Inter-Asset Comparisons of Betas and Returns to Small and Large Firms' Stocks

Scott Gilbert*
Southern Illinois University Carbondale
and
Petr Zemčík
CERGE-EI, Prague†

March 2006

Abstract

In finance, differences among assets' mean returns are often accounted for by differences in betas in asset return regressions. We test whether betas and groups of betas differ *statistically* across assets in a multifactor asset pricing model. We use the individual and joint inter-assets beta equality tests to analyze why expected returns on portfolios of small and large firms differed prior to 1982 but not since. The disappearance of the difference is due to differing sensitivities to variables other than the market e.g. default premium or consumption growth.

KEY WORDS: multifactor asset pricing model; cross-equation restrictions; hypothesis test; firm size

JEL CLASSIFICATION: G10-G12, C315, C30

* *Correspondence address:* Scott Gilbert, Assistant Professor, Department of Economics, Mailcode 4515, Southern Illinois University Carbondale, Carbondale, IL 62901-4515, Tel: (618) 453-5065, Fax: (618) 453-2717, E-mail: gilberts@siu.edu.

† CERGE-EI is a joint workplace of the Center for Economic Research and Graduate Education, Charles University, and the Economics Institute of the Academy of Sciences of the Czech Republic.

I. Introduction

The main purpose of asset pricing models is to explain differences in expected 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), such performance differences arise due to differing sensitivities ('betas') to some economic variables, either some explicit source(s) of risk or some underlying state variable(s). We illustrate how formal tests of equality of betas across assets can be used to interpret time series and cross sectional behavior of expected returns. We focus on the model of returns in the spirit of Chen, Roll and Ross (1986) and combine beta equality tests with point beta estimates and tests for zero intercepts (see Gibbons, Ross, and Shanken 1989) to account for differences between large and small firms' stock returns.

There is a vast number of studies involving beta estimation and some of the studies do take into account standard errors of beta estimates. To our surprise though, we have not been able to find a study, which compares *formally* betas across assets. The present work attempts to fill this gap in the literature. While informal comparison of point estimates maybe sufficient in some applications, it is not when we attempt to evaluate the effect of a group of factors rather then just that of the market beta (market vs. non-market factors, economic vs. statistical, etc.). Rather then merely stating that a multifactor model is needed to explain time series and cross-sectional behavior of expected returns,[1] we would like to uncover more about the nature of these differences. Hence we propose various joint tests for equality of factor sensitivities across assets. Applying these tests to portfolios of stocks of large and small firms yields new insights. Namely, we are able to explain the somewhat puzzling recent disappearance of a measurable difference in mean returns on stocks of small and large firms.

[1]Cochrane (1999) lists this observation as one of the consensual results in recent financial research.

To test formally for differences among particular betas, we make use of panel (time-series + cross-section) regression models of excess returns. This is a standard modelling framework for defining and discussing sensitivities of excess returns on covariates/regressors, e.g. risk factors and/or state variables. For example, estimation of such a model using panel data is the main focus of the influential papers by Fama and French (1993, 1996). With sensitivities defined via the slope parameters in the model, differences in sensitivity are clearly in the domain of formal, testable hypotheses. The financial data typically contain some form of heteroskedasticity and autocorrelation (see for example French, Schwert, and Stambaugh 1987 and Campbell, Lo, and MacKinlay, 1997, Ch. 2), which calls for the use of robust tests. The heteroskedasiticity and autocorrelation consistent methods (HAC) built on earlier work by White (1980) are developed in Newey and West (1987, 1994), Andrews (1991) and Andrews and Monahan (1992), and are further studied by den Haan and Levin (1996, 1997).

Robust tests often suffer from distortions arising due to the test rule, which relies on the asymptotic distribution of test statistics, and which generally differs from rules based on the exact (but unknown) finite-sample distribution.[2] To find limitations of these methods in our testing framework, we examine performance of Wald and Hansen (1982) tests with HAC estimates of covariance matrices of residuals and compare it with that of classical F tests. In simulation the HAC Hansen tests distort less than the F test and HAC Wald tests, and simple pre-whitening is as good or better than other methods of handling serial correlation. Consequently, we report results of estimation conducted using the Hansen method with simple pre-whitening (see den Haan and Levin 1996, 1997).

---

[2]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 (1997) report on test distortions for a variety of HAC tests in a single equation context (see also Cushing and McGarvey, 1999) and Cochrane (2001, Ch. 15) studies the zero intercepts hypothesis in the multi-equation context, using HAC methods of Newey and West.

We use the afore-mentioned methods to test for differences in the economic sensitivity of small and large firm excess returns measured by the top and bottom deciles of the CRSP Capitalization Indices since 1959. The potential for such differences has long been recognized,[3] and as covariates we include standard economic risk factors (market return and consumption growth) as well as other standard economic variables (default premium, term structure, industrial production, inflation, and money growth) related to the economy. Some researchers use instead covariates consisting of size and book-to-market related portfolios (see Fama and French, 1993, and Chan and Chen, 1991), or other statistical factors (Lehmann and Modest, 1988, and Connor and Korajczyk, 1988), but since statistical and economic covariates appear to have similar predictive power with respect to stock returns (see Ferson and Korajczyk, 1995), we use just the latter, similarly to Chen, Roll and Ross (1986).

We first focus on the standard CAPM. Prior to 1982, the sample mean returns for small and large firms are statistically different, indicative of the size effect. We then test for equality of market betas and find that this difference in performance can be in part accounted for by the difference in the level of risk of the two considered portfolios measured by the market beta. The remaining part is due to the positive intercept for small firms' stock returns and corresponds to rejection of the CAPM by the test for zero intercepts, confirming the small size effect. Since 1982, the mean excess returns for the small and large returns are formally indistinguishable with returns on small firms' stocks being smaller. However, the market betas are statistically different and the market beta for stocks of small firm is actually the smaller one. With the CAPM not rejected, this suggests that investment in small firms is less risky and has the same (or greater if one considers only the point estimates) return, making it clearly the better investment opportunity.

