A skeptical appraisal of asset pricing tests ☆

Jonathan Lewellen [a,*], Stefan Nagel [b], Jay Shanken [c]

[a] Tuck School of Business, Dartmouth College, Hanover, NH 03755, USA
[b] Graduate School of Business, Stanford University, Stanford, CA 94305, USA
[c] Goizueta Business School, Emory University, Atlanta, GA 30322, USA

ARTICLE INFO

ABSTRACT

It has become standard practice in the cross-sectional asset pricing literature to evaluate models based on how well they explain average returns on size-B/M portfolios, something many models seem to do remarkably well. In this paper, we review and critique the empirical methods used in the literature. We argue that asset pricing tests are often highly misleading, in the sense that apparently strong explanatory power (high cross-sectional R^2s and small pricing errors) can provide quite weak support for a model. We offer a number of suggestions for improving empirical tests and evidence that several proposed models do not work as well as originally advertised.

© 2009 Elsevier B.V. All rights reserved.

1. Introduction

The finance literature has proposed a wide variety of asset pricing models in recent years, motivated, in part, by the well-known size and book-to-market (B/M) effects in stock returns. The models suggest new risk factors to help explain expected returns, including labor income (Jagannathan and Wang, 1996; Heaton and Lucas, 2000; Jacobs and Wang, 2004; Santos and Veronesi, 2006), growth in macroeconomic output and investment (Cochrane, 1996; Vassalou, 2003; Li, Vassalou, and Xing, 2006), growth in luxury, durable, and future consumption (Ait-Sahalia, Parker, and Yogo, 2004; Bansal, Dittmar, and Lundblad, 2005; Parker and Julliard, 2005; Yogo, 2006;

Hansen, Heaton, and Li, 2006), innovations in assorted state variables (Brennan, Wang, and Xia, 2004; Campbell and Vuolteenaho, 2004; Petkova, 2006), and liquidity risk (Pastor and Stambaugh, 2003; Acharya and Pedersen, 2005). In addition, the literature has suggested new variables to summarize the state of the economy, including the surplus consumption ratio (Campbell and Cochrane, 1999), the consumption-to-wealth ratio (Lettau and Ludvigson, 2001; Duffee, 2005), the housing-collateral ratio (Lustig and Van Nieuwerburgh, 2004), and the labor income-to-consumption ratio (Santos and Veronesi, 2006).

Empirically, many of the proposed models seem to do a good job explaining the size and B/M effects, an observation at once comforting and disconcerting—comforting because it suggests that rational explanations for the anomalies are readily available, but disconcerting because it provides an embarrassment of riches. Reviewing the literature, one gets the uneasy feeling that it seems a bit too easy to explain the size and B/M effects, a concern reinforced by the fact that many of the new models have little in common economically with each other.

* Corresponding author.

E-mail address: jonathan.lewellen@dartmouth.edu (J. Lewellen).

Our paper is motivated by the suspicion above. Specifically, our goal is to explain why, despite the seemingly strong evidence that many proposed models can explain the size and B/M effects, we remain unconvinced by the results. We offer a critique of the empirical methods that have become popular in the asset pricing literature, a number of prescriptions for improving the tests, and evidence that several proposed models do not work as well as the original evidence suggested.

The heart of our critique is that the literature has inadvertently adopted a low hurdle to meet in claiming success: high cross-sectional R^2s (or low pricing errors) when explaining average returns on Fama and French's (FF, 1993) 25 size-B/M portfolios. This hurdle is low because the size-B/M portfolios are well known to have a strong factor structure, in particular, FF's factors explain more than 90% of the time variation in the portfolios' realized returns and more than 80% of the cross-sectional variation in their average returns. Given those features, we show that obtaining a high cross-sectional R^2 in other models is easy because loadings on almost any proposed factor are likely to line up with expected returns—basically all that is required is for a factor to be (weakly) correlated with SMB or HML but not with the tiny, idiosyncratic three-factor residuals of the size-B/M portfolios.

The problem we highlight is not just a sampling issue, i.e., it is not solved by getting standard errors right. In population, if returns have a covariance structure like that of size-B/M portfolios, loadings on a proposed factor will line up with true expected returns so long as the factor correlates only with the common sources of variation in returns. The problem is also not solved by using a stochastic discount factor (SDF) approach. Under the same conditions that give a high cross-sectional R^2, the pricing errors in an SDF specification will also be small, a result that follows immediately from the close parallel between the regression and SDF approaches (see, e.g., Cochrane, 2001).

This is not to say that sampling issues are unimportant. Indeed, the covariance structure of size-B/M portfolios also implies that, even if we do find factors that explain little of the cross-sectional variation in true expected returns, we are still reasonably likely to estimate a high cross-sectional R^2 in sample. As an illustration, we simulate artificial factors that are correlated with returns but produce zero true cross-sectional R^2s for FF's size-B/M portfolios. We show that a sample adjusted R^2 might need to be as high as 44% to be statistically significant in models with one factor, 62% in models with three factors, and 69% in models with five factors. Further, with three or five factors, the power of the tests is extremely small: the sampling distribution of the adjusted R^2 is almost the same when the true R^2 is zero and when it is as high as 70% or 80%. In short, the high R^2s reported in the literature are not as impressive as they might, on the surface, appear.

The obvious question is then: What can be done? How can asset pricing tests be improved to make them more convincing? We offer four suggestions. First, since the problems are caused by the strong factor structure of size-B/M portfolios, one simple solution is to include other portfolios in the tests, sorted, for example, by industry, beta, or other characteristics. Second, because the problems are exacerbated when a model's risk premia are estimated as free parameters, another simple solution is to impose restrictions on the risk premia when theory provides appropriate guidance. For example, zero-beta rates should be close to the risk-free rate, the risk premium on a factor portfolio should be close to its average excess return, and the cross-sectional slopes in conditional models should be determined by the volatility of the equity premium (see Lewellen and Nagel, 2006). Third, we argue that the problems are likely to be less severe for generalized least squares (GLS) than for ordinary least squares (OLS) cross-sectional regressions, so another, imperfect, solution is to report the GLS R^2. An added benefit is that the GLS R^2 has a useful economic interpretation in terms of the relative mean-variance efficiency of a model's factor-mimicking portfolios (this interpretation builds on and generalizes the results of Kandel and Stambaugh, 1995). The GLS R^2 provides an imperfect solution because it suffers from the same problems as the OLS R^2 in some situations (as do other common asset pricing statistics).

Finally, because the problems are exacerbated by sampling issues, a fourth partial solution is to report confidence intervals for test statistics, not rely just on point estimates and p-values. We describe how to do so for the cross-sectional R^2 and other, more formal statistics based on the weighted sum of squared pricing errors, including Shanken's (1985) cross-sectional T^2 statistic, Gibbons, Ross, and Shanken's (1989) F-statistic, and Hansen and Jagannathan's (1997) HJ distance. For the latter three statistics, the confidence intervals again have a natural economic interpretation in terms of the relative mean-variance efficiency of a model's factor-mimicking portfolios.

Our suggestion to report confidence intervals has two main benefits, even apart from the other issues considered in this paper. The first is that confidence intervals can reveal the often high sampling error in the statistics—by showing the wide range of true parameters that are consistent with the data—in a way that is more direct and transparent than p-values or standard errors (standard errors are insufficient because asset pricing statistics are typically biased and skewed). The second advantage of confidence intervals over p-values is that they avoid the somewhat tricky problem of deciding on a null hypothesis. In economics, researchers typically set up tests with the null hypothesis being that a model does not work, or does not work better than existing theory, and then look for evidence to reject the null. (For example, in event studies the null is that stock prices do not react to the event.) Asset pricing tests often reverse the approach: the null is that a model works perfectly (zero pricing errors) which is 'accepted' as long as we do not find evidence to the contrary. This strikes us as a troubling shift in the burden of proof, particularly given the limited power of many tests. Confidence intervals mitigate this problem because they simply show the range of true parameters

that are consistent with the data without taking a stand on the right null hypothesis.

We apply these prescriptions to a handful of proposed models from the recent literature. The results are disappointing. None of the five models that we consider performs well in our tests, despite the fact that all seemed promising in the original studies.

The paper proceeds as follows. Section 2 formalizes our critique of asset pricing tests. Section 3 offers suggestions for improving the tests, and Section 4 applies the prescriptions to several recently proposed models. Section 5 concludes.

