Using Stocks or Portfolios in Tests of Factor Models*

Andrew Ang†
Columbia University and NBER

Jun Liu‡
CKGSB, SWUFE, and UCSD

Krista Schwarz§
Columbia University

This Version: 20 November 2008

JEL Classification: G12
Keywords: Specifying Base Assets, Cross-Sectional Regression, Estimating Risk Premia, APT, Efficiency Loss

*We thank Rob Grauer, Bob Hodrick, Raymond Kan, Bob Kimmel, Yuhang Xing, and Xiaoyan Zhang for helpful discussions and seminar participants at Columbia University and the CRSP forum for comments.

†Columbia Business School, 3022 Broadway 413 Uris, New York NY 10027, ph: (212) 854-9154; email: aa610@columbia.edu; WWW: http://www.columbia.edu/∼aa610.

‡Jun Liu is on leave from UC San Diego and visiting Cheung-Kong Graduate School of Business (CKGSB) and Southwestern University of Finance and Economics (SWUFE) of China. Rady School of Management, Otterson Hall, 4S148, 9500 Gilman Dr, #0553, La Jolla, CA 92093-0553; ph: (858) 534-2022; email: junliu@ucsd.edu; WWW: http://rady.ucsd.edu/faculty/directory/liu/.

§Columbia Business School, 3022 Broadway 311 Uris, New York NY 10027; email: kbs2101@columbia.edu

Using Stocks or Portfolios in Tests of Factor Models

Abstract

We examine the asymptotic efficiency of using individual stocks or portfolios as base assets to test asset pricing models using cross-sectional data. The literature has argued that creating portfolios reduces idiosyncratic volatility and allows factor loadings, and consequently risk premia, to be estimated more precisely. We show analytically and demonstrate empirically that the more efficient estimates of betas from creating portfolios do not lead to lower asymptotic variances of factor risk premia estimates. Instead, the standard errors of factor risk premia estimates are determined by the cross-sectional distribution of factor loadings and residual risk. Creating portfolios shrinks the dispersion of betas and leads to higher asymptotic standard errors of risk premia estimates.

1 Introduction

In cross-sectional factor models, expected excess returns are a linear function of factor loadings. This relation holds for all assets, whether these assets are individual stocks or whether the assets are portfolios formed from individual stocks. The literature has taken two different approaches in specifying the universe of base assets for cross-sectional regression tests. First, researchers have followed Black, Jensen and Scholes (1972) and Fama and MacBeth (1973), among many others, to group stocks into portfolios and then run factor model tests using portfolios as base assets. An alternative approach is to estimate cross-sectional risk premia using the entire universe of stocks following Litzenberger and Ramaswamy (1979) and others. Perhaps due to the easier availability of portfolios constructed by Fama and French (1993) and others, the first method of using portfolios as test assets is the more popular approach in recent empirical work.

Blume (1970) gave the original motivation for creating test portfolios of assets as a way to reduce the errors-in-variables problem. Cross-sectional regressions specify estimated betas as regressors. If the errors in the estimated betas are imperfectly correlated across assets then the estimation errors would tend to offset each other when the assets are grouped into test portfolios. Thus, using portfolios as test assets allows for more efficient estimates of factor loadings. Blume argues these more precise estimates of factor loadings enable factor risk premia to also be estimated more precisely. On the other hand, an argument stated by Litzenberger and Ramaswamy (1979) for using individual stocks as test assets is that generally far fewer than 100 portfolios, often as few as 10-25 portfolios, are often used as test assets. In contrast, in standard empirical applications with U.S. data, the number of individual stocks is currently above 5000. Thus, the number of individual stocks is usually two orders of magnitude greater than the number of portfolios commonly used in empirical tests leading to a potentially severe loss of efficiency for using portfolios as base assets.

In this paper we study the relative efficiency of using individual stocks or portfolios in tests of cross-sectional factor models. We focus on theoretical results in a very simple one-factor setting applicable to the original CAPM, but also extend our results to multifactor models and models with characteristics as well as factor loadings. We derive analytical forms using maximum likelihood but stress the intuition from these formulae apply to all types of standard errors.[1] Maximum likelihood estimators have the advantage of obtaining the Cramér-Rao lower

[1] Jobson and Korkie (1982), Huberman and Kandel (1987), MacKinlay (1987), Zhou (1991), Velu and Zhou (1999), among others, derive small-sample or exact finite sample distributions of various maximum likelihood statistics but do not consider efficiency using different test assets.

bound, so maximum likelihood standard errors serve as a useful benchmark to measure efficiency for other standard errors. Shanken (1992) shows the more commonly used two-pass methodology of Fama and MacBeth (1973) is asymptotically equivalent to the one-step approach of maximum likelihood. The Cramér-Rao lower bound can be computed with any set of consistent estimators, and since the two-pass regression estimators tend to perform best in small samples (see Chen and Kan, 2004; Shanken and Zhou, 2007), we use the point estimates from two-pass estimators for computing efficiency losses in simulations as well as actual data.[2] We also consider the effect of using portfolios versus individual stocks with pooled and Shanken (1992) standard errors, in addition to maximum likelihood standard errors, in data.

Forming portfolios dramatically reduces the standard errors of factor loadings due to decreasing idiosyncratic risk, but we show the more precise estimates of factor loadings do *not* lead to more efficient estimates of factor risk premia. In a setting where all stocks have the same idiosyncratic risk, the idiosyncratic variances of portfolios decline linearly with the number of stocks in each portfolio but the variance of the risk premia estimates increase when portfolios as used compared to the case when all stocks are used. Thus, creating portfolios to reduce estimation error in the factor loadings does not lead to smaller estimation error of the factor risk premia. Nor do we find that it is simply greater power by using a larger number of assets for individual stocks compared to using portfolios that makes estimates from employing individual stocks as test assets more efficient.

The most important determinant of the standard variances of risk premia is the cross-sectional distribution of risk factor loadings scaled by the inverse of idiosyncratic variance. Intuitively, the more disperse the cross section of betas, the more information the cross section contains to estimate risk premia. More weight is given to stocks with lower idiosyncratic volatility as these observations are less noisy. Aggregating stocks into portfolios causes the information contained in individual stock betas to become more opaque and shrinks the cross-sectional dispersion of betas. This causes estimates of factor risk premia to be less efficient when portfolios are created. We show these results with maximum likelihood by analytically computing the efficiency losses with portfolios when stock betas are normally distributed and idiosyncratic risk is constant across stocks. The same intuition applies for all types of standard errors. Furthermore, we demonstrate these results also hold when idiosyncratic volatility is stochastic and correlated

[2] Chen and Kan (2004) argue that maximum likelihood estimators do not have finite means leading to their poor small sample performance. This fact is irrelevant for the existence of the Cramér-Rao lower bound and for measuring efficiency losses relative to the Cramér-Rao lower bound, especially since we use two-pass estimators which do have good small sample performance.

with betas in Monte Carlo exercises. When betas need to be estimated to form portfolios, rather than forming portfolios on true betas, efficiency losses further increase.

Finally, we empirically verify that using portfolios leads to wider standard error bounds in estimates of one-factor and three-factor models using the CRSP database of stock returns. We find that for both a one-factor market model and the Fama and French (1993) multifactor model estimated using the full universe of stocks, the market risk premium estimate is positive and highly significant. In the one-factor model, we fail to reject the hypothesis that the market cross-sectional risk premium is equal to the time-series average of the market excess return. In contrast, using portfolios often produces insignificant and sometimes negative point estimates of the market risk premium in both one- and three-factor specifications.

From our theoretical and empirical results, the most efficient estimates of cross-sectional risk premia in linear factor models are obtained using individual stocks and there are potentially large efficiency losses using portfolios. But, we stress that our results do not mean that portfolios should never be used to test factor models. In particular, any non-linear procedure especially involving GMM necessitates using a small number of portfolios and portfolios must be constructed to measure investable returns. Our analysis is from an econometric, rather than from an investments, perspective. Our setting also considers only efficiency and we do not examine power. A large literature discusses how to test for factors in the presence of spurious sources of risk (see, for example, Kan and Zhang, 1999; Kan and Robotti, 2006; Hou and Kimmel, 2006; Burnside, 2007) holding the number of test assets fixed. From our results, efficiency will increase in all these settings when individual stocks are used. Other authors like Zhou (1991) and Shanken and Zhou (2007) examine the small-sample performance of various estimation approaches under both the null and alternative.[3] These studies do not discuss the relative efficiency of the test assets employed in cross-sectional factor model tests.

Two closely related papers that examine the effect of different portfolio groupings in testing asset pricing models are Berk (2000) and Grauer and Jamaat (2004). Berk (2000) addresses the issue of grouping stocks on a characteristic known to be correlated with expected returns and then testing an asset pricing model on the stocks within each group, rather than using all stocks or using portfolios constructed from the groups. Berk (2000) argues that this practice, as

[3] Other authors have presented alternative estimation approaches to maximum likelihood or the two-pass methodology such as Brennan, Chordia and Subrahmanyam (1998), who run cross-sectional regressions on all stocks using risk-adjusted returns as dependent variables, rather than excess returns, with the risk adjustments involving estimated factor loadings and traded risk factors. This approach cannot be used to estimate factor risk premia.

done by Daniel and Titman (1997), leads to spurious rejections of a factor model.[4] We examine the relative efficiency of portfolios formed by different groupings, where all portfolios are used, rather than just a subset of stocks or portfolios within a group that Berk (2000) examines. Grauer and Janmaat (2004) show that portfolio grouping under the alternative when a factor model is false may cause the model to appear correct. Neither Berk (2000) nor Grauer and Janmaat (2004) discuss the efficiency of using tests assets of portfolios versus individual securities or address the relative efficiency of different numbers of portfolios as test assets.

The rest of this paper is organized as follows. Section 2 presents the econometric theory and derives asymptotic standard errors concentrating on the one-factor model. We analytically characterize the efficiency loss for using portfolios as opposed to individual stocks. Section 3 compares the performance of portfolios versus stocks in simulations and in the CRSP database. Finally, Section 4 concludes.

2 Econometric Setup

Section 2.1 defines the model and estimators. We focus on a one-factor model which highlights the intuition behind the results. Section 2.2 derives asymptotic standard errors. We discuss the creation of portfolios and how they affect estimates of factor loadings in Section 2.3. Section 2.4 compares portfolios and individuals stocks as test assets. We briefly describe the multivariate case in Section 2.5.

2.1 The Model

We work with the following one-factor model:

$$R_{it} = \alpha + \beta_i \lambda + \beta_i (R_{mt} - \mu_m) + \sigma_i \varepsilon_{it}, \tag{1}$$

where R_{it}, $i = 1, ..., N$ and $t = 1, ..., T$, is the excess (over the risk-free rate) return of stock i at time t, R_{mt} is the excess return of the market index, and the parameters α, μ_m, β_i, and σ_i are constant across time. We specify the shocks ε_{it} to be IID $N(0, 1)$ over time t and uncorrelated across stocks.[5] We concentrate on the one-factor case as the intuition is easiest to see and

[4] Lo and MacKinlay (1990) point out a data-snooping bias in sorting firms by characteristics which are known to be correlated with returns in sample. We do not address this issue here.

[5] Our results also extend to the case where the error terms are cross-sectionally correlated but the notation is more complicated. We choose to work with the factor model setting where idiosyncratic risk is uncorrelated across

present results for multiple factors, such as Fama and French (1993), in Section 2.5.[6] In vector notation we can write equation (1) as

$$R_t = \alpha 1_N + \beta\lambda + \beta(R_{mt} - \mu_m) + \Sigma_\varepsilon^{1/2}\varepsilon_t, \tag{2}$$

where R_t is an $N \times 1$ vector of stock returns, α is a scalar, 1_N is a $N \times 1$ vector of ones, $\beta = (\beta_1 \ \ldots \ \beta_N)'$ is an $N \times 1$ vector of betas, Σ_ε is a $N \times N$ diagonal matrix with elements σ_i^2, and ε_t is an $N \times 1$ vector of idiosyncratic shocks.

Equation (1) states that the risk premium, or the expected excess return, of asset i is a linear function of stock i's beta:

$$\mathrm{E}(R_{it}) = \alpha + \beta_i\lambda. \tag{3}$$

This is the beta representation of Connor (1984), which is estimated by Black, Jensen and Scholes (1972) and Fama and MacBeth (1973).

Asset pricing theories impose various restrictions on α and β in equation (3). If the risk premium is given by the Arbitrage Pricing Theory or the CAPM, then

$$\alpha = 0. \tag{4}$$

If the market factor is priced with a risk premium, then

$$\lambda \neq 0. \tag{5}$$

In addition, if the risk premium is given by the CAPM,

$$\lambda = \mu_m > 0. \tag{6}$$

Most linear asset pricing models involve at least one of the restrictions imposed by equations (4)-(6). Note that α, λ, and β_i are all estimated from data and the relation between the parameters is non-linear in equation (3).

A complementary view presented in standard finance textbooks labels equation (3) the empirical Security Market Line (SML). Under the SML implied by the CAPM, a graph of expected excess returns on the y-axis versus beta on the x-axis should yield a straight line. The SML's

stocks as this is the main case examined in the literature in both cross-sectional tests, such as Shanken (1992), as well as time-series tests, such as Gibbons, Ross and Shanken (1989).

[6] Multifactor extensions can also handle a conditional CAPM as long as the conditional CAPM is estimated using an unconditional factor model test with additional factors resulting from parameterizing the time variation in risk premia or betas by linear functions of predictive instruments. The models of Jagannathan and Wang (1996), Cochrane (2001), and Lettau and Ludvigson (2001), among many others, fall into this category.

intercept should be the origin and the slope of the line should be the market risk premium. The empirical SML in equation (3) allows for two deviations from CAPM theory: a potentially non-zero intercept term, motivated from the zero-beta Black (1972) model, and the slope of the SML can be different from the average market excess return.

We derive the statistical properties of the estimators of α, λ, and β_i in equations (1)-(3). We use maximum likelihood rather than directly working with the more commonly used two-pass procedures developed by Fama and MacBeth (1973) for two reasons. First, the maximum likelihood estimators are unbiased, asymptotically efficient, and analytically tractable. We derive in closed-form the Cramér-Rao lower bound, which achieves the lowest standard errors of all consistent estimators. This represents a natural benchmark from which to measure efficiency losses. GMM standard errors, derived by Shanken (1992), Cochrane (2001), and Jagannathan, Skoulakis and Wang (2002), among others, achieve the Cramér-Rao lower bound only with additional assumptions.

Second, our results also apply to the two-pass estimators. Shanken (1992) shows that maximum likelihood and two-pass estimators are asymptotically equivalent under our standard regularity assumptions of IID error terms. Cochrane (2001) shows that the Fama-MacBeth (1973) estimates are also numerically identical to pooled time-series maximum likelihood estimates in a balanced panel with constant betas, which is the setting we use in equation (1). Shanken and Zhou (2007) also argue that maximum likelihood estimators are unbiased in small samples similar to those used in empirical work.

The log-likelihood of R_{it} is given by

$$L = -\sum_i \sum_t \frac{1}{2\sigma_i^2}\Big(R_{it} - \alpha - \beta_i\lambda - \beta_i(R_{mt} - \mu_m)\Big)^2, \tag{7}$$

ignoring the constant and the determinant of the covariance terms.[7] For notational simplicity, we assume that μ_m, σ_m, and σ_i for all i are known.[8] As argued by Merton (1980), variances

[7] Gibbons (1982) and Shanken (1985) work with an alternative empirical time-series specification of the CAPM:

$$R_{it} = \alpha_i + \beta_i(R_{mt} - \mu_m) + \sigma_i\varepsilon_{it},$$

where the CAPM imposes the restriction that $\alpha_i = \beta_i\mu_m\ \forall i$. This is a special case of our set-up with $\lambda = \mu_m$. Note the model

$$R_{it} = \alpha_i + \beta_i\lambda + \beta_i(R_{mt} - \mu_m) + \sigma_i\varepsilon_{it},$$

which allows for a stock-specific intercept term, does not allow λ to be identified and the Hessian term for λ is undefined. This arises because there is no common cross-sectional mean to identify λ.

[8] It can be verified that the maximum likelihood estimators of the parameters we do not consider are given by

are estimated very precisely at high frequencies and are much easier to estimate than means. Furthermore, the market risk premium μ_m and market volatility σ_m can be estimated separately using time-series data on the market index return.

The assumption that μ_m is known is innocuous as the role of this parameter is to make the factor shocks mean zero. We show below that the time-series mean μ_m does not enter the Hessian, even though it is present in the second derivative of the log likelihood, because the sample mean of the excess market return converges in probability to μ_m. We are also especially interested in the cross-sectional parameters α and λ rather than the time-series parameter μ_m. The cross-sectional parameters are identified only using the cross section of stock returns, not the time series of the market return. Thus, our parameters of interest are $\Theta = (\alpha, \lambda, \beta_i)$, $i = 1, \ldots, N$.

