Sizing up the errors in models of risk

Scott Gilbert[1]
Southern Illinois University Carbondale
and
Petr Zemčík[2]
CERGE-EI, Prague

July 12, 2008

Abstract

In models of risk, including the capital asset pricing model (CAPM) and the arbitrage pricing theory (APT), a key construct is investor *expectation* of returns on assets at a given date, given information on systematic risks at that date. The difference between actual and expected return is the model's *error*. In application the magnitude of such errors is important, and the present work develops statistics for sizing errors up, overall, across multiple assets and time periods. The statistics describe model performance, and like the fitted model they are sample-dependent, so we characterize their sampling distribution, which is non-standard. We apply the methods to CAPM and multi-factor models, using Fama-French stock data, and find that some previously reported differences between CAPM and multi-factor models seem exaggerated.

Key Words (+ JEL Codes): Asset Pricing (G12), Econometric/Statistical Estimation (C13), Models with Panel Data (C33)

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

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

1 Introduction

The stock market is a chance for ordinary households to share in firms' profits. The history of the stock market suggests that profits of diverse companies are related to each other, despite differences in industry, region, etc., thereby generating a systematic risk to firm values. For households, existence of inter-firm, macroeconomic, systematic risk means that a well-diversified stock portfolio is risky, though less risky than any individual stock.

Economic models, including the capital asset pricing model (CAPM) and the arbitrage pricing theory (APT), provide an opportunity to characterize rational responses to systematic risk. Both the CAPM and APT yield risk-conditioned expectations of the return on assets, and in rational, efficient markets these expectations should be unbiased, or nearly so, for most assets. The difference between actual and expected return is the model's error, which includes bias – if any exists – plus random, idiosyncratic variation in returns. Economic research on the CAPM and APT has focused on evidence of bias, and hence deviations from rationality or efficiency, based on statistical tests such as likelihood ratio, Wald, Lagrange multiplier, GMM, and F tests. Such tests are statements about statistical significance, not necessarily practical signficance.

We focus on the typical *size* of model errors, over time and across assets, for models in which systematic risk is represented by returns on one or more observable funds. We estimate total error and also its bias and variance components. Large estimated bias often suggests the existence of extraordinary returns on some assets, and in a highly influential paper Fama and French (1993) use this principle to choose among models – discarding as incredible models with large bias. Fama and French also prefer models with smaller error variance, and overall they proceed as if they care about the magnitude of total error, including bias and variance, with preference for small error.

We estimate the size of errors in models of systematic risk. In addition to point point estimates, we obtain closed-form expressions and computational formulas for distribution moments, under the assumption of normally distributed asset returns. By comparison, Geweke and Zhou (1996) compute numerically Bayesian posterior distributions for bias in an APT model with unobserved factors. We check the performance of our moment estimators via bootstrap simulation based on real stock return data, and find they work well despite apparent non-normality in the data.

The remainder of this work is organized as follows. Section 2 provides an economic context for the proposed methods, Section 3 proposes some statistics, Section 4 provides distributional theory for the new statistics, and Section 5 applies the methods to the stock market.

2 Economic Theory

Many economic models imply formulas for the expectations held by economic agents, and for each such expectation their is an associated *error*, equal to the difference betweeen realized value and predicted value. We want to examine, via econometrics, the size of such errors, in the case where investors hold expectations about the returns on capital assets. Even in this special setting there are numerous economics models, with acronyms like CAPM, APT, ICAPM, and CCAPM. Among these, the capital asset pricing model (CAPM) and the arbitrage pricing theory (APT) are static (one-period) models in which systematic risk plays a natural role. We will focus on errors in this sort of model, with portfolio choice among assets during a single period, and random asset returns which have a normal probability distribution.

Let $R_1, ..., R_n$ be the returns on n risky assets, with a probability distribution which is jointly normal $N(\mu, \Omega)$, with mean vector μ having typical element μ_i given by $E[R_i]$, and variance-covariance matrix Ω with typical element Ω_{ij} given by $\text{cov}(R_i, R_j)$. There is a riskless asset with interest rate R_0. Each investor has some endowment of assets, and chooses portfolio weights $\pi_0, \pi_1, ..., \pi_n$, which together sum to 1, earning a return on investment R_π equal to $\pi_0 R_0 + \pi_1 R_1 + \cdots + \pi_n R_n$. Investor's preferences among portfolios are represented by some function U of the portfolio return's mean μ_π and variance σ_π^2, and is not otherwise affected by the comovement of R_π with other factors, such as labor income. This setting is consistent with the CAPM model of Sharpe (1964) and Lintnter (1965), and the APT model of Ross (1976), but is inconsistent with the intertemporal CAPM (ICAPM) models of Merton (1973), Breeden (1979) and Campbell (1993).

Suppose that some investors' endowments yield them less utility than is possible via trade in financial markets. Let F be any commonly traded fund, composed of risky assets, purchased or sold by some investors to raise the utility of their portfolios. For example, F could be a market basket of stocks purchased by an investor with no endowment of stock, or it could be a basket of stocks for those firms having small market capitalization ("small cap"), etc. We treat information on such funds F as sufficient to represent systematic risk, and do not explicitly consider other risks such as fluctuation in economic growth, or inflation.

A fund F provides a return on investment, call it R_F. Let $\mathcal{F}$ be the collection of assets which are included in the fund. The return R_i on any asset i in $\mathcal{F}$ is linked to the fund return R_F, because the asset is included in the fund. To further describe the link between systematic risk F and underlying asset return R_i, consider the expectation $E[R_i|R_F]$ of R_i conditional on an observed value for fund return R_F, or a collection of such fund returns. If the fund achieves, via diversification across $\mathcal{F}$, near elimination of all risk, then R_F may be nearly constant, in which case the conditional expecation $E[R_i|R_F]$ may be close to the unconditional expectation $E[R_i]$. If so, the link between systematic risk and asset returns is weak. On the other hand, if R_F contains substantial risk, not diversified via $\mathcal{F}$, then $E[R_i|R_F]$ may frequently be far from $E[R_i]$, supporting a strong link between systematic risk and the underlying assets.

Conditional expected returns $E[R_i|R_F]$ are predictions of returns R_i, given fund performance R_F, and as such incur errors, equal to the difference $R_i - E[R_i|R_F]$. It is helpful to distinguish between ordinary mathematical expectations E and what we will call model predictions, denoted E^*, which incorporate some restrictions on (μ, Ω) implied by economic theory. For the latter, we will suppose that investor actions either lead to a fully optimal, utility-maximizing, socially efficient market outcome, or at least guarantee an outcome which is deemed satisfactory or sufficient. In the CAPM theory, investor actions are fully optimal, while in the APT theory, they need only preclude arbitrage – a sufficient outcome. For expectations E^*, obtained from some economic model, errors δ_i are as follows:

$$\delta_i = R_i - E^*[R_i|R_F] \tag{1}$$

for each i in $\mathcal{F}$. The quantity δ_i is a random variable having some mean value α_i. If model predictions are *unbiased* then $\alpha_i = 0$ for each i in $\mathcal{F}$, otherwise some expections are biased. Without loss of generality, model errors take the form:

$$\delta_i = \alpha_i + \varepsilon_i \tag{2}$$

for $i = 1, 2, ..., n$, where ε_i is a random error for which $E[\varepsilon_i|R_F] = 0$.

The present work is devoted to estimating the magnitude of model errors, empirically, for given funds F, with focus on aggregate, or overall, error magnitude observed across all the assets in the relevant set $\mathcal{F}$. We are interested both in bias α and the random error ε, that together generate model errors via equation (2).

We next illustrate the above-described framework of model errors, in the CAPM and APT models. To summarize, in the CAPM and APT both the bias and random error components of δ_i are of economic interest: large biases suggest instances of *extraordinary* asset performance, and small random errors suggest importance of systematic risk.

2.1 CAPM

In the CAPM model, as presented in Sharpe (1964) and Lintner (1965), all investors have the same beliefs regarding the distribution of asset returns, a distribution which is assumed multivariate normal with mean vector μ and covariance matrix Ω. Asset markets are in competitive equilibrium, and each investor maximizes utility by forming an optimal portfolio of risky and riskless assets. In equilibrium, investors each hold the market portfolio of risky assets, and to achieve this outcome they buy and sell assets, if needed. For investors with no endowment of stocks, the market portfolio represents a fund F that is purchased to maximize utility. The fund F represents the whole market, so the collection $\mathcal{F}$ of assets included in F is the entirety of all traded assets. The fund's return R_F is the market return, denoted here as R_m, and for each asset i the CAPM expected return, conditional on R_m, is:

$$E^*[R_i|R_m] = R_0 + \beta(R_m - R_0) \tag{3}$$

with 'beta' value $\beta = \text{cov}(R_i, R_m)/V(R_m)$. In the CAPM, actual returns R_i will almost surely differ from the expectation shown in equation (3): this difference is represented by the quantity δ_i in equations (1) and (2), with the latter equation decomposing δ_i into bias α_i and random error ε_i. If the CAPM model is correct, bias α_i is zero; if not then model prediction can differ from the more generally valid expectation formula which, under the asssumed normality of returns and the non-stochastic nature of R_0, can be expressed as follows:[3]

$$E[R_i|R_m] = \alpha_i + R_0 + \beta_i(R_m - R_0) \tag{4}$$

In equation (4), also called the market model, the 'alpha' value α_i takes on an additional meaning, connoting 'extraordinary' returns. To make this clearer we can compare equations (3) and (4), and observe that α_i equals the difference between generally-vaid expectation E and CAPM model prediction E^*:

$$\alpha_i = E[R_i|R_m] - E^*[R_i|R_m]$$

Hence α_i represents a surprise to investors, associated with revising expectation away from that formed in CAPM equilbrium, for the i-th asset. If such surprises are rare then α_i may represent an extraordinary return. Estimates of alpha, for the market model, are widely reported and of general interest to financial economists. In the present work we seek to estimate the typical magnitude of bias α_i across all relevant assets i.

The market portfolio's random return represents a risk, in the CAPM, for those who share in the purchase of the market fund, and more generally for all investors since each ends up holding some combination of the riskless asset and the market portfolio of risky assets. How important is this risk? The CAPM theory itself has no answer to this question: systematic risk can be trivial and unimportant if all returns R_i are mutually independent. In particular, if risky returns are IID normal then the market portfolio is an equal-weighted average $\bar{R}$ of all these returns, and for large n the correlation between any asset return and $\bar{R}$ converges to zero. On the other hand, systematic risk can be important if important if returns exhibit unequal variance or strong correlation among each other.

To further discuss risk in the CAPM, we use equations (2) and (3) to relate returns R_i to market return R_m and random error ε_i, yielding the standard CAPM equation:

$$R_i = R_0 + \beta_i(R_m - R_0) + \varepsilon_i \tag{5}$$

for $i = 1, 2, ..., n$. If asset returns are IID normal then $R_m = \bar{R}$, $\beta_i = 1$ for each i, and $V[\varepsilon_i] = V[R_i](1 - \frac{1}{n})$. If, further, n gets large then risk has vanishing importance and:

$$\frac{1}{n}\sum_{i=1}^{n} V[R_i] \approx \frac{1}{n}\sum_{i=1}^{n} V[\varepsilon_i] \tag{6}$$

That is, if risk is totally idiosyncratic, diversifiable, and unimportant then the errors in the CAPM equation (5) should have the same variability as the asset returns themselves. The

[3]With R_i and R_m distributed multivariate normal, $E[R_i|R_m] = \gamma_i + R_0 + \beta_i(R_m - R_0)$, with $\gamma_i = E[R_i] - R_0 - \beta_i(E[R_m] - R_0)$. Also, from equations (1), (2), and (3), $\alpha_i = E[R_i - R_0 - \beta_i(R_m - R_0)]$, hence $\gamma_i = \alpha_i$.

approximation (6) remains valid if we allow returns to have unequal mean values or unequal variances, so long as such between-asset differences remain bounded as $n \to \infty$.

If some risk is systematic and hence important then equation (6) must not hold, a situation implied by the following condition:

$$\frac{1}{n}\sum_{i=1}^{n} V[\varepsilon_i] \ << \ \frac{1}{n}\sum_{i=1}^{n} V[R_i] \tag{7}$$

That is, if the bound (7) holds then systematic risk is important in the CAPM, as can be further demonstrated by decomposing return variance $V[R_i]$ into systematic and idiosyncratic components:

$$V[R_i] \ = \ \beta_i^2 V[R_m] + V[\varepsilon_i] \tag{8}$$

for each asset i. For a given value of return variance $V[R_i]$, if idiosyncratic variance component $V[\varepsilon_i]$ is relatively small then systematic variance component $\beta_i^2 V[R_m]$ must be relatively large.

