The Cost of Equity Capital for Regulated Companies:

A Review for Ofgem

Stephen Wright
Birkbeck College

Andrew Smithers
Smithers & Co Ltd

Executive Summary

1. We have been asked by Ofgem to review the methodology for assessing the real equity market return, or (market cost of equity capital) as an input to the calculation of the weighted average cost of capital (WACC) in the context of RIIO price controls.

2. An important impetus for this review has been the view expressed by the Competition Commission (CC), in its provisional determination for Northern Ireleand Electricity, that more weight should be given to contemporary market evidence. This contrasts to the methodology originally set out in Smithers & Co's 2003 report (Mason, Miles & Wright, 2003) to a consortium of UK regulators, and employed fairly consistently since then, that the real market cost of capital should be assumed constant, on the basis of data from long-term historic averages of realised stock returns.

3. The range of values proposed in Mason *et al*'s original report was between 6½% and 7½% (arithmetic average returns) based on a figure of around 5 ½% (geometric average) derived with reference both to long-term returns on the UK stock market and an average of major markets. Ofgem subsequently opted to assume a return of 7 ¼%, in the upper end of this range.

4. Even without any change in methodology, a case can be made for two, relatively minor, downward adjustments to this figure:

   - Some case can be made for reducing the assumed return in the light of realised returns since 2000 (the terminal date for calculations in Mason *et al*), simply because the extra years of data (up until 2012) have lowered long-term averages for both UK and global markets, by very similar amounts (by around 0.4 % points). However it should be noted that the lowering of historic averages largely reflects the unwinding of overvalued markets in 2000, which had been anticipated, and hence had already been at least somewhat factored into Mason *et al*'s long-term return estimates.
   - A quite strong case can also be made for adjusting returns downwards for the "formula effect", since the Office for National Statistics project that this will increase RPI inflation by around 0.4 percentage points, and hence lower required returns by the same amount, compared to what it would have been on the previous methodology. But, for the sake of transparency and comparability with historic returns it would be preferable for this adjustment to be made separately.

5. **We conclude that, with unchanged methodology the assumed real market cost of capital feeding into WACC calculations would be lowered by around ½% point (or at most ¾ % point). Based on Ofgem's previous assumptions, this would bring it down to around 6¾ %, or (at the lowest) 6 ½%. This figure is at the very top of the CC's assumed range of 5 to 6½%.**

6. The primary factor explaining this difference appears to be that the CC has given at least some weight to a model in which the expected market return is assumed to have been pulled down by falls in the risk-free rate. In Mason *et al* we argued against this model, pointing to the lack of any historical stability in the risk-free rate, and hence in estimates of the market equity premium. We believe that recent events have simply added to the weight of evidence against this approach.

7. In contrast the Mason *et al/*Ofgem approach implies a counter-cyclical equity premium, which is consistent with some more recent academic research, and with recent patterns in observable proxies for risk premia such as corporate bond spreads. It also has the advantage of providing stability in the regulatory process.

8. **We conclude that there is no plausible case for any further downward adjustment in the assumed market cost of equity based on recent movements in risk-free rates (or indeed any other "recent market evidence")**

9. We should however note an important caveat to this conclusion: it does *not* mean that we conclude that Ofgem has necessarily been getting the overall CAPM-based cost of equity right for the DNOs. Ofgem's consultation document itself points to evidence of a market valuation premium in recent years, which suggests that it may have set the DNOs' cost of equity too high. **Our core conclusions stated above imply that it would be incorrect to ascribe any valuation premium to an incorrect assumption on the assumed market cost of equity. As a direct implication, this suggests that other aspects of the assumed cost of equity merit further investigation.**

1. Updating Estimates of the Long-Term Real Market Return

Mason, Miles & Wright (2003, hereafter MMR)[1] proposed a methodology in which the real market cost of equity (that is, the expected real return on investments in the equities of a firm with a CAPM β of precisely one), should be assumed constant, and set in the light of realised historic real returns over long samples. Since the Competition Commission (amongst others) has signalled a departure (albeit modest) from this approach, we revisit below (in Section 2) the rationale for this methodology. But before doing so, in this section we first consider whether data from the last 11 years or so implies any required updates in the numerical estimates derived using Ofgem's existing methodology.

It should be stressed at the outset that the figures proposed by MMR were not derived by a mechanistic formula, eg based on the historic average real return in a single country. Instead, we considered evidence (see MMR Section 2.4) from a range of major markets, exploiting both the database of Dimson, Marsh & Staunton[2] of returns on stocks, bonds and bills for the major markets since 1899, and the longer dataset (from 1802) constructed by Jeremy Siegel for the United States.[3]

The rationale for using long-term average returns.

The real market cost of equity is an expected return, and hence inherently unobservable (an issue we revert to in more detail in Section 2, below). However, we can conceptually split *realised* returns on any asset into two terms, namely

$$\text{Realised Return} = \text{Expected Return} + \text{``Surprise''}$$

where, by letting the surprise term be defined as the realised return minus the expected return, we are simply adding and subtracting the expected return.

Over a long enough sample we would hope that surprises should cancel out, thus revealing the *average* expected return. But there are three important caveats to this approach that complicate such calculations:

- Not all surprises cancel out;
- Large valuation shifts (eg those in recent years) can change average returns nontrivially even over very long samples;
- The average return itself may not be stable, implying a similar lack of stability in the average *expected* return.

MMR considered both issues in depth; here we briefly revisit the key issues, to the extent that they are relevant to more recent data.

[1] Robin Mason, David Miles and Stephen Wright *A Study into Certain Aspects of the Cost of Capital for Regulated Industries in the UK*, February 2003 (http://www.oftel.gov.uk/publications)

[2] Dimson, E., P. Marsh, and M. Staunton (2001): *Triumph of the Optimists*. Princeton University Press.

[3] Siegel, J. J. (1998): *Stocks for the Long Run*. McGraw-Hill.

Evidence from US markets

The US dataset plays, and continues to play, an important role in providing evidence of the relative stability of the real stock return, compared to other asset classes, over a very long sample. Figure 1.1 updates this evidence using data since 2000 (the terminal date for data used in MMR).[4]

