ELSEVIER

Journal of Empirical Finance 8 (2001) 427–449

Journal of EMPIRICAL FINANCE

www.elsevier.com/locate/econbase

Tests of asset-pricing models: how important is the iid-normal assumption?☆

Nicolaas Groenewold [a,*], Patricia Fraser [b]

[a] *Department of Economics, University of Western Australia, 35 Stirling Highway, Crawley, WA 6009, Australia*

[b] *Department of Accountancy, University of Aberdeen, Aberdeen, Scotland, UK*

Accepted 8 June 2001

Abstract

Financial data are typically not iid-normal. Yet standard tests of asset-pricing models are based on this assumption. We address the question of how sensitive the tests are to violations of iid-normality and use Australian data to compare the standard test results with those of the GMM-based *J* test and bootstrap-based tests. We find that, in contrast to US evidence, standard test results are robust to the iid-normality assumption. This is true for all tests and for all three asset-pricing models analysed.

JEL classification: G120; C120

Keywords: Iid-normal; Asset-pricing models; *J* test; Bootstrap; CAPM

1. Introduction

The mean-variance model is central to modern financial economics—it provides the foundation for the Capital Asset Pricing Model (CAPM) which continues to be the most widely used model of asset-pricing in practice. The mean-variance

☆ A first draft of this paper was completed while the second author was a visitor in the Department of Economics at the University of Western Australia. We are grateful for the Departmental Research Grant which made this visit possible. The paper has benefited greatly from the comments by two anonymous referees.

* Corresponding author. Tel.: +61-8-9380-3345; fax: +61-8-9380-1016.

E-mail address: ngroenew@ecel.uwa.edu.au (N. Groenewold).

0927-5398/01/$ - see front matter
PII: S0927-5398(01)00035-4

model and the CAPM are firmly based on the assumption that asset returns are identically, independently and normally distributed (iid-normal). The models are usually derived from these assumptions and by far the majority of tests of the models require the iid-normal assumption for their (asymptotic) validity.

Yet it has become a stylised fact of empirical finance that the iid-normal assumption is systematically violated by asset returns in practice—returns typically exhibit skewness, excess kurtosis, (conditional) heteroskedasticity and temporal dependence.[1]

A pressing question in applied research in asset-pricing is, therefore, the sensitivity of the standard tests of asset-pricing models to the violation of the iid-normal assumption and, in particular, the extent to which the typically ambivalent test outcomes are the results of inappropriate tests. There is little evidence of whether the type of test used matters for the test outcomes although such evidence as there is suggests that the test results may not be robust to the iid-normal assumption.[2]

It is important to know whether test sensitivity is confined to isolated incidents or whether it is a general phenomenon applying to a range of countries, time periods and models. In this paper we contribute to the very limited literature on the subject and provide further evidence on this question by using Australian share-market data to compare the results obtained from standard tests to those obtained from tests which do not depend on the iid-normal assumption. The overall aim of the study is to contribute to the exploration of the question of when deviations from the iid-normal assumption matter and when they do not.

To achieve our aim we begin by testing the CAPM (or the mean-variance efficiency of our chosen market portfolio) using standard Wald and F tests and compare the outcome of these to the results of the J test associated with the generalised methods of moments (GMM) estimator, which is robust to a wide range of departures from the iid-normal assumption. A second alternative to the standard procedures which we use is to compare the standard W and F statistics to critical values based on the actual (non-iid-normal) characteristics of the data using the bootstrapping method. We go on to extend the literature by applying the tests to two additional asset-pricing models, both of which involve the use of macroeconomic instrumental variables. Such an investigation allows us to assess whether or not our conclusions are model-specific and hence the extent to which test outcomes are likely to be influenced by issues concerning the testing of joint hypotheses.

[1] See, e.g. Richardson and Smith (1993) for US evidence, Mills and Coutts (1996) for the UK and results presented in this paper for Australia.

[2] See Affleck-Graves and McDonald (1989) who present Monte Carlo evidence, MacKinlay and Richardson (1991) who use the J test associated with the GMM estimator applied to US data as do Chou and Zhou (1997) who also use standard tests with bootstrapped critical values also using US data. Faff and Lau (1997) apply the MacKinlay and Richardson analysis to Australian data.

Consistent with existing evidence, we find widespread departures from the iid-normal property in both returns and in market-model errors. However, in contrast to the US and Australian evidence cited above, we find no instance of sensitivity of the results to the test used. Thus for all the results—for all three models and for all three tests—we find consistent outcomes. In evaluating the results we first point out differences in the properties of our market portfolio compared with the US data used in earlier studies. We go on to argue that even, given this difference, the contrast between our results and those reported in the existing literature is more apparent than real since, on the whole, we find differences in prob values between standard and alternative tests similar in magnitude to those found previously. The different test outcomes are the result of the fact that in the US the CAPM restrictions are usually close to being rejected so that small changes in the prob values may change the outcome of the test. For our Australian data set, however, we find that the CAPM is very far from being rejected in almost all cases so that prob values can change very substantially as a result of using a different test without changing the outcome of the test at conventional significance levels. We arrive at the conclusion that the use of tests which take into account departures from the iid-normal assumption do indeed affect prob values. However, changes in prob values are likely to affect the test outcome only if the standard test produces a test statistic that is close to the critical value.

The structure of the remainder of the paper is as follows. We set out the three asset-pricing models to be tested in the next section. Section 3 presents the tests used in the paper. The data are described and tests of the iid-normal assumption are reported in Section 4. The following three sections discuss the results: first for the unconditional CAPM, then the conditional CAPM and finally the macro-factor version of the APT. Conclusions are drawn in the final section.

2. The asset-pricing models

We begin with the standard (unconditional) CAPM. Assume that there exists a risk-free asset with return R_f and N risky assets with returns R_i $(i = 1, 2, \ldots, N)$. Denote the return to the market portfolio of risky assets by R_m. Then the model states that

$$E(R_i) = R_f + \left[E(R_m) - R_f\right]\beta_i, \quad i = 1,2,\ldots,N \tag{1}$$

where $E(\cdot)$ is the unconditional expectations operator and $\beta_i = \mathrm{cov}(R_m, R_i)/\mathrm{var}(R_m)$. Given that R_f is non-stochastic, the model can be written in terms of excess returns,

$$E(r_i) = E(r_m)\beta_i, \quad i = 1,2,\ldots,N \tag{2}$$

where $r_i \equiv R_i - R_f$ and $r_m \equiv R_m - R_f$. Tests of the above model can be interpreted as tests of CAPM only if data on the return to the market portfolio are available. If this is not the case and a proxy such as the return to a stock market index is used, the test is generally interpreted as the test of the mean-variance efficiency (MVE) of the market portfolio. We continue to refer to the test as one of the CAPM but keep this qualification in mind.

