

Estimating the Cost of Capital Through Time: An Analysis of the Sources of Error

Wayne E. Ferson • Dennis H. Locke

*Department of Finance and Business Economics, Box 353200, University of Washington, Seattle, Washington 98195
Moss Adams Advisory Services L.L.P., 1001 Fourth Avenue, Suite 2700, Seattle, Washington 98154-1199*

Practitioners needing estimates of a firm's equity cost of capital have long relied on the Capital Asset Pricing Model (CAPM). Recent evidence casts renewed doubt on the validity of the CAPM and beta. However, there is not much evidence to gauge the importance of the rejections of the CAPM in a practical decision-making context. This paper presents evidence on the sources of error in estimating required returns over time. We use a number of proxies for the true mean variance efficient portfolio, allowing that the CAPM is the "wrong" model. The analyst is assumed to rely on a standard market index. We find that the great majority of the error in estimating the cost of equity capital is found in the risk premium estimate, and relatively small errors are due to the risk measure, or beta. This suggests that analysts should improve estimation procedures for market risk premiums, which are commonly based on historical averages. This can be done by using regression models, such as have appeared in the recent finance literature, or by purchasing forecasts from firms that specialize in producing them.

(Cost of Capital; Capital Budgeting; Risk Premium; Beta; Capital Asset Pricing Model; Forecasting; Business Valuation)

1. Introduction

Practitioners needing estimates of a firm's equity cost of capital, or expected rate of return, have long relied on the Capital Asset Pricing Model (CAPM, Sharpe 1964) as a framework for producing such estimates. The model is appealing because it has a strong theoretical basis and is relatively easy to use. However, recent evidence, notably the work of Fama and French (1992) and others, has cast renewed doubt on the validity of the CAPM and its famous beta coefficient. If beta is "dead," then its death has left many practitioners in the lurch, as no theoretically justified and easily implementable way to estimate the cost of capital seems to be waiting in the wings.¹

Of course, the fact that the CAPM can be rejected by data should not be surprising. It is a highly stylized theoretical model, and all such models are in some sense "wrong." The important questions, from a practical perspective, are whether the use of a model leads to systematically poor decisions and whether another model can help improve decision making. The answer to these questions is likely to depend on the application to which the model is put.

In this paper, we focus on the problem of estimating a cost of equity capital. This application is important in a number of settings. For example, required return estimates are a crucial input in public utility rate regulation, the valuation of closely held enterprises, and corporate capital budgeting problems. However, there is

¹ One natural successor to the CAPM is the Arbitrage Pricing Theory (APT) of Ross (1976), which replaces the market beta with a number of factor betas. However, studies suggest that such multiple-beta models do not resolve the CAPM's problems in accurately predicting av-

erage expected returns (see Lehmann and Modest 1988 and MacKinlay 1995). It is more difficult to estimate a number of betas and risk premiums, which has limited the appeal of multiple-beta models in practice (see the survey evidence in Bruner et al. 1996).

little evidence on how well or poorly the CAPM actually performs in these settings (Kaplan and Ruback 1995 is one recent exception).

The performance of a model in practice depends on how the model is implemented. An imperfect model, expertly implemented, may lead to better decisions than a superior model that is poorly implemented. The objective of our analysis is therefore to evaluate the relative magnitudes of the various sources of errors that are likely to be made, given current practices for estimating a required return.

Our study is directly motivated by problems such as public utility rate hearings and the valuation of closely held companies by financial advisors. In these applications one wants to establish precise values for a cost of capital, given the state of information at a point in time. These may be considered as "single-project" decisions made repeatedly over time. As such, errors over time are crucial. We therefore focus our analysis on the errors through time, as measured by statistics such as the time-series forecast error variance and mean absolute error. Errors in estimating the market risk premium are likely to be very important in such a context. In contrast, consider an analyst whose objective is to rank multiple projects at a point in time. Since at any point in time the market risk premium is fixed across the projects, errors in the market premium are unlikely to be crucial for project ranking decisions.

Although our focus is required-return errors through time, our results are important for capital budgeting applications that involve project selection. Errors in the risk premium will affect the cost-of-capital estimates for all of the projects under consideration at a given time. If these estimates are used to determine the overall size of the capital budget, then the time-series errors are important. Firms that repeatedly make poor capital funding decisions are likely to pass up profitable investments and take on bad projects, placing themselves at a competitive disadvantage.

Using simulations we find that errors in the cost of equity capital are reduced by about 40 percent, for small stocks and most industries, when the standard practice of using historical averages to estimate the premium on a market index is replaced with a simple estimate using current market indicators. Errors that are likely to arise from using the "wrong" beta have a much smaller im-

pact. This is especially interesting in view of the recent controversies over the CAPM and beta. To some extent, our paper renders these controversies unimportant from the perspective of intertemporal variations in the cost of capital.

The paper proceeds as follows. Section 2 describes the models and our approach to analyzing of the sources of error in required returns. Section 3 describes the data. Sections 4–7 describe the simulation approach. In §8 we present the results. Section 9 is the conclusion. An appendix provides technical details.

2. Modeling the Cost of Capital

Basic portfolio theory describes mean variance efficient portfolios, which maximize expected return for a given level of risk (variance of return). In the CAPM, the "market portfolio" is supposed to be mean-variance efficient in equilibrium. This ex ante mean variance efficiency of the market portfolio justifies using the CAPM in cost-of-capital calculations. However, the evidence (e.g., Fama and French 1992) suggests that standard market indexes are not mean variance efficient. If the CAPM is "wrong," then it is difficult to assess the sources of error in cost of capital estimates, unless the "right" model that determines the cost of capital is known. To address this problem, we use a *latent variables* model. A latent variables model has the advantage that it does not require us to take a stand on what is the correct equilibrium model.

Roll (1977) showed that even when the CAPM fails, expected returns must be explained by betas relative to a mean-variance efficient portfolio. In other words, R_p is the return of a mean-variance efficient portfolio if and only if the following expression holds:

$$\begin{aligned} E(R_i) &= E(\gamma_0) + \beta_{ip}E[R_p - \gamma_0], \\ \beta_{ip} &= \text{Cov}(R_i; R_p) / \text{Var}(R_p), \end{aligned} \quad (1)$$

where $E(R_i)$ is the expected return or cost of capital for security i , $E(\gamma_0)$ is the expected return on a zero-beta portfolio to R_p and $E[R_p - \gamma_0] > 0$ is the risk premium of the mean-variance efficient portfolio. If there is a risk-free asset, with $\beta_{ip} = 0$, then $E(\gamma_0) = R_f$ is the risk-free rate of return. A cost-of-capital estimate should be forward looking; thus, all the expected values, variances, and covariances in Equation (1) should depend on the

information available at the time the expectations are formed.

If we could identify the mean-variance efficient portfolio R_p ex ante, we could use it to estimate required returns. However, no one knows which portfolio is ex ante mean variance efficient.² We, therefore, treat the unobserved mean-variance efficient portfolio's expected return as a latent variable. A simple development of the latent variable model is now presented.

Assume that we have a risk-free asset with return R_f , so that $E(\gamma_0) = R_f$. Consider a market index such as the SP500 and denote its return as R_m . We choose the SP500 as the benchmark premium to be consistent with current practice. According to the latent variable model, the SP500 is just another portfolio that should satisfy Equation (1):

$$\begin{aligned} E(R_m) &= R_f + \beta_{mp}E[R_p - R_f], \\ \beta_{mp} &= \text{Cov}(R_m; R_p) / \text{Var}(R_p). \end{aligned} \quad (2)$$

Now, assuming that β_{mp} is not zero, taking the ratio of the Equations (1) and (2) produces a simple result:

$$\begin{aligned} E(R_i) &= R_f + C_i E[R_m - R_f], \\ C_i &= \text{Cov}(R_i; R_p) / \text{Cov}(R_m; R_p). \end{aligned} \quad (3)$$

Equation (3) provides a general expression for $E(R_i)$, the required return or cost of equity capital. The term C_i is the *C coefficient* for asset i , which is the measure of risk in the latent variable model. Note that the *C* coefficient depends on the unobserved, mean variance efficient portfolio.