2. Interpreting asset pricing tests

In what follows, R is a vector of excess returns on N test assets (in excess of the risk-free rate) and F is a vector of K risk factors that perfectly explain expected returns, meaning that $\mu_R \equiv E[R]$ is linear in the $N \times K$ matrix of stocks' loadings on the factors, $B \equiv \text{cov}(R,F)\text{var}^{-1}(F)$. For simplicity, and without loss of generality, we assume that the mean of F equals the cross-sectional risk premium on B. Thus, our basic model is $\mu_R = B\mu_F$, or

$$R = BF + e, \tag{1}$$

where e is a vector of mean-zero residuals with $\text{cov}(e,F)=0$. The only assumption we make at this point about the covariance matrix of e is that it is nonsingular, so the model is completely general. [Eq. (1) has no economic content since an appropriate F can always be found; for example, any K portfolios that span the tangency portfolio would work.]

We follow the convention that all vectors are column vectors unless otherwise noted. For generic random variables x and y, $\text{cov}(x,y) \equiv E[(x-\mu_x)(y-\mu_y)']$. We use ι to denote a conformable vector of ones, 0 to denote a conformable vector or matrix of zeros, and I to denote a conformable identity matrix. M denotes the matrix $I - \iota\iota'/d$ that transforms, through pre-multiplication, the columns of any matrix with row dimension d into deviations from their means.

The factors in F can be thought of as a 'true' model that is known to price assets; it serves as a benchmark but we are not interested in it per se. Instead, we want to test a proposed model P consisting of J factors, with associated factor loadings $C \equiv \text{cov}(R,P)\text{var}^{-1}(P)$. We will say that P perfectly explains the cross section of expected returns if $\mu_R = C\gamma$ for some risk-premium vector γ (which, ideally, would be determined by theory).

A common way to test whether P is a good model is to estimate a cross-sectional regression of expected returns on factor loadings:

$$\mu_R = z\iota + C\lambda + \alpha, \tag{2}$$

where λ denotes a $J \times 1$ vector of regression slopes. In principle, we could test three features of Eq. (2): (i) z should be roughly zero (that is, the zero-beta rate should be close to the risk-free rate); (ii) λ should be non-zero and might be restricted by theory; and (iii) α should be zero and the cross-sectional R^2 should be one. In practice, empirical tests often focus on the restrictions that $\lambda \neq 0$

and the cross-sectional R^2 is one (the latter is sometimes treated only informally). The following observations consider the conditions under which P appears to be well specified in such tests.

Observation 1. *Suppose F and P have the same number of factors and cov(F, P) is nonsingular. Then P perfectly explains the cross section of expected returns if cov(e, P)=0, i.e., if P is correlated with R only through the common variation captured by F, even if P has arbitrarily small (but nonzero) correlation with F and explains little of the time variation in returns.*

Proof. If $\text{cov}(e,P)=0$, then stocks' factor loadings on P are linearly related to their loadings on F: $C \equiv \text{cov}(R,P)$ $\text{var}^{-1}(P)=B\,\text{cov}(F,P)\text{var}^{-1}(P)=BQ$, where $Q \equiv \text{cov}(F,P)\,\text{var}^{-1}(P)$. It follows that $\mu_R = B\mu_F = C\lambda$, where $\lambda = Q^{-1}\mu_F$. □

Observation 1 says that, if P has the same number of factors as F, testing whether expected returns are linear in betas with respect to P is nearly the same as testing whether P is uncorrelated with e, a test that does not seem to have much economic meaning in recent empirical applications. In particular, in tests with size-B/M portfolios, we know that R_M, SMB, and HML (the model F in our notation) explain nearly all of the time variation in returns (more than 92%), so the residual in $R=BF+e$ is both small and largely idiosyncratic. In that setting, we do not find it surprising that many macroeconomic factors are correlated with returns primarily through R_M, SMB, and HML—indeed, we would be more surprised if $\text{cov}(e,P)$ was not close to zero. In turn, we are not surprised that many proposed models seem to explain the cross section of expected size-B/M returns about as well as R_M, SMB, and HML. The strong factor structure of size-B/M portfolios makes it likely that their loadings on almost any proposed factor will line up with expected returns.[1]

Put differently, Observation 1 provides a skeptical interpretation of recent asset pricing tests, in which unrestricted cross-sectional regressions (or equivalently SDF tests; see Cochrane, 2001) have become the norm. In our view, many empirical tests say only that a number of proposed factors are correlated with SMB and HML, a fact that might have some economic content but seems like a fairly low hurdle to meet in claiming that a proposed model explains the size and B/M effects.

It is important to note that, in Observation 1, pricing errors for the proposed model are exactly zero and the model works perfectly, in population, based on any metric of performance. Thus, our concern about tests with size-B/M portfolios is not just a critique of the OLS R^2 but also applies to formal statistics such as Hansen's (1982) J-test, Shanken's (1985) T^2 statistic, and Hansen and Jagannathan's (1997) HJ distance. We will discuss

---

[1] This argument works most cleanly if P has three factors, matching FF's model. It should also apply when P has two factors because size-B/M portfolios all have multiple-regression market betas close to one in FF's model (see Fama and French, 1993). In essence, FF's model is really just a two-factor model for the purposes of explaining cross-sectional variation in expected returns (on size-B/M portfolios).

later, however, a number of reasons to believe the problem is less severe in practice for formal asset pricing statistics.

Observation 2. *Suppose returns have a strict factor structure with respect to F, i.e., var(e) is a diagonal matrix. Then any set of K assets perfectly explains the cross section of expected returns so long as the K assets are not asked to price themselves (that is, the K assets are not included on the left-hand side of the cross-sectional regression and the risk premia are not required to equal their expected returns). The only restriction is that $cov(F, R_K)$ must be nonsingular.*

Proof. Let $P=R_K$ in Observation and re-define R as the vector of returns for the remaining $N-K$ assets and e as the residuals for these assets. The strict factor structure implies that $cov(e, R_K)=cov(e, B_K F+e_K)=0$. The result then follows immediately from Observation 1. □

Observation 2 is useful for a couple of reasons. First, it provides a simple illustration of our argument that, in some situations, it is easy to find factors that explain the cross section of expected returns: under the common arbitrage pricing theory (APT) assumption of a strict factor structure, any collection of K assets would work. Obtaining a high cross-sectional R^2 is not very difficult when returns have a strong factor structure, as they do in many empirical applications.

Second, Observation 2 illustrates that it can be important to take restrictions on the cross-sectional slopes seriously. In particular, the result hinges on the fact that the risk premia on R_K are not required to equal the factors' expected returns, as theory would require. To see why, Observation 1 (proof) shows that the cross-sectional slopes on C are $\lambda = Q^{-1}\mu_F$, where Q is the matrix of slope coefficients when F is regressed on R_K. In the simplest case with one factor, λ simplifies to $\mu_K/cor^2(R_K, F)$, which clearly does not equal μ_K unless $cor(R_K, F) = 1$ (the formula for λ follows from the definition of Q and the fact that $\mu_K = B_K\mu_F$). The implication is that the problem highlighted by Observations 1 and 2—that 'too many' factors explain the cross section of expected returns—would be less severe if the restriction on λ was imposed (R_K would price the cross section only if $cor(R_K, F) = 1$).

Observations 1 and 2 are rather special since, in order to get clean predictions, they assume that a proposed model P has the same number of factors as the known model F. The intuition goes through when $J<K$ because, even in that case, we would typically expect the loadings on proposed factors to line up (imperfectly) with expected returns if returns have a strong factor structure. The next observation generalizes our results, at the cost of changing the definitive conclusion in Observations 1 and 2 into a probabilistic statement.

Observation 3. *Suppose F has K factors and P has J factors, with $J \leq K$. Assume, as before, that P is correlated with R only through the variation captured by F, meaning that $cov(e, P) = 0$ and that $cov(F, P)$ has rank J. In a generic sense, made precise below, the cross-sectional R^2 in a regression of μ_R on C is expected to be J/K.*

Proof. By a generic sense, we mean that nothing is known about the proposed factors and, thus, we treat the

loadings on P as randomly related to those on F (a similar result holds if we treat the factors themselves as randomly related). More specifically, suppose F is normalized to make $B'MB/N=I_K$, i.e., the loadings on F are cross-sectionally uncorrelated and have unit variances, so loadings on all factors have the same scale. Observation 1 shows that $C=BQ$, where Q is a $K \times J$ matrix. A generic sense means that we view the elements of Q as randomly drawn from a $N[0, \sigma_q^2]$ distribution. The proof then proceeds as follows: In a regression of μ_R on ι and C, the R^2 is $\mu_R' MC(C'MC)^{-1}C'M\mu_R/\mu_R' M\mu_R$. Substituting $\mu_R=B\mu_F$ and $C=BQ$, and using the assumption that $B'MB/N=I_K$, the R^2 simplifies to $\mu_F' Q(Q'Q)^{-1}Q'\mu_F/\mu_F' \mu_F$. The result then follows from observing that $E[Q(Q'Q)^{-1}Q']$ is a diagonal matrix with J/K on the diagonal,[2] so $E[R^2] = \mu_F' [(J/K)I_K]\mu_F/\mu_F' \mu_F = J/K$. □

Observation 3 generalizes Observations 1 and 2. Our earlier results show that, if a K-factor model explains both the cross section of expected returns and much of the time-series variation in returns, it should be easy to find other K-factor models that also explain the cross section of expected returns. The issue is messier with $J<K$. Intuitively, the more factors that are in the proposed model, the easier it should be to find a high cross-sectional R^2 as long as the proposed factors are correlated with the 'true' factors. Thus, we are not surprised if a proposed three-factor model explains the size and B/M effects about as well as the FF factors, nor are we surprised if a one- or two-factor model has some explanatory power. We *are* impressed if a one-factor model works as well as the FF factors, since this requires a single factor to capture the pricing information in both SMB and HML. (We note again that size-B/M portfolios all have FF three-factor market betas close to one, so the model can be thought of as a two-factor model consisting of SMB and HML for the purposes of explaining cross-sectional variation in expected returns.)