An alternative reaction to the unobservability of the return to the market portfolio is to treat this as a latent variable and, following Gibbons and Ferson (1985) and Campbell (1987), relate it linearly to a set of observable variables, say, $z_1, \ldots, z_k$. In that case

$$r_m = \lambda_0 + \lambda_1 z_1 + \ldots + \lambda_k z_k + \eta \tag{3}$$

where η is a random error term. Substituting this into Eq. (2), we define the conditional CAPM as:

$$E(r_i) = \beta_i(\lambda_0 + \lambda_1 z_1 + \ldots + \lambda_k z_k), \quad i = 1,2,\ldots,N \tag{4}$$

where the model restricts the way in which the observable variables $z_1, \ldots, z_k$ enter the asset-pricing equations.

Eq. (4) is not unlike a multi-factor asset-pricing equation such as might be obtained from the Arbitrage Pricing Theory (APT) with macroeconomic or aggregate factors of the type first put forward by Chen et al. (1986). It has been applied to US data by Chen, Roll and Ross and to UK data by Beenstock and Chan (1988), Clare et al. (1993) and Clare and Thomas (1994). If we use the z variables as the macroeconomic factors in an APT model, the restrictions implied by the APT may be imposed on the equation for the ith asset as:[3]

$$E(r_i) = \gamma_{i0} + \gamma_{i1} z_1 + \gamma_{i2} z_2 + \ldots + \gamma_{ik} z_k, \quad i = 1,2,\ldots,N \tag{5}$$

where $\gamma_{i0} = (\pi_1 \gamma_{i1} + \pi_2 \gamma_{i2} + \ldots + \pi_k \gamma_{ik})$ for some constants, π_j.

3. Estimation and testing

Consider the standard (unconditional) CAPM first. Written in terms of excess returns, Eq. (2) suggests a test based on the market model also in terms of excess returns:

$$r_{it} = \alpha_i + \beta_i r_{mt} + \varepsilon_{it}, \quad i = 1,2,\ldots,N \text{ and } t = 1,2,\ldots,T \tag{6}$$

where the CAPM imposes the restriction that $\alpha_i = 0$ for all i. If we make the standard assumption that the ε_{it}s are iid-normal, we can test $H_0: \alpha_i = 0$ using a

[3] See McElroy et al. (1985) for a derivation. Applications using US data are by McElroy and Burmeister (1988), to Singapore by Ariff and Johnson (1990) and to Australian data by Groenewold and Fraser (1997).

Wald test based on the statistic

$$W = \mathbf{a}'\left[\text{var}(\mathbf{a})\right]^{-1}\mathbf{a} \tag{7}$$

where $\boldsymbol{a}$ is the vector of estimated α_is which has covariance matrix var($\boldsymbol{a}$). If var($\boldsymbol{a}$) is replaced by a consistent estimator, W is asymptotically χ^2-distributed with N degrees of freedom under H_0. An alternative to the Wald test is the GRS test due to Gibbons et al. (1989) who demonstrated that an adjusted version of W has an exact F distribution under H_0 and the iid-normal assumption:

$$F = \left[(T - N - 1)/TN\right]W \sim F_{(N,T-N-1)} \tag{8}$$

Since this test does not rely on large-sample properties, a comparison of W and F test outcomes will give us an indication of the effect of applying an asymptotic test to a finite sample. This is particularly important in tests of asset-pricing models that are typically tested over relatively short periods in order to minimise the effect of parameter changes over time.

Both the above tests rely on the iid assumption, and may therefore not be robust to heteroskedasticity, some evidence for which was found in our analysis of the data and the market-model residuals in the previous section. An increasingly popular alternative is to use Hansen's (1982) J test associated with the GMM estimator. MacKinlay and Richardson (1991), using US data, found that test outcomes may change if the J statistic is used instead of standard W and F statistics. Similar findings are reported for the US by Chou and Zhou (1997) and for Australia by Faff and Lau (1997).

If we denote by $\boldsymbol{z}_t$ the vector $(1, r_{mt})'$ and by $\boldsymbol{\varepsilon}_t$ the vector of errors for period t, the GMM estimator chooses the parameters to minimise the following quadratic criterion function:

$$\left[1/T\Sigma\boldsymbol{\varepsilon}_t \otimes \mathbf{z}_t\right]'\mathbf{W}^{-1}\left[1/T\Sigma\boldsymbol{\varepsilon}_t \otimes \mathbf{z}_t\right] \tag{9}$$

where $\mathbf{W}$ is a weighting matrix which in practice takes various forms. In the case where there are more moment conditions than parameters, the resulting over-identifying restrictions can be tested using the optimised value of the criterion function. Hansen shows that under quite weak conditions the resulting J statistic is asymptotically distributed χ^2 with N degrees of freedom. This therefore provides a method of testing the asset-pricing model in the face of non-iid-normal errors and a comparison of the W and J tests will provide us with an indication of the effect of the violation of the underlying iid assumption of asymptotic tests.

An alternative to the use of the GMM-J test is to use the bootstrapping procedure to derive critical values for standard test statistics. Bootstrapping proceeds by drawing repeated samples with replacement from the residuals of the model. These residuals are then used to generate new "data" which are used to re-estimate the model and re-compute the test statistic. With a large number of repeated samples it is possible to build up an empirical distribution of the test

statistic which is used for inference. Since the bootstrapping method is based on sampling from the actual residuals, it overcomes both the small-sample problem and the non-normality of the errors.

There are several ways of proceeding to bootstrap tests of asset-pricing models. They differ both as to resampling scheme used and statistic to be sampled. Little is known about the theoretical properties of the various alternatives and Monte Carlo evidence is only beginning to emerge. Based on the discussion in Efron and Tibshirani (1993) and Li and Maddala (1996), we begin by using a standard bootstrapping method based on the following scheme:

1. Estimate the market model, Eq. (6), using OLS, reserve the residuals and compute the Wald and F statistics
2. Draw a sample with replacement from the residuals reserved in step (1) and generate a new set of return series under the null hypothesis (i.e. assuming that $\alpha_i = 0$ for all i and with the β_is set at the values estimated in step 1).
3. Estimate the market model based on the generated returns from step (2) and calculate the Wald and F statistics.
4. Repeat steps (2) and (3) 1000 times.
5. Build up an empirical distribution of the Wald and F statistics from the results in (4) and calculate critical values to which the values of the test statistics from step (1) are compared.

Some experimentation with more than 1000 replications was carried out but the results were not found to be sensitive to the number of replications in the 1000–10,000 range.

The bootstrapping procedure just outlined is based on the implicit assumption that the model errors which are sampled are independent. If this is not the case, then the basic bootstrap method is inappropriate since it fails to preserve the interdependence present in the data. A common alternative is the block bootstrap in which the residuals in step 1 above are divided into B sets of L consecutive observations or blocks of residuals ($BL = T$) and a sample of B blocks is drawn with replacement. The sampled blocks are then combined to create a new series of T observations and the remainder of the steps are undertaken as before. Various more complicated alternatives have been suggested; for a recent discussion see Berkowitz and Kilian (1996) and Davison and Hinkley (1997, Chap. 8). Clearly there is a trade-off in choosing block length, L, between blocks long enough to preserve the dependence in the data and having a sufficient number of blocks from which to sample so as to generate sufficient variability in the generated series. There is little guidance in the literature in the choice of block size and we experiment with various alternatives.

