TESTS OF MEAN STATIONARITY FOR AUSTRALIAN SHARE MARKET RETURNS DATA

Dr Martin Lally
Capital Financial Consultants Ltd

2 June 2022

CONTENTS

EXECUTIVE SUMMARY

The AER currently estimates the Market Risk Premium for Australia by applying primary weight to the Ibbotson approach (involving averaging of annual excess returns) with data from 1988. However, data is available back to 1883, and this raises the question of whether some or all of the earlier data should be used. Accordingly, this paper examines whether the population means for the excess returns are the same in all of these years, i.e., mean stationarity prevails. In addition, the nominal and real returns are also tested for mean stationarity. The conclusions are as follows.

Firstly, in testing for a time trend, the null hypothesis of no time trend cannot be rejected for each of the three return series. Secondly, upon splitting the data into subperiods, the null hypothesis of no difference in the population means across the subperiods cannot be rejected for each of the three return series. Thirdly, Augmented Dickey Fuller tests for a unit root are not useful for testing for the mean stationarity of returns, although they are useful for a wide range of other economic and financial time series. Fourthly, mean ergodicity is a very similar concept to mean stationarity, and the subtle distinctions between the two concepts do not seem to be relevant to returns; accordingly, separate tests do not seem warranted. Fifthly, the ability of the first two tests to detect non-stationarity is mitigated by the fact that changes in population means induce simultaneous changes in realised outcomes in the opposite direction. Subject to this caveat, this analysis supports the conclusion that returns are mean stationary.

1. Introduction

The AER currently estimates the MRP for Australia by applying primary weight to the Ibbotson approach (involving averaging of annual excess returns) with data from 1988. However, data is available back to 1883, and this raises the question of whether some or all of the earlier data should be used. Accordingly, this paper examines whether the population means for the excess returns are the same in all of these years, i.e., mean stationarity prevails. In addition, the nominal and real returns are also tested for mean stationarity.

2. Mean Stationarity Tests of Excess Returns

Tests for mean stationarity should reflect the possible types of departures from stationarity. One such possibility is a gradual drift downwards in the population mean for excess returns (as investors have become more diversified and the cost of forming a well-diversified portfolio has fallen). The natural test for this would involve_regressing excess returns on time.[1] The result is a coefficient on time of 0.005% and this is not statistically significant (*p* = 0.88). So, the hypothesis of no time trend can't be rejected. However, unlike most economic and financial time series, returns reflect not only events that have occurred in the period in question but revised expectations about the future. So, if the population mean excess return (the MRP) declines, the asset price simultaneously rises, thereby raising the excess return. Thus, as the mean population mean excess return falls over time, the realized excess returns tend to be drawn from above the population mean, and this latter effect reduces the downward drift in realized excess returns, thereby making it harder to detect the downward drift in the population mean excess return from the regression test.

A second possible source of non-stationarity in excess returns is that the population mean (the MRP) experiences occasional changes (regime shifts). The natural test for this is to partition the data into subsets and test for the statistical significance of the differences in sample means across the subsets. Wahab and Lashgari (1993, pp. 244-245) use two subsets in testing for stationarity in means for stock returns. Pagan and Schwert (1990, page 167), and Loretan and Phillips (1994, page 218), do likewise in testing for stationarity in variances for stock returns.

[1] The excess return for a year is the capital gain, dividend yield and the product of the imputation credit yield and the utilization rate for credits (with the latter value set at 0.65), with data supplied by the AER.

Generalising this, I split the Australian excess returns data into two, three, four, and five equal sized subsets (first and second half of the data; then first, second and third parts; etc), and the resulting sample means are shown in the second column of Table 1. For each of the four cases, the sample means are similar. The standard test for differences in the true means is the ANOVA test (Mood et al, 1974, pp. 435-438), involving a test statistic that has the *F* distribution if the null hypothesis (that the true means are equal) is true.[2] The results are shown in the third column of Table 1. In all four cases, the differences in the sample means are not statistically significant at even the 10% level. This is consistent with the hypothesis that the true mean has not changed over time.