2.2 APT

In the Ross (1976) APT, returns on assets conform to a strict *factor* model:

$$R_i = \mu_i + \gamma_i z + \varepsilon_i \tag{9}$$

where $\mu_i = E[R_i]$ as earlier, z is a $K \times 1$ vector z of unobserved random factors, for some $K \geq 0$, γ_i is a $1 \times K$ vector of constants – 'factor loadings', and ε_i is a random error.[4] The factors $z_1, ..., z_K$ each have zero mean, and they have a variance/covariance matrix, say Ω_z, which, without loss of generality, equals the $K \times K$ identity matrix I_K. Each error ε_i has zero mean, positive variance, and zero covariance with z. Distinct errors ε_i and ε_j are uncorrelated, and since we assume asset returns are distributed jointly normal, without loss of generality we suppose that z is a standard normal vector, and that $\varepsilon_1, \varepsilon_2, ...$ are mutually independent normal variables, all independent of z.

The factor model, as specified by equation (9), does not explicitly link asset returns R_i to risk represented by the return R_F on a commonly traded fund(s) F. Instead it links returns R_i to a vector z of unobserved factors. To incorporate F, a possibility is to suppose that z is linearly related to a vector R_F of fund returns, with each fund composed of all assets, via:

$$z = \nu + \Phi R_F \tag{10}$$

for some K-vector ν and some $K \times K$ invertible matrix Φ. Equation (10) is not generally consistent with the factor model (9). This is easy to confirm in the case $n = 2$ and $K = 1$, with ε_1 and ε_2 standard normal variables, and $\mu_1 = \mu_2 = \gamma_1 = \gamma_2 = 1$, $R_F = \pi_1 R_1 + \pi_2 R_2$,

[4]The coincidence of notation ε_i in equation (9) and earlier ones like (2) is intentional and is discussed below.

since here the covariance of R_1 and R_F must be equal the covariance of R_2 and R_F: $\Phi(2\pi_1 + \pi_2) = \Phi(\pi_1 + 2\pi_2)$, implying that $\pi_1 = \pi_2$, in which case errors ε in equation (10) are the same as those in regression of returns on their average $\bar{R} = (R_1 + R_2)/2$, yielding $R_i = \bar{R} + \varepsilon_i$, for $i = 1, 2$, in which case errors ε_1 and ε_2 have negative covariance: $-\frac{1}{2}$, a contradiction. However, if we have instead $n = 3$ then this error covariance becomes $-\frac{1}{3}$, closer to zero, and it approaches 0 as $n \to \infty$, so violations of (10) are not necessarily important in large samples.

The APT theory is itself a large-sample theory, with the celebrated implication:[5]

$$E[R_i] = \alpha_i + R_0 + \beta_i \tau \tag{11}$$

for a $K \times 1$ vector τ and some scalars $\alpha_1, ..., \alpha_m$ for which:[6]

$$\frac{1}{n} \sum_{i=1}^{n} \alpha_i^2 \to 0 \quad \text{as} \quad n \to \infty \tag{12}$$

Applying equations (9), (10) and (11), we find that the vector τ in equation (11) has typical element $\tau_k = E[R_{Fk}] - R_0$,[7] in which case:

$$R_i = \alpha_i + R_0 + \sum_{k=1}^{K} \beta_{ik}(R_{Fk} - R_0) + \varepsilon_i \tag{13}$$

Under the APT restriction (12), the typical magnitude of α_i should be small when n is large, in which case we can interpret the large-n prediction $E^*[R_i|R_F]$ to be:

$$E^*[R_i|R_F] = R_0 + \sum_{k=1}^{K} \beta_{ik}(R_{Fk} - R_0) \tag{14}$$

By comparison, a generally valid expectation formula, under assumed normality of asset returns, is:

$$E[R_i|R_F] = \alpha_i + R_0 + \sum_{k=1}^{K} \beta_{ik}(R_{Fk} - R_0) \tag{15}$$

In the framework specified by equations (13)-(15), we can now endow the symbols α_i and ε_i with the meaning given earlier in equation (2), where α_i is the bias component of the model error $\delta_i = R_i - E^*[R_i|F]$, and ε_i is the random component. As in the CAPM, bias α_i in the APT can be interpreted as an extraordinary return.

The importance of risk, in the APT, is unclear a proiri. Consider, as in the CAPM discussion earlier, the large-n no-risk case where asset returns are IID normal. From the APT regression (13), the covariance between asset returns is:

---

[5] Coincidence of ‘alpha’ notation α_i in equations (11) and (2) is explained below.

[6] In equation (12), the average of squared alpha values converges to 0, for large m: somewhat more generally, the APT stipulates that $\sum_{i=1}^{m} \alpha_i^2$ remains bounded as $m \to \infty$.

[7] For discussion on this point see Huberman, Kandel, and Stambaugh (1987).

$$COV[R_i, R_j] = \beta_i' \Omega_{R_F} \beta_j \tag{16}$$

for each i and j, where β_i is the K-vector of factor loadings for asset i, and Ω_{R_F} is the variance-covariance matrix for R_F. With Ω_{R_F} a positive definite matrix, there can be at most K assets for which β_i has non-zero elements and such that, whenever paired, the i, j covariance in equation (16) equals zero. Hence, for large n, if returns are IID normal then β_i equals the $K \times 1$ zero vector 0_K for almost all n, and in the APT regression (13) the typical variance of returns R_i is about the same as the typical variance of errors ε_i. That is, equation (6) holds, just as in the CAPM, in the absence of systematic risk. As earlier, (6) holds even if we allow non-constant mean and variance values for returns R_i, provided that such between-asset differences remain bounded as n gets large.

If some risk is systematic in the APT then equation (6) must fail, just as in the CAPM, and if this risk is large then the bound (7) should hold. As earlier, we can usefully decompose return variance into systematic and idiosyncratic components:

$$V[R_i] = \beta_i' \Omega_{R_F} \beta_i + V[\varepsilon_i] \tag{17}$$

for each asset i, in which case smaller idiosyncratic risk $V[\varepsilon_i]$ implies a larger systematic risk $\beta_i' \Omega_{R_F} \beta_i$, for each given value of return variance $V[R_i]$.

3 Statistics

3.1 Measures of error magnitude

We want to size up the errors δ_i in models of systematic risk. In the aggregate, across assets $i = 1, 2, ..., n$, a simple measure of the magnitude of model error is the expected mean square error, or equivalently the root mean squared error (RMSE), which we denote as ϕ:

$$\phi = \sqrt{\frac{1}{n} E\left[\sum_{i=1}^{n} \delta_i^2\right]} \tag{18}$$

For the economist that wants an economic model to have small error, a small value of ϕ may be desired. However, the typical investor may hold a different view about ϕ: to see why, first decompose each error δ_i as in equation (2), and express the i-th expected square error $E[(\delta_i)^2]$ as:

$$E\left[\delta_i^2\right] = \alpha_i^2 + \sigma_i^2$$

with squared bias term α_i^2 and variance term where $\sigma_i^2 = V[\varepsilon_i]$. We can then write RMSE as the contribution of two components:

$$\phi = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \alpha_i^2 + \frac{1}{n} \sum_{i=1}^{n} \sigma_i^2} \tag{19}$$

We can further define constants χ_1 and χ_2:

$$\chi_1 = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\alpha_i^2}\,, \quad \chi_2 = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\sigma_i^2}$$

in which case $\phi = \sqrt{\chi_1^2 + \chi_2^2}$. The constant χ_1 is the root mean square of biases, and hence represents the model's *non-stochastic* error magnitude, whereas χ_2 is the root mean square of standard deviations, and is therefore the *stochastic* error magnitude.

If the typical investor views the existence of shared, or systematic, risk as bad, they may prefer, for given values of asset return variances $V[R_i], i = 1, 2, ..., n$, that the stochastic error magnitude ψ_2 be relatively *large*, since this suggests a relatively small contribution of systematic risk to $V[R_i]$ (for relevant discussion see the previous section on the CAPM). But if so then the investor may prefer to see a larger RMSE value ϕ, for the economic model, under some conditions. At the least, it seems reasonable to seek information about RMSE components χ_1 and χ_2, when one is investigating RMSE itself.

Concerning model bias (α), if bias is bad then presumably a smaller value of root-mean-square bias (χ_1) is preferred. But since bias is evaluated across n separate investments, the individual biases $\alpha_1, ..., \alpha_n$ may or may not be large on average.[8] To address this possibility we can define two more constants ψ_1 and ψ_2:

$$\psi_1 = |\bar{\alpha}|\,, \quad \psi_2 = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\alpha_i - \bar{\alpha})^2}$$

where $\bar{\alpha}$ is the average bias: $\frac{1}{n}\sum_{i=1}^{n}\alpha_i$, in which case $\chi_1 = \sqrt{\psi_1^2 + \psi_2^2}$. The constant ψ_1 is the unsigned pooled bias, while ψ_2 is the bias *heterogeneity* across assets. So, for a given magnitude of bias – measured by χ_1 – the relative contributions of ψ_1 and ψ_2 may be of some interest.

Heterogeneity of bias, ψ_2, can be viewed as ex ante source of heterogeneity in model errors, while stochastic error magnitude χ_2 can can be viewed as ex post heterogeneity. We can then define *total heterogeneity* component ω, of model errors as:

$$\omega = \sqrt{\psi_2^2 + \chi_2^2}$$

By construction, RMSE can be expressed as $\phi = \sqrt{\psi_1^2 + \omega^2}$,[9] in which case the model's overall error magnitude, measured by RMSE, can be attributed to common (ψ_1) and heterogeneous (ω_2) components.

To further interpret the RMSE (ϕ) and its various components χ, ψ, ω, note that all of these constants are non-negative. It then follows that ϕ is greater than or equal to each

[8] In the market model (4), with market portfolio return R_m equal to a weighted sum $\sum_i w_i R_i$ of asset returns, the *weighted* average alpha necessarily equals zero: $\sum_i w_i \alpha_i = 0$, though the simple average $n^{-1}\sum_i \alpha_i$ may be non-zero. The same result holds for the APT regression equation (13).

[9] Note that, from equation (19), $\phi^2 = \chi_1^2 + \chi_2^2 = \psi_1^2 + \psi_2^2 + \chi_2^2$.

component value (χ, ψ, ω). Similarly, non-stochastic error magnitude χ_1 is no less than its component values ψ_1 and ψ_2. Each of these error measures is computable in terms of bias α and error variance σ, as shown in Table 1.

[INSERT Table 1 about here]

3.2 Estimation

To apply the RMSE value ϕ, and its components, note that all of these constants can be computed if bias α_i and error variance σ_i^2 is known for each asset i, as implied by Table 1. It then suffices to have good estimators $\hat{\alpha}$ of α and $\hat{\sigma}$ of σ. We can write α_i as:

$$\alpha_i = E[\, E[R_i|R_F] - E^*[R_i|R_F]\,] \tag{20}$$

where as earlier R_i is the return on an asset, R_F is the return on a commonly traded fund(s), E represents mathematical expectation, and E^* is prediction from the economic model. To obtain equation (20) apply (1) and (2) and note that, by the law of iterated expectations, $E[E[R_i|R_F]] = E[R_i]$. Earlier we assumed joint normality of asset returns, in which case mathematical expectation can be written:

$$E[R_i|R_F] = \beta_{i0} + R_0 + \sum_{j=1}^{K} \beta_{ij}(R_{Fj} - R_0) \tag{21}$$

with, as earlier, R_0 the risk-free rate of return, for some intercept values β_{i0} and slopes β_{ij}, $i = 1, 2, ..., n$, $j = 1, 2, ..., K$, with K the number of funds. model prediction E^* imposes on E some economic restrictions which, from equation (21), is some restriction(s) on β. In the CAPM and APT economic models, which we earlier interpreted as models of systematic risk – the general class of models we are interested in, the economic restriction on β is that the intercepts β_{i0} each equal 0, and with this approach we have:

$$E^*[R_i|R_F] = R_0 + \sum_{j=1}^{K} \beta_{ij}(R_{Fj} - R_0) \tag{22}$$

Model bias α_i, for asset i, then coincides with the intercept β_{i0} in equation (21), and we can re-state (21) as a set of regression equations:

$$R_i = \alpha_i + R_0 + \sum_{j=1}^{K} \beta_{ij}(R_{Fj} - R_0) + \varepsilon_i \tag{23}$$

with regression errors ε_i which are normal, with mean 0 and variance σ_i^2.[10]

[10] The symbol ε_i was earlier used in equation (2) to represent the random component of the error made by an *economic* model, and is justified in regression (21) due to the intercept identification $\beta_{0i} = \alpha_i$.

To estimate the regression system (21), we suppose that for all asset and fund returns there are observations at successive times, denoted $t = 1, 2, ..., T$, such that the return probability distribution exhibits independence over time. With assumed normality, the maximum likelihood estimator (MLE) of (α, β) is the same as the ordinary least squares (OLS) estimator applied to each separate instance $i = 1, 2, ..., n$ of (21). Denoting thess MLEs by $\hat{\alpha}, \hat{\beta}$, which are unbiased estimators, let $\hat{\sigma}_i^2$ be the OLS regression standard error (of the estimate), which is a bias-corrected version of the MLE for σ^2.

Plugging in $\hat{\alpha}$ and $\hat{\sigma}$ into the above formulas for the economic model's magnitude of error, we obtain consistent estimates $\hat{\phi}, \hat{\chi}, \hat{\psi}, \hat{\omega}$ of RMSE ϕ and its components χ, ψ, ω.[11]

4 Econometric Theory

The proposed RMSE statistics have sampling distributions which are unknown. These distributions can be useful, and we therefore derive features of them.