So far we have dealt only with the unconditional CAPM but similar W, F and J tests and bootstrapping procedures may be applied to the conditional CAPM and

the APT. Consider first the conditional CAPM. For this model Eqs. (2) and (3) imply the modified market model:[4]

$$r_{it} = \alpha_i + \beta_i(\lambda_0 + \lambda_1 z_{1t} + \ldots + \lambda_k z_{kt}) + \varepsilon_{it} \quad (10)$$

Not all of the parameters of this model can be identified and we follow Campbell (1987) and set $\alpha_i = 0$ for all i and $\beta_1 = 1$. Then the restrictions can be tested in the framework:

$$r_{it} = \gamma_{i0} + \gamma_{i1} z_{1t} + \ldots + \gamma_{ik} z_{kt} + \varepsilon'_{it} \quad (11)$$

by testing H_0: $\gamma_{i0}/\gamma_{10} = \ldots = \gamma_{ik}/\gamma_{1k}$ for $i = 2,3,\ldots, N$. We again use three tests: the Wald, GRS and GMM-J tests. It should be noted that the GRS results were not derived in the framework of a model such as the conditional CAPM and therefore the distributional result for the GRS F statistic is not strictly valid. We apply it nevertheless and interpret the GRS statistic as a general small sample correction of the W statistic.

In the case of the macro-factor version of the APT, the unrestricted model is again given by Eq. (11) but the restrictions on the coefficients are $\gamma_{i0} = (\pi_1\gamma_{i1} + \pi_2\gamma_{i2} + \ldots + \pi_k\gamma_{ik})$ for all $i = 1, 2,\ldots, N$ and for some constants, $\pi_1, \pi_2,\ldots, \pi_k$ implying $N–k$ restrictions. The GRS test will again be interpreted as a small-sample correction of the W test.

4. The data

Following recent work for the US and the UK (see for example Engel et al. (1995), for the US and Clare et al. (1997), for the UK), we use sectoral share-price index data. Returns are calculated for five sectors. All data are monthly and for the period February 1973 to June 1996.

The equity prices consist of value-weighted market and sector price indices taken from the Australian section of the Datastream Global Indices series. Using Financial Times Actuary classifications, the market is broken down into its constituent industrial sectors. The sectors and their components are

- Mineral Extraction (Gold Mining, Other Mining, Oil Integrated, Oil Exploration and Production)
- General Industries (Building and Construction, Building Materials and Merchandising, Chemicals, Diversified Industries, Electronic and Electrical, Engineering, Paper and Print)
- Consumer Goods (Brewers, Spirit, Wines, CID, Food Producers, Health Care, Pharmaceuticals, Tobacco)

[4] Note that r_{it} refers to the return obtained from holding and asset from period t to $t+1$ while z_t refers to period t so that the z_ts may be considered known when the expectation of r_{mt} is being formed and so they may be thought of as lagged instruments.

- Services (Distributors, Leisure and Hotels, Media, Retailers, Food, Retail, General Support Services, Transport, Other Businesses)
- Financial (Banks, Retail, Insurance, Other Financial, Property)

The return for the market portfolio was derived from a market share price index calculated as a value-weighted average of the five sectoral indexes. The return for each sector as well as the market was calculated on a continuous-compounding basis and included both the capital gain and the dividend yield.

The aggregate macroeconomic variables experimented with following work on testing the macro-factor version of the APT reported in Groenewold and Fraser (1997) consisted of the following variables: 90-day Bank-Accepted Bill Rate, Trade Weighted Exchange Rate Index, US$/A$ Exchange Rate, Yen/A$ Exchange Rate, Current Account Balance, Manufacturing Price Index, Unemployment Rate, Employment, M3, M1. Data for all these variables were obtained from the Reserve Bank of Australia Monthly Bulletin section of the dX EconData database. In addition to the above variables, we also used a measure of the return on the US stock market obtained from Datastream and calculated in a manner similar to the way in which the return to the Australian market portfolio was calculated. Finally, excess returns were calculated by subtracting the 90-day Bill Rate as a measure of the Australian risk-free rate.

Summary statistics for the portfolios are reported in Table 1 and Table 2 reports diagnostic statistics for the excess returns in Panel A and for the residuals from the excess-return form of the market model in Panel B.[5] The latter are included because the test of the CAPM are generally dependent on the distribution of the errors rather than the returns themselves. The estimated autocorrelation coefficients and the $Q(6)$ statistics indicate little evidence of autocorrelation in excess returns or the residuals from the market model. None of the Q statistics for the errors is significant at the 5% level (or at the 10% level). Of the individual autocorrelations only 7% are significant at the 5% level, which is close enough to the significance level to be consistent with the absence of autocorrelation in the model errors.

The results of the test for heteroskedasticity are mixed although the problem seems to be more severe in the residuals than in the excess returns themselves. Thus, while there is no evidence of intertemporal dependence of the AR type, there is evidence that in some sectors the errors are dependent in their second moments. We return to this matter later in the paper when considering the appropriate bootstrapping method.

The remaining statistics are relevant to the normality of the variables. In the case of the excess returns, there is clear evidence of skewness and excess kurtosis

[5] The diagnostics reported in Table 2 were also computed for returns and residuals from the market models estimated with returns rather than excess returns. The results were almost identical to those reported in Table 2.

Table 1
Sector returns: summary statistics

Sector	Average market value	Percentage of total	Mean excess return	Standard deviation
1. Mineral Extraction	21,893	24.25	−0.0021	0.0875
2. General Industries	27,983	31.25	0.0014	0.0689
3. Consumption Goods	6,908	7.72	0.0022	0.0590
4. Services	13,564	15.15	0.0038	0.0730
5. Financial	19,194	21.44	−0.0001	0.0728
Total (Market)	89,544	100.00	0.0004	0.0680

for all sectors and, not surprisingly, the Jarque–Bera test for normality has very large values of the test statistic relative to the 5% critical values reported in the notes to the table. The alternative normality test, the goodness-of-fit test, has a significant test statistics for four of the six cases. There is thus strong evidence of non-normality in the excess returns. The same is true of the residuals from the market model although there is less evidence of skewness and the goodness-of-fit

Table 2
Excess returns and market model residuals: diagnostics

Sector	ρ_1	ρ_2	ρ_3	ρ_4	ρ_5	ρ_6	Q (6)	ARCH (6)	Sk	Ku	JB	GF
Panel A: excess returns												
1	0.05	−0.06	−0.02	0.12	0.03	−0.13	11.32	12.425	−9.2455	32.264	1128.17	31.659
2	0.07	−0.05	0.00	0.04	−0.02	−0.06	3.71	4.1455	−5.4553	14.852	235.202	16.957
3	0.03	0.03	0.00	0.04	−0.08	0.03	2.87	1.0068	−14.067	50.901	2643.20	32.167
4	0.09	−0.10	−0.03	0.02	−0.06	0.02	7.11	0.5993	−12.101	43.749	1951.91	41.428
5	0.04	−0.07	−0.01	0.01	−0.02	−0.08	3.79	53.636	−9.2482	30.142	940.754	42.268
Total	0.07	−0.08	−0.04	0.08	0.01	−0.06	6.43	4.0121	−2.6940	30.512	965.486	26.322
Panel B: market model residuals												
1	0.05	0.01	−0.01	0.00	−0.13	−0.08	7.58	11.263	−1.7221	6.9140	47.269	21.289
2	0.01	−0.03	−0.04	−0.01	0.11	−0.11	7.32	15.774	5.7737	12.374	176.08	20.860
3	0.09	0.04	0.02	0.11	0.05	0.08	8.58	5.1424	−1.4326	3.9965	16.592	18.592
4	0.07	−0.17	0.03	−0.01	−0.01	0.04	10.16	21.925	−1.1816	2.5212	7.0424	13.358
5	0.06	0.01	0.03	0.01	−0.04	−0.06	2.86	1.4836	−15.666	71.262	5045.2	28.822