Table 1: ANOVA Tests on Sample Mean Excess Returns

Partition	Sample Means (%)	Observed *F* Value	*P* Value
Halves	5.9, 6.9	0.15	> 0.10
Triples	6.4, 5.7, 7.1	0.09	> 0.10
Quarters	5.9, 6.0, 7.2, 6.5	0.05	> 0.10
Fifths	6.7, 6.0, 5.7, 6.3, 7.3	0.04	> 0.10
Regime Changes	6.7, 6.0, 6.5	0.02	> 0.10

These results are consistent with visual examination of the 30-year rolling average of excess returns, as calculated by the AER, i.e., there is very little variation in the 30-year average around the overall mean excess return of 6.4%.

An alternative approach for ANOVA tests would be to split the data at points considered or suspected to correspond to changes in the true mean (regime changes). The AER (2021, pp. 39-40) suggests a number of possibilities. In respect of 1937, this is because a broader stock index was used from that point. In respect of 1988, this is because dividend imputation was introduced then.[3] I therefore additionally split the data into 1883-1936, 1937-1987, and

[2] The test statistic is the sum of the squared differences between the subperiod means and the overall mean, divided by the sum of the estimated variances for the subperiods, with constants reflecting the sample sizes and the number of subperiods. So, if the true mean shifts over time, the numerator of this ratio will tend to increase, thereby increasing the chance of it exceeding the critical *F* value.

[3] The AER (ibid) also refers to 1958 and 1980 as possible regime changes, because the indexes were calculated from those points in real time rather than retrospectively. I do not consider that this would lead to a change in

1988-2021. The results are shown in the last row of Table 1, and the differences in the sample means are not statistically significant at even the 10% level. Again, this is consistent with the hypothesis that the true mean has not changed over time.

In so far as the regime shifts correspond to events that could change prices (such as the switch to dividend imputation in 1987 but not the use of a different index from 1937), these ANOVA tests for regime shifts are subject to the same problem as the test for time trend: at the moment the population mean excess return falls (rises), the realized excess return rises (falls), which makes it harder to detect the regime shifts by examining the excess returns.

A third possible source of non-stationarity in time series is that the process is autoregressive (outcomes are linearly related to past outcomes) with a unit root, and this can be tested for using tests such as the Augmented Dickey Fuller (ADF) test. The QTC (2022, page 13) conduct this test on real returns for nonstationarity and reject that hypothesis. The same test might be applied to excess returns. However it is inconceivable that excess returns would be autoregressive with a unit root. For example, letting ER_t denote the excess return in year t and e_t denote white noise in year t, an autoregressive process of order 1 (denoted AR(1)) with a unit root would require that the excess return in a year was equal to that in the preceding year plus white noise (e_t for year t):

$$ER_t = ER_{t-1} + e_t$$

This implies that the best predictor of the excess return in a year is the previous year's outcome. This would require an unprecedented degree of informational inefficiency in a market, involving expected excess returns (i.e., true MRPs) that would be negative whenever the preceding year's excess return was negative and true MRPs that were very large whenever excess returns in the previous year were very large. Such a situation would also be incompatible with any version of the CAPM, in which the true MRP is a reward for bearing risk.[4] Thus, conducting the ADF test on excess returns would seem to be pointless.

---

the true mean. The AER (ibid) also refers to 1958 because data on short-term government securities were available from that point, but such data is not used in determining excess returns here (these being based on ten-year bond yields).

[4] Regressing Australian excess returns for 1883-2021 on their counterpart in the previous year yields a coefficient on the previous year's excess return that is mildly negative (-0.17) rather than positive, let alone 1. One explanation for this is that the bad (good) news that induces unusually low (high) returns in a year causes the MRP to rise (fall) thereby typically yielding higher (lower) returns next year. An additional possibility is

Unsurprisingly, upon conducting it on real returns, the QTC (2022, page 13) reject the null hypothesis of a unit root. There are many processes in economics and finance for which a unit root might be present, including asset prices, but not nominal, real, and excess returns. To illustrate the point that prices (but not returns) could have a unit root, suppose that dividends on an asset arise annually (D_t at end year t) and follow an AR(1) process with a unit root:

$$D_t = D_{t-1} + e_t$$

Let P_0 denote the asset price at time 0 (now), just after D_0. With discount rate k, the price then will be

$$P_0 = \frac{E(D_1)}{1+k} + \frac{E(D_2)}{(1+k)^2} + \cdots = \frac{E(D)}{k} = \frac{D_0}{k}$$