In the next step, we investigate bivariate models with the market as one factor and one of

[3]See Schwert (1983) for a review of early theories, and Fama and French (1992, 1993) and Cochrane (1999) for further discussion.

the above mentioned macroeconomic variables as the other. We also consider a model with all seven risk factors. Implications of the bivariate and multivariate models are similar. The point estimates of the market betas in both sub-periods do not change much as compared with the CAPM. The formal test of their equality also shows that they are statistically different at both sub-samples. Before 1982, test results lead to similar conclusions as in the CAPM. Namely, the small firms' stocks are somewhat riskier mainly due to differing market betas and the risk premium is greater than it would be accounted for by the multifactor asset pricing model. Since 1982, implications of our tests are very different from those of the CAPM. Consistent with findings of Horowitz, Loughran and Savin (2000) and Fama and French (1993), we find that the size effect has either disappeared or has been reversed in favor of the large firms. Results of the joint test of beta equality for all variables but the market return signal that there are other sources of risk differences then market beta. This impression is confirmed by individual tests of beta equality; more specifically, all our variables with the exception of industrial production have statistically different sensitivities for the two portfolios. A brief look at point estimates indicates higher sensitivity of large firms to the market but smaller to the other economic variables, which explains why mean returns are similar since 1982.

Our analysis indicates that formal testing of betas, jointly and individually, can be a useful source of information for a financial economist in addition to tests for zero intercepts and point beta estimates. A possible strategy for comparison of groups of assets would be: (i) Get point beta estimates and test for zero intercepts; (ii) Formally compare market betas; (iii) Formally compare sensitivities of other available factors, jointly and individually. Using this approach, we can draw conclusions regarding the disappearance of the statistical difference in expected returns on stocks of small and large firms since 1982. The disappearance is due to previously minor but lately significant differences in sensitivities of expected returns to variables other than the market. Since the early 1980's, the risk exposure to those variables works in the opposite direction than market risk. Consequently, the mean returns on portfolios of small and large firms are similar.

The rest of the paper is organized as follows. Section II presents the formal test of equality of betas in a time-series regression model and Section III discusses the asymptotic properties of the HAC robust Wald and Hansen tests. Section IV addresses the data selection and lists data sources. Section V studies the finite sample properties of the tests in a simulation exercise. Section VI applies the tests to the data and Section VII concludes.

II. 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. We recall that a unifying implication of the finance theories listed above is the following restriction (see Cochrane, 2001):

$$Er_{it} = \beta_i \lambda, \quad i = 1, ..., n, \tag{1}$$

where β_i is a $1 \times K$ vector of betas (sensitivities) for asset i with respect to risk factors, and λ a $K \times 1$ vector of risk premia. It is obvious here that if the mean values $Er_{it}, i = 1, ..., n$ are not all the same then neither are the betas $\beta_i, i = 1, ..., n$.

To estimate the risk premia in (1), one needs to estimate betas first. This first step is common for both the two-pass method and for the Fama and MacBeth (1973) empirical method[4]. Both methods use the linear regression model of asset returns of the form:

$$r_{it} = \alpha_i + \beta_i x_t + \varepsilon_{it}, \quad i = 1, ..., n, \quad t = 1, ..., T, \tag{2}$$

where x_t is a $K \times 1$ vector of covariates (risk factors and/or state variables), β_i is the same beta as in (1), α_i is 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 one

[4]Shanken (1992) relates the time-series and cross-sectional regressions (2) and (1) with respect to these two methods.

unit increase in the covariate x_{tk}, while $\alpha_i = E[r_{it}|x_t = (0, ..., 0)']$, e.g. α_i is the expected excess return when each covariate equals 0. The model is linear in the parameters α and β, but x_t itself may be non-linear in some underlying variables which themselves may be non-contemporaneous with r_t, hence the model may be both non-linear and dynamic in some underlying variables (see Ferson and Harvey, 1999, for a recent example).

In the Sharpe-Lintner CAPM version of the model, x is the excess return on the market portfolio, and the betas measure sensitivity to market risk. 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 V for a detailed discussion). Such models offer an explanation of differences among average returns for various assets, provided that sensitivities differ among assets. Informal comparisons of betas across assets are widespread in the industry, facilitated by point and interval estimates, but formal comparison via hypothesis tests has not received attention in the literature.

The hypotheses of present interest take the form of linear restrictions on β. To concisely express such hypotheses for the purpose of testing, for each equation i we denote by $\theta_{[i]}$ the $(K+1) \times 1$ vector $(\alpha_i, \beta_{i1}, ..., \beta_{iK})'$, and let θ be the $n(K+1) \times 1$ vector $(\theta'_{[1]}, \theta'_{[2]}, ..., \theta'_{[n]})'$. The intercept α_i will be unrestricted with the exception of our simulation exercise. 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: $A\,\theta = 0_p$.

For testing differences in slopes across equations, and the relevant restriction is of the form:

$$D\,\theta_{[i]} = D\,\theta_{[j]}, \qquad i, j = 1, ..., n, \tag{3}$$

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:

$$A = J_n \otimes D, \tag{4}$$

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, and $\otimes$ is the Kronecker product operator. For example, for the test of equality of slopes across equations for $n = 2$ and $K = 1$, we have $p = 1$, $A = \begin{bmatrix} 0, & 1, & 0, & -1 \end{bmatrix}$, $D = [0,\ 1]$, and $J_2 = [1,\ -1]$.

III. Tests

In this section we describe methods of hypothesis testing based on HAC Wald and Hansen tests which we later study as alternatives to the F test. The HAC robust tests are prone to small sample distortion and we can only use in our models since we focus on just two portfolios. In the standard modelling framework where a large number of assets is studied, these robust methods would be impractical. Typically, a researcher attempting HAC estimation would have to assume a certain form of heteroskedasticity and autocorrelation.

HAC Test Statistics

To conduct generalized Wald tests we let $\hat{\theta}$ denote the ordinary least squares (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 test statistic is:

$$W \ = \ \hat{\theta}' A' \left(A \hat{V}_{\hat{\theta}} A' \right)^{-1} A \, \hat{\theta}. \tag{5}$$

The 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. Under the null hypothesis, W is distributed as chi square asymptotically, with p degrees of freedom.

To conduct generalized Hansen tests, for any parameter values α_i and β_i define the regression residuals for the model (2):

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

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 $\hat{V}_m$ an estimator (specified below) of the variance-covariance matrix of $m(\hat{\theta})$, the Hansen test statistic is:

$$S = \min_{\theta \in H_0} \; m(\theta)' \, \hat{V}_m^{-1} \, m(\theta). \tag{6}$$

The Hansen 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. Like the Wald statistic, S is distributed chi square (asymptotically) under the null hypothesis, with p degrees of freedom.

