Appendix A: Conduct of the reference

Introduction

1. This is the first reference of a NATS licence modification decision under the Transport Act 2000 (TA 2000).

2. We note that the Special Reference Group has discretion to determine the appropriate process for this reference.

3. Throughout this investigation, we have had regard to the duties of the Secretary of State and the CAA in s.1 and s.2 of TA 2000.

Core submissions from the Parties

4. The core submissions from the Parties were as follows:

   - CAA submitted the Reference to the CMA of the NERL RP3 price controls and Notice of Reference on 19 November 2019 (CAA Reference).

   - NERL submitted its Statement of Case and supporting documents on 28 November 2019 (NERL Statement of Case).

   - CAA sent its response to NERL's Statement of Case on 18 December 2019 (CAA Response).

   - NERL submitted its Reply to the CAA Response, and supporting documents, on 30 December 2019 (NERL Reply).

5. The Parties supplied additional information during the course of the reference when requested by the CMA.

Evidence from third parties

6. We invited representations from third parties, following publication of the Statement of Case. We received representations and supporting documents from 19 third parties.

7. Where we considered a hearing would be helpful in clarifying or understanding the representations, or where we had specific questions concerning elements of the price control or RP3 process, we invited third parties to a hearing (in person or via a call).

8. In addition, we had calls with the CCWG Co-chairs and with the authors of the UKRN report on cost of capital for regulated companies, to ensure full understanding of their reports.

Site visit and hearings

9. Before the reference was formally made, we held technical teach-ins with the Parties. We attended a site visit at NERL's Swanwick air traffic control centre. We then held main hearings in London over two days in February with both Parties.

Provisional findings

10. We published our provisional findings on 24 March 2020, and invited representations from third parties and the main parties.

11. We received responses from the CAA and NERL, and from 15 third parties.

Second consultation

12. We published a consultation on how to proceed with the reference, in light of the impact of COVID-19, on 24 June 2020, and invited comments from parties.

13. We received responses from the CAA and NERL, and from eight third parties.

Final report

14. In line with requirements of the Transport Act 2000, we sent our final report on the reference to the CAA on 23 July 2020.

Transparency

15. We published all the core documents from the Parties on the CMA case page. We also published all the main representations from third parties.

16. Following the consultation on our provisional findings published in March, we have published the responses from the Parties, and the main responses from third parties, on the CMA case page.

17. We published our second consultation. We have also published the responses to that consultation.

18. We published the summary of our final report, and will publish the full non-confidential version of the document in due course.

Appendix B: Regulated revenue and charges in the CAA RP3 Decision

How allowances are set

Spending allowances

1. The CAA sets spending allowances for NERL during the price control period. These are:

   (a) The amount NERL can spend on en-route operational expenditure: ‘opex’. This includes all day-to day expenditure on running the en-route services, and payments into the pension funds. It is reduced or ‘offset’ by an amount set to reflect expected income from non-regulated activities.

   (b) The amount NERL can spend on capital investment: ‘capex’. This includes business-as-usual investments in buildings and equipment including IT, and the investment needed to deliver the airspace modernisation strategy.

   (c) The amount NERL can spend on opex to deliver the London Approach services.

   (d) The amount NERL can spend on capex to deliver the London Approach services.

   (e) The amount NERL can spend on opex to deliver the Oceanic services;

   (f) The amount NERL can spend on capex to deliver the Oceanic services.

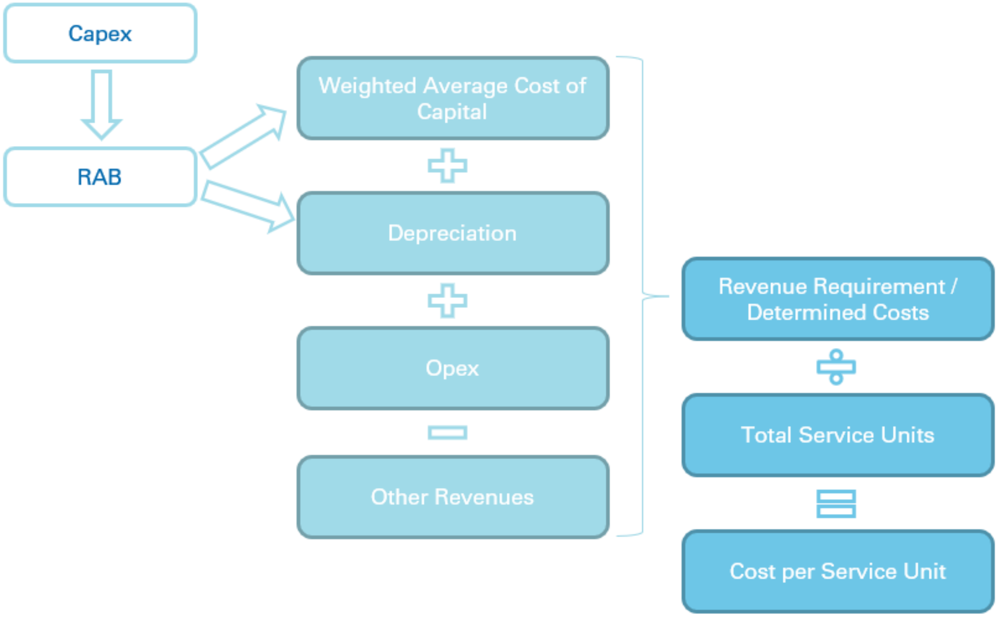
Revenue allowances

2. These spending allowances feed into the decision on how much NERL can charge its customers: the revenue allowances. These, like the spending allowances, are calculated for each calendar year.

3. To determine the revenue allowances, CAA has calculated NERL’s charges for RP3 on the basis of a ‘building block’ approach. This is the approach typically used in UK economic regulation.

4. The building block approach for NERL is illustrated below.

Figure B-1: the building blocks of NERL's revenue allowances



Source: Adapted from the CAA's Reference, page 25, Figure 4

The capex element

5. The first building block is the 'Regulatory Asset Base' at the start of the Price Control Period (PCP). This is the value of NERL's capital assets.

6. The capex allowance is not directly funded by revenues when it is incurred. Instead the expected increase in the asset base, resulting from the capex spend, is added to the starting RAB to give the RAB for each year in the PCP.

7. The capex revenue allowance is made of two elements:

   (a) Revenue to cover the cost of financing investments. NERL is permitted to charge an amount to cover the 'Weighted Average Cost of Capital' or WACC, applied to existing and new assets in the RAB, each year, over the lifetime of each asset.

   (b) Revenue to cover depreciation costs associated with the assets in the RAB.

The opex element

8. The opex revenue allowance is determined directly by the CAA, based on its assessment of NERL's efficient costs over the period (the opex element of the spending allowance).

Revenue requirement or ‘determined costs’

9. Adding the capex revenue allowance to the opex revenue allowance, net of the other revenues (such as forecast income from non-regulated services), produces the total revenue requirement or ‘determined costs’) for regulated services, for each year in the PCP.

Charges

10. Air traffic service users pay charges per ‘service unit’. This is a standard definition based on the weight of the aircraft and the distance it travels within the controlled airspace.

11. To determine the charge per service unit, CAA estimates the total number of ‘service units’ that will be used each year, using the traffic forecast to give ‘total service units’ or TSUs.

