

2003

Some evidence on the performance benchmarking of Australian fixed interest funds

David E. Allen
Edith Cowan University

Victor Soucik

Follow this and additional works at: <https://ro.ecu.edu.au/ecuworks>



Part of the [Finance Commons](#)

This is an Author's Accepted Manuscript of: Allen, D. E., & Soucik, V. (2003). Some evidence on the performance benchmarking of Australian fixed interest funds. Proceedings of Modelling and Simulation Society of Australia and New Zealand International Congress on Modelling and Simulation. (pp. 1221-1226). Townsville, Qld. Modelling and Simulation Society of Australia and New Zealand. Available [here](#)

This Conference Proceeding is posted at Research Online.

<https://ro.ecu.edu.au/ecuworks/3161>

Some evidence on the Performance Benchmarking of Australian Fixed Interest Funds

David E Allen and Victor Soucik

School of Accounting, Finance and Economics, Edith Cowan University, Perth.

Abstract: In this paper we analyse the performance of Australian fixed interest managed funds by examining the relative effectiveness of various indices of bond performance which are combined with various measures of: interest rate fluctuations, economic fundamentals, maturity risk, default risk, and equity market returns, in an attempt to find an 'optimum' index. Our dataset is sourced from the Australian fund-rating agency ASSIRT. We show that a correct combination of a bond market variable, a mixture of interest rate factors and economic factors as well as a proxy for movements in the equity markets yield the optimal benchmark.

Keywords: Performance Benchmarking, Fixed Interest Managed Funds.

Introduction

In this paper we analyse a number of combinations of benchmarks suitable for benchmarking fixed interest funds. Previous studies have produced a wide range of conflicting results. Lehman and Modest (1987) and Grinblatt and Titman (1994) examined various models and benchmarks suggesting that the choice of the performance measure and the reference benchmark have a profound influence on the excess returns observed. Robson (1986) came to similar conclusions in an Australian study. Friend et al. (1970) cautioned "against using a benchmark that effectively tricks the alpha calculation by overweighting [certain] returns", thus highlighting the issue of a 'fair' benchmark definition (Ippolito, 1993). We analyse a set of benchmarks for fixed interest funds and assess their comparative results as well as their individual efficiencies. We use the ASSIRT database on Australian managed funds. Our sample period is divided into two five-year time frames between 1990 and 1999. This permits a number of inter-temporal analyses of our results.

The remainder of this paper is organised as follows: section II describes the research method and data, followed by Section III which presents the results. Section IV concludes.

II. Research Method and dataset

Fama and French (1996) and Carhart (1997) argue that the influence on fund returns does not arise from a single source only, as typically represented by the single factor market model regression. We apply an extension to this model using multiple factors. In addition to the single index market proxy - $r_{m,t}$ we include a vector of factors $\Omega_{BM,t}$.

$$r_{i,t} = \alpha_i + \beta_i \times \Omega_{BM,t} + \varepsilon_{i,t} \quad (1)$$

where $r_{i,t}$ is the excess return (the raw return minus the risk free rate) on fund i in the month t ; α_i represents the abnormal performance of the fund i ; β_i represents the beta risk of fund i to each factor; $r_{m,t}$ is a measure of excess returns on the benchmark market index and $\varepsilon_{i,t}$ is the error term with expected characteristics of a white noise (such as a mean of zero).

Our definitions of the market returns proxy include the All Ordinaries Index, and a 500 stock Value-Weighted Index. Our multifactor proxies include factors designed to capture the effects of interest rate fluctuations, term spreads, default spreads, and GDP growth. The All Ordinaries Index is a value-weighted index tracking top firms listed on the Australian Stock Exchange and may be biased towards a small number of large, well-established companies. We also construct a value-weighted index. We construct this benchmark as an open-ended index, which thus eliminates any survivorship bias and non-trading bias.

$$VWI_{t+1} = VWI_t \times \left[\sum_{i=1}^{N_t} \left[\frac{MV_{i,t}}{\sum_{j=1}^{N_t} (MV_{j,t})} \times \left(\frac{P_{i,t+1}}{P_{i,t}} - 1 \right) + 1 \right] \right] \quad (2)$$