If the discount rate k does not change, the price at time 1 will be

$$P_1 = \frac{D_1}{k} = \frac{D_0 + e_1}{k} = \frac{D_0}{k} + \frac{e_1}{k} = P_0 + \frac{e_1}{k}$$

and this is an AR(1) process with a unit root. By contrast, the rate of return for the first year is

$$R_1 = \frac{D_1 + P_1}{P_0} - 1 = \frac{D_0 + e_1 + P_0 + \frac{e_1}{k}}{P_0} - 1 = \frac{D_0}{P_0} + \frac{e_1}{P_0}\left(1 + \frac{1}{k}\right) = k + \frac{e_1}{P_0}\left(1 + \frac{1}{k}\right)$$

Since e_1 is white noise, and therefore mean zero, the rate of return R_1 will be mean stationary (with mean equal to the discount rate k) but its variance will change over time in accordance with the asset price. If the risk-free rate is also mean stationary, then excess returns will also be mean stationary.[5]

Consistent with these comments, ADF tests on various macroeconomic time series and asset prices are common (for example, see Nelson et al, 2005, section 1) but are only very rarely

that prices overreact to both good and bad news, leading to partial reversal of prices in the following year and therefore the negative correlation in returns.

[5] If the process for dividends had instead been $D_t = D_{t-1}(1 + e_t)$, then the process for prices would be $P_t = P_{t-1}(1 + e_t)$, which is not quite AR(1), and the rate of return process would be $R_t = k + e_t(1 + k)$, which is both independent over time and stationary in all respects.

conducted on returns and seem to always result in rejection of the null hypothesis of a unit root (for example, Aggarwal and Kyaw, 2005, section IV).

3. Mean Stationarity Tests of Real Returns

The trend test is repeated on real returns, yielding a coefficient on time of -0.008% and this is not statistically significant (p = 0.82). So, the hypothesis of no time trend can't be rejected.

The ANOVA tests are also repeated on real returns and the results are shown in Table 2. As with excess returns, the hypothesis that the true mean has not changed over time cannot be rejected even at the 10% level.

Table 2: ANOVA Tests on Sample Mean Real Returns

Partition	Sample Means (%)	Observed F Value	P Value
Halves	8.2, 9.0	0.11	> 0.10
Thirds	9.6, 6.6, 9.7	0.53	> 0.10
Quarters	8.7, 7.8, 8.8, 9.2	0.05	> 0.10
Fifths	10.5, 9.5, 5.4, 8.7, 9.1	0.37	> 0.10
Regime Changes	10.4, 6.3, 9.2	0.79	> 0.10

However, there is considerably more variation over time in the sub-period sample means for real returns than for excess returns. Focusing upon the split into fifths, for which variation over time would typically be greatest, the range from lowest to highest outcome for real returns is 5.1% (5.4% to 10.5%) whilst the corresponding figure for excess returns is only 1.6% (5.7% to 7.3%). This is consistent with the comparison of the 30-year rolling averages generated by the AER, which show real returns dipping noticeably in the middle. It is also consistent with the observed F values for real returns being generally higher than for excess returns (averaging 0.37 for real returns versus 0.07 for excess returns).

As with excess returns, ADF tests are not useful for testing for the mean stationarity of real returns.

4. Mean Stationarity Tests of Nominal Returns

The trend test is repeated on nominal returns, yielding a coefficient on time of 0.03% and this is not statistically significant (p = 0.82). So, the hypothesis of no time trend cannot be rejected.

The ANOVA tests are also repeated on nominal returns and the results are shown in Table 3. As with excess returns and real returns, the hypothesis that the true mean has not changed over time cannot be rejected even at the 10% level.