12. However, not all service units are liable to charges under the price control. NERL manages airspace use by military aircraft under a priced contract. Some airspace use is exempt from charges, and funding for managing this is provided by government. Using forecast estimates of these types of use, CAA arrives at an estimated forecast of ‘chargeable service units’ or CSUs.

13. Dividing the total determined costs by the forecast CSUs gives a ‘determined unit cost’ or DUC. This is the basis of the charge per service unit that NERL charges in each year.

14. Under the EC performance scheme, charges per service unit are compared across national ANSPs on the assumption that all TSUs are chargeable. In reporting its price control decisions, the CAA makes an upward adjustment to the total determined costs/revenue requirement, so that the reported charge on a TSU basis is the same as the charge on a CSU basis set for NERL.

15. The revenue allowance decision produces a basic charge per CSU for each year of the PCP. The actual charge in each year is an adjusted version of this charge.

Adjustments to Oceanic charges

16. Oceanic charges are levied per flight in either the Atlantic sector, or the Tango sector.[1] The maximum average charge for either sector is based on a nominal

[1] The Oceanic service is sometimes used to provide ATS to traffic flying around French airspace where there is congestion or other operational reasons. This area is referred to as Tango. As these flights only use a small part

value, set in the licence. The maximum average charge can be adjusted for two factors:

(a) Variations in inflation, compared with the levels assumed at the time the price control was set;

(b) A fixed addition to each of the maximum average Tango charge, and the maximum average Atlantic charge, for periods in which the satellite ADS-B service is available.

17. The core price control for the Oceanic service does not have a Traffic risk-sharing mechanism. NERL will gain or lose all of the revenue changes resulting from traffic being higher or lower than forecast, in respect of the core charge. In respect of the ADS-B data charge for Tango routes service, there is an element of traffic risk sharing as NERL has agreed a fixed charge with Aireon for ADS-B data for Tango flights in RP3. If actual flights are greater than forecast, the data charge will be less per flight, and if actual flights are lower than forecast, the data charge will be greater per flight. For the Atlantic service the data charge is fixed (in real terms) per flight.

Adjustments to en route and London Approach unit charges, including incentives and the Traffic Risk-sharing Mechanism (Traffic RSM)

18. Each year, amounts are added to or subtracted from the total determined costs before the charge is calculated. There are various adjustments:

(a) To account for any variation in inflation, compared with the inflation that was assumed when the determined costs for that year was set (INF_t in the formula below, and $LINF_t$ in the London Approach formula);

(b) To account for specified restructuring costs (ReS_t in the formula below. $LReS_t$ in the London Approach formula – both these amounts are £0 throughout the RP3 period);

(c) To take account of any cost changes, for costs that are exempt from being solely at NERL's risk, and that were unforeseen and out of NERL's control. For example, this includes costs arising from changes to relevant laws or international agreements, and certain changes to pensions costs. Most cost changes are recovered in the following PCP or PCPs (CSM_t and $LCSM_t$);

of Oceanic airspace in the South East corner of Shanwick airspace, they are subject to separate ADS-B charges that are much lower than those for the North Atlantic.

(d) To put right over- or under-recoveries in previous years, including a 'clawback' of charge income that was ring-fenced for particular items, and underspent (MOD_t and $LMOD_t$, $Tvar_t$ and $LTvar_t$, FAS_t);

(e) To ensure costs that have been funded from sources other than regulated charges are taken out of the charge calculation ($INEA_t$,VFR_t and $LVFR_t$);

(f) To give effect to performance incentives in the licence (FI_t and LFI_t); and

(g) To implement the Traffic Risk-sharing mechanism (Traffic RSM), for the en route and London approach services, but not the Oceanic service. (TRS_t and $LTRS_t$).

19. Full definitions of all these terms are given in the CAA's draft licence modifications, Appendix H to the CAA RP3 Decision.[2]

20. The CMA's decisions on service quality targets, and on traffic forecasts, will have financial implications for NERL through the last two adjustments above, for performance incentives and traffic risk-sharing.

21. The formula below highlights the incentives, and Traffic RSM, in red. The charge per unit is given in terms of TSUs, but it is equal to the charge per chargeable service unit, as explained above (see paragraphs 10 to 15).

Figure B-2: Charging formula in the price control

$$Maximum\ Charge_t = \frac{DC_t + INF_t + ReS_t + TRS_t + CSM_t + FI_t + MOD_t + Tvar_t - VFR_t - INEA_t - FAS_t}{ForecastTSU_t} - DISCOUNT$$

Source: Adapted from CAA Decision, CAP1830a, Appendix H

22. The 'discount' is a figure that may be set by NERL at its discretion. It is not determined in the price control decision, nor in the CMA investigation.

Incentives

23. If NERL earns incentive payments, or is penalised, under the service targets and incentives schemes, then its charges are adjusted, for the year 2 years after the year in which the payment or penalty was incurred. The charge per CSU is therefore changed upwards or downwards, so that customers pay

[2] CAA, CAP1830a, RP3 Decision Appendices

more for service that is above the target, and less for service that is below target.

24. The capex incentives in the CAA Price Control Decision, subject to any modifications following this review, will take effect through adjustments to the RAB, at the end of the PCP. They will not affect charges during this PCP.

The Traffic RSM

25. In any given year, the actual CSUs will be above or below the forecast CSUs. NERL will therefore earn more or less than its determined costs.

26. This creates a risk for customers, and a risk for NERL:

(a) For customers, if traffic is higher than the forecast, they are collectively paying NERL more than it costs to provide the service;

(b) For NERL, if traffic is lower than forecast, then income from charges will not cover its costs.

27. The price control for en route and London approach services contain a 'Traffic Risk-sharing Mechanism' (Traffic RSM), designed to mitigate these risks to both customers and NERL. This works through adjustments to charges, similar to the adjustment for service quality incentives.

28. Charges are adjusted, according to a formula in the licence, for the year two years after the traffic variation from the forecast occurred. This adjustment leads to higher unit charges (if DC is adjusted upwards), or lower unit charges (if it is adjusted downwards), in the year that the adjustment applies.

29. The EU performance regulation sets a 'default' Traffic RSM. National supervisory authorities (NSAs) may set a different Traffic RSM, provided that this different Traffic RSM does not give more revenue protection to the ANSP than the default Traffic RSM. The CAA proposed a Traffic RSM that is the same as the default Traffic RSM for this PCP.

30. Under the Traffic RSM:

(a) NERL bears all traffic risk, when traffic varies within ±2% of the forecast used for RP3. That is, if traffic is lower by up to 2%, NERL has to absorb the shortfall in revenues; if traffic is higher by up to 2%, customers continue to pay the same level of charge and NERL benefits from higher revenues. The ±2% is called the 'deadband'. There are no adjustments to charges.

(b) NERL bears 30% (the 'risk sharing rate') of the incremental cost, and receives 30% of the incremental increase in revenues, when traffic varies between ±2% and ±10% of the forecast. Customers benefit from, or pay to make good, 70% of the revenue effect of this traffic variation. This is given effect by adjusting the per-unit charges upwards if traffic was lower, and downwards if traffic was higher, so that the final effect on NERL is 30% of what it would have been if charges had remained the same. This band is 8% wide (10%-2%), so the maximum effect of variations in this band is 30% of ±8%, which is ±2.4%.

