6-2001

Testing for Differences in Risk Exposure Among Assets

Scott Gilbert
Southern Illinois University Carbondale

Petr Zemcik
Southern Illinois University Carbondale

Testing for Differences in Risk Exposure Among Assets

Scott Gilbert* and Petr Zemčík

Department of Economics
Southern Illinois University at Carbondale

June 2001

Abstract

One purpose of asset pricing models is to explain empirical differences in the time-averaged returns among risky assets. Of interest is whether differences in risk exposure can explain differences in average returns. In the framework of asset return regression systems, the problem is to test equality of parameter values across equations. We examine the performance of Wald and score tests of cross-equation restrictions, with robustness to empirically documented residual heteroskedasticity and autocorrelation. The tests display distortions, but in simulation they perform well when applied parsimoniously. We use the tests to examine the risk exposure of stocks sorted by firm size.

Key Words: Asset pricing; Multivariate regression; Heteroskedasticity; Autocorrelation; Finite sample properties

JEL classification: C15; C30; G12

*Corresponding author: Department of Economics, Mailcode 4515, Southern Illinois University at Carbondale, Carbondale, IL 62901-4515, Tel:(618) 453-5065, Fax: (618) 453-2717, E-mail: gilberts@siu.edu.

1. Introduction

A main purpose of asset pricing models is to explain empirical differences in the average returns among stocks and other risky assets. In the tradition of the Sharpe (1964) and Lintner (1965) capital asset pricing model (CAPM) and its generalizations via the Merton (1973) and Breeden (1979) intertemporal equilibrium models and the Ross (1976) arbitrage pricing theory (APT), a higher average return is compensation for greater exposure to some source(s) of risk.

The linear regression model of excess returns on risk factors provides the framework for much applied work, including recent contributions by Fama and French (1993, 1996) who propose two ways to judge the success of such models: The magnitude of the fitted model's prediction errors, and the proximity of estimated parameters to hypothesized values. Test statistics, specifically t statistics and F statistics, provide a formal metric for measuring proximity of unrestricted parameter estimates to the maintained hypothesis, and Gibbons, Ross and Shanken (1989) and Fama and French (1993, 1996) apply these statistics to the hypotheses of zero intercept and zero slope restrictions.

In the regression model, it is possible to formally test whether differences in mean return could be due to differences in risk exposure. The relevant null hypothesis is that differences in risk exposure have no impact on differences in mean return, e.g. that the slope parameters (betas) for a given risk factor are the same for each risky asset under study. The hypothesis of equal slopes across assets is more general than the hypothesis of zero slopes, and it is possible to encounter risk factors which add explanatory power to the model, with slopes 'far' from zero, yet which are unrelated to differences in mean returns across some interesting sets of assets. Previous empirical work has shed some light on the issue of beta equality across assets (including Fama and French 1993, 1996, Li and Hu 1998), and Perez-Quiros and Timmerman (2000) test equality of parameters for some non-linear dynamic models of returns on small and large firms, for example.

The present work studies the problem of testing equality of parameters across equations in a linear regression system, with robustness to some econometric features of financial data. While classic t statistics and F statistics provide a valuable metric for judging proximity of the unrestricted parameter estimates to their equality-restricted counterparts, inference based on these statistics is not robust to residual heteroskedasticity or serial correlation. Aware of the potential importance of conditional heteroskedasticity for asset pricing applications, MacKinlay and Richardson (1991), Ferson and Foerster (1994), Ferson and Korajczyk (1995) and He, Kan, Ng, and Zhang (1996) suggest the use of heteroskedasticity-robust tests. The presence of autocorrelation in asset returns is well-documented (see, for example, Campbell, Lo and MacKinlay 1997, Ch. 2), and even in cases where assets returns have little autocorrelation themselves, when multiplied by risk factors the resulting series may be strongly autocorrelated, and this insidious phenomenon is important for testing restrictions on the regression parameters. We find highly significant autocorrelation and heteroskedasticity in model residuals (as in the CAPM, reported later), and for robustness we suggest the use of heteroskedasticity and autocorrelation consistent (HAC) test methods.

We examine the performance of HAC test methods, in which residual heteroskedasticity and autocorrelation play no explicit role in model specification, but which are allowed to be present, in general form. The methods, due to Newey and West (1987, 1994), Andrews (1991), Andrews and Monahan (1992) and den Haan and Levin (1997), build on earlier work by White (1980) on heteroskedasticity-robust tests, and on the Hansen (1982) generalized method of moments (GMM). The HAC Wald and score tests, which are equivalent in large samples under the null and local alternative hypotheses, can nevertheless exhibit great disparity and distortion in finite samples.

We simulate the behavior of the proposed tests for models calibrated to financial data, documenting noticeable distortions except in sufficiently parsimonious cases. The source of distortions lies in the test rule, which relies on the large-sample (chi square) distribution of test statistics, and which generally differs from rules based on the exact (but unknown) finite-sample distribution. MacKinnon and White (1985) acknowledge the distortions problem for heteroscedasticity-robust tests, proposing corrective methods, and Ferson and Foerster (1994) examine the importance of distortions for heteroskedasticity-robust tests of some financial models. Den Haan and Levin (1996) report on test distortions for a variety of HAC-type tests in a single equation context (see also Cushing and McGarvey 1999), and the present work describes the multi-equation context. No correction method is available to solve the distortion problem in our (highly complex) situation, and for useful application we suggest that the practitioner limit the number of assets and parameter restrictions under study. In some cases such limitations on models size and complexity will be too confining, and a possible recourse is to build into the regression model a parsimonious specification of heteroskedasticity and autocorrelation, to be estimated jointly with the regression parameters. With this caveat, we recommend the use of HAC Wald and score tests, and our simulations suggest that the HAC Wald test, with a simple pre-whitening method for parameter estimate covariance estimation, is a good choice among these tests.

We use the methods to test for differences in risk exposure for small and large firms. Differences in risk exposure are one possible explanation for observed differences in mean return for the two types of firms, and a host of other possible explanations have been offered (see Schwert 1983 for a review of early theories and Fama and French 1992 for more on empirical evidence). Some researchers attempt to link the underlying risk factors to size and book-to-market related portfolios (see Fama and French 1993 and Chan and Chen 1991). Explanations of systematic differences among stock of different capitalization are also based on statistical factors as in Lehman and Modest (1988) and Connor and Korajczyk (1988) or on factors loosely connected to economic theory as in Chen, Roll and Ross (1986). Since statistical and economic factors appear to have similar predictive power with respect to stock returns (see Ferson and Korajczyk 1995), we use just the economic factors, with which we form a variety of parsimonious one-factor and two-factor models.

Using a variety of one-factor models, we find a number of significant differences in risk exposure for small and large firms, in terms of the market return, default premium, consumption growth, and inflation surprise. To check for stability of risk exposure over time (as discussed in Ghysels 1998, for example), we split the sample and find that the gap in risk exposure appears to shift over time, in a way that is not explained by previous literature. Using a variety of two-factor models, in which the first factor is the market excess return, we find some significant differences in risk exposure for large and small firms, in terms of the market return and other factors (default premium, term premium, consumption growth and inflation surprise), and this pattern also changes from the first to second sub-sample. The overall impression is that some differences in risk exposure, for small and large firms, are large in the metric of the proposed statistical tests, and appear to evolve over time.

