{L}$ is the maximised value of the likelihood function.

In no case was the "best" model a D-BEKK(1,1). Results of the maximum likelihood estimation are contained in the tables below:

Table C1 – GARCH model summary for BT

Model	*Diag VECH*	*BEKK*	*CC*	*DCC*	*Asymmetric DCC*	*T-BEKK*	*D-BEKK*	*Cholesky*	*Full VECH*	*Diagonal*
LogL	-15128.35	-15137.44	-15150.88	-15128.40	-15133.69	-15138.19	-15161.16	-15129.95	-15052.01	-15872.93
N	4733	4733	4733	4733	4733	4733	4733	4733	4733	4733
Parameters	11	13	9	10	11	11	9	9	23	8
BIC	30349.79	30384.89	30377.93	30341.43	30360.46	30369.47	30398.47	30336.06	30298.66	31813.56

Table C2 - GARCH model summary for NG

Model	Diag VECH	BEKK	CC	DCC	Asymmetric DCC	T-BEKK	D-BEKK	Cholesky	Full VECH	Diagonal
LogL	-13302.38	-13291.66	-13329.22	-13294.97	-13294.73	-13293.64	-13319.30	-13305.73	-13239.05	-13879.04
N	4733	4733	4733	4733	4733	4733	4733	4733	4733	4733
Parameters	11	13	9	10	11	11	9	9	23	8
BIC	26697.85	26693.34	26734.61	26674.57	26682.55	26680.37	26714.76	26687.63	26672.74	27825.77

Table C3 - GARCH model summary for UU

Model	*Diag VECH*	*BEKK*	*CC*	*DCC*	*Asymm DCC*	*T-BEKK*	*D-BEKK*	*Cholesky*	*Full VECH*	*Diagonal*
LogL	-13611.33	-13593.47	-13685.74	-13608.68	-13608.45	-13597.63	-13623.89	-13614.16	No convergence	-14116.53
N	4733	4733	4733	4733	4733	4733	4733	4733	4733	4733
Parameters	11	13	9	10	11	11	9	9	23	8
BIC	27315.74	27296.94	27447.64	27301.98	27309.98	27288.34	27323.95	27304.49		28300.75

Table 1 - GARCH model summary for SSE

Model	Diag VECH	BEKK	CC	DCC	Asymm DCC	T-BEKK	D-BEKK	Cholesky	Full VECH	Diagonal
LogL	-13596.48	-13567.46	-13674.72	-13592.41	-13592.41	-13573.05	-13606.54	-13609.61	No convergence	-14094.23
N	4733	4733	4733	4733	4733	4733	4733	4733	4733	4733
Parameters	11	13	9	10	11	11	9	9	23	8
BIC	27286.04	27244.92	27425.61	27269.45	27277.91	27239.18	27289.24	27295.39		28256.16

Table C5 - GARCH model summary for PNN

Model	Diag VECH	BEKK	CC	DCC	Asymm DCC	T-BEKK	D-BEKK	Cholesky	Full VECH	Diagonal
LogL	-14091.64	-14088.75	-14180.42	-14087.36	-14086.01	-14094.03	-14103.17	-14128.66	-14001.37	-14439.72
N	4733	4733	4733	4733	4733	4733	4733	4733	4733	4733
Parameters	11	13	9	10	11	11	9	9	23	8
BIC	28276.36	28287.51	28437.00	28259.33	28265.10	28281.15	28282.51	28333.48	28197.38	28947.13

Table C6 - GARCH model summary for SVT

Model	Diag VECH	BEKK	CC	DCC	Asymm DCC	T-BEKK	D-BEKK	Cholesky	Full VECH	Diagonal
LogL	-13925.48	-13903.23	-14002.61	-14002.61	-13909.90	-13908.67	-13939.26	-13927.90	-13832.59	-14360.39
N	4733	4733	4733	4733	4733	4733	4733	4733	4733	4733
Parameters	11	13	9	10	11	11	9	9	23	8
BIC	27944.04	27916.47	28081.38	28089.84	27912.88	27910.43	27954.68	27931.96	27859.81	28788.48

The "statistically best" models are summarised in the table below:

Table C7 - Summary of preferred models

Model	***Number of lowest BIC cases***
T-BEKK (Triangular BEKK)	2
Full VECH (Half-vectorisation)	4

The very flexible – and often hard to estimate – Full VECH model is the most often preferred. We were only able to get these models to converge on the full 2000-2018 data set – sub-periods lacked sufficient data for success. The T-BEKK (Triangular BEKK – the matrices that pre- and post-multiply the errors and variances are lower-triangular rather than diagonal) form is second most often preferred.

We now turn to the beta estimates from these 6 GARCH models. Average values are reported in the table below and the daily values and a 500 day moving average (with the OLS rolling 500 day regression as a point of comparison) are then presented in a series of graphs.

Table C8 - Average of daily betas from preferred GARCH models

Stock	***Average of daily GARCH betas***
BT	0.99
NG	0.61
UU	0.55
SSE	0.54
PNN	0.49
SVT	0.53

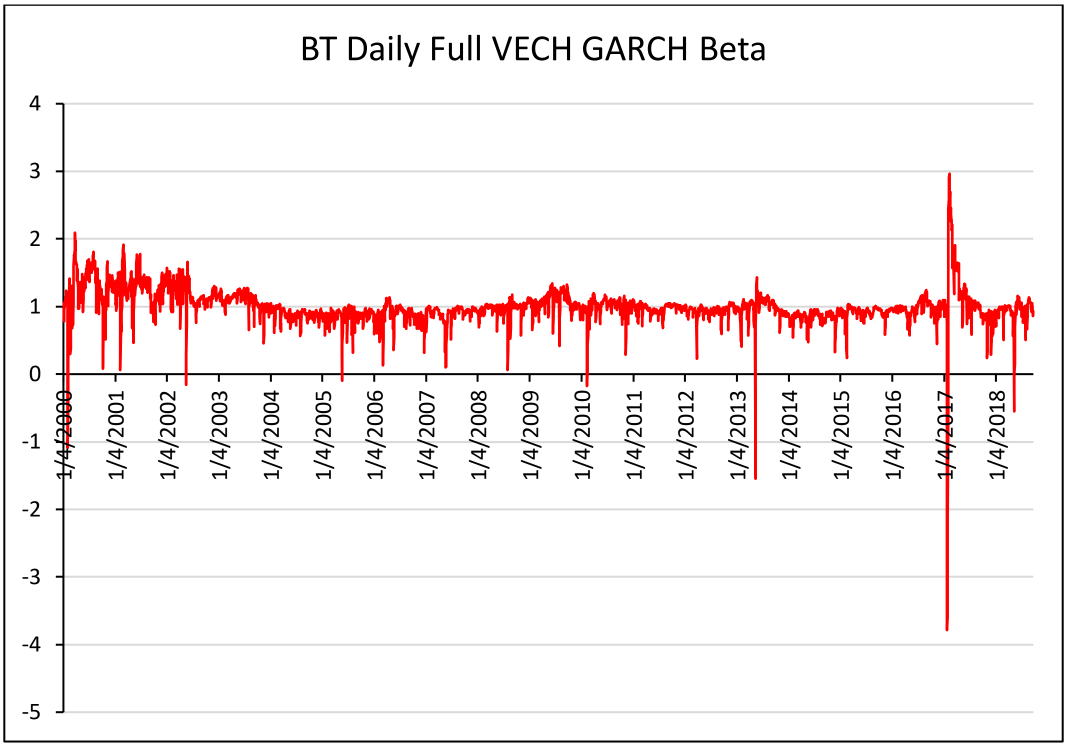


Figure C1 - BT daily GARCH beta

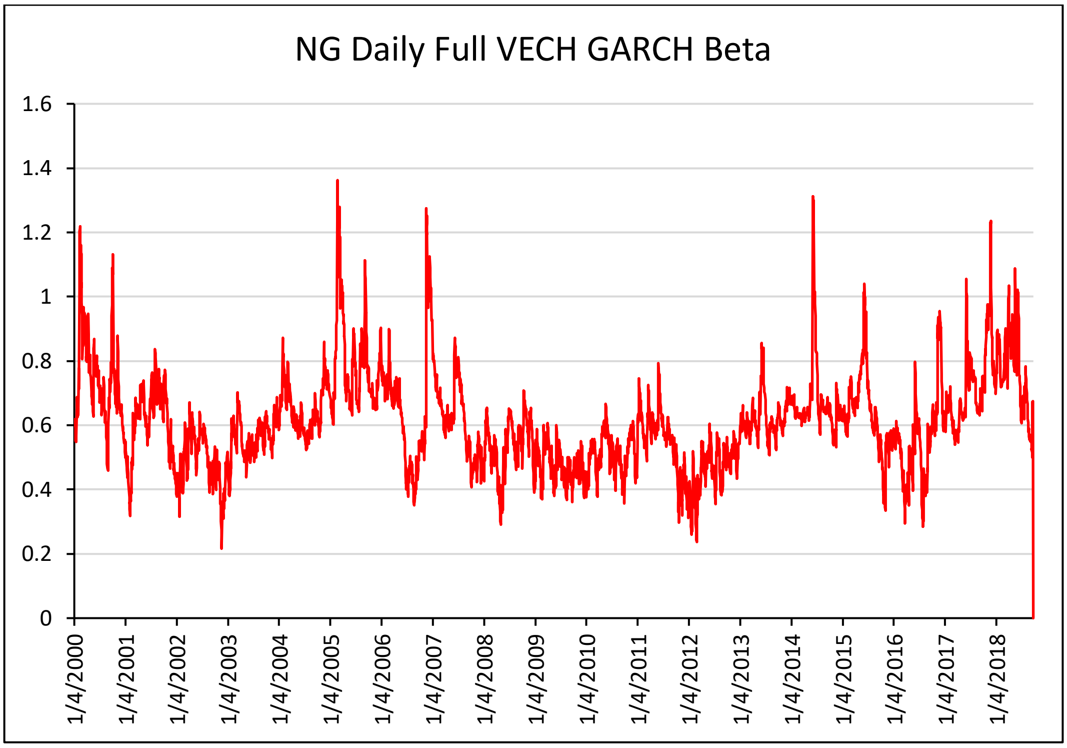


Figure C2 - NG daily GARCH beta

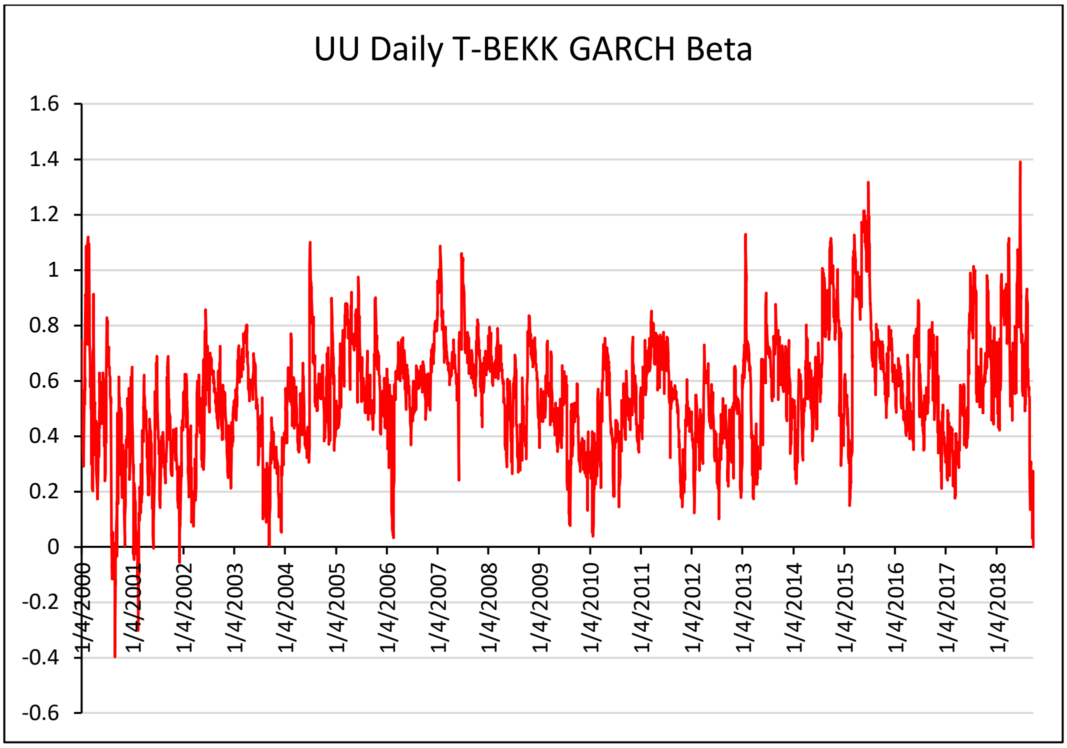


Figure C3 - UU daily GARCH beta

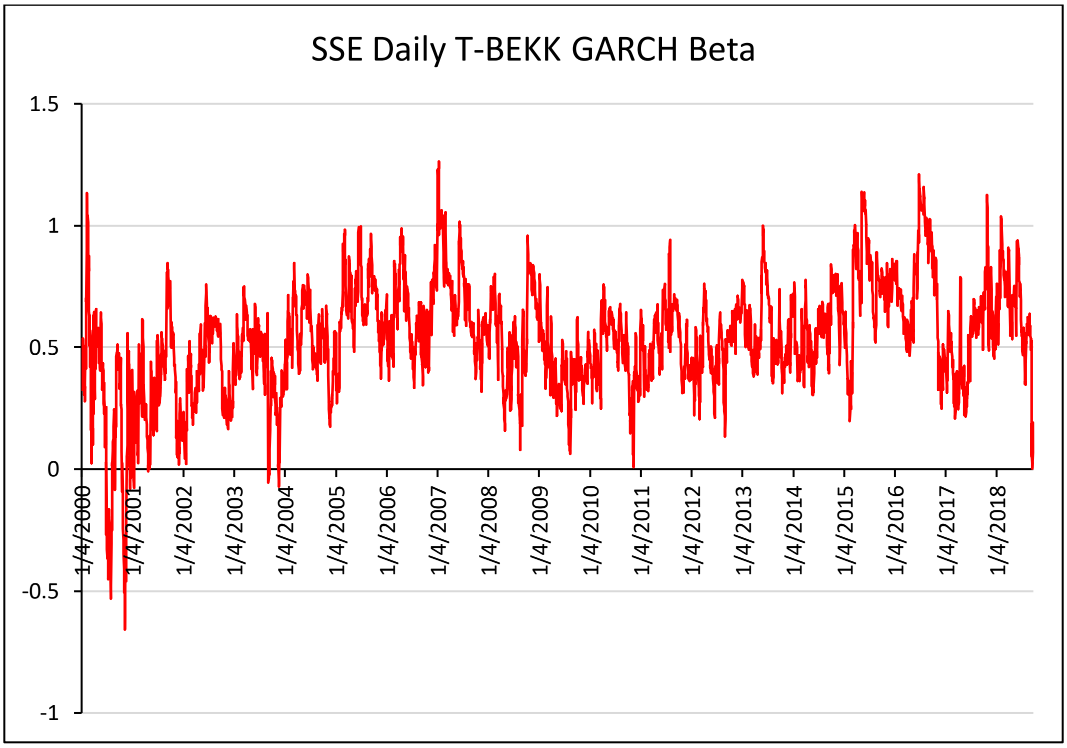


Figure C4 - SSE daily GARCH beta

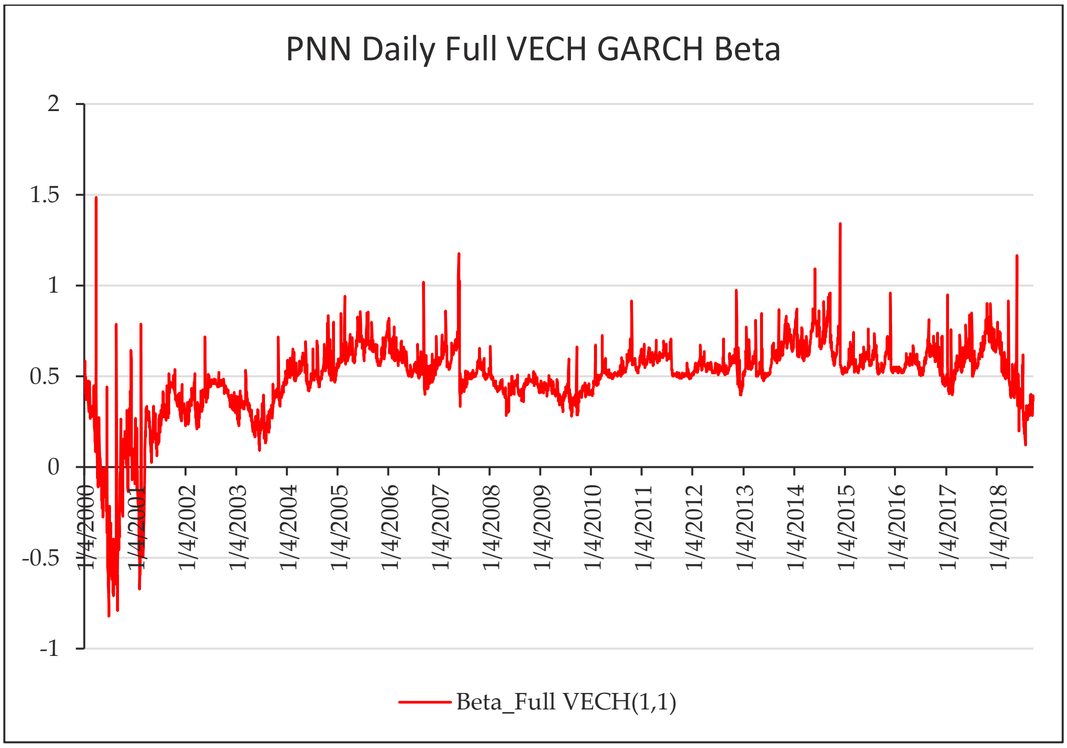


Figure C5 - PNN daily GARCH beta

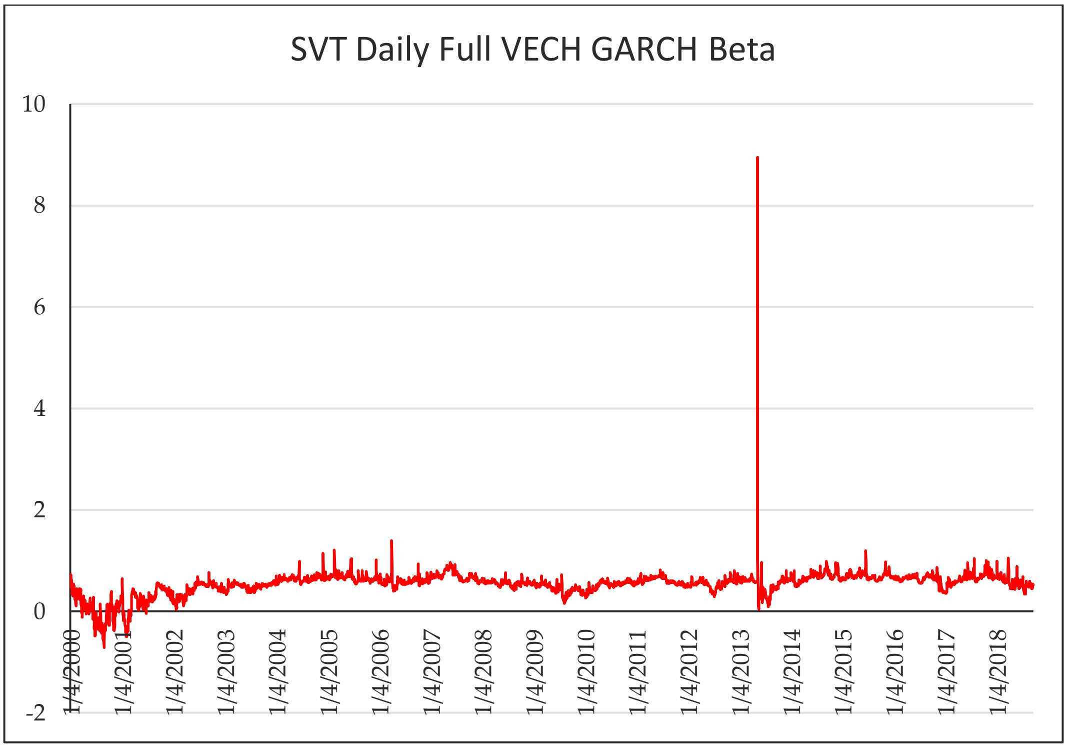


Figure C6 - SVT daily GARCH beta

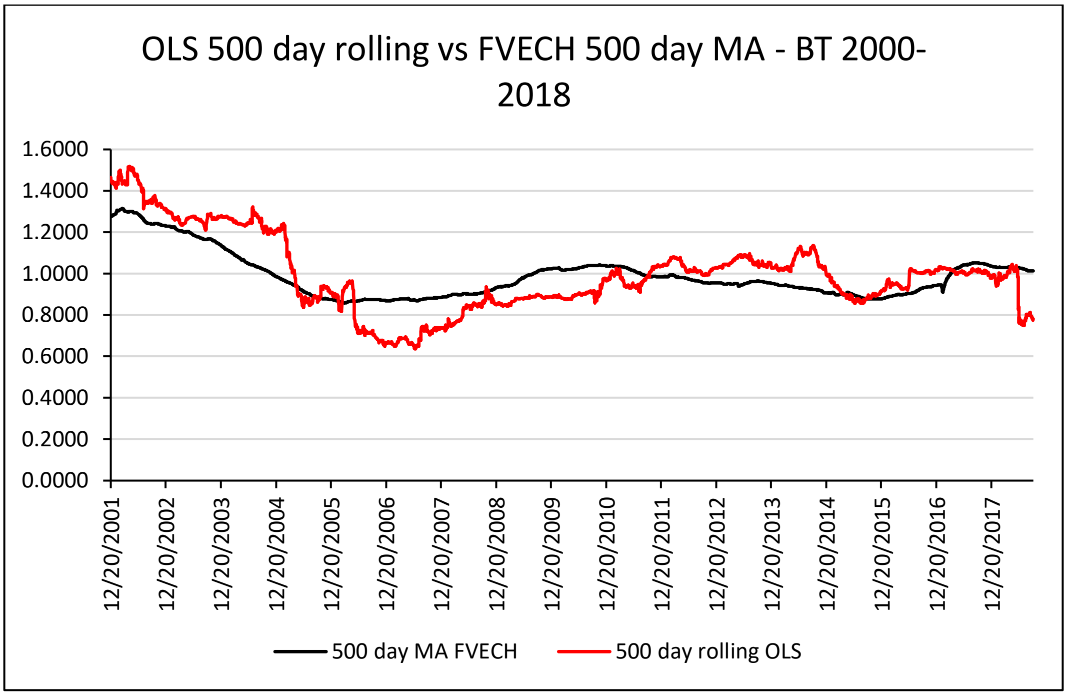


Figure C7 - BT moving average GARCH beta and rolling OLS beta

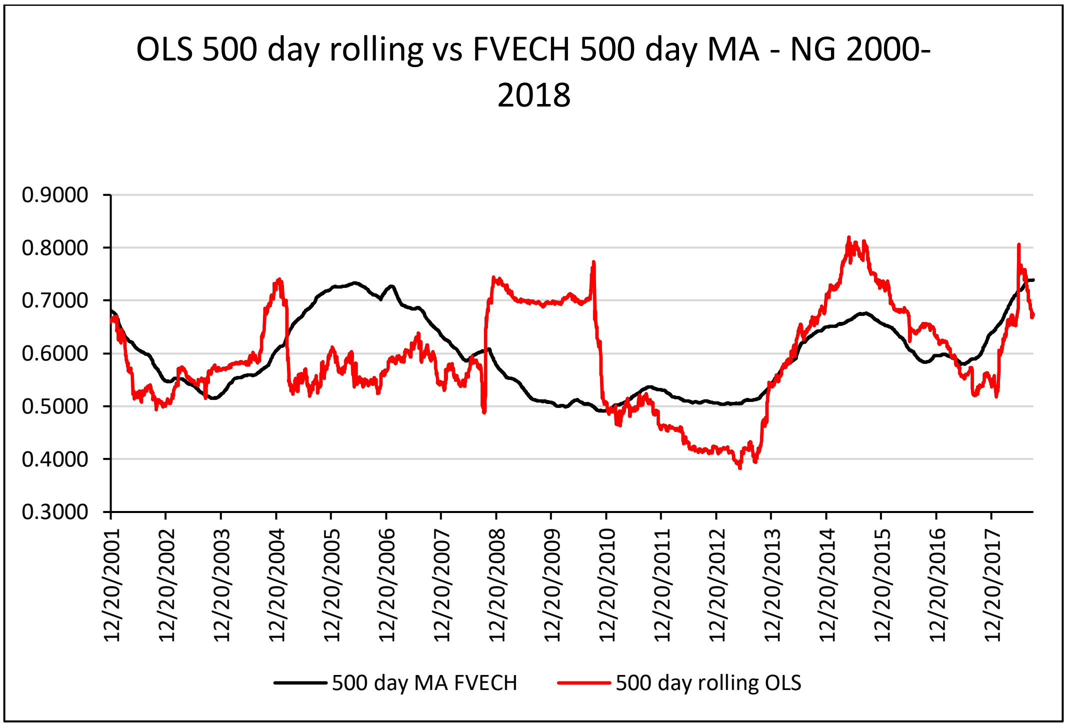


Figure C8 - NG moving average GARCH beta and rolling OLS beta

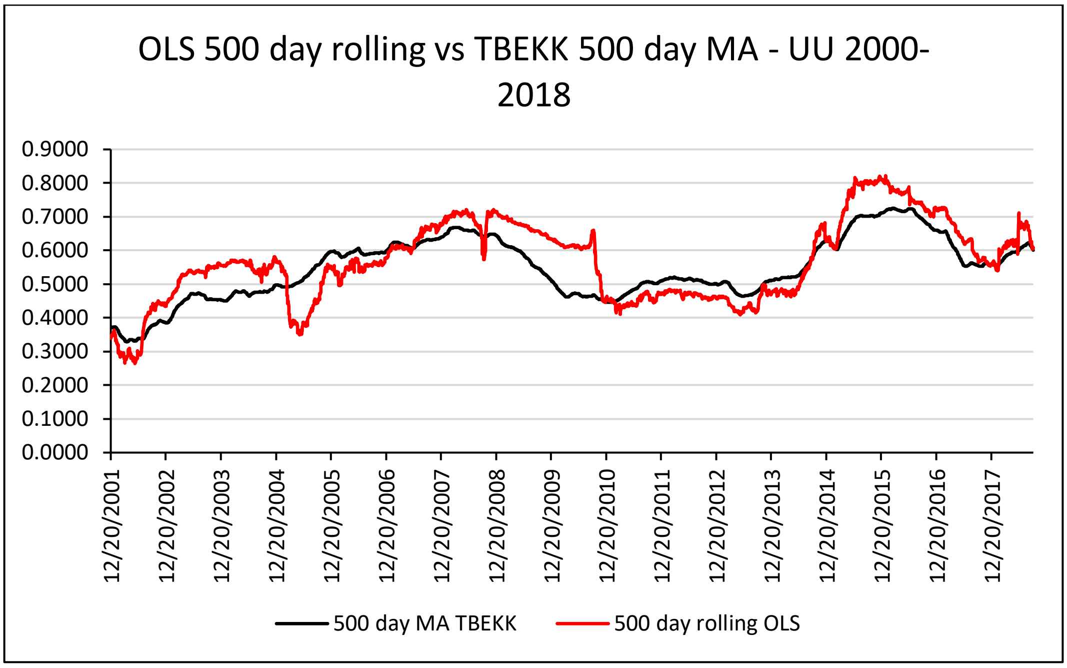


Figure C9 - UU moving average GARCH beta and rolling OLS beta

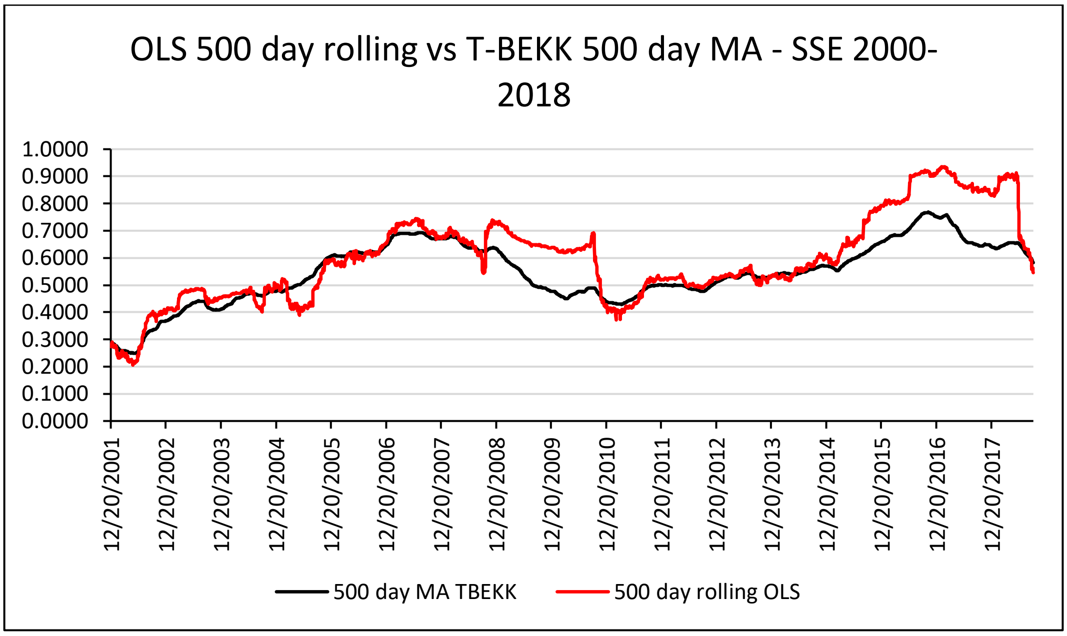


Figure C10 - SSE moving average GARCH beta and rolling OLS beta

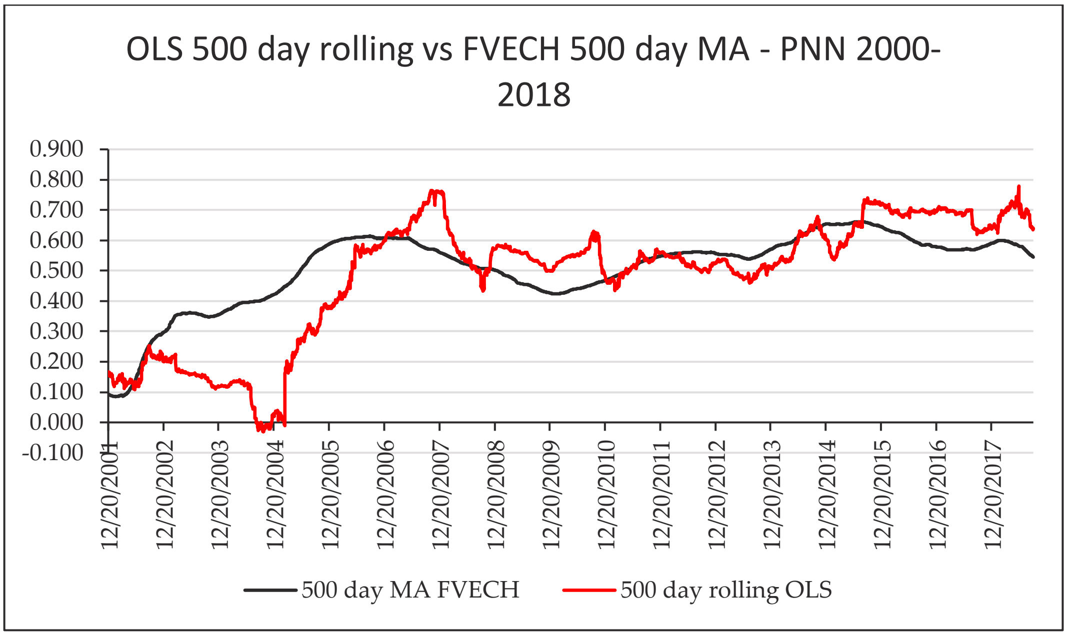


Figure C11 - PNN moving average GARCH beta and rolling OLS beta

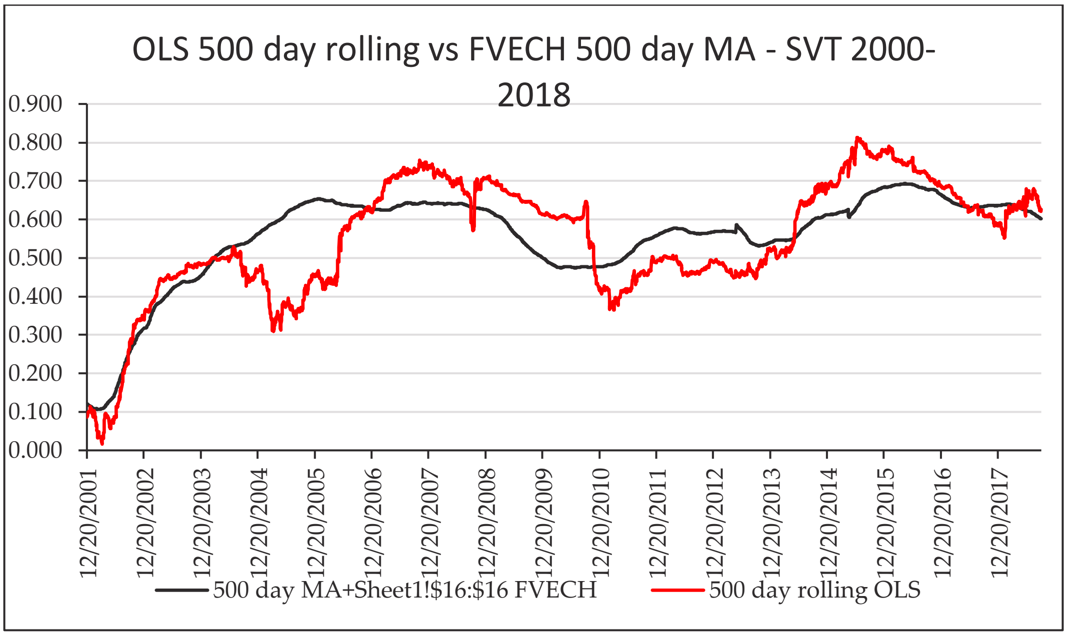


Figure C12 - SVT moving average GARCH beta and rolling OLS beta

C.4 Discussion

The moving average of the GARCH betas tracks the rolling OLS result quite closely, although it irons out the more extreme fluctuations of the latter.

Bibliography

ENGLE, R. F. (1982): "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation," *Econometrica*, 50, 987-1007.

ENGLE, R. F., AND K. F. KRONER (1995): "Multivariate Simultaneous Generalized Arch," *Econometric Theory*, 11, 122-150.

ROBERTSON, D. (2018): "Estimating B."

SCHWARZ, G. (1978): "Estimating the Dimension of a Model," *Ann. Statist.*, 6, 461-464.

Annex C1 Impact of GARCH order on beta estimates

CC1.1 Introduction

Our initial investigations of GARCH models – like those of Robertson (2018) – considered only GARCH(1,1) models. An obvious question is whether higher-order models offer a better fit. This annex investigates this issue for the six utility stocks.

Data and Methodology

Data are daily returns from Jan 2000 to Sep 2018 calculated from Bloomberg data supplied by Ofgem. This yields 4733 data points. The independent variable is the return on the ASX.

All estimation was performed in RATS v10. Testing of model residuals (standardised) was undertaken using @mvarchtest with lags of 5 and 10 and @archtest with lags of 5 and 10: the first function performs the multivariate LM test while the second is used on the two sets of standardised residuals (market return and stock return) separately.

Alternative models were compared using both the Bayesian Information Criterion (BIC) and the Akaike Information Criterion (Akaike, H. 1974, Schwarz, G. 1978). The BIC is preferred as it penalises overfitting more aggressively.