(c) If traffic varies by more than 10% from the forecast, in either direction, the charge per unit is adjusted in the opposite direction so that customers' total costs, and NERL's income, are held at the levels they would be with a variation of ±10%.

31. This means that in any year, the amount of revenue benefit, or shortfall, is capped at ±4.4% (±2% plus ±2.4%).

32. As noted above, any changes to charges come into effect 2 years after the year when the variation occurred. Any over- or under-recovery that occurred in the two years before this change is taken into account by the inclusion of $Tvar_t$ in the formula.

Appendix C: Timeline of RP3 process[3]

1. April 2017: CAA published discussion document on desired main objectives for RP3

   CAA RP3 discussion document

2. September 2017: CAA launched consultation on its business plan guidance for NERL

   CAA business plan guidance consultation

3. January 2018: CAA published Guidance for NERL in preparing business plan for RP3

   CAA business plan guidance

4. April 2018 NERL submitted its Initial Business Plan (IBP) to the CAA

5. May 2018 CAA sent letter to NERL indicating its view that IBP had ‘fallen short of expectations as put forward in the business plan guidance’

   CAA letter to NERL

6. May-September 2018 Customer Consultation Working Group, culminating in a report by the Group’s Co-chairs (Co-chairs Report)

   CCWG Co-chairs Report

7. October 2018 NERL published Revised Business Plan (RBP)

   NERL Revised Business Plan

8. February 2019 CAA published Draft Decision, followed by consultation period

   CAA RP3 Draft Decision

9. 29 August 2019: CAA published its Final Decision on RP3 (CAA Decision)

   CAA CAP 1830 UK RP3 CAA Decision document

10. 10 September 2019: NATS rejected the CAA RP3 Decision

    NATS, letter from Martin Rolfe to Richard Moriarty, 10 September 2019

---

[3] CAA Reference, paragraph 1.38 and B50

11. 19 November 2019: CAA makes the Reference to the CMA

CAA Notice of Reference

CAA CAP 1857 Reference to the CMA of the NERL RP3 price controls (CAA Reference)

12. 25 February 2020: CAA sends Notice of Variation to the CMA

CAA Notice of Variation

13. 6 May 2020: CAA grants extension of the statutory deadline to 17 November 2020

CAA notice of extension to the statutory deadline

Appendix D: Technical note on betas and gearing

1. This appendix provides a technical assessment of the reasons why the beta for NERL varies with the assumption on gearing, and provides context to the decision on the level of gearing to be used in setting the cost of capital.

Purpose

2. In the NATS case, we have identified a particular concern which arises as a result of the CAA's decision relating to gearing. The CAA and NATS have assumed a notional 60% gearing in setting the cost of capital, although the comparator companies have much lower gearing. The highest comparator is around 40% geared (Fraport), the direct comparator (ENAV) has 0% gearing, and the majority of other comparators have gearing around 20%.

3. The CAA and NATS have then used a standard approach used by regulators to 'de-gear' and 're-gear' the betas of comparator firms, on the assumption of an asset beta which is constant with gearing.

4. We have some concerns with the consequences of the standard regulatory approach to 're-gearing' in this case. Our concerns start from the analytical finding that the cost of capital increases by around 0.5% as a result of the assumed higher gearing of NERL (60%) relative to gearing assumption based on the gearing of comparators (30%), which is not consistent with either finance theory or with our understanding of how actual financing models work. In this calculation we have adjusted for the changing proportion of embedded and new debt.

5. Our understanding is that other regulators currently apply the same approach of using an asset beta and then re-gearing using a formula comparable to that used by the CAA. We have found one reference to the problem identified in this case, which is in the CC's review of the price control for airports in 2007.[4] At that time the CC's response was to use a debt beta which was at that time a divergence from regulatory precedent, and which was sufficient to address the concerns that the WACC would otherwise increase with gearing. In this appendix, we consider the potential options for addressing the similar concerns in this case.

[4] CC Heathrow Airport Ltd and Gatwick Airport Ltd price control review, paragraphs 83 to 90

Background

6. According to standard finance theory, the cost of capital (WACC) does not, at least in a theoretical model, vary with gearing, other than for tax reasons. This finding was specified by Modigliani and Miller (MM) in their seminal paper from 1958, 'The Cost of Capital, Corporate Finance, and the Theory of Investment', and is still referred to in corporate finance texts and papers today.[5]

7. In MM's paper, WACC is found to be independent of gearing for reasons which are not linked to the model for the cost of equity. In particular, MM do not make the assumption that the Capital Asset Pricing Model (CAPM) is used for the assessment of the cost of equity. The CAPM was developed by Sharpe and others independently of Modigliani and Miller's analysis.[6]

8. In very simple terms, the MM theory holds on the assumption that individual investors can borrow at the same rate as the firms, and therefore if the cost of capital were not constant, then investors would buy/sell the shares and adjust their own leverage, thus adjusting the share price to the level where MM holds. MM's paper recognises that their finding that the WACC is independent of gearing is based on some simplifying assumptions, and that work will need to follow to understand how frictions and imperfect markets will drive actual behaviour.

9. MM concluded that their findings should however challenge assumptions that would imply that there was a systematic and persistent difference between actual markets and the model assumed by MM.[7] The standard theory then, and, arguably, now, is that actual financial markets operate in a 'U shape' where the cost of capital falls with gearing up to an optimal level, and then starts to rise above that optimal gearing. MM described this as:

> Although the falling, or at least U-shaped, cost-of-capital function is in one form or another the dominant view in the literature, the ultimate rationale of that view is by no means clear.

The regulatory model

10. The model used by the regulators is based on the following assumptions:

---

[5] Modigliani-Miller
[6] There is a summary of the development of the CAPM in Fama-French's 1996 paper
[7] Modigliani-Miller, page 22

- Embedded debt is recovered at its actual cost (potentially subject to efficiency adjustments). This is relevant to the extent that this implies that the cost of debt will change to reflect the level of gearing.

- The cost of new debt can be estimated based on observable market data for the cost of debt.

- The cost of equity can be modelled based on the CAPM, based on first measuring an asset beta based on comparators, and second, assuming that the asset beta is independent of gearing

11. In the NATS case, these assumptions are not consistent with MM, as the cost of capital calculated using this model increases with gearing.

12. This is because the observable cost of debt, used to calculate the cost of new debt, breaches an assumption underlying the CAPM, which is that the investors' required levels of financing costs of debt and equity are explainable entirely by regressing a single factor – the total market return. For this 'one-factor' model to work, the following equation would need to hold for the cost of debt:

$$R_D = R_{rf} + \beta_D\left(R_m - R_{rf}\right) + ELP$$ [8]

13. In respect of the current NATS cost of debt, this equation does not hold in practice. Europe Economics estimated the debt beta on this basis, and the implied debt beta was concluded by the CAA, and also by us, to be implausibly high.[9] There are a number of reasons why this equation may not hold. The 'liquidity premium' referred to by the CC in 2007 may have increased. Alternatively, the debt premium may have increased to offset the ultra-low returns on government bonds by comparison to historical averages. In either case, the additional risk premium appears to be a systematic or market risk factor, in that a comparable risk premium applies across the corporate bond markets, and is not therefore diversifiable by corporate bond investors.