For econometric testing of linear restrictions H_0 on linear regression systems, Hansen tests are seldom used while F and Wald tests are popular, whereas for nonlinear problems the Hansen test is common, as in Hansen (1982) and Ferson and Foerster (1994). Yet our simulations (reported later) suggest a useful role for HAC Hansen tests of parameter equality across equations in linear systems.

Computation

To compute the test statistics we apply formulas (4), (5) and (6), with various specifications for the covariance matrix estimators $\hat{V}_{\hat{\theta}}$ and $\hat{V}_m$. For the HAC Wald and Hansen tests, we use a variety of HAC covariance estimators. Among these are the Bartlett kernel and the data-dependent Newey and West (1994) bandwidth, with and without pre-whitening (denoted NW and NW-P, respectively), 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. Further, we include the simple pre-whitening method (denoted VARHAC) with parametric, vector autoregressive, adjustment for serial correlation, studied by den Haan and Levin (1996, 1997). Finally, for comparison purposes we include the White covariance estimator (WH) which is robust to heteroskedasticity but not serial correlation. Since the technical details of covariance estimators are neatly summarized in Campbell, et al. (1997) and Cushing and McGarvey (1999), we omit them for brevity.

To carry out the minimization (6) required for the Hansen statistic S, we use the GMM (simultaneous-iteration) routine, which at each iteration stage simultaneously solves for updated parameter and covariance matrix estimates, as in Hansen, Heaton and Yaron (1996).[5]

IV. Data

We examine excess returns on stocks of firms ranked by capitalization. We use the industry standard CRSP Stock File Capitalization Decile Indices, monthly time series based on portfolios rebalanced annually. To limit the number of dependent variables (and the potential for test distortions, reported later), we use one return for Decile 1 portfolio and one for Decile 10 portfolio, respectively corresponding to the largest and smallest companies. In all cases, 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 - 2003:12, are in Table 1. The starting period of the data series is determined by availability of the consumption

[5]We use the econometrics software Eviews 3.1 for all our calculations. The relevant code is available upon request.

series (defined below). We further split the sample in two sub-samples, 1959:02-1982:10 and 1982:11-2003:12, enabling us to examine stability of regression parameters. October 1982 marks the approximate ending of the Paul Volcker's war on inflation in the early eighties. In all considered sample periods, the excess return on small caps tends to be more volatile, in accord with Malkiel and Xu (1997). A comparison of the sample means for excess returns reveals that the excess return on the large capitalization portfolio is greater than the excess return on the small-cap portfolio (by 9.93% annually) but the gap is much smaller in the second sub-sample (2.07% annually), consistent with Fama and French (1993) and Horowitz, at al. (2000) (in fact, for the Cap based portfolios, the large firms have actually outperformed the small ones in the second sub-sample). To confirm the impression based on a simple comparison we also conduct formal t-tests for differences in means, which account for the covariance between the two portfolios. The t-statistics and p-values are respectively 2.29/0.02 overall, 2.43/0.02 for the first sub-sample, and 0.59/0.55 for the second sub-sample, which substantiates our conclusions.

As covariates in the model (see Table 1 for summary statistics of covariates, 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, Roll and Ross (1986) and use data on the stock market, bond market, the business cycle and inflation, and 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 Timmermann, 2000).

To describe the stock market we use 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 (see Table 2), and since the large-cap firms account for most of the market value, this is not surprising.[6]

[6]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 (denoted 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. The long-term government bond return-to-maturity is from the 5-year Treasury Bonds, obtained from the web site of the Board of Governors of the Federal Reserve System (BGFRS). 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 (g_{IP}) and real per capita consumption (g_{CONS}). We obtain industrial production data (series INDPRO, seasonally adjusted) the BGFRS web site, 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 Bureau of Economic Analysis.

The consumer price index is used to as an inflation variable. Since the null hypothesis of the unit root cannot be rejected in some sub-samples of this series, we use the first difference in our analysis. For money growth, we use the growth rate of the seasonally adjusted monetary base (g_{MON}), obtained from the St. Louis Fed's web site (series AMBSL, seasonally adjusted).

V. Simulation

In this section, we attempt to find the limit of a sensible employment of the HAC methods to see exactly what level of model complexity they can handle. We use computer simulation, based on a calibrated model of asset returns, to assess test performance. Of interest are rejection rates under the null hypothesis and under the alternative. If the nominal distribution (F distribution for the F test, chi square distribution for the HAC tests) is an accurate approximation then the tests should reject under H_0 at a rate near the theoretical test size; otherwise, the tests will exhibit noticeable distortions.

To set up the simulation, we define a first-order vector autoregressive (VAR) process for covariates x_t:

$$x_t = c + \Phi\, x_{t-1} + u_t, \tag{7}$$

where c is a $K \times 1$ vector of constants, Φ is an $K \times K$ matrix of coefficients, and u_t is a $K \times 1$ vector of random variables which are independent over time and normally distributed with zero mean and cross-sectional variance-covariance matrix Λ.

To see what range of values might be realistic for the parameters of the x_t process, we estimate (7) for $K = 4$ by OLS using $x_t = (r_{VWNY}, r_{TERM}, g_{CONS}, g_{MON})'$. Estimates of elements the matrix Φ range from -0.30 to 0.43. We also try several other combinations of explanatory variables and while estimates differ to a large extent, the diagonal elements tend to be greater then offdiagonal ones, which are often close to 0. Therefore, for our simulation we set $\Phi_{ij} = 0.10$ for $i = j$ and $\Phi_{ij} = 0$ for $i \neq j$. Estimates of the constant term tend to be small relative to elements of Φ, 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 much smaller, hence we let Λ be a diagonal matrix with each diagonal entry equal to 0.0001.

For the regression errors ε_{it} in (2), we posit a dynamic model with serial correlation and generalized autoregressive conditional heteroskedasticity (GARCH), as follows:

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

with η standard normal noise. Parameter ψ_1 specifies the autocorrelation, and parameters ψ_2 and ψ_3 specify the conditional heteroskedasticity. We choose ψ so that the autocorrelation of the error term ε_{it}, as well as its variance relative to that of x's, 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 $\psi_1 = .1$, $\psi_2 = .003$ and $\psi_3 = .2$. The cross-sectional empirical covariance of η_{it} is sometimes positive and sometimes negative, and we specify the population covariance between η_{1t} and η_{2t} to be 0.