The regression of every fund in every time frame for every model against every benchmark and every factor was performed. In specifying an appropriate performance measurement model, due consideration must be given to the trade-off between the model's ability to explain variance in assets' returns, and its parsimony aimed at improving forecasting accuracy. The first step in the process of forming a new performance measure

is therefore a study of the explanatory power contained in each benchmark specification. In this context it is important to recognise the two-dimensional nature of benchmark information – the ability to explain *temporal* changes in return series and the capacity to explain cross-sectional returns variations *across* individual funds. Initially we examine the explanatory quality of indices *within* each category. Since the incremental information can have a compounding effect in explaining fund returns as we add more factors into a benchmark, we compute *all* benchmarks comprising *all* combinations of one through to n factors, where n is the total number of factors in the given category. For the *aggregate bond returns* category (see above), for example, this resulted in 127 benchmarks, seven of which were in a one-factor group, twenty-one in two-factor group through to one benchmark comprising all seven factors¹. Time series for these benchmarks are regressed against each fund in the sample, producing a sequence of regression coefficients and additional statistics such as the adjusted-R² values. Armed with these results we then set out to examine the explanatory power of different factors and factor combinations in temporal as well as APT sense². Our time-series methodology involves detailed examination of R² and adjusted-R² statistics resulting from the above regressions³. We analyse these results at three independent levels. First, we examine the *average* explanatory power offered by groups with different numbers of factors, thus creating an $n \times n$ matrix of t-statistics and p-values. Second, we conduct an F-test of a joint hypothesis that the explanatory power is equal amongst benchmarks with a *given* number of factors, thus reflecting on the substitutability of factors. Third, we formulate an $m \times m$ matrix of t-statistics and p-values, where m is the total number of benchmarks defined for a given category⁴. This permits us to look at the differential explanatory power of individual pairs of factor compilations. If the factors are perfect substitutes, information

¹ Sum of combinations:
 $\binom{7}{1} + \binom{7}{2} + \binom{7}{3} + \binom{7}{4} + \binom{7}{5} + \binom{7}{6} + \binom{7}{7} = 7 + 21 + 35 + 35 + 21 + 7 + 1 = 127$

² Models based on the Arbitrage Pricing Theory of Ross (1976) were formulated to explain the cross-sectional behaviour of returns on alternative bonds by pricing each of the factors contributing to the observed variation. See, also Elton et al (1995) for their application of APT to analysis of bond fund returns.

³ We examine the explanatory power indicated by R² in the context of different number of independent factors, as well as Adjusted-R² which already takes into account the loss in degrees of freedom as more independent variables are introduced.

⁴ For example, the *aggregate bond returns* category, $m = 127$ as previously calculated

content for any combination should not only be identical relative to each other, but also to each factor individually.

In the cross-sectional analysis we develop two separate tests. A first test looks at the proportion of cross-sectional variation explained by each benchmark using a method similar to Elton, Gruber and Blake (1995), see equation (3).

$$R^2 = 1 - \frac{\sum_{i=1}^n (r_{i,t} - r_{i,t}^e)^2}{\sum_{i=1}^n (r_{i,t} - \bar{r}_i)^2} \quad (3)$$

We define the unexplained return for each fund at each point in time as the difference between realized return, $r_{i,t}$ and the expected return from equation (1), $r_{i,t}^e$. Next we record the R² from the regression of realised returns on the unexplained returns. This is the proportion of variation *not* explained by the benchmark. Taking one minus this figure therefore suggests the proportion of cross-sectional variation explained by the factors in the benchmark. Unlike Elton et al. (1995), our formulation of the R² as a time-series based on cross-sectional stacks allows us to attribute a significance level to each mean as well as to comparison of means between alternative benchmarks. We then perform the same three-level analysis as for the time-series. The second test of cross-sectional data determines the individual contribution by each factor towards the overall estimation of fund returns, using equations (4A) and (4B).

$$w_{(k|BM),j} = \frac{\beta_{k,j} \times \bar{r}_{k,j}}{\sum_{i=1}^n \beta_{i,j} \times \bar{r}_{i,j}} \quad (4A)$$

$$\bar{w}_{(k|BM)} = \frac{w_{(k|BM),j}}{\sum_{j=1}^s w_{(k|BM),j}} \quad (4B)$$

The weight w for each factor k forming a part of benchmark BM is defined as the proportional product of factor coefficient β and the average return \bar{r} on that factor over the regression time frame (eq 4A). Weights are first calculated for every firm j to permit computation of the series variance and hence the significance level attributed to the average weight $\bar{w}_{(k|BM)}$ (eq 4B)⁵. In line

⁵ See also Elton, Gruber and Blake (1995) on their estimation of factor contributions in a more limited collection of benchmarks applied to the sphere of ME bond funds. In their analysis, however, the authors derive the weights directly from averages of factor betas, and are thus unable to attribute significance levels to the resulting weights.

with our hypothesis, if n factors are perfect substitutes then each should contribute $1/n$ towards the formation of returns expectation.