4.1 Quadratic Forms

Nominally, the proposed statistics are measures of goodness-of-fit. Standard measures of goodness-of-fit are based on sums of squares, or more generally quadratic forms, and suitable transformations thereof. The proposed statistics fit in this category, also.

To proceed, in terms of notation let x be the $T \times K$ matrix of data on the excess returns of funds F, and let y be the $T \times n$ matrix of data on the excess returns on assets. Given y, let vector $\mathbf{y}$ be the "vec" version of y, this being the $nT \times 1$ column vector consisting of the columns of y stacked on top of each other, starting at the left of y. We introduce a *semi-norm*:

$$||y||_S = \sqrt{\mathbf{y}' S \mathbf{y}} \tag{24}$$

for some square $nT \times nT$ symmetric positive semi-definite matrix S, with rank r which is user-specified, and which we will call the *kernel*. We restrict attention to the situation where S is a known function of x.

Semi-norm $||y||_S$ represents the magnitude, or size, of some r-dimensional feature of the $nT \times 1$ vector y. A semi-norm is a generalization of the concept norm, such that $||z|| = 0$ is possible for $z \neq 0$ when $|| \cdot ||$ is a semi-norm, but not a norm – see for example Horn and Johnson (1999, p. 259). Matrices of type S can always be factored $S = A'A$, for some $r \times nT$ matrix A, so we can view $||y||_S$ as the (Euclidean) length of the $r \times 1$ vector $z = A\mathbf{y}$, and intepret z as the feature measured by $||y||_S$. For example, let $A = C'$, with C the Cholesky root of S. There is more than one such factorization, as $S = A'B'BA$ for any $r \times r$ matrix B for which $B' \times B = I_r$, where I_r is the $r \times r$ identity matrix.

[11]Note that ϕ represents the RMSE of the *economic* model, inclusive of bias, whereas in regression the standard RMSE statistic describes model fit without bias.

For any two features z_1 and z_2, if $A_1A_2' = 0_{r_1,r_2}$, with $0_{r_1,r_2}$ the $r_1 \times r_2$ matrix consisting of 0s, then the features are orthogonal in the sense that the mappings $\mathbf{y} \to A_1\mathbf{y}$ and $\mathbf{y} \to A_2\mathbf{y}$ have derivatives $\frac{d}{d\mathbf{y}}A_1\mathbf{y} = A_1$ and $\frac{d}{d\mathbf{y}}A_2\mathbf{y} = A_2$, such that the rows of A_1 are vectors orthogonal to those of A_2. We call quadratic forms $Q_1 = \mathbf{y}S_1\mathbf{y}$ and $Q_2 = \mathbf{y}S_2\mathbf{y}$ orthogonal if $S_1S_2 = 0_{nT,nT}$. It follows that if z_1 and z_2 are orthogonal then so are Q_1 and Q_2. Conversely, if Q_1 and Q_2 are orthogonal then z_1 and z_2 are orthogonal. To see this note that if $S_1S_2 = 0_{n,n} = A_1'A_1A_2'A_2$ then the rank of S_1S_2 equals 0, and this is no more than the maximum of the ranks of matrices: A_1', A_1A_2', and A_2 which, when multiplied, yield S_1S_2, with A_1' and A_2 having ranks r_1 and r_2, respectively, implying that A_1A_2' has rank 0 and hence equals $0_{r_1,r_2}$. If the elements of y are independent and identically distributed (IID) normal then features are orthogonal if and only if they are stochastically independent. To see this note that

$$cov(z_1, z_2) = E[(z_1 - Ez_1)(z_2 - Ez_2)'] = E[A_1(\mathbf{y} - E\mathbf{y})(\mathbf{y} - E\mathbf{y})'A_2']$$

Hence $cov(z_1, z_2) = A_1V_{\mathbf{y}}A_2'$, with $V_{\mathbf{y}}$ the variance-covariance matrix of $\mathbf{y}$. If $\mathbf{y}$ is IID then $V_{\mathbf{y}}$ is a multiple of the identity matrix, and so $cov(z_1, z_2) = 0$ if and only if $A_1A_2' = 0_{r_1,r_2}$. In the CAPM and APT models discussed earlier, the IID assumption can fail if, for example, different assets i have different error variances $V[\varepsilon_i]$, in which case orthogonal features z_1 and z_2 may be stochastically dependent.

Table 2 presents various statistics expressible as quadratic forms $||y||_S^2 = \mathbf{y}'S\mathbf{y}$.[12] This table present some formulas that involve matrix producs, and inverses, and direct products – also called Kronecker products.[13]. Included are three univariate statistics: the average squared value: $T^{-1}y'y$, the squared average: $T^{-1}\hat{y}'\hat{y}$, and the average squared deviation: $T^{-1}(y-\hat{y})'(y-\hat{y})$. The rank r of the corresponding S matrices is T, 1, and $T-1$, respectively. In the case of squared average, the semi-norm $||y||_S = |\bar{y}|$ represents the magnitude of a 1-dimensional feature of the data, represented by the average $\bar{y}$ which we can write as $z_1 = A_1y$ with $A_1 = n^{-1}1_n$. For the average squared deviation, the feature of interest has dimension $T - 1$, and can be represented via a vector of contrasts $z_2 = A_2y$, where, with the aid of Helmert's transformation, we can select A_2 and z_2 via:

$$\begin{aligned}
&z_{21} = \tfrac{1}{\sqrt{2n}}(y_1 - y_2),\\
&z_{22} = \tfrac{1}{\sqrt{6n}}(y_1 + y_2 - 2y_3), ...,\\
&z_{2,n-1} = \tfrac{1}{\sqrt{n(n-1)n}}(y_1 + y_2 + \cdots + y_{n-1} - (n-1)y_n)
\end{aligned}$$

Other specifications of z_2 are possible, for example by permuting elements y_i of y in each of the above formulas for the elements of z_2. The average and the contrast vector are orthogonal features of y, a fact expressible as $S_1S_2 = 0$ for respective kernels S_1 and S_2, which from Table 1 are $S_1 = T^{-2}1_{T,T}$ and $S_2 = T^{-1}(I_T - T^{-1}1_{T,T})$. For IID normal data $y_t, t = 1, 2, ..., T$, this implies the well-known fact that $(\bar{y})^2$ is independent of the sample variance s_y^2. On the

[12] The results in Table 2 are readily derived from standard least-squares formulas, as contained in econometric texts like Greene (2003) and Ruud (2000).

[13] See Horn and Johnson (1999), Magnus and Neudecker (2002).

other hand if, say, the covariance matrix V_y of y is of diagonal form with non-constant entries on the diagonal, then $(\bar{y})^2$ and s_y^2 are dependent, as can be shown by computing $A_1 V_y A_2$ and verifying that it is not equal to the zero matrix $0_{T,T}$.

[INSERT Table 2 about here]

In the context of single-equation regression models, also addressed in Table 2, the (scaled) magnitude of fitted value $\hat{y}$ and residual $y - \hat{y}$ are orthogonal features whose magnitudes take the form $||y||_S$. In the multi-equation context (Table 2, bottom), most relevant to our study, these features are present again, though passage to multiple equations complicates matters somewhat, in terms of the S formula.

To cast $RMSE$ estimator $\hat{\phi}$ and its components $(\hat{\chi}, \hat{\psi}, \hat{\omega})$ as semi-norms of the form $\mathbf{y}'S\mathbf{y}$, we apply the expectation model (21) to data observed at times $t = 1, 2, ..., T$, as follows:

$$y_{it} = \alpha_i + x_t \beta + \varepsilon_{ti} \tag{25}$$

where $y_{it} = R_{ti} - R_{t0}$ is excess return on the i-th asset at time t, x_t is the $1 \times K$ vector of excess returns $R_{tFk} - R_{t0}$ for factors $k = 1, 2, ..., K$, at time t, and β is a $K \times 1$ vector of factor loadings. For OLS estimators $(\hat{\alpha}_i, \hat{\beta}_i)$ of (α_i, β_i), we have:

$$\hat{\alpha}_i = \bar{y}_i - \bar{x}\hat{\beta}_i \tag{26}$$

with $\bar{y}_i = \frac{1}{T} 1_T' y_i$, and $\hat{\beta}_i = (x^{*\prime} x^*)^{-1} x^{*\prime} y_i^*$, where y_i is the $T \times 1$ vector with typical element y_{it}, x^* and y_i^* are de-meaned (centered) versions of x and y_i, and 1_T is the T-vector with all entries equal to 1. We can write $y_i^* = (I_T - \frac{1}{T} 1_{T,T}) y_i$, with $1_{T,T}$ the $T \times T$ matrix consisting of 1s, in which case equation (26) yields:

$$\hat{\alpha}_i = a_x' y_i$$

for each $i = 1, ..., n$, with $T \times 1$ vector a_x as follows:

$$a_x = \frac{1}{T} 1_T \; - \; (I_T - \frac{1}{T} 1_{T,T}) \, x^* (x^{*\prime} x^*)^{-1} \, \bar{x}$$

We can then write the $n \times 1$ vector $\hat{\alpha}$ as

$$\hat{\alpha} = \begin{bmatrix} a_x' y_1 \\ \vdots \\ a_x' y_n \end{bmatrix}$$

In terms of the $nT \times 1$ vector $\mathbf{y}$, we then have:

$$\hat{\alpha} = (I_n \otimes a_x')\,\mathbf{y}$$

Theorem 1 *The kernel matrix, S, for the quadratic forms underlying sample RMSE and its components are as shown in Table 3.*

[INSERT Table 3 about here]

Corollary 1 *The kernel matrices S associated with sample RMSE and its components have the following relationships:*

$$\begin{array}{lll} i. & S_\phi = S_{\chi_1} + S_{\chi_2}, & S_{\chi_1} S_{\chi_2} = 0 \\ ii. & S_\phi = S_{\psi_1} + S_\omega, & S_{\psi_1} S_\omega = 0 \\ iii. & S_{\chi_1} = S_{\psi_1} + S_{\psi_2}, & S_{\psi_1} S_{\psi_2} = 0 \\ iv. & S_\omega = S_{\chi_2} + S_{\psi_2}, & S_{\chi_2} S_{\psi_2} = 0 \end{array}$$

From Corollary 1, components $\hat{\chi}_1$ and $\hat{\chi}_2$ of the sample RMSE ($\hat{\phi}$) are *orthogonal* in the sense that their associated quadratic forms are orthogonal. Likewise, $\hat{\psi}_1$ and $\hat{\psi}_2$ are orthogonal, as are $\hat{\psi}_1$ and $\hat{\omega}$, and $\hat{\chi}_2$ and $\hat{\psi}_2$. If, conditional on X, the joint distribution of vector $\mathbf{y}$ is IID normal it follows that orthogonal statistics are also independent. In application to stock returns the IID assumption requires CAPM regression errors to have equal variance across all assets, for example, and this is likely unrealistic. Hence orthogonal components of RMSE may be stochastically dependent.

4.2 Moments

For point estimates of model error measures $(\phi, \chi, , \psi, \omega)$, we would like to know their probability distribution. For a quadratic form expressible as $Q = \mathbf{y}'S\mathbf{y}$, with normally distributed random vector $\mathbf{y}$, the probability distribution of Q is the same as that of a weighted sum of independent chi square random variables (Imhof (1961)). This distribution depends on the associated weights, and generally can not be tabulated independently of them. For example, in our setup if $K = 1$ then the squared non-stochastic error $\hat{\chi}_1^2$ is just the sum of squared sample averages $\sum_{i=1}^n (\bar{y}_i)^2$, and its distribution depends on means, variances, and covariances among the y variables, hence while we can view this distribution as a weighted sum of chi squares, the weights are unknown a priori. Morever, there is no tractable way to 'standardize' such statistics, via some transformation $Q \to Q^*$, with Q^* having a known distribution (standard normal, chi square, etc.), even if the moments of y are known. This remains true in large samples, despite availability of the centeral limit theorem.