2. Model

For a collection of n risky assets, each earning a return during periods $t = 1, 2, ..., T$, let r_{it} denote the excess return to the i-th asset. The linear regression model of asset returns takes the form:

$$r_{it} = \alpha_i^* + \beta_i^* x_t + \varepsilon_{it}, \quad i = 1, ..., n, \tag{1}$$

where x_t is a $K \times 1$ vector of factors, β_i^* is the true value of the i-th $1 \times K$ vector of slopes ('betas'), α_i^* is the true value of the i-th intercept, and the errors ε_{it} have conditional expectation $E[\varepsilon_{it}|x_t] = 0$. In this model, β_{ik} is the expected increase in the excess return r_{it}, given a 1 unit increase in the factor x_{tk}, while $\alpha_i = E[r_{it}|x_t = (0, ..., 0)']$, e.g. α_i is the expected excess return when each factor equals 0. The model is linear in the parameters α and β, but x_t itself may be non-linear in some underlying state variables which themselves may be non-contemporaneous with r_t, hence the model may be both non-linear and dynamic in the state variables, as in Ferson and Harvey (1999), for example.

In the Sharpe-Lintner CAPM version of the model, x is the excess return on the market portfolio, the betas measure exposure to market risk, and the alphas are each 0. Other candidates for x include consumption growth, as in the Breeden (1979) consumption-based CAPM, and other variables, possibly instruments for some latent factors (see Section 4 for a detailed discussion). Estimation of the model (1) is the the first step in the Fama and MacBeth (1973) empirical method (see Ferson and Harvey 1999 for a recent example) and is also used for tests of various hypotheses (see Fama and French 1993, 1996, Gibbons et al. 1989, and Li and Hu 1998).

The hypotheses of present interest take the form of linear restrictions on β and/or α. To concisely express such hypotheses for the purpose of testing, we denote by θ the column vector with entries (top to bottom) $\alpha_1, \beta_{11}, ..., \beta_{1K}, ..., ..., \alpha_n, \beta_{n1}, ..., \beta_{nK}$. With 0_p the column vector consisting of p entries each equal to 0, and with A some user-specified $p \times n(K+1)$ matrix, each linear restriction on the model parameters takes the form:

$$H_0\colon\ A\,\theta^* = 0_p.$$

As a familiar example, testing the zero intercepts ($\alpha_i^* = 0$, $i = 1, ..., n$) restriction is a common means of testing asset pricing models versus some *unspecified* 'anomalies' (as, for example, in Gibbons et al. 1989 and Fama and French 1996). For testing the predictive power of some factors, the relevant hypothesis is that these factors have zero betas, and *specific* asset pricing 'anomalies' arise as the unexpected rejection of zero betas for some factors (as in Fama and French 1993, 1996). When testing the zero intercepts or zero betas null hypothesis, the restriction is of the form:

$$C\,\theta_i^* = 0_q, \quad i = 1, ..., n, \tag{2}$$

for some $q \times (K+1)$ matrix C and some number q of restrictions, in which case the appropriate form of the matrix A in H_0 is:

$$A = I_n \otimes C, \tag{3}$$

where I_n is the $n \times n$ identity matrix, and $\otimes$ is the Kronecker product operator. Hypothesis tests for asset return regressions have typically targeted restrictions of the form (2)-(3).

The practical relevance of 'anomalies' is tied to their implied expectations of asset returns, partly conveyed by:

$$E\, r_{it} = \alpha_i^* + \beta_{i1}^*\, E\, x_{t1} + \cdots + \beta_{iK}^*\, E\, x_{tK}, \qquad i = 1, ..., n, \quad t = 1, ..., T, \tag{4}$$

in which case differences in unconditional expected returns, across assets, are entirely due to differences in intercepts and/or differences in slopes. Regardless of possible 'anomalous' non-zero intercepts or non-zero betas, if these parameters do not vary across assets then they contribute nothing to differences in unconditional expected returns. To test for differences in intercepts and/or slopes across equations, the relevant restriction is of the form:

$$D\, \theta_i^* = D\, \theta_j^*, \qquad i, j = 1, ..., n, \tag{5}$$

for some $r \times (K + 1)$ matrix D, some number r of restrictions, and all assets i, j. The appropriate form of the matrix A in H_0 is then:

$$A = J_n \otimes D, \tag{6}$$

where J_n is the $(n-1) \times n$ matrix with entries $J_{ni1} = 1$, $J_{n,i,i+1} = -1$, and $J_{nij} = 0$ otherwise, $i, j = 1, ..., n$. We are primarily interested in testing restrictions of the form (5)-(6).

3. Tests

In this section we describe methods for testing null hypotheses of the form H_0, with robustness to residual heteroskedasticity and autocorrelation.

3.1. Statistics

A standard method of hypothesis testing in the Sharpe-Lintner CAPM and related models is the Wald test. This method estimates the model under the alternative (unrestricted) hypothesis, and compares the estimates to theoretical values. Since each regression equation in (1) contains the same explanatory variables, a suitable estimation method is ordinary least squares (OLS), equation-by-equation. We let $\hat{\theta}$ denote the OLS estimator, and we let $\hat{V}_{\hat{\theta}}$ denote an estimator, further described below, of the variance-covariance matrix for $\hat{\theta}$. For each given choice of $\hat{V}_{\hat{\theta}}$, the Wald-type test of H_0 is:

$$W \;=\; \hat{\theta}' A' \left(A \hat{V}_{\hat{\theta}} A' \right)^{-1} A\, \hat{\theta}. \tag{7}$$

statistic W measures the distance (in R^p, with norm $||v|| = v'(A\hat{V}_{\hat{\theta}}A')^{-1}\, v$) between the vector $A\,\hat{\theta}$ and the value 0_p hypothesized under H_0, hence larger values of W suggest larger departures of the data from H_0. Wald tests include t-tests of individual intercepts and/or betas, as well as F-tests of joint restrictions on several parameters. These classic tests, which use the standard OLS estimator for $\hat{V}_{\hat{\theta}}$, can perform well in the absence of residual heteroskedasticity and serial correlation, but suffer distortions in the presence of these effects, even when the sample size (T) is large. Hence, due to evidence of these effects (documented later) for robustness we propose to use HAC estimators $\hat{V}_{\hat{\theta}}$, as discussed in Section 3.2.

A second method for hypothesis testing is the score test. To obtain this test, for any parameter values α_i and β_i define the regression residuals for the model (1):

$$e_{it} = r_{it} - \alpha_i - \beta_i\, x_t, \quad i = 1, ..., n, \quad t = 1, \ldots, T.$$

The relevant sample moments comprise the $n(K+1) \times 1$ vector $m(\theta)$, given by:

$$m(\theta) = \frac{1}{T} \sum_{t=1}^{T} z_t \otimes e_t,$$

where z_t is the $(K+1) \times 1$ vector $(1, x_t')'$. Denoting by $\hat{V}_{m(\theta)}$ a HAC estimator (specified below) of the variance-covariance matrix of $m(\theta)$, the score test statistic is:

$$S = \min_{\theta \in H_0} \; m(\theta)' \, \hat{V}_m^{-1} \, m(\theta). \tag{8}$$

To carry out the minimization required for S, we iterate over repeated trials, at each stage simultaneously solving for updated parameter and covariance matrix estimates, as in Hansen, Heaton and Yaron (1996).