Although Equation (3) may look similar to the CAPM, it is important to understand the differences. First the benchmark portfolio, which we choose to be the SP500, is not the "market portfolio" of the CAPM. Second, the C_i coefficient is used to measure risk, instead of the beta coefficient of the CAPM. The evidence that the SP500 is not mean variance efficient implies that the C_i coefficients are not the same as the usual betas. (If the CAPM was correct, in the sense that R_m is the

mean-variance efficient portfolio R_p , then the C_i coefficients reduce to the usual betas.)

As described in subsequent sections, we use the latent variable model (3) as a convenient way to structure our simulations. It is also possible to estimate the model directly to generate estimates of the cost of capital. We, therefore, consider the performance of a hypothetical analyst who adopts this sophisticated approach, even though it is not representative of current practice.

2.1. Sources of Error in Cost of Capital Estimation

Using the latent variable model, we can frame the problem of estimating the cost of capital in a way that makes it easy to break out the sources of error. Three things go into the formula, as given by Equation (3). The first is the risk-free rate of return. Typically, a Treasury market yield is used in practice for R_f . The second input is the expected risk premium on the SP500, denoted by $E(R_m - R_f)$. Finally, the risk measure, C_i , is needed. The latter two inputs are likely to be the most troublesome.³ We therefore focus on errors in the estimates of risk and errors in estimating the SP500 premium.⁴

Define $r_i = R_i - R_f$ as the excess rate of return for asset i . The excess return on the SP500 is $r_m = R_m - R_f$. Equation (3) says that the expected excess return for asset i is $E(r_i) = C_i E(r_m)$. (Working in excess returns, we replace R_i with r_i in the definition of C_i , Equation (3).) We use estimates, denoted by \hat{C}_i and $\hat{E}(r_m)$ to get $\hat{E}(r_i) = \hat{C}_i \hat{E}(r_m)$. The error in estimating the cost of capital can therefore be written as:

$$\begin{aligned} \epsilon_i &= \hat{E}(r_i) - E(r_i) \\ &= \hat{C}_i [\hat{E}(r_m) - E(r_m)] + E(r_m) [\hat{C}_i - C_i]. \end{aligned} \quad (4)$$

Equation (4) expresses the error in the required return for asset i in terms of the error in the risk measure and the error in estimating the SP500 risk premium.

³ To see this, rewrite Equation (3) as $E(R_i) = (1 - C_i)R_f + C_i E(R_m)$. Errors in R_f are likely to have a small impact for two reasons. First, the variance of the errors in measuring R_f is likely to be small in relative terms. Second, when the C_i s are close to 1.0, errors in R_f get relatively small weight. Similar to beta, the weighted average of assets' C_i coefficients is 1.0 (see also Table 2).

⁴ We ignore the impact of alphas, valuation discounts or premiums, or other subjective elements, which obviously could reduce (or increase) the forecast errors. See Pratt (1993) for a practical discussion.

² Most empirical tests of asset pricing models can be thought of as testing whether a portfolio is mean variance efficient (see Huberman, Kandel, and Stambaugh 1987, Grinblatt and Titman 1987, or Shanken 1987 for discussions). So far, the literature has failed to identify the ex ante efficient portfolio, R_p .

Equation (4) can be used to derive the following expression for the error variance.⁵

$$\begin{aligned} \text{Var}(\epsilon_i) = & \{\bar{E}(\hat{C}_i)\}^2 \text{Var}[\hat{E}(r_m) - E(r_m)] \\ & + \{\bar{E}(r_m)\}^2 \text{Var}[\hat{C}_i - C_i] \\ & + \text{interaction effects}, \end{aligned} \quad (5)$$

where $\bar{E}(\cdot)$ is the average (or unconditional) expectation.

Equation (5) breaks the total error variance into the sum of three effects. The first term is the contribution of the variance of the errors in estimating the expected risk premium on the SP500. This is the variance of the expected risk premium error multiplied by the square of the expected C coefficient. The latter is a number close to 1.0 (see Table 2 and Figure 1). The second term is the contribution of the variance of the errors in the estimates of the risk, C_i . Its effect is the variance of the risk measurement error multiplied by the square of the average expected premium. The average premium is on the order of $\frac{1}{2}\%$ per month, so the multiplier for the risk estimation effect is about $(0.005)^2$, a number much smaller than 1.0. The third term captures interaction effects that may arise when the errors in the estimates of C_i and the errors in forecasting the expected market premium are correlated. We show below that this term is relatively small. Thus, Equation (5) suggests the logic of our result that errors in estimating the risk measures are likely to be relatively unimportant, while errors in estimating the market-wide premium are very important.

3. The Data

We use a sample of monthly data for common stock returns, interest rates, and a set of market indicators. The stocks are value-weighted portfolios formed from New York and American Stock Exchange common

stocks, and returns are measured in excess of the one month U.S. Treasury bill rate. The data are provided by the Center for Research in Security Prices (CRSP). CRSP forms portfolios of stocks grouped on market equity capitalization at the beginning of each year. We use the returns for the smallest decile of firms to approximate closely held firms that are usually the subject of business valuation problems. We also use 12 value-weighted industry portfolios, grouped by two-digit SIC codes, following Ferson and Harvey (1991). Industry comparables are often used in practice to value closely held firms for which historical data are not available. One of the industries represents public utilities, which are of interest as the subject of rate regulation proceedings.

To define the true risk premiums to be used in our simulations, we investigate several methods for forecasting the SP500. One method is to use the forecasts provided by econometric forecasting services, such as Data Resources, Inc. (DRI). We obtained DRI forecasts monthly, starting in July of 1985. DRI forecasts the level of the SP500 one to six months ahead and three years ahead. Each month we use the nearest term forecast that is at least one month ahead and convert the forecast to a monthly return.⁶ As DRI forecasts the index level, which is net of dividends, we combine their forecasts with the actual monthly dividend yield to study forecast rates of return. Using the actual yield makes the DRI forecasts look better, but the effect is probably small since the aggregate dividends are relatively easy to predict one month ahead.

Recent research suggests that expected returns and risk, conditioned upon the current values of certain market indicators or state variables, differ from the unconditional moments of the return distributions. By including both the security returns and the market indicators in our sample and using the joint empirical distribution, we are able to capture in our analysis this important time-variation in risk and return.

We represent information about the state of the economy at a point in time using a collection of variables suggested by previous research (see the review article by Ferson 1995). The variables are: (1) the lagged level

⁵ Equation (5) follows from a Taylor series expansion for ϵ_i viewed as a function of two random variables: the risk premium error and the risk measurement error. The interaction effects include the covariance between the two. Equation (5) assumes that the forecasts are unbiased and efficient, which implies that the forecast errors should be mean zero and independent of the forecasts. To the extent that this assumption fails, the resulting errors will be attributed to the interaction effects.

⁶ To do this, we assume a flat term structure of expected returns. For example, if the forecast is a two-month return, r_2 , we form the one-month return as $r_1 = (1 + r_2)^{1/2} - 1$.

of the one-month Treasury bill yield from CRSP, (2) the lagged dividend yield of the CRSP value-weighted NYSE + AMEX stock index, (3) the lagged slope of the U.S. Government bond term structure, measured by the yield difference for 10-year versus three month-to-maturity bonds from Citibase, (4) a quality spread in the corporate bond market, measured as the difference between the Citibase yields for BAA and AAA rated corporate bonds, and (5) a dummy variable for the month of January. These data are available to us on a monthly basis, starting in February of 1947. Our sample ends in December of 1989.

