Estimation of the market risk premium and its relationship to the risk free rate in the context of regulation of electricity and gas energy networks: A report to the Australian Energy Regulator Consumer Reference Group

Glenn Boyle
Director, Sapere Research Group
Christchurch, New Zealand

Kieran Murray
Managing Director, Sapere Research Group
Wellington, New Zealand

February 25, 2022

1 Introduction

1. In this report, we discuss, examine and assess various methods proposed for estimation of the market risk premium (MRP), with particular reference to the Australian electricity and gas sectors that are supervised by the Australian Energy Regulator (AER). We summarise and review many issues that have arisen in past AER determinations (e.g., AER, 2018), and some that have not. Our aim is to provide an independent perspective on MRP estimation questions relevant to the AER and the Consumer Reference Group (CRG).

2. The approach we follow is based on first principles questions: what exactly is the MRP, why is it important to regulators, what constraints does that importance place on regulator choices and objectives, and how can theory and evidence from financial economics shed light on the best mechanisms for achieving these objectives? In attempting to answer these questions, we encounter many issues familiar from past MRP reviews, plus a few that are often ignored or glossed over. Although we attempt to make recommendations where it is possible to do so, our approach is not overly-prescriptive as, unfortunately, MRP theory and evidence have not evolved to the point where this is possible: the relevant theory is sometimes vague or based on unrealistic assumptions and the necessary data are often limited or unavailable or yield estimates that seem implausible. Instead, we focus on identifying the tradeoffs involved in attempting to resolve the various issues.

3. In the next section, we define and interpret the MRP, from both an economic and statistical perspective. We also distinguish between unconditional and conditional forms of the MRP, their implications for estimation, and the different information they provide to regulators. Sections 3–7 outline and assess the various methods proposed for empirical estimation of the MRP: the justifications and rationale for each, their implementation, their pros and cons, and, to the extent possible, offer

recommendations. Section 9 provides some concluding remarks.

2 MRP basics

4. The MRP has an important and ubiquitous role in financial economics. In asset pricing theory, it frequently arises as a pricing kernel; in investments it serves as a guide to asset allocation decisions and as a performance benchmark; and in corporate finance it is often fundamental to the choice of investment hurdle rate.

5. For regulators, the MRP is important because of its central role in determining the allowed, or target, rate of return for regulated entities. Many regulators (including the AER) use a version of the Sharpe-Lintner Capital Asset Pricing Model (CAPM) to estimate the cost of equity component of the allowed return. This states that the expected return on equity for entity i, $E[R_i]$ (i.e., the cost of i's equity funding) is given by:

$$E[R_i] = R_f + \beta_i\{E[R_m] - R_f\} \tag{1}$$

where R_f is the riskless rate of interest, R_m is the rate of return on the market portfolio of risky assets, and β_i is asset i's beta, i.e., the sensitivity of i's equity returns to market portfolio returns.[1]

6. In equation (1), MRP is the name given to the term $E[R_m] - R_f$, and so the CAPM can be re-expressed as:

$$E[R_i] = R_f + \beta_i\{MRP\} \tag{2}$$

[1] In a CAPM world, R_f is both a beginning-of-period yield and the end-of-period realised return. Such equivalence need not hold in real world applications in which the riskless asset proxy has a different maturity to the investment horizon of interest.

which highlights the importance of MRP to regulators: in order to compute entity i's cost of equity via the CAPM, a necessary condition is that they know MRP.[2]

7. In practice, MRP (and β_i) are unknown and so must be estimated in some way. Regulators therefore estimate entity i's cost of equity $\hat{E}[R_i]$ as:

$$\hat{E}[R_i] = R_f + \hat{\beta}_i\{M\hat{R}P\} \tag{3}$$

where a ˆ denotes an estimate. In words, regulatory estimates of an entity's cost of equity equals the riskless rate of interest (typically able to be observed via a proxy such as a government bond yield or return) plus the product of (i) an estimate of the entity's beta and (ii) an estimate of the market risk premium.[3]

8. It is important to distinguish between the market risk premium MRP $\equiv E[R_m] - R_f$ and the expected market return ERM $\equiv E[R_m]$, i.e.:

$$MRP = ERM - R_f$$

This relationship suggests two alternative approaches to estimating MRP. First, as encapsulated in (3), it can be estimated directly. Second, it can be estimated indirectly by first estimating ERM and then subtracting the riskless interest rate (or some suitable proxy) to obtain an estimate of MRP. In the latter case:

$$M\hat{R}P = E\hat{R}M - R_f$$

and

$$\hat{E}[R_i] = R_f + \hat{\beta}_i\{E\hat{R}M - R_f\}$$

[2]The market risk premium is sometimes denoted E[MRP] in order to emphasize its forward-looking nature. Because the market risk premium is by definition the mean of future excess returns, the E[.] component is, in our view, redundant, i.e., $MRP \equiv E[R_m - R_f]$, so $E[MRP] = MRP$.

[3]In this report, we assume the riskless rate proxy is readily observable, and so problems in estimating the cost of equity emanate solely from estimation of MRP (and β_i).

9. In this report, we shall consider both the direct and indirect approaches to estimating MRP and their implications for cost of equity estimates. But before turning to specific methods that are available for implementing these approaches, it is useful to consider some fundamental economic and statistical properties of MRP and ERM as these properties help inform assessment of the estimation methods.

2.1 Economic interpretation of MRP and ERM

10. Equation (2) states that the cost of equity for entity i equals the riskless rate of interest plus an additional term equal to the product of β_i and MRP. Alternatively, we can rewrite (2) as:

$$E[R_i] - R_f = \beta_i\{MRP\} \tag{4}$$

which states that the difference between entity i's cost of equity and the riskless rate of interest is proportional to the MRP.

11. The economic interpretation of terms like $E[R_i] - R_f$ is that of a *risk premium* — the additional *expected* return required by the marginal investor as compensation for being exposed to the risk of asset i. From a consumer's perspective, if entity i is a regulated network, $E[R_i] - R_f$ is the compensation he/she effectively has to pay (in terms of retail prices) to the network for assuming the risk of providing service.

12. Similarly, MRP is the additional *expected* return required by the marginal investor as compensation for being exposed to the risk of the market portfolio. Intuitively, asset prices are set so that the marginal investor in risky asset i obtains the expected return compensation that make him indifferent between holding that asset and the riskless asset. Because the market portfolio is simply a linear combination of all the individual risky assets, this indifference criterion also applies to the

market portfolio.

13. Equation (4) thus states that the risk premium applicable to every asset i is proportional to the MRP, where the factor of proportionality is the fraction of market risk to which asset i is exposed, i.e., β_i. For example, if $\beta_i = 0.5$, then asset i has half as much risk as the market portfolio and hence i's risk premium is half that of MRP. This illustrates the importance of MRP to all regulators who employ the CAPM: the risk premium applicable to all regulated entities is simply a fraction of MRP.

14. In principle, MRP can be positive, negative or zero. If investors are *risk-averse* (i.e., require compensation for bearing risk), then MRP is strictly positive; if they are *risk-neutral* (i.e., indifferent towards risk), then MRP is exactly zero; if investors are *risk-loving* (i.e., are willing to offer compensation for bearing risk), then MRP is strictly negative. Ever since the pioneering 17th century work of Bernoulli (1954), economists have typically assumed that, at least when it comes to the selection and pricing of risky assets, investors display risk-averse behaviour and hence MRP must be strictly positive.[4]

15. In settling on MRP, the task for regulators is thus to identify a positive-valued number that reflects the additional expected return investors require over and above the riskless rate of interest in order to willingly hold the market portfolio and its associated risk. Importantly, this is the same for all investors. Potential investors in regulated networks may believe that the regulator chooses an MRP that differs from the true MRP, but this has no effect on the expected return they require on their network investments as it is the true MRP that determines their opportunity cost of capital. Put another way, an investor who is considering

[4]When defined with respect to a long-term riskless bond whose value fluctuates over its life, MRP represents the difference in expected returns on two risky assets and so could become negative in the short term (but not over the period corresponding to the maturity of the bond).

investing in a regulated network will take the regulator's likely estimate of the MRP into account when forecasting future cash flows from the investment, but the benchmark against which these expected cashflows are measured (the opportunity cost of capital) is a function of the true MRP only.

16. The ERM story is similar. If investors are risk averse, they will require a positive expected return as compensation for foregoing the riskless zero return obtainable from leaving their money under the mattress.

17. In settling on ERM, the task for regulators is thus to identify a positive-valued number that reflects the total expected return investors require in order to willingly hold the market portfolio and its associated risk.

18. Unfortunately, investor expectations are not directly observable and hence neither are MRP and ERM. Moreover, theoretical models link MRP and ERM to variables like risk and investor risk preferences which are also not observable. As a result, estimates of MRP and/or ERM must be inferred indirectly, typically from data. This necessitates consideration of the statistical interpretation of MRP and ERM.

2.2 Statistical interpretation of MRP

19. In the CAPM, the riskless interest rate is known and certain over the time period covered by the MRP, so $E[R_f] = R_f$. Thus:

$$\begin{aligned} MRP &= E[R_m] - R_f \\ &= E[R_m - R_f] \qquad (5a) \\ &= E[R_e] \qquad (5b) \end{aligned}$$

which states that MRP is the expected value, or mean, of the realised *excess* market portfolio return $R_e \equiv R_m - R_f$. Similarly, ERM $\equiv E[R_m]$ is the expected value, or mean, of the realised *total* market portfolio return R_m. As MRP = ERM

$-R_f$, it follows that MRP can also be represented as the mean of R_m minus the observable value of R_f.

20. Estimating MRP and ERM thus involve estimating the mean of a probability distribution: of the R_e distribution when estimating MRP, and of the R_m distribution when estimating ERM.

21. Estimating probability distribution parameters (such as the mean) require an estimator, or method, for generating the estimate. In general, a "good" estimator is unbiased and has low variance. That is, it generates estimates that are right "on average" and place tight bounds on the range of possible estimates, i.e., do not systematically differ from the true value and have a low standard error.[5]

22. Unfortunately, the ex-post unobservability of MRP and ERM creates challenges in applying these statistical criteria. Once the MRP for the next year is set at date t, all we subsequently observe is the realised return at data $t+1$. But as Gu et al. (2020) point out:

    > "...market efficiency forces return variation to be dominated by unforecastable news that obscures risk premiums."

    That is, actual returns are primarily influenced by events that were unanticipated at the date the MRP was set and so provide little information about what that MRP actually was. In other words, we cannot in 2022 say "Ah hah, with the benefit of hindsight we now know that the true MRP in 2018 was X": actual returns in the intervening period are dominated by unforecastable shocks that drown out the ex ante MRP. Thus, except in special cases, estimators of MRP and ERM are unverifiable.

[5] Alternatively, estimators can be compared using, for example, the root mean square error statistic. This tends to be highly correlated with the standard error, at least for unbiased estimates.

23. Faced with this difficulty, it seems reasonable to also assess MRP estimators according to their *regulatory* quality. Three criteria stand out. First, from a regulator's perspective, an estimator must be feasible. That is, the information needed to generate the estimate is both available and reliable, and the method is reasonably straightforward to apply. Second, if both consumers and networks are risk averse, they will dislike large swings in prices: such swings make it difficult for consumers to budget and for networks to confidently invest. Thus, MRP or ERM estimators that generate relatively stable estimates of the cost of equity over time, without big changes from review date to review date, are likely to be preferred by regulators to estimators that do not have this property, all else equal. Third, estimators of MRP or ERM should, as far as possible, be consistent with the underlying pricing framework being used, which for most regulators is the CAPM. Thus, an MRP estimator that, directly or indirectly, implies or involves a phenomenon or relationship that does not appear in the CAPM should be viewed with some suspicion by regulators. If an empirical regularity is not captured by a model, and that regularity is believed to be important, then the correct solution is to update the model so that it does capture the regularity; attaching an ad-hoc adjustment to the existing model runs a high risk of increasing, rather than decreasing, estimation error.

2.3 Conditional vs unconditional MRP and ERM

24. The observation that MRP (ERM) is the mean of the future excess (total) returns distribution requires distinguishing between unconditional and conditional means, and hence between the unconditional and conditional MRP.[6] In practical terms for regulators, the conditional MRP can be thought of as the time-varying risk

[6]Our discussion of this matter is designed to capture the essential intuition rather than the mathematical rigour.

premium applicable solely to the next review period; the unconditional MRP is (roughly) the constant risk premium applicable on average to a large number of future review periods.

25. To illustrate the difference between unconditional and conditional means, consider a weather example. Over a year the average daily temperature in Sydney is approximately 19 degrees Celsius, but is 22 degrees in December. So if on 1 January you wish to forecast Sydney temperatures over the rest of the year, 19 degrees is the best prediction. But if you wish to do so on 1 December, then 22 degrees is a better prediction. The former is the *unconditional* mean or forecast temperature; the latter is the *conditional* forecast that varies with the time of year.

26. Much discussion of whether the MRP is constant or time-varying is essentially asking whether there are "seasons" in excess stock returns and thus whether the unconditional MRP differs from the conditional MRP. The answer depends on what form of stochastic process best describes returns. Our discussion of this point initially focuses on excess returns R_e for concreteness, but exactly the same principles apply to total returns R_m.

27. If excess stock returns are independently and identically distributed (iid) through time, then:

$$R_{et} = \mu + \epsilon_t$$

where μ is a constant and ϵ is a zero-mean random variable. That is, excess returns at each date equal a constant value (μ) plus a "surprise". Then:

$$MRP = E[R_{et}] = \mu$$

is the same at all dates t.

28. In an iid world, excess returns fluctuate unpredictably around a fixed value and so the MRP is a time-invariant constant (equal to μ); moreover, the unconditional

and conditional MRPs are equivalent. However, much evidence suggests that, at least in US data, returns are predictable and that the MRP is therefore time-varying, e.g., LeRoy and Porter (1981), French et al. (1987), Campbell and Shiller (1988), Fama and French (1988), Cochrane (2008, 2011), Chen et al. (2013), Martin (2017). This casts doubt on the validity, and practical utility, of the iid assumption.[7]

29. A less restrictive (than iid) assumption is that excess returns follow a *stationary* process: the conditional mean (and possibly other parameters) changes over time, but eventually converges back on a long-run (unconditional) value, i.e., time variation is transitory rather than permanent, as in the weather seasons example of para 25. In this case:

$$R_{et} = \mu_{t-1} + \epsilon_t$$

where $E[\mu_{t-1}] = \mu$. So the conditional mean of R_{et} is $E_{t-1}[R_{et}] = \mu_{t-1}$, which depends on date $t-1$ conditions and so varies over time. But the unconditional mean is $E[R_{et}] = E[\mu_{t-1}] = \mu$.

30. A plausible story for why the MRP might vary over time is that it is affected by the business cycle, being high when economic conditions are depressed and investors require significant compensation in order to take on more risk, and low when economic conditions are buoyant and investors face a low marginal cost of saving. As Cochrane (2013) puts it:

> "December 2008 was a recent time of low price/dividend ratios. Is it not plausible that the average investor, like our endowments, said, 'sure, I

[7] The extent of return predictability is uncertain, due to weak statistical properties (e.g., Goyal and Welch, 2008) and apparent variation across countries (Rangvid et al., 2014). But as Cochrane (2008) forcefully points out, observed variation in dividend-price ratios must predict either future cash flows or future returns, and the evidence for the former is much weaker than for the latter. Thus, return predictability seems to be an empirical fact even if its magnitude is not.

know stocks are cheap, and the long-run return is a bit higher now than it was. But they are about to foreclose on the house, repossess the car, take away the dog, and I might lose my job. I can't take any more risk right now.' Conversely, in the boom, when people 'reach for yield', is it not plausible that people say 'yeah, stocks aren't paying a lot more than bonds. But what else can I do with the money? My business is going well. I can take the risk now."

31. Many discussions of the MRP observe that MRPs clearly vary through time and conclude from this that estimating a "long-run" MRP is inaccurate, or at least unhelpful. As explained in para 24, the truth is more subtle. If excess returns are stationary, there exists both a *conditional* MRP (that reflects current conditions such as the state of the business cycle) and an *unconditional* MRP (that describes the premium required over the "long-run"). Thus, at least in principle, regulators can choose between two alternative MRPs when excess returns are stationary (but not iid).

32. Finally, excess returns could follow a *non-stationary* process. In this case, time variation in the mean (and possibly other parameters) is permanent rather than transitory and trends upwards or downwards over time rather than reverting to a long-run value. As a result, a conditional mean exists, but the unconditional mean does not and the best estimate of the MRP is the current excess return.

