Chapter 5
Validating Return-Generating Models

Marshall E. Blume, Mustafa N. Gültekin, and N. Bülent Gültekin

Return-generating models and the assessment of conditional expected returns underlie many important applications in finance. Jensen's (1969) measures of investment performance, as well as those based upon the Ross' (1976) Arbitrage Pricing Theory (APT), compare the realized return of a portfolio to a benchmark return. The benchmark return is an expected return conditional on some set of publicly available information. The assessment of a conditional expected return requires the specification of some stochastic process to characterize realized returns. Likewise, studies of the effect of an announcement of an unanticipated event often measure this effect by the difference between the realized return at the time of the announcement and some conditional expected return. Again, this measurement requires the specification of some stochastic process.

An assumption underlying many studies is that the market model, or more generally a model with one factor common to all securities, generates realized returns. In such a one-factor model, realized returns are the sum of an asset's response to a stochastic factor common to all assets and a factor unique to the individual asset. In the last decade, there has been much interest in models with more than one common stochastic factor, using either pre-specified factors, like Fama and French (1993) 3-factor model, or factors identified through factor analysis or similar multivariate techniques.[1]

[1]Factor analysis and similar factor analytic techniques have on occasion played an important role in the analysis of returns on common stocks and other types of financial assets. Farrar (1962)

M.E. Blume • N.B. Gültekin
The Wharton School, University of Pennsylvania, 3730 Walnut St., Philadelphia, PA 19104, USA

M.N. Gültekin (✉)
The Kenan-Flagler Business School, The University of North Carolina at Chapel Hill,
300 Kenan Dr, Chapel Hill, NC 27599, USA
e-mail: gultekin@unc.edu

J.B. Guerard, Jr. (ed.), *Portfolio Construction, Measurement, and Efficiency*,
DOI 10.1007/978-3-319-33976-4_5

A typical way to use a return-generating model is to estimate the model with data from one period of time and then employ the estimated model to calculate conditional expected returns in a different period of time—often the immediately following period. Implicit in this use of a model is the assumption that the underlying model is stationary over time. In fact, it is highly unlikely that any economic model, except for the most trivial, is stationary over time. The question is not whether a model is stationary, but rather the degree of sensitivity to non-stationarity since the accuracy of a predictive model hinges upon the “degree” of non-stationarity.

This paper will explore the effects of such non-stationarities upon the accuracy of conditional expectations assessed for time periods following the estimation period. To this end, this paper will assess the relative accuracy of the conditional expectations of various commonly used models with data different from those used in estimating the models. In psychometrics, evaluating the accuracy of a model in terms of how it is used is termed the validation of a model.

The principal finding of this paper is that, when the criterion of accuracy is the mean-squared forecast error, multi-factor models estimated with factor analytic techniques provide more accurate out-of-sample forecasts than the Fama–French 3-factor model and the usual market model. The predictive accuracy of the market model itself depends critically on the choice of the index—equal-weighted or value-weighted. The paper also examines one model that includes the pre-specified macro variables that Chen et al. (1986) have used in a prior study. The empirical evidence indicates that a model based solely upon these macro variables provides less accurate forecasts than the usual market model. Overall, the multi-factor models provide the most accurate forecasts of those models examined.

The organization of the paper is as follows. The first section describes the design of the empirical tests and proposes the mean-squared error of the forecasts as a natural statistic to analyze in the context of performance measurement and announcement studies. The second section examines various factor models to validate the number of required factors. The third section compares the accuracy

may have been the first to use factor analysis in conjunction with principal component analysis to assign securities into homogeneous correlation groups. King (1966) used factor analysis to evaluate the role of market and industry factors in explaining stock returns. These two studies sparked an interest in multi-index models, and a rich body of empirical work soon emerged. Examples include Elton and Gruber (1971, 1973), Meyer (1973), Farrell (1974), and Livingston (1977), among others. The major goal of these earlier studies was to establish the smallest number of “indexes” needed to construct efficient sets.

Factor models have been used in the tests of arbitrage pricing theory and its variants. See, for example, Rosenberg (1974), Rosenberg and Marathe (1979), Roll and Ross (1980), Chen (1983), Brown and Weinstein (1983), Dhrymes et al. (1984), Dhrymes et al. (1985a,b), Gültekin and Rogalski (1985), and Cho et al. (1984), to cite a few from the large literature. A four-factor model constructed with the Dhrymes, Friend, Gültekin, and Gültekin (1985b) methodology was used in conjuction with the Bloch, Guerard, Markowitz, Todd, and Xu (1993) stock selection model to construct efficient portfolios in the U.S., See Guerard, Gültekin, and Stone (1997).

of factor analytic models to the usual market model and those using pre-specified macro variables. The final section contains concluding remarks.[2]

5.1 The Design of the Experiment

The analysis in this paper for the most part follows a two-step procedure. The first step assumes the validity of specific return-generating models and utilizes one sample of data to estimate the parameters of these models. The second step uses data in a subsequent period to validate the estimated models.

5.1.1 The Validation Criterion

There are numerous ways to validate a statistical model. The specific method of validating a model hinges upon how a researcher plans to use the model. The focus of this paper is on the use of return-generating functions in performance evaluation studies and in analysis of the reaction of stock prices to unanticipated

[2] A prior and related paper is that of Conway and Reinganum (1988). The primary purpose of their paper is to assess the adequacy of the likelihood ratio test to determine the number of factors. They use as their validation criterion the accuracy of the implied variance–covariance matrix from a factor model estimated on one sample with the estimated variance–covariance matrix from a different sample in contrast to the focus of this paper on the mean-squared forecast error of the conditional predictions. These two validation criteria are clearly related, but the one used in this paper addresses directly the way in which researchers use return-generating models in event studies and performance evaluation. The reader is referred to Chen (1988) and Stambaugh (1988) for a further discussion of the differences in these two methods of validation.

Additionally, this study shows that the number of factors and the variance–covariance matrix of returns vary substantially over time, even over the July 1962–December 1972 time period that Conway and Reinganum examine. This study explicitly adjusts summary statistics for these non-stationarities.

There is also a significant difference in the selection of the estimation and validation period between this study and that of Conway and Reinganum. For the most part, Conway and Reinganum break their sample into even and odd days, using one set of days to estimate the model and the other to validate the model. This is appropriate under their assumption that the underlying variance–covariance matrix is stationary over time. They do present one analysis using the first five years to estimate a model and the second five years to validate it. This is closer to the spirit of this paper, but it still does not parallel as closely the usage of return-generating models in studies of events and performance evaluation where the prediction period is usually much shorter than the estimation period.

Finally, a major purpose of this study is to compare factor models with estimated factors, factor models with pre-specified factors, and variants of the usual market model, which was not a goal of Conway and Reinganum.

events, frequently termed "event" or "CAR" (cumulative average residual) studies. A measure consistent with these uses is the mean-squared forecast error.[3]

In his seminal article, Jensen (1969) proposes a measure of investment performance that relies upon the validity of the Capital Asset Pricing Model (CAPM) and, with the additional assumption of a one-factor generating model, shows how to estimate this measure with a least-squares regression. Implicit in least-squares regression is the objective of minimizing squared deviations.[4] Connor and Korajczyk (1988) show that Jensen's intuition generalizes to the APT and a multi-factor model. Similarly, event or CAR studies compare realized returns to conditional predicted returns, and then test the significance of the residuals using t-tests, which again use a metric based on mean-squared errors.

5.1.2 Conditional Expectations

The first part of this section develops the formulas for assessing conditional expected returns, assuming that the return-generating process for securities is jointly normal, stationary, and independent over time and that the parameters of the joint distribution are known. The second part incorporates factor models into the formulas and interprets factor models as placing restrictions on the estimated covariance matrix.