With some work we can obtain moments of the distribution of model error statistics. For a quadratic form $Q = \mathbf{y}'S\mathbf{y}$, with vector $\mathbf{y}$ having mean vector μ and covariance matrix V, and square matrix S, the expected value μ of Q is:[14]

$$E(Q) = tr[SV] + \mu' S\mu \tag{27}$$

with "tr" meaning the trace operator – summing a matrix's diagonal elements. The expectation E is interpreted as conditional on data x for the independent variables in the regression model. Under the additional assumption that $\mathbf{y}$ is normally distributed, the variance of Q is:[15]

$$V(Q) = 2\,tr[SVSV] + 4\mu' SVS\mu \tag{28}$$

To apply the mean formula (27) and variance formula (28) for quadratic forms in normal variables, we need the relevant mean μ and variance V values. In our application, the variable of interest is y, and we use the distribution of y conditional on x. With $y_{ti} = \alpha_i + \beta_i' x_t + \varepsilon_{ti}$, introduce notation: $z_t = (1, x_t')'$, and $\theta_i = (\alpha_i, \beta_i')'$, $i = 1, ..., n$, each of which is a $(K+1)\times 1$ vector. Then $y_{ti} = \theta_i' z_t + \varepsilon_{ti}$, and the conditional mean of y_{ti}, given x_t, is $\theta_i' z_t$. With $(K+1)\times n$ matrix θ having typical row θ_i, and $T \times (K+1)$ matrix z having typical column z_t, we then have the conditional mean $\mu_{\mathbf{y}}$ of the $nT \times 1$ vector $\mathbf{y}$:

$$\mu_y = vec(z\theta)$$

To get the variance/covariance of y, conditional on x, using $y_{ti} = z_t\theta_i + \varepsilon_{ti}$ we observe that the conditional variance of y_{it} is the error variance $V(\varepsilon_{ti})$. Let V_ε denote the variance-covariance matrix for $(\varepsilon_{ti}, i = 1, 2, ..., n)$, and any fixed t.[16] When then have a formula for the variance V_y of vector $\mathbf{y}$:

$$V_y = V_\varepsilon \otimes I_T$$

A practical challenge, in applying equations 27 and 28 is that they require multiplication of huge matrices, unless n and T are small, with the number of calculations proportional to $(nT)^2$. Fortunately, the matrices being multiplied are 'sparse', and can be simplified, as follows.

Theorem 2 *The expected value formula (27) for the RMSE statistic and its components can be expressed in a form that is computationally tractable. For this, Table 4 provides formulas for $tr[SV_y]$, and Table 5 provides formulas for $\mu_y' S\mu_y$.*

[14] See for example Wiley (1970), for a presentation of this classic result.

[15] See Wiley (1970). As with expectation E, variance V is computed conditional on x.

[16] We are assuming the data is independent and identically distributed, over time.

[INSERT Tables 4 and 5 about here]

Theorem 3 *The variance formula (28) for the RMSE statistic and its components can be expressed in a form that is computationally tractable. For this, Table 6 provides formulas for* $tr[SV_ySV_y]$*, and Table 7 provides formulas for* $\mu_y' SV_yS\mu_y$*.*

[INSERT Tables 6 and 7 about here]

In Theorem 2 and 3, the required number of matrix computational steps is proportional to $n + K^2T$, as compared to $(nT)^2$ needed for direct evaluation of equations 27 and 28.[17]

For non-normal data, with finite variance, the mean formula (27) for quadratic forms remains valid but the variance formula (28) may not. For our model error statistics $\hat{\phi}, \hat{\chi}, \hat{\psi}, \hat{\omega}$, some of the variance formulas in Tables 6 and 7 remain valid in large samples even if data are non-normal. If we assume that $\hat{\alpha}$ is *asymptotically* normal, then for statistics expressible as quadratic forms in $\hat{\alpha}$ the proposed variance formulas are (asymptotically) valid. From Table 1, these 'robust' statistics are $\hat{\chi}_1, \hat{\psi}_1$, and $\hat{\psi}_2$. The remaining statistics: $\hat{\phi}, \hat{\chi}_2$, and $\hat{\omega}$ each are functions of sample (error) variances, and the (population) variance of sample variance is well-known to be sensitive to data non-normality, fat-tailed distributions, etc.

4.3 Confidence Intervals

For quadratic forms Q we calculated in Theorems 2 and 3 their mean and variance, and we can then construct intervals of the form:

$$\left(E[Q] \pm k\sqrt{V[Q]}\right) \tag{29}$$

for some number $k \geq 0$. By Chebychev's inequality, this interval contains the value Q with probability:

$$P(|Q - \mu| \leq k\sigma) \geq 1 - \frac{1}{k^2} \tag{30}$$

Hence, the bound (30) renders formula (29) a conservative confidence interval for Q, with probability at least $1 - 1/k^2$ of containing the true value. If Q is normally distributed, or nearly so, then (29) with $k = 2$ will have approximately 95% probability coverage, and with

[17] Computer code, for the computations in tables 4,..., 7 are available from the authors' website, as EViews programs.

$k = 3$ will have about 99% coverage. This compares to conservative coverage levels 75% for $k = 2$ and 89% for $k = 3$.

The RMSE statistic, and its components, is of the form $\sqrt{Q}$, with Q one of the quadratic forms described earlier. Each Z is non-negative, in which case the transformation $Q \to \sqrt{Q}$ is monotonic (increasing). This provides a conservative confidence interval for $\sqrt{Q}$ also:[18]

$$\left(\sqrt{\mu - k\sigma}, \sqrt{\mu + k\sigma} \right) \tag{31}$$

To use the formulas 29 and 31, we substitute unkowns μ and σ with plug-in values $\hat{\mu}$ and $\hat{\sigma}$. To do this we apply the formulas for μ and σ appearing in Theorems 2 and 3, and plug in OLS estimates for regression constants $\alpha, \beta, V_\varepsilon$. This yields approximate (conservative) confidence intervals, with probability coverage that approaches the true level in large samples due to the consistency of estimators $\hat{\mu}$ and $\hat{\sigma}$.

5 Application

5.1 Descriptives

Model error (ϕ) and its components describe on the ability of systematic-risk models to 'fit' data an asset returns. We illustrate the use of the RMSE statistics using an updated version of the dataset from Fama and French (1993). The data has been downloaded from the web site of Kenneth R. French, see Appendix B for a data description.[19] The sample is extended from the original 1963:7-1991:12 to 1946:1-2007:12 to cover the post WWII period. The returns to be modelled are for the 5 x 5 = 25 Fama and French portfolios, with 5 (quintile) categories of firm 'size' – measured by market capitalization, and 5 categories of firm 'value' – measured by the ratio of book value to market value. The systematic risk factors include the market (excess) return, the Fama and French SMB (small minus big) factor – measuring difference in returns for small vs. big firms, and the HML (high minus low) factor – measuring difference in returns for high-value versus low-value firms, see Appendix D for further details.

Table 8 summarizes the historical performance of the 5 x 5 portfolios, in terms of monthly excess return. There are six statistics, describing aspects of portfolio performance, each labelled with a symbol ($\hat{\phi}, \hat{\chi}_1$, etc.) which corresponds to symbols used later to describe model performance. The typical magnitude or 'size' of excess returns is described by the normed return vector $||y||$, which from Table 8 equals 5.39. Of this value 5.39, the normed mean vector $||\bar{y}||$ and the pooled variability $\hat{\sigma}_y$ can be viewed components, equal to 1.19 and 5.26 respectively, whose root mean square value equals pooled variability: $5.39 = \sqrt{1.19^2 + 5.26^2}$. The normed mean vector $||\bar{y}||$ itself has components: pooled mean and mean heterogeneity, equal to 1.17 and 0.20, whose root mean square equals 1.19. Total heterogeneity, among returns, is composed of mean heterogeneity and pooled variability, and equals 5.26. We can

[18] If $\mu - k\sigma < 0$ then we can replace the expression $\sqrt{\mu - k\sigma}$ with the value 0 in formula 31.

[19] See http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

further interpret the magnitude $||y||$ of excess returns in terms of pooled mean (=1.17) and total heterogeneity (=5.26) components, whose root mean square equals $||y||$. On the whole, excess returns are rather small in mean value, relative to their variability, and there is little heterogeneity in mean value across portfolios, relative to the typical mean value.

5.2 CAPM and APT models

Table 9 reports model errors, for the CAPM model and seven APT models. For each error statistic a useful reference point is the descriptive statistic in Table 8 having the corresponding symbol (ϕ, χ_1, etc.). Note that each *column* value in Table 8 equals a *row* value in Table 9, for the row that reports on the APT model with no factors: in this model systematic risk is precluded, and the APT prediction is that all assets earn the riskless return on average, over time. This model's error (RMSE), equal to 5.39, is necessarily larger than that of the CAPM and alternative APT models. The correspondence between columns in Table 8 and the 'factorless' APT row in Table 9 motivates our use of symbols ϕ, χ, etc. in Table 8.

[INSERT Table 9 about here]

The CAPM model's RMSE error equals 2.89, mostly attributable to stochastic error (=2.83) rather than non-stochastic error (=0.58). The latter is attributable more to pooled bias (=0.52) than to bias heterogeneity (=0.25). Also, in reference to Table 8, the CAPM model's error is considerably less than the magnitude $||y||$ of excess returns (=5.39).

Based on the economic theory discussion in Section 2, if the CAPM holds then the model's true non-stochastic error (χ_1) should equal 0, and the sample value $\hat{\chi}_1 = 0.58$ may or may not be viewed as 'close' to zero. Concerning risk, if risk is unimportant and returns are mutually independent with (nearly) equal variance then, from equation (6) Section 2, stochastic error ($\chi_2 = 2.83$) should be about equal to pooled variability (=5.26, see Table 8), but is instead more consistent with the bound (7) from Section 2, whereby stochastic error is much less than pooled variability. That is, systematic risk seems important in the CAPM.

The remaining six APT models take the form of regression models orginally reported in Fama and French (1993).[20] In Table 9, these models are reported in the last six rows, with the final row ("all") including all 3 Fama-French factors – this being the well-known "3-factor" model of stock returns. The non-stochastic error χ_1 is estimated at 0.38 for the 3-factor model, compared to 0.58 for the CAPM, and consists more of pooled bias (=0.35) than bias heterogeneity (=0.13). The estimate of stochastic error χ_2 equals 2.83 for the CAPM, versus 1.64 for the 3-factor model. Across the board, for each category of error (ϕ, χ, etc.), the 3-factor model has smaller estimated error.

If the APT holds then, as discussed in Section 2, the true non-stochastic error (χ_1) should be equal to zero, or close to it. For the 3-factor model, the estimate $\hat{\chi}_1 = 0.38$ may or may not be viewed as 'close' to zero: it is about two-thirds the value of the CAPM $\hat{\chi}_1 = 0.58$, so

[20] In their paper, Fama and French do not define these as APT models.

if CAPM pricing errors were deemed far from zero – on the whole – it would be hard to say that the 3-factor model's pricing errors were close to zero. Also, sampling variability may account for some of the difference between these $\hat{\chi}_1$ values, an issue we address below.

The data suggests that risk is important in the 3-factor APT: stochastic error ($\hat{\chi}_2 = 1.64$) is small in comparison to pooled variability (=5.26), consistent with the bound (7) in Section 2. This compares to $\hat{\chi}_2 = 2.83$ in the CAPM.

5.3 Error Distributions

For each of the model error statistics shown in Table 9, we report in Tables 10 and 11 some distribution characteristics of the quadratic forms underlying each statistic.Shown are estimates of mean value, standard deviation, and intervals of form: mean ± 3 std. dev., using the methods spelled out in Section 4.

[INSERT Tables 10 and 11 about here]

From the tables, the effect of sampling variability on model error statistics is sometimes substantial, particularly in reference to non-stochastic error (χ_1) and its components ψ_1 and ψ_2, where the estimated standard deviation $\sqrt{\hat{V}(Q)}$ is not negligible in comparison to the estimated mean value $\hat{E}(Q)$. Earlier we noted that the CAPM and 3-factor Fama-French model do not differ by a great margin, in terms of non-stochastic error *estimates*. Table 9 further suggests that whatever difference exist may be largely attributable to sampling variability, rather than to a true difference in model performance. On the other hand, the two models seem convincingly different in terms of stochastic error (χ_2) and total error (ϕ).

5.4 Simulation

To check our theory-driven error distributions, we simulate error statistics via Monte Carlo. As indicated in Tables 12 and 13, simulation is consistent with theory (Tables 10 and 11) for those error statistics which involve alpha estimates only: $\hat{\chi}_1, \hat{\psi}_1, \hat{\psi}_2$. For the other statistics: $\hat{\phi}, \hat{\chi}_2, \hat{\omega}$, which involve variance estimates, simulation indicates greater variability than suggested by the theory. This contrast is consistent with data non-normality which, as mentioned in Section 4, does not have a large-sample impact on alpha-only statistics, but can impact the other statistics.[21]

[INSERT Tables 12 and 13 about here]

For purposes of comparing errors across models, in the case of the CAPM and 3-factor APT models the simulation results give the same impression as did the theoretical distributions reported earlier: the models differ convincingly in terms of stochastic error (χ_2), but not in terms of non-stochastic error (χ_1).

[21]Estimates of kurtosis suggest fat-tailed, non-normal distributions for the 5x5 return data: with average kurtosis equal to 5.534.

6 Discussion

6.1 Methods

The proposed measures of model error, as presented in Table 1, are intended for models of risk, as in the CAPM and APT, with risks proxied by portfolio returns. These (population) measures, and their sample analogues, can be extended to other risk models as well, specifically the Merton (1973) intertemporal capital asset pricing model (ICAPM). For this it is necessary to allow for time-varying alphas and betas, perhaps by constructing annual alphas and betas from daily data. Also, the econometric theory in Section 4 – for the distribution of sample errors – would need to be extended to allow serial correlation in the data.

Our measures of model error are based of simple averages, across assets, of regression intercept values, or regression errors. Value-weighted averages may be an attractive alternative, and these are easy to compute so long as the relevant weights are available. The probability distribution, for the weighed averages, will generally differ from that described in Section 4, but Theorems 1, 2, and 3 can be extended to handle this case.