Taking the first derivative of the log-likelihood we obtain

$$\begin{aligned}
\frac{\partial L}{\partial \alpha} &= \sum_{i,t} \frac{1}{\sigma_i^2}\Big(R_{it} - \alpha - \beta_i\lambda - \beta_i(R_{mt} - \mu_m)\Big) \\
\frac{\partial L}{\partial \lambda} &= \sum_{i,t} \frac{1}{\sigma_i^2}\Big(R_{it} - \alpha - \beta_i\lambda - \beta_i(R_{mt} - \mu_m)\Big)\beta_i \\
\frac{\partial L}{\partial \beta_i} &= \sum_{t} \frac{1}{\sigma_i^2}\Big(R_{it} - \alpha - \beta_i\lambda - \beta_i(R_{mt} - \mu_m)\Big)(\lambda + R_{mt} - \mu_m). \qquad (8)
\end{aligned}$$

These equations lead to the following maximum likelihood estimators:

$$\hat{\alpha} = \frac{1}{T}\frac{\sum_{i,t}\frac{1}{\sigma_i^2}\Big(R_{it} - \hat{\beta}_i\hat{\lambda} - \hat{\beta}_i(R_{mt} - \mu_m)\Big)}{\sum_i \frac{1}{\sigma_i^2}} \qquad (9)$$

$$\hat{\lambda} = \frac{1}{T}\frac{\sum_{i,t}\frac{\hat{\beta}_i}{\sigma_i^2}\Big(R_{it} - \hat{\alpha} - \hat{\beta}_i(R_{mt} - \mu_m)\Big)}{\sum_i \frac{\hat{\beta}_i^2}{\sigma_i^2}} \qquad (10)$$

$$\hat{\beta}_i = \frac{\sum_t (R_{it} - \hat{\alpha})(\hat{\lambda} + R_{mt} - \mu_m)}{\sum_t (\hat{\lambda} + R_{mt} - \mu_m)^2}. \qquad (11)$$

From equations (9)-(11) we make the following observations:

---

the formulas

$$\begin{aligned}
\hat{\mu}_m &= \frac{1}{T}\sum_t R_{mt} \\
\hat{\sigma}_m^2 &= \frac{1}{T}\sum_t (R_{mt} - \mu_m)^2 \\
\hat{\sigma}_i^2 &= \frac{1}{T}\sum_t (R_{it} - \hat{\alpha} - \hat{\beta}_i\hat{\lambda} - \hat{\beta}_i(R_{mt} - \mu_m))^2.
\end{aligned}$$

Comment 2.1 *The maximum likelihood parameters impose the constraints under the null.*

Although the betas are defined in the data generating process (1) as

$$\beta_i = \frac{\text{cov}(R_{it} - \text{E}(R_{it}), R_{mt} - \mu_m)}{\text{var}(R_{mt})},$$

the maximum likelihood estimator of the betas in equation (11) is not the regular OLS estimator. The pricing restrictions of the expected return are imposed to gain more efficient beta estimates. Given the betas, equations (9) and (10) take a similar form as a weighted least squares (WLS) cross-sectional regression, as noted by Cochrane (2001):

$$\hat{\lambda}_{WLS} = (\hat{B}\Sigma_\varepsilon^{-1}\hat{B})^{-1}\hat{B}'\Sigma_\varepsilon^{-1}(\bar{R} - \hat{\alpha}),$$

for $\hat{\lambda}_{WLS}$ a 2×1 vector containing the WLS estimates of α and β, $\hat{B} = [1_N\ \hat{\beta}]$ corresponds to the vector notation in equation (2) with $\hat{\beta}$ being the vector of maximum likelihood estimates of β_i satisfying equation (11), $\bar{R} = (1/T)\sum_t R_t$, and we set μ_m equal to the sample mean of R_{mt}. However, we see below that a regular WLS standard error for $\hat{\lambda}$ does not apply under maximum likelihood because of the restrictions under the null.

The non-linear equations (9)-(11) can be solved iteratively (see Gibbons, 1982) or in one step (see Shanken, 1985). Shanken (1992) shows that both the maximum likelihood estimators and the more popular two-pass Fama-MacBeth (1973) cross-sectional estimators are both asymptotically efficient as $T \to \infty$ and thus are asymptotically equivalent. Because the two-pass estimators are most often used in the literature and the small sample performance of the maximum likelihood estimators and the two-pass estimators are very similar in small samples (see Shanken and Zhou, 2007), we use first-pass OLS estimates of betas and estimate risk premia coefficients in a second-pass cross-sectional regression in our empirical work. However, we derive appropriate standard errors with maximum likelihood as these achieve the Cramér-Rao lower bound. These are valid with any consistent estimators of α, λ, and β_i.

Comment 2.2 *The estimators $\hat{\alpha}$ and $\hat{\lambda}$ are negatively correlated, all else being equal.*

This is shown directly by equations (9) and (10). The earliest study of the CAPM by Douglas (1969) found that the SLM intercept term was positive and its estimated slope was lower than the average market excess return. Black, Jensen and Scholes (1972) also found that the slope of the SLM was lower than the average market excess return. Equations (9) and (11) imply that $\hat{\alpha}$ and $\hat{\beta}_i$ are negatively correlated, all else being equal. This is also observed in equation (1) as any over-estimation of beta in the panel will result in an under-estimation of alpha and vice versa.

2.2 Asymptotic Standard Errors

To derive asymptotic standard errors for the parameters Θ, the second derivative of the log-likelihood is:

$$\frac{\partial^2 L}{\partial\Theta\partial\Theta'} =$$

$$\left(\begin{array}{ccc} -T\sum_i \frac{1}{\sigma_i^2} & -T\sum_i \frac{\beta_i}{\sigma_i^2} & -\sum_t \frac{\lambda+R_{mt}-\mu_m}{\sigma_i^2} \\ -T\sum_i \frac{\beta_i}{\sigma_i^2} & -T\sum_i \frac{\beta_i^2}{\sigma_i^2} & -\sum_t \frac{\alpha+2\beta_i(\lambda+R_{mt}-\mu_m)-R_{it}}{\sigma_i^2} \\ -\sum_t \frac{\lambda+R_{mt}-\mu_m}{\sigma_i^2} & -\sum_t \frac{\alpha+2\beta_i(\lambda+R_{mt}-\mu_m)-R_{it}}{\sigma_i^2} & -\sum_t \frac{(\lambda+R_{mt}-\mu_m)^2}{\sigma_i^2} \end{array} \right).$$

The Hessian is then given by:

$$\left(\mathrm{E}\left[-\frac{\partial^2 L}{\partial\Theta\partial\Theta'}\right]\right)^{-1} = \frac{1}{T}\left(\begin{array}{ccc} \sum_i \frac{1}{\sigma_i^2} & \sum_i \frac{\beta_i}{\sigma_i^2} & \frac{\lambda}{\sigma_i^2} \\ \sum_i \frac{\beta_i}{\sigma_i^2} & \sum_i \frac{\beta_i^2}{\sigma_i^2} & \frac{\beta_i\lambda}{\sigma_i^2} \\ \frac{\lambda}{\sigma_i^2} & \frac{\beta_i\lambda}{\sigma_i^2} & \frac{\lambda^2+\sigma_m^2}{\sigma_i^2} \end{array} \right)^{-1}, \qquad (12)$$

where under the null $\frac{1}{T}\sum_t R_{mt} \to \mu_m$ and $\frac{1}{T}\sum_t R_{it} \to \alpha + \beta_i\lambda$.

We define the following cross-sectional sample moments, which we denote with a subscript c to emphasize they are cross-sectional moments:

$$\begin{aligned} \mathrm{E}_c(\beta/\sigma^2) &= \frac{1}{N}\sum_j \frac{\beta_j}{\sigma_j^2} \\ \mathrm{E}_c(\beta^2/\sigma^2) &= \frac{1}{N}\sum_j \frac{\beta_j^2}{\sigma_j^2} \\ \mathrm{E}_c(1/\sigma^2) &= \frac{1}{N}\sum_j \frac{1}{\sigma_j^2} \\ \mathrm{var}_c(\beta/\sigma^2) &= \left(\frac{1}{N}\sum_j \frac{\beta_j^2}{\sigma_j^4}\right) - \left(\frac{1}{N}\sum_j \frac{\beta_j}{\sigma_j^2}\right)^2 \\ \mathrm{cov}_c(\beta^2/\sigma^2, 1/\sigma^2) &= \left(\frac{1}{N}\sum_j \frac{\beta_j^2}{\sigma_j^4}\right) - \left(\frac{1}{N}\sum_j \frac{\beta_j^2}{\sigma_j^2}\right)\left(\frac{1}{N}\sum_j \frac{1}{\sigma_j^2}\right). \end{aligned} \qquad (13)$$

The first three expressions in equation (13) are the cross-sectional sample averages of β/σ^2, β^2/σ^2, and $1/\sigma^2$, respectively, and the last two expressions are the cross-sectional sample variance of β/σ^2 and the sample covariance between β^2/σ^2 and $1/\sigma^2$, respectively. From the last two definitions, we can write

$$\left(\sum_j \frac{\beta_j^2}{\sigma_j^2}\right)\left(\sum_j \frac{1}{\sigma_j^2}\right) - \left(\sum_j \frac{\beta_j}{\sigma_j^2}\right)^2 = N^2\Big(\mathrm{var}_c(\beta/\sigma^2) - \mathrm{cov}_c(\beta^2/\sigma^2, 1/\sigma^2)\Big). \qquad (14)$$

From the Hessian in equation (12), the asymptotic variances of $\hat{\alpha}$, $\hat{\lambda}$, and $\hat{\beta}_i$ are:

$$\text{var}(\hat{\alpha}) = \frac{1}{NT}\frac{\sigma_m^2+\lambda^2}{\sigma_m^2}\frac{\mathrm{E}_c(\beta^2/\sigma^2)}{\text{var}_c(\beta/\sigma^2)-\text{cov}_c(\beta^2/\sigma^2,1/\sigma^2)} \tag{15}$$

$$\text{var}(\hat{\lambda}) = \frac{1}{NT}\frac{\sigma_m^2+\lambda^2}{\sigma_m^2}\frac{\mathrm{E}_c(1/\sigma^2)}{\text{var}_c(\beta/\sigma^2)-\text{cov}_c(\beta^2/\sigma^2,1/\sigma^2)} \tag{16}$$

$$\text{var}(\hat{\beta}_i) = \frac{1}{T}\frac{\sigma_i^2}{(\sigma_m^2+\lambda^2)}\left(1+\frac{\lambda^2}{N\sigma_i^2\sigma_m^2}\frac{\mathrm{E}_c(\beta^2/\sigma^2)-2\beta_i\mathrm{E}_c(\beta/\sigma^2)+\beta_i^2\mathrm{E}_c(1/\sigma^2)}{\text{var}_c(\beta/\sigma^2)-\text{cov}_c(\beta^2/\sigma^2,1/\sigma^2)}\right). \tag{17}$$

The proof of equations (15) to (17) can be found in Appendix A. The analytical expressions of the asymptotic variances in equations (15)-(17) enable us to make several observations.

Comment 2.3 *Cross-sectional variation in betas is necessary to identify α and λ.*

The variance of $\hat{\alpha}$ and $\hat{\lambda}$ in equations (15) and (16) are not defined when stock returns are identically distributed with the same beta and idiosyncratic risk. This is intuitive. We can identify α and λ, which constitute the cross-sectional risk premium, only from the cross section of individual stocks. When all stocks are identical, there is no cross-sectional variation in expected returns and we cannot identify α and λ.

Comment 2.4 *The asymptotic variance of $\hat{\alpha}$ and $\hat{\lambda}$ depend on the cross-sectional distributions of betas and idiosyncratic volatility.*

Equations (15) and (16) reveal the cross-sectional distribution of betas scaled by idiosyncratic volatility determines the asymptotic variance of $\hat{\alpha}$ and $\hat{\lambda}$. Some intuition for these results can be gained from considering a standard OLS regression in a panel with independent observations exhibiting heteroskedasticity. In this case WLS is optimal and this can be implemented by dividing the regressor and regressand of each observation by residual volatility. Not surprisingly, in our setting this leads to the variances of $\hat{\alpha}$ and $\hat{\lambda}$ involving moments of $1/\sigma^2$. Intuitively, scaling by $1/\sigma^2$ places more weight on the asset betas estimated more precisely corresponding to those stocks with lower idiosyncratic volatilities. Unlike standard WLS, the regressors are estimated and not exogenous and the parameters β_i and λ enter non-linearly in the data generating process (1). These assumptions under the null are imposed on the maximum likelihood estimators and cause the maximum likelihood standard errors to be different from regular WLS.

Comment 2.5 *Cross-sectional and time-series data are useful for estimating α and λ but primarily only time-series data is useful for estimating β_i.*

In both equations (15) and (16), the variance of $\hat{\alpha}$ and $\hat{\lambda}$ depend on N and T. Under the IID error assumption, increasing the data by one time period yields another N cross-sectional observations to estimate α and λ. Thus, the standard errors follow the same convergence properties as a pooled regression with IID time-series observations, as noted by Cochrane (2001). In contrast, the variance of $\hat{\beta}_i$ in equation (17) depends primarily on the length of the data sample, T. The stock beta is specific to an individual stock, so the variance of $\hat{\beta}_i$ converges at rate $1/T$ and the convergence of $\hat{\beta}_i$ to its population value is not dependent on the size of the cross section. The standard error of $\hat{\beta}_i$ depends on a stock's idiosyncratic variance, σ_i^2, and intuitively stocks with smaller idiosyncratic variance have smaller standard errors for $\hat{\beta}_i$.

However, the cross-sectional distribution of betas and idiosyncratic variance does enter the variance of $\hat{\beta}_i$, but the effect is second order. Equation (17) has two terms. The first term involves the idiosyncratic variance for a single stock i. The second term involves cross-sectional moments of beta and idiosyncratic volatilities. The second term arises because α and λ are estimated, and the sampling variation of $\hat{\alpha}$ and $\hat{\lambda}$ contributes to the standard error of $\hat{\beta}_i$. Note that the second term is of order $1/N$ and when the cross section is large enough tends to zero.[9]

Comment 2.6 *Sampling error of the factor loadings affects the standard errors of $\hat{\alpha}$ and $\hat{\lambda}$.*

Appendix A shows that the term $(\sigma_m^2 + \lambda^2)/\sigma_m^2$ in equations (15) and (16) arise through the estimation of the betas and increases the terms involving the cross-sectional distribution of betas and idiosyncratic volatilities. This term also plays a role in the tests of Gibbons, Ross and Shanken (1989) and Shanken (1992), which take into account the estimation of the betas. For comparison, suppose that α is known or not estimated. Then, var($\hat{\lambda}$) simplifies to

$$\frac{1}{NT}\frac{\sigma_m^2 + \lambda^2}{\sigma_m^2}\frac{1}{\mathrm{E}_c(\beta^2/\sigma^2)}. \tag{18}$$

In this same setting with $\alpha = 0$, the Shanken (1992) standard variance of a WLS two-pass estimator of λ is

$$\frac{1}{T}\left(\frac{\sigma_m^2 + \lambda^2}{\sigma_m^2}(\beta'\Sigma_\varepsilon^{-1}\beta)^{-1} + \sigma_m^2\right) = \frac{1}{NT}\frac{\sigma_m^2 + \lambda^2}{\sigma_m^2}\frac{1}{\mathrm{E}_c(\beta^2/\sigma^2)} + \frac{1}{T}\sigma_m^2, \tag{19}$$

[9] It is important to note that the estimators are not N-consistent as emphasized by Jagannathan, Skoulakis and Wang (2002). That is, $\hat{\alpha} \nrightarrow \alpha$ and $\hat{\lambda} \nrightarrow \lambda$ as $N \to \infty$. The maximum likelihood estimators are only T-consistent in line with a standard Weak Law of Large Numbers. With T fixed, $\hat{\lambda}$ is estimated ex post, which Shanken (1992) terms an ex-post price of risk. As $N \to \infty$, $\hat{\lambda}$ converges to the ex-post price of risk. Only as $T \to \infty$ does $\hat{\alpha} \to \alpha$ and $\hat{\lambda} \to \lambda$.

which is also rederived by Cochrane (2001) and Jagannathan, Skoulakis and Wang (2002). The Shanken (1992) standard variance has an additional term involving the market variance which is due to using the regular OLS moment conditions to estimate the factor loadings. This term is not present in the maximum likelihood variance of $\hat{\lambda}$ because the OLS moment conditions implicitly use stock-specific constant terms to estimate the OLS betas whereas maximum likelihood imposes that the constant term is shared across all stocks from the null in equation (3).

Comment 2.7 *In the presence of characteristics, the asymptotic variance of $\hat{\alpha}$ and $\hat{\lambda}$ depend on the joint cross-sectional distribution of factor loadings and characteristics.*

We stress that we do not focus on the question of the most powerful specification test of the factor structure in equation (1) (see, for example, Daniel and Titman, 1997; Jagannathan and Wang, 1998; Lewellen, Nagel and Shanken, 2007) or whether the factor lies on the efficient frontier (see, for example, Roll and Ross, 1994; Kandel and Stambaugh, 1995). Our focus is on testing whether the model intercept term is zero and whether the factor is priced given the model structure. Nevertheless, many authors have used additional firm-specific characteristics, such as firm size and book-to-market ratios, as additional determinants of expected returns. If equation (1) is extended to

$$R_{it} = \alpha + \beta_i \lambda + z_i \gamma + \beta_i (R_{mt} - \mu_m) + \sigma_i \varepsilon_{it},$$

to allow for a firm-specific characteristic z_i so that betas alone do not fully account for the cross section of expected returns, then var($\hat{\alpha}$) and var($\hat{\lambda}$) now involve the joint cross-sectional distribution of betas and characteristics. This case is examined in Appendix B. While we leave the empirical examination of this extension to future work, we note that the same results in Section 2.4 hold for estimating the coefficient on the firm characteristic on individual stocks versus portfolios. Grouping into portfolios destroys cross-sectional information and inflates the standard error of $\hat{\alpha}$, $\hat{\lambda}$, and $\hat{\gamma}$.

Finally, the off-diagonal terms in the Hessian in equation (12) lead to the following asymptotic covariances:

$$\begin{aligned}
\mathrm{cov}(\hat{\alpha}, \hat{\lambda}) &= \frac{1}{NT} \frac{\sigma_m^2 + \lambda^2}{\sigma_m^2} \frac{-\mathrm{E}_c(\beta/\sigma^2)}{\mathrm{var}_c(\beta/\sigma^2) - \mathrm{cov}_c(\beta^2/\sigma^2, 1/\sigma^2)} && (20) \\
\mathrm{cov}(\hat{\alpha}, \hat{\beta}_i) &= \frac{1}{NT} \frac{\lambda}{\sigma_m^2} \frac{\beta_i \mathrm{E}_c(\beta/\sigma^2) - \mathrm{E}_c(\beta^2/\sigma^2)}{\mathrm{var}_c(\beta/\sigma^2) - \mathrm{cov}_c(\beta^2/\sigma^2, 1/\sigma^2)} && (21) \\
\mathrm{cov}(\hat{\lambda}, \hat{\beta}_i) &= \frac{1}{NT} \frac{\lambda}{\sigma_m^2} \frac{\mathrm{E}_c(\beta/\sigma^2) - \beta_i \mathrm{E}_c(1/\sigma^2)}{\mathrm{var}_c(\beta/\sigma^2) - \mathrm{cov}_c(\beta^2/\sigma^2, 1/\sigma^2)}. && (22)
\end{aligned}$$

From equation (20) we observe:

Comment 2.8 *The correlation between $\hat{\alpha}$ and $\hat{\lambda}$ is negative.*

This is also demonstrated by the maximum likelihood estimates in equations (9) and (10). Thus, positive estimates of α will be correlated with low slope estimates of λ, which the early studies testing the CAPM found.