The ρ_i are autocorrelation coefficients and have a standard error of 0.06.
$Q(6)$ is the Box–Pierce–Ljung statistic for first- to sixth-order autocorrelation; it has a $\chi^2_{(6)}$ distribution that has a 5% critical value of 12.5916.
ARCH(6) is Engle's test for sixth-order ARCH; it has a $\chi^2_{(6)}$ distribution with a 5% critical value of 12.5916.
Sk and Ku are tests for skewness and excess-kurtosis and are both distributed $N(0,1)$.
JB is the Jarque–Bera test for normality and is $\chi^2_{(2)}$-distributed with a 5% critical value of 5.9915.
GF is the goodness-of-fit test for normality based on 17 partitions; it is $\chi^2_{(17)}$-distributed with a 5% critical value of 27.5871.

test provides little evidence of non-normality in contrast to the Jarque–Bera statistic which is again large, on the whole, compared to its 5% critical value.

We may conclude, therefore, that neither the excess returns nor the residuals from the market model satisfy the iid-normal assumptions. In particular, there is strong evidence of non-normality across the board, and evidence for two sectors of temporal dependence in the second moments of the model errors, although there is no autocorrelation in the errors themselves.

5. Results: unconditional CAPM

5.1. Standard test results

We begin by discussing tests of the traditional (unconditional) CAPM. The results are reported in Table 3.

The CAPM betas are clearly all precisely estimated and the model has considerable explanatory power as measured by the R^2s. All intercept coefficients are individually insignificant as predicted by the excess-return form of the CAPM.

The more formal tests of the CAPM restrictions are reported in Panel B of the table. The first prob value for W is obtained from the χ^2 tables and indicates that the null cannot be rejected. A comparison to the prob value for the F test indicates that there is little effect on the prob value and no effect on test outcome of making the small-sample correction incorporated in the GRS F test. Clearly, the application of an asymptotic test to a finite sample is not a problem in this case.

Table 3
Tests of unconditional CAPM

Panel A: the market model: $r_{it} = \alpha_i + \beta_i r_{mt} + \varepsilon_{it}$, $i = 1, 2, \ldots, 5$, $t = 1, 2, \ldots, 282$

Industry	α_I	β_I	R^2
Mineral Extraction	−0.0023 (1.09)	1.1850 (38.83)	0.8434
General Industries	0.0009 (0.81)	0.9752 (60.06)	0.9280
Consumption Goods	0.0017 (0.80)	0.6904 (22.00)	0.6340
Services	0.0032 (1.46)	0.9270 (28.74)	0.7468
Financial	−0.0005 (0.18)	0.8496 (21.84)	0.6301

Panel B: tests of CAPM

Test	Test statistic	Prob	Bootstrapped prob
W	3.1470	0.677	0.666
F	0.6160	0.688	0.666
J	3.392	0.639	–
τ	35.229	0.000	–

Absolute value of the t-statistic in parentheses.

These results clearly support the CAPM restrictions. This finding is broadly consistent with those reported by Faff (1991) when he uses a value-weighted market portfolio (as we do) and by Wood (1991) when he uses industry portfolios and continuously compounded rates of return (as we do). They are at variance, however, with the results reported by Faff and Lau (1997) for their industry portfolios/value-weighted market portfolio cases (which correspond most closely to ours).[6,7]

5.2. The GMM-J test

The third row of panel B reports the value for the J test associated with the GMM estimates of the model. The J statistic is asymptotically distributed $\chi^2_{(5)}$ under the null and the value for J reported indicates that the null cannot be rejected. The prob value is similar to that obtained from the previous two tests. In particular, a comparison of the J and W test outcomes provide strong evidence that the departures from the iid-normal assumption which are accommodated by the J test are not important for the outcome of the test. The outcomes of the GMM-J test are consistent with those reported by Faff and Lau (1997) for their RGMM test on the data set with a value-weighted market index and industry portfolios. It is interesting to note that they use 24 industries compared to our five, so that the similarity of our results provides some evidence that the number of sectors is not crucial to the test outcome.

In the last line of panel B we report the value for the τ test based on Hoffman and Pagan (1989) and Ghysels and Hall (1990), which tests for the structural stability of the model under GMM estimation. It splits the sample in two parts and is based on the evaluation of the orthogonality conditions in the second part of the sample based on parameter estimates from the first part of the sample. We split the sample arbitrarily at its mid-point, September 1984. The test statistic is χ^2 distributed with degrees of freedom equal to the number of orthogonality conditions, 10 in this case. The stability hypothesis is clearly rejected.

5.3. The bootstrapped tests

The second prob value for the W statistic is obtained from the bootstrapping procedure. Given the relatively weak evidence of intertemporal dependence in the

[6] Unfortunately, Faff and Lau do not make anything of nor do they attempt to reconcile the differences between their results and those reported in the earlier paper by Faff (1991).

[7] Note that Faff and Lau (1997) also report marked differences in their results depending on whether they test the "zero-beta" form or the "Sharpe–Lintner" form of the CAPM. We do not find similar differences. When we use a test of the Gibbons (1982) type (Faff and Lau's "zero-beta form") we obtain a prob value for a Wald test of 0.7429 and for the likelihood ratio test of 0.6757 compared to our original prob value of 0.6773.

model errors, we initially use a standard bootstrapping procedure that samples individual errors. We explore later in the paper the effect on our results of taking account of this dependence. The bootstrapped prob value is very similar to the asymptotic prob value indicating that the combination of small-sample problems and non-iid properties of the data do not affect the Wald test in this sample. This is a surprising result given that substantial departures from the iid-normal assumption are evident in the data and in the residuals from the market model as reported in Table 2. The prob value for the F statistic is identical to that for the Wald statistic as one would expect—Eq. (8) makes clear that F is simply a multiple of W so that the ranking of the W statistics generated by the 1000 replications will be the same as that of the generated F statistics and hence the probs identical.

5.4. Sub-sample results

It is common practice in tests of asset-pricing models to test the models over relatively short periods (often 5 years) in view of the evidence of parameter instability in the market model, which is consistent with our earlier finding of a lack of stability in the model. We proceed, therefore, to test the model over 5-year sub-periods. The results are reported in Table 4.

The results show the expected variation across periods. In only one period, however, are the CAPM restrictions rejected at the 5% level of significance, viz.