4. Overview of the Method

The simulation strategy proceeds in three steps, which are described in the next three sections of the paper. An appendix at the end of the paper provides further technical details. In the first step, we define the *true* risk premiums for the SP500. The true required returns for the various assets are then constructed to conform to a particular latent variable model, as in Equation (3). (We use several specifications of the model to check the sensitivity of the results.) In the second step, we generate samples of artificial data. Each of the artificial samples contains 515 monthly returns for our representative assets and the SP500 index, matched to the February 1947 through December 1989 period. The artificial returns are equal to the true required returns plus noise, which noise is generated by resampling from the actual data. In the third step, we model a hypothetical analyst who observes a sequence of artificial returns for the assets and the SP500, and at each date forms cost of capital estimates. The analyst does not know the true efficient portfolio, the true risk premiums, or the true risk measures.

The experiment is repeated 1,000 times for each set of assumptions about which the model generates the data and what the analyst does with the data. In each experiment the analyst sequentially processes a sample with 515 observations, generating a monthly sequence of forecasts. We then analyze the errors in the estimates, aggregating across the 1,000 trials.⁷

⁷ In those cases where the analyst runs a latent variable model, it requires a highly iterative search for each month. Repeated for each

5. Defining the True Expected Excess Returns

Our first step is to specify the *true* values for the expected risk premiums. This requires a specification of the expected premium for the SP500 index and the true values of the *C* coefficients. Together, these determine the true values of the expected premiums for the representative assets via Equation (3).

5.1. The Expected SP500 Premiums

Since our simulations may be sensitive to the specification for the true expected premium for the SP500, we use finance theory and the results of recent empirical research, augmented by our own analysis, to guide the specification. We also vary the assumptions to check the sensitivity of the results.

The concept of semi-strong form market efficiency is well established in finance research (e.g., Fama 1970). It implies that the expected market premium fully reflects publicly available information at the time the forecasts are made. It also implies that the best forecast using only publicly available information is the closest to the true expected premium.

Table 1 compares forecasts of the SP500 monthly returns from several expected return models, using data for July 1985–December 1989. The analysis begins in July of 1985 because the DRI data are available from that date. The first column of the table records the results using the DRI forecasts. The remaining columns summarize four alternative models.

Two of the models use the historical, arithmetic average excess returns of the SP500 to form forecasts. Bruner et al. (1996) provide a survey of the methods used to estimate cost of capital in a sample of highly regarded firms and financial advisors. Most of the firms used historical averages to estimate the market risk premium, despite a general agreement that the market premium should theoretically reflect current market expectations. Two versions of this historical average approach are used in Table 1. One uses all of the historical data (starting in 1947) that would be available on the date that a

month, each simulation trial, and for each version of the simulations, it takes a lot of computer time. It takes almost *two months* of run time (using Gauss on a 486-50 PC) to get 100 trials. Therefore, we run only 100 trials for those simulations.

Table 1 Forecast Errors for the Future Standard and Poors 500 Returns: Three Expected Return Models (July 1985–December 1989)

Forecast:	DRI	Historical averages		Regression models	
		60-month	All past	60-month	All past
Total SP500 Variance =	0.002822*	0.002787	0.002787	0.002787	0.002787
Error Variance	0.002300	0.002876	0.002794	0.002845	0.002757
+ Var (forecast)	0.000427	0.000011	<0.000001	0.000233	0.000080
+ systematic forecast error (MAE) ²	0.000095	-0.000100	-0.000007	-0.000291	-0.000050
Average Forecast Error	0.00027	0.00144	0.00151	0.00146	0.00149
MSE	-0.00992	-0.00156	-0.00394	0.00645	-0.00552
	0.00235	0.00283	0.00276	0.00283	0.00274

Notes. DRI stands for the Data Resources International forecasts. MSE denotes the mean square forecast error, MAE is the mean absolute forecast error; the units for the average errors are monthly decimal fractions. The variances and MSE are decimal fraction per month squared. The forecasts errors are broken down as follows: Let r_m be the future return of the SP500 index and E be a forecast. Define the forecast error as ϵ , where $r_m = E + \epsilon$. Then $\text{Var}(r_m) = \text{Var}(E) + \text{Var}(\epsilon) + 2\text{Cov}(\epsilon; E)$, where $\text{Var}(r_m)$ is the Total SP500 Variance, $\text{Var}(\epsilon)$ is the error variance, $\text{Var}(E)$ is the Variance of the forecast (denoted $\text{Var}(\text{forecast})$), and $2\text{Cov}(\epsilon; E)$ is the systematic forecast error.

* The total SP500 variance in the "DRI" column differs from the others because the series for the SP500 provided by DRI differs slightly from the SP500 return series provided by CRSP.

forecast is made. This approach is supported by case law in a number of applications, such as utility rate regulation. Using a long historical average effectively assumes that the expected premium is a fixed parameter. The second approach uses only the most recent 60 months of data to form an average. This reflects the view that the older data are less relevant.

Two additional cases in Table 1 follow recent studies on the predictability of stock returns. The idea is to use information on the joint distribution of the stock market return and the lagged market indicators, or state variables. The conditional expectation of the return, given the current values of the state variables, serves as the forecast. Following recent studies, we estimate a regression of the stock return on the most recently known lagged values of the state variables. The fitted regression equation, plugging in the currently available market indicators, is used as the expected premium. There are two versions of this approach summarized in Table 1. One uses only the last 60 months of data to estimate the regression and the other uses all of the historical data, starting in 1947, up to the forecast date.

The first row of Table 1 shows the total variance of the SP500 excess return. The second row shows the error

variance of the five forecasts. The last three rows of the table summarize the average forecast errors, the mean squared errors (MSE), and the mean absolute errors (MAE). To keep the units comparable, we report the square of the MAE. By any of these measures, the DRI forecasts were the best over this period. Their forecasts explain about 16 percent of the variance of the future return. Using either error variance or MSE, the second best forecast is delivered by the regression models using the entire history of the available data. (Using MAE, the 60-month historical average is second best, but the differences in the MAEs, excluding DRI, are relatively small.) By error variance or MSE, the historical average using only the last 60 months of returns is the worst forecast.

Rows 2 through 4 of Table 1 break the SP500 variance down into the sum of three terms for each forecast. These terms, as explained in the notes to the table, are the forecast error variance, the variance of the forecast, and the *systematic forecast error*. The latter component indicates if the forecast fails to make efficient use of information by measuring the relation between the forecast and the forecast error. (If the forecast error is related to the forecast, and *if* this relation were known in advance, the forecast could be improved.) The systematic

forecast error numbers do not show any glaring inefficiencies, which supports the assumptions used to derive Equation (5). Four of the five average forecast errors are negative, which indicates that the SP500 outperformed most expectations over this period. The most variable forecast is the DRI forecast, followed by the 60-month regression model.

Since the DRI forecasts in Table 1 have the smallest forecast errors, market efficiency suggests that the DRI forecasts provide the best model of the expected SP500 premium, under the assumption that DRI uses only publicly available information to generate its forecasts. However, it is possible that DRI uses more information than is readily available in the market. Moreover, it is difficult for us to mimic the process by which DRI arrives at its forecasts in a simulation.⁸ The second best model is the regression model using all of the historical data. This model uses only readily available data, and in a simple mechanical way. We therefore use this model to represent the true expected premiums in our simulations. We regress the SP500 excess return on the market indicators:

$$r_{mt} = \delta_0 + \delta'_m Z_{t-1} + u_{mt} \quad t = 1, \dots, T, \quad (6)$$

where Z_{t-1} is the vector of the most recent lagged values of the state variables. We use time series data for February 1947 through December 1989, and take the fitted values of the regression as our model of the "true" expected premium.

Our model for the true expected premium on the SP500 is likely to lead us to understate the importance of risk premium errors in the cost-of-capital calculations. The DRI forecasts vary more from month to month than the regression forecasts. If the DRI numbers are closer to the true expected premiums, then we use smoother premiums than we should in the simulations. A smoother expected premium is likely to appear to be easier to estimate. Given our finding that risk premium error is important, the regression model is a conservative choice.

⁸ The details of DRI's approach are proprietary, but our inquiries at DRI revealed that they use a number of variables similar to the ones we use in our regression models in forming their forecasts.