2.3 Portfolios and Factor Loadings

From the properties of maximum likelihood, the estimators using all stocks are most efficient with asymptotic variances given by equation (15)-(17). If we use only P portfolios as test assets, what is the efficiency loss? This analysis has two goals. First, we examine an analytical distribution of beta to develop intuition on how forming portfolios affects the efficiency loss. Second, we ask under these settings how many portfolios are required for the efficiency loss to be negligible?

Let the portfolio weights be ϕ_{pi}, where $p = 1, \ldots, P$ and $i = 1, \ldots, N$. The returns for portfolio p are given by:

$$R_{pt} = \alpha + \beta_p \lambda + \beta_p (R_{mt} - \mu_m) + \sigma_p \varepsilon_{pt}, \tag{23}$$

where we denote the portfolio returns with a superscript p to distinguish them from the underlying securities with subscripts i, $i = 1, \ldots, N$, and

$$\begin{aligned}
\beta_p &= \sum_i \phi_{pi} \beta_i \\
\sigma_p &= \left(\sum_i \phi_{pi}^2 \sigma_p^2 \right)^{1/2} \\
\varepsilon_{pt} &= \frac{1}{\sigma^p} \sum_i \phi_{pi} \sigma_i \varepsilon_{it}.
\end{aligned} \tag{24}$$

The literature forming portfolios as test assets has predominantly used equal weights with each stock assigned to a single portfolio (see for example, Jagannathan and Wang, 1996). Typically, each portfolio contains an equal number of stocks. We follow this practice and form P portfolios, each containing N/P stocks with $\phi_{pi} = 1/P$ for stock i belonging to portfolio p and zero otherwise. Each stock is assigned to only one portfolio usually based on an estimate of a factor loading or a stock-specific characteristic. In our theoretical framework, we assume that the true betas are known; we deal with estimation error in the factor loadings in the simulation results of Section 3.1.

2.3.1 The Approach of Fama and French (1992)

An approach that uses all individual stocks but computes betas using test portfolios is Fama and French (1992). This approach would seem to have the advantage of more precisely estimated factor loadings, which come from portfolios, with the greater efficiency of using all stocks as observations. Fama and French run cross-sectional regressions using all stocks, but they use portfolios to estimate factor loadings. First, they create P portfolios and estimate betas, $\hat{\beta}_p$, for each portfolio p. Fama and French assign the estimated beta of an individual stock to be the fitted beta of the portfolio to which that stock is assigned. That is,

$$\hat{\beta}_i = \hat{\beta}_p \qquad \forall\, i \in p. \tag{25}$$

The Fama-MacBeth (1973) cross-sectional regression is then run over all stocks $i = 1, \ldots, N$ but using the portfolio betas instead of the individual stock betas. In Appendix C we show that in the context of estimating only factor risk premia, this procedure results in exactly the same risk premium coefficients as running a cross-sectional regression using the portfolios $p = 1, \ldots, P$ as test assets. Thus, estimating a pure factor premium using the approach of Fama and French (1992) on all stocks is no different from estimating a factor model using portfolios as test assets. Consequently, we do not need to separately consider the Fama and French (1992) approach in our analysis.

2.3.2 Estimates of Factor Loadings

The literature's principle motivation for grouping stocks into portfolios is that "estimates of market betas are more precise for portfolios" (Fama and French, 1993, p430). This is due to the diversification of idiosyncratic risk in portfolios. In the context of our maximum likelihood setup, equation (17) shows that the variance for $\hat{\beta}_i$ is directly proportional to idiosyncratic volatility, ignoring the small second term if the cross section is large. Going from one stock with $\beta_i = 1$ and an idiosyncratic volatility of 50% to an equally-weighted portfolio of 100 such stocks approximately decreases var($\hat{\beta}_i$) by a ratio of 100.

We can also illustrate this effect in the context of a time-series regression to estimate betas. Consider a typical small stock with $\beta_i = 1$ and $\sigma_i = 0.50$. For this stock, the R^2 of a time-series regression to estimate β_i is

$$1 - \frac{(0.50)^2}{(0.15)^2 + (0.50)^2} = 0.08.$$

with $\sigma_m = 0.15$. In contrast, consider an equally-weighted portfolio of 100 stocks all with $\beta_i = 1$ each having an idiosyncratic volatility of 50%. The idiosyncratic variance of the portfolio is

$\sigma_p = \sqrt{\sigma_i^2/100} = 0.05$. The R^2 of the time-series regression of portfolio returns on the market factor is now

$$1 - \frac{(0.05)^2}{(0.15)^2 + (0.05)^2} = 0.90.$$

Thus, portfolios dramatically decrease measurement error in the betas.

However, this substantial reduction in the standard errors of portfolio betas does not mean that the variance of $\hat{\alpha}$ and $\hat{\lambda}$ are smaller. In fact, we now show that aggregating information into portfolios generally increases the variance of $\hat{\alpha}$ and $\hat{\lambda}$ and that we can only attain the efficiency of using all stocks only in very special cases.

2.4 Comparisons of Portfolios and Individual Stocks as Test Assets

Since the maximum likelihood estimates achieve the Cramér-Rao lower bound creating subsets of this information can only do worse.[10] What drives the identification of α and β is the cross-sectional distribution of betas. Intuitively, if the individual distribution of betas is extremely diverse, there is a lot of information in the betas of individual stocks and aggregating stocks into portfolios causes the information contained in individual stocks to become more opaque. Thus, we expect the efficiency losses of creating portfolios to be largest when the distribution of betas is very disperse. Naturally, the actual cross section of factor loadings is an empirical question, which we investigate in Section 3. In this section we examine analytically a benchmark case where stock betas are normally distributed. We assume that σ_i is the same across stocks and equal to σ. In this case the asymptotic variances of $\hat{\alpha}$ and $\hat{\lambda}$ simplify to

$$\begin{aligned} \text{var}(\hat{\alpha}) &= \frac{\sigma^2}{NT} \frac{\sigma_m^2 + \lambda^2}{\sigma_m^2} \frac{\text{E}_c(\beta^2)}{\text{var}_c(\beta)} \\ \text{var}(\hat{\lambda}) &= \frac{\sigma^2}{NT} \frac{\sigma_m^2 + \lambda^2}{\sigma_m^2} \frac{1}{\text{var}_c(\beta)}. \end{aligned} \tag{26}$$

Assume that beta is normally distributed with mean μ_β and standard deviation σ_β. We create portfolios by partitioning the beta space into P sets, each containing an equal proportion of stocks. We assign all portfolios to have $1/P$ of the total mass. Denoting $N(\cdot)$ as the cumulative

[10] Berk (2000) also makes the point that the most effective way to maximize the cross-sectional differences in expected returns is to not sort stocks into groups. However, Berk focuses on first forming stocks into groups and then running cross-sectional tests within each group. In this case the cross-sectional variance of expected returns within groups is lower than the cross-sectional variation of expected returns using all stocks. Our results are different because we consider the efficiency losses of using portfolios created from all stocks, rather than just using stocks or portfolios within a group. Appendix D details a special case where creating portfolios can attain the same efficiency as using individual stocks, but it is of limited empirical application.

distribution function of the standard normal, the critical points δ_p corresponding to the standard normal are

$$N(\delta_p) = \frac{p}{P}, \quad p = 1, ..., P-1.$$

The points ζ_p, $p = 1, \ldots, P-1$ that divide the stocks into different portfolios are given by

$$\zeta_p = \mu_\beta + \sigma_\beta \delta_p. \tag{27}$$

Grouping stocks into portfolios has two effects on $\text{var}(\hat{\alpha})$ and $\text{var}(\hat{\lambda})$. First, the idiosyncratic volatilities of the portfolios change. However, the factor σ^2/N using all individual stocks in equation (26) remains the same using P portfolios since each portfolio contains equal mass $1/P$ of the stocks:

$$\frac{\sigma_p^2}{P} = \frac{(\sigma^2 P/N)}{P} = \frac{\sigma^2}{N}.$$

Thus, when idiosyncratic risk is constant, forming portfolios shrinks the standard errors of factor loadings, but this has no effect on the efficiency of the risk premium estimate. In fact, the formulas (26) involve the total amount of idiosyncratic volatility diversified by all stocks and forming portfolios does not change the total composition.[11]

Second, the variance of the portfolio betas changes compares to the variance of the individual stock betas. In particular, forming portfolios destroys some of the information in the cross-sectional dispersion of beta making the portfolios less efficient. When idiosyncratic risk is constant across stocks, the only effect that creating portfolios has on $\text{var}(\hat{\lambda})$ is to reduce the cross-sectional variance of beta compared to using all stocks, that is $\text{var}_c(\beta_p) < \text{var}_c(\beta)$.

Denoting the asymptotic variances of $\hat{\alpha}$ and $\hat{\lambda}$ computed using portfolios as $\text{var}_p(\hat{\alpha})$ and $\text{var}_p(\hat{\lambda})$, respectively, we compute the variance ratios

$$\frac{\text{var}_p(\hat{\alpha})}{\text{var}(\hat{\alpha})} \quad \text{and} \quad \frac{\text{var}_p(\hat{\lambda})}{\text{var}(\hat{\lambda})} \tag{28}$$

in forming P portfolios. The analytical expressions for the efficiency losses are derived in Appendix E. We note neither of these variance ratios involve the idiosyncratic variance of stocks.[12]

[11] This result is similar to Kandel and Stambaugh (1995) and Gauer and Janmaat (2008) who show repackaging the tests assets by linear transformations (or forming portfolios) does not change the position of the mean-variance frontier.

[12] Appendix E provides some intuition for the variance ratio $\text{var}_p(\hat{\lambda})/\text{var}(\hat{\lambda})$, which takes the form of the inverse of a numerical approximation of $\text{var}(Z^2)$ for $Z \sim N(0,1)$. This approximation evaluates the integral using non-equally spaced rectangles lying below the normal curve and the inverse of this approximation is always greater than one.

Figure 1 plots the variance ratios in equation (28) for betas drawn from a normal distribution with mean $\mu_\beta = 1.2$ and standard deviation $\sigma_\beta = 0.8$. When there are $P = 5$ portfolios, $\text{var}_p(\hat{\alpha})$ is 1.08 times larger than $\text{var}(\hat{\alpha})$ and $\text{var}_p(\hat{\lambda})$ is 1.11 times larger than $\text{var}(\hat{\lambda})$. For $P = 10$ portfolios, the ratio $\text{var}_p(\hat{\lambda})/\text{var}(\hat{\lambda})$ is 1.04 and even at $P = 20$ portfolios the variance ratios for both $\hat{\alpha}$ and $\hat{\lambda}$ remain above 1.01.

Figure 1 may suggest that there is very little lost in using the standard 25 portfolios (Fama and French, 1993) or 100 portfolios (Fama and French, 1992) in cross-sectional tests often employed in the literature. This is not true. While most of these portfolios have significant variation in expected returns, this is not due to forming the portfolios strictly on factor loadings. Nor is this variation in expected returns necessarily highly correlated with factor loading dispersion. For example, the 10×10 portfolios created by Fama and French (1992) and Jagannathan and Wang (1996) rank stocks on beta and size. Size is correlated with beta and other factor loadings, but the correlation is low (see Daniel and Titman, 1997). Thus, there are effectively little more than 10 portfolios ranked only on beta. In the 25 portfolios of Fama and French (1993), portfolios are formed on size and book-to-market ratios without any role for beta. These portfolios deliver very low beta dispersion. More recently, Pástor and Stambaugh (2003) use only 10 portfolios sorted on a liquidity factor loading. Thus, for many studies Figure 1 suggests the efficiency losses in creating portfolios may be significant.

We illustrate the shrinking estimation errors of beta in Figure 2, which plots two standard error bars in vertical lines for the case of a sample size of $T = 60$ with $N = 1000$ stocks. We graph various percentiles of the true beta distribution with circles. For individual stocks, the typical standard error of $\hat{\beta}_i$ is around 0.38. When we create portfolios, equation (17) shows that $\text{var}(\hat{\beta}_i)$ shrinks by approximately the number of stocks in each portfolio, which is N/P. Figure 2 graphs two standard error bars of five portfolio betas in crosses linked by the solid line. These are graphed at the mid-point percentiles of each portfolio. The standard errors for $\hat{\beta}_p$ are much smaller, at around 0.04, but Figure 2 also clearly shows the cross-sectional dispersion of β_p is smaller than the cross-sectional dispersion of all stock betas. It is this shrinking of the cross-sectional dispersion of betas that causes $\text{var}(\hat{\alpha})$ and $\text{var}(\hat{\beta})$ to increase when portfolios are used.

2.5 Multivariate Case

For completeness we extend the results to the following K multifactor model:

$$R_{it} = \alpha + \beta_i'\lambda + \beta_i'F_t + \sigma_i\varepsilon_{it}, \tag{29}$$

for the $K \times 1$ factor vector F_t and the vector of factor loadings β_i is now $K \times 1$. We assume without loss of generality that F_t has mean zero and covariance matrix $\mathrm{E}(F_t F_t') = \Sigma_F$. We assume the $K \times K$ matrix Σ_F and the scalar idiosyncratic volatility σ_i are known and we are interested in estimating the intercept scalar α and the $K \times 1$ vector of cross-sectional risk premia λ. In this setting with $i = 1, \ldots, N$ stocks, the asymptotic variance of $\hat{\alpha}$ and $\hat{\lambda}$ are

$$\mathrm{var}(\hat{\alpha}) = \frac{1}{NT} \frac{(\mathrm{E}_c(1/\sigma^2) - \mathrm{E}_c(\beta/\sigma^2)'[\mathrm{E}_c(\beta\beta'/\sigma^2)]^{-1}\mathrm{E}_c(\beta/\sigma^2))^{-1}}{1 - \lambda'(\lambda\lambda' + \Sigma_F)^{-1}\lambda} \tag{30}$$

$$\mathrm{var}(\hat{\lambda}) = \frac{1}{NT} \frac{(\mathrm{E}_c(\beta\beta'/\sigma^2) - \mathrm{E}_c(\beta/\sigma^2)[\mathrm{E}_c(1/\sigma^2)]^{-1}\mathrm{E}_c(\beta/\sigma^2)')^{-1}}{1 - \lambda'(\lambda\lambda' + \Sigma_F)^{-1}\lambda}, \tag{31}$$

where the cross-sectional moments are given by

$$\begin{aligned}
\mathrm{E}_c(\beta/\sigma^2) &= \frac{1}{N}\sum_j \frac{\beta_j}{\sigma_j^2} \quad [K \times 1] \\
\mathrm{E}_c(\beta\beta'/\sigma^2) &= \frac{1}{N}\sum_j \frac{\beta_j\beta_j'}{\sigma_j^2} \quad [K \times K] \\
\mathrm{E}_c(1/\sigma^2) &= \frac{1}{N}\sum_j \frac{1}{\sigma_j^2} \quad [1 \times 1],
\end{aligned} \tag{32}$$

with the dimensions given in square brackets. Equations (30) and (31) are derived in derived in Appendix F and simplify to the one-factor model case (15) and (16) for $K = 1$ with $F_t = R_{mt} - \mu_m$ and $\Sigma_F = \sigma_m^2$. The same intuition that creating portfolios shrinks the cross-sectional dispersions of factor loadings holds in the multifactor case, where $\mathrm{E}_c(\beta_p/\sigma^2)$ and $\mathrm{E}_c(\beta_p\beta_p'/\sigma^2)$ are now in matrix form relative to the scalar case in equation (32).

3 Empirical Work

In this section we characterize the increase in standard errors resulting from using portfolios versus individual stocks to estimate a cross-sectional factor model. Section 3.1 reports results of Monte Carlo simulations that extend the analytical characterization of the previous section. We compare estimates of a one-factor market model on the CRSP universe in Section 3.2 and the Fama-French (1993) three-factor model in Section 3.3.

3.1 Monte Carlo Simulations

Although Section 2.4 demonstrates that creating portfolios may result in large efficiency losses relative to using individual stocks, there are two remaining issues we investigate with Monte

Carlo simulations. First, we allow idiosyncratic volatility to be stochastic and correlated with betas. Second, we previously assumed that portfolios are created ranking on true betas whereas in practice betas must be estimated. We show the estimation error in the betas further contributes to efficiency losses.

We consider the following data generating process in which the CAPM holds:

$$R_{it} = \beta_i \mu_m + \beta_i (R_{mt} - \mu_m) + \sigma_i \varepsilon_{it}. \tag{33}$$

We simulate data at a monthly frequency where the market excess returns $R_{mt} \sim N(\mu_m, \sigma_m^2)$, where $\mu_m = 0.06/12$ and $\sigma_m = 0.15/\sqrt{12}$. We specify a joint normal distribution for $(\beta_i, \ln \sigma_i)$:

$$\begin{pmatrix} \beta_i \\ \ln \sigma_i \end{pmatrix} \sim N \left(\begin{pmatrix} 1.09 \\ -1.03 \end{pmatrix}, \begin{pmatrix} (0.77)^2 & (0.43)(0.77)(0.58) \\ (0.43)(0.77)(0.58) & (0.58)^2 \end{pmatrix} \right) \tag{34}$$

with the $\ln \sigma_i$ parameters set for an annual frequency. To obtain monthly σ_i values we employ the transformation $\exp(v)/\sqrt{12}$ for v generated from the $\ln \sigma_i$ process in equation (34). All of these parameters are calibrated to the sample 1961-2005 described below in Section 3.2. From this generated data, we compute the standard errors of $\hat{\alpha}$ and $\hat{\lambda}$ in the estimated process (1), which are given in equations (15) and (16).

We simulate small samples of size $T = 60$ months with $N = 5000$ stocks in the cross section. We use OLS beta estimates to form portfolios using the ex-post betas estimated over the sample. Note that these portfolios are formed ex post at the end of the period and are not tradable portfolios. We also form portfolios using the true betas of each small sample following the analytical characterization in Section 2.4. Then, we compute the variance ratios in equation (28) using the true simulated parameter values in each small sample because these are the actual efficiency losses. We simulate $M = 10,000$ small samples and report the mean, median and standard deviation of variance ratio statistics across the generated small samples. Table 1 reports the results. In all cases the mean and medians are very similar and the standard deviations of the variance ratios are very small at less than 1/10th the value of the mean or median.