The score test measures the distance (in $R^{n(K+1)}$, with the norm $||v|| = v'\hat{V}_m^{-1} v$) between the vector $m(\theta)$ of sample moments and the value $0_{n(K+1)}$ hypothesized under H_0, hence larger values of S suggest larger departures from H_0. For testing linear restrictions H_0 on linear regression systems, the score test is seldom used and the Wald test is standard (whereas for nonlinear problems the score test is common, as in Hansen 1982 and Ferson and Foerster 1994). For HAC-robust Wald tests the exact sampling distribution is unknown and reliance on asymptotic theory can lead to over-rejection under the null hypothesis, as reported in Section 5. For this reason, we examine the score test as a possible companion or alternative to the Wald test.

3.2 Computation

To compute the tests we use the GMM (simultaneous-iteration) routine in EViews 3.1, with a variety of choices for the HAC-robust covariance estimation method. As options in this routine we include covariance estimators based on the Bartlett kernel and the data-dependent Newey and West (1994) bandwidth, with and without pre-whitening (denoted NW and NW-P, respectively). We also include the quadratic spectral kernel with the Andrews (1991) data-dependent bandwidth (without prewhitening, denoted A), and the Andrews and Monahan (1992) method (denoted AM) with pre-whitening. Finally, we include the simple pre-whitening method (denoted VARHAC), studied by den Haan and Levin (1996, 1997). Since the technical details of HAC-robust covariance estimation are neatly summarized in Campbell, et al. (1997) (see also Cushing and McGarvey 1999), we omit them for brevity.

3.3. Decision Rule

The HAC Wald and score tests have unknown distributions in finite samples, even under classical conditions, and we take a standard approach (as in Ferson and Foerster 1994 and Campbell et al. 1997) which is to base test decisions on the asymptotic (chi square) properties of such tests. For the HAC Wald tests, the asymptotic chi square distribution follows from asymptotic normality of OLS regression estimators and consistency of HAC covariance estimators (see, for example, Cushing and McGarvey 1999). For the HAC score tests, the asymptotic distribution can also be shown chi square by invoking suitable assumptions, as we now briefly discuss.

To further justify the presumed asymptotic properties of the Wald and score tests, we assume that the covariance estimators $\hat{V}_{\hat{\theta}}$ and $\hat{V}_m$, when multiplied by the number of time periods T, converge as follows:

$$T\,\hat{V}_{\hat{\theta}} \xrightarrow{p} \Omega_\theta, \quad T\,\hat{V}_m \xrightarrow{p} \Omega_m, \tag{9}$$

where $\xrightarrow{p}$ denotes convergence in probability, in which case Ω_θ and Ω_m are the large-T limits of T times the variance-covariance matrix for $\hat{\beta}$, and for $m(\theta^*)$, respectively. Simplified versions S^* and W^* of the Wald and score statistics are then:

$$W^* = \hat{\theta}' A' \left(A \frac{\Omega_\theta}{T} A' \right)^{-1} A\,\hat{\theta}, \tag{10}$$

and

$$S^* \;=\; \min_{\theta \in H_0} \; m(\theta)' \left(\frac{\Omega_m}{T}\right)^{-1} m(\theta), \tag{11}$$

in which case we assume that:

$$W = W^* + o_p(1), \quad S = S^* + o_p(1), \tag{12}$$

where $o_p(1)$ denotes a term converging to 0 in probability.

Under the null hypothesis and suitable regularity conditions (stationarity, finite moments, mixing, etc., as in White 2000 and Davidson 1994, 2000, for example), the statistics W^* and S^* are distributed asymptotically as chi square variables with p degrees of freedom (see, for example Harris and Mátyás 1999), and hence W and S also have this property when (12) holds. Using these asymptotic distributions, the decision rule for testing H_0 is to reject if the test statistic exceeds the relevant critical value from the chi square distribution.

Despite asymptotic equivalence of the chi square tests S and W, in finite samples they can behave very differently. In the classical single-equation regression model, with regression parameters and covariance parameters estimated via maximum likelihood, the inequality of score (Lagrange multiplier) and Wald test statistics (Buse 1982, Engle 1984) is:

$$S \leq W, \tag{13}$$

and in practice it is possible that S is far smaller than W. Hence, by using the same (chi square) decision rule for both statistics it is possible to reach different conclusions from the two tests. The problem arises due to test *distortions*, caused by use of inaccurate chi square approximations to the true sampling distribution.

For HAC Wald and score tests of equality between parameters in a regression system, the inequality (13) generally fails, and test distortions are more complex. To summarize the problem faced in applying the chi square decision rule, we note here simply that test distortions can be severe when the number p of hypothesized restrictions is large, for a given sample size T. To describe the magnitude of this problem, in Section 5 we report the results of simulations for testing differences in risk exposure among stocks.

4. Data

As dependent variables in the regression model, we use excess returns on stocks of firms ranked by capitalization. We use CRSP NYSE Cap-Based Portfolio Indices, monthly time series based on portfolios rebalanced quarterly. Frequently, cross-sectional differences among stock returns are investigated using decile indices; however, to limit the number of dependent variables (and the potential for test distortions, reported later), we use one return for portfolios combining Deciles 1 through 5, and a second return for deciles 6 through 10, where the largest companies are in Decile 1 portfolio and the smallest in portfolio 10. These returns are produced by CRSP, and we calculate excess returns using the 30-Day Treasury Bill return, also provided by CRSP. We denote the excess returns as r_{LARGE} and r_{SMALL}, respectively.

Summary statistics, for monthly excess returns in the period 1959:02 - 1999:12, are in Table 1. The starting period of the data series is determined by availability of the consumption series (defined below). We further split the sample in two sub-samples, 1959:02-1979:12 and 1980:01-1999:12, enabling us to examine stability of regression parameters. A comparison of the sample means for excess returns, for the sample period from 1959 to 1979, reveals that the excess return on the large-cap portfolio (3.02% annually) is far less then the excess return on the small-cap portfolio (8.08% annually), but the gap in average excess returns changes sign in the second sub-sample (9.98% for the large-caps vs. 8.34% for the small-caps, respectively), consistent with Fama and French (1993) and Horowitz, Loughran and Savin (2000). In all considered sample periods, the excess return on small caps tends to be more volatile, in accord with Malkiel and Xu (1997).

As independent variables in the model (see Table 1 for summary statistics, and Table 2 for correlations with dependent variables), we choose ones likely to affect the stochastic discount rate and/or the expected stream of cash flows. We follow Chen, et al. (1986) and use data on the stock market, bond market, the business cycle and inflation. We augment the dataset by the growth of monetary base to address the issue of asymmetric reaction of firms of different capitalization to restrictive monetary policy (see Gertler and Gilchrist 1994, Li and Hu 1998 and Perez-Quiros and Timmermmann 2000). We do not use portfolios constructed by Chan and Chen (1991) or by Fama and French (1993) since there would be size related variables on both sides of the regression equation, potentially resulting in spurious estimates, especially in the present model with limited number of excess returns.

