SOUTH AUSTRALIAN
**CENTRE FOR ECONOMIC STUDIES**

ADELAIDE & FLINDERS UNIVERSITIES

Mr Jim Hancock, Deputy Director
Telephone: (+61-8) 8303 5515 Facsimile: (+61-8) 8232 5307
Email: jim.hancock@adelaide.edu.au http://www.adelaide.edu.au/saces

11 September, 2006

Dr John Tamblyn
Chairman
Australian Energy Market Commission
PO Box H166
AUSTRALIA SQUARE NSW 1215

Dear Dr Tamblyn,

## Submission on Draft Rule Determination and Second Draft Rule

I refer to your Draft Rule Determination and second Draft National Electricity Amendment (Economic Regulation of Transmission Services) Rule 2006.

### *Market risk premium*

I note that if the Draft Rule is adopted the market risk premium (MRP) will be deemed to be 6 per cent (following Section 6A.6.2 of the Draft Rule).

The true value of the MRP is an uncertain quantity but the evidence suggests that it has fallen in recent decades. The analysis in the attached report, *The Market Risk Premium for Australian Regulatory Decisions*, which has been prepared for the Advocacy Panel, concludes that the evidence from historic excess returns suggests that the most likely value of the 1-year MRP is in the range 5 to 6 per cent, but with considerable uncertainty still attached. Among the issues raised in the report is that reductions in securities transaction costs have had a downward influence on the MRP of the order of 1 per cent since the early 1980s.

Thus the analysis in the report does not argue strongly against your proposed parameter value but suggests that the contemporary MRP value is more likely to be below 6 per cent than above it.

### *Equity beta*

I note that if the Draft Rule is adopted the equity beta will be deemed to be 1.0 (following Section 6A.6.2 of the Draft Rule).

In the Draft Rule Determination you say that:

> The equity beta is the most difficult parameter to estimate, as it cannot be measured accurately from empirical data that is available. The Commission understands that the value of 'one' that was adopted in the SRP represents a compromise between the difficulties of estimation and the consequent need to err on the side of caution. Regulators have applied equity betas above and below 'one', but 'one' has come to represent the most widely accepted practice.

In my view you have not given due recognition to the accumulation of knowledge since the SRP. In the SRP it was said that:

> The code clearly provides for consideration of market data, and the emerging data suggests the appropriate equity beta for TNSPs may be less than 1. However current statistical methods for estimating the equity beta from market data tend to produce varying confidence interval (and sample average) estimates. The ACCC also notes that the time period of the market data is not long enough to satisfy the ACCC that market derived equity betas would not systematically under-compensate the TNSPs. That is, the current decline in the measures of market derived equity betas may reflect a short term deviation from normal trend. In saying this, the ACCC will continue to use market evidence to check the reasonableness of a TNSP's equity beta.

A footnote to that paragraph noted that:

> The estimated re-levered equity betas from a sample of comparable Australian energy firms have fallen from around 1 in 2000 to around 0.3 in 2003. (see ESCOSA, Electricity Distribution Price Review: Return on Assets-Preliminary Views, January 2004, p.56). This analysis is consistent with the ACCC's estimates of market derived equity betas shown in recent regulatory decisions.

A submission on your first Draft Rule from the Government of South Australia said that "an equity beta for Australian utilities of substantially lower than 1 (possibly as low as 0.5) may be reasonable". That submission also drew to your attention work by Professor Lally advising that an equity beta of no more than 0.8 was justified. That advice was provided in May 2005 as part of a submission to an appeal over an Essential Services Commission of South Australia (ESCOSA) determination which set an equity beta of 0.8 for the South Australian electricity distribution assets. Submissions were also received from Professors Gray and Officer arguing for a beta of 1.0. In its final decision ESCOSA set a beta of 0.9. The decision did not indicate any retreat from the view that a beta of 0.8 was to be preferred on the evidence, but instead was explained by ESCOSA on the ground that its original decision "did not give adequate weight to the requirement of clause 6.10.3(e)(6) of the NEC to provide reasonable certainty and consistency over time in the outcomes of regulatory processes". Thus the view put to you by the South Australian Government draws on an extensive and more recent review process than the SRP, yet in your Draft Determination you reach your conclusions on equity beta without addressing the South Australian Government submission or its foundations, while carrying out no substantive analysis of your own in the Draft Determination. As your decision will effectively lock in a contentious beta estimate until 2011 it is highly desirable to take on board now what has been learned since the SRP was prepared.

I would be happy to discuss these matters further at your convenience.

Yours sincerely,

Jim Hancock
**Deputy Director**

SOUTH AUSTRALIAN
CENTRE FOR ECONOMIC STUDIES

ADELAIDE & FLINDERS UNIVERSITIES

# The Market Risk Premium for Australian Regulatory Decisions

Prepared for:

**The Advocacy Panel**

Prepared by:

**The South Australian Centre for Economic Studies**

**July 2006**

PO BOX 125, RUNDLE MALL, SOUTH AUSTRALIA 5000
PHONE (+61-8) 8303 5555 FAX (+61-8) 8232 5307
FOURTH FLOOR, LIGERTWOOD BUILDING, UNIVERSITY OF ADELAIDE

# Contents




This report was prepared by the following researcher:

***Jim Hancock, Deputy Director, SACES.***



**Acknowledgement:** I am very grateful to The Advocacy Panel which provided the essential funding support for this study. Bob Officer kindly provided me with his long series of Australian excess returns. I have benefited greatly from discussions with Robert Schwarz about regulatory cost of capital issues over many years. During the preparation of this study I also received helpful comments from Steven Bishop, Jasper Mikkelsen, Bob Officer, Jeff Washusen and Ralf Zurbrugg. I would like to thank all for their input. This does not imply that they agree with the content of this report and of course the responsibility for any errors remains with the author.

# Executive Summary

This is the report of an investigation into what value of the market risk premium (MRP) should be used to set regulated rates of return for Australian electricity utilities.

Of course the MRP is a generic parameter, so that much of what is said herein is also relevant for setting regulatory rates of return for other long-lived infrastructure.

The MRP $\pi$ is defined as

$$\pi = E(R_M) - 1 - i$$

where $E(R_M)$ is the expected gross return on the market and $i$ is the return on a riskfree asset (ie. a pure interest rate). Although time subscripts have not been included, the MRP exists at a point it time. There is nothing in its construction that requires that it be unchanging through time. However, constancy has often been assumed, particularly in estimation exercises.

In Australian electricity regulatory decisions the most common assumption regarding the value of the MRP is 6 per cent. However, there is a divergence of views on the appropriateness of this judgment. On the one hand it is argued that a long series of historical returns shows a 1-year average MRP of over 7 per cent, and that the use of even a 6 per cent MRP has scant statistical support. On the other hand, it is argued that plausible views as to growth prospects for dividends, and surveys of financial market practitioners' expectations, both suggest an MRP below 6 per cent.

This study is primarily concerned with the examination of historic excess returns. Under certain assumptions the average of historic excess returns can be used to estimate the market risk premium. The key finding of this report is that the best estimate of the contemporary MRP can be made by considering monthly excess returns over the last 30 years. After allowing for known biases in this data, the central estimate of the MRP is in the range 5 to 6 per cent. This finding rests on a rejection of the view that the MRP has been stable over the last 122 years.

In this study three series of historic excess returns are considered – a long series of 122 annual return annual observations from Professor Officer, a 35 year series of annual observations from Ibbotson Associates, and a 30 year series comprising 372 monthly observations from the Australian Graduate School of Management. The Officer series and the AGSM series are of most interest and have competing claims to be the better estimator. The Officer dataset with 122 observations has the largest available series of non-overlapping 1-year excess returns. However, the AGSM data has 361 overlapping 12-month excess returns and also has 372 monthly excess return observations which can be used to estimate mean 12-month excess returns with the use of another moment condition. In each case an adjustment has been made to the raw data to incorporate the value of franking credits into the excess return series. The series have been boosted by 0.6 per cent for the years since the introduction of dividend imputation.

If one believes that the MRP changes over time, then there are obvious attractions in giving greater weight to more recent data. However, an MRP estimate based on just 31 annual observations (which is the span of the AGSM data to the end of 2004) forms a relatively small sample and therefore is quite imprecise.

A key finding in this report is that it is possible to make the estimate for the last 31 years with much greater precision than has been done in the past. It is possible, using monthly data for the last 31 years, to estimate average 1-year returns with a 95 per cent confidence interval spanning about 4½ percentage points. Estimates made using just annual observations, which is the standard past practice, produce a confidence interval with width of about 16 percentage points when calculated on just 30 years data. The reason for the difference is that there are 360 monthly observations available over 30 years but just 30 non-overlapping 12-month observations.

It is possible to construct a weighted estimator comprising an estimator based on 92 years of annual data and an estimator based on 30 years of monthly data. This weighted estimate of the 1-year MRP is 6.4 per cent. It is very precise, with a standard error of just 0.9 percentage points. Its 95 per cent confidence interval spans just 3½ percentage points.

However, the use of such a measure rests on the idea that the distribution of excess returns has been stable over time. It has previously been found by Gray (2001) that there is not "strong" evidence of a structural change in average excess returns. The problem is, as has been pointed out by the Victorian Essential Services Commission, that these tests lack "power". Monte Carlo simulations in this report confirm that view: in the simulation performed, conventional t-tests would detect a 5 per cent reduction in the MRP only 40 per cent of the time (ie. they wrongly allow the "no change" hypothesis to stand 60 per cent of the time). Yet a change in the MRP of this order would be of massive economic significance. The conclusion is that the fact that conventional t-tests do not detect a change in mean excess returns cannot be regarded as a strong rebuttal that such a change has occurred.

Extensive tests are conducted over the question of whether the long series can be regarded as stable over its 122 years. These tests seek to allow for autocorrelation in returns, which complicates the analysis. Tests for structural breaks were then carried out by incorporating dummy variables in an ARMA model. Breaks were considered for each year from 1960 to 1985. The non-parametric tests rejected the "no break" hypothesis at a 5 per cent significance level in most cases and in every case at a 7 per cent significance level.

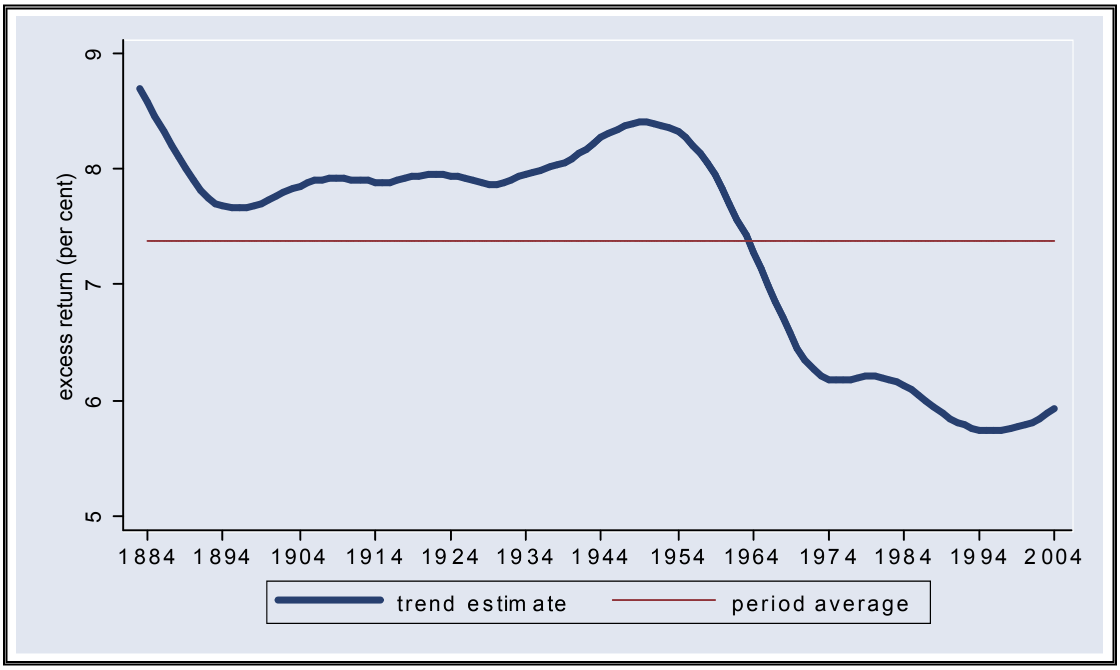
On the basis of these considerations the hypothesis of a constant MRP over the last 122 years is difficult to accept. Of course the hypothesis does not in any case have particularly strong support in any economic theory; rather, it is a convenient assumption for statistical inference. The data appear to be inconsistent with it.

Various smoothing techniques were applied to the data to produce trend estimates of the MRP which allow for it to change over time. One of these was the Hodrick-Prescott filter, which is commonly used to filter high frequency noise out of macroeconomic time series. The smoothed series is shown in Figure E.1 below, along with the simple average. It is readily apparent that filtering the data in this way suggests that the MRP has dipped substantially, and has recently been at all time lows around 6 per cent. (This study uses data only until the end of 2004; had the data for 2005 been used it would show a stronger rise towards the end of the period).

The predictive power of a range of filters was measured by considering the prediction accuracy of rolling predictions over a 60 year period. The best prediction performance was achieved by a lagged value of the Hodrick-Prescott filter. However, the simple average also performed relatively well.

The author's view is that the assumption that the mean 1-year excess return has been stable over the last 122 years is very tenuous. This tends to rule against the very long term average as an estimator. Now that higher-precision estimates are available for the past 30 years, it seems reasonable to put more weight on that period.

**Figure E.1**
**Average excess returns and Hodrick-Prescott trend estimate of excess returns**




Some recent studies have thrown new light on the interpretation of excess return data by taking into account taxation effects and the effects of transaction costs and, related to transaction costs, illiquidity in investors' portfolios. In this view, the market risk premium needs to be seen as more than a pure risk-compensation element of equity returns. It will also include components relating to the relative generosity of the tax treatment of equities versus bonds, and components relating to transaction costs and illiquidity in asset markets.

With this distinction in mind it may be sensible to differentiate two concepts that have been treated as synonymous in much of the Australian discussion. The equity premium, which is the expected excess return on equities, will include a pure risk premium, a tax-compensation element and a transactions costs/liquidity premium. With these three components in mind, it becomes apparent that what is known as the "market risk premium" in Australia is not just a pure risk premium but also includes these tax and transaction cost/liquidity elements.

A further implication of this breakdown is changes in taxation and transaction costs and the liquidity characteristics of markets have the potential to drive changes in the expected value of excess returns (i.e. the market risk premium).

To date, there has been only limited empirical research on these issues.

Some "experimental" estimates are presented here for tax effects. They indicate some upward pressure on excess returns over the last 30 years, reflecting that lower inflation creates a much less tax-disadvantaged environment for bonds and that capital gains tax has been introduced.

These influences are offset somewhat by the introduction of dividend imputation which has benefited domestic equity investors. However, it has not been possible here to make good allowance for changes in the vehicles via which investors enter bond and equity markets, and this may have important implications for relative tax rates. Therefore the estimates should be seen very much as a tentative step towards understanding the impact of changing taxation arrangements on excess returns.

Rough estimates have also been made of transaction costs and liquidity premium effects drawing on parameters from the literature. The price of equity market transactions has undoubtedly fallen since the early 1980s, but at the same time turnover has risen, with the result that transaction costs are not necessarily much lower. However, lower transaction prices mean that markets are more liquid and that it is much less costly to optimise portfolio structure. When one factors portfolio adjustment costs into equity return calculations, there are grounds to include some "liquidity premium" as a component of the total equity premium. It is estimated in this paper, using parameters from other studies and market data, that there has been a reduction in this liquidity premium of at least 1 per cent in Australia since the early 1980s. And while the data do not exist to make comparable estimates in the more distant past, there is a possibility that the liquidity premium earlier in the 20$^{th}$ century was even greater than in the 1980s – which offers some explanation for the apparent fall in excess returns since the 1950s.

The final section of the report considers the potential for biases in excess returns. It is concluded that there may have been a small upward bias in excess returns as a result of a small and unanticipated downward trend in gross real rates of return. This trend reflects falls in real interest rates and the (il)liquidity premium for equities, offset to some extent by less favourable tax treatment. The combination of these factors may have lent an upward bias to the excess return realisations of the order of ½ per cent – obviously a small amount and subject to some considerable uncertainty.

There is still considerable uncertainty about the true value of market risk premium that should be entered into regulatory pricing models. This report demonstrates that the use of a very long time series of historic data probably confounds the analysis by mixing together periods in which the market risk premium was different. Although the statistical evidence is not overwhelming, it is suggestive of an average around 6 per cent over the last 30 years or so. Trends in transaction costs will have given downward impetus to the MRP over that period, although it is possible that tax effects have had some offsetting influence. Taking all these factors into account, the most likely value of the MRP is in the range 5 to 6 per cent, but with significant uncertainty still attached.

# 1. Introduction

## 1.1 Structure of this report

In Australian electricity regulatory decisions the most common assumption regarding the value of the market risk premium (MRP) is 6 per cent. However, there is a divergence of views on the appropriateness of this judgment. On the one hand it is argued that a long series of historical excess returns shows a 1-year average MRP of over 7 per cent, and that the use of even a 6 per cent MRP has scant statistical support. On the other hand, it is argued that plausible views as to growth prospects for dividends, and surveys of financial market practitioners' expectations, both suggest an MRP below 6 per cent.

This study is primarily concerned with the re-examination of historic excess returns. Section 1.2 of this introduction briefly sets out the connection between the MRP and regulatory cost of capital and considers how the MRP might be estimated from historical data.

An important question is whether or not the MRP can be regarded as a stable quantity through time. If it can, then estimation of the MRP can be carried out with the largest sample possible, making use of indicators of the MRP from many years ago. On the other hand, if the MRP changes through time, then it will be desirable to discount or even ignore data from many years ago.

Section 2 of this report considers the historic excess return data that is available for Australia. Although three series are considered, two warrant the most attention: the Officer series which provides 122 annual observations over the long period from 1883 to 2004, and the AGSM series which provides 360 monthly observations spanning the period from January 1974 to December 2003. Estimates of mean excess returns are made for different periods, including an efficient weighted estimator which estimates a 122 year mean with greater precision than the methods used in the past. The important issue of bias is also explored.

Section 3 turns to the critical foundational assumption for employing long data series to estimate the MRP: the assumption that the MRP is stable over long periods of time. Standard t-tests have in the past been unable to reject the hypothesis of "constant MRP", but it has been recognised that the tests may lack power. In Section 3 the issue of power issues is explored with simulations. The power of t-tests is found to be very weak. To further investigate the question of structural stability, the stability of the variance of excess returns is explored, and it is found that the distribution has not been stable in this respect. The stability of mean excess returns is then further investigated with allowance for autocorrelation. Autocorrelation can undermine the power of conventional statistical tests, and by making allowance for it more efficient tests may be carried out. After autocorrelation is allowed for, the weight of evidence against a constant mean is reinforced.

Although there is evidence for a structural break in the MRP, the values of it through time are unknown. Section 4 applies alternative smoothing techniques to isolate an underlying trend in the equity premium. The alternative techniques are compared with each other by assessing their out-of-sample prediction performance.

Section 5 raises two important issues which have hitherto had limited attention in analysis of the excess returns – taxes on the investor and transaction costs and liquidity premiums. It is common to ignore these items and to equate excess returns with a pure market risk premium. But in fact they can be expected to combine with the pure market risk premium to generate the investor's expected excess return. It follows that as taxes at the investor level and transactions costs and liquidity premiums change, then the expected value of excess returns is likely to change too.

Section 6 considers the potential for bias in excess return series and its possible influence on the more recent Australian data.

Appendices cover the use of lognormal estimators of the mean, the properties of Officer's long data series, and the construction of weighted estimators of maximal efficiency.

## 1.2 Conceptual foundations

***Relevance of efficient costs***

It is common in Australia and elsewhere for governments to regulate the prices charged by utilities. The rationale for price regulation is to prevent those utilities from exploiting their monopoly positions to set prices in excess of what is required for the recovery of reasonable costs. Such regulation is widespread in industries such as electricity, gas, telecommunications, water and transport, particularly in respect of network infrastructure.

To regulate in this way, regulators need to make assessments of what efficient costs actually are. One of the particular challenges that arises is the need to measure an efficient cost of capital in respect of capital investments.

***Definition of the cost of capital***

Capital is provided by investors. In its simplest form investment involves an up-front payment by an investor in return for which she receive the rights to a future income stream. The amount of the future income stream is usually uncertain (although there are some exceptions such as certain government securities).

Define the simple gross return on an investment, R, in terms of the investment I and the payoff P:

$$R = \frac{P}{I} \tag{1.1}$$

The simple net return then is $R-1$.

In the case where an investment's returns are uncertain, there is a variety of outcomes which could eventually occur, with differing returns attached to each. However, if the investor can form a view about the range of possible outcomes and their likelihood, then she can form an expectation over the set of possible returns. This is just the probability-weighted average of the return under each possible outcome. It is:

$$\mathrm{E}(R) = \sum_j prob(outcome = j) R_j \tag{1.2}$$

Where $prob\,(outcome = j)$ is the probability that outcome j occurs and the summation is over all the possible outcomes.

$\mathrm{E}(R)$ is known as the "ex ante" gross rate of return and is also a measure of the ex ante "price" paid for capital and as such is a measure of the cost of capital.

The expectation in Equation 1.2 can also be expressed in terms of the payoff term in Equation 1.1. Summing over each of the possible payoffs $P_j$ :

$$\begin{aligned}\mathrm{E}(R) &= \sum_j prob(outcome = j)\frac{P_j}{I} \\ &= \frac{\sum_j prob(outcome = j)P_j}{I}\end{aligned} \qquad (1.3)$$

For any given distribution of payoffs, an increase in the price of the investment, I, implies a lower expected return and thus a lower cost of capital, and conversely a decrease in I implies a higher expected return and a higher cost of capital.

Each of the possible return outcomes $R_j$ can be expressed as the sum of an ex ante interest rate $i$, which is common to all the possible return outcomes and therefore is not subscripted, and an "excess return" outcome $\pi_j$

$$R_j = 1 + i + \pi_j \qquad (1.4)$$

Taking expectations across the terms of Equation 1.3, and noting that the expectation of something that is known with certainty is just itself, we get

$$E(R) = 1 + i + E(\pi) \qquad (1.5)$$

Thus the expected return is equal to 1 plus the interest rate plus the expected excess return.

### *The cost of capital for the "market"*

It is easy to apply these simple relations to the case where the investment under consideration is an investment in the whole market. Letting $R_{Mj}$ denote the return on the market under outcome j and $\pi_{Mj}$ the excess return on the market under outcome j, substitution into Equation 1.4 gives

$$E(R_M) = 1 + i + E(\pi_M) \qquad (1.6)$$

The return on the market $R_M$ is not known with certainty and as such it has a risk attached to it. The quantity $1+i$ is known with certainty and can thus be regarded as a riskfree element of the expected return. The component $E(\pi_M)$ is risky. It is called the "market risk premium" (MRP).

As was noted above, for a given expectation over future cashflows, the expected return on the market and the market risk premium depend on the price of an investment. The same is true of the market. If the demand for investments in the market exceeds the supply at some particular value of the MRP, then the price of the market is bid up and the MRP is reduced; the converse is true if the supply exceeds the demand.

### *A framework for estimating the MRP*

The MRP is defined as the expected value of excess returns. The MRP at any point in time is therefore a parameter pertaining to an unobservable statistical distribution. The only evidence that is seen of the MRP is a single ex post realisation, and it is not possible on the basis of this to infer much about the underlying distribution and its parameters.