To get a sense for the behavior of the F test and 'robust' HAC tests, we first generate results for the case $n = 2$, with $K = 2$ and, alternatively, $K = 4$, using 500 simulated time series for r_{it}, $i = 1, 2$, with 250 and 500 observations, roughly corresponding to one half of our sample and the whole sample of our historical monthly data, respectively. We conduct a Monte-Carlo experiment based on a calibrated model, rather than a bootstrap method as in Ferson and Foerster (1994), for two reasons: First, the calibrated model allows us to identify the source of test success or failure; second, the regression errors have posited dynamics which would not be replicated by standard bootstrap sampling.[7] 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 of cross-equation equality for all coefficients, e.g. the case where the restriction defining matrix D in Section II equals the $p \times p$ identity matrix. We calibrate all β values to equal to 1, and all α values to equal 0. Our simulations show a serious tendency for distortion in most but not all tests. Specifically, the F test and the HAC Wald tests over-reject[8], and the two of the Hansen tests (Newey-West and Newey-West with pre-whitening) under-reject the null hypothesis. On the other hand, three of the Hansen tests (VARHAC, Andrews and Andrews-Monahan) show

[7] Alternatively, one could employ a block-bootstrap method, as in Cochrane (2001, Ch. 15).
[8] For similar results see Cushing and McGarvey (1999) and Cochrane (2001, Ch. 15).

minimal distortion, and of these three the VARHAC test is by far the simplest to compute and interpret. We have examined the Hansen VARHAC test in numerous other simulations exercises: For $n = 2$, we gradually increase the number of covariates K by two up to $K = 8$, and rejection rates fall toward 0.03 and 0.04 for sample sizes 250 and 500. Since many studies consider decile indices, we also look at $n = 10$ and increase the number of covariates from two to eight, in which case the rejection rates for the VARHAC Hansen tests are respectively 0.01 and 0.02 for the two sample sizes. For no other test method do we find less distortion than for the VARHAC Hansen test, and our results suggest that a researcher attempting to investigate the relationship between various variables and asset returns is 'safer' when the number of assets is smaller since the asymptotic and finite sample distributions of the test statistic are closer.

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

$$\begin{aligned} r_{1t} &= x_{1t} + x_{2t} + \ldots + x_{Kt} + \varepsilon_{1t}, \\ r_{2t} &= x_{1t} + x_{2t} + \ldots + x_{\frac{K}{2},t} + (1 + \tfrac{0.2}{K})(x_{(\frac{K}{2}+1),t} + x_{(\frac{K}{2}+2),t} + \ldots + x_{Kt}) + \varepsilon_{2t}. \end{aligned}$$

Table 4 reports rejection rates under the alternative hypothesis for $K = 2$ and $K = 4$, with relatively high rejection rates for the F test, and with higher rejection rates for the HAC Wald test than for the corresponding HAC Hansen test. Among the HAC Hansen tests, the Andrews, Andrews-Monahan and VARHAC methods reject more frequently than the others. These results describe the frequency with which an economist would correctly reject the null hypothesis, using the nominal (F or chi square) distribution of the relevant statistic. A related, but different, issue is the frequency of correct rejection for an economist who knows and uses the exact test distribution. The latter power calculations are not interesting here because the economist does not know the exact distribution, and it is impossible to concisely report on this distribution in a way that would be broadly useful for asset return regression.

We have nevertheless done such power calculations, with the same rankings described above, for the various tests.

Overall, the simulations reveal some serious problems with the F test and with the 'robust' HAC Wald tests, in terms of over-rejection under the null hypothesis, whereas three of the HAC Hansen tests avoided serious distortions and were also best among Hansen tests under the alternative hypothesis. Among these favored three we recommend the VARHAC Hansen test, with it's simple, parametric pre-whitening approach to serial correlation adjustment. For the range of sample sizes under study, the VARHAC Hansen test performance under null and alternative hypotheses suggests that for a small number of assets, $n = 2$, we can have as many as 8 covariates and still avoid major test distortions. In cases of $n = 10$ asset returns, the number of covariates in a restricted econometric model should be kept small, perhaps no more than 4 or 6. In cases where larger models and a greater number of restrictions are desired, larger sample sizes (weekly rather than monthly data, for example) may be necessary for satisfactory results.

VI. Empirical results

Having scrutinized a variety of test methods, we turn now to the problem of testing for differences in sensitivity among firms of different size. As our simulations warn against the use of overly large models, we only use up to seven explanatory variables in our two-asset model. To save space we report only the Hansen-type tests with parametric VARHAC adjustment for residual serial correlation and heteroskedasticity, as these tests showed relatively little distortion in simulation, and are generally in agreement with the other tests for the models we analyze. The tests are formulated by defining the matrix D in Section II accordingly.

We first examine the CAPM. Table 5 gives results for the full monthly sample (1959-2003) and two sub-samples. The test of equality of market betas suggests significant difference in risk exposure, for large and small firms, in both sub-samples but not overall. This is a result of changing beta for small firms, which is 1.28 in the first but only 0.81 in the second sub-sample. We also test for whether the intercepts are each 0; since the only factor in the CAPM is the market excess return, the test of zero intercepts is essentially a HAC robust version of the standard F-test commonly applied in testing the CAPM.[9] The CAPM is rejected overall and in the first sub-sample but not in the second sub-sample.

In our discussion of the results, we will focus on the two sub-samples since the results for the whole time period are (more) likely to reflect changing betas over time. The theory, on which CAPM is based, offers the following interpretation. Higher betas in the first sub-sample are not sufficient to explain higher mean returns for the same period. The CAPM is rejected and the intercept for small firms is significantly positive, suggesting that investment in small firms delivers a premium higher than accounted for by the CAPM. This is in fact the so called firm size effect. In this case, statistically different betas do not provide us with information that makes a difference. The recommendation here is clear: invest in small firms. The theory of CAPM gives no such recommendation for the second sub-sample, where any differences in expected returns are simply due to differing market betas. However, statistically different betas (at 10% level of significance) are in contrast with mean returns, which do not differ for small and large firms from a statistical point of view. In other words, the mean returns are approximately the same but the large firms are riskier (recall that small firms' beta is now 0.81). Using this interpretation, the recommendation is again clear: buy stocks of small firms.