Robertson (2018) cites three ways of estimating beta from a Diagonal BEKK GARCH model (all parameters are, obviously, the estimated values thereof):

Equation 3 - Long-run beta in D-BEKK(1,1) model

$$\beta_{LR} = \frac{m_{21}/(1 - a_{11}a_{22} - b_{11}b_{22})}{m_{11}/(1 - a_{11}^2 - b_{11}^2)}$$

Equation 4 - GARCH beta estimated from average covariance and variance

$$\beta_{avs} = \frac{\frac{1}{T}\sum \sigma_{iM,t}}{\frac{1}{T}\sum \sigma_{M,t}^2}$$

Equation 5 - GARCH beta estimate as average of daily values

$$\beta_{SR} = \frac{1}{T}\sum \frac{\sigma_{iM,t}}{\sigma_{M,t}^2}$$

The first result depends on the particular form of the Diagonal BEKK model whereas the second and third do not. We thus present estimates based on Equation 4 and Equation 5.

Results

Firstly, does allowing higher order GARCH change model choice? The tables below show the BIC values for the various stocks.

Table CA1.1 - BIC values for various GARCH models: UU

Model	*Order (1,1)*	*Order (2,2)*
Diagonal VECH	27315.74	27346.92
BEKK	27296.94	27335.87
CC	27447.64	27473.80
DCC	27301.98	27326.67
Asymmetric DCC	27309.98	27334.78
T-BEKK	27288.34	27310.21
D-BEKK	27323.95	
Cholesky	27304.49	27294.09
Full VECH	27292.09	
Diagonal	28300.75	28300.28

Missing values in this and all the following tables indicate a failure of the estimation algorithm to converge. It is readily obvious that the addition of order (2,2) options does not change the model choice for UU – it remains the Triangular BEKK(1,1) model

Table CA1.2 - BIC values for various GARCH models: SVT

Model	*Order (1,1)*	*Order (2,2)*
Diagonal VECH	27944.04	27928.70
BEKK	27916.47	27915.29
CC	28081.38	
DCC	28089.84	27876.46
Asymmetric DCC	27912.88	27881.28
T-BEKK	27910.43	27925.30
D-BEKK	27954.68	
Cholesky	27931.96	27947.58
Full VECH	27859.81	
Diagonal	28788.48	28759.00

Opening up higher order options does not cause a shift in “preferred” model for SVT.

Table CA1.3 - BIC values for various GARCH models: BT

Model	*Order (1,1)*	*Order (2,2)*
Diagonal VECH	30349.79	30376.30
BEKK	30384.89	30322.45
CC	30377.93	30405.89
DCC	30341.43	30368.35
Asymmetric DCC	30360.46	30385.41
T-BEKK	30369.47	30316.51
D-BEKK	30398.47	30362.76
Cholesky	30336.06	30350.03
Full VECH	30295.89	
Diagonal	31813.56	31841.67

The "preferred" model for BT remains order (1,1) even given the availability of higher order GARCH options.

Table CA1.4 - BIC values for various GARCH models: NG

Model	*Order (1,1)*	*Order (2,2)*
Diagonal VECH	26697.85	26701.20
BEKK	26693.34	26690.42
CC	26734.61	26730.41
DCC	26674.57	26671.59
Asymmetric DCC	26682.55	26679.84
T-BEKK	26680.37	26668.95
D-BEKK	26714.76	26715.56
Cholesky	26687.63	26667.72
Full VECH	26672.74	
Diagonal	27825.77	27825.31

The "preferred" model for NG changes when higher order options are available. The Cholesky model, however, has the undesirable property that estimation order – whether the model is (ASX, utility) or (utility, ASX) – affects results, rendering it somewhat problematic.

Table CA1.5 - BIC values for various GARCH models: SSE

Model	*Order (1,1)*	*Order (2,2)*
Diagonal VECH	27286.04	27270.55
BEKK	27244.92	27268.21
CC	27425.61	
DCC	27269.45	
Asymmetric DCC	27277.91	27256.12
T-BEKK	27239.18	27252.38
D-BEKK	27289.24	
Cholesky	27295.39	27283.23
Full VECH	27258.88	
Diagonal	28256.16	28233.33

The "preferred" model for SSE is unchanged by the inclusion of (2,2) options.

Table CA1.6 - BIC values for various GARCH models: PNN

Model	*Order (1,1)*	*Order (2,2)*
Diagonal VECH	28276.36	28170.20
BEKK	28287.51	28270.33
CC	28437.00	28294.43
DCC	28259.33	28144.53
Asymmetric DCC	28265.10	28150.21
T-BEKK	28281.15	28264.55
D-BEKK	28282.51	28264.66
Cholesky	28333.48	28225.60
Full VECH	28197.38	
Diagonal	28947.13	28831.83

The "preferred" model for PNN changes when (2,2) options are available.

In the cases of NG and PNN, the extension of options to include second order GARCH results in a switch of chosen model as shown in the table below.

Table CA1.7 - Impact of greater model order options on preferred model

Stock	*Best model restricted to (1,1)*	*Best model including (2,2)*
BT	Full VECH(1,1)	Full VECH(1,1)
NG	Full VECH(1,1)	Cholesky(2,2)
UU	T-BEKK(1,1)	T-BEKK(1,1)
SSE	T-BEKK(1,1)	T-BEKK(1,1)
PNN	Full VECH(1,1)	DCC(2,2)
SVT	Full VECH(1,1)	Full VECH(1,1)

In many cases, the differences in BIC values across models are small and other specifications might be regarded as acceptable on that basis. Here we work solely and strictly with the "lowest BIC" criterion.

As seen above, the form of GARCH model chosen varies according to the level of order permitted. Results are summarised in the following table:

Table CA1.8 - Model forms chosen by order

Model type	*Only order (1,1)*	*Order (1,1) and (2,2)*
Triangular BEKK	2	2
DCC		1
Full VECH	4	2
Cholesky		1

What of the (standardised) residuals from the chosen models: do they exhibit any residual (G)ARCH character? Tests are reported below:

Table CA1.9 - Tests for residual (G)ARCH behaviour for BT: p-values

Test	*BT Full VECH(1,1)*
Multivariate ARCH – 5 lags	0.20
Multivariate ARCH – 10 lags	0.44
Univariate ARCH – market – 5 lags	0.59
Univariate ARCH – market – 10 lags	0.53
Univariate ARCH – stock – 5 lags	0.95
Univariate ARCH – stock – 10 lags	0.99

There is no evidence of (G)ARCH behaviour in the standardised residuals from the BT model.

Table CA1.10 - Tests for residual (G)ARCH behaviour for NG: p-values

Test	*NG Full VECH(1,1)*	*NG Cholesky(2,2)*
Multivariate ARCH – 5 lags	0.59	0.01
Multivariate ARCH – 10 lags	0.80	0.04
Univariate ARCH – market – 5 lags	0.66	0.47
Univariate ARCH – market – 10 lags	0.69	0.63
Univariate ARCH – stock – 5 lags	0.54	0.74
Univariate ARCH – stock – 10 lags	0.67	0.86

In the “preferred” order (2,2) model there is a conflict between the univariate and multivariate (G)ARCH tests for the standardised residuals. We would interpret this as favouring a finding of no residual (G)ARCH behaviour.

Table CA1.11 - Tests for residual (G)ARCH behaviour for UU: p-values

Test	*UU T-BEKK(1,1)*
Multivariate ARCH – 5 lags	0.00
Multivariate ARCH – 10 lags	0.00
Univariate ARCH – market – 5 lags	0.39
Univariate ARCH – market – 10 lags	0.35
Univariate ARCH – stock – 5 lags	0.01
Univariate ARCH – stock – 10 lags	0.10

There is some evidence of residual (G)ARCH behaviour in the standardised residuals from the UU model: an order (1,2) or (2,1) model might be tried.

Table CA1.12 - Tests for residual (G)ARCH behaviour for SSE: p-values

Test	*SSE T-BEKK(1,1)*
Multivariate ARCH – 5 lags	0.00
Multivariate ARCH – 10 lags	0.35
Univariate ARCH – market – 5 lags	0.15
Univariate ARCH – market – 10 lags	0.17
Univariate ARCH – stock – 5 lags	0.10
Univariate ARCH – stock – 10 lags	0.47

There is limited conflict between the multivariate and univariate tests: we would interpret this as suggesting an acceptable model for SSE.

Table CA1.13 - Tests for residual (G)ARCH behaviour for PNN: p-values

Test	*PNN Full VECH(1,1)*	*PNN DCC(2,2)*
Multivariate ARCH – 5 lags	0.17	0.04
Multivariate ARCH – 10 lags	0.17	0.04
Univariate ARCH – market – 5 lags	0.44	0.39
Univariate ARCH – market – 10 lags	0.42	0.35
Univariate ARCH – stock – 5 lags	0.28	0.27
Univariate ARCH – stock – 10 lags	0.65	0.56

While there remains some conflict between the multivariate and univariate tests for the DCC(2,2) model, the overall impression is of an acceptably corrected error process.

Table CA1.14 - Tests for residual (G)ARCH behaviour for SVT: p-values

Test	*SVT Full VECH(1,1)*
Multivariate ARCH – 5 lags	0.95
Multivariate ARCH – 10 lags	0.95
Univariate ARCH – market – 5 lags	0.57
Univariate ARCH – market – 10 lags	0.47
Univariate ARCH – stock – 5 lags	0.99
Univariate ARCH – stock – 10 lags	0.99

For the SVT model, the residual behaviour seems most acceptable.

Does model order actually have a noticeable impact on estimated beta values? The tables below show the results for the various models:

Table CA1.15 - Estimates of beta from various models: BT

Model	*Beta (s.e. in parentheses where available)*
OLS	1.040 (0.023)
OLS (HAC errors)	1.040 (0.030)
LAD	1.000 (0.023)
Beta from averages[7] Full VECH(1,1)	1.010
Average of betas[8] Full VECH(1,1)	0.991

To one decimal place, the estimates are all 1.0. To two decimal places, they range from 0.99 to 1.04.

Table CA1.16 - Estimates of beta from various models: NG

Model	*Beta (s.e. in parentheses where available)*
OLS	0.614 (0.015)
OLS (HAC errors)	0.614 (0.027)
LAD	0.559 (0.015)
Beta from averages Full VECH(1,1)	0.584
Average of betas Full VECH (1,1)	0.610
Beta from averages Cholesky(2,2)	0.576
Average of betas Cholesky(2,2)	0.576

To one decimal place, the estimates are all 0.6. To two decimal places, they range from 0.56 to 0.61.

Table CA1.17 - Estimates of beta from various models: UU

Model	*Beta (s.e. in parentheses where available)*
OLS	0.570 (0.016)
OLS (HAC errors)	0.570 (0.022)
LAD	0.571 (0.016)
Beta from averages T-BEKK(1,1)	0.564
Average of betas T-BEKK(1,1)	0.546

To one decimal place, the estimates are 0.6, 0.6, 0.6, 0.6 and 0.5. To two decimal places, they range from 0.55 to 0.57.

[7] See Equation 4

[8] See Equation 5

Table CA1.18 - Estimates of beta from various models: SSE

Model	*Beta (s.e. in parentheses where available)*
OLS	0.569 (0.016)
OLS (HAC errors)	0.569 (0.027)
LAD	0.544 (0.016)
Beta from averages T-BEKK(1,1)	0.568
Average of betas T-BEKK(1,1)	0.537

To one decimal place, the estimates are 0.6, 0.6, 0.5, 0.6 and 0.5. To two decimal places, they range from 0.54 to 0.57.

Table CA1.19 - Estimates of beta from various models: PNN

Model	*Beta (s.e. in parentheses where available)*
OLS	0.449 (0.018)
OLS (HAC errors)	0.449 (0.025)
LAD	0.438 (0.018)
Beta from averages Full VECH(1,1)	0.454
Average of betas Full VECH(1,1)	0.486
Beta from averages DCC(2,2)	0.433
Average of betas DCC(2,2)	0.490

To one decimal place, the estimates are 0.4, 0.4, 0.4, 0.5, 0.5, 0.4 and 0.5. To two decimal places, they range from 0.43 to 0.50.

Table CA1.20 - Estimates of beta from various models: SVT

Model	*Beta (s.e. in parentheses where available)*
OLS	0.529 (0.017)
OLS (HAC errors)	0.529 (0.026)
LAD	0.524 (0.017)
Beta from averages Full VECH(1,1)	0.523
Average of betas Full VECH (1,1)	0.533

If we round to one decimal place we have 0.5 throughout. To two decimal places, the range is from 0.52 to 0.53.

It would thus appear that neither choice of estimation strategy nor order of GARCH model permitted has a major impact on the estimated beta. See the table below:

Table CA1.21 - Variability of beta estimate by estimation method

Stock	*Beta range (2 d.p.)*	*Beta range (1 d.p.)*
BT	0.99-1.04	1.0-1.0

Stock	*Beta range (2 d.p.)*	*Beta range (1 d.p.)*
NG	0.56-0.61	0.6-0.6
UU	0.55-0.57	0.5-0.6
SSE	0.54-0.57	0.5-0.6
PNN	0.43-0.50	0.4-0.5
SVT	0.52-0.53	0.5-0.5

It should be noted that the OLS adjusted R^2 values are not high even though these are time series models (SVT - 0.166, BT - 0.299, NG - 0.254, UU - 0.212, SSE - 0.205, PNN - 0.112). These relatively low numbers could be taken to suggest that the CAPM is far from a perfect model to understand daily returns relationships.

Similarly, the GARCH models reported have a simple model for the mean return in each series: it is constant. More complex models of the time series mean within the GARCH model are feasible, including both ARIMA and regression forms. Time constraints in the current project preclude consideration of all options, although it is deemed probable that the impact of further elaboration of the GARCH models on usable beta estimates will be negligible given the limited impact seen above.

Finally, we look at the daily beta estimates and their one-year (250 observation) moving averages for the various models.

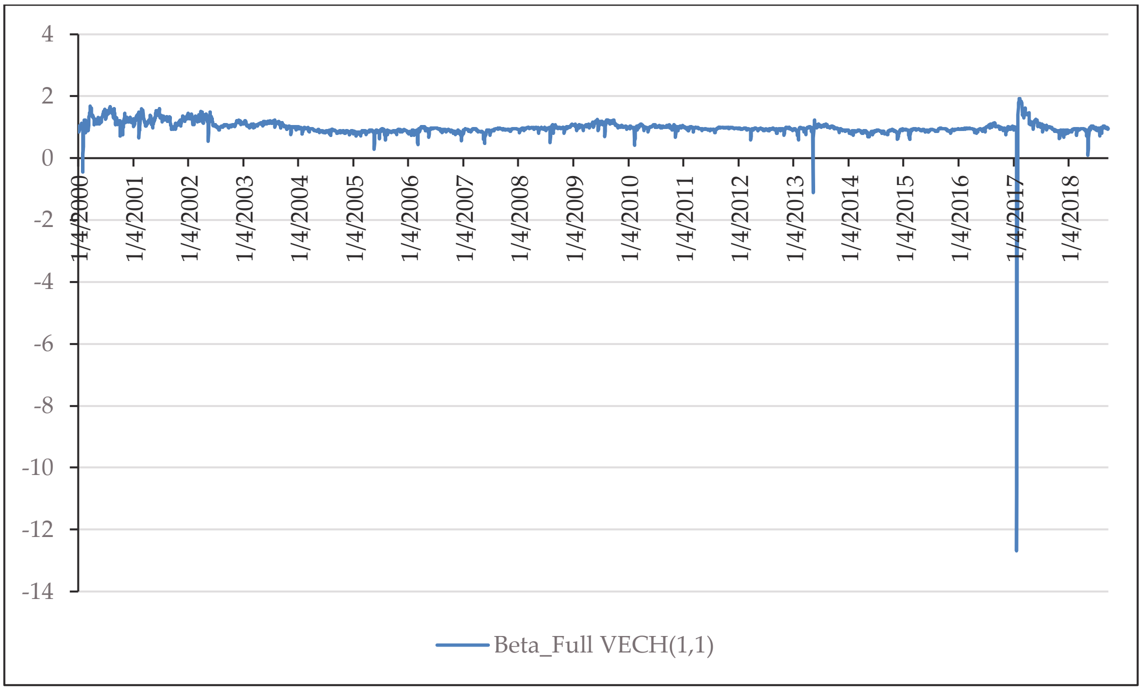


Figure CA1.1 - BT daily betas from Full VECH(1,1) model

The sharp dips in beta in May 2013 and January 2017 are interesting: it is unclear whether they are an artefact of the estimation process or represent some real event involving BT. The 2017 dip is profound enough that it shows up in the MA 250 representation as well.

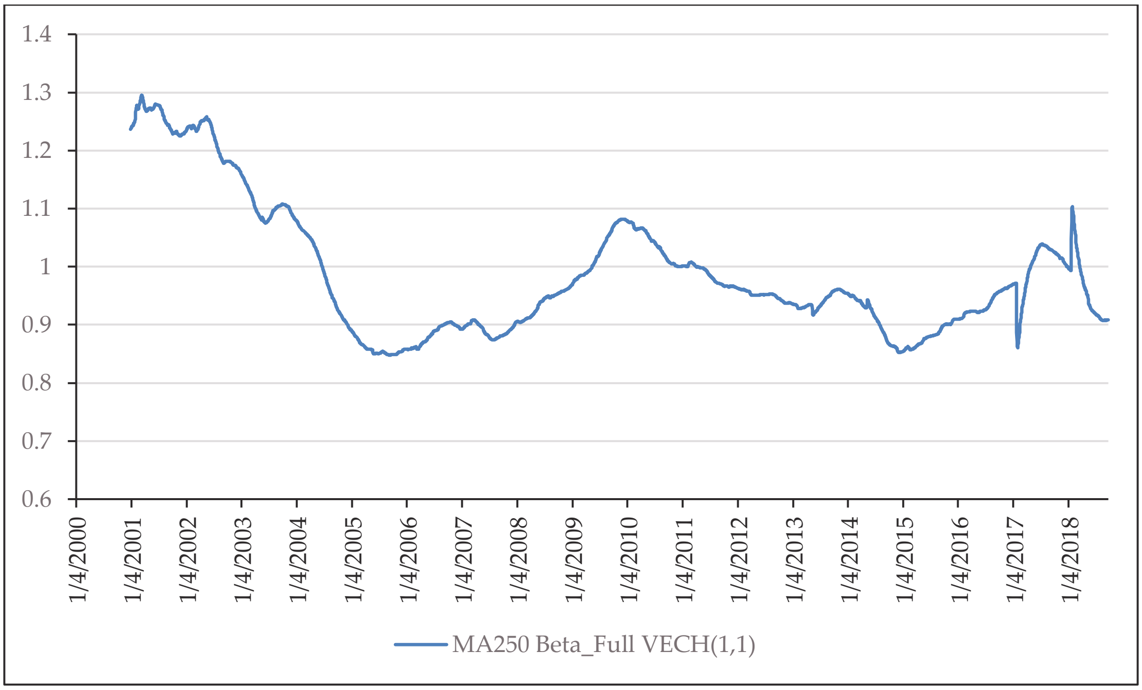


Figure CA1.2 - BT MA 250 beta from Full VECH(1,1) model

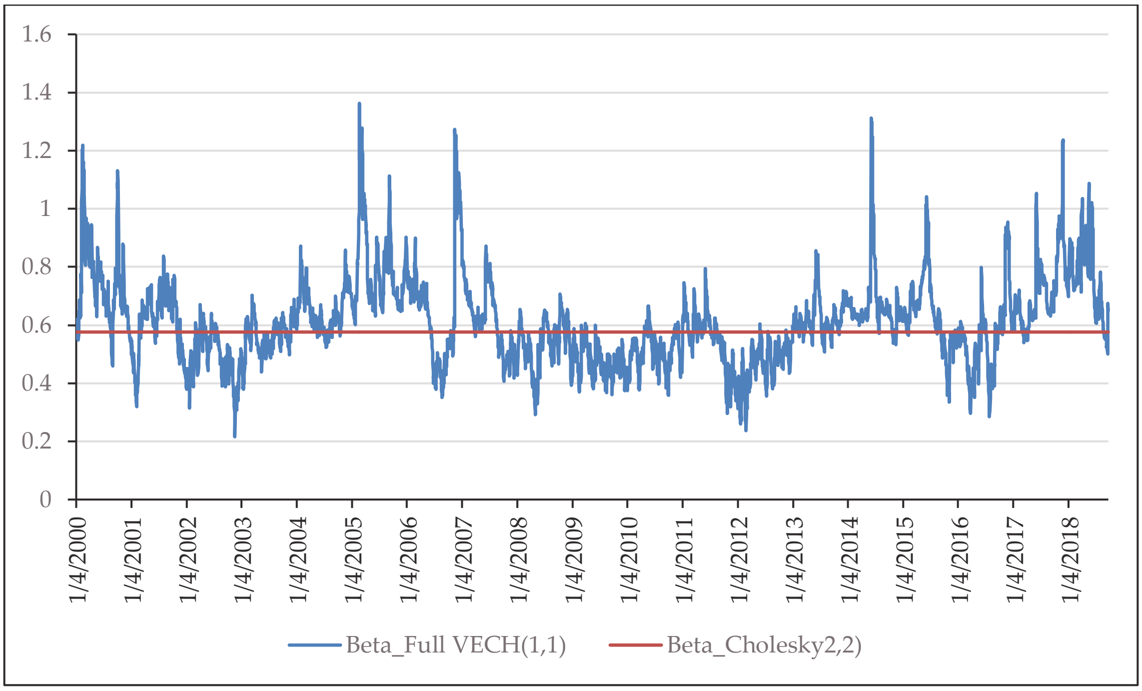


Figure CA1.3 - NG daily betas from Full VECH(1,1) and Cholesky(2,2) models

The behaviour of the Cholesky(2,2) daily beta is strange – it is constant in its first two decimal places. The model has been estimated twice and the results are the same each time. It would appear that this is a feature of the Cholesky model – the estimate is similarly all-but-constant (at a different level to NG) when a Cholesky(2,2) model is estimated for UU.

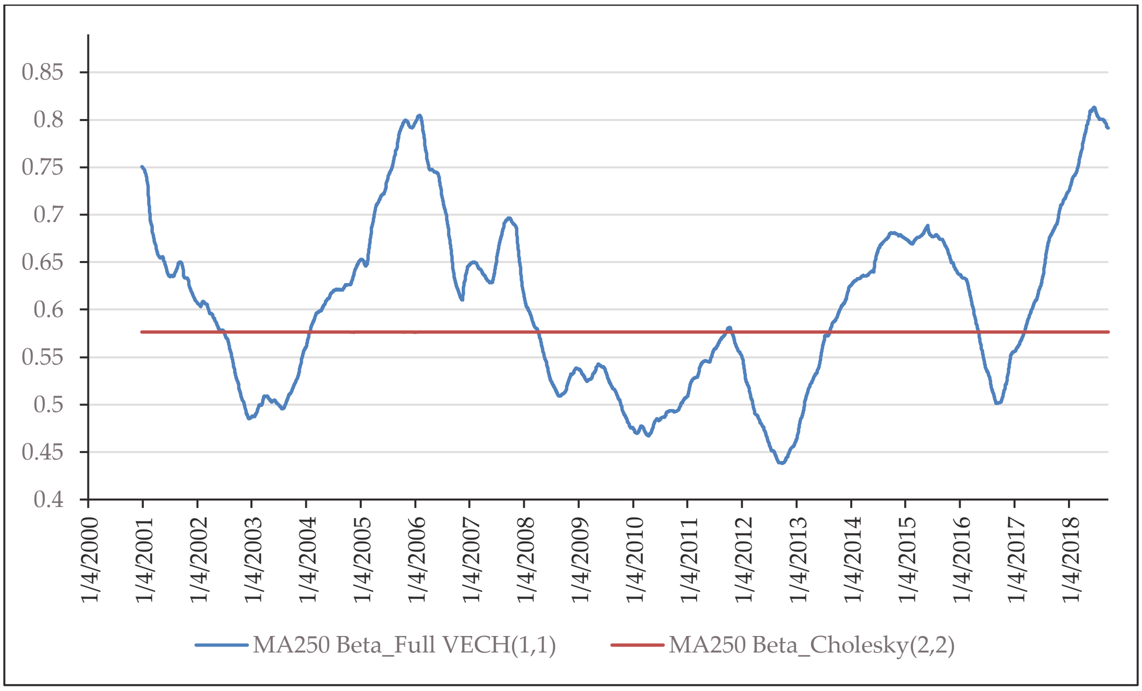


Figure CA1.4 - NG MA 250 betas from Full VECH(1,1) and Cholesky(2,2) models

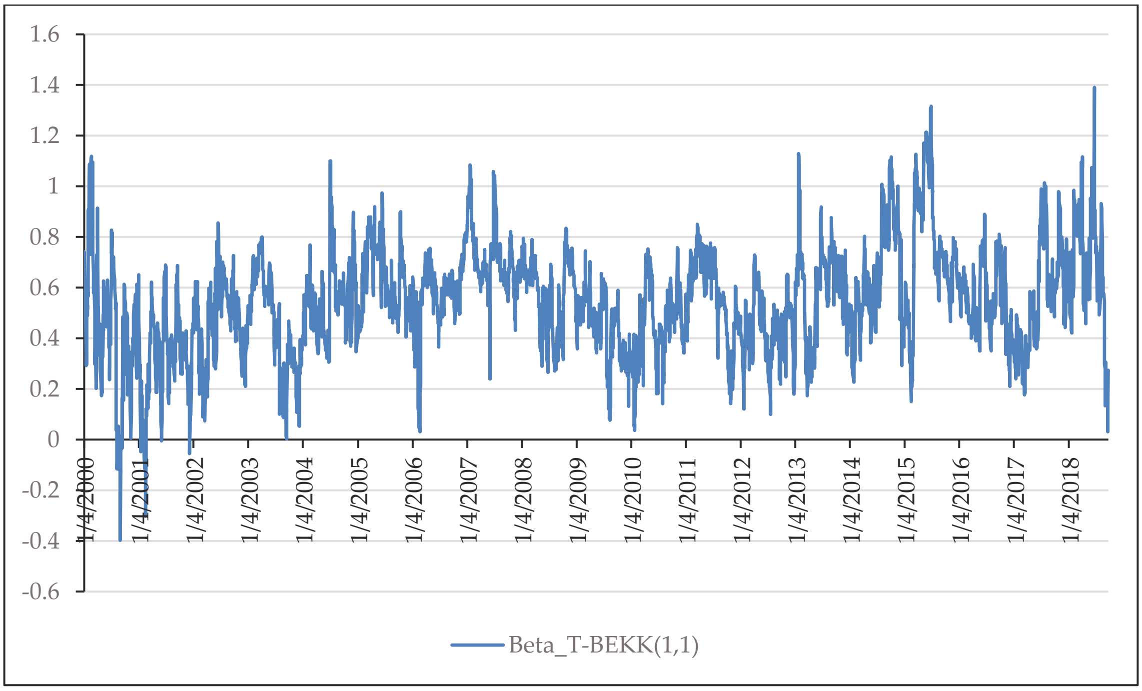


Figure CA1.5 - UU daily betas from T-BEKK(1,1) model

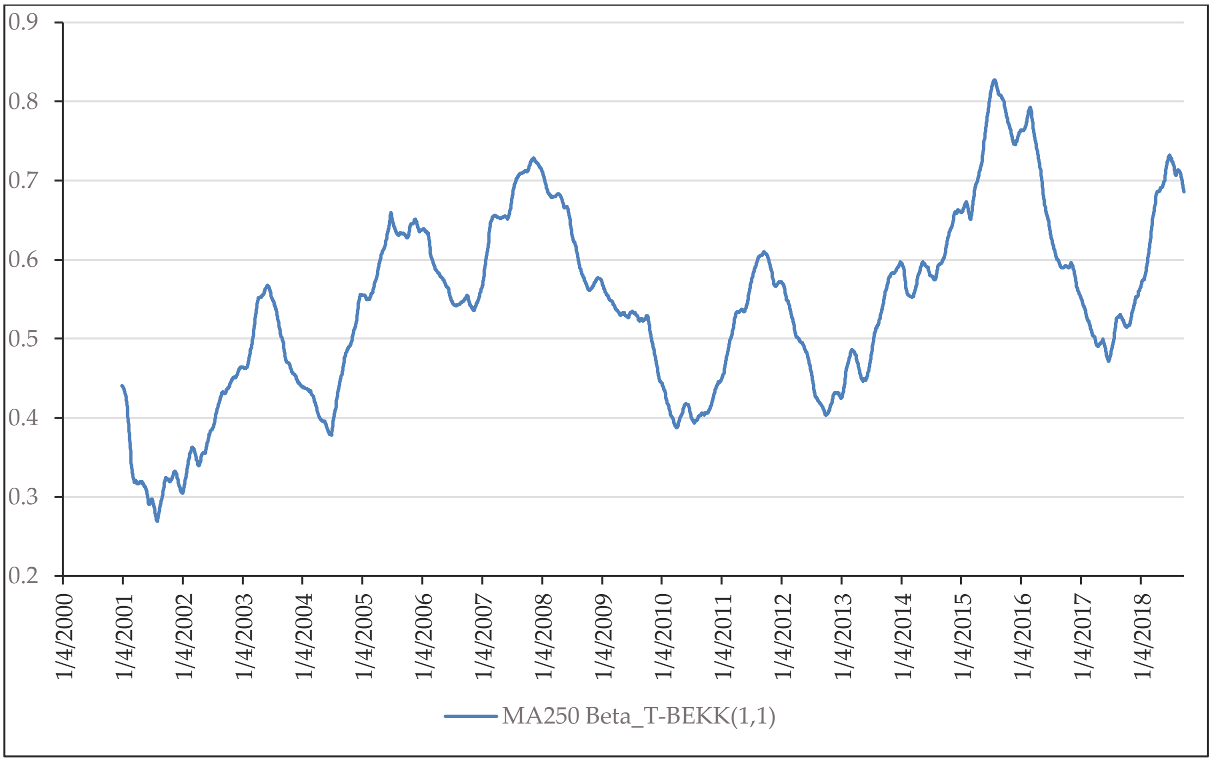


Figure CA1.6 - UU MA 250 beta from T-BEKK(1,1) model

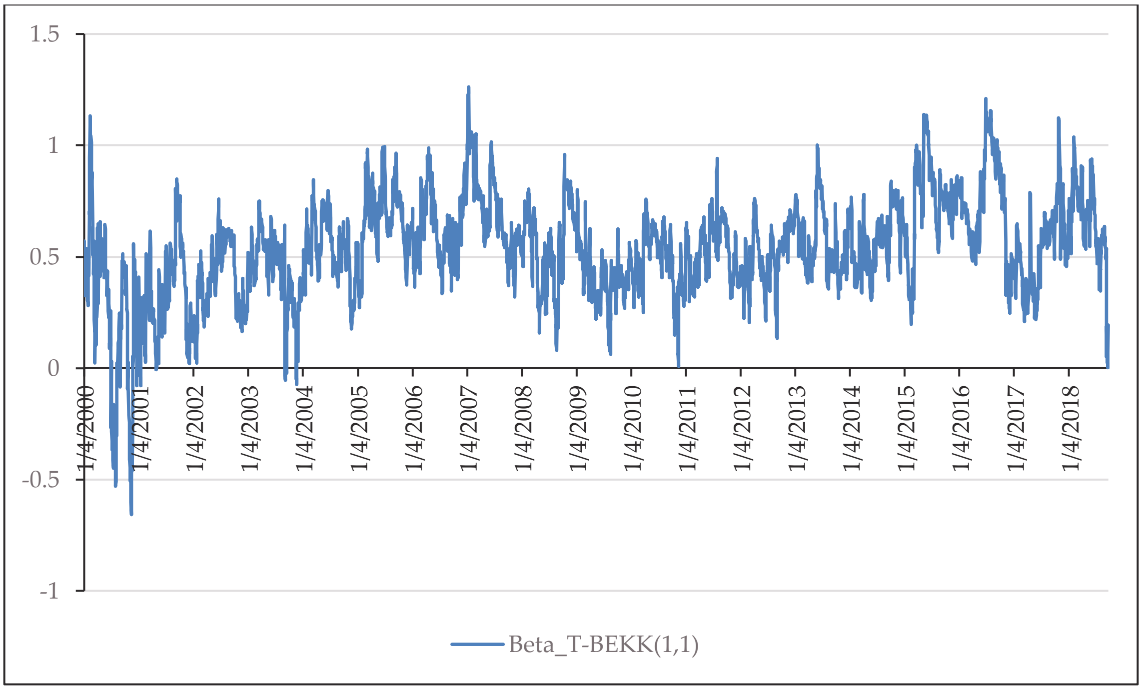


Figure CA1.7 - SSE daily betas from T-BEKK(1,1) model

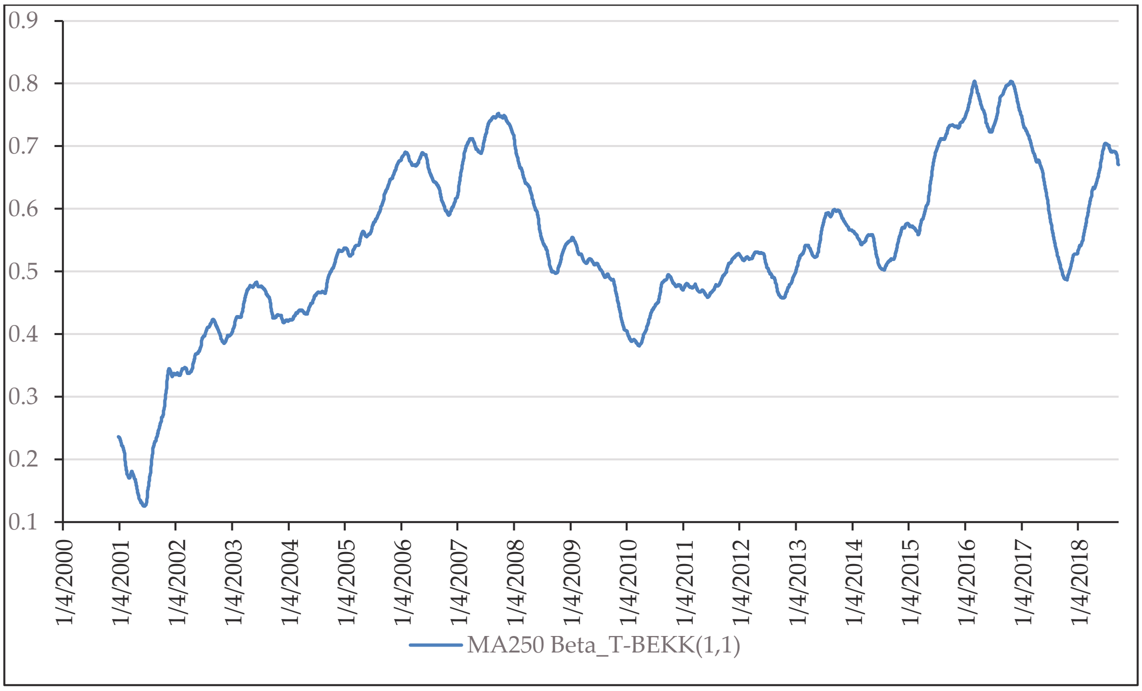


Figure CA1.8 - SSE MA 250 beta from T-BEKK(1,1) model

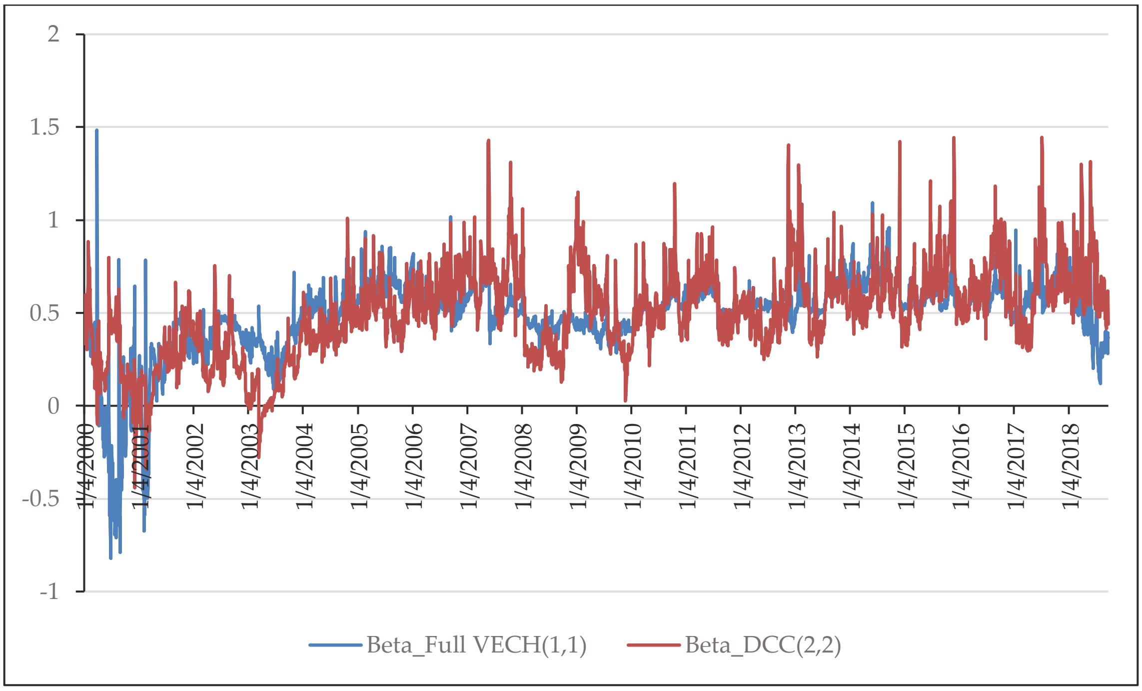


Figure CA1.9 - PNN daily betas from Full VECH(1,1) and DCC(2,2) models

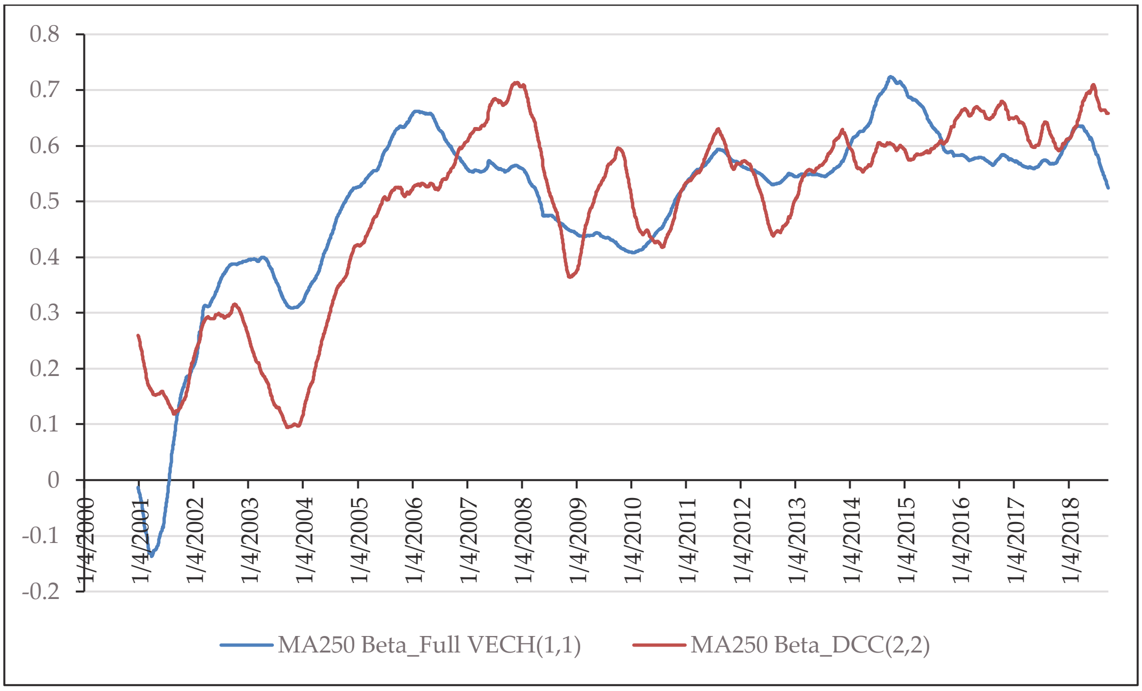


Figure CA1.10 - PNN MA 250 betas from Full VECH(1,1) and DCC(2,2) models

The moving average representations of the two PNN models are really rather similar.