To describe the stock market we employ the CRSP NYSE value-weighted index. Again, we use returns in excess of the 30-Day Treasury Bill, denoting the results by r_{VW}. The correlation with the large-cap return is close to one, and since the large-cap firms account for most of the market value, this is not surprising (see Table 2, Fama and French 1996 report a similar correlation).

We consider two bond market variables. The effect of unanticipated changes in bond risk premia is measured by the difference (r_{DEF}) between interest rates on the low grade bonds and long-term government securities. The low grade bond interest rate is measured by the Seasoned Baa Corporate Bond Yield, collected by Moody's Investors Service and available at the St. Louis Federal Reserve bank's website. The long-term government bond return-to-maturity is from the 5-year Treasury Bonds, also obtained from the St. Louis Fed website. To describe the term structure we use the difference between the one-period holding return on the 5-year Treasury Bond, collected by CRSP, and the first lag of the return on a 30-Day Treasury Bill. This term premium (r_{TERM}) proxies for the influence of changes in the term structure on equity returns.

As measures of real economic activity, we include the growth rates of industrial production and real per capita consumption. We obtain industrial production data (Market Groups, series b50001, seasonally adjusted) from the Federal Reserve Board's website, and we obtain consumption data (series PCEND, non-durables, series PCES, services, POP, population, series CPIAUCSL, Consumer Price Index For All Urban Consumers, All Items 1982-84=100, all series seasonally adjusted), from the St. Louis Fed's website.

To get a one-period forecast of the inflation rate, we run a regression of the inflation rate, measured by the above consumer price index, on a constant, its lagged value, the lagged value of a Treasury Bill rate and a moving average term (see Fama and Gibbons 1984 for a similar procedure). The unexpected inflation (π_{UI}), in the style of Chen et al. (1986), is defined as the difference between actual inflation and forecasted inflation. Chen et al. (1986) also use the change in the expected inflation; since this variable is highly correlated with the unexpected inflation and implied test results are practically indistinguishable, we omit them for brevity. For money growth, we use the growth rate of the seasonally adjusted monetary base (g_{MON}), obtained from the St. Louis Fed's website (series AMBSL, seasonally adjusted).

5. Simulation

We use computer simulation to assess performance of HAC Wald and score tests of cross-equation restrictions. Of interest are rejection rates under the null hypothesis H_0 and under the alternative, and we report the rejection rates of tests for models calibrated to stock returns. If the chi square distribution is an accurate approximation then the tests W and S should reject under H_0 at a rate near the theoretical test size; otherwise, the tests exhibit noticeable distortions. When testing equality of parameters across equations, the number of restrictions is $p = r(n-1)$, with r the number of parameters restricted in each equation, in which case large r and/or large n can give rise to major test distortions.

To set up the simulation, we let x_t follow a first-order VAR process

$$x_{it} = c + \phi_i \, x_{t-1} + u_{it}, \qquad i = 1, 2, \tag{14}$$

where c is a $K \times 1$ vector of constants, ϕ_i is an $1 \times K$ vector of coefficients and the variables u_{it} are normally distributed with mean zero, independence over time, and some cross-sectional variance-covariance matrix Λ.

We estimate (14) by OLS using for x_t variables in the set $\{r_{VWNY}, r_{TERM}, g_{CONS}, g_{MON}\}$, to see what range of values might be considered realistic for the parameters of the x_t process. For the matrix Φ, consisting of row vectors $\phi_1, ..., \phi_n$, estimates of its elements range from -0.25 to 0.32, and for our simulation we set $\Phi_{ij} = 0.10$ for $i = j$ and $\Phi_{ij} = 0$ for $i \neq j$. While estimates of the constant term tend to be small relative to elements of Φ, they are generally significantly positive, and we set $c = 0.002$ in our simulation exercise. The diagonal elements of the estimated residual covariance matrix $\hat{\Lambda}$ are typically of order 0.0001, and the off-diagonal elements are typically of a lower order, hence we let Λ be a diagonal matrix with each diagonal entry equal to 0.0001.

For the regression errors ε_{it} in (1), we posit a dynamic model with serial correlation and generalized autoregressive conditional heteroskedasticity (GARCH), as follows:

$$\varepsilon_{it} = \psi_1 \varepsilon_{i,t-1} + \psi_2 \sqrt{1 + \psi_3 \, \varepsilon_{i,t-1}^2} \, \eta_{it}, \qquad i = 1, 2,$$

with η standard normal white noise. Parameter ψ_1 specifies the residual autocorrelation, and parameters ψ_2 and ψ_3 specifies the residual conditional heteroskedasticity. We choose ψ's so that the autocorrelation if the error term ε_{it}, as well as its variance relative to that of x's, roughly corresponds to what we observe in historical data series, with r_{1t} and r_{2t} excess returns on portfolios of small and large firms, respectively. In this case, we set $\psi_1 = .1$, $\psi_2 = .003$, $\psi_3 = .2$. The cross-sectional empirical covariance of η_{it} is sometimes positive and sometimes negative empirically, and we assume that the population covariance between η_{1t} and η_{2t} is 0. We generate results for the case $n = 2$, for $K = 2$ and $K = 4$, using 500 simulated time series for r_{it}, $i = 1, 2$, with 240 and 492 observations, corresponding to 20 and 41 years of our historical monthly data, respectively. We use simulation, rather than a bootstrap method as in Ferson and Foerster (1994), to generate the psuedo-data because the regression errors have posited dynamics which would not be replicated by standard bootstrap sampling. We record the number of rejections of the null hypothesis using the chi square critical values at the 5% level of significance.

Table 3 reports rejection rates under the null hypothesis H_0 of equality of all regression parameters across equations, e.g. the case where the restriction defining matrix D in Section 2 equals the $p \times p$ identity matrix. Results are of similar nature when testing equality of intercepts only, or slopes only (simulation results available on request). We calibrate all β values to equal to 1, and all α values to equal 0. The results show a tendency for distortion in both Wald and score tests, for some of the HAC methods, and these effects can be considerably greater than those of classic t and F tests, for which test distortions are described in Gibbons, et al. (1989) and Campbell, et al. (1997, Ch. 5, 6). The VARHAC method, as well as the AM method, do comparatively well. Distortions appear worse in smaller samples, and in the larger model ($K = 4$) with more restrictions ($p = 5$) appearing under the null hypothesis.

To describe performance under the alternative hypothesis, we generate the times series for excess returns via:

$$r_{1t} \;=\; x_{1t} + x_{2t} + \varepsilon_{1t}, \qquad r_{2t} \;=\; x_{1t} + 1.1\, x_{2t} + \varepsilon_{2t},$$

for $K = 2$, and:

$$r_{1t} \;=\; x_{1t} + x_{2t} + x_{3t} + x_{4t} + \varepsilon_{1t}, \qquad r_{2t} \;=\; x_{1t} + x_{2t} + 1.05\,(x_{3t} + x_{4t}) + \varepsilon_{2t},$$

for $K = 4$.

Table 4 reports rejection rates under the alternative hypothesis, with lower rejection rates for the score test than for the Wald test. The disparity can be dramatic in the $K = 4$ setup, for some of the HAC methods.