Table 3: ANOVA Tests on Sample Mean Nominal Returns

Partition	Sample Means (%)	Observed *F* Value	*P* Value
Halves	10.0, 13.7	2.4	> 0.10
Thirds	10.6, 10.3, 14.6	1.0	> 0.10
Quarters	9.6, 10.3, 15.1, 12.2	0.8	> 0.10
Fifths	10.2, 10.8, 9.8, 16.4, 11.8	0.7	> 0.10
Regime Changes	10.9, 12.5, 12.3	0.1	> 0.10

However, there is considerably more variation over time in the sub-period sample means for nominal returns than for real returns and excess returns. Focusing upon the split into fifths, for which variation over time would typically be greatest, the range from lowest to highest outcome for nominal returns is 6.6% (9.8% to 16.4%) whilst the corresponding figures for real returns and excess returns are 5.1% and 1.6% respectively. This is consistent with the comparison of the 30-year rolling averages generated by the AER, which shows nominal returns rising noticeably in the second half of the series and then falling. It is also consistent with the observed F values for nominal returns being generally higher than their counterparts for real and excess returns (and averaging 1.0 for nominal returns versus 0.37 for real returns and 0.07 for excess returns).

As with real and excess returns, ADF tests are not useful for testing for the mean stationarity of nominal returns.

5. Ergodicity

A closely related issue to that of mean stationarity is mean ergodicity. The latter means that the sample distribution will approach the population distribution as the sample size becomes very large, and therefore the sample mean will converge on the population mean. Ergodicity guarantees stationarity, but stationarity does not guarantee ergodicity. However, contrary cases are not characteristic of returns. For example, suppose the first observation in a process is drawn from a normal distribution with a mean of 10 and a variance of 8, and that observation is 5, and all subsequent values for the process are 5, i.e., the process remains stuck at the first outcome. The expected outcome for all points in the process is therefore 10, and therefore the process is mean stationary. However, the sample distribution will not converge on the population distribution and therefore the sample mean (of 5 in this case) will not converge on the population mean (of 10) as the sample size becomes very large. So, the process is not ergodic. This example has no relevance to the process generating returns. So, for returns, it is sufficient to test just for mean stationarity.

6. Conclusions

This paper has examined whether the population means for excess returns, nominal returns, and real returns in the Australian share market are the same in all of the years 1883-2021, i.e., mean stationarity prevails. The conclusions are as follows. Firstly, in testing for a time trend, the null hypothesis of no time trend cannot be rejected for each of the three return series. Secondly, upon splitting the data into subperiods, the null hypothesis of no difference in the population means across the subperiods cannot be rejected for each of the three return series. Thirdly, Augmented Dickey Fuller tests for a unit root are not useful for testing for the mean stationarity of returns, although they are useful for a wide range of other economic and financial time series. Fourthly, mean ergodicity is a very similar concept to mean stationarity, and the subtle distinctions between the two concepts do not seem to be relevant to returns; accordingly, separate tests do not seem warranted. Fifthly, the ability of the first two tests to detect non-stationarity is mitigated by the fact that changes in population means induce simultaneous changes in realised outcomes in the opposite direction. Subject to this caveat, this analysis supports the conclusion that returns are mean stationary.

References

AER, 2021. *Rate of Return*: *Overall Rate of Return, Equity and Debt Omnibus, Final Working Paper*, www.aer.gov.au.

Aggarwal, R., and Kyaw, N., 2005. "Equity Market Integration ion the NAFTA Region: Evidence from Unit Root and Cointegration Tests", *International Review of Financial Analysis*, vol. 14, pp. 393-406.

Loretan, M., and Phillips, P., 1994. "Testing the Covariance Stationarity of Heavy-Tailed Series", *Journal of Empirical Finance*, vol. 1, pp. 211-248.

Mood, A., Graybill, F., and Boes, D., 1974, *Introduction to the Theory of Statistics*, 3rd edition, McGraw-Hill, New York.

Nelson, C., Piger, J., and Zivot, E., 2000. "Markov Regime-Switching and Unit Root Tests", Federal Reserve System International Finance Discussion Paper 683, https://www.federalreserve.gov/Pubs/Ifdp/2000/683/ifdp683.pdf.

Pagan, A., and Schwert, G., 1990. "Testing for Covariance Stationarity in Stock Market Data", *Economic Letters*, vol. 33, pp. 165-170.

QTC, 2022, *Submission on the 2022 Rate of Return Instrument*, www.aer.gov.au.

Wahab, M., and Lashgari, M., 1993. "Covariance Stationarity of International Equity Market Returns: Recent Evidence", *Financial Review*, vol. 28 (2), pp. 239-260.