Panel A forms P portfolios on true betas and shows that forming as few as $P = 5$ portfolios leads to standard variances 2.99 and 3.10 times larger for $\hat{\alpha}$ and $\hat{\lambda}$, respectively. These are substantially higher than the setting of Section 2.4 where idiosyncratic risk was constant across stocks and betas were normally distributed, where the corresponding variance ratios were 1.08 and 1.11 for $P = 5$ portfolios. Even when 2500 portfolios are used with each portfolio containing two stocks, the variance ratios are 1.60 for both $\hat{\alpha}$ and $\hat{\lambda}$. This substantial increase can be traced to two sources. First, we work with a small sample of $N = 5000$ stocks rather than an entire distribution of stocks as in Section 2.4. The effect of this channel is very small because $N = 5000$ is more than enough to cover the normal distribution of betas and idiosyncratic volatility very well. Second, there is now cross-sectional variation in σ_i and this is positively correlated with betas. Creating portfolios causes the significantly shrinks the $-\text{cov}_c(\beta^2/\sigma^2, 1/\sigma^2)$ term in equations (15) and (16) causing the standard variances using portfolios to substantially increase. When the correlation of beta and $\ln \sigma$ is set higher than our value of 0.43, there can be substantial increases in the efficiency losses from using portfolios.

In Panel B, we form portfolios on OLS estimated betas.[13] When the betas are estimated, creating portfolios further increases the efficiency losses. For $P = 25$ portfolios the mean variance ratio $\text{var}_p(\hat{\lambda})/\text{var}(\hat{\lambda})$ is 5.14 in Panel B compared to 3.02 in Panel A when portfolios are formed on the true betas. For $P = 100$ portfolios formed on estimated betas, the mean variance ratio for $\hat{\lambda}$ is 4.95. Thus, the efficiency losses increase considerably once portfolios are formed on estimated betas. More sophisticated approaches to estimating betas, such as Avramov and Chordia (2006) and Meng, Hu and Bai (2007), do not make the performance of using portfolios any better because these methods can be applied at both the stock and the portfolio level.

When betas are estimated, the cross section of estimated betas is wider, by construction, than the cross section of true betas. These estimation errors are diversifiable in portfolios, which is why the $P = 5$ and $P = 10$ portfolio variance ratios are slightly lower than the moderately large $P = 25$ or $P = 50$ cases. For example, the variance ratio for $\hat{\lambda}$ is 4.61 for $P = 5$ when we sort on estimated betas, but 5.14 using $P = 25$ portfolios. Interestingly, the efficiency losses are greatest for using $P = 25$ portfolios, a number often used in empirical work. As the number of portfolios further increases, the diversification of beta estimation error becomes minimal, but this is outweighed by the increasing dispersion in the cross section of (noisy) betas causing the variance ratios to decrease. These two offsetting effects cause the slight hump-shape in the variance ratios in Panel B.

In summary, when idiosyncratic volatility is correlated with betas, the efficiency losses associated with using portfolios instead of individual stocks in asset pricing tests are even larger than when idiosyncratic volatility is constant across stocks. When portfolios are formed based

[13] We confirm Shanken and Zhou (2007) that the maximum likelihood estimates are very close to the two-pass cross-sectional estimates and portfolios formed on maximum likelihood estimates give very similar results to portfolios formed on the OLS betas.

on estimated, rather than true betas, the efficiency losses are further magnified.

3.2 Empirical Estimates of a One-Factor Model

In this section we estimate a one-factor market model using the CRSP universe of individual stocks or using portfolios. Our empirical strategy mirrors the data generating process (1) and looks at the relation between realized factor loadings and realized average returns. We take the CRSP value-weighted excess market return to be the single factor but do not assume that its mean, μ_m, is equal to λ. We do not claim that the unconditional CAPM is appropriate or truly holds, rather our purpose is to illustrate the differences on parameter estimates and the standard errors of $\hat{\alpha}$ and $\hat{\lambda}$ when the entire sample of stocks is used compared to creating test portfolios.

3.2.1 Distribution of Betas and Idiosyncratic Volatility

We work in non-overlapping five-year periods, which is a trade-off between a long enough sample period for estimation but over which an average true (not estimated) stock beta is unlikely to change drastically (see comments by Lewellen and Nagel, 2006; Ang and Chen, 2007). Our first five-year period is from January 1961 to December 1965 and our last five-year period is from January 2001 to December 2005. We consider each stock to be a different draw from equation (1). All our data is at the monthly frequency and we take all stocks listed on NYSE, AMEX, and NASDAQ with share type codes of 10 or 11. In order to include a stock in our universe it must be traded at the end of each five-year period and must have data for at least three out of five years. Our stock returns are in excess of the Ibbotson one-month T-bill rate. In our empirical work we report regular OLS estimates of betas and use second-pass estimates of α and λ to construct standard errors.

Table 2 reports summary statistics of the beta and idiosyncratic volatilities across firms. The full sample contains 29,096 firm observations. As expected, betas are centered around one with the beta distribution having a mean of 1.093 and a standard deviation of 0.765. The average annualized idiosyncratic volatility is 0.425 with a standard deviation of 0.278. Average idiosyncratic volatility has generally increased over the 1961-2000 period beginning at 0.278 and ending at 0.563, as Campbell et al. (2001) find, but declines at the end of 2005 consistent with Bekaert, Hodrick and Zhang (2008). The cross-sectional dispersion of σ and $\ln \sigma$ has also increased over the sample. Stocks with high idiosyncratic volatilities tend to be stocks with high betas, with the correlation between beta and $\ln \sigma$ equal to 0.430.

In Figure 3, we plot empirical histograms of beta (top panel) and $\ln \sigma$ (bottom panel) over all firm observations. The distribution of beta is positively skewed, at 0.783 and fat-tailed with an excess kurtosis of 3.412. This implies there is valuable cross-sectional dispersion information in the tails of betas which forming portfolios may destroy. The distribution of $\ln \sigma$ is fairly normal, with almost zero skew at 0.0161 and little excess kurtosis of 0.326. The behavior of near-normal residuals for $\ln \sigma$ is most commonly seen in a time-series context like the stochastic volatility models of Jacquier, Polson and Rossi (1994) and others who specify $\ln \sigma$ as a stochastic process, but Figure 3 shows that the cross-sectional distribution of $\ln \sigma$ is also well-approximated by a normal distribution.

3.2.2 Using All Stocks

Panel A of Table 3 reports the estimates of α and λ in equation (1) using all 29,096 firm observations. The estimates are produced by the two-pass methodology so OLS betas are estimated for each stock over each five-year period. The factor model in equation (1) implies a relation between realized firm excess returns and realized firm betas. Thus, we stack all stocks excess returns from each five-year period into one panel and run the second cross-sectional regression using realized firm excess returns over each five-year period as the regressand and the estimated betas as the regressor.

Using the two-pass consistent estimates we compute various standard errors and t-statistics. We compute the maximum likelihood standard errors (equations (15) and (16)) in the column labeled "Max Lik." The columns labeled "Pooled" report robust pooled standard errors where the clustering is done at the firm or portfolio level in each five-year period. The next set of columns labeled "Shanken" report Shanken (1992) standard errors. The last three columns of Table 3 report statistics of the cross-sectional dispersion of beta: the cross-sectional standard deviation, $\sigma_c(\hat{\beta})$, and the beta values corresponding to the 5%- and 95%-tiles of the cross-sectional distribution of beta.

Using all stocks produces annualized estimates of $\hat{\alpha} = 6.14\%$ and $\hat{\lambda} = 5.24\%$. The maximum likelihood standard errors are 0.11 and 0.12, respectively, with t-statistics of 56.5 and 44.2. The pooled standard errors allow for serial autocorrelation and this reduces the t-statistics to 21.0 and 20.2, respectively. Finally, the Shanken (1992) standard errors are produced using a different set of moment conditions not imposing the cross-sectional structure of the model on the first-stage estimated betas. Shanken t-statistics corresponding to $\hat{\alpha}$ and $\hat{\lambda}$ are 14.5 and 6.60, respectively. All of these t-statistics reject the CAPM since the hypothesis $\alpha = 0$ is rejected.

Clearly while the CAPM is rejected, we also reject that $\lambda = 0$ so the market factor is priced. In fact, over 1961-2005, the market excess return is $\mu_m = 5.76\%$ per annum, which is very close to the estimate $\hat{\lambda} = 5.24\%$ per annum. We fail to reject the hypothesis that $\hat{\lambda} = \mu_m$ using pooled and Shanken standard errors.[14]

The t-statistics in Panel A are large compared to most of the literature for two main reasons. First, we test the relation of realized returns with realized betas over the same sample period on individual stocks. The magnitudes of these t-statistics are comparable, especially for the Shanken (1992) standard errors, to other studies with same experimental design like Ang, Chen and Xing (2006). Second, the literature often reports t-statistics using portfolios, particularly portfolios sorted on predicted rather than realized betas. Our theoretical results show there should be a potentially large loss of efficiency in using portfolios. We examine portfolio sorts in Panels B and C of Table 3. What is important is not the differences across the various standard errors (maximum likelihood versus pooled versus Shanken), rather, we should focus on the increase in the standard errors, or the decrease in the absolute values of the t-statistics, over each type of standard error as we form portfolios. We now investigate these effects.

3.2.3 "Ex-Post" Portfolios

We form "ex-post" portfolios in Panel B of Table 3. Over each five-year period we group stocks into P portfolios based on realized OLS estimated betas over those five years. All stocks are equally weighted at the end of the five year period within each portfolio. Thus, these portfolios are formed ex post and are not tradeable. Nevertheless, they represent valid test assets to estimate the cross-sectional model (1) as we can still measure the relation between realized covariances with the market and realized average returns. In all cases, $\hat{\alpha}$ and $\hat{\lambda}$ estimated using the ex-post portfolios are very close to the estimates computed using all stocks.

However, the standard errors using portfolios are much larger than the standard errors computed using all stocks. For example, for $P = 25$ portfolios the maximum likelihood standard error on $\hat{\lambda}$ is 1.90 compared with 0.12 using all stocks. The corresponding Shanken (1992) standard errors are 1.85 using $P = 25$ portfolios and 0.79 using all stocks. As P increases, the standard errors decrease (and the t-statistics increase) to approach the values using individual stocks. But, this convergence is slow. Even at $P = 100$ portfolios the maximum likelihood

[14] The parameter μ_m is treated as fixed in this test. Taking the market excess return alone, the standard error of $\hat{\mu}_m$ is $\sigma_m/\sqrt{T}$ assuming the market return is IID. This is not related at all to the standard error of $\hat{\lambda}$ because λ is identified solely on cross-sectional, not time-series, data.

standard error for $\hat{\lambda}$ is 0.93 with a variance efficiency loss of $\text{var}_p(\hat{\lambda})/\text{var}(\hat{\lambda}) = 67$. The corresponding variance efficiency loss for pooled and Shanken standard variances are 15 and 2.5, respectively. Thus, forming portfolios ex post results in significant losses of efficiency.

The loss in efficiency results from the shrinking cross-sectional standard deviation of beta, which is reported in the last three columns of Table 3. Over all stocks, the cross-sectional standard deviation of beta $\sigma_c(\hat{\beta}) = 0.77$. For $P = 25$ and $P = 100$ portfolios the cross-sectional standard deviation of beta is $\sigma_c(\hat{\beta}_p) = 0.69$. Forming portfolios shrinks the cross-sectional dispersion of beta, which destroys information and leads to efficiency losses.

3.2.4 "Ex-Ante" Portfolios

In Panel C of Table 3 we form "ex-ante" tradeable portfolios. We group stocks into portfolios at the beginning of each calendar year ranking on the market beta estimated over the previous five years. Equally-weighted portfolios are created and the portfolios are held for twelve months to produce monthly portfolio returns. The portfolios are rebalanced annually. The first estimation period is January 1956 to December 1959 to produce monthly returns for the calendar year 1961 and the last estimation period is January 2000 to December 2004 to produce monthly returns for 2005. Thus, the sample period is exactly the same as Panels A and B with all stocks and the ex-post portfolios. After the ex-ante portfolios are created, we compute realized OLS market betas of each portfolio in each non-overlapping five-year period and then run a second-pass cross-sectional regression to estimate α and λ. Thus, the same realized beta–realized return relation is tested as in Panels A and B, except the test portfolios are different.

Panel C shows the estimates of α and λ from these ex-ante portfolios are very different from Panels A and B. Using the ex-ante portfolios produces an estimate of α approximately around 10-11% and an estimate of λ close to zero. With the ex-ante portfolios we would reject the CAPM ($\alpha = 0$ and $\lambda = \mu_m$) and we also cannot reject the hypothesis that the market factor is not priced with all the t-statistics corresponding to $\hat{\lambda}$ being close to zero for all three types of standard errors.

The ex-ante portfolios produce such a markedly different $\hat{\alpha}$ and $\hat{\lambda}$ because ranking on pre-formation betas estimated over the previous five years dramatically shrinks the post-formation realized distribution of beta. It is the realized distribution of betas that is important for testing the factor model. As an example, take $P = 10$ portfolios. The average pre-formation beta for each stock in each portfolio, averaging the beginning of each calendar year, ranges from 0.245 for decile 1 to 2.332 to decile 10. The average realized post-formation beta for each portfolio, averaging across all five-year periods, ranges from 0.661 to 1.696. Thus, this portfolio formation has significantly decreased the cross-sectional dispersion of beta and this produces a very low value of $\hat{\lambda}$. Put another way, the ex-ante portfolios have a much smaller spread in realized betas to identify λ. Note that the ex-post betas in Panel B have larger beta dispersions because the portfolios are created at the end of each period, rather than at the beginning of each year. Effectively, the ex-ante portfolios have damped the information in the long tails of the beta distribution in Figure 3 even more than the ex-post portfolios.

The last three columns of Table 3, Panel C dramatically show the shrinking dispersion of the cross section of betas compared to all stocks in Panel A and the ex-post portfolios in Panel B. For $P = 25$ ex-post portfolios, the cross-sectional standard deviation of beta is only $\sigma_c(\hat{\beta}_p) = 0.37$ for the ex-post portfolios compared to $\sigma_c(\hat{\beta}) = 0.77$ using all stocks and $\sigma_c(\hat{\beta}_p) = 0.69$ for the $P = 25$ ex-ante portfolios. The 5%- and 95%-tiles show that the ex-post portfolios remove a lot of information in the tails of the beta distribution, with the 5% and 95%-tiles for the beta distributions from the $P = 25$ ex-post portfolios being 0.50 and 1.71, respectively, compared to 0.05 and 2.42 for all stocks.

3.3 Empirical Estimates of the Fama-French (1993) Model

Our final analysis considers estimates of the Fama and French (1993) model,

$$R_{it} = \alpha + \beta_{MKT,i}\lambda_{MKT} + \beta_{SMB,i}\lambda_{SMB} + +\beta_{HML,i}\lambda_{HML} + \sigma_i\varepsilon_{it}, \qquad (35)$$

where MKT is the excess market return, and SMB and HML are zero-cost portfolios formed on size and book-to-market ratios, respectively. We follow the same procedure as Section 3.2 in estimating the cross-sectional coefficients α, λ_{MKT}, λ_{SMB}, and λ_{HML} by working in non-overlapping five-year periods and stacking all observations into one panel.

3.3.1 Factor Loadings

Panel A of Table 4 reports summary statistics of the first-pass factor loadings $\hat{\beta}_{MKT}$, $\hat{\beta}_{SMB}$, and $\hat{\beta}_{HML}$. Market betas are centered around one. The SMB and HML factor loadings are not centered around zero even though SMB and HML are zero-cost portfolios, as the break points used by Fama and French (1993) to construct SMB and HML are based on NYSE stocks rather than on all stocks. Small stocks tend to skew the SMB and HML loadings to be positive, especially for the SMB loadings which have a mean of 0.805. Panel A also shows the cross-sectional correlations of the factor loadings are small. Because of the low correlations, univariate cross-sectional regressions (not reported) where each factor loading is estimated separately yield almost identical results to the multivariate regression coefficients we report below.

We report statistics on the cross-sectional distribution of the factor loadings in Panel B of Table 4: the cross-sectional standard deviation, σ_c, and the factor loadings corresponding to the 5%- and 95%-tiles of the cross-sectional distribution. Across all stocks, factor loadings of SMB and HML are more disperse than market betas, with cross-sectional standard deviations of 1.059 and 0.716, respectively, compared to a cross-sectional standard deviation of 0.716 for $\hat{\beta}_{MKT}$.

We form $n \times n \times n$ ex-post portfolios by grouping stocks into equally-weighted P portfolios based on realized estimated factor loadings at the end of the period. These are independent sorts, formed by taking intersections of n portfolios sorted on $\hat{\beta}_{MKT}$, n portfolios sorted on $\hat{\beta}_{SMB}$, and n portfolios sorted on $\hat{\beta}_{HML}$. The $n \times n \times n$ ex-ante portfolios are formed by grouping stocks into portfolios at the beginning of each calendar year ranking on the estimated factor loadings estimated over the previous five years and then the portfolios are held for twelve months to produce monthly portfolio returns. The portfolios are rebalanced annually at the beginning of each calendar year. Like the estimates of the one-factor model, with both the ex-post and ex-ante portfolios we test the relation between realized factor loadings and realized returns.

Panel B of Table 4 shows that the cross-sectional dispersion of the factor loadings decrease modestly for the ex-post portfolios. For example, for the $5 \times 5 \times 5$ portfolios, the $\hat{\beta}_{SMB}$ $\hat{\beta}_{HML}$ cross-sectional standard deviations are 0.969 and 0.946, respectively, compared to all stocks at 1.059 and 1.086, respectively. However, the ex-ante portfolios dramatically shrink the cross-sectional dispersion, with the SMB and HML factor loadings for the ex-ante $5 \times 5 \times 5$ portfolios having cross-sectional standard deviations of 0.586 and 0.390, respectively. In particular, the ex-ante portfolios significantly reduce the left-hand tail of HML factor loadings, with the 5%-tile going from -1.548 for all stocks to -0.288 for the $5 \times 5 \times 5$ ex-post portfolios. Since the cross-sectional dispersion is so much smaller for the ex-post portfolios, we might expect the second-pass cross-sectional factor risk premia estimates may be quite different for the ex-ante portfolios versus the estimates using all stocks and the ex-post portfolios. Table 5 demonstrates this is indeed the case.