In the formulas developed following notation is used; r_i is the return on asset i less its unconditional expectation, σ_{ii} is the variance of the return on asset i, and σ_{ij} is the covariance between the returns of asset i and asset j; there are N assets.

Under normality, the expectation of r_i conditional on the returns of the remaining $(N-1)$ assets is a linear function of these remaining returns. Specifically, if R^i is the vector of returns with the return of asset i deleted, the conditional expected return is given by

$$E[r_i|R^i] = \sum_{k \neq i} w_k r_k \ , \tag{5.1}$$

where w_k are weights appropriate to asset k. From normal theory, the weights themselves are given by

---

[3]Other uses would suggest different criteria. An index arbitrageur might want to use a return-generating model to construct a portfolio of a limited number of stocks to mimic the S&P 500 index. In this case, a natural way to evaluate a model is to use the estimated model to form a portfolio of securities that maximizes the correlation of its return with the S&P 500 and at the same time matches the variance of the S&P 500. One way to validate such a model is to compare in a subsequent period the profits from an arbitrage strategy using the mimicking portfolio with those using all 500 stocks.

[4]A Bayesian justification of the use of a mean-squared error rests upon an investor loss function. If an investor's loss function is quadratic, the natural measure of loss is the mean-squared error.

$$W^i = (\Sigma^i)^{-1} C^i \tag{5.2}$$

where W^i is a column vector of the $(N-1)$ weights, C^i is a column vector of the covariances of the returns of asset i with respect to each of the other $(N-1)$ assets, and Σ^i is a square matrix with dimension $(N-1)$ obtained by deleting the ith row and ith column of the full covariance matrix of all N securities.

The weights, given by (5.2), have the important property that they minimize the variance of r_i conditional on R^i.[5] This is not a surprising result since these weights are nothing more than the expected value of the estimated coefficients of a regression of r_i on the returns of the remaining $(N-1)$ assets. The essence of least-squares regression is to minimize mean-squared errors.

Thus, the process of estimating a least-squares regression can be viewed as consisting of two steps: First, estimate the covariance matrix of the dependent and independent variables. Second, use this estimated matrix to estimate the regression coefficients, which can then be used as the weights in Eq. (5.1). Viewing a regression this way helps clarify the role of factor models in forming conditional expectations.

Using a factor model to assess conditional expected returns is similar to a regression but with an important exception: Factor models place restrictions on the structure of the covariance matrix of returns, whereas the usual least-squares regression places no restrictions on this matrix. To develop these restrictions, consider the factor model:

$$r_{it} = \sum_{k=1}^{K} \lambda_{ik} f_{kt} + \eta_{it} \ , \tag{5.3}$$

where K is the number of factors, λ_{ik} is the so-called factor loading of asset i on factor k, f_{kt} is the score or value of factor k during interval t, and η_{it} is a mean-zero independent disturbance. The expected value of f_{kt} is zero, and it is scaled so that $\sigma(f_{kt})$ is 1.0. In addition, estimation of a factor model requires some assumption about the covariances between the different factors. The usual assumption, which is also made in this paper, is that $\text{Cov}(f_{kt}, f_{jt}) = 0, k \neq j$.

From (5.3), the variance of the return of asset i is

[5]The vector of weights W^i are those that minimize

$$E(r_i - W^i R^i)^2,$$

which can be rewritten as

$$\sigma^2(r_i) - W^{i'} C^i + W^{i'} \Sigma^i W^i.$$

Minimizing this expression with respect to W^i yields Eq. (5.2) above.

$$\sigma^2(r_{it}) = \sum_{k=1}^{K} \lambda_{ik}^2 + \sigma^2(\eta_{it}) , \tag{5.4}$$

and the covariance between the returns of assets i and j is

$$\text{Cov}(r_{it}, r_{jt}) = \sum_{k=1}^{K} \lambda_{ik}\lambda_{jk} . \tag{5.5}$$

Within the estimation period and within the class of linear estimators, estimates of the conditional expected returns for asset i that place no restrictions on the estimated covariance matrix will mathematically produce the minimum mean-squared errors. However, outside the estimation period, there is no guarantee that such an unrestricted estimate of the covariance matrix will yield the minimum mean-squared errors, or even the minimum expected mean-squared errors. If the restrictions that factor models impose on the covariance matrix are valid, it is possible that calculating conditional expected returns using a covariance matrix estimated with restrictions will yield lesser mean-squared errors in the prediction period than using an unrestricted estimate.

Non-stationarities complicate the story. Without restrictions, an estimate of the covariance matrix may "discover" non-existent relations among the returns. With restrictions, an estimate of the covariance matrix may be less prone to discover non-existent relations. In turn, it is possible that restrictions, even if not perfectly true, may improve the accuracy of conditional expectations out of the estimation period. Validating various models with different data from those used in estimating the models provides some insight into these two issues: restrictions on the estimated covariance matrix and the effect of non-stationarities.

5.2 The Experiment

The first part of this section describes the data. The second and third parts analyze the conditional expected returns, or predictions, based upon these models. The fourth part examines the impact of a January seasonal on the factor results. The fifth part decomposes the mean-squared forecast errors into the sources of the errors. The final part compares the predictions of factor models and standard market models with models that use prespecified macro variables of Chen et al. (1986).

5.2.1 Data

The empirical analyses use monthly returns of 82 sets of size-ranked portfolios of NYSE stocks constructed from the CRSP file. The first set consists of all securities in the CRSP files with complete data for the six years 1926 through 1931.

These securities were ranked by their market value as of December 1930 and then partitioned into twenty size-ranked portfolios with as close to an equal number of securities as possible. This process was repeated year by year to 2012. The sixth year in each set will be used to identify the set, so that the first set is the 1931 set and the last set is the 2013 set. The total number of securities used in the analysis starts at 361 for the 1931 set, increases to 763 for the 1949 set, and then gradually reaches 1790 for the 2012 set. In anticipation of the validation tests, the first five years of each data set will be used to estimate a model, and the sixth year will be used to validate the model.

An analysis of the basic data discloses dramatic changes in the variability of the returns of the portfolios over time. The variability is greatest in the 1930s, but even in the later years, the variability does change somewhat from one year to the next (Fig. 5.1). For most years, the smaller portfolios display greater variability in returns than the larger portfolios.[6] These changes in variability make summary measures of mean-squared errors misleading without some adjustment for these changes, and such adjustments will be made as discussed below.

For the Fama–French 3-factor model we use the data provided by Kenneth French.[7] The factors $R_m - R_f$, *SMB*, and *HML* are constructed from six size/book-to-market benchmark portfolios that do not include hold ranges and do not incur transaction costs.

1. $R_m - R_f$, the excess return on the market, is the value-weighted return on all NYSE, AMEX, and NASDAQ stocks (from CRSP) minus the one-month Treasury bill rate (from Ibbotson Associates).
2. *SMB* (Small Minus Big) is the average return on three small portfolios minus the average return on three big portfolios, $SMB = 1/3(SmallValue + SmallNeutral + SmallGrowth) - 1/3(BigValue + BigNeutral + BigGrowth)$.
3. *HML* (High Minus Low) is the average return on two value portfolios minus the average return on two growth portfolios, $HML = 1/2(SmallValue + BigValue) - 1/2(SmallGrowth + BigGrowth)$.

5.2.2 Factor Models

We use the maximum likelihood method to estimate the factor models; the usual way to assess the number of required factors is to rerun the procedure, successively increasing the number of factors until the χ^2 test for the goodness of fit developed by

[6]Interestingly, there is little change over time in the relative size of the portfolio consisting of the largest stocks, even though the market value of all of the portfolios increased almost tenfold from 1930 through 2012. In 1930, the market value of the portfolio with the largest companies is 51 % of the total market value of all twenty portfolios. By 2012, this number is 43 %.