For the sample sizes under study, test performance under null and alternative hypotheses suggests that the number p of tested restrictions should be kept small, perhaps no more than 3 in models with $K \leq 2$ risk factors. In cases of $n > 2$ asset returns, which we have also simulated but omit for brevity, the number $p = r(n-1)$ of cross-equation restrictions is larger and the distortions and the disparity in score and Wald tests is often greater than when $n = 2$. In such cases we continue to find the rule $p \leq 3$ useful for $T \geq 240$ observations and $K \leq 2$.

In cases where larger models and a greater number of restrictions are desired, larger sample sizes (weekly rather than daily data, for example) may be necessary for useful application of the HAC Wald and score tests. Also possible is to parsimoniously model and estimate the form of residual heteroskedasticity and autocorrelation in the regression model.

6. Empirical results

We apply the proposed methods to the problem of testing for differences in risk exposure among firms of different size (market capitalization). We conduct tests of equality of parameters, equality of slopes for a specified risk factor, equality of intercepts, and intercepts being equal to zero. The tests are formulated by defining matrices C and D in Section 2 accordingly. We first report on the Sharpe-Lintner CAPM specification the asset return regression system, including parameter estimates and their HAC standard errors, tests for residual heteroskedasticity and autocorrelation, tests for parameter equality across assets, and tests of the zero intercepts hypothesis. For these results we use a variety of technical specifications (described earlier) for the HAC Wald test and score test, and the results are broadly similar for each method. We then turn to tests of various other one-factor models, as well as two-factor models that include the market excess return as a risk factor, and for these results we report HAC Wald tests (with VARHAC parameter covariance estimate), which showed relatively good power and modest distortions in our simulations.

We report, in Table 5, parameter estimates and standard errors (for each of five HAC parameter covariance estimators) for the CAPM model, with $n = 2$ excess returns (on small-cap and large-cap stocks), $K = 1$ risk factor (value-weighted market portfolio), and three sample periods (1959-1999, 1959-1979, 1980-1999). Intercept and slope estimates are OLS, equation-by-equation. For the whole sample 1959-1999 and first sub-sample 1959-1979, the results are consistent with a market beta in excess of 1 for the small-cap firms, a beta approximately equal to 1 for large-cap firms, and small intercepts for both firm types. A similar relationship between betas and the market excess return is reported for instance by Chan, Chen and Hsieh (1985) and Fama and French (1993). For the second sub-sample, the results are suggestive of equality among betas across firm types, reflecting the small difference between mean excess returns on firms of different capitalization in that period. Table 6 reports residual tests for the CAPM regressions, including tests for cross-equation error correlation, residual heteroskedasticity and residual autocorrelation. The residuals indicate strong evidence of error correlations and heteroskedasticity, in which case the HAC variance-covariance estimates and hypotheses tests are highly appropriate.

We next report hypothesis tests for the market model - see Tables 7 and 8. The test statistics seem to behave as expected, based on our simulation analysis. The Wald statistic is always greater than the score statistic, and the two are reasonably close, with only minor differences in implied p-values. When testing the joint equality of intercepts and slopes for both assets, at the 5% significance level we reject in the first sub-sample but not in the second sub-sample, consistent with descriptive evidence in Tables 1 and 5. The test of equality of betas suggests significant difference in risk exposure, for large and small firms, in the first but not second sub-sample, while the test of equal intercepts fails to find a significant difference during either sub-sample. We also test for whether the intercepts are each 0, with mixed results (rejection on the second sub-sample but not the first one). The test of zero intercepts for the market model is essentially a HAC version of the standard F-test commonly applied in testing the CAPM. As pointed out in Gibbons et al. (1989), the test of the CAPM is equivalent to the test of *ex-ante* mean-variance efficiency of a particular portfolio and the test statistic (either S or W) can be interpreted as a measure of distance from the mean-variance frontier. The performance of the market model could be viewed as evidence against the CAPM since only p-values for the period from 1959 to 1979 are larger than standard significance values.

Table 9 reports hypothesis tests for a variety of one-factor models other than the market model, using the Wald test statistic with the VARHAC parameter covariance estimator. Tests of equal parameters, for small and large firms, reject the null hypothesis for consumption growth and inflation, both in the full sample and each sub-sample, while for the default premium the null is rejected in the second sub-sample, but not the first. With the exception of the default premium, the apparent source of this parameter heterogeneity is risk exposure (beta), in the full sample and each sub-sample. These results correspond to findings of Chan et al. (1985) who compare similar explanatory variables to portfolios ranked by size for the sample period 1958-1977, roughly our first sub-sample. The relationship between the excess return for firms and the default and term premiums is somewhat weaker than indicated in Fama and French (1993) who use a finer division of firms based not only on the market capitalization but also on the book-to-market-ratio. Similarly to Li and Hu (1998), betas do not differ across firms for industrial production and money variables. Tests for a zero intercept tend to reject the null during the second sub-sample, but not the first sub-sample. The overall impression is that covariance of excess returns with the various risk factors is often significantly different for small and large firms, and, in the case of the default premium, seems to change over time.

Table 10 reports hypothesis tests for a variety of two-factor models in which the first factor is the stock market excess return. For each version of the model, tests reject parameter equality, for small and large firms, in the whole sample and first sub-sample, but not in the second sub-sample. The apparent source of parameter heterogeneity is mostly difference in betas (rather than alphas), particularly market beta. Exposure to risk represented by the second of the two factors shows in each case no significant differences in the first sub-sample, but frequently shows such differences in the second sub-sample (when the second factor is the default premium, consumption growth, or unexpected inflation) or overall (term premium). Tests for zero intercepts tend to reject the null in the whole sample and second sub-sample, but not in the first sub-sample. The test of zero-intercepts is comparable to the F-test of Gibbons et. al (1989) conducted in Fama and French (1996), with size-related and book-to-market related portfolios added to the excess return on the market proxy.

The empirical results suggest importance of risk factors, such as the market excess return, consumption growth, inflation, the default premium and term premium, in explaining differences in small and large firm performance. We found considerable instability in risk exposure over time, which is consistent with the unstable gap in time-averaged returns on small and large firms (as in Table 1). Future work could attempt to identify models which explain such instability, using suitably robust test methods such as the HAC tests that we have studied.

7. Conclusion

In this paper, we have examined the problem of formally comparing assets according to their risk exposure. In the standard framework of linear regression systems, the problem is one of testing linear restrictions across equations in the system. Scrutiny of regression residuals suggests heteroskedasticity and autocorrelation, and hence we proposed the use of

(HAC) tests robust to these data features. HAC Wald tests showed distortions in simulation, as did score tests, but both performed well when used parsimoniously. The HAC Wald test, with parameter estimate covariance estimated by simple prewhitening, showed particular promise, having relatively good power and distortions not much greater than that of other tests.

In application to stocks sorted by firm size, we used the tests to address the possibility that differences in mean returns on small- and large-sized firms coincided with differing exposure to risk. Among various sources of risk studied, it was market risk (proxied by the value-weighted market portfolio), consumption growth, inflation surprises, the default premium and term premium that showed some significant differences in risk exposure for small and large firms. While we focused on stocks sorted by firm size, future work could apply the proposed HAC testing methods to stocks sorted by industry or book-to-market ratio, for example, or to both stock and bond data. As a parallel to the HAC style of robust testing, it would also be helpful to investigate tests based on fully-specified probability models with built-in provision for residual heteroskedasticity and autocorrelation.