Figure 1.1
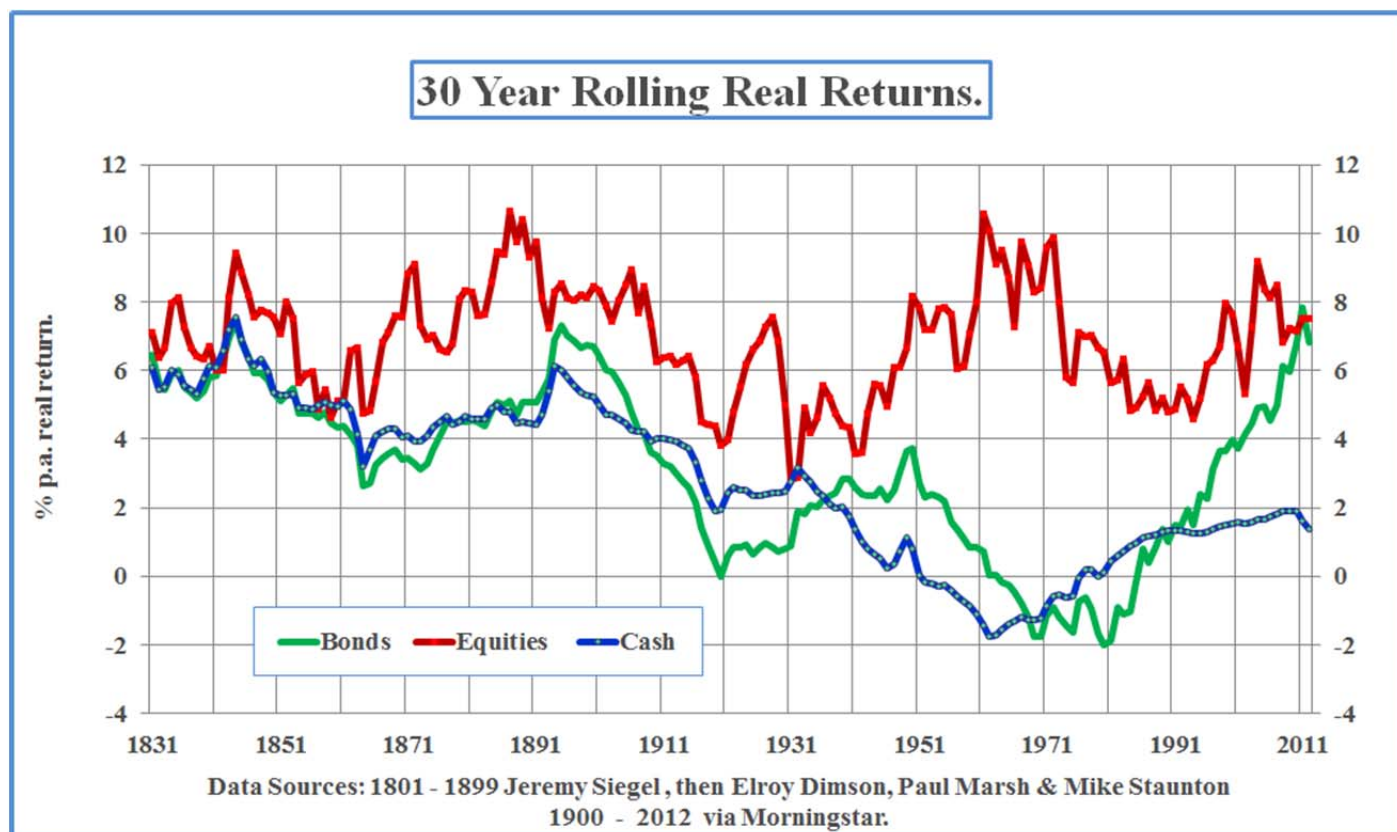
An extra 12 years of data change the overall profile of the chart relatively little, but if anything reinforce its overall message, of a remarkable degree of stability of the long-term real return on US stock markets. This stability is remarkable both in absolute terms, and relative to the *lack* of stability of realised returns on competing asset classes. We discuss below the crucial implications of this latter feature for any attempt to estimate the market risk premium.

We continue to regard evidence from the US stock market as extremely important, both because it is such an important market in global terms, and because its history is so well-documented. However we noted at the time two features of long-term returns in the USA might lead to an overstatement of the market cost of equity:

- **Survivor Bias**: The US market represents the return from what is now the dominant global market. But at the start of the dataset (indeed even at the start of the Dimson dataset at the

---

[4]The three series shown are: "equities", the return on a broad-based equity portfolio; "bonds", the holding return on a portfolio of long-dated government bonds; "cash", the return on short-dated treasury bills. The figure shows compound average rolling real returns over a thirty year horizon, thus includes underling annual return data since 1802, to limit the impact of short-run volatility. All three series are treated symmetrically.

turn of the 20th century) it would not have been obvious that this would be the case. Thus the realised return on the US market may plausibly contain an element of "one-sided surprise", and thus may well have exceeded the return that was expected.

- **The 1990s Boom:** We noted in MMR that data to 2000 were still affected to a nontrivial extent by the stock market boom of the late 1990s. We therefore considered (MMR Section 2.4.5) estimates that attempted to correct for the overvaluation. These suggested that the historic average long-term return might overstate the true average (corrected for overvaluation) by up to 0.8% percentage point. Subsequent data appear to support the need for a correction: the historic average compound average return from 1802-2012 was 6.7%, compared to a figure of 7.1%, using data only to 2000, reflecting the weak realised returns since 2000 (also clearly visible even in the 30-year rolling return shown in Figure 1.1). Even by 2013, valuation criteria had not reverted to long-term historic averages[5] so it is quite possible that the larger downward adjustment proposed in MMR may yet prove more accurate. MMR did however also note that evidence of predictability of stock returns from valuation criteria was (and remains) contentious.

Evidence from global markets

While the unwinding of the 1990s stock market boom has done quite a lot to remove the second source of bias in long-term US stock market returns, the survivor bias problem remains. For this reason in MMR we also looked for comparative purposes at returns from other markets, despite these being over shorter samples.

Figure 1.2 summarises the evidence from other markets, using data from the Dimson et al database, based on returns over the period 1900 to 2012.

Figure 1.2 shows compound average real returns both over the full sample and splitting the data into two roughly equal halves. In so doing it provides evidence that, while being aware of the problem of survivor bias in the US stock market, we also need to be aware of the possibility of what might be termed an offsetting "disaster bias" in return estimates for some other markets. Returns for the first six countries in the chart show a distinct pattern: very low real returns in the first sub-sample (which includes both world wars) and consistently higher returns in the postwar sample. It can be argued [6] that this pattern reflects the extreme events that hit this group of countries in particular during the first sub-sample, most notably, extensive destruction of physical capital in two world wars. In the second sub-sample, returns for these "unlucky countries" were much closer to those of the remaining "lucky" countries.[7]

---

[5] For recent estimates, see Smithers & Co Report no. 434. Data on the cyclically adjusted P/E are regularly updated on Robert Shillers's website, http://aida.wss.yale.edu/~shiller/data.htm; data on q are also regularly published on Smithers & Co's website, www.smithers.co.uk

[6] See Andrew Smithers, *Wall Street Revalued,* Chapter 11.

[7] Returns for Italy remain a clear outlier even in the second sub-sample. We do not claim to have any idea why this should be.

Figure 1.2

Real $ Equity Market Returns
(Geometric Averages, Source DMS)
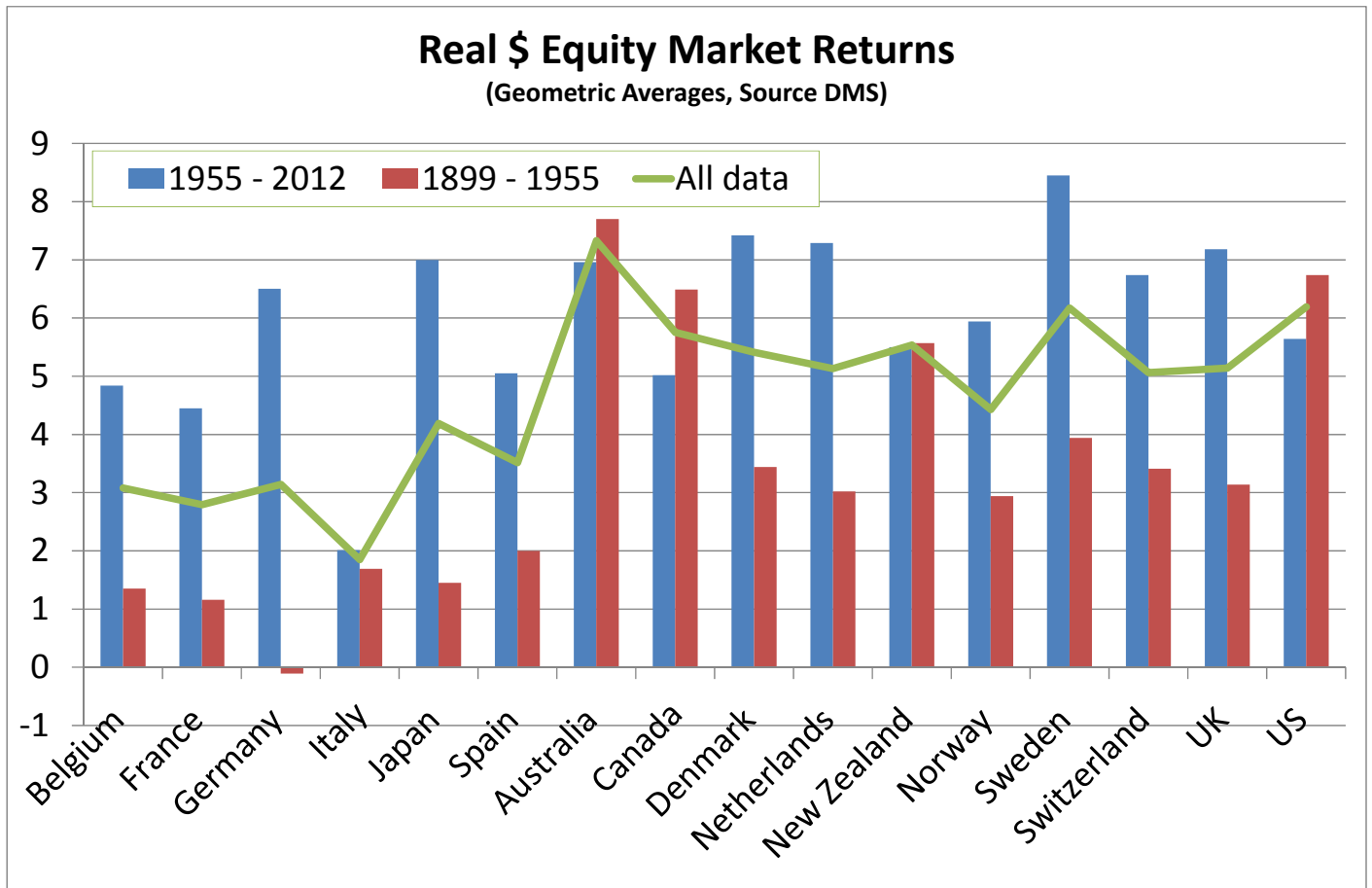
Is UK Evidence Special?