5.2. Defining the True Risk Measures

The true C coefficients depend on the unobserved, mean variance efficient portfolio. We use three widely different proxies for the mean variance efficient portfolio. If the results are similar in all three cases, we build confidence in their relevance to the real world. The first proxy is the CRSP value-weighted common stock index. The second is a combination of the stock index and a 20-year U.S. Treasury bond weighted as 90 percent bond, 10 percent stock. The third proxy we call "ex-post optimal." We form a combination of the size and industry portfolios, using their returns data over the entire sample period, which maximizes the average excess return of the portfolio divided by the sample standard deviation of the portfolio.

Table 2 summarizes the values of the *true* risk measures, C_i . There are three columns corresponding to the three different assumptions about the mean variance efficient portfolio. From Equation (3), the C coefficient for an asset i equals its conditional beta on the mean variance efficient portfolio, divided by the conditional beta of the SP500 on the mean variance efficient portfolio: $C_i = \beta_{ip} / \beta_{mp}$. To form the coefficients in Table 2 we estimate conditional betas for the asset and for the SP500 under the assumption that the conditional betas are fixed parameters.⁹ Our estimate of C_i is the ratio of the estimates for β_{ip} and β_{mp} . Table 2 shows that the C_i are substantially different using the different mean variance portfolio proxies; especially, the ex post optimal proxy is different from the other two. Therefore, comparing the simulations that use the different proxies should provide useful information on the robustness of the results.

⁹ If r_p is the assumed mean-variance efficient portfolio excess return, we regress r_p over time on a constant and the lagged market indicators to obtain residuals, u_p :

$$r_{pt} = \delta_{op} + \delta'_{ip} Z_{t-1} + u_{pt} \quad t = 1, \dots, T.$$

We then regress the relevant size or industry portfolio or the SP500 return, denoted r_{it} , on a constant and u_p :

$$r_{it} = a_i + \beta_{ip} u_{pt} + \epsilon_{it} \quad t = 1, \dots, T.$$

Under standard regression model assumptions, the second regression delivers a consistent estimate β_{ip} for the conditional beta, so the ratio of two regression slopes is a consistent estimate of the C coefficient.

Table 2 Values of the Risk Measures Used in the Simulations

Mean-variance efficient portfolio: Risk measures (C_i) for the following:		VW	Bond/Stock	Ex post
1	Standard and Poors 500 (SP500)	1.00	1.00	1.00
2	Small stock portfolio	1.16	1.16	1.36
3	Petroleum Industry Portfolio	0.99	0.98	1.26
4	Finance/Real Estate Portfolio	1.01	1.02	1.02
5	Consumer durable goods producers	1.11	1.10	1.07
6	Basic Industries	1.08	1.07	1.02
7	Food and Tobacco Firms	0.86	0.87	1.12
8	Construction Industry	1.18	1.18	0.85
9	Capital Goods producers	1.07	1.06	1.02
10	Transportation Industry	1.19	1.19	0.94
11	Utilities	0.64	0.65	0.82
12	Textiles and Trading firms	1.03	1.04	0.95
13	Services Portfolio	1.21	1.21	1.24
14	Leisure Industry portfolio	1.18	1.19	1.14

Notes. The table shows the "true" values of the C_i coefficients, which measure risk in the simulations, when different portfolios are assumed to be the unobserved mean-variance efficient portfolio.

VW refers to the value-weighted CRSP stock index of NYSE and AMEX firms. Bond/Stock refers to a portfolio composed of 90 percent long-term U.S. government bonds and 10% the value-weighted CRSP stock index of NYSE and AMEX firms. Ex post refers to a portfolio of the Small stock index and 12 industry portfolios, chosen to be mean variance optimal on the basis of the full historical sample of returns data.

Each of the cases summarized in Table 2 makes the assumption that the true C coefficients are fixed parameters over time. As the simulation results may be sensitive to that assumption, we conduct additional experiments, described below, in which the true C coefficients are assumed to be time-varying.

6. Generating the Artificial Data

The difference between the excess returns in the actual data and the true expected premiums, as determined above, define a sample of error terms. Each error term is an N vector, where N is the number of assets including the SP500. We demean these errors to produce a sample of "unexpected" monthly return vectors. Following Efron (1982), the unexpected returns represent the population distribution for the simulations. We randomly resample, with replacement, from this "urn" of unexpected return vectors. This implies that the conditional covariance matrix of the N returns in the "population" for our simulations is equal to the sample conditional covariance matrix that we estimate in the actual

data. By resampling we also preserve any nonnormality that may be present in the data.

The artificial excess returns for asset i are formed for a given "date" in the simulations as the sum of the true expected excess return for that date, $C_i E(r_m)$, plus a randomly resampled unexpected return, u_i :

$$r_i = C_i E(r_m) + u_i \quad i = 1, \dots, N. \quad (7)$$

Each artificial sample contains 515 observations for the N excess returns, matched to the February 1947 through December 1989 period.

7. Modeling the Analyst's Behavior

In the simulations the hypothetical analyst observes a sequence of artificial returns data for the SP500 and the other assets, but does not know what mean variance efficient portfolio generates the true expected returns. In each simulation trial, the analyst "lives" on a single sample path of the artificial data. The analyst estimates the cost of capital for the small stock portfolio and the 12 industries at each date, using only data that would

be available on that "date." Depending on the method that the analyst adopts, as many as 120 monthly observations are used to initialize the model. We therefore exclude the first 120 months when we compare the various approaches, evaluating the simulated forecast errors for the remaining 395 "months."

7.1. Forecasting the SP500 Premium

The first forecast for the SP500 used by our hypothetical analyst is the historical average excess return of the SP500, given the series of artificial data known to the analyst on the forecast date. As we described earlier, this is representative of current practice. In a second approach, a more sophisticated analyst estimates a regression model for the excess return of the SP500 on the predetermined market indicator variables. In this case the analyst uses the artificial data for the SP500, and we allow him/her to observe the actual sample values of the predetermined indicators.

We re-use the actual market indicators in the simulations for several reasons. This approach implies that given an infinite sample, the analyst would discover the true expected return model. By combining the actual state variables with randomly resampled residuals, the only relation between the predetermined variables and the artificial returns is through the true expected returns, consistent with the market efficiency assumption. The approach ensures that this relation is the same as the one we observe in our actual data. A cost of this approach is that the simulations do not reflect randomness in the conditioning variables. However, since the variance of returns is dominated by the unexpected returns, this does not seem to be a serious limitation.

As described above, we are unable to evaluate an analyst who purchases forecasts from a service like DRI, because we would have to simulate what they do to generate their forecasts. However, since DRI uses variables similar to those in our regression models, our simulations in which the analyst uses regression models may provide a reasonable approximation to the results that would be obtained by an analyst who purchases forecasts from a service like DRI.

A third approach used by our hypothetical analyst is to estimate a latent variable model explicitly, using the Generalized Method of Moments (GMM, see Hansen 1982) as described in the appendix. Since a latent vari-

able model is used by us to generate the "true" risk premiums, it seems reasonable to consider an analyst who also uses state-of-the-art financial econometrics.

7.2. Estimating Risk

We compare three approaches that an analyst might use for estimating the risk measures, C_i . These are designed to capture a wide range of variation such as might be experienced in practice. In the first approach the analyst uses the previous 60 months of artificial data and estimates simple regression betas against the SP500. (In practice, analysts may purchase similar beta estimates from third party vendors or estimate them in house—see Bruner et al. (1996) for relevant survey evidence.) The second approach uses the "sophisticated" estimates of the C_i coefficients that are produced by GMM estimation of a latent variable model. The third approach is the simplest and crudest. In this case, the analyst simply assumes that all of the betas are equal to one. This means that the analyst's expected return of the SP500 is assigned as the cost of capital.