References

Andrews, D., 1991. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. Econometrica 59, 817-858.

Andrews, D. and J. Monahan, 1992. An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator. Econometrica 60, 953-966.

Breeden, D., 1979. An intertemporal asset pricing model with stochastic consumption and investment opportunities. Journal of Financial Economics 7, 256-296.

Buse, A., 1982. The likelihood ratio, Wald, and Lagrange multiplier tests: An expository note. The American Statistician 36, 153-157.

Campbell, J., A. Lo and A. MacKinlay, 1997. The Econometrics of Financial Markets (Princeton University Press: Princeton)

Chan, K. and N. Chen, 1991. Structural and return characteristics of small and large firms. The Journal of Finance 46, 1467-1484.

Chan, K., N. Chen and David A. Hsieh, 1985. An exploratory investigation of the firm size effect. Journal of Financial Economics 14, 451-471.

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

Connor, G. and R. Korajczyk, 1988. Risk and return in an equilibrium apt: Application of a new test methodology. Journal of Financial Economics 21, 255-289.

Cushing, M. and M. McGarvey, 1999. Covariance matrix estimation. in L. Matyas, ed., Generalized Method of Moments Estimation (Cambridge University Press: Cambridge) 63-94.

Davidson, J., 1994. Stochastic limit theory (Oxford University Press: Oxford).

Davidson, J., 2000. Econometric Theory (Blackwell: New York).

den Haan, W. and A. Levin, 1996. Inferences from parametric and non-parametric covariance matrix estimation procedures. National Bureau of Economic Research Technical Paper No. 195.

den Haan, W. and A. Levin, 1997. A practitioner's guide to robust covariance matrix estimation. in G. Maddala and C. Rao (eds.) Handbook of Statistics, Volume 15 (Elsevier Science, North-Holland: Amsterdam).

Engle, R., 1984. Wald, likelihood ratio and Lagrange multiplier tests in econometrics. in Z. Griliches and M. Intriligator (eds.) Handbook of Econometrics (North-Holland, Amsterdam), Ch. 13.

Fama, E. and K. French, 1992. The cross section of expected stock returns. The Journal of Finance 47, 427-465.

Fama, E. and K. French, 1993. Common risk factors in the returns on stocks and bonds. Journal of Financial Economics 33, 3-56.

Fama, E. and K. French, 1996. Multifactor explanations of asset pricing anomalies. The Journal of Finance 51, 55-84.

Fama, E. and M. Gibbons, 1984. A comparison of inflation forecasts. Journal of Monetary Economics 13, 327-348.

Fama, E. and J. MacBeth, 1973. Risk, return, and equilibrium: Empirical tests. Journal of Political Economy 81, 607-636.

Ferson, W. and S. Foerster, 1994. Finite sample properties of the generalized method of moments in tests of conditional asset pricing models. Journal of Financial Economics 36, 29-55.

Ferson, W. and C. Harvey, 1999. Conditioning variables and the cross section of stock returns. The Journal of Finance 54, 1325-1357.

Ferson, W. and R. Korajczyk, 1995. Do arbitrage pricing models explain the predictability of stock returns? Journal of Business 68, 309-349.

Gertler, M. and S. Gilchrist, 1994. Monetary policy, business cycles, and the behavior of small manufacturing firms. The Quarterly Journal of Economics 109, 309-340.

Gibbons, M., S. Ross and J. Shanken, 1989. A test of the efficiency of a given portfolio. Econometrica 57, 1121-1152.

Hansen, L., 1982. Large sample properties of the generalized method of moments estimator. Econometrica 50, 1029-1054.

Hansen, L., J. Heaton and A. Yaron, 1996. Finite-sample properties of some alternative gmm estimators. Journal of Business and Economic Statistics 14, 262-280.

Harris, D. and L. Mátyás, 1999. Introduction to the Generalized Method of Moments. in L. Mátyás (ed) Generalized Method of Moments Estimation (Cambridge University Press: Cambridge), Ch. 1.

He, J., R. Kan, L. Ng and C. Zhang, 1996. Tests of the relations among marketwide factors, firm-specific variables, and stock returns using a conditional asset pricing model. The Journal of Finance 51, 1891-1908.

Horowitz, J., T. Loughran and N. Savin, 2000. Three analyses of the firm size premium. Journal of Empirical Finance 7, 143-153.

Lehmann, B. and D. Modest, 1988. The empirical foundations of the arbitrage pricing theory. Journal of Financial Economics 21, 213-254.

Li, L. and Z. Hu, 1998. Responses of the stock market to macroeconomic announcements across economic states. International Monetary Fund Working Paper No. 79.

Lintner, J., 1965. Security prices, risk and maximal gains of diversification. The Journal of Finance 20, 587-615.

MacKinlay, A. and M. Richardson, 1991. Using generalized method of moments to test mean-variance efficiency. The Journal of Finance 46, 511-527.

MacKinnon, J. and H. White, 1985. Some heteroskedasticity consistent covariance matrix estimators with improved finite sample properties. Journal of Econometrics 29, 305-325.

Malkiel, B. and Y. Xu, 1997. Risk and return revisited. The Journal of Portfolio Management 23, 9-14.

Merton, R., 1973. An intertemporal capital asset pricing model. Econometrica 41, 867-887.

Newey, W. and K. West, 1987. A simple positive-definite heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica 55, 703-708.

Newey, W. and K. West, 1994. Automatic Lag Selection in Covariance Matrix Estimation. Review of Economic Studies 61, 631-653.

Perez-Quiros, G. and A. Timmermann, 2000. Firm size and cyclical variations in stock returns. The Journal of Finance 55, 1229-1262.

Ross, S., 1976. The arbitrage pricing theory of capital asset pricing. Journal of Economic Theory 13, 341-360.

Schwert, G., 1983. Size and stock returns, and other empirical anomalies. Journal of Financial Economics 12, 3-12.

Sharpe, W., 1964. Capital asset prices: A theory of market equilibrium under conditions of risk. Journal of Finance 19, 425-442.

White, H., 1980. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. Econometrica 48, 817-838.

White, H., 2000. Asymptotic Theory for Econometricians, second edition (Academic Press: New York).