14. The result of this difference between the implied CAPM cost of debt and the actual cost of debt is that the cost of capital calculated by the regulators' model increases with gearing (g). The increase is equal to the difference between the actual (observable) cost of new debt, and the CAPM cost of debt (R_D) implied by the equation above.

[8] Where R_D is cost of debt, R_{rf} is risk free rate, β_D is debt beta, R_m is total market return and ELP is expected loss premium.

[9] CAA Decision Appendices, Appendix E, paragraph E138

15. In the UKRN paper, the authors also noted this effect, describing the regulatory model as the CAPM(E)-WACC, by comparison to a model based purely on CAPM parameters, which they characterised as a CAPM-WACC. The authors highlighted the difference could go either way.[10]

16. In practice, the difference between the actual cost of new debt and the implied cost of new debt from the CAPM is currently material and positive. This difference, multiplied by the change in gearing, directly translates to the increase in WACC with gearing.

An alternative model consistent with MM

17. The Modigliani-Miller model assumes that cost of capital is independent of gearing. For this to be true, as gearing increases, the following needs to hold:

- $Increase\ in\ total\ R_D = Reduction\ in\ total\ R_E$[11]
- $Total\ interest\ cost\ at\ higher\ g - Total\ interest\ cost\ at\ lower\ g = Total\ R_E\ at\ lower\ g - \ Total\ R_E\ at\ higher\ g$

18. Assuming that the cost of debt is approximately constant (or the change is measurable from market data) for a small increase in gearing from g to g*, this means:

$$R_E \times g * - R_E \times g = (g * - g) \times R_D$$

19. If we assume that this equation holds, ie that the WACC is independent of gearing, and if we can estimate R_D from market data, then we only need to know the cost of equity at any one level of gearing, to be able to work out the cost of equity at any other level of gearing.

20. Therefore, if we can use the CAPM to estimate the cost of equity at one level of gearing, for example, the gearing of the comparator firms, we can imply the cost of equity for all other levels of gearing.

Illustration: scenarios for the cost of equity

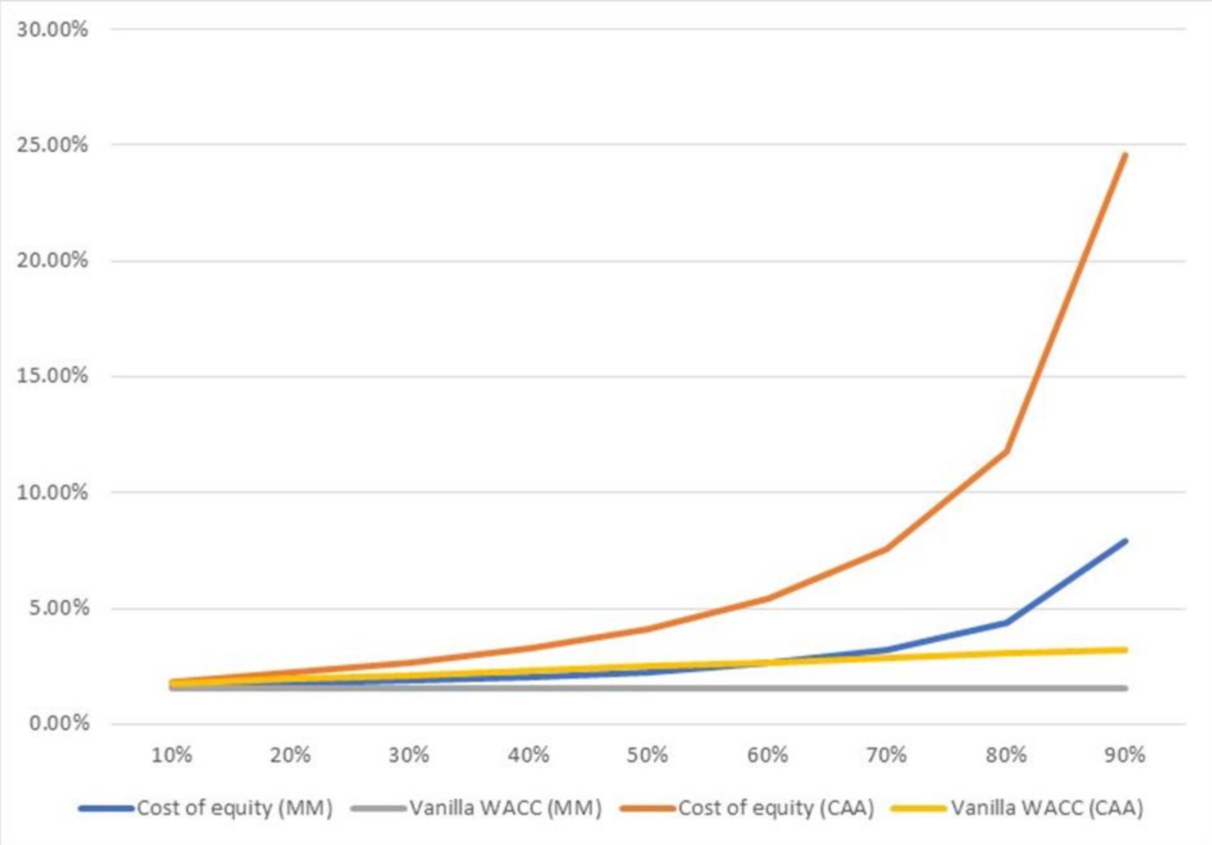
21. We illustrate below the consequence for the implied cost of equity for NATS at different levels of gearing of using the approach described above of a WACC which is constant with gearing. The graph below demonstrates the effect of this approach on the pace at which the cost of equity increases for NATS, if

---

[10] UKRN report, page 23-24
[11] Where R_D is cost of debt and $E[R_i]$ is cost of equity.

the starting point for both is the CAA's asset beta, based on a comparator with 0% gearing (ENAV).

Figure D-1: Cost of equity and Vanilla WACC at different gearing levels, under the CAA's model and an alternative MM model which assumes constant WACC with gearing.



Source: CAA RP3 Decision and CMA analysis.

22. Figure D-1 illustrates that the cost of equity can increase with gearing and also that the cost of capital can be consistent with MM, but that implies that asset betas change with gearing. Given the current level of the cost of new debt, the consequential effect is in practice that asset betas would reduce with gearing. This is consistent with the observations of Wright and co-authors in the UKRN paper referred to above:[12]

> [3.1] Note that in general the asset beta (β_A) will be a function of the leverage of the company g. We typically assume that β_A is a U-shaped function of g, reaching a minimum at the optimal level of gearing, g*. The CAPM in turn implies that the same property must hold for the WACC.

[12] UKRN report, page 23-24

23. The MM scenario including a constant WACC seems more consistent than CAA's model with the observed practice that the WACC is constant or reduces as gearing increases towards an optimal gearing, which most regulators have assumed to be around 60% when setting the cost of capital.

24. Figure D-1 illustrates that if the WACC is to be constant with gearing, then the asset beta and total market return are not sufficient to explain the cost of equity, as the equity beta increases at a slower rate than under a model where WACC is constant, the consequence of which is that asset beta varies (and in fact, falls) with gearing.