[7]The data is publicly available from: http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

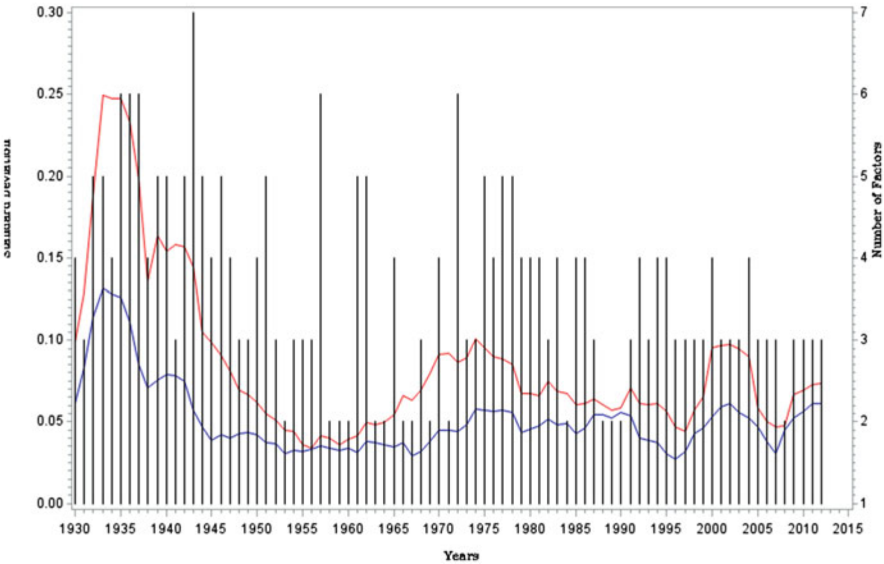

Fig. 5.1 *Portfolio return volatility and number of factors.* This figure shows the unconditional standard deviation of returns of the smallest and the largest capitalization portfolios and the adequacy of a k-factor model that generates the monthly stock returns on twenty-size ranked portfolios based on Bartlett's chi-squared test at 5X level of significance. Standard deviation and number of factors are estimated every year using the observations from the previous five years. Returns are monthly and measured as percentage changes

Bartlett (1954) indicates that the number of factors is sufficient. To use this criterion, one must specify the level of significance, often arbitrarily set at 1 or 5 %. The level of significance is important since there is a direct relation between the level of significance and the number of significant factors. However, there is no direct relation between this arbitrary level of significance and the criterion of minimizing the mean-squared errors in the forecast period.

To address the arbitrariness of setting a particular level of significance, this paper replicated the analysis for three levels of significance: 5, 10, and 20 %. For reasons to be discussed, the general nature of the results is the same whichever level of significance is used. To conserve space, the text presents only the results that use a significance level set at 5 %.

The number of required factors varies over time (Fig. 5.1). More factors are required at the beginning and the end of the 1930–2012 period than in the mid-part. Further analysis of the required number of factors reveals a positive relation between the number of factors and the variability of returns during the estimation period.[8]

[8] The Spearman's rank correlation between number of factors and the standard deviation of the equally weighted market portfolio over the sample period is 0.563, which is significant at any conventional level.

A rationale for this finding is that during periods of relatively low volatility, most of the volatility is firm-specific and it is difficult to identify the common factors. In more volatile times, the common factors are relatively more important than the firm-specific factors, making it easier to identify them.

The changing number of factors over times is strongly suggestive that the factor models are non-stationary. We conducted a series of simple Chow F-tests to formally test for stationarity.[9] We do not report these tests for brevity. The results confirm that the F-test rejects stationarity more often than could be attributed to chance and the χ^2 statistics are consistent with this impression.[10]

Since there appear to be significant non-stationarities in these factor models, validating the model with data different from those used in estimating the model is a useful tool in gaining insight into the usefulness of the model. As mentioned above, the process of validation used in this paper involves calculating conditional expectations following the estimation periods using Eqs. (5.1) and (5.2) and then analyzing the forecast errors.

The magnitudes of the mean-squared errors vary substantially over time and with portfolio size. Like the variability of the monthly returns, average mean-squared errors vary substantially for each twelve-month predictive period as a function of both time and the number of factors.[11] In view of this substantial variation, any summary measure of these mean-squared errors over time or portfolio size would be misleading without some form of scaling or normalization.

The scale factor used in this study is the mean-squared error associated with a naive forecast. The naive forecast is the average return for each portfolio in the estimation period, that is, an estimate of the unconditional expectation. An analysis of the scaled mean-squared errors shows that this normalization removes a large portion of the time trends in the annual mean-squared errors over time for a given portfolio size. However, substantial differences still remain among the size-ranked

[9]For alternative tests of non-stationarities and exploration of span of stationarity, see Hsu (1982, 1984).

[10]The stationarity test utilizes an F-statistic as proposed by Chow (1960). Although the Chow F-test was originally developed for linear regressions, it can be applied in a similar way to factor models. Specifically, estimate the factor model on the ten-year period consisting of the first five years 1926–1930 of the 1931 data set and the first five years 1931–1935 of the 1936 data set and then reestimate the factor model on the first half of the data and then on the second half, and so on. If the factor models are non-stationary, the fit of the estimated models in either half of the data will tend to be better than for the models estimated over the entire period. Specifically, if the factor model is stationary, the variances of the disturbances, η_{it} in (5.3), should be the same for all three models. The test of equality of sample variances is an F-test. Under the null hypothesis that the model is stationary and if the F-statistics are independent across portfolios, the probabilities of the F-statistics should be uniformly distributed. A χ^2-test rejects this hypothesis of an uniform distribution.

[11]For example, the average mean-squared errors for a twelve-month period ranged from 0.412 in 1944 to 26.931 in 1933 for the largest portfolio and from 1.446 in 1977 to 377.605 in 1935 for the smallest portfolio.

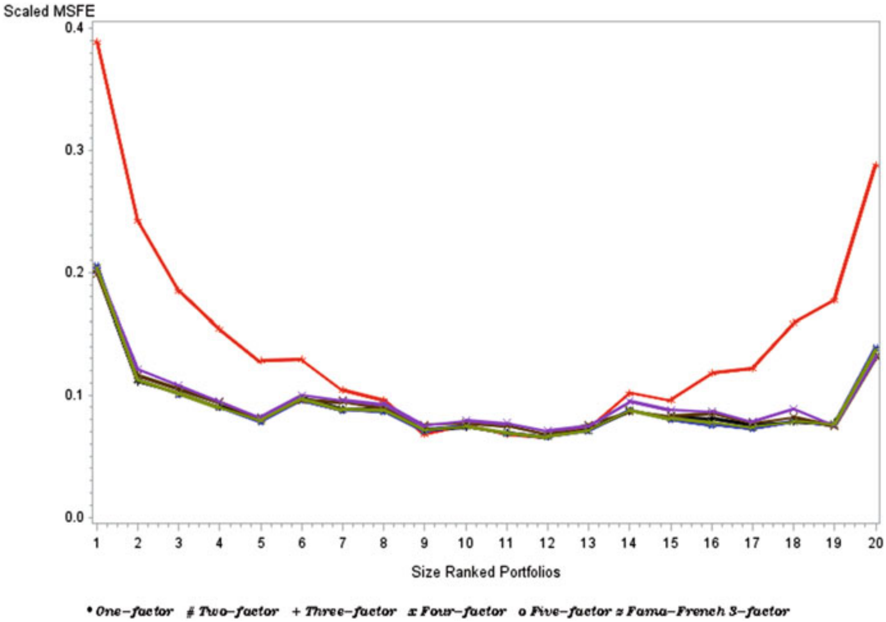

Fig. 5.2 *Scaled mean squared forecast errors for factor models* This figure shows the mean squared forecast errors scaled by the naive forecasts for size ranked portfolios. Portfolio 1 contains the smallest firms and portfolio 20 contains the largest firms. Scaled MSFE are averaged over the period 1931 to 2013

portfolios. As a consequence, the following tables and figures present summary statistics aggregated over time but not across portfolios of different sizes.