In MMR we argued strongly that, in highly integrated global stock markets, we should not look exclusively at data from the UK, which is (at least in recent data) a relatively small market in global terms. At the same time we argued (MMR Section 2.4.6) that the relatively stable experience of the UK economy – with an absence either of significant positive or negative surprises, and (as illustrated by Figure 1.2) a return close to the global average, meant that it might (largely coincidentally) be regarded as a good proxy for global expected returns.

We continue to regard this argument as valid, but it should not be over-interpreted, as implying that *only* evidence from UK markets should be considered in setting a figure for the market cost of equity. This caveat is of particular importance given the possibility (discussed below) that changes in the measurement of UK RPI inflation may introduce new distortions. Since this is a UK-specific problem, it becomes all the more important to take into account evidence from other markets.

Summarising International Evidence

Table 1.1 provides a summary of long-term evidence on compound (i.e. geometric) average stock returns, and illustrates how recent data have affected the average.

Table 1.1. Summary of Long-Term Real Returns on UK and Global Stock Markets

	Compound Average Real Returns, % p.a.				
	UK, £	UK, $	World, $	World, excl US, $	US, $
1899 to 2012	5.23	5.23	5.01	4.42	6.26
1899-2000	5.78	5.61	5.42	4.68	6.89
2000-2012	0.67	2.08	1.55	2.27	1.08
1955 to 2012	6.58	7.32	5.58	6.04	5.78

Source: Dimson, Marsh & Staunton/Credit Suisse Global Returns database

We would bring out the following key points from this table:

- On a comparable basis (ie, in dollar terms), the long-term real return over the full sample (1899-2012) on the UK stock market has been very close (within roughly 20 basis points) to the global average return.

- This feature is evident whether returns are measured in real sterling terms (ie, using a UK measure of prices – here, the RPI) or in real dollar terms (i.e, using a US measure of prices, the US consumer price index). Indeed the similarity in returns is particularly striking over the full sample of data.[8] Thus, over long samples, movements in the sterling/dollar rate have cancelled out movements in relative prices. (But note that this has *not* been the case over shorter samples – viz, the lack of equality of the two measures of returns over the past 12 years)

- In contrast, as discussed above, US returns have been distinctly higher than the global average over the full sample, and thus have nontrivially boosted the long-term global average over the full sample.

- It is worth noting, however, that in the postwar sample (as illustrated in Figure 1.2), global markets excluding the USA actually achieved *higher* returns than the US markets. Thus a large part of the long-term out-performance of the US market reflects the exceptionally *low* returns of the "unlucky countries" in the first half of the twentieth century.

- Returns on all markets have been low, relative to historic averages, since 2000, and thus have depressed the average using all data by between 0.25 and 0.6 of a percentage point. The extent of this negative impact has been, as expected, largest for the US market (given that, on most measures, it appeared considerably more overvalued at the end of the 1990s) and least so for the global measure excluding the US.

---

[8] The two figures are actually different, but only beyond the second decimal place.

- It should be stressed, however, that this pattern of returns was largely anticipated in MMR's original analysis, and that the relatively low figure of 5.5% (compound average) assumed therein included, at least informally, a degree of adjustment for overvaluation. It is notable that the global average return in the postwar sub-sample 1955-2012 (which arguably is less contaminated by the impact of either capital destruction or over-valuation) is extremely close to MMR's original assumed figure.

- **Thus on the basis of evidence on long-term compound average returns alone we would argue that a downward adjustment of 40 basis points in the assumed compound average market return (in line with the shift in both the UK and global average return) would be the *most* that would be warranted in light of recent data.**

Geometric vs Arithmetic Averaging

In MMR (Section 2.4.2) we argued for a methodology of looking at evidence from compound average returns (or the virtually identical average log returns), and then making an adjustment for the impact of arithmetic averaging, which takes into account the volatility of returns, rather than measuring arithmetic returns directly. There are several arguments for this approach:

- On *a priori* grounds we would expect the distribution of returns to be more stable in log terms than in terms of the absolute return, given the asymmetry of the distribution of absolute returns.

- If returns have any predictability – based either on their own past, or from valuation indicators – then short-run estimates of stock return volatility may overstate volatility of returns over longer samples, which are more relevant to investments in regulated companies. But since evidence of predictability is contentious, we argued that regulators should explicitly consider whether they wish to take this factor into account.

- When returns are compared or aggregated across markets they are usually expressed in a common currency (usually dollars). This introduces an additional source of short-term volatility, from exchange rate movements, and thus when we compare arithmetic average returns in, say, a global market index in dollars (as above) with returns on a single market like the UK, which is uncontaminated by exchange rate volatility, we are not comparing like with like. However compound average returns, measured over long samples, are unaffected by this distortion.

Table 1.2 illustrates.

Table 1.2: Arithmetic vs Compound Average (Geometric) Returns

	UK, £	UK, $	World, $	World, excl US, $	US, $
Arithmetic Average Returns, % p.a.					
1899 to 2012	7.08	7.70	6.50	6.14	8.24
1899-2000	7.65	7.96	6.80	6.24	8.86
2000-2012	2.31	5.51	3.92	5.35	2.96
1955-2012	9.03	10.39	7.34	8.08	7.61
Arithmetic minus compound average returns,% p.a.					
1899 to 2012	1.86	2.47	1.49	1.73	1.98
1899-2000	1.87	2.35	1.38	1.56	1.97
2000-2012	1.64	3.42	2.37	3.08	1.89
1955-2012	2.46	3.07	1.76	2.03	1.83
Standard deviations of returns, % p.a.					
1899 to 2012	19.85	23.22	17.49	19.18	20.12
1899-2000	20.03	23.04	17.02	18.56	20.21
2000-2012	18.30	25.69	21.73	24.81	19.36
1955-2012	22.90	25.26	17.15	19.70	17.43

Source: Dimson, Marsh & Staunton/Credit Suisse Global Returns database

While the table shows that arithmetic averaging returns are always (as would be expected) significantly higher than compound average (or geometric[9]) returns, it also illustrates that this difference varies considerably depending on the index used, and sometimes for largely spurious reasons:

- Most strikingly, while, as noted above, compound average returns for the UK market over the full sample are virtually identical, whether measured in sterling or dollars, the arithmetic average dollar return is distinctly higher. But this difference is entirely spurious, since it simply reflects additional volatility from exchange rate movements.

- The gap is also much larger, whether UK returns are measured in sterling or dollars, in the postwar period, reflecting a higher volatility of returns. This feature is consistent with the finding in our 2006 report to Ofgem[10] that the CAPM β of the UK market on the global market appeared to have risen. But volatilities of returns on both global and US markets in this sub-sample were equal to or lower than in the full sample, so, incorporating this higher volatility into an arithmetic average would arguably again be spurious.