Typically the way around this difficulty is to assume that a time series of historic excess returns represents repeated draws from a stable distribution. If that assumption can be maintained, then an estimate of the MRP and, just as important, the confidence intervals pertaining to an estimate, can be estimated from the sample of excess returns.

The simplest estimator of the mean 1-period return is the arithmetic average of a sample of 1-period returns. An arithmetic average is an unbiased estimator of the mean. As such the arithmetic mean is an obvious starting point in the search for an estimator of 1-period returns. The precision of the estimate obtained will depend on the sample size used for estimation.

The geometric mean, defined as the $T^{th}$ root of a series of T returns, is not an unbiased estimator of the 1-period expected return. This is simply a statistical truism.

However, it can be shown that if returns are drawn from a lognormal distribution, the expected value of returns is given by

$$E(R) = e^{(\mu+\sigma^2/2)} \tag{1.7}$$

where $\mu$ and $\sigma^2$ are respectively the mean and variance of $\ln R$. It can be shown that the quantity $e^{\mu}$ is the geometric average of the series of returns. This is an alternative estimator of the expected value of returns. It can then be used to estimate the expected value of the MRP by deducting interest rates. The disadvantages of this estimator (over and above its relative obscurity!) are that it relies on normality and that it does not have the unbiasedness property of a simple arithmetic average. Its great advantage is that it enables the generation of multi-period expected returns and associated statistical confidence intervals from 1-period expected returns, in a way that an arithmetic average of 1-period returns cannot. For instance, as will be seen, it can be used to generate estimates of the 1-year MRP from monthly return data. Moreover, multiperiod estimates can be made which include exact adjustments for the distorting effects of autocorrelation. These issues have been explored in work by Blume (1974), Cooper (1996) and Jacquier, Kane and Marcus (2003). There is further discussion of them in Appendix A. There are also more accessible discussions in Wright, Mason and Miles (2003) and Lally (2004).

More complicated alternatives relax the assumption of constancy in the underlying distribution of excess returns, and instead allow it to evolve over time in a deterministic way. Conditional expectations such as the mean, the variance and covariances may then vary over time. The conditioning could be on exogenous variables or on the past history of the excess returns themselves. The attraction of such estimators is that allowing for a time-varying MRP may allow both better prediction at a point in time and lower estimation errors with consequently greater precision of estimates.

Rather than seeking to further explore these ideas here, without direct context, they are discussed further when they arise in the report.

# 2. Estimating the Market Risk Premium from Long Historical Data Assuming Stability of Excess Return Distributions

## 2.1 Description of Datasets

In this section three historic excess returns series are analysed to draw out evidence on the market risk premium. In each case the excess return is calculated as the difference between the equity return over a period and the 10-year bond rate at the beginning of the period. The three series are:

- Professor Officer's long series of Australian excess returns, comprising 122 annual observations from 1883 to 2004;
- a series based primarily on Ibbotson Associates (2004) series of Australian excess returns, comprising 35 annual observations from 1970 to 2004; and
- a series based on the Australian Graduate School of Management's *Share Price and Price Relativities Data File,* comprising 372 monthly excess return observations from January 1974 to December 2004.

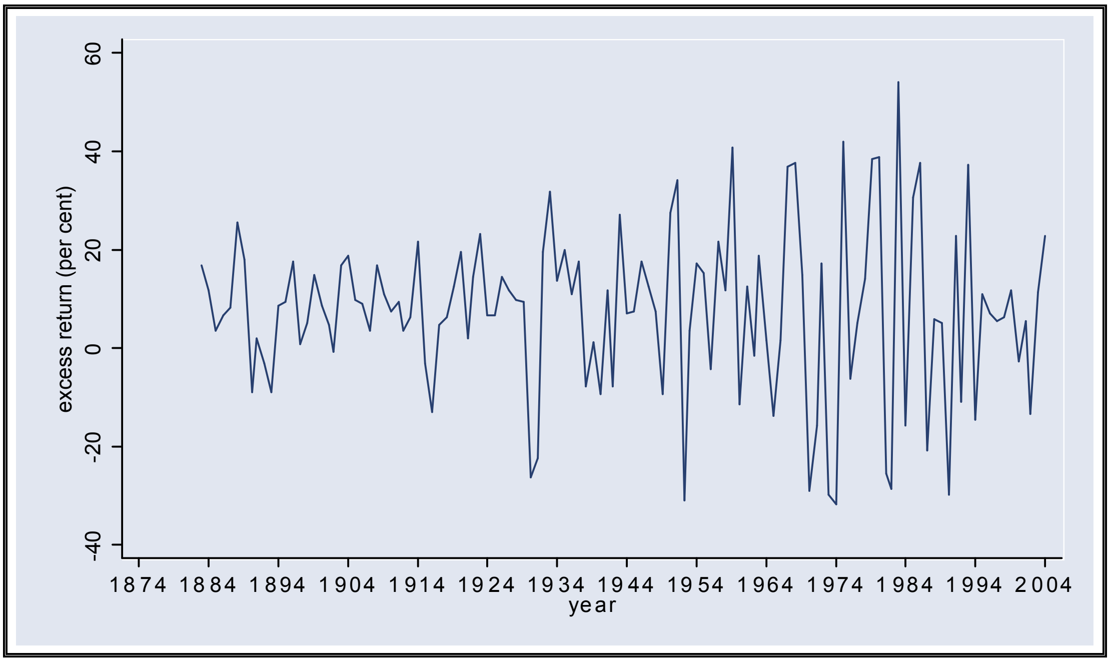
Each of these series is based on broad-based Australian equity accumulation indexes which capture returns in the form of dividends paid and capital gains. Essential Services Commission of Victoria (2005) points out that allowance should also be made for the value of franking credits which have been available since the middle of 1987 and suggests that a figure of 0.6 per cent per annum is appropriate. This figure is similar to Hathaway's (2005) estimate of 0.53 points per annum. The ESC suggestion has been adopted and the excess return data from these series has been increased by a value of 0.6 per cent per annum from July 1987 onward.

Professor Officer kindly provided the data which he used to produce summary statistics published in Victorian Essential Services Commission (2002b); in turn they on build on his 1989 study of long term Australian equity returns. The data that he provided covered the period 1883 to 2000, and are the longest available consistent series of broad Australian equity market returns. For this report, observations for the years 2001 to 2004 were constructed from Australian Stock Exchange and Reserve Bank of Australia data and appended to Officer's data. This gives a series of 122 annual excess return observations for the years 1883 to 2004. They are plotted in Figure 2.1, and Appendix B provides more information about the dataset.

Ibbotson Associates data is based on MSCI indices. The "long horizon" series has been used, which uses a bond rate as the riskfree interest rate proxy. The Ibbotson data cover the period to the end of 2003, and a 2004 observation has been added, calculated as the difference between the 10-year bond rate and the ASX 200 accumulation index for 2004.

The AGSM data covers the period 1974 to 2003, and it has been supplemented by including observations for each of the months of 2004, calculated as the difference between the 10-year bond rate and the ASX 200 accumulation index. The AGSM data has the greatest breadth of coverage of companies. This is an advantage because the greater the breadth of coverage the more likely a measure is to correlate with the universe of available investment opportunities. As Lally (1995) shows, indexes which incompletely cover the market portfolio may bias estimates of cost of capital. The fact that the AGSM series is available monthly means that it also provides a large number of relatively recent observations; with 360 observations it is the largest sample that we have. Annual equity returns were compiled from the monthly data by compounding monthly relatives. To get the annual excess return, the beginning of period bond rate was then subtracted. This procedure yielded 361 overlapping and 31 non-overlapping annual excess return observations.

**Figure 2.1**
**Excess Returns from 1883 to 2004**




## 2.3 Arithmetic average estimators of mean 1-year excess returns

Many equities are common to each of these indexes, and that commonality is even more pronounced in value-weighted terms. Therefore it is to be expected that the indexes will produce similar results when considered over common periods, and the data in Table 2.1 bear this out. Take the period 1974 to 2004, for instance. The Officer dataset had an average excess return of 6.8 per cent, which compares with 6.4 per cent for Ibbotson Associates. The average of AGSM excess returns for the calendar years 1974 to 2004 was 6.7 per cent. The average of AGSM 12-month excess returns for the 361 months December 1974 to December 2004 was 6.6 per cent (but this involves overlapping observations).

The correlation coefficients between each series are either 0.99 or 0.98, which are very high. This is not surprising; many stocks are common across the datasets.

Table 2.1 also includes standard errors for the average excess returns. Over the period 1974 to 2004, for the 3 non-overlapping cases, the standard errors are 4.2, which means that the associated 95-per cent confidence intervals for mean excess returns range from about –2 per cent to 15 per cent – a 17 per cent range, which is very wide.[1]

[1] This reckoning is based on the 2 standard errors rule for a 95-per cent confidence interval.

**Table 2.1**
**Comparison of excess returns from Officer, Ibbotson Associates and AGSM datasets**

| | Officer data | Ibbotson Associates data | AGSM data non-overlapping | AGSM data overlapping |
|---|---|---|---|---|
| **1883 to 2004:** | | | | |
| Average | 7.4 | n.a. | n.a. | n.a. |
| Std Dev of Observations | 16.9 | n.a. | n.a. | n.a. |
| Std Error of Average | 1.5 | n.a. | n.a. | n.a. |
| **1970 to 2004:** | | | | |
| Average | 4.3 | 3.9 | n.a. | n.a. |
| Std Dev of Observations | 23.6 | 23.4 | n.a. | n.a. |
| Std Error of Average | 4.0 | 4.0 | n.a. | n.a. |
| Corr coeffs | | | | |
| vs Officer | - | 0.99 | n.a. | n.a. |
| Std Dev | | | | |
| **1974 to 2004:** | | | | |
| Average | 6.8 | 6.4 | 6.7 | 6.6 |
| Std Dev of Observations | 23.0 | 23.0 | 23.1 | 21.4 |
| Std Error of Average | 4.2 | 4.2 | 4.2 | 1.1(1)<br>3.1(2) |
| Corr coeffs | | | | n.a. |
| vs Officer | - | 0.99 | 0.99 | n.a. |
| vs Ibbotson | - | - | 0.98 | n.a. |

(1) Conventional calculation.
(2) Newey-West standard error.

The overlapping AGSM data is interesting because, by virtue of the fact that it uses monthly data, it has many more observations than the other datasets. Moreover, they are relatively recent. The conventional standard errors have been calculated at 1.2 percentage points. However, the overlaps introduce autocorrelation to the data, which means that the conventional standard errors are not valid. Therefore Newey-West standard errors, which are asymptotically valid, have been presented too. They suggest a 95-per cent confidence interval ranging from around zero to about 13 per cent. Obviously this is still a very wide range.

## 2.4 Lognormal estimator of mean 1-year excess returns

While the Newey-West standard errors make allowance for autocorrelation (and heteroskedasticity) of unknown form in the series of 12-month excess returns, it is possible that an approach that makes explicit allowance for that autocorrelation would do better. We know that the overlapping nature of the observations builds in autocorrelation, and direct analysis of the monthly returns would avoid that problem.

Unfortunately the relation between monthly and 12-monthly excess returns is messy. More straightforward is the relation between a 12-month return and the 12 single monthly returns. It is exactly

$$R(12) = \prod_{t=1}^{12} R_t \qquad (2.1)$$

where $R(12)$ is a 12-month return and the $R_t$ are 12 monthly returns that it is made up from.

Let $i_t$ and $\gamma_t$ denote interest and excess return elements which combine multiplicatively to give the return in period t

$$R_t = (1+i_t)(1+\gamma_t) \tag{2.2}$$

Then Equation 2.1 can be rearranged to get

$$\begin{aligned} R(12) &= \prod_{t=1}^{12} R_t \\ &= \prod_{t=1}^{12} (1+i_t)(1+\gamma_t) \\ &= \prod_{t=1}^{12} (1+i_t) \prod_{t=1}^{12} (1+\gamma_t) \\ &= (1+i(12))(1+\gamma(12)) \end{aligned} \tag{2.3}$$

where $i(12)$ and $\gamma(12)$ are multiplicative interest and risk premium elements making up the 12-month return.

In contrast, the "excess return" over 12 months, denoted $\pi(12)$ is given by

$$\pi(12) = R(12) - i(12) \tag{2.4}$$

and because this relationship is additive it is not possible to express $\pi(12)$ as a product (or for that matter a sum) of monthly excess returns $\pi_t$. For this reason it is more fruitful to build an estimate of the 12-month average excess return from multiplicative monthly risk premia.

Equation 2.3 lends itself to a logarithmic decomposition

$$\begin{aligned} \ln R(12) &= \sum_{t=1}^{12} \ln R_t \\ &= \sum_{t=1}^{12} \ln(1+i_t) + \sum_{t=1}^{12} \ln(1+\gamma_t) \end{aligned} \tag{2.5}$$

Turn now to the case where $i_t$ and $\gamma_t$ are stochastic. If we assume that $1+i_t$ and $1+\gamma_t$ are lognormally distributed and independent of each other and that their logs have respective means $\mu_i$ and $\mu_\gamma$ and variances $\sigma_i^2$ and $\sigma_\gamma^2$ then $R_t$ will also be stochastic with a lognormal distribution, and its log will have mean $\mu_R = \mu_i + \mu_\gamma$ and variance $\sigma_R^2 = \sigma_i^2 + \sigma_\gamma^2$.

Furthermore, $R(12)$ is also stochastic. So long as there is no autocorrelation in its constituents, then it too will be lognormally distributed and its log will have a mean

$$\mu_{R(12)} = 12\mu_R$$
$$= 12(\mu_i + \mu_\gamma)$$

and a variance

$$\sigma^2_{R(12)} = 12\sigma^2_R$$
$$= 12(\sigma^2_i + \sigma^2_\gamma)$$

If the log of a lognormally distributed variable has mean $\mu$ and variance $\sigma$, it can be shown that the mean of the variable (which is by definition its expected value) is

$$E(R) = \exp(\mu + \sigma^2 / 2) \qquad (2.6)$$

It can be shown that the expected value of $R(12)$ is then

$$E(R(12)) = \exp(\mu_{R(12)} + \sigma^2_{R(12)} / 2) \qquad (2.6)$$

By taking estimates of parameters, the expected value of $R(12)$ and its confidence intervals can be obtained.

In making such a calculation it is convenient to recognise that the risk free interest rate for any 12 month period can be observed at the outset with virtual certainty. For a given interest rate then, $\sigma^2_i$ is zero and $\sigma^2_{R(12)} = 12\sigma^2_\gamma$. But the estimate of the 12-month expected return, and ultimately the 12-month market risk premium, will depend on the interest rate at which the expectations are evaluated. An obvious candidate is the average interest rate available over the 372 months for which excess returns observations are taken.

This estimator is described as a "lognormal estimator from monthly returns". The calculation is shown in Table 2.2. The data are presented in decimal form. Converting to percentages, the estimate of the 12-month expected excess return is 6.4 per cent. The lower boundary of the 95 per cent confidence interval is 4.2 per cent and the upper bound is 8.6 per cent.

Thus, by making use of the monthly return data from AGSM with a lognormal estimator, we get a narrower error bound on the estimate of the MRP than is possible with the long Officer series. In each case we estimate mean 12-month excess returns. The "2 standard errors" range for the "monthly data estimator" is 4.5 percentage points. The 2 standard error range for the Officer data is 6.2 percentage points.

It is interesting to note that when the lognormal estimator is applied to the Officer data series, it estimates the mean for that series at 7.37 per cent which is very close to the arithmetic average estimate of 7.29 per cent. The lognormal estimator behaves perfectly sensibly in that case.

Are there any drawbacks to the lognormal estimator? Potential limitations are:

- It is an asymptotic estimator. While it is unbiased asymptotically, it may be biased in finite samples. However, the biases get smaller as samples get larger. It is notable that the value of the lognormal estimator of the 12-month excess return, 6.4 per cent, is close to the values of two alternative unbiased estimates – 6.7 per cent for the non-overlapping average and 6.6 per cent for the overlapping average calculated out of the AGSM dataset.

- It does rely on there being no autocorrelation in the monthly (multiplicative) excess returns. The only hint of autocorrelation in the correlogram and Breusch-Godfrey tests was at 1 lag, at a significance level of around 15 per cent. A variance ratio test at 12 months found no significant evidence of autocorrelation. (Appendix C presents details of an autocorrelation adjustment, but it has not been carried out here.)
- The monthly excess log returns are not normal (the Jarque-Bera test strongly rejects this), which violates an underlying assumption with the lognormal estimator. The consequences of this are unknown.
- It is assumed that there is no structural discontinuity in the data. Such an assumption is common to all the procedures which involve averaging of historic excess returns, such as the widely used averages of the long Officer series.

**Table 2.2**
**Lognormal estimator of mean 12-month returns and their 95 per cent confidence interval**

| **Item** | **Calculation basis** | **Value** |
|---|---|---|
| **Input data** | | |
| $\ln(E(1+i))$ | Log of sample average of monthly long interest rate relatives | 0.00781983 |
| $\hat{\mu}_\gamma$ | Sample average of logs of monthly $\gamma_t$ | 0.00331662 |
| $\hat{\sigma}_\gamma^2$ | Sample variance of logs of monthly $\gamma_t$ | 0.00282259 |
| $N$ | Number of monthly observations | 372 |
| **Calculations:** | | |
| $E(1+i(12))$ | $=(E(1+i))^{12}$ | 1.09838181 |
| **Central estimate of expected return on market:** | | |
| $E(R)$ = 1-month expected | $E(R)=\exp(\ln(E(1+i))+\hat{\mu}_\gamma+\hat{\sigma}_\gamma^2/2)$ | 1.01263065 |
| $E(R(12))$ = 12-month expected | $E(R(12))=(E(R))^{12}$ | 1.16255320 |
| Variance of $R(12)$ | $VarR(12)=\exp(12(2(\ln E(1+i)+\hat{\mu}_\gamma)+\hat{\sigma}_\gamma^2))(\exp(\hat{\sigma}_\gamma^2)-1)$ | 0.04668954 |
| Variance of estimator | $VarR(12)/N$ | 0.00012551 |
| Standard error of estimator | Square root of variance of estimator | 0.01120310 |
| **Estimates of MRP:** | | |
| Estimate of 12-month MRP | $MRP(12)=E(R(12))-E(1+i(12))$ | 0.06417139 |
| Lower 95% confidence interval | $=MRP(12)-1.96 xs \tan darderror$ | 0.04221330 |
| Upper 95% confidence interval | $=MRP(12)+1.96 xs \tan darderror$ | 0.08612947 |

While there are some remnant uncertainties about the properties of the lognormal estimator, there is no evidence that it is fundamentally flawed, and its estimate of the expected value of 12-month excess returns coheres with two more traditional estimates of that quantity. Its major advantage is that it has quite a narrow confidence interval.

If the lognormal estimator based on monthly data is used, the conclusion that arises from it is that the mean 1-year MRP over the period 1974 to 2004 is "about" 6½ per cent. This is a result close to 6 per cent, and the test does not offer strong support for 6½ per cent in preference to 6 per cent. If the "long series" estimator is used, it implies an MRP of "about" 7½ per cent over the last 122 years. However, a 6 per cent 1-year MRP lies inside 1 standard error of this estimate.

## 2.5 A weighted estimator

The AGSM data, available since 1974, can be combined with data from prior to its inception, such as the pre-1974 data from the Officer series. So long as one maintains the assumption that the distribution is stable, the parameters of the distribution can be estimated and, by virtue of the increased sample, the standard errors potentially can be reduced.

An agglomerated estimator $Z$ of a population mean can be constructed from two independent estimators, $X$ and $Y$, by combining them with weights depending on their respective variances. Except in the trivial case where $X$ or $Y$ have a variance of zero (which would imply that the mean is already known with certainty), the variance of the estimator $Z$ will be smaller than the variance of either of the estimators $X$ and $Y$. Thus the estimator $Z$ will be a more efficient estimator of the mean. Appendix C provides more details.

Under the null hypothesis of a constant mean for the period 1883 to 2004, it is possible to construct two independent estimators as follows. The first estimator is an arithmetic average of the 1-year returns for the period 1883 to 1973 from the Officer series. The second estimator is a lognormal estimator based on AGSM monthly data for the period January 1974 to December 2004. The estimates of the mean and the variances of those estimates are shown in Table 2.3 under "Input Data". The optimal weighted estimate and its variance are also shown. The weighting scheme is set out in Appendix C.

The weighted estimate of the mean, based on 92 observations from the Officer series (1883 to 1973 and then 2004) and 372 observations from the AGSM series (January 1974 to December 2004) is 6.9 per cent. It has a standard error of just 0.9 percentage points which is considerably smaller than the 1.5 percentage point standard deviation on the full Officer series and moderately smaller than the standard error of 1.1 percentage points when the logarithmic estimator is applied to the AGSM monthly data. However, it must be remembered that these calculations are predicated on a constant mean throughout the estimation period.

**Table 2.3**
**Efficient agglomerated estimator of mean 1-year returns and their 95 per cent confidence interval**

| Item | Calculation basis | Value |
|---|---|---|
| **Input data:** | | |
| X | Average of annual excess returns 1873 to 1973 from Officer series | 7.737391 |
| Var X | Variance of average excess returns 1873 to 1973 from Officer series | 2.248630 |
| Y | Logarithmic estimator of 12-month return from AGSM data 1974 to 2004 | 6.417139 |
| Var Y | Variance of logarithmic estimator of 12-month return from AGSM data 1974 to 2004 | 1.255095 |
| **Weighted estimate:** | | |
| Z | Calculated as per Appendix C | 6.890076 |
| Var Z | Calculated as per Appendix C | 0.805498 |
| Standard error of Z | | 0.897496 |

# 3. Is There Evidence of Structural Change in Excess Return Distributions?

One of the difficulties with estimation models which assume stability of underlying distributions is that the assumption is intuitively difficult to justify. As Merton (1992) notes, if the risk associated with the market varies then the MRP could be expected to vary too. And if aggregate risk aversion, which could be expected to reflect preferences and the distribution of wealth, changes slowly over time, which seems a reasonable assumption, then it would have an impact on the MRP too. In addition, if personal taxes, transactions costs, liquidity premia vary over time they may also have consequences for the distribution of excess returns.

Moreover, investors just do not have perfect knowledge of the distribution of returns, and their perceptions will be moulded by emerging information. For instance, is it plausible to assume that investors expectations about returns were unaffected by the Depression? Or is the Depression more realistically interpreted as, in the words of Cogley and Sargent (2004 – p.27), an episode "shattering the representative consumer's beliefs about the likelihood of expansions and contractions"?[2]

The difficulty with identifying a structural change, however, is that there are no strong a priori grounds to suggest when structural changes might have occurred. And even if events can be identified that are so fundamentally important as to be likely to cause structural change, they may take effect over quite protracted periods. Entering a dummy variable at the time of the fundamental event may not capture the influence of the structural change very well at all.

## 3.1 Tests for Equality of Means

One of the simplest forms of structural change is a once-and-for-ever change in the mean.[3] For instance, we could consider the case where there is a change in the mean occurring in Year X. If we let A be the set of excess returns occurring before Year X and B the set of observations in Year X and afterward

$$\pi_t = \begin{array}{ll} \pi_A + \varepsilon_i & t \in A \\ \pi_B + \varepsilon_i & t \in B \end{array} \tag{3.1}$$

In that case we test for equality of $\pi_A$ and $\pi_B$ .