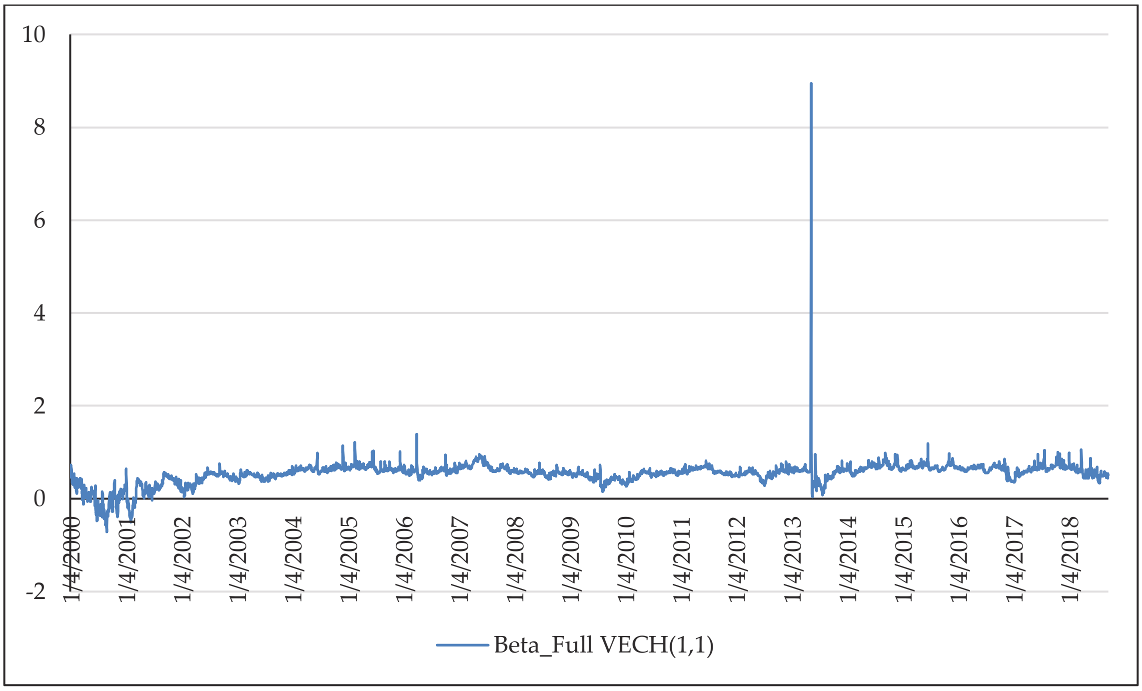


Figure CA1.11 - SVT daily betas from Full VECH(1,1) model

It is unclear what happened in May 2013 – whether it is a statistical artefact or represents an event with impact on SVT.

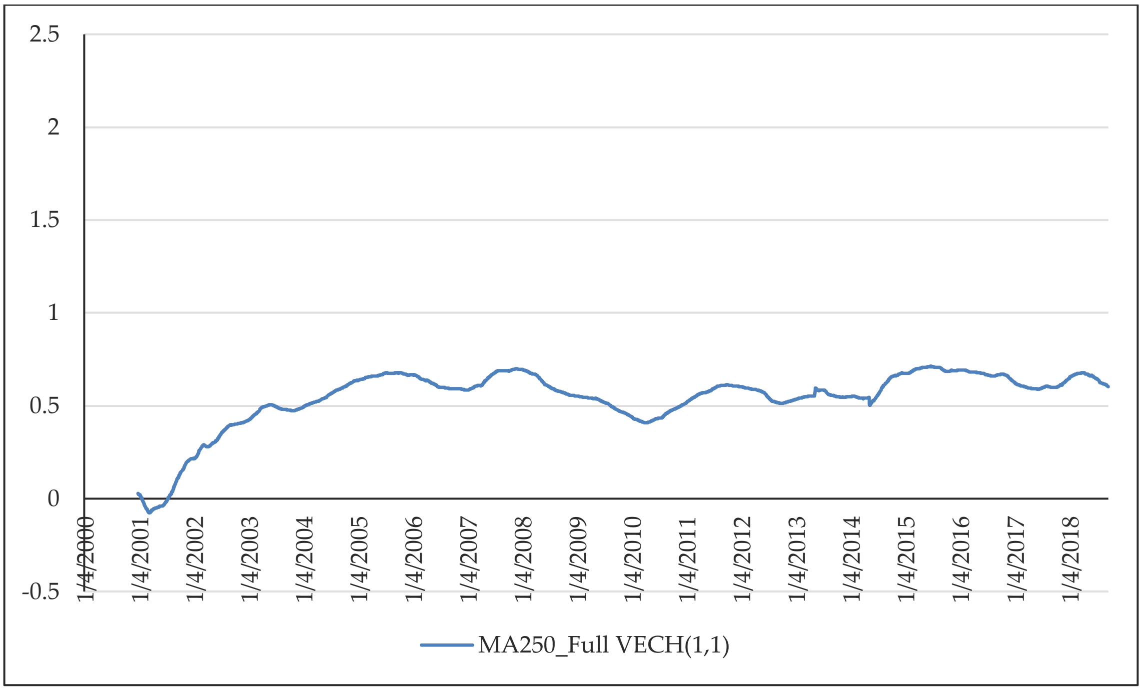


Figure CA1.12 - SVT MA 250 beta from Full VECH(1,1) model

CC1.2 Discussion

We have seen that choice of estimation approach and of GARCH order have little impact at the level of actionable results for beta. However, it is obviously best to arrive at a completely stable model with no evidence of residual (G)ARCH behaviour.

When choice is restricted to order (1,1) models, the "best" (minimum BIC) models for the six utilities are always Full VECH or Triangular BEKK. When the range is extended to include order (2,2) models we see that in four cases – BT, UU, SSE and SVT – the "best" model remains the same whereas in two cases – NG and PNN – the preferred model changes to one from the (2,2) order group.

In almost all cases the standardised residuals (divided by the estimated GARCH variances) show no further evidence of (G)ARCH behaviour, suggesting that the models have mostly successfully addressed the issue.

Finally, it should be noted that some authors have issues with the DCC model – as chosen for PNN when order (2,2) is included (see, e.g., Caporin, M. and M. McAleer 2013). Further, issues arise with the Cholesky model which means we are predisposed to dropping this from the set of specifications. As such, in the main report we will focus only on order (1,1) models.

In this research, we have considered only statistical "fit" as measured by the BIC as a primary indicator of appropriateness, with tests of residual ARCH behaviour as a secondary indicator.

Bibliography

AKAIKE, H. (1974): "A New Look at the Statistical Model Identification," *IEEE Transactions on Automatic Control*, 19, 716-723.

CAPORIN, M., AND M. MCALEER (2013): "Ten Things You Should Know About the Dynamic Conditional Correlation Representation," *Econometrics*, 1.

ROBERTSON, D. (2018): "Estimating B (Report to Ofgem)."

SCHWARZ, G. (1978): "Estimating the Dimension of a Model," *Ann. Statist.*, 6, 461-464.

Annex C2 GARCH model choice and estimation period

CC2.1 Aim

To assess whether the estimation period affects the choice of "best" GARCH model.

CC2.2 Methodology

All models were run in RATS v10. Data were daily returns on the ASX and the six utilities, calculated from Bloomberg data provided by Ofgem. Estimation periods were 2000-2018, 2008-2018 and 2013-2018. Model choice was based on the Bayesian Information Criterion (BIC) (Schwarz, G. 1978).

CC2.3 Results

First-choice models for the three periods are given in the table below where highlighting represents the choice of the same model form in successive periods:

Table CA2.1 - Model choice (unconstrained)

Company	***2000-2018***	***2008-2018***	***2013-2018***
UU	TBEKK	TBEKK	Cholesky
SVT	FVECH	DCC	DCC
BT	FVECH	Cholesky	Cholesky
NG	FVECH	TBEKK	TBEKK
SSE	TBEKK	TBEKK	Cholesky
PNN	FVECH	DCC	Cholesky

It is worrying that the Cholesky form is chosen five times – this form produces different results depending on the order that the time series enter the model. If we exclude it on those grounds, we have the following model choices:

Table CA2.2 - Model choice (constrained to exclude Cholesky)

Company	2000-2018	2008-2018	2013-2018
UU	TBEKK	TBEKK	DCC
SVT	FVECH	DCC	DCC
BT	FVECH	BEKK	DVECH
NG	FVECH	TBEKK	TBEKK
SSE	TBEKK	TBEKK	TBEKK
PNN	FVECH	DCC	DBEKK

We see that the TBEKK form occurs frequently – it is “top choice” in 7 out of the 15 cases. The DCC and FVECH models also occur often: 4 times each.

What, then, of choices within the three estimation periods? Which models were first, second and third? The following tables present the results, again first with Cholesky as an option and then without.

Table CA2.3 - Unconstrained model choice 2000-2018

Company	First	Second	Third
UU	TBEKK	FVECH	BEKK
SVT	FVECH	TBEKK	ADCC
BT	FVECH	Cholesky	DCC
NG	FVECH	DCC	TBEKK
SSE	TBEKK	BEKK	FVECH
PNN	FVECH	DCC	ADCC

The first choice is the 23 parameter FVECH model in a surprising four cases, with Triangular BEKK in the other two. Second and third choice models also draw on other families like DCC and Cholesky.

Table CA2.4 - Unconstrained model choice 2008-2018

Company	First	Second	Third
UU	TBEKK	BEKK	DCC
SVT	DCC	TBEKK	Cholesky
BT	Cholesky	BEKK	TBEKK
NG	TBEKK	BEKK	DCC
SSE	TBEKK	BEKK	FVECH
PNN	DCC	DVECH	TBEKK

For 2008-2018 the Triangular BEKK model is first choice in three cases, including two where it had also been first choice in the whole sample. Second choices tend to be from the BEKK family while third choices vary more widely.

Table CA2.5 - Unconstrained model choice 2013-2018

Company	First	Second	Third
UU	Cholesky	DCC	TBEKK
SVT	DCC	ADCC	Cholesky
BT	Cholesky	DVECH	DCC
NG	TBEKK	DBEKK	Cholesky
SSE	Cholesky	TBEKK	BEKK
PNN	Cholesky	DBEKK	TBEKK

In the shortest estimation period, the rather strangely-behaved Cholesky model is first choice in four cases.

We now turn to model choice where the available models are constrained to exclude the Cholesky form.

Table CA2.6 - Constrained model choice 2000-2018

Company	First	Second	Third
UU	TBEKK	FVECH	BEKK
SVT	FVECH	TBEKK	ADCC
BT	FVECH	DCC	DVECH
NG	FVECH	DCC	TBEKK
SSE	TBEKK	BEKK	FVECH
PNN	FVECH	DCC	ADCC

First choices are unchanged and the one instance of a Cholesky (BT) is replaced by a Diagonal VECH model.

Table CA2.7 - Constrained model choice 2008-2018

Company	First	Second	Third
UU	TBEKK	BEKK	DCC
SVT	DCC	TBEKK	BEKK
BT	BEKK	TBEKK	DVECH
NG	TBEKK	BEKK	DCC
SSE	TBEKK	BEKK	FVECH
PNN	DCC	DVECH	TBEKK

Table CA2.8 - Constrained model choice 2013-2018

Company	First	Second	Third
UU	DCC	TBEKK	DBEKK
SVT	DCC	ADCC	DVECH
BT	DVECH	DCC	DBEKK
NG	TBEKK	DBEKK	DCC

SSE	TBEKK	BEKK	DBEKK
PNN	DBEKK	TBEKK	BEKK

The large number of Cholesky models are replaced by a variety of forms – with several Diagonal BEKK models – for the period 2013-2018.

CC2.4 Discussion

Overall, the Triangular BEKK (Engle, R. F. and K. F. Kroner 1995) and DCC (Dynamic Conditional Correlation) (Engle, R. F. 2002) forms are chosen as statistically "best" in many cases when the estimation period is reduced from 2000-2018 to 2008-2018. The Diagonal BEKK form appears quite often in the constrained 2013-2018 results.

The evidence for stability across estimation horizons is not particularly strong but it is there: in 5 cases the same models were chosen in two periods:

- UU and SSE have the same form for 2000-2018 and 20098-2018
- SVT, BT and NG have the same form for 2008-2018 and 2013-2018.

Bibliography

ENGLE, R. F. (2002): "Dynamic Conditional Correlation," *Journal of Business & Economic Statistics*, 20, 339-350.

ENGLE, R. F., AND K. F. KRONER (1995): "Multivariate Simultaneous Generalized Arch," *Econometric Theory*, 11, 122-150.

SCHWARZ, G. (1978): "Estimating the Dimension of a Model," *Ann. Statist.*, 6, 461-464.

Annex C3 Beta estimates from GARCH models chosen for different estimation windows

CC3.1 Aim

To assess the extent to which beta estimates change when using only recent data.

CC3.2 Method

Cholesky models were excluded from consideration for the usual reasons of asymmetry. For the chosen GARCH models, the daily betas were calculated from the estimated daily variances and covariances. These daily betas were averaged while a second beta measure was calculated using the average of the individual variances and covariances.

The OLS rolling 500 day betas were averaged over both the periods 2008-2018 and 2013-2018 and also 2010-2018 and 2015-2018, the latter being appropriate if all data before January 2008/January 2013 is to be ignored.

CC3.3 Results

The results for 2008-2018 are summarised in the following table:

Table CA3.1 - Beta results 2008-2018

Stock	*GARCH: average of dailies*	*GARCH: average of variances*	*OLS average of rolling regressions 2008-2018*	*OLS average of rolling regressions 2010-2018*	*OLS single estimate 2008-2018*
BT	0.92	0.95	0.96	0.98	0.94
UU	0.58	0.59	0.60	0.58	0.60
NG	0.58	0.60	0.60	0.59	0.63
SSE	0.57	0.63	0.65	0.65	0.64
PNN	0.61	0.56	0.59	0.60	0.55
SVT	0.66	0.58	0.61	0.59	0.60

Typically, differences between the GARCH and various OLS estimates are small.

Results for 2013-2018 are summarised below:

Table CA3.2 - Beta results 2013-2018

Stock	*GARCH: average of dailies*	*GARCH: average of variances*	*OLS average of rolling regressions 2013-2018*	*OLS average of rolling regressions 2015-2018*	*OLS single estimate 2013-2018*
BT	0.93	0.97	0.99	0.95	0.97
UU	0.73	0.70	0.62	0.69	0.69
NG	0.68	0.67	0.62	0.67	0.66
SSE	0.69	0.77	0.71	0.80	0.77
PNN	0.68	0.64	0.63	0.67	0.67
SVT	0.73	0.69	0.64	0.69	0.69

In most cases, the two GARCH estimates and the 2015-2018 rolling OLS (that uses only data after Jan 2013) are quite similar.

For comparison, the 2000-2018 simple OLS beta, the average of the 500 day rolling OLS betas and the GARCH daily average beta are given in the table below:

Table CA3.3 - Beta results 2000-2018

Stock	*GARCH: average of dailies*	*GARCH: average of variances*	*OLS single estimate*	*OLS average of rolling regressions 2002-2018*
BT	0.99	1.01	1.04	1.00
UU	0.55	0.56	0.57	0.57
NG	0.61	0.58	0.61	0.59
SSE	0.54	0.57	0.57	0.60
PNN	0.49	0.45	0.45	0.49
SVT	0.53	0.52	0.53	0.55

The betas of all but BT seem to have risen in the most recent 5 years of data.

The following five tables compare the results for each approach across the three time periods.

Table CA3.4 - Single shot OLS estimates

Stock	*2000-2018*	*2008-2018*	*2013-2018*
BT	1.04	0.94	0.97
UU	0.57	0.60	0.69
NG	0.61	0.63	0.66
SSE	0.57	0.64	0.77
PNN	0.45	0.55	0.67
SVT	0.53	0.60	0.69

Table CA3.5 – Average of rolling OLS using only data after initial date of period

Stock	*2000-2018*	*2010-2018*	*2015-2018*
BT	1.00	0.98	0.95
UU	0.57	0.58	0.69
NG	0.59	0.59	0.67
SSE	0.60	0.65	0.80
PNN	0.49	0.60	0.67
SVT	0.55	0.59	0.69

Table CA3.6 – Average of rolling OLS using data from before January of initial year

Stock	*2000-2018*	*2008-2018*	*2013-2018*
BT	N/A[9]	0.96	0.99
UU	N/A	0.60	0.62
NG	N/A	0.60	0.62
SSE	N/A	0.65	0.71
PNN	N/A	0.59	0.63
SVT	N/A	0.61	0.64

Table CA3.7 – Average of daily GARCH betas

Stock	*2000-2018*	*2008-2018*	*2013-2018*
BT	0.99	0.92	0.93
UU	0.55	0.58	0.73
NG	0.61	0.58	0.68
SSE	0.54	0.57	0.69
PNN	0.49	0.61	0.68
SVT	0.53	0.66	0.73

Table CA3.8 – GARCH betas from average covariances and variances

Stock	*2000-2018*	*2008-2018*	*2013-2018*
BT	1.01	0.95	0.97
UU	0.56	0.59	0.70
NG	0.58	0.60	0.67
SSE	0.57	0.63	0.77
PNN	0.45	0.56	0.64
SVT	0.52	0.58	0.69

[9] Would involve creating a new dataset from 1998 forwards

CC3.4 Discussion

Choosing just the last five years of data or the last ten years of data for estimation will typically lead to higher beta estimates – no matter what approach is taken to estimation. In previous analysis in this project we have already noted the time-varying nature of beta and that several of the stocks have seen a rising trend in recent years.

The differences between beta estimates for the period 2000-2018 and those for 2008-2018 or 2013-2018 are broadly similar across estimation approaches.

It is obvious – and always has been – that choice of estimation period matters. Given the non-stationary nature of the time series of daily betas, a decision must be taken as to how much of the available information about that temporal variation is to be included in the chosen estimate. I would argue for the longest reasonable data series, with a forecast based on ARIMA models of the beta series (or, perhaps, Fourier decompositions) with the proviso that, should the ARIMA forecast be below the full-period average it be replaced with that average value.

Annex C4 Testing for ARCH errors at various time horizons

CC4.1 Aim

To assess whether evidence for ARCH in OLS residuals varies with estimation period.

CC4.2 Method

OLS models of utility return vs ASX return at a daily level were run in RATS v10 for the following timeframes:

- 2000-2018
- 2008-2018
- 2013-2018
- 2016-2018
- 2000-2004
- 2003-2004

The RATS @archtest function was used to conduct the Engle test (Engle, R. F. 1982). The test null (H_0) is that residuals are homoscedastic. As there is not complete consensus on the number of lags to include, we used all values between 1 and 20. As data are daily this allows ARCH effects to have a lag up to about a month, which seems more than long enough. The test was run in its LM form rather than the F-test form.

CC4.3 Results

Results are summarised in the following table – for a discussion of the patterns see the successive section of this document:

Table CA4.1 ARCH residuals across different time frames

Stock	*2000-2018*	*2008-2018*	*2013-2018*	*2016-2018*	*2000-2004*	*2003-2004*
BT	H_0 rejected at all lag lengths	H_0 rejected from lag 9 to lag 20 – not rejected from lag 1 to 8	H_0 **not** rejected at all lag lengths	H_0 **not** rejected at all lag lengths	H_0 rejected at all lag lengths	H_0 rejected from lags 2 to 20 – not rejected at lag 1
NG	H_0 rejected at all lag lengths	H_0 rejected at all lag lengths	H_0 rejected from lag 1 to lag 9 – not rejected from lag 10 to lag 20	H_0 rejected from lag 1 to lag 5 – not rejected from lag 6 to lag 20	H_0 rejected at all lag lengths	H_0 rejected at all lag lengths
UU	H_0 rejected at all lag lengths	H_0 rejected at all lag lengths	H_0 **not** rejected at all lag lengths	H_0 **not** rejected at all lag lengths	H_0 rejected at all lag lengths	H_0 rejected at all lag lengths
SSE	H_0 rejected at all lag lengths	H_0 rejected at all lag lengths	H_0 rejected from lag 2 to lag 7 – not rejected at other lag lengths	H_0 rejected at lags 3 and 4 – not rejected at other lag lengths	H_0 rejected at all lag lengths	H_0 **not** rejected at all lag lengths
PNN	H_0 rejected at all lag lengths	H_0 rejected at all lag lengths	H_0 rejected from lag 1 to lag 14 – not rejected from lags 15 to 20	H_0 rejected from lag 1 to lag 6 – not rejected from lags 7 to 20	H_0 rejected at all lag lengths	H_0 rejected at all lag lengths
SVT	H_0 rejected at all lag lengths	H_0 rejected at all lag lengths	H_0 rejected only from lag 18 to lag 20 – not rejected in all other cases	H_0 rejected from lag 1 to lag 9 – not rejected from lags 10 to 20	H_0 rejected at all lag lengths	H_0 rejected from lag 11 to lag 20 – not rejected at lags 1 to 10

CC4.4 Discussion

There are obviously variations in the detection of departures from homoscedasticity with that variation driven by time period and company:

- BT and UU show no evidence of possible ARCH in either 2013-2018 or 2016-2018
- All companies show consistent evidence of possible ARCH at all lag lengths in the period 2000-2018

- All companies except BT show consistent evidence of possible ARCH at all lag lengths in the period 2008-2018: BT shows evidence at shorter lag lengths, which are likely to be more indicative of ARCH issues when daily data is employed
- An overall trend comparing 2000-2018 to 2008-2018, 2013-2018 and 2016-2018 seems to be that ARCH appears less as the time period gets shorter: SVT is something of a counterexample with stronger ARCH evidence in 2016-2018 than in 2013-2018
- A counterargument to the trend is presented by considering the 5 years 2000-2004: all companies show evidence of possible GARCH at all lag lengths
- Considering only the years 2003 and 2004 also provides general support for the idea that length of estimation period is not linked systematically to the presence of evidence for deviations from heteroscedasticity: BT, NG, UU and PNN show strong evidence for possible ARCH in just a two-year estimation window while SVT shows some evidence and, surprisingly given its performance over other estimation windows, SSE shows no evidence of possible ARCH

We can conclude that it is not universally true that shorter time periods mean less evidence of ARCH. The strength of evidence for ARCH depends on both exactly which years of data are chosen and the company in question – there is no overarching (pun intended) rule.

Rejection of the test can also indicate a structural break rather than ARCH errors, but prior analysis of structural breaks has tended to argue that purely “statistical” breaks for which a good “business story” cannot be told should largely be ignored. We thus read the Engle test results as indicative of possible ARCH issues in the OLS residuals.

Bibliography

ENGLE, R. F. (1982): "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation," *Econometrica*, 50, 987-1007.

Ofgem

Beta Study – RIIO-2

Appendices D – H

December 2018

Indepen Limited, 4th Floor, 1 Fore Street, London, EC2Y 9DT
www.indepen.uk.com

Indepen is a management consultancy working with clients facing the challenges of regulation, deregulation, competition and restructuring. We help investors, boards and senior managers identify and assess political and regulatory risk and to develop and implement internal and external strategies to manage their exposure.

Our clients are the organisations involved in financing, constructing, managing and regulating built and natural infrastructure – water, energy, transport, land and property. We have constructive relationships with relevant government departments and agencies.

Our team combines experience of public policy, regulation, corporate finance, communication and engagement and organisational development. We complement this with input from our associates – CEOs and chairs of FTSE and privately owned companies, regulators, government ministers and academics.

The Indepen Forum provides the opportunity for investors, government and business leaders to debate, under the Chatham House Rule, issues that if mishandled could undermine well-intentioned policy initiatives.

Table of Contents

Important Notice

Appendix D LAD and OLS

D.1 Introduction

This appendix compares LAD – Least Absolute Deviation regression - and OLS results for beta at various frequencies. LAD is a form of "robust" regression in that it is less sensitive to outliers than OLS: OLS minimises the sum of squared residuals and outliers increase this sum strongly as the deviation is squared whereas LAD minimises the sum of absolute deviations. As an example, consider a data point with a deviation of 10 – in OLS this will contribute 100 (10^2) to the minimand whereas in LAD it contributes just 10. Data used are from 2000 to 2018.

D.2 Daily

Daily results are reported below:

Table D1 - Daily beta estimates (single calculation over period 2000-2018)

Stock	***OLS (s.e. in parentheses)***	**LAD (s.e. in parentheses)**
BT	1.04 (0.02)	1.00 (0.02)
NG	0.61 (0.02)	0.56 (0.02)
UU	0.57 (0.02)	0.57 (0.02)
SSE	0.57 (0.02)	0.54 (0.02)
PNN	0.45 (0.02)	0.44 (0.02)
SVT	0.53 (0.02)	0.52 (0.02)

The differences between LAD and OLS are small – if rounded to one decimal place, the estimates are the same. The largest absolute difference is seen with NG, where there is a 0.05 difference, or almost a 10% difference. However, both estimates would be 0.6 if rounded to one decimal place so the difference should not concern us unduly.

D.3Weekly

We present results only for Wednesday-to-Wednesday returns – other days are available upon request.

Table D2 - Weekly beta estimates (single period estimate 2000-2018)

Stock	**OLS (s.e. in parentheses)**	**LAD (s.e. in parentheses)**
BT	0.93 (0.05)	0.98 (0.05)
NG	0.56 (0.03)	0.56 (0.03)
UU	0.48 (0.04)	0.56 (0.04)
SSE	0.50 (0.04)	0.53 (0.04)
PNN	0.45 (0.04)	0.40 (0.04)
SVT	0.45 (0.04)	0.47 (0.04)

As with the daily data, differences between LAD and OLS are small in absolute size. In the weekly case, the one decimal place rounded estimates would differ in the case of UU – 0.5 from OLS and 0.6 from LAD. The most striking thing about the results would seem to be the very limited impact in most cases of switching to robust regression as a way to offset the impact of outliers – implying, perhaps, that outliers have little impact in these data.

D.4 Monthly

We present results only for the returns from first trading day of the month to first trading day of the month. Results for other definitions of monthly return (those assuming a 20-day trading month) are available upon request.

Table D3 - Monthly beta estimates (single period estimate 2000-2018)

Stock	**OLS (s.e. in parentheses)**	**LAD (s.e. in parentheses)**
BT	1.11 (0.11)	1.28 (0.11)
NG	0.56 (0.08)	0.60 (0.08)
UU	0.50 (0.08)	0.59 (0.08)
SSE	0.42 (0.08)	0.37 (0.08)
PNN	0.47 (0.10)	0.41 (0.10)
SVT	0.43 (0.09)	0.43 (0.09)

Absolute differences between OLS and LAD are larger at the monthly level. Three of the series would have different betas if rounded to one decimal place:

- BT: 1.1 vs 1.3
- UU: 0.5 vs 0.6
- PNN: 0.5 vs 0.4

Intuitively, returns month-to-month are almost surely larger in absolute value than those measured week-to-week or day-to-day. This may account for the greater apparent impact of outliers – difference between OLS and lad – at this frequency.

D.5 Discussion

Our general preference is for analysis at the level of daily data – a situation in which there is very little difference between OLS and LAD.

An alternative to LAD might be to calculate the leverage of each observation and then to assess whether high leverage leads to its being an *influential point* – that is, one that has a strong influence on the estimated coefficients. This is a form of outlier analysis.

Appendix E Debt betas

E.1 Background

The Competition Commission (CC) in its Heathrow and Gatwick Q5 inquiry 2007 explicitly considered the relevance of debt beta, a measure of the systematic risk of corporate debt, akin to its equity counterpart in determining the cost of capital for the regulated airports in the UK. It considered the traditional position of UK regulators until then to assume that debt beta had a value of 0 and rejected this position since it argued that in the specific circumstances of the airports concerned assuming zero debt beta resulted in the implausible relationship that increasing gearing led to increase in pre-tax weighted average cost of capital (WACC) whereas a positive debt beta would, consistent with theoretical expectations, result in falling pre-tax WACC.

A case for non-zero debt beta also rests on theoretical grounds and debt market pricing practices which suggest that promised returns to debt securities increase with increasing risk of those securities. It also rests on the view that debt securities cannot be immune to economy-wide forces and their returns are likely to co-vary with such forces.

Since the landmark decision of the CC in the Heathrow-Gatwick inquiry several UK sectoral regulators, the CC itself and its successor the Competition and Markets Authority (CMA) have explicitly considered the relevance of debt beta to their determination of cost of capital and not assumed it away with a zero value.

In a recent judgement in 2016 the Delaware Chancery Court in the US has suggested that debt beta should be taken into account in determining the cost of capital for a regulated business. This is in line with the recent evolution of UK regulatory practice.

This annex re-visits the intellectual case for debt beta and identifies the conceptual and computational issues that make estimation of debt beta in practice a challenging task for regulators. It also provides some empirical data from some regulatory determinations in the UK and evidence available from academic studies to illustrate the complexity of the issues involved. It highlights the need for and the scope of future research to help manage the computational task in practice.

E.2 How does debt beta affect the Cost of Capital?

Empirical studies are fairly consistent in reporting that as leverage increases credit risk and debt premia increase. As a higher proportion of a firm's assets are financed with debt rather than equity, more of the asset risk should be borne by debt. At the extreme, if all assets were financed by debt, debt beta would converge on the asset beta. The shift in risk sharing from equity holders to debt holders is a consequence of increasing gearing levels and cannot be ignored in arriving at realistic estimates of regulatory cost of capital.

Several empirical studies from the US provide evidence that there are systematic risk factors driving corporate bond yields. Fama and French (1993) identify two bond market factors:

1. term structure i.e. unexpected changes in interest rates and
2. general default risk i.e. change in the likelihood of default.

Fama and French find that these systematic factors capture most of the corporate bond returns. The two factors are, however, not independent of the three FF factors – market beta, size and book to market -that capture equity returns.

Driessen (2005) reports that common risk factors account for a significant part of corporate bond returns. They dominate other factors such as firm-specific risk factors and liquidity. I discuss this study in more detail below.

Elton et al (2001) argue that 'if common equity receives a risk premium for systematic risk, then corporate bonds must also earn a risk premium' (p272). Elton et al estimate the different components of observed corporate bond returns by the decomposition method (see below for further discussion of this method). They find that systematic risk accounts for a large part of bond returns and that the three FF factors may proxy for systematic risk factors. The impact of these systematic risk factors is much stronger for BBB rated bonds than for higher quality, AA or A, bonds. They conclude:

> ***This is strong evidence of the existence of a risk premium of a magnitude that has economic significance and provides an explanation as to why spreads on corporate bonds are so large (p272)***

The intellectual case for ignoring debt beta i.e. assuming zero debt beta is, therefore, unsustainable. The definitional and computational issues however remain.

E.3 Definitional Issues

Historically cost of debt has been estimated differently from cost of equity. Since the advent of the CAPM, it has been extensively used in academic research and by corporates as a tool for estimating cost of equity. Other models have also been used to this end:

- dividend growth models
- Price earnings ratio based models
- Fama-French 3-factor (FF3F) models
- Four factor (FF3F+ momentum factor) models
- Multi-factor models based on the arbitrage pricing theory (APT).

In contrast to the widespread use of model-based procedures for estimating cost of equity, the procedures for estimating cost of debt have been distinctly *ad hoc* and estimates have relied on the market debt premium over the risk free rate i.e. MDP, determined in relation to the investment rating of debt by rating agencies such as S&P, Moody's and Fitch. One of the major reasons for this approach is that debt is not as widely traded as equity. Some of the debt e.g. bank debt is not traded at all. So estimating debt beta from traded prices of debt is difficult. However, MDP may differ from the CAPM market premium for debt.