3.3.2 Factor Risk Premia

Table 5 reports estimates of the Fama-French (1993) factor risk premia with pooled, maximum likelihood, and Shanken (1992) standard errors and t-statistics. Panel A using all stocks rejects the Fama-French model by rejecting that $\alpha = 0$ with all three standard errors. However, the market risk premium is priced, with $\hat{\lambda}_{MKT} = 3.50\%$ per annum and is highly significant with a maximum likelihood t-statistic of 27.7, a pooled t-statistic of 11.6 and a Shanken (1992) t-statistic of 4.28. The size premium, $\hat{\lambda}_{SMB} = 3.90\%$ per annum is slightly larger than the market risk premium, but the value premium has a negative coefficient of $\hat{\lambda}_{HML} = -2.48\%$ per annum. The negative HML coefficient is surprising given the strong evidence of a book-to-market effect across long samples, but note the high returns accrue to stocks with high book-to-market ratios rather than stocks with high HML loadings (see Daniel and Titman, 1997). Over all stocks, the correlation of realized HML loadings and realized returns is negative. This is in marked contrast to the positive cross-sectional coefficients on book-to-market ratio characteristics found in many studies like Fama and French (1992).

Panels B and C report results using ex-post and ex-ante portfolios. In both cases, there are marked efficiency losses observed in the lower magnitudes of the t-statistics for all standard errors. For example, the maximum likelihood t-statistic for $\hat{\lambda}_{MKT}$ is 1.59 for the $3 \times 3 \times 3$ ex-post portfolios and 0.57 for the $3 \times 3 \times 3$ ex-ante portfolios compared to 27.7 using all stocks. The efficiency loss ratio var$_p(\hat{\lambda}_{MKT})/$var$(\hat{\lambda}_{MKT})$ is 4.66 using Shanken (1992) standard errors for the $3 \times 3 \times 3$ ex-post portfolios. Note that the point estimates of the risk premia for the ex-post portfolios are similar to using all stocks. This is not surprising as Table 4 shows there is little decrease in the cross-sectional dispersion of the factor loadings in the ex-post portfolios compared to the dispersion in all stocks.

In contrast, the ex-ante portfolios yield very different estimates of the Fama-French (1993) factor risk premia using all stocks and the ex-post portfolios in Panels A and B, respectively. For the ex-ante portfolios in Panel C, the market risk premium is estimated to be -0.52% per annum for the $3 \times 3 \times 3$ ex-ante portfolios with a maximum likelihood t-statistic of -0.57. This has the opposite sign to the market risk premium estimate $\hat{\lambda}_{MKT} = 3.50\%$ per annum with all stocks. For the $5 \times 5 \times 5$ ex-ante portfolios, $\hat{\lambda}_{MKT} = 0.60\%$ per annum with a maximum likelihood t-statistic of 1.05. For the ex-ante portfolios the HML premium is now strongly positive, at 2.82% per annum for the $3 \times 3 \times 3$ ex-ante portfolios and 2.02% for the $5 \times 5 \times 5$ ex-ante portfolios while $\hat{\lambda}_{HML} = -2.48\%$ per annum for all stocks.[15] The dramatically different results

[15] Using the 5×5 Fama and French (1993) portfolios sorted on size and book-to-market ratio characteristics, as

are driven by the much smaller cross-sectional dispersions for the ex-ante portfolios reported in Table 4. In the very truncated cross-sectional distribution of factor loadings produced by the ex-post portfolios, realized HML loadings are positively correlated with realized returns even though they are negatively correlated in the universe of all stocks.

In summary, using portfolios instead of individual stocks also results in efficiency losses for factor risk premia estimates in the Fama-French (1993) multifactor cross-sectional regression. In particular, estimates of the market risk premium are positive using all stocks, but are insignificantly different from zero at the 95% level using ex-ante portfolios. The sign of the HML factor premium is also not robust across all stocks and using portfolios.

4 Conclusion

The finance literature takes two approaches to specifying base assets in tests of cross-sectional factor models. One approach is to aggregate stocks into portfolios for test assets. Another approach is to use the whole stock universe and run cross-sectional tests directly on all individual stocks. The motivation for creating portfolios is originally stated by Blume (1970) that betas are estimated with error and this estimation error is diversified away by aggregating stocks into portfolios. Numerous authors, Black, Jensen and Scholes (1972), Fama and MacBeth (1973), and Fama and French (1993) have used this motivation to use portfolios as base assets in factor model tests. The literature suggests that more precise estimates of factor loadings should translate into more precise estimates, and lower standard errors, of factor risk premia.

We show analytically and confirm empirically that this motivation is wrong. The sampling uncertainty of factor loadings is markedly reduced by grouping stocks into portfolios, but this does not translate into lower standard errors for factor risk premia estimates. The most important determinant of the standard variance of risk premia is the cross-sectional distribution of risk factor loadings. Intuitively, the more disperse the cross section of betas, the more information the cross section contains to estimate risk premia. Aggregating stocks into portfolios causes the information contained in individual stock betas to become more opaque and tends to shrink the cross-sectional dispersion of betas. Thus, in creating portfolios, estimates of factor loadings become more precise, but the cross-sectional dispersion of factor loadings shrinks. It is the loss of information in the cross section of beta when stocks are grouped into portfolios that

opposed to factor loadings as is done here, also results in a positive HML premium of 4.59% per annum with a maximum likelihood t-statistic of 13.6.

contributes to potentially large efficiency losses in using portfolios versus individual stocks. Simulations show the efficiency losses are magnified when portfolios are formed on estimated, rather than true, factor loadings.

In data the point estimates of the cross-sectional market risk premium using individual stocks are positive and highly significant. This is true for both a one-factor market model and the three-factor Fama and French (1993) model. For the one-factor model using all stocks, the cross-sectional market risk premium estimate of 5.24% per annum is very close to the time-series average of the market excess return, at 5.76% per annum. In contrast, the market risk premium is insignificant, and sometimes has a negative sign, when portfolios constructed on estimated factor loadings at the beginning of the period are used as base assets. Thus, using stocks or portfolios can result in very different conclusions regarding whether a particular factor carries a significant price of risk.

The most important message of our results is that using individual stocks permit more efficient tests of whether factors are priced. When just two-pass cross-sectional regression coefficients are estimated there should be no reason to create portfolios and the asset pricing tests should be run on individual stocks. The use of portfolios in cross-sectional tests should be carefully motivated and be restricted to settings where economic models apply directly to portfolios, such as industries, or portfolios should be used only in non-linear econometric tests necessitating a parsimonious number of base assets.

Appendix

A Derivation of Asymptotic Variances

We restate the inverse of the Hessian here for convenience:

$$T\begin{pmatrix} \sum_i \frac{1}{\sigma_i^2} & \sum_i \frac{\beta_i}{\sigma_i^2} & \frac{\lambda}{\sigma_i^2} \\ \sum_i \frac{\beta_i}{\sigma_i^2} & \sum_i \frac{\beta_i^2}{\sigma_i^2} & \frac{\beta_i\lambda}{\sigma_i^2} \\ \frac{\lambda}{\sigma_i^2} & \frac{\beta_i\lambda}{\sigma_i^2} & \frac{\lambda^2+\sigma_m^2}{\sigma_i^2} \end{pmatrix}. \tag{A-1}$$

To invert this we partition the matrix as:

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} = \begin{pmatrix} Q^{-1} & -Q^{-1}BD^{-1} \\ -D^{-1}CQ^{-1} & D^{-1}(I + CQ^{-1}BD^{-1}) \end{pmatrix},$$

where $Q = A - BD^{-1}C$, and

$$A = \begin{pmatrix} \sum_i \frac{1}{\sigma_i^2} & \sum_i \frac{\beta_i}{\sigma_i^2} \\ \sum_i \frac{\beta_i}{\sigma_i^2} & \sum_i \frac{\beta_i^2}{\sigma_i^2} \end{pmatrix}, \quad B = \begin{pmatrix} \frac{\lambda}{\sigma_i^2} \\ \frac{\beta_i\lambda}{\sigma_i^2} \end{pmatrix}, \quad C = \begin{pmatrix} \frac{\lambda}{\sigma_i^2} & \frac{\beta_i\lambda}{\sigma_i^2} \end{pmatrix}, \quad D = \frac{\lambda^2 + \sigma_m^2}{\sigma_i^2}.$$

In our case,

$$Q = \sum_i \begin{pmatrix} \frac{1}{\sigma_i^2} & \frac{\beta_i}{\sigma_i^2} \\ \frac{\beta_i}{\sigma_i^2} & \frac{\beta_i^2}{\sigma_i^2} \end{pmatrix} - \frac{\lambda^2}{\sigma_m^2 + \lambda^2} \sum_i \begin{pmatrix} \frac{1}{\sigma_i^2} & \frac{\beta_i}{\sigma_i^2} \\ \frac{\beta_i}{\sigma_i^2} & \frac{\beta_i^2}{\sigma_i^2} \end{pmatrix} = \frac{\sigma_m^2}{\sigma_m^2 + \lambda^2} \sum_i \begin{pmatrix} \frac{1}{\sigma_i^2} & \frac{\beta_i}{\sigma_i^2} \\ \frac{\beta_i}{\sigma_i^2} & \frac{\beta_i^2}{\sigma_i^2} \end{pmatrix}.$$

Note that we only list the beta for one stock i in the Hessian in equation (A-1), but there are N such equations. In the above equation, this yields the summation over i in the second term.

The inverse of Q is

$$Q^{-1} = \frac{\sigma_m^2 + \lambda^2}{\sigma_m^2} \frac{1}{\left(\sum_i \frac{\beta_i^2}{\sigma_i^2}\right)\left(\sum_i \frac{1}{\sigma_i^2}\right) - \left(\sum_i \frac{\beta_i}{\sigma_i^2}\right)^2} \sum_i \begin{pmatrix} \frac{\beta_i^2}{\sigma_i^2} & -\frac{\beta_i}{\sigma_i^2} \\ -\frac{\beta_i}{\sigma_i^2} & \frac{1}{\sigma_i^2} \end{pmatrix}. \tag{A-2}$$

This gives the variance of $\hat{\alpha}$ and $\hat{\lambda}$ in equations (15) and (16), and the covariance of $\hat{\alpha}$ and $\hat{\lambda}$ in equation (20).

To compute the term $D^{-1}(I + CQ^{-1}BD^{-1})$ we evaluate

$$\begin{aligned} D^{-1}CQ^{-1}BD^{-1} &= \frac{\lambda^2}{\sigma_m^2(\lambda^2+\sigma_m^2)} \frac{1}{\left(\sum_j \frac{\beta_j^2}{\sigma_j^2}\right)\left(\sum_j \frac{1}{\sigma_j^2}\right) - \left(\sum_j \frac{\beta_j}{\sigma_j^2}\right)^2} \\ &\qquad \times \begin{pmatrix} 1 & \beta_i \end{pmatrix} \begin{pmatrix} \sum_j \frac{\beta_j^2}{\sigma_j^2} & -\sum_j \frac{\beta_j}{\sigma_j^2} \\ -\sum_j \frac{\beta_j}{\sigma_j^2} & \sum_i \frac{1}{\sigma_j^2} \end{pmatrix} \begin{pmatrix} 1 \\ \beta_i \end{pmatrix} && \text{(A-3)} \\ &= \frac{\lambda^2}{\sigma_m^2(\lambda^2+\sigma_m^2)} \frac{\sum_j \frac{\beta_j^2}{\sigma_j^2} - 2\beta_i \sum_j \frac{\beta_j}{\sigma_j^2} + \beta_i^2 \sum_j \frac{\beta_j^2}{\sigma_j^2}}{\left(\sum_j \frac{\beta_j^2}{\sigma_j^2}\right)\left(\sum_j \frac{1}{\sigma_j^2}\right) - \left(\sum_j \frac{\beta_j}{\sigma_j^2}\right)^2}. && \text{(A-4)} \end{aligned}$$

Thus,

$$D^{-1} + D^{-1}CQ^{-1}BD^{-1} = \frac{\sigma_i^2}{(\lambda^2 + \sigma_m^2)} \left(1 + \frac{\lambda^2}{\sigma_i^2\sigma_m^2} \frac{\sum_j \frac{\beta_j^2}{\sigma_j^2} - 2\beta_i \sum_j \frac{\beta_j}{\sigma_j^2} + \beta_i^2 \sum_j \frac{1}{\sigma_j^2}}{\left(\sum_j \frac{\beta_j^2}{\sigma_j^2}\right)\left(\sum_j \frac{1}{\sigma_j^2}\right) - \left(\sum_j \frac{\beta_j}{\sigma_j^2}\right)^2} \right). \tag{A-5}$$

This gives the variance of $\hat{\beta}_i$ in equation (17).

To compute the covariances between $(\hat{\alpha}, \hat{\lambda})$ and $\hat{\beta}_i$, we simplify

$$
\begin{aligned}
-Q^{-1}BD^{-1} &= -\frac{\lambda}{\sigma_m^2} \frac{1}{\left(\sum_j \frac{\beta_j^2}{\sigma_j^2}\right)\left(\sum_j \frac{1}{\sigma_j^2}\right) - \left(\sum_j \frac{\beta_j}{\sigma_j^2}\right)^2} \begin{pmatrix} \sum_j \frac{\beta_j^2}{\sigma_j^2} & -\sum_j \frac{\beta_j}{\sigma_j^2} \\ -\sum_j \frac{\beta_j}{\sigma_j^2} & \sum_i \frac{1}{\sigma_j^2} \end{pmatrix} \begin{pmatrix} 1 \\ \beta_i \end{pmatrix} \\
&= \frac{1}{N} \frac{\lambda}{\sigma_m^2} \frac{1}{\mathrm{var}_c(\beta/\sigma^2) - \mathrm{cov}_c(\beta^2/\sigma^2, 1/\sigma^2)} \begin{pmatrix} -\mathrm{E}_c(\beta^2/\sigma^2) + \beta_i \mathrm{E}_c(\beta/\sigma^2) \\ \mathrm{E}_c(\beta/\sigma^2) - \beta_i \mathrm{E}_c(1/\sigma^2) \end{pmatrix} \qquad \text{(A-6)}
\end{aligned}
$$

This yields the covariances in equations (21) and (22).

B Factor Risk Premia and Characteristics

Consider the data generating process

$$
R_{it} = \alpha + \beta_i \lambda + z_i \gamma + \beta_i (R_{mt} - \mu_m) + \sigma_i \varepsilon_{it}, \qquad \text{(B-1)}
$$

where z_i is a firm-specific characteristic and ε_{it} is IID $N(0,1)$. Assume that α, σ_i, μ_m, and σ_i are known and the parameters of interest are $\Theta = (\lambda \; \gamma \; \beta_i)$. We assume the intercept term α is known just to make the computations easier. The Hessian is given by

$$
\left(\mathrm{E}\left[-\frac{\partial^2 L}{\partial \Theta \partial \Theta'}\right]\right)^{-1} = \frac{1}{T} \begin{pmatrix} \sum_i \frac{\beta_i^2}{\sigma_i^2} & \sum_i \frac{\beta_i z_i}{\sigma_i^2} & \frac{\beta_i \lambda}{\sigma_i^2} \\ \sum_i \frac{\beta_i z_i}{\sigma_i^2} & \sum_i \frac{z_i^2}{\sigma_i^2} & \frac{z_i \lambda}{\sigma_i^2} \\ \frac{\beta_i \lambda}{\sigma_i^2} & \frac{z_i \lambda}{\sigma_i^2} & \frac{\lambda^2 + \sigma_m^2}{\sigma_i^2} \end{pmatrix}^{-1}. \qquad \text{(B-2)}
$$

Using methods similar to Appendix A, we can derive $\mathrm{var}(\hat{\lambda})$ and $\mathrm{var}(\hat{\gamma})$ to be

$$
\begin{aligned}
\mathrm{var}(\hat{\lambda}) &= \frac{1}{NT} \frac{\sigma_m^2 + \lambda^2}{\sigma_m^2} \frac{\mathrm{E}_c(z^2/\sigma^2)}{\mathrm{var}_c(z\beta/\sigma^2) - \mathrm{cov}_c(\beta^2/\sigma^2, z^2/\sigma^2)} \\
\mathrm{var}(\hat{\gamma}) &= \frac{1}{NT} \frac{\sigma_m^2 + \lambda^2}{\sigma_m^2} \frac{\mathrm{E}_c(\beta^2/\sigma^2)}{\mathrm{var}_c(z\beta/\sigma^2) - \mathrm{cov}_c(\beta^2/\sigma^2, z^2/\sigma^2)}, \qquad \text{(B-3)}
\end{aligned}
$$

where we define the cross-sectional moments

$$
\begin{aligned}
\mathrm{E}_c(z^2/\sigma^2) &= \frac{1}{N} \sum_j \frac{z_j^2}{\sigma_j^2} \\
\mathrm{E}_c(\beta^2/\sigma^2) &= \frac{1}{N} \sum_j \frac{\beta_j^2}{\sigma_j^2} \\
\mathrm{var}_c(z\beta/\sigma^2) &= \left(\frac{1}{N} \sum_j \frac{z_j^2 \beta_j^2}{\sigma_j^4}\right) - \left(\frac{1}{N} \sum_j \frac{z_j \beta_j}{\sigma_j^2}\right)^2 \\
\mathrm{cov}_c(z^2/\sigma^2, \beta^2/\sigma^2) &= \left(\frac{1}{N} \sum_j \frac{z_j^2 \beta_j^2}{\sigma_j^4}\right) - \left(\frac{1}{N} \sum_j \frac{z_j^2}{\sigma_j^2}\right)\left(\frac{1}{N} \sum_j \frac{\beta_j^2}{\sigma_j^2}\right). \qquad \text{(B-4)}
\end{aligned}
$$

C The Approach of Fama and French (1992)

In the second-stage of the Fama and MacBeth (1973) procedure, excess returns, R_i, are regressed onto estimated betas, $\hat{\beta}_i$ yielding a factor coefficient of

$$
\hat{\lambda} = \frac{\mathrm{cov}(R_i, \hat{\beta}_i)}{\mathrm{var}(\hat{\beta}_i)}.
$$

In the approach of Fama and French (1992), P portfolios are first created and then the individual stock betas are assigned to be the portfolio beta to which that stock belongs, as in equation (25). The numerator of the Fama-MacBeth coefficient can be written as:

$$\begin{aligned}
\text{cov}(R_i, \hat{\beta}_i) &= \frac{1}{N}\sum_i (R_i - \bar{R})(\hat{\beta}_i - \bar{\beta}) \\
&= \frac{1}{P}\sum_p \left(\frac{1}{(N/P)} \sum_{i \in p} (R_i - \bar{R}) \right) (\hat{\beta}_p - \bar{\beta}) \\
&= \frac{1}{P}\sum_{p=1}^{P} (\hat{R}_p - \bar{R})(\hat{\beta}_p - \bar{\beta}) \\
&= \text{cov}(\hat{R}_p, \hat{\beta}_p),
\end{aligned} \tag{C-1}$$

where the first to the second line follows because of equation (25). The denominator of the estimated risk premium is

$$\begin{aligned}
\text{var}(\hat{\beta}_i) &= \frac{1}{N}\sum_i (\hat{\beta}_i - \bar{\beta})^2 \\
&= \frac{1}{P}\sum_p \frac{1}{(N/P)} \sum_{i \in p} (\hat{\beta}_i - \bar{\beta})^2 \\
&= \frac{1}{P}\sum_{p=1}^{P} (\hat{\beta}_p - \bar{\beta})^2 \\
&= \text{var}(\hat{\beta}_p),
\end{aligned} \tag{C-2}$$

where the equality in the third line comes from $\hat{\beta}_p = \hat{\beta}_i$ for all $i \in p$, with N/P stocks in portfolio p having the same value of β_p for their fitted betas. Thus, the Fama and French (1992) procedure will produce the same Fama-MacBeth (1973) coefficient as using only the information from $p = 1, \ldots, P$ portfolios.