- Returns on global indices have lower volatility, due to the gains from international diversification (which more than offsets the additional volatility due to exchange rate movements) , and thus have the smallest gaps between arithmetic and geometric returns. But the gap is also affected by whether the US market is included in the average: the global return

---

[9] Strictly speaking the compound average return is the geometric average of (1+return), minus 1, but, as in MMR we take the terms compound average and geometric average to be equivalent. For precise formulae for the different averaging techniques see MMR Section 2.4.2.

[10] Stephen Wright, Robin Mason, Meltem Baskaya, Ken Hori and Steve Satchell, *Report on the Cost of Capital* September 2006 (http://www.ofgem.gov.uk)

excluding the US is more volatile, since it is both less diversified, and more strongly affected by exchange rate movements.

In light of the distortions introduced by direct arithmetic averaging, we continue to advocate deriving return estimates from compound average returns. A deliberate decision then needs to made on how much to adjust for the impact of return volatility on the arithmetic average. In MMR we argued for an adjustment of 1 to 2 percentage points, depending on how much account regulators wish to take on predictability of returns. We see no reason to change this recommendation.

Adjusting for the impact of increased RPI bias

Ofgem's consultation document (Appendix 2, paras 2.5 onwards) notes that changes in the way the ONS has calculated the Retail Prices Index since 2010 appear to have increased the so-called "formula effect" by around 0.4 percentage points. The ONS and OBR project that this is likely to increase measured RPI inflation by this amount on a sustained basis.

The long-term UK returns summarised in Table 1.1 above were calculated by Dimson *et al* using RPI inflation (from 1947 onwards) and a much more narrowly defined cost of living index in earlier data. Thus, since regulated asset values will be continue to be indexed in line with the RPI, which is now expected to grow systematically more rapidly than it has in the past, this would suggest a *prima facie* case for an adjustment to the assumed market cost of equity of the same amount, namely 0.4 percentage points.

While we would agree that some adjustment can be justified on these grounds, we would note that there are some caveats to this conclusion:

- This is by no means the first change that has been made to the RPI. Such changes have not been very well documented. As an example, recent analysis by the Office for Budget Responsibility of the "long-run" difference between RPI and CPI Inflation does not actually address the question on any reasonable measure of the long run, since it uses data only since 1998 or later.[11] We therefore simply do not know whether, for example, this new source of bias may simply offset the impact of other biases in earlier data.

- As an example, the inflation measure used by Dimson et al in the period before 1947 differs nontrivially from a new measure introduced recently by the ONS[12], with the result that even over the full sample, average measured inflation on the ONS measure is around ¼ % point higher (and hence real returns are an equivalent amount lower). But, given available information it is not possible to tell which of these alternative measures of inflation is most consistent with the methodology used for measuring RPI inflation from 1947 onwards, which would give the ideal degree of consistency over time.

- One cross-check (albeit a rough-and-ready one) is to side-track the issue of measuring UK prices altogether, and look at long-term real returns for a dollar investor investing in the UK market, where the correction into real terms uses US, rather than UK price indices. The

---

[11] Ruth Miller, "The long-run difference between RPI and CPI inflation", OBR *Working paper* No. 2, November 2011

[12] The ONS publish a "Long term indicator of prices of consumer goods and services" (ONS code CDID) in their publication: MM23 Consumer Price Indices

former have arguably been constructed on a more consistent basis (the Bureau of Labour Statistics publishes US CPI data as far back as 1913, on a consistent basis). While any such calculation requires the additional assumption that long-run exchange rate movements offset movements in relative prices at countries at a similar level of development, this calculation does offer some reassurance that Dimson et al's real return estimate is in the right ballpark, since (as noted above) the dollar real return is virtually identical to the sterling real return. But, while this offers some reassurance it is clearly only very rough and ready.

Given these caveats, we would argue that, other things being equal, a cautious approach should be applied. We would therefore argue for a downward correction due to RPI bias of *at most* **0.4 percentage points, with a preference for a smaller adjustment on grounds of caution. We would also argue that any such correction should be made as an explicit adjustment, rather than simply incorporated into the assumed long-term average.[13]**

Summary of recommended adjustments to estimates of average long-term market returns on unchanged methodology.

In summary:

- In MMR we proposed a central estimate of 5 ½ % for our estimate of the compound average real market return

- Recent data supports a downward adjustment of *at most* 40 basis points to this estimate due to recent market movements in both UK and global markets.

- But since our original estimate had already informally factored in some expected correction to market overvaluation (and was based also on longer-term data from the USA) we would argue for a lesser adjustment: a cautious figure would be a downward adjustment of 25 basis points, to 5¼%.

- We see no reason to change our view on the impact of arithmetic vs geometric averaging, which in MMR we set at 1 to 2 percentage points, depending on the view Ofgem wishes to take on predictability of returns.

- **Thus we would suggest a modified range of 6.25 to 7.25% for the assumed arithmetic average real market return on a consistent RPI basis.**

- **We would also argue that any correction for the assumed future "RPI formula effect" should be explicitly indicated. The adjustment should again be** *at most* **of 40 basis points. A cautious approach would suggest a further downward correction of 25 basis points to correct for the formula effect.**

---

[13] This analysis also clearly begs the question of whether Ofgem should continue to use RPI as a basis for indexation, rather than, for example, the CPI, as targeted by the Bank of England. But this issues goes beyond the scope of the current consultation.

2. Should Estimates of the Real Market Cost of Equity take into account "Recent Market Evidence"?

The analysis of the previous section restricted itself deliberately to the task of using long-term data on realised stock market returns to attempt to derive an estimate of the inherently unobservable average *expected* market return. We briefly review here the arguments (covered in greater depth in MMR Section 2) for using such averages as estimates of the real market cost of equity in WACC calculations. We contrast it, in particular, with the alternative approach of building up the real market cost of equity from the risk-free rate plus an assumed market risk premium.

The cost of equity is an expected return, and hence inherently unobservable

The price of a share should in theory be equal to the risk-adjusted present discounted value of expected future payoffs per share. The risk-adjusted discount rate *is* the cost of capital: a lower discount rate would increase the price, for a given expected payoff, and hence allow the company to raise money more cheaply. The discount rate should also be equal to the expected *return* to a rational investor. But since it is an expectation it is inherently unobservable.

In contrast *realised* returns, after a share has been bought, incorporate both changes in expected cashflows *and* changes in discount rates, and hence expected returns. To introduce a concrete example: how are we to interpret the realised negative returns in global stock markets in, eg, 2008 and 2009? If we assumed that expectations of future cashflows to shareholders were unchanged, these falls would indicate a *rise* in discount rates, not a fall. But arguably expected cashflows (or equivalently, expected returns on capital employed) arguably also fell over this period. **Thus it is impossible to make any clear-cut inference about the cost of equity based on just a few years' returns.**

The risk-free rate (at a given maturity) **is** *observable*

Returns on risk-free assets provide a very marked contrast. As long as we focus on default-free bonds, the yield on a zero coupon bond is by definition the expected return if the bond is held to maturity. Barring minor problems that zero coupon bond prices must be inferred from the prices of coupon bonds, we can therefore directly observe the risk-free return at any given maturity. **While we noted this (to us), rather self-evident point in the analysis of MMR (Section 2.5.2), it is worth stressing, because of the observed tendency of both Ofgem and most other UK regulators** *not* **to feed observable measures of risk-free rates into WACC calculations.**

The market equity premium is inherently unobservable

This point follows by basic arithmetic, given the definition of a risk premium as the gap between the (observable) risk-free rate, at any given maturity, and the (unobservable) expected market return. **Thus all statements about the market cost of equity or the market equity premium are guesses, or assumptions, or typically some combination of the two.**

Best Guesses vs Market Expectations

Even if we focus solely on "best guesses", ie. forecasts of returns based on econometric evidence, we can only equate these to market expected returns by making some additional assumptions. Thus,

suppose we make some econometric forecast of returns, based on some indicator: for example a valuation indicator like the cyclically adjusted P/E, popularised by Robert Shiller (2000) or *q*, as used by Smithers & Wright (2000) or some variant of the Dividend Discount Model (eg, Fama & French, 2001).[14] Can we simply assume that these "informed best guesses" are equal to market expectations? In practice any such assumption is subject to major caveats:

- There is considerable econometric uncertainty about any such forecasts. The academic debate about the predictability of stock returns continues, but just about the only point of agreement in the literature is that *no* model has anything more than a marginal ability to predict returns.