Table 1
Summary statistics (annualized, percentages)

	r_{LARGE}	r_{SMALL}	r_{VW}	r_{DEF}	r_{TERM}	g_{IP}	g_{CONS}	π_{UI}	g_{MON}
				1959:01-1999:12					
Mean	6.41	8.21	6.35	1.85	1.40	3.43	2.09	0.00	6.68
Median	9.78	12.17	9.73	1.81	0.31	4.38	2.39	-0.18	6.57
Maximum	198.80	389.06	195.37	3.93	112.93	71.97	21.60	16.36	32.40
Minimum	-261.90	-354.16	-266.67	0.14	-79.19	-50.96	-21.58	-10.69	-11.98
Std. Dev.	50.48	67.14	50.73	0.80	18.82	10.54	5.41	2.51	4.86
Skewness	-0.40	-0.18	-0.43	0.18	0.23	-0.10	-0.21	0.52	0.16
Kurtosis	5.22	7.32	5.31	2.26	7.06	9.03	4.49	7.61	5.15
				1959:01-1979:12					
Mean	3.02	8.08	3.29	1.42	0.17	4.12	2.28	0.19	5.86
Median	4.11	9.67	4.92	1.26	-0.30	4.66	2.45	-0.16	5.98
Maximum	198.80	389.06	195.37	3.54	54.81	71.97	21.60	16.36	22.22
Minimum	-144.16	-253.47	-145.93	0.14	-70.98	-50.96	-21.58	-6.89	-10.40
Std. Dev.	49.57	72.57	50.20	0.73	15.33	12.39	5.92	2.59	4.16
Skewness	-0.05	0.40	-0.06	0.85	-0.16	-0.11	-0.19	1.08	0.03
Kurtosis	4.13	6.48	4.07	2.97	6.24	8.48	4.29	8.28	3.96
				1980:01-1999:12					
Mean	9.98	8.34	9.57	2.29	2.68	2.70	1.89	-0.20	7.54
Median	12.38	14.69	12.76	2.25	2.29	3.91	2.24	-0.19	7.32
Maximum	149.87	167.19	149.07	3.93	112.93	25.60	17.78	7.97	32.40
Minimum	-261.90	-354.16	-266.67	0.78	-79.19	-30.09	-14.68	-10.69	-11.98
Std. Dev.	51.27	61.08	51.19	0.60	21.86	8.12	4.81	2.41	5.38
Skewness	-0.75	-1.19	-0.81	0.31	0.27	-0.40	-0.32	-0.24	0.04
Kurtosis	6.45	8.46	6.72	2.78	6.30	4.35	4.38	6.19	5.34

Table 2
Correlations

		r_{SMALL}	r_{LARGE}
1959:02-1999:12	r_{VW}	0.87	1.00
	r_{DEF}	0.19	0.20
	r_{TERM}	0.15	0.24
	g_{IP}	-0.03	-0.02
	g_{CONS}	0.20	0.15
	π_{UI}	-0.22	-0.20
	g_{MON}	-0.06	-0.03
1959:02-1979:12	r_{VW}	0.88	1.00
	r_{DEF}	0.20	0.20
	r_{TERM}	0.17	0.21
	g_{IP}	0.04	0.06
	g_{CONS}	0.22	0.21
	π_{UI}	-0.21	-0.22
	g_{MON}	-0.05	-0.09
1980:01-1999:12	r_{VW}	0.87	1.00
	r_{DEF}	0.27	0.18
	r_{TERM}	0.16	0.26
	g_{IP}	-0.15	-0.14
	g_{CONS}	0.17	0.09
	π_{UI}	-0.23	-0.16
	g_{MON}	-0.07	0.00

Table 3
Rejection rates under null hypothesis

			Covariance Matrix Estimator				
K	sample size	test	NW	NW-P	A	AM	VARHAC
2	240	score	0.03	0.02	0.05	0.05	0.05
		wald	0.08	0.07	0.07	0.07	0.07
2	492	score	0.05	0.05	0.06	0.05	0.05
		wald	0.08	0.07	0.07	0.06	0.06
4	240	score	0.02	0.01	0.04	0.04	0.04
		wald	0.13	0.13	0.10	0.08	0.08
4	492	score	0.03	0.03	0.04	0.04	0.04
		wald	0.08	0.08	0.06	0.05	0.05

Table 4
Rejection rates alternative hypothesis

			Covariance Matrix Estimator				
K	sample size	test	NW	NW-P	A	AM	VARHAC
2	240	score	0.78	0.77	0.85	0.84	0.84
		wald	0.89	0.89	0.88	0.87	0.87
2	492	score	0.99	0.99	0.99	0.99	0.99
		wald	1.00	1.00	1.00	1.00	1.00
4	240	score	0.26	0.26	0.43	0.48	0.48
		wald	0.62	0.61	0.61	0.58	0.58
4	492	score	0.76	0.74	0.82	0.84	0.84
		wald	0.87	0.87	0.88	0.88	0.88

Table 5
Estimation of the market model

				standard errors				
years	f. size	prm.	estimate	NW	NW-P	A	AM	VARHAC
59-99	small	α	0.000822	0.00138	0.00138	0.00128	0.00127	0.00127
	small	β	1.148570	0.04282	0.04322	0.04970	0.05018	0.05029
	large	α	6.57E-05	9.61E-05	9.67E-05	8.28E-05	8.53E-05	8.53E-05
	large	β	0.993957	0.00267	0.00268	0.00257	0.00262	0.00262
59-79	small	α	0.003406	0.00185	0.00185	0.00175	0.00174	0.00174
	small	β	1.266719	0.07075	0.07256	0.08201	0.08955	0.08957
	large	α	-0.000198	0.00011	0.00012	0.00010	0.00011	0.00011
	large	β	0.986656	0.00380	0.00390	0.00371	0.00415	0.00415
80-99	small	α	-0.00135	0.00186	0.00189	0.00190	0.00192	0.00192
	small	β	1.03827	0.05451	0.05457	0.05389	0.05439	0.05441
	large	α	0.00031	0.00013	0.00013	0.00012	0.00012	0.00012
	large	β	1.00045	0.00358	0.00358	0.00350	0.00358	0.00358

Table 6
Tests for residual heteroskedasticity and correlations

Residuals are calculated using the market model; Pearson test = a chi-square test for correlation; White test = F test with no cross terms; Q test = Q statistic for testing 12 lags of autocorrelation; P-values in parentheses

residual property		f. size	test	59-99	59-79	80-99
correlation	across equations		Pearson test	-0.839	-0.921	-0.762
				(0.000)	(0.000)	(0.000)
	across time	small	Q-test	42.388	64.828	9.879
				(0.000)	(0.000)	(0.627)
		large	Q-test	76.814	88.868	18.852
				(0.000)	(0.000)	(0.092)
heteroskedasticity		small	White test	14.074	27.126	3.796
				(0.000)	(0.000)	(0.000)
		large	White test	31.056	45.867	5.997
				(0.000)	(0.000)	(0.000)

Table 7
Market model: Hypothesis testing, Part I

p-values in parentheses

			Covariance Matrix Estimator				
hypothesis	years	test	NW	NW-P	A	AM	VARHAC
equal parameters	59-99	score	8.87	8.77	8.40	8.10	8.18
			(0.01)	(0.01)	(0.02)	(0.02)	(0.02)
		wald	12.17	11.96	9.37	8.94	8.93
			(0.00)	(0.00)	(0.01)	(0.01)	(0.01)
	59-79	score	11.32	11.14	10.41	10.07	10.14
			(0.00)	(0.00)	(0.01)	(0.01)	(0.01)
		wald	18.25	17.64	13.84	11.53	11.50
			(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
	80-99	score	0.78	0.77	0.75	0.74	0.75
			(0.68)	(0.68)	(0.69)	(0.69)	(0.69)
		wald	0.88	0.87	0.86	0.85	0.85
			(0.64)	(0.65)	(0.65)	(0.65)	(0.65)
equal slopes	59-99	score	8.37	8.31	7.72	7.65	7.71
			(0.00)	(0.00)	(0.01)	(0.01)	(0.01)
		wald	11.78	11.56	8.86	8.68	8.64
			(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
	59-79	score	10.47	10.31	8.88	8.99	8.96
			(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
		wald	14.28	13.57	10.76	9.01	9.01
			(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
	80-99	score	0.39	0.39	0.40	0.39	0.39
			(0.53)	(0.53)	(0.53)	(0.53)	(0.53)
		wald	0.43	0.43	0.44	0.43	0.43
			(0.51)	(0.51)	(0.51)	(0.51)	(0.51)