Our econometric theory assumes that asset returns are normally distributed. For non-normal data we can generalize variance equation (28) and Theorem 3 in Section 4. The theory already calls for rather involved computations, and generalizing it will surely add more. Also, simulation – reported in Section 5 – suggested that non-normality may not matter much in terms of the theory's usefulness.

6.2 Application

Section 5 applied the proposed methods to return data for the Fama and French (1993) 5x5 collection of portfolios. As a 'universe' of returns, this collection is highly aggregated, a fact that may bias the application of CAPM and APT models. Specifically, in assessing the importance of risk in the stock market, aggregation across assets may smooth out idiosyncratic risk, causing estimates of stochastic error (χ_2) to look smaller than they would if had used a more disaggregated menu of assets. In other words, we may be understating the importance of idiosyncratic risk, and overstating the importance of systematic risk, in the stock market. As a remedy, we can disaggregate the data and again apply our methods, though some care is needed in using the econometric theory, as well-behaved estimates of the stochastic error (ε) variance/covariance matrix require that the number n of assets is reasonably small in relation to the number of time periods T. The modelling approach of Campbell, Lettau, Malkiel and Xu (2001) may be helpful here.

In applying the APT theory to Fama-French regressions, we assumed that the APT factor model in Section 2 was valid. The model assumes zero correlation among errors ε_i appearing in regressions for different assets i. The essential results of the APT remain valid if such correlation is mild (Chamberlain and Rothschild (1983)). Pairwise sample correlation is 0.152, on average, for 3-factor model residuals, not too large, and compares to 0.375 for the CAPM model.

Our sample period covers January 1947 through December 2007, for the Fama-French data set. The original data, studied by Fama and French (1993), covered July 1963 through December 1991. Our earlier remarks apply also to the original sample.[22] Specifically, the following three facts are valid for both samples. First, heterogeneity in historical mean values – across the 5x5 Fama-French portfolios – is small relative to return values: on the original sample the mean heterogeneity statistic equals 0.23, small in comparison to normed return value 5.87. Second, the CAPM and 3-factor models are broadly comparable in terms of model bias (alpha) statistics: on the original sample non-stochastic error estimates equal to 0.80 and 0.58, respectively, for the two models. Third, the 3-factor model has much smaller stochastic error (=1.45) than does the CAPM (=2.68), as was reported earlier on the full sample.

Fama and French (1993) report that the 3-factor regression model outperforms the CAPM model on two fronts – pricing errors (α) are smaller and explanatory power is greater. In terms of explanatory power, they find the 3-factor model to be superior both in terms of capturing variation over time for specific portfolios, and for capturing differences in historical mean return across portfolios.

In terms of pricing errors, unlike Fama and French (1993) we find that pricing errors are similar, on the whole, for the CAPM and 3-factor models. In terms of explanatory power, heterogeneity in historical mean values – across the 5x5 Fama-French portfolios – is small relative to return values. So, there is not much for the economic models to capture, in terms of mean heterogeneity. The 3-factor model does have smaller stochastic error (χ_2) than does the CAPM, and this implies a larger measured value for systematic risk, as discussed in Section 2. With a larger systematic component, the 3-factor model is capturing more variability in returns, and most of this return variation is over time. Hence, the 3-factor model appears superior to the CAPM in capturing time series variation of returns on the 5x5 portfolios.

On the whole, we find that previous evidence for superior 'fit' of the 3-factor model versus the CAPM model to be somewhat overstated, as pricing errors are comparable across models, and there is little explained cross-sectional variation. The superior fit of time-series variation, within each given portfolio's history, does seem compelling, though both models create the impression that sytematic risk makes an important contribution to this temporal variation in asset performance.

[22] See the authors' website for tables corresponding to this sample.

REFERENCES

Campbell, John Y. (1993), Intertemporal Asset Pricing without Consumption Data American Economic Review 3, 487-512.

Campbell, John Y., Martin Lettau, Burton G. Malkiel, and Yexiao Xu (2001), Have Individual Stocks Become More Volatile," Journal of Finance 56, pp. xxx.

Chamberlain, G., and Michael Rothschild (1983), "Arbitrage, Factor Structure, and Mean-Variance Analysis on Large Asset Markets," Econometrica 51, 1281-1304.

Chung, Y. Peter, Herb Johnson and Michael J. Schill (2006), "Asset Pricing When Returns Are Nonnormal: Fama-French Factors versus Higher-Order Systematic Comoments," The Journal of Business 79, 923-940.

Cochrane, John (2001), Asset Pricing. Princeton University Press.

Cochrane, John (2008), "A Mean-Variance Benchmark for Intertemporal Portfolio Theory," unpublished manuscript, University of Chicago.

Connor, John (1984), "A Unified Beta Pricing Theory," Journal of Economic Theory 34, 13-31.

Dyvbig, Phillip (1983), "An explicit bound on deviations from APT pricing in a finite economy," Journal of Financial Economics 12, 483-96.

Fama, Eugene F. and Kenneth R. French (1992), "The cross-section of expected stock returns," Journal of Finance 47, 427-465.

Fama, Eugene F. and Kenneth R. French (1993), "Common risk factors in the returns on stocks and bonds," Journal of Financial Economics, 33, 3-56.

Fama, Eugene F. and Kenneth R. French (1995), "Multifactor explanations of Asset pricing anomalies," Journal of Finance 50, 131-155.

Fama, Eugene F. and Kenneth R. French (1996), "Size and Book-to-Market Factors in Earnings and Returns," Journal of Finance 51, 55-84.

Fama, Eugene F. and Kenneth R. French (1998), "Value versus growth: the international evidence," Journal of Finance 53, 1975-1999.

Geweke, John, and Guofu Zhou (1996), "Measuring the pricing error of the arbitrage pricing theory," Review of Financial Studies 9, 557-587.

Gibbons, Michael R., Stephen A. Ross, and Jay Shanken (1989), "A Test of the Efficiency of a Given Portfolio," Econometrica 57, 1121-1152.

Grinblatt, Mark, and Sheridan Titman (1987), "Factor pricing in a finite economy," Journal of Financial Economics 12, 497-507.

Greene, William H. (2003), Econometric Analysis, 5th edition. Prentice-Hall.

Harvey, Campbell R. and Akhtar Siddique (2000), "Conditional Skewness in Asset Pricing Tests," Journal of Finance 55, 1263-1295.

Horn, Roger A. and Charles R. Johnson (1999), Matrix Analysis. Cambridge University Press.

Imhof, J. P. (1961), "Computing the Distribution of Quadratic Forms in Normal Variables," Biometrika 48, 419-426.

Jagannathan, Ravi and Zhenyu Wang (1996), "The Conditional CAPM and the Cross-Section of Expected Returns," The Journal of Finance 51, 3-53.

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

Magnus, Jan R. and Heinz Neudecker (2002), Matrix Differential Calculus with Applications in Statistics and Econometrics. Wiley.

Merton, Robert C. (1973), "An intertemporal capital asset pricing model," Econometrica 41, 867-887.

Mossin, Jan (1966), "Equilibrium in a Capital Asset Market," Econometrica 34, 768-783.

Ross, Stephen A. (1976), "The Arbitrage Theory of Capital Asset Pricing," Journal of Economic Theory 13, 341-360.

Rubinstein, M., (1973), "The Fundamental Theorem of Parameter-preference Security Valuation," Journal of Financial and Quantitative Analysis 8, 61-69.

Ruud, P. (2000), An Introduction to Classical Econometric Theory. Oxford University Press.

Searle, Shayle R. (1971), Linear Models. John Wiley.

Sharpe, William F. (1964), "Capital asset prices: A theory of market equilibrium under conditions of risk," Journal of Finance 19, 425-442.

Singleton, Kenneth (2006) Empirical Dynamic Asset Pricing: Model Specification and Econometric Assessment. Princeton University Press.

Appendix A: Proofs of Theorems

Proof of Theorem 1:
For the χ_1 result in Table 3, we have:

$$\tfrac{1}{n}\sum_i \hat{\alpha}_i^2 = \tfrac{1}{n}\hat{\alpha}'\hat{\alpha} = \tfrac{1}{n}\mathbf{y}'(I_n \otimes a_x')'(I_n \otimes a_x')\,\mathbf{y}.$$

To evaluate $(I_n \otimes a_x')'$ in the above equation, for any matrix of the form $M_1 \otimes M_2$, its transpose is $M_1' \otimes M_2'$,[23] hence $(I_n \otimes a_x')' = I_n \otimes a_x$. Consequently, we can write the product $(I_n \otimes a_x')'(I_n \otimes a_x')$ as $(I_n \otimes a_x)(I_n \otimes a_x')$, which can be re-written as $I_n \otimes a_x a_x'$.[24] Plugging this in, above, we obtain part (i) of the theorem.

To get the rank of kernel S in the χ_1 case, note that S is of the form AA', with $A = n^{-1/2} a_x \otimes I_n$. The rank of S is then equal to the rank of A (see Magnus and Neudecker (2002, page 9)). By construction, the matrix A has dimensions $nT \times n$, hence has rank at most n. Also, the columns of n are orthogonal to each other, hence span a space of dimension n, so A has rank n, as does S.

For the χ_2 result in Table 3: See Table 2, last row, for a formula for $(nT)^{-1}\sum(y_{it} - \hat{y}_{it}))^2$.

For the ϕ result: It suffices to apply our results for the kernel of components χ_1 and χ_2, and note that the sum $S_{\chi_1} + S_{\chi_2}$ equals the kernel S_ϕ of ϕ.

For the ψ_1 result: Note that $(\sum_i \hat{\alpha})^2 = \hat{\alpha}' 1_{n,n} \hat{\alpha}$. Also, from the proof of Theorem 1, $\hat{\alpha} = (I_n \otimes a_x')y$. Therefore, $S_{sae} = \frac{1}{n^2}(I_n \otimes a_x')' 1_{n,n} (I_n \otimes a_x')$. Also, $(I_n \otimes a_x')' = I_n \otimes a_x$, as noted earlier. With $1_{n,n} = 1_n 1_n'$, and a_x a $T \times 1$ vector, we can write $(I_n \otimes a_x)1_n = 1_n \otimes a_x$, likewise $1_n'(I_n \otimes a_x') = 1_n' \otimes a_x'$. Hence $(\sum_i \hat{\alpha})^2 = (1_n \otimes a_x)(1_n' \otimes a_x')$. The last expression is of the form $(M_1 \otimes M_2)(M_3 \otimes M_4)$, which simplifies $(M_1 M_3) \otimes (M_2 M_4)$ when the products $M_1 M_3$ and $M_2 M_4$ are well-defined (see Magnus and Neudecker...), in which case to $(1_n \otimes a_x)(1_n' \otimes a_x') = (1_n 1_n') \otimes (a_x a_x')$. Since $1_n 1_n' = 1_{n,n}$, the result follows.

For the ψ_2 result: Note that $\sum_i(\hat{\alpha}_i - (\frac{1}{n}\sum_i \hat{\alpha}_i))^2 = (\sum_i \hat{\alpha}_i^2) - n(\frac{1}{n}\sum_i \hat{\alpha}_i)^2$. Applying Theorem 1.1 to $\sum_i \hat{\alpha}_i^2$ and Theorem 1.4 to $(\frac{1}{n}\sum_i \hat{\alpha}_i)^2$, we get $\sum_i(\hat{\alpha}_i - (\frac{1}{n}\sum_i \hat{\alpha}_i))^2 = I_n \otimes (a_x a_x') - \frac{1}{n} 1_{n,n} \otimes (a_x a_x')$. Simplifying, the result follows.

For the ω result: It suffices to apply our earlier kernel results for the components χ_2 and ψ_2 of ω, and note that the sum $S_{\chi_2} + S_{\psi_2}$ of these kernels equals the kernel S_ω for ω.

[23] See Magnus and Neudecker (2002,p. 28).

[24] Each of these block matrices is a Kronecker sum, and their product is also a Kronecker sum with square block $a_x a_x'$ repeating on the diagonal.

PROOF of Corollary 1
Here and later, we will make use of:

Lemma 1 (Magnus and Neudecker (2002, page 28)): For matrices M_1, M_2, M_3, M_4 such that matrix products M_1M_3 and M_2M_4 are well-defined, the following holds: $(M_1 \otimes M_2)(M_3 \otimes M_4) = (M_1M_3) \otimes (M_2M_4)$.