Table 4
Sub-sample tests of unconditional CAPM

Test	Test statistic	Prob	Bootstrapped prob
1973–1996			
W	3.1470	0.677	0.666
F	0.6160	0.688	0.666
1973–1977			
W	3.0669	0.689	0.689
F	0.6003	0.699	0.689
1978–1982			
W	2.4392	0.786	0.795
F	0.4878	0.793	0.795
1983–1987			
W	12.9967	0.023	0.027
F	2.5440	0.029	0.027
1988–1992			
W	5.8154	0.325	0.362
F	1.1383	0.3403	0.362

the period 1983–1987, which includes the October 1987 Crash in addition to including a period of substantial de-regulation in the Australian financial system. In no period, however, does the use of bootstrapped critical values change the outcome of the test, irrespective of whether the W or the F test is used. It is interesting in particular to note that the comparison between W and F test results is largely unaffected by the smaller samples used in the sub-periods. Even in samples as small as 60, the exact F test produces very similar outcomes to those of the asymptotic W test.

5.5. Bootstrapping dependent residuals

We return now to the matter of the bootstrapping procedure. The bootstrapped results reported in Tables 3 and 4 were all derived from a bootstrap based on the assumption that the model errors are independent. While we found no evidence of autocorrelation, there was evidence of ARCH in two of our five sectors. Hence, it is appropriate to assess the sensitivity of our results to the independence assumption. We do this initially by re-computing the bootstrapped prob values for the full sample using a block bootstrapping procedure.

We begin by considering block length. Since there is little guidance in the literature on the choice of block length, we experimented with blocks of 3, 6, 12 and 24 observations (adjusting the length of the sample to 276 for the first three and 264 for the last of these). For each block length we sampled blocks, created 1000 new time-series and examined the AR and ARCH characteristics of these artificial data by computing the means of the AR and ARCH statistics over the 1000 replications. The results are reported in Table 5. The Wald statistics and prob values for the various block lengths are reported in Table 6.

Table 5
The effects of block length on AR and ARCH

Statistic	Block length					
	Data	1	3	6	12	24
$Q(6)$ Sector 1	7.1133	6.0147	7.0015	9.4789	9.2163	6.8561
$Q(6)$ Sector 2	7.4089	6.1208	7.0689	9.0315	15.8203	15.0157
$Q(6)$ Sector 3	7.9414	6.0233	6.2952	8.2669	10.7652	12.9717
$Q(6)$ Sector 4	9.6310	6.0823	10.1936	13.9962	19.3538	17.8233
$Q(6)$ Sector 5	2.7773	5.6491	7.6527	7.4196	9.8722	9.2851
ARCH(6) Sector 1	10.6149	5.3383	6.2762	12.6474	13.7131	11.8116
ARCH(6) Sector 2	14.0681	4.7833	6.6676	9.2516	18.1763	20.7351
ARCH(6) Sector 3	4.9456	5.4190	5.5934	5.2509	5.2449	4.4453
ARCH(6) Sector 4	20.9422	5.7075	8.6161	23.0526	24.3075	20.4776
ARCH(6) Sector 5	1.4466	3.1419	5.6844	13.2326	13.5646	12.8531

Both Q and ARCH statistics are distributed $\chi^2_{(6)}$ under the null hypothesis of no AR(1–6) and no ARCH(6) respectively. The 5% critical value is 12.5916.

Table 6
Block-bootstrapped Wald test statistics

Statistic/prob	Block length				
	1	3	6	12	24
Wald	4.2522	4.2522	4.2522	4.2522	4.4121
Prob	0.5137	0.5137	0.5137	0.5137	0.4917
Bootstrapped prob	0.5360	0.5390	0.5790	0.6560	0.6790

In Table 5 the first column of figures provides the AR and ARCH statistics for the data. These differ somewhat from those reported in Table 2 reflecting the slightly shorter sample period. There is no evidence of AR in any of the sectors and evidence of ARCH in sectors 2 and 4. The AR and ARCH statistics for the bootstrapped data do not reproduce the ARCH characteristics of the data at all well for short block lengths as expected. As block length increases, the ARCH statistics tend toward those calculated from the data so that it seems that the longer block length is preferable. However, the longer block length also introduces AR into the artificial data as is obvious from the $Q(6)$ statistics: for a block length of 24, three of the five sectors have significant AR, in contrast to the original data where none of the Q statistics are significant. Thus, the appropriate block length is not clear-cut.

Table 6 reports the Wald statistics and the bootstrapped prob values for the various block lengths. We see that the difference between the theoretical and bootstrapped probs increases with the block length. However, in all cases the model restrictions are not rejected and increasing the block length increases the bootstrapped prob value, thus making it unlikely that experimentation with longer block lengths would produce a reversal of the test outcome. Hence, we may conclude that our earlier finding that the test outcome is unaffected by the use of bootstrapped prob values still stands even when we take account of dependence in the data by using a block bootstrapping procedure. The result is not sensitive to choice of block size.

If we turn to the sub-sample results which we discussed previously and reported in Table 4, it will be recalled that for only one of the sub-sample is it at all likely that the use of an alternative bootstrapping procedure might result in a different test outcome, viz. the 1983–1987 period for which the prob values were between 2% and 3% using the independent bootstrapping procedure. We therefore re-computed the prob values for the Wald test for this period by using the block bootstrapping method. In this case a block length of six observations reproduces both AR and ARCH characteristics of the data quite closely so that a block length of six was the only one used. The Wald statistic for the sub-sample was 14.1463 with a theoretical prob value of 0.0147 and a bootstrapped prob value of 0.038. Again, the test outcome is not affected by the use of either standard or block-

bootstrapping. In all cases the model restrictions are clearly rejected at the 5% level of significance.

An alternative approach to the use of block-bootstrapping for taking account of temporal dependence in the residuals is to use a "post-blackening" procedure in which the temporal dependence is explicitly modelled and the sampling is done from the residuals after the dependence has been removed.[8] In the case of AR residuals the procedure is relatively straightforward: one estimates the model and estimates an appropriate AR process for the residuals either simultaneously or sequentially and one then samples from the residuals of the AR model (which will be independent given that the AR model has removed the autocorrelation), feeding them into the estimated AR model to produce artificial residuals for the original model with the appropriate AR characteristics.

In contrast to block-bootstrapping, this procedure requires one to identify the nature of the temporal dependence and to estimate a model to capture it. In addition, in the case of ARCH residuals, there are two practical problems. The first is that the model for the residuals generates only squared residuals (presuming that one captures the ARCH characteristics by regressing the squared residuals on their own lagged values) for the returns equation and it is quite possible that it will generate some negative values. The second problem is that even if one deals with the negative squared residuals, one must assign a sign to them after taking their square root and before using them in the returns equation to generate artificial returns data for the bootstrapping exercise.

We found the problem of negative generated squared residuals to be a significant one for ARCH processes of a higher order than ARCH(1) and therefore restricted our experimentation to ARCH(1) models. In this case the relatively few negative numbers were set equal to zero. We overcame the second problem by sampling from both the ARCH residuals and the original residuals. A "runs test" indicated that the signs of the original residuals were consistent with independence so that we used signs of the sampled residuals to sign the square-roots of the generated squared residuals obtained from the estimated ARCH model of the residuals. The prob value for the Wald statistic obtained from this procedure was very close to both the theoretical prob and the prob obtained from a standard bootstrapping method.