MDP may cover a range of risks:

- Systematic risk in the CAPM sense i.e. covariation with the market risk;
- Default loss risk – the potential loss of principal and any accrued interest following default;
- Credit default risk – the probability of default;
- Illiquidity risk i.e. due to lack of a liquid secondary market in debt;
- Market distortion risk e.g. due to supply and demand imbalances;

Based on the above we can write,

Market Debt Premium = Nominal debt cost – risk free rate

= market risk premium (MRP) for the debt + default loss premium + default risk premium + liquidity premium

Promised rate of interest (the coupon rate) = RFR + MDP

Expected return on debt (in the CAPM world) = RFR + β_d MRP

where MRP is the market risk premium.

There are questions about the nature of some of these risks comprised in MDP i.e. whether they are systematic or non-systematic and, therefore, diversifiable in the CAPM sense. MDP compensates for both systematic and non-systematic risk factors. Welch (2017, Section 11.6) suggests that the default loss may be reflected in the borrower's cash flows in which case it will not be reflected in the Expected return but in this note we treat it as included in MDP.

Many academic studies have shown that corporate bond returns are partly determined by credit risk factors but a substantial part of these returns are related to non-credit risk factors (e.g. Fama and French, 1993; Schaefer and Strebulaev, 2007). These extra credit risk factors include the FF3F! This may suggest that bond returns are determined by systematic factors. Nevertheless, Fama and French (1993) show that these three factors may be substituting for term structure and liquidity factors. These factors, while possibly systematic, are different from the CAPM beta i.e. market sensitivity factor. To the extent that default risk reflects economy-wide factors e.g. in a recession defaults are likely to be widespread it may be included in the systematic risk factor β_d.

E.4 Estimation issues

The definitional issues cannot be disentangled from the computational issues. There are broadly two procedures:

- **Indirect method by decomposing** the observed MDP into premium for its various components; this involves subtracting estimates of premia for risks other than beta risk and using this residual risk premium to estimate the debt beta (the decomposition method);
- **Direct method** by regressing observed excess bond returns over the risk free rate on MRP; this parallels the way CAPM beta is estimated for equity (the CAPM method);

The decomposition method

The decomposition method requires a number of estimates to be made before arriving at the residual risk premium and the derivation of debt beta from that residual premium. The quality of the derived beta therefore depends on the quality of estimates of these risk premia. Major challenges are:

- The identification of the components of the MDP and
- Deriving reasonable estimates of the risk premia for these components.

Debt beta based on default probabilities and loss given default rates

In many UK inquiries into the cost of debt for regulated businesses, firms and their advisers like OXERA, Europe Economics (EE) etc have used this method but have differed in including or excluding some of the MDP components. For example, in its report 'PR19 – Initial Assessment of the Cost of Capital' (December 2017), EE has decomposed MDP into just two components: systematic risk and default loss risk (p53) which takes into account both the probability of default by the borrower firm and the loss given default (LGD). Other proponents have argued for inclusion of a liquidity premium since debt securities are not frequently traded especially the low-rates ones.

If only the default loss risk is taken into account, then $ß_d$ is estimated from

$ß_d$ = (MDP/ MRP) – [P_D x (loss of Interest + % of principal lost)/MRP]

where P_D is default probability.

In its PR19 report, EE assumes RFR = 200bps, MDP = 162bps, MRP = 675bps, Coupon = R_f + MDP =362 bps. P_D = 0.2% and LGD = 20%

$ß_d$ = (1.62/ 6.75) – [0.002 x 20)/ 6.75]

= 0.23

With P_D of 0.1%, $ß_d$ = 0.24. As default risk falls, the systematic component rises.

It is clear that the systematic component and hence $ß_d$ become smaller as many more components of MDP are stripped out.

Debt beta based on default probabilities and loss given default rates and liquidity premium

In its estimation of debt beta for the Heathrow and Gatwick airports for Q5, the CC included both default loss risk and liquidity risk as shown in the following table (see Table 5 of Appendix F to the report). The CC then settled for a debt beta of 0.10.

Table E1: Debt beta calculation

Beta estimation (data in basis points)			
		Low	High
MDP	R1	110	110
Liquidity premium	R2	41	30
Default premium	R3	38	14
Implied systematic risk premium	R4 = R1-R2-R3	31	66
MRP	R5	350	350
Debt beta	(R4/R5)/100	0.09	0.19

Driessen and De Jong (2005) (cited in Ivo Welch, *Corporate* Finance, 4th edition 2017, pp266-267) break the promised (quoted) yields on debt securities rated AAA down to CCC into the Default premium, liquidity premium, tax premium and the risk premium proxy for systematic risk premium. For their sample US securities during 1985 to 2003, they estimated the default loss premium at about 40bps (250bps) for AAA/ A (B/CCC) rated debt. The liquidity premium was 50bps (100-150bps) for highly rated (junk) bonds.

In a recent paper Bongaerts, de Jong and Driessen (2017) find 'a strong effect of the liquidity level and equity market liquidity risk on expected corporate bond returns, while there is little evidence that corporate bond liquidity risk exposures explain expected' corporate bond returns, even during the recent financial crisis. Thus liquidity premium seems to be an important component of the MDP and needs to be taken into account in using the decomposition method to estimate debt beta.

Based on credit default swap (CDS) premia

A CDS contract is a derivative instrument to provide protection against default. The seller of the protection agrees to pay the nominal value of a bond in the event of default by corporate bond issuer for periodic (generally quarterly) payments of agreed premia. The buyer of the protection i.e. the bond investor, in return for the premia payments, is assured of avoiding a loss on the investment. The size of the CDS premium is a function of the credit risk faced by the bond investor. Therefore, we would expect a high correlation between CDS premium and bond yield spread. According to OXERA, credit risk includes both default premium and default risk premium.

How well do CDS premia track credit spreads?

Blanco, Brennan and Marsh (2005) empirically establish for 33 US and European bond issuers with CDS data from January 2001 to June 2002 that CDS rates lead the yield spreads in price discovery i.e. they are lead indicators and more efficiently price credit risk. They compare 5 year CDS rates and 5 year corporate yield spreads.

Longstaff, Mithal and Neis (2005) use the CDS rates as a means of splitting the default and non-default components e.g. liquidity of corporate yield spreads. For US data, they report that for A rated bonds, CDS rates account for 53% to 60% of the yield spread whereas for BBB rated bonds they account for 68% to 74% of the spread. Thus a very high proportion of yield spreads in A or BBB bonds is attributable to default premium. These authors find that non-default component of the yield spread is significant but it remains relatively constant across rating categories. For A rated bonds the default premium may be just about 1.2 time the non-default component but for BBB this rises to about 2.5 times (these ratios are based on my visual interpretations of Figure 5 in their paper since they do not report exact numbers).

From these analyses we can infer that CDS rates

- reflect credit risk well and
- are made up mostly of default premium and to lesser extent non-default premium.

Impact of varying assumptions about MRP

Since the MRP is the denominator in the decomposition of the MDP, if MRP varies it can have a dramatic effect on the estimated systematic risk premium for debt and the estimation of debt beta. For example, the CC in the Heathrow-Gatwick inquiry assumed a MRP of 350 bps whereas EE has assumed 650 bps. Increasing the MRP from 350 to 650 bps would nearly halve the debt beta in the table in section 4.3.

E.5 The direct method

The CAPM method also suffers from the fact that debt trading is infrequent. This reduces the power of the statistical techniques used to estimate debt beta. There are very few academic studies that report debt betas in the UK in contrast to numerous studies that have estimated equity betas.

This method can raise severe estimation problems. In its Heathrow-Gatwick inquiry cited above, the CC said:

> *The direct method caused us the greater difficulty due to estimation problems including the relatively poor quality of the data that we have on returns to debt holders, the poor statistical properties of the regressions and the difference between historical and assumed Q5 gearing levels (discussed further from paragraph 92 onwards below).*
>
> *These problems, and others such as thin trading, affect debt beta estimates more seriously than equity beta estimates even for large firms. These factors have led us to favour the indirect, decomposition method, where we can be much more confident that we are correctly observing how much compensation lenders are asking for in exchange for bearing systematic risk*

Jonathan Berk and Peter DeMarzo state that 'data for the historical returns of debt securities are much more difficult to obtain, making a direct calculation of the beta for debt problematic' and recommend the decomposition method (*Corporate Finance*, Pearson International Edition, 2007, p443).

In addition to the data problems, the modelling issues discussed in this report in the context of estimation of equity beta are likely to emerge in a more intractable form. In his estimation of debt betas for debt securities with a range of ratings from AAA to D, using 5 years of monthly data ending in the four quarters in 2015, Clifford Ang (2017) reports that for the AAA securities the debt beta is negative and falls from -0.41 in March 2015 to -0.28 in December 2015 estimate whereas for the BBB debt it increases from 0.01 to 0.08. The negative beta is counter-intuitive and the beta estimates are also unstable. However, for the BBB and lower rated debt beta is positive and increases exponentially as rating declines. Overall, the direct method needs to be rigorously tested to overcome the econometric and data issues. The approaches developed in this report in the context of equity beta estimation are likely to be of value in addressing many of these issues.

E.6 Conclusions

The relevance of debt beta has increasingly been recognised in the UK regulatory practices, US judicial pronouncements on regulatory cost of capital and US regulatory practices. The conceptual rationale for taking it into account in determining regulatory cost of capital is also now accepted widely. But the challenges of estimating debt beta in a robust way remain. Both the direct method of estimating it through the CAPM-type modelling and the indirect method of estimating it by decomposing the observed or promised interest rates on debt securities raise conceptual, data-related and methodological issues which need to be addressed in order to make robust estimates of debt beta for regulatory determination of cost of debt and hence cost of capital for regulated firms. Further research addressing these issues identified in this annex will contribute to reliable and robust estimation of debt beta and enhance the precision of regulatory cost of capital determination.

Appendix F Gearing evidence

As discussed in section 4, the capital structure of the business is important for understanding the difference between the observed equity β and the underlying business risk as measured by the asset β. Section 4 also sets out three different measures of gearing, all of which can be relevant for different questions and consequently need to be understood and used in a consistent manner. In this appendix we set out some of the different measures that we have calculated or which have been calculated recently in studies for other regulators.

EE (2017) sets out the three measures of gearing and provides estimates of each for the three listed water companies (reported in section 4 of our report). Figure F1 below reports the RCV gearing estimate calculated and reported in EEs' report.

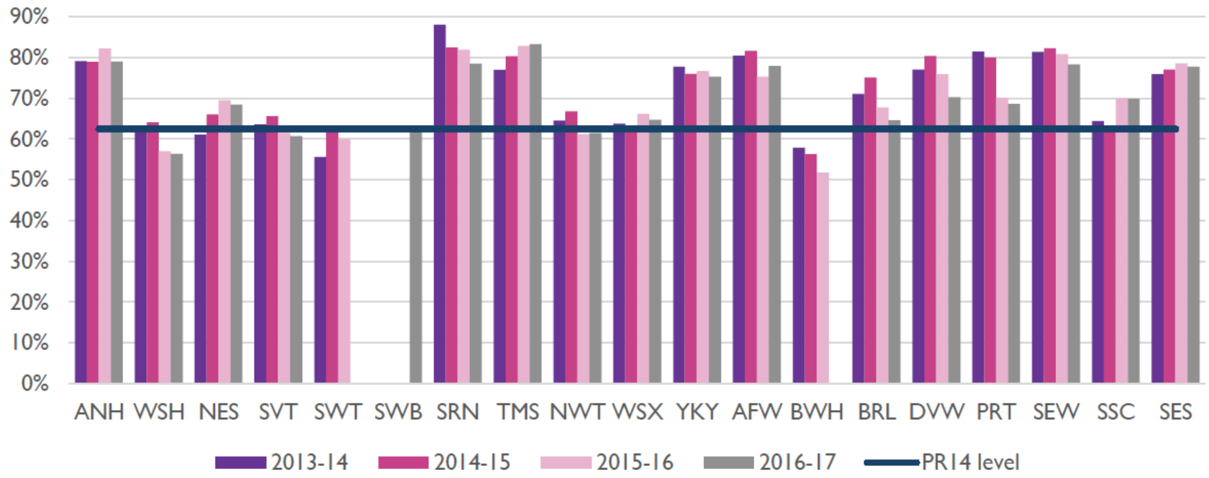


Figure F1 Net debt to RCV, 2014-2017

Source: EE (2017) Figure 6.1

As can be seen from the figure, the majority of the companies were relatively close to, or above, the notional efficient level of gearing set at PR14. Four of the larger Water and Sewerage companies, Anglian, Southern, Thames and Yorkshire have RCV gearing closer to 75-80%.

This can be compared with the Enterprise Value of Gearing which uses the market value of equity in its calculation. Table F1 provides annual data on this measure from 2001 to 2018.

As can be seen, these measures of gearing tend to be lower as the market value of equity tends to be greater than the book value of equity – even when that is inflation adjusted as is the case for RCV gearing. The relationship between the market and book value is captured in the market to asset ratio (MAR).

Table F1 Enterprise Value Gearing, 2001-2018

	BT	NG	UU	SSE	PNN	SVT
2001	45.4%	32.0%	45.1%	20.1%	52.5%	50.3%
2002	36.1%	47.9%	47.2%	17.1%	50.7%	48.8%
2003	41.3%	53.8%	50.7%	18.2%	58.7%	50.4%
2004	35.5%	48.8%	46.8%	19.4%	55.6%	51.1%
2005	28.5%	47.3%	54.6%	16.1%	47.2%	47.8%
2006	28.2%	42.0%	40.8%	18.2%	47.3%	43.2%
2007	23.0%	35.8%	37.4%	14.3%	42.8%	48.6%
2008	35.9%	51.2%	32.0%	23.1%	43.8%	51.1%
2009	67.3%	63.3%	62.0%	33.3%	57.2%	61.9%
2010	53.2%	57.8%	57.6%	36.3%	50.7%	58.3%
2011	39.7%	47.8%	55.6%	30.3%	46.4%	53.8%
2012	35.4%	46.5%	57.4%	32.5%	45.1%	52.8%
2013	28.1%	43.1%	55.3%	27.9%	46.9%	51.9%
2014	19.2%	41.0%	52.4%	29.0%	44.4%	50.8%
2015	13.7%	40.8%	50.2%	23.6%	40.0%	49.1%
2016	19.8%	39.4%	51.8%	31.2%	42.6%	48.3%
2017	25.2%	30.6%	51.3%	30.7%	42.2%	47.8%
2018	32.2%	44.7%	60.3%	39.1%	50.9%	55.5%

Source: Ofgem, Bloomberg and Indepen analysis

If we focus on the regulated utilities that form the core of the analysis in this report, more recent evidence on both the RCV and Enterprise Value measures of gearing is provided in Table F2.

Table F2 Comparison of gearing measures, October 2018

	NG	UU	SSE	PNN	SVT
RCV Gearing	65.3%	62.4%	100.9%	88.2%	60.6%
Enterprise Value Gearing	44.9%	58.5%	41.7%	50.1%	54.6%

Source: Ofgem, Bloomberg and Indepen analysis

This illustrates that, at least for two of the water companies, the MAR is close to 1 as the two estimates of gearing are close. That is not the case for Pennon or the energy companies. As such, care needs to be taken when considering evidence for those companies and the relationship between the asset and equity βs.

Although our use of international comparators is limited, it is useful to provide some estimate of the gearing value. We only report Enterprise Value Gearing in table F3 below. Other measures could be generated but would require more time and resources than available under this project. Some of this information is used for the decomposition of National Grid's β, described in Appendix H.

Table F3 Enterprise Value Gearing estimates for international comparators, 2001 to 2018

	EDP	REE	TRN	ACE	NTGY	SRG	ENG	ES	ED	UTL
2001	33.9%	23.4%	NA	16.1%	30.8%	NA	NA	57.8%	43.4%	49.7%
2002	42.5%	25.8%	NA	42.0%	30.1%	39.2%	NA	63.3%	40.8%	49.0%
2003	61.9%	59.9%	NA	60.7%	17.3%	34.3%	41.8%	69.0%	45.6%	54.2%
2004	50.7%	54.2%	NA	46.5%	17.6%	29.1%	38.2%	60.7%	42.4%	45.0%
2005	53.2%	46.6%	31.0%	37.6%	20.9%	24.1%	34.6%	63.4%	40.6%	46.1%
2006	44.5%	44.1%	34.1%	33.2%	24.1%	39.5%	26.8%	56.8%	42.7%	46.1%
2007	37.3%	36.7%	29.8%	32.2%	15.6%	38.8%	25.1%	41.6%	39.9%	53.4%
2008	49.7%	31.2%	32.3%	42.2%	16.1%	43.3%	31.7%	54.1%	46.0%	51.6%
2009	58.9%	48.3%	46.3%	53.3%	64.4%	44.0%	51.1%	57.1%	48.1%	66.0%
2010	57.8%	41.6%	37.7%	59.1%	60.9%	43.1%	43.3%	50.8%	46.6%	55.1%
2011	61.6%	49.8%	40.6%	54.6%	57.6%	43.4%	46.7%	44.8%	41.8%	58.3%
2012	68.2%	52.6%	48.5%	72.4%	56.7%	47.3%	50.1%	46.2%	37.6%	54.3%
2013	68.2%	50.9%	49.9%	75.3%	50.9%	50.3%	46.1%	40.0%	39.9%	47.8%
2014	58.9%	40.7%	47.7%	50.7%	40.5%	47.7%	43.8%	39.4%	44.0%	44.5%
2015	57.6%	39.7%	50.3%	46.0%	44.9%	45.8%	39.4%	37.9%	41.3%	41.5%
2016	60.9%	34.2%	44.7%	43.5%	46.8%	41.3%	38.5%	35.6%	38.5%	42.4%
2017	58.9%	37.5%	45.4%	45.6%	42.6%	43.6%	45.3%	37.1%	39.7%	39.8%
2018	56.2%	38.1%	44.4%	46.2%	39.5%	47.6%	47.6%	42.6%	40.8%	39.7%

Source: Ofgem, Bloomberg and Indepen analysis

Given the impact that different measures of gearing can have, to aid consistency we would recommend making it clear on what basis an asset β has been de-geared so that a consistent approach is more likely to be applied – or at least any inconsistency is obvious.

Appendix G International evidence

G.1 USA

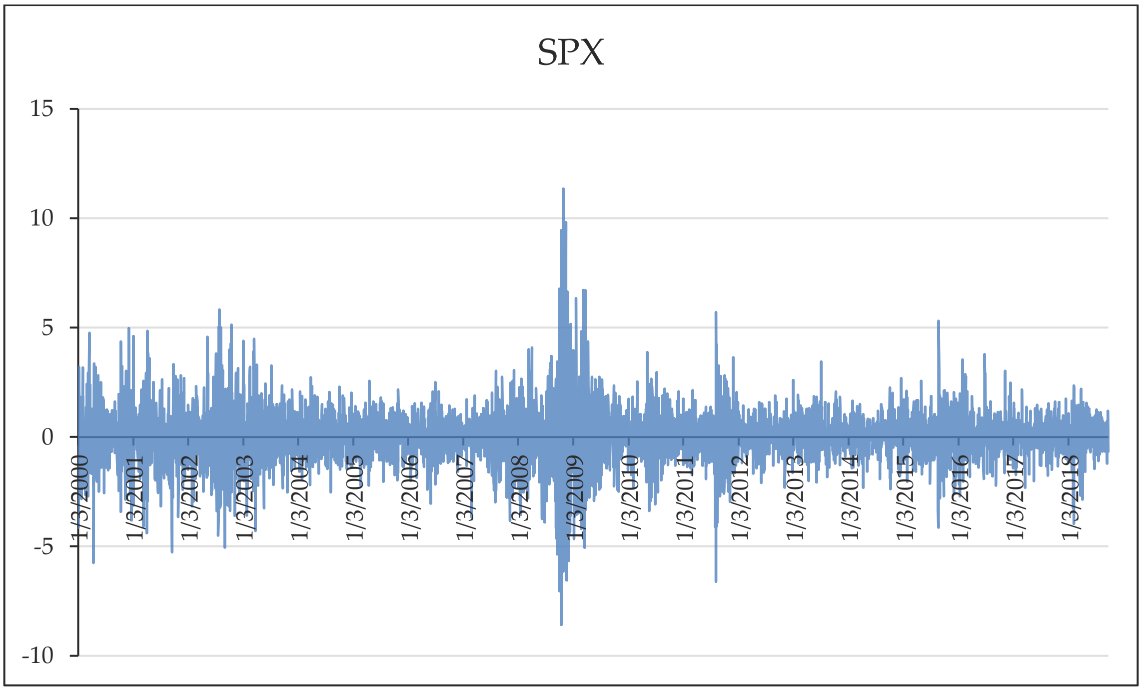


Figure G1 - SPX daily returns 2000-2018

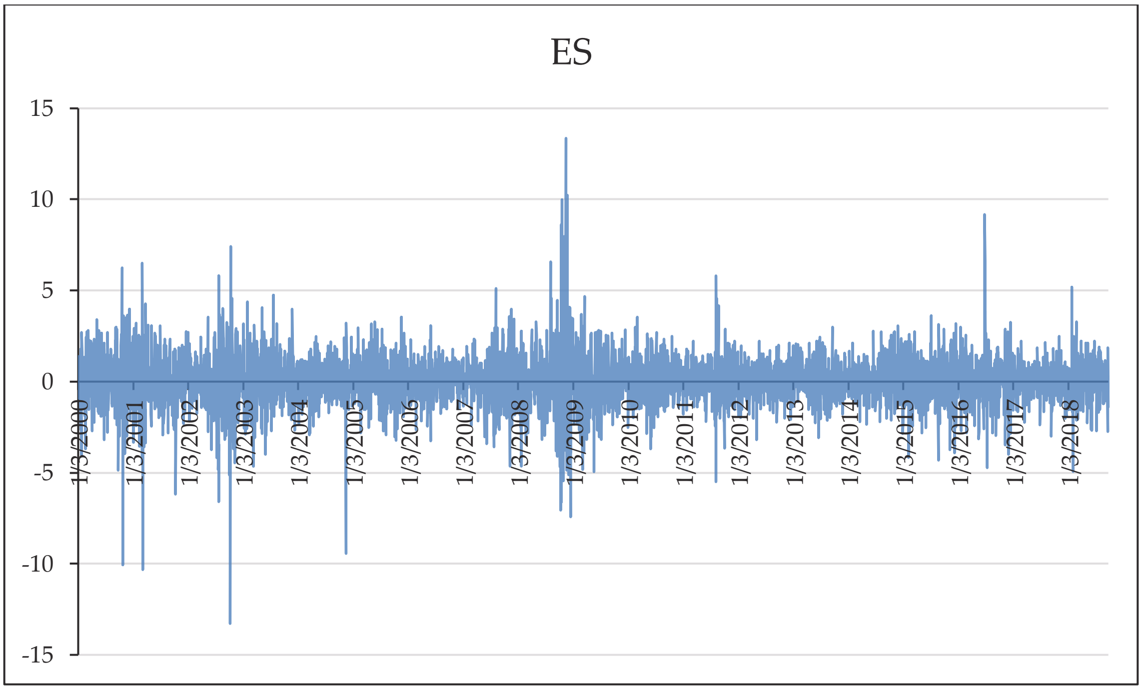


Figure G2 - ES daily returns 2000-2018

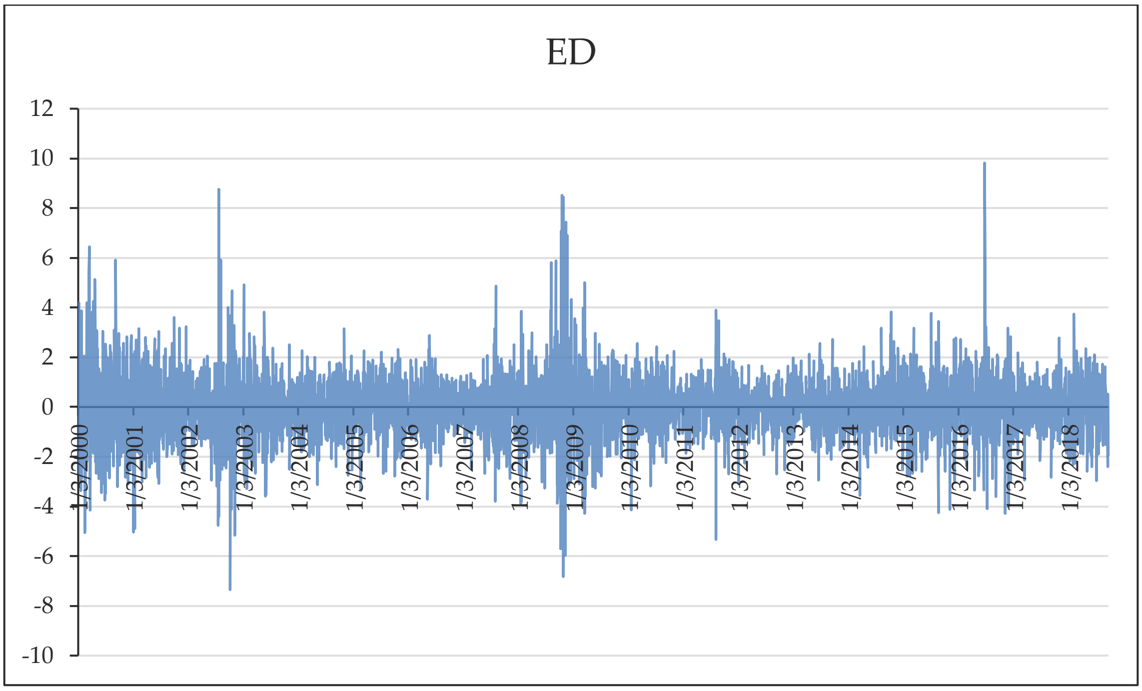


Figure G3 - ED daily returns 2000-2018

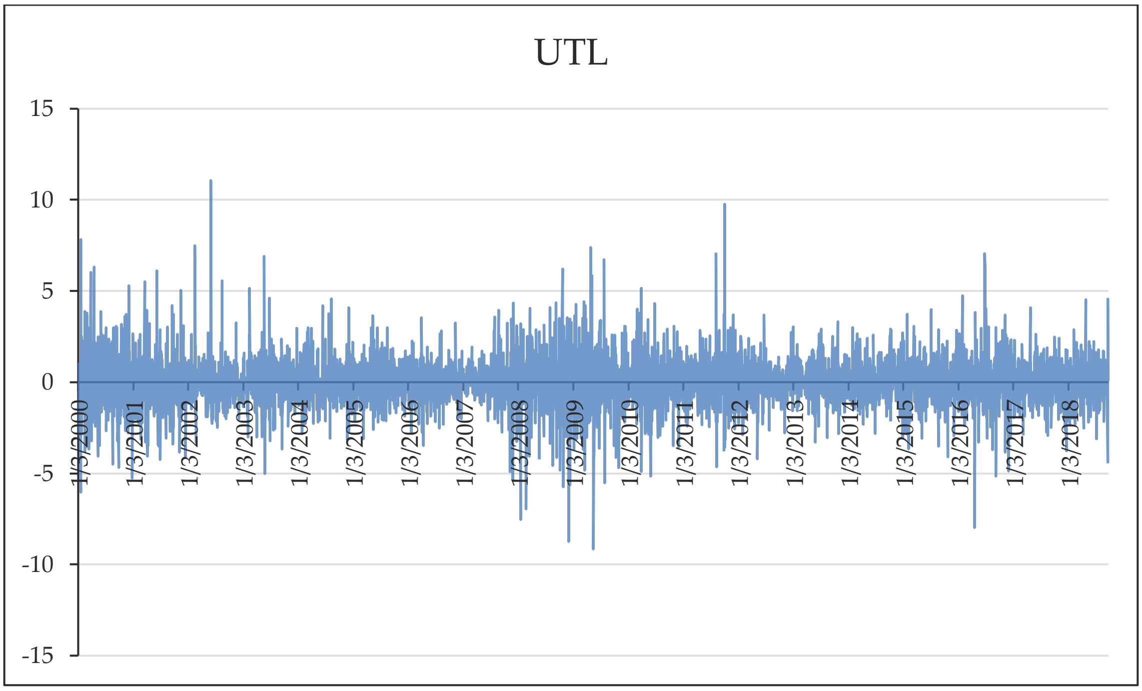


Figure G4 - UTL daily returns 2000-2018

Table G1 - OLS and LAD betas for entire period 2000-2018

Stock	*OLS (s.e. in parentheses)*	*OLS (HAC – s.e. in parentheses)*	*LAD (s.e.in parentheses)*
ES	0.598 (0.014)	0.598 (0.022)	0.607 (0.014)
ED	0.518 (0.012)	0.518 (0.019)	0.545 (0.012)
UTL	0.353 (0.015)	0.353 (0.023)	0.387 (0.015)

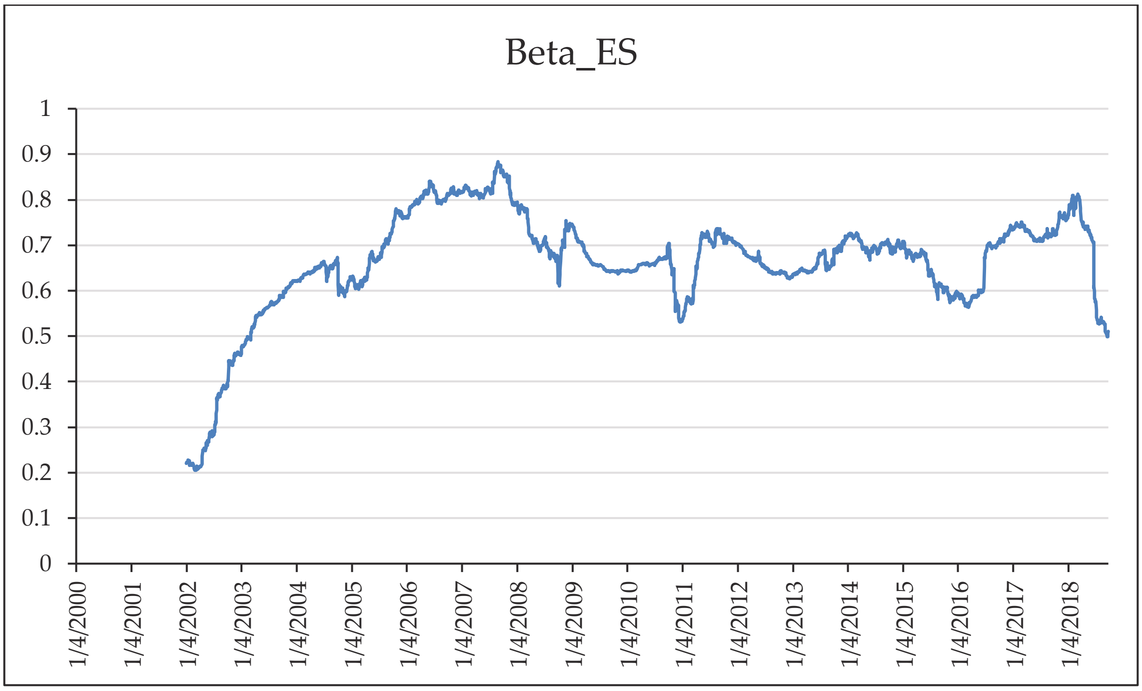


Figure G5 - ES rolling 500 day OLS beta

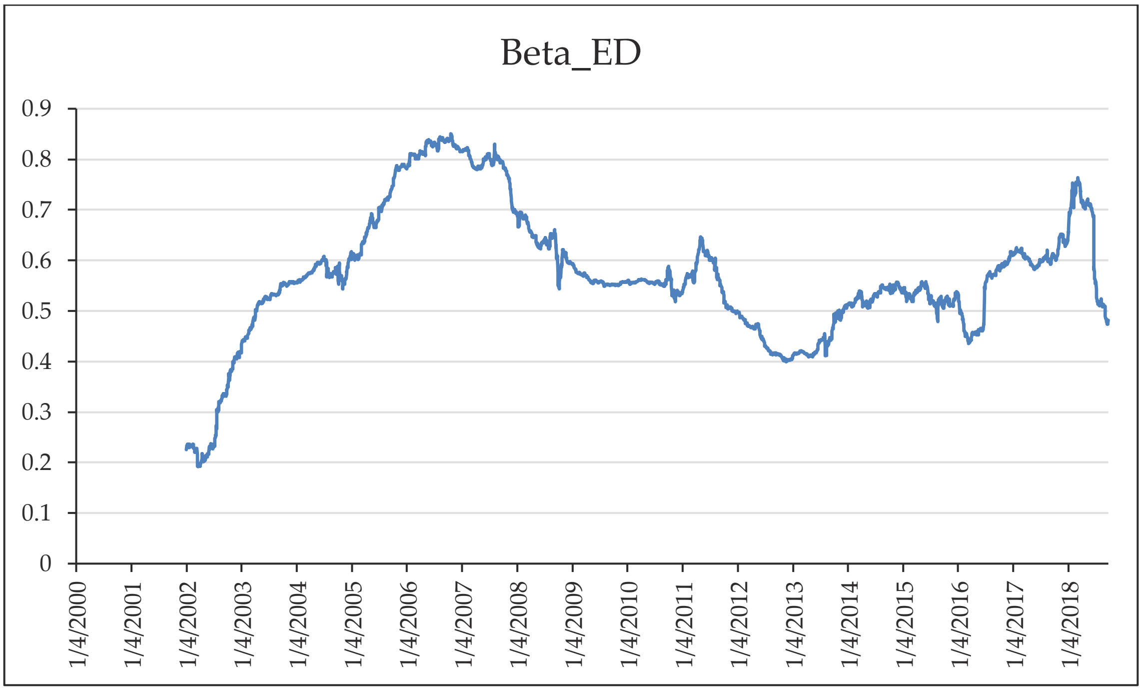


Figure G6 - ED rolling 500 day OLS beta

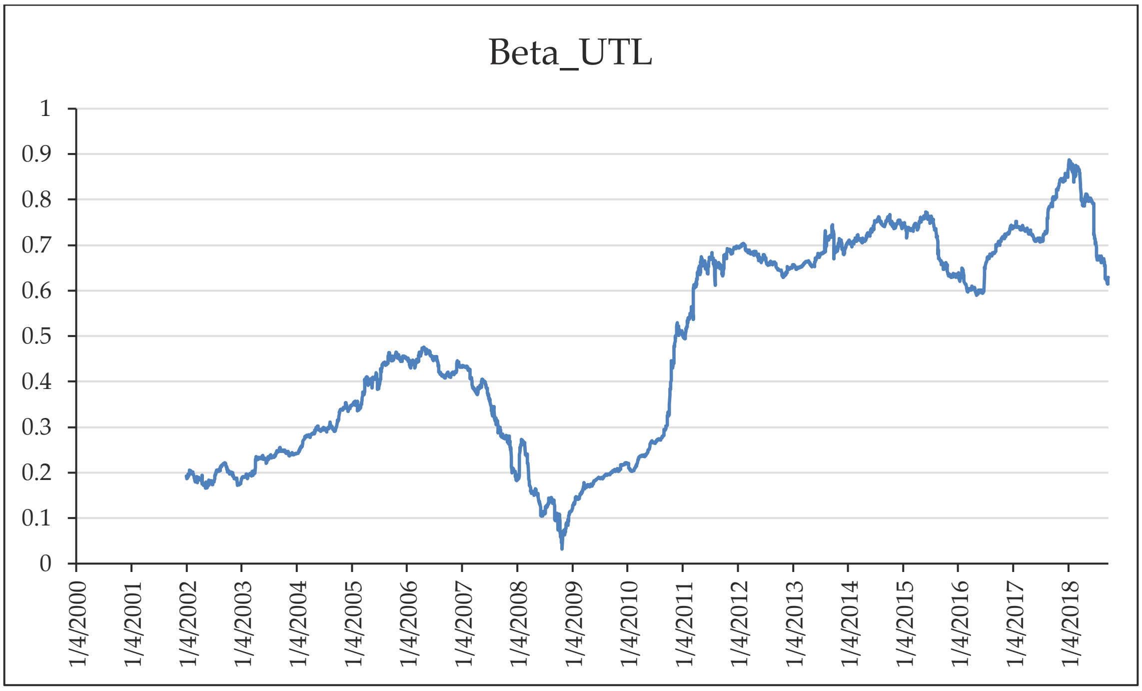


Figure G7 - UTL rolling 500 day OLS beta

GARCH model choices:

- ES – Full VECH(1,1)
- ED – Full VECH(1,1)
- UTL – DCC(1,1) – Full VECH(1,1) could not be made to converge but would probably be better

<u>ES GARCH model residuals - tests</u>

MV ARCH test lags 5 p=0.61

MV ARCH test lags 10 p = 0.93

Univariate:

SPX residual

Lags 5 p=0.01

Lags 10 p=0.06

ES residual

Lags 5 p=0.77

Lags 10 p=0.90

Overall, ES residuals look OK

ED GARCH model residuals tests

MV ARCH test lags 5	p=0.02
MV ARCH test lags 10	p=0.32
Univariate:	
SPX residual	
Lags 5	p=0.01
Lags 10	p=0.07
ED residual	
Lags 5	p=0.42
Lags 10	p=0.77

The multiple variants of the GARCH model has ramifications for model specification and selection – but getting a good GARCH model is not a turnkey operation – it takes time to optimise the model. As we have seen that model choice typically makes little difference to beta estimates we will go with this model for now.