33. Figure 1 illustrates the difference between a stationary and non-stationary series. In the top picture (stationary), the mean may vary over time, but there also appears to be a long-run value. In the bottom picture (non-stationary), by contrast, the mean clearly trends down, especially in the latter third of observations.

34. The lack of theoretical and empirical support for non-stationarity in stock returns means we can safely ignore the possibility when thinking about how to

Figure 1: Stationary vs Non-Stationary Time Series

best estimate MRP. Nevertheless, for completeness we occasionally refer to any implications that non-stationarity might give rise to.

35. Everything in paras 24-34 also applies to ERM. Whether R_m is iid, stationary or non-stationary determines the available set of means, and therefore the options available for estimating ERM.

36. To summarise, if excess (total) returns are iid, then there is a single MRP (ERM) that holds at all points in time; if excess (total) returns are stationary but not iid, then there is both an unconditional/long-run and a conditional/current MRP (ERM); if excess (total) returns are non-stationary, then there is only a conditional

MRP (ERM) that varies through time and does not revert to any long-run value. Which property holds in practice determines what MRP and/or ERM choices are available to regulators. If excess (total) returns are non-stationary, then attempts to estimate a constant "long-run" MRP (ERM) are futile and wrong, but stationarity implies that both a constant long-run/unconditional and a time-varying current/conditional MRP (ERM) are potentially available to regulators.

37. Which MRP estimate is more relevant to regulators? On the one hand, the MRP of direct interest to regulators is the conditional MRP, since this reflects market risk pricing conditions at the date the allowed return is set and thus provides networks with the right incentives. On the other hand, real world imperfections and practical constraints in *estimating* the conditional MRP could imply a regulatory preference for "looking through" short-run changes in MRP and instead estimating the MRP that holds in the "long run", i.e., the unconditional MRP. In this case, the MRP is chosen to provide basic compensation to networks.

38. Four reasons for relying on the unconditional MRP stand out. First, methods for estimating the conditional MRP may be infeasible or unreliable. Second, time variation in the MRP may reflect irrational under- and over-pricing, not rational risk pricing. Third, use of the conditional MRP, if variable enough, might induce large swings in the allowed return (see para 23). Fourth, in the case of the AER, the MRP is set for four years and any attempt to impose a conditional MRP that is correct today will, by definition, be incorrect for a network facing a new determination in, say, 3.5 years time; if the true conditional MRP has in the interim switched from above to below, or below to above, the unconditional MRP, use of the MRP set 3.5 years ago will be less accurate than the unconditional MRP. Thus, regulators can be faced with choosing between a poor estimate of the right MRP and a good estimate of the wrong MRP.

39. Throughout this report, we are careful to distinguish between unconditional and conditional MRPs (and their corresponding ERM and cost of equity counterparts). Different estimation approaches differ in whether they offer unconditional or conditional estimates and this distinction is not always clear in MRP discussions. In general, a necessary condition for feasible estimation of the conditional MRP is that current observed variables can reliably predict future returns, while estimates of the unconditional MRP rely on (testable) assumptions about the nature of the distribution of excess returns. Neither can be ruled out *a priori* from a regulator's perspective and both potentially have roles to play.

2.4 Real vs nominal MRP and ERM

40. Our discussion to date has discussed MRP and ERM without mentioning whether we are referring to nominal or real (i.e., inflation adjusted) returns. This is because the issues raised are equally applicable to both.[8]

41. In general, the same continues to be true throughout this report — the criteria by which different methods assess the various approaches to estimating ERM and MRP apply regardless of whether it is nominal or real values being considered. Of course, the practical tradeoffs in using one or the other may differ. For example, real returns may appear more "stable" than nominal returns, thus enabling greater estimation precision under some approaches. But because the final allowed return must be expressed in nominal terms, working with real returns introduces the practical problem of having to reconvert to nominal by using expected inflation forecasts, which are often inaccurate (e.g., Kliesen, 2015) and are highly variable across sources (e.g., Verbrugge and Zaman, 2021).[9]

[8]The same applies to before- and after-tax returns.

[9] In practice, AER uses current inflation as its forecast of future inflation, consistent with inflation following a random walk. For some evidence supporting this view, see, for example, Canova (2007).

42. Thus, in our terminology, the total market return R_m will be used generically to refer to both the nominal return and the real return, i.e., the nominal return minus an adjustment for inflation over the same period. Similarly, the expected market return ERM will be used generically to refer to both the expected nominal return and the expected real return, i.e., the expected nominal return minus an adjustment for expected inflation over the same period.

43. For excess returns R_e and the market risk premium MRP, the issue is redundant since any inflation adjustment should affect both stock and bond returns equally, leaving the difference unaffected. Thus, the nominal and real values of R_e and MRP are equivalent.[10]

44. Armed with these basic principles of MRP and ERM estimates, we now turn to a description and assessment of the methods most commonly used to obtain such estimates.

3 Estimating the MRP: Method I – Historical Averaging of Excess Returns ("Historical-MRP")

3.1 Description and justification

45. The traditional method for arriving at an estimate $\hat{MRP}$ of the market risk premium is the historical averaging procedure pioneered by Ibbotson and Sinquefield

[10]Strictly speaking, the nominal return factor (1 + nominal rate of return) should be deflated by the inflation factor (1 + rate of inflation) to obtain the real return factor, which would drive a wedge between nominal and real excess returns and between nominal and real MRP. However, the difference will be small in most cases. More importantly, all our subsequent analysis can easily accommodate this adjustment and so can be thought of as applying to either nominal or real values.

(1976), which we refer to as the **Historical-MRP Approach**. As its name suggests, this involves using a data sample of length T from a broad-based stock market index (such as the ASX 200), together with a proxy for riskless asset returns over the same sample period, to compute:[11]

$$M\hat{R}P = (\frac{1}{T})\sum_{t=1}^{T}(R_{mt} - R_{ft}) = (\frac{1}{T})\sum_{t=1}^{T}R_{et} \equiv \overline{MRP} \tag{6}$$

In words, (i) observe historical *excess* returns for each date in a sample consisting of T periods (usually years), and (ii) calculate the average of this series. For example, if data on excess returns are available annually from 1951 to 2020 (70 years), then MRP can be estimated by a simple average of those 70 excess returns.

46. Depending on the investment horizon, either short-term government bills or long-term government bonds can be used to proxy for the riskless asset in the calculation of excess returns. The n-year review period typically employed by regulators implies that an n-year bond is more appropriate in regulatory settings.[12]

47. Essentially, the Historical-MRP Approach assumes that the past is a reliable guide to the future, and is justified as a "forward-looking" estimate (i.e., a forecast of the future) by the following set of conditions:

    - Stationary (or iid) excess returns, i.e., the past distribution of excess returns is the same as the future distribution.

[11]The use of a stock market index for this purpose is discussed further in paras 117-119

[12]A more general version of the CAPM requires only a zero-beta asset, not a riskless one. However, a government bond with maturity matched to the regulatory review period will (approximately) have this property. Another minor complication, not addressed here, involves the choice of end-of-year bond returns or beginning-of-year bond yields when proxying for the riskless rate. The former is more consistent with the use of stock returns, but yields are more readily available and are more commonly used in typical applications of the CAPM. Views are varied: for example, Bishop et al (2018) and Lally (2019) recommend the use of yields, but Damodoran (2021) and Dimson et al (2002) advocate for returns.

- The distribution of excess returns follows what is known as an an ergodic process, i.e., the sample average converges to the unconditional mean in large samples.[13]
- The market is informationally efficient as in Fama (1970), i.e., the market of investors perceive the true distribution of returns and so historical excess returns fluctuate randomly around the true mean.

Historical-MRP in long-period samples

48. Together, the three conditions in para 47 ensure that, at least in large samples, the sample average will approximate the true mean of excess returns, i.e., the MRP. Thus, at a minimum, regulators considering employing the historical average approach should use the longest-available sample of data and test for the stationarity of excess returns in that sample, and probably for ergodicity as well. If these can be rejected, the Historical-MRP Approach is invalid.

49. The Historical-MRP Approach is sometimes subject to criticism on the grounds that a considerable body of evidence indicates that MRP varies through time and is not constant (see para 28). But this ignores the distinction between conditional and unconditional MRPs. The statistical justification for $\overline{MRP}$ is as an estimate of the *unconditional* MRP that, by definition, does not vary over time. If for some reason a regulator is interested in the unconditional MRP, then the observation that the *conditional* MRP varies through time is irrelevant. Of course, if the regulator is actually more interested in the conditional MRP (as would typically, but not always, be the case — see paras 37-38), then the relevant question is whether the Historical-MRP Approach is capable of providing sufficiently accurate estimates of the conditional MRP to be useful for that purpose. We consider this point at various points below.

[13] See Anonymous (2019) for a more formal and detailed, but still relatively succinct, description of ergodicity and its relation to the Law of Large Numbers.

Historical-MRP in short-period samples

50. Unfortunately, in shorter-period samples intended to reflect current conditions, historical averages can provide very misleading estimates, e.g., Elton (1999) notes that there have been periods in US data where $\overline{MRP}$ has been implausibly high and other times where it has been implausibly low.[14] Also, Damadoran (2021, Table 5) finds that $\overline{MRP}$ estimates obtained over even a relatively long time period (25 years) have standard errors that equal or exceed the estimates themselves, even in developed markets such as Canada, France, Germany and the UK.[15] In other words, the estimates are statistically indistinguishable from zero, which is unlikely to be very useful to a regulator.

51. This reflects a fundamental property of $\overline{MRP}$ that precludes its use over short time periods. Suppose in the course of a year that the conditional MRP declines. This increases stock prices and hence R_e is high, which in turn raises (slightly) $\overline{MRP}$. So the estimate of MRP (i.e., $\overline{MRP}$) rises exactly when the true (conditional) MRP falls. This is effectively a reversal of the return predictability evidence of Campbell and Shiller (1988) and Cochrane (2008, 2011). In that body of research, a series of high actual returns is evidence of high *expected* returns in the past. When applying Historical-MRP to a short time series, a series of high actual returns is assumed to be evidence of high expected returns in the future. The former reflects thinking like an economist; the latter reflects naive extrapolation from the past.

52. Figure 2 illustrates this point and the resultant dangers involved in attempting to use Historical-MRP over a short period. In this (extreme) example, the true

[14]While time variation in the conditional MRP would be expected, negative estimates, or estimates exceeding 20%, seem unlikely to be accurate.

[15]Australia does not appear in the table, but other data suggest it is similar — see para 54.

conditional MRP is initially 3%, but gradually declines to 1% over 20 years.[16] This drives stock prices up, and therefore returns. As a result, the historical average of excess returns (Historical-MRP) gradually *rises* over the 20-year period, from 3% to over 7%. Attempting to use historical averaging to estimate the conditional MRP is a fool's errand: when the true conditional MRP falls, the estimated conditional MRP rises, and vice versa. In the long run, this balances out, but not in the short run.[17]

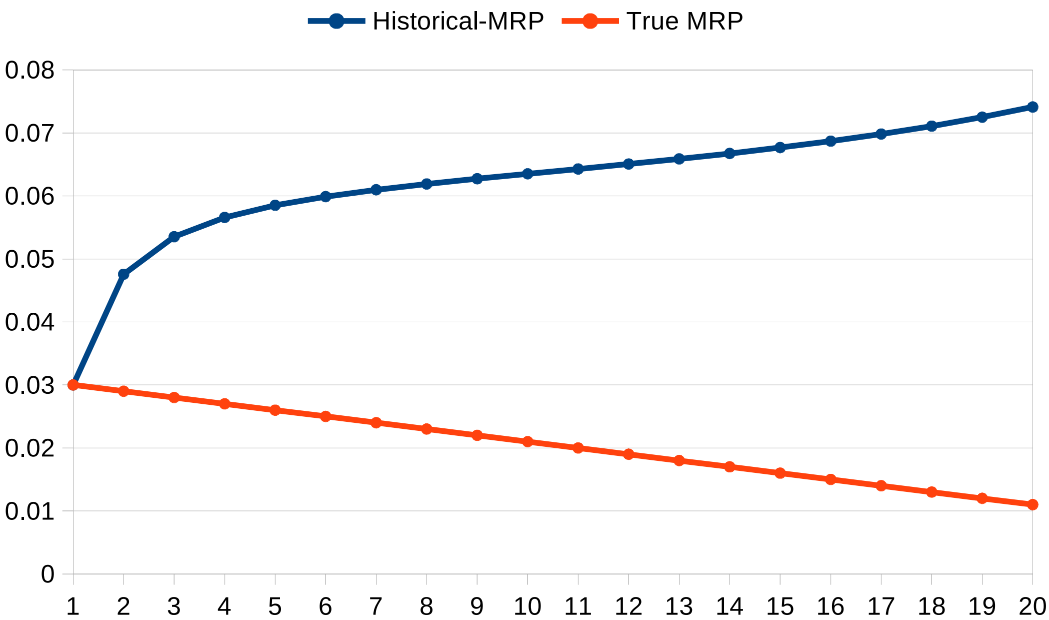

Figure 2: Divergence between True and Historical MRP

[16]The Dividend Growth Model — see section 5 — is used to calculate prices and returns. The initial dividend is set equal to \$1, the dividend growth rate to 2%, and R_f to 2%.

[17]The sole exception to this rule arises in the unlikely event that excess returns are non-stationary. In such a case, the best estimate of a future MRP is the current excess return. If the objective is to obtain the MRP prevailing over the next 5 years, then "current' could legitimately be interpreted as the average excess return over the last 5 years.

Acceptable sample period length for Historical-MRP

53. This begs a related question — how long does the time period have to be for $\overline{MRP}$ to be a precise estimate of the unconditional MRP and thus potentially useful to regulators? "Very" seems to be the answer, given the observed volatility in stock returns. To illustrate, consider Table 6 in Damodoran (2021), taken from Dimson et al. (2018), that provides estimates of MRP obtained from the 117-year period 1900–2017. For Australia, $\overline{MRP} = 6.6\%$ with a standard deviation of 18.1%. So the standard error is $18.1/\sqrt{117} = 1.7\%$ and the 95% confidence interval for $\overline{MRP}$ is [3.3%, 9.9%]. Even 117 years of data is insufficient to obtain a very precise estimate.

54. Over a shorter period of 30 years, the corresponding confidence interval is [0.1%, 13.1%], which is so wide as to be of little value for practical purposes. If plus or minus 2 percentage points (i.e., a confidence interval of [4.6%, 8.6%]) was deemed to be the minimum-acceptable level of precision, then 315 years of data would be required![18]

AER implementation of Historical-MRP

55. AER (2018, p252) reports that it uses several data samples of differing length, including at least one quite short one:

> "Estimates from all five periods should be considered. While the longer periods are likely to be more statistically robust, the most recent period of 1987 onwards is most likely to provide an estimate commensurate to the current market."

56. We are sceptical about such an approach. Historical averaging is justified only over long sample periods where the sample average converges to the *unconditional*

[18]Setting the confidence bounds to ±2% requires $\frac{18.1}{\sqrt{T}} = \frac{2}{1.96}$, which yields $T = 315$.

mean. No corresponding small-sample property exists for the *conditional mean*, so stating that post-1987 data provides an estimate "commensurate with" the current MRP has no basis in fact. As para 50 points out, historical averaging can be very misleading over short time periods. An MRP estimate obtained from a sample of recent data that produces a lower average but with a high standard error (which renders the recent average statistically indistinguishable from the long-sample average) is essentially meaningless — the lower average is almost certainly due to sampling variation, not to any fundamental change in the MRP.

Historical-MRP summary

57. To reiterate:

- Historical averaging has desirable statistical properties only in a large sample covering a long time period, assuming the underlying distribution is stationary. In that case, it provides an estimate of the unconditional or "long-run" MRP.
- In general, historical averaging over short time periods cannot provide an accurate estimate of the conditional or "current" MRP except by sheer chance. The only exception occurs in the unlikely case where R_e is non-stationary, which would then rule out the use of historical averaging over a long period.
- If the excess returns distribution is stationary and an estimate of the current MRP is required, then either (i) use the long-MRP estimate (and hope that time variation in the conditional MRP is not too great) or (ii) disregard historical averaging in favour of an approach that estimates the current MRP directly. Applying historical averaging to a short sample period is not the answer.

3.2 Implications for the regulated cost of equity

58. Recall that MRP is important to regulators because of its contribution to the cost of equity for every regulated entity. If $\overline{MRP}$ is used to estimate MRP, then the estimated cost of equity for entity i is calculated using the CAPM as:

$$\begin{aligned} \hat{E}[R_i] &= R_f + \hat{\beta}_i\{\hat{MRP}\} \\ &= R_f + \hat{\beta}_i\{\overline{MRP}\} \end{aligned} \tag{7}$$

which, because $\overline{MRP}$ is a constant, has the property that cost of equity estimates move 1-for-1 with the riskless rate R_f. That is, if R_f rises by x percentage points, then (ignoring any possible changes in β_i) $E[R_i]$ also rises by x percentage points. Similarly, if R_f falls by y percentage points, then (again ignoring any possible changes in β_i) $E[R_i]$ also falls by y percentage points.