Gray (2001) considered such a model using the Officer data set with updates to 2000. He sought to identify a series break by breaking the data set in two and comparing the means of the various pre- and post-break periods. He considered breaks for each individual year from 1960 through 1985. Using t-statistics to test for equality of means, he found that in no case was there a statistically significant difference at the conventional 5 per cent significance level. Indeed in only one case could a significant difference be identified even at a 10 per cent significance level.

2 As they observe, "It certainly shattered prevailing opinions among economists. For example, witness the evolution of Keynes's thinking as he passed from the orthodoxy of A Tract on Monetary Reform (1923) to The General Theory of Employment, Interest and Money (1936)."

3 Another very simple alternative is to allow for a linear time trend in the ex ante excess return, but this makes little sense as it implies that it either (a) gets larger and larger or (b) becomes negative and more and more so, ad infinitum.

Essential Services Commission Victoria (2002b) noted that, in assigning weight to Gray's evidence:

> A general problem with the application of empirical tests to actual equity returns in general is that the variability in annual equity returns makes it very difficult to discern changes to the underlying market risk premium. Specifically in relation to the test performed by Professor Gray, this meant that while a lot of comfort could have been taken if the hypothesis of 'no change' had been rejected, the failure to reject the hypothesis does not provide much positive for the proposition that the mean has remained unchanged. [p. 326]

It notes that, in statistical terms, these t-tests lack "power", which means that they run a potentially large risk of accepting the hypothesis that there has been no change in the market risk premium even when there has.

To investigate this issue further, we carried out a Monte Carlo study to clarify the power of these tests. The purpose of the Monte Carlo study was to answer the following question: If there was a fall in the market risk premium of x percentage points, what is the chance that the t-tests would confirm that this had happened? To answer that question we considered the case of a constant market risk premium for a 90 year period, and then an x per cent reduction to a new, constant market risk premium for the following 32 years. In terms of the actual dataset this corresponds most directly to an assessment of the power of a test for a structural break occurring in 1973, but similar results could be expected for other break points. We considered several sizes of break – reflected in the value of x – ranging from 1 to 10 percentage points.[4]

To implement the Monte Carlo study 10,000 sets of 122 observations were drawn (with replacement) from the set of excess returns observed from 1883 to 2004.[5] For each of these samples, the value of each observation was then reduced by x per cent from 1973 onward, for different values of x. Thus, by construction, we know that the mean for the period 1973 to 2004 is x per cent lower; the question is how effective the t-tests are in detecting it.

If a t-test of one of these data sets rejects the hypothesis that there has been a reduction in the average, then the test has failed as, by construction, we know that there actually was a reduction of x per cent. For each of the 10,000 simulations, t-statistics were calculated and, for a range of significance levels, the decision was made whether to accept or reject the null

---

4 We commence the Monte Carlo study with the null hypothesis that there is a constant equity premium. Under this assumption we can draw (hypothetical) alternative realisations of the ex post equity premium for the years 1883 to 2004 by drawing (with replacement) 122 values from the actual ex post equity premia. This was done with a random number generator and a 1-in-122 chance of selection for each observation in each drawing. The sample forms a sequence, and the first 90 observations in the sequence are taken as is, but x percentage points are deducted from the last 32 observations. We then test the hypothesis that the mean is lower for the last 32 (which have been adjusted) than for the first 90. We choose a significance level at which we will accept or reject the null hypothesis of no-change. If we accept the null hypothesis, then the test has failed because we know that, in fact, the mean has changed. We carried out 10,000 iterations of this procedure to calculate the power at different size structural breaks and significance levels.

5 The dataset includes a franking credit value of 0.6 per cent per annum after the introduction of dividend imputation in mid 1987.

hypothesis of "no change".[6] For each significance level the number of rejections was counted over all 10,000 simulations. The number of rejections divided by 10,000 then gives the proportion of occasions on which the "no change" hypothesis was rejected, and this proportion measures the power of the test. A power of 100 per cent would mean that the test always worked, while a power of zero would mean that it never worked.

Table 3.1 summarises the results. The first column shows the power of the test when the statistical testing uses a 5 per cent significance level, which is common in statistical analysis and is what was employed by Gray. If the reduction in the market risk premium is 1 percentage point, the t-test detects this structural break only 8 per cent of the time. Small variations are of course hard to detect. But if there was a 5 percentage point break the test would detect that a change had occurred only 40 per cent of the time. While such a change might not be statistically significant, it certainly is economically significant.

The second and third columns of Table 3.1 show the power when lower significance levels are adopted. As would be expected a lower significance level does improve the power of tests. But power generally remains low. Consider the case where the data contain a 5 per cent reduction in the market risk premium. At a 10 per cent significance level, the test succeeds only 55 per cent of the time, while at a 25 per cent significance level it succeeds 77 per cent of the time.

A key conclusion that emerges from these power tests is that t-tests at the conventional 5 per cent statistical significance level on the annual historic returns series are unlikely to detect some changes in the MRP which are of considerable economic significance.

**Table 3.1**
**Power of tests of a changed market risk premium**

| | **significance level** | | |
|---|---|---|---|
| | **5 per cent** | **10 per cent** | **25 per cent** |
| | **power (per cent)** | **power (per cent)** | **power (per cent)** |
| 1 percentage point | 8 | 15 | 33 |
| 2 percentage points | 13 | 22 | 44 |
| 3 percentage points | 20 | 32 | 56 |
| 4 percentage points | 29 | 43 | 67 |
| 5 percentage points | 39 | 55 | 77 |
| 6 percentage points | 51 | 66 | 84 |
| 7 percentage points | 63 | 75 | 90 |
| 8 percentage points | 73 | 83 | 94 |
| 9 percentage points | 81 | 90 | 97 |
| 10 percentage points | 88 | 94 | 98 |

6 The statistic is calculated as $t = \dfrac{\hat{\pi}_A - \hat{\pi}_B}{\sqrt{\dfrac{S_P^2}{n_A} + \dfrac{S_P^2}{n_B}}}$, where $S_P^2 = \dfrac{(n_A - 1)S_A^2 + (n_B - 1)S_B^2}{n_A + n_B - 2}$ is the pooled variance estimator, $\hat{\pi}_A$ and $\hat{\pi}_B$ are the means, $S_A^2$ and $S_B^2$ are the variances, and $n_A$ and $n_B$ are the sample sizes for sub-samples A and B. It could be argued that Satterthwaite's procedure should be used to allow for unequal variances because, as will be shown subsequently, there is strong evidence against a constant variance over time. Power tests were also conducted on this basis. The power of the t-tests using that procedure was very similar to the power under the "pooled variance" approach.

## 3.2 Tests for Equality of Variances

This testing can be extended to check for a structural break in variances. It is apparent even from visual inspection of Figure 2.1 that excess returns became more volatile after about World War II. We carried out tests for equality of variances in different time periods, in a similar style to Gray's tests for equality of means. The tests overwhelmingly reject the hypothesis of constant variances. Results are presented for 1904 through 1984 at 10 year intervals in Table 3.2. Each of the tests but that for 1984 reject the hypothesis of constant variance; by 1984 the early period included the volatile years from the late 1940s to 1983 and thus there was no apparent change in variance for this comparison. The rejections for 1904 through 1974 were highly significant and in some cases extremely so. For instance, the F-test for 1944 rejects the hypothesis of constant variances at a significance level of 0.00006 per cent. The issue of non-constant variances will be revisited in a consideration of the time series properties of the data.

**Table 3.2**
**Tests for a Break in the Variance of Excess Returns**

| Date of Break | F-stat | Prob(F) |
|---|---|---|
| 1904 | 4.05 | 0.0003997 |
| 1914 | 5.80 | 0.0000004 |
| 1924 | 5.09 | 0.0000001 |
| 1934 | 3.45 | 0.0000050 |
| 1944 | 3.66 | 0.0000006 |
| 1954 | 3.25 | 0.0000031 |
| 1964 | 3.48 | 0.0000010 |
| 1974 | 2.68 | 0.0001832 |
| 1984 | 1.23 | 0.2493631 |

A balance needs to be struck between the significance and the power of tests. The natural tendency is to accept the conventional 5 per cent significance level as it is common in statistical practice, possibly without giving much consideration to the power of tests. But it is desirable to strike a balance having in mind the consequences of the two types of error that can arise – accepting the hypothesis of a lower market risk premium when in fact it has not changed (Type I error) and rejecting the hypothesis of a lower market risk premium when it has in fact changed (Type II error) – and the costs associated with each. Only if one believed that setting the cost of capital too low vastly outweighed the costs of setting it too high could one place so much weight on significance and so little on power. Washusen (2004) reports that MRP parameters of 3 to 4 per cent have been set in the UK without a drying-up of utility investment.

## 3.3 Tests for Autocorrelation and Other Serial Dependence

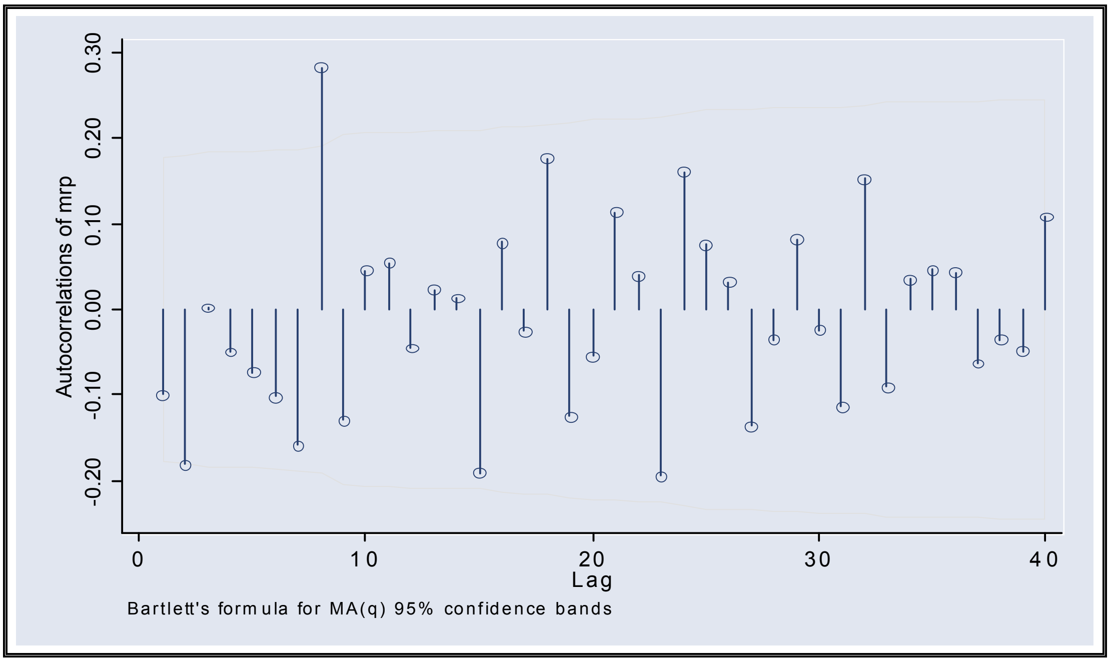
In the analysis thus far the issue of serial dependence in the data has been ignored. However, if there is serial dependence it may have important implications for parameter estimates – both means and standard errors. The time series characteristics of the data, and consequent modifications to the estimation strategy are explored in this section.

A number of influential studies have reported the existence of autocorrelation in both stock market returns and volatilities. For instance in the US Poterba and Summers (1989) found evidence of positive autocorrelations in returns over short periods and negative autocorrelations over long periods. Also in the US, Fama and French (1988) report large negative autocorrelations for return horizons beyond a year, and estimate that 25 to 45 per cent of the variation in three to five year stock returns is predictable from past returns. While the validity of these findings has since been challenged, for instance in Richardson (1991) and Lamoureux and Zhou (1996), they remain influential and there are differences of view as to the existence of autocorrelation in returns.

If autocorrelation exists, then a model structure that allows for it is likely to fit better than one which ignores it. If such a model can significantly reduce the unexplained variation in returns it should have more power to detect structural breaks than a naïve "white noise" model. It is useful therefore to further investigate the time series properties of the Officer data to determine whether autocorrelation is present.

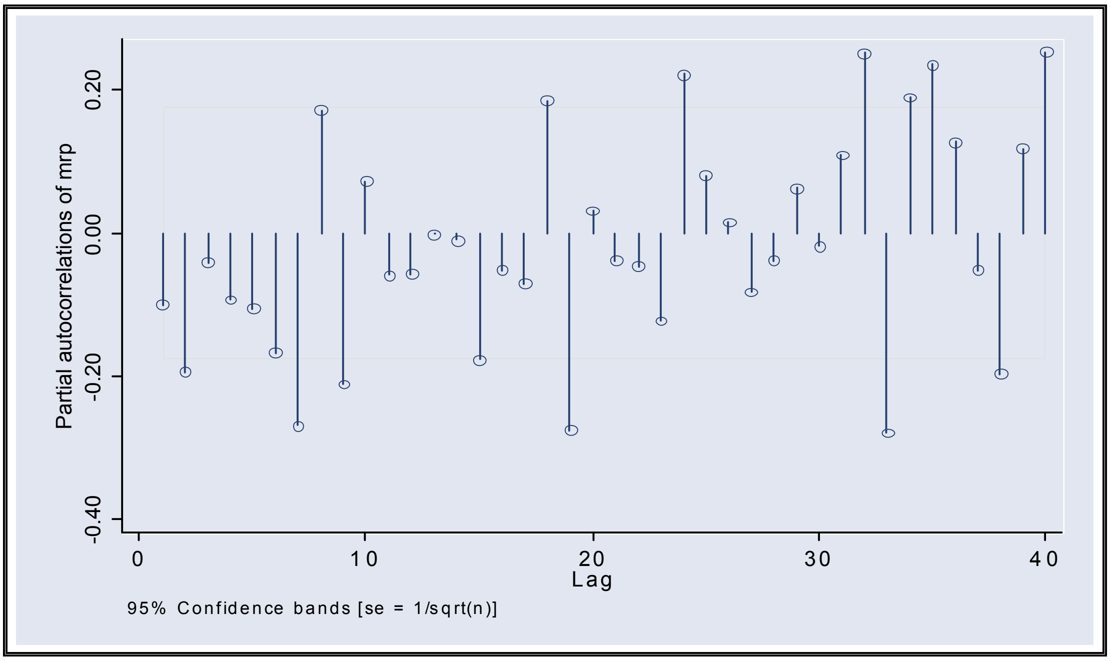
Figures 3.1a and 3.1b show the sample autocorrelation (AC) function and partial autocorrelation (PAC) function for the annual excess return data with 95 per cent confidence bands.[7] These functions can sometimes give a preliminary insight into the nature of autocorrelation. The AC coefficient at 2 lags is negative and almost significant at a 5 per cent level and the AC coefficient at 8 lags is significantly positive. Otherwise the AC coefficients are insignificant. The PAC coefficient at 2 lags is significantly negative. Moreover the first 7 AC coefficients are generally negative.

**Figure 3.1a**
**Autocorrelations of excess returns at different lag lengths with confidence intervals**




[7] These sample autocorrelation coefficients suffer a negative bias. However, most would remain negative if the Fuller bias correction were made, as they would be increased by less than 0.01 (see Campbell, Lo and MacKinlay 1997 p. 46).

**Figure 3.1b**
**Partial autocorrelations of excess returns at different lag lengths with confidence intervals**




At a conventional 5 per cent significance level these patterns do not indicate significant autocorrelation. However, if interpreted literally, the preponderance of negative values of the sample autocorrelation coefficient at short lags is suggestive of a time series process with negative-coefficient moving average elements, with the weak significance of the estimates meaning that not much weight can be placed on this.[8] Such a process could be produced if a positive or negative deviation from mean excess return in one period tends to be followed by deviations of opposite sign over the next few periods. Such a pattern would be observed if stock prices exhibit a degree of mean reversion. However, the sample AC and PAC functions cannot conclusively resolve whether autocorrelation is present. Further tests are needed.

Three commonly used diagnostic tests are the Q ("portmanteau") test developed by Box and Pierce and modified by Ljung and Box, the Lagrange multiplier (LM) test developed by Breusch and Godfrey, and Durbin's "alternative" test.[9] These are tests for strict white noise – i.e., that errors are uncorrelated and have constant variance. It follows that in interpreting the test results we need to be careful not to attribute to serial correlation what could in fact be caused by some form of heteroskedasticity; this problem is considered subsequently.

Table 3.3 shows the (Ljung-Box) Q statistic at different lags and the probability of attaining such values if there truly is no dependence in the data. For the first 7 lags there is not much evidence of autocorrelation. However, at 8 or more lags there is in every instance rejection at

8 Discussions of the visual interpretation of the autocorrelation function can be found in Harvey 1993 Ch. 2, Hamilton 1994 pp. 48-52, Enders 1995 pp. 78-82, and Gujarati pp. 840-845.

9 Perhaps the best known test for autocorrelation is the Durbin-Watson test. However, the Durbin-Watson test is of use only in detecting first order autocorrelation and is thus too limited for our purposes. Greene (2003, pp. 268-271) discusses the LM, Q and DW tests. Harvey (1993, p.77) discusses the connection between the LM and Q tests. Under the null hypothesis the LM and Q statistics share a common asymptotic $X^2$ distribution, but they assume different values in finite samples because of a finite sample adjustment included in the Ljung-Box variant of the Q test. Maddala (2001) cites research that finds Q tests may have weak power and he argues that the LM test generally performs better.

a 5 per cent significance level and in some cases rejection at a 1 per cent significance level.[10] Table 3.3 also shows Breusch-Godfrey (LM) statistics for different lags. There is no evidence of dependence at the 5 per cent level up to 6 lags (although dependence can be rejected at the 10 per cent level for 2 lags). But the LM test shows quite strong evidence of autocorrelation at 7 lags (significant at 0.9 per cent) and at longer lag structures.[11] The Table also shows the Durbin test, and its results are broadly consistent with the Ljung-Box and Breusch-Godfrey tests: strict white noise is rejected at the 5 per cent significance level at 7 or more lags.

**Table 3.3**
**Tests of autocorrelation and autoregressive conditional heteroskedasticity in excess returns**

| | Ljung-Box test | | Breusch-Godfrey test | | "Robust" Breusch-Godfrey test | | Durbin's alternative test | | "Robust" Durbin test | |
|---|---|---|---|---|---|---|---|---|---|---|
| **LAG** | **$X^2$ statistic** | **Prob > stat** | **F statistic** | **Prob > stat** | **F statistic** | **Prob > stat** | **F statistic** | **Prob > stat** | **F statistic** | **Prob > stat** |
| 1 | 1.2513 | 0.2633 | 1.223 | 0.2710 | 0.612 | 0.4356 | 1.215 | 0.2725 | 0.621 | 0.4323 |
| 2 | 5.3694 | 0.0682 | 2.865 | 0.0610 | 1.465 | 0.2353 | 2.933 | 0.0571 | 2.012 | 0.1383 |
| 3 | 5.3699 | 0.1466 | 1.942 | 0.1267 | 0.984 | 0.4031 | 1.974 | 0.1218 | 1.336 | 0.2662 |
| 4 | 5.6814 | 0.2242 | 1.685 | 0.1583 | 0.852 | 0.4954 | 1.712 | 0.1523 | 1.233 | 0.3010 |
| 5 | 6.3723 | 0.2717 | 1.577 | 0.1723 | 0.837 | 0.5265 | 1.604 | 0.1647 | 1.142 | 0.3425 |
| 6 | 7.7216 | 0.2592 | 1.845 | 0.0969 | 1.190 | 0.3171 | 1.917 | 0.0844 | 1.789 | 0.1080 |
| 7 | 10.996 | 0.1388 | 2.705 | 0.0127 | 1.908 | 0.0753 | 3.013 | 0.0063 | 2.340 | 0.0291 |
| 8 | 21.564 | 0.0058 | 2.623 | 0.0117 | 1.617 | 0.1287 | 2.961 | 0.0050 | 2.243 | 0.0298 |
| 9 | 23.787 | 0.0047 | 2.771 | 0.0060 | 1.720 | 0.0936 | 3.241 | 0.0017 | 2.682 | 0.0077 |
| 10 | 24.074 | 0.0074 | 2.529 | 0.0092 | 1.711 | 0.0882 | 2.945 | 0.0028 | 2.573 | 0.0081 |
| 11 | 24.487 | 0.0108 | 2.310 | 0.0144 | 1.537 | 0.1302 | 2.672 | 0.0048 | 2.316 | 0.0142 |
| 12 | 24.768 | 0.0160 | 2.122 | 0.0220 | 1.506 | 0.1351 | 2.435 | 0.0082 | 2.106 | 0.0232 |
| 13 | 24.841 | 0.0242 | 1.941 | 0.0347 | 1.446 | 0.1533 | 2.201 | 0.0150 | 2.022 | 0.0268 |
| 14 | 24.865 | 0.0359 | 1.771 | 0.0548 | 1.391 | 0.1730 | 1.980 | 0.0277 | 1.819 | 0.0470 |
| 15 | 30.01 | 0.0119 | 1.820 | 0.0434 | 1.382 | 0.1730 | 2.078 | 0.0179 | 1.909 | 0.0321 |
| 16 | 30.882 | 0.0139 | 1.702 | 0.0606 | 1.368 | 0.1762 | 1.923 | 0.0281 | 1.827 | 0.0394 |
| 17 | 30.977 | 0.0201 | 1.600 | 0.0816 | 1.297 | 0.2134 | 1.789 | 0.0423 | 1.824 | 0.0373 |
| 18 | 35.493 | 0.0082 | 1.615 | 0.0744 | 1.332 | 0.1894 | 1.832 | 0.0340 | 2.060 | 0.0144 |
| 19 | 37.759 | 0.0064 | 1.747 | 0.0442 | 1.285 | 0.2160 | 2.076 | 0.0125 | 2.279 | 0.0056 |
| 20 | 38.206 | 0.0084 | 1.656 | 0.0592 | 1.230 | 0.2524 | 1.948 | 0.0193 | 2.155 | 0.0084 |

In their standard forms the Ljung-Box, Breusch-Godfrey and Durbin alternative tests are not robust in the presence of heteroskedasticity. However, heteroskedasticity-robust variants can be constructed for the Breusch-Godfrey and Durbin alternative tests (see, respectively, Wooldridge 2001 and 2003). While these robust tests are asymptotic tests, with uncertain

[10] Sensitivity to outliers was tested by replacing the highest and the lowest excess return observations with the mean. When this was done, the Q tests did not reject independence at the 5 per cent level but still rejected at 10 per cent. This illustrates that there is considerable sensitivity to a few observations in the data set. The tests were also carried out for two halves of the data set. For the period 1883 to 1943 serial independence was not rejected, but it was at the 5 per cent level for the period 1944 to 2004. There are 61 observations in each of these two sub-samples, and less when one introduces a lag structure; these sizes may be sufficiently low to hamper statistical inference.

[11] Exclusion of the highest and lowest excess return observations did not materially change the results of the LM test. The test was also applied to two halves of the sample (1883 to 1943 and 1944 to 2004); the picture emerging from this differed from the Q tests in that serial dependence was significant at the 5 per cent level for 1883 to 1943 but not for 1944 to 2004. Of course halving the sample increases the standard errors and can make it more difficult to say things about the data.

small sample properties, it is common to employ them where the data under consideration is known to be heteroskedastic.