We next examine bivariate models, with covariates given by the market return and one of the remaining five economic variables, with results reported in Table 6. The properties of the market betas are not changed with addition of another covariate, i.e. the market slopes differ in the first sub-sample as well as in the second one. Sensitivity to the second covariate shows

[9]As pointed out in Gibbons, Ross, and Shanken (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 F, S or W) can then be interpreted as a measure of distance from the mean-variance frontier.

in each case no significant differences in the first sub-sample. In the second sub-sample, there are now significant differences in slopes for the default premium and consumption. We also conduct the joint test for zero intercepts, which can be loosely interpreted as a test of our asset pricing model[10] In the first period, three out of six models have non-zero intercepts and all have either an insignificant or positive intercept for small firms. In the second period, only one out of six indicates non-zero intercepts (the default premium being the second variable), in this case with a positive intercept for large firms. The recommendation for the period from 1959:02 till 1982:10 stands: statistical differences in mean returns are driven by differences in market betas and there is extra premium on investment in small firms, which dominates investment in large ones. For the second period, the situation has become more complex. We have statistically undistinguishable mean returns and statistically differing market betas. The smaller market beta for small firms indicates a better investment opportunity. However, we also have other variables whose beta differ, namely the default premium and consumption. The results with default premium even indicate that it might be the large firms, which have become a bargain. So, while performance of the two portfolios is similar in the second sub-sample there are significant differences in risk exposure between the large and small firms in addition to previously identified differences in market betas. These differences indicate that returns on small firms' stock is more sensitive (i.e. riskier) to variables other than the market excess return.

Results from bivariate models call for a model with more explanatory variables. We examine the model in which all seven covariates are included at once. Table 7 reports parameter estimates and their standard errors, computed via the VARHAC method.[11] To further describe the model we report in Table 8 residual diagnostic tests. As indicated, there

[10] Loosely because not all the variables can be interpreted as asset returns.

[11] The reported estimates are also in accord with other studies on firm-size effects, which use multi-factor models - see Fama and French (1993, market beta, betas for the default and term premia), Chan, Chen, and Hsieh (1985, all but the money supply beta), and Li and Hu (1998, industrial production and money supply betas).

is strong evidence of both residual heteroskedasticity and autocorrelation, in which case our use of HAC test methods is highly appropriate.[12] Finally, Table 9 reports results of tests of cross-section restrictions.

Table 7 indicates that market betas still differ in both sub-periods and the point estimates are similar to those in the uni- and bivariate regressions. For other sensitivities, with the exception of the monetary slope, the differences in the point estimates have grown in the second sub-sample. The test for zero intercepts interestingly indicates non-zero intercepts in the second rather then the first sub-period. Table 9 shows slopes for the default premium and consumption betas statistically different in the period from 1982:11 to 2003:12 and they are joined, for this sub-sample, by sensitivities to the term premium, inflation and money supply. Consumption betas differ also for the first period. In addition, a joint test of beta equality for all but the market variables results in rejection of the null in the second sub-sample.

We will now attempt to interpret our results. Let us start with the period from 1959:02 to 1982:10. Risk measured by betas is different mainly due to differences in the market beta, which partly explains differing mean returns. Stocks of small firms appear underpriced, and hence are a better investment opportunity. The situation is more complex in the period from 1982:11 to 2003:12 where the recommendation based solely on market betas is essentially reversed when one uses additional factors. While investment in small firms appears less risky based on market betas, it is riskier as measured by other betas. Being underpriced, the large firms seem to be the bargain here.

VII. Conclusion

As finance theory suggests, differing sensitivities (betas) of economic factors translate in differing asset performance (unless they affect expected returns in opposite directions). While this connection is widely recognized, almost no attention has been paid to *formal*

[12] We found similarly strong evidence of conditional heteroskedasticity and serial correlation in a majority of the univariate and bivariate models we studied.

differences in betas. The present work explicitly acknowledges this link by testing for statistical differences of betas across assets and by considering implications of these differences for mean excess returns.

We first provide a general regression framework, which can be easily used to test for equality of betas across equations. A number of methods can be used to conduct such tests. We consider the standard F, Wald, and Hansen tests. The advantage of the Wald and Hansen methods is relatively simple accommodation of robustness to general forms of autocorrelation and heteroskedasticity, often present in the financial data. The price for generality in this case is potential distortion of the tests in larger regression systems. In a simulation exercise tailored to our data application, we find that finite sample distortions are relatively minor and that the Hansen method with parametric pre-whitening outperforms the other methods.

We illustrate the usefulness of formal comparison of betas in application to stocks sorted by firm capitalization. Namely, a simple t-test indicates that small firms outperformed the larger firms prior to 1982:10 but not since then. We attempt to shed some light on this empirical observation by carefully analyzing betas of several economic factors. We find that the market beta difference is the main source of differing mean returns before 1982:10. However, while the market beta differ also since 1982:10, the mean returns do not. Moreover, it is the small firms, which appear safer. Testing of beta equality of factors other than the market reveals the reason behind this seeming inconsistency. The other sensitivities (especially to the default premium) also differ but are higher (in absolute value) for the small firms, making them riskier from this perspective.

Overall, our empirical analysis suggests that formally comparing, individually and jointly, market betas and betas of other macroeconomic variables can be helpful in explaining behavior of expected returns and can lead to investment recommendations. While individual statistical comparison of betas may be redundant at times, joint comparisons are useful in any case as they summarize information contained in a number of point beta estimates. They can be combined with mean returns, regression intercepts and point beta estimates to form a clearer basis for judging investment opportunities. A by-product of our calculations is confirmation of the now widely accepted need for multi-factor models. Default premium and consumption growth seem to be two important sources of risk differences other than the market.

There are several directions for future research. As our results indicate, any two assets or a group of assets can be compared by formally testing equality of betas across equations. For instance, one could revisit the influential Fama and French (1993) paper to evaluate whether the stocks sorted by size and book-to-market ratios really differ as the point beta estimates suggest. The importance of formal comparison of betas across assets also leads to the possibility of uncertainty in betas being priced by the market. This possibility could be investigated by the Fama and MacBeth (1973) method using some measure of uncertainty in the market beta as one of the factors in the time-series regression. One such measure could be the difference in publicly available beta estimates. The cross-sectional regression would then reveal whether this factor is priced or not.

References

Andrews, D.W.K. (1991).‘ Heteroskedasticity and autocorrelation consistent covariance matrix estimation,’ *Econometrica* 59, 817-858.

Andrews, D.W.K. and Monahan, J.C. (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.

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

Chan, K.C. and Chen, N. (1991). ‘Structural and return characteristics of small and large firms,’ *The Journal of Finance* 46, 1467-1484.

Chan, K.C., Chen, N. and Hsieh, D.A. (1985). ‘An exploratory investigation of the firm size effect,’ *Journal of Financial Economics* 14, 451-471.