It is easy to verify, via Table 3, that $S_\phi = S_{\chi_1} + S_{\chi_2} = S_{\psi_1} + S_\omega$, $S_{\chi_1} = S_{\psi_1} + S_{\psi_2}$, $S_\omega = S_{\chi_2} + S_{\psi_2}$. To show that $S_{\chi_1}S_{\chi_2} = 0_{nT,nT}$, first note that:

$$\alpha_x'(I_{nT} - I_n \otimes (X(X'X)^{-1}X')) = 0_{T,T} \tag{32}$$

this being a standard result which indicates, in terse matrix terms, that the regression residuals, if themselves regressed on x, must generate an intercept (vector) equal to 0. Consequently:

$$S_{\chi_1}\, S_{\chi_2} = (I_n \otimes (\tfrac{1}{n}a_xa_x'))\,(I_{nT} - I_n \otimes (X(X'X)^{-1}X') =$$

$$I_n \otimes (\tfrac{1}{n}a_xa_x') - I_n \otimes ((\tfrac{1}{n}a_xa_x')(\tfrac{1}{nT}(X(X'X)^{-1}X')) =$$

$$I_n \otimes ((\tfrac{1}{n}a_x(a_x'(I_{nT} - \tfrac{1}{nT}(X(X'X)^{-1}X'))) = I_n \otimes 0_{T,T} = 0_{nT,nT}$$

with the last line following from equation 32 and Lemma 1.

We can similarly verify that $S_{\psi_1}S_{\psi_2} = 0_{nT,nT}$ since, from Table 3,

$$S_{\psi_1}S_{\psi_2} = \left(\tfrac{1}{n}1_{n,n} \otimes (a_xa_x')\right)\,\left(\tfrac{1}{n}(I_n - \tfrac{1}{n}1_{n,n}) \otimes (a_xa_x')\right)$$

Applying Lemma 1, this expression simplifies to:

$$S_{\psi_1}S_{\psi_2} = \left(\tfrac{1}{n}1_{n,n}\tfrac{1}{n}(I_n - \tfrac{1}{n}1_{n,n}) \otimes (a_xa_x')^2\right) = 0_{n,n} \otimes (a_xa_x')^2 = 0_{nT,nT}$$

For the result $S_{\psi_1}S_\omega = 0_{nT,nT}$, we have $S_\omega = S_{\chi_2} + S_{\psi_2}$, and since we showed above that $S_{\psi_1}S_{\psi_2} = 0_{nT,nT}$, it suffices to show that $S_{\chi_2}S_{\psi_1} = 0_{nT,nT}$. For this we can apply Table 3 and equation 32, applying arguments similar to those used earlier to show $S_{\chi_1}S_{\chi_2} = 0_{nT,nT}$. Likewise, we can verify that $S_{\chi_2}S_{\psi_2} = 0_{nT,nT}$.

Proof of Theorem 2
i. Table 4 results. To get the χ_1 result, with $S_{\chi_1} = \frac{1}{n}I_n \otimes (a_xa_x') = I_n \otimes (\frac{1}{n}a_xa_x')$ and $V_y = V_\varepsilon \otimes I_T$, we have $S_{\chi_1}V_y = (I_n \otimes (\frac{1}{n}a_xa_x'))(V_\varepsilon \otimes I_T)$. This is of the form $(M_1 \otimes M_2)(M_3 \otimes M_4)$, with matrices $M_1, ..., M_4$ for which M_1M_3 and M_2M_4 are defined, so by Lemma 1 the expression equals $M_1M_3 \otimes M_2M_4$. Hence $S_{\chi_1}V_y = V_\varepsilon \otimes \frac{1}{n}a_xa_x'$. Since $tr[M_1 \otimes M_2] = tr[M_1]tr[M_2]$ for any square matrices M_1 and M_2 (see Magnus and Neudecker 2002, equation 8, page 28), $tr[S_{\chi_1}V_y] = tr[V_\varepsilon]tr[\frac{1}{n}a_xa_x']$.

To get the χ_2 result, we apply the same 'tricks' used above, but here using $S_{\chi_2} = \frac{1}{n(T-K-1)}(\,I_{nT} - (I_n \otimes (X(X'X)^{-1}X'))\,)$ rather than S_{χ_1}. First,

$$S_{\chi_2} V_y = (\tfrac{1}{n(T-K-1)}(\, I_{nT} - (I_n \otimes (X(X'X)^{-1}X'))\,)(V_\varepsilon \otimes I_T),$$

which we can write as:

$$S_{\chi_2} V_y = \tfrac{1}{n(T-K-1)}(V_\varepsilon \otimes I_T - (I_n \otimes X(X'X)^{-1}X')(V_\varepsilon \otimes I_T)).$$

The expression $(I_n \otimes X(X'X)^{-1}X')(V_\varepsilon \otimes I_T)$, is of the form $(M_1 \otimes M_2)(M_3 \otimes M_4)$, in the sense discussed above, hence this expression simplifies to $V_\varepsilon \otimes X(X'X)^{-1}X'$. Combining this with the above results, we get

$$S_{\chi_2} V_y = V_\varepsilon \otimes (\tfrac{1}{n(T-K-1)}(I_T - X(X'X)^{-1}X')).$$

Proceeding as we did with the form $S_{\chi_1} V_y$ earlier, we then get

$$tr[S_{\chi_2} V_y] = tr[V_\varepsilon] tr[(\frac{1}{n(T-K-1)}(I_T - X(X'X)^{-1}X'))].$$

To get the ϕ result, first note that $S_\phi V_y = (S_{\chi_1} + S_{\chi_2})V_y$. Applying formulas for $S_{\chi_1} V_y$ and $S_{\chi_2} V_y$, from the preceeding discussion, we get

$$\begin{aligned} tr[S_\phi V_y] &= tr[(V_\varepsilon \otimes \tfrac{1}{n}\, a_x a_x' + V_\varepsilon \otimes \tfrac{1}{(n(T-K-1))^2}(I_T - X(X'X)^{-1}X'))] \\ &= tr[S_{\chi_1} V_y] + tr[S_{\chi_2} V_y]. \end{aligned}$$

Deriving the ψ_1 and ψ_2 results is analogous to deriving the χ_1 and χ_2 results. The ω result is likewise similar.

ii. Table 5 results. To derive the χ_1 result, recall that $S_{\chi_1} = I_n \otimes (\frac{1}{n} a_x a_x')$ is a block-diagonal $(nT \times nT)$ matrix with the $T \times T$ matrix $\frac{1}{n} a_x a_x'$ on its diagonal. The expression for $\mu_y' S_{\chi_1} \mu_y$ then readily follows. Similarly for the χ_2 result, where the $T \times T$ matrix on the diagonal is $(I_T - X(X'X)^{-1}X'))$. This stems from the observation that

$$\begin{aligned} S_{\chi_2} &= \tfrac{1}{n(T-K-1)}(\, I_{nT} - (I_n \otimes (X(X'X)^{-1}X'))\,) \\ &= \tfrac{1}{n(T-K-1)}(\, I_n \otimes I_T - I_n \otimes (X(X'X)^{-1}X')) \\ &= \tfrac{1}{n(T-K-1)}(\, I_n \otimes (I_T - (X(X'X)^{-1}X'))\,. \end{aligned}$$

With this, the fact that $S_\phi = S_{\chi_1} + S_{\chi_2}$ provides the ϕ result in Table5. To get the ψ_1 result, note that $S_{\psi_1} = \frac{1}{n}\, 1_{n,n} \otimes (a_x a_x')$ is a matrix consisting of $n \times n$ matrix-blocks $\frac{1}{n} \otimes (a_x a_x')$ with dimensions $(T \times T)$. Therefore, the summation is over both i and j indexes. To arrive at ψ_2 result, it is sufficient to observe that $S_{\psi_2} = \dfrac{1}{n}\,(I_n - \dfrac{1}{n} 1_{n,n}) \otimes (a_x a_x') = S_{\chi_2} - \dfrac{1}{n}\, S_{\psi_1}$. Similarly for the ω result, where $S_\omega = S_{\chi_2} + S_{\psi_2}$.

Proof of Theorem 3

i. Table 6 results

To get the χ_1 result, we simply extend the proof of Theorem 2 where we showed that $S_{\chi_1}V_y = V_\varepsilon \otimes \frac{1}{n}a_x a_x'$. Likewise, $S_{\chi_1}V_y S_{\chi_1}V_y = (V_\varepsilon)^2 \otimes (\frac{1}{n}a_x a_x')^2$. Again using properties of Kronecker products, $tr[S_{\chi_1}V_y S_{\chi_1}V_y] = tr[(V_\varepsilon)^2]tr[(\frac{1}{n}a_x a_x')^2]$.

To get the χ_2 result, we build on the expression $S_{\chi_2}V_y$ from the proof of Theorem 2. Proceeding as we did with the form $S_{\chi_1}V_y$ earier, we get $tr[S_{\chi_2}V_y] = tr[V_\varepsilon^2]tr[(\frac{1}{nT}(I_T - X(X'X)^{-1}X'))^2]$. Noting that the matrix M given by $M = I_T - X(X'X)^{-1}X'$ is idempotent (e.g. $M^2 = M$),[25] we then get $tr[S_{\chi_2}V_y] = tr[V_\varepsilon^2]tr[(\frac{1}{(n(T-K-1))^2}(I_T - X(X'X)^{-1}X'))]$, as desired.

To get the ψ_1 result, first note that $S_\phi V_y = (S_{\chi_1} + S_{\chi_2})V_y$. Applying formulas for $S_{\chi_1}V_y$ and $S_{\chi_2}V_y$, from the preceeding discussion, we get

$$tr[S_\phi V_y S_\phi V_y] = tr[(V_\varepsilon \otimes \frac{1}{n}a_x a_x' + V_\varepsilon \otimes \frac{1}{(n(T-K-1))^2}(I_T - X(X'X)^{-1}X'))^2].$$

Using same properties of the trace and Kronecker operators as above, this expression can be further written as:

$$tr[((V_\varepsilon)^2 \otimes (\frac{1}{n}a_x a_x')^2 + (V_\varepsilon)^2 \otimes (\frac{1}{n(T-K-1)}(I_T - X(X'X)^{-1}X'))^2 + \\ 2((V_\varepsilon)^2 \otimes (\frac{1}{n}a_x a_x'(\frac{1}{n(T-K-1)}(I_T - X(X'X)^{-1}X')^2)].$$

We now make use of the fact that matrix $M = I_T - X(X'X)^{-1}X'$ is idemponent (as noted earlier). Also, we have $a_x a_x' M = 0_{T,T}$ because $a_x'M = 0$, as discussed earlier. Applying these two restrictions to the above form for $tr[S_\phi V_y S_\phi V_y]$, the result follows.

The ψ_1 and ψ_2 results are derived in a manner similar to the χ_1 and χ_2 results, and the ω result then follows.

ii. Table 7 results

For the χ_1 result, applying the 'tricks' used earlier we can write $S_{\chi_1}V_y S_{\chi_1} = V_\varepsilon \otimes (\frac{1}{n}a_x a_x')^2$. This matrix, having dimensions $Tn \times Tn$, has as its upper-left $T \times T$ block $V_{\varepsilon,11}(\alpha_x \alpha_x')^2$, its lower-right $T \times T$ block $V_{\varepsilon,nn}(a_x a_x')^2$, and so on, with the common $T \times T$ block 'repeating' itself $(a_x a_x')^2$ a total of n times in each row, and in each column. At the same time, the $Tn \times 1$ vector μ_y has as its upper-most $T \times 1$ sub-vector the expression $X\theta_1$, with θ_1 the first column of θ; its next $T \times 1$ sub-vector is $X\theta_2$, and so on. Matching $T \times T$ blocks with $T \times 1$ vectors, during multiplications represented by $\mu_y' S_{\chi_1}V_y S_{\chi_1}\mu_y$, the common form $X'(a_x a_x')^2 X$ 'repeats' itself, multiplied by differing vectors on the left and right. Simplifying, the result follows.

For the χ_2 and ϕ results, the same 'tricks' apply, but with $(\frac{1}{n}a_x a_x')^2$ replaced by

[25] This is a standard regression result, see Goldberger (1991, Ch. 14.2).

$\frac{1}{(n(T-K-1))^2}(I_T - X(X'X)^{-1})X'$ and $((\frac{1}{n}a_x a_x')^2 + \frac{1}{(n(T-K-1))^2}(I_T - X(X'X)^{-1}X')$, respectively. Also, for the ϕ result use the fact that $S_{\chi 1} V_y S_{\chi 2} = 0$, in which case $\mu_y' S_\phi V_y S_\phi \mu_y = \mu_y' S_{\chi 1} V_y S_{\chi 1} \mu_y + \mu_y' S_{\chi 2} V_y S_{\chi 2} \mu_y$.

To derive the ψ_1 result, note that $S_{\psi_1} V_y S_{\psi_1} = (1_{n,n} V_\varepsilon 1_{n,n}) \otimes (\frac{1}{n}a_x a_x')^2$. This is an $nT \times nT$ matrix with elements equal to $(\sum_{i,j=1}^n V_{\varepsilon,ij} \frac{1}{n} a_x a_x')^2$. Multiplying by μ_y from both sides delivers the result. Similarly for the ψ_2 result with $S_{\psi_2} V_y S_{\psi_2} = \Omega \otimes (\frac{1}{n}a_x a_x')^2$ where $\Omega = (I_n - \frac{1}{n}1_{n,n}) V_\varepsilon (I_n - \frac{1}{n}1_{n,n})$. Deriving the ω result resembles deriving the corresponding result in Table 5: $S_{\psi_2} V_y S_{\chi 2} = 0$ and $\mu_y' S_{\omega_2} V_y S_{\omega_2} \mu_y = \mu_y' S_{\psi_2} V_y S_{\psi_2} \mu_y + \mu_y' S_{\chi 2} V_y S_{\chi 2} \mu_y$.