- Additionally, even if a given model is taken as the true model, further assumptions are required to equate econometric forecasts with market expected returns: first, that market expectations are equal to mathematical expectations (ie, minimum variance forecasts based on available information); and second, that the information set we are using is the same as market participants. These are both very strong assumptions.[15]

As a practical example of the importance of these caveats, in the late 1990s a range of authors (including those cited above) used econometric and other evidence to conclude that prospective returns on US and global stock markets were likely to be below average (see MMR Sections 2.4.5 and 2.5.1). As discussed in Section 1, this evidence was indeed taken into account by MMR in assessing long-term averages of market returns.

It is however a nontrivial leap to assume that evidence of any econometric predictability of market returns was automatically incorporated into market expectations at the time. Indeed, it did not represent even an academic consensus view (counter-examples of eminent academics who took the contrary view were, eg, Robert Hall and Edward Prescott[16]). More crucially, survey evidence of market expectations in the late 1990s suggested that many participants expected *above*-average returns.[17]

Thus even while there is evidence of predictability (albeit extremely limited) and hence variation over time in econometric predictions of market returns, this does *not* necessarily translate to market expectations of the market return, which is what we should be seeking to measure. There is thus no straightforward, systematic, transparent and replicable way of incorporating "recent evidence" into estimates of the market cost of equity.

What should we assume to be stable?

In light of the difficulties in assessing true market expectations, we argued in MMR that regulators need to choose between competing *assumptions* about the market cost of equity, based on historic averaging over long samples. Specifically, should we make an assumption about the expected

---

[14]Robert Shiller (2000) *Irrational Exuberance*, Princeton University Press; Andrew Smithers & Stephen Wright (2000) *Valuing Wall Street*, McGraw-Hill; Eugene Fama., and Kenneth French (2001): "The Equity Premium," *Journal of Finance*

[15] It includes, for example, the implicit assumption that we can apply the Law of Iterated Expectations: which essentially requires that that are no Keynesian "Beauty Contests".

[16] See Robert Hall (2000): "`e-Capital': The Link Between the Stock Market and the Labor Market in the 1990s" *Brookings Papers on Economic Activity*; .Ellen McGrattan and Edward Prescott (2001): "Is the Stock Market Overvalued?" *Working Paper* 8077, NBER

[17]See, for example, Sushil B. Wadhwani, "The US stock market and the global economic crisis" *National Institute Economic Review* January 1999 vol. 167 no. 1 86-105

market return itself, or should we build this assumption up by making an assumption about the market risk premium, and adding it to the risk-free rate?

A minimal requirement for assuming some magnitude is constant in expectation is that it should historically have been stable, ex post. A glance at Figure 1.1, above, demonstrates very clearly, on the basis of US data, that real stock returns have shown a remarkable degree of stability over more than two centuries. This is consistent – given the analysis of Section 1 – with the average *expected* return having been stable. In marked contrast, there is no such stability evident in either the bond return or the risk-free return. MMR also showed strong evidence that the same pattern of relative stability is evident in the other markets in the Dimson database

This point was made quite forcefully in MMR (Section 2.5.2); and experience since then has strongly reinforced it. At the time of writing, in 2003, it was at least possible to point to a consensus amongst monetary policymakers – based largely around the "Taylor Rule" for monetary policy – that the real risk-free rate "should" be around 2%. But even at the time MMR noted that the empirical basis for this assumption was quite weak – only in the relatively short period roughly from 1980 until the early 2000s did this figure correspond to an actual historic average. Average ex-post risk-free rates in the US and UK over the twentieth century as a whole were very much closer to zero.

Events since 2003, and especially since the onset of the financial crisis in 2008, have provided a very clear reminder of the dangers of making guesses about either the risk-free rate or bond yields. Short-term risk-free rates have fallen to zero, or near-zero in most countries, while inflation has continued to be mainly positive, so that on any reasonable short-term forecast of inflation, real short-term rates have been nontrivially negative. At the same time quantitative easing has pushed up bond prices, thus giving extremely strong bond returns, *ex post* (Figure 1.1. shows that in the recent past long-term real returns on bonds have actually been higher than those on equities in the USA), but, in so doing, have pushed *expected* returns on bonds downwards. As noted by Ofgem in their consultation document (and as shown in Figure 2.1, overleaf) real yields on UK index-linked bonds have in the past few years turned negative, for the first time since their introduction.

Figure 2.1

Yields on Indexed-Linked Bonds: Maturities 2.5 to 25 years
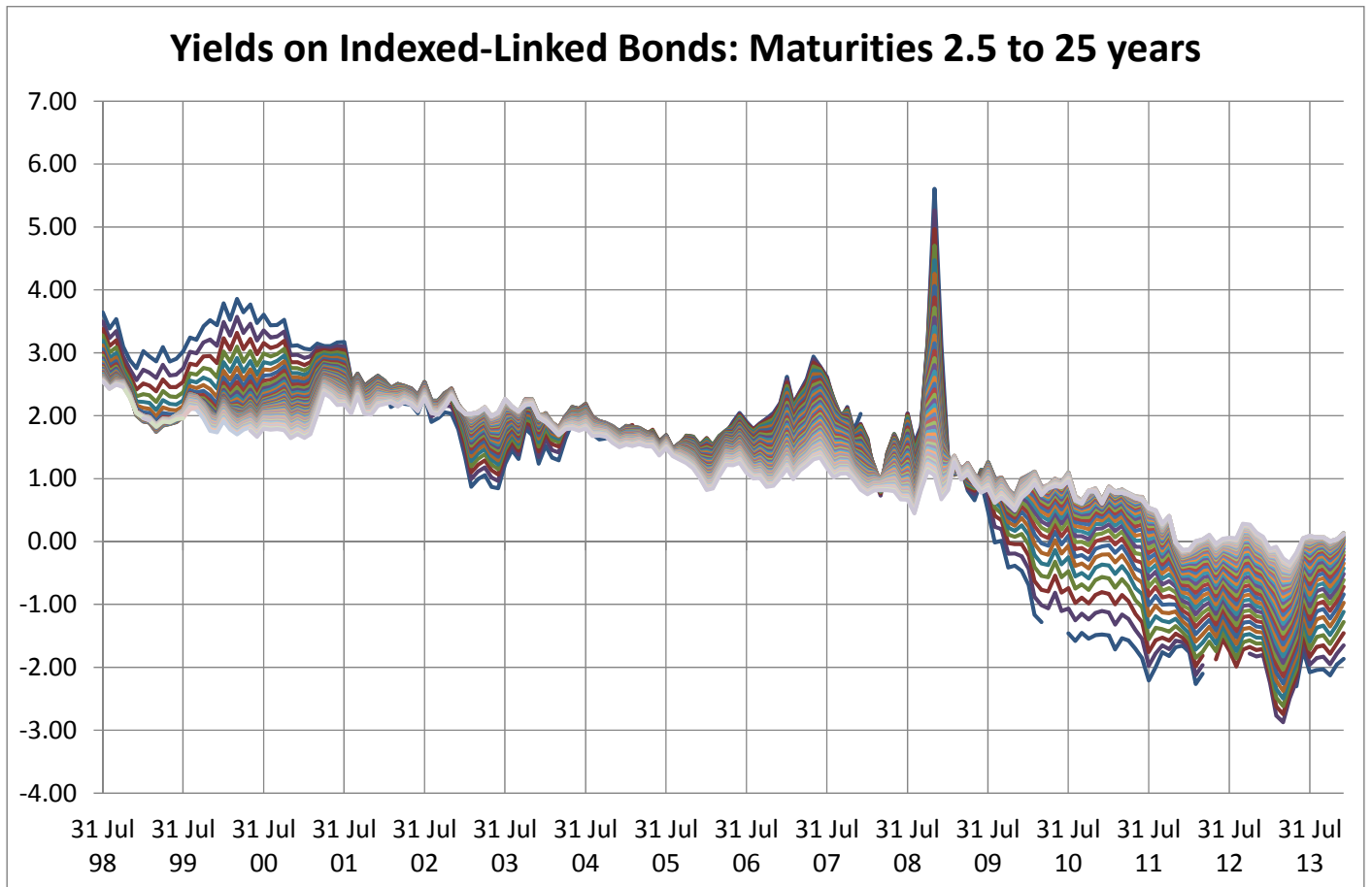
Thus both historical and more recent evidence point to the same conclusion: in contrast to the stock return there is no evidence of stability in the risk-free rate, at any maturity. As a direct implication, there is no evidence of stability of the market equity premium. Without such evidence, there is no empirical basis for the assumption that falls in risk-free rates should translate to falls in expected market returns.