Chen, N., Roll, R. and Ross, S.A. (1986). ‘Economic forces and the stock market,’ *Journal of Business* 59, 383-403.

Cochrane, J.H. (1999). ‘New facts in finance,’ Working Paper 7169, NBER.

Cochrane, J.H. (2001). *Asset Pricing* (Princeton University Press: Princeton, NJ).

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

Cushing, M.J. and McGarvey, M.G. (1999). ‘Covariance matrix estimation,’ in L. Mátyás, ed., *Generalized Method of Moments Estimation*, Cambridge University Press: Cambridge, MA.

den Haan, W. and Levin, A. (1996). 'Inferences from parametric and non-parametric covariance matrix estimation procedures,' Working paper t0195 , NBER.

den Haan, W. and Levin, A. (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, Holland.

Fama, E.F. and French, K.R. (1992). 'The cross section of expected stock returns,' *The Journal of Finance* 47, 427-465.

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

Fama, E.F. and French, K.R. (1996). 'Multifactor explanations of asset pricing anomalies,' *The Journal of Finance* 51, 55-84.

Fama, E. F. and Gibbons, M.R. (1984). 'A comparison of inflation forecasts,' *Journal of Monetary Economics* 13, 327-348.

Fama, E.F. and MacBeth, J.D. (1973). 'Risk, return, and equilibrium: Empirical tests,' *Journal of Political Economy* 81, 607-636.

Ferson, W.E. and Foerster, S. (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.E. and Harvey, C. (1999). 'Conditioning variables and the cross section of stock returns,' *The Journal of Finance* 54, 1325-1357.

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

French, K.R., Schwert, G.W., and Stambaugh, R.S. (1987). 'Expected stock returns and volatility,' *Journal of Financial Economics* 19, 3-29.

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

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

Hansen, L.P. (1982). 'Large sample properties of the generalized method of moments estimator,' *Econometrica* 50, 1029-1054.

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

He, J., Kan, R., Ng, L. and Zhang, C. (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.L., Loughran, T. and Savin, N. E. (2000). 'Three analyses of the firm size premium,' *Journal of Empirical Finance* 7, 143-153.

Lehmann, B.N. and Modest, D.M. (1988). 'The empirical foundations of the arbitrage pricing theory,' *Journal of Financial Economics* 21, 213-254.

Li, L. and Hu, Z.F. (1998). 'Responses of the stock market to macroeconomic announcements across economic states,' Working paper, International Monetary Fund.

Lintner, J., (1965). 'The valuation of risky assets and the selection of risky investments in stock portfolios and capital budgets,' *Review of Economics and Statistics* 47, 13-37.

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

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

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

Merton, R.C. (1973). 'An intertemporal capital asset pricing model,' *Econometrica* 41, 867-887.

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

Newey, W.K. and West, K.D. (1994). 'Automatic lag selection in covariance matrix estimation,' *Review of Economic Studies* 61, 631-653.

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

Ross, S.A. (1976). 'The arbitrage pricing theory of capital asset pricing,' *Journal of Economic Theory* 13, 341-360.

Schwert, G.W. (1983). 'Size and stock returns, and other empirical anomalies,' *Journal of Financial Economics* 12, 3-12.

Shanken, J. (1992). 'On the estimation of beta pricing models,' *Review of Financial Studies* 5, 1-34.

Sharpe, W. (1964). 'Capital asset prices: A theory of market equilibrium under conditions of risk,' *The 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.

TABLE 1

Summary Statistics

	r_{SMALL}	r_{LARGE}	r_{VW}	r_{DEF}	r_{TERM}	g_{IP}	g_{CONS}	π	g_{MON}
				1959:02-2003:12					
Mean	11.22	4.98	5.67	23.76	3.15	3.16	2.02	4.12	6.60
Median	8.25	6.83	9.32	22.68	2.39	3.52	2.16	3.57	6.48
Max	694.93	210.29	195.37	57.36	113.86	71.98	21.63	21.53	47.03
Min	-367.96	-241.97	-266.66	1.68	-86.59	-43.36	-21.58	-6.58	-32.16
St.Dev.	78.56	49.92	50.98	10.87	19.48	9.81	5.24	3.65	5.49
Skewness	-0.29	1.20	-0.39	0.38	0.22	0.03	-0.18	0.97	0.67
Kurtosis	4.73	15.61	4.98	2.62	6.58	9.67	4.54	4.64	18.38
				1959:02-1982:10					
Mean	11.84	1.91	3.21	18.29	0.74	3.22	1.95	5.13	5.98
Median	6.86	4.63	4.69	15.84	0.57	3.60	2.29	3.99	6.08
Max	694.93	210.29	195.37	47.16	113.86	71.98	21.63	21.53	23.48
Min	-321.65	-157.32	-145.90	1.68	-86.59	-43.36	-21.58	-3.90	-5.56
St.Dev.	89.24	49.51	51.84	9.50	21.11	12.07	6.01	4.26	3.96
Skewness	1.61	0.06	0.00	0.74	0.47	0.00	-0.17	0.69	0.01
Kurtosis	15.40	4.25	3.90	2.70	7.98	7.77	3.95	3.16	4.04
				1983:11-2003:12					
Mean	10.51	8.44	8.43	29.90	5.85	3.09	2.10	2.99	7.30
Median	9.77	10.93	12.20	28.56	5.81	3.51	2.00	2.94	7.09
Max	291.21	156.09	149.07	57.36	56.35	23.83	16.03	11.35	47.03
Min	-367.96	-241.97	-266.66	13.92	-40.34	-14.49	-14.78	-6.58	-32.16
St.Dev.	64.68	50.24	49.95	8.84	17.12	6.42	4.22	2.34	6.74
Skewness	-0.19	-0.67	-0.87	0.72	-0.06	0.15	-0.10	-0.20	0.56
Kurtosis	9.53	5.48	6.58	2.97	3.06	3.31	4.71	5.14	16.29

Notes: r_{SMALL} and r_{LARGE} denote respectively the excess returns on the small-cap and large-cap portfolios, r_{VW} is the excess return on the market portfolio, r_{DEF} and r_{TERM} are the default and risk premium, respectively, g_{IP} and g_{CONS} are growth rates of industrial production and per capita consumption, respectively, π measures the inflation rate and g_{MON} the growth rate of the money supply, respectively. All reported numbers are annualized, in percentages.