The results of this two-pass analysis suggests the 'preferred' factor(s) chosen from each category based on its (their) joint contribution towards explaining temporal and cross-sectional variation of returns. Consideration is given to the trade-off between model's explanatory power improved by adding more factors, and its parsimony characterised by fewer factors. Given a statistically insignificant difference between the information content of two alternative factor groups, the group with fewer factors is preferred. The method is then reapplied to *all* benchmark combinations formed from the preferred factors. An outcome of this final two-pass test is the benchmark that uses the fewest number of factors to achieve the maximum explanatory power in both dimensions of fund returns⁶. Since the computation of several benchmark factors is specific to the Australian market, we peruse the actual asset allocations of every fund classified as *interest bearing* [ASSIRT code *IB*], and approve only those funds, which principally invest in Australian fixed interest securities. This results in 168 funds entering our sample in the first period and 537 funds in the second. Our two periods of analysis are: 1990-1994 and 1995-1999. The bulk of our data is taken from the Datastream database compiled by Primark, cross-checked (and when necessary, supplemented) by the Australian Stock Exchange electronic data requests.

Our factors are constructed as follows: the monthly return series for the UBS Warburg Composite Bond Index, Salomon Smith Barney WGBI Index, JP Morgan Bond Return and the JP Morgan Bond Price Index were taken from the Datastream database. The Datastream All Maturities Bond Index is formed by Primark Corporation and represents a composite of bond yields covering the full spectrum of maturities. The Value Weighted Index and the Equally Weighted Index of managed fixed-interest fund returns were computed from the monthly return series contained in the ASSIRT Library.

All of the interest rate and yield series, including the 90-Day Treasury note rates, 10-Year Government bond rates and the composite Datastream indices of government bond yields from different maturity segments, were from the Datastream database. This source was also used to obtain monthly information on Australian inflation position, and the quarterly Gross Domestic Product (GDP) reports which are our economic factors. To concur with our monthly frequency requirement,

⁶ It should be noted that the explanatory power offered by this benchmark might not be itself an absolute maximum, but rather be statistically indifferent from the absolute maximum.

we have interpolated official GDP figures to fill the intra-quarter estimates under the assumption of progressive growth from one quarter to the next. In addition, all GDP series have undergone an orthogonality transformation against inflation data to highlight their differential information content.

The remaining two bond indices, the Lehman Brothers High Yield index and the WDR index of Asset Backed Securities, were extracted from Datastream as was the All Ordinaries Accumulation Index. We paired the UBS Warburg Composite Bond Index, Salomon Smith Barney WGBI Index, JP Morgan Bond Return Index and the JP Morgan Bond Price Index with their respective US counterparts⁷. In the same spirit we match up the Datastream All Maturities Bond Index with its corresponding US series.

III Results

We start by reviewing the information content of factors within each category. Winners from each category are then earmarked for selection into the final round where the preferred factors across all categories are tested. We look at both the ability to explain temporal as well as cross-sectional variations in the returns series in two time frames: 1990-94 and 1995-99.

Table 1 presents a summary of results derived from the information efficiency tests carried out with factors representing aggregate bond market returns. Panels A1 and A2 (not reported because of space constraints) present the time-series explanatory power for the 1990-94 and 1995-99 periods and are based on adjusted R^2 values and tests on their group means. (Two further panels, also not included, explain the data in cross section for the same two periods).