D A Special Case when Portfolios Have the Same Efficiency

We examine a special case where certain portfolios attain the same efficiency as using all stocks. Suppose that α is known and we only need to estimate λ. The variance of $\hat{\lambda}$ using all stocks is

$$\left(\text{E}\left[-\frac{\partial^2 L}{\partial \lambda^2} \right] \right)^{-1} = \frac{1}{T \sum_i \frac{\beta_i^2}{\sigma_i^2}} = \frac{1}{NT} \frac{1}{\text{E}_c(\beta^2/\sigma^2)}$$

Suppose we have a portfolio with weight proportional to β_i/σ_i^2, thus the portfolio weight on stock i is

$$\phi_i = \frac{\frac{\beta_i}{\sigma_i^2}}{\sum_i \frac{\beta_i}{\sigma_i^2}}.$$

The beta and variance of this portfolio are

$$\beta_\phi = \frac{\sum_i \frac{\beta_i^2}{\sigma_i^2}}{\sum_i \frac{\beta_i}{\sigma_i^2}} \quad \text{and} \quad \sigma_\phi^2 = \sum_i \frac{\frac{\beta_i^2}{\sigma_i^4}}{(\sum_i \frac{\beta_i}{\sigma_i^2})^2} \sigma_i^2 = \frac{\sum_i \frac{\beta_i^2}{\sigma_i^2}}{(\sum_i \frac{\beta_i}{\sigma_i^2})^2}.$$

With this single portfolio, we can estimate λ from the time-series mean of the portfolio return as there is no cross section used. Since

$$T\frac{\beta_\phi^2}{\sigma_\phi^2} = T\sum_i \frac{\beta_i^2}{\sigma_i^2},$$

this portfolio produces the same standard error for $\hat{\lambda}$ as using all stocks together. What underlies this result is that weighting by β_i/σ_i^2 efficiently captures the same information in each cross section at time t.

By similar reasoning, in the case where λ is known and we need to estimate only α, using a single portfolio with weight proportional to $1/\sigma_i^2$ yields the same standard variance for $\hat{\alpha}$ as using all stocks together. These examples are unrealistic empirical cases because no cross sectional information is used (only one portfolio is created).

E Efficiency Results for Betas Drawn from a Normal Distribution

If beta is normally distributed with mean μ_β and standard deviation σ_β, the relevant cross-sectional moments are:

$$\begin{aligned} \mathrm{E}_c(\beta^2) &= \sigma_\beta^2 + \mu_\beta^2 \\ \mathrm{var}_c(\beta^2) &= \sigma_\beta^2. \end{aligned}$$

The P portfolios are partitioned by the points ζ_p defined in equation (27), where

$$N(\delta_p) = \frac{p}{P}, \quad p = 1, ..., P-1.$$

and we define $\delta_0 = -\infty$ and $\delta_P = +\infty$. The beta of portfolio p, β_p, is given by:

$$\beta_p = \frac{\int_{\delta_{p-1}}^{\delta_p} (\mu_\beta + \sigma_\beta \delta) e^{-\frac{\delta^2}{2}} \frac{d\delta}{\sqrt{2\pi}}}{\int_{\delta_{p-1}}^{\delta_p} e^{-\frac{\delta^2}{2}} \frac{d\delta}{\sqrt{2\pi}}} = \mu_\beta + \frac{P\sigma_\beta}{\sqrt{2\pi}} \left(e^{-\frac{\delta_{p-1}^2}{2}} - e^{-\frac{\delta_p^2}{2}} \right).$$

Therefore, the cross-sectional moments for the P portfolio betas are:

$$\begin{aligned} \mathrm{E}_c[\beta_p] &= \mu_\beta \\ \mathrm{E}_c[\beta_p^2] &= \frac{1}{P} \sum_{p=1}^{P} \left(\mu_\beta + \frac{P\sigma_\beta}{\sqrt{2\pi}} \left(e^{-\frac{\delta_{p-1}^2}{2}} - e^{-\frac{\delta_p^2}{2}} \right) \right)^2 \\ &= \mu_\beta^2 + P \frac{\sigma_\beta^2}{2\pi} \sum_{p=1}^{P} \left(e^{-\frac{\delta_{p-1}^2}{2}} - e^{-\frac{\delta_p^2}{2}} \right)^2 \\ \mathrm{var}_c[\beta_p] &= P \frac{\sigma_\beta^2}{2\pi} \sum_{p=1}^{P} \left(e^{-\frac{\delta_{p-1}^2}{2}} - e^{-\frac{\delta_p^2}{2}} \right)^2. \end{aligned} \tag{E-1}$$

The ratio of the standard variance of $\hat{\alpha}$ using the P portfolios compared to the standard variance using all stocks is:

$$\begin{aligned} \frac{\mathrm{var}_p(\hat{\alpha})}{\mathrm{var}(\hat{\alpha})} &= \frac{\mathrm{E}_c(\beta_p^2)/\mathrm{var}_c(\beta_p)}{\mathrm{E}_c(\beta^2)/\mathrm{var}_c(\beta^2)} \\ &= \frac{\frac{\mu_\beta^2}{\sigma_\beta^2} \left(P \frac{1}{2\pi} \sum_{p=1}^{P} \left(e^{-\frac{\delta_{p-1}^2}{2}} - e^{-\frac{\delta_p^2}{2}} \right)^2 \right)^{-1} + 1}{\frac{\mu_\beta^2}{\sigma_\beta^2} + 1}, \end{aligned} \tag{E-2}$$

where we use the subscript p to denote the variance of the estimator computed using the P portfolios. Similarly, we can compute

$$\begin{aligned} \frac{\mathrm{var}_p(\hat{\lambda})}{\mathrm{var}(\hat{\lambda})} &= \frac{1/\mathrm{var}_c(\beta_p)}{1/\mathrm{var}_c(\beta)} \\ &= \frac{1}{P \frac{1}{2\pi} \sum_{p=1}^{P} \left(e^{-\frac{\delta_{p-1}^2}{2}} - e^{-\frac{\delta_p^2}{2}} \right)^2}. \end{aligned} \tag{E-3}$$

As expected, as $P \to \infty$, $\text{var}_p(\hat{\alpha}) \to \text{var}(\hat{\alpha})$ and $\text{var}_p(\hat{\lambda}) \to \text{var}(\hat{\lambda})$ since as $P \to \infty$,

$$P \frac{1}{2\pi} \sum_{p=1}^{P} \left(e^{-\frac{\delta_{p-1}^2}{2}} - e^{-\frac{\delta_p^2}{2}} \right)^2 \to 1.$$

Note that

$$\frac{1}{2\pi} \left(e^{-\frac{\delta_{p-1}^2}{2}} - e^{-\frac{\delta_p^2}{2}} \right)^2 = \Big(f(\delta_p) - f(\delta_{p-1}) \Big)^2 \approx \Big(df(\delta_p) \Big)^2 = \Big(f'(\delta_p) \Big)^2 (d\delta_p)^2,$$

where $f(\cdot)$ is the probability density function of the standard normal. From Equation (27), we have

$$\frac{1}{P} = N(\delta_p) - N(\delta_{p-1}) \approx N'(\delta_p) d\delta_p = f(\delta_p) d\delta_p.$$

Combining the above two equations, we obtain

$$P \frac{1}{2\pi} \sum_{p=1}^{P} \left(e^{-\frac{\delta_{p-1}^2}{2}} - e^{-\frac{\delta_p^2}{2}} \right)^2 \approx \sum_{p=1}^{P} \frac{\Big(f'(\delta_p) \Big)^2}{f(\delta_p)} d\delta_p = \sum_{p=1}^{P} e^{-\delta_p^2/2} \delta_p^2 \frac{d\delta_p}{\sqrt{2\pi}} \to 1.$$

F Multifactor Case

From equation (29) define the $(1 + K + KN) \times 1$ parameter vector $\Theta = (\alpha, \lambda', \beta_1', ..., \beta_N')'$. Assuming λ, Σ_F, and σ_i are known, the log-likelihood is

$$L = -\sum_t \sum_i \frac{1}{2\sigma_i^2} (R_{it} - (\alpha + \beta_i'\lambda + \beta_i' F_t))^2. \tag{F-1}$$

The first derivative of the log likelihood is

$$\frac{\partial L}{\partial \Theta} = \begin{pmatrix} \sum_{i,t} \frac{R_{it} - (\alpha + \beta_i'\lambda + \beta_i' F_t)}{\sigma_i^2} \\ \sum_{i,t} \frac{R_{it} - (\alpha + \beta_i'\lambda + \beta_i' F_t)}{\sigma_i^2} \beta_i \\ \sum_t \frac{R_{1t} - (\alpha + \beta_1'\lambda + \beta_1' F_t)}{\sigma_1^2} (\lambda + F_t) \\ \vdots \\ \sum_t \frac{R_{Nt} - (\alpha + \beta_N'\lambda + \beta_N' F_t)}{\sigma_N^2} (\lambda + F_t) \end{pmatrix}.$$

The second derivative of the log-likelihood is

$$\begin{aligned} \frac{\partial^2 L}{\partial \Theta \partial \Theta'} &= -\begin{pmatrix} \sum_{i,t} \frac{1}{\sigma_i^2} & \sum_{i,t} \frac{\beta_i'}{\sigma_i^2} & \sum_t \frac{(\lambda+F_t)'}{\sigma_1^2} & \cdots & \sum_t \frac{(\lambda+F_t)'}{\sigma_N^2} \\ \sum_{i,t} \frac{\beta_i}{\sigma_i^2} & \sum_{i,t} \frac{\beta_i \beta_i'}{\sigma_i^2} & \sum_t \frac{\beta_1 (\lambda+F_t)'}{\sigma_1^2} & \cdots & \sum_t \frac{\beta_N (\lambda+F_t)'}{\sigma_N^2} \\ \sum_t \frac{(\lambda+F_t)}{\sigma_1^2} & \sum_t \frac{(\lambda+F_t)\beta_1'}{\sigma_1^2} & \sum_t \frac{(\lambda+F_t)(\lambda+F_t)'}{\sigma_1^2} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum_t \frac{(\lambda+F_t)}{\sigma_N^2} & \sum_t \frac{(\lambda+F_t)\beta_N'}{\sigma_N^2} & 0 & \cdots & \sum_t \frac{(\lambda+F_t)(\lambda+F_t)'}{\sigma_N^2} \end{pmatrix} \\ &\to \quad -T \begin{pmatrix} \sum_i \frac{1}{\sigma_i^2} & \sum_i \frac{\beta_i'}{\sigma_i^2} & \frac{\lambda'}{\sigma_1^2} & \cdots & \frac{\lambda'}{\sigma_N^2} \\ \sum_i \frac{\beta_i}{\sigma_i^2} & \sum_i \frac{\beta_i \beta_i'}{\sigma_i^2} & \frac{\beta_1 \lambda'}{\sigma_1^2} & \cdots & \frac{\beta_N \lambda'}{\sigma_N^2} \\ \frac{\lambda}{\sigma_1^2} & \frac{\lambda \beta_1'}{\sigma_1^2} & \frac{\lambda\lambda' + \Sigma_F}{\sigma_1^2} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{\lambda}{\sigma_N^2} & \frac{\lambda \beta_N'}{\sigma_N^2} & 0 & \cdots & \frac{\lambda\lambda' + \Sigma_F}{\sigma_N^2} \end{pmatrix}. \end{aligned} \tag{F-2}$$

We define the following partitions:

$$-\frac{\partial^2 L}{\partial\Theta\partial\Theta'} = T\begin{pmatrix} A & B \\ B' & D \end{pmatrix} \tag{F-3}$$

for

$$A = \begin{pmatrix} \sum_i \frac{1}{\sigma_i^2} & \sum_i \frac{\beta_i'}{\sigma_i^2} \\ \sum_i \frac{\beta_i}{\sigma_i^2} & \sum_i \frac{\beta_i\beta_i'}{\sigma_i^2} \end{pmatrix}, B = \begin{pmatrix} \frac{\lambda'}{\sigma_1^2} & \cdots & \frac{\lambda'}{\sigma_N^2} \\ \frac{\beta_1\lambda'}{\sigma_1^2} & \cdots & \frac{\beta_N\lambda'}{\sigma_N^2} \end{pmatrix}, D = \begin{pmatrix} \frac{\lambda\lambda'+\Sigma_F}{\sigma_1^2} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{\lambda\lambda'+\Sigma_F}{\sigma_N^2} \end{pmatrix}.$$

Then,

$$\left(-\frac{\partial^2 L}{\partial\Theta\partial\Theta'}\right)^{-1} = \frac{1}{T}\begin{pmatrix} Q^{-1} & -Q^{-1}BD^{-1} \\ -D^{-1}B'Q^{-1} & D^{-1}+D^{-1}B'Q^{-1}BD^{-1} \end{pmatrix},$$

where $Q = A - BD^{-1}B'$.

Note that

$$\begin{aligned} BD^{-1}B' &= \begin{pmatrix} \frac{\lambda'}{\sigma_1^2} & \cdots & \frac{\lambda'}{\sigma_N^2} \\ \frac{\beta_1\lambda'}{\sigma_1^2} & \cdots & \frac{\beta_N\lambda'}{\sigma_N^2} \end{pmatrix} \begin{pmatrix} \frac{\lambda\lambda'+\Sigma_F}{\sigma_1^2} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{\lambda\lambda'+\Sigma_F}{\sigma_N^2} \end{pmatrix}^{-1} \begin{pmatrix} \frac{\lambda}{\sigma_1^2} & \frac{\lambda\beta_1'}{\sigma_1^2} \\ \vdots & \vdots \\ \frac{\lambda}{\sigma_N^2} & \frac{\lambda\beta_N'}{\sigma_N^2} \end{pmatrix} \\ &= \sum_i \frac{\lambda'(\lambda\lambda'+\Sigma_F)^{-1}\lambda}{\sigma_i^2}\begin{pmatrix} 1 & \beta_i' \\ \beta_i & \beta_i\beta_i' \end{pmatrix}, \end{aligned}$$

so

$$Q = (1-\lambda'(\lambda\lambda'+\Sigma_F)^{-1}\lambda)A = (1-\lambda'(\lambda\lambda'+\Sigma_F)^{-1}\lambda)\begin{pmatrix} \sum_i \frac{1}{\sigma_i^2} & \sum_i \frac{\beta_i'}{\sigma_i^2} \\ \sum_i \frac{\beta_i}{\sigma_i^2} & \sum_i \frac{\beta_i\beta_i'}{\sigma_i^2} \end{pmatrix}. \tag{F-4}$$

We invert Q to obtain

$$\begin{aligned} Q^{-1} &= \frac{1}{(1-\lambda'(\lambda\lambda'+\Sigma_F)^{-1}\lambda)} \\ &\times \begin{pmatrix} \left(\left(\sum_i \frac{1}{\sigma_i^2}\right) - \left(\sum_i \frac{\beta_i'}{\sigma_i^2}\right)\left(\sum_i \frac{\beta_i\beta_i'}{\sigma_i^2}\right)^{-1}\left(\sum_i \frac{\beta_i}{\sigma_i^2}\right)\right)^{-1} & \frac{-\left(\sum_i \frac{\beta_i'}{\sigma_i^2}\right)\left(\sum_i \frac{\beta_i\beta_i'}{\sigma_i^2}\right)^{-1}}{\left(\left(\sum_i \frac{1}{\sigma_i^2}\right)-\left(\sum_i \frac{\beta_i'}{\sigma_i^2}\right)\left(\sum_i \frac{\beta_i\beta_i'}{\sigma_i^2}\right)^{-1}\left(\sum_i \frac{\beta_i}{\sigma_i^2}\right)\right)} \\ \frac{-\left(\sum_i \frac{\beta_i'}{\sigma_i^2}\right)\left(\sum_i \frac{\beta_i\beta_i'}{\sigma_i^2}\right)^{-1}}{\left(\left(\sum_i \frac{1}{\sigma_i^2}\right)-\left(\sum_i \frac{\beta_i'}{\sigma_i^2}\right)\left(\sum_i \frac{\beta_i\beta_i'}{\sigma_i^2}\right)^{-1}\left(\sum_i \frac{\beta_i}{\sigma_i^2}\right)\right)} & \left(\left(\sum_i \frac{\beta_i\beta_i'}{\sigma_i^2}\right) - \left(\sum_i \frac{\beta_i}{\sigma_i^2}\right)\left(\sum_i \frac{1}{\sigma_i^2}\right)^{-1}\left(\sum_i \frac{\beta_i'}{\sigma_i^2}\right)\right)^{-1} \end{pmatrix}. \end{aligned}$$

Thus,

$$\mathrm{var}(\hat{\alpha}) = \frac{1}{T}\frac{1}{(1-\lambda'(\lambda\lambda'+\Sigma_F)^{-1}\lambda)}\left(\left(\sum_i \frac{1}{\sigma_i^2}\right) - \left(\sum_i \frac{\beta_i'}{\sigma_i^2}\right)\left(\sum_i \frac{\beta_i\beta_i'}{\sigma_i^2}\right)^{-1}\left(\sum_i \frac{\beta_i}{\sigma_i^2}\right)\right)^{-1} \tag{F-5}$$

$$\mathrm{var}(\hat{\lambda}) = \frac{1}{T}\frac{1}{(1-\lambda'(\lambda\lambda'+\Sigma_F)^{-1}\lambda)}\left(\left(\sum_i \frac{\beta_i\beta_i'}{\sigma_i^2}\right) - \left(\sum_i \frac{\beta_i}{\sigma_i^2}\right)\left(\sum_i \frac{1}{\sigma_i^2}\right)^{-1}\left(\sum_i \frac{\beta_i'}{\sigma_i^2}\right)\right)^{-1}. \tag{F-6}$$

Defining the cross-sectional moments as in equation (32) leads to equations (30) and (31). In order for var($\hat{\alpha}$)or var($\hat{\lambda}$) to be defined, the number of linearly independent β_i's must be strictly greater than K. This implies the number of individual stocks must be at least K.