59. As we discuss in section 3.6, there may be good reasons for scepticism about such a simple deterministic relationship. Moreover, if the true conditional MRP is negatively related to R_f, then cost of equity estimates obtained from (7), which assume a constant MRP, have an unattractive feature. If R_f is below its long-term mean, then the cost of equity is set too low (relative to its conditional value); if R_f is above its long-term mean, then the cost of equity is set too high. While these errors will even out in the long run (so long as R_f is stationary), this may not be a property that can be relied on by regulators. For example, if R_f is above or below its long-run mean for extended periods of time, then the cost of equity estimated for a given set of network assets may be persistently too high or too low over a sizeable chunk of assets' lives.

60. This problem arises because equation (7) combines an estimate of the *long-run* (i.e., unconditional) MRP with a proxy for the *current* riskless interest rate, and

so all variation in the cost of equity is driven by changes in the riskless interest rate.

3.3 A simple adjustment: Historical-MRP-$\mathbf{R_f}$

61. One solution is, when undertaking cost of equity calculations like (7), to replace the current riskless interest rate with its long-term mean. That is:

$$\hat{E}[R_i] = \overline{R_f} + \hat{\beta}_i\{\overline{MRP}\} \tag{8}$$

where $\overline{R_f} = (\frac{1}{T})\sum_{t=1}^{T} R_{ft}$ is the average value of R_f in a sample of size T, i.e., the long-run sample average of R_f.

62. This Historical-MRP-R_f approach for estimating the cost of equity has some obvious advantages over its standard counterpart. First, it would greatly reduce regulatory uncertainty for both networks and consumers as $\hat{E}[R_i]$ would be very stable over time. Second, its mechanical nature would greatly reduce the need for lengthy and expensive discussions about the appropriate cost of equity.

63. Of course, the use of (8) to estimate the cost of equity also requires that R_f be stationary and ergodic so that the unconditional mean exists and is approached by the sample average in a large sample. Although the steady decline in interest rates over the last 30 years might appear to make this unlikely, Campbell and Viceira (2002, p24) point out that interest rates have shown no evidence of long-term trends over the last 200 years. Like other stationarity questions, the answer can only be obtained empirically. Until then, the Historical-MRP-R_f cost of equity estimate described by (8) should not be dismissed out of hand.

64. Even if bond returns pass the necessary tests, the Historical-MRP-R_f approach might also be objected to on the basis that it takes no account at all of current market conditions and that a partly right estimate of the current cost of equity

(i.e., one that varies through time with interest rates) is better than one that is completely wrong (i.e., one that doesn't vary through time at all). While an intuitively appealing argument, it isn't necessarily correct, as we now demonstrate.

Cost of equity estimation error: Historical MRP vs Historical-MRP-R_f

65. Suppose a regulator wishes to estimate the conditional/current cost of equity, but has only the historical average $\overline{MRP}$ available and hence must choose between the Historical-MRP and Historical-MRP-R_f approaches, i.e., between equations (7) and (8). Under the former approach (equation (7)), the estimation error (the difference between the cost of equity estimate and its true value) is:[19]

$$\begin{aligned} \text{Historical-MRP estimation error} &= \text{estimated cost of equity - true cost of equity} \\ &= (R_f + \beta_i(\overline{MRP})) - (R_f + \beta_i(MRP)) \\ &= \beta(\overline{MRP} - MRP) \end{aligned} \tag{9}$$

while for the Historical-MRP-R_f approach (equation (8)) the corresponding error is:

$$\begin{aligned} \text{Historical-MRP-}R_f \text{ estimation error} &= \text{estimated cost of equity - true cost of equity} \\ &= (\overline{R_f} + \beta_i(\overline{MRP})) - (R_f + \beta_i(MRP)) \\ &= (\overline{R_f} - R_f) + \beta(\overline{MRP} - MRP) \end{aligned} \tag{10}$$

66. Comparing (9) and (10), we can see that the former error will tend to be greater (in absolute value) than the latter if MRP and R_f are negatively correlated, i.e., the riskless rate is above (below) its long-run average at the same time the MRP is below (above) its long-run average. In such a case, the two sources of error in (10) offset each other and so the overall error is lower.

[19] We assume β_i is known.

67. To illustrate, suppose the true (conditional) MRP is 7.5%, the current riskless rate is 2%, the long-run average MRP is 6%, the long-run average riskless rate is 2.5%, and the β is 0.5. Then the true (conditional) cost of equity is:

$$E[R_i] = 2\% + 0.5(7.5\%) = 5.75\%$$

The cost of equity estimate based on equation (7) is:

$$\hat{E}[R_i] = 2\% + 0.5(6\%) = 5\%$$

and the cost of equity estimate based on equation (8) is:

$$\hat{E}[R_i] = 2.5\% + 0.5(6\%) = 5.5\%$$

The equation (7) error is 0.75 percentage points, compared to only 0.25 percentage points when using equation (8), despite the latter taking no account of current conditions. This arises because the positive error in the riskless rate estimate when using (8) partly offsets the negative error in MRP estimates. Thus, the Historical-MRP-R_f approach partly mitigates the problem with the standard approach described in para 59.

68. However, it would be dangerous to push this point too far because, as we show in the Appendix, a negative correlation between MRP and R_f is necessary but not sufficient for the Historical-MRP-R_f approach to produce smaller average errors than the standard Historical-MRP approach. Specifically, this occurs if and only if

$$b_{MRP} < \frac{-1}{2\beta}$$

where b_{MRP} is the sensitivity of MRP to R_f shocks, i.e., the change in MRP associated with a 1 percentage point change in R_f. In 2018, AER set $\beta = 0.6$, so the above condition becomes:

$$b_{MRP} < -0.83$$

That is, the Historical-MRP-R_f approach produces lower average errors for the cost of equity only so long as every percentage point decline in the riskless interest rate is associated with a more than 0.83 percentage point increase in the MRP. This requires that changes in R_f and MRP be close to offsetting each other.

69. The intuitive explanation for this requirement is that if R_f and MRP have a strong negative relationship (i.e., $b_{MRP} < -0.83$), then the error in the estimated cost of equity from using the standard Historical-MRP approach (which assumes $b_{MRP} = 0$) is sufficiently great that an approach that makes no adjustments at all for interest rates (the Historical-MRP-R_f approach) generates smaller average errors. If the true MRP-R_f relationship is weaker, then the reverse is true, i.e., ignoring movement in the riskless rate generates larger estimation errors on average..

3.4 Historical-ERM Approach

70. An alternative statement of the CAPM is:

$$E[R_i] = \beta_i R_f + (1 - \beta_i)E[R_m] \tag{11}$$

which expresses the expected return required on i's equity as a weighted sum of the riskless rate and ERM, where the weights are determined by β_i. This makes explicit the two-fund separation property of the CAPM: all assets i can be replicated by, and hence must have the same expected return as, a so-called "tracking portfolio" with a proportion β_i of its wealth invested in the market portfolio and the remaining proportion 1-β_i invested in a *single* risk-free (or zero-beta) asset.[20] The usual version of the CAPM (i.e., equation (1)) is just an algebraic rearrangement of this property.

[20]For detailed discussions on separation theorems, see, for example, Huang and Litzenberger (1988) or Brennan (1989).

71. Motivated by the possible problems with Historical-MRP noted in para 59, some regulatory participants have suggested a historical averaging approach based on (11) where averaging is applied to total market returns R_m instead of excess returns, thereby obtaining an estimate of "long-run" ERM. That is:

$$E\hat{R}M = (\frac{1}{T})\sum_{t=1}^{T} R_{mt} \equiv \overline{ERM} \tag{12}$$

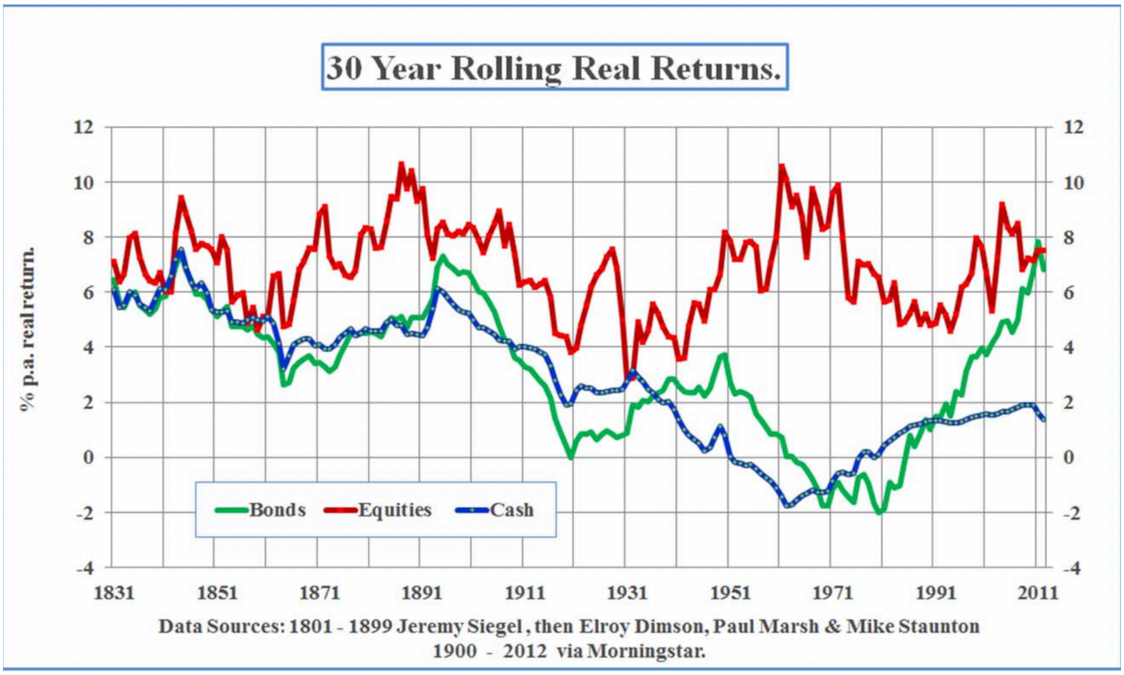

Figure 3: Stability of Total Stock Returns (red line)

72. This calculation, which we call the Historical-ERM Approach, is used by UK regulators and is typically argued for on the basis of pictures like Figure 3 that depict long-run "stability" of total US stock returns relative to bond returns.[21]

73. However, such pictures are akin to Venezuelan elections, in that the alternative candidate (excess returns) is not shown. Moreover, historical averaging of R_m is justified if and only if total returns satisfy formal tests of stationarity and ergodicity, not just that they look "stable". If stationarity and ergodocity cannot be rejected, then the sample average of R_m (i.e., $\overline{ERM}$) in large samples will approximate the true (unconditional) ERM.

[21]See, for example, Wright and Smithers (2013).

74. Just as a constant MRP underpins the Historical-MRP approach, a constant ERM does the same for the Historical-ERM approach. How plausible is the assumption of a constant expected total return? Referring to nominal returns, Brealey et al. (2020) claim that (p338) "Investors are not likely to demand the same return each year on an investment in common stocks". Similarly, Lally (2005) states that assuming ERM to be time-invariant is "untenable". Even for real returns, it might seem unlikely that investors would require the same expected return on risky assets in the midst of a depression as during the height of a boom all else equal.

75. However, all else may not be equal: interest rates are likely to be low during the depression and high during the boom, thus offsetting any cyclical movements in risk premia. As a result, a near-constant ERM cannot be ruled out.

76. In any event, a time-varying *conditional* expected return does not preclude the existence of a constant, "long-run", or *unconditional* expected return. If R_m data satisfy stationary and ergodic requirements, then $\overline{ERM}$ will provide a reliable estimate of the *unconditional* market return in large samples. Again though, small samples utilising short time series are of no value in estimating the short-run, or *conditional*, market return. As with excess returns, historical averages of total returns can provide very misleading estimates when computed over shorter time periods. And for the same reasons. Suppose in the course of a year that the conditional ERM declines, i.e., discount rates fall. This increases prices and hence R_m is high, which in turn raises (slightly) $\overline{ERM}$. So the estimate of R_m (i.e., $\overline{ERM}$) rises exactly when the true (conditional) ERM falls, and vice versa.

77. Recall that

$$M\hat{R}P \equiv \hat{E}[R_m - R_f]$$

Since the Historical-ERM Approach sets $\hat{E}[R_m] = \overline{ERM}$, its estimate of the MRP is:[22]

$$\hat{MRP} = \overline{ERM} - R_f \tag{13}$$

which, in contrast to $\overline{MRP}$, varies in an inverse 1-to-1 manner with R_f. That is, according to the Historical-ERM approach, if R_f rises by x percentage points, then the MRP estimate falls by x percentage points. Similarly, if R_f falls by y percentage points, then the MRP estimate rises by y percentage points. The Historical-ERM approach thus generates an estimate of the conditional MRP, where all time variation is attributable to variation in R_f.

78. The behaviour of MRP and ERM estimates under the Historical-MRP and Historical-ERM approaches is essentially orthogonal. Under the Historical MRP approach, $\hat{ERM}$ goes up and down with R_f while $\hat{MRP}$ is independent of R_f; under the Historical-ERM approach, $\hat{MRP}$ varies inversely with R_f while $\hat{ERM}$ is independent of R_f.

3.5 Implications for the regulated cost of equity

79. If $\overline{ERM}$ is used to estimate ERM, then the estimated cost of equity for entity i is

$$\hat{E}[R_i] = (1 - \hat{\beta}_i)R_f + \hat{\beta}_i\overline{ERM} \tag{14}$$

80. Because $\overline{ERM}$ is a constant, the Historical-ERM approach has the property that cost of equity estimates move in proportion $(1 - \hat{\beta}_i)$ with the riskless rate R_f. That is, if R_f rises by x percentage points, then $E[R_i]$ rises by $(1 - \hat{\beta}_i)x$ percentage points. Similarly, if R_f falls by y percentage points, then $E[R_i]$ falls by

[22] If $\hat{E}[R_f]$ is equal to the long-run average $\overline{R_f}$, then the MRP under the Historical-ERM approach is equal to $\overline{MRP}$, as in para 61. To be consistent with the usual implementation of the Historical-ERM Approach, we set $\hat{E}[R_f] = R_f$, where R_f is defined to be the current riskless interest rate (or some deterministic function of it).

$(1 - \hat{\beta}_i)y$ percentage points.[23] For any β_i between 0 and 1, the estimated cost of equity is thus less sensitive to changes in R_f under Historical-ERM than under Historical-MRP (see para 58).

Cost of equity estimation error: Historical-ERM vs Historical-MRP

81. At first glance, Historical-ERM appears to have an attractive feature relative to its Historical-MRP counterpart. Suppose R_f is below its long-term mean. All else equal, this lowers the current (conditional) cost of equity. However, if R_f and MRP are negatively correlated (a possibility we discuss in section 3.6), this direct effect of lower R_f is partly offset by the indirect effect of a higher MRP. The Historical-ERM approach captures this indirect relationship while the Historical-MRP approach does not.

82. It might, therefore, be tempting to conclude that the Historical-ERM is preferred because it generates lower cost of equity estimation errors. However, it turns out this is not always the case.

83. To see why, return to the example in para 67, but suppose that the true conditional MRP is 6.2% rather than 7.5%. Then, the true (conditional) cost of equity is:

$$E[R_i] = 2\% + 0.5(6.2\%) = 5.1\%$$

With $\overline{MRP} = 6\%$ and a long-term riskless interest rate of 2.5%, the implied $\overline{ERM}$ is 2.5% + 6% = 8.5%, and the cost of equity estimate based on the Historical-R_m approach is:

$$\hat{E}[R_i] = 0.5(2\%) + 0.5(8.5\%) = 5.25\%$$

The cost of equity estimate based on the Historical-MRP approach is:

$$\hat{E}[R_i] = 2\% + 0.5(6\%) = 5\%$$

[23] We assume that changes in β_i are unrelated to R_f.

In this case, the Historical-MRP error is -0.1 percentage points while the Historical-ERM error is 0.15 percentage points, which is larger in absolute value. This occurs because the true MRP-R_f relationship is relatively minor and so the relationship assumed by the Historical-ERM approach over-corrects the estimate.