Focusing first on the *average* explanatory power offered by the market benchmark in the time series sense as presented in the first half of Panels A1 the information content increases relatively uniformly from an average of 64.4% in 1990-94 when only a single factor is used, to a peak of 77.8% in 1990-94 when all factors are combined. The variability in the goodness of fit of individual combinations *within* each level (i.e. given a number of factors) varies substantially. The F-Test shows that the increment in 1990-94 is relatively uniform across the combinations with the test statistics approximating unity at all levels. This can be also confirmed in the matrix of level differences showing the only significant *average* difference to exist between the first (one factor)

⁷ On the basis of responses from various institutions we feel assured that the principal differences in definitions applicable to Australia and the ME lie in procedural methods of data compilation, leaving the substantive nature of these index counterparts equivalent.

and the last (all factor) levels. This is supported by an F-Test statistic that is significant at a 1% level for all levels. Given this finding of factor non-substitutability it is therefore important to identify which factor(s) perform the best. Turning attention to the individual performances we find consistent dominance of the indices based on managed funds themselves.

TABLE 1
Two-Pass Analysis of Factors Representing Aggregate Bond Market Returns

Presented is a summary of statistics resulting from the two-pass analysis of the seven factors chosen in this study to proxy the movements of the bond market. Results from temporal tests, presented in Panels A1, are derived from R^2 values of time series regressions $r_{i,t} = \alpha_i + \beta_i \times \Omega_{BM,t} + \varepsilon_{i,t}$. The section shown is based on the average explanatory power attributable to combinations of n -factors and thus reflects the incremental benefit derived from adding more independent variables. Whilst the last column presents the group averages together with an F-Test results of benchmark substitutability, the first set of columns relay a comparative matrix. Second section of each panel shows the individual performance of each benchmark, as well as a comparison to the maximum R^2 obtained when all benchmarks are combined. This reflects on how well the more parsimonious combination of factors is able to perform against a peak that is achieved by non-parsimonious inclusion of every factor in the category.

PANEL A1: Tests of Time-Series Variance (1990-1994)

Category [†]	T-Test of Difference in Group Means						F-Test of Grp Means
	1F	2F	3F	4F	5F	6F	
1F							0.644 (0.920 E*)
2F	0.03 1 (0.6 34)						0.675 (0.955 E*)
3F	0.05 9 (0.3 71)	0.02 7 (0.6 77)					0.702 (0.991 E**)
4F	0.08 3 (0.2 08)	0.05 1 (0.4 35)	0.02 4 (0.7 15)				0.726 (0.999 E**)
5F	0.10 3 (0.1 19)	0.07 2 (0.2 79)	0.04 4 (0.5 04)	0.020 (0.761)			0.746 (1.000 E**)
6F	0.12 0 ^{D*} (0.0 70)	0.08 8 (0.1 82)	0.06 1 (0.3 57)	0.037 (0.577)	0.01 7 (0.8 00)		0.763 (0.989 E**)
7F	0.13 5 ^{D**}	0.10 4	0.07 6	0.052 (0.435	0.03 2	0.01 5	0.778 (N/A)

[†] Corresponds to groups of benchmarks incorporating one through to seven factors.

Factor [†]	T-Test of Difference in Means of Benchmark Pairs						
	VW 1	EW 2	DS ₃ 3	UBS W ₄ 4	SSB 5	JPM R ₆ 6	All Factors 7
V							-0.053 (0.136)
W ₁							
E	0.01 5 (0.6 44)						-0.042 (0.209)
W ₂							
DS ₃							
3	0.01 9 (0.5 67)	0.03 4 (0.3 02)					0.076 ^{D**} (0.023)
UB							
SW	0.01 2 (0.7 21)	0.02 7 (0.4 16)	0.00 7 (0.8 31)				-0.059 ^{D*} (0.094)
SS							
B ₅	0.02 0 (0.5 49)	0.03 5 (0.2 90)	0.00 1 ^{E**} (0.9 79)	-0.008 (0.811)			0.077 ^{D**} (0.021)
JP							
M	0.02 0 (0.5 31)	0.03 6 (0.2 78)	0.00 2 ^{E**} (0.9 57)	-0.009 (0.790)	0.00 1 ^{E**} (0.9 79)		0.077 ^{D**} (0.020)
RI ₆							
JP							
MP	0.01 7 (0.5 94)	0.03 3 (0.3 21)	0.00 1 ^{E**} (0.9 69)	-0.006 (0.862)	0.00 2 ^{E*} (0.9 48)	0.00 3 ^{E*} (0.9 27)	0.074 ^{D**} (0.025)

However, in the cross sectional results such difference is only restricted to groups of two to four factors, with *adjacent* levels showing insignificant difference across both time frames.