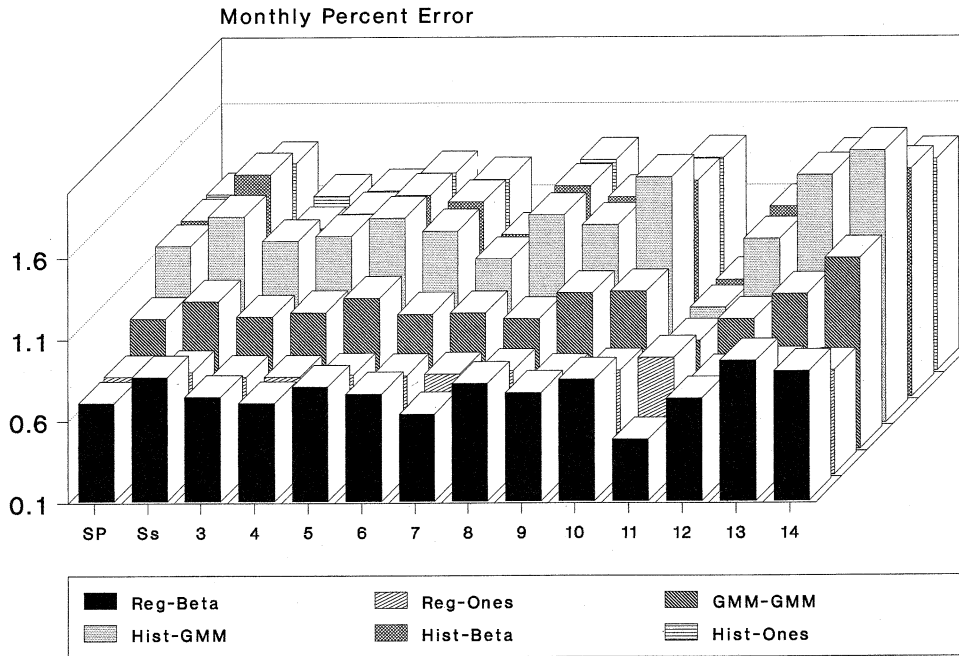
8. The Simulation Results

With three ways to estimate the market risk premium, three ways to estimate risk, and three proxies for the unobserved efficient portfolio, there are 27 combinations. We consider the GMM market premium estimates only in combination with the GMM risk estimates, which reduces the number of cases to 21.

8.1. Errors in Estimating Required Return

Figures 1–3 summarize the total errors in the cost of capital estimates. Each figure uses a different assumption about the unobserved mean-variance efficient portfolio, and includes one set of columns for each the 12 industries, the small firm stock portfolio, and the SP500. In Figure 1 the value-weighted index is the mean-variance efficient portfolio. The smallest overall errors are obtained when the regression model is used to estimate the SP500 premium, and the largest errors occur using the past historical average. The GMM approach is better than the historical averages, but not as good as the regression methods. The method used to estimate the SP500 premium matters more than the method used to estimate risk. For example, the first two rows of the figure show that simply assuming $C_i = 1$ works better

Figure 1 Required Return Estimation Errors: Efficient Portfolio—Value Weighted



Notes for Figures 1–3. Required return estimation errors are standard deviations of the difference between the analyst’s forecast of the required return and the actual required return in the simulation, averaged over the simulation trials. The approaches are denoted in the legend as PREMIUM METHOD—RISK ESTIMATION METHOD, where the premium method is either REG (regression model), HIST (historical averages), or GMM (Generalized Method of Moments estimate). The risk estimation method is either BETA (60-month regression betas using the Standard and Poores 500), GMM (using a latent variables model) or ONES (set all risk measures equal to 1.0). The size and industry portfolios are shown on the horizontal axis, denoted as follows: SP = Standard and Poores 500; Ss = Small stock portfolio (bottom two deciles on NYSE + AMEX); 3 = Petroleum Industry Portfolio; 4 = Finance/Real Estate Portfolio; 5 = Consumer durable goods producers; 6 = Basic Industries; 7 = Food and Tobacco Firms; 8 = Construction Industry; 9 = Capital Goods producers; 10 = Transportation Industry; 11 = Utilities; 12 = Textiles and Trading firms; 13 = Services Portfolio; 14 = Leisure Industry Portfolio.

than regression betas on the SP500 for some industries and worse for others, but the average performance across the industries of the two methods is close.¹⁰

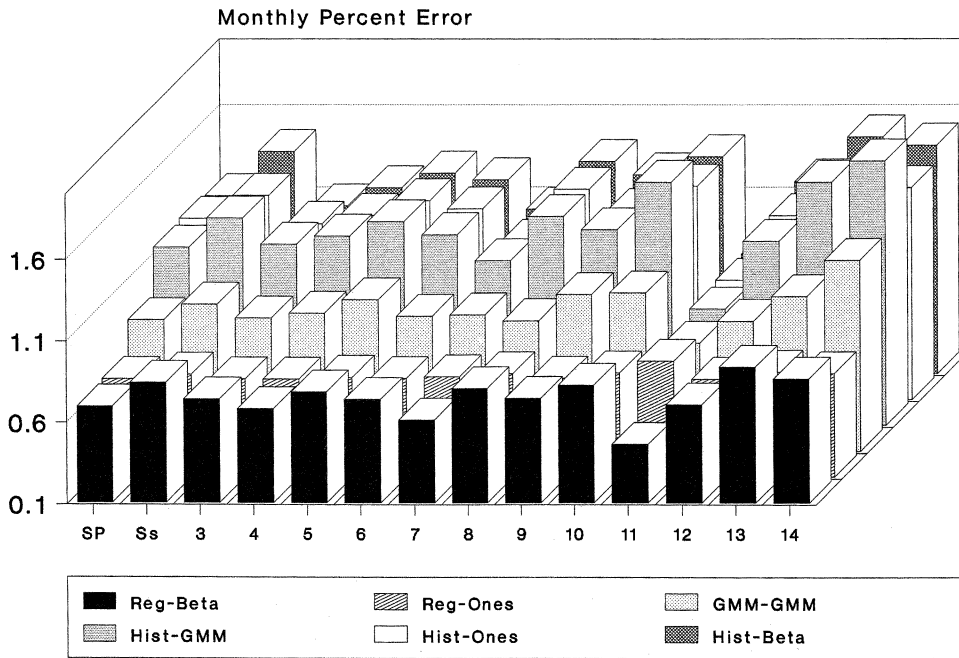
Figure 2 summarizes the results when the efficient portfolio is assumed to be 90 percent bonds and 10 percent stocks. The smallest errors are obtained using a regression model for the SP500 premium and the 60-month regression betas for risk. The worst results are obtained when the analyst uses historical averages to

estimate the SP500 premium and assumes that the betas are all equal to 1.0. The method used to estimate risk does not matter as much as the method used to estimate risk premiums. The GMM approach performs better than the historical premium estimates, but not as well as the regression approach. The GMM approach is a relatively complicated nonlinear procedure. It is interesting to find that a simpler regression approach produces smaller errors, out-of-sample.

Figure 3 illustrates the results when the efficient portfolio is ex post optimal, and again the results are similar. The fact that we get similar results using three very different proxies for the mean-variance efficient portfolio is comforting. It suggests that similar results are relevant in the real world, even though the ex ante mean-variance efficient portfolio is unknown.

¹⁰ We repeat the simulations behind Figure 1 (excepting the models using the GMM) but we use MAE to measure the total forecast errors, and we find similar results: The largest errors are produced using the historical average to estimate the SP500 premium (using ones is worse than regression betas) and the smallest errors are made using the regression to estimate the premium (either type of beta similar).