Fig. 1 illustrates these results using simulations with FF's 25 size-B/M portfolios, moving beyond the specific assumptions underlying Observations 1–3. We calculate quarterly excess returns on the 25 portfolios from 1963 to 2004 and explore, in several simple ways, how easy it is to find factors that explain expected returns (or, put differently, how rare it is to find factors that do not). The figure treats the average returns and sample covariance matrix as population parameters. Thus, like Observations 1–3 the figure focuses on explaining expected returns in population, not on sampling issues (which we consider later).

The three panels generate different types of artificial factors to explain expected returns. The simulations in Panel A match the assumptions underlying

---

[2] Let q_i' be the ith row of Q. The off-diagonal terms of $Q(Q'Q)^{-1}Q'$ can be expressed as $q_i' (Q'Q)^{-1}q_j$ for $i \neq j$, and the matrix $Q'Q$ equals $\Sigma_i q_i q_i'$. q_i is independent of q_j for $i \neq j$ and is mean independent of $q_i q_i'$ (because of normality), implying that $E[q_i|Q'Q$ and $q_{j\neq i}]=0$ and, thus, $E[q_i' (Q'Q)^{-1}q_j]=0$. Also, the diagonal elements of $Q(Q'Q)^{-1}Q'$ must all have the same expected values since the elements of Q are assumed to be identically distributed. It follows that $E[Q(Q'Q)^{-1}Q']=(1/K)$ $E[tr(Q(Q'Q)^{-1}Q')I_K]=(1/K) E[tr(Q'Q(Q'Q)^{-1})]I_K=(J/K)I_K$.
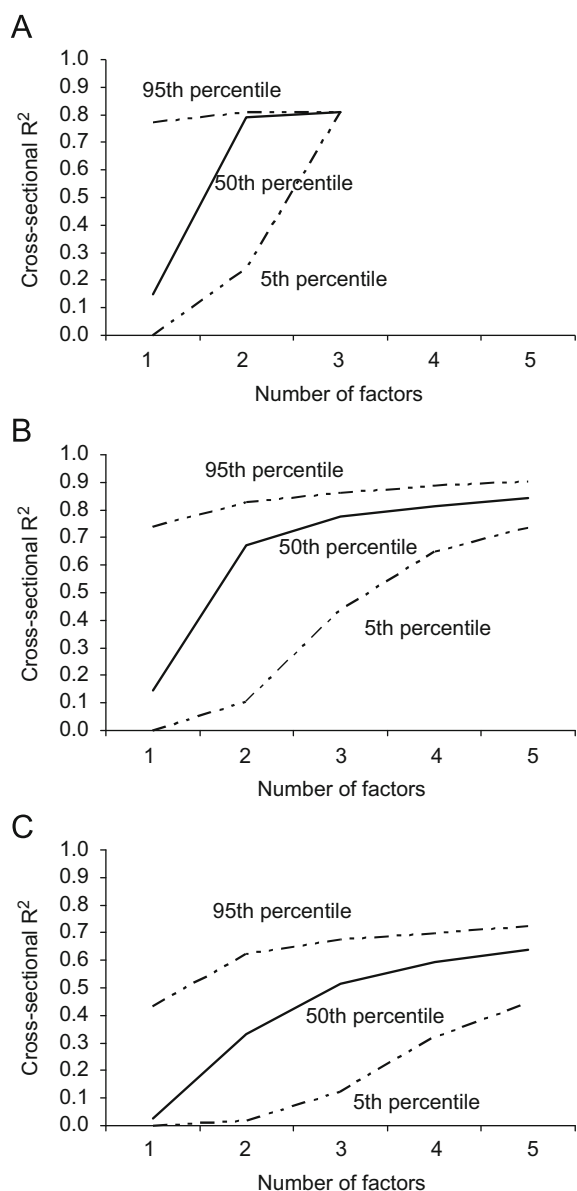
A

B

C

Fig. 1. Population R^2s for artificial factors. We randomly generate artificial factors and report the simulated distribution of population R^2s when expected returns on the Fama-French 25 size-B/M portfolios are regressed on their artificial factor loadings (quarterly average returns, variances, and covariances from 1963 to 2004 are treated as population parameters in the simulations). In Panel A, macro factors are generated by randomly drawing a 3×1 vector of weights, w, from a standard normal distribution, defining a factor $P=F'w$, where $F=[R_M,$ SMB, HML]. In Panel B, factors are constructed as zero-investment combinations of the size-B/M portfolios by randomly drawing a 25×1 vector of weights from a standard normal distribution. In Panel C, the procedure for Panel B is repeated but only factors with (roughly) zero expected returns are retained. The plots show the 5th, 50th, and 95th percentiles of the simulated distribution based on 5,000 draws of up to five factors at a time.

Observations 1–3 most closely: we generate artificial 'macro' factors that are correlated with R_M, SMB, and HML but not with the three-factor residuals of the size-B/M portfolios. Specifically, we randomly draw 3×1 vectors of

weights, w_i, from a standard normal distribution, each defining a factor $P_i=w'_iF+v_i$, where $F=[R_M,$ SMB, HML] and v_i is an arbitrary random variable independent of returns. The covariance between returns and P_i is then $g_i=\text{cov}(R,P_i)=\text{cov}(R,F)w_i$. We repeat this five thousand times, generating up to three artificial factors at a time, and report the cross-sectional R^2 when size-B/M portfolios' expected returns are regressed on the g_i. The simulations capture the idea that a proposed factor correlates at least somewhat with the common factors in returns (the FF factors) but, consistent with it being macroeconomic, is orthogonal to the idiosyncratic residuals of the size-B/M portfolios.

The results in Panel A suggest that it can be easy to find factors that help explain expected returns on the size-B/M portfolios. Taken individually, half of our artificial factors produce an OLS R^2 greater than 0.15 and 23% produce an R^2 greater than 0.50 (the latter is not reported in the figure). Using two factors, the median R^2 is 0.79 and a remarkable 89% of the models have an R^2 greater than 0.50. These high values reflect the fact that FF's model is basically a two-factor model for the purpose of explaining cross-sectional variation in expected returns. Finally, with three factors, the simulated R^2 always matches the cross-sectional R^2 of FF's model, 0.81.

In Panel B, we relax the assumption that our artificial factors are completely uncorrelated with the residuals of the 25 size-B/M portfolios. Specifically, we generate factors that are simply random combinations of the size-B/M portfolios by drawing a 25×1 vector of portfolio weights from a standard normal distribution (the weights are shifted and re-scaled to have a mean of zero and to have one dollar long and one dollar short). The logic is that any asset pricing factor can always be replaced by an equivalent mimicking portfolio of the test assets—equivalent in the sense that the model's pricing errors do not change—and these simulations explore how easy it is to stumble across mimicking portfolios that explain the cross section of expected returns. In a sense, the simulations also explore how special the FF factors are: How much worse does a random combination of the size-B/M portfolios perform relative to R_M, SMB, and HML in terms of cross-sectional explanatory power?

Panel B again suggests that it is easy to find factors that explain expected returns on the size-B/M portfolios. The median R^2 using one factor is 0.15, jumping to 0.67, 0.78, 0.82, and 0.84 for models with two, three, four, and five factors, respectively. More than 71% of our artificial two-factor models and 92% of our artificial three-factor models explain at least half of the cross-sectional variation in expected returns. Ten percent of the two-factor models and 30% of the three-factor models actually explain more cross-sectional variation than the FF factors.[3]

Finally, Panel C repeats the simulations in Panel B with a small twist: we keep only artificial factors (random

---

[3] These facts in no way represent an indictment of Fama and French (1993) since one of their main points was precisely that returns on the size-B/M portfolios could be summarized by a small number of factors. Our simulations just indicate that the factors they constructed are far from unique in their ability to explain cross-section variation in expected returns on the size-B/M portfolios.

combinations of the 25 size-B/M portfolios) that have roughly zero expected returns. These simulations illustrate how important it can be to impose restrictions on the cross-sectional slopes. Theoretically, the risk premia on the artificial factors in Panel C should be zero, matching their expected returns, but the panel ignores this restriction and just searches for the best possible fit in the cross-sectional regression. Thus, the simulated R^2s differ from zero only because we ignore the theoretical restrictions on the cross-sectional slopes and intercept.