The validation of the factor models confirms the inferences based upon the χ^2 criterion that more than one factor is needed to represent the stochastic process generating returns for size-ranked portfolios. As one moves from a one-factor to a two-factor model, the mean-squared errors drop dramatically for both large and small portfolios, while there is little change for the mid-size (Fig. 5.2 and Table 5.1). As one moves to the three- or possibly four-factor model, the mean-squared errors for the large and small portfolios drop further, though only slightly. In addition, the minimum mean-squared error for the mid-size portfolios tends to occur with fewer factors than for the large or small portfolios. While we observe similar patterns for the Fama–French 3-factor model the mean-squared errors are consistently between two-factor and three-factor models.

The mean-squared errors in the forecast period for the factor models selected by the χ^2 criterion are slightly greater than the mean-squared errors associated with the best performing factor model in the forecast period for each portfolio size. The behavior of the mean-squared errors as a function of the number of factors leads to the conjecture that the arbitrary selection of two or three factors for mid-size

Table 5.1 Scaled-mean squared forecast errors for factor models for size ranked portfolios forecast periods 1931–2012

Estimation period								Forecast period								
	Factor models							Factor models						Model choice		
Portfolio size (1)	$k = 1$ (2)	$k = 2$ (3)	$k = 3$ (4)	$k = 4$ (5)	$k = 5$ (6)	$k = 6$ (7)	χ^2 at 5 % (8)	$k = 1$ (9)	$k = 2$ (10)	$k = 3$ (11)	$k = 4$ (12)	$k = 5$ (13)	FF-3 (14)	χ^2 at 5 % (15)	Min. MSFE (16)	Median k-factor (17)
1-small	0.231	0.100	0.090	0.085	0.080	0.094	0.086	0.389	0.205	0.200	0.199	0.202	0.203	0.202	0.202	4
2	0.166	0.072	0.067	0.063	0.059	0.069	0.064	0.242	0.113	0.111	0.116	0.121	0.112	0.117	0.121	3
3	0.135	0.067	0.061	0.058	0.055	0.063	0.059	0.185	0.101	0.102	0.105	0.108	0.101	0.107	0.109	3
4	0.108	0.060	0.056	0.053	0.050	0.058	0.055	0.154	0.090	0.090	0.094	0.095	0.090	0.091	0.095	3
5	0.095	0.059	0.057	0.055	0.052	0.058	0.055	0.128	0.078	0.081	0.081	0.082	0.079	0.082	0.082	3
6	0.082	0.058	0.054	0.051	0.048	0.056	0.053	0.129	0.096	0.098	0.096	0.100	0.097	0.097	0.100	3
7	0.073	0.055	0.053	0.049	0.047	0.054	0.051	0.104	0.088	0.089	0.095	0.096	0.088	0.090	0.096	3
8	0.070	0.056	0.053	0.050	0.048	0.054	0.052	0.096	0.087	0.089	0.090	0.093	0.088	0.088	0.092	3
9	0.061	0.053	0.051	0.049	0.045	0.052	0.049	0.068	0.071	0.072	0.076	0.075	0.071	0.074	0.075	2
10	0.055	0.052	0.048	0.045	0.044	0.050	0.047	0.076	0.075	0.074	0.077	0.079	0.075	0.077	0.079	2
11	0.051	0.049	0.045	0.042	0.040	0.047	0.044	0.068	0.069	0.069	0.075	0.077	0.069	0.070	0.076	2
12	0.053	0.048	0.044	0.043	0.040	0.046	0.043	0.066	0.066	0.066	0.068	0.071	0.066	0.071	0.071	2
13	0.056	0.049	0.045	0.043	0.041	0.047	0.044	0.073	0.071	0.072	0.076	0.075	0.071	0.074	0.075	2
14	0.066	0.051	0.048	0.045	0.042	0.049	0.045	0.102	0.088	0.087	0.087	0.095	0.088	0.088	0.095	3
15	0.080	0.055	0.050	0.047	0.044	0.052	0.049	0.096	0.080	0.082	0.083	0.088	0.081	0.084	0.088	3
16	0.084	0.049	0.046	0.043	0.041	0.047	0.045	0.118	0.076	0.081	0.085	0.087	0.078	0.084	0.087	2
17	0.092	0.049	0.047	0.045	0.043	0.048	0.046	0.122	0.073	0.076	0.077	0.078	0.074	0.077	0.078	2
18	0.099	0.049	0.045	0.042	0.040	0.046	0.044	0.159	0.079	0.078	0.082	0.089	0.079	0.079	0.089	3
19	0.131	0.053	0.046	0.043	0.040	0.048	0.044	0.178	0.076	0.077	0.075	0.076	0.076	0.077	0.075	3
20-large	0.202	0.088	0.077	0.071	0.065	0.080	0.074	0.288	0.139	0.131	0.131	0.132	0.137	0.136	0.130	3

(continued)

Table 5.1 (continued)

Factor model forecasts are made within and out of the sample or estimation period. Columns (2) through (9) are within sample forecast errors and Columns (8) through (16) are out of sample forecast errors. For the out of the sample period, forecasts for the month s of the year t are obtained by $r_{it+s} = \lambda_i f^{(i)}_{t+s}, s = 1, 2, \ldots, 12$, where λ_i is a $k \times 1$ vector of the factor loadings for the ith portfolio, and $f^{(i)}_{t+s}$ is a k-element column vector containing the factor scores for the prediction period $t + s$. Factor scores, $f^{(i)}_{t+s}$, are computed using the factor loadings from the estimation period and realized returns at time $t + s$. Formally, $f^{(i)}_{t+s} = \Lambda'^{(i)}[(\Lambda^{(i)}\Lambda'^{(i)} + \Omega^{(i)})]^{-1} R^{(i)}_{t+s}$, where $\Lambda^{(i)}$ is a $19 \times k$ matrix of factor loadings, $\Omega^{(i)}$ is a 19×19 diagonal matrix of specific variances, and $R^{(i)}_{t+s}$ is a 19×1 vector containing portfolio returns at time $t + s$. Superscript (i) denotes that the ith element of a vector or matrix corresponding to the ith portfolio is deleted. Returns are in deviation form from the sample mean over the estimation period. Factor model parameters, Λ and Ω are re-estimated every year using the previous 5 years' data, estimation period; forecasts and forecast errors are estimated for the next 12 months. The procedure is repeated by updating the parameter estimates of the factor models every year as a 5-year moving window. The first estimation period is from 1926 to 1930 with the corresponding forecast year being 1931. The last year of forecasts is 2013. Squared differences between realized returns and forecasts are defined as squared forecast errors. Mean squared forecast errors (MSFE) are the average of squared forecast errors for the period from 1931 to 2013

Forecast errors within sample period are the residuals of the fitted model within the same period. Within sample period forecast errors are also re-estimated every year from 1926 to 2012. Naive forecast for a year is defined as the mean portfolio returns over the estimation period

Portfolios 1 through 20 [Column (1)] contain equal number of size ranked companies. Portfolio 1 contains firms with the smallest market value while portfolio 20 the largest. Market values at the end of estimation period are used for ranking. MSFEs in Columns (2) through (7) and (9) through (14) are for one to five-factor and FF-3 models that are assumed to generate the returns. Each of these factor models is separately estimated. MSFEs in other columns are obtained from: (a) χ^2 criterion [Column (8) and (15)], a k-factor model is chosen every year based on the Bartlett's χ^2 test during the first stage of the estimation as described in Fig. 5.1; (b) Min. MSFE [Column (16)] is average of forecast errors for the factor model which produced the smallest MSFE during the estimation period. Median k-Factor model is the median number of factors that yields minimum forecast errors for out of sample period [Column (16)]; within sample period the five-factor model produces the smallest MSFEs portfolios and three or four factors for the largest and smallest portfolios leads to lesser mean-squared errors than using the standard χ^2 test.