We conclude that the failure of the bootstrapping procedure to reverse the outcomes of any of the model tests was not because of ignored temporal dependence in the model residuals.

5.6. Comparison with previous results

On the basis of the tests of the CAPM (or mean-variance efficiency of the market portfolio) reported in this section, we may conclude that the outcome of the

[8] See Davison and Hinkley (1997), Section 8.2.3 for a brief description of this procedure.

tests is not at all sensitive to whether the test is a small-sample one based on iid-normal errors (the F test), or an asymptotic test based on iid-normal errors (the Wald test), or an asymptotic test based on a more general error process (the J test) or whether the test uses bootstrapped critical values. Thus, neither departures from iid-normal nor small samples appear to be a problem in testing this model on our particular data set.

This conclusion is very surprising in the light of the violations of the iid-normal assumption reported in Section 4 and in the light of contrary evidence for the US and Australia. MacKinlay and Richardson (1991) in their application of GMM to MVE tests reported that the test outcomes were sensitive to the nature of the test used, as do Faff and Lau (1997) in their application of the MacKinlay and Richardson analysis to Australian data. Chou and Zhou (1997), in an application of bootstrapping to tests of asset-pricing models, found that bootstrapping frequently produced different results.

However, the contrast between our evidence and existing results is not as strong as appears at first sight as will be clear from a more detailed comparison. Both MacKinlay and Richardson (1991) and Faff and Lau (1997) compare the Wald and GRS tests as we do and they also find that both produce almost identical prob values; similarly, in their comparison of the Wald and J tests, MacKinlay and Richardson (1991) find only small differences in prob values (except for one case) —often in the range of 0.02 to 0.04 and therefore similar to those reported in Table 3 above. For their data set, however, these differences in prob values produce some reversals of the test outcomes because their original prob values are close to conventional significance levels. It is interesting to note that their results produce no outcome reversals if a 10% significance level is used instead of a 5% level. Faff and Lau (1997), on the other hand, report some large differences between the prob values for Wald and F tests and those for the GMM-J test. It is interesting that their prob values for the GMM-J test are similar to ours but generally their Wald and F prob values are very much smaller.

A similar comparison can be made between our results and those reported by Chou and Zhou (1997) with broadly similar conclusions, although Chou and Zhou's (1997) results are more varied despite apparently using a data set which is identical to MacKinlay and Richardson's (1991) (except for a somewhat longer sample period).[9] They do, however, use a wider range of tests and consider shorter sub-samples than those of MacKinlay and Richardson (1991)—10 years compared to 30 years. Again, the Wald and GRS prob values are very similar and they find no outcome reversals at the 10% level but two reversals at the 5% level. The Chou

[9] Both use monthly data for 10 value-ranked portfolios obtained from the CRSP data tape for a period starting in 1926 and ending in 1988 for MacKinlay and Richardson (1991) and in 1995 for Chou and Zhou (1997). MacKinlay and Richardson (1991) experiment with both a value-weighted market return and an equally weighted alternative whereas Chou and Zhou (1997) use only a value-weighted one.

and Zhou (1997) comparison of the Wald and J tests produces greater prob-value differences than ours or MacKinlay and Richardson (1991) and the test outcome is affected in two of the seven sub-samples considered and also in the aggregate. Perhaps this can be accounted for by the shorter sub-samples used by Chou and Zhou (1997)—unfortunately they do not attempt to replicate the MacKinlay and Richardson (1991) results despite using what appears to be an identical data set. In their comparison of the Wald to the bootstrapped Wald results, prob-value differences are again very modest although somewhat larger than ours and they report one case where the test outcome is reversed under bootstrapping.

Hence, while our results are at variance with those reported for the US, on closer comparison the differences are not so marked. Nevertheless, differences between our results remain and there are various possible (not mutually exclusive) explanations.

The first possibility we consider is the characteristics of the market portfolio for which we test the MVE property. MacKinlay and Richardson (1991) show that if the model errors are multivariate-*t*-distributed rather than multivariate normal, as assumed by standard tests, the bias resulting from the application of the standard Wald test is positively related to the square of the Sharpe index, μ/σ, for the market return. They report the value of this index for their market portfolio for the full sample and a number of sub-samples and find that the difference between the values of the Wald and GMM statistics are greater, the greater is the estimated value of the index. Faff and Lau (1997) report similar evidence. The value of $(\mu/\sigma)^2$ for our full sample and sub-samples is uniformly smaller than the values reported by MacKinlay and Richardson[10], which would lead us to expect smaller differences between the prob values for standard statistics and those which account for departures from the iid-normal assumption.

A second possibility is that we have used different sample periods and sub-samples of different length. MacKinlay and Richardson (1991) use sub-samples of 30 years and Chou and Zhou (1997) of 10 years while we have 5-year sub-periods. Given that our sample period starts in 1973, we cannot match any of MacKinlay and Richardson's (1991) sub-periods but we were able to match two of Chou and Zhou's (1997) periods: 1976–1985 and 1986–1995. The results we obtained for these periods generally supported our earlier results: prob values for GRS, Wald, bootstrapped Wald and J tests are 0.4311, 0.3952, 0.4060 and 0.3575, respectively, for 1976–1985 and 0.9242, 0.9170, 0.9240 and 0.9167, respectively, for 1986–1995. Thus the length of our sub-periods does not seem to be able to account for our different results.

[10] The MacKinlay and Richardson (1991) values for $(\mu/\sigma)^2$ range from 0.013 to 0.21 while ours range from 0.0006 to 0.115. The values reported by Faff and Lau (1997) are also larger than ours on average.

A third possibility lies in the number of portfolios used and the method of their construction. Both MacKinlay and Richardson (1991) and Chou and Zhou (1997) use 10 value-ranked portfolios while our data are defined in terms of five industry-based portfolios and Faff and Lau use 24 industry-based portfolios and two alternative definitions of the market portfolio (one a value-weighted and the other an equally weighted portfolio). It is possible that the extent of aggregation in our five portfolios has a bearing on our results although the similarity of our GMM-*J* results to those of Faff and Lau when they use a value-weighted market portfolio similar to ours, but 24 industry portfolios suggests that the *number* of portfolios itself is not crucial.

On the method of portfolio *construction*, there is conflicting evidence. Recent work by Clare et al. (1997) provides evidence based on UK data that the method of portfolio construction matters for tests of asset-pricing models. Further, the earlier works by Faff (1991) and Wood (1991) which report standard tests of the CAPM suggests that the method of portfolio construction and the definition of the market portfolio (value-weighted or equally weighted) may both have an effect on the test outcome. On the other hand, MacKinlay and Richardson (1991) and Chou and Zhou (1997) use apparently identical data sets but the difference between their results are as large as the difference between ours and theirs, so that the method of portfolio construction is clearly not the only factor involved. Nevertheless, we conjecture that the influence of the method of portfolio construction will be worth further research, a project that will require a different data set to ours that consists of industry-based share-price indexes.