The additional degrees of freedom turn out to be very important, especially with multiple factors. The median R^2s are 0.03, 0.52, and 0.64 for models with one, three, and five factors, respectively, while the 95th percentiles are 0.53, 0.67, and 0.72. Twenty-four percent of the artificial three-factor models and 54% of the artificial five-factor models explain at least half of the cross-sectional variation in expected returns, even though properly restricted R^2s would be zero.

These results show that it may be easy to explain expected returns, in population, when assets have a covariance structure like that of the size-B/M portfolios. Our final observation suggests that the problem can be worse taking sampling error into account.

Observation 4. *The problems are exacerbated by sampling issues: If returns have a strong factor structure, it can be easy to find a high sample cross-sectional R^2 even in the unlikely scenario that the population R^2 is small or zero.*

Observation 4 is intentionally informal and, in lieu of a proof, we offer simulations using FF's 25 size-B/M portfolios to illustrate the point. The simulations differ from those in Fig. 1 because, rather than study the population R^2 for artificial factors, we now focus on sampling variation in estimated R^2s *conditional on a given population R^2*. The simulations have two steps. First, we fix a true cross-sectional R^2 that we want a model to have and randomly generate a matrix of factor loadings, C, which produces that R^2. Factor portfolios, $P = w'R$, are constructed to give those factor loadings, i.e., we find portfolio weights, w, such that $cov(R, P)$ is linear in C.[4] Second, we bootstrap artificial time series of returns and factors by sampling, with replacement, from the historical time series of size-B/M returns (quarterly, 1963–2004). We then estimate the sample cross-sectional adjusted R^2 for the artificial data by regressing average returns on estimated factor loadings. The second step is repeated four thousand times to construct a sampling distribution of the adjusted R^2. In addition, to make sure the particular matrix of loadings generated in step 1 is not crucial, we repeat that step ten times, giving us a total sample of 40,000 adjusted R^2s corresponding to an assumed true R^2.

Fig. 2 shows results for models with one, three, and five factors. The left-hand column plots the distribution of the sample adjusted R^2 (5th, 50th, and 95th percentiles) corresponding to true R^2s of 0.0–1.0 for the simulations described above. The right-hand column repeats the exercise but uses factors that are imperfectly correlated with returns, as they are in most empirical applications. We start with the portfolio factors used in the left-hand panels and add noise equal to three times their variance. Thus, for the right-hand plots, a maximally correlated combination of the size-B/M portfolios would have a time-series R^2 of 0.25 with each factor.

The figure suggests that a sample adjusted R^2 must be quite high to be statistically significant, especially for models with several factors. Focusing on the simulations in the right-hand column, the 95th percentile of the sampling distribution is 44% using one factor, 62% using three factors, and 69% using five factors when the true cross-sectional R^2 is zero. Thus, even if we could find factors that have no true explanatory power (something that seems unlikely given our population results above), it would still not be too unusual to find fairly high R^2s in sample. Further, with multiple factors, the ability of the sample R^2 to discriminate between good and bad models is quite limited since the distribution of the sample R^2 is similar across a wide range of true R^2s. For example, with five factors, a sample R^2 greater than 73% is needed to reject that the true R^2 is 30% or less (at a 0.05 one-sided significance level), but that outcome is unlikely even if the true R^2 is 70% (probability of 0.17) or 80% (probability of 0.26). The bottom line is that, in both population and sample, the cross-sectional R^2 does not seem to be a very useful metric for distinguishing among models.

2.1. Related research

Our appraisal of asset pricing tests overlaps with a number of studies. Roll and Ross (1994) and Kandel and Stambaugh (1995) argue that, in tests of the capital asset pricing model (CAPM), the cross-sectional R^2 is not very meaningful because, as a theoretical matter, it tells us little about the location of the market proxy in mean-variance space. We reach a similarly skeptical conclusion about the R^2, but our main point—that it can be easy to find factors that explain expected returns when assets have a covariance structure like that of the size-B/M portfolios—is quite different. The closest overlap comes from our simulations in Panel C of Fig. 1, which show that factor portfolios with zero mean returns might still produce high R^2 in unrestricted cross-sectional regressions. These portfolios are far from the mean-variance frontier by construction (they have zero Sharpe ratios) yet often have high explanatory power, consistent with the results of Roll and Ross and Kandel and Stambaugh.

Kan and Zhang (1999) study cross-sectional tests with so-called useless factors, defined as factors that are independent of asset returns. They show that regressions with useless factors can be misleading because the usual

---

[4] Specifically, for a model with J factors, we randomly generate J vectors, g_j, that are uncorrelated with each other and which individually have explanatory power of $c^2 = R^2/J$ (and, thus, the correct combined R^2). Each vector is generated as $g_j = c\mu_s + (1-c^2)^{1/2}e_j$, where μ_s is the vector of expected returns on the size-B/M portfolios, shifted and re-scaled to have mean zero and standard deviation of one, and e_j is generated by randomly drawing from a standard normal distribution (e_j is transformed to have exactly mean zero and standard deviation of one, to be uncorrelated with μ_s, and to make $cov(g_i, g_j) = 0$ for $i \neq j$). The factor portfolios in the simulations have covariance with returns (a 25×1 vector) given by the g_j.
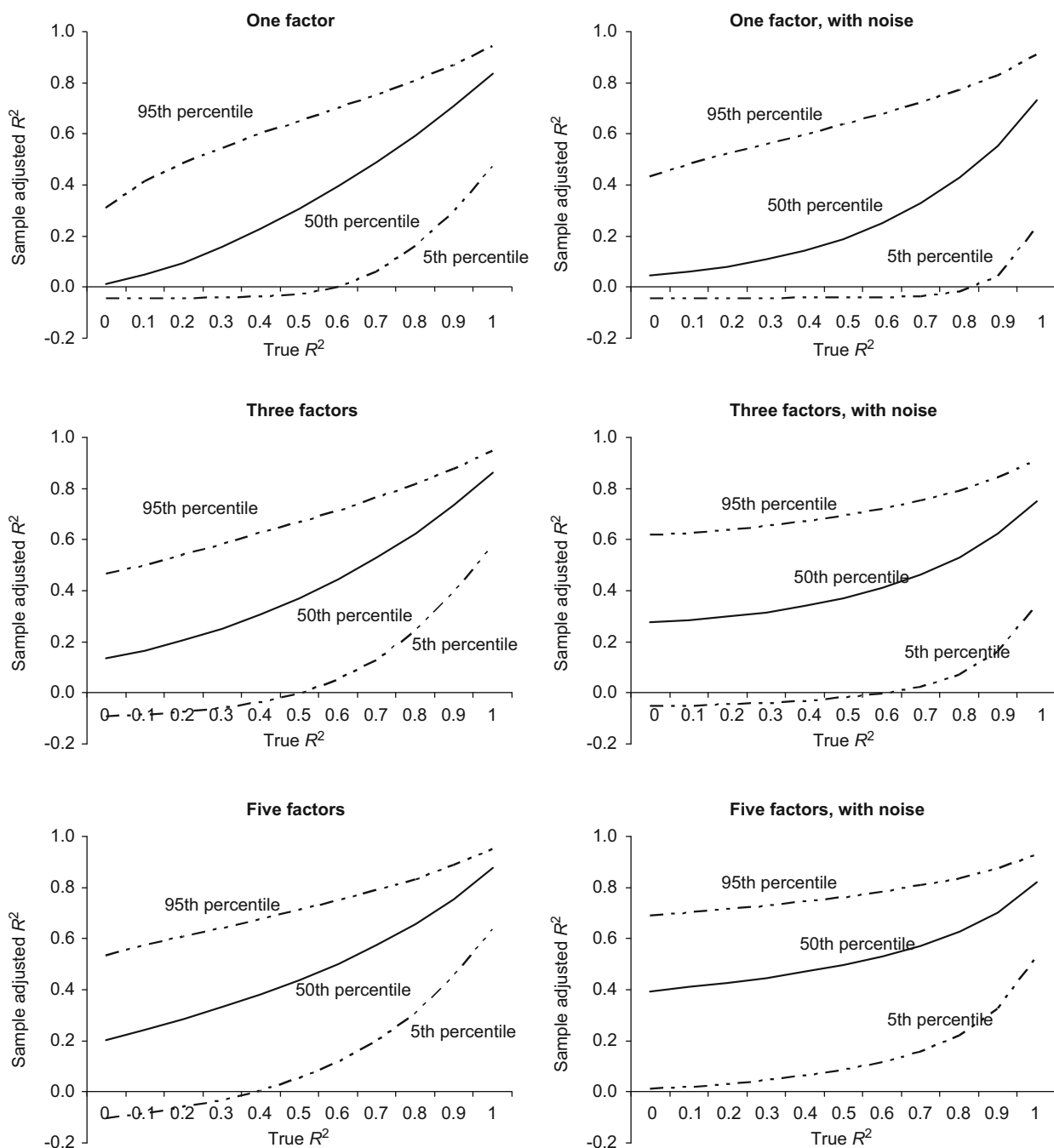
Fig. 2. The figure shows the sample distribution of the cross-sectional adjusted R^2 (average returns regressed on estimated factor loadings) as a function of the true R^2 for artificial asset-pricing models with one to five factors, using Fama and French's 25 size-B/M portfolios as test assets (quarterly returns, 1963–2004). In the left-hand panels, the factors are constructed as combinations of the 25 size-B/M portfolios (the weights are randomly drawn, as described in the text, to produce the true R^2 reported on the x-axis). In the right-hand panels, noise is added to the factors equal to three times their variance, to simulate factors that are not perfectly spanned by returns. The plots are based on 40,000 bootstrap simulations (ten sets of random factors; 4,000 simulations with each).