The χ^2 criterion yields little difference in the mean-squared errors among different levels of commonly used significance, because any criterion that points to two to five factors leads to similar mean-squared errors. With a significance level of 5 %, the median number of factors over the 82 estimation periods is 4; with a significance level of 10 %, the median number is also 4; and with a significance level of 20 %, the median number is 3.

For comparison with the predictions in the forecast period, Table 5.1 also contains the average mean-squared errors for the conditional expectations within the estimation period. In contrast to the predictions in the validation period, the average mean-squared errors decrease monotonically for each portfolio as the number of factors increases from one to five. On the surface, this result suggests that the greater the number of factors the better. However, the validation of the models with additional data shows that there is little difference between models with anywhere from two to five factors.

5.2.3 The Market Model

If more than one factor in the process generates returns, the mean-squared errors from factor models should be smaller than those from the usual market model, given by

$$r_i = \beta_i r_m + \epsilon_i \tag{5.6}$$

where r_m is the return on a market index, again with all returns measured from their unconditional expectations. The measure of the market is alternatively a value-weighted or an equally weighted index of NYSE stocks. The associated covariance matrix for the market model is

$$\begin{bmatrix} \sigma^2(r_i) & \beta_i\sigma^2(r_m) \\ \beta_i\sigma^2(r_m) & \sigma^2(r_m) \end{bmatrix} \tag{5.7}$$

Applying Eq. (5.2) yields the conditional expectation $E(r_i|r_m)$ as $[\mathrm{Cov}(r_i, r_m)/\sigma^2(r_m)]r_m$, the usual conditional forecast for the market model. It should be noted that asset i is included in the market portfolio, and thus the conditional forecast of r_i is partially conditioned by itself, a fact of importance in explaining the behavior of the mean-squared errors for the portfolio with the largest companies.

As with the factor models, there is substantial evidence of non-stationarity in the market models using either the equally weighted index or the value-weighted index. In view of this possible non-stationarity, it is appropriate to validate either variant of the market model with subsequent data. Generally, the mean-squared errors for

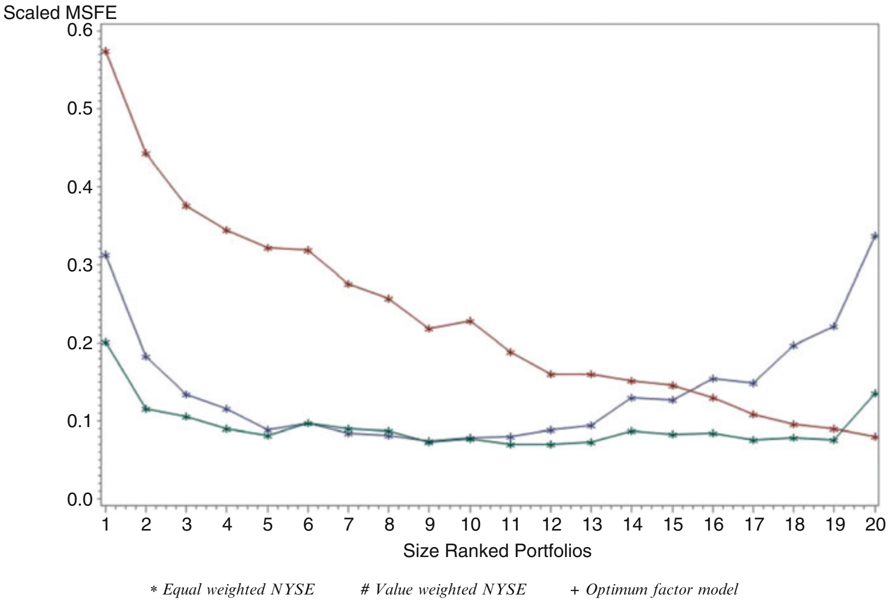

Fig. 5.3 *Scaled mean squared forecast errors for the market model and the factor model.* This figure compares the mean squared forecast errors scaled by the naive forecasts for market model and the factor model for size ranked portfolios. For market model equally and value weighted New York Stock Exchange index are used. Choice of factor model is based upon the chi–square tests of factor analysis. Scaled forecast errors are averaged over the period 1931–2012

the market model are greater than those for the factor models (Fig. 5.3). The glaring exception is the largest portfolio using a value-weighted index. Since the stocks in the largest portfolio represent an extremely large proportion of a value-weighted index of NYSE stocks and since this index is used to forecast the returns of this portfolio, this result is not surprising. Except for the largest five portfolios, the mean-squared errors associated with the equal-weighted index are less than those associated with the value-weighted index.

5.2.4 A January Seasonal

A large body of literature shows that the distribution of stock returns in January is different from the distribution of stock returns in other months. Keim (1983) found significant differences in the returns of small and large stocks in January. Tinic and West (1984) showed that virtually all of the relation between returns and betas in tests of the Capital Asset Pricing Model is due to a January seasonal. Gültekin and Gültekin (1987) demonstrated that the same is true for the two-stage tests of the Arbitrage Pricing Model.

Likewise, the factor models estimated in this paper display a January seasonal in the mean-squared errors. For every size portfolio, the mean-squared errors for January are uniformly greater than those for the other months of the year. In the case of the smallest portfolio, the mean-squared errors for January are over three times as great as the mean-squared errors for the remaining months.

This January seasonal raises the question of whether the better forecasting characteristics of a multi-factor model may be due solely to the returns in January. To answer this question, we reestimated the factor models excluding the January returns in each estimation period. According to the χ^2 criterion at a level of 5 %, the median number of required factors drops from four to three.

Even with January excluded, the minimum scaled mean-squared forecast errors still tend to occur with more than one factor (Table 5.2). For the forecast errors for February through December, two-factor models yield smaller scaled mean-squared errors than one-factor models in all cases except one mid-size portfolio. Although January was excluded in the estimating period, the estimated models still can be used to forecast January returns. For these January returns, two-factor models yield smaller scaled mean-squared errors than one-factor models in all but four cases. Thus, the presence of more than one factor is not due just to a January seasonal.[12]

5.2.5 *Biases and Inefficiencies*

Theil's decomposition shows that most of the mean-squared forecast error is random, except for the smallest portfolio, regardless of which forecasting model is used (Table 5.3).[13] The random component almost always accounts for over 90 % of the mean-squared forecast errors, and frequently accounts for over 95 % for all but the smallest portfolio.

Still, some differences among the various models warrant mention. The biases associated with the market model using an equal-weighted index of NYSE stocks are smallest for the mid-size portfolios and increase as the size of the stocks in the portfolio becomes more extreme—either larger or smaller. The largest bias, 7.6 %, is associated with the smallest portfolio. The biases for the market model using a value-weighted portfolio of NYSE stocks are similar for the mid-size portfolios

[12]To determine the importance of a January seasonal, we replicated the early analysis including, in addition to the returns on the twenty portfolios, a variable with a value of 1.0 for the months of January and 0.0 otherwise. According to the χ^2 criterion at a level of 5 %, the median number of required factors is four as before. However, there is virtually no improvement in the mean-squared errors. Again the number of factors that minimize the mean-squared error is less for the mid-size portfolios than the large or small portfolios.

Although there is a January seasonal, directly incorporating such a variable does not improve the mean-squared errors. The returns themselves already capture this seasonal. Thus, a January seasonal of itself does not account for the presence of more than one factor.