84. This reflects a general property. As we show in the Appendix, the Historical-MRP estimation error variance minus the Historical-ERM estimation error variance is proportional to:

$$\frac{1}{2} - b_{ERM}$$

where b_{ERM} is the sensitivity of ERM to R_f shocks. That is, the Historical-ERM approach produces lower errors on average for the cost of equity so long as every percentage point decline in the riskless interest rate is associated with a less than 0.5 percentage point decline in ERM. Intuitively, this applies because the Historical-MRP approach assumes $b_{ERM} = 1$ while the Historical-ERM approach assumes $b_{ERM} = 0$; an actual b_{ERM} value of less than 0.5 is closer to the latter assumption and so that assumption yields lower estimation errors on average. If $b_{ERM} > 0.5$, then the reverse is true.

85. In the Appendix, we also show that the Historical-MRP-R_f estimation error variance minus the Historical-ERM estimation error variance is strictly positive for any $\beta_i \leq 1$. That is, for regulated networks, Historical-ERM is almost certain to produce lower estimation errors on average than Historical-MRP-R_f.

Cost of equity estimate stability: Historical-ERM vs Historical-MRP

86. We can also compare Historical-MRP and Historical-ERM in terms of their ability to generate "stability" in cost of equity estimates over time. As discussed in para 23, both consumers and networks are likely to prefer, all else equal, that the cost of equity not vary greatly from review date to review date.

87. To undertake such a comparison, we compute the difference in the variance of estimates under the two approaches. Let $\hat{\phi}_M^2$ denote the variance of cost of equity estimates for entity i when the Historical-MRP approach is used and $\hat{\phi}_E^2$ denote the equivalent when the Historical-ERM approach is used. In the Appendix, we show that

$$\frac{\hat{\phi}_E^2 - \hat{\phi}_M^2}{\hat{\phi}_M^2} = \beta_i(\beta_i - 2) \tag{15}$$

from which it follows that Historical-ERM generates less volatile cost of equity estimates than Historical-MRP so long as $0 < \beta_i < 2$, which seems extremely likely for regulated networks.[24]

88. The intuition for (15) is that Historical-ERM, in contrast to Historical-MRP, allows for time variation in the MRP estimate, which in turn introduces two additional sources of volatility in the estimated cost of equity. Changes in MRP have (i) a direct impact on the estimated cost of equity in proportion to β_i^2; (ii) an inverse indirect effect (negative correlation between MRP and R_f) in proportion to $2\beta_i$. The first term dominates when β_i is sufficiently large (or negative).

89. In its 2018 determination, the AER set beta equal to 0.6. From (15), this implies that Historical-ERM estimates of the cost of equity would have just 16% of the intertemporal volatility of their Historical-MRP counterparts.[25]

AER assessment of Historical-ERM

90. AER (2018) dismisses the use of Historical-ERM (which they refer to as the Wright approach) on three grounds: (i) that there is no theoretical justification for a constant expected return, (ii) that the perfect negative correlation with R_f

[24] Because it uses long-run averages of both R_f and MRP, the Historical-MRP-R_f estimate exhibits approximately perfect stability over time.

[25] If volatility is measured in terms of standard deviation rather than variance, this factor would equal 40% ($0.4^2 = 0.16$).

is extreme, and (iii) the approach is not used by Australian practitioners. On the first point, there is, equally, no theoretical justification for a constant MRP either. On the second point, we agree that a perfectly negative correlation is probably unrealistic, but then so is the Historical-MRP alternative of a zero correlation. As we discuss in the next subsection, there is not, and largely cannot, be any definitive evidence on the existence and magnitude of this correlation. On the third point, the observation that practitioners do not use a particular model is not in and of itself evidence against it (just as their use of a model is not evidence for it) — what matters is how well it conforms to the data when compared to alternatives.

3.6 Correlation between R_f and MRP

91. In our discussion of the properties of the Historical-MRP and Historical-ERM approaches, the correlation between R_f and MRP (and/or ERM) has played an important role. The different assumptions about these correlations are summarised in Table 1 (where the Historical-MRP-R_f approach is also included for completeness) in terms of b_x coefficients — the sensitivity of x to a 1-unit change in R_f.

Table 1: Sensitivity of $\hat{MRP}$ and $\hat{ERM}$ to R_f under the Historical-MRP, Historical-MRP-R_F and Historical-ERM Approaches

	Historical MRP	Historical ERM	Historical MRP-R_f
b_{ERM}	1	0	0
b_{MRP}	0	-1	0

Notes: b_{MRP} is the sensitivity of $\hat{MRP}$ to R_f shocks, i.e., the change in the MRP estimate associated with a 1 percentage point change in R_f. b_{ERM} is the corresponding sensitivity for the expected return estimate $\hat{ERM}$.

92. Historical-MRP and Historical-ERM are inked by the following simple relationship:[26]

$$b_{ERM} = 1 + b_{MRP} \tag{16}$$

93. The important question is: where on this continuum does the truth lie? Is it at one of the extremes (i.e., either Historical-MRP or Historical-ERM) or is it somewhere in between?

What does theory say about the MRP-R_f correlation?

94. The MRP and R_f are both financial market prices and hence are determined endogenously as functions of a potentially-large number of factors, e.g., investor sentiment, perceived risk, tradeoff between current and future consumption, monetary and fiscal policies, and so on. Asset pricing models attempt to distill these effects into a simple and approximate form.

95. As has been noted by a number of authors (e.g., Friend and Blume, 1975; Merton, 1980; Huang and Litzenberger, 1988; Boyle, 2006), MRP is not a free parameter in the CAPM and instead must satisfy:

$$MRP = \gamma \times \sigma_e^2 \tag{17}$$

where γ is the average risk aversion of all investors holding the market portfolio and σ^2_{MRP} is the variance of excess market returns R_e. Both are exogenous, fixed, paramteters in the CAPM, as is R_f, so a correlation between MRP and R_f cannot arise within the CAPM framework. This implies $b_{MRP} = 0$ and $b_{ERM} = 1$.

96. However, similar equations to (17) also arise in more general asset pricing models such as the Intertemporal CAPM (Merton, 1973) and the Consumption CAPM (Breeden, 1978). In these models, shifts in R_f can affect both risk and risk

[26] See the Appendix for proof.

aversion in a variety of ways. As a result, standard asset pricing models are unable to provide a definitive guide on the MRP-R_f relationship and leave open the possibility of a non-zero correlation.

97. There would seem to be at least two good reasons to suspect a negative correlation between MRP and R_f. First, conventional monetary policy typically forces interest rates down when economic conditions are weak. Such conditions are also likely to be characterised by high risk and/or high investor risk aversion, resulting in higher risk premia. Similarly, interest rates are typically high when economic conditions are buoyant and risk premia are low. This suggests that intertemporal movements in R_f are likely to be at least partly offset by opposite movements in MRP and hence the cost of equity estimate moves by *less than* 1-for-1 with the riskless rate, i.e., $b_{MRP} < 0$ and $b_{ERM} < 1$.

98. Second, any phenomenon that causes a portfolio rebalancing between stocks and bonds, and hence drives stock and prices in different directions, must induce a negative relationship between MRP and R_f. For example, an adverse banking or financial shock frequently induces a so-called "flight to safety" where investors sell stocks and buy government bonds. This forces riskless bond yields down at the same time risk premia rise.

Empirical evidence on the MRP-R_f correlation

99. What do the data say about b_{ERM} and/or b_{MRP}? Ang and Bekaert (2007) find some evidence of a negative relationship between that short-term interest rates and future excess returns, but only over short horizons; whether long-term bond yields predict the 5-year horizon returns of relevance to regulators is not addressed. Campbell and Thompson (2008) report that long-term government bond yields have only low predictive power for future excess returns, but this at least in part reflects the disproportionate effect of unanticipated events on realised returns (see

para 22).

100. A more recent study by Harris and Marston (2013) takes a different approach to estimating b_{ERM}. Instead of asking whether bond yields can predict future returns, they ask whether bond yields can explain ERM estimates obtained from the Dividend Growth Model (DGM, which we discuss in more detail in section 5). Specifically, they investigate the relationship between 3-5 year estimated ERMs and 30-year bond yields. Across a variety of specifications and data sets, they consistently estimate b_{ERM} to be in the 0.3-0.4 range.[27] Since such estimates are closer to 0 than 1, they suggest that Historical-ERM is likely to be closer to reality than Historical-MRP ((i.e. $b_{ERM} < 0.5$) and thus support use of the former.

101. Using Australian data, CEPA (2021) follow a similar approach and obtain estimates of b_{ERM} ranging from 0.53 to -0.26. Although highly variable, these estimates confirm that R_f is negatively correlated with DGM estimates of MRP.

102. However, there is an important problem with the Harris and Marston (2013) and CEPA (2021) approaches that makes it difficult to set too much store by their results. Specifically, their approach implicitly assumes the MRP is given by

$$M\hat{R}P_I = MRP + \eta$$

where $M\hat{R}P_I$ is the MRP estimate implied by the DGM and μ is an error term with zero mean that is uncorrelated with MRP. That is, that MRP estimates obtained from the DGM fluctuate around the true MRP in an unbiased and random manner. As we discuss in section 5, this may or may not be true. If it is not true, and there is no evidence either way, then the negative relationship observed by Harris and Marston and CEPA could just reflect a negative correlation between R_f and η. Unfortunately, we have no way of ruling this out.

[27] Harris and Marston actually estimate b_{MRP}, which we convert to b_{ERM} using equation (16).

103. The fundamental problem with attempts to empirically estimate the MRP-R_f correlation is that the true MRP (or ERM) is not observable, even ex-post, which precludes usual methods for estimating economic relationships statistically. Instead, proxies for MRP (or ERM) must be used. One is actual, or realised, returns. Unfortunately, as already mentioned, realised returns reflect far more than just their expectation at the beginning of the period and so are a very noisy proxy indeed. For example, suppose that today (date 0), the true MRP is 4% and that in a year's time it is 5%. However, because of unanticipated events in the course of each year, actual annual returns (at dates 1 and 2) may bear little or no relationship to these expected returns.

104. The other possible proxy is implied-MRP from an asset pricing model such as the DGM, which is the approach adopted by Harris and Marston (2013) and CEPA (2021). However, precisely because the true MRP is unobservable, it is impossible to test the validity and accuracy of such models.

105. Tellingly, the CEPA (2021) estimates have some odd features. First, their two principal DGM-implied estimates imply that ERM is negatively related to R_f. This seems unlikely. Second, in alternative regressions they use realised returns as a proxy for MRP. This yields a much weaker negative relationship between MRP and R_f overall, but one that appears to become stronger after 1993 when inflation targeting was introduced, just as the discussion in para 97 suggests would be expected. By contrast, no post-1993 difference is observable when using the DGM proxy. These quirks suggest that CEPA's estimates are a doubtful guide to the true MRP-R_f correlation.

106. Overall, absent some methodological breakthrough, we are pessimistic about the potential for empirical analysis to provide reliable and accurate estimates of b_{ERM} and b_{MRP}. The choice between Historical-MRP and Historical-ERM seems un-

likely to be resolved in this way.

3.7 Overall: Historical-MRP vs Historical-ERM vs Historical-MRP-R_f

107. The Historical-MRP and Historical-ERM approaches both seek to forecast expected future market returns: excess returns in the case of the former and total returns in the case of the latter. Historical-MRP thus estimates the MRP and Historical-ERM estimates the expected market return.

108. If the underlying excess and total return series are stationary and ergodic (and the stock market is efficient), then both approaches will produce estimates that converge on their true values if sufficient historical data are available.[28]

> Before using either Historical-MRP or Historical-ERM, suitable tests of stationary and ergodicity should be undertaken.

109. Both Historical-MRP and Historical-ERM generate reliable estimates of unconditional means in large samples. Using a subset of recent data in an attempt to better estimate current (conditional) means can yield extreme values and is not recommended.

> Historical-MRP and Historical-ERM should be applied to the largest sample of returns data available. They should *not* be applied to short time series unless returns are believed to be non-stationary.

[28]For US evidence of total and excess return stationarity, see, for example, Shapiro and Horner (2014).

110. Either approach can yield lower average estimation errors for the cost of equity, depending on the correlation between expected market returns and the riskless interest rate.

If the relationship between expected market returns and the riskless interest rate is above a critical value, then Historical-MRP yields estimation errors that are lower on average. If the relationship between expected market returns and the riskless interest rate is below the same critical value, then Historical-ERM does.

111. Either approach can yield cost of equity estimates that are more stable over time, depending on the beta of the entity under consideration.

If the entity's beta exceeds 2 (or is negative), then Historical-MRP yields less volatile cost of equity estimates. If the entity's beta is positive but less than 2, then Historical-ERM does.

112. In cost of equity calculations, Historical-MRP-R_f replaces R_f with the same interest rate proxy used in estimating $\overline{MRP}$, i.e., the average interest rate observed over a long time period. This results in cost of equity estimates that are very stable over time, but have relatively large estimation errors on average.

113. Finally, when it comes to allowed regulatory returns, the choice of historical estimator to apply to vital parameters such as MRP and ERM is possibly less important than a commitment to stick with whatever approach is chosen. Incentives for regulatory parties to engage in opportunistic behaviour by lobbying for self-interested changes in approach over time would then be negated.[29]

[29] Networks would prefer Historical-MRP when interest rates are high and Historical-ERM when

Regulators should choose an estimation approach and commit to standing by it through multiple regulatory cycles. Any switch to a different approach should require compelling new evidence and, in the presence of uncertainty about implementation, be gradual.

114. Our discussion and assessment of the various approaches, in particular Historical-ERM, has been implicitly in terms of nominal returns. One thing changes if real returns are used instead: an adjustment for expected inflation must be made prior to computing the cost of equity. Since expected inflation is not observable, this introduces an additional source of estimation error, and possibly bias — see para 9. This could potentially change the rankings in Table 2 below. We do not pursue this matter here, but it should be kept in mind.

The choice of nominal vs real returns involves a tradeoff. Although real total and excess returns are likely to be less volatile, and possibly better fit stationarity criteria, than their nominal counterparts, their usage necessitates that expected inflation be estimated, which is not necessarily a simple task.

115. Table 2 summarizes the properties and rankings of the various historical estimation approaches:

---

interest rates are low; consumers would prefer the opposite.

Table 2: Summary Rankings for the Historical-MRP, Historical-MRP-R_f, Historical-ERM, and Historical-Smart Approaches

	Historical MRP	Historical ERM	Historical MRP-R_f
Assumptions	Stationarity/ergodicity of R_e	Stationarity/ergodicity of R_m	Stationarity/ergodicity of R_e and R_f
Statistical properties (given assumptions)	Unbiased (in large samples)	Unbiased (in large samples)	Unbiased (in large samples)
Evidence	Unknown	Unknown	Unknown
Cost of equity estimation error	1 or 2	1 or 2	3 (likely)
Cost of equity intertemporal stability	3 (almost certainly)	2 (almost certainly)	1
Regulator consistency with status quo	1	2=	2=
Long-run cost	2 =	2 =	1

Notes: 1 denotes top-ranked, 2 denotes 2nd-ranked, and 3 denotes bottom-ranked.

3.8 Additional issues common to both Historical-MRP and Historical-ERM approaches

116. The various historical-averaging approaches are also subject to additional issues or questions that affect all of them but do not alter the choice between them. So before turning to other potential methods for estimating MRP or ERM, we first briefly consider these extra matters.

3.8.1 ***Market portfolio isn't known***

117. In the CAPM, the market portfolio is the portfolio of *all* risky assets, including so-called non-traded assets such as works of art, vintage cars and other durable goods, real estate, and, most ubiquitously of all, human capital. However, because the exact composition of, and the returns on, this wider portfolio are unknown,

standard practice typically employs a broad-based stock market index to proxy the market portfolio, e.g., the ASX 200.

118. Such a procedure implicitly assumes either (i) the portfolio of non-traded assets is sufficiently small relative to the portfolio of stock market assets that the former's omission makes little difference to the true MRP or (ii) returns on the portfolio of non-traded assets are highly correlated with those on the portfolio of stock market assets so that they command a similar risk premium. Since (i) seems untenable, that leaves (ii). Such a requirement is by no means guaranteed. For example, Boyle and Guthrie (2004) find that the correlation between labour income and stock returns is close to zero in all 11 countries for which the necessary data are available.

119. In the CAPM, the expected return on the true market portfolio exceeds that on the stock market portfolio proxy if and only if the variance of true market portfolio returns is greater than the variance of stockmarket returns — see equation (17). On the one hand, much lower liquidity suggests the variance of non-traded asset returns is likely to be higher than that of stock market returns, which increases the variance of true market portfolio returns (relative to stock market returns). On the other hand, imperfect correlation between stockmarket and non-traded returns lowers the variance of true market portfolio returns. Thus, the stockmarket proxy can over- or under-estimate the true market portfolio return.