asymptotic theory breaks down, due to the fact that the cross-sectional spread in estimated loadings goes to zero as the time series grows. Our results, in contrast, focus on population R^2s and apply to factors that are correlated with asset returns.

Ferson, Sarkissian, and Simin (1999) also critique asset pricing tests, emphasizing how difficult it can be to distinguish a true 'risk' factor from an irrationally priced factor portfolio. They show that loadings on a factor portfolio can be cross-sectionally correlated with expected

returns even if the factor simply picks up mispricing. Our analysis is different because it applies to general macroeconomic factors, not just return factors, and highlights the difficulties created by the strong covariance structure of the size-B/M portfolios.

Some of our results are reminiscent of the literature on testing the APT and multifactor models (see, e.g., Shanken, 1987, 1992a; Reisman, 1992). Most closely, Nawalkha (1997) derives results like our Observations 1 and 2, though with a much different message. In particular, he emphasizes that, in the APT, well-diversified variables (those uncorrelated with idiosyncratic risks) can be used in place of the true factors without any loss of pricing accuracy. We generalize his theoretical results to models with $J < K$ proposed factors, consider sampling issues, and emphasize the empirical implications for recent tests using size-B/M portfolios.

Finally, our critique is similar in spirit to a contemporaneous paper by Daniel and Titman (2005). They show that, even if expected returns are determined by firm characteristics such as B/M, a proposed factor can appear to price characteristic-sorted portfolios simply because loadings on the factor are correlated with characteristics in the underlying population of stocks (and forming portfolios tends to inflate the correlation). Our ultimate conclusions about using characteristic-sorted portfolios are similar, but we highlight different concerns, emphasizing the importance of the factor structure of size-B/M portfolios, the impact of using many factors and not imposing restrictions on the cross-sectional slopes, and the role of both population and sampling issues.

3. Improving asset pricing tests

The theme of Observations 1–4 is that, in situations like those encountered in practice, it may be easy to find factors that explain the cross section of expected returns. Finding a high cross-sectional R^2 and small pricing errors often has little economic meaning and, in our view, does not, by itself, provide much support for a proposed model. The problem is not just a sampling issue—it cannot be solved by getting standard errors right—though sampling issues exacerbate the problem. Here, we offer a few suggestions for improving empirical tests.

Prescription 1. Expand the set of test portfolios beyond size-B/M portfolios.

Empirical studies often focus on the size and B/M effects because of their importance. This practice is understandable but problematic, since the concerns highlighted above are most severe when a couple of factors explain nearly all of the time variation in returns, as is true for size-B/M portfolios. One simple solution, then, is to use additional portfolios that do not correlate as strongly with SMB and HML. Reasonable choices include portfolios sorted by industry, beta, volatility, or factor loadings (the last being loadings on the proposed factor; an alternative would be to use individual stocks, but errors-in-variables problems could make this impractical, or statistically based portfolios, such as those described by Ahn, Conrad,

and Dittmar, 2009). Bond portfolios might also be used. The idea is to price all portfolios at the same time, not in separate cross-sectional regressions.

Two points are worth emphasizing. First, the additional portfolios do not need to offer a big spread in expected returns; the goal is simply to relax the tight factor structure of size-B/M portfolios. A different way to say this is that adding portfolios can be useful as long as they exhibit variation in either expected returns, on the left-hand side of the cross-sectional regression, or in risk, on the right-hand side. One is necessary, not both.

Second, we acknowledge the concern that no model is perfect and will explain all patterns in the data. This truism makes it tempting to view a model as successful if it explains even one or two anomalies, such as the size and B/M effects. The problem with this view, however, is that tests with size-B/M portfolios alone do not provide a sufficient test of a model, for all of the reasons discussed in Section 2. It just does not seem practical to judge a model as 'successful' if it works only on those assets. Put differently, we expect many models to price the size-B/M portfolios about as well as the Fama-French factors, so tests with size-B/M portfolios alone do not provide a meaningful way to distinguish among the theories (though some of our suggestions below can help).

Fig. 3 illustrates the benefits of expanding the set of test assets. We replicate the simulations in Fig. 1 but, instead of using only size-B/M portfolios, we augment them with FF's 30 industry portfolios. As before, we explore how well artificial factors explain, in population, the cross section of expected returns (average returns, variances, and covariances from 1963 to 2004 are treated as population parameters). The artificial factors are generated in three ways. In Panel A, the factors are constructed by randomly drawing 3×1 vectors of weights, w_i, from a N[0,1] distribution, defining a factor $P_i = w'_i F + v_i$, where $F = [R_M, SMB, HML]$ and v_i is orthogonal to returns. In Panel B, the factors are constructed as zero-investment combinations of the size-B/M-industry portfolios by randomly drawing a 55×1 vector of portfolio weights from a N[0,1] distribution. In Panel C, we repeat the simulations of Panel B but use only factor portfolios with expected returns of zero. The point in each case is to explore how easy it is to find factors that produce a high population cross-sectional R^2. (Section 2 provides a more complete description of the simulations.)

Fig. 3 suggests that it is much 'harder,' using artificial factors, to explain expected returns on the 55 portfolios than on the 25 size-B/M portfolios (the median and 95th percentiles for the latter are repeated from Fig. 1 for comparison). For example, in models with three factors, the median R^2s in Panels A, B, and C are 35%, 21%, and 11%, respectively, for the full set of 55 portfolios, compared with median R^2s of 81%, 78%, and 52% for the size-B/M portfolios. The difference between the size-B/M portfolios and the full set of 55 portfolios rises substantially for models with multiple factors, consistent with the factor structure of size-B/M portfolios being important. In short, explaining returns on the full set of portfolios seems to provide a higher hurdle for a proposed model.
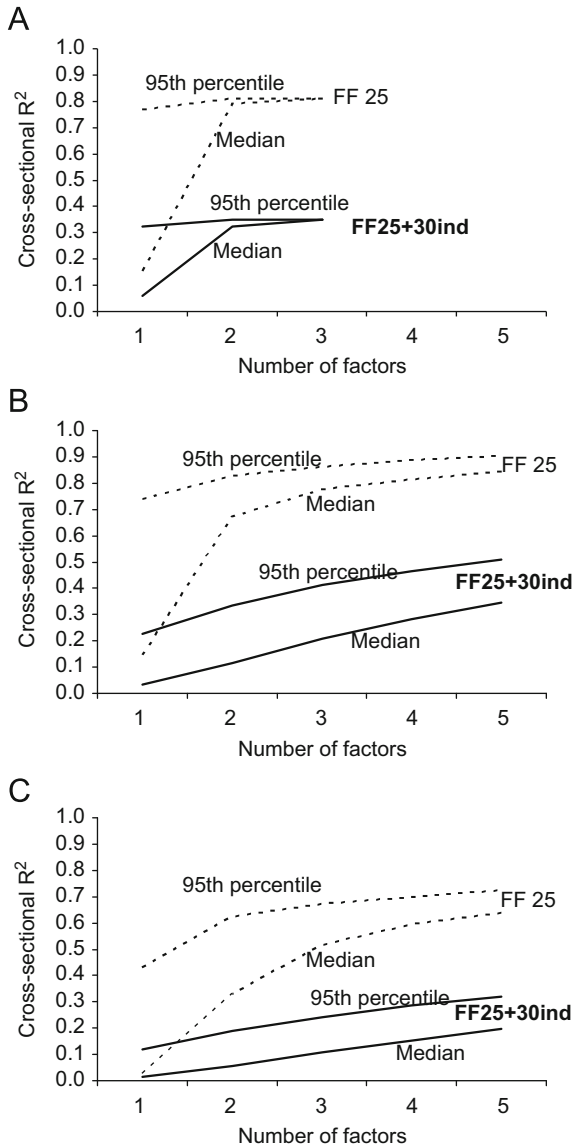
A