25. As described above in the extract from the UKRN report, one way of looking at this is that the asset beta, whilst it in theory is based only on the risk of the assets, is in practice not constant as the way those assets are financed (eg. in the choice of the level of gearing) changes. Another way of describing the effect is that it is consistent with academic practice which considers models for returns on capital which go beyond the CAPM, normally called 'multi-factor models'.

26. Our understanding is that multi-factor models have been rejected for use by regulators not because they are wrong – academic evidence suggests they are better in explaining actual returns to investors. They have been rejected because they are hard to populate in practice. The most popular multi-factor model, the Fama-French model, has been considered by some regulators for use but it was concluded that it was not feasible to populate the model.[13]

Additional note: the concept of 'optimal gearing'

27. In practice, we understand the WACC is likely to not to be constant, but to reduce as gearing increases towards 'optimal gearing'. A model where the WACC falls with increasing gearing would imply that the increase in the cost of equity as gearing rises is even more shallow than implied by the MM model, which assumes that WACC is constant with gearing. In other words, the rate increase in the cost of equity between zero gearing and 'optimal gearing' is even more distinct from the assumption in the regulatory model, ie a model which has an increasing WACC.

28. Our analysis of the literature, including the UKRN report, suggests that there is no theoretical problem with using a model which implies the WACC falls with higher gearing. Such a model implies that, as observed by financial market observers, the low cost of debt means that the most companies are

[13] Also discussed in the Fama-French 2004 review

more valuable if they gear up as far as debt markets will allow at the current low cost of new debt.

29. For the purpose of setting a regulatory determination, a possible implication is that when estimating the cost of capital, we should recognise that the difference between the cost of capital at notional gearing and ‘optimal gearing’ represents another source of uncertainty. A model based on measuring the cost of capital at one level of gearing may overstate the cost of capital at ‘optimal gearing’. However, the scale of the difference between the cost of capital at notional gearing and ‘optimal gearing’ is not directly measurable using a standard regulatory model that assumes a constant asset beta.

Appendix E: Estimating historical returns

Inflation measures

1. There are four main inflation indices that have been used to deflate historic equity returns. These are the retail price index (RPI), the consumer price index (CPI), the Cost of Living Index (COLI) and the Consumption Expenditure Deflator (CED).[14] We provide a brief description of each inflation measure, in terms of its availability, ie the time period for which it is available, its robustness as a measure of inflation, and its consistency over time.

RPI

Availability

2. RPI was first calculated in 1947 and was the headline measure of inflation in the UK until 2013. Although it was stripped of its National Statistics status in March 2013, RPI is still collected and published by the ONS due to its use in indexing a broad range of prices, including gilts, pensions, student loans etc.

Robustness

3. RPI is not a robust measure of inflation. As the ONS explains:

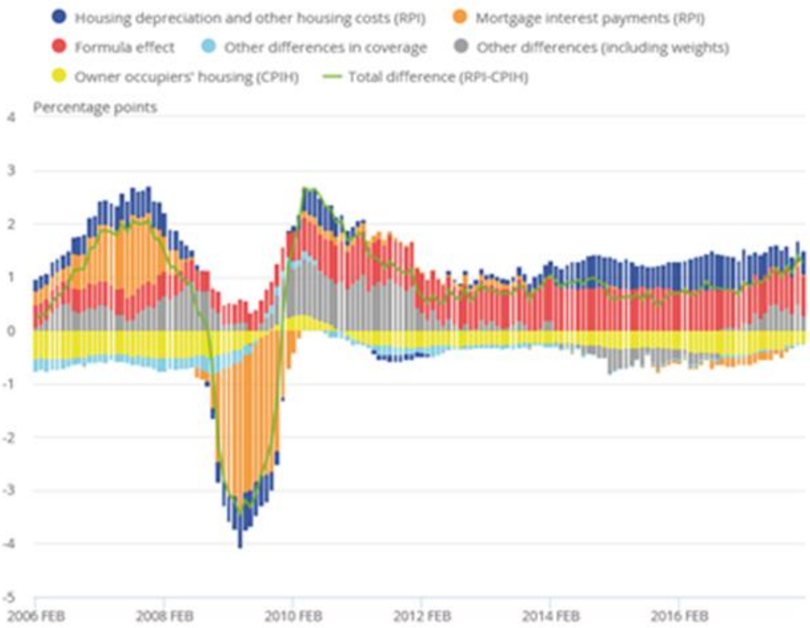
> Overall, RPI is a very poor measure of general inflation, at times greatly overestimating and at other times underestimating changes in prices and how these changes are experienced.
>
> In 2013, the RPI lost its status as a National Statistic. Our position on the RPI is clear: we do not think it is a good measure of inflation and discourage its use. There are other, better measures available and any use of RPI over these far superior alternatives should be closely scrutinised.[15]

4. Overall, RPI tends to overestimate inflation due to its use of the Carli formula (giving rise to 'the formula effect'), which is an arithmetic average of price relatives, rather than the Jevons formula, which gives a geometric average price change. However, there are also issues with the index due to the data

[14] In addition, there is a dataset that is sometimes used to cover the period 1900 to 1914, estimated by Feinstein (1991). However, given the relatively short period covered by this inflation series, and therefore the minor impact using it would have on our estimates of TMR, we do not propose to consider it in detail.
[15] ONS, Shortcomings of the retail prices index as a measure of inflation

source of the weights (coming from the Living Costs and Food Survey only), population coverage (excluding the highest-earning 4% of households, as well as pensioner households that derive at least 75% of their income from state benefits, institutional households and foreign visitors to the UK) and treatment of some goods, such as owner occupiers housing.[16] Figure 1, below, shows how these issues have contributed to a ‘wedge’ between RPI and CPIH in recent years.

Figure E-1: Causes of the difference between the RPI and CPIH inflation rates, 2006 to 2018



Source: Office for National Statistics

5. In 2012, the National Statistician consulted on changing the RPI to address some of its flaws. However, the consultation[17] concluded ‘there is significant value to users in maintaining the continuity of the existing RPI’s long time series without major change, so that it may continue to be used for long-term indexation and for index-linked gilts and bonds in accordance with user

[16] For example, the Johnson Review states that:

> As we stressed above it is generally hard in this area to come to absolute conclusions. But it is our strong view that the use of the Carli is inappropriate and that the RPI is upwardly biased because of its use. In light of this, ONS has introduced an additional inflation measure – RPIJ – which is essentially the same as the RPI except that it uses the Jevons method wherever the RPI uses the Carli… But it is not just the use of the Carli which is problematic in the construction of the RPI as a measure of consumer price inflation. Issues with the data source of the weights, population coverage and treatment of some goods (like insurance and owner occupiers housing costs) make the RPI less suitable as a measure of overall inflation. RPIJ is problematic for all the same reasons.

Similarly, Wright et al (2018) note that “[T]he elementary price aggregation methods in RPI create significant and unstable biases between recorded inflation and what it is attempting to measure.”, pg D-109.