The same type of analysis in time series and cross-section was applied to each of the factor variables considered as candidates for our multi-factor benchmark. In the case of factors representing interest rate fluctuations, the DataStream Government Bond Index and the lagged version thereof proved to be the best choice. The influence of economic factors on fixed interest managed funds appeared to be best-captured by the measures of inflation and GDP, whilst term and maturity risk appeared to be optimally reflected in the composite Datastream Government Bond index which reflects a maturity premium difference between instruments of one to three years and over ten years. The Lehman Brothers High Yield Index optimally captured the influence of default risk, a reflection of the

[†] VW₁ and EW₂ refer to the Value Weighted Index and Equally Weighted Indices of Managed Bond Fund Returns, DS₃ refers to the Datastream All Maturities Bond Index, UBSW₄ to the UBS Warburg Composite Index of government, semi-government and corporate fixed interest securities, SSB₅ to the Salomon Smith Barney Government Bond Index, JPMR₆ and JPMR₇ to the JP Morgan Bond Return Index and the JP Morgan Bond Price Index, respectively.

importance of including non-investment grade bonds in the measure. Finally, the Australian All Ordinaries Index appeared to best reflect the influence of equity market movements on Australian fixed interest managed funds. (The results are available on request from the authors).

Once the most appropriate index to reflect a particular economic factor has been isolated, its composite performance, in-conjunction with all the other factors needs to be assessed. Does it make a significant contribution to the overall model? Tests were undertaken on the equally weighted fund-based index (EW), economic proxies for inflation (INFL) and the orthogonalised measure of GDP growth (GDP), the index for high yield non-investment grade bond securities (LBHYI), the All Ordinaries Index (AOI), the DataStream medium term interest rate factor (DSGBI) and the lagged variant thereof (ΓDS_{GBI}) and finally the term premium between the long term and medium term fixed interest securities (δDS_{LM}). The objective of the joint analysis was to search whether the peak R^2 of this group could be achieved in a more parsimonious manner with fewer factors.

The temporal analysis of adjusted R^2 averages reveals that whilst the addition of extra factors contributed significantly to the information content carried by the benchmark, such increments do experience diminishing returns. In fact, combining more than six factors to form a benchmark has no real benefit. This is clearly demonstrated where the increase in R^2 as a result of using all eight factors instead of six factors increases the average explanatory power by 7.8% with a p-value of 0.259 in the 1990-94 period. This are set against peak R^2 values of 95.2%. When combinations of six or less factors are formed, however, significant informational differences are evident between the resulting benchmarks. F-Tests significantly reject the null hypothesis of benchmark equality for all levels up to six factors (6F) in both periods. As such analysis of individual factors and combinations thereof is warranted.

The equally weighted index of managed fund returns takes the lead amongst single factor benchmarks. Coming next are the term and risk premium variables, as well as the equity market proxy. The remaining four variables representing interest rate and economic influences lie on the other side of the spectrum with low R^2 values. Whilst the individual performance of these factors is weak, they team up strongly with other factors, particularly the aggregate market factor. The leading pair of factors combines the aggregate market index with the interest rate proxy in 1990-94 and the lagged variant of this proxy in 1995-99. Although such a combination substantially improves the information content of such

benchmark, it too still falls significantly short of the peak.

We perform one further test of robustness for the information content of the selected factors. The selection process followed several steps that included picking a winner factor from each category and then finding the most suitable combinations. We test whether combining non-selected factors from one category with non-selected factors from another could produce a more informative result than the combination of winners. We retest all alternatives for each of the factors that enter our pre-selected benchmark and find their temporal and cross sectional explanatory power. To make the exercise more manageable we take note of the fundamental pairing of factors within two of the categories. First, inflation and GDP growth have been inherently linked in the economic fundamentals category, as has been a pair of their estimation errors. Similarly, spot and lagged variants of the interest rate proxies have been closely tied together. Consequently, in this test we test seven aggregate bond market factors against two alternative pairs of economic factors, two alternative pairs of equity market proxies and two alternative pairs of interest rate proxies. Table II summarises the results.