[13]See Theil (1966), and Mincer and Zarnovitz (1969) for details.

Table 5.2 Scaled mean squared forecast errors for factor models for size ranked portfolios: out of sample period forecasts adjusted for January seasonal

Forecast periods 1931–2013															
	January only					February–December					All months				
	k-factor model					k-factor model					k-factor model				
Portfolio size (1)	$k = 1$ (2)	$k = 2$ (3)	$k = 3$ (4)	$k = 4$ (5)	$k = 5$ (6)	$k = 1$ (7)	$k = 2$ (8)	$k = 3$ (9)	$k = 4$ (10)	$k = 5$ (11)	$k = 1$ (12)	$k = 2$ (13)	$k = 3$ (14)	$k = 4$ (15)	$k = 5$ (16)
1-small	1.549	0.571	0.552	0.529	0.537	0.282	0.178	0.179	0.177	0.177	0.387	0.210	0.210	0.207	0.207
2	0.871	0.219	0.212	0.199	0.200	0.186	0.104	0.105	0.110	0.114	0.243	0.114	0.114	0.117	0.121
3	0.588	0.198	0.197	0.237	0.218	0.149	0.098	0.100	0.103	0.108	0.186	0.106	0.108	0.115	0.117
4	0.500	0.135	0.120	0.113	0.123	0.125	0.084	0.085	0.088	0.092	0.157	0.088	0.088	0.090	0.095
5	0.273	0.102	0.114	0.125	0.132	0.117	0.078	0.080	0.081	0.083	0.130	0.080	0.083	0.085	0.087
6	0.309	0.136	0.136	0.129	0.137	0.114	0.093	0.095	0.094	0.100	0.130	0.096	0.099	0.096	0.103
7	0.135	0.136	0.144	0.136	0.135	0.101	0.085	0.089	0.095	0.097	0.104	0.089	0.093	0.099	0.100
8	0.133	0.102	0.098	0.100	0.119	0.093	0.085	0.088	0.090	0.092	0.096	0.087	0.089	0.091	0.094
9	0.088	0.130	0.126	0.120	0.111	0.068	0.068	0.072	0.073	0.075	0.069	0.073	0.077	0.077	0.078
10	0.101	0.107	0.111	0.107	0.112	0.075	0.073	0.072	0.077	0.079	0.077	0.076	0.075	0.080	0.082
11	0.086	0.092	0.085	0.126	0.117	0.066	0.068	0.068	0.072	0.077	0.068	0.070	0.070	0.076	0.080
12	0.145	0.118	0.116	0.122	0.105	0.059	0.062	0.065	0.067	0.071	0.066	0.067	0.069	0.072	0.073
13	0.133	0.120	0.112	0.123	0.129	0.067	0.067	0.069	0.071	0.073	0.073	0.071	0.072	0.076	0.077
14	0.217	0.150	0.147	0.126	0.124	0.093	0.087	0.088	0.089	0.096	0.103	0.092	0.093	0.092	0.098
15	0.185	0.113	0.099	0.100	0.112	0.088	0.077	0.081	0.082	0.090	0.096	0.080	0.083	0.084	0.092
16	0.380	0.138	0.146	0.165	0.164	0.093	0.071	0.075	0.080	0.081	0.117	0.077	0.081	0.087	0.088
17	0.214	0.087	0.099	0.110	0.097	0.112	0.072	0.073	0.076	0.078	0.121	0.073	0.075	0.079	0.079
18	0.498	0.117	0.120	0.122	0.112	0.134	0.079	0.081	0.088	0.089	0.164	0.083	0.084	0.091	0.091
19	0.569	0.103	0.117	0.101	0.102	0.145	0.074	0.076	0.075	0.077	0.181	0.077	0.079	0.077	0.079
20-large	1.014	0.225	0.218	0.236	0.222	0.231	0.133	0.127	0.127	0.126	0.296	0.140	0.135	0.136	0.134

Table 5.2 (Continued)

Parameters of the factor models are estimated by excluding the January return during the estimation period. Forecasts for the month s of the year t are obtained by $r_{it+s} = \lambda_i' f_{t+s}^{(i)}$, where λ_i is a $k \times 1$ vector of the factor loadings for the ith portfolio, and $f_{t+s}^{(i)}$ is a k-element column vector containing the factor scores for the prediction period $t + s$. Factor scores, $f_{t+s}^{(i)}$, are computed using the factor loadings from the estimation period and realized returns at time $t + s$. Formally, $f_{t+s}^{(i)} = \Lambda^{`(i)}[(\Lambda^{(i)}\Lambda^{`(i)} + \Omega^{(i)}]^{-1} R_{t+s}^{(i)}$, where $\Lambda^{(i)}$ is a $19 \times k$ matrix of factor loadings, $\Omega^{(i)}$ is a 19×19 diagonal matrix of specific variances, and $R_t^{(i)}$ is a 19×1 vector containing portfolio returns at time $t + s$. Superscript (i) denotes that the ith element of a vector or matrix corresponding to the ith portfolio is deleted. Returns are in deviation form from the sample mean over the estimation period. Factor model parameters, Λ and Ω are re-estimated every year using the previous 5 years' data, estimation period; forecasts and forecast errors are estimated for the next 12 months. The procedure is repeated by updating the parameter estimates of the factor models every year as a 5-year moving window. The first estimation period is from 1926 to 1930 with the corresponding forecast year being 1931. Last year of forecasts is 2013. Squared differences between realized returns and forecasts are defined as squared forecast errors. Mean squared forecast errors (MSFE) are the average of squared forecast errors for the period from 1931 to 2013. Forecast errors for the year are scaled by the root mean squared naive forecast error. Naive forecast for a year is defined as the mean portfolio returns over the estimation period

Portfolios 1 through 20 [Column (1)] contain equal number of size ranked companies. Portfolio 1 contains firms with the smallest market value while portfolio 20 the largest. Market values at the end of estimation period are used for ranking. Number of factors [Columns (2) through (16)] indicate the factor model which is assumed to generate the returns. Each of the factor models, one to five, is separately estimated

Columns (2)–(6) present mean squared forecast errors for January only, Columns (7)–(11) for 11 months from February to December, and Columns (12)–(16) for all months

Table 5.3 Decomposition of scaled mean-squared forecast errors into sources of errors forecast periods 1931–2013