The results of these robust tests are also presented in Table 3.3. The robust version of the LM test indicates that the evidence for autocorrelation is substantially weaker than is suggested by the test in its basic form. None of the lags test positive for autocorrelation at a 5 per cent significance level, although there are some positive results at a 10 per cent level. However, the robust form of the Durbina alternative test continues to indicate the presence of autocorrelation at conventional test values. The divergence between the two sounds a cautionary note. It may be due to differences in the small-sample properties of the tests – the Durbin alternative test, for instance, effectively operates on regression parameters which are biased in finite samples. Moreover, for a given sample size, the extent of that bias potentially increases as the number of lags increases, and this may explain the significant test results at long lags with the Durbin alternative approach. The Breusch-Godfrey test is probably to be preferred under the circumstances.

The conclusion to emerge is that there is some weak evidence of autocorrelation in the excess returns data when we consider several years of lags. In drawing this conclusion, it is assumed that the results of the Durbin alternative tests cannot be relied on. This means that the primary evidence is the Breusch-Godfrey tests which do not reject the "no autocorrelation" hypothesis at a 5 per cent level, but do in some instances reject it at a 10 per cent level.

### *Testing the variance structure of excess returns*

The previous tests have in common that they investigate the autocorrelation structure of the excess returns data. But it is also possible to test for autocorrelation by analysing variances over different periods.

Variance ratio tests start from the insight that if there truly is no autocorrelation and the variance is constant through time, then an implication is that the variance of returns is proportional to the period over which they are calculated. For instance, the variance of 2-year returns would be twice as large as the variance of 1-year returns. And the variance of q-year returns would be q times as large as the variance of 1-year returns.[12] This observation leads to the construction and testing of so-called "variance ratios".[13]

The H-period variance ratio is defined as

$$VR(H) = \frac{\sigma_q^2 / H}{\sigma_1^2} \qquad (3.2)$$

[12] More generally, if we calculate q-period returns, then in a large population in the absence of autocorrelation, their variance should be equal to the sum of the variances of each of the constituent 1-year returns. This is simply a reflection of the fact that if Z is the sum of two random variables X and Y, then the variance of Z, $\sigma_Z^2$, is given by

$$\sigma_Z^2 = \sigma_X^2 + \sigma_Y^2 + 2\sigma_{XY}$$

where $\sigma_X^2$ and $\sigma_Y^2$ are the variances of X and Y respectively and $\sigma_{XY}$ is their covariance. If there is no autocorrelation the covariance is zero, and the expression reduces to the sum of the two variances. Kreyszig (1970 ) p.137 provides a demonstration.

[13] See Campbell, Lo and MacKinlay 1997, pp. 68-74 for a discussion of the variance ratio methodology.

In the case where there is no autocorrelation and the variance of 1-year returns is constant (i.e. "homoskedasticity") the variance of a q-period return is given by

$$\sigma_q^2 = H\sigma_1^2 \qquad (3.3)$$

which means that $VR(H) = 1$. If there is negative autocorrelation $VR(H) < 1$, and conversely if there is positive autocorrelation $VR(H) > 1$.

An attraction of the variance ratio approach is that it does not rely on a stable autocorrelation structure. For instance, if an above average return is likely to be followed by below average returns over the next few years, but with that mean reversion having a somewhat unstable timing, it might be easier to detect with variance ratios than autocorrelation coefficients.[14] For example, Cochrane (1988) notes, in the context of US GNP, that where mean reversion exists, "that reversal is likely to be slow, loosely structured and not easily captured in a simple parametric model" [p. 898]. In the context of Australian excess returns, it also seems desirable to admit the possibility that mean reversion may also be loosely structured, and to employ tests that are robust to such a situation.

The testing strategy employed here follows the approach set out in Campbell, Lo and MacKinlay (1997). It involves making an estimate of the variance ratio, $\overline{VR}(H)$, which is then used to construct a test statistic, $\psi(H)$, which is asymptotically normal and can be used to test the hypothesis of no autocorrelation. Importantly, Campbell, Lo and MacKinlay set out a testing framework which can accommodate overlapping multi-period returns and which includes adjustments for small sample bias.

This methodology was applied to the Officer series of Australian excess returns data for the period 1883 to 2004. Table 3.4 presents values of the bias-adjusted variance ratio $VR(H)$, the test statistic $\psi(H)$ (which is the test-statistic for the null hypothesis that the variance ratio is 1 under the assumption of homoskedastic returns), and the probability values associated with $\psi(H)$ (which asymptotically has a standard normal distribution) for return periods from 2 to 20 years in length.[15] The variance ratios are each less than 1, although not much different in the case of a 2-year return. Interpreted literally, these variance ratios would suggest that there is substantial mean reversion over a period of several years. For instance, 7-year returns have a variance which is only half as large as it would be if there were no autocorrelation. Autocorrelation of this magnitude certainly would be economically significant. But before drawing any strong conclusions we must question whether these variance ratios differ from 1 in a statistically significant way. The tests reject the null hypothesis of no autocorrelation at a 5 per cent significance level for 7, 8, 9 10 and 11-year returns and reject that null at a 10 per cent significance level for most other return period lengths.

---

14 It can be shown that the variance ratio is related to the autocorrelation coefficients in the following way:

$$VR(H) = 1 + 2\sum_{k=1}^{H-1}\left(1 - \frac{k}{H}\right)\rho(k)$$

This means that if we knew the parameters of the distribution of returns, the variance ratio would tell us nothing new. But usually we do not know those parameters, and must estimate them. When we are estimating to test for autocorrelation, the variance ratio will be useful if the variance of the summation over the $\rho(k)$ is less than the sum of the variances of the terms in $\rho(k)$.

15 The case of a 1-year return is not presented because, by arithmetic identity, this variance ratio is simply 1.

However, as was noted previously, the assumption of constant variance (homoskedasticity) is difficult to maintain in the case of the Officer data. Moreover, it would generally be regarded as a tenuous assumption in respect of financial market returns. Heteroskedasticity seems highly likely, in which case the test results described may simply constitute a rejection of homoskedasticity.

Campbell, Lo and MacKinlay (1997) set out a heteroskedasticity-robust test statistic which can be used to test for autocorrelation in the presence of heteroskedasticity of unknown form. This "robust" test statistic, $\psi^*(q)$, is also presented in Table 3.4 along with associated probability values. These tests suggest that the evidence for autocorrelation is weaker; only the 8-year variance ratio differs from 1 at a 10 per cent significance although for most other return period lengths a difference is statistically significant at a 20 per cent level.

**Table 3.4**
**Variance ratio tests for autocorrelation in excess returns**

| Periods | $\overline{VR}(q)$ | $\psi(q)$ | Prob > $\lvert\psi(q)\rvert$ | $\psi^*(q)$ | Prob > $\lvert\psi^*(q)\rvert$ |
|---|---|---|---|---|---|
| 2 | 0.910 | -0.991 | 0.3216 | -0.702 | 0.4828 |
| 3 | 0.764 | -1.733 | 0.0830 | -1.250 | 0.2112 |
| 4 | 0.699 | -1.763 | 0.0779 | -1.290 | 0.1971 |
| 5 | 0.640 | -1.802 | 0.0715 | -1.333 | 0.1826 |
| 6 | 0.570 | -1.905 | 0.0568 | -1.423 | 0.1546 |
| 7 | 0.493 | -2.030 | 0.0423 | -1.531 | 0.1257 |
| 8 | 0.391 | -2.257 | 0.0240 | -1.717 | 0.0860 |
| 9 | 0.381 | -2.111 | 0.0348 | -1.615 | 0.1064 |
| 10 | 0.347 | -2.117 | 0.0343 | -1.627 | 0.1037 |
| 11 | 0.330 | -2.066 | 0.0388 | -1.594 | 0.1110 |
| 12 | 0.324 | -1.974 | 0.0483 | -1.526 | 0.1269 |
| 13 | 0.315 | -1.889 | 0.0589 | -1.463 | 0.1434 |
| 14 | 0.313 | -1.778 | 0.0754 | -1.350 | 0.1770 |
| 15 | 0.309 | -1.783 | 0.0747 | -1.388 | 0.1652 |
| 16 | 0.283 | -1.725 | 0.0846 | -1.316 | 0.1883 |
| 17 | 0.254 | -1.787 | 0.0739 | -1.398 | 0.1622 |
| 18 | 0.237 | -1.689 | 0.0912 | -1.281 | 0.2003 |
| 19 | 0.243 | -1.672 | 0.0946 | -1.290 | 0.1970 |
| 20 | 0.221 | -1.717 | 0.0860 | -1.354 | 0.1758 |

## 3.4 Testing for Structural Breaks With ARMA Models

Autocorrelation in the excess returns series means that potentially useful information is not used. Therefore autoregressive moving average (ARMA) specifications were considered in an attempt to better model the data.

A basic model of excess returns specifies the excess return in period t, $\pi_t$, as the sum of its conditional mean in period t, $\hat{\pi}_t$, plus an error term $u_t$ with mean zero and uncorrelated over time

$$\pi_t = \hat{\pi}_t + u_t \qquad (3.4)$$

The conditional mean in $\hat{\pi}_t$ could be a constant, or it could be related to a set of exogenous explanatory variables. For instance the conditional mean could be given by the model

$$\hat{\pi}_t = a + bD_t \tag{3.5}$$

where $D_t = 0, t < k$ and $D_t = 1, t \geq k$. In this model $\hat{\pi}_t = a$ prior to time period $k$ and $\hat{\pi}_t = a + b$ form time period $k$ onward, i.e. there is a structural break at period $k$.

If the coefficient $b$ is significantly different from zero, then a structural change has occurred. If the error terms $u_t$ are autocorrelated, then while ordinary least squares regression will continue to produce unbiased estimates, it will no longer produce the most efficient estimates (i.e. the minimum variance estimates). This means that the power of tests to detect a structural break encapsulated in the coefficient $b$ on the dummy variable will not be maximised. This problem can be addressed by including a specification which allows for autocorrelation in the data.

Equation 3.4 can be rearranged to

$$\pi_t - \hat{\pi}_t = u_t \tag{3.6}$$

The autoregressive moving average (ARMA) model allows for autocorrelation by incorporating lagged values of the dependent variable and past prediction errors as follows

$$(\pi_t - \hat{\pi}_t) = \alpha_1(\pi_{t-1} - \hat{\pi}_{t-1}) + \ldots + \alpha_p(\pi_{t-p} - \hat{\pi}_{t-p}) + \varepsilon_t + \beta_1\varepsilon_{t-1} + \ldots + \beta_q\varepsilon_{t-q} \tag{3.7}$$

This then leads to the question of what lag structure to fit. If the excess returns are mean reverting over a period of several years, then a moving average process would provide a literal representation of this. However, autoregressive components also have corresponding moving average processes (at least so long as the data are stationary), and it is possible that these provide a more parsimonious representation of the moving average process. Therefore autoregressive elements should not be ignored in the specification search.

Three "general to specific" specification searches were carried out. These searches are summarised in Box 3.1 and Table 3.5. The 2 viable models were tested for structural breaks with dummy variables over the period 1960 to 1985. The break parameters were not significant for the MA(8) specification and the MA(2/8 constrained) model produced spurious results.

In fact there are potential biases in parameter estimates when ARMA models are estimated in small samples. "Small" is of course a vague concept, but there is certainly a potential for problems with an 8-lag structure in 122 observations.

To explore the extent of these biases Monte Carlo simulations were carried out. The simulations involved taking a sample of 122 observations, drawn randomly with replacement from the Officer data, and then estimating the parameters of an MA(8) ARMA model with the 8 MA coefficients constrained to be equal. This procedure was repeated 5,000 times to generate a distribution of the ARMA model parameters in the case where there truly is no autocorrelation.

**Box 3.1**
**Specification searches for an ARMA model**

**Nesting 1:** The first specification considered was a fairly parsimonious one – an ARMA(2,2). This is presented in Table 3.5 under "Nesting 1". The AR(1/2) MA(1/2) version was highly significant overall, but not all coefficients were at the 5-per cent level. The least significant coefficient was the AR(1) term, and this was eliminated. The new specification was still highly significant overall and the Schwarz Information Criterion (SIC) improved, meaning that the elimination of AR(1) was justified on grounds of parsimony. The remaining variables all were significant. There are risks of "common factors" in models with AR and MA terms, and a model was considered with the AR(2) term excluded. The SIC deteriorated marginally and the model became insignificant even at a 10 per cent level.

**Nesting 2:** A 10 lag moving average structure – MA(1/10) – was considered. An MA (1/9) lag structure was superior according to the Schwarz Information Criterion and an MA(1/8) was superior to this. However, an MA(1/7) structure was inferior to the MA(1/8) structure. With all 8 lags in place the specification was highly significant but not all the moving average terms had significant coefficients. The coefficients for lags 1 to 7 were generally negative but the 8 year lag had a significantly positive coefficient. Piecewise deletion of insignificant terms led to the deletion of all of lags 1 to 7. Only the 8 year lag remained and was highly significant. This is a very odd lag structure. It seems likely that it reflects features specific to this quite small dataset rather than a resilient aspect of the data generating process.

**Nesting 3:** In Nesting 3 constrained lag structures are considered against the unconstrained MA(1/8) alternative. The first variant involves constraining the 1$^{st}$ to 8$^{th}$ MA terms to have the same coefficient values. the coefficients for the 2$^{nd}$ through 7$^{th}$ year MA lags were constrained to be equal. An MA(1/8) model was then estimated. The MA(1) terms was not significant and was eliminated. The common coefficient for the MA(2/7) terms was negative and highly significant and the MA(8) term had a positive and highly significant coefficient. Finally a very general specification, an MA(8) model with all MA coefficients constrained to equality was estimated. This specification nests both the MA(2/7) and MA(8) models, both of which are superior according to the SIC.

The significance of an ARMA model can be tested with a Wald statistic, which theoretically follows a Chi-2 distribution. In the simulations it was found that for 15 per cent of the simulations the Chi-2 value for 5 per cent significance was exceeded. Clearly there is a major tendency to over-accept the significance of ARMA models involving 122 draws form the Officer dataset. This means that the Chi-2 statistic cannot be used in the usual way to infer the significance of the results.

However, it is possible instead to use the simulations to form a non-parametric test. The Wald statistic for the constrained MA(8) model on the actual Officer data was 6.2 per cent down from the top. This suggests that the MA(8) structure in the true series of historic excess returns is significant at a 6.2 per cent level, just outside the conventional 5 per cent.

These simulations tell us nothing about the power of the non-parametric test. Given that the MA(8) specification is almost significant at the 5 per cent level, and given that substantial power problems were evident with simple comparisons of means, it is still worthwhile looking for structural breaks by including year-of-break dummies in an MA(8) specification. This was done and the results are reported in Table 3.6.

**Table 3.5**
**ARMA Modelling Results**

| | Constant | AR(1) | AR(2) | MA(1) | MA(2) | MA(3) | MA(4) | MA(5) | MA(6) | MA(7) | MA(8) | sigma | Wald | SIC |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Nesting 1:** | | | | | | | | | | | | | | |
| AR(1/2) MA(1/2) | | | | | | | | | | | | | | |
| coeff | 7.314 | 0.237 | 0.231 | -0.418 | -0.434 | | | | | | | 16.03 | 124.2 | 1,047.7 |
| std err | 0.452 | 0.255 | 0.211 | 0.197 | 0.197 | | | | | | | 1.088 | | |
| prob | 0.000 | 0.353 | 0.273 | 0.034 | 0.027 | | | | | | | 0.000 | 0.000 | |
| AR(2) MA(1/2) | | | | | | | | | | | | | | |
| coeff | 7.307 | | 0.389 | -0.216 | -0.610 | | | | | | | 16.06 | 84.3 | 1,043.4 |
| std err | 0.461 | | 0.191 | 0.072 | 0.132 | | | | | | | 1.109 | | |
| prob | 0.000 | | 0.042 | 0.003 | 0.000 | | | | | | | 0.000 | 0.000 | |
| MA(1/2) | | | | | | | | | | | | | | |
| coeff | 7.212 | | | -0.181 | -0.275 | | | | | | | 16.36 | 3.4 | 1,042.7 |
| std err | 0.823 | | | 0.138 | 0.158 | | | | | | | 1.144 | | |
| prob | 0.000 | | | 0.190 | 0.082 | | | | | | | 0.000 | 0.183 | |

**Table 3.5 (continued)**
**ARMA Modelling Results**

| | Constant | AR(1) | AR(2) | MA(1) | MA(2) | MA(3) | MA(4) | MA(5) | MA(6) | MA(7) | MA(8) | sigma | Wald | SIC |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Nesting 2:** | | | | | | | | | | | | | | |
| MA(1/8) | | | | | | | | | | | | | | |
| coeff | 7.261 | | | -0.116 | -0.268 | -0.103 | -0.056 | -0.139 | -0.073 | -0.150 | 0.236 | 15.26 | 32.1 | 1,055.6 |
| std err | 0.496 | | | 0.109 | 0.102 | 0.138 | 0.151 | 0.198 | 0.164 | 0.148 | 0.128 | 0.982 | | |
| prob | 0.000 | | | 0.285 | 0.008 | 0.456 | 0.709 | 0.483 | 0.655 | 0.310 | 0.065 | 0.000 | 0.000 | |
| MA(1/3,5/8) | | | | | | | | | | | | | | |
| coeff | 7.251 | | | -0.120 | -0.282 | -0.110 | | -0.133 | -0.074 | -0.180 | 0.241 | 15.28 | 30.6 | 1,051.1 |
| std err | 0.509 | | | 0.106 | 0.097 | 0.127 | | 0.190 | 0.164 | 0.113 | 0.126 | 0.980 | | |
| prob | 0.000 | | | 0.257 | 0.004 | 0.384 | | 0.485 | 0.652 | 0.112 | 0.056 | 0.000 | 0.000 | |
| MA(1/3,5,7/8) | | | | | | | | | | | | | | |
| coeff | 7.235 | | | -0.129 | -0.286 | -0.098 | | -0.168 | | -0.192 | 0.231 | 15.30 | 27.0 | 1,046.8 |
| std err | 0.531 | | | 0.107 | 0.098 | 0.142 | | 0.136 | | 0.106 | 0.130 | 0.972 | | |
| prob | 0.000 | | | 0.227 | 0.003 | 0.490 | | 0.216 | | 0.071 | 0.077 | 0.000 | 0.000 | |
| MA(1/2,5,7/8) | | | | | | | | | | | | | | |
| coeff | 7.240 | | | -0.153 | -0.300 | | | -0.195 | | -0.170 | 0.186 | 15.37 | 26.6 | 1,042.8 |
| std err | 0.552 | | | 0.101 | 0.102 | | | 0.131 | | 0.107 | 0.088 | 0.961 | | |
| prob | 0.000 | | | 0.133 | 0.003 | | | 0.138 | | 0.111 | 0.033 | 0.000 | 0.000 | |
| MA(1/2,7/8) | | | | | | | | | | | | | | |
| coeff | 7.189 | | | -0.139 | -0.295 | | | | | -0.187 | 0.176 | 15.66 | 19.0 | 1,042.2 |
| std err | 0.804 | | | 0.142 | 0.162 | | | | | 0.110 | 0.097 | 1.055 | | |
| prob | 0.000 | | | 0.328 | 0.069 | | | | | 0.088 | 0.071 | 0.000 | 0.001 | |
| MA(2,7/8) | | | | | | | | | | | | | | |
| coeff | 7.218 | | | | -0.226 | | | | | -0.144 | 0.184 | 15.80 | 26.3 | 1,039.4 |
| std err | 1.174 | | | | 0.128 | | | | | 0.122 | 0.124 | 1.021 | | |
| prob | 0.000 | | | | 0.076 | | | | | 0.237 | 0.139 | 0.000 | 0.000 | |
| MA(2,8) | | | | | | | | | | | | | | |
| coeff | 7.289 | | | | -0.190 | | | | | | 0.241 | 15.95 | 19.0 | 1,037.0 |
| std err | 1.505 | | | | 0.130 | | | | | | 0.095 | 1.051 | | |
| prob | 0.000 | | | | 0.144 | | | | | | 0.011 | 0.000 | 0.000 | |
| MA(8) | | | | | | | | | | | | | | |
| coeff | 7.338 | | | | | | | | | | 0.285 | 16.21 | 6.9 | 1,036.2 |
| std err | 1.854 | | | | | | | | | | 0.108 | 1.080 | | |
| prob | 0.000 | | | | | | | | | | 0.009 | 0.000 | 0.009 | |

**Table 3.5 (continued)**
**ARMA Modelling Results**

| | Constant | AR(1) | AR(2) | MA(1) | MA(2) | MA(3) | MA(4) | MA(5) | MA(6) | MA(7) | MA(8) | sigma | Wald | SIC |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Nesting 3:** | | | | | | | | | | | | | | |
| MA(1/8) | | | | | | | | | | | | | | |
| coeff | 7.261 | | | -0.116 | -0.268 | -0.103 | -0.056 | -0.139 | -0.073 | -0.150 | 0.236 | 15.26 | 32.1 | 1,055.6 |
| std err | 0.496 | | | 0.109 | 0.102 | 0.138 | 0.151 | 0.198 | 0.164 | 0.148 | 0.128 | 0.982 | | |
| prob | 0.000 | | | 0.285 | 0.008 | 0.456 | 0.709 | 0.483 | 0.655 | 0.310 | 0.065 | 0.000 | 0.000 | |
| MA(1/8), MA(2=3=4=5=6=7) | | | | | | | | | | | | | | |
| coeff | 7.284 | | | -0.096 | -0.138 | -0.138 | -0.138 | -0.138 | -0.138 | -0.138 | 0.264 | 15.44 | 23.9 | 1,034.4 |
| std err | 0.527 | | | 0.140 | 0.041 | 0.041 | 0.041 | 0.041 | 0.041 | 0.041 | 0.110 | 0.997 | | |
| prob | 0.000 | | | 0.494 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.016 | 0.000 | 0.000 | |
| MA(2/8), MA(2=3=4=5=6=7) | | | | | | | | | | | | | | |
| coeff | 7.294 | | | | -0.148 | -0.148 | -0.148 | -0.148 | -0.148 | -0.148 | 0.259 | 15.49 | 19.3 | 1,030.6 |
| std err | 0.574 | | | | 0.035 | 0.035 | 0.035 | 0.035 | 0.035 | 0.035 | 0.092 | 1.007 | | |
| prob | 0.000 | | | | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.005 | 0.000 | 0.000 | |
| MA(2/7), MA(2=3=4=5=6=7) | | | | | | | | | | | | | | |
| coeff | 7.407 | | | | -0.128 | -0.128 | -0.128 | -0.128 | -0.128 | -0.128 | | 16.12 | 15.1 | 1,035.0 |
| std err | 0.413 | | | | 0.033 | 0.033 | 0.033 | 0.033 | 0.033 | 0.033 | | 1.030 | | |
| prob | 0.000 | | | | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | | 0.000 | 0.000 | |
| MA(8) | | | | | | | | | | | | | | |
| coeff | 7.338 | | | | | | | | | | 0.285 | 16.21 | 6.9 | 1,036.2 |
| std err | 1.854 | | | | | | | | | | 0.108 | 1.080 | | |
| prob | 0.000 | | | | | | | | | | 0.009 | 0.000 | 0.009 | |
| MA(1/8), MA(2=3=4=5=6=7=8) | | | | | | | | | | | | | | |
| coeff | 7.363 | | | -0.087 | -0.087 | -0.087 | -0.087 | -0.087 | -0.087 | -0.087 | -0.087 | 16.51 | 14.41 | 1,040.5 |
| std err | 0.531 | | | 0.023 | 0.023 | 0.023 | 0.023 | 0.023 | 0.023 | 0.023 | 0.023 | 1.109 | | |
| prob | 0.000 | | | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | |

The first column of Table 3.6 shows the estimated break coefficient for dummies commencing in each year from 1960 to 1985. The literal interpretation of the -1.95 coefficient for 1960 is that the average excess return from 1960 onward was 1.95 per cent smaller than in the years before 1960. The question is, how meaningful is this result?