3. Implications of a constant market return assumption for the cyclical properties of the (assumed) market equity premium

In Section 2 we reiterated the conclusion of MMR, that regulators should assume that the (unobservable) real market cost of equity is constant. We also noted that the risk-free rate *is* observable (although we also noted that regulators seem strangely unwilling to exploit this observability).

It is therefore an application of simple arithmetic to conclude that, applying our methodology, the (assumed) market risk premium and the risk-free rate must move in opposite directions: ie, must be perfectly negatively correlated. This phenomenon cannot, of course, be directly observed, since the true MRP is also, as noted above inherently unobservable. However, it is reasonable to ask whether this pattern of behaviour in the market equity premium is at least both plausible on *a priori* grounds, and consistent at least with indirect evidence on risk premia. We argue below that it is.
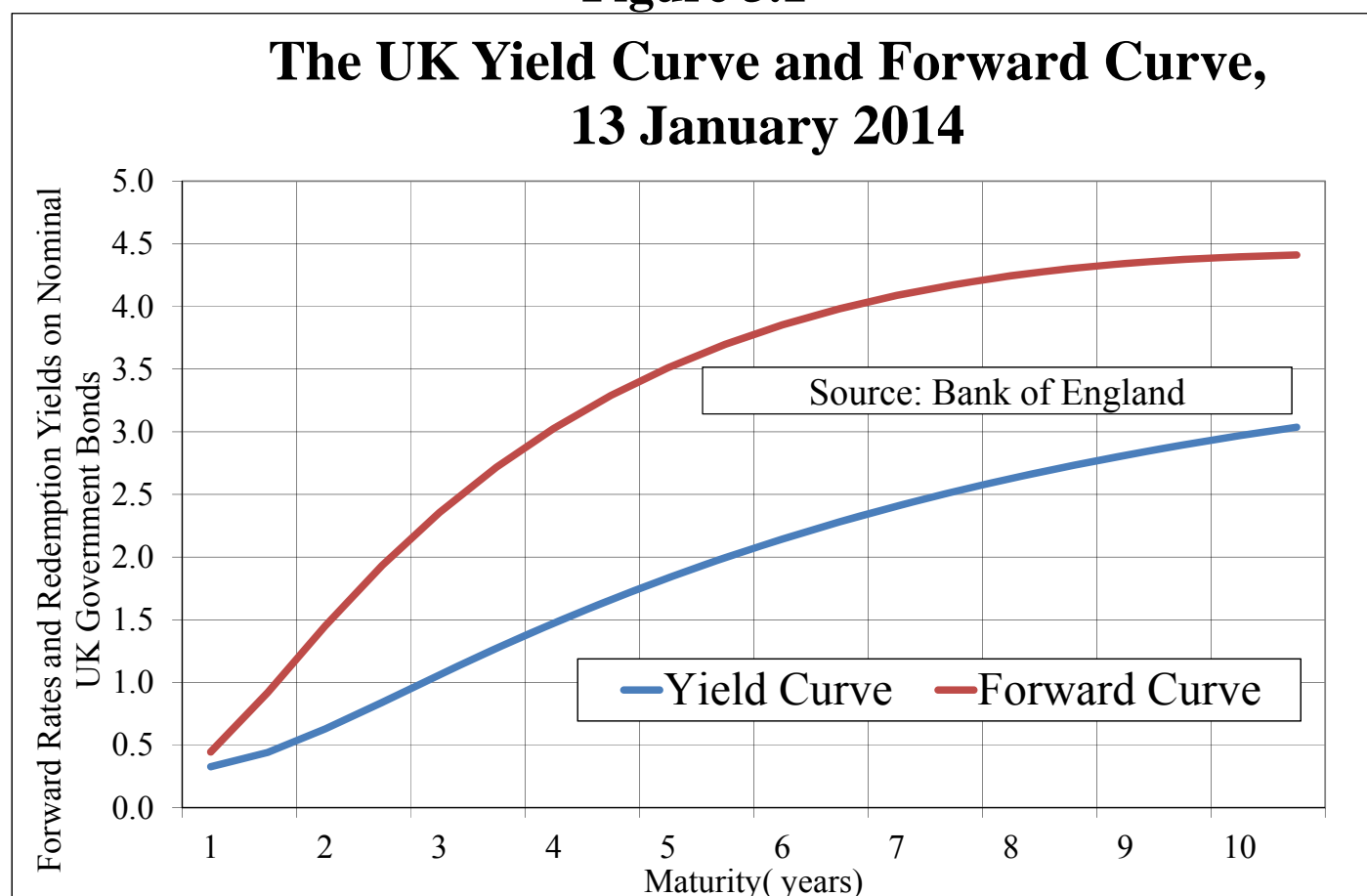
What determines the risk-free rate?

It is fairly evident that in most major economies risk-free rates, set by the central bank, move pro-cyclically. This is indeed the essence of the "Taylor Rule" that has dominated thinking on monetary policy in most major economies in recent decades. Thus, in recent years, a combination of high unemployment and relatively subdued inflationary risks has resulted in risk-free rates in most major economies being pushed virtually to zero in nominal terms, and well below zero in real terms. In the UK, having effectively reached the zero bound for nominal rates at the short maturities associated in the past with monetary policy, the Bank of England's programme of Quantitative Easing has in effect simply shifted the maturity of the operations of monetary policy along the yield curve.

But, based both on public pronouncements and past experience, this is clearly not a permanent state of affairs. During the era since the early 1990s, in which inflation has been broadly stable, the policy rate set by the Bank of England (as well as most other major central banks) has moved up when either inflation or the economy has recovered, and has thus moved procyclically. It is reasonable to assume that this pattern will continue if the UK economy continues to recover from recession. This expectation appears to be shared both by the Bank of England and – from the evidence of forward rates, at least – by bond market participants (See Figure 3.1, overleaf)

It should be stressed (in light of the discussion above) that we do *not* need to assume that the risk-free rate has a a stable mean for this to be the case: we simply need to assume that *changes* in the risk-free rate are positively correlated with changes in output and inflation.

Figure 3.1

The UK Yield Curve and Forward Curve, 13 January 2014

Source: Bank of England

—Yield Curve   —Forward Curve

(Y-axis: Forward Rates and Redemption Yields on Nominal UK Government Bonds; X-axis: Maturity(years))

What determines risk premia?

Given the clear procyclical pattern of risk-free rates, the assumption of a constant market cost of equity implies that the assumed market equity premium is *counter*cyclical. It is reasonable to ask if this is consistent with what we think we know about risk premia.

While there are a range of competing models of risk premia, they share a number of common features.

The first is that any given asset pricing model should apply across all asset classes. Thus what determines the risk premium on equities should also in principle determine risk premia on any other asset. Specifically, in the benchmark model of modern finance, there should be a common "stochastic discount factor" that applies to all assets; with risk premia on any given asset being determined by its correlation with the discount factor (or, less technically, by how much systematic risk the asset carries). Thus, information about risk premia on one class of assets (eg bonds) should in principle convey some information about risk premia on another class (eg equities).

A second common feature of asset pricing models is that risk premia in general should reflect two factors: the *quantity* of systematic risk (ie, overall market volatility) and the market *price* of that risk. There is quite strong evidence that it is the *price* of risk that varies most, and most persistently.