	Market model				One-factor model				Five-factor model			
		Percentage source of error				Percentage source of error				Percentage source of error		
Portfolio size (1)	Scaled MSFE (2)	Bias (3)	Inefficiency (4)	Random errors (5)	Scaled MSFE (6)	Bias (7)	Inefficiency (8)	Random errors (9)	Scaled MSFE (10)	Bias (11)	Inefficiency (12)	Random errors (13)
1-small	0.313	7.64	12.05	80.32	0.389	7.26	15.94	76.80	0.202	3.19	11.15	85.66
2	0.183	4.80	5.37	89.82	0.242	5.02	9.27	85.71	0.121	0.16	3.22	96.62
3	0.135	6.25	3.54	90.21	0.185	6.61	6.91	86.48	0.108	0.79	0.81	98.40
4	0.117	3.10	2.47	94.43	0.154	4.29	5.59	90.13	0.095	0.05	0.67	99.29
5	0.089	1.48	1.04	97.48	0.128	2.51	3.26	94.23	0.082	0.02	0.63	99.34
6	0.097	0.47	2.28	97.25	0.129	1.32	4.57	94.11	0.100	0.04	1.55	98.41
7	0.085	0.00	1.52	98.47	0.104	0.33	3.50	96.18	0.096	0.29	1.13	98.59
8	0.082	0.00	0.19	99.81	0.096	0.37	1.08	98.54	0.093	0.63	0.42	98.96
9	0.075	0.20	0.05	99.75	0.068	0.11	0.47	99.42	0.075	0.15	0.05	99.81
10	0.079	0.60	2.52	96.88	0.076	0.01	3.96	96.03	0.079	0.09	2.36	97.55
11	0.080	0.90	0.92	98.18	0.068	0.08	1.43	98.49	0.077	0.04	0.88	99.08
12	0.090	4.11	0.37	95.52	0.066	2.78	0.68	96.53	0.071	1.06	1.25	97.69
13	0.094	3.47	1.43	95.10	0.073	2.08	2.05	95.88	0.075	0.34	2.02	97.64
14	0.130	3.11	0.42	96.47	0.102	2.03	0.35	97.62	0.095	0.26	0.00	99.74
15	0.128	3.75	2.85	93.40	0.096	2.77	3.33	93.90	0.088	0.24	2.45	97.31
16	0.155	2.71	1.55	95.75	0.118	1.81	1.44	96.76	0.087	0.05	0.98	98.97
17	0.149	1.94	2.44	95.62	0.122	1.04	2.22	96.74	0.078	0.24	1.29	98.47
18	0.197	3.71	0.74	95.55	0.159	2.79	0.51	96.71	0.089	0.03	0.01	99.96
19	0.222	4.77	3.04	92.19	0.178	4.09	2.67	93.24	0.076	0.22	1.86	97.93
20-large	0.337	3.02	3.14	93.84	0.288	2.34	2.61	95.06	0.132	0.32	0.45	99.23

This table shows the decomposition of scaled mean squared forecast errors (MSFE) generated by market model (using equally weighted NYSE) and one- and five-factor models. Using Theil's method, scaled MSFEs from these models (in columns 2, 6, and 10) are decomposed into three sources of errors: Bias in columns 3, 7, and 11; Inefficiency in columns 4, 8, and 12; and Random Error in columns 5, 9, and 13. Formally: $\frac{1}{N}\sum(A_t - F_t)^2 = (\overline{A} - \overline{F})^2 + (1-\beta)^2\sigma_F^2 + (1-\rho^2)\sigma_A^2$ or, MSFE = Bias + Inefficiency + Random error, where σ_A^2 is the variance for actual returns (A) and σ_F^2 is the variance for the forecasted returns (F), ρ is the correlation between actual returns and their forecasts, and β is the slope coefficient of the regression of actual returns on forecasts

MSFE are the average of squared forecast errors for the period from 1931 to 2013. Forecast errors for the year are scaled by the root mean squared naive forecast error. Naive forecast for a year is defined as the mean portfolio returns over the estimation period (not reported in the tables). However, the biases are substantially larger for the large portfolios than for the smallest portfolios.[14] The behavior of the biases for the one-factor model is similar. With two or more factors, the biases are minimal with the exception of the smallest portfolio. But even for the smallest portfolio, the percentage biases for a five-factor model (as well as a two-, three-, or four-factor model) are nearly half of those for either the market or a one-factor model.

For all the models, the forecast errors for the small portfolios display the greatest inefficiency. As the number of factors increases to two or more, the inefficiency of the forecasts declines markedly. Again, the multi-factor models' forecasting characteristics are better than either the market model or a one-factor model.

5.2.6 Macroeconomic Variables

A growing body of research uses prespecified macroeconomic variables to estimate conditional moments of stock returns. Prespecifying macroeconomic variables overcomes one of the major difficulties of factor analysis: how to associate the estimated factors with observable and economically meaningful variables. As an example, Chen et al. (1986) used some directly observable macroeconomic variables as proxies for factors in the two-stage tests of the multi-factor pricing models in a way analogous to the use of instrumental variables in regression models.

Models incorporating macroeconomic variables can be validated in much the same way as validating the market model. Estimate the model using one set of data and validate it with a different set. For each portfolio, a regression of a time series of returns on the macro variables provides the estimated model. As before, all variables are measured from their unconditional means as estimated in the estimation period, and the validation of the estimated models uses data from the 12 months following the estimation period.

Chen, Roll, and Ross provide a detailed discussion of the selection of their macro variables. Their final list of variables is the following:

1. the equal- or value-weighted NYSE index
2. the monthly growth rate of the industrial production index, measured as $\log(I\!PI_{t+1}/I\!PI_t)$, where $I\!PI_t$ is the industrial production index for the month t
3. unanticipated inflation, measured as the difference between the realized inflation for the month t and the monthly T-Bill rate at the beginning of the month (see Fama and Gibbons (1984) for details)
4. the change in the term structure, measured by the difference between the return of a portfolio of long-term government bonds and the T-Bill rate

[14] The percentage biases are 9.6, 15.6, and 16.3 % for portfolios 18, 19, and 20, respectively, and 3.0 % for the smallest portfolio.

5. changing risk premia, measured by the return on BAA rated non-convertible corporate bonds less the return on a portfolio of long-term government bonds

The validation tests in this section use these same variables. Since the industrial production data are available only after 1946 and the estimation period requires five years of data, the first forecast year is 1952 and the last 2013. For comparison purposes, some of the earlier analyses have been replicated for these years.[15]

The validation process suggests that the macro variables by themselves have no forecasting power (Table 5.4), with the scaled mean-squared errors in the validation period ranging from 1.134 to 1.271. Since these statistics have been scaled by the mean-squared errors of the naive forecasts, a statistic greater than one indicates that the naive forecasts are more accurate than those using just the macro variables. Within the estimation period, the macro variables by themselves do have some explanatory power, with the scaled mean-squared errors ranging from 0.577 to 0.909. These two results imply that the regression in the estimation period found a relation that was not there, or that any relation in the estimation was not sufficiently stationary to provide forecasting power, or some combination of the two.

Adding either the equal-weighted or value-weighted index of NYSE stocks to the macro variables leads to a substantial reduction in the scaled mean-squared errors for every portfolio. As an example, the average scaled mean-squared errors for the largest portfolio is 1.160 with just the macro variables, but drops to 0.461 with the addition of the equal-weighted index. Even more accurate are the forecasts that drop the macro variables and include just a stock market index, suggesting that the macro variables merely add noise to the forecasts. Again, the multi-factor models generally yield smaller scaled mean-squared errors than either version of the market model.

5.3 Conclusions

The goal of this paper was to validate various stochastic return-generating models on data different from those used in estimating the models. The specific models analyzed were factor models, the traditional market model, and models incorporating prespecified macroeconomic variables. The principal conclusion of this paper is that factor models with two to five factors yield more accurate predictions than either the traditional market model or a one-factor model.

A model that included the prespecified macroeconomic variables used by Chen et al. (1986) had no predictive power. Thus, at least for the macro economic variables considered here, there is no gain to adding these variables to the traditional market model. But importantly, the predictions of a multi-factor model were more accurate than the market model.

Acknowledgements We thank Craig MacKinlay and Jennifer Conrad for their careful and thoughtful comments.

[15] The stationarity tests again tend to reject stationarity.