Appendix B: Data description

The dependent variables to be explained are excess stock returns on 25 portfolios, sorted on size and (independently) on book-to-market equity (BE/ME). The portfolios are constructed by Fama and French (1993) as follows. The quintile breakpoints for size in a given year are based on market capitalization of NYSE stocks in June of the same year. The quintiles for book-to-market ratios are calculated using NYSE stocks using BE and ME from December of the previous year. The portfolios are then formed using stocks from NYSE, AMEX and NASDAQ, for which there is a positive book equity (from COMPUSTAT) available from December of the previous year and market equity available in June of the given year and December of the previous year. Finally, value-weighted monthly portfolio returns are computed starting in July of the current year and ending in June of the following year (stock prices are from CRSP). The excess returns are calculated using the one-month Treasury bill rate (from Ibbotson Associates).

The explanatory variables are the market excess return plus the two additional empirically motivated factors – SMB and HML – related to size and book-to-market ratios. The market excess return is defined as the value-weight return on all NYSE, AMEX, and NASDAQ stocks (from CRSP) minus the one-month Treasury bill rate (from Ibbotson Associates). The two latter factors are constructed from six portfolios, again sorted on size and book-to-market equity. The algorithm to construct these portfolios is the same as above, with size breakpoint being the median and book-to-market equity breakpoints being respectively the 30th and the 70th NYSE percentiles.

To construct the SMB factor, all available stocks are divided into two groups based on median market equity (size), *Small* and *Big*. For the HML factor, the stocks are grouped by their book-to-market equity ratios (BE/ME), and the breakpoints are the 30th and the 70th BE/ME percentiles, resulting in three BE/ME categories: *High, Medium* and *Low*. High BE/ME is consistently associated with low earnings on assets (the so called value stocks) and vice versa (the growth stocks). The returns on SMB and HMB are respectively calculated as

$$\begin{aligned} SMB &= 1/3\,(Small\ High + Small\ Medium + Small\ Low) \\ &- 1/3\,(Big\ High + Big\ Medium + Big\ Low). \end{aligned} \tag{33}$$

and

$$\begin{aligned} HML &= 1/2\,(Small\ High + Big\ High) \\ &- 1/2\,(Small\ Low + Big\ Low). \end{aligned} \tag{34}$$

Table 1: Measures of Model Error

measure	symbol	formula
total error (RMSE)	ϕ	$\sqrt{\frac{1}{n}\sum_{i=1}^{n}\alpha_i^2 + \sigma_i^2}$
non-stochastic error	χ_1	$\sqrt{\frac{1}{n}\sum_{i=1}^{n}\alpha_i^2}$
stochastic error	χ_2	$\sqrt{\frac{1}{n}\sum_{i=1}^{n}\sigma_i^2}$
(unsigned) pooled bias	ψ_1	$\lvert\bar{\alpha}\rvert$
bias heterogeneity	ψ_2	$\sqrt{\frac{1}{n}\sum_{i=1}^{n}(\alpha_i - \bar{\alpha})^2}$
total heterogeneity	ω	$\sqrt{\frac{1}{n}\sum_{i=1}^{n}(\alpha_i - \bar{\alpha})^2 + \sigma_i^2}$

Table 2: Known Statistics Expressible as $\mathbf{y}'S\mathbf{y}$

type	statistic	S	rank(S)
univariate, descriptive	$\frac{y'y}{T}$	$\frac{1}{T}I_T$	T
	$(\bar{y})^2$	$\frac{1}{T^2}1_{T,T}$	1
	$\frac{1}{T}(y-\bar{y})'(y-\bar{y})$	$\frac{1}{T}(I_T - \frac{1}{T}1_{T,T})$	$T-1$
univariate, regression	$\frac{\hat{y}'\hat{y}}{T}$	$\frac{1}{T}X(X'X)^{-1}X'$	$K+1$
	$\frac{1}{T}(y-\hat{y})'(y-\hat{y})$	$\frac{1}{T}(I_T - X(X'X)^{-1}X')$	$T-(K+1)$
multivariate	$\frac{\mathbf{y}'\mathbf{y}}{nT}$	$\frac{1}{nT}I_{nT}$	nT
	$\frac{\hat{\mathbf{y}}'\hat{\mathbf{y}}}{nT}$	$\frac{1}{nT}(\, I_n \otimes \left(X(X'X)^{-1}X'\right)\,)$	$n(K+1)$
	$\frac{(\mathbf{y}-\hat{\mathbf{y}})'(\mathbf{y}-\hat{\mathbf{y}})}{nT}$	$\frac{1}{nT}(\, I_{nT} - (I_n \otimes \left(X(X'X)^{-1}X'\right))\,)$	$T-n(K+1)$

Note: y denotes an $T \times 1$ vector $(y_1, ..., y_T)'$, as does $\hat{y}$, where $\hat{y}_t = \hat{\alpha} + \hat{\beta}' x_t$, with $\hat{\alpha}, \hat{\beta}$ the OLS estimators of α and β. β is a $K \times 1$ vector, x_t is a $K \times 1$ vector, and $X = [1_T, x]$. $\mathbf{y}$ denotes the $nT \times 1$ vector $(y_{11}, ..., y_{1T}, ..., y_{n1}, ..., y_{nT})'$.

Table 3: Kernel S for quadratic forms underlying RMSE and components

Note: As described in Table 1, ϕ = total error (RMSE), χ_1 = stochastic error, χ_2 = non-stochastic error, ψ_1 = pooled bias, ψ_2 = bias heterogeneity, ω = total heterogeneity.

error component	kernel S	rank of S
ϕ	$\frac{1}{n}(I_n \otimes a_x a_x') + \frac{1}{nT}(I_T - I_n \otimes (X(X'X)^{-1}X'))$	$T - nK$
χ_1	$\frac{1}{n}(I_n \otimes a_x a_x')$	nK
χ_2	$\frac{1}{nT}(I_{nT} - I_n \otimes (X(X'X)^{-1}X'))$	$T - n(K+1)$
ψ_1	$\frac{1}{n} 1_{n,n} \otimes (a_x a_x')$	1
ψ_2	$\frac{1}{n}(I_n - \frac{1}{n}1_{n,n}) \otimes (a_x a_x')$	$n - 1$
ω	$\frac{1}{nT}(I_T - I_{nT} \otimes (X(X'X)^{-1}X')) + \frac{1}{n}(I_n - \frac{1}{n}1_{n,n}) \otimes (a_x a_x')$	$T - nK - 1$

Table 4: Formulas for $tr[SV_y]$, in computing mean of RMSE Distributions

kernel S	$tr[SV_y]$
$S_{\chi 1}$	$tr[V_\varepsilon]\ tr[\frac{1}{n}\, a_x a_x']$
$S_{\chi 2}$	$tr[V_\varepsilon]\ tr[(\frac{1}{(n(T-K-1))}(I_T - X(X'X)^{-1}X')]$
S_ϕ	$tr[S_{\chi_1}V] + tr[S_{\chi_2}V]$
S_{ψ_1}	$tr[(1_{n,n}V_\varepsilon)]\ tr[\frac{1}{n}a_x a_x']$
S_{ψ_2}	$tr[(I_n - \frac{1}{n}1_{n,n})V_\varepsilon)]\ tr[\frac{1}{n}a_x a_x']$
S_ω	$tr[S_{\psi_2}V_y] + tr[S_{\chi_2}V_y]$

Note: $tr[\frac{1}{n}\, a_x a_x'] = \frac{1}{n}\sum_{t=1}^{T} a_{x,t}^2$

Table 5: Formulas for $\mu_y' S \mu_y$, in computing mean of RMSE distributions

kernel S	$\mu_y' S \mu_y$
$S_{\chi 1}$	$\sum_{i=1}^{n} \theta_i'\, X'(\frac{1}{n}a_x a_x')X\, \theta_i,$
$S_{\chi 2}$	$\sum_{i=1}^{n} \theta_i'\, X' \left(\frac{1}{n(T-K-1)}(I_T - X(X'X)^{-1}X'\right) X\, \theta_i$
S_ϕ	$\mu_y' S_{\chi 1}\mu_y + \mu_y' S_{\chi 2}\mu_y$
S_{ψ_1}	$\sum_{i,j=1}^{n} \theta_i'\, X'(\frac{1}{n}a_x a_x')X\, \theta_j$
S_{ψ_2}	$\mu_y' S_{\chi 1}\mu - \frac{1}{n}\mu' S_{\psi_1}\mu_y$
S_{ω_2}	$\mu' S_{\chi 2}\mu + \mu' S_{\psi_2}\mu$

Table 6: Formulas for $tr[SV_ySV_y]$, in computing variance of RMSE Distributions

kernel S	$tr[SV_ySV_y]$
S_{χ_1}	$tr[V_\varepsilon^2]\ tr[(\frac{1}{n}\, a_x a_x')^2]$
S_{χ_2}	$tr[V_\varepsilon^2]\ tr[(\frac{1}{(n(T-K-1))^2}(I_T - X(X'X)^{-1}X')]$
S_ϕ	$tr[S_{\chi_1} V S_{\chi_1} V] + tr[S_{\chi_2} V S_{\chi_2} V_y]$
S_{ψ_1}	$tr[(1_{n,n} V_\varepsilon)^2]\ tr[(\frac{1}{n} a_x a_x')^2]$
S_{ψ_2}	$tr[(I_n - 1_{n,n}) V_\varepsilon)^2]\ tr[(\frac{1}{n} a_x a_x')^2]$
S_ω	$tr[S_{\chi_2} V_y S_{\chi_2} V_y] + tr[S_{\psi_2} V_y S_{\psi_2} V_y]$

Table 7: Formulas for $\mu_y' SV_y S\mu_y$, in computing variance of RMSE distributions

kernel S	$\mu_y' SV_y S\mu_y$
S_{χ_1}	$\sum_{i,j=1}^{n} V_{\varepsilon,ij}\, \theta_i'\, X'(\frac{1}{n} a_x a_x')^2 X\, \theta_j$
S_{χ_2}	$\sum_{i,j=1}^{n} V_{\varepsilon,ij}\, \theta_i'\, X'(\frac{1}{(n(T-K-1))^2}(I_T - X(X'X)^{-1}X')X\, \theta_j$
S_ϕ	$\sum_{i,j=1}^{n} V_{\varepsilon,ij}\, \theta_i'\, X'(\,(\frac{1}{n} a_x a_x')^2 + \frac{1}{(n(T-K-1))^2}(I_T - X(X'X)^{-1}X')X\, \theta_j$
S_{ψ_1}	$\sum_{i,j=1}^{n} \theta_i'\, X'(\sum_{l,m=1}^{n} V_{\varepsilon,ml} \frac{1}{n}(a_x a_x')^2) X\, \theta_j$
S_{ψ_2}	$\sum_{i,j=1}^{n} \Psi_{ij}\, \theta_i'\, (X'(\frac{1}{n} a_x a_x')^2 X)\, \theta_j$
S_ω	$\sum_{i,j=1}^{n} \Psi_{ij}\, \theta_i'\, (X'(\frac{1}{n} a_x a_x')^2 X)\, \theta_j$ $+ \sum_{i,j=1}^{n} V_{\varepsilon,ij}\, \theta_i'\, (X'(\frac{1}{(n(T-K-1))^2}(I_T - X(X'X)^{-1}X')X)\, \theta_j$

Table 8: Descriptive Statistics for Fama-French 5x5 Stock Portfolios

Notes: (i) For 5x5 = 25 portfolios, observed on common history 1947:01 - 2007:12, y_{it} denotes the monthly excess return on portfolio i at time t, y denotes the the $nT \times 1$ vector $(y_{11}, ..., y_{1T}, ..., ..., y_{n1}, ..., y_{nT})$, $\bar{y}_i$ denotes the sample mean return for the i-th portfolio, $\bar{y}$ denotes the n-vector of mean returns, and $\hat{\sigma}^2_{y_i} = T^{-1} \sum_{t=1}^{T} (y_{it} - \bar{y}_i)^2$ is the Gaussian maximum likelihood estimator of sample variance for the i-th portfolio. (ii) $\phi = \sqrt{\chi_1^2 + \chi_2^2} = \sqrt{\psi_1^2 + \omega^2}$, $\chi_1 = \sqrt{\psi_1^2 + \psi_2^2}$.

statistic	formula	symbol	value
normed return vector	$\lvert\lvert y \rvert\rvert = \sqrt{\frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} y_{it}^2}$	ϕ	5.39
normed mean vector	$\lvert\lvert \bar{y} \rvert\rvert = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (\bar{y}_i)^2}$	χ_1	1.19
pooled variability	$\hat{\sigma}_y = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \hat{\sigma}^2_{y_i}}$	χ_2	5.26
pooled mean	$\bar{\bar{y}} = \frac{1}{n} \sum_{i=1}^{n} \bar{y}_i$	ψ_1	1.17
mean heterogeneity	$\sqrt{\frac{1}{n} \sum_{i=1}^{n} (\bar{y}_i - \bar{\bar{y}})^2}$	ψ_2	0.20
total heterogeneity	$\sqrt{\chi_2^2 + \psi_2^2}$	ω	5.26