3.8.2 ***World or domestic market portfolio?***

120. The need to use a stock market index to proxy for the market portfolio raises another question arises — what index? In particular, is the appropriate index domestic (e.g., ASX 200) or global (e.g., FTSE All-World)?[30]

[30] This section is adapted from Boyle and Murray (2021).

121. As is widely recognised, the theoretical answer is that it depends on whether individual-country capital markets are *integrated* (i.e., effectively form one global market) or *segmented* (i.e., are many separate markets, each with their own pricing factors). If the former, then the appropriate market portfolio proxy is a world index, because all investors hold that portfolio in an international version of the CAPM; if the latter, a domestic index is more suitable.

122. The question of whether world capital markets are integrated or segmented has long been a matter of debate amongst researchers. As Ibbotson et al. (1982, p82) put it almost 40 years ago:

> Since international investment occurs, markets cannot be totally segmented. But, interest rates and equity returns ... appear to differ substantially from country to country. We view the world market as partly segmented and partly integrated.

123. Despite considerable liberalisation of capital markets in the following decades, Boyle (2009) reaches much the same conclusion 27 years later, as, more recently still, does Orlowski (2020) for the European Union. Moreover, attempts to deepen financial market integration continue to interest policymakers (e.g., European Central Bank, 2020), suggesting they continue to see a not-insignificant amount of segmentation.

124. Even if stock markets are largely integrated, business and interest rate cycles can still diverge, often substantially, and a domestic index is likely to be more correlated with domestic economic conditions than a world index. Thus, use of a domestic index may also be preferred for this reason.

125. A domestic index also better meets any desire for simplicity and transparency. World index returns are calculated in a specific currency, so an exchange rate adjustment will typically be required when used in a different country. Moreover,

there are a range of extant world indices, all constructed in different ways, so a choice would have to be made between them; by contrast, the choice of domestic index is usually more restricted and its construction easily understood. Finally, a world index can only roughly replicate the tax situation faced by investors in different countries.

126. There is also little evidence that the use of a world market index (and the associated international CAPM) is any more accurate than the simpler domestic approach. For example, Harris et al. (2003) find that the domestic and international CAPMs fit US data equally well and hence conclude that the choice of domestic or world index is largely immaterial. Koedijk et al. (2002) and Bruner et al. (2008) arrive at a similar conclusion from analyses of developed and emerging countries.

127. Overall, given the lack of empirical support for the view that use of a world market index better explains observed stock market returns, the greater simplicity (and consistency with the imperfect integration of global capital markets) afforded by a domestic index should be preferred.

3.8.3 ***Arithmetic or geometric average?***

128. In equations (6) and (12), we obtain historical estimates of MRP and ERM respectively using what are known as *arithmetic* average returns, i.e., excess or total returns observed for each date in the sample period are added together and then divided by the number of dates. An alternative approach is to compute *geometric* average returns, as follows:

$$M\hat{R}P = \prod_{t=1}^{T}(1 + R_{et})^{1/T} - 1 \tag{18}$$

$$E\hat{R}M = \prod_{t=1}^{T}(1 + R_{mt})^{1/T} - 1 \tag{19}$$

With geometric averaging, the T return factors are first multiplied together and then the T-th root is taken. The operation is thus non-linear whereas arithmetic averaging is linear.

129. An important distinction between arithmetic and geometric averages is that, for the same set of data, the former always exceeds the latter, with the difference being approximately equal to half the variance of returns.[31] Thus, in regulatory hearings, regulated networks might be expected to prefer the arithmetic average and consumer representatives the geometric average.

130. The answer as to which of arithmetic or geometric averaging is better depends on the context in which the question is asked. If the objective is to assess performance over some historical period, then the geometric average is superior. To see why, consider a \$100 investment whose value rises to \$200 a year later (100% return) before falling back to \$100 after another (-50% return). At the end of two years, the average annual (and the total) return is obviously 0%. Applying the arithmetic and geometric calculations, we get:

$$\text{Arithmetic Average} = \frac{100\% + -50\%}{2} = 25\%$$

$$\text{Geometric Average} = \sqrt{(1 + 100\%)(1 - 50\%)} - 1 = 0\%$$

and so the arithmetic average grossly overstates the true value, which equals the geometric average.

131. However, the context of interest for a regulator tasked with setting an allowed return is different. The intention is not to measure performance over some known past, but instead to estimate the mean of a future return that takes some unknown path. That is, we wish to estimate the probability-weighted *sum* of all possible future returns. This requirement favours the arithmetic average, for the following reasons.

[31] See, for example, Mindlin (2011).

132. First, as shown by Mindlin (2011), if returns are serially uncorrelated (as in Cochrane, 2014, p34), the arithmetic average converges to the true expected return in large samples, whereas the geometric average does not.[32] Similarly, Mehra and Prescott (2008) show that the future value of an investment computed at the arithmetical average return tends to the expected value of the investment. Brealey et al. (2018) argue that the arithmetic average return correctly estimates the ERM (or MRP) but that the geometric average understates it, leading them to conclude:

> "If the cost of capital is estimated from historical returns or risk premiums, use arithmetic averages, not (geometric averages)."[33]

133. The assumption of zero autocorrelation in returns is questionable at longer horizons (see, for example, Fama and French, 1988). When returns are autocorrelated, both average and geometric returns are biased estimators of the true MRP (or ERM), but Cooper (1996) finds that the bias is smaller for the former. However, Indro and Lee (1997) show that a weighted average of arithmetic and geometric averages is both a less biased and a more efficient estimator than either alone.

134. Thus, if the relevant returns data are iid, then the arithmetic average is the correct approach. If the data are serially correlated, then a combined estimate may be more appropriate.

---

[32] A simple example of this point, adapted from Brealey et al (2018), appears in the Appendix.

[33] Jacquier et al. (2005) show that the arithmetic average return can have a large bias if the period over which the MRP is being estimated is long relative to the data period available, but this is of little consequence to regulators so long as the 1-5 year return forecasts (in which regulators are typically interested) are obtained from historical data sets covering a large number of years, i.e., the large sample requirement is met.

3.8.4 ***Interim dividend reinvestment***

135. Conventions in the computation of market portfolio returns can result in these returns being over-stated. As a result, so will any estimates based on an average of these returns, whether arithmetic or geometric.

136. Fried et al. (2021) point out that typical calculations of market portfolio returns assume that interim (e.g., quarterly or semi-annual) dividends are reinvested in the market portfolio and hence earn the return R_m. As they point out, there are two problems with this. First, most dividends are not reinvested in the market portfolio. Second, it is not possible for investors as a group to do so because most reinvested shares have to be repurchased from other investors. Most investors must therefore reinvest dividends in lower-yielding assets and so do not earn the excess return quoted.

137. The obvious implication is that if market portfolio returns are over-stated, then so are any historical averages based on these returns, i.e., MRP and ERM. Fried et al. (2021) report that the upward bias can be substantial in US data, ranging from 1 to 3 percentage points in the MRP.[34] Clearly, if even just approximately correct, this could have major implications for cost of equity estimates in regulatory settings.

138. It would, therefore, seem appropriate for research to be undertaken determining the extent to which the Fried et al. (2021) effect is present in Australian data.

[34] If additional stockholder cash flows such as repurchases and issuances are included, the estimated bias is greater still.

3.8.5 ***Survivorship bias***

139. Survivorship bias is the name given to the phenomenon whereby returns calculated for indexes or markets overstate true performance because of the exclusion of failed companies or markets.

140. Various forms of survivorship bias have been proposed by researchers (see, for example, Rietz, 1988; Brown et al., 1995; Jorion and Goetzmann, 1999; Barro, 2006), but for the purpose of estimating MRP or ERM in a single market, two types stand out: we refer to these as ex-post and ex-ante survival bias.

141. Ex-post bias refers to the mechanical exclusion of failed companies from the computation of index returns, thereby biasing realised market returns, and hence their historical average, upwards. This can happen in two ways. One is that a company was originally in the market index, but dropped out after poor performance commenced. In this case, its negative returns after exclusion from the index are ignored. However, those prior to exclusion, which presumably were also poor, are not, and so so the bias may be small. Moreover, any ignored negative returns are counter-balanced by the exclusion of highly-positive returns from successful companies in their early-growth phase prior to the point where they become large enough to be included in the index.

142. The other possible form of ex-post bias is that a failed company never becomes large enough to be included in the index, in which case all of its (poor) returns are ignored. However, companies of this size have a minuscule weighting (even in aggregate) in the true market portfolio and so the amount of bias is likely to have only a tiny impact on historical-based estimates of MRP and ERM.

143. Ex-ante bias refers to the non-appearance of “disaster”-type events in the historical record when these were in fact accorded positive probability by investors in

setting prices and expected returns ex-ante. As Cochrane (2005, p461) puts it:

> "Think of the things that did not happen in the last 50 years. We had no banking panics, and no depressions; no civil wars, no constitutional crises; we did not lose the Cold War, no missiles were fired over Berlin, Cuba, Korea, or Vietnam. If any of these things had happened, we might well have seen a calamitous decline in stock values, and I would not be writing about the equity premium puzzle."

144. The idea underlying ex-ante bias is that the average return from any sample that does not include a "disaster" will be larger than the ex-ante expected return. A simple example illustrates this point. Suppose there are three possible stock market states of the world at all future dates: a 30% return with 5% probability, a 10% return with 94% probability, and a -90% return with 1% probability. The true expected return is $0.05 \times 30\% + 0.94 \times 10\% + 0.01 \times -90\% = 10\%$ However, the historical average over any long time period in which the -90% return does not eventuate is $(0.05/0.99) \times 30\% + (0.94/0.99) \times 10\% = 11\%$.

145. Unfortunately, investor expectations about disaster events are unobservable, so the practical extent of ex-ante survival bias is uncertain and empirical research is ambiguous. Li and Xu (2002) argue that the probability of "disaster" would have to be implausibly high to have any significant effect on MRP. Others find the bias to be substantial: using plausible parameter values, Barro (2006) calibrates a model in which ex-ante bias reduces the true MRP by over 2 percentage points. Also, van Binsbergen et al. (2020) estimate that the ex ante bias reduces the Historical-MRP estimate by about 1/3.

146. One possible solution to survivorship bias is to compute an "average of the averages", i.e., use the average historical-MRP from a range of countries, including those in which a "disaster" event occurred, e.g., Germany, Japan, Russia. How-

ever, countries that experienced a collapse in stock prices also tended to suffer a collapse in bond prices, so the effect on average excess returns is likely to be very small.

3.8.6 ***Overall assessment of additional issues in historical data estimates of MRP and ERM***

147. Interim dividend re-investment assumptions, reliance on historical average (if returns are not iid) and survivorship bias all point in one direction — that average historical returns over-state the true MRP and ERM.

148. While estimation of the empirical magnitude of these issues may ultimately turn out to be impractical, their existence implies that any historical average should be considered an upper bound on the unconditional value of MRP (or ERM), and therefore that any claims advocating additional allowances to the MRP (or ERM) over and above its historical average should be treated with considerable scepticism and caution.

4 Estimating the MRP: Method II — Surveys

149. Instead of inferring the expectations of past investors from historical data, an alternative approach to estimating MRP is to directly ask current investors about their current expectations. With this approach, a sample of individuals — private investors, market professionals, CFOs or academics — are asked what they think the market risk premium is over some specified time period. The responses are then summarised in some simple way (i.e., sample average or median) and reported as the "consensus" estimate of the current MRP.

150. Following this approach, the estimated cost of equity for entity i is:

$$\hat{E}[R_i] = R_f + \hat{\beta}_i\{M\hat{R}P_s\}$$

where $M\hat{R}P_s$ is the survey-based estimate of MRP.

151. The survey approach has obvious attractions. First, in contrast to the historical approaches, it provides a direct estimate of the current (conditional) MRP, and so could be expected to more accurately reflect current conditions. Second, because respondents are active investors, the estimate it provides should be a well-informed one. Third, in contrast to the historical approaches, it automatically incorporates any correlation between MRP and R_f and so avoids the need to estimate this separately.

152. However, there is also an obvious disadvantage — because individual responses are both anonymous and generally unremunerated, the incentives faced by those participating in surveys are weak. As a result, the presumption that survey responses reflect high-information expectations is doubtful.

153. Even when the incentives are stronger, survey forecasts do not seem to perform well. For example, Ince and Molodtsova (2017) report widespread biases in survey-based exchange rate forecasts. Similarly, Easton and Summers (2007) document a significant upwards bias in the returns forecast by surveyed analysts, suggesting these tell us more about "hoped-for" returns than true risk premia. As Easton and Summers note, the tendency of analysts' forecasts to be optimistic means that estimates of the expected rate of return based on such forecasts are likely to be higher than the true MRP and ERM.

154. Survey forecasts also seem to contain a rookie error. High current prices and returns reflect a fall in expected returns (discount rates), all else equal; low current prices and returns reflect a rise in expected returns. That is, MRP and ERM tend

to have a counter-cyclical property. But as revealed by Adam et al. (2017) and Greenwood and Shleifer (2014), survey forecasts exhibit precisely the opposite property: survey expected returns are higher following high realized stock market returns and in times of high price-dividend ratios, and vice versa. Similarly, Damadoran (2021) notes that survey forecasts of MRP were at their peak in 1999 during the height of the tech boom and then fell away after the market collapse, leading him to conclude:

> "(I)t is ... likely that these survey premiums will be more reflections of the recent past rather than good forecasts of the future."

155. In other words, stand-alone survey-based estimates of MRP seem to provide roughly the same information as a historical approach based on a recent sample of excess returns which, as emphasized in para 50 , is likely to be highly misleading.

156. AER (2018) suggest some (unspecified) weight should be given to surveys because they provide an additional, partly-independent source of MRP data. This suggests a possible role for surveys as part of a combined (with other estimates of the conditional MRP) estimator, a topic we discuss further in section 8.

Survey-based estimates of MRP and ERM are likely to contain considerable noise and possible bias, but may be useful when combined with other estimates.

5 Estimating the MRP: Method III — DGM-Implied

157. A third approach to estimating MRP is to infer it from current market prices. That is, determine what MRP (or ERM) is "implied" by observable stock prices. Using market prices to infer unobservable parameters has a venerable history in financial economics, perhaps most notably in the use of option prices to obtained implied estimates of stock volatility, e.g., Hull (2009, pp 296-297).

158. Obviously, applying such an approach in this context requires some model linking stock prices to the MRP, and the usual candidate for this purpose is the Dividend Growth Model (DGM), also sometimes known as the Gordon Growth Model due to its association with Gordon (1959).

159. The DGM proceeds by assuming that a stock's expected dividend growth rate (g) and expected return (ERM) are perpetual constants, i.e., whatever they are today, they then remain at that level forever. Under these assumptions, it is straightforward to show that the current market value P is given by:[35]

$$P = \frac{D(1+g)}{ERM - g}$$

where D is the current market dividend. This can be rearranged to yield:

$$ERM = \frac{D(1+g)}{P} + g \tag{20}$$

which "implies" the ERM consistent with current market prices and dividends. Because g is unobservable and must be estimated ($\hat{g}$), estimates for ERM and MRP are given by:

$$E\hat{R}M_I = \frac{D(1+\hat{g})}{P} + \hat{g} \tag{21}$$

[35]See, for example, Brealey et al. (2020)

which reveals that the ERM estimate implied by the DGM is a linear function of the contemporaneous dividend-price ratio, with intercept $\hat{g}$ and slope 1-$\hat{g}$. Also, since MRP = ERM - R_f:

$$M\hat{R}P_I = \frac{D(1+\hat{g})}{P} + \hat{g} - R_f \tag{22}$$

160. Equations (21) and (22) suggest a beguilingly simple way to estimate ERM and MRP: observe the current dividend-price ratio, gross it up a bit to allow for future dividend growth, and out pops an estimate of ERM or MRP. We call this the DGM-Implied Approach to estimating MRP (or ERM).

161. An attractive feature of the DGM-Implied approach is that, in contrast to the historical averaging approach, it takes account of current market and economic conditions via the dividend-price ratio term, i.e., as the dividend-price ratio moves up or down, so do the MRP and ERM estimates. That is, it provides an estimate of the conditional (current) MRP rather than the unconditional (long-run) MRP. In contrast to the survey approach, the predicted movement in MRP conforms with standard economics: a high price-dividend ratio is associated with low ERM and MRP, and vice versa. Also, similar to the survey approach but in contrast to the historical approaches, the DGM-implied approach automatically incorporates any correlation between MRP and R_f and so avoids the need to estimate this separately.