Figure 2 Required Return Estimation Errors: Efficient Portfolio—Bond/Stock (90/10)



In Figure 3 the standard error of the SP500 premium forecast errors are reduced from 1.2 percent per month, using the historical average, to about 0.7 percent using the

regression models. The cost of capital errors for most of the other portfolios is reduced by similar magnitudes. These figures suggest that it should be possible to

Figure 3 Required Return Estimation Errors: Efficient Portfolio—Ex-post Optimal

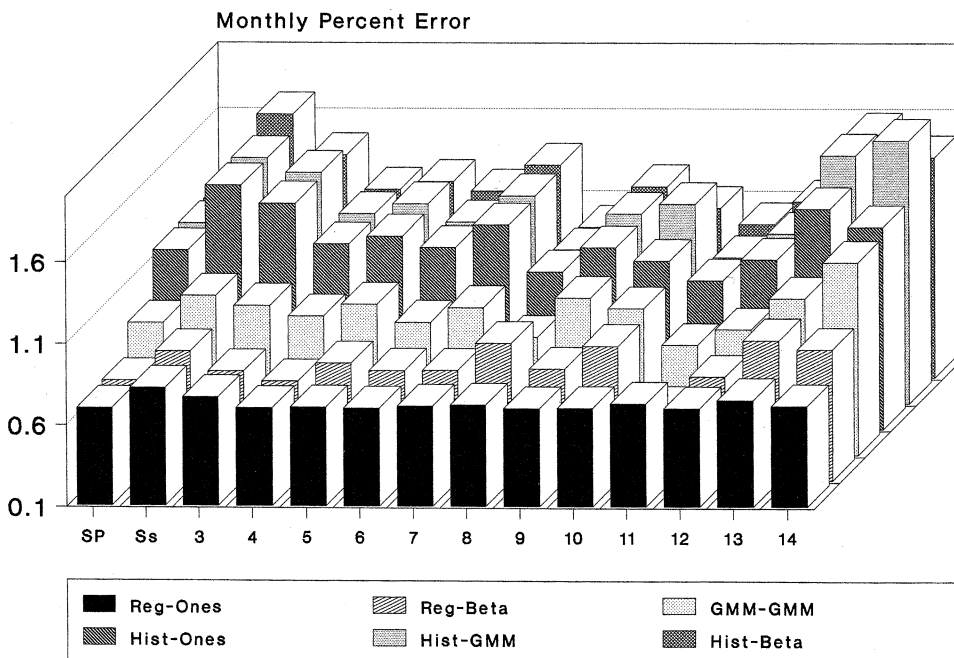
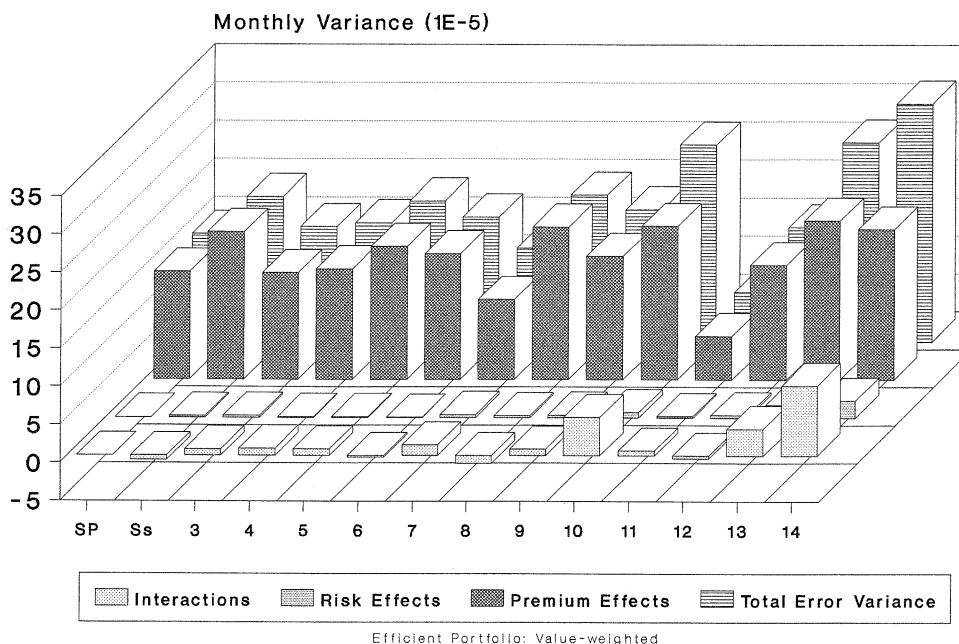


Figure 4 Decomposition of Error Sources: Premium—Historical Average; Risks—GMM Estimates



Notes for Figures 4–6. The back row depicts the variances of the difference between the analyst's forecast of the required return and the actual required return in the simulation, averaged over the simulation trials. This total error variance is broken down as in Equation (5) of the text into three components, depicted in the front three rows. Risk effects are the component due to errors in estimating risks. Premium effects are the component due to errors in estimating the expected risk premium of the Standard and Poors 500 index. Interactions reflect the correlation between the first two components and approximation errors. The size and industry portfolios are indicated on the horizontal axis as in Figure 1–3.

substantially reduce errors in cost of capital estimates by using more accurate measures of the expected SP500 premium.

8.2. Breaking Down the Sources of Errors

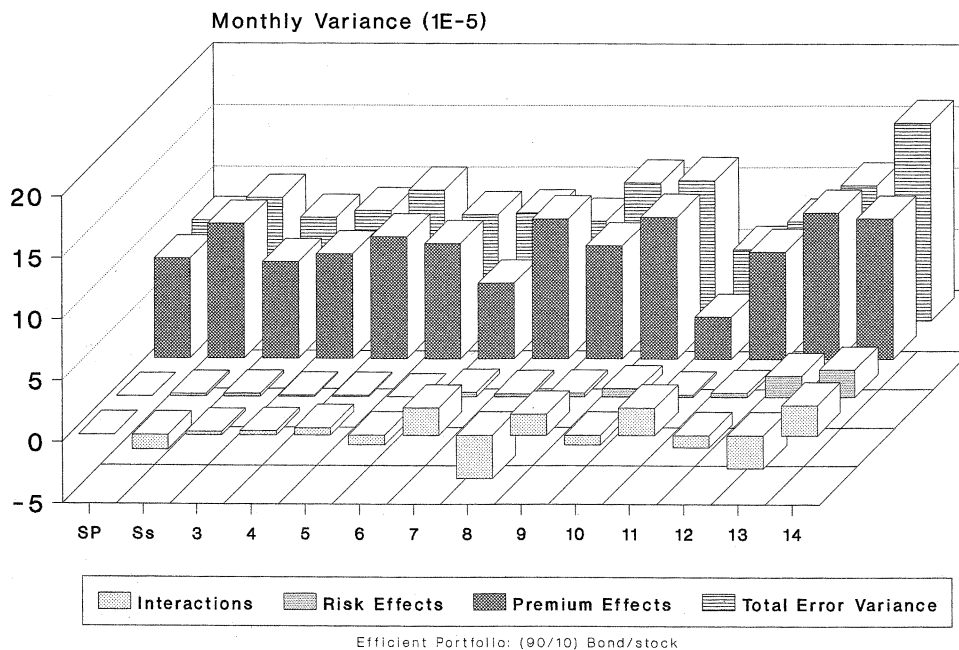
Figures 4–6 break down the total errors in the cost of capital estimates according to Equation (5). There are 21 cases; we select three representative examples for the figures. Each figure uses a different proxy for the mean-variance efficient portfolio. The row of columns in the back of each figure depicts the total forecast error variance. The components of the error variance are shown in the three foreground rows. For example, the left-most columns illustrate that 100 percent of the error is due to risk premium estimation error for the SP500, since its C coefficient is known to be 1.0. The relative heights of the columns depict the relative contributions of the various sources of error. For example, in Figure 4 the results for the leisure industry (portfolio 14) indicate that 64 percent of the total error

comes from the risk premium, 7 percent from risk estimation, and 29 percent from interaction effects.

A remarkable result, which appears in almost all of the experiments, is that the risk premium estimation error is the most important effect. The direct contributions from using the wrong risk measures are usually small. The risk estimation error accounts for 7 percent, or less, of the total error in a cost of capital estimate, no matter what method is used to estimate the risks, in all except three cases in Figures 4–6 (the exceptions yield 7.2 percent, 15.7 percent, and 13.9 percent, respectively).

The direct risk estimation effect may understate the importance of risk estimation error, because the interaction effects reflect correlation between the errors in the estimates of risk and errors in the risk premiums. However, even if we attribute the entire interaction effect to risk estimation error, we find that the combined effect is smaller than the direct risk premium estimation error most of the time.