UTL GARCH model residuals tests

MV ARCH test lags 5	p=0.00
MV ARCH test lags 10	p=0.00
Univariate:	
SPX residual	
Lags 5	p=0.00
Lags 10	p=0.03
UTL residual	
Lags 5	p=0.00
Lags 10	p=0.00

This is not a good performance – looking at the rolling beta above we see an odd pattern over time.

Table G2 - OLS and GARCH betas

Stock	*OLS*	*Average of dailies*	*Beta from averages*
ES	0.598 (0.014)	0.611	0.592
ED	0.518 (0.012)	0.556	0.521
UTL	0.353 (0.015)	0.496	0.351

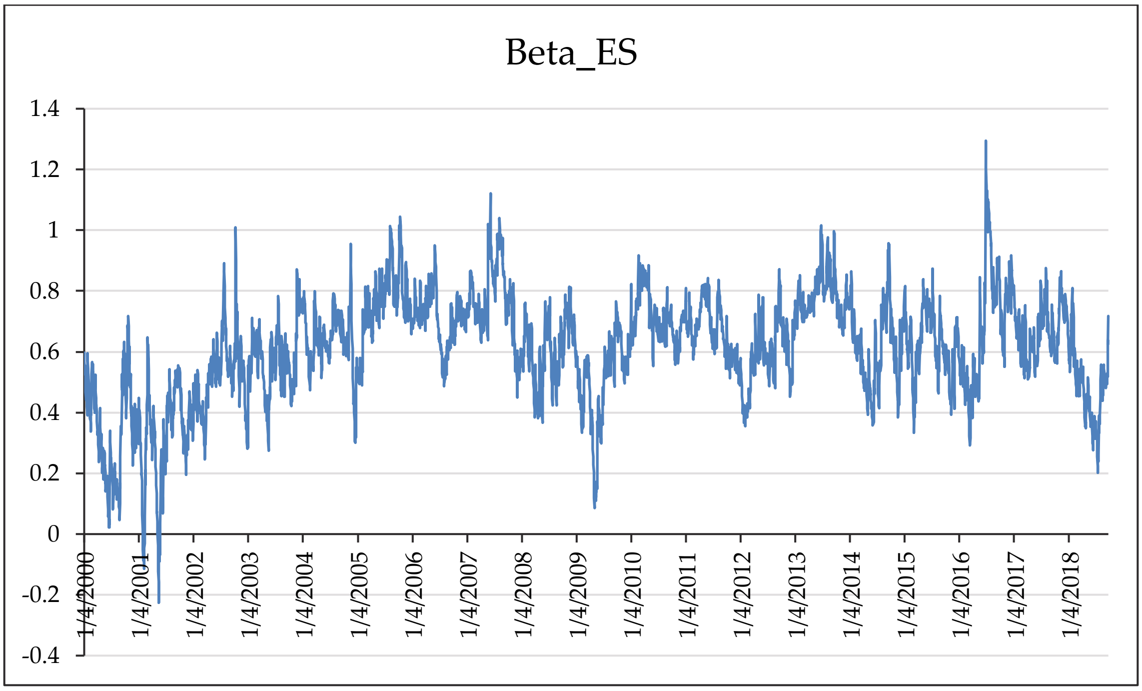


Figure G8 - ES daily GARCH betas

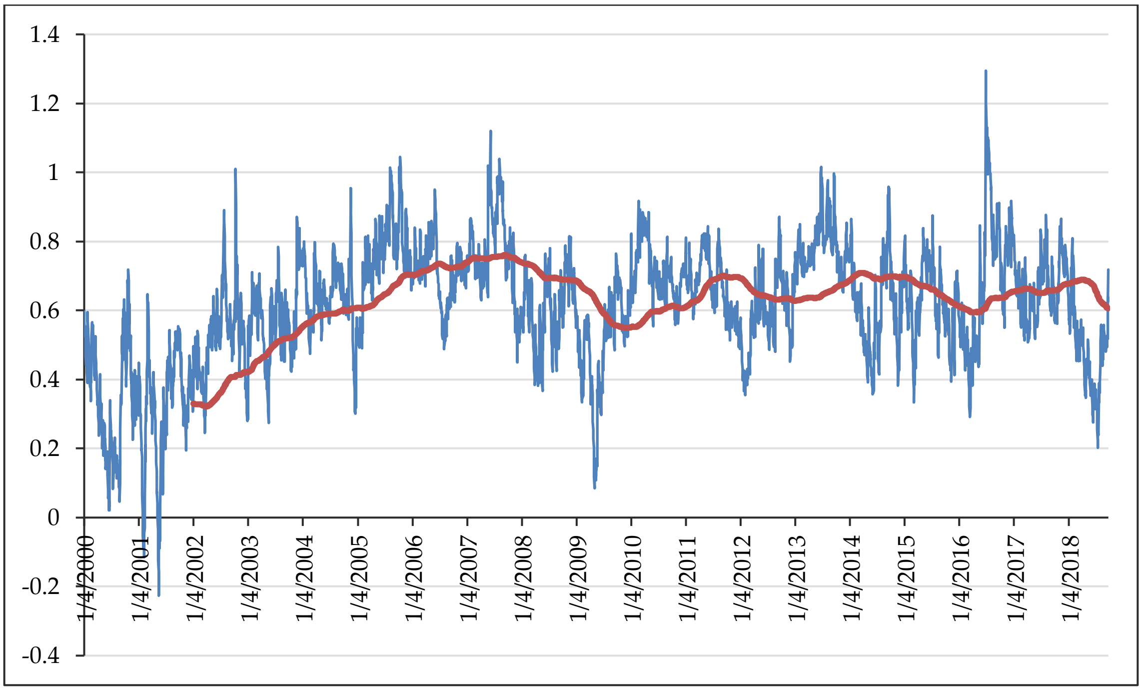


Figure G9 - ES daily GARCH betas and 500 day MA

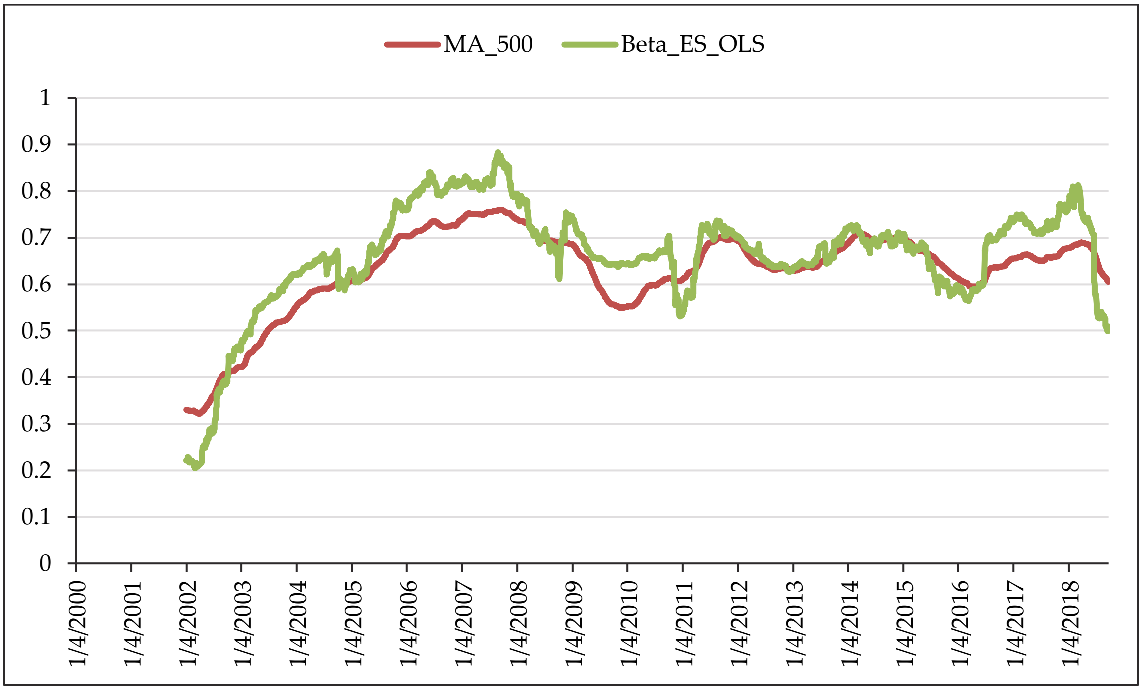


Figure G10 - ES daily GARCH beta 500 day MA and rolling 500 day OLS beta (green)

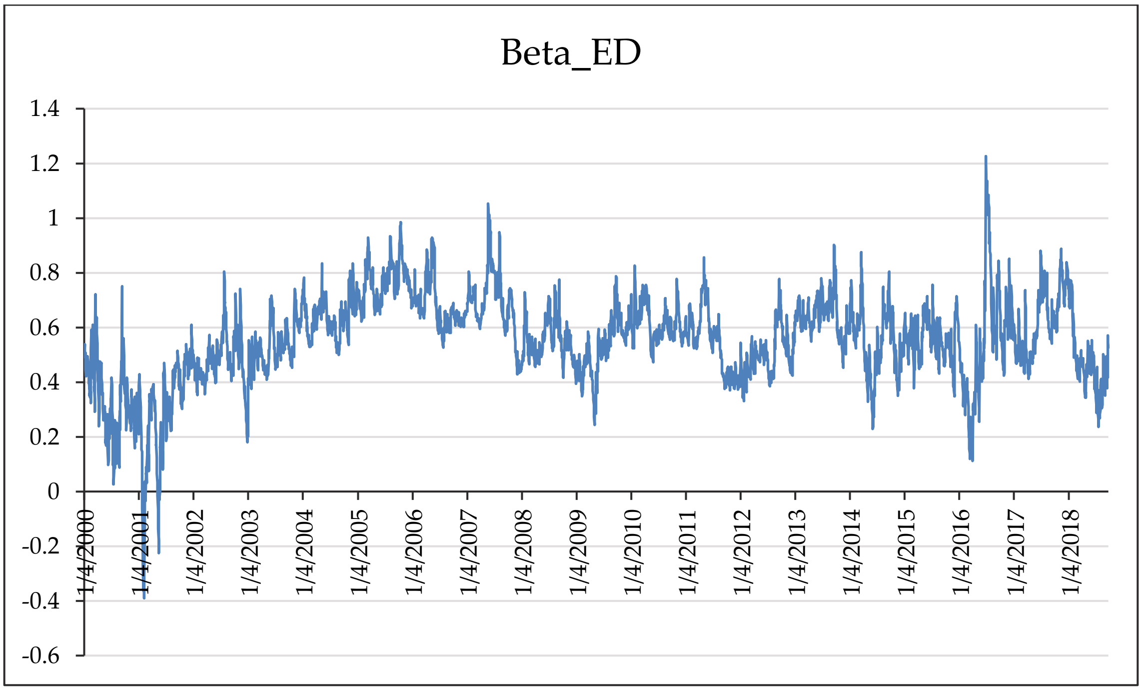


Figure G11 - ED daily GARCH betas

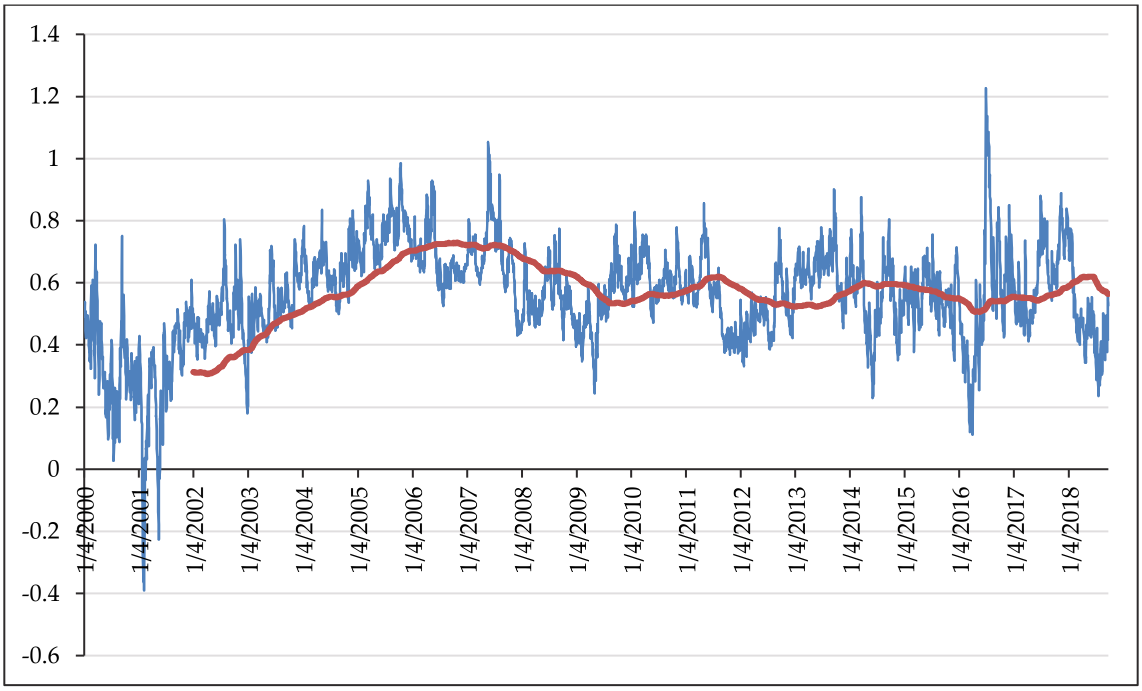


Figure G12 - ED daily GARCH betas and 500 day MA

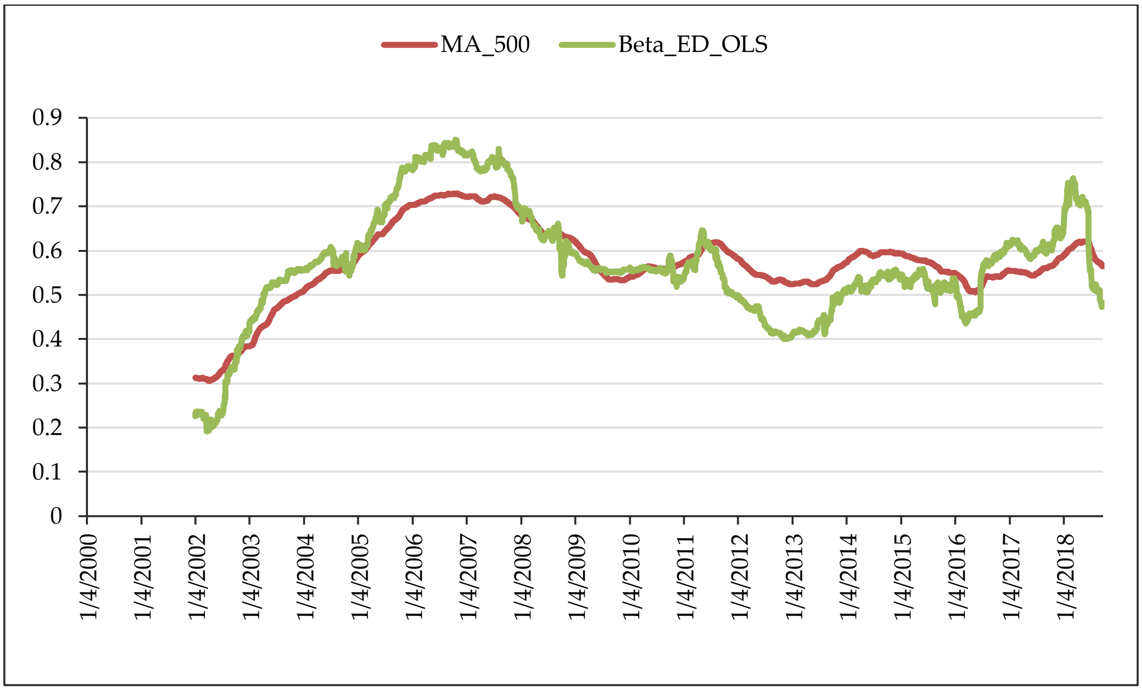


Figure G13 – ED daily GARCH beta 500 day MA and rolling 500 day OLS beta (green)

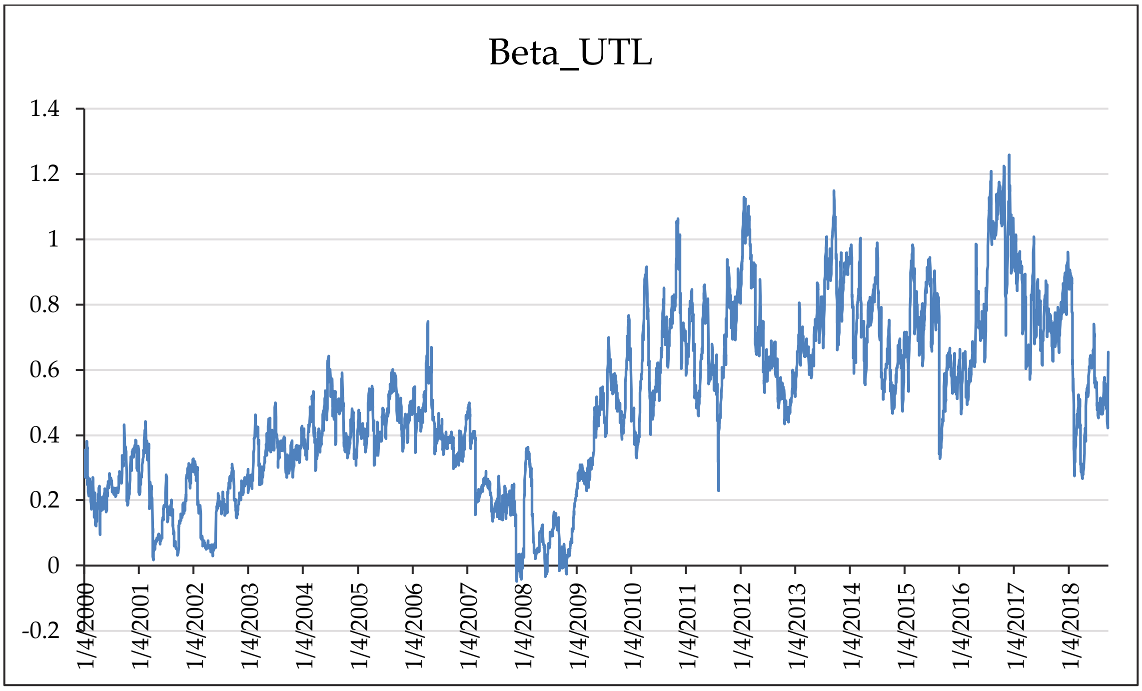


Figure G14 - UTL daily GARCH betas

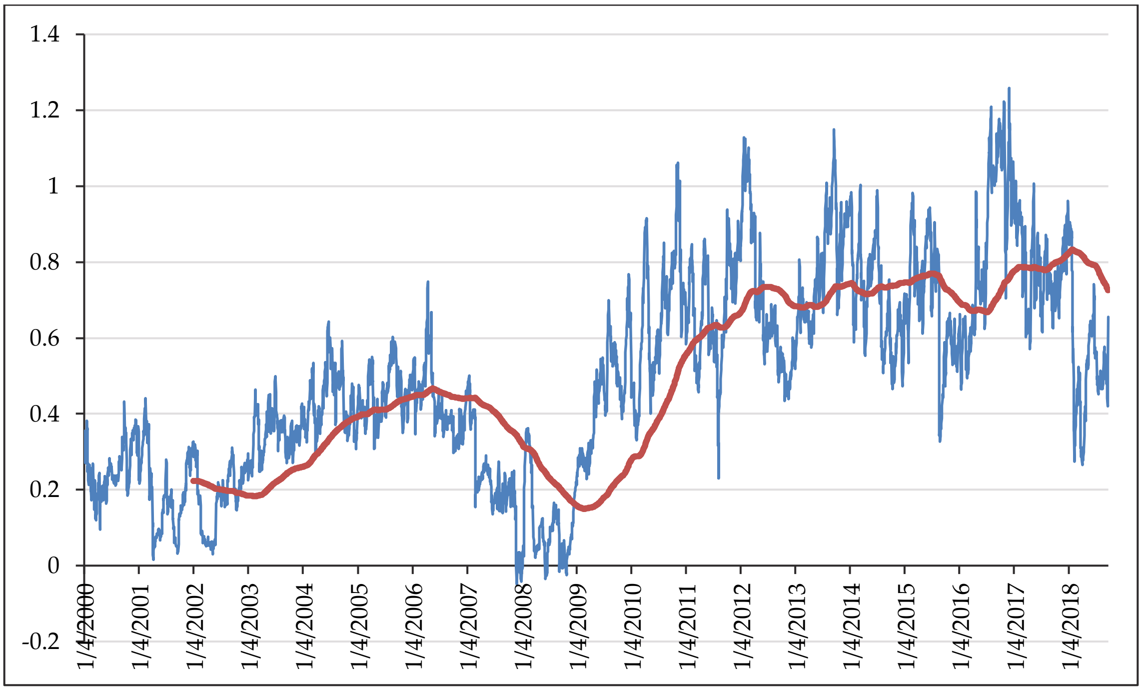


Figure G15 - UTL daily GARCH betas and 500 day MA

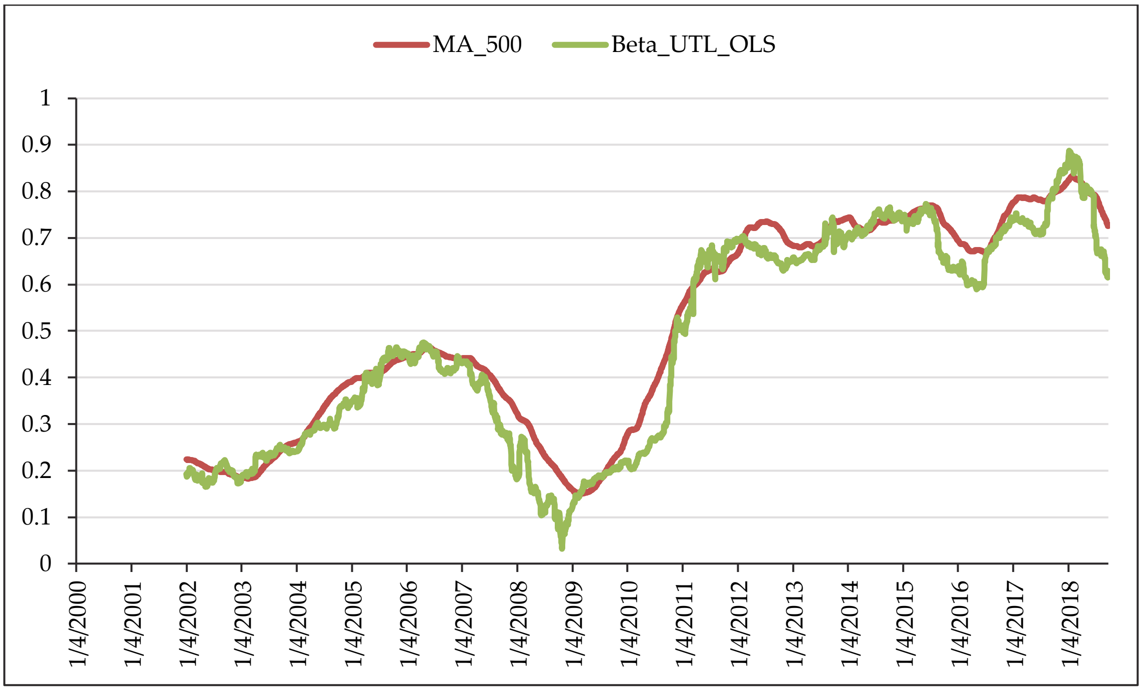


Figure G16– UTL daily GARCH beta 500 day MA and rolling 500 day OLS beta (green)

A.1 Italy

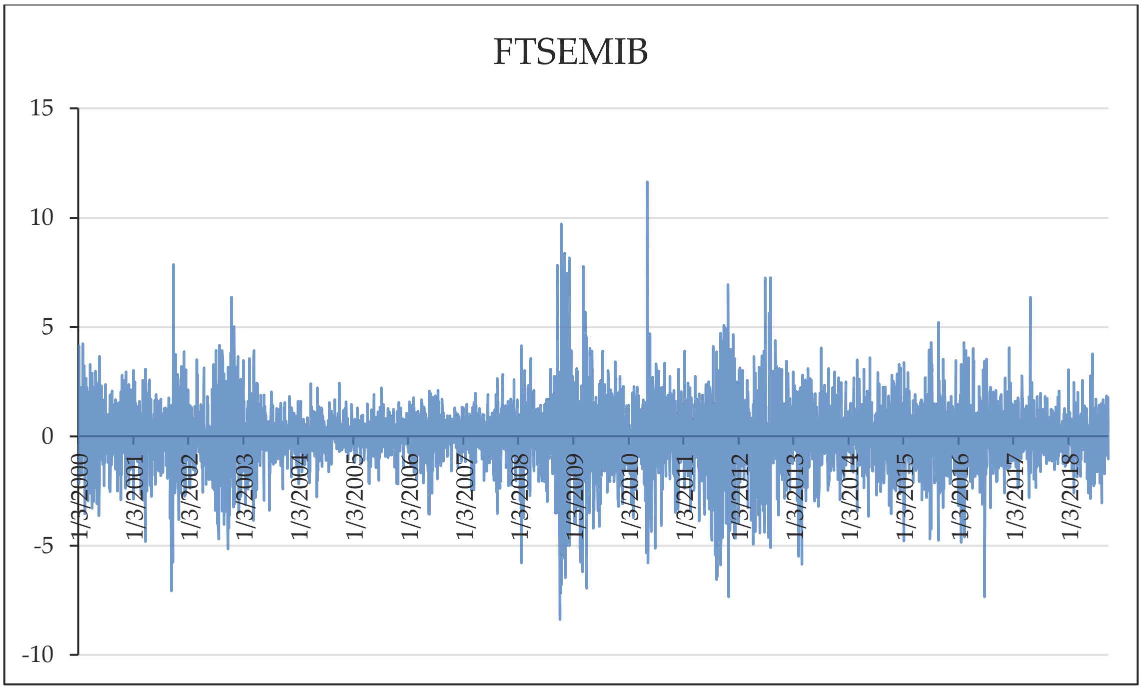


Figure G17 - FTSE-MIB daily returns 2000-2018

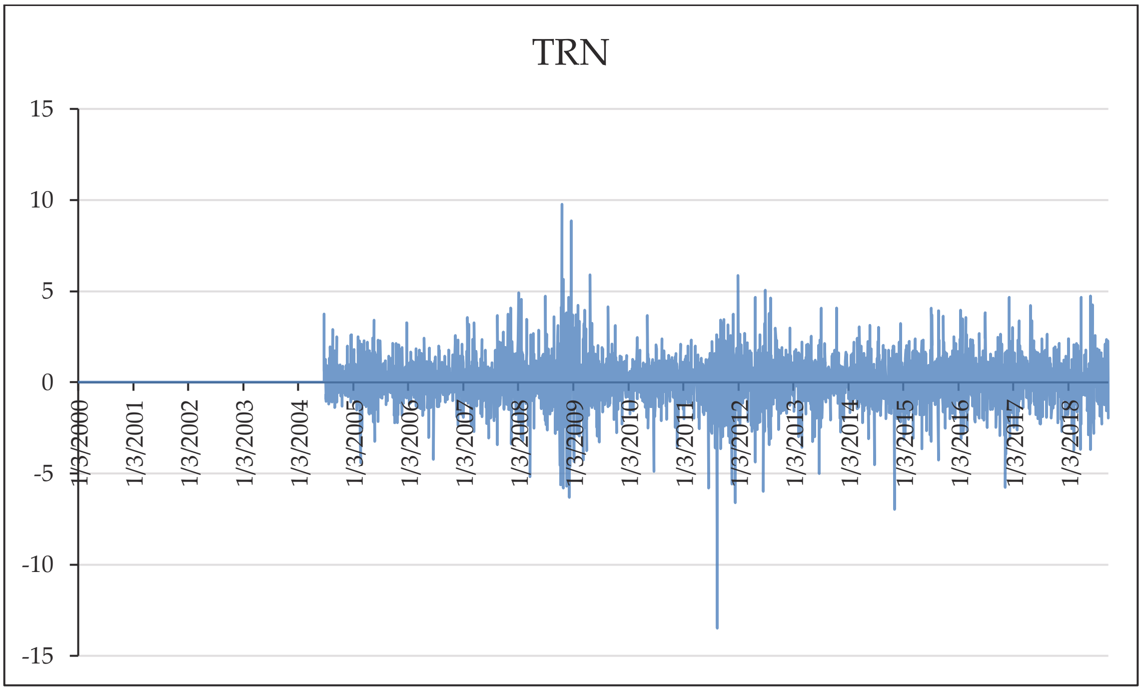


Figure G18 - TRN daily returns 2000-2018

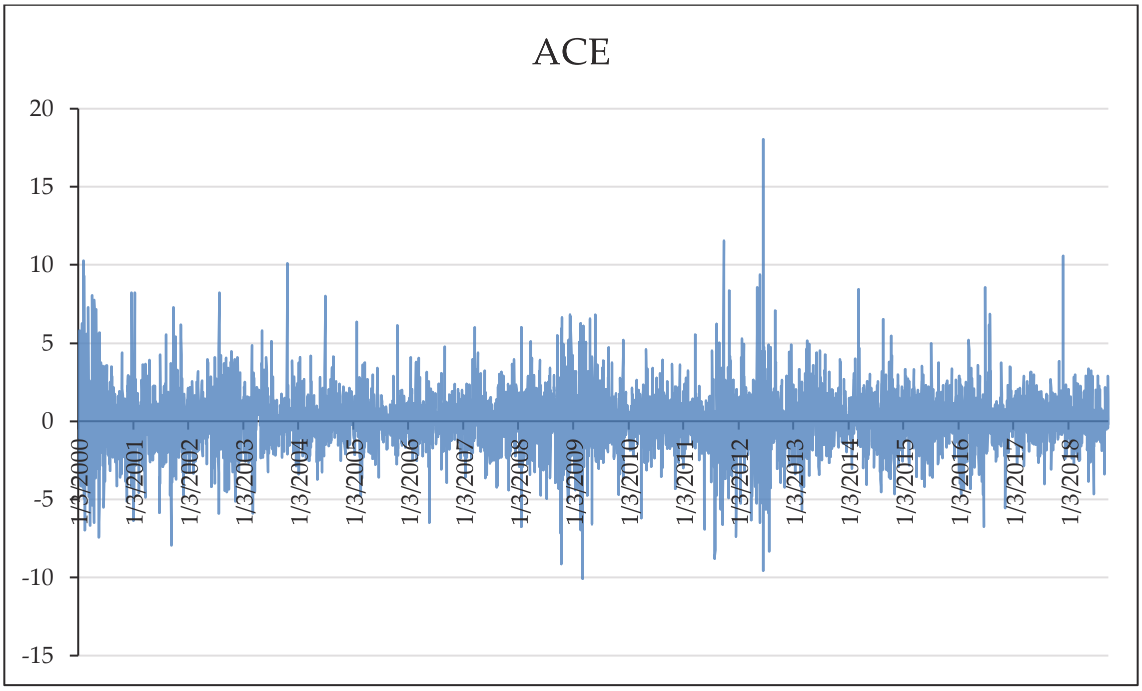


Figure G19 - ACE daily returns 2000-2018

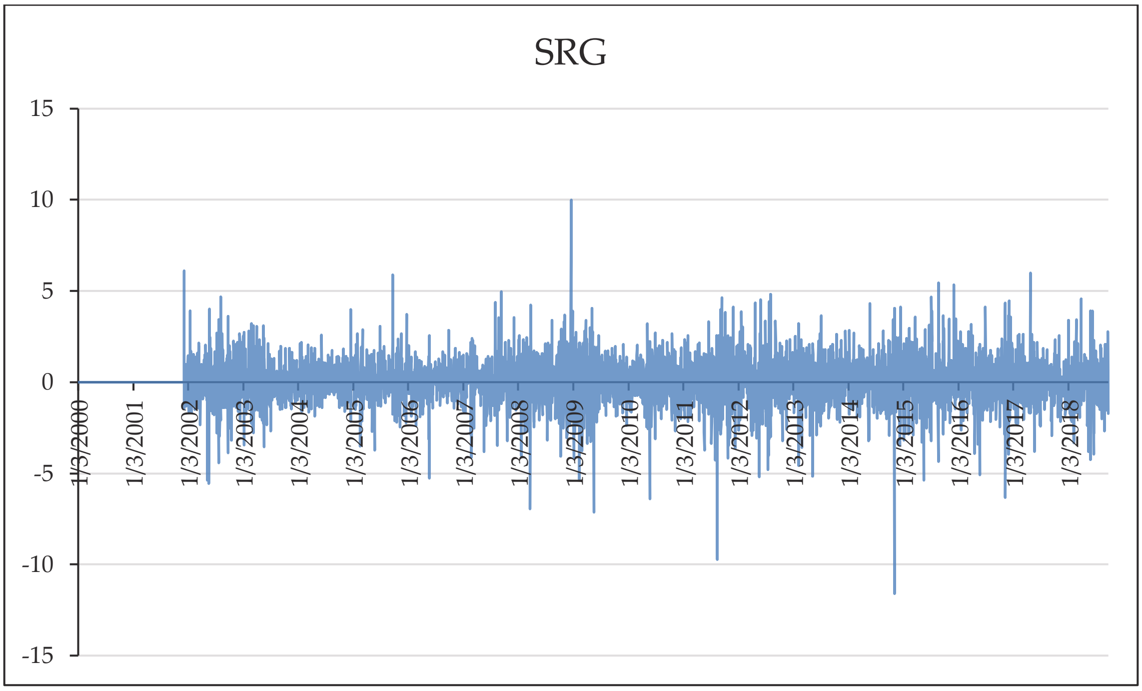


Figure G20 - SRG daily returns 2000-2108

Table G3 - OLS and LAD betas for entire period 2000-2018

Stock	*OLS (s.e. in parentheses)*	*OLS (HAC – s.e. in parentheses)*	*LAD (s.e.in parentheses)*
TRN	0.497 (0.011)	0.497 (0.016)	0.509 (0.012)
ACE	0.637 (0.016)	0.637 (0.020)	0.601 (0.016)
SRG	0.433 (0.011)	0.433 (0.016)	0.462 (0.011)