Table 5.4 Comparison of scaled mean-squared forecast errors using macroeconomic variables, market model and factor model

Forecast periods 1952–2013						
Portfolio size (1)	Macro variables (2)	Macro & equal-weighted index (3)	Equal-weighted index (4)	Macro & value-weighted index (5)	Value weighted index (6)	Optimal factor model (7)
1-small	1.247	0.386	0.334	0.759	0.625	0.204
2	1.222	0.218	0.203	0.590	0.489	0.124
3	1.227	0.175	0.157	0.520	0.425	0.108
4	1.195	0.150	0.132	0.489	0.394	0.096
5	1.202	0.117	0.104	0.443	0.370	0.091
6	1.226	0.132	0.108	0.469	0.368	0.100
7	1.271	0.132	0.096	0.460	0.328	0.103
8	1.213	0.129	0.100	0.389	0.318	0.110
9	1.215	0.093	0.069	0.335	0.253	0.075
10	1.198	0.099	0.085	0.340	0.276	0.085
11	1.181	0.098	0.083	0.285	0.229	0.079
12	1.212	0.101	0.081	0.248	0.175	0.069
13	1.214	0.098	0.088	0.214	0.165	0.070
14	1.187	0.131	0.119	0.220	0.167	0.091
15	1.134	0.158	0.132	0.187	0.155	0.088
16	1.175	0.171	0.148	0.182	0.137	0.079
17	1.140	0.191	0.158	0.144	0.120	0.081
18	1.172	0.251	0.226	0.146	0.115	0.087
19	1.160	0.282	0.248	0.118	0.101	0.083
20-large	1.160	0.461	0.391	0.116	0.097	0.168

This table compares MSFE produced by four forecasting models. The first model uses a set of macro economic variables to make conditional forecasts (column 2). The second model includes market return as an additional exogenous variable to macro economic variables (column 3 for equally weighted NYSE index and column 5 for value weighted NYSE index). The third model forecasts are conditional on market returns (column 4 for equal weighted NYSE index and column 6 for value weighted NYSE index). The fourth model uses the optimal factor model based on the Bartlett's chi-square test during estimation period (column 7)

Model Parameters are re-estimated every year using the previous 5 years' data, estimation period; forecasts and forecast errors are estimated for the next 12 months. The procedure is repeated by updating the parameter estimates of the models every year as a 5-year moving window. The first estimation period is from 1947 to 1951 with the corresponding forecast year being 1952. Last year of forecasts is 2013. Forecast errors for the year are scaled by the root mean squared naive forecast error. Naive forecast for a year is defined as the mean portfolio returns over the estimation period. Mean squared forecast errors are the average of squared forecast errors for the period from 1952 to 2013

The macro variable are

1. Unanticipated inflation.
2. Monthly growth rate of industrial production.
3. Yield differential between BAA rated corporate bonds and long-term government bonds.
4. Yield differential between long-term government bonds and T-Bills.

The indexes are for all NYSE stocks

References

Bartlett. M. S. (1954). A note on the multiplying factors for various chi-squared approximations. *Journal of the Royal Statistical Society, 16*, 296–298.

Bloch, J. B. G., Jr., Markowitz, H. M., Todd, P., & Xu, G. L. (1993). A comparison of some aspects of the U.S. and Japanese equity markets. *Japan & the World Economy, 5*, 3–26.

Brown, S. J., & Weinstein, M. (June 1983). A new approach to testing asset pricing theories: the bilinear paradigm. *Journal of Finance, 38*, 711–743.

Chen, N. -f. (December 1983). Some empirical tests of the theory of arbitrage pricing. *Journal of Finance, 38*, 1392–1414.

Chen, N. -f. (January 1988). Stable factors in security returns: identification using cross-validation - comment. *Journal of Business & Economic Statistics, 6* , 16.

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

Cho, D. C., Elton, E. J., & Gruber, M. J. (March 1984). 'On the robustness of the Roll and Ross APT methodology. *Journal of Financial and Quantitative Analysis, 19* , 1–10.

Chow, G. C. (July 1960). Tests of equality between sets of coefficients in two linear regressions. *Econometrica, 28* , 591–605.

Connor, G., & Korajczyk, R. A. (September 1988). Risk and return in an equilibrium APT: application of a new test methodology. *Journal of Financial Economics, 21* , 255–289.

Conway, D. A., & Reinganum, M. R. (January 1988). Stable factors in security returns: identification using cross-validation. *Journal of Business & Economic Statistics, 6* , 1–15.

Dhrymes, P. J., Friend, I., & Gültekin, N. B. (June 1984). A critical reexamination of the empirical evidence on the arbitrage pricing teory. *Journal of Finance, 39* , 323–346.

Dhrymes, P. J., Friend, I., Gültekin, N. B., & Gültekin, M. N. (March 1985a). An empirical examination of the implications of arbitrage theory. *Journal of Banking and Finance, 9* , 73–99.

Dhrymes, P. J., Friend, I., Gültekin, N. B., & Gültekin, M. N. (July 1985b). New tests of the APT and their implications. *Journal of Finance, 40*, 659–675.

Elton, E. J., & Gruber, M. J. (October 1971). Improved forecasting through the design of homogeneous groups. *Journal of Business, 44*, 432–450.

Elton, E. J., & Gruber, M. J. (December 1973). Estimating the dependence structure of share prices–implications for portfolio selection. *Journal of Finance, 8*, 1203–1232.

Fama, E. F., & Gibbons, M. (1984). A comparison of inflation forecasts. *Journal of Monetary Economics, 13*, 327–348.

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

Farrar, D. E. (1962). *The investment decision under uncertainty*. Englewood Cliffs, NJ: Prentice-Hall, Inc.

Farrell, J. (April 1974). Analyzing covariation of returns to determine homogeneous stock groupings. *Journal of Business, 47*, 186–207.

Guerard, J. B. Jr., Gultekin, M., & Stone, B. K. (1997). The role of fundamental data and analysts' earnings breadth, forecasts, and revisions in the creation of efficient portfolios. In: Chen, A. (Ed.), *Research in finance*, vol. 15, pp. 69–92.

Gültekin, M. N., & Gültekin, N. B. (December 1987). Stock return anomalies and the tests of the arbitrage pricing theory. *Journal of Finance, 42*, 1213–1224.

Gültekin, N. B., & Rogalski, R. J. (March 1985). Government bond returns, measurement of interest rate risk, and the arbitrage pricing theory. *Journal of Finance, 40*, 43–61.

Hsu, D. A. (March 1982). A Bayesian robust detection of shift in the risk structure of stock market returns. *Journal of the American Statistical Association, 77*, 29–39.

Hsu, D. A. (March 1984). The behavior of stock: Is it stationary or evolutionary? *Journal of Financial and Quantitative Analysis, 19*, 11–28.

Jensen, M. (April 1969). Risk, the pricing of capital assets, and the evaluation of investment portfolios. *Journal of Business, 42*, 167–247.

Keim, D. B. (1983). Size related anomalies and stock return seasonality: further empirical evidence. *Journal of Financial Economics, 12*, 13–32.

King, B. F. (January 1966). Market and industry factors in stock price behavior. *Journal of Business, 39*, 139–190.

Livingston, M. (June 1977). Industry movements of common stocks. *Journal of Finance, 32*, 861–874.

Meyer, S. (June 1973). A re-examination of market and industry factors in stock price behavior. *Journal of Finance, 8*, 645–709.

Mincer, J., & Zarnovitz, V. (1969). The valuation of economic forecasts. In J. Mincer (Ed.), *Economic forecasts and expectations*. New York: National Bureau of Economic Research.

Roll, R., & Ross, S. A. (December 1980). An empirical investigation of the arbitrage pricing theory. *Journal of Finance, 35*, 1073–1103.

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

Rosenberg, B. (1974). Extra-Market Components of Covariance in Security Returns. *Journal of Financial and Quantitative Analysis, 9,* 263–274.

Rosenberg, S. A., & Marathe, V. (1979). Tests of capital asset pricing hypotheses. In H. Levy (Ed.), *Research in finance* (Vol. 1). Greenwich: JAI Press.

Stambaugh, R. F. (January 1988). Stable factors in security returns: identification using cross-validation - comment. *Journal of Business & Economic Statistics, 6*, 20–21.

Theil, H. (1966). *Applied economic forecasting*. Amsterdam: North-Holland Publishing Company.

Tinic, S. M., & West, R. R. (1984). Risk and return: January vs. the rest of the year. *Journal of Financial Economics, 13*, 561–574.