Figure 5 Decomposition of Error Sources: Premium—GMM Estimates; Risks—GMM Estimates



8.3. Time-Varying Risk

The dramatic reductions in the total errors resulting from improved SP500 premium estimates in Figures 1–

3 are consistent with the small impact of the risk estimation errors, as illustrated in Figures 4–6. It is interesting to note that the small impact of the risk

Figure 6 Decomposition of Error Sources: Premium—Regression Model; Risks—Regression Estimates

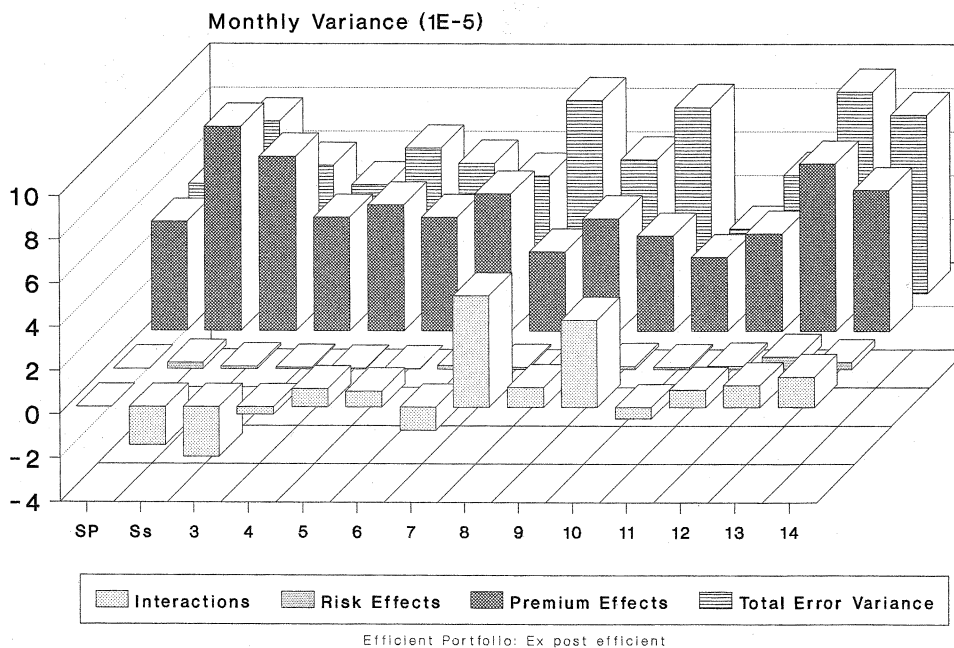
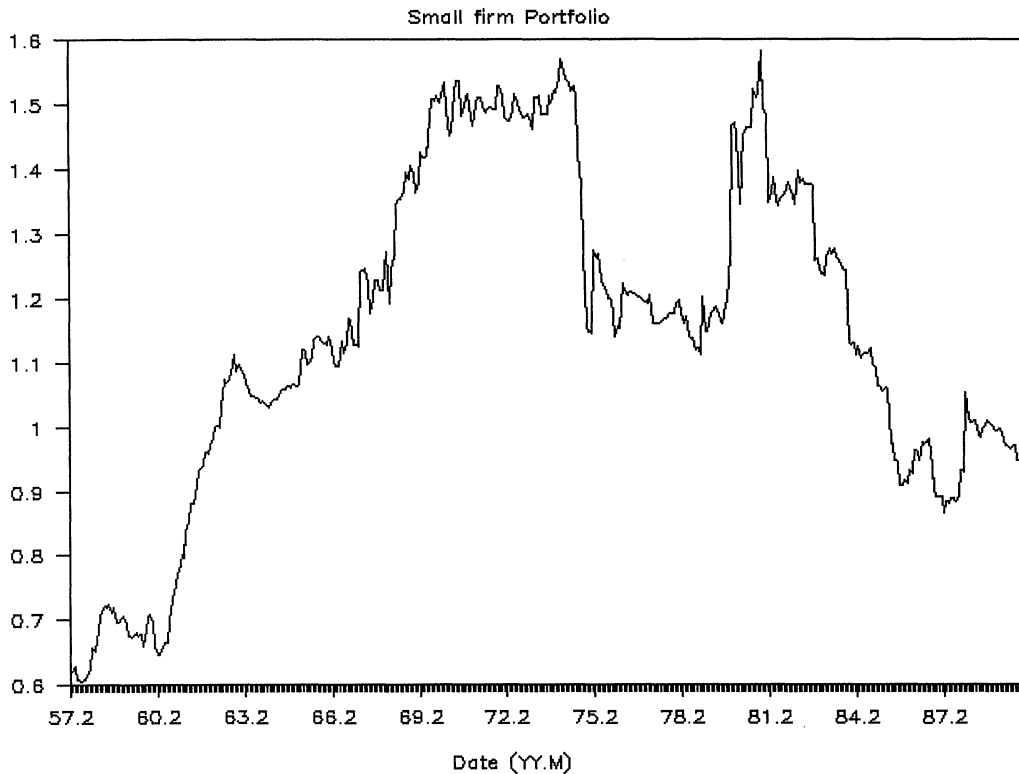


Figure 7 Sixty-Month Regression Beta Estimates



Each beta is the ordinary least squares regression slope coefficient, using the previous 60 months of data, for the small firm portfolio return regressed on a constant and the return of the Standard and Poors 500 index. The values correspond to the February 1957 through December 1989 period.

estimation errors is not simply an artifact of risk estimates that do not vary much over time. Figure 7 is a time-series plot of 60-month regression beta estimates for the evaluation period of a typical simulation. It is clear that the beta estimates have a substantial amount of variation over time. Since this figure relates to a simulation in which the true risk measure is a fixed parameter, it shows that there is a substantial amount of risk estimation error. It is therefore interesting to find that this error contributes relatively little to the overall cost of capital estimation error.

In the simulations summarized by Figures 1–6 the true risk measures are assumed to be constant over time. We repeat some of the simulation experiments, using true risk measures that are assumed to be time varying. We determine the true values in these experiments by first regressing each excess return over time on the lagged market indicators. We then model the true

C_i coefficient by the ratio of the fitted values of two regression functions. The numerator uses the regression for the excess return of asset i , and the denominator uses the regression for the SP500 return.¹¹ Figure 8 presents a time-series plot of these time-varying C_i coefficients for the small-firm portfolio. The risk measures, con-

¹¹ The two regressions are:

$$r_{it} = \delta_{0i} + \delta'_{1i}Z_{t-1} + u_{it}$$

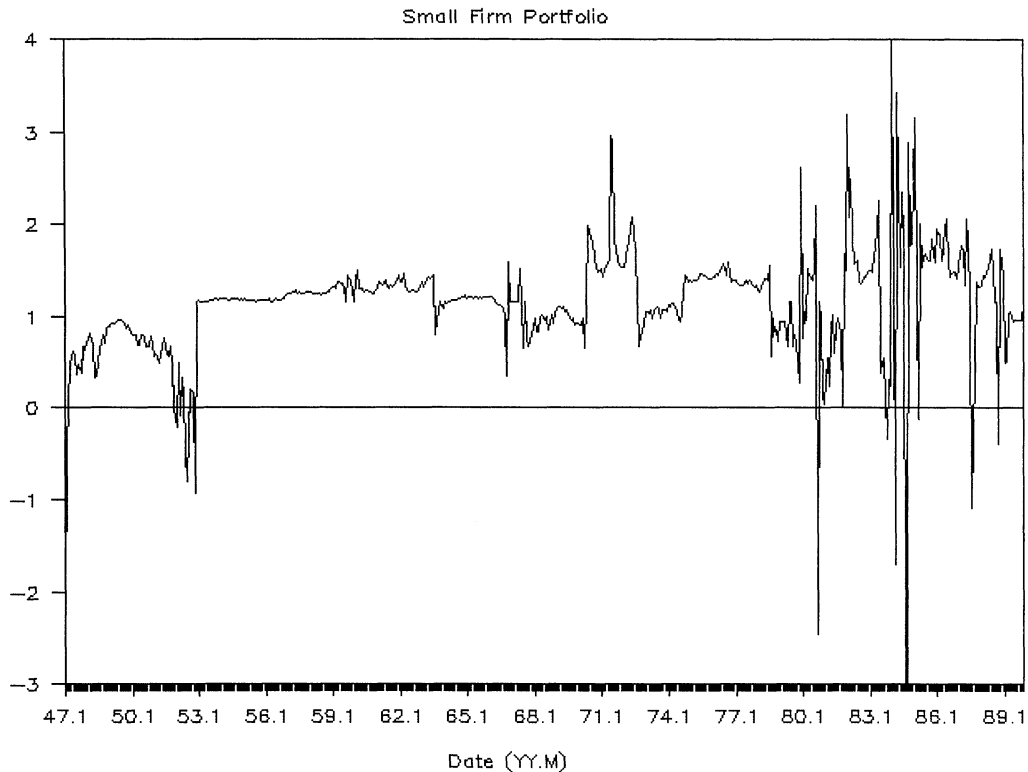
$$r_{mt} = \delta_{0m} + \delta'_{1m}Z_{t-1} + u_{mt}$$