TABLE 2

Correlations

		r_{SMALL}	r_{LARGE}
1959:02-2003:12	r_{VW}	0.69	0.99
	r_{DEF}	0.16	0.12
	r_{TERM}	0.11	0.19
	g_{IP}	-0.01	0.00
	g_{CONS}	0.20	0.16
	π	-0.12	-0.18
	g_{MON}	-0.07	-0.01
1959:02-1982:10	r_{VW}	0.74	0.98
	r_{DEF}	0.21	0.21
	r_{TERM}	0.18	0.22
	g_{IP}	0.05	0.07
	g_{CONS}	0.20	0.18
	π	-0.13	-0.19
	g_{MON}	-0.05	-0.01
1982:11-2003:12	r_{VW}	0.63	0.99
	r_{DEF}	0.19	-0.03
	r_{TERM}	-0.01	0.13
	g_{IP}	-0.18	-0.14
	g_{CONS}	0.22	0.14
	π	-0.12	-0.13
	g_{MON}	-0.10	-0.03

Notes: See notes in Table 1 for variable definitions.

TABLE 3

Rejection Rates Under the Null Hypothesis

				Covariance Matrix Estimator					
K	Sample Size	Test		WH	NW	NW-P	A	AM	VARHAC
2	250	F	0.08						
		Hansen		0.07	0.03	0.02	0.05	0.05	0.05
		Wald		0.08	0.08	0.07	0.07	0.07	0.07
2	500	F	0.07						
		Hansen		0.07	0.05	0.05	0.06	0.05	0.05
		Wald		0.08	0.08	0.07	0.07	0.06	0.06
4	250	F	0.07						
		Hansen		0.05	0.02	0.01	0.04	0.04	0.04
		Wald		0.10	0.13	0.13	0.10	0.08	0.08
4	500	F	0.06						
		Hansen		0.05	0.03	0.03	0.04	0.04	0.04
		Wald		0.06	0.08	0.08	0.06	0.05	0.05

Notes: We simulate 500 times the series $r_{it} = x_{1t} + x_{2t} + \ldots + x_{Kt} + \varepsilon_{it}, i = 1, 2, t = 1, ..., T, T = 250$ or 500, $K = 2$ or $K = 4$. $x_{jt} = 0.002 + 0.10\, x_{j,t-1} + u_{jt}$, $j = 1, 2, \ldots, K$, where $u_{1t}, ..., u_{Kt}$ are mutually independent and i.i.d. normally distributed with zero mean and variance 0.0001. $\varepsilon_{it} = 0.1\varepsilon_{i,t-1} + 0.003\times \sqrt{1 + 0.2\, \varepsilon^2_{i,t-1}}\eta_{it}$, $i = 1, 2$, with η standard normal noise. We test the null hypothesis of equality of all parameters across the two assets.

TABLE 4

Rejection Rates Under the Alternative Hypothesis

				Covariance Matrix Estimator					
K	Sample Size	Test		WH	NW	NW-P	A	AM	VARHAC
2	250	F	0.89						
		Hansen		0.87	0.78	0.77	0.85	0.84	0.84
		Wald		0.90	0.89	0.89	0.88	0.87	0.87
2	500	F	1.00						
		Hansen		0.99	0.99	0.99	0.99	0.99	0.99
		Wald		1.00	1.00	1.00	1.00	1.00	1.00
4	250	F	0.58						
		Hansen		0.51	0.26	0.26	0.43	0.48	0.48
		Wald		0.61	0.62	0.61	0.61	0.58	0.58
4	500	F	0.89						
4	500	Hansen		0.86	0.76	0.74	0.82	0.84	0.84
		Wald		0.89	0.87	0.87	0.88	0.88	0.88

Notes: We simulate 500 times series

$$\begin{aligned} r_{1t} &= x_{1t} + x_{2t} + \ldots + x_{Kt} + \varepsilon_{1t}, \\ r_{2t} &= x_{1t} + x_{2t} + \ldots + x_{\frac{K}{2},t} + (1 + \tfrac{0.2}{K})(x_{(\frac{K}{2}+1),t} + x_{(\frac{K}{2}+2),t} + \ldots + x_{Kt}) + \varepsilon_{2t}, \end{aligned}$$

$t = 1, 2, \ldots, T$, $T = 250$ or 500, $K = 2$ or $K = 4$. $x_{jt} = 0.002 + 0.10\, x_{j,t-1} + u_{jt}$, $j = 1, 2, \ldots, K$, where $u_{1t}, ..., u_{Kt}$ are mutually independent and i.i.d. normally distributed with zero mean and variance 0.0001. $\varepsilon_{it} = 0.1\, \varepsilon_{i,t-1} + 0.003\sqrt{1 + 0.2\, \varepsilon_{i,t-1}^2}\, \eta_{it}$, $i = 1, 2$, with η standard normal noise. We test the null hypothesis of equality of all parameters across the two assets.

TABLE 5

Tests of the CAPM

Hypothesis	Years	r_{VW}
equal slopes	59:02-03:12	1.16
		(0.28)
	59:02-82:10	5.32
		(0.02)
	82:11-03:12	3.25
		(0.07)
zero intercepts	59:02-03:12	5.07
		(0.08)
	59:02-82:10	5.73
		(0.06)
	82:11-03:12	1.94
		(0.38)

Notes: The model is $r_{it} = \alpha_i + \beta_i x_t + \varepsilon_{it},\ i = 1, 2$, where α_i is the i-th intercept, β_i is the i-th slope, $r_{1t} = r_{SMALL}$, $r_{2t} = r_{LARGE}$, x_t is r_{VW} (see notes for Table 1 for variables' definitions) and ε_{it} is the regression error. Reported are HAC Hansen statistics (VARHAC method) for testing equality of slopes and zero values for intercepts. p-values are in parentheses.