The period immediately after the financial crisis saw a sharp rise in stock market volatility; but these movements have in large part unwound, which is a common feature of volatility (technically, time-varying volatility has "low persistence"). However, the market *price* of risk can display considerably greater persistence over time. One very common explanation of this greater persistence is that it reflects some measure of the state of the economy, with a weak economy frequently associated with increased risk aversion.
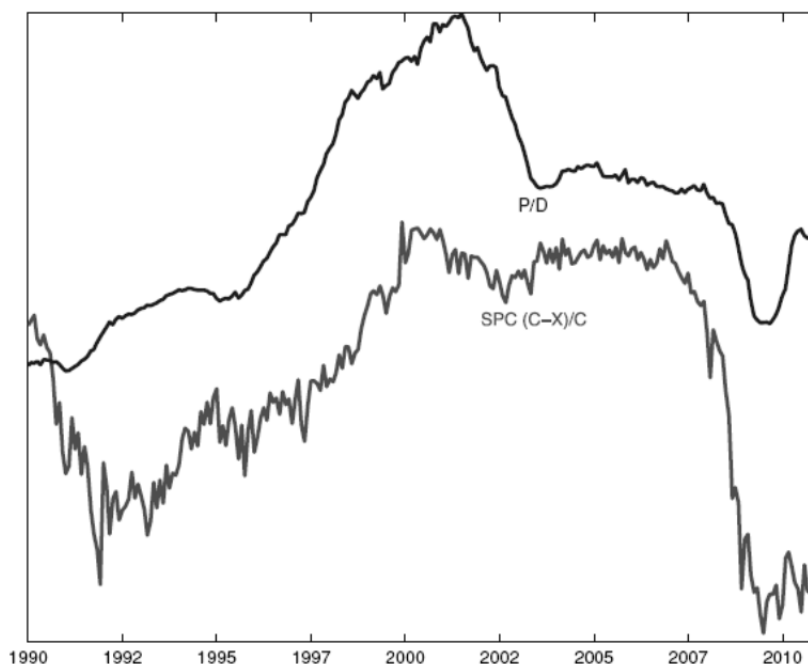
Figure 3.2

Figure 9. **Surplus-consumption ratio and price-dividend ratio.** The price-dividend ratio is that of the CRSP NYSE value-weighted portfolio. The surplus consumption is formed from monthly real nondurable consumption using the Campbell and Cochrane (1999) specification and parameters, multiplied by three to fit on the same scale.

As an example, Figure 3.2 is taken from John Cochrane's presidential address to the 2011 American Finance Association,[18] which plots the US "surplus consumption ratio"[19] (which is assumed to be negatively related to risk aversion) relative to a measure of US stock market value, the price/dividend ratio.[20] The chart shows that during the crisis surplus consumption fell sharply, and has remained low, implying that risk aversion has remained high. Cochrane points out that the chart shows that this was strongly associated with the fall in stock prices, which he thus attributes primarily to a rise in desired returns in the equity market. Given that at the same time, for reasons given above, the risk-free rate was falling sharply, the implication is that the equity premium was rising all the more sharply. (Of course this is only an *implication* – it does not require any claim that the equity premium can be measured directly.)

---

[18] *Journal of Finance,* Vol LXVI (4), pp 1047-1108
[19] This is derived from the "habit formation" model of consumer behaviour, as implemented in Campbell & Cochrane (1999), *Journal of Political Economy,* 107, pp 205-251. It is defined as the gap between actual consumption and some assumed minimum level of consumption – as consumption falls towards this minimum level, measured risk aversion increases. In practice, the implied minimum level is estimated as a slow-moving weighted average of actual consumption.
[20] We would argue that this is not actually a very good valuation indicator. But its short term movements over the crisis shown in the chart were fairly similar to those of other, more reliable indicators of value.

Indirect evidence of counter-cyclical risk premia in government bonds

Recent research by Ludvigson and Ng (2009),[21] provides historical evidence suggesting that risk premia on long-dated government bonds have also displayed counter-cyclical patterns. Figure 3.3, taken from their paper, summarises this evidence, showing that there has historically been a strong tendency for (their estimates of) risk premia to rise during US recessions (based on NBER data).[22]

Figure 3.3
Yields, Expected Returns and Implied Risk Premia on US 5-Year Bonds

Shadings denote months designated as recessions by the National Bureau of Economic Research.
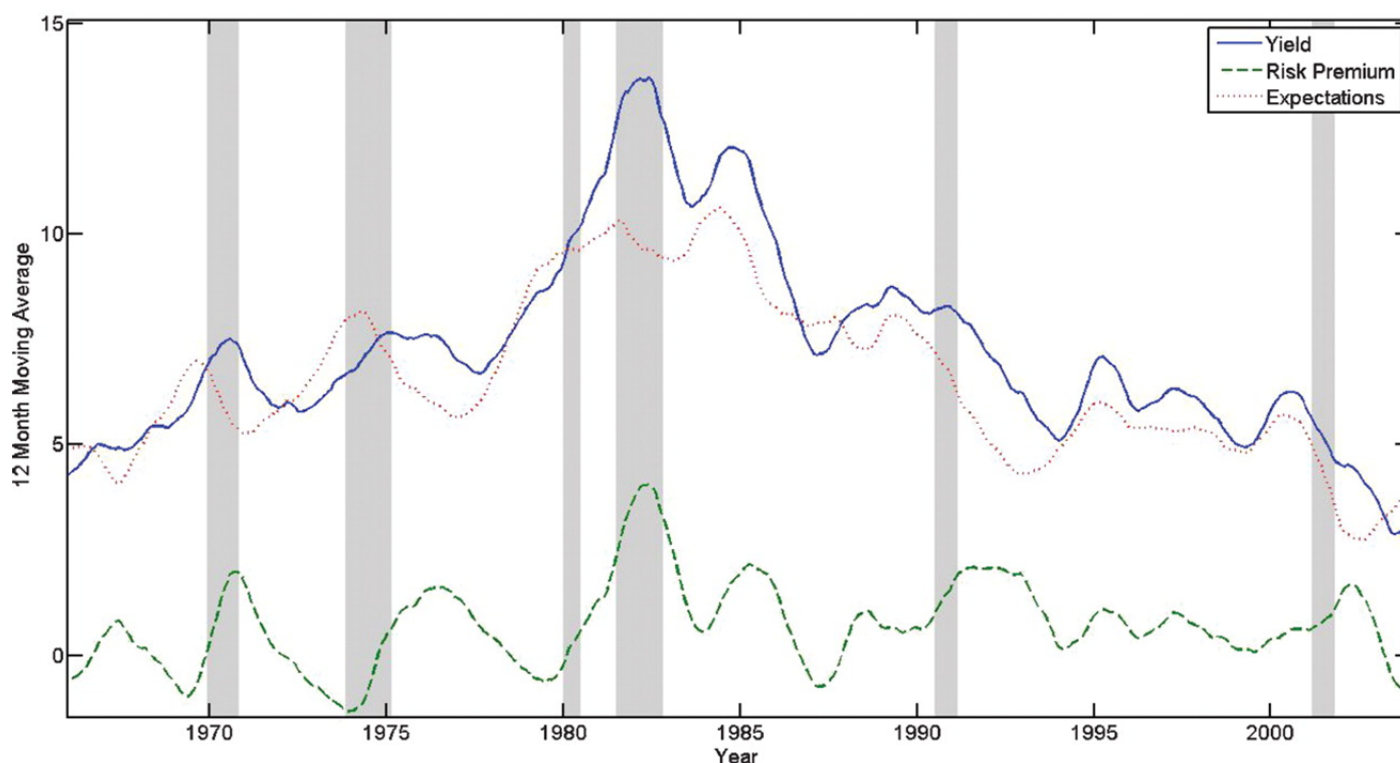
Source, Ludvigsen & Ng (2009)