[17] ONS review of RPI

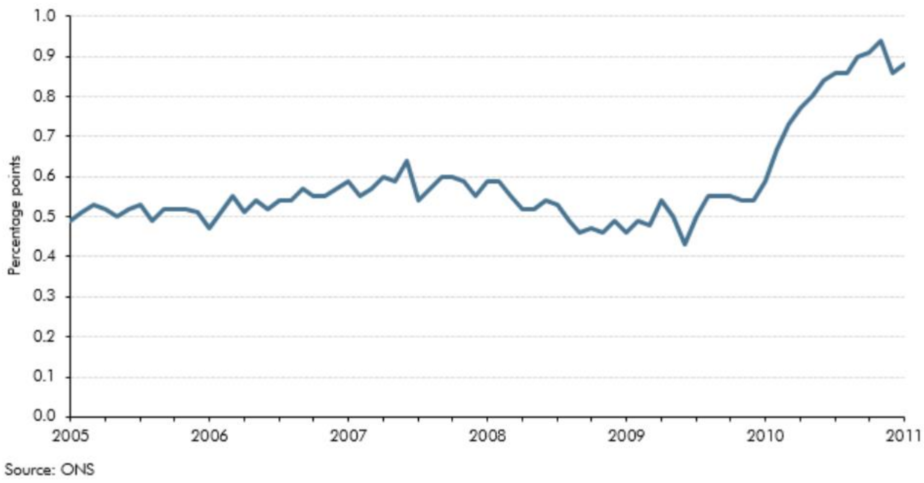
expectations'. This reflected the fact that some users valued the continuity of the index, despite its flaws.

6. While the ONS has committed to continue producing RPI as currently defined, any future changes to the index will be limited to issues such as the annual update of the basket and weights, improvements to data validation and quality assurance etc.[18]

Consistency

7. RPI is not a consistent index insofar as changes to the underlying methodology used to calculate the RPI mean that it is not comparable over time. The clearest example of this was the significant increase in the formula effect in 2010 as a result of a change to the way that clothing prices were collected. This increase in the formula effect, from around 0.5 percentage points to 0.8-0.9 percentage points, is shown in the chart below.

Figure E-2: Contribution of the difference between RPI and CPI from the formula effect



Source: OBR website

8. Oxera highlights 5 (further) key changes in the RPI methodology since 1947:

(a) In 1956 the RPI experienced a range of important methodological improvements, in particular, all wage-earning households were included—not only the working class, the index took its weights from the more recent 1953 expenditure survey, rather than the pre-war late-1930s survey, and owner-occupier housing costs were included for the first time.

(b) From 1962, expenditure weights were updated on an annual basis.

---

[18] Johnson Review, page 53.

(c) In 1968, prices of food and drink purchased in restaurants were introduced.

(d) In 1975, mortgage interest payments were introduced to represent owner-occupiers' housing costs.[19]

(e) In 1986 it was decided to exclude the top 4% of households, based on their household income (before this, households earning more than a certain amount were excluded). In the following years, holidays started being included as well.[20]

CPI

Availability

9. CPI was first published in 1996 and it replaced RPI as the headline measure of inflation from 2013. However, growth rates for the CPI and its main component indices are available for the period from 1989 until the present day.[21] The ONS has sought to estimate CPI from 1949-1987, otherwise known as the 'back-cast' by using the historic RPI series and an estimate for the RPI-CPI 'wedge'. See paragraphs 18 to 20 for further details of how this 'backcast' has been carried out.

Robustness & consistency

10. CPI is prepared according to international best practice, predominantly using the Jevons formula for unweighted averages (ie based on geometric average price changes), covering all private and institutional households and drawing on a full range of sources from the National Accounts. However, CPI does not include any measure of owner occupier housing costs, which are an important element of household budgets. For this reason, the ONS intends to replace CPI with CPIH as its headline measure of inflation in the UK.[22]

11. The CPI 'backcast' has been estimated by creating an 'RPI adjusted' (RPIA) series which uses RPI data and applies the RPI calculation formula but mimics the CPI coverage and weighting.

12. It then uses RPIA to compute the formula effect (RPIA – CPI) for the period with overlapping data for the RPI and CPI series (1989 to 2011). This allows

---

[19] This was prompted by concerns that equivalent rents did not measure housing costs for owner-occupiers well, especially given recent rises in interest rates and the growth of owner occupation (meaning that more than half of all households fell into this category)
[20] Estimating RPI-adjusted equity market returns, Oxera, 2 August 2019
[21] Modelling a back-series for the consumer price index, ONS
[22] Users and uses of consumer price inflation statistics

the researchers to identify the 'pure' formula effect (ie excluding differences arising due to differences in coverage and weighting). The researchers use the computed formula effect for the overlapping period, together with other economic information (including on RPI, GDP etc), to estimate a time series (ARIMA) model, which is then used for out-of-sample prediction, ie to backcast the formula effect to 1949. Finally, they derive CPI estimates using the backcast formula effects and the adjusted RPI series.

13. The approach to backcasting CPI ensures consistency across the CPI figures over the 1949 to 2020 period. However, the researchers who carried out the backcast highlighted that

14. [t]he method provides only approximate results and there is no way to determine how accurate our method is as sufficient data to calculate the CPI do not exist prior to 1987. The modelled estimates described in this paper provide an estimate of a consistent series for the primary inflation measure used in the UK over a period for which no such measure was previously available. Because of the assumptions made in their construction, these estimates are not National Statistics.[23]

Consumption Expenditure Deflator

15. This is the implied deflator for consumers' expenditure derived from estimates of consumers' expenditure valued at current and constant prices taken from the unofficial national accounts of the United Kingdom, prepared by the Department of Applied Economics at Cambridge University (source: C H Feinstein, National Income, Expenditure and Output of the United Kingdom 1855-1965, 1972, tables 24 and 25).

Availability

16. It is available for the period between 1870 and 1947.

Robustness & consistency

17. The CED is not a price index but rather an implied deflator. Feinstein notes that the margin of error in terms of consumption levels is likely to be less than 5% for the period from 1914 onwards, and between 5% and 15% for the period from 1890 to 1913. However, O'Donoghue et al note that the implied

[23] Modelling a back-series for the consumer price index, ONS

deflator is likely to be subject to a smaller margin of error than the underlying expenditure data as it is based on relative rather than absolute levels.[24]

18. Due to its method of calculation the CED is a Paasche index rather than a Laspeyres, ie it uses current-period quantity weightings while the latter uses base-period quantity weightings. This means that the index takes into consideration (changes in) consumption patterns within period. As a result, it will tend to understate the changes in price because the index already reflects changes in consumption patterns when consumers respond to price changes and adopt substitutes.

19. The CED has been estimated on a consistent basis over the 1870 to 1947 period.

Cost of Living index[25]

Availability

20. This was first prepared by the UK Government in 1914 and continued to be produced until 1947.

Robustness & consistency

21. The COLI was designed to measure the costs faced by working class households with the aim of helping to protect workers from the economic consequences of the First World War. To that end, the COLI sought to monitor changes in prices of the main items purchased by working-class households. These changes were weighted according to spending on the different items to produce an overall index.