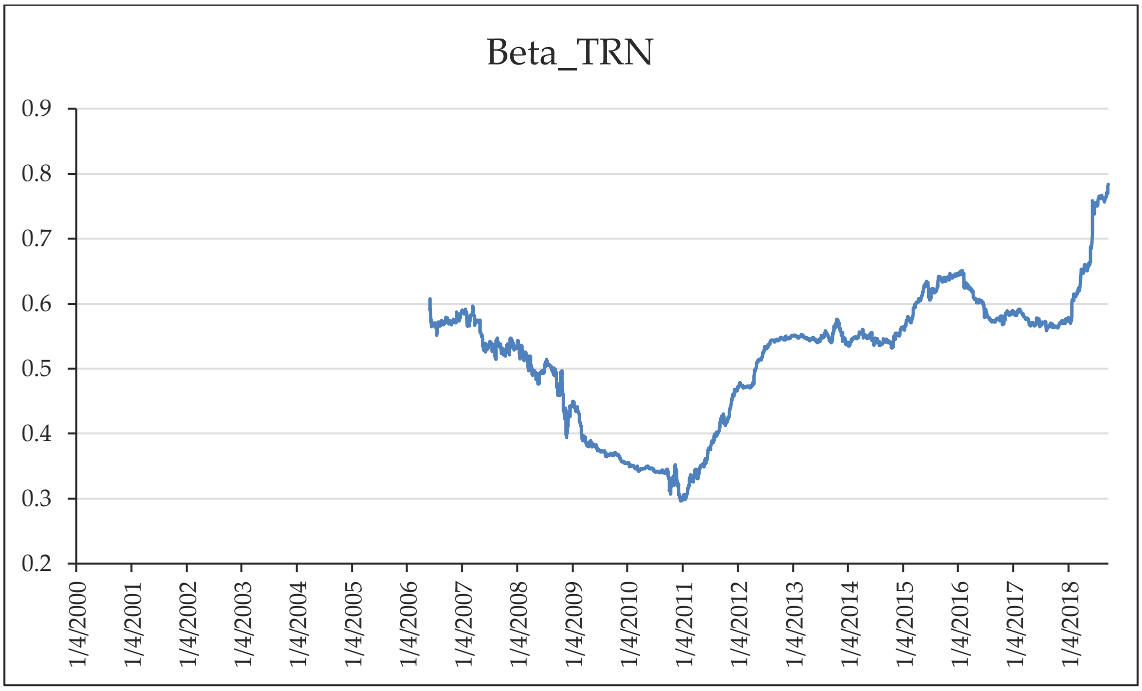


Figure G21 - TRN rolling 500 day OLS beta

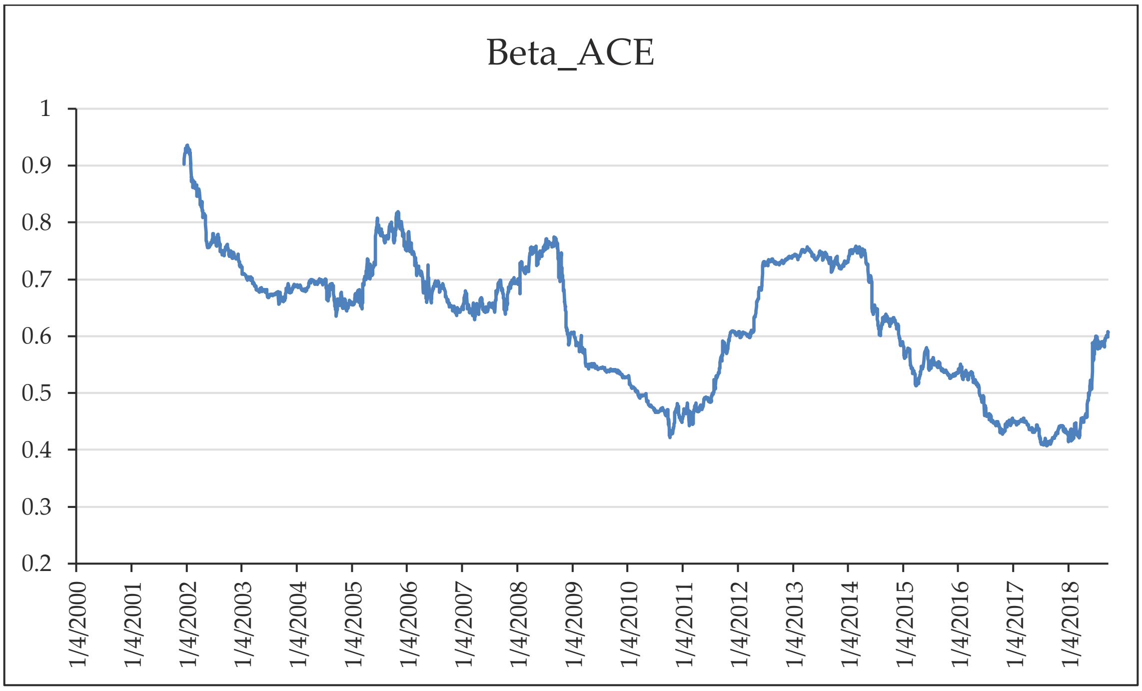


Figure G22 – ACE rolling 500 day OLS beta

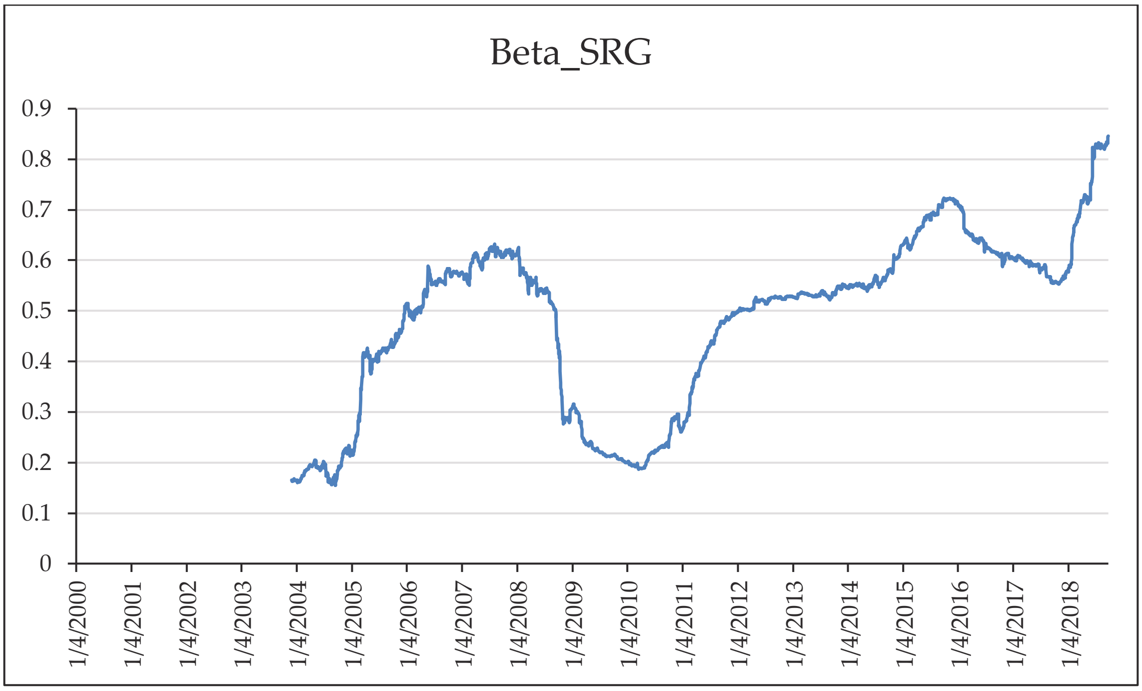


Figure G23 - SRG rolling 500 day OLS beta

GARCH model choices:

- TRN – Asymmetric DCC(1,1) – Full VECH would not converge
- ACE – Full VECH(1,1)
- SRG – Asymmetric DCC(1,1) – Full VECH was a close second in BIC

<u>TRN GARCH model residuals tests</u>

MV ARCH test lags 5 p=0.00

MV ARCH test lags 10 p=0.00

Univariate:

MIB residual

Lags 5 p=0.00

Lags 10 p=0.00

TRN residual

Lags 5 p=0.84

Lags 10 p=0.99

The TRN residuals are acceptable – those for MIB are presumably the cause of the multivariate test rejection.

<u>ACE GARCH model residuals tests</u>

MV ARCH test lags 5	p=0.25
MV ARCH test lags 10	p=0.67
Univariate:	
MIB residual	
Lags 5	p=0.01
Lags 10	p=0.06
ACE residual	
Lags 5	p=0.89
Lags 10	p=0.93

The residuals are acceptable.

<u>SRG GARCH model residuals tests</u>

MV ARCH test lags 5	p=0.00
MV ARCH test lags 10	p=0.00
Univariate:	
MIB residual	
Lags 5	p=0.00
Lags 10	p=0.01
SRG residual	
Lags 5	p=0.70
Lags 10	p=0.88

The residuals for SRG seem acceptable – those for the MIB series are less so.

Table G4 - OLS and GARCH betas

Stock	*OLS*	*Average of dailies*	*Beta from averages*
TRN	0.497 (0.011)	0.579	0.475
ACE	0.637 (0.016)	0.646	0.641
SRG	0.433 (0.011)	0.526	0.414

The impact of the asymmetric DCC model formulation – for TRN and SRG – seem to be reflected in the differences between the averages of the daily betas and the beta created from the average variance and covariance.

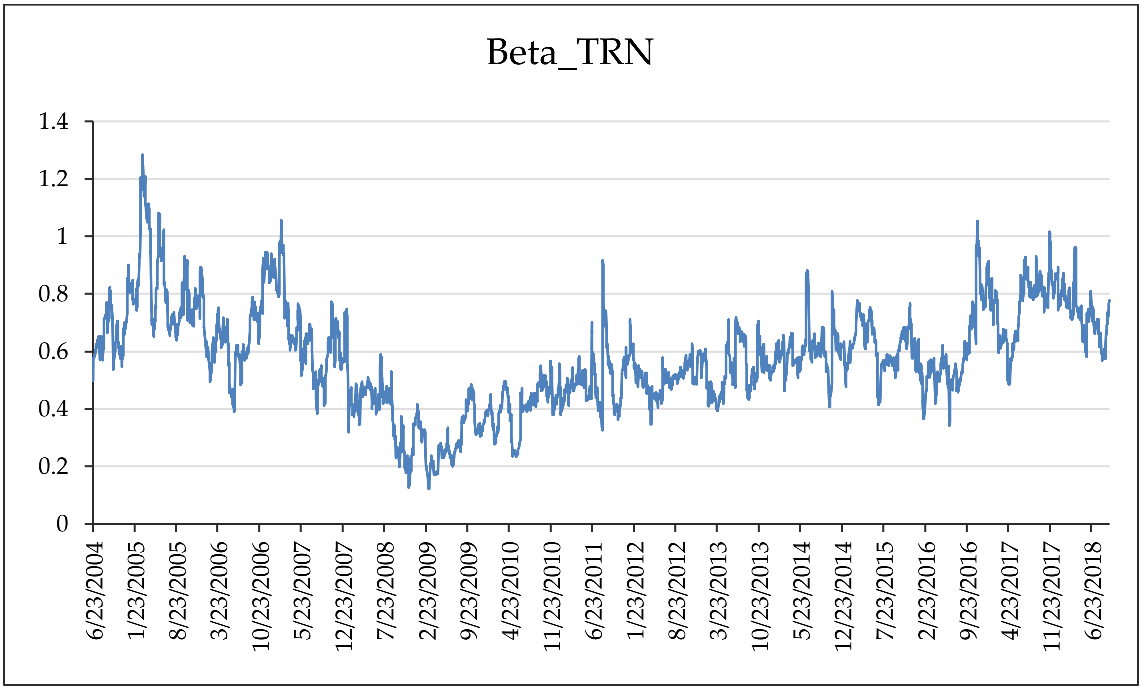


Figure G24 – TRN daily GARCH betas

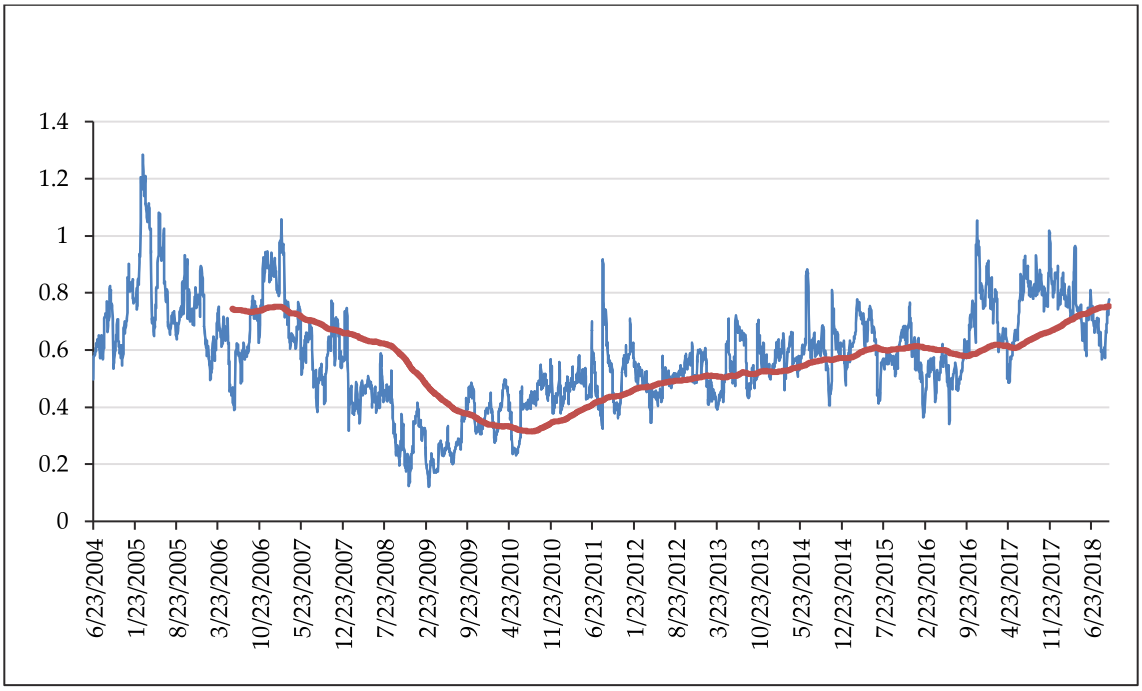


Figure G25 – TRN daily GARCH betas and 500 day MA

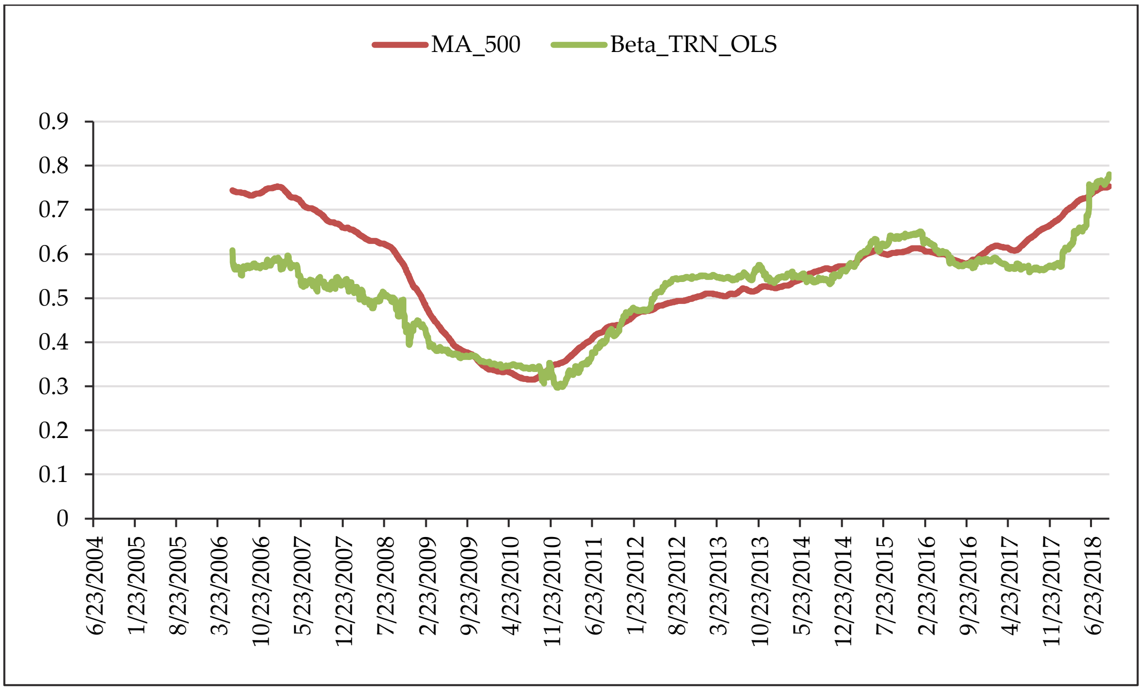


Figure G26 - TRN daily GARCH beta 500 day MA and rolling 500 day OLS beta (green)

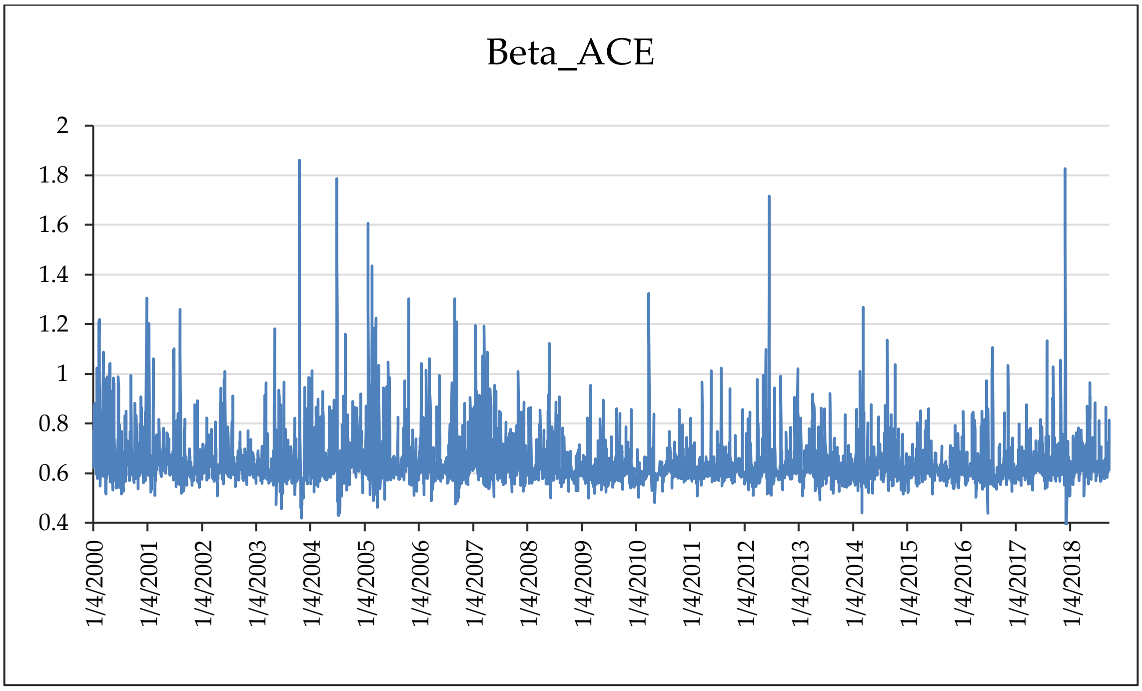


Figure G27 - ACE daily GARCH betas

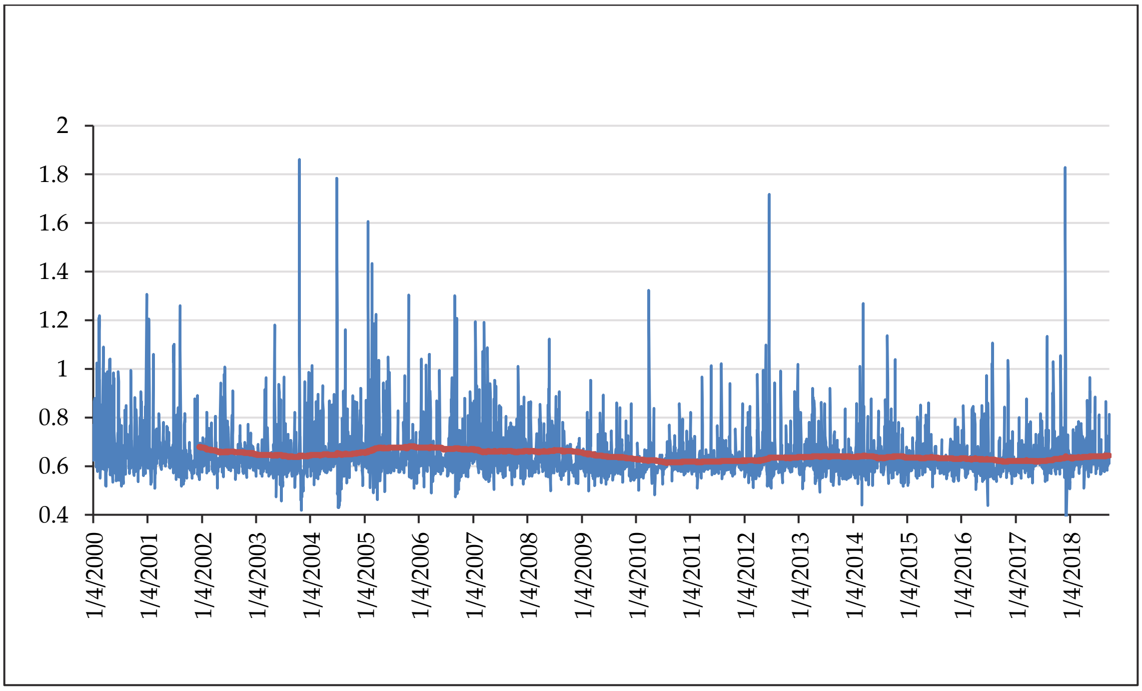


Figure G28 - ACE daily GARCH betas and 500 day MA

Figure G29 - ACE daily GARCH beta 500 day MA and rolling 500 day OLS beta (green)

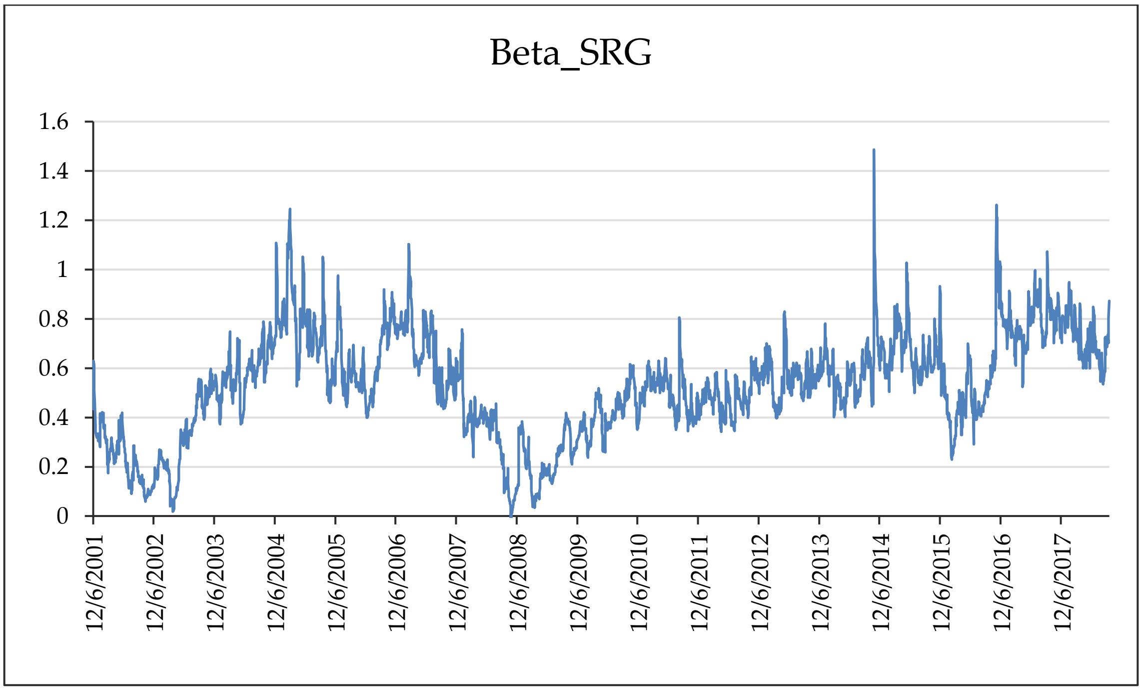


Figure G30 - SRG daily GARCH betas

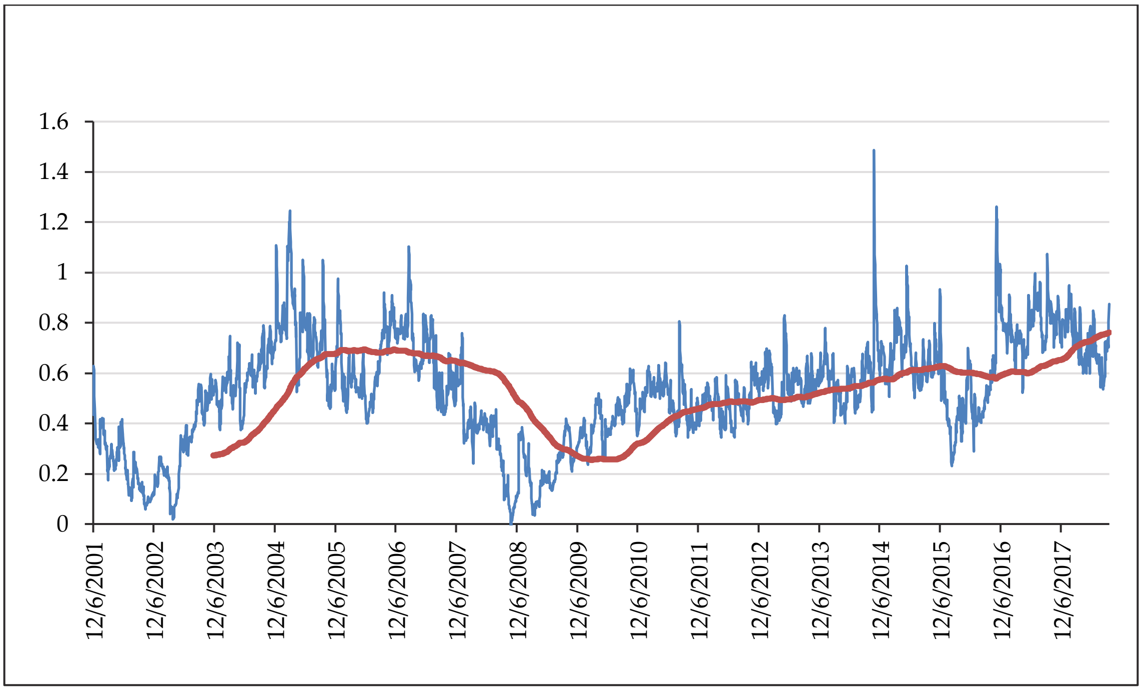


Figure G31 - SRG daily GARCH betas and 500 day MA

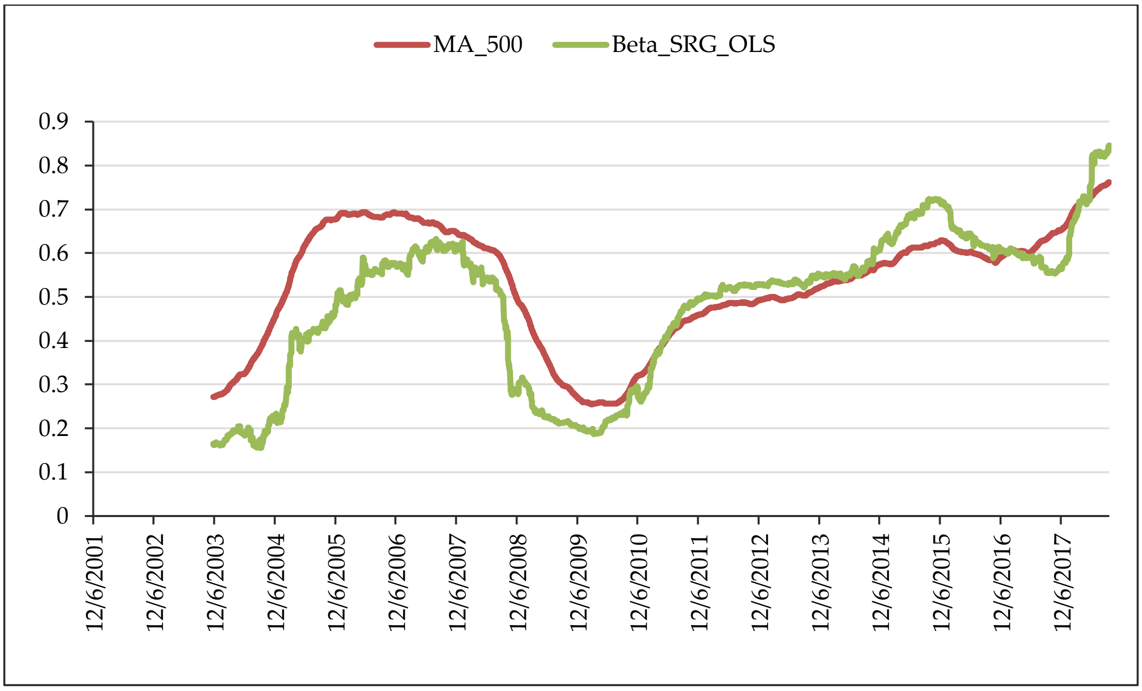


Figure G32 - SRG daily GARCH beta 500 day MA and rolling 500 day OLS beta (green)