References

[1] Ang, A., and J. Chen, 2007, "CAPM Over the Long-Run: 1926-2001," *Journal of Empirical Finance*, 14, 1-40.

[2] Ang, A., J. Chen, and Y. Xing, 2006, "Downside Risk," *Review of Financial Studies*, 19, 1191-1239.

[3] Avramov, D., and T. Chordia, 2006, "Asset Pricing Models and Financial Market Anomalies," *Review of Financial Studies*, 19, 1001-1040.

[4] Bekaert, G., R. J. Hodrick, and X. Zhang, 2008, "Is there a Trend in Idiosyncratic Volatility?" working paper, Cornell University.

[5] Berk, J. B., 2000, "Sorting Out Sorts," *Journal of Finance*, 55, 407-427.

[6] Black, F., 1972, "Capital Market Equilibrium with Restricted Borrowing," *Journal of Business*, 45, 444-455.

[7] Black, F., M. Jensen, and M. Scholes, 1972, "The Capital Asset Pricing Model: Some Empirical Tests," in Jensen, M., ed., *Studies in the Theory of Capital Markets*, Praeger, New York, pp79-121.

[8] Blume, M. E., 1973, "Portfolio Theory: A Step Toward Its Practical Application," *Journal of Business*, 43, 152-73.

[9] Brennan, M. J., T. Chordia, and A. Subrahmanyam, 1998, "Alternative Factor Specifications, Security Characteristics, and the Cross-Section of Expected Stock Returns," *Journal of Financial Economics*, 49, 345-373.

[10] Burnside, A. C., 2007, "Empirical Asset Pricing and Statistical Power in the Presence of Weak Risk Factors," NBER working paper 13357.

[11] Campbell, J. Y., M. Lettau, B. G. Malkiel, and Y. Xu, 2001, "Have Individual Stocks Become More Volatile? An Empirical Exploration of Idiosyncratic Risk," *Journal of Finance*, 56, 1-43.

[12] Chen, R., and R. Kan, 2004, "Finite Sample Analysis of Two-Pass Cross-Sectional Regressions," working paper, University of Toronto.

[13] Cochrane, J. H., 2001, *Asset Pricing*, Princeton University Press, NJ.

[14] Connor, G., 1984, "A Unified Beta Pricing Theory," *Journal of Economic Theory*, 34, 13-39.

[15] Daniel, K., and S. Titman, 1997, "Evidence on the Characteristics of Cross Sectional Variation in Stock Returns," *Journal of Finance*, 52, 1-34.

[16] Douglas, G. W., 1969, "Risk in the Equity Markets: An Empirical Appraisal of Market Efficiency," *Yale Economic Essays*, 9, 3-45.

[17] Fama, E. F., and K. R. French, 1992, "The Cross-Section of Expected Stock Returns," *Journal of Finance*, 47, 2, 427-465.

[18] Fama, E. F., and J. D. MacBeth, 1973, "Risk, Return, and Equilibrium: Empirical Tests," *Journal of Political Economy*, 71, 607-636.

[19] Gibbons, M. R., 1982, "Multivariate Tests of Financial Models," *Journal of Financial Economics*, 10, 3-27.

[20] Gibbons, M. R., S. Ross, and J. Shanken, 1989, "A Test of the Efficiency of a Given Portfolio," *Econometrica*, 57, 1121-1152.

[21] Grauer, R. R., and J. A. Janmaat, 2004, "The Unintended Consequences of Grouping in Tests of Asset Pricing Models," *Journal of Banking and Finance*, 28, 2889-2914.

[22] Grauer, R. R., and J. A. Janmaat, 2008, "Cross-Sectional Tests of the CAPM and Fama-French Three-Factor Model," working paper, UBC.

[23] Hou, K., and R. L. Kimmel, 2006, "On Estimation of Risk Premia in Linear Factor Models," working paper, Ohio State University.

[24] Huberman, G., and S. Kandel, 1987, "Mean-Variance Spanning," *Journal of Finance*, 42, 873-888.

[25] Jagannathan, R., G. Skoulakis, and Z. Wang, 2002, "The Analysis of the Cross Section of Security Returns," forthcoming Aït-Sahalia, Y., and Hansen, L. P., eds., *Handbook of Financial Econometrics*, Elsevier.

[26] Jagannathan, R., and Z. Wang, 1996, "The Conditional CAPM and the Cross-Section of Expected Returns," *Journal of Finance*, 51, 3-53.

[27] Jagannathan, R., and Z. Wang, 1998, "An Asymptotic Theory for Estimating Beta-Pricing Models using Cross-Sectional Regression," *Journal of Finance*, 53, 12851309.

[28] Jacquier, E., N. G. Polson, and P. E. Rossi, 1994, "Bayesian Analysis of Stochastic Volatility Models," *Journal of Business and Economic Statistics*, 12, 371-389.

[29] Jobson, D., and R. Korkie, 1982, "Potential Performance and Tests of Portfolio Efficiency," *Journal of Financial Economics*, 10, 433-436.

[30] Kan, R., and C. Zhang, 1999, "Two-Pass Tests of Asset Pricing Models with Useless Factors," *Journal of Finance*, 54, 203-35.

[31] Kan, R., and C. Robotti, 2006, "Specification Tests of Asset Pricing Models Using Excess Returns," Federal Reserve Bank of Atlanta Working Paper 2006-10.

[32] Kandel, S., and R. Stambaugh, 1995, "Portfolio Inefficiency and the Cross-Section of Expected Returns," *Journal of Finance*, 50, 157184.

[33] Lettau, M., and S. Ludvigson, 2001, "Consumption, Aggregate Wealth, and Expected Stock Returns," *Journal of Finance*, 56, 815-849.

[34] Lewellen, J., S. Nagel, and J. Shanken, 2007, "A Skeptical Appraisal of Asset-Pricing Tests," working paper, Stanford University.

[35] Lewellen, J., and S. Nagel, 2006, "The Conditional CAPM Does Not Explain Asset-Pricing Anomalies," *Journal of Financial Economics*, 82, 289-314.

[36] Litzenberger, R. H., and K. Ramaswamy, 1979, "The Effects of Personal Taxes and Dividends on Capital Asset Prices: Theory and Empirical Evidence," *Journal of Financial Economics*, 7, 163-195.

[37] Lo, A., W., and A. C. MacKinlay, 1990, "Data-Snooping Biases in Tests of Financial Asset Pricing Models," *Review of Financial Studies*, 3, 431-468.

[38] MacKinlay, A. C., 1987, "On Multivariate Tests of the CAPM," *Journal of Financial Economics*, 18, 341-372.

[39] Meng, J. G., G. Hu, and J. Bai, 2007, "A Simple Method for Estimating Betas when Factors are Measured with Error," working paper, NYU.

[40] Merton, R. C., 1980, "On Estimating the Expected Return on the Market," *Journal of Financial Economics*, 8, 323-361.

[41] Pástor, L., and R. F. Stambaugh, 2003, "Liquidity Risk and Expected Stock Returns," *Journal of Political Economy*, 111, 642-685.

[42] Roll, R., and S. A. Ross, 1994, "On the Cross-Sectional Relation between Expected Returns and Betas," *Journal of Finance*, 49, 101-121.

[43] Shanken, J., 1985, "Multivariate Tests of the Zero-Beta CAPM," *Journal of Financial Economics*, 14, 327-348.

[44] Shanken, J., 1992, "On the Estimation of Beta-Pricing Models," *Review of Financial Studies*, 5, 1-33.

[45] Shanken, J., and G. Zhou, 2007, "Estimating and Testing Beta Pricing Models: Alternative Methods and Their Performance in Simulations," *Journal of Financial Economics*, 84, 40-86.

[46] Velu, R., and G. Zhou, 1999, "Testing Multi-Beta Asset Pricing Models," *Journal of Empirical Finance*, 6, 219-241.

[47] Zhou, G., 1991, "Small Sample Tests of Portfolio Efficiency," *Journal of Financial Economics*, 30, 165-191.

Table 1: Variance Ratio Efficiency Losses in Monte Carlo Simulations

	Number of Portfolios P							
	5	10	25	50	100	250	1000	2500
Panel A: Sorting on True Betas								
Alpha Efficiency Variance Ratios $\text{var}_p(\hat{\alpha})/\text{var}(\hat{\alpha})$								
Mean	2.99	2.99	2.97	2.94	2.89	2.74	2.23	1.60
Median	2.96	2.96	2.96	2.92	2.87	2.73	2.23	1.60
Stdev	0.17	0.17	0.16	0.16	0.16	0.14	0.10	0.06
Lambda Efficiency Variance Ratios $\text{var}_p(\hat{\lambda})/\text{var}(\hat{\lambda})$								
Mean	3.10	3.07	3.02	2.97	2.90	2.75	2.23	1.60
Median	3.09	3.05	3.00	2.95	2.89	2.74	2.23	1.60
Stdev	0.16	0.16	0.15	0.15	0.15	0.13	0.10	0.06
Panel B: Sorting on Estimated Betas								
Alpha Efficiency Variance Ratios $\text{var}_p(\hat{\alpha})/\text{var}(\hat{\alpha})$								
Mean	5.09	5.55	5.78	5.74	5.53	4.95	3.24	1.87
Median	5.06	5.52	5.76	5.71	5.51	4.91	3.22	1.86
Stdev	0.49	0.57	0.59	0.57	0.53	0.47	0.26	0.10
Lambda Efficiency Variance Ratios $\text{var}_p(\hat{\lambda})/\text{var}(\hat{\lambda})$								
Mean	4.61	4.96	5.14	5.11	4.95	4.49	3.07	1.83
Median	4.57	4.92	5.11	5.08	4.93	4.47	3.04	1.83
Stdev	0.39	0.45	0.47	0.46	0.43	0.39	0.23	0.10

The table reports the efficiency loss variance ratios $\text{var}_p(\hat{\theta})/\text{var}(\hat{\theta})$ for $\theta = \alpha$ or λ where $\text{var}_p(\hat{\theta})$ is computed using P portfolios and $\text{var}(\theta)$ is computed using all stocks. We simulate 10,000 small samples of $T = 60$ months with $N = 5,000$ stocks using the model in equations (33) and (34). Panel A sorts stocks by true betas in each small sample and Panel B sorts stocks by estimated betas. Betas are estimated in each small sample by regular OLS, but the standard variances are computed using the true cross-sectional betas and idiosyncratic volatilities. All the portfolios are formed equally weighting stocks at the end of the period.

Table 2: Summary Statistics of Betas and Idiosyncratic Volatilities

	Means			Stdev			Correlations		
	$\hat{\beta}$	$\hat{\sigma}$	$\ln\hat{\sigma}$	$\hat{\beta}$	$\hat{\sigma}$	$\ln\hat{\sigma}$	$(\hat{\beta},\hat{\sigma})$	$(\hat{\beta},\ln\hat{\sigma})$	No Obs
1960-1965	1.192	0.278	-1.395	0.575	0.153	0.460	0.279	0.354	1434
1965-1970	1.342	0.350	-1.139	0.542	0.151	0.423	0.553	0.610	1821
1970-1975	1.316	0.399	-0.997	0.548	0.164	0.398	0.570	0.559	2210
1975-1980	1.276	0.338	-1.183	0.548	0.160	0.438	0.562	0.630	2054
1980-1985	1.098	0.331	-1.188	0.534	0.139	0.403	0.421	0.457	1943
1985-1990	1.057	0.381	-1.075	0.463	0.190	0.472	0.287	0.365	3670
1990-1995	0.984	0.437	-1.007	0.918	0.281	0.603	0.163	0.227	4935
1995-2000	0.935	0.563	-0.772	0.774	0.382	0.647	0.589	0.605	5723
2000-2005	1.114	0.438	-1.039	1.002	0.301	0.670	0.597	0.600	5306
Overall	1.093	0.425	-1.026	0.765	0.278	0.580	0.390	0.430	29096

The table reports the summary statistics of estimated betas ($\hat{\beta}$) and idiosyncratic volatility ($\hat{\sigma}$) over each five year sample and over the entire sample. We estimate betas and idiosyncratic volatility in each five-year non-overlapping period using time-series regressions of monthly excess stock returns onto a constant and monthly excess market returns. The idiosyncratic stock volatilities are annualized by multiplying by $\sqrt{12}$. The last column reports the number of stock observations.

Table 3: Estimates of a One-Factor Model

			Max Lik		Pooled		Shanken		$\hat{\beta}$ Cross Section		
Num Ports P		Estimate (%)	SE	|t-stat|	SE	|t-stat|	SE	|t-stat|	$\sigma_c(\hat{\beta})$	5%	95%
Panel A: All Stocks											
	$\hat{\alpha}$	6.14	0.11	56.5	0.29	21.0	0.42	14.5	0.77	0.05	2.42
	$\hat{\lambda}$	5.24	0.12	44.2	0.26	20.2	0.79	6.60			
Panel B: "Ex-Post" Portfolios											
5	$\hat{\alpha}$	5.20	4.75	1.09	1.75	2.98	3.61	1.44	0.65	0.06	2.15
	$\hat{\lambda}$	4.88	4.37	1.12	1.82	2.68	3.31	1.47			
10	$\hat{\alpha}$	5.08	3.29	1.54	1.73	2.94	2.80	1.81	0.67	0.21	2.37
	$\hat{\lambda}$	4.99	3.04	1.64	1.71	2.92	2.59	1.92			
25	$\hat{\alpha}$	4.99	2.04	2.45	1.56	3.20	1.96	2.55	0.69	0.17	2.33
	$\hat{\lambda}$	5.06	1.90	2.67	1.46	3.48	1.85	2.74			
50	$\hat{\alpha}$	4.99	1.42	3.51	1.34	3.71	1.53	3.25	0.69	0.17	2.30
	$\hat{\lambda}$	5.07	1.33	3.82	1.22	4.15	1.51	3.35			
100	$\hat{\alpha}$	4.98	0.99	5.02	1.11	4.47	1.21	4.12	0.69	0.15	2.32
	$\hat{\lambda}$	5.07	0.93	5.45	1.00	5.06	1.26	4.02			
Panel C: "Ex-Ante" Portfolios											
5	$\hat{\alpha}$	11.0	1.88	5.84	1.96	5.61	3.57	3.08	0.36	0.55	1.70
	$\hat{\lambda}$	-0.17	1.85	0.09	1.67	0.10	3.58	0.05			
10	$\hat{\alpha}$	10.9	1.38	7.94	1.26	8.65	2.56	4.28	0.37	0.52	1.67
	$\hat{\lambda}$	-0.11	1.34	0.08	1.06	0.11	2.62	0.04			
25	$\hat{\alpha}$	10.9	0.91	12.0	0.78	13.9	1.61	6.74	0.37	0.50	1.71
	$\hat{\lambda}$	-0.06	0.88	0.06	0.64	0.09	1.73	0.03			
50	$\hat{\alpha}$	10.7	0.68	15.6	0.67	15.9	1.16	9.18	0.37	0.49	1.71
	$\hat{\lambda}$	0.11	0.66	0.17	0.55	0.20	1.33	0.08			
100	$\hat{\alpha}$	10.4	0.53	19.5	0.56	18.6	0.86	12.1	0.38	0.50	1.72
	$\hat{\lambda}$	0.34	0.51	0.65	0.47	0.71	1.09	0.31			

Note to Table 3

The point estimates of α and λ in equation (1) are reported over all stocks (Panel A) and various portfolio sortings (Panels B and C). The betas are estimated by running a first-pass OLS regression of monthly excess stock returns onto monthly excess market returns over non-overlapping five-year samples beginning in January 1961 and ending in December 2005. All of these stock returns in each five-year period are stacked and treated as one panel. We use a second-pass cross-sectional regression to compute $\hat{\alpha}$ and $\hat{\lambda}$. Using these point estimates we compute the various standard errors (SE) and absolute values of t-statistics (|t-stat|). We compute the maximum likelihood standard errors (equations (15) and (16)) in the columns labeled "Max Lik." The columns labeled "Pooled" report robust pooled standard errors where the clustering is done at the firm or portfolio level in each five-year period. The columns labeled "Shanken" report Shanken (1992) standard errors. The three last columns labeled "$\hat{\beta}$ Cross Section" report various statistics of the cross-sectional beta distribution: the cross-sectional standard deviation, $\sigma_c(\hat{\beta})$, and the beta values corresponding to the 5%- and 95%-tiles of the cross-sectional distribution of beta. In Panel B we form "ex-post" portfolios, which are formed in each five-year period by grouping stocks into equally-weighted P portfolios based on realized estimated betas over those five years. In Panel C we form "ex-ante" portfolios, which are formed by grouping stocks into portfolios at the beginning of each calendar year ranking on the estimated market beta over the previous five years. Equally-weighted portfolios are created and the portfolios are held for twelve months to produce monthly portfolio returns. The portfolios are rebalanced annually at the beginning of each calendar year. The first estimation period is January 1956 to December 1960 to produce monthly returns for the calendar year 1961 and the last estimation period is January 2000 to December 2004 to produce monthly returns for 2005. Thus, the sample period is exactly the same as using all stocks and the ex-post portfolios. After the ex-ante portfolios are created, we follow the same procedure as Panels A and B to compute realized OLS market betas in each non-overlapping five-year period and then estimate a second-pass cross-sectional regression. In both Panels B and C, the second-pass cross-sectional regression is run only on the P portfolio test assets. All estimates $\hat{\alpha}$ and $\hat{\lambda}$ are annualized by multiplying the monthly estimates by 12.