162. So the DGM-Implied approach is simplicity itself, requires only the estimation of g to implement, and yields an economically-intuitive estimate of the current MRP and ERM. It seems almost too good to be true. And of course it is.

163. One potential problem lies with the DGM itself. Any MRP value that is implied by a theoretical model calibrated to market data is only as good as the model itself. As already noted, the DGM assumes that dividends can be expected to grow at a constant rate for an extremely long time and that the discount rate

remains the same over the same extremely long time. Such firms (or in this case a market index) are not known to exist.[36]

164. This is by no means a fatal objection. As Friedman (1953) long ago pointed out, the validity of any theory is determined not by the realism of its assumptions but by the accuracy of its predictions. Like all models, the DGM is an approximation of reality, and so the MRP or ERM it implies can also only be an approximation to its true value. The relevant question is how good an approximation. In a recent critique, Heaton (2021, pp5-6) is pessimistic:

> "No academic research supports the use of the DGM as a reliable implied cost of capital model in practice. The reason is that the DGM's assumptions are too far from reality for any firm, even as an approximation. There is no known example of a firm whose dividends have grown at a constant rate over a very long time period and with a discount rate that has remained constant as well."

165. Even if the DGM is a good model of stock prices, in that it closely approximates how investors set prices in general, the data it is applied to may not be so helpful. Specifically, use of the DGM in the manner described above assumes that the market price P accurately reflects underlying economic fundamentals on the day (or average of days) that the DGM is applied. If market prices are too high relative to fundamentals, then the dividend-price ratio is downwardly-biased and the ERM (or MRP) is set too low; if market prices are too low relative to fundamentals, then the dividend-price ratio is upwardly-biased and the ERM (or MRP) is set too high.

[36] The DGM can be extended to the case where expected dividend growth is temporarily high before reverting to a long-run constant value — see Damadoran (2021, pp85-86) — although this increases mathematical complexity without resolving the fundamental problem. But see para 175.

166. In short, attempting to infer ERM and MRP from current market prices (using the DGM or any theoretical model) assumes a very pure form of market efficiency: not merely that "you can't beat the market" (markets set prices in a way that don't deviate systematically from fundamental values), but also that "the market is always right" (markets set prices to always equal fundamental values). While there is very broad professional support for the former hypothesis, the reverse is true for the latter — see, for example, Thaler (2014).[37]

167. This might not be much of a problem if ERM and MRP were relatively insensitive to the dividend-price ratio. But inspection of (21) reveals that ERM moves with D/P on a greater than 1-to-1 basis. For example, suppose the true D/P is 0.04, the observed D/P is 0.05 (i.e., prices are below what is justified by fundamentals), and g is known to be 0.05. Then the true ERM (assuming the DGM is true) is

$$0.04 \times 1.05 + 0.05 = 0.092$$

but the DGM-implied ERM is:

$$0.05 \times 1.05 + 0.05 = 0.1025$$

which is too high by more than 1 percentage point. Because D/P varies a lot in practice (see para 177 below), and so can potentially deviate from fundamental value by a significant amount, use of the DGM to estimate ERM and MRP risks large deviations from their true values.

168. This reflects a further general problem with the DGM. In contrast to, say, the CAPM, it is not a model of the risk-return tradeoff reflecting market equilibrium

[37]Of course, this problem could be somewhat mitigated by applying DGM-Implied not to just a single date but to a range of dates. For example, the implied ERM and MRP could be estimated on a daily/monthly/quarterly basis for, say, the last five years and then averaged to obtain a single estimate. This would help reduce errors created by market mis-pricing, but would also weaken the supposed "forward-looking" characteristic of the DGM-Implied approach.

or the absence of arbitrage opportunities. Instead, it is simply a mathematical identity made tractable by the simplifying assumptions of perpetually constant expected returns and dividend growth. As a result, it makes no distinction between expected return movements due to rational risk pricing and those due to irrational shocks to investor sentiment (irrational exuberance or gloom). The former are relevant to the regulator's task, but the latter are not — and the DGM is unable to distinguish between them.

169. If the DGM is a good model and the market exhibits super-efficiency, is the DGM-Implied approach home free? Unfortunately not, as g still has to be estimated. This is important because inspection of (21) reveals that ERM also moves with g on a greater than 1-to-1 basis. For example, suppose the true and observed D/P is 0.04, the true g is 0.04, and the estimated g is 0.05. Then the true ERM is

$$0.04 \times 1.04 + 0.04 = 0.0816$$

but the DGM-implied ERM is:

$$0.04 \times 1.05 + 0.05 = 0.0.092$$

which is too high by more than 1 percentage point — even fairly small mis-estimates of g can result in a large error for ERM and MRP.

Estimating g

170. A standard way of estimating g is to use analyst forecasts of dividend growth over a finite future period, e.g., 3-5 years. However, as discussed in para 153, such forecasts are typically optimistic. Moreover, even when a relatively high value of g is justified in the short-run, this does not justify the same for the perpetual value of g required by the DGM-Implied approach. As a US court has put it:[38]

[38]Merion Cap., L.P. v. 3M Cogent, Inc., No. CV 6247-VCP, 2013 WL 3793896, at *21 (Del. Ch. July 8, 2013) (internal citations and quotations omitted), judgment entered sub nom. Merion Cap., L.P v. 3M Cogent, Inc. (Del. Ch. 2013).

"(g) should not be greater than the nominal growth rate for the United States economy, because if a company is assumed to grow at a higher rate indefinitely, its cash flow would eventually exceed America's (gross national product).

171. Analyst estimates of g can seriously misinform. Heaton (2021) describes a US case, eventually settled out of court, where an expert witness who relied on analysts' forecasts of g arrived at a cost of equity for a public utility that was over a percentage point higher than for the market as a whole, and arguably twice as high as the utility's true cost of equity.

172. Other methods could be used to estimate g. One simple method would be to use the historical average of market dividend growth. Alternatively, Rozeff (1984) argues that dividend growth should converge on the riskless rate of interest in the long-run and so $\hat{g} = R_f$, which implies:

$$E\hat{R}M_I = \frac{D}{P} \times (1 + R_f) + R_f$$

$$M\hat{R}P_I = \frac{D}{P} \times (1 + R_f)$$

which states that the MRP estimate is a product of the dividend-price ratio and the riskless return factor. Although there is strong evidence, at least in US data, that D/P does indeed positively predict MRP (see Cochrane, 2008 and 2011), a positive relationship between MRP and R_f is difficult to explain (see paras 97-98), and contrary to the empirical evidence of Harris and Marston (2013) obtained using the DGM.

173. Another possible approach to estimating g, untried to the best of our knowledge, is to exploit an unconditional version of (21):

$$ERM_{LR} = (\frac{D}{P})_{LR} \times (1 + g) + g \quad (23)$$

where ERM_{LR} is the unconditional mean of market returns (i.e., the "long-run" value of ERM) and $(\frac{D}{P})_{LR}$ is the is the unconditional dividend-price mean.[39] Assuming stationarity and ergodicity conditions are satisfied, we can use historical averages to estimate long-run means, and so write (23) as:

$$\overline{ERM} = \overline{(\frac{D}{P})} \times (1+g) + g$$

where $\overline{(\frac{D}{P})}$ is the historical average of the dividend-price ratio.[40] This can be rearranged to obtain an estimate for g:

$$\hat{g} = \frac{\overline{ERM} - \overline{(\frac{D}{P})}}{1 + \overline{(\frac{D}{P})}} \tag{24}$$

which could then be substituted back into (21) to obtain an estimate of the current ERM.

174. To illustrate, suppose $\overline{ERM}$ is 10.27% (Rangvid et al., 2014) and $\overline{(\frac{D}{P})}$ is 4% (Rangvid et al.). Then the estimate of g is:

$$\hat{g} = \frac{0.1027 - 0.04}{1.04} = 6.02\%$$

If the current D/P is 0.02 (i.e., prices are high), then

$$E\hat{R}M = 0.02(1.0602) + 0.0602 = 8.14\%$$

from which the current riskless interest rate could be subtracted to obtain $M\hat{R}P$.

175. This approach to estimating g is easily adapted to multi-stage adaptions of the DGM (where there are short-to-medium term periods of expected dividend growth that vary from the long-run value) by using the "unconditional" estimate of g in the terminal stage.[41]

[39] No "LR" subscript is required for g since this is already defined as a long-run value.

[40] Alternatively, one could write $\overline{ERM} = \overline{DY} + g$ where $\overline{DY}$ is the unconditional mean of the dividend yield (i.e., next period's expected dividend divided by current price), so that $\hat{g} = \overline{ERM} - \overline{DY}$

[41] AER (2018) adopts a 3-stage approach.

Contrast with MRP-Historical

176. A tangential concern with DGM-Implied is that, at least in the last 60-odd years of US data, it has almost always produced lower MRP estimates than MRP-Historical — see, for example, Fama and French (2002) and Damodoran (2021, Fig 9).[42] This may just indicate that the unconditional and conditional MRPs have generally diverged over the last 60 years. Or it could, as Fama and French argue, indicate a 60-year decline in the true MRP that isn't captured by historical averaging (as illustrated in Figure 2). But it could also indicate a fundamental downward bias in DGM-Implied estimates due to some systematic flaw in the DGM. Unfortunately, which of these is the correct answer is unknown.

Volatility of the dividend-price ratio

177. Implausible assumptions and/or difficult-to-implement features arguably matter less to a regulator if a model generates accurate and stable estimates. A final objection to the DGM-Implied approach is thus a purely practical one — it generates volatile estimates of MRP and the cost of equity. To see why, note first that the DGM states that ERM equals scaled-up D/P + a constant (equation (21)), so volatility of D/P translates directly into volatility of ERM. As Figure 1 in Cochrane (2011) reveals, D/P has fluctuated between a little over 1% and a little under 7% in 1950-2010 US data. This has resulted in similarly high volatility in DGM-Implied estimates of MRP — see Damodoran (2021, Figure 9). The evidence of Rangvid et al. (2014, Table 1) suggests lower (about 2/3) but still significant D/P volatility in Australia.

178. High volatility in D/P estimates translates into high volatility for cost of equity

[42] AER (2018) appears to be more concerned about DGM-Implied estimates being higher than historical estimates in Australian data.

estimates. Even with a beta of 0.6 (as determined by AER, 2018), a shift in D/P from 1% to 7% would see a 3.6 percentage point change in the estimated cost of equity. If Australian D/P volatility is 2/3 as great, this corresponds to a 2.4 percentage point shift in cost of equity estimates. Changes of this magnitude are at odds with the stability objective (see para 23) and are unlikely to be welcomed by either consumers or networks.

179. This effect could be partly mitigated by the existence of a negative relationship between D/P and R_f. That is, when D/P rises, ERM rises as well (pushing up the estimated cost of equity) but R_f falls (pushing down the cost of equity). However, Cochrane (2008, Table 1) finds that D/P has very little predictive power for interest rates. Whether or not the same is true for Australia is unknown.

Summary

180. The fundamental problem with relying on the DGM (or any other method for estimating the conditional MRP) is that the impossibility of observing the MRP means the model is unverifiable. That is, the DGM predictions of MRP cannot be compared with ex-post realisations of MRP and so its ability to match the data is unknown.

181. This lack of verification, together with the high volatility in estimates it seems likely to give rise to, suggests to us that the DGM-Implied approach should be viewed with considerable caution. This comes with two caveats. First, an approach based on extracting information from market prices is in principle a powerful one, so it should not be entirely dismissed. More general, yet-to-be developed, models based on risk and return rather than a mathematical identity might eventually offer more confidence. Second, D/P does seem to be able to predict (albeit very noisily) future excess returns, so the DGM's central feature of a direct link between D/P and MRP does have limited empirical support (see Campbell and

Thompson, 2008). It could thus be a useful estimator when used in conjunction with others.

182. AER (2018) arrive at a similar conclusion, albeit partly for different reasons. As with surveys, the best possible use of the DGM approach would seem to be as part of a combined estimator, a topic we discuss in section 8.

DGM-Implied estimates of MRP and ERM are volatile, unverifiable, and rely on strong assumptions. Improved (currently unknown) models could eventually offer hope for the future, but the current best use is likely to be as part of a combined estimator.

6 Estimating the MRP: Method IV — Theoretical

183. Standard implementation of the CAPM assumes that the market risk premium is a free parameter and hence must be estimated from data. But, as noted in para 95, equilibrium in the riskless asset market implies equation (17) in a CAPM world:

$$MRP = \gamma \times \sigma_e^2 \tag{17}$$

where γ is the average risk aversion of all investors holding the market portfolio and σ_e^2 is the variance of excess market returns R_e.

184. Equation (17) has an obvious interpretation: to hold the market portfolio, investors require an expected return premium equal to the compensation required per unit of risk (γ) times the number of units of risk (σ_e^2). It also seems to offer promise for estimating MRP — several well-established ways exist for estimating

σ_e^2, while Campbell and Viceira (2002) point out that 200 years of financial market history suggest that γ is a constant. Thus, all that's required is to first estimate γ and σ_e^2, then multiply them together to get an estimate of MRP that is internally consistent with the framework (CAPM) in which it is applied.

185. Unfortunately, this is a chimera. First, little is known about the true value of γ and estimates vary widely. Second, σ_e^2 is itself highly volatile — see, for example, Schwert (1989) and Moreira and Muir (2019). Unless γ is set far below what is normally considered plausible (i.e., close to risk neutrality), this would translate into extremely large shifts indeed in the MRP estimate over time, and therefore in the cost of equity.

7 Estimating the MRP: Method V — Empirical

186. Even if (17) is not a directly practical method for estimating MRP, its logic seems sound — MRP *should* reflect risk and risk aversion. Moreover, French et al. (1987), Bishop et al. (2011) and Kassa et al. (2021) report evidence consistent with a strong positive relationship between MRP and predicted volatility.[43] This suggests another possibility — an empirical approach where the relationship between excess returns and variables proxying for risk and risk aversion is estimated and then the fitted value of the empirical model is used as an estimate of MRP.

187. Specifically, this approach first estimates the regression:

$$R_e = \alpha + \lambda X + \eta \tag{25}$$

[43] The latter two studies also report results from Australian data. Surprisingly, however, the research literature on this relationship is by no means conclusive, with a number of authors reporting a zero or even negative relationship. However, as Kassa et al. (2021) point out, this work typically fails to distinguish between predicted and unexpected volatility.

where α and λ are constants, X is a predictor or explanatory variable, and η is an error term. This model estimates the extent to which future returns can be forecast by a currently observable variable X.[44]

188. If (25) is estimated using data from data at dates t-1,...,0, then the date t MRP estimate can be computed as:

$$M\hat{R}P_t = \hat{\alpha} + \hat{\lambda} \cdot X_t \tag{26}$$

189. What variables are candidates for X? Campbell and Thompson (2008) suggest:

- Difference between corporate and government bond yields – as a proxy for increased risk pricing;
- Long-term government bond yield – to allow for the effects of monetary policy and other broader economic factors that have a contemporaneous impact on stock and bond markets
- Dividend- and earnings-to-price ratios;
- Book-to-market;
- Consumption-to-wealth ratio.

190. Regardless of what exactly is chosen for X, the purpose of estimating (25) is to obtain the best possible prediction of MRP. That is, it is a forecasting exercise, not a hypothesis test or tests; the objective is to explain as much variation in MRP (more precisely, its proxy) as possible, not to test the sign and/or size of particular relationships between MRP and X. In statistics terminology, what is required is a high R^2, both in-sample and *out-of-sample*.

[44] Using, as is sometimes suggested, the DGM-implied MRP as the left-hand side variable is of little help here as it would only reveal the extent to which **X** is able to explain variation in the DGM-implied MRP and not the true MRP.

191. As Welch and Gao (2008) argue, successful development of a robust forecasting model, at least to a level deemed acceptable, is likely to be challenging as existing candidates yield low out-of-sample R^2 values. Campbell and Thompson (2008) claim some success:

> "We show that simple restrictions on predictive regressions, suggested by investment theory, improve the out-of-sample performance of key forecasting variables and imply that investors could have profited by using market timing strategies."

In other words, their model has predictive power for excess returns. Moreover, it often outperforms return forecasts based on the long-run historical mean of stock returns. But, as they admit, the improvement is small. While use of the model could generate excess returns for long-run investors (who have the flexibility to adjust portfolios), the gains would be of little value to a regulator wishing to obtain a precise estimate of MRP today.

192. Rapach et al. (2010) extend this approach by combining estimates from multiple candidates for X, while Gu et al. (2020) introduce machine learning. Both report additional explanatory power, but the gains are still small, and in the case of Gu et al. the estimator is complex.