First, the model does not appear to produce significant biases in the dummy coefficients. The average from the simulations was about +0.1 and never more than 0.2. There is no reason therefore to think that the model produces biased estimates out of the data.

Second, the z-statistics (in the second column) are high, being at least 3.6 in every instance (ignoring the minus signs). If interpreted as a normal variate (which they are in the third column), the z-statistic would lead to strong rejections of the "no structural break" hypothesis. However, the z-statistics produced in the simulations do not follow a normal distribution, and it seems unreasonable therefore to assume that the z-statistic from the true series would do so either.

Third, it is possible to observe where each dummy coefficient sits in the distribution of coefficients generated in the simulations. Their percentile is shown in the fourth column. They sit toward the bottom of the range, but not at the extreme ends of it. For instance, the dummy coefficient for a structural break in 1960, -1.95 per cent, sits just above the 25$^{th}$ percentile. All of the dummies for the true data sit above the 17$^{th}$ percentile. Fourth, however, this comparison makes no allowance for the explanatory power of the MA terms. If they have good explanatory power, they reduce the (robust) standard errors of the dummy coefficients, thus boosting the z-statistics. The z-statistics do make such an allowance. The 5$^{th}$ column shows the z-statistics percentile in the ranking of the simulated z-statistics. And they do sit very low in this ranking. All are below the 7$^{th}$ percentile and many are below the 5$^{th}$ percentile.

The key conclusion to emerge from this table is that allowing for the autocorrelation in the data – probably due to mean reversion – provides stronger evidence for the view that there has been a structural break in the mean sometime over the last 40 years or more. The evidence of a structural break is generally stronger at later dates.

However, it is difficult to accept this evidence with a high degree of confidence. The MA(8) specification actually implies larger structural breaks than a specification with just dummies alone. It is to be expected that the MA(8) component would reduce the variance of residuals and thus allow interpretation of dummy coefficients with a greater degree of confidence. The fact that it accentuates the apparent size of structural breaks is an interesting finding, but in the absence of a good explanation, it leaves yet another question mark over the ARMA tests.

**Table 3.6**
**Tests for a structural beak in an MA(8) model**

| | Difference | Robust z-stat | Conventional prob | Difference percentile | Z-rank percentile |
|---|---|---|---|---|---|
| 1960 | -1.94977 | -3.6691 | 0.000182 | 0.2554 | 0.0606 |
| 1961 | -1.95361 | -3.67634 | 0.000177 | 0.262 | 0.0632 |
| 1962 | -2.01929 | -3.79993 | 0.000114 | 0.2488 | 0.0542 |
| 1963 | -2.04761 | -3.85322 | 9.4E-05 | 0.2438 | 0.0546 |
| 1964 | -2.11849 | -3.9866 | 5.75E-05 | 0.2338 | 0.0516 |
| 1965 | -2.15424 | -4.05387 | 4.47E-05 | 0.2402 | 0.0484 |
| 1966 | -2.14996 | -4.04582 | 4.61E-05 | 0.239 | 0.053 |
| 1967 | -2.16808 | -4.07992 | 4.06E-05 | 0.2398 | 0.0526 |
| 1968 | -2.29339 | -4.31573 | 1.64E-05 | 0.2322 | 0.047 |
| 1969 | -2.41737 | -4.54905 | 6.45E-06 | 0.2182 | 0.0368 |
| 1970 | -2.49119 | -4.68796 | 3.65E-06 | 0.2134 | 0.0326 |
| 1971 | -2.43874 | -4.58927 | 5.48E-06 | 0.2244 | 0.0414 |
| 1972 | -2.43598 | -4.58407 | 5.59E-06 | 0.233 | 0.0454 |
| 1973 | -2.5466 | -4.79223 | 2.37E-06 | 0.2182 | 0.035 |
| 1974 | -2.48663 | -4.67938 | 3.78E-06 | 0.221 | 0.0422 |
| 1975 | -2.40008 | -4.51651 | 7.36E-06 | 0.2314 | 0.0552 |
| 1976 | -2.61533 | -4.92157 | 1.37E-06 | 0.2122 | 0.0408 |
| 1977 | -2.62169 | -4.93353 | 1.3E-06 | 0.21 | 0.0434 |
| 1978 | -2.67076 | -5.02588 | 8.77E-07 | 0.2168 | 0.0456 |
| 1979 | -2.76898 | -5.21072 | 3.92E-07 | 0.217 | 0.0424 |
| 1980 | -3.00288 | -5.65087 | 5.4E-08 | 0.1996 | 0.0352 |
| 1981 | -3.2796 | -6.1716 | 4.65E-09 | 0.1846 | 0.025 |
| 1982 | -3.18653 | -5.99646 | 1.07E-08 | 0.1942 | 0.031 |
| 1983 | -3.03984 | -5.72043 | 3.92E-08 | 0.211 | 0.04 |
| 1984 | -3.51576 | -6.61602 | 5.3E-10 | 0.1798 | 0.0268 |
| 1985 | -3.46007 | -6.51122 | 8.9E-10 | 0.188 | 0.0336 |

# 4. Predicting the Contemporary Market Risk Premium from Historical Data

## 4.1 Conceptual issues

This section considers alternative methods of predicting the contemporary MRP from historical data and compares the forecast accuracy of these methods. The methods compared are all univariate, in the sense that they use only past excess returns data to generate MRP predictions. They do not include models in which the MRP is determined by other exogenous variables.

With the exception of the simple average, each of these models places a premium on more recent excess return observations. The rationale for such an approach is as follows. Suppose that the MRP does change over time. Suppose also that we could observe it (ie. that we could strip out the very large "noise" element which causes an individual year's excess return to differ from the MRP in each year). In that case we could fit a suitable time series model of the MRP. Most of the useful information would be in very recent observations. In the case where the market risk premium follows a random walk (and a constant MRP is a degenerate variant of this), the only relevant information would be the most recent value of the market risk premium. But it is possible to imagine generating processes for the market risk premium in which other recent observations would also have useful information content. For instance, if fundamental explanations for the value of the market risk premium lie in some trending macroeconomic variables, then the market risk premium could also be expected to show trending behaviour of its own.

As it is, we do not know the true historic values of the market risk premium, and must instead estimate them from excess returns. If we had the option of sampling repeatedly at each point in time, we could improve the precision of estimates for each year's market risk premium. Assuming that there were no limit on the size of the sample that could be drawn, then the true market risk premium could be estimated with as high a degree of precision as was wanted.

However, the reality is that for each year there is only one excess return observation available. To deal with this problem, the standard approach is to make an assumption about the deterministic structure of the market risk premium, with the time series of excess returns then treated as repeated draws over this structure and used to estimate parameters for this deterministic structure.

A simple example of this approach is the case where the conditional value of the market risk premium is assumed to be constant, in which case a simple average of the longest possible series of data is used to estimate the market risk premium. The validity of this model can also be tested using the data available. Sections 2 and 3 of this report have been primarily concerned with doing just that: testing whether the assumption of a constant market risk premium over more than 100 years is sustainable. The general conclusion is that there is evidence of a downward shift in the market risk premium, but that it is statistically weak.

But of course a "constant market risk premium" model is not the only possible prior assumption. The idea that the market risk premium could change over time is, intuitively, strongly appealing. Why would it stay constant forever? Peoples endowments, preferences, and indeed the riskiness of the market, all change over time, and each of these could be expected to affect the market risk premium.

For this reason there is some attraction to weighting schemes which place more weight on recent excess return observations. To compare the consequences of alternative weighting schemes, the following three alternatives were considered:

- equal-weighted moving averages;
- exponentially weighted moving averages; and
- the Hodrick-Prescott filter.

### *Moving average*

The equal weighted moving average really just truncates the sample. Moving average periods of 25 and 50 years were considered. The effect is to discard any information outside the length of the moving average.

### *Exponentially weighted moving average*

Exponentially weighted moving averages geometrically decrease the weights applied to observations in the more distant past. A parameter α determines just how quickly the weighting falls away. Let $c_i, i = 0,1,2,...$ be the weights to apply to values of the equity premium. The weighted average equity premium for period t is then

$$\hat{\pi}_t = c_0\pi_t + c_1\pi_{t-1} + c_2\pi_{t-2}... \qquad (4.1)$$

If the weights are geometric, we require that $c_i = c_0(1-\alpha)^i$. $\alpha$ is thus a measure of how much past values are discounted. A low value of $\alpha$ means a gradual reduction in the weights attached to past observations and a high value of $\alpha$ means a rapid reduction in the weight given to past observations. It makes sense to require that the weights sum to 1, and when this is the case it can be shown that $c_i = \alpha(1-\alpha)^i$. Substituting this back into 4.1 gives the exponentially weighted average equity premium

$$\hat{\pi}_t = \alpha\pi_t + \alpha(1-\alpha)\pi_{t-1} + \alpha(1-\alpha)^2\pi_{t-2}... \qquad (4.2)$$

This can be reduced to a recurrence form which is easy to use in calculations

$$\begin{aligned}\hat{\pi}_t &= \alpha\pi_t + \alpha(1-\alpha)\pi_{t-1} + \alpha(1-\alpha)^2\pi_{t-2}...\\ &= \alpha\pi_t + (1-\alpha)(\alpha\pi_{t-1} + \alpha(1-\alpha)\pi_{t-2} + \alpha(1-\alpha)^2\pi_{t-3}...) \qquad (4.3)\\ &= \alpha\pi_t + \hat{\pi}_{t-1}\end{aligned}$$

This recurrence form expresses the exponentially weighted average as a weighted sum of the current period value of the equity premium and the preceding period exponentially weighted average. It is suited to a certain sort of non-stationary process. Since it is possible that the equity premium is non-stationary over the sample period that we have available, the exponentially weighted moving average is an interesting predictor to consider.[16]

[16] See Chatfield (1996) pp 68-70 for further discussion.

***Hodrick-Prescott filter***

The third moving average considered is the Hodrick-Prescott filter. The underlying premise of this filter is that for some variables the rate of change of the variable is likely to have momentum and therefore should not change too quickly. In visual terms this translates to the requirement that the data series move around smoothly over time.[17] There is thus a trade-off between contemporaneity and smoothness. Smoothness is achieved by giving more weight to more distant observations. To apply the HP filter a "smoothness" parameter $\lambda$ is selected. The details of the filter and algorithms to compute it are not given here as they are complex and messy. Calculations were carried out with an HP filter written for STATA by Sorensen (2005).

In the case of the HP filter the matter of revisions must also be taken into consideration. Trend estimates are revised as new data becomes available, and the revisions are generally largest for the most recent trend estimates. It may therefore be preferable to use a filter value that is a few years old as a predictor.

## 4.2 Prediction performance of models

Predictions were made using the simple average, moving average, exponentially weighted moving average, and Hodrick-Prescott filters.

Construction of a simple average over a period $t=1, \ldots, T$ is a straightforward matter. However, as noted above, there is no unique moving average, exponential smoother or Hodrick-Prescott filter. Parameters must be selected to implement those methods. For a moving average, the averaging window needs to be selected. In the case of exponential smoothing, the parameter $\alpha$, which corresponds to the degree of weight given to more recent observations in the averaging process, must be selected.

Common sense would suggest that the value of $\alpha$ in an exponential smoothing should not be less than the inverse of the number of observations; otherwise we weight recent observations less than observations in the more distant past. But theory does not in general provide a strong guide as to the appropriate values of these parameters. Consequently we have considered a range of values. This means that we have several forecasting models to consider: one "simple average", several "equal-weighted moving average" models, several "exponentially weighted moving average" models based on different values of $\alpha$, and several HP filters based on different lag lengths.[18]

Figure 4.1 plots excess returns and trend excess returns calculated using the Hodrick-Prescott filter. The volatility of the excess returns is such that, on this scale, it is difficult to discern much movement at all in the trend series.

Figure 4.2 plots the trend series along with the arithmetic average of excess returns in the Officer series. Certainly the trend estimates are strongly suggestive of a downward move after the late 1950s.

[17] Pedersen (2001) describes the filter as one "which removes a smooth trend as one would draw it with a free hand drawing".

[18] The results reported here all employ a $\lambda$ of 1,600. Calculations were also carried out with $\lambda$ = 6,400, but the forecast performance was inferior.

**Figure 4.1**
**Excess returns and Hodrick-Prescott trend estimate of excess returns**

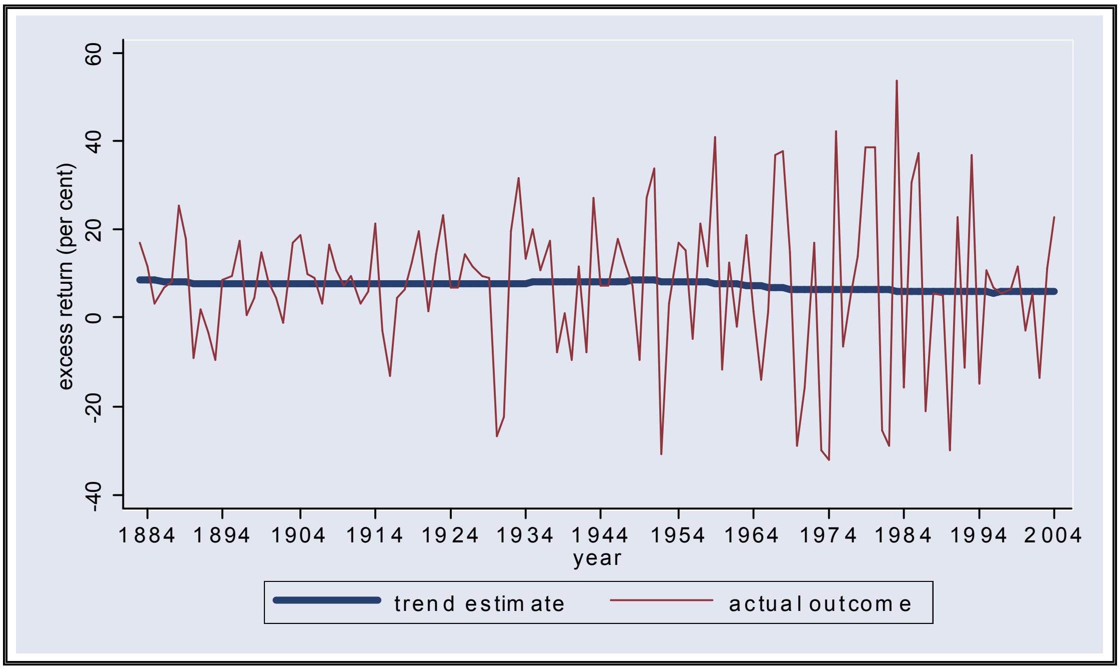
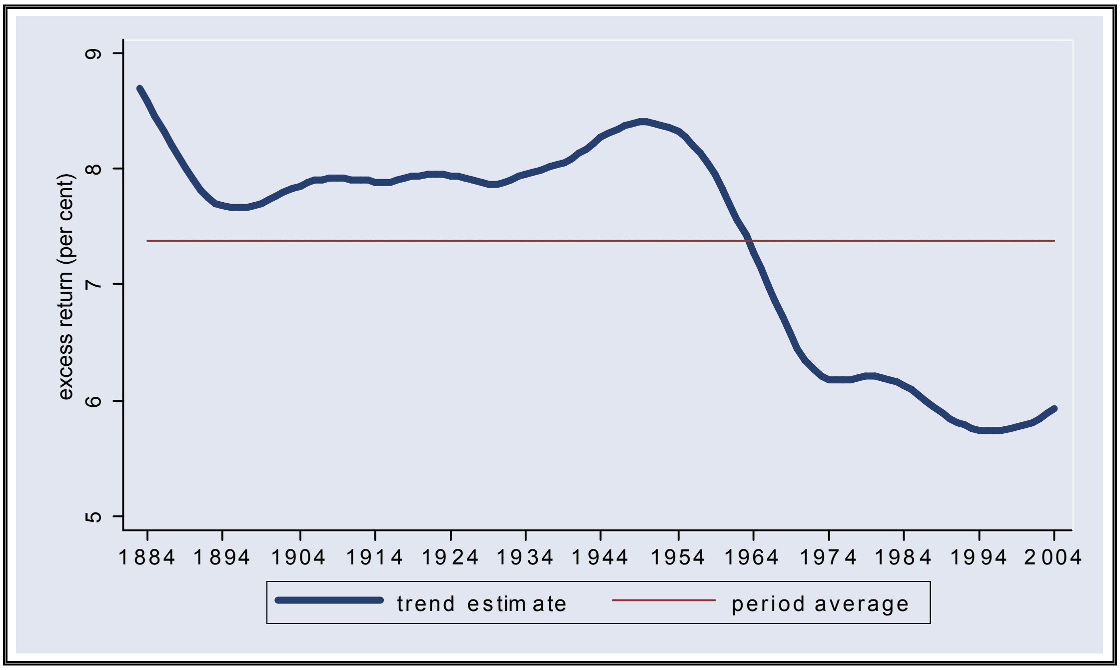



**Figure 4.2**
**Average excess returns and Hodrick-Prescott trend estimate of excess returns**




One commonly used method to compare the performance of alternative forecasting models is to compare the mean square errors of forecasts. Where the intention is to infer which model might perform best in the future, it is generally accepted that the evaluations should be performed on out of sample forecast performance. This means that when we estimate a model to forecast the parameter of interest in time period t, we estimate the parameters of that model using only information that is available up to t-1.

We considered a series of rolling forecasts, and evaluate the performance of each model. The Officer dataset was used for this purpose. The first prediction is for 1946, making use of excess returns data to the end of 1945. These predictions are then compared with the observed excess return for 1946 and the squared forecast error is calculated. The process is repeated, producing predictions for 1947 using data available up to the end of 1946 and again the squared forecast errors are calculated. And the process is repeated year by year until data to the end of 2003 is used to produce a forecast for 2004 and to calculate a squared forecast error.[19] These forecast errors are then averaged over the period 1970 to 2004 to generate mean squared errors (MSEs) for each prediction method. These mean square errors are reported in Table 4.1, along with the latest estimates of the market risk premium.

**Table 4.1**
**Mean square forecast errors**

| Forecasting model | Mean square error | Latest forecast |
|---|---|---|
| Simple average 1883 onward | 575 | 7.4 |
| Moving averages: | | |
| 25 year | 594 | 5.9 |
| 50 year | 582 | 6.6 |
| Exponential smoothing | | |
| $\alpha = 0.1$ | 627 | 6.2 |
| $\alpha = 0.05$ | 598 | 6.3 |
| $\alpha = 0.04$ | 592 | 6.4 |
| $\alpha = 0.03$ | 586 | 6.6 |
| $\alpha = 0.02$ | 581 | 6.9 |
| Hodrick Prescott filter | | |
| value last year | 663 | 5.9 |
| value 5 years ago | 633 | 5.8 |
| value 10 years ago | 604 | 5.7 |
| value 15 years ago | 585 | 5.8 |
| value 20 years ago | 573 | 6.1 |
| value 25 years ago | 566 | 6.2 |

A striking feature of the table is that the simple average is one of the better performing out-of-sample forecasting methods. Over the sample considered, it is matched only by the trend estimates produced by the Hodrick-Prescott filter at quite long lags. Moving averages perform a little worse and exponential smoothing performs worst of all.

The lowest prediction of the contemporary market risk premium is 5.7 per cent, based on a 10-year lagged value from the Hodrick Prescott filter. The highest prediction is 7.4 per cent, coming from the simple average of past excess returns.

19 Data to the end of 2004 could be used to calculate a 2005 forecast, but we are unable to calculate the forecast error for 2005 until the actual outcome is known.

# 5. The influence of taxes on investors, transaction costs and liquidity premiums

The accumulation indexes that are used to measure equity returns make no allowance for personal income taxation, and make no allowance for transaction costs and illiquidity.[20] But for a risk averse investor, the relevant comparison of expected returns on equities and bonds must be a comparison after taxes and transaction costs and any costs associated with illiquidity – i.e. the investor will want to compare net returns rather than gross returns.

If the expected rates of taxes and transaction costs on equities and bonds are the same, then there will be no tax-induced difference in gross and net expected returns. But if they differ, the equity premium gross of taxes and transaction costs will differ from the equity premium measured on a net basis. Furthermore, if there are changes over time in the difference between, on the one hand, taxes and transaction costs on equities, and, on the other hand, taxes and transaction costs on bonds, then those changes will cause the gross equity premium to vary even if the net equity premium stays constant.

If one regards the equity premium as a quantity which derives from individuals' risk aversion – albeit that individuals' often delegate asset management to intermediaries such as superannuation funds – then it will arguably be more fruitful to regard the gross equity premium as an amalgam of a net equity premium, differences in personal income taxes across bonds and equities, and differences in transaction costs across bonds and equities. In this approach, an estimate of the gross equity premium would be arrived at by estimating the net equity premium, and the net differences in personal taxes and taxation costs across equities and bonds, and then summing them. To do this, it is necessary to make estimates of the differences in personal taxes and transaction costs across the two asset classes.

## 5.1 Excess taxes

The excess of taxes on equities over taxes on bonds can be expected to flow into the equity premium. Experimental estimates of this excess were made for the period 1974 to 2004 using fairly simple scenarios for both the holdings of superannuation funds and holdings of direct personal equity investment. Basically this involved:

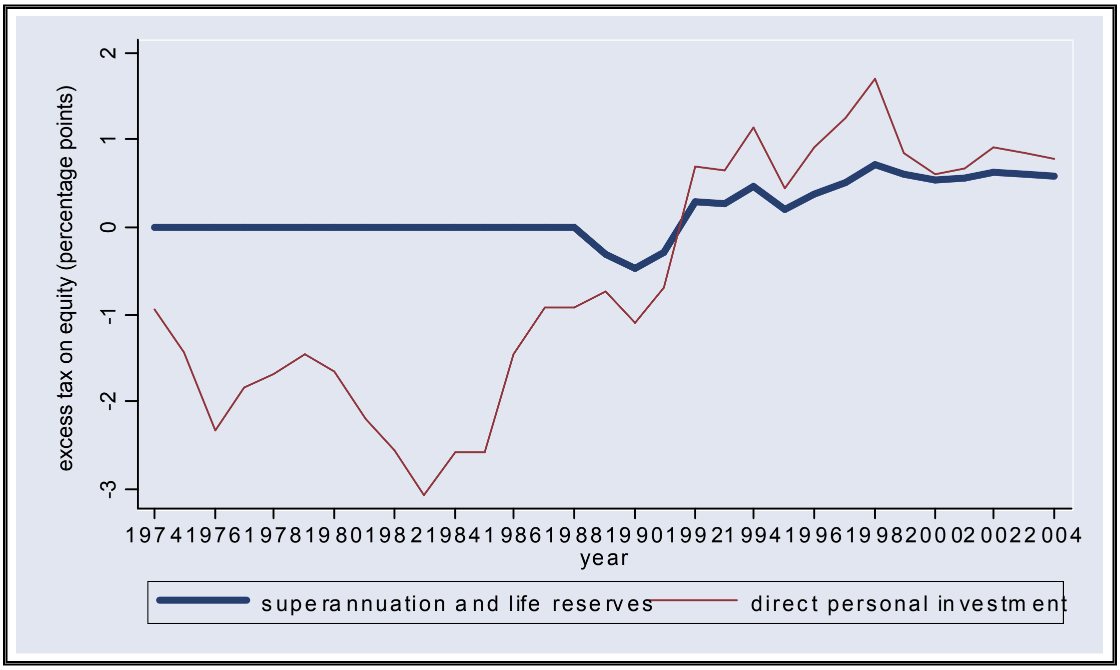
- assuming an after-tax equity premium of 6 per cent for investments through each of these two vehicles;
- calculating the after-tax return on a bond using the actual bond rate (which is pre-tax);
- setting the after-tax ex ante equity return equal to the after-tax bond return plus 6 per cent;
- calculating the after-tax dividend yield by deducting an estimate of tax from reported dividend yield figures;
- calculating the after-tax franking credit yield by deducting an estimate of tax from an estimated 0.6 per cent gross franking credit value;
- deducting the after-tax dividend yield and franking credit yield from the after-tax equity return, which yields an after-tax capital gain;

[20] The indexes are constructed on an after company tax basis.