22. The inflation basket included food and drink, rent and rates, clothing, fuel and lighting. The weightings for the items in the basket were set in 1914 and never updated. The excerpt below, taken from the Interim Report of the Cost of Living Advisory Committee outlines the categories and weights assigned in 1914.[26]

---

[24] O'Donoghue et al, 'Consumer price inflation since 1750', ONS Economic Trends, March 2004
[25] https://publications.parliament.uk/pa/ld201719/ldselect/ldeconaf/246/24604.htm, paragraphs 23 to 27
[26] Ministry of Labour and National Service, Interim Report of the Cost of Living Advisory Committee, Cmd 7077, March 1947

Figure E-3: Comparison of components of Cost of Living Index, 1914 and 1937-38

COMPARISON OF WEIGHTS USED IN COMPILING THE EXISTING COST-OF-LIVING INDEX AND THOSE SHOWN BY THR 1937-38 BUDGETS

Groups of Items	*Cost-of-Living Index Weights based on 1914 Expenditure*	*Weights based on Expenditure in the 1937–38 Budgets*
Food	60(1)	40·1
Rent, Rites, &c.	16	12·7(2)
Clothing	12	9·5
Fuel and Light	8	7·6(3)
Other items—		
In Index	4.	8·1
Not in Index	...	22·0
	100	100·0

(1) Expenditure on food was based on an inquiry made by the Board of Trade in 1904. In the returns obtained in that in uiry food accounted for 22*s*. 6*d*. in a total of 36*s*. 10*d*. In 1914 it was estimate8 that food expenditure had risen to 23*s*. 6*d*. in a total of 42s.

(2) This weight for rent, rates, &c., is based on total expenditure on housing, including payments by households purchasing their dwellings and the (relatively small) payments by households who had already completed the purchase of their dwellings.

(3) Including electricity, coke and firewood, which are not covered by the official index.

Source: Interim Report of the Cost of Living Advisory Committee, March 1947, Appendix I

23. Due to the design of the index, the range of products included was limited, as was its coverage of the population.[27] In addition, the weights were not changed over time despite spending patterns changing, such that they became increasingly out of date. The index also made judgements as to how working-class households should spend their money; beer was excluded and the weight for tobacco did not reflect how much was actually spent on it.[28]

24. The COLI was prepared on a consistent basis over the 1914 to 1947 period.

Datasets

25. There are two main datasets of recognised historical inflation series that have been used by academics, regulators and other parties to evaluate TMR. These are the Credit Suisse Yearbook, prepared by Dimson, Marsh and Staunton ('DMS' dataset) and the Bank of England Millennium dataset. Both use a combination of the above inflation measures. We provide a brief description of each below.

---

[27] O'Donoghue et al, 'Consumer price inflation since 1750', ONS Economic Trends, March 2004

[28] See Johnson Review, page 46.

Dimson Marsh and Staunton (DMS)

26. The Credit Suisse Global Investment Returns Yearbook produced by Dimson, Marsh and Staunton (DMS) of London Business School, is an annual study of global historic investment returns.

27. The underlying inflation indices used to compile the DMS published long run inflation data series have changed over the recent past. This has a significant impact on the resulting estimated historic "real" returns calculated.

28. In 2016, DMS introduced CPI into their inflation dataset for the period 1988 onwards, motivated by continuing concern about the upward bias in the RPI.[29] A further change in the latest 2019 Yearbook[30] saw a move to using CPI 'backcast' data for the period from 1949. The composition of the inflation series is summarised below.

    - Pre 2016 Yearbook: Cost of Living Index from 1900 to 1948 and RPI data from 1949 onwards.

    - 2016 Yearbook: Cost of Living Index from 1900 to 1948 and RPI data from 1949 to 1987 and CPI from 1988 onwards.

    - 2019 Yearbook: Cost of Living Index from 1900 to 1948, ONS's 'back cast' measure of CPI since 1949, and CPI from 1988 onwards.

29. As a result, when the use of DMS data, whether real returns or historical inflation series, has been cited, it is important to understand from which Yearbook the data has been taken as this would significantly impact the results.

Bank of England Millennium Dataset

30. A group of Bank of England staff alongside academics have sought to collate a body of evidence to create a long run series of historic inflation: The Millennium dataset. This dataset comes with the caveat that it is not an official Bank of England data source and that it has been compiled for use by students and researchers.[31]

31. The dataset itself is not intended to be static, with data being added as necessary and errors corrected, therefore it is advised that the dataset should be viewed as 'work in progress'.

---

[29] Credit Suisse Global Investment Returns Yearbook 2017, February 2017, page 212
[30] Credit Suisse Global Investment Returns Yearbook 2019, February 2019, page 212
[31] A millennium of macroeconomic data for the UK, Bank of England

32. There are the following versions of the inflation series:

- CPI 'original' method (version 1): Composite price index series from a paper by O' Donoghue et al (2004) which uses ONS's Consumption Expenditure Deflator (CED) up to 1949. ONS's back-cast measure of CPI since 1949, and CPI from 1988 onwards.

- CPI 'preferred' measure: This is identical to the 'original' series apart from the period 1900-1914, which instead of the O'Donoghue et al. (2004) series is based on a series from a paper by Feinstein (1991).[32]

- RPI measure - uses the same composite price index series from O' Donoghue et al (2004)[33] up to 1949 and published RPI data from 1949.

Averaging techniques and holding periods

33. Under the assumptions that expected returns are constant over time, and that returns in each period are independent of each other, the arithmetic average of realised returns is an unbiased measure of the constant expected return. A simple approach to measuring historical returns is therefore to calculate an arithmetic average of historical returns.

34. The length of the period over which the return to be averaged is measured is a complex issue. The relevant period would seem to be the period for which investors expect to be invested in the market (we describe this as the holding period). It seems very unlikely that this is as short as one year. Because of their price variability, equities are usually regarded as a long-term investment.

35. Blume has shown that, if the holding period is longer than one year, the arithmetic mean of one-year returns is an upwards-biased measure of the true expected return (assuming that returns are independently and identically distributed around the true expected return).[34] Blume suggested a number of unbiased measures if the holding period is longer than one year. Assuming a holding period of h years, expressed as equivalent annual returns, these included:

[32] C. H. Feinstein 'A new look at the cost of living', in Foreman-Peck J. ed. 'New perspectives on the late Victorian Economy', Cambridge University Press, 1991
[33] O'Donoghue et al, 'Consumer price inflation since 1750', ONS Economic Trends, March 2004
[34] Blume, M, 'Unbiased estimators of long-run expected rates of return', Journal of the American Statistical Association, 1979.

(a) We describe this as the 'simple' estimator of the average return for a holding period of h years.[35] The DMS data covers 120 years and if we wish to use all of this data we are limited to values of h which are factors of 120. However, the number of non-overlapping observations drops off rapidly as the holding period increases — there are only 12 observations for a holding period of ten years, 6 for a holding period of 20 years and two for a holding period of 60 years.

(b) The arithmetic mean of returns for all overlapping periods of h years.[36] This greatly increases the number of observations (the data gives 111 such observations for a ten-year holding period): intuitively, we might expect accuracy to be increased by extending the observations even though these observations are not independent of each other, but Blume's simulations tended to suggest that the overlapping mean tends to be a less efficient estimator than the non-overlapping mean.

(c) A weighted average of the arithmetic and geometric means[37] where the weight on the arithmetic mean is $\frac{(120-h)}{(t-1)}$ and the weight on the geometric mean $\frac{(h-1)}{(t-1)}$ where t is the length of time for which we have data. We describe this as the Blume estimator. For a holding period of one year, this is the arithmetic mean which, as noted above, is unbiased for a holding period of one year; and for a holding period equal to t (120 years for our data), this is equal to the geometric mean which is an unbiased estimator for this length of holding period (albeit one based on a single observation of the expected return over 120 years).