B

C

Fig. 3. Population R^2s for artificial factors: simulations using size-B/M and industry portfolios. We randomly generate artificial factors and report the simulated distribution of population R^2s when expected returns on the Fama-French 25 size-B/M portfolios alone (FF25; dotted lines) or augmented with the Fama-French 30 industry portfolios (FF25+30ind; solid lines) are regressed on their artificial factor loadings (quarterly average returns, variances, and covariances from 1963 to 2004 are treated as population parameters in the simulations). In Panel A, macro factors are generated by randomly drawing a 3×1 vector of weights, w, from a standard normal distribution, defining a factor $P=Fw$, where $F=[R_M, SMB, HML]$. In Panel B, factors are constructed as zero-investment combinations of the size-B/M-industry portfolios by randomly drawing a 55×1 vector of weights from a standard normal distribution. In Panel C, the procedure for Panel B is repeated but only factors with (roughly) zero expected returns are retained. The plots show the median and 95th percentile of the simulated distribution based on 5,000 draws of up to five factors at a time.

Prescription 2. Take the magnitude of the cross-sectional slopes seriously.

The literature sometimes emphasizes a model's high cross-sectional R^2 but does not consider whether the

estimated slopes and zero-beta rates are reasonable. Yet theory often provides guidance for both that should be taken seriously, i.e., the theoretical restrictions should be imposed ex ante or tested ex post. Most clearly, theory says that the zero-beta rate should equal the risk-free rate. A possible retort is that Brennan's (1971) model relaxes this constraint if borrowing and lending rates differ, but this argument is not convincing in our view: (riskless) borrowing and lending rates are not sufficiently different, perhaps 1–2% annually, to justify the extremely high zero-beta estimates in many papers. An alternative argument is that the equity premium is anomalously high, à la Mehra and Prescott (1985), so it is unfair to ask a consumption-based model to explain it. However, it does not seem reasonable to accept a model that cannot explain the level of expected returns.

A related restriction, mentioned earlier, is that the risk premium associated with a factor portfolio should be the factor's expected excess return. For example, the cross-sectional price of market-beta risk should equal the aggregate equity premium, and the price of interest-rate risk, captured by movements in long-term Tbond returns, should equal the expected Tbond return over the risk-free rate. In practice, this type of restriction could be tested in cross-sectional regressions or imposed ex ante by focusing on time-series regression intercepts (Jensen's alphas). Below, we discuss ways to incorporate the constraint into cross-sectional regressions (see also Shanken, 1992b).

As a third example, conditional models generally imply concrete restrictions on the cross-sectional slopes, a point emphasized by Lewellen and Nagel (2006). For example, Jagannathan and Wang (1996) show that a one-factor conditional CAPM implies a two-factor unconditional model: $E_{t-1}[R_t]=\beta_t\gamma_t \rightarrow E[R]=\beta\gamma+\text{cov}(\beta_t, \gamma_t)$, where β_t and γ_t are the conditional beta and equity premium, respectively, and β and γ are their unconditional means. The cross-sectional slope on $\text{cov}(\beta_{it}, \gamma_t)$ in the unconditional regression should be one but that constraint is often neglected in the literature. Lewellen and Nagel discuss this issue in detail and provide empirical examples from recent tests of both the simple and consumption CAPMs. For tests of the simple CAPM, the constraint can be imposed using the conditional time-series regressions of Shanken (1990), if the relevant state variables are all known, or the short-window approach of Lewellen and Nagel, if they are not.

Prescription 3. Report the GLS cross-sectional R^2.

The literature typically favors OLS over GLS cross-sectional regressions. The rationale for neglecting GLS regressions appears to reflect concerns with the statistical properties of GLS and the apparent difficulty of interpreting the GLS R^2, which, on the surface, simply tells us about the model's ability to explain expected returns on 're-packaged' portfolios, not the basic portfolios that are of direct interest (if μ and B are expected returns and loadings for the test assets, OLS regresses μ on $[\iota \, B]$ while GLS regresses $V^{-1/2}\mu$ on $V^{-1/2}[\iota \, B]$, where $V=\text{var}(R)$). We believe these concerns are misplaced, or at least

overstated, and that GLS actually has some advantages over OLS.

The statistical concerns with GLS are real but not prohibitive. The main issue is that, because the covariance matrix of returns must be estimated, the exact finite-sample properties of GLS are generally unknown and the asymptotic properties of textbook econometrics can be a poor approximation, especially when the number of assets is large relative to the length of the time series (Shanken, 1985, provides examples in a closely related context; see also Shanken and Zhou, 2007). But we see little reason this problem cannot be overcome using simulation methods or, in special cases, the finite-sample results of Shanken (1985) or Gibbons, Ross, and Shanken (1989).

The concern that the GLS R^2 is hard to interpret also seems misplaced. In fact, Kandel and Stambaugh (1995) show that the GLS R^2 is, in some ways, a more meaningful statistic than the OLS R^2: when expected returns are regressed on betas with respect to a factor portfolio, the GLS R^2 is completely determined by the factor's proximity to the minimum-variance boundary while the OLS R^2 has, in general, little connection to the factor's location in mean-variance space (see also Roll and Ross, 1994; this result assumes the factor is spanned by the test assets). Thus, if a market proxy is nearly mean-variance efficient, the GLS R^2 is nearly one but the OLS R^2 can, in principle, be anything. A factor's proximity to the minimum-variance boundary is not the only metric for evaluating a model, but it does seem to be both economically reasonable and easy to understand. The broader point is that, while the OLS R^2 might be relevant for some questions—for example, asking whether a model's predictions of expected returns are accurate for a given set of assets (subject to the limitations discussed in Section 2)—the GLS R^2 is probably more relevant for other questions—for example, asking how well a model explains the risk-return opportunities available in the market.

The same ideas apply to models with non-return factors. In this case, Appendix A shows that a GLS regression is equivalent to using maximally-correlated mimicking portfolios in place of the actual factors *and* imposing the constraint that the risk premia on the portfolios equal their excess returns (in excess of the zero-beta rate if an intercept is included). The GLS R^2 is determined by the mimicking portfolios' proximity to the minimum-variance boundary, i.e., the distance from the boundary to the 'best' combination of the mimicking portfolios. Again, this distance seems like a natural metric by which to evaluate a model since any linear asset pricing model boils down to a prediction that the factor-mimicking portfolios span the mean-variance frontier. The appendix also shows that the GLS R^2 is closely linked to formal asset pricing tests, such as Shanken's (1985) cross-sectional test of linearity or the HJ distance (see also Kan and Zhou, 2004).

One implication of these facts is that obtaining a high GLS R^2 would seem to be a more rigorous hurdle than obtaining a high OLS R^2: a model can produce a high OLS R^2 even though its factor mimicking portfolios are far from mean-variance efficient, while the GLS R^2 is high only if the model can explain the maximum Sharpe ratio available on the test assets.

The implicit restrictions imposed by GLS are not a full solution to the problems discussed in Section 2. Indeed, Observations 1 and 2 apply equally to OLS and GLS regressions (both R^2s are one given the stated assumptions). But simulations with artificial factors, which relax the strong assumptions of the formal propositions, suggest that finding a high population GLS R^2 in practice is much less likely than finding a high OLS R^2. Fig. 4 illustrates this result. The figure shows GLS R^2s for the same simulations reported in Fig. 1, using artificial factors to explain expected returns on FF's size-B/M portfolios (treating their sample moments from 1963 to 2004 as population parameters; the OLS plots are repeated for comparison). The plots show that, while artificial factors have some explanatory power in GLS regressions, the GLS R^2s are dramatically lower than the OLS R^2s. The biggest difference is in Panel C, which constructs artificial factors that are random, zero-cost combinations of the 25 size-B/M portfolios, imposing the restriction that the factors' Sharpe ratios are zero. The GLS R^2s are appropriately zero because the risk premia on the factors match their expected returns (zero), while the OLS R^2 are often 50% or more in models with multiple factors.