and

$$C_{it} = (\delta_{0i} + \delta'_{1i}Z_{t-1}) / (\delta_{0m} + \delta'_{1m}Z_{t-1}).$$

This approach creates a problem when the fitted value in the denominator takes values close to zero. When the denominator is less than 10 basis points, we replace the time-varying C coefficient for that date by the average C coefficient in Table 2.

Figure 8 "True" Time-Varying C_i Coefficients



Time series plots of the "true" time-varying C_i coefficients used in the simulations for the small firm portfolio. Each C_i coefficient is estimated by the ratio of two regression functions, as described in the text. The values correspond to the February 1947 through December 1989 sample period (515 monthly observations).

structured in this way, have substantial time-variation. When we use these in the simulations as the true risk measures, we find that the basic findings are robust. Even when the true risk measures vary over time, risk premium error remains the largest component of the errors in the required return estimates.

9. Conclusions

To calculate a cost of equity capital, analysts often use historical average returns as an estimate of the risk premium on a market index, combined with estimates of beta according to the CAPM. Our simulations suggest that the accuracy of cost of capital figures is likely to benefit from improving the estimates of the expected premium on a standard market benchmark, even if the CAPM is the "wrong" model. Estimates

that reflect the current state of the economy may be purchased from specialists or generated in house, for example, using straightforward regression analysis. When errors in the cost of capital over time are the issue, improving the market premium estimate is much more important than are concerns about using the "wrong" beta.

Recent evidence that expected risk premiums vary with the state of the economy raises a host of issues about the practice of capital budgeting, and this article has merely scratched the surface. For example, our analysis followed the common practice of using monthly data to develop the estimates, even though the results may be used to evaluate cash flows many years into the future. However, studies find that market indicators like the ones we use have higher explanatory power for long-horizon returns than for short-horizon returns

(e.g., Fama and French 1989). Therefore, state dependence in required returns is likely to be very important when formally incorporated into long-term capital budgeting problems. Future research is needed to deepen our understanding of these important issues.¹²

¹² We are grateful to Robert Chiang, Chip Haley, Campbell Harvey, Paul Malatesta, Jessica Rutledge, Larry Schall, and participants at the 1993 American Society of Appraisers Conference for helpful discussions. We especially appreciate the extensive comments of a *Management Science* referee and the editor, which have greatly improved the paper.

Appendix

Under the assumption that the C coefficient risk measure is a fixed parameter for each asset, the following system of equations describes the latent variable model:

$$r_{1t} = \delta_1' Z_{t-1} + u_{1t},$$

$$r_{jt} = C_j(\delta_j' Z_{t-1}) + u_{jt} \quad t = 1, \dots, T; j = 2, \dots, N. \quad (A1)$$

where r_1 is the excess return of a reference asset, which we take to be the SP500. The remaining asset excess returns are r_2, \dots, r_N , and C_2, \dots, C_N are the C coefficients. The vector of predetermined market indicator variables is denoted by Z_{t-1} . The expected excess return of the reference asset is assumed to follow a linear regression on the lagged instruments and $\delta_1' Z_{t-1}$ is the conditional expected return.

The system (A1) is a multivariate regression model with nonlinear cross-equation restrictions. The system is estimated by the GMM (Hansen 1982), using an iterated version of the approach as recommended by Ferson and Foerster (1994). The model implies that the combined error term $u_t = (u_{1t}, u_{2t}, \dots, u_{Nt})$ satisfies the restriction $E(u_t | Z_{t-1}) = 0$. This implies, by the law of iterated projections, that if we define a sample orthogonality condition, using data for $t = 1, \dots, T$, by $g_T = T^{-1} \sum_t (u_t \otimes Z_{t-1})$, then $E(g_T) = 0$. The GMM chooses the parameter values for the model so as to minimize the quadratic criterion $g_T' W g_T$, where W is a nonsingular weighting matrix. (See Ferson and Foerster 1994 for further details.)

References

Bruner, Robert F., Kenneth M. Eades, Robert S. Harris, and Robert C. Higgins, "Best Practices in Estimating the Cost of Capital: Survey

- and Synthesis," Working Paper, University of Virginia, Charlottesville, VA, and University of Washington, 1996.
- Efron, B., *The Jackknife, the Bootstrap, and Other Resampling Plans*, Society for Industrial and Applied Mathematics, Philadelphia, PA, 1982.
- Fama, E. F., "Efficient Capital Markets: A Review of Theory and Empirical Work," *J. Finance*, 25 (1970), 383-417.
- and K. R. French, "Business Conditions and Expected Stock Returns," *J. Financial Economics*, 25 (1989), 23-50.
- and —, "On the Cross-section of Expected Stock Returns," *J. Finance*, 47 (1992), 427-466.
- Ferson, W. E., "Theory and Empirical Testing of Asset Pricing Models," in Jarrow, Maksimovic, and Ziemba (Eds.), *Finance, Handbooks in Operations Research and Management Science*, Elsevier, (1995), 145-200.
- and S. R. Foerster, "Finite Sample Properties of the Generalized Method of Moments in Tests of Conditional Asset Pricing Models," *J. Financial Economics*, 36 (1994), 29-56.
- and C. R. Harvey, "The Variation of Economic Risk Premiums," *J. Political Economy*, 99 (1991), 385-415.
- Grinblatt, M. and S. Titman, "On the Relation Between Mean-variance Efficiency and Arbitrage Pricing," *J. Business*, 60 (1987), 97-112.
- Hansen, L., "Large Sample Properties of the Generalized Method of Moments Estimators," *Econometrica*, 50 (1982), 1029-1054.
- Huberman, G., S. A. Kandel, and R. F. Stambaugh, "Mimicking Portfolios and Exact Arbitrage Pricing," *J. Finance*, 42 (1987), 1-10.
- Kaplan, S. and R. Ruback, "The Valuation of Cash Flow Forecasts: An Empirical Analysis," *J. Finance*, 50 (1995), 1059-1094.
- Lehmann, B. N. and D. M. Modest, "The Empirical Foundations of the Arbitrage Pricing Theory," *J. Financial Economics*, 21 (1988), 213-54.
- MacKinaly, A. Craig, "Multifactor Models Do Not Explain Deviations from the CAPM," *J. Financial Economics* (1995).
- Pratt, Shannon P., "Valuing Small Businesses and Professional Practices," in *Business One*, Irwin, Homewood, IL, 1993.
- Roll, R., "A Critique of the Asset Pricing Theories' Tests, Part I: On Past and Potential Testability of the Theory," *J. Financial Economics*, 4 (1977), 129-76.
- Ross, S. A., "The Arbitrage Theory of Capital Market Asset Pricing," *J. Economic Theory*, 13 (1976), 341-360.
- Sharpe, W. F., "Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk," *J. of Finance*, 19 (1964), 425-42.

Accepted by Robert Heinkel; received May 1994. This paper has been with the authors 26 months for 2 revisions.