36. Jacquier, Kane and Marcus (JKM) extended Blume's work under the assumption that returns were lognormally distributed.[38] JKM proposed a general class of estimators of annualized returns taking the form: $e^{(m+0.5vk)}$ where m is the arithmetic mean and v is the variance of annual returns; and k is a parameter depending on h and t. In particular, JKM proposed:

[35] The mean is calculated from the formula $\frac{\Sigma\left(R_t+\frac{h}{R_t}\right)}{\frac{120}{h}}\,\frac{1}{h}$ where h is holding period, R_t is value of returns index at the end of year t and the expression is summed for $\frac{120}{h}$ values of t for which non-overlapping data is available.

[36] The mean is calculated from the formula $\frac{\Sigma\left(R_t+\frac{h}{R_t}\right)}{120-h+1}\,\frac{1}{h}$ where h is holding period, R_t is value of returns index at the end of year t and the expression is summed for values of t for which overlapping data is available.

[37] The geometric mean of annual return indices is equal to the compound annual growth rate in returns over the period.

[38] Jacquier, E, Kane, A and Marcus, A J, 'Optimal estimation of the risk premium for the long run and asset allocation: a case of compounded estimation risk', Journal of Financial Econometrics, 2005.

(a) an unbiased estimator, where $k = 1 - \frac{h}{t}$; and

(b) a further estimator, where $k = 1 - \frac{3h}{t}$. JKM show that this minimizes the difference between the estimate and the true value in small samples (is small sample efficient), even though it is not unbiased.[39] This is useful because our sample of independent observations becomes small as h increases.

Results

37. Table E-1 shows the results of our analysis using the various inflation measures and estimators discussed in this appendix.

Table E- 1: CMA estimates of real returns, 1900 to 2019

		Inflation series		
	Holding period	CED/CPI	CED/RPI	COLI/RPI
Arithmetic mean	1 year	7.0%	6.7%	7.1%
Geometric mean	120 years	5.2%	5.0%	5.4%
Blume (1974)	10 years	6.8%	6.6%	7.0%
	20 years	6.7%	6.4%	6.8%
JKM (2005) unbiased estimator	10 years	6.9%	6.6%	7.0%
	20 years	6.7%	6.5%	6.9%
JKM (MSE)	10 years	6.6%	6.3%	6.7%
	20 years	6.1%	5.9%	6.3%
Overlapping	10 years	6.6%	6.4%	6.7%
	20 years	6.7%	6.4%	6.7%
Non-overlapping	10 years	6.8%	6.5%	6.7%
	20 years	7.2%	6.8%	7.1%

Source: CMA analysis, DMS returns data

[39] Blume assumed that returns were normally distributed, implying that the return index can take a negative value; the lognormal assumption avoids this implication and is more analytically tractable.

Appendix F: Licence modifications

Summary

1. This appendix sets out modifications to the licence that we consider should be made to reflect our determination. We expect that some of our conclusions will be dealt with outside the licence, and these have not been re-presented below.

2. We note that the CAA may also wish to make further modifications to the licence, in line with its statutory duties.

Condition 10

Efficiency incentive

3. Condition 10 of NERL's Licence should be modified to include an enhanced role for the Independent Reviewer, and a requirement for quarterly service and investment plan (SIP) updates, in line with the CAA RP3 Decision.

4. Condition 10 of NERL's Licence should be modified to include reference to:

   (a) a DIWE test as being the basis upon which the CAA would determine whether capex should be disallowed from NERL's regulatory asset base (RAB) following an ex-post efficiency review; and

   (b) a Regulatory Policy Statement setting out how the CAA expects to apply the DIWE test.

5. The Regulatory Policy Statement (RPS) concerning application of the DIWE test, should be published by the CAA alongside the licence modification above.

6. The RPS should be substantially consistent with the draft RPS the CAA submitted as part of its response to our provisional findings,[40] other than where changes to the draft RPS can be shown to be justified in order to address issues raised in the CAA's subsequent consultation that had not already been considered as part of our assessment.

7. The capex delivery incentive proposed in the CAA RP3 Decision should not be introduced.

---

[40] CAA PF response, Appendix B.

8. Condition 10 of NERL's Licence should be modified to include:

(a) Reference to the introduction of a new capex incentive based on the quality of NERL's engagement, and actions in response to engagement;

(b) Reference to a guidance document setting out the process through which, and the basis upon which, the CAA would assess NERL's performance under the new incentive, and determine the level of penalty (if any) to be applied;

(c) Details of how the penalty cap that would apply to the incentive would be calculated. This should provide that the level of the penalty cap will be calculated using an approach that is, and assumptions that are, consistent with that used by the CAA when calculating the £36 million capex delivery incentive penalty cap proposed in the CAA RP3 Decision, other than that the penalty cap would be determined on the basis of NERL's actual RP3 capex rather than on the level of NERL's RP3 capex allowance.

9. The guidance document referred to in 8(b) should be published by the CAA alongside the licence modification and be substantially consistent with the draft guidance the CAA submitted as part of its response to our provisional findings,[41] subject to the conclusions and comments in chapter 9 of this report.

Condition 21

10. Based on our decision, the levels of the en route controls will change, and this will affect the following terms within Condition 21:

(a) Determined Costs (DC_t), expressed in nominal terms, in Condition 21.1.

11. Our revised price control indicates the following changes to Determined Costs:

Table F-1 Changes to Determined costs

Year t	DC_t (CAA decision)	DC_t (CMA determination)
2018	589,585,024	589,585,024
2019	579,006,611	579,006,611
2020	678,457,133	689,955,378
2021	661,997,446	674,270,832
2022	675,862,804	688,739,423

[41] CAA PF response, Appendix C.

2023	643,690,078	n/a
2024	646,694,954	n/a

12. All other numbers in Condition 21 are unaffected by the CMA's determination, other than to the extent that all numbers are now only required for years up to 2022, with the subsequent years being assessed as part of the CAA's forthcoming review.

Condition 22

13. Based on our decision, the levels of the Oceanic controls will change, and this will affect the following terms within Condition 22:

    (a) Unit Charges (U_t), expressed in nominal prices, in Condition 22.1.

    (b) The price charged per Atlantic Flight for the use of the ADS-B service (ADA_t), expressed in nominal prices, also in Condition 22.1.

14. Our revised price control indicates the following changes to Unit Charges:

Table F-2 Oceanic unit charges

Year t	U_t (CAA decision)	U_t (CMA determination)
2020	56.04	56.56
2021	54.74	55.21
2022	54.80	55.26
2023	51.30	n/a
2024	49.88	n/a

15. Our revised price control indicates the following changes to the price charged per flight for the use of the ADS-B service:

Table F-3 Atlantic ADS-B charges

Year t	ADA_t (CAA decision)	ADA_t (CMA determination)
2020	31.64	33.30
2021	32.27	33.97
2022	32.92	34.65
2023	33.57	n/a
2024	34.24	n/a

Table F-4 Tango ADS-B charges

Year t	ADT_t (CAA decision)	ADT_t (CMA determination)
2020	4.90	5.15
2021	4.83	5.08
2022	4.76	5.02
2023	4.56	n/a
2024	4.51	n/a