The advantage of GLS over OLS regressions in the simulations seems to come from two sources. The first, discussed above, is that GLS forces the risk premium on a factor (or the factor's mimicking portfolio, in the case of non-return factors) to equal its expected return, and the GLS R^2 is determined solely by the factor's location in mean-variance space. The second is that, in practice, FF's factors have much less cross-sectional explanatory power in a GLS regression than in an OLS regression using the size-B/M portfolios: the GLS R^2 is just 19.5%, compared with an OLS R^2 of 80.9%. The implication is that a proposed model must do more than simply piggy-back off SMB and HML if it is to achieve a high GLS R^2. Both issues suggest that, in practice, obtaining a high GLS R^2 represents a more stringent hurdle than obtaining a high OLS R^2.

Prescription 4. If a proposed factor is a traded portfolio, include it as one of the test assets on the left-hand side of the cross-sectional regression.

Prescription 4 builds on Prescription 2, in particular, the idea that the cross-sectional price of risk for a factor portfolio should be the factor's expected excess return. One simple way to build this restriction into a cross-sectional regression is to ask the factor to price itself, that is, to test whether the factor itself lies on the estimated cross-sectional regression line.

Prescription 4 is most important with GLS regressions. When a factor portfolio is included as a left-hand-side asset, GLS forces the regression to price the asset perfectly: the risk premium on the factor exactly equals the factor's average return in excess of the estimated zero-beta rate (in essence, the asset is given infinite weight in the regression). Thus, a GLS cross-sectional regression, when a traded factor is included as a
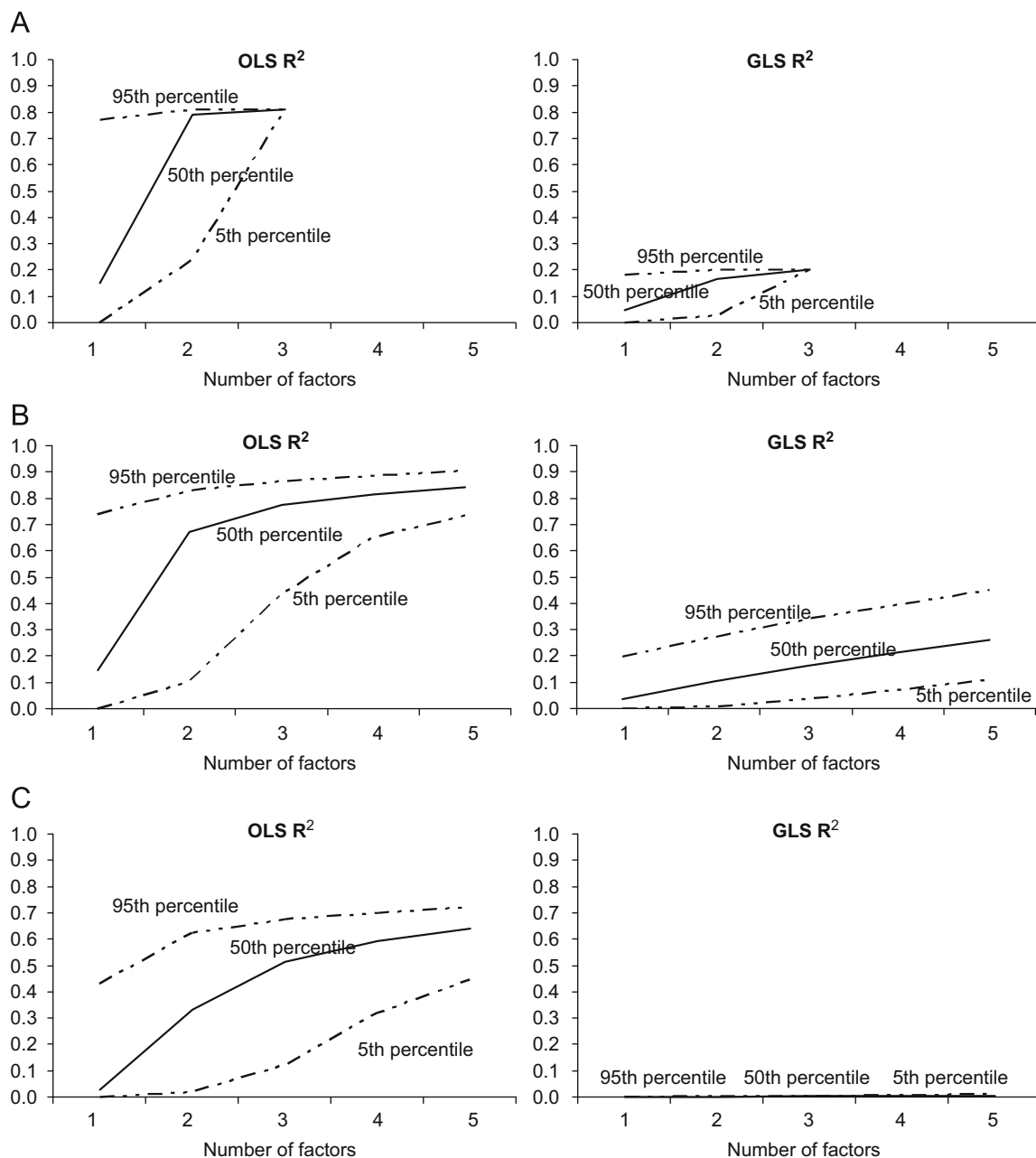
A

B

C

Fig. 4. Population OLS and GLS R^2s for artificial factors. We randomly generate artificial factors and report the simulated distribution of population OLS and GLS R^2s when expected returns on the Fama-French 25 size-B/M portfolios are regressed on their artificial factor loadings (quarterly average returns, variances, and covariances from 1963 to 2004 are treated as population parameters in the simulations). In Panel A, macro factors are generated by randomly drawing a 3×1 vector of weights, w, from a standard normal distribution, defining a factor $P = F'w$, where $F = [R_M, SMB, HML]$. In Panel B, factors are constructed as zero-investment combinations of the size-B/M portfolios by randomly drawing a 25×1 vector of weights from a standard normal distribution. In Panel C, the procedure for Panel B is repeated but only factors with (roughly) zero expected returns are retained. The plots show the 5th, 50th, and 95th percentiles of simulated distributions based on 5,000 draws of up to five factors at a time.

test asset, is similar to the time-series approach of Black, Jensen, and Scholes (1972) and Gibbons, Ross, and Shanken (1989).

Prescription 5. Report confidence intervals for the cross-sectional R^2.

Prescription 5 is less a solution to the problems highlighted above—it does nothing to address the concern that many factor models will produce high population R^2s—than a way to make the sampling issues more transparent. We suspect researchers would put less weight on the cross-sectional R^2 if the extremely high

sampling error in it was clear (extremely high when using size-B/M portfolios, though not necessarily with other assets). More generally, we find it surprising that papers often emphasize this statistic with little regard for its sampling properties (exceptions include Jagannathan, Kubota, and Takehara, 1998; Bansal, Dittmar, and Lundblad, 2005).

The distribution of the sample R^2 can be derived analytically in special cases but we are not aware of a general formula that incorporates first-stage estimation error in factor loadings. An alternative is to use simulations like those in Fig. 2, one panel of which is repeated in Fig. 5. By inspection, the simulations indicate that the sample OLS R^2 is often significantly biased and skewed by an amount that depends on the true cross-sectional R^2. These properties suggest that reporting a confidence interval for R^2 is more meaningful than reporting just a standard error.

The easiest way to get confidence intervals is to 'invert' Fig. 5, an approach suggested by Stock (1991) in a different context. In the figure, the sample distribution of the estimated R^2, for a given true R^2, is found by slicing the picture along the x-axis (fixing x, then scanning up and down). Conversely, a confidence interval for the true R^2, given a sample R^2, is found by slicing the picture along the y-axis (fixing y, then scanning across). For example, a sample R^2 of 0.50 implies a 90% confidence interval for the population R^2 of roughly [0.25, 1.00], depicted by the dotted line in the graph. (Formally, the confidence interval represents all values of the true R^2 for which the estimated R^2 falls within the 5th and 95th percentiles of the sample distribution.) The extremely wide interval in this example illustrates just how uninformative the sample R^2 can be.

Prescription 6. Report confidence intervals for the weighted sum of squared pricing errors.
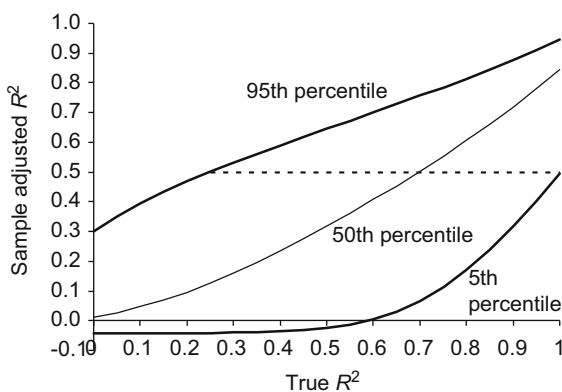

Fig. 5. Sample distribution of the cross-sectional adjusted R^2. This figure repeats the 'One factor' panel of Fig. 2. It shows the sample distribution of the cross-sectional adjusted R^2 as a function of the true cross-sectional R^2 for a model with one factor using Fama and French's 25 size-B/M portfolios as test assets (quarterly returns, 1963–2004). The simulated factor is a combination of the size-B/M portfolios (the weights are randomly drawn to produce the given true R^2, as described in Section 2). The plot is based on 40,000 bootstrap simulations (ten sets of simulated factors; 4,000 simulations with each).