Table 8
Market model: Hypothesis testing, Part II

p-values in parentheses

			Covariance Matrix Estimator				
hypothesis	years	test	NW	NW-P	A	AM	VARHAC
equal intercepts	59-99	score	0.27	0.27	0.32	0.32	0.32
			(0.60)	(0.61)	(0.57)	(0.57)	(0.57)
		wald	0.27	0.27	0.32	0.32	0.32
			(0.60)	(0.61)	(0.57)	(0.57)	(0.57)
	59-79	score	2.94	2.94	3.63	3.63	3.62
			(0.09)	(0.09)	(0.06)	(0.06)	(0.06)
		wald	3.38	3.40	3.81	3.84	3.83
			(0.07)	(0.07)	(0.05)	(0.05)	(0.05)
	80-99	score	0.69	0.66	0.65	0.65	0.65
			(0.41)	(0.42)	(0.42)	(0.42)	(0.42)
		wald	0.71	0.69	0.69	0.67	0.68
			(0.40)	(0.41)	(0.41)	(0.41)	(0.41)
zero intercepts	59-99	score	5.36	5.28	6.54	6.35	6.37
			(0.07)	(0.07)	(0.04)	(0.04)	(0.04)
		wald	6.07	5.94	6.80	6.52	6.53
			(0.05)	(0.05)	(0.03)	(0.04)	(0.04)
	59-79	score	2.94	2.94	3.89	3.63	3.62
			(0.23)	(0.23)	(0.14)	(0.16)	(0.16)
		wald	3.39	3.40	3.89	3.85	3.85
			(0.18)	(0.18)	(0.14)	(0.15)	(0.15)
	80-99	score	6.68	6.74	7.69	8.66	8.69
			(0.04)	(0.03)	(0.02)	(0.01)	(0.01)
		wald	8.93	8.89	9.57	10.22	10.24
			(0.01)	(0.01)	(0.01)	(0.01)	(0.01)

Table 9
Wald Tests (VARHAC method) of assorted univariate models

p-values in parentheses

		independent variable					
hypothesis	years	r_{DEF}	r_{TERM}	g_{IP}	g_{CONS}	π_{UI}	g_{MON}
equal parameters	59-99	2.75	1.90	0.90	15.90	9.26	2.45
		(0.25)	(0.39)	(0.64)	(0.00)	(0.01)	(0.29)
	59-79	4.59	3.99	3.85	12.44	5.58	3.27
		(0.10)	(0.14)	(0.15)	(0.00)	(0.06)	(0.20)
	80-99	9.27	3.78	1.18	7.03	8.24	1.82
		(0.01)	(0.15)	(0.55)	(0.03)	(0.02)	(0.40)
equal slopes	59-99	1.96	1.26	0.08	14.13	8.54	1.31
		(0.16)	(0.26)	(0.77)	(0.00)	(0.00)	(0.25)
	59-79	2.01	0.28	0.01	7.18	3.06	0.00
		(0.16)	(0.60)	(0.94)	(0.01)	(0.08)	(0.99)
	80-99	9.09	3.28	0.78	6.68	8.12	1.54
		(0.00)	(0.07)	(0.38)	(0.01)	(0.00)	(0.21)
equal intercepts	59-99	0.88	1.01	0.80	0.04	0.95	2.42
		(0.35)	(0.31)	(0.37)	(0.84)	(0.33)	(0.12)
	59-79	0.33	3.30	2.26	0.90	3.58	0.89
		(0.56)	(0.07)	(0.13)	(0.34)	(0.06)	(0.34)
	80-99	9.12	0.17	0.13	2.24	0.70	0.45
		(0.00)	(0.68)	(0.72)	(0.13)	(0.40)	(0.50)
zero intercepts	59-99	9.67	6.27	7.19	2.33	8.27	5.15
		(0.01)	(0.04)	(0.03)	(0.31)	(0.02)	(0.08)
	59-79	7.08	3.34	2.27	1.27	3.81	2.29
		(0.03)	(0.19)	(0.32)	(0.53)	(0.15)	(0.32)
	80-99	10.00	8.81	15.92	10.61	11.34	4.75
		(0.01)	(0.01)	(0.00)	(0.00)	(0.00)	(0.09)

Table 10
Wald Tests (VARHAC method) of assorted bivariate models

p-values in parentheses

		independent variable (in addition to market return)					
hypothesis	years	r_{DEF}	r_{TERM}	g_{IP}	g_{CONS}	π_{UI}	g_{MON}
equal parameters	59-99	9.48	11.24	9.63	22.13	14.70	10.83
		(0.02)	(0.01)	(0.02)	(0.00)	(0.00)	(0.01)
	59-79	11.83	11.50	14.61	18.74	11.67	11.50
		(0.01)	(0.01)	(0.00)	(0.00)	(0.01)	(0.01)
	80-99	12.62	4.25	1.79	8.77	12.26	3.51
		(0.01)	(0.24)	(0.62)	(0.03)	(0.01)	(0.32)
equal slopes (market)	59-99	8.15	9.69	8.73	6.90	6.97	8.71
		(0.00)	(0.00)	(0.00)	(0.01)	(0.01)	(0.00)
	59-79	8.82	8.60	8.83	7.80	8.52	9.00
		(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)
	80-99	0.04	1.08	0.33	0.20	0.10	0.46
		(0.84)	(0.30)	(0.56)	(0.66)	(0.75)	(0.50)
equal slopes (other)	59-99	0.83	4.88	0.07	9.07	4.97	2.14
		(0.36)	(0.03)	(0.80)	(0.00)	(0.03)	(0.14)
	59-79	0.70	0.37	0.17	1.46	0.23	0.23
		(0.40)	(0.54)	(0.68)	(0.23)	(0.63)	(0.63)
	80-99	10.63	3.25	0.75	5.86	9.89	2.46
		(0.00)	(0.07)	(0.39)	(0.02)	(0.00)	(0.12)
equal intercepts	59-99	0.33	0.44	0.33	0.25	0.38	2.50
		(0.56)	(0.50)	(0.57)	(0.62)	(0.54)	(0.11)
	59-79	0.02	3.84	3.13	1.58	3.97	0.45
		(0.88)	(0.05)	(0.08)	(0.21)	(0.05)	(0.50)
	80-99	11.77	0.47	0.29	3.07	0.92	0.45
		(0.00)	(0.49)	(0.59)	(0.08)	(0.34)	(0.50)
zero intercepts	59-99	0.84	6.40	6.25	5.70	6.72	9.31
		(0.66)	(0.04)	(0.04)	(0.06)	(0.03)	(0.01)
	59-79	3.85	3.85	3.15	1.71	4.03	0.85
		(0.15)	(0.15)	(0.21)	(0.42)	(0.13)	(0.65)
	80-99	18.02	9.48	9.11	15.71	10.17	10.91
		(0.00)	(0.01)	(0.01)	(0.00)	(0.01)	(0.00)