Table 9: Sample Errors of CAPM and APT Models, for Fama-French 5x5 Stock Portfolios

Notes:
(i) For 5x5 = 25 portfolios, sample 1947:01 - 2007:12, a model for monthly excess returns (in percent form) for a portfolio i at time t is given by $y_{it} = \alpha_i + \beta_i x_t + \varepsilon_{it}$, $i = 1, 2, ..., 25$, $t = 1, 2, ..., T$.
(ii) ϕ = model error (RMSE) = $\sqrt{\chi_1^2 + \chi_2^2} = \sqrt{\psi_1^2 + \omega^2}$, $\chi_1 = \sqrt{\psi_1^2 + \psi_2^2}$, $\omega = \sqrt{\psi_1^2 + \chi_2^2}$. Definitions of ϕ, χ, etc. appear in Table 1.

model	systematic risk	total error ϕ	non-stochastic error χ_1	stochastic error χ_2	pooled bias ψ_1	bias heterog. ψ_2	total heterog. ω
CAPM	market	2.89	0.58	2.83	0.52	0.25	2.84
APT	none	5.39	1.19	5.26	1.17	0.20	5.26
	smb	4.56	1.08	4.43	1.06	0.21	4.43
	hml	5.23	1.30	5.06	1.29	0.14	5.06
	smb, hml	4.43	1.11	4.29	1.10	0.11	4.30
	market, smb	2.15	0.57	2.08	0.51	0.26	2.09
	market, hml	2.60	0.43	2.56	0.41	0.14	2.57
	all	1.68	0.38	1.64	0.35	0.13	1.65

Updated: Petr, May 22, 2008

Table 10: Theoretical Distribution of Model RMSE and its Components, based on the Sample 1947:01 2007:12, Part 1/2

Notes:
(i) ϕ = model RMSE = $\sqrt{\chi_1^2 + \chi_2^2} = \sqrt{\psi_1^2 + \omega^2}$, $\omega = \sqrt{\psi_1^2 + \chi_2^2}$, $\chi_1 = \sqrt{\psi_1^2 + \psi_2^2}$. Definitions of ϕ, χ, etc. appear in Table 1.
(ii) The mean and variance of the statistics are calculated as $E(Q) = tr[SV] + \mu' S\mu$ and $V(Q) = 2\,tr[SVSV] + 4\mu' SVS\mu$ with elements of these formulas given in Tables 4-7.
(iv) The interval $\hat{E}(Q) \pm 3\sqrt{\hat{V}(Q)}$ has, from Chebychev's inequality, has at least 89 percent chance (=100$(1 - \frac{1}{3^2})$) of containing the true Q value, in large samples.

model	risks	$\hat{E}(Q)$	$\sqrt{\hat{V}(Q)}$	$\hat{E}(Q) - 3\sqrt{\hat{V}(Q)}$	$\hat{E}(Q) + 3\sqrt{\hat{V}(Q)}$
			total error $\hat{\phi}^2 = \frac{1}{n}\sum_i \hat{\alpha}_i^2 + \frac{1}{n}\sum_i \hat{\sigma}_i^2$		
CAPM	mkt	8.34	0.25	7.60	9.08
APT	none	29.03	1.26	25.25	32.81
	smb	20.74	0.90	18.02	23.45
	hml	27.55	1.26	23.78	31.33
	smb, hml	19.62	0.90	16.90	22.33
	mkt, smb	4.63	0.12	4.28	4.97
	mkt, hml	6.75	0.22	6.07	7.42
	all	2.82	0.04	2.69	2.96
			non-stochastic error $\hat{\chi}_1^2 = \frac{1}{n}\sum_i \hat{\alpha}_i^2$		
CAPM	mkt	0.35	0.07	0.14	0.56
APT	none	1.45	0.41	0.23	2.67
	smb	1.19	0.31	0.25	2.14
	hml	1.80	0.47	0.40	3.21
	smb, hml	1.25	0.33	0.26	2.24
	mkt, smb	0.33	0.05	0.18	0.49
	mkt, hml	0.20	0.05	0.04	0.36
	all	0.15	0.02	0.10	0.19
			stochastic error $\hat{\chi}_2^2 = \frac{1}{n}\sum_i \hat{\sigma}_i^2$		
CAPM	mkt	7.99	0.24	7.29	8.70
APT	none	27.58	1.19	24.00	31.16
	smb	19.54	0.85	17.00	22.08
	hml	25.75	1.17	22.25	29.25
	smb, hml	18.37	0.84	15.84	20.89
	mkt, smb	4.29	0.10	3.99	4.60
	mkt, hml	6.55	0.22	5.89	7.20
	all	2.68	0.04	2.55	2.80

Updated: Petr, June 3, 2008

Table 11: Theoretical Distribution of Model RMSE and its Components, based on the Sample 1947:01 2007:12, Part 2/2

Notes:
(i) ϕ = model RMSE = $\sqrt{\chi_1^2+\chi_2^2} = \sqrt{\psi_1^2+\omega^2}$, $\omega = \sqrt{\psi_1^2+\chi_2^2}$, $\chi_1 = \sqrt{\psi_1^2+\psi_2^2}$. Definitions of ϕ, χ, etc. appear in Table 1. Each statistic is of the form $\sqrt{Q}$, for a qudratic form $Q = \mathbf{y}'S\mathbf{y}$.
(ii) The mean and variance of the statistics are calculated as $E(Q) = tr[SV] + \mu' S\mu$ and $V(Q) = 2\,tr[SVSV] + 4\mu' SVS\mu$ with elements of these formulas given in Tables 4-7.
(iv) The interval $\hat{E}(Q) \pm 3\sqrt{\hat{V}(Q)}$ has, from Chebychev's inequality, has at least 89 percent chance (=100$(1-\frac{1}{3^2})$) of containing the true Q value, in large samples.

model	risks	$\hat{E}(Q)$	$\sqrt{\hat{V}(Q)}$	$\hat{E}(Q) - 3\sqrt{\hat{V}(Q)}$	$\hat{E}(Q) + 3\sqrt{\hat{V}(Q)}$
				pooled bias ψ_1^2	
CAPM	mkt	6.89	1.62	2.03	11.74
APT	none	35.10	10.21	4.47	65.73
	smb	28.62	7.90	4.94	52.31
	hml	44.39	11.60	9.60	79.19
	smb, hml	30.83	8.29	5.97	55.70
	mkt, smb	6.56	0.94	3.75	9.36
	mkt, hml	4.35	1.24	0.64	8.07
	all	3.11	0.41	1.89	4.33
				bias heterogeneity ψ_2^2	
CAPM	mkt	0.07	0.02	0.01	0.13
APT	none	0.05	0.01	0.00	0.09
	smb	0.05	0.02	0.00	0.10
	hml	0.03	0.01	-0.01	0.06
	smb, hml	0.02	0.00	0.00	0.03
	mkt, smb	0.07	0.02	0.01	0.13
	mkt, hml	0.02	0.01	0.01	0.04
	all	0.02	0.01	0.01	0.03
				total heterogeneity ω^2	
CAPM	mkt	8.06	0.24	7.36	8.77
APT	none	27.63	1.19	24.05	31.21
	smb	19.59	0.85	17.05	22.13
	hml	25.78	1.17	22.28	29.28
	smb, hml	18.38	0.84	15.86	20.91
	mkt, smb	4.36	0.10	4.05	4.68
	mkt, hml	6.57	0.22	5.92	7.23
	all	2.70	0.04	2.57	2.82

Updated: Petr, June 3, 2008

Table 12: Bootstrapped Distribution of Model RMSE and its Components, 10,000 simulations, based on the Sample 1947:01 - 2007:12, 1/2

model	risks	mean	stdev	1%	5%	10%	90%	95%	99%
		total error: $\hat{\phi}^2 = \frac{1}{n}\sum_i \hat{\alpha}_i^2 + \frac{1}{n}\sum_i \hat{\sigma}_i^2$							
CAPM	market	8.34	0.54	7.19	7.50	7.67	9.04	9.26	9.77
APT	none	29.03	1.90	24.93	26.05	26.64	31.50	32.30	33.81
	smb	20.75	1.18	18.17	18.88	19.26	22.27	22.74	23.68
	hml	27.28	1.88	23.24	24.34	24.94	29.75	30.53	32.09
	smb, hml	19.63	1.16	17.11	17.83	18.16	21.14	21.60	22.51
	capm, smb	4.62	0.26	4.05	4.21	4.29	4.97	5.08	5.27
	capm, hml	6.74	0.43	5.86	6.09	6.21	7.31	7.49	7.86
	all	2.82	0.10	2.59	2.66	2.69	2.96	3.00	3.08
		non-stochastic error: $\hat{\chi}_1^2 = \frac{1}{n}\sum_i \hat{\alpha}_i^2$							
CAPM	market	0.35	0.07	0.21	0.24	0.26	0.44	0.47	0.53
APT	none	1.45	0.40	0.66	0.84	0.96	1.97	2.15	2.50
	smb	1.20	0.31	0.56	0.72	0.81	1.61	1.74	2.00
	hml	1.72	0.45	0.82	1.04	1.16	2.30	2.49	2.87
	smb, hml	1.25	0.33	0.59	0.75	0.84	1.68	1.83	2.10
	capm, smb	0.33	0.05	0.22	0.25	0.27	0.40	0.42	0.47
	capm, hml	0.20	0.05	0.10	0.12	0.13	0.27	0.29	0.35
	all	0.15	0.02	0.11	0.12	0.13	0.17	0.17	0.19
		stochastic error: $\hat{\chi}_2^2 = \frac{1}{n}\sum_i \hat{\sigma}_i^2$							
CAPM	market	7.99	0.52	6.90	7.20	7.34	8.67	8.88	9.38
APT	none	27.58	1.93	23.45	24.57	25.19	30.11	30.93	32.36
	smb	19.55	1.15	17.04	17.74	18.11	21.05	21.54	22.39
	hml	25.56	1.89	21.56	22.63	23.21	28.03	28.83	30.38
	smb, hml	18.38	1.12	15.95	16.61	16.97	19.85	20.31	21.19
	capm, smb	4.29	0.25	3.76	3.90	3.98	4.61	4.72	4.90
	capm, hml	6.55	0.41	5.70	5.92	6.04	7.09	7.26	7.62
	all	2.68	0.10	2.46	2.52	2.55	2.81	2.85	2.93

Updated: Petr, June 3, 2008

Table 13: Bootstrapped Distribution of Model RMSE and its Components, 10,000 simulations, based on the Sample 1947:01 - 2007:12, 2/2

model	risks	mean	stdev	1%	5%	10%	90%	95%	99%
				pooled bias: ψ_1^2					
CAPM	market	6.86	1.63	3.60	4.37	4.84	8.99	9.68	11.09
APT	none	35.04	10.04	15.14	19.75	22.71	48.22	52.73	61.59
	smb	28.65	7.83	12.92	16.83	18.87	39.04	42.31	49.08
	hml	42.35	11.03	20.09	25.56	28.58	56.75	61.48	70.79
	smb, hml	30.84	8.23	14.28	18.33	20.60	41.71	45.32	52.19
	capm, smb	6.54	0.93	4.51	5.10	5.39	7.78	8.16	8.86
	capm, hml	4.34	1.24	1.84	2.45	2.83	5.95	6.48	7.64
	all	3.11	0.41	2.24	2.48	2.60	3.65	3.81	4.13
				bias heterogeneity: ψ_2^2					
CAPM	market	0.07	0.02	0.03	0.04	0.05	0.10	0.11	0.13
APT	none	0.05	0.01	0.02	0.03	0.03	0.07	0.07	0.09
	smb	0.05	0.02	0.02	0.03	0.03	0.07	0.08	0.09
	hml	0.03	0.01	0.01	0.01	0.01	0.04	0.05	0.06
	smb, hml	0.02	0.00	0.01	0.01	0.01	0.02	0.02	0.03
	capm, smb	0.07	0.02	0.03	0.04	0.05	0.10	0.10	0.12
	capm, hml	0.02	0.01	0.01	0.01	0.02	0.03	0.04	0.05
	all	0.02	0.01	0.01	0.01	0.01	0.03	0.03	0.04
				total heterogeneity: ω^2					
CAPM	market	8.06	0.52	6.97	7.27	7.41	8.74	8.96	9.44
APT	none	27.63	1.93	23.49	24.62	25.24	30.16	30.98	32.40
	smb	19.60	1.16	17.08	17.78	18.15	21.11	21.59	22.44
	hml	25.59	1.89	21.58	22.65	23.24	28.06	28.85	30.40
	smb, hml	18.39	1.12	15.97	16.63	16.98	19.86	20.33	21.21
	capm, smb	4.36	0.25	3.82	3.97	4.05	4.69	4.79	4.98
	capm, hml	6.57	0.42	5.72	5.94	6.06	7.12	7.29	7.64
	all	2.70	0.10	2.48	2.54	2.57	2.83	2.87	2.95

Updated: Petr, June 3, 2008