6. Results: conditional CAPM

Before the conditional CAPM model can be implemented and tested, we need to specify the aggregate factors denoted by the z variables in Eq. (11). The set of variables originally entertained has been set out in Section 4. They were chosen to include both domestic and open-economy variables, both real and nominal, and followed earlier work on the macro-factor version of the APT reported in Groenewold and Fraser (1997).[11] Not all variables could be included in the equation due to estimation limitations, so that the next step was to narrow down the range of variables by examining the correlation between a factor and the next period's market excess return as measured by the return to the stock market index. Secondly, we considered the correlation between the variables to eliminate variables that were potentially capturing similar sources of risk. This process left us

[11] While there are no existing Australian estimates or tests of the conditional CAPM, the paper by Faff and Brooks (1998) does introduce aggregate variables into the market model but it does so by substituting them for a time-varying beta rather than for the unobservable market return as we do.

with two factors—the 90-day bill rate and the US stock-market return. A third variable of potential importance with which we have carried out only limited experimentation is the US$ exchange rate. It is recognised that the subsequent tests of the model depend on this factor-search procedure. This is necessarily the case, however, since finance theory provides very little guidance to the choice of factors, so that in the literature generally ad hoc search procedures are used; e.g. Pesaran and Timmermann (1997) use a recursive modelling technique while Campbell (1993) bases his choice on preliminary analysis of a VAR.

The results of estimating and testing the model are reported in Table 7. The intercept term in each of the equations of the unrestricted model is insignificantly different from zero. Both regressors are consistently significant and the equations have a value of R^2 in the 5–10% range, indicating that the proxy variables are considerably less successful in explaining the variation in excess returns than the return to the stock market index is. It should be recalled, though, that the regressors in the present case are lagged while in the unconditional CAPM the return to the market portfolio is contemporaneous. The signs and magnitudes of the regressors are remarkably similar across equations, providing informal evidence in favour of the CAPM restrictions that the aggregate factors enter the equations in the same way.

More formal evidence in favour of the null hypothesis is provided in Panel B of the table. Both the Wald and F statistics are very small and do not allow one to reject the restrictions implied by the conditional CAPM. As was the case in tests of the unconditional CAPM, the use of bootstrapped critical values does not alter

Table 7
Tests of conditional CAPM

Panel A: the unrestricted model: $r_{it} = \gamma_{i0} + \gamma_{i1}\Delta \mathrm{BB90}_{t-1} + \gamma_{i2} R_t^{\mathrm{US}} + \varepsilon_{it}$, $i = 1, 2, \ldots, 5$, $t = 1, 2, \ldots, 282$

Industry	γ_{i0}	γ_{i1}	γ_{i2}	R^2	Cumby test
Mineral Extraction	−0.0050 (0.96)	−17.690 (3.07)	0.2862 (2.44)	0.0533	2.955
General Industries	−0.0014 (0.35)	−16.090 (3.59)	0.2703 (2.95)	0.0732	4.315
Consumption Goods	0.0002 (0.06)	−16.241 (4.25)	0.1920 (2.46)	0.0808	4.720
Services	0.0008 (0.18)	−21.208 (4.53)	0.2946 (3.08)	0.0988	4.974
Financial	−0.0027 (0.61)	−15.343 (3.21)	0.2489 (2.55)	0.0580	3.563

Panel B: tests of CAPM

Test	Test statistic	Prob	Bootstrapped prob
W	0.5028	0.999	0.981
F	0.0602	0.993	0.981
J	6.063	0.640	–
τ	16.837	0.329	–

this conclusion—the bootstrapped prob values are very close to the ones taken from the χ^2 and F tables, suggesting again that the non-normality observed in the data does not undermine traditional tests. The J test leads to the same conclusion as the W and F tests, i.e. that the null cannot be rejected. There is, however, an interesting contrast to the results reported above for the unconditional CAPM—there the J test prob values were very close to those obtained for the Wald test, while in the present case the use of the J test gives rise to a substantially different prob value, suggesting that the ability of the J test to accommodate a wider class of error behaviour is important in the case of the conditional CAPM even though the magnitude of the effect is not sufficient to reverse the test outcome. Moreover, the τ test for structural stability indicates that structural stability does not need to be rejected as it was in the previous case.

Finally, returning to panel A, we report the results of a further test for parameter stability due to Cumby (1990). Recall that the unconditional CAPM requires that the relative βs are constant. Cumby suggests a two-stage test, the first of which tests the constancy of the covariances and the second tests their relative constancy if the first test suggests time-varying covariances. In panel A the column headed "Cumby" shows the results of the first stage. The test statistic is $\chi^2_{(2)}$ distributed, which has a 5% critical value of 5.9915. Clearly, the value for the test statistic for each of the sectors falls below the critical value, so that we cannot reject the null hypothesis of covariance stability and do not proceed to the second stage of the test.

7. Results: the APT

We consider now the APT variant of the conditional model discussed in the previous section by interpreting the model as an APT model with macroeconomic factors along the lines first suggested by Chen et al. (1986). In this case the restrictions on the model of Eq. (11) can be written:

$$\gamma_{i0} = \pi_1 \gamma_{i1} + \pi_2 \gamma_{i2}, \quad i = 1,2,\ldots,5 \tag{12}$$

where π_1 and π_2 are free parameters, so that there are three restrictions.

The results of estimating this model and testing the restrictions are reported in Table 8. The estimated γ coefficients are similar to those of the unrestricted model reported in Table 7. Thus, it appears that the APT restrictions are not a serious constraint on the model. These results are consistent with our earlier tests of the APT model with macro factors reported in Groenewold and Fraser (1997), where the test statistic has a prob value in excess of 90%. This is supported by the insignificant estimates of the π parameters. Formal tests of the model are given in panel B of the table. Each of the three tests statistics has a very low value and correspondingly high prob value, indicating that the restrictions cannot be rejected. The prob values for the W and F tests are identical, while that for the J tests is

Table 8
Tests of APT

Panel A: the restricted model: $r_{it} = \pi_1\gamma_{i1} + \pi_2\gamma_{i2} + \gamma_{i1}\Delta BB90_{t-1} + \gamma_{i2}R_t^{US} + \varepsilon_{it}$, $i = 1, 2, \ldots, 5$, $t = 1, 2, \ldots, 282$; $\pi_1 = -0.0037$ (0.2684), $\pi_2 = -0.2814$ (0.2755)

Industry	γ_{i1}	γ_{i2}	R^2
Mineral Extraction	−18.3591 (3.38)	0.2620 (2.55)	0.0559
General Industries	−16.9369 (3.96)	0.2350 (2.83)	0.0745
Consumption Goods	−15.4500 (4.15)	0.2142 (2.89)	0.0830
Services	−21.2981 (4.59)	0.2841 (2.99)	0.1008
Financial	−15.6226 (3.38)	0.2271 (2.49)	0.0590

Panel B: tests of APT

Test	Test statistic	Prob	Bootstrapped prob
W	0.0057	1.000	0.986
F	0.0011	1.000	0.986
J	1.0771	0.7826	–
τ	135.273	0.0000	–

somewhat smaller, although not by as great a margin as in the conditional CAPM case. Again, the outcome of the test is not affected by these differences. As in previous cases, the bootstrapped prob values differ little from their theoretical counterparts and we did not experiment with block-bootstrapping. Finally, the test for structural stability of the model estimated by the GMM estimator provides strong evidence of absence of stability.