G.3 Spain

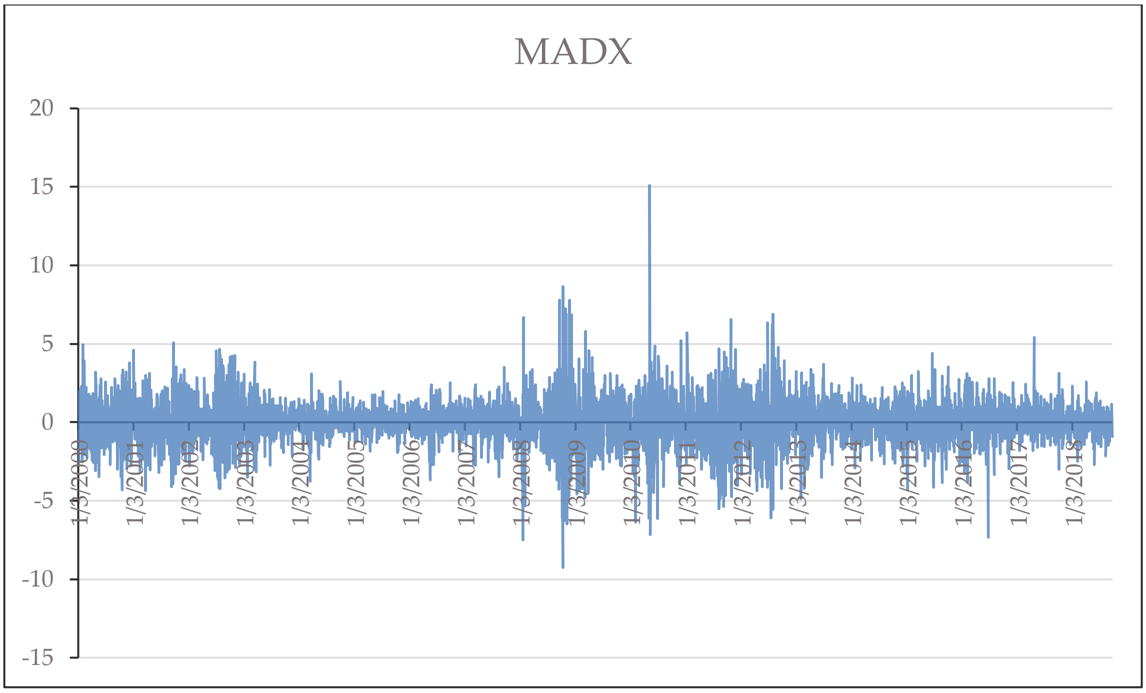


Figure G33 - MADX daily returns 2000-2018

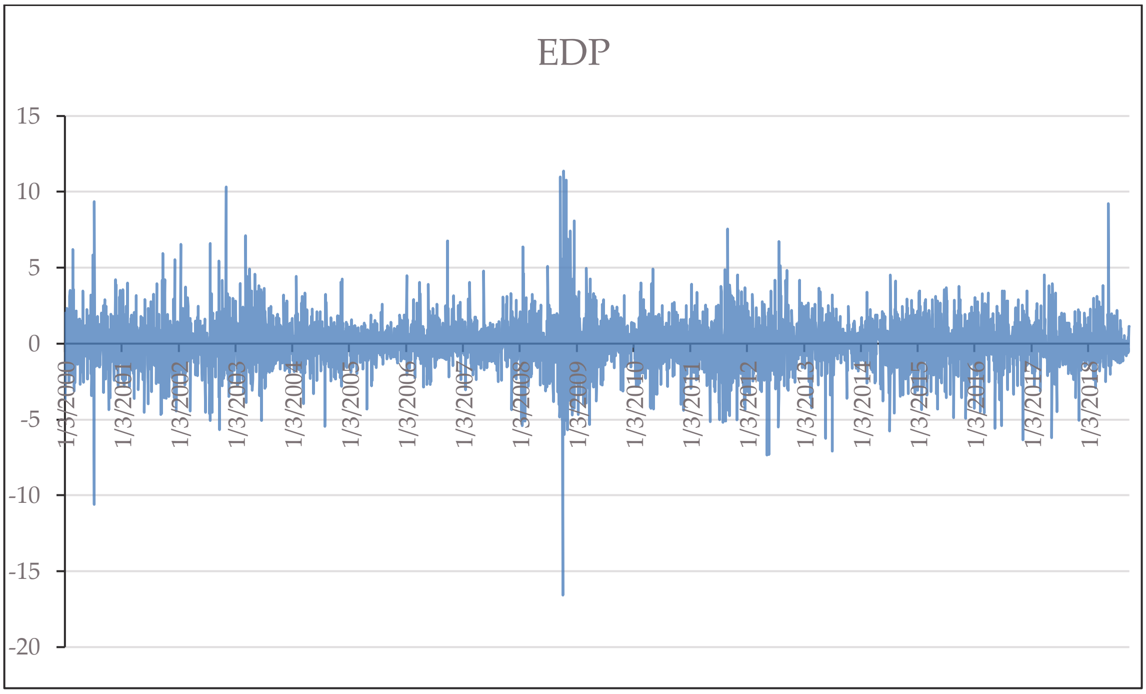


Figure G34 - EDP daily returns 2000-2018

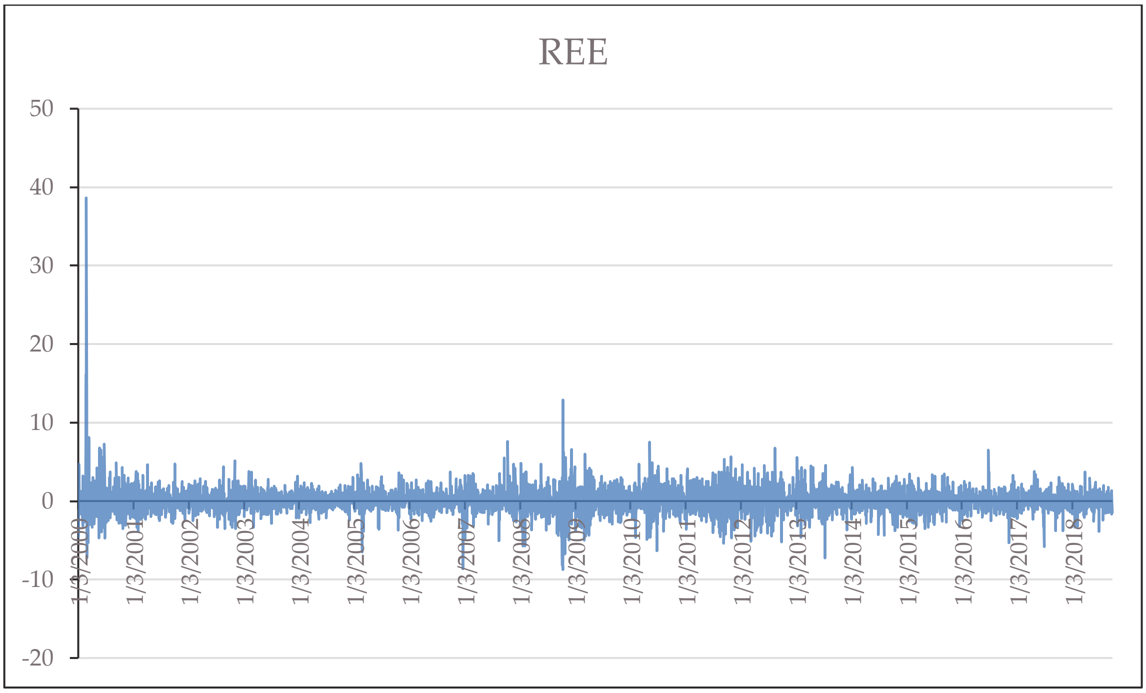


Figure G35 - REE daily returns 2000-2018

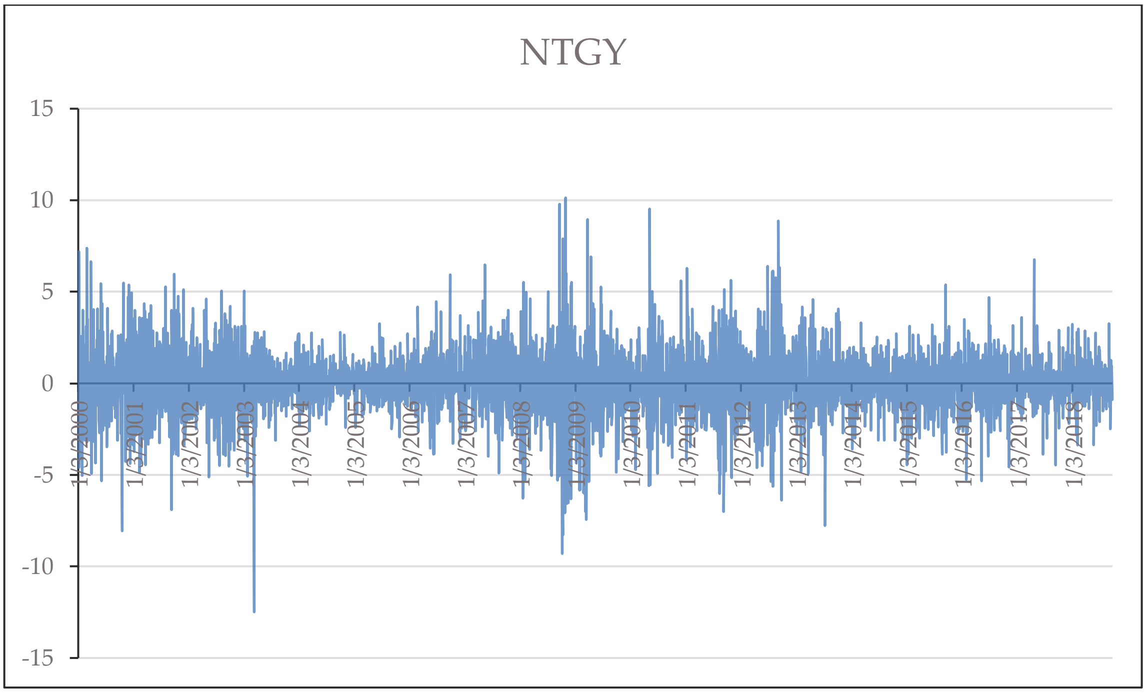


Figure G36 - NTGY daily returns 2000-2108

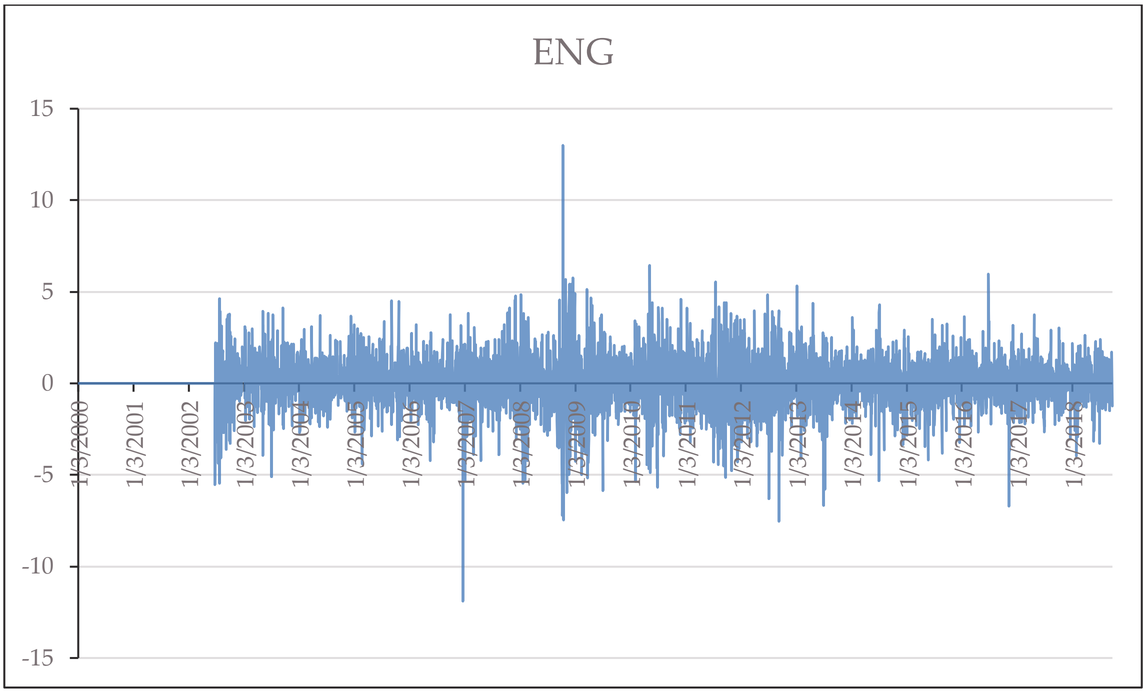


Figure G37 - ENG daily returns 2000-2018

Table G5 - OLS and LAD betas for entire period 2000-2018

Stock	*OLS (s.e. in parentheses)*	*OLS (HAC – s.e. in parentheses)*	*LAD (s.e.in parentheses)*
EDP	0.591 (0.014)	0.591 (0.021)	0.564 (0.014)
REE	0.588 (0.014)	0.588 (0.017)	0.563 (0.014)
NTGY	0.770 (0.013)	0.770 (0.018)	0.784 (0.013)
ENG	0.606 (0.013)	0.606 (0.018)	0.623 (0.013)

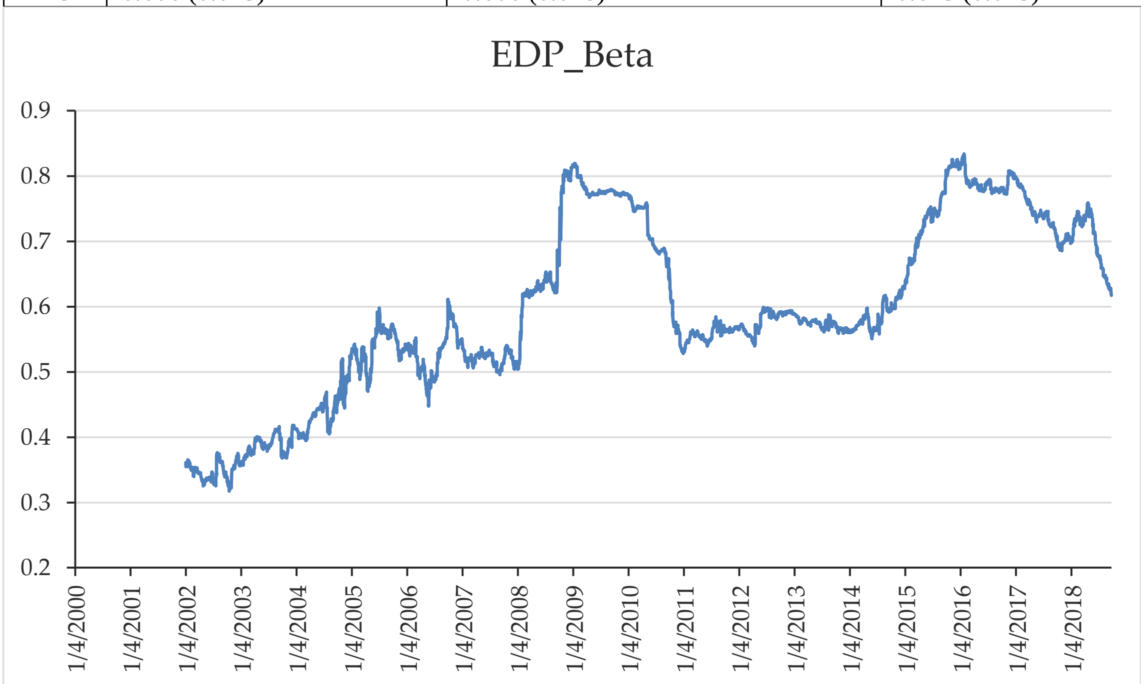


Figure 1 - EDP rolling 500 day OLS beta

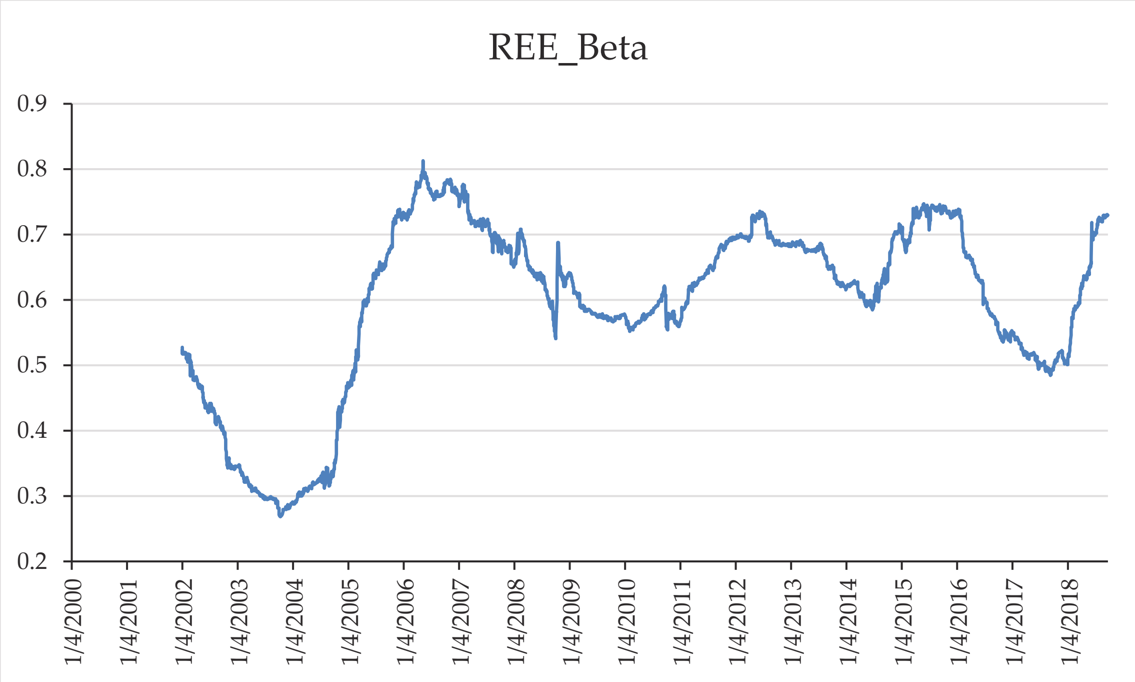


Figure G39 – REE rolling 500 day OLS beta

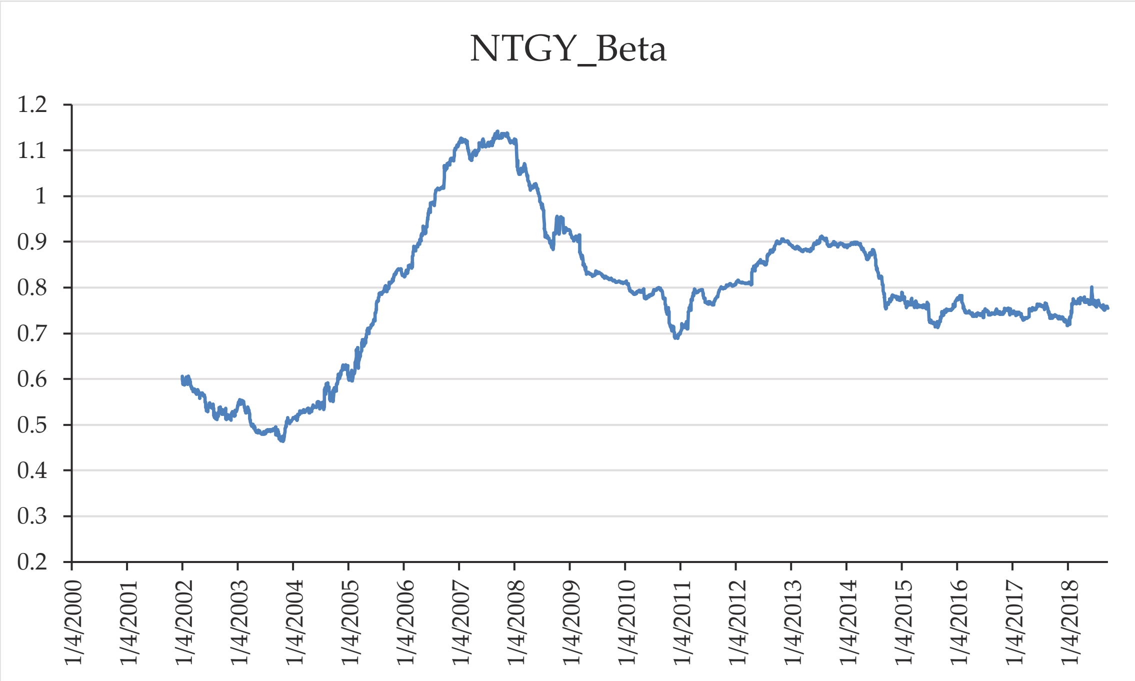


Figure G40 - NTGY rolling 500 day OLS beta

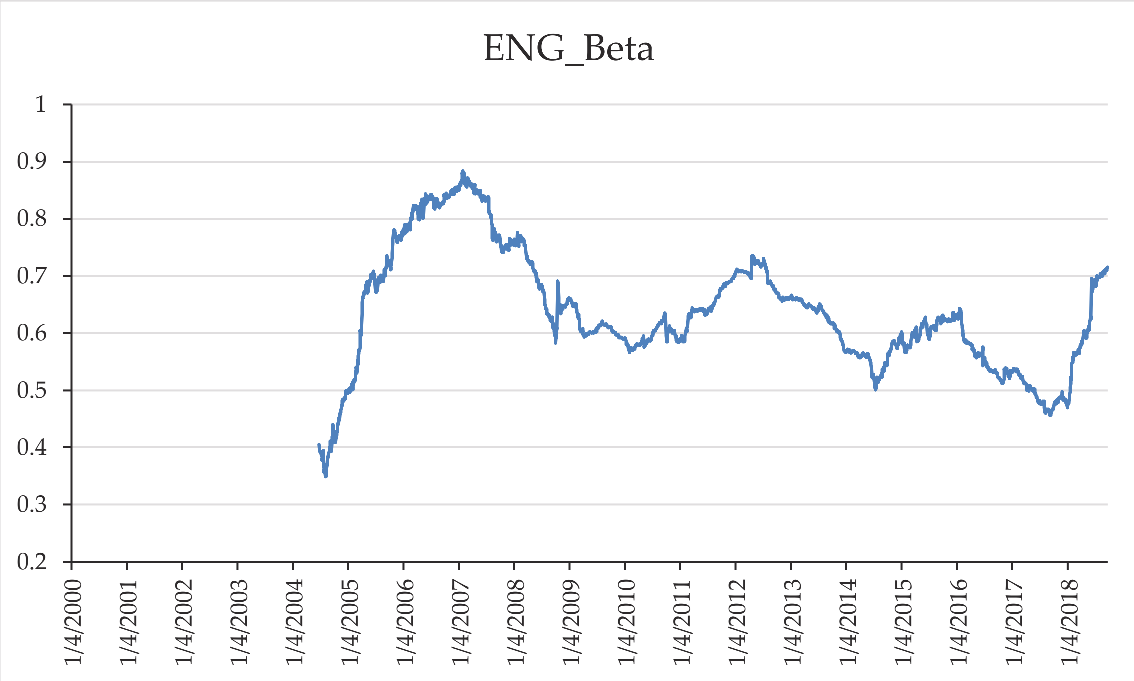


Figure G41 - ENG rolling 500 day OLS beta

GARCH model choices:

- EDP – DCC(1,1) – Cholesky had slightly better BIC but Cholesky properties are strange and I chose to disregard the model (and Full VECH converged)
- REE – Full VECH(1,1)
- NTGY – Asymmetric DCC(1,1) – even though Full VECH converged!
- ENG – DCC(1,1) – even though Full VECH converged

It is potentially of interest – if the result is more than chance – that different models seem to fit better in different markets/countries: the general strength of the Full VECH form outside Spain is not repeated in this market.

EDP GARCH model residuals tests

MV ARCH test lags 5 p=0.00

MV ARCH test lags 10 p=0.10

Univariate:

MADX residual

Lags 5 p=0.00

Lags 10 p=0.00

EDP residual

Lags 5	p=0.83
Lags 10	p=0.86

EDP results are mixed – the residuals of the EDP series look acceptable while those of MADX are less satisfactory.

REE GARCH model residuals tests

MV ARCH test lags 5	p=0.00
MV ARCH test lags 10	p=0.00
Univariate:	
MADX residual	
Lags 5	p=0.00
Lags 10	p=0.00
REE residual	
Lags 5	p=0.00
Lags 10	p=0.00

The residual results are not satisfactory – perhaps a higher order model is appropriate? Project time constraints preclude further investigation.

NTGY GARCH model residuals tests

MV ARCH test lags 5	p=0.00
MV ARCH test lags 10	p=0.00
Univariate:	
MADX residual	
Lags 5	p=0.00
Lags 10	p=0.00
NTGY residual	
Lags 5	p=0.21

Lags 10 p=0.11

The residuals for NTGY are acceptable, those for the market are not.

ENG GARCH model residuals tests

MV ARCH test lags 5 p=0.00

MV ARCH test lags 10 p=0.00

Univariate:

MADX residual

Lags 5 p=0.00

Lags 10 p=0.00

ENG residual

Lags 5 p=0.36

Lags 10 p=0.11

Again, the utility residuals are acceptable while those for the market are much less so.

Table G6 - OLS and GARCH betas

Stock	***OLS***	***Average of dailies***	***Beta from averages***
EDP	0.591 (0.014)	0.613	0.587
REE	0.588 (0.014)	0.619	0.590
NTGY	0.770 (0.013)	0.818	0.740
ENG	0.606 (0.013)	0.660	0.603

Despite the residual error issues, estimates are quite consistent with each other and OLS.

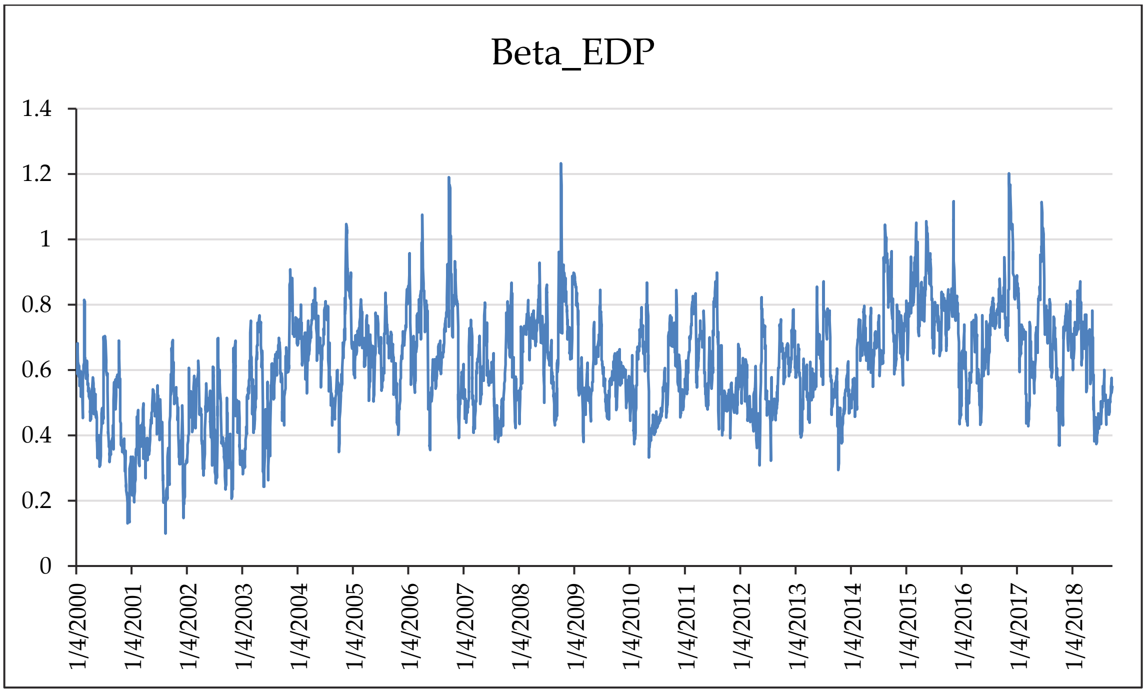


Figure G42 - EDP daily GARCH betas

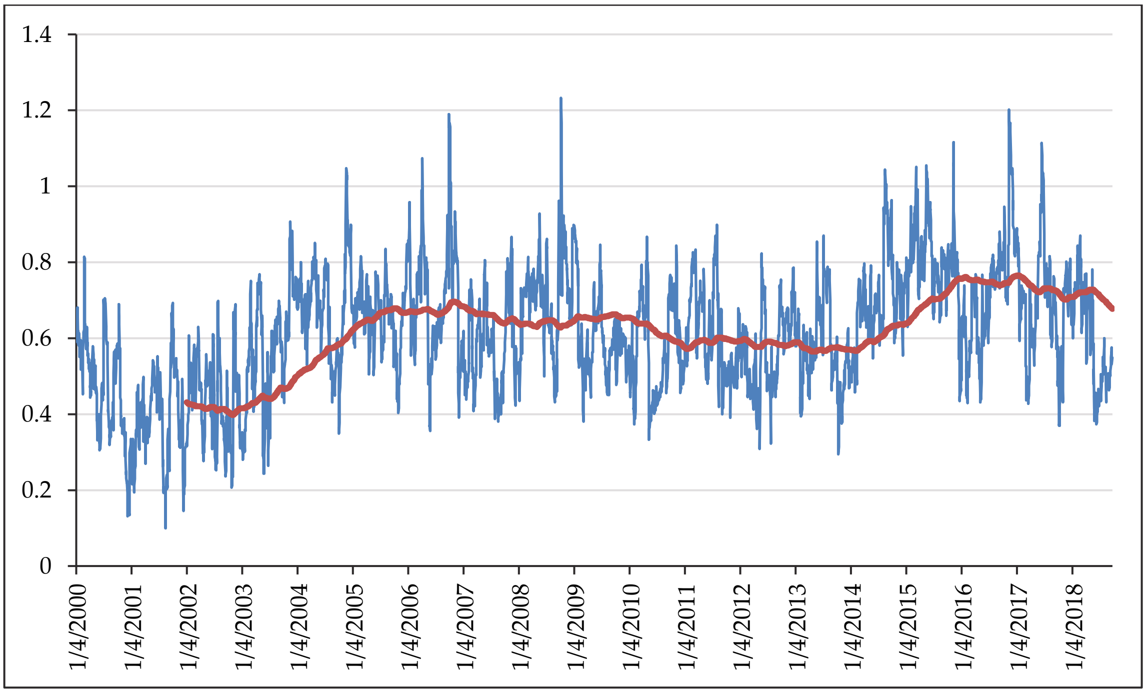


Figure G43 - EDP daily GARCH betas and 500 day MA

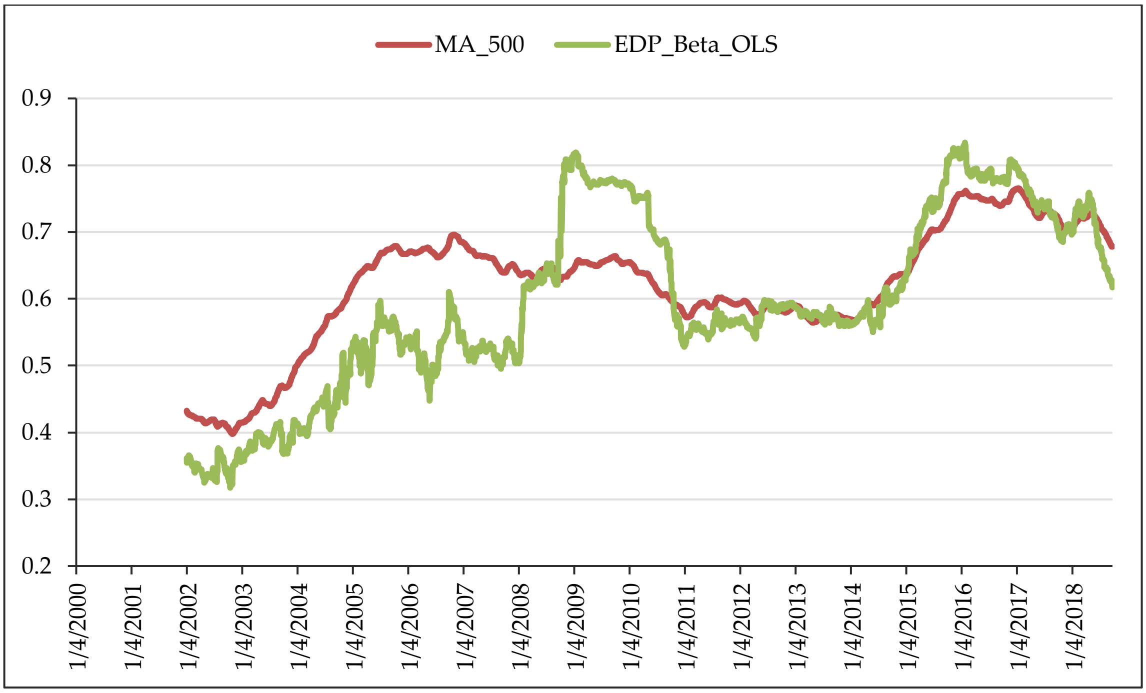


Figure G44 – EDP daily GARCH beta 500 day MA and rolling 500 day OLS beta (green)

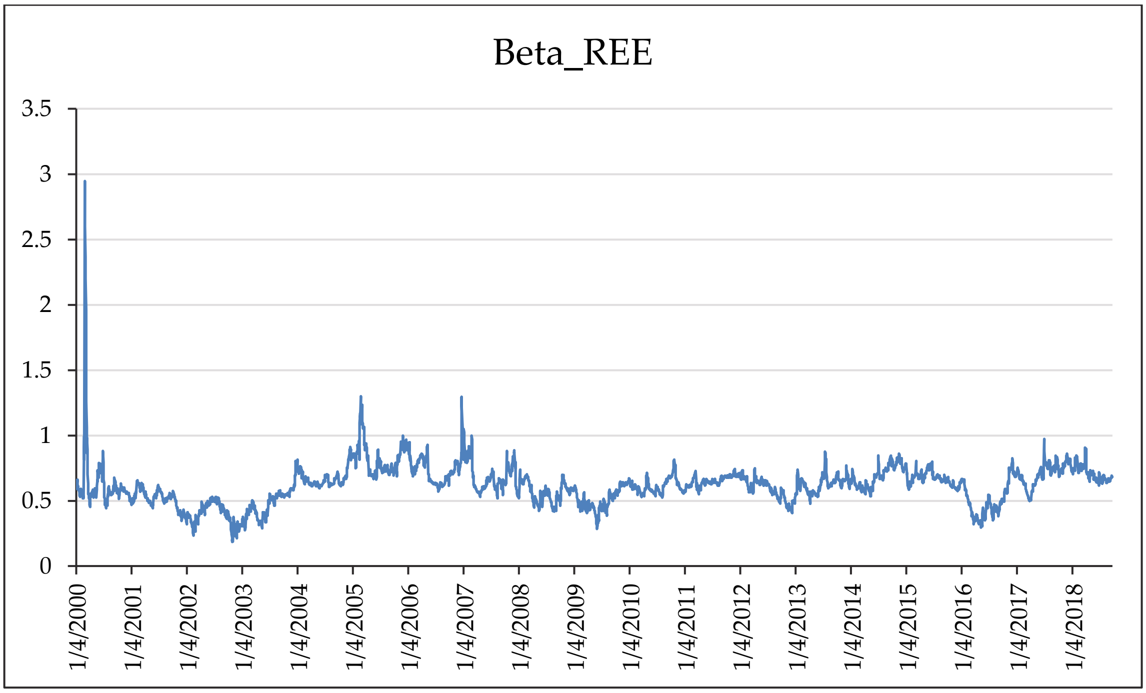


Figure G45 - REE daily GARCH betas

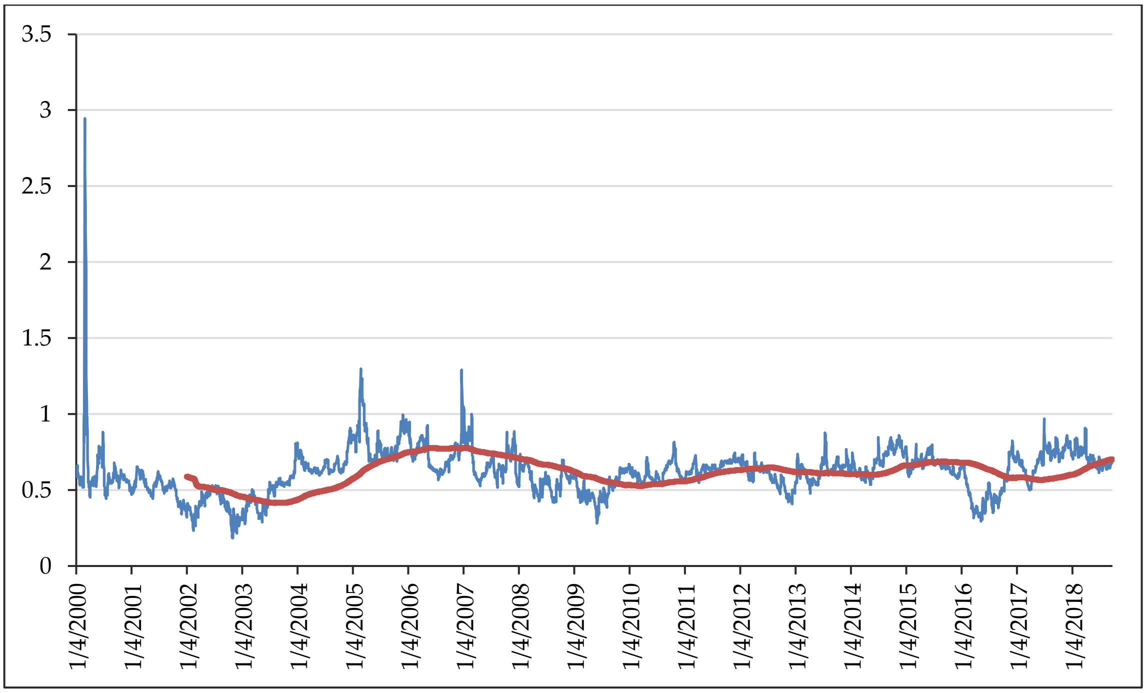


Figure G46 - REE daily GARCH betas and 500 day MA

Figure G47 – REE daily GARCH beta 500 day MA and rolling 500 day OLS beta (green)