TABLE II
Comparison of Winning Factor Combinations

Presented below are the results of information content across time (Panel A) and in cross section (Panel B) achieved from combinations of benchmark factors according to the winning framework. Below is the legend of constituting factors for the four-digit COMBINATION code [ABCD]:

- A (Aggregate bond returns, 1F)*
1. Value Weighted Index of Fixed Interest Managed Funds
 2. Equally Weighted Index of Fixed Interest Managed Funds
 3. UBS Warburg Composite Index
 4. Salomon Smith Barney WGBI Index
 5. Datastream All Maturities Index
 6. JP Morgan Bond Return Index
 7. JP Morgan Bond Price Index
- B (Economic variables, 2F)*
1. Inflation + GDP Growth
 2. E^e (Inflation + GDP Growth)
- C (Equity Market Returns, 1F)*
1. All Ordinaries Accumulation Index
 2. 500 Value Weighted Index
- D (Interest Rates, 2F)*
1. 90-Day Treasury Note Rate + Lagged Variant
 2. Datastream GBI with one to three year maturities + Lagged Variant

PANEL A: Time-Series Winners					
1990-1994			1995-1999		
Combina	R^2	Δ p-value	Combinati	R^2	Δ p-value
2112	93.9%		2112	94.9%	
1112	89.8%	0.000	1112	91.2%	0.000
2111	88.9%	0.003	2111	90.1%	0.000
2211	88.6%	0.110	2121	88.6%	0.000
2122	88.0%	0.064	1111	88.5%	0.432

PANEL B: Cross-Sectional Winners

1990-1994			1995-1999		
Combina- tion	R ²	Δ p- value	Combinati- on	R ²	Δ p- value
2112	88.5%		2112	88.6%	
1112	86.1%	0.002	1112	87.1%	0.013
2111	86.0%	0.364	2211	86.7%	0.042
1111	85.4%	0.021	2121	86.5%	0.396
2121	85.4%	0.931	1111	86.1%	0.124

Referring to the legend for the table it is immediately clear that the six factors as prescribed by the above analysis take the lead in both dimensions.

4. SUMMARY

We have set out to find an optimum performance measure for fixed interest managed funds. The choice of benchmark proves to be a major influence on the final results. We find it critical to include a factor representing aggregate bond returns, a proxy for interest rates, economic factors and an index representing equity market returns for a benchmark to be informative both across time and in cross section. The optimum benchmark includes the Equally Weighted Index, a medium term interest rate proxy such as the one to three year government bond index compiled by DataStream and the lagged variant thereof, an inflation variable coupled with an orthogonalised GDP Growth measure and finally an All Ordinaries Accumulation Index representing the movements of equity markets. Our results bear some similarity to Blake, Elton and Gruber's (1993) suggestion that "bond returns can be explained by no more than three, and possibly two factors", but we find that four factors are optimal.

REFERENCES

Blake C.R., Elton E.J. and Gruber M.J., 1993, The performance of bond mutual funds, *Journal of Business* 66, 371-403

Carhart M.M., 1997, On Persistence in Mutual Fund Performance, *Journal of Finance* 52, 57-82

Elton E.J., Gruber M.J., and Blake C.R., 1995, Fundamental Economic Variables, Expected Returns, and Bond Fund Performance, *Journal of Business* 69, 133-157

Fama E.F. and French K.R., 1996, Multifactor Explanations of Asset Pricing Anomalies, *Journal of Finance* 51, 55-84

Friend I., Blume M. and Crockett J., 1970, *Mutual Funds and Other Institutional Investors*, New York: McGraw Hill

Grinblatt M. and Titman S., 1994, A Study of Monthly Mutual Fund Returns and Performance Evaluation Techniques, *Journal of Financial and Quantitative Analysis* 29, 419-444

Ippolito R.A., 1993, On Studies of Mutual Fund Performance 1962-1991, *Financial Analysts Journal*, January - February, 42-50

Lehman B. and Modest D., 1987, Mutual Fund Performance Evaluation: A Comparison of Benchmarks and Benchmark Comparisons, *Journal of Finance* 42, 233-265

Robson G.N., 1986, The Investment Performance of Unit Trusts and Mutual Funds in Australia for the Period 1969 to 1978, *Journal of Accounting and Finance*, 55-79

Ross S., 1976, The Arbitrage Pricing Theory of Capital Asset Pricing, *Journal of Economic Theory*, 13, 341-360

White H., 1980, A Heteroscedasticity Consistent Covariance Matrix Estimator and a Direct Test for Heteroscedasticity, *Econometrica* 48, 817-838