8. Conclusions

This paper set out to investigate the sensitivity of tests of asset-pricing models to the iid-normal assumption on which standard tests are based. We used two alternatives to these tests. The first was the J test based on the GMM estimator. The second alternative was to compute the critical values for the W and F tests using a bootstrapping technique.

Our conclusion can be very simply stated. We find differences between prob values based on the iid-normal assumption and those derived from the alternatives of a similar magnitude to those reported in the earlier literature and can conclude, therefore, that the iid-normal assumption does matter for tests of asset-pricing models. However, for the US the original prob values are generally close to conventional significance levels, so that small changes in prob values can easily reverse the test outcomes. In contrast, our original prob values are, almost without exception, large so that even substantial changes in prob values associated with

different tests have no effects on test outcomes. Hence, for our data set, the test used makes no difference to the outcome of the test. This was true for all comparisons, for all three asset-pricing models and for all but one of the sub-samples considered.

We explored several reasons for the differences in results and conclude, on the basis of earlier Australian work as well as UK evidence reported in Clare et al. (1997), that the method of portfolio construction seems to have an important bearing on the results, although it cannot be at the root of all the differences. We conjecture that it will, nevertheless, be a fruitful avenue for future research on the sensitivity of tests of asset-pricing models to the violation of the iid-normal assumption. Such research will, however, require a data set different to ours consisting as it does of industry-based share-price indexes.

References

Affleck-Graves, J., McDonald, B., 1989. Non-normalities and tests of asset pricing theories. Journal of Finance 44, 889–908.

Ariff, M., Johnson, L.W., 1990. Ex ante risk premia on macroeconomic factors in the pricing of stocks: an analysis using arbitrage pricing theory. In: Ariff, M., Johnson, L.W. (Eds.), Securities Markets and Stock Pricing. Longman, Singapore, pp. 194–204.

Beenstock, M., Chan, K., 1988. Economic forces in the London stock market. Oxford Bulletin of Economics and Statistics 50, 27–39.

Berkowitz, J., Kilian, L., 1996. Recent developments in bootstrapping time series. Finance and Economics Discussion Paper Series No. 1996-45, Division of Research and Statistics and Monetary Affairs. Federal Reserve Board, Washington, DC.

Campbell, J.Y., 1987. Stock returns and the term structure. Journal of Financial Economics 18, 373–399.

Campbell, J.Y., 1993. Intertemporal asset pricing without consumption data. American Economic Review 83, 487–512.

Chen, N., Roll, R., Ross, S.A., 1986. Economic forces and the stock market. Journal of Business 59, 383–403.

Chou, P.H., Zhou, G., 1997. Using bootstrap to test portfolio efficiency. Paper Presented to the Fourth Asia-Pacific Finance Association Conference, Kuala Lumpur.

Clare, A.D., Thomas, S.N., 1994. Macroeconomic factors, the APT and the UK stock market. Journal of Business Finance and Accounting 21, 309–330.

Clare, A., O'Brien, R., Thomas, S., Wickens, M., 1993. Macroeconomic shocks and the domestic CAPM: evidence from the UK stock sarket. Centre for Empirical Research in Finance, Brunel University, Discussion Paper 93-02.

Clare, A.D., Smith, P.N., Thomas, S.H., 1997. UK stock returns and robust tests of mean variance efficiency. Journal of Banking and Finance 21, 641–661.

Cumby, R.E., 1990. Consumption risk and international equity returns. Journal of International Money and Finance 9, 182–192.

Davison, A.C., Hinkley, D.V., 1997. Bootstrap Methods and Their Application. Cambridge Univ. Press, Cambridge.

Efron, B., Tibshirani, R., 1993. An Introduction to the Bootstrap. Chapman & Hall, New York.

Engel, C., Frenkel, J.A., Froot, K.A., Rodrigues, A.P., 1995. Tests of conditional mean-variance efficiency of the US stock market. Journal of Empirical Finance 2, 3–18.

Faff, R.W., 1991. A likelihood ratio test of the zero-beta CAPM in Australian equity returns. Accounting and Finance 31, 88–95.

Faff, R.W., Brooks, R.D., 1998. Time varying beta risk for Australian industry portfolios: an exploratory analysis. Journal of Business Finance and Accounting 25, 721–745.

Faff, R.W., Lau, S., 1997. A generalised methods of moments test of mean variance efficiency in the Australian stock market. Pacific Accounting Review 9, 2–16.

Ghysels, E., Hall, A., 1990. A test for structural stability of Euler condition parameters estimated via the generalized method of moments estimator. International Economic Review 31, 355–364.

Gibbons, M.R., 1982. Multivariate tests of financial models: a new approach. Journal of Financial Economics 10, 3–27.

Gibbons, M., Ferson, W., 1985. Tests of asset pricing models with changing expectations and an unobservable market portfolio. Journal of Financial Economics 14, 217–236.

Gibbons, M., Ross, S., Shanken, J., 1989. Testing the efficiency of a given portfolio. Econometrica 57, 1121–1152.

Groenewold, N., Fraser, P., 1997. Share prices and macroeconomic factors. Journal of Business Finance and Accounting 24, 1367–1383.

Hansen, L., 1982. Large sample properties of generalized methods of moments estimators. Econometrica 50, 1269–1286.

Hoffman, D., Pagan, A., 1989. Post-sample prediction tests for generalized methods of moments estimators. Oxford Bulletin of Economics and Statistics 51, 333–343.

Li, H., Maddala, G.S., 1996. Bootstrapping time series models. Econometric Reviews 15, 115–158.

MacKinlay, A.C., Richardson, M.P., 1991. Using generalized methods of moments to test mean-variance efficiency. Journal of Finance 46, 511–527.

McElroy, M.B., Burmeister, E., 1988. Arbitrage pricing theory as a restricted nonlinear multivariate regression model: iterated nonlinear seemingly unrelated regression estimates. Journal of Business and Economic Statistics 6, 29–42.

McElroy, M.B., Burmeister, E., Wall, K.D., 1985. Two estimators for the APT when factors are measured. Economics Letters 19, 271–275.

Mills, T.C., Coutts, J.A., 1996. Misspecification testing and robust estimation of the market model: estimating betas for the FT-SE industry baskets. European Finance Journal 2, 319–331.

Pesaran, M.H. and A. Timmermann, 1997, A Recursive Modelling Approach to Predicting UK Stock Returns, Cambridge University and University of Southern California.

Richardson, M.P., Smith, T., 1993. A test for multivariate normality in stock returns. Journal of Business 66, 295–321.

Wood, J., 1991. A cross-sectional regression tests of the mean variance efficiency of an Australian value weighted market portfolio. Accounting and Finance 31, 96–107.