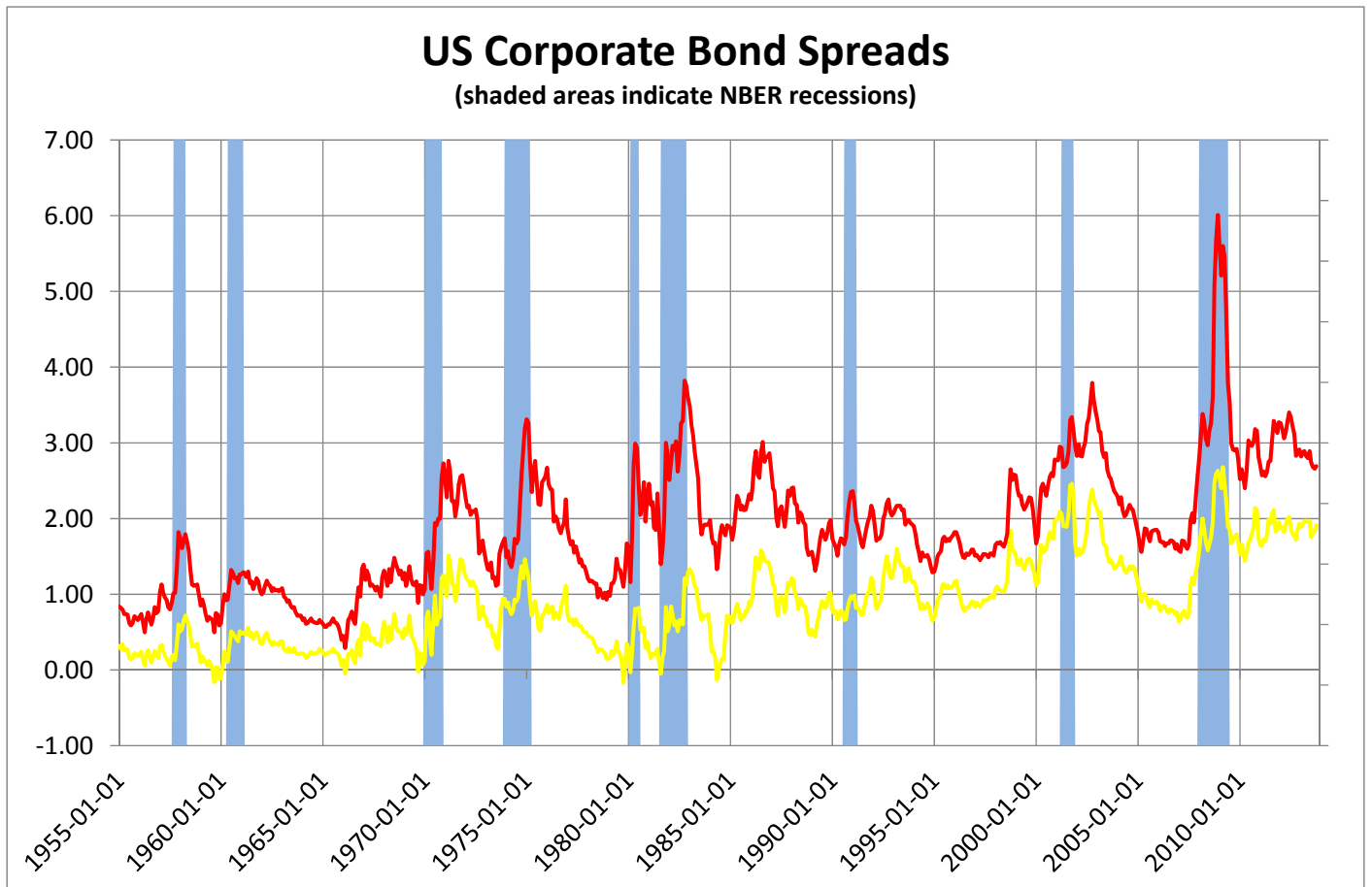*Indirect evidence of counter-cyclical risk premia in corporate bonds*

There is also a considerable body of evidence suggesting that corporate bond spreads contain a strong risk premium element. Elton *et al* (2001), for example, provide evidence that observed fluctuations in the corporate bond spread (see Figure 3.4, overleaf) cannot be explained by other factors such as default risk, tax differences, or liquidity[23]. Since, as Figure 3.4. illustrates, corporate bond spreads strongly counter-cyclical there is a clear implication that these short-term movements largely represent counter-cyclical movements in risk premia.

---

[21] Ludvigson, S and Ng, S, "Macro Factors in Bond Risk Premia", *Review of Financial Studies*, 2009, pp 5027-5067
[22] A caveat, in light of our discussion in Section 2, is that the estimates they use equate econometric forecasts of excess returns to market expectations, and hence to market risk premia.
[23] See, for example, Elton, E, Gruber, M, Agrawal, D and Mann, C (2001) "Explaining the Rate Spread on Corporate Bonds", *Journal of Finance* LVI, pp 247-278.

Figure 3.4

US Corporate Bond Spreads
(shaded areas indicate NBER recessions)

Given the necessary link between risk premia on all asset classes, both of these features of risk premia in other markets suggest strong indirect evidence that the market equity risk premium is countercyclical.

This does *not* mean that we claim that the actual (and unobservable) market risk premium will move one-for-one (and with opposite sign) with the risk-free rate, but we do claim that the general pattern of risk premia is consistent with this assumed countercyclical pattern, which follows from the assumption of a constant expected market return on equity.

4. Conclusions and Further Implications

Our key conclusions are:

- We do not believe that there is any case for changing the methodology for assessing the market cost of equity, as originally recommended by Mason, Miles and Wright (2003) and as implemented thus far by Ofgem and most other UK regulators. We thus disagree with the new approach, suggested by the Competition Commission, of giving more weight to contemporary market evidence in assessing the market cost of equity.

- There is a case for (small) adjustments in the assumed figure, due to a combination of the impact of recent market returns on the average, and the impact of the "formula effect". But *neither* of these adjustments implies any change in the methodology itself.

- We have argued that the counter-cyclical pattern of the implied market equity premium that follows, of necessity, from our preferred approach of assuming a constant market cost of equity is at least qualitatively consistent with both the underlying economic logic, and the available indirect evidence, on the generic properties of risk premia.

- We should perhaps stress that our arguments do *not* rule out using contemporary market evidence on the risk-free rate, which enters the cost of equity calculation in its own right; we simply argue that the current risk-free rate should not play a role in estimating the market cost of equity.

While we have focussed in this report on conclusions that arise within the relatively narrow remit of the consultation exercise, these conclusions themselves have implications that we believe Ofgem should bear in mind in its future deliberations on the cost of capital for DNOs.

While we disagree with the Competition Commission on their approach to the assumed market cost of equity, this does *not* mean that we endorse every aspect of Ofgem's approach to setting the cost of equity for DNOs. Indeed we believe that arguments in favour of reducing the assumed market cost of equity have largely arisen due to a quite widely shared perception that the assumed cost of equity has been set at too high a level.

As evidence of this, Ofgem's own consultation document (para 1.6) notes that there is evidence (albeit only indirect) that "the market has valued regulated networks at more than their regulated asset values, and a valuation premium has persisted for a number of years". We are not in a position to comment on the strength of the evidence for a valuation premium. But we trust it will be evident that in our view, if there *is* a valuation premium, this does *not* reflect an incorrect assumption on the market cost of equity.

However, if there is a premium, and it is *not* due to the assumed market cost of equity being set at too high a level, this clearly begs the question: what explains the valuation premium?

To answer this question thoroughly would take us beyond the remit of the consultation exercise. But the most obvious alternative explanation – that was indeed pointed out in our earlier analysis for

Ofgem[24] is that it has arisen from Ofgem's decision to assume a value of the CAPM "beta" close to one, despite strong evidence that it is much closer to around one half, thus artificially raising the cost of equity.

It is worth noting that this alternative explanation is also consistent with the observation that valuation premia have been persistent phenomena. In contrast, if the valuation premium were attributable to the market cost of equity having fallen due to recent falls in real risk-free rates (as the CC has argued), the evidence would only have emerged relatively recently.

We would recommend that having completed this consultation exercise, Ofgem should focus its attention on these other criticisms of its methodology for setting the cost of equity capital. Having done so, it may well conclude that the assumed cost of equity capital for DNOs should be reduced. But it should do so for the right reasons, *not* in response to criticisms – which we believe are ill-founded - of its methodology for setting the assumed market cost of equity.

---

[24] See Stephen Wright, Robin Mason, Meltem Baskaya, Ken Hori and Steve Satchell, *Report on the Cost of Capital* September 2006 (http://www.ofgem.gov.uk)