193. Once again, this approach runs headlong into the problem that the true MRP cannot be observed even ex-post and so the extent to which an empirical model is successful in estimating MRP is unknown. Nevertheless, such models do seem to predict, albeit noisily, future returns and so could provide a useful tool for regulators in seeking to estimate the conditional MRP, primarily in conjunction with other conditional MRP estimators, e.g., DGM, surveys.

194. AER (2018) appear to consider variables (which they refer to as "conditioning variables") that are potentially X candidates, but, at least as far as we can tell,

only in a qualitative manner. For example, they interpret low values of implied volatility as being consistent with a lower MRP, but do not attempt to quantify such relationships using a regression like (25). In our view, the more ambitious, quantitative, approach is worth pursuing as an estimator to be used in combination with others.

A theoretical, CAPM-based approach to estimating MRP is unlikely to be feasible. An empirical "conditioning variables" approach is worthy of further investigation.

8 Combined Estimators

195. Ever since the seminal paper of Bates and Granger (1969), economists have been aware that, under certain conditions, combinations of individual estimators can outperform the individual estimators themselves. Intuitively, this arises because error in one estimator is likely to be imperfectly correlated with, or even independent of, error in another estimator. Thus, the combined estimator can generate a more precise estimate of the parameter of interest.

196. The extensive literature on combined estimators is beyond the scope of this report, as is consideration of the very large number of possible combinations. Instead, we briefly outline and discuss two possible combined MRP estimators that could, be adopted by the AER.

8.1 Historical-Smart

197. Recall the section 3.6 discussion of possible correlation between R_f and MRP and the differing assumptions about this correlation made by the Historical-MRP and

Historical-ERM estimators. This suggests the use of the following combined MRP estimator:

$$M\hat{R}P_s = b_{ERM}\overline{MRP} + (1 - b_{ERM})(\overline{ERM} - R_f)$$

and, therefore, a combined estimator for entity i's cost of equity:

$$E_s[R_i] = R_f + \beta_i M\hat{R}P_s \tag{27}$$

which, for conciseness, we call the "Historical-Smart Approach". That is, the Historical-Smart estimate of the MRP is a weighted average of the Historical-MRP and Historical-ERM estimates, where the weights are b_{ERM} and $1 - b_{ERM}$ respectively. For example, if $b_{ERM} = 0.5$, which is halfway between the Historical-MRP and Historical-ERM assumptions (see Table 1), then the Historical-Smart approach is a simple average of the two standard approaches. If, on the other hand, $b_{ERM} = 0.8$, which is much closer to the Historical-MRP approach, then Historical-Smart allocates 80% of the weight to Historical-MRP. Note that, as would be expected, $b_{ERM} = 0$ causes Historical-Smart to be equivalent to Historical-ERM while $b_{ERM} = 1$ makes it equivalent to Historical-MRP.

198. The properties of the Historical-Smart approach are straightforward, but uninteresting, to demonstrate, so we omit the details. First, Historical-Smart has lower average cost of equity estimation errors than both Historical-MRP and Historical-ERM so long as b_{ERM} lies between 0 and 1 (as would be expected). Second, Historical-Smart generates more stable cost of equity estimates than Historical-MRP. Third, however, if β_i is high relative to b_{ERM}, then Historical-ERM can generate more stable estimates than Historical-Smart. Intuitively, this happens because if b_{ERM} is low then the true cost of capital is close to being given by Historical-ERM which, if β_i is close to 1, effectively neutralises the effect of interest rate changes on cost of equity estimates. As a result, combining Historical-MRP with Historical-ERM produces more volatile estimates than just using the latter alone.

199. In practice, neither Historical-MRP nor Historical-ERM are likely to perfectly reflect reality and Historical-Smart acknowledges this point by generating a weighted average of the two, with the weights reflecting the MRP-R_f correlation and hence the extent to which the two approaches approximate market reality. If the MRP-R_f correlation can be estimated accurately, this results in lower cost of equity estimation errors and (usually) greater estimate stability, but inaccurate estimation of the MRP-R_f correlation can have the opposite effect.

200. The crucial input to Historical-Smart is thus the estimation of b_{ERM} which, as discussed in section 3.6 is problematical. As a practical matter, bearing in mind the recommendation in para 113, and the assumed desire that allowed returns not vary too much over time, the best approach might be to set b_{ERM} close to 1 (i.e., close to Historical-MRP) and adopt a watching brief. Regardless, we recommend that the chosen weights be explicitly quantified so that regulatory judgement is transparent.

Historical-Smart can potentially improve estimation accuracy and stability, but some caution is advisable due to uncertainty about optimal weights.

8.2 Historical-Plus

201. The Historical-Smart estimator combines an unconditional MRP estimator (Historical-MRP) with a conditional MRP estimator (Historical-ERM) where all time variation in the latter is driven by changes in R_f. An alternative combined estimator, which we call Historical-Plus, combines Historical-MRP with a combined estimator of the conditional MRP. It thus involves a double combination.

202. The intuition is straightforward. For various reasons, we have expressed reserva-

tions about all of the individual conditional MRP estimators — DGM, surveys, empirical models. But even if each of these estimators is a poor estimator of the level of the MRP, it is reasonable to expect them to perform better at identifying the direction of movement. Moreover, for the usual reasons, this property should be enhanced by combining the individual estimators.

203. Let $M\hat{R}P_c$ denote a combined estimator obtained from the individual conditional MRP estimators and let $\overline{MRP}_c$ be its long-run sample average. Then we have:

$$M\hat{R}P_+ = \overline{MRP} + (M\hat{R}P_c - \overline{MRP}_c)$$

and, therefore, a combined estimator for entity i's cost of equity:

$$E_+[R_i] = R_f + \beta_i M\hat{R}P_+ \tag{28}$$

That is, the Historical-Plus estimate of MRP takes the Historical-MRP estimate and adjusts it upwards or downwards depending on whether a combined estimator of the conditional MRP is above or below its sample mean.

204. To illustrate, suppose $\overline{MRP}$ is 6%, while $M\hat{R}P_c$ equals 6.7% which is below its sample mean of 7%. Then:

$$M\hat{R}P_+ = 6\% + (-0.3\%) = 5.7\%$$

That is, Historical-MRP is adjusted downwards to reflect the below-average estimate of conditional MRP.

205. If $\overline{MRP}_c$ is considerably different to $\overline{MRP}$, then a modification to allow for percentage, rather than absolute, deviations in the conditional estimate might be preferred:

$$M\hat{R}P_+ = \overline{MRP}(1 + (\frac{M\hat{R}P_c - \overline{MRP}_c}{\overline{MRP}_c}))$$

The example above would then become:

$$M\hat{R}P_+ = 6\%(1 + (-(0.3/7)) = 5.74\%$$

206. Overall, Historical-Plus generates an intuitive estimate of the conditional MRP without requiring the specification of any unidentifiable weights. However, implementation would require extensive prior work on the properties of $M\hat{R}P_c$ and $\overline{MRP}_c$ to ensure compliance with regulatory objectives. For example, large swings in $M\hat{R}P_c$ relative to $\overline{MRP}_c$ could undermine any objective for stability in allowed returns.

A sensible combination of historical and "forward-looking" estimators potentially provides the best overall approach to estimating MRP, although more work on their properties in Australian data would be required prior to implementation. If adopted, the weights used in combining estimators should be explicitly quantified, rather than simply reporting that "some" weight was given to a particular individual estimator.

8.3 Foreign MRP estimates

207. A crucial feature of most combined estimators is that the individual estimators making up the combination all relate to a single parameter of interest. For example, combining test batting averages with imports of used cars would generate a more precise estimate of a meaningless parameter (some "average" of batting success and car imports).

208. Sometimes the combined parameter can be of legitimate interest. To illustrate, suppose one is interested in estimating the mean size of Cocker Spaniels. Data are collected on Spaniel size and a sample average calculated. However, the number of available Spaniels is small, so the estimate has a high standard error. Happily

though, there are a large number of Great Danes available, so the original estimate could be joined with another estimate estimate generated by the sample of Great Danes so as to obtain a more precise "combined" estimate. As an estimate of Spaniel size, this is obviously nonsense, and illustrates the importance of not mixing up "apples and oranges" in creating combined estimators. But suppose a Great Dane jumps the fence and mates with the Spaniel across the road. Then one might well be interested in estimating the likely size of the offspring, in which case the combined estimator is exactly what is required.

209. This little parable is relevant to MRP estimation. It is commonly claimed (e.g., Lally, 2019) that estimates of MRP from Australian data should be supplemented with estimates of MRP from other countries to reduce sampling error and obtain a more precise estimate. However, this also risks introducing intrinsic variation (i.e., the true MRP differs across countries) and the net effect on the standard error is ambiguous. Moreover, the combined parameter being estimated is some proxy for a world MRP, not the Australian MRP. To return to the dog parable, the problem is that the Australian MRP could be akin to a Spaniel while foreign MRPs are closer to a Great Dane: combining them does not lead to a better estimate of the former, only of their combination.

210. This illustrates the problems with combining local and foreign MRP estimators. In general, the one exception to this is when one has strong priors that there is no intrinsic variation, i.e., that the true MRP is the same in all countries. Then the combination is indeed one of "apples with apples" and a better-quality estimate of apples is the result. However, international differences in market size, liquidity, tax systems, agricultural vs industrial vs technological bases and so on make such a position difficult to sustain.

211. It is sometimes claimed that capital mobility will drive MRPs together, but this is

incorrect. Capital mobility will equalize expected returns on *identical* assets, but individual country market portfolios are likely to be far from identical. Even if capital markets are perfectly integrated, each country's MRP will be proportional to its beta with the world market portfolio. Because of differences in size, liquidity, tax, and economic base, there is no obvious reason to suppose that each country's beta will be the same.

212. Overall, we advise caution in the use of MRP estimates from foreign countries. At best, these might be used to identify any significant cross-country divergences that emerge over time and potentially warrant further investigation.

9 Concluding Comments

213. As extensively discussed, and as summarised in Table 3 below, all MRP estimators have their drawbacks. Historical averaging can provide a robust estimate of the unconditional MRP but can also be very misleading if applied to short periods of data. The DGM-implied approach is sound in principle, but is dependent on the pricing model used and the DGM model (including its various extensions) relies on strong assumptions and has little predictive power for excess returns; the latter reservation also applies to an empirical "conditioning variables" approach. Surveys suffer from fundamental shortcomings in both design and outcomes.

214. If excess returns are stationary and ergodic, which are testable conditions, then we know that the sample average obtained from a long period of data must closely approximate the true unconditional MRP. No such comfort exists for the conditional MRP: because stock price changes are dominated by events that were ex ante unforeseen, the true MRP is unobservable. As a result, estimators of the conditional MRP are unable to be compared with true values and use of such estimators must therefore be justified by faith alone.

215. Finally, the AER practice of keeping MRP fixed between instrument settings creates problems for all estimators: an MRP set today is potentially outdated by the time a new allowed return is set for a particular firm (which could be up to almost four years later). If time variation in the true MRP is fast, this favours use of an unconditional MRP estimator (i.e., Historical-MRP); if time variation in the true MRP is slow, this favours use of a conditional MRP estimator.

216. Overall, given the current development of finance theory and practice, the advice of Goyal and Welch (2008) and Dimson et al. (2011, p13) — that investors cannot do better than a historical average obtained from a long period of data when it comes to forecasting excess returns — remains prudent. Nevertheless, it may be possible to obtain superior estimates via the use of suitably-chosen combined estimators that integrate the known properties of historical averaging with the unknown, but diversified, properties of other methods that allow for time variation in MRP. Whether any such improvement is sufficiently strong to justify a switch of approach, or is currently feasible, is uncertain, but, at least in in our view, that is the road along which further investigations should proceed.

Table 3: Summary of MRP Estimators

Estimator	MRP	Pros	Cons	Comments
A. Individual				
Historical MRP	Unconditional	• Known statistical properties • Straightforward to compute	• Cost of equity moves 1-1 with R_f • Misleading in short samples unless R_e non-stationary	•Requires R_e ergodicity •Base component of any MRP estimator
Historical MRP-R_f	Unconditional	• Minimal regulatory uncertainty and cost	•Cost of equity independent of R_f • Likely higher estimation errors	•Requires R_f ergodicity • Very blunt instrument
Historical ERM	Conditional	• Time-varying • Straightforward to compute	•Cost of equity independent of R_f	• Requires R_m ergodicity • Most useful in Historical-Smart
DGM	Conditional	• Plausible link to true MRP via D/P • Exploits current market information	• Requires implausible assumptions • Requires full market efficiency • Very weak empirical support	• Possible role in combination estimator
Surveys	Conditional	• Forecasts from informed participants • Exploits current market information	• Weak participant incentives • Often reflects extrapolation from past	• Possible role in combination estimator
Empirical	Conditional	• Links MRP to observable variables	• Weak evidence for link	• Possible role in combination estimator
B. Combined				
Historical-Smart	Conditional	• Straightforward to compute • Time-varying: captures MRP-R_f correlation • Low cost of equity errors • Stable cost of equity estimates	• Weights difficult to estimate • All time variation due to R_f	• Requires R_e and R_m ergodicity • Cons suggest 'gentle' introduction
Historical-Plus	Conditional	• Exploits all possible estimators	•Depends on quality of $M\hat{R}P_c$	• Full information estimator • Requires more work on practical effects

References

Adam, K., A. Marcet and J. Beutel, 2017. Stock price booms and expected capital gains. American Economic Review 107, 2352–2408

AER, 2018. Rate of return instrument explanatory statement. Available at `https://www.aer.gov.au/networks-pipelines/guidelines-schemes-models-reviews/rate-of-return-instrument-2018/final-decision`

Ang, A., and G. Bekaert, 2007. Stock return predictability: is it there? Review of Financial Studies 20, 651–707

Anonymous, 2019. Ergodic theory. Available at `http://bactra.org/notebooks/ergodic-theory.html`

Barro, R., 2006. Rare disasters and asset markets in the twentieth century. Quarterly Journal of Economics 121, 823-866.

Bates, J. and C. Granger, 1969. The combination of forecasts. Operational Research Quarterly 20, 451–468

Bernoulli D., 1954. Exposition of a new theory on the measurement of risk. Econometrica 22, 23–36.

Bishop, S., M. Fitzsimmons and R. Officer, 2011, Adjusting the market risk premium to reflect the global financial crises. JASSA The Finsia Journal of Applied Finance 1, 8–14.

Bishop, S., T. Carlton and T Pan, 2018. Market risk premium: australian evidence. Macquarie University. Available at `https://www.charteredaccountantsanz.com/-/media/5d58bf969ac94767a21c8eb6bd458dad.ashx`

Boyle, G., 2006. Risk, expected return, and the cost of equity capital. New Zealand Economic Papers 39, 181-194

———, 2009. Capital market integration: a review of the issues and an assessment of New Zealand's position. Ministry of Economic Development and the Capital Market Development Taskforce. Available at `https://researcharchive.vuw.ac.nz/xmlui/handle/10063/4034`

——— and G. Guthrie, 2004. Human capital and popular investment advice. Review of Finance 9, 139–164

———- and K. Murray, 2021. Systematic risk and the role and measurement of equity beta: A report to the AER Consumer Reference Group. Available at `https://www.aer.gov.au/system/files/CRG%20-%20Submission%20-%20Overall%20rate%20of%20return%2C%20Equity%20and%20Debt%20-%203%20September%202021%20-%20Sapere%20Systematic%20risk%20and%20equity%20beta%20report.pdf`

Brealey, R., S. Myers and F. Allen, 2018. Principles of Corporate Finance (12th ed.): New York: McGraw-Hill.

Brealey, R., S. Myers and A. Marcus, 2020. Fundamentals of Corporate Finance (10th ed.): New York: McGraw-Hill.

Breeden, D., 1979. An intertemporal asset pricing model with stochastic consumption and investment opportunities. Journal of Financial Economics 7, 265–296.

Brennan, M., 1989. Capital asset pricing model. The New Palgrave Dictionary of Economics: Finance. New York and London: W.W. Norton, 81–85

Brown, S., W. Goetzmann, and S. Ross, 1995. Survival. Journal of Finance 50, 853–873.

Bruner, R., W. Lia, M. Kritzman, S. Myrgrenc and S. Page, 2008. Market integration in developed and emerging markets: Evidence from the CAPM. Emerging Markets Review 9, 89-103

Campbell, J. and R. Shiller, 1988. The dividend-price ratio and expectations of future dividends and discount factors. Review of Financial Studies 1, 195–228

——— and S. Thompson, 2008. Predicting excess stock returns out of sample: can anything beat the historical average? Review of Financial Studies 21, 1509–1531.

——— and L. Viceira, 2002. Strategic Asset Allocation. New York: Oxford University Press.