Table 4: Cross-Sectional Fama-French (1993) Factor Loadings

Panel A: Means and Correlations using All Stocks

		$\hat{\beta}_{MKT}$	$\hat{\beta}_{SMB}$	$\hat{\beta}_{HML}$
Mean		1.003	0.805	0.258
Correlations				
	β_{MKT}	1.000	0.011	0.195
	β_{SMB}		1.000	0.111
	β_{HML}			1.000

Panel B: Cross-Sectional Dispersion

		$\sigma_c(\hat{\beta})$	5%	95%
All Stocks	$\hat{\beta}_{MKT}$	0.716	0.013	2.183
	$\hat{\beta}_{SMB}$	1.059	-0.488	2.715
	$\hat{\beta}_{HML}$	1.086	-1.548	1.811
"Ex-Post" Portfolios				
$3 \times 3 \times 3$	$\hat{\beta}_{MKT}$	0.556	0.178	1.911
	$\hat{\beta}_{SMB}$	0.898	-0.256	2.426
	$\hat{\beta}_{HML}$	0.868	-1.039	1.674
$5 \times 5 \times 5$	$\hat{\beta}_{MKT}$	0.599	0.046	2.077
	$\hat{\beta}_{SMB}$	0.969	-0.430	2.555
	$\hat{\beta}_{HML}$	0.946	-1.305	1.845
"Ex-Ante" Portfolios				
$3 \times 3 \times 3$	$\hat{\beta}_{MKT}$	0.191	0.645	1.286
	$\hat{\beta}_{SMB}$	0.539	0.022	1.809
	$\hat{\beta}_{HML}$	0.328	-0.158	0.818
$5 \times 5 \times 5$	$\hat{\beta}_{MKT}$	0.225	0.619	1.353
	$\hat{\beta}_{SMB}$	0.586	-0.010	1.880
	$\hat{\beta}_{HML}$	0.390	-0.288	0.930

Note to Table 4

The table reports cross-sectional summary statistics of estimated Fama-French (1993) factor loadings, $\hat{\beta}_{MKT}$, $\hat{\beta}_{SMB}$, and $\hat{\beta}_{HML}$. The factor loadings are estimated by running a multivariate OLS regression of monthly excess stock returns onto the monthly Fama-French (1993) factors (MKT, SMB, and HML) over non-overlapping five-year samples beginning in January 1961 and ending in December 2005. All of the factor loadings in each five-year period are stacked and treated as one panel. Panel A reports cross-sectional means and correlations over all stocks. The total number of observations is exactly the same as the breakdown reported in Table 2 at 29,096. Panel B reports cross-sectional dispersion statistics: the cross-sectional standard deviation of the estimated factor loadings, $\sigma_c(\hat{\beta})$, and and the estimated factor loadings corresponding to the 5%- and 95%-tiles of the cross-sectional distribution. The "ex-post" portfolios are formed in each five-year period by grouping stocks into equally-weighted P portfolios based on realized estimated factor loadings over those five years. We form $n \times n \times n$ portfolios using the intersections of independent sorts of n portfolios ranked on each of the Fama-French factor loadings at the end of each five-year period. The "ex-ante" portfolios are formed by grouping stocks into portfolios at the beginning of each calendar year ranking on the estimated factor loadings over the previous five years. Equally-weighted portfolios are created and the portfolios are held for twelve months to produce monthly portfolio returns. The portfolios are rebalanced annually at the beginning of each calendar year. The first estimation period is January 1956 to December 1960 to produce monthly returns for the calendar year 1961 and the last estimation period is January 2000 to December 2004 to produce monthly returns for 2005. Thus, the sample period is exactly the same as using all stocks and the ex-post portfolios. After the ex-ante portfolios are created, we follow the same procedure as the all stocks and ex-post portfolios to compute realized OLS market betas in each non-overlapping five-year period.

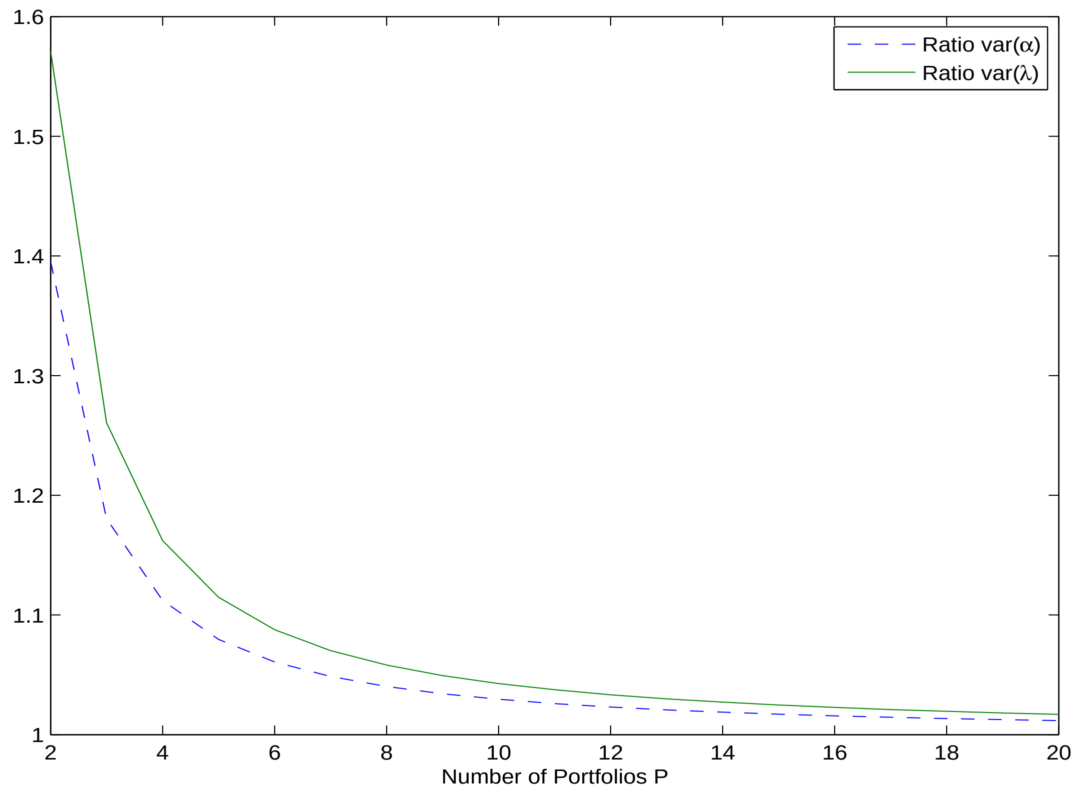
Table 5: Estimates of the Fama-French (1993) Model

Num Ports P		Estimate (%)	Max Lik SE	Max Lik |t-stat|	Pooled SE	Pooled |t-stat|	Shanken SE	Shanken |t-stat|
Panel A: All Stocks								
	$\hat{\alpha}$	5.86	0.12	50.4	0.34	17.4	0.47	12.5
	$\hat{\lambda}_{MKT}$	3.50	0.13	27.7	0.30	11.6	0.82	4.28
	$\hat{\lambda}_{SMB}$	3.90	0.10	39.8	0.22	17.7	0.95	4.09
	$\hat{\lambda}_{HML}$	-2.48	0.11	23.5	0.22	11.3	0.65	3.81
Panel B: "Ex-Post" Portfolios								
$3 \times 3 \times 3$	$\hat{\alpha}$	6.26	2.13	2.95	0.96	6.54	1.75	3.58
	$\hat{\lambda}_{MKT}$	3.25	2.04	1.59	0.99	3.28	1.77	1.83
	$\hat{\lambda}_{SMB}$	2.09	1.30	1.60	0.45	4.68	1.39	1.51
	$\hat{\lambda}_{HML}$	-2.07	1.42	1.46	0.40	5.19	1.33	1.56
$5 \times 5 \times 5$	$\hat{\alpha}$	6.62	0.94	7.03	0.77	8.59	0.98	6.78
	$\hat{\lambda}_{MKT}$	3.26	0.90	3.62	0.70	4.68	1.13	2.90
	$\hat{\lambda}_{SMB}$	1.73	0.58	2.97	0.36	4.84	1.06	1.63
	$\hat{\lambda}_{HML}$	-2.15	0.63	3.40	0.38	5.71	0.85	2.52
Panel C: "Ex-Ante" Portfolios								
$3 \times 3 \times 3$	$\hat{\alpha}$	7.23	0.86	8.38	2.09	3.45	3.32	2.18
	$\hat{\lambda}_{MKT}$	-0.52	0.91	0.57	2.12	0.25	3.35	0.16
	$\hat{\lambda}_{SMB}$	4.04	0.43	9.43	0.91	4.46	1.85	2.18
	$\hat{\lambda}_{HML}$	2.82	0.60	4.73	1.16	2.42	2.41	1.17
$5 \times 5 \times 5$	$\hat{\alpha}$	6.60	0.54	12.2	1.18	5.59	1.50	4.40
	$\hat{\lambda}_{MKT}$	0.60	0.57	1.05	1.20	0.50	1.66	0.36
	$\hat{\lambda}_{SMB}$	3.59	0.29	12.2	0.57	6.29	1.19	3.01
	$\hat{\lambda}_{HML}$	2.02	0.39	5.25	0.74	2.75	1.14	1.77

Note to Table 5

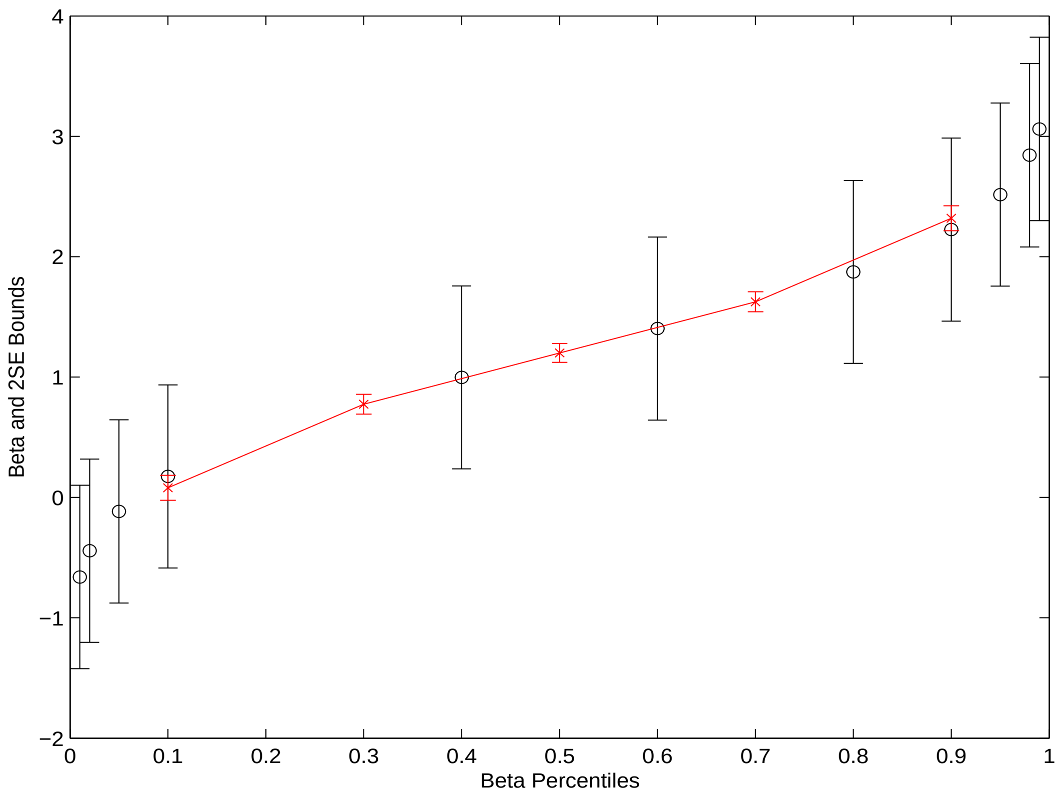
The point estimates $\hat{\alpha}$, $\hat{\lambda}_{MKT}$, $\hat{\lambda}_{SMB}$, and $\hat{\lambda}_{HML}$ in equation (35) are reported over all stocks (Panel A) and various portfolio sortings (Panels B and C). The betas are estimated by running a first-pass multivariate OLS regression of monthly excess stock returns onto the monthly Fama-French (1993) factors (MKT, SMB, and HML) over non-overlapping five-year samples beginning in January 1961 and ending in December 2005. All of these stock returns in each five-year period are stacked and treated as one panel. We use a second-pass cross-sectional regression to compute the cross-sectional coefficients. Using these point estimates we compute the various standard errors (SE) and absolute values of t-statistics (|t-stat|). We compute the maximum likelihood standard errors (equations (30) and (31)) in the columns labeled "Max Lik." The columns labeled "Pooled" report robust pooled standard errors where the clustering is done at the firm or portfolio level in each five-year period. The columns labeled "Shanken" report Shanken (1992) standard errors. In Panel B we form "ex-post" portfolios, which are formed in each five-year period by grouping stocks into equally-weighted P portfolios based on realized estimated factor loadings over those five years. We form $n \times n \times n$ portfolios using the intersections of independent sorts of n portfolios ranked on each of the Fama-French factor loadings at the end of each five-year period. In Panel C we form "ex-ante" portfolios, which are formed by grouping stocks into portfolios at the beginning of each calendar year ranking on the estimated factor loadings over the previous five years. Equally-weighted portfolios are created and the portfolios are held for twelve months to produce monthly portfolio returns. The portfolios are rebalanced annually at the beginning of each calendar year. The first estimation period is January 1956 to December 1960 to produce monthly returns for the calendar year 1961 and the last estimation period is January 2000 to December 2004 to produce monthly returns for 2005. Thus, the sample period is exactly the same as using all stocks and the ex-post portfolios. After the ex-ante portfolios are created, we follow the same procedure as Panels A and B to compute realized OLS factor loadings in each non-overlapping five-year period and then estimate a second-pass cross-sectional regression. In both Panels B and C, the second-pass cross-sectional regression is run only on the P portfolio test assets. All estimates are annualized by multiplying the monthly estimates by 12.

Figure 1: Asymptotic Variance Ratios of $\hat{\alpha}$ and $\hat{\lambda}$ using Portfolios versus All Stocks



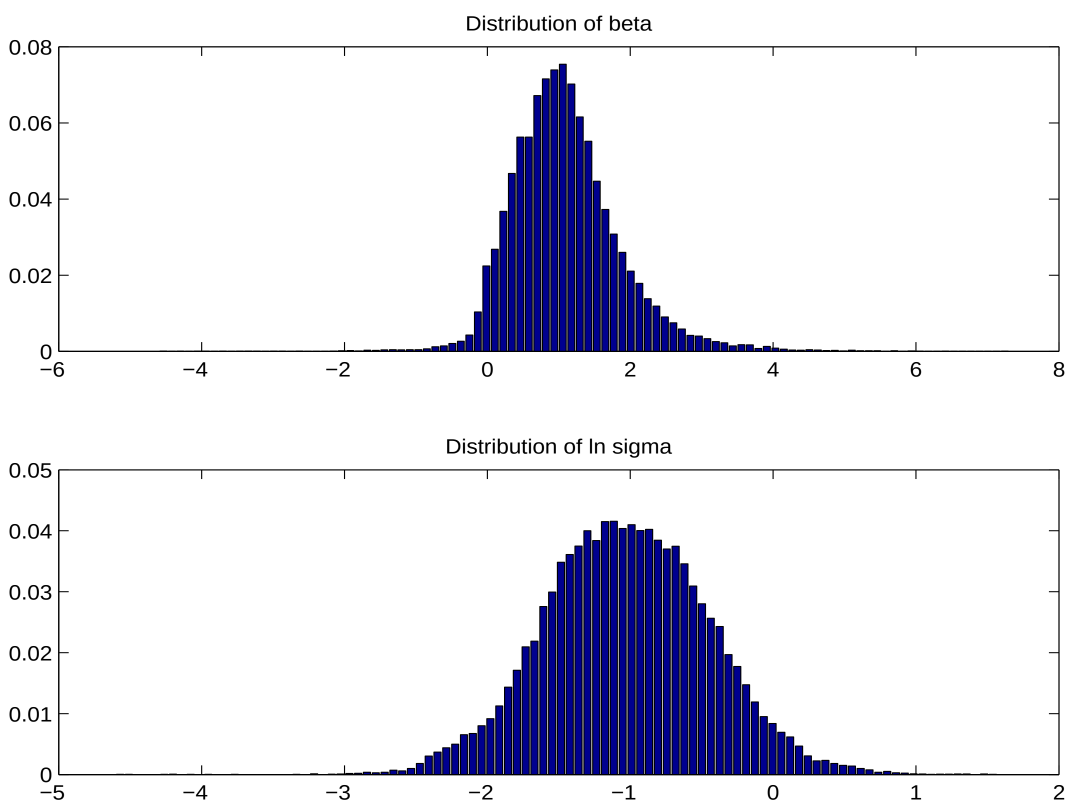
We graph the ratio of the asymptotic variance of $\hat{\alpha}$ and $\hat{\lambda}$ computed using portfolios to using all stocks, that is $\text{var}_p(\hat{\theta})/\text{var}(\hat{\theta})$, where $\theta = \alpha$ or λ, $\text{var}_p(\hat{\theta})$ is the variance computed using P portfolios and $\text{var}(\hat{\theta})$ is the variance computed using all stocks. We assume a normal distribution for beta with mean $\mu_\beta = 1.2$ and standard deviation $\sigma_\beta = 0.8$. The formulas for the variance ratios are given in Appendix E.

Figure 2: Standard Errors for $\hat{\beta}$ Using All Stocks or Five Portfolios



We assume that beta is drawn from a normal distribution with mean $\mu_\beta = 1.2$ and standard deviation $\sigma_\beta = 0.8$ and idiosyncratic volatility across stocks is constant at $\sigma_i = \sigma = 0.5/\sqrt{12}$. We assume a sample of size $T = 60$ months with $N = 1000$ stocks. We graph two standard error bars of $\hat{\beta}$ for the various percentiles of the true distribution marked in circles for percentiles 0.01, 0.02, 0.05, 0.1, 0.4, 0.6, 0.8, 0.9, 0.95, 0.98, and 0.99. The standard error bands for the portfolio betas for $P = 5$ portfolios are marked with crosses and connected by the line. These are graphed at the percentiles 0.1, 0.3, 0.5, 0.7, 0.9 which correspond to the mid-point percentiles of each portfolio. The formula for var($\hat{\beta}$) is given in equation (17) and the computation for the portfolio moments are derived in Appendix E.

Figure 3: Empirical Distributions of Betas and Idiosyncratic Volatilities



The figure plots an empirical histogram over the 29,096 firms in non-overlapping five year samples from 1960-2005, computed by OLS estimates. Panel A plots the histogram of market betas while Panel B plots the histogram of annualized log idiosyncratic volatility.