- applying a tax formula, including an inflation component, to derive a pre-tax capital gain;
- adding to the pre-tax capital gain the pre-tax dividend yield and the pre-tax franking credit yield to get a pre-tax equity rate of return; and
- deducting the pre-tax bond yield from this pre-tax equity rate of return to get a pre-tax equity premium.

This exercise was conducted separately both for investment through superannuation funds and for direct personal investment. The results are shown in Figure 5.1, and as can be seen there are some significant differences in the results for the two different investment channels. These differences relate mainly to differences in tax rates and different timing in the variation of tax rates over time. It is notable that early in the period the excess tended to be negative, meaning that the tax treatment of equities was in absolute terms more favourable than the treatment of bonds – which was due mainly to the fact that the inflation-compensation component of bond returns was fully taxable whereas the inflation-compensation component of equity returns was not if it was delivered as a capital gain. Super funds were not taxed on their investment returns at all until the late 1980s.[21] The differentials have also been affected by the introduction of capital gains tax, dividend imputation, and changes to the capital gains tax.

**Figure 5.1**
**Excess of investor's tax on equities over tax on bonds – alternative investment channels**

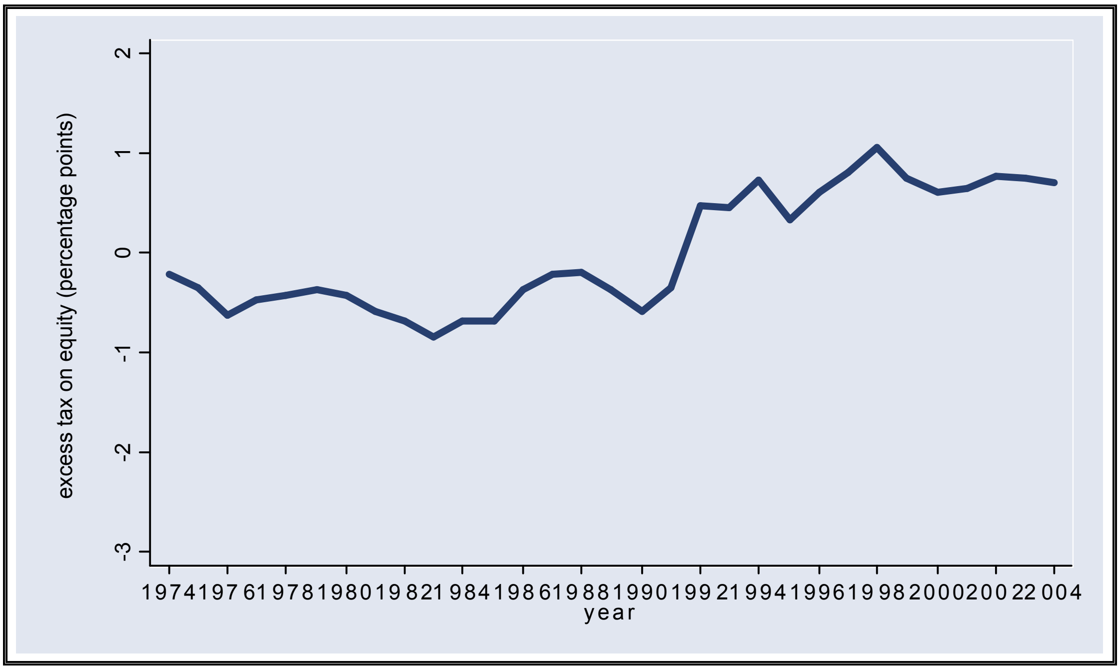



The experimental estimates of the excess tax on equities were combined to form a weighted average. The weights were calculated by adding together average direct equity holdings of super funds and life offices and of households for the period 1988 to 2004, and then calculating the proportions in, on the one hand, superannuation funds and life offices and, on

[21] There have at times been regulations on the portfolio allocations of super funds and life funds, which could be viewed as quasi-taxes – e.g. the so-called 30/20 ratio, introduced in the early 1960s and operating , which provided strong incentives for life and superannuation offices to allocate at least 20 per cent of their portfolios to Commonwealth government securities (see Grant 1974 pp. 183-185 and Covick and Lewis (1993) p. 166).

the other, households. It must be acknowledged that this procedure is very rough, and it would be better to have a more sophisticated weighting pattern that allowed for changes over time and, moreover, allowed for different allocations of bonds and equities across the two investment channels, but the data that would be necessary to do this are not readily available. No allowance has been made for switching in investment channels, although one would expect investors to take advantage of changing patterns of tax advantages by adjusting their holdings.[22] The resulting weighted average is shown in Figure x.2. It indicates that over the period 1974 to 2004 the tax rate payable on equities may have shifted up by about 1 per cent relative to the tax rate payable on bonds. However, it must be emphasised that the estimate is very much of an experimental character and is really only a first step towards resolving the influence of changes in tax arrangements on gross excess returns.

**Figure 5.2**
**Excess of investor's tax on equities over tax on bonds – weighted across investment channels**




## 5.2 Transaction costs and liquidity premiums

In pure risk-based pricing models for the equity premium transaction costs are ignored. It is assumed that investors costlessly maintain an optimal portfolio structure – i.e.. one in which at all times diversifiable risk is entirely diversified away and the quantum of systematic risk is attuned exactly in line with the investor's preferences. This is achieved through instantaneous portfolio rebalancing. A model resting on these assumptions will be satisfactory so long as they are a reasonable approximation to the reality. However, recent work suggests that transaction costs are significant and that their implications need more explicit attention. A pathbreaking paper in this regard is Constantinides (1986).

22 Drake (1985) argues that "differences in the tax treatment of dividends are such that the individual Australian shareholder fares worse than the life office, the superannuation funds and the overseas investor", and offers this as one explanation for a substantial decline in household equity ownership between 1950 and 1980 (pp. 284-286).

Constantinides points out that although there is a demand for transactions arising from emerging imbalances in investors' portfolios, the transactions will be enacted only once the imbalance is sufficiently large, and the associated cost to the investor sufficiently large, to warrant the costs that are incurred when rebalancing transactions are made. When the investor carries out these transactions a cost is incurred. But until the investor makes the transaction, there is a cost to the investor associated with holding a sub-optimal portfolio; Constantinides describes this as a "liquidity premium".[23] Thus the investor will factor transaction costs and a liquidity premium into his costs and, because these costs are unavoidable for investors, they will need to be incorporated into the gross equity premium.

At least two recent papers argue that transaction costs and/or a liquidity premium go some way, and indeed a long way, to resolving the "equity premium puzzle". In a study of US historical data McGrattan and Prescott (2003) use equity mutual-fund costs as a measure of the costs of holding a diversified equity portfolio. A chart in their article shows these costs varying between about 1¼ and 2½ per cent per annum over the period 1980 to 2000.[24] An interesting feature of the data presented by McGrattan and Prescott, of limited relevance to their analysis but highly relevant to this analysis, is that equity mutual-fund costs declined substantially over the period 1980 to 2000. In recent work Swan (2005) argues that there is an "invisible cost" in the form of substantial gains from trading equities (i.e. portfolio rebalancing) that are forgone because of transaction costs, and that these invisible costs are "15 times higher than all the observed costs of trading, such as spreads and commissions, combined" [p. 3].

If these authors are correct in their assessments, it seems likely that changes in these costs over time will also have implications for the gross equity premium.

It is helpful at this point to consider the demand for transactions in an asset market using a diagrammatic analysis. This analysis will serve to illustrate the link between transaction prices, the quantity of transactions, transaction costs, and the liquidity premium. A key point is that the costs to an individual of holding an asset include both transaction costs and a "liquidity premium" for portfolio balances that are not corrected.

Transaction costs reflect both price and quantity elements. The price of a transaction is typically measured in "round-trip" terms, defined as the total cost of buying and selling a security as a proportion of the its value. Fisher (1994) models transaction costs with a bid-ask spread, and argues that three elements of the bid-ask spread need to be captured: (1) the market buy/sell quotations, (2) broker commission plus taxes, and (3) the costs of gathering information, market impact and/or management fees. Transaction quantities can be captured with a "turnover ratio", which is the ratio of the value of trades in a year to market capitalisation.[25]

Figure 5.3 presents a hypothetical demand schedule for equity transactions. The schedule shows that with a round-trip transaction cost of 6 per cent there will be a turnover rate of 11.6 per cent. If the round-trip transaction cost falls to 3 per cent, turnover rises to 26.8 per cent.

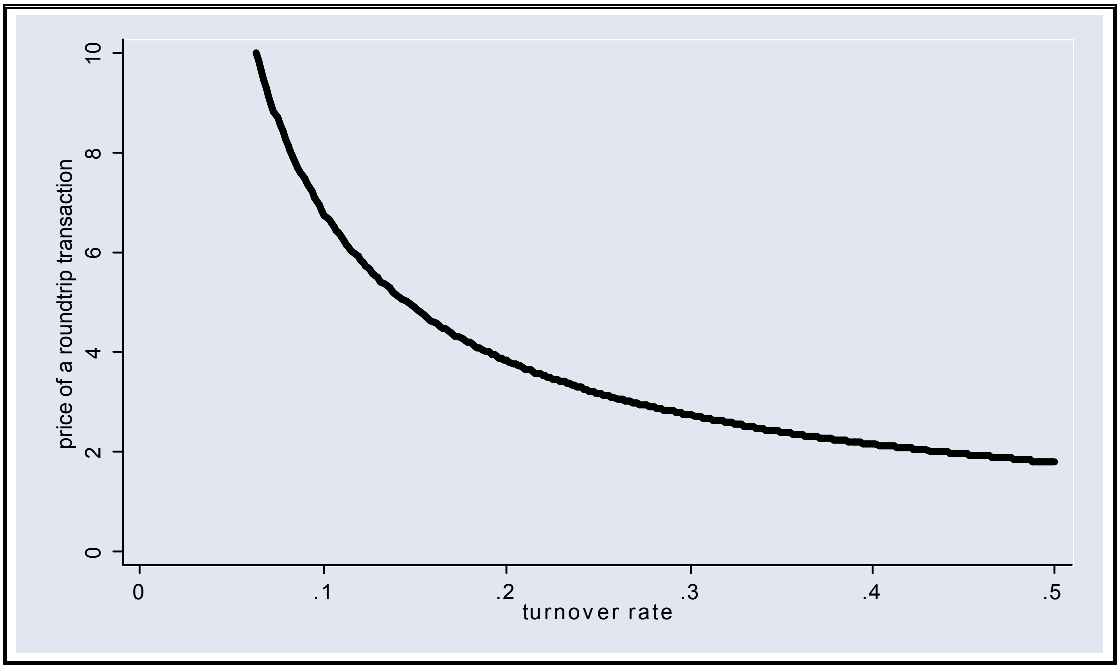
---

[23] Constantinides simulations on US historical data lead him to conclude that "a small liquidity premium is sufficient to compensate an investor for deviating significantly from the target portfolio proportions" [p. 843].

[24] McGrattan and Prescott argue that in fact equity mutual-fund costs understate diversification costs because they do not include brokerage charges.

[25] In some analyses numbers of trades are modelled.

**Figure 5.3**
**The demand for securities transactions**



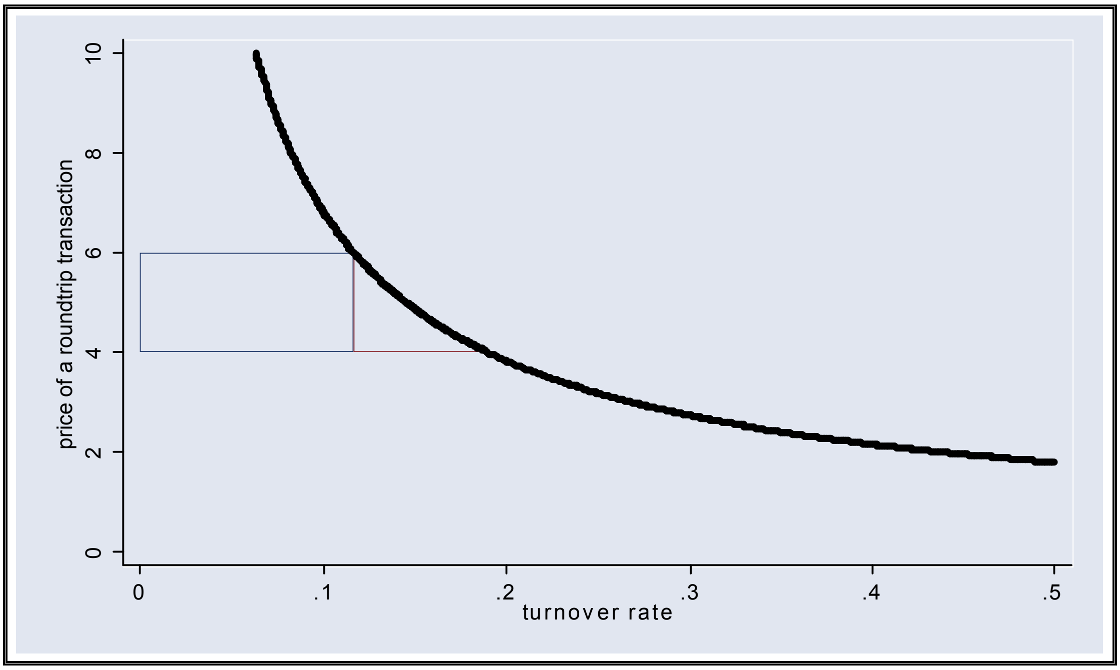

Transaction costs are given by the product of the transaction price and the transaction volume. At a transaction price of 5 per cent, the turnover rate is 11.6 per cent and therefore the transaction costs are 0.7 per cent per annum. At a transaction price of 2.5 per cent, the turnover rates is 26.8 per cent and therefore the transaction costs are 0.8 per cent per annum. It is notable, and counterintuitive, that the reduction in the round-trip price of a transaction has actually led to an increase in transaction costs! The reason is that in the example the demand for transactions has an elasticity greater than unity – an elasticity of 1.2.

It is important to understand that although transaction costs have risen, the investor is not worse off. The area under the schedule and to the right of the investor's turnover rate is the required liquidity premium, and when the turnover rises this falls. The investor could have maintained his turnover rate at 14 per cent when the transaction price fell from 5 per cent to 2.5 per cent, and would have enjoyed a reduction in transaction costs of 0.35 per cent. But the investor actually increased his turnover, taking advantage of the lower price. Thus the gain to the investor is actually equal to the shaded area in Figure 5.4.

The challenge then is to assemble the data necessary to estimate how the liquidity premium has changed in Australia. The data on Australian transaction costs over time are in fact very limited. In a study of a large dataset of stock market transactions, Swan (1994) estimated that the round-trip transaction cost on the ASX in 1993-94 was 2.4 per cent, made up of a round-trip brokerage charge of 0.8 per cent, a bid-ask spread of 1 per cent, and round-trip stamp duty of 0.6 per cent.[26] Thus stamp duties accounted for about 25 per cent of transaction costs. But as recently as 1982 (the beginning of Swan's study period) transaction costs were estimated to account for just 9 per cent of transaction costs, implying an average round trip transaction cost of 6.7 per cent. Clearly there was a large fall in round-trip costs over the period 1982 to 1994 (the deregulation of brokerage in 1984 and the introduction of new information technologies are likely explanations).

[26] The estimate has a heavy weighting of institutional clients, and the costs would be larger for small investors.

**Figure 5.4**
**The reduction in the liquidity premium**




In addition, there have almost certainly been further falls since 1994. Stamp duties were halved (to 0.3 per cent on a round-trip) in 1995 and then abolished in 2000. There have potentially been some technology related reductions too and also a likely reduction in the bid-ask spread as market turnover has risen (from under 30 per cent in 1994 to around 50 per cent in recent years – see ASX data).

An estimate was made of the reduction in the liquidity premium using the following assumptions:

- round-trip transaction costs 6.7 per cent in 1982;
- round-trip transaction costs then declining linearly to reach 2.4 per cent in 1994;
- round-trip transaction costs then declining to 2.1 per cent in 1995; and
- round-trip transaction costs then declining to 1.8 per cent in 2000.

The transaction price movements implied in this schedule were applied, year by year, to the turnover rate on the Australian Stock Exchange to get a series of annual changes in the liquidity premium for equities. The calculations are shown in Table 5.1. The conclusion is that between 1982 and 2000 the liquidity premium on equities fell by about 1.1 per cent, due to the influence of lower transaction costs and improvements in portfolio rebalancing.

To calculate the effect of reductions in the price of transactions on excess returns, one would ideally make allowance for the impact of changed transaction prices on liquidity premia in the bond market. This has not been done because the data that are needed to do so are not readily available. However, there are grounds to believe that transaction prices in the bond market are and have for long been significantly lower in the bond market than the equity market and that therefore the gains to be had from small price reductions are considerably smaller than in the equity market.

**Table 5.1**
**Change in annual liquidity premium for Australian equities 1982 to 2004**

| Year to Dec | Market capitalisation ($bn) | Turnover ($bn) | Turnover rate | Round-trip cost (%) | Annual change in % annual liquidity premium | Cumulative change in % annual liquidity premium |
|---|---|---|---|---|---|---|
| 1982 | 54* | 7* | 0.13* | 6.67 | #N/A | |
| 1983 | 60* | 12* | 0.21* | 6.31 | -0.04 | -0.04 |
| 1984 | 84* | 15* | 0.18* | 5.96 | -0.07 | -0.12 |
| 1985 | 130* | 31* | 0.24* | 5.60 | -0.07 | -0.18 |
| 1986 | 180 | 40 | 0.22 | 5.24 | 0.00 | -0.18 |
| 1987 | 191 | 81 | 0.43 | 4.89 | -0.08 | -0.26 |
| 1988 | 215 | 49 | 0.23 | 4.53 | -0.15 | -0.41 |
| 1989 | 230 | 57 | 0.25 | 4.18 | -0.08 | -0.49 |
| 1990 | 197 | 51 | 0.26 | 3.82 | -0.09 | -0.58 |
| 1991 | 261 | 60 | 0.23 | 3.47 | -0.09 | -0.67 |
| 1992 | 389 | 62 | 0.16 | 3.11 | -0.08 | -0.76 |
| 1993 | 477 | 100 | 0.21 | 2.76 | -0.06 | -0.81 |
| 1994 | 450 | 129 | 0.29 | 2.40 | -0.07 | -0.89 |
| 1995 | 546 | 133 | 0.24 | 2.10 | -0.09 | -0.97 |
| 1996 | 615 | 185 | 0.30 | 2.10 | 0.00 | -0.97 |
| 1997 | 777 | 229 | 0.30 | 2.10 | 0.00 | -0.97 |
| 1998 | 879 | 256 | 0.29 | 2.10 | 0.00 | -0.97 |
| 1999 | 845 | 307 | 0.36 | 2.10 | 0.00 | -0.97 |
| 2000 | 1,006 | 391 | 0.39 | 1.80 | -0.11 | -1.08 |
| 2001 | 1,110 | 476 | 0.43 | 1.80 | 0.00 | -1.08 |
| 2002 | 994 | 543 | 0.55 | 1.80 | 0.00 | -1.08 |
| 2003 | 1,099 | 567 | 0.52 | 1.80 | 0.00 | -1.08 |
| 2004 | 1,326 | 709 | 0.53 | 1.80 | 0.00 | -1.08 |

<u>Note</u>: * figures relate to year ending in June of the following year (i.e. 1982 refers to 1982-83 data).

# 6. Potential biases in excess return series

Before inferring anything about the market risk premium from these averages of historic excess returns, one should first ask whether there are any obvious biases which might affect them.

There are two main types of bias which are of concern – unanticipated changes in the cost of capital and measurement problems in indexes..

## 6.1 Changes in the cost of capital

If an investor's rate of return for a particular type of asset falls, while the earnings outlook remain constant, then the investor's valuation of the asset will rise. The return in that period will then be boosted. As a consequence, the average of historic excess returns will then be boosted in periods when there are unanticipated reductions in required rates of return (and conversely depressed in periods when there are increases in investor's required rates of return). Changes that are anticipated should have affected excess returns at the time that they came into anticipation, so changes that occur but were already anticipated have no effect.

If one conceives of an equity return as comprising five element – a real interest component, an inflation-compensation component, a market risk premium, a tax component and a liquidity premium – then unanticipated changes in these elements may affect excess returns. Over the period 1974, the influence of these factors may be summarised as follows.

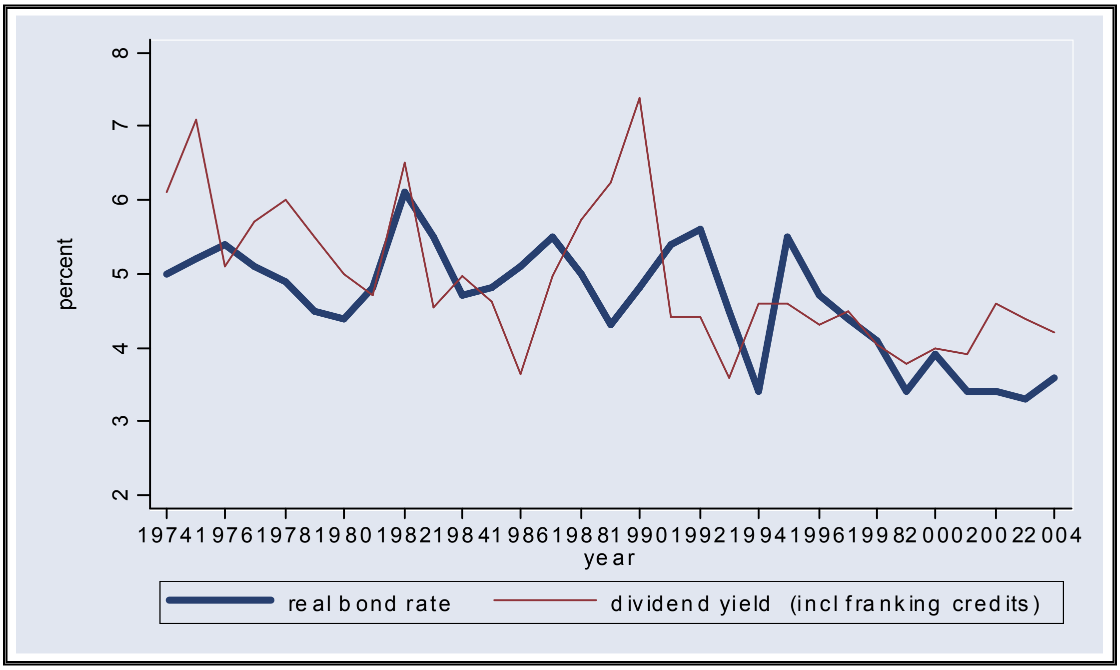
i) One possible influence – as has been argued in respect of the US by Fama and French (2002) – is that there has been a long term downward move in discount rates, that this has led to unanticipated capital gains on stocks, and that average stock returns over the last half century have therefore been higher than expected. If that is so, then excess returns would incorporate biases.