Prescription 6 has the same goal as Prescription 5: to provide a better measure of how well a model performs. Again, Prescription 6 does not address our concern that it is easy to find factors that explain expected returns on size-B/M portfolios. But confidence intervals should at least make clear when a test has low power: we might not reject that a model works perfectly, but we also will not reject that the pricing errors are large. Conversely, confidence intervals can reveal when a model is rejected because the pricing errors are estimated precisely, not because they are large. In short, confidence intervals allow us to better assess the economic significance of the results.

The weighted sum of squared pricing errors is an alternative to the cross-sectional R^2 as a measure of performance. The literature offers several versions of such statistics, including Shanken's (1985) cross-sectional T^2 statistic, Gibbons, Ross, and Shanken's (GRS, 1989) F statistic, Hansen's (1982) J statistic, and Hansen and Jagannathan's (1997) HJ distance. Confidence intervals for any of these can be obtained using an approach similar to Fig. 5, plotting the sample distribution as a function of the true parameter. We describe here how to get confidence intervals for the GRS F test, the cross-sectional T^2 (or asymptotic χ^2) statistic, and the HJ distance, all of which have useful economic interpretations and either accommodate or impose restrictions on the zero-beta rate and risk premia.

The GRS F statistic tests whether the time-series intercepts are all zero when excess returns are regressed on a set of factor portfolios, $R=\alpha+BR_P+e$. Let $\hat{a}$ be the OLS estimate of α given a sample for T periods. The covariance matrix of $\hat{a}$, conditional on the realized factors, is $\Omega=c\Sigma$, where $\Sigma\equiv\text{var}(e)$, $c\equiv(1+s_P^2)/T$, and s_P^2 is the sample maximum squared Sharpe ratio attainable from combinations of P. GRS show that, under standard assumptions, the statistic $S=c^{-1}\hat{a}'\Sigma_{OLS}^{-1}\hat{a}$, is asymptotically χ^2 and, if e is multivariate normal conditional on R_P, the statistic $F=S\times(T-N-K)/[N(T-K-1)]$ is small-sample F with non-centrality parameter $c^{-1}\alpha'\Sigma^{-1}\alpha$ and degrees of freedom N and $T-N-K$. The term $\alpha'\Sigma^{-1}\alpha$ equals the model's unexplained squared Sharpe ratio, the difference between the population squared Sharpe ratio of the tangency portfolio (θ_τ^2) and that attainable from P (θ_P^2). Thus, an exact confidence interval for $\theta_z^2\equiv\alpha'\Sigma^{-1}\alpha$ can be found by inverting a graph similar to Fig. 5, showing the sample distribution of F as a function of θ_z^2 (i.e., finding the set of θ_z^2 for which F is between the chosen percentiles of an F distribution with noncentrality parameter $c^{-1}\theta_z^2$).

Fig. 6 illustrates the confidence-interval approach for testing the unconditional CAPM. The test is based on FF's 25 size-B/M portfolios from 1963 to 2004, and our market proxy is the Center for Research in Security Prices (CRSP) value-weighted index. The size and B/M effects are strong during this sample (the average absolute quarterly alpha is 0.96% across the 25 portfolios), and the GRS F statistic, 3.49, strongly rejects the CAPM with a p-value of 0.000. The graph shows, moreover, that we can reject that the squared Sharpe ratio on the market is within 0.21 of the squared Sharpe ratio of the tangency portfolio: a 90% confidence interval for θ_z^2 is [0.21, 0.61]. Interpreted
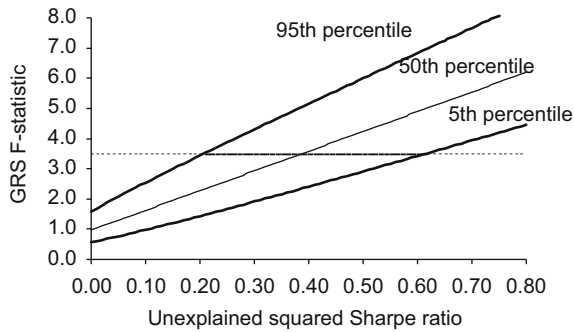
Fig. 6. Sample distribution of the GRS F statistic and confidence interval for θ_z^2. This figure provides a test of the CAPM using quarterly returns on Fama and French's 25 size-B/M portfolios from 1963 to 2004. The sample distribution of the F statistic, for a given value of the true unexplained squared Sharpe ratio, θ_z^2, is found by slicing the graph along the x-axis (fixing θ_z^2 then scanning up to find percentiles of the sample distribution). A confidence interval for θ_z^2, given the sample F statistic, is found by slicing along the y-axis (fixing F then scanning across). θ_z^2 is the squared Sharpe ratio of the tangency portfolio minus that of the CRSP value-weighted index. $F = c^{-1}\hat{a}'\Sigma_{OLS}^{-1}\hat{a}(T-N-1)/[N(T-2)]$, where $N=25$, $T=168$, and $c \approx 1/T$. It has an F distribution with noncentrality parameter $c^{-1}\theta_z^2$ and degrees of freedom N and $T-N-1$. The sample F is 3.49, and the corresponding 90% confidence interval for θ_z^2 is depicted by the dark dotted line.

differently, following MacKinlay (1995), there exists a zero-beta portfolio that, with 90% confidence, has a quarterly Sharpe ratio between 0.46 ($=0.21^{1/2}$) and 0.78 ($=0.61^{1/2}$). This compares with a quarterly Sharpe ratio for the market portfolio of 0.18 during this period. The confidence interval provides a good summary measure of just how poorly the CAPM works.

Shanken's (1985) T^2 test is like the GRS F test but focuses on pricing errors, or residuals, in the cross-sectional regression of expected returns on factor loadings, $\mu_R = z\imath + B\lambda + \alpha$. The T^2 test can be used with non-return factors and does not restrict the zero-beta rate to be the risk-free rate (unless the intercept is omitted). The test is based on the traditional two-pass methodology: Let b be the matrix of factor loadings estimated in the first-pass time-series regression and let $x = [\imath \; b]$ be regressors in the second-pass cross-sectional regression (with average returns as the dependent variable). The estimated pricing errors, $\hat{a}$, have asymptotic variance $\Sigma_a = (1 + \lambda'\Sigma_F^{-1}\lambda)y\Sigma y/T$, where $y \equiv I - x(x'x)^{-1}x'$ and the term $1 + \lambda'\Sigma_F^{-1}\lambda$ accounts for estimation error in b.[5] The T^2 statistic is then $T^2 \equiv \hat{a}'S_a^+\hat{a}$, where S_a^+ is the pseudoinverse of the estimated Σ_a based on consistent estimates of the parameters (the pseudoinverse is required because Σ_a is singular). Appendix A shows that T^2 is asymptotically χ^2 with degrees of freedom $N-K-1$ and noncentrality parameter $\alpha'\Sigma_a^+\alpha = \alpha'(y\Sigma y)^+\alpha[T/(1 + \lambda'\Sigma_F^{-1}\lambda)]$. The quadratic, $q \equiv \alpha'(y\Sigma y)^+\alpha$, again has an economic interpretation, in this case measuring how far factor-mimicking portfolios are from the mean-variance frontier. Specifically, let R_P be

K portfolios formed from the test assets that are maximally correlated with the factors, and let $\theta(z)$ be a 'generalized' Sharpe ratio, using the zero-beta rate, $r_f + z$, in place of the risk-free rate. The appendix shows that $q = \theta_\tau^2(z) - \theta_P^2(z)$, the difference between the maximum generalized squared Sharpe ratio on any portfolio and that attainable from R_P. (The zero-beta rate in this definition is the one that minimizes q and turns out to be the GLS zero-beta rate.) Therefore, as with the GRS F test, a confidence interval for q can be found by plotting the sample distribution of the T^2 statistic as a function of q, using either the asymptotic χ^2 distribution or a simulated small-sample distribution (a third possibility, applying the logic of Shanken, 1985, would be to replace the asymptotic χ^2 distribution with an approximate finite-sample F distribution).

The final test we consider, the HJ distance, focuses on SDF pricing errors, $\alpha = E[w(1+R)-1]$, where $w = g_0 + g_1'P$ is a proposed SDF and we now define R as an $N+1$ vector of total (not excess) returns including the riskless asset. Let m be any