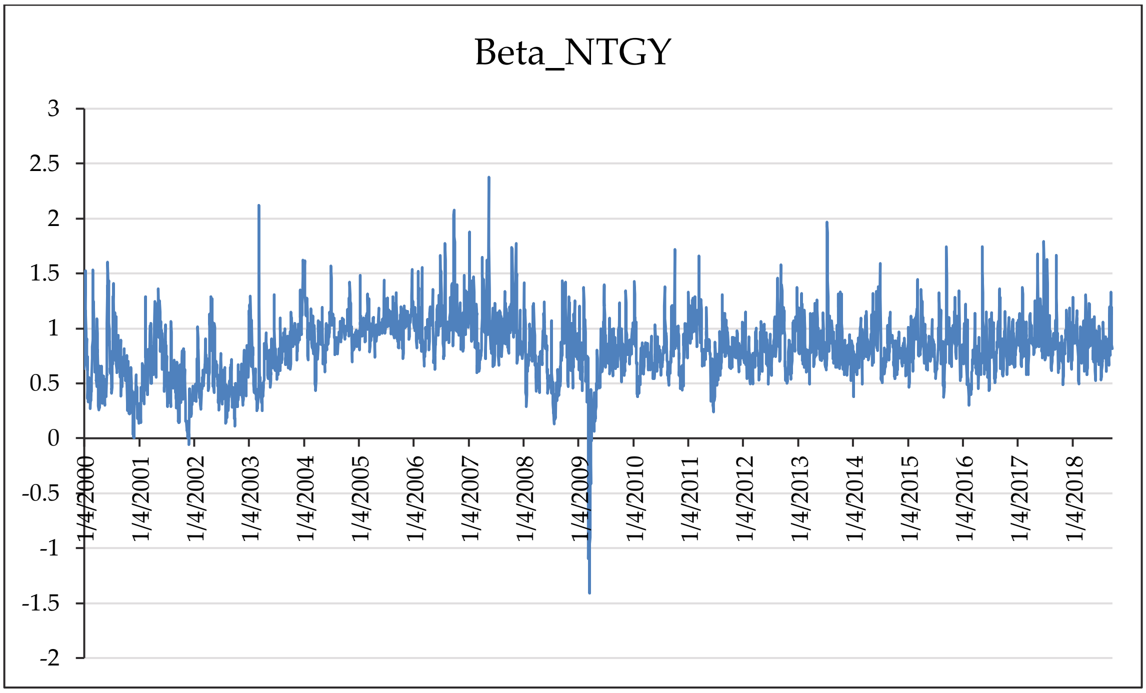


Figure G48 - NTGY daily GARCH betas

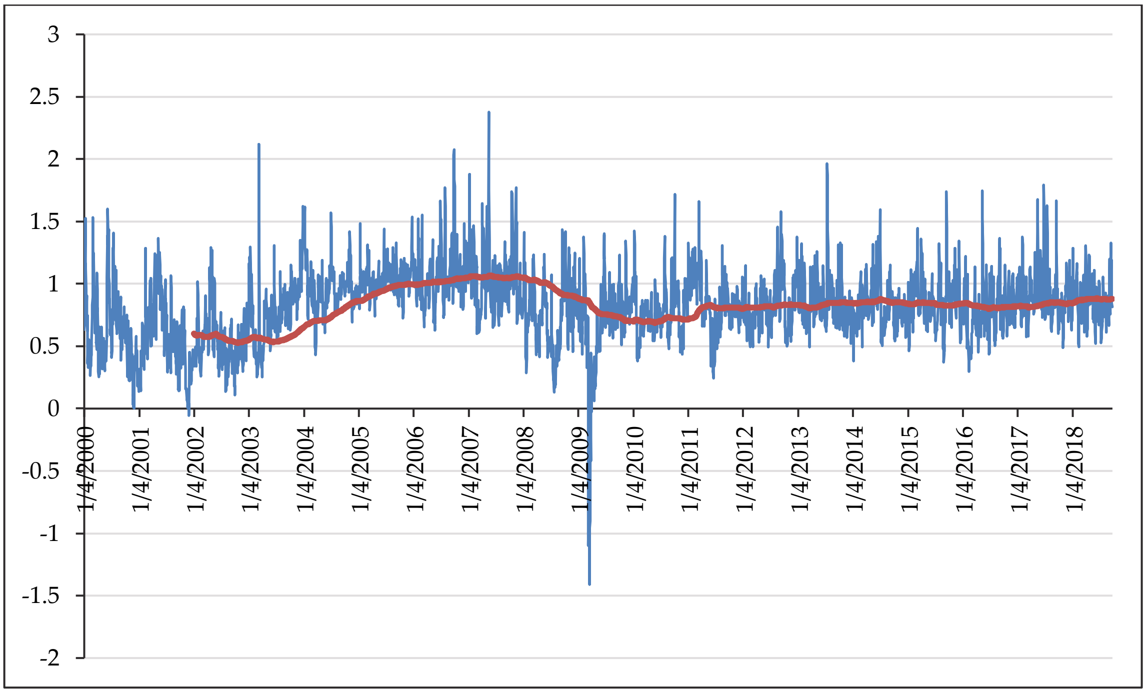


Figure G49 - NTGY daily GARCH betas and 500 day MA

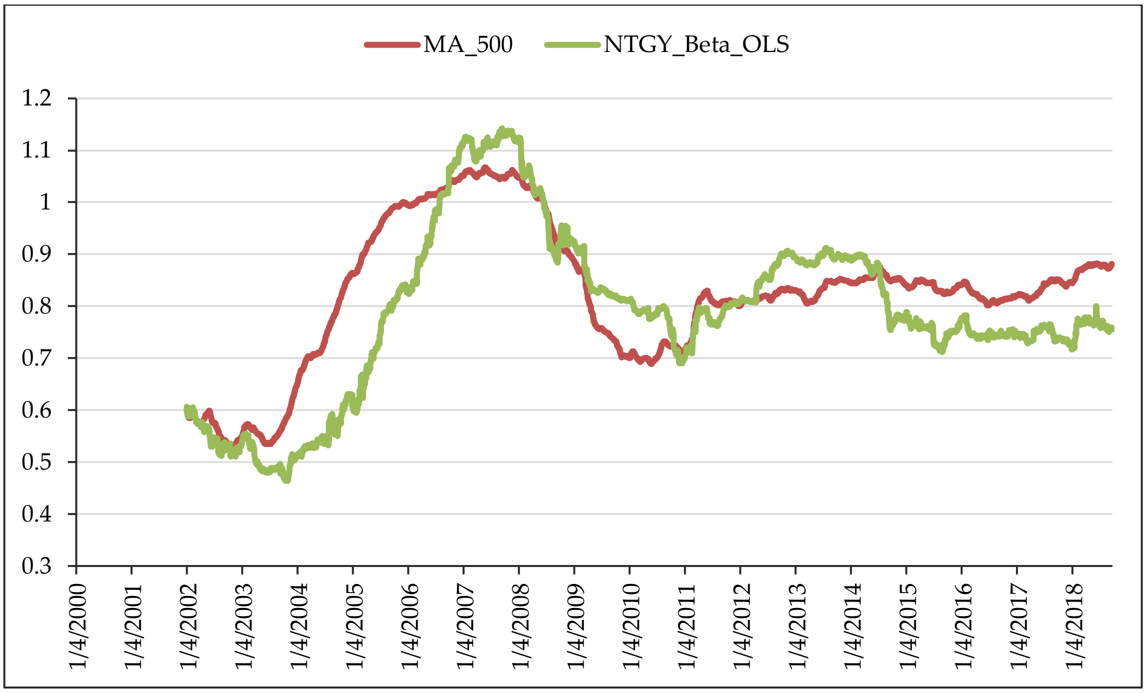


Figure G50 –NTGY daily GARCH beta 500 day MA and rolling 500 day OLS beta (green)

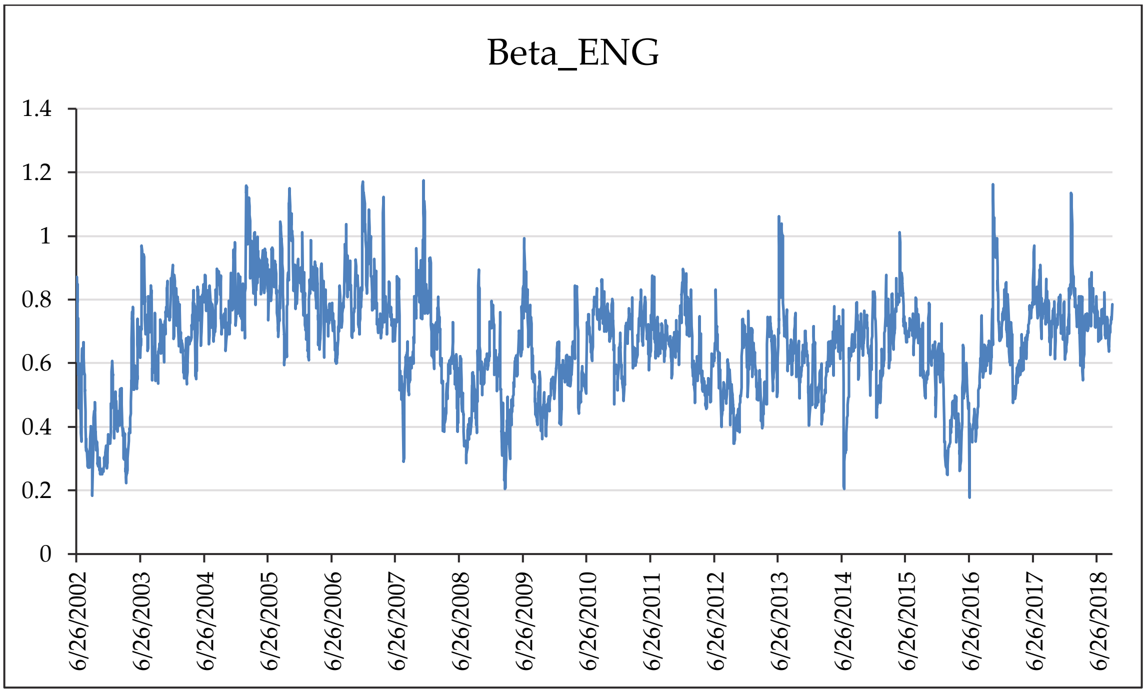


Figure G51 - ENG daily GARCH betas

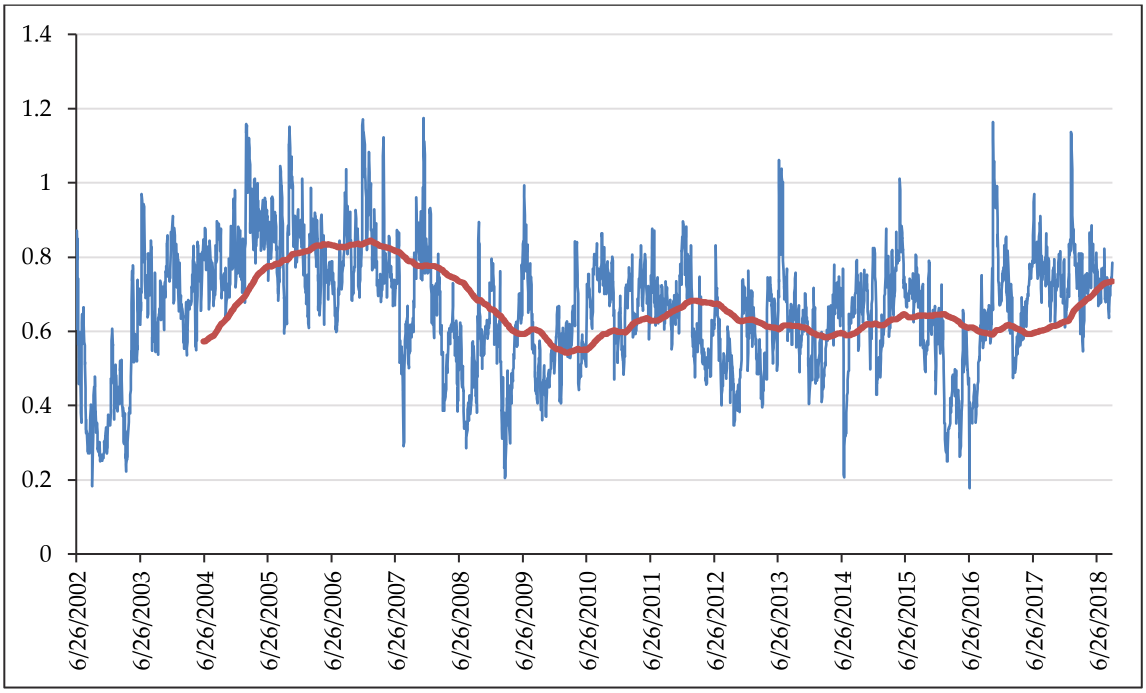


Figure G52 - ENG daily GARCH betas and 500 day MA

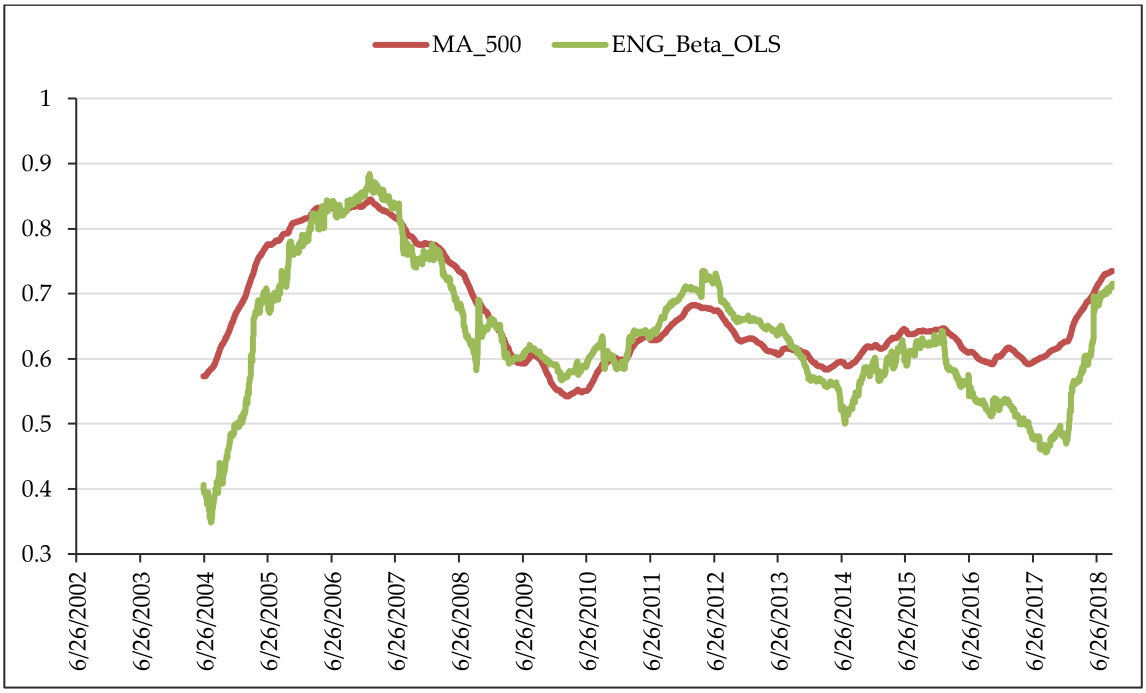


Figure G53 - ENG daily GARCH beta 500 day MA and rolling 500 day OLS beta (green)

Appendix H Forecasting with the GARCH βs

H1 Aim

To produce two years of daily beta forecasts from the time series created by the GARCH models.

H2 Method

The forecasts are created using the *auto.arima* function from the R *forecast* package (Hyndman, R. J. and Y. Khandakar 2008). The BIC is used for model selection (Schwarz, G. 1978). The forecast mean was written to a csv file.

Sample R code (data to be in R time series format):

```
uu_mod<-auto.arima(uu_ts, ic=c("bic"), stepwise=FALSE, approximation=FALSE)

uu_mod

Series: uu_ts

ARIMA(0,1,2)

Coefficients:

         ma1      ma2

     -0.0428  -0.0635

s.e.   0.0146   0.0151


sigma^2 estimated as 0.003996:  log likelihood=6352.66

AIC=-12699.32   AICc=-12699.32   BIC=-12679.94

uu_fc<-forecast(uu_mod, h=500)

write.csv(uu_fc$mean, "D:/uu_fc.csv")
```

H3 Results

The following model specifications were chosen:

Table H1 - Models chosen by auto.arima

Series	*ARIMA Model*
BT Full VECH(1,1)	(0,1,5)
NG Cholesky(2,2)	(0,0,0)
UU Triangular BEKK(1,1)	(0,1,2)
SSE Triangular BEKK(1,1)	(0,1,3)
PNN DCC(2,)	(1,1,1)
SVT Full VECH(1,1)	(1,1,2)
NG Full VECH(1,1)	(0,1,0)
NG Triangular BEKK(2,2)	(0,1,3)
PNN Full VECH(1,1)	(1,1,1)

The mean values pre-forecast and for the forecast period are presented below:

Table H2 - Mean beta pre- and post-forecast

Series	*Pre-forecast mean*	*Forecast mean*
BT Full VECH(1,1)	0.99	0.94
NG Cholesky(2,2)	0.58	0.58
UU Triangular BEKK(1,1)	0.55	0.25
SSE Triangular BEKK(1,1)	0.54	0.17
PNN DCC(2,)	0.49	0.50
SVT Full VECH(1,1)	0.53	0.50
NG Full VECH(1,1)	0.61	0.65
NG Triangular BEKK(2,2)	0.59	0.44
PNN Full VECH(1,1)	0.49	0.35

The data are presented graphically below:

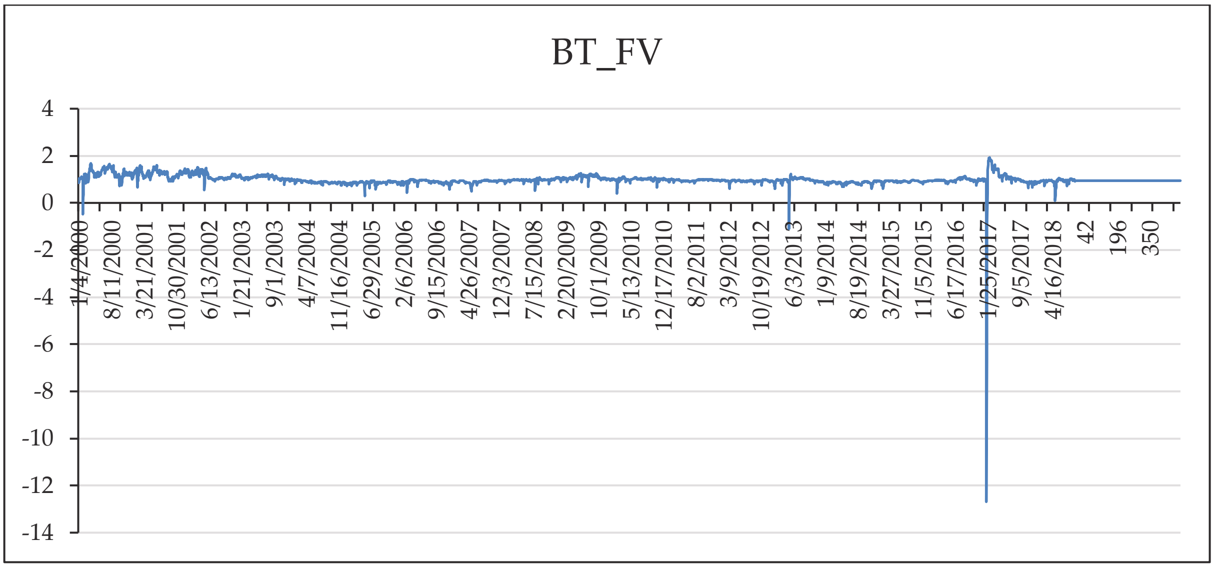


Figure H1 - BT Full VECH model with 500 day forecast

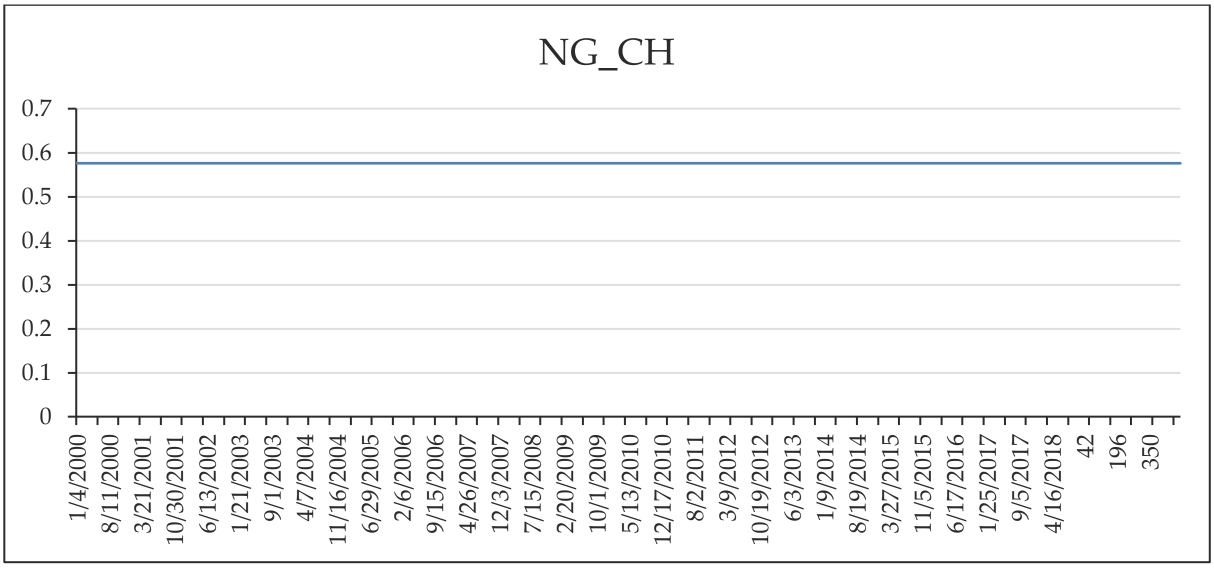


Figure H2 - NG Cholesky model with 500 day forecast

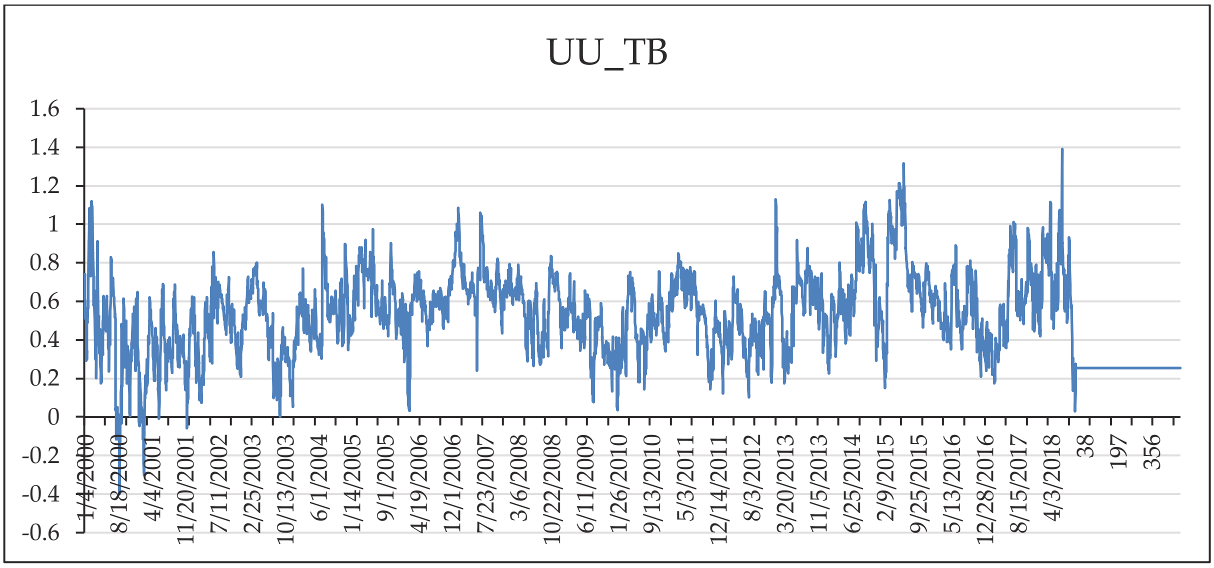


Figure H3 - UU Triangular BEKK model with 500 day forecast

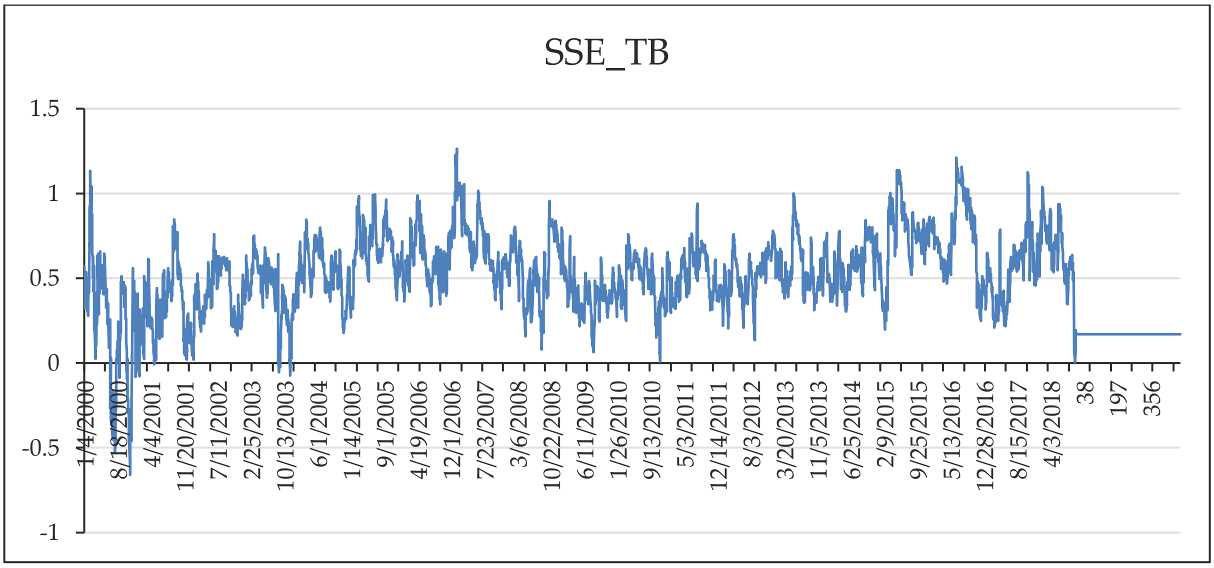


Figure H4 - SSE Triangular BEKK model with 500 day forecast

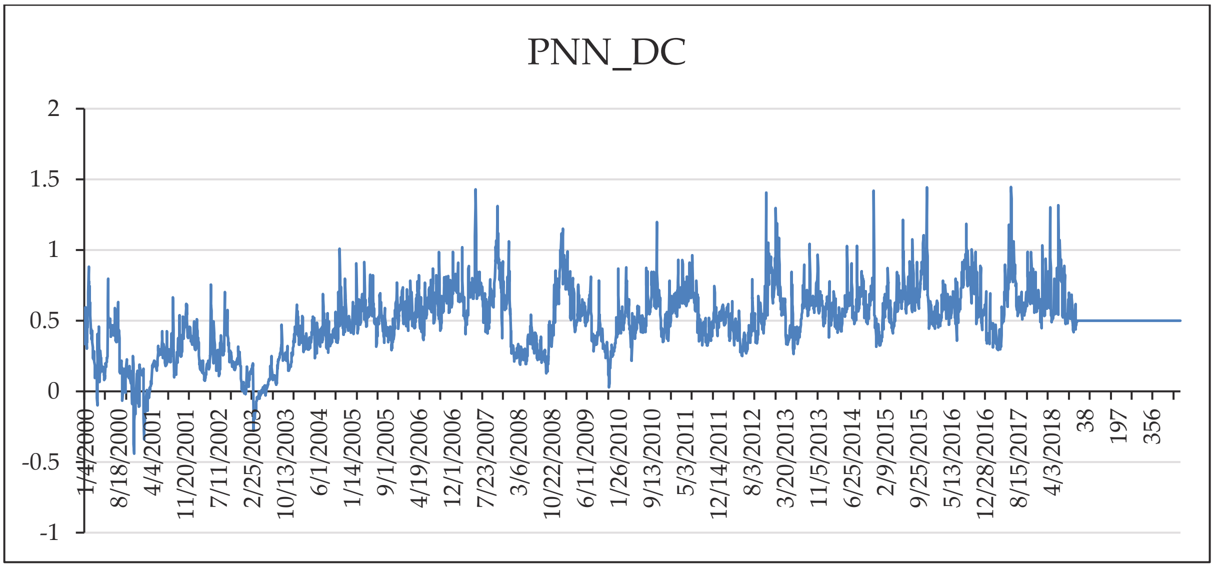


Figure H5 - PNN DCC model with 500 day forecast

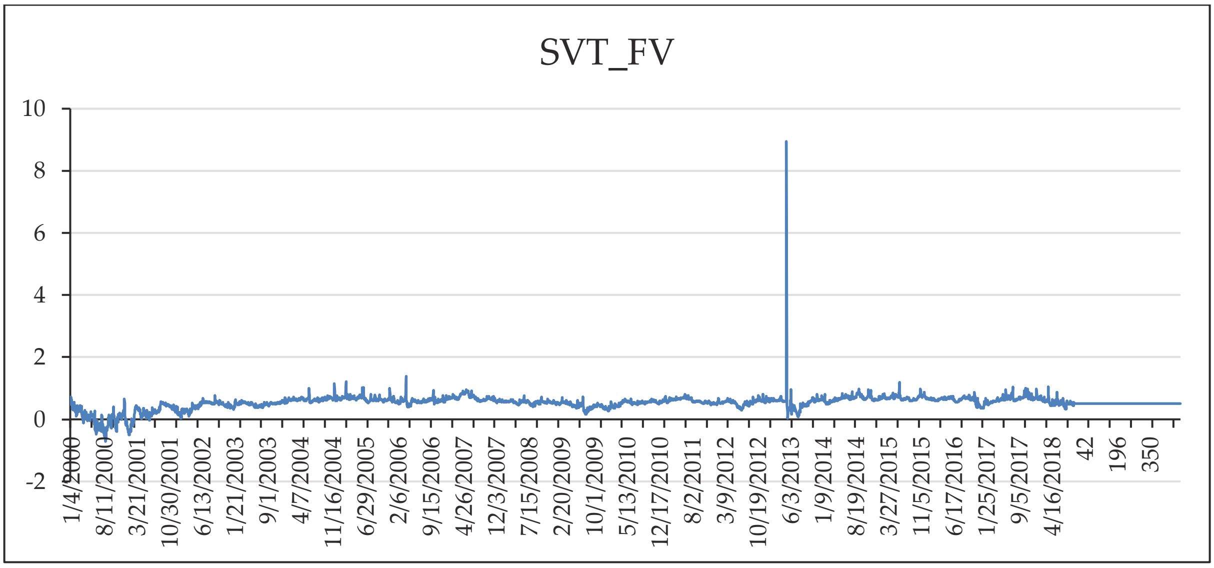


Figure H6 - SVT Full VECH model with 500 day forecast

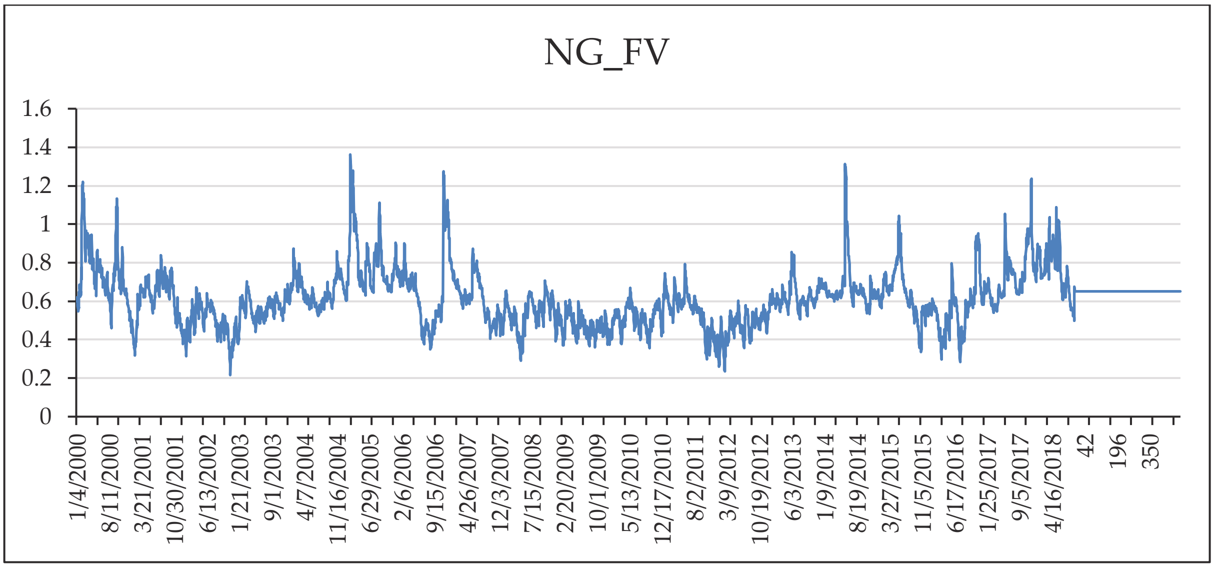


Figure H7 - NG Full VECH model with 500 day forecast

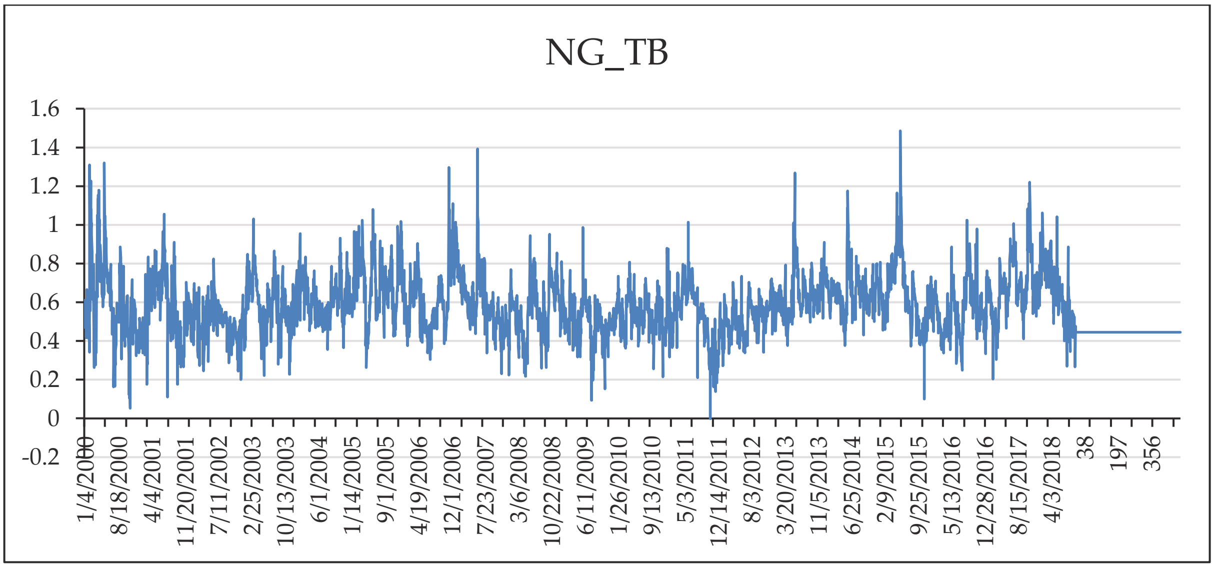


Figure H8 - NG Triangular BEKK model with 500 day forecast

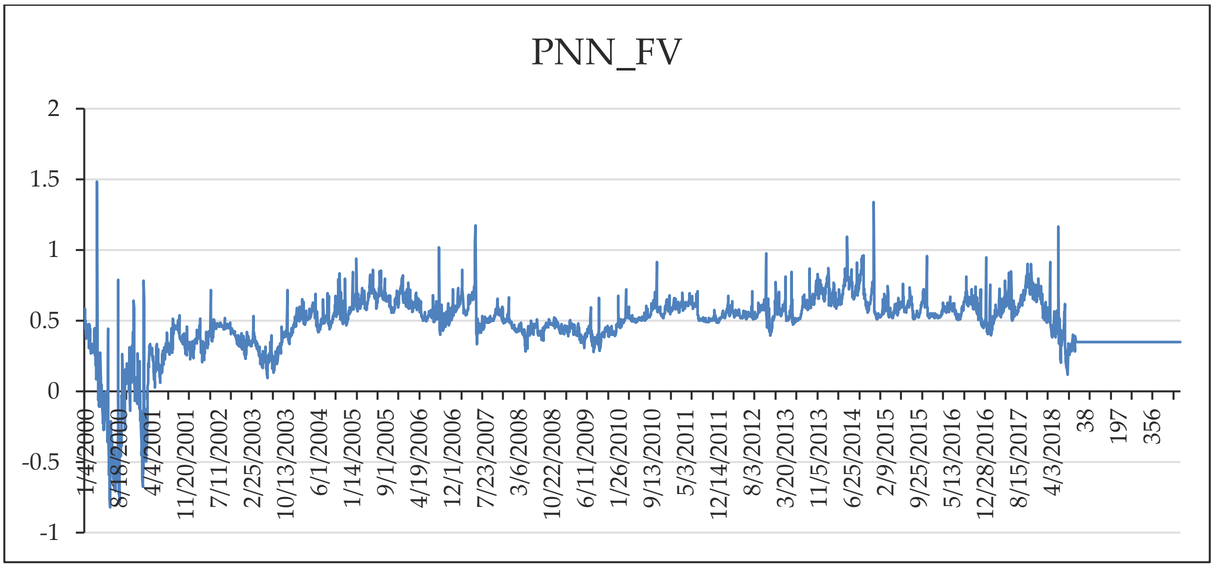


Figure H9 - PNN Full VECH model with 500 day forecast

H4 Discussion

It is interesting that the three Triangular BEKK models produce the least credible forecasts – and similar ARIMA models (0,1,2), (0,1,3) and (0,1,3). One assumes – without the time to do further research in the current project – that this reflects the stochastic properties of the Triangular BEKK model. Or, perhaps, the R routine has trouble building ARIMA models for such processes?

Alternatively, perhaps "low" forecasts tend to be obtained when the early series values are low – it may be worth estimating ARIMA on, say, the last 10 years to see what difference it makes. Current project timing precludes such investigation at this point.

On balance, as the forecast means for the other models are similar to the in-sample means, I would suggest using in-sample means for forecast periods.

Bibliography

HYNDMAN, R. J., AND Y. KHANDAKAR (2008): "Automatic Time Series Forecasting: The Forecast Package for R," *Journal of Statistical Software*, 26, 1-22.

SCHWARZ, G. (1978): "Estimating the Dimension of a Model," *Ann. Statist.*, 6, 461-464.