Figure 6.1 shows Australian real long term interest rates over the last 30 years, and it is clear that there has been a downward move. Roughly, real interest rates fell from around 5 per cent in the mid 1970s to about 3½ by 2004.

A similar decline can be seen in dividend yields (including franking credits). Dividend yields are only a part of equity yields. While it is possible that there has been a corresponding increase in ex ante capital gains, it is by no means clear that this is so.

ii) While it seems quite likely that the pure risk premium has changed over time – for instance because of changes in perceived or actual levels of risk and because of changes in investors' attitudes to risk – the measurement difficulties with identifying the pure risk premium make it difficult to quantify this.

**Figure 6.1**
**Australian real long term interest rates**




Note: These real interest rates are calculated by deducting estimates of inflation expectations from Government bond rates. The data are calculated as (1+nominal bond rate)/(1+inflation expectation)-1. The bond rate and the inflation expectation are measured with data from Australian Bureau of Statistics (2005), which is based on the Commonwealth Treasury's TRYM model. The bond rate is "Interest Rate: 10 Year Bonds" taken from Table 31, and the inflation expectation is "Full Information Inflation Expectations" taken from Table 23. The inflation expectation is measured by the difference between the yield on non-indexed and indexed 10-year Treasury bonds (Commonwealth Treasury 1996).

iii) There have been changes to taxation rates, to inflation rates (which interact with tax rates) and to investment channels which make it difficult confidently to say much about the impact of tax changes. In an earlier version of this report emphasis was placed on the introduction of dividend imputation and its impact on excess returns, but this approach neglected other changes such as the introduction of capital gains tax which potentially had a negative effect on excess returns. It is very uncertain what the overall impact has been, but the experimental estimates in the previous section suggest that tax changes may have been of the order of a 1 per cent increase.

iv) It has been argued in the previous section that the liquidity premium for holding equities has almost certainly fallen – by about 1 per cent it is estimated.

The influences (i), (iii) and (iv), taken together, have a downward impact of about 1½ per cent on the gross equity return. Assuming (purely for the purposes of this rough arithmetic, the exact number is not particularly important) that the market risk premium was 6 per cent throughout the period, then the expected equity return fell 1.5 per cent from (say) 10.5 per cent to 9 per cent. Such a change would, on an unchanged earning outlook, boost stock values by about 15 per cent. Thus, over the period 1974 to 2004, this might have added something like ½ per cent per annum to excess returns. However, the uncertainties around this are obviously large.

## 6.2 Mismeasurement in stock indexes

A second type of bias is the potential for bias in the actual equity accumulation indexes, for instance so-called "success" and "survivorship" biases, by which is meant the tendency to include disproportionately investments that succeed or survive. To avoid bias, excess return calculations should be representative of what an investor might have done when making an investment. The index portfolio that results will be different from one based on relative performance later on: investments that do well will be relatively more important later on and investments that did poorly will be relatively less important.[27]

However, while these flaws have been found to be serious in various overseas studies, they are less likely to be a problem with the data sets under review here. The early period of the Officer data comes from indices assembled by Lamberton (1958). Lamberton was certainly aware of the need to avoid the selection biases that can arise and he went to some length to do so (see pp. 49-50 of his work). Although he was at pains to point out that his work had not perfectly addressed every possible problem, there is no reason to believe that it contained large biases.

Even if stock indices are well constructed, it is possible that success biases affect countries as a whole. For instance, if all Australian investments have succeeded more than expected, this would mean that realised excess returns are still a biased estimate of the market risk premium. Success and survival bias at the country level cannot easily be estimated.[28]

Taken together, it seems likely that the introduction of dividend imputation and declining discount rates have combined to add about 1 percentage points per annum to Australian excess returns over the period 1974 to 2003. This would suggest that the average of excess returns over this period, which is around 5½ to 6 per cent depending on the dataset, is upward-biased by about 1 per cent as an estimator of the market risk premium. Thus central estimates of the 1-year equity premium over the last 30 years are in the 4½ to 5 per cent range. The most precise estimate has a 95 per cent confidence interval of approximately plus or minus 2 percentage points.

27 For discussion see Dimson, Marsh and Staunton (2002), especially Chapter 3.

28 Recent theoretical work by Li and Xu (2002) suggests that survival biases cannot be particularly large over long periods. In their study of the issue Brown, Goetzmann and Ross (1995) noted that, inter alia, the Amsterdam, Berlin, Brussels, Copenhagen, Frankfort, Madrid and Tokyo stock exchanges have had trading interruptions since 1901. Data assembled by Dimson, Marsh and Staunton (2002) show that average excess returns over the 20th century are not consistently lower for this group than for among the English-speaking countries. Yet these data are potentially misleading. In countries where there was significant unanticipated inflation (e.g. countries that had hyperinflations), nominal bond yields will be grossly inadequate as measures of expected risk free rates of return.

# References

Australian Bureau of Statistics (2005), *Modeller's Database.* Cat. No. 1364.0.15.003.

Blume, M. (1974), 'Unbiased estimation of long-run expected rates of return', *Journal of the American Statistical Association,* vol. 69, pp. 634-638.

Brealey, Richard A. and Stewart C. Myers (2003), *Principles of corporate finance,* 7th ed. McGraw-Hill.

Brown, Stephen J., William N. Goetzmann and Stephen A. Ross (1995), 'Survival', *Journal of Finance,* vol. 50 no. 3, pp. 853-873.

Campbell, John Y., Andrew W. Lo and A. Craig MacKinlay (1997), *The Econometrics of Financial Markets.* Princeton University Press.

Cochrane, John H. (1988), 'How Big Is the Random Walk in GNP?', *Journal of Political Economy,* vol. 96 no. 5, pp. 893-920.

Cogley, Timothy and Thomas J. Sargent (2004), 'The market price of risk and the equity premium: a legacy of the Great Depression?', Working Paper [December 2004] http://www.econ.ucdavis.edu/faculty/twcogley/

Commonwealth Treasury (1996), *The Macroeconomics of the TRYM Model of the Australian Economy.*

Constantinides, George M. (1986), 'Capital market equilibrium with transaction costs', *Journal of Political Economy,* vol. 94 no. 4, pp. 842-862.

Cooper, I. (1996), 'Arithmetic versus geometric mean estimators: setting discount rates for capital budgeting', *European Financial Management,* vol. 2, pp.157-67.

Covick and Lewis (1993), 'Insurance, superannuation and collective investments', in M. K. Lewis and R. H. Wallace (eds.) *The Australian Financial System,* pp. 136-213.

Davis, Lance E. and Robert E. Gallman (2001), *Evolving financial markets and international capital flows: Britain, the Americas and Australia, 1865-1914.*

Dimson, Elroy, Paul Marsh and Mike Staunton (2002), *Triumph of the Optimists: 101 Years of Global Investment Returns.* Princeton University Press.

Drake, P. J. (1985), 'Stock and share markets', in M. K. Lewis and R. H. Wallace (eds.) *Australia's Financial Institutions and Markets,* pp. 247-304.

Enders, Walter (1995), *Applied Econometric Time Series.* John Wiley & Sons, Inc.

Essential Services Commission Victoria (2002b), *Review of Gas Access Arrangements: Final Decision.*

Fama, Eugene F. and Kenneth R. French (1988), 'Permanent and transitory components of stock prices', *Journal of Political Economy,* vol. 81, pp. 246-273.

Fama, Eugene F. and Kenneth R. French (2002), 'The equity premium', *Journal of Finance,* vol. 57 no. 2, pp. 637-659.

Fisher, Stephen J. (1994), 'Asset trading, transaction costs and the equity premium', *Journal of Applied Econometrics,* vol. 9, supplement: Special Issue on Calibration Techniques and Econometrics, pp. S71-S94.

Grant, J. McB. (1974), 'Life assurance and superannuation funds', in R. R. Hirst and R. H. Wallace (eds.) *The Australian Capital Market,* pp. 166-206.

Gray, Stephen (2001), *Issues in Cost of Capital Estimation.* Annexure D in United Energy (2001), *Submission to the Office of the Regulator-General: 2003 Review of Gas Access Arrangements – Position Paper.*

Greene, William H. (2003), *Econometric Analysis, 5th edn.* Prentice Hall.

Gujarati, Damodar N. (2003), *Basic Econometrics, 4th ed.* McGraw Hill.

Hamilton, James D. (1994), *Time Series Analysis.* Princeton University Press.

Harvey, Andrew C. (1993), *Time Series Models, 2nd ed.* Harvester Wheatsheaf.

Hodrick, Robert J. and Edward C. Prescott (1997), 'Postwar U.S Business Cycles: An Empirical Investigation', *Journal of Money, Credit and Banking,* vol. 29 no. 1, pp1-16.

Jacquier, Eric, Alex, Kane and Alan J. Marcus (2002), 'Optimal Forecasts of Long-Term Returns: Geometric, Arithmetic, or Other Means' (October 2002). *Boston College Working Paper.* http://ssrn.com/abstract=353242

Jacquier, Eric, Alex, Kane and Alan J. Marcus (2002), 'Geometric or Arithmetic Mean: A Reconsideration', *Financial Analysts Journal,* vol 59 no 6, pp. 46-53.

Kreyszig, Erwin (1970), *Introductory Mathematical Statistics: Principles and Methods.* John Wiley & Sons.

Lally, Martin (1995), 'The accuracy of CAPM proxies for estimating a firm's cost of equity', *Accounting and Finance,* vol. 35 no. 2, pp. 63-72.

Lally, Martin (2004), *The cost of capital for regulated entities.* Report prepared for the Queensland Competition Authority.

Lamberton, D. McL (1958), *Share Price Indices in Australia.* The Law Book Co. of Australasia.

Li, Haitao and Xu Yuewu (2002), 'Survival Bias and the Equity Premium Puzzle', *Journal of Finance,* vol. 57 no. 5, pp.1981-1995.

Mehra, Rajnish and Edward C. Prescott (1985), 'The equity premium: a puzzle', *Journal of Monetary Economics,* vol. 15, pp. 145-161.

Merton, Robert C. (1980), 'On estimating the expected return on the market', *Journal of Financial Economics,* vol 8, pp. 323-361.

McGrattan, Ellen R. and Edward C. Prescott (2003), 'Average debt and equity returns: puzzling?' *American Economic Review,* vol. 93, no. 2, pp. 392-97.

Officer, R. R. (1989), 'Rates of Return to Shares, Bond Yields and Inflation Rates: An Historical Perspective', in Ray Ball, Philip Brown, Frank J. Finn and R. R. Officer (eds.), *Share Markets and Portfolio Theory: Readings and Australian Evidence.* University of Queensland Press.

Pedersen, Torben Mark (2001), 'The Hodrick-Prescott filter, the Slutzky effect, and the distortionary effect of filters', *Journal of Economic Dynamics and Control,* vol.25 iss. 8, pp. 1081-1101.

Poterba, J. and L. Summers (1988), 'Mean reversion in stock prices: evidence and implications', *Journal of Financial Economics,* vol. 22, pp. 27-59.

Reserve Bank of Australia (2004), *Bulletin Statistical Tables.* http://www.rba.gov.au/. [downloaded ]

Richardson (1991), 'Temporary components of stock prices: a skeptics view', *Journal of Business and Economic Statistics,* vol. 11(2), pp. 199-207.

Sharpe, William F. (2005), *Macro-Investment Analysis.* http://www.stanford.edu/~wfsharpe/mia/mia.htm [downloaded 12 April 2005].

Sorensen, Erik O. (2005), *Hodrick-Prescott filter for Stata.* [Downloaded 12/01/2005 from http://homepage.mac.com/sameos/research/stata/StateHP.html]

Standard & Poor's (2005), *Results for: S&P/ASX 200 Index.* http://www2.standardandpoors.com/ [downloaded 5 January 2005].

Swan, Peter (1994), 'Responsiveness of share trading to transaction costs', *ASX Perspective,* vol.1, pp. 19-26.

Swan, Peter L. (2005), 'Optimal portfolio balancing under conventional preferences and transactions costs explains the equity premium puzzle', Working Paper, School of Banking and Finance, Faculty of Commerce, UNSW, Sydney. Version dated December 10, 2005.

Verbeek, Marno (2004), *A Guide to Modern Econometrics, 2nd ed.* John Wiley & Sons.

Wooldridge, Jeffrey M. (2001), *Econometric Analysis of Cross Section and Panel Data.* MIT Press.

____________________ (2003), *Introductory Econometrics: A Modern Approach, 2e.* Thomson South-Western.

Wright, Stephen, Robin Mason and David Miles (2003), *A study into certain aspects of the cost of capital for regulated utilities in the U.K.* [Downloaded from http://www.psc.gov.uk/documents/competition/Websiteversion.pdf]

Wackerly, Mendenhall and Scheaffer (2002), *Mathematical Statistics with Applications.* 6th ed. Duxbury.

Washusen, Jeff (2004), *Victorian Electricity Distribution Price Review Implications for customers – Preliminary Indications.* Presentation to VIC Consumer Group Meeting, Melbourne, 21 December.

# Appendix A

## Multi-period returns

When it is necessary to estimate a 1-period return, the arithmetic average of a sample of 1-period returns from a stable distribution gives an unbiased estimate of the mean. But the precision of an estimate of the mean depends on the number of observations in the sample. If returns are available over shorter periods then a larger sample can be obtained by taking observations over shorter periods. The challenge then is to build up from those short 1-period observations to a period which is of greater interest. An obvious example is the case of building up an estimate of 1-year expected returns from monthly or quarterly data.

An important point in such an exercise is that autocorrelation in the data can seriously affect such an exercise. In Section A.1 the point is illustrated with a simple example in which a 2-year return is considered as the outcome of two 1-year returns. It is shown that the stochastic properties of the 1-year returns affect the expected 2-year rate of return.

In Section A.2 a more general case is considered in which returns are distributed lognormally at all return lengths. It is shown why lognormality is a more reasonable assumption than normality. Then the relation between H-period expected returns and 1-period expected returns is explored, including in the presence of autocorrelation.

Section A.2 provides exact relations between 1-period and H-period returns when the parameters of the underlying probability distributions are known. But usually those parameters are not known and must be estimated. For instance, this study is fundamentally concerned with returns to risk capital in Australia, and certainly the underlying probability distributions for returns are not known with certainty. Studies by Blume (1974), Cooper (1996) and Jacquier, Kane and Marcus (2002, 2003) have shown that unbiased estimates of 1-period returns do not generally produce unbiased estimates of H-period returns. In Section A.3 these bias issues and possible adjustments are considered.

### A.1 Numerical example

In this section a 2-year return is decomposed into two 1-year returns and the relationship between 1-year and 2-year returns is explored. The numerical example draws on Campbell (2001), but a wider range of issues is illustrated here.

The discussion commences with a consideration of deterministic returns, then turns to a consideration of stochastic returns and the implications for the distribution of returns, and then unifies the two by introducing autocorrelation as a measure of the extent of determinism in returns.

### *Deterministic returns*

First, take the deterministic case in which returns are known with certainty. Suppose an investor knows that he can make an investment in year 1 and earn a grow return of 1.500, make an investment in year 2 and earn a gross rate of return of 0.667 (ie. a net loss of 1/3). Various aspects of the return structure of this investment are summarised in "Deterministic case" in Table A.1.

**Table A.1**
**Expected 2-year returns under alternative stochastic properties of 1-year returns**

| Case | Probability | Actual returns (R) | | | Arithmetic average of 1-year actual returns |
|---|---|---|---|---|---|
| | | Year 1 | Year 2 | 2 Years | |
| **Deterministic case** | | | | | |
| Scenario: | 1.00 | 1.500 | 0.667 | 1.000 | 1.083 |
| Expected value: | | 1.500 | 0.667 | 1.000 | |
| **Stochastic independent case** | | | | | |
| Scenario 1 | 0.25 | 1.500 | 1.500 | 2.250 | 1.500 |
| Scenario 2 | 0.25 | 1.500 | 0.667 | 1.000 | 1.083 |
| Scenario 3 | 0.25 | 0.667 | 1.500 | 1.000 | 1.083 |
| Scenario 4 | 0.25 | 0.667 | 0.667 | 0.444 | 0.667 |
| Expected value | | 1.083 | 1.083 | 1.174 | |
| **Stochastic perfect dependence case** | | | | | |
| Scenario 1 | 0.00 | 1.500 | 1.500 | 2.250 | 1.500 |
| Scenario 2 | 0.50 | 1.500 | 0.667 | 1.000 | 1.083 |
| Scenario 3 | 0.50 | 0.667 | 1.500 | 1.000 | 1.083 |
| Scenario 4 | 0.00 | 0.667 | 0.667 | 0.444 | 0.667 |
| Expected value | | 1.083 | 1.083 | 1.000 | |
| **Stochastic imperfect dependence case** | | | | | |
| Scenario 1 | 0.10 | 1.500 | 1.500 | 2.250 | 1.500 |
| Scenario 2 | 0.40 | 1.500 | 0.667 | 1.000 | 1.083 |
| Scenario 3 | 0.40 | 0.667 | 1.500 | 1.000 | 1.083 |
| Scenario 4 | 0.10 | 0.667 | 0.667 | 0.444 | 0.667 |
| Expected value | | 1.083 | 1.083 | 1.069 | |

The 2-year gross rate of return is 1.000, which is equal to the product of the two 1-year gross rates of return (1.000 = 1.500 x 0.667).

The expected value of year 1 returns is 1.500, the expected value of the year 2 return is minus 0.667, and the expected value of the 2-year return is 1.000.

The arithmetic average of the 1-year net rates of return is 1.083 per cent. However, if this arithmetic average is compounded it does not give the 2-year return (compounding the arithmetic average twice gives a gross return of 1.174 = 1.0833 x 1.0833).

The geometric 1-year average return is, by definition, the square root of the 2-year return (alternatively, the geometric average is that return which when compounded gives the 2-year return). Thus the geometric average return is 1.000 per annum.

What is the cost of capital? Remember that there is (by assumption) no uncertainty about these returns. The Year 1 cost of capital is 1.500 and the Year 2 cost of capital is 0.667. The 2-year cost of capital is 1.000.

### *Stochastic independent returns*

Now consider the case in which the 1-year returns and the 2-year return are not deterministic but are in fact random (or "stochastic"). Suppose that there are two possible outcomes in year 1 and year 2, a return of 1.500 and a return of 0.667, each with a probability of 0.5, and that the outcomes are independent of each other. "Independent" in this context means that the probability of a particular outcome in year 2 is the same regardless of what happened in year 1 (and vice versa).

In this case there are four possible 2-year outcomes – (1.500, 1.500), (1.500, 0.667), (0.667, 1.500) and (0.667, 0.667). The respective 2-year returns are 2.250, 1.000, 1.000 and 0.444 per cent (see Table A.1).

The probability of each outcome is 0.5 x 0.5 = 0.25. This allows us to take the expectation across these outcomes, to calculate expected returns for year 1, year 2 and 2 years. The expected returns for years 1 and 2 are 1.083, and the expected 2-year return is 1.174 per cent. Note that, unlike the deterministic case, this expected 2-year return is equal to the product of the expected 1-year returns.

### *Stochastic perfectly dependent returns*

Now consider the case in which the 1-year returns are random but their expected values are dependent (in which case we say that returns are "autocorrelated"). Take the case where the return in year 1 is random, and that there are two possible returns: 1.500 and 0.667, each with a probability of 0.5. But now assume that the outcome in year 2 depends on the outcome in year 1.

Consider a very strong form of dependence in which (a) the probability of a 1.500 return in year 2 is 0.0 if there was a 1.500 return in year 1, and 1.0 if there was a 0.667 return in year 1 and (b) the probability of a 0.667 return in year 2 is 1.0 if there was a 1.500 return in year 1, and 0.0 if there was a 0.667 return in year 1.

This means that the probabilities attached to the 4 outcomes are: Prob(Outcome 1) = 0.0; Prob(Outcome 2) = 0.5; Prob(Outcome 3) = 0.5; and Prob(Outcome 4) = 0.0 (meaning that Outcomes 1 and 4 definitely will not happen).

These probabilities can then be applied to the returns associated with each outcome to get expected values. The expected 1-year returns are 1.083 (the same as in the purely stochastic case) but the expected 2-year return is 1.000 (not the same as the stochastic case, but the same as the deterministic case).

### *Autocorrelated and stochastic returns*

The "stochastic independent returns" and "stochastic perfectly dependent returns" cases are at opposite ends of a spectrum. A more likely situation is one in which there is imperfect dependence in returns.

Suppose that the possible returns in years 1 and 2 are still 1.500 and 0.667. In year 1 the probability of each is 0.5. But in year 2 (a) the probability of a 1.500 return is 0.2 if there was a 1.500 return in year 1, and 0.8 if there was a 0.667 return in year 1 and (b) the probability of a 0.667 return is 0.8 if there was a 1.500 return in year 1, and 0.2 if there was a 0.667 return in year 1.

This means that the probabilities of the 4 possible outcomes are: Prob(Outcome 1) = 0.1; Prob(Outcome 2) = 0.4; Prob(Outcome 3) = 0.4; and Prob(Outcome 4) = 0.1.

The expected 1-year returns are still 1.083. The expected 2-year return is now 1.069. This is greater than the expected 2-year return of 1.000 under full dependence but less than the expected 2-year return of 1.174 under independence (no autocorrelation).

***Implications***

All these comments relate of course to a very simple example. But the example does illustrate that the connection between a 1-period expected return and a 2-period expected return depends on the extent of any dependence in the sequence of returns. If the returns are stochastic and independent, the 1-period return can be compounded to give the 2-period expected return. But if the returns are not independent – in particular if there is autocorrelation – then compounding the 1-period expected returns will overstate the 2-period expected return.

## A.2 Multi-period returns under lognormality

Consider an $H$-period gross return $R(H)$, made up of $H$ 1-period returns $R_i$, $i = 1, ..., H$. The relationship between $R(H)$ and the $R_i$ is:

$$R(H) = \prod_{i=1}^{H} R_i \qquad \text{(A.1)}$$

By taking logs, the multiplicative process in Equation A.1 can be converted to an additive process:

$$\ln R(H) = \sum_{i=1}^{H} \ln R_i \qquad \text{(A.2)}$$

If we make the assumption that the 1-period returns are normally distributed, it follows that they will not be normally distributed for any return length other than 1 period. Taking the case of $H$ = 2, for instance,

$$\begin{aligned} 1 + R(2) &= (1 + R_1)(1 + R_2) \\ &= 1 + R_1 + R_2 + R_1 R_2 \end{aligned}$$

While the terms $R_1$ and $R_2$ will be normally distributed, the term $R_1 R_2$ will not be, and consequently $R(2)$ will not be. This means that while there might be some length of period for which normality holds, it cannot be expected to hold in general. There is therefore an internally contradictory character to an assumption of normality in returns.

On the other hand, a sum of normal random variables is itself normal. The log of the $H$-period gross return is simply the sum of $H$ 1-period gross returns. So if the 1-period returns are lognormal, the $H$ -period return is lognormal too.