TABLE 6

Tests of Assorted Bivariate Models

Hypothesis	Years	independent variable (in addition to market return)					
		r_{DEF}	r_{TERM}	g_{IP}	g_{CONS}	$\Delta\pi$	g_{MON}
equal slopes	59:02-03:12	0.83	1.23	1.16	0.62	1.17	1.15
(market)		(0.36)	(0.27)	(0.28)	(0.43)	(0.28)	0.28
	59:02-82:10	5.28	4.90	5.09	4.63	5.15	5.04
		(0.02)	(0.03)	(0.02)	(0.03)	(0.02)	(0.02)
	82:11-03:12	3.34	2.71	4.02	3.90	3.04	3.59
		(0.07)	(0.10)	(0.04)	(0.05)	(0.08)	(0.06)
equal slopes	59:02-03:12	1.82	3.13	0.06	7.67	0.00	0.68
(other)		(0.18)	(0.08)	(0.81)	(0.01)	(0.98)	(0.41)
	59:02-82:10	0.64	0.01	0.04	0.80	0.19	0.02
		(0.42)	(0.92)	(0.85)	(0.37)	(0.66)	(0.88)
	82:11-03:12	13.42	2.61	1.52	5.15	0.04	0.27
		(0.00)	(0.11)	(0.22)	(0.02)	(0.84)	(0.60)

Notes: The model is $r_{it} = \alpha_i + \beta_i x_t + \varepsilon_{it},\ i = 1, 2$, where α_i is the i-th intercept, β_i is the i-th vector of slopes, $r_{1t} = r_{SMALL}$, $r_{2t} = r_{LARGE}$, elements of x_t are r_{VW} and of one the following variables: r_{VW}, r_{DEF}, r_{TERM}, g_{IP}, g_{CONS}, $\Delta\pi$, g_{MON} (see notes for Table 1 for variables' definitions) and ε_{it} is the regression error. Reported are HAC Hansen statistics (VARHAC method) for testing equality of slopes. p-values are in parentheses.

TABLE 7

Estimated Model with Seven Covariates

		independent variable							
Years	Size	Intcpt.	r_{VW}	r_{DEF}	r_{TERM}	g_{IP}	g_{CONS}	$\Delta\pi$	g_{MON}
59:02-03:12	small	-0.005	1.036	0.660	-0.185	-0.097	1.442	1.139	-1.058
		(0.005)	(0.079)	(0.248)	(0.138)	(0.269)	(0.499)	(0.713)	(0.608)
	large	0.000	0.973	-0.081	-0.005	0.047	-0.156	0.123	0.171
		(0.001)	(0.011)	(0.036)	(0.022)	(0.035)	(0.071)	(0.107)	(0.067)
59:02-82:10	small	0.003	1.247	0.584	-0.125	0.014	0.985	0.643	-1.342
		(0.007)	(0.126)	(0.495)	(0.209)	(0.304)	(0.548)	(0.874)	(1.870)
	large	0.000	0.950	-0.048	-0.058	0.036	-0.109	0.071	0.007
		(0.001)	(0.018)	(0.063)	(0.030)	(0.041)	(0.088)	(0.141)	(0.206)
82:11-03:12	small	-0.022	0.778	1.260	-0.387	-1.042	2.306	2.554	-0.909
		(0.010)	(0.099)	(0.405)	(0.232)	(0.759)	(1.071)	(1.031)	(0.484)
	large	0.004	1.000	-0.219	0.067	0.096	-0.218	0.116	0.201
		(0.001)	(0.011)	(0.059)	(0.027)	(0.064)	(0.119)	(0.154)	(0.058)

Notes: The estimated model is: $r_{it} = \alpha_i + \beta_i x_t + \varepsilon_{it},\ i = 1, 2$, where α_i is the i-th intercept, β_i is the i-th vector of slopes, $r_{1t} = r_{SMALL}$, $r_{2t} = r_{LARGE}$, $x_t = (r_{VW}, r_{VW}, r_{DEF}, r_{TERM}, g_{IP}, g_{CONS}, \Delta\pi, g_{MON})'$ (see notes for Table 1 for variables' definitions) and ε_{it} is the regression error. Reported are OLS estimates of the model parameters with VARHAC standard errors in parentheses.

TABLE 8

Tests for Residual Heteroskedasticity and Correlation

Residual Property		Size	Test	from 59:02 to 03:12	59:02 82:10	82:11 03:12
correlation	across equations		Pearson	–0.70	-0.70	-0.57
				(0.14)	(0.02)	(0.03)
	across time	small	Q	98.11	66.42	45.47
				(0.00)	(0.00)	(0.00)
		large	Q	41.47	25.14	14.80
				(0.000)	(0.01)	(0.25)
heteroskedasticity		small	White	4.61	5.88	3.35
				(0.00)	(0.00)	(0.00)
		large	White	4.77	6.17	2.17
				(0.00)	(0.00)	(0.01)

Notes: The estimated model is: $r_{it} = \alpha_i + \beta_i x_t + \varepsilon_{it},\ i = 1, 2$, where α_i is the i-th intercept, β_i is the i-th vector of slopes, $r_{1t} = r_{SMALL}$, $r_{2t} = r_{LARGE}$, $x_t = (r_{VW},\ r_{VW},\ r_{DEF},\ r_{TERM},\ g_{IP},\ g_{CONS},\ \Delta\pi,\ g_{MON})'$ (see notes for Table 1 for variables' definitions) and ε_{it} is the regression error. Residuals are calculated using OLS estimates, equation by equation ; Pearson = chi-square test for correlation; White test = F test with no cross terms; Q = Q statistic for testing 12 lags of autocorrelation; p-values in parentheses.

TABLE 9

Tests of Seven Covariate Model

	period		
Hypothesis	1959:02 2003:12	1959:02 1982:10	1982:11 2003:12
equal slopes:			
r_{VW}	0.51	3.84	4.78
	(0.48)	(0.05)	(0.03)
r_{DEF}	6.84	1.16	10.57
	(0.01)	(0.28)	(0.00)
r_{TERM}	1.36	0.08	2.81
	(0.24)	(0.78)	(0.09)
g_{IP}	0.24	0.00	1.88
	(0.62)	(0.95)	(0.17)
g_{CONS}	7.77	3.01	3.59
	(0.01)	(0.08)	(0.06)
$\Delta\pi$	1.86	0.37	4.39
	(0.17)	(0.54)	(0.04)
g_{MON}	2.99	0.40	3.52
	(0.08)	(0.53)	(0.06)
equal slopes:	14.99	5.60	20.89
(all but mkt.)	(0.02)	(0.47)	(0.00)

Notes: The model is $r_{it} = \alpha_i + \beta_i x_t + \varepsilon_{it},\ i = 1, 2$, where α_i is the i-th intercept, β_i is the i-th vector of slopes, $r_{1t} = r_{LARGE}$, $r_{2t} = r_{SMALL}$, x_t = $(r_{VW}, r_{DEF}, r_{TERM}, g_{IP}, g_{CONS}, \pi_{UI}, g_{MON})'$ (see notes for Table 1 for variables' definitions) and ε_{it} is the regression error. Reported are respectively statistics for the Hansen tests (VARHAC method) of equality of slopes across equations for a given covariate and of equality of all slopes with the exception of the market. p-values are in parentheses.