Canova, F., 2007. G-7 inflation forecasts: random walk, Phillips curve or what else? Macroeconomic Dynamics 11, 1–30.

CEPA, 2021.Relationship between RFR and MRP. Available at https://www.aer.gov.au/system/files/Report%20to%20the%20AER%20-%20Relationship%20between%20RFR%20and%20MRP%20-%2016%20June%202021.pdf

Chen,L., D. Zhi and X. Zhao, 2013. What drives stock price movements? Review of Financial Studies 26, 841–876

Cochrane, J., 2005. Asset Pricing. Princeton, New Jersey and Oxford: Princeton University Press.

———, 2008. The dog that did not bark: a defense of return predictability. Review of Financial Studies 21, 1533–1575.

———, 2011. Discount rates. Journal of Finance 66, 1047–1108.

———, 2013. Bob Shiller's Nobel. Available at https://johnhcochrane.blogspot.com/2013/10/bob-shillers-nobel.html

———, 2014. Week 1 notes. Available at https://static1.squarespace.com/static/5e6033a4ea02d801f37e15bb/t/5f5a8e111341366ceee6655a/1599770132177/week_1_notes.pdf

Cooper, I., 1996. Arithmetic versus geometric mean estimators: setting discount rates for capital budgeting. European Financial Management 2, 157-167.

Damodaran, A., 2021. Equity risk premiums (ERP): determinants, estimation, and implications – the 2021 edition . Available at https://ssrn.com/abstract=3825823.

Dimson, E., P. Marsh and M. Staunton, 2002. Triumph of the Optimists 101 Years of Global Investment Returns. Princeton, New Jersey and Oxford: Princeton University Press.

——————————————, 2011. Equity Premia around the World. Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1940165

Easton, P., and G. Sommers, 2007. Effect of analysts' optimism on estimates of the expected rate of return implied by earnings forecasts. Journal of Accounting Research 45, 983–1015

Elton, E., 1999. Expected return, realized return, and asset pricing tests. Journal of Finance 54, 1199–1220.

European Central Bank, 2020. ECB pushes for EU capital markets integration and development. Available at https://www.ecb.europa.eu/press/pr/date/2020/html/ecb.pr200303~5eaf4c119d.en.html

Fama, E., 1970. Efficient capital markets: a review of theory and empirical work Journal of Finance 25, 383–417

Fama, E. and K. French, 1988. Permanent and temporary components of stock prices. Journal of Political Economy 96, 246–273.

————————, 2002. The equity premium. Journal of Finance 57, 637–659

French, K., G. Schwert and R. Stambaugh, 1987. Expected stock returns and volatility. Journal of Financial Economics 19, 3–29.

Fried, J., P. Ma and C. Wang, 2021. Stock investors' returns are exaggerated. European Corporate Governance Institute. Available at https://papers.ssrn.com/sol3/

papers.cfm?abstract_id=3964177

Friedman, M., 1953. Essays in Positive Economics. Chicago: University of Chicago Press.

Friend, I. and M. Blume, 1975. The demand for risky assets. American Economic Review 65, 900-922.

Gordon, M., 1959. Dividends, earnings, and stock prices. Review of Economics and Statistics, 41, 99–105

Goyal, A., and I. Welch. 2008. A comprehensive look at the empirical performance of equity premium prediction. Review of Financial Studies 21, 1455–1508

Greenwood, R. and A. Shleifer, 2014. Expectations of returns and expected returns. Review of Financial Studies 27, 714–746.

Gu, S., B. Kelly, and D. Xiu, 2020. Empirical asset pricing via machine learning. Review of Financial Studies 33, 2223–2273

Harris, R., F. Marston, D. Mishra and T. O'Brien, 2003. Ex ante cost of equity estimates of S&P 500 firms: the choice between global and domestic CAPM. Financial Management 32, 51-66

Harris, R. and F. Marston, 2013. Changes in the market risk premium and the cost of capital: implications for practice. Journal of Applied Finance 23, 1-14.

Heaton, J., 2021. Unjust and unreasonable: misuse of the dividend growth model in public utility rate setting. Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3947360

Huang, C. and R. Litzenberger, 1988. Foundations for Financial Economics. North-Holland.

Hull, J., 2009. Options, Futures and Other Derivatives (7th edition). Pearson Prentice Hall.

Ibbotson, R., R. Carr and A. Robinson, 1982. International equity and bond returns. Financial Analysts Journal 38, 61-83.

Ibbotson, R. and R. Sinquefield, 1976. Stocks, bonds, bills and inflation: year-by-year historical returns (1926-1974). Journal of Business 49, 11–47.

Ince, O. and T. Molodtsova, 2017. Rationality and forecasting accuracy of exchange rate expectations: evidence from survey-based forecasts. " Journal of International Financial Markets, Institutions and Money 47, 131–151.

Indro, D. and W Lee, 1997. Biases in arithmetic and geometric averages as estimates of long-run expected returns and risk premia. Financial Management 26, 81–90

Jacquier, E., A. Kane and A. Marcus, 2005. Optimal estimation of the risk premium for the long run and asset allocation: a case of compounded estimation risk. Journal of Financial Econometrics 3, 37-55

Jorion, P. and W. Goetzmann, 1999. Global stock markets in the twentieth century. Journal of Finance 54, 953–980

Kassa,H., F. Wang, and X. Yan, 2021. Expected stock market returns and volatility: three decades later. Critical Finance Review, forthcoming. Available at `https://cfr.pub/forthcoming/papers/kassa2021expected.pdf`

Kliesen, K., 2015. How accurate are measures of long-term inflation expectations? Federal Reserve Bank of St Louis Economic Synopses. Available at `https://research.stlouisfed.org/publications/economic-synopses/2015/05/06/how-accurate-are-measures-of-long-term-inflation-expectations/`

Koedijk, K., C. Kool, P. Schotman, and M. Van Dijk, 2002. The cost of capital in international financial markets: Local or global? Journal of International Money and Finance, 21, 905-929

Lally, M., 2005. The weighted average cost of capital for electricity lines businesses. Available at https://comcom.govt.nz/__data/assets/pdf_file/0027/88362/The-Weighted-Average-Cost-of-Capital-for-Electricity-Lines-Businesses.pdf

———, 2019. Estimation of the tamrp. Available at https://comcom.govt.nz/__data/assets/pdf_file/0043/189889/Dr-Martin-Lally-Estimation-of-the-TAMRP-26-September-2019.pdf

LeRoy, S. and R. Porter, 1981. The present-value relation: tests based on implied variance bounds. Econometrica 49, 555–574.

Li, H. and Y. Xu, 2002. Survival Bias and the Equity Premium Puzzle. Journal of Finance 57, 1981–1995

Martin, I., 2017. What is the expected return on the market? Quarterly Journal of Economics 132, 367–433

Mehra, R. and E. Prescott, 2008. The equity premium: ABCs. Handbook of the Equity Risk Premium (ed. R. Mehra). Amsterdam: Elsevier.

Merton, R., 1973. An intertemporal capital asset pricing model. Econometrica 41, 867–87.

———-, 1980. On estimating the expected return on the market. Journal of Financial Economics 8, 323-361.

Mindlin, D., 2011. On the relationship between arithmetic and geometric returns. Available at https://ssrn.com/abstract=2083915

Moreiraa,A. and T. Muir, 2019. Should long-term investors time volatility? Journal of Financial Economics 131, 507–527

Orlowski, L., 2020. Capital markets integration and economic growth in the European Union. Journal of Policy Modeling 42, 893-902

Rangvid, J., M. Schmeling and A. Schrimpf, 2014. Dividend predictability around the world. Journal of Financial and Quantitative Analysis 49, 1255–1277

Rapach, D., J. Strauss and G. Zhou, 2010. Out-of-sample equity premium prediction: combination forecasts and links to the real economy. Review of Financial Studies 23, 821–862

Rietz, T., 1988, 'The equity risk premium: a solution?' Journal of Monetary Economics 21, 117–132.

Rozeff, M., 1984. Dividend yields are equity risk premiums, Journal of Portfolio Management, 11, 68-75.

Schwert, G., 1989. Why does stock market volatility change over time? Journal of Finance 44, 1115–1153.

Shapiro, S. and S. Horner, 2014. Stationarity of Ibbotson Associates equity risk premiums. Avaialble at `https://www.aeaweb.org/conference/2014/retrieve.php?pdfid=705`

Thaler, R., 2014. What scientific idea is ready for retirement? Available at `https://www.edge.org/response-detail/25293`

van Binsbergen, J., S. Hua and J. Wachter, 2020. Is the United States a lucky survivor: a hierarchical Bayesian approach.

Verbrugge, R. and S. Zaman, 2021. Whose inflation expectations best predict inflation? Federal Reserve Bank of Cleveland Economic Commentary. Available at `https://`

www.clevelandfed.org/en/newsroom-and-events/publications/economic-commentary/2021-economic-commentaries/ec-202119-whose-inflation-expectations-best-predict-inflation.aspx

Wright, S. and A. Smithers, 2013. The cost of equity capital for regulated companies: a review for Ofgem. Available at https://www.ofgem.gov.uk/sites/default/files/docs/2014/02/wright_smithers_equity_market_return.pdf

Appendix A Error variance comparison of cost of equity approaches: equation (7) vs equation (8)

From (9), we have:

$$\text{Historical-MRP estimation error for } E[R_i] \; = \beta(\overline{MRP} - MRP)$$

which has variance $\hat{\sigma}_M^2$ given by:

$$\hat{\sigma}_M^2 = \beta_i^2 \sigma_{MRP}^2 \tag{29}$$

where σ_{MRP}^2 is the variance of MRP. Similarly, from (10) we have:

$$\text{Historical-MRP-}R_f \text{ estimation error for } E[R_i] \; = (\overline{R_f} - R_f) + \beta(\overline{MRP} - MRP)$$

which has variance $\hat{\sigma}_A^2$ given by:

$$\hat{\sigma}_A^2 = \sigma_{R_f}^2 + \beta_i^2 \sigma_{MRP}^2 + 2\beta_i \sigma_{fp} \tag{30}$$

where $\sigma_{R_f}^2$ is the riskless rate variance and σ_{fp} is the covariance of R_f and MRP. Comparing (29) and (30) reveals that $\hat{\sigma}_M^2 < \hat{\sigma}_A^2$ if and only if

$$b_{MRP} < \frac{-1}{2\beta}$$

where $b_{MRP} \equiv \frac{\sigma_{fp}}{\sigma_{R_f}^2}$ is the sensitivity of MRP to R_f shocks, i.e., the change in MRP associated with a 1 percentage point change in R_f.[45]

[45] Strictly speaking, b_{MRP} is a regression coefficient in the linear equation

$$MRP = a + b_{MRP} R_f + \epsilon$$

and so is a linear sensitivity.

Appendix B Error variance comparison of cost of equity approaches: Historical-ERM vs Historical-MRP

As before, the cost of equity estimation error associated with Historical-MRP is:

$$\text{Historical-MRP estimation error for } E[R_i] = \beta(\overline{MRP} - MRP)$$

which has variance $\hat{\sigma}_M^2$ given by:

$$\hat{\sigma}_M^2 = \beta_i^2 \sigma_{MRP}^2 \tag{31}$$

Similarly, the cost of equity estimation error for the Historical-ERM approach is:

$$\begin{aligned}
\text{Historical-ERM estimation error for } E[R_i] &= \text{estimated cost of equity - true cost of equity} \\
&= (R_f + \beta_i(\overline{R_m} - R_f)) - (R_f + \beta_i(ERM - R_f)) \\
&= \beta_i(\overline{R_m} - ERM)
\end{aligned}$$

which has variance $\hat{\sigma}_E^2$ given by:

$$\hat{\sigma}_E^2 = \beta_i^2 \sigma_{ERM}^2 \tag{32}$$

where σ_{ERM}^2 is the variance of ERM. Comparing (31) with (32), we see that the difference in average estimation errors is proportional to $(\sigma_{MRP}^2 - \sigma_{ERM}^2)$. Noting that MRP $= ERM - R_f$, it follows that:

$$\sigma_{MRP}^2 = \sigma_{ERM}^2 + \sigma_{R_f}^2 - 2\sigma_{fe}$$

where σ_{fe} is the covariance of R_f and ERM. Hence:

$$\sigma_{MRP}^2 - \sigma_{ERM}^2 = \sigma_{R_f}^2 - 2\sigma_{fe}$$

Dividing by $\sigma_{R_f}^2$ and rearranging, we see that $\sigma_{MRP}^2 - \sigma_{ERM}^2$ has the sign of:

$$\frac{1}{2} - b_{ERM}$$

where $b_{ERM} \equiv \frac{\sigma_{fe}}{\sigma^2_{R_f}}$ is the sensitivity of ERM to R_f shocks, i.e., the change in ERM associated with a 1 percentage point change in R_f. Thus, Historical-ERM generates lower average errors than Historical-MRP if and only if the R_f sensitivity of ERM is strictly less than 0.5.

Following some simple but tedious algebra, an analogous computation reveals that:

$$\hat{\sigma}^2_A - \hat{\sigma}^2_E = \sigma^2_{R_f}[\beta^2_i(1 - 2b_{ERM}) + (1 + 2\beta_i b_{MRP})]$$

Since $b_{fp} = b_{fm} - 1$, this becomes;

$$\hat{\sigma}^2_A - \hat{\sigma}^2_E = \sigma^2_{R_f}(\beta_i - 1)[(\beta_i - 1) - 2\beta_i b_{ERM}]$$

which is strictly positive for any $\beta_i \leq 1$, which is likely to be the case for any regulated network. Thus, Historical-ERM almost certainly generates lower average estimation errors on average than Historical-MRP-R_f.

Appendix C Stability comparisons of cost of equity approaches: Historical-ERM vs Historical-MRP

Let $\hat{\phi}^2_M$ denote the variance of cost of equity estimates for entity i when the Historical-MRP approach is used. From (7), we have:

$$\hat{\phi}^2_M = \sigma^2_{R_f}$$

Let $\hat{\phi}^2_E$ denote the variance of cost of equity estimates for entity i when the Historical-ERM approach is used. From (14), we have:

$$\hat{\phi}^2_E = (1 - \beta_i)^2 \sigma^2_{R_f}$$

So

$$\frac{\hat{\phi}^2_E}{\hat{\phi}^2_M} = (1 - \beta_i)^2$$

and therefore the percentage difference is:

$$\frac{\hat{\phi}_E^2 - \hat{\phi}_M^2}{\hat{\phi}_M^2} = \beta_i(\beta_i - 2)$$

Letting $\hat{\phi}_{MR}^2$ denote the variance of cost of equity estimates for entity i when the Historical-MRP-R_f approach is used, note that because $\overline{MRP}$ and $\overline{R_f}$ are effectively constant over time, we have:

$$\hat{\phi}_{MR}^2 \approx 0$$

Appendix D Arithmetic or Geometric average return: Brealey et al. (2018)

Suppose that the value of Neverland's stock market index is \$100. There is an equal chance that at the end of the year it will be worth \$90, \$110, or \$130, and so the *expected* cashflow is \$110 (assuming Neverland companies do not pay dividends). The future rate of return is -10%, +10%, or +30% and so the *expected* return is (1/3)(-10+10+30) = 10%.

If we discount the expected cash flow by the expected rate of return, we obtain the correct value of Neverland's stock market index:

$$PV = \frac{110}{1.1} = \$100$$

The expected return of 10% is therefore the correct rate at which to discount the expected cash flow for Neverland's market portfolio. It is also the opportunity cost of capital for investments that have the same degree of risk as the Neverland market, i.e., ERM = 10% for Neverland.

Now suppose that we observe the returns on the Neverland index over T years. If the odds are unchanged, the return will be -10% in $T/3$ years, +10% in $T/3$ years, and

+30% in the remaining $T/3$ years. The arithmetic average of these yearly returns is:

$$\frac{(-10+10+30)}{3} = 10\%$$

Thus the arithmetic average of the returns correctly estimates the ERM for Neverland.

By contrast, the geometric average annual return for Neverland is

$$((0.9)^{T/3} \times (1.1)^{T/3} \times (1.3)^{T/3})^{1/T} - 1 = (0.9 \times 1.1 \times 1.3)^{1/3} = 0.088 \text{ or } 8.8\%$$

which is less than than Neverland's true ERM (10%). Using the geometric average return understates the true cost of equity and results in inefficient investment decisions.

Appendix E Proof of equation (16)

By definition:

$$ERM = R_f + MRP$$

which implies:

$$\sigma_{fm} = \sigma^2_{R_f} + \sigma_{fp}$$

and therefore, dividing through by $\sigma^2_{R_f}$ yields:

$$b_{ERM} = 1 + b_{MRP}$$