Lognormality of returns is therefore a characteristic which could apply across all return period lengths (whether it does apply is of course an empirical question). It is also consistent with the multiplicative process for aggregating sequential returns. Lognormality is therefore a much better candidate than normality for the task of modelling returns. Jacquier, Kane and Marcus (2002) set out some of the properties of lognormally distributed returns and consider the structure of expected returns over different horizons, and these are now summarised and developed.

The lognormal distribution is asymmetric. If $R$ is lognormally distributed and $\ln R$ has mean $\mu$ and variance $\sigma^2$, then the expected value of $R$ is given by

$$E(R) = e^{(\mu+\sigma^2/2)} \qquad \text{(A.3)}$$

Using $\mu(H)$ to designate the mean of $\ln R(H)$ and $\sigma(H)^2$ to designate its variance, then by substitution in Equation A.3

$$E(R(H)) = e^{(\mu(H)+\sigma(H)^2/2)} \qquad \text{(A.4)}$$

Two important results from mathematical statistics are that if $X_1, X_2, ..., X_N$ are random variables each with mean $\overline{X}_i$ and variance $Var(X_i)$ then the sum of the random variables $Y$ is also a random variable with the following properties:[29]

a) The expected value of $Y$ is the sum of the each of the expected values of $X_i$

$$E(Y) = \sum_{i=1}^{N} E(X_i)$$

b) The variance of $Y$ is the sum of the variances and covariances of $X_i$

$$Var(Y) = \sum_{i=1}^{N} \sum_{j=1}^{N} \text{cov}(X_i, X_j)$$

This then allows a consideration of the implications for $E(R(H))$ of different assumptions about the distribution of the 1-period returns $R_i$.

Case 1: One possibility is that the $R_i$ are independent and identically distributed (i.i.d.). In that case, $\mu(H) = H\mu$ and $\sigma(H)^2 = H\sigma^2$. These can be incorporated into Equation 5.5 to give the standard result that the expected value of the $H$-period return is equal to the expected value of the 1-period return raised to the $H^{th}$ power:

[29] For details see Wackerly, Mendenhall and Scheaffer (2002) pp. 255-258.

$$\begin{aligned} E(R(H)) &= e^{(H\mu + H\sigma^2/2)} \\ &= \left[ e^{(\mu+\sigma^2/2)} \right]^H \\ &= (E(R))^H \end{aligned} \quad \text{(A.5)}$$

Case 2: Another possibility is that the 1-period returns are autocorrelated. For instance, it is often argued that equity returns exhibit a form of autocorrelation known as "mean reversion", which means that above average returns tend to be followed by below average returns. Where mean reversion exists, $\sigma(H)^2 < H\sigma^2$. Define the "variance ratio" as

$$VR(H) = \frac{\sigma(H)^2}{H\sigma^2} \quad \text{(A.6)}$$

in which case $\sigma(H)^2 = VR(H)H\sigma^2$. This can then be incorporated into Equation 5.5 to give

$$\begin{aligned} E(R(H)) &= e^{(H\mu + VR(H)H\sigma^2/2)} \\ &= e^{H(\mu+\sigma^2/2+(VR(H)-1)\sigma^2/2)} \\ &= \left[ e^{(\mu+\sigma^2/2)} e^{(VR(H)-1)\sigma^2/2} \right]^H \\ &= \left[ E(R) e^{(VR(H)-1)\sigma^2/2} \right]^H \end{aligned} \quad \text{(A.7)}$$

The *H*-period expected return is thus the compound of a term comprising the 1-period expected return and an autocorrelation adjustment. In the case where $VR(H) = 1$, which means there is no autocorrelation, the adjustment factor equals 1, and thus has no effect. In the case where $VR(H) = 0$, which means that the *H*-period return is deterministic even though 1-period returns are stochastic, the expected value of the *H*-period return is

$$\begin{aligned} E(R(H)) &= e^{(H\mu + VR(H)H\sigma^2/2)} \\ &= e^{H\mu} \\ &= [e^{\mu}]^H \end{aligned} \quad \text{(A.8)}$$

## A.3 Estimating expected multi-period returns

The discussion in Section A.2 produces exact analytical relations between 1-period expected returns and H-period expected returns, making use of the parameters of underlying return distributions. However, in practice the parameters of the return distribution are not known with certainty and must be estimated. While it is tempting to think that unbiased parameter estimates will produce unbiased estimates of expected H-period returns, this is generally not the case.

For the case where a series of returns is assumed to be independently, identically normally distributed, Blume (1974) illustrates that while the sample average will be an unbiased estimate of the 1-period expected return, raising it to the *H*th power will produce an upward biased estimate of the expected *H*-period return. On the other hand the sample geometric average raised to the *H*th power will produce a downward biased estimate of the expected *H*-

period return. He proposes two alternative estimators, each of which is based on a weighted average of the arithmetic and geometric sample averages.

Cooper (1996) considers the biases that can arise when estimating multi-period discount factors.

Jacquier, Kane and Marcus (2003) consider the estimation of multi-period expected returns under lognormality. In particular, they are concerned with the consequences of estimation error in the $e^{(H\mu+H\sigma^2/2)}$ component of Equation A.5 (but their concerns also apply to Equation A.7). Their concern is that while errors in unbiased estimates of $\mu$ and $\sigma^2$ will be symmetrical, errors in the exponential of them will generally be asymmetrical, which means that estimates of expected multi-period returns will be upward biased. They focus on the consequences of error in estimates of $\mu$, ignoring the consequences of error in estimates of $\sigma^2$ on the ground that $\sigma^2$ can be estimated arbitrarily accurately by increasing sampling frequency.[30] With a stable lognormal distribution it is possible to derive the exact bias in the estimate of an expected *H*-period return where the 1-period return is estimated from a sample of size T.

Take the case where $R$ is identically (but not independently) lognormally distributed:

$$\ln R \sim N(\mu, \sigma^2)$$

Now consider a sample of T observations of $R$. Each $R_i$ can be expressed as:

$$\ln R_i = \mu + \sigma\varepsilon_i$$

where $\varepsilon_i \sim N(0,1)$.

The following estimator $\hat{\mu}$ is an unbiased estimate of $\mu$

$$\begin{aligned}\hat{\mu} &= \frac{1}{T}\sum_{i=1}^{T}\ln R_i \\ &= \frac{1}{T}\left(\mu T + \sigma\sum_{i=1}^{T}\varepsilon_i\right) \qquad \text{(A.9)}\end{aligned}$$

The variance of $\hat{\mu}$ is $\sigma^2/T$. Therefore we can write $\hat{\mu} = \mu + \omega\sigma/\sqrt{T}$ where $\omega \sim N(0,1)$.

Let $A(H)$ be the arithmetic average estimator of expected *H*-period returns which is derived by using $\hat{\mu}$ to estimate $\mu$:

30 Increasing the frequency of sampling requires reducing the return durations in the sample. Although better estimates of $\sigma^2$ can be obtained with such a strategy, the precision of $\mu$ depends on the return length and therefore increased sampling frequency cannot improve estimates of $\mu$.

$$
\begin{aligned}
A(H) &= e^{(\hat{\mu}+\sigma^2/2)H} e^{(VR(H)-1)H\sigma^2/2} \\
&= e^{(\mu+\omega\sigma/\sqrt{T}+\sigma^2/2)H} e^{(VR(H)-1)H\sigma^2/2} \\
&= e^{(\mu+\sigma^2/2)H} e^{(VR(H)-1)H\sigma^2/2} e^{w\sigma H/\sqrt{T}} \\
&= E(R(H))e^{\omega\sigma H/\sqrt{T}}
\end{aligned}
\qquad (A.10)
$$

Then the expected value of $A(H)$ is

$$
\begin{aligned}
E(A(H)) &= E(R(H))E(e^{w\sigma H/\sqrt{T}}) \\
&= E(R(H))e^{\frac{1}{2}\sigma^2 H^2/T}
\end{aligned}
\qquad (A.11)
$$

The term $e^{\frac{1}{2}\sigma^2 H^2/T}$ is the upward bias factor. Note that when the power term is equal to zero it has no effect. The power term equals zero when $\sigma = 0$. It also approaches zero as the number of observations in the sample, T, approaches $\infty$.

A bias-adjusted estimator of $E(R(H))$, designated $U(H)$ can then be calculated:

$$
\begin{aligned}
U(H) &= e^{(\hat{\mu}+\sigma^2/2+(VR(H)-1)\sigma^2/2)H} e^{-\frac{1}{2}\sigma^2 H^2/T} \\
&= e^{(\hat{\mu}+(1-\frac{H}{T})\sigma^2/2)H} e^{(VR(H)-1)H\sigma^2/2}
\end{aligned}
\qquad (A.12)
$$

This has the same basic form as Equation A.6, but with the addition of the term in *–H/T*. The larger is the sample size relative to the return horizon the less is the impact of the adjustment factor.

# Appendix B

## The Officer Data Set

### B.1 Updating the Data

One of the most widely cited studies of the market risk premium in Australia is Officer's (1989) study covering the period 1882 to 1987. The data set was updated to 1999 and summary statistics were published in the Essential Services Commission Victorian (2002b).

Officer's data includes, year by year for the period 1882 to 1999, 10-year Treasury bond yields at the last available date in December and nominal equity returns to the end of December. Officer then calculates excess returns as the difference between the equity return to December and the bond yield in the previous December; thus the first observed excess return is for 1883. He then uses averages of these excess returns to generate estimates of the market risk premium, and uses their standard deviations to estimate the standard errors of those estimates of the market risk premium.

Professor Officer has kindly provided us with the underlying data set, and we have updated it to the end of 2004. Herein we allude to this extended data set when we refer to the "Officer data". Our update proceeded in the following way:

- Bond yields for 2000 to 2003 were incorporated from the Reserve Bank of Australia (2004) Table F2 using the end of December observations for 1999 to 2003.
- Officer's equity return series, which is an accumulation index, is a patched series drawing on different data sets in different periods (see Officer 1989 p. 211). The variant supplied to us uses the All Ordinaries Accumulation Index for 1982 to 1999. Therefore we updated the series by taking values of the end-December All Ordinaries Accumulation (gross) Index from Standard and Poors (2005). This introduced a revised 1999 index value and consequently a revision to the 1999 equity return (to 16.1 per cent from 19.3 per cent). Data were also incorporated for 2000 to 2004.

As would be expected our modifications had a minimal impact on descriptive statistics for the whole data set; we added 4 extra observations to increase its size to 122 observations and revised one of the existing observations. While these additions might not seem particularly valuable if one takes an equal-weighted view of the data set, they are potentially more important if one decides to accord more weight to recent observations.

### B.2 Basic Descriptive Statistics for the Officer Data

The arithmetic average excess return over the period was 7.3 per cent.

As can be seen in Figure 2.1, there is quite considerable variation from year to year in the excess return. The standard deviation is 17.0, the minimum value is -32.0 and the maximum is 53.8. The standard error of the mean was 1.5 per cent and the simple 95-percent confidence interval for the excess return was from 4.2 to 10.3 per cent (Table B.1).

The geometric average excess return over the period was 5.9 per cent, which is appreciably smaller than the arithmetic average.[31]

Table B.1 also shows the average excess return and its simple confidence interval for four (rather arbitrary) sub-periods. At face value the averages suggest a reduction in the market risk premium, but the confidence intervals are large (especially in later periods). Moreover, these findings are sensitive to the choice of endpoints. The arithmetic average for the period 1974 to 2004 was 6.4 per cent, compared with 7.8 per cent for the period 1943 to 1973, but its 95-percent confidence interval was from –4.1 to 13.6 per cent.

The fact that confidence intervals become progressively wider through later sub-periods reflects increases in the standard deviation of excess returns. This is also suggested by visual inspection of Figure 2.1: the series appears to be more variable after (say) World War Two.

Table B.1 also shows geometric average returns. These geometric averages are each lower than their corresponding arithmetic averages. However, it is notable that the differences get larger over time. For instance, for the period 1883 to 1912 the geometric average was 7.8 per cent, just 0.3 percentage points less than the arithmetic average. But over the period 1974 to 2004 the geometric average of 3.9 per cent was a large 2.5 percentage points lower than the arithmetic average of 6.4 per cent. The reason for the growing divergence was an increase in the variance (i.e. the volatility) of returns. For 1883 to 1912 the multiplicative standard deviation was just 1.4 per cent, whereas for 1974 to 2004 it had risen to 4.1 per cent.

**Table B.1**
**Summary Statistics for Excess Returns Over Different Time Periods**

| Period | Average | Standard error | 95% lower conf. int. bound | 95% upper conf. int. bound |
|---|---|---|---|---|
| Arithmetic averages: | | | | |
| 1883 to 2004 | 7.3 | 1.5 | 4.2 | 10.3 |
| | | | | |
| 1883 to 1912 | 8.1 | 1.5 | 5.1 | 11.1 |
| 1913 to 1942 | 6.8 | 2.4 | 1.8 | 11.8 |
| 1943 to 1973 | 7.8 | 3.5 | 0.7 | 15.0 |
| 1974 to 2004 | 6.4 | 4.2 | -2.2 | 15.0 |
| | | | | |
| Geometric averages:* | | | | |
| 1883 to 2004 | 5.9 | 1.5 | 2.8 | 9.1 |
| | | | | |
| 1883 to 1912 | 7.8 | 1.4 | 4.8 | 10.9 |
| 1913 to 1942 | 5.9 | 2.5 | 0.7 | 11.3 |
| 1943 to 1973 | 6.0 | 3.5 | -1.2 | 13.8 |
| 1974 to 2004 | 3.9 | 4.1 | -4.4 | 12.9 |

Note: * Geometric standard errors have been applied multiplicatively because they are calculated as additive quantities on logs.

The increasing divergence between geometric and arithmetic averages is entirely consistent with the increase in the variance of returns over time. For a given geometric average, higher variance in the dataset implies a higher arithmetic average.

[31] If returns are lognormally distributed, the arithmetic average exceeds the geometric average by half the variance of returns (see Campbell 2001 p.3). Half the variance of returns for the Officer series is 1.4 per cent, which reasonably closely matches the observed difference of 1.2 per cent.

Figure B.1 shows a histogram of excess returns. It also includes a superimposed normal curve (a mean and a variance are sufficient to identify the normal curve and the sample values are used for this purpose). Visual inspection suggests that the histogram has the bell shaped character of the normal distribution, but that it is "fat" in the tails and middle of the distribution (and correspondingly "thin" between the middle and the tails).[32]

However, a Jarque-Bera test does not reject the assumption of normality. The J-B statistic is 0.91, and there is a 64 per cent probability of obtaining this under the null hypothesis.

Normality in the arithmetic excess returns is a surprising phenomenon and difficult to accept. If we have normality of 1-period returns, then the distribution of n-period returns will be given by a product of n normal distributions, and the product of normal distributions is not itself a normal distribution. Normality of 1-period returns would then suggest non-normality of returns over shorter and longer durations. Moreover, normality would strictly require that returns of unlimited negative size are possible, whereas there are in fact limits – a negative excess return cannot exceed 100 plus the bond rate.

In contrast, lognormal returns do not suffer these defects. They can prevail across all return durations. That is, lognormality of 1-period returns is quite consistent with lognormal returns over n periods, where n may be greater than or less than one. For instance, lognormality of 1-year returns is consistent with lognormality of 5 year returns and of 1 month returns. And log returns approach minus infinity as the return approaches zero.

However, when confronted with the actual data – see Figure B.2 – the hypothesis that annual excess returns are lognormally distributed (i.e. that the log of 1 plus the excess return is normally distributed) is resoundingly rejected. The Jarque-Bera statistic is 11.6, and there is only a 0.3 per cent chance of achieving this. The hypothesis of lognormality is thus easily rejected at a 5-per cent and even a 1-per cent significance level. How much the deviations from lognormality matter is another question. Where it is necessary to construct multi-period returns from 1-period returns, and to carry out statistical inference on these, the assumption of lognormality is conceptually much better founded than an assumption of normality (see Appendix A).

[32] The implication is that extreme values have occurred rather more often than would occur if excess returns truly were normally distributed. In fact such a result is not surprising: it is commonly observed and could be explained, for instance, by changes in the variance. One possible cause of fat tails is drawing from a mixture of conditionally normal distributions with constant means but differing variances. See Campbell, Lo and MacKinlay (1997) pp. 13-20.

**Figure B.1**
**Histogram of Excess Returns**
**Annual Data for 1883 to 2004**

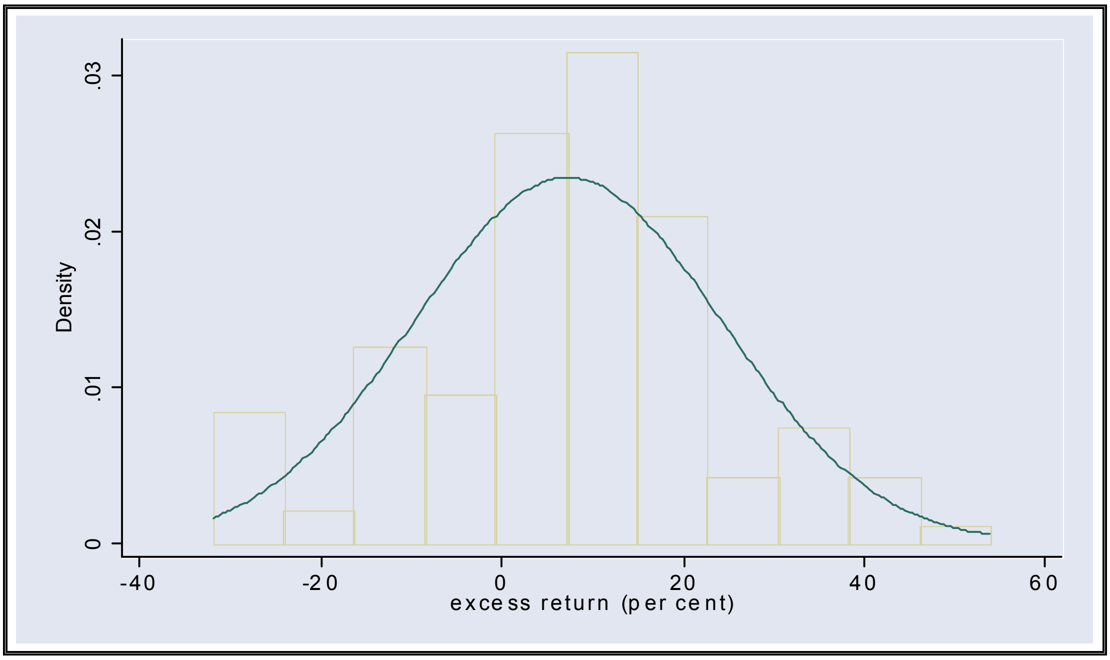
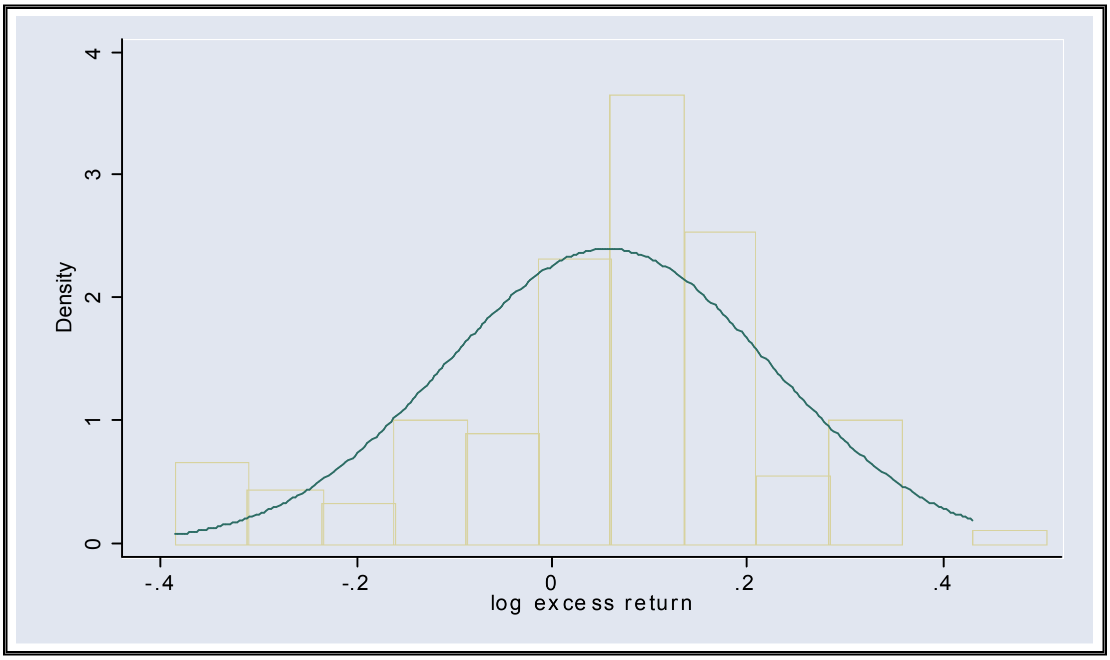



**Figure B.2**
**Histogram of Log Excess Returns**
**Annual Data for 1883 to 2004**



# Appendix C

## Deriving an optimal estimator from two independent estimators

Let $X$ and $Y$ be two estimators of a population mean $\mu$. An weighted estimator $Z$ can be formed as follows

$$Z = aX + (1-a)Y$$

By virtue of the fact that the weights sum to 1

$$\begin{aligned} E(Z) &= aE(X) + (1-a)E(Y) \\ &= a\mu + (1-a)\mu \\ &= \mu \end{aligned}$$

So the estimator $Z$ is unbiased.

If $X$ and $Y$ are independent of each other and $\sigma_X^2$ and $\sigma_Y^2$ are their respective variances, then the variance of $Z$ is

$$\sigma_Z^2 = a^2\sigma_X^2 + (1-a)^2\sigma_Y^2$$

To minimise $\sigma_Z^2$ differentiate with respect to a

$$\frac{\mathrm{d}\sigma_Z^2}{da} = 2a\sigma_X^2 - 2(1-a)\sigma_Y^2$$

Evaluating at zero gives the result that, for a minimum (or maximum)

$$a = \frac{\sigma_Y^2}{\sigma_X^2 + \sigma_Y^2}$$

And to confirm that it is a minimum, consider the second derivative

$$\frac{\mathrm{d}^2\sigma_Z^2}{da^2} = 2\sigma_X^2 + 2\sigma_Y^2$$

This second derivative must be positive, which implies that $\sigma_Z^2$ reaches a minimum when the first order condition is met.

To conclude, the most efficient estimator (i.e. the minimum variance estimator) given two independent estimators X and Y is given by

$$Z = \left(\frac{\sigma_Y^2}{\sigma_X^2 + \sigma_Y^2}\right)^2 X + \left(\frac{\sigma_X^2}{\sigma_X^2 + \sigma_Y^2}\right)^2 Y$$

and its variance is

$$\sigma_Z^2 = \left( \frac{\sigma_Y^2}{\sigma_X^2 + \sigma_Y^2} \right)^2 \sigma_X^2 + \left( \frac{\sigma_X^2}{\sigma_X^2 + \sigma_Y^2} \right)^2 \sigma_Y^2$$

$$= \frac{(\sigma_Y^2)^2 \sigma_X^2 + (\sigma_X^2)^2 \sigma_Y^2}{(\sigma_X^2 + \sigma_Y^2)^2}$$

$$= \frac{\sigma_Y^2 \sigma_X^2 (\sigma_X^2 + \sigma_Y^2)}{(\sigma_X^2 + \sigma_Y^2)^2}$$

$$= \frac{\sigma_Y^2 \sigma_X^2}{\sigma_X^2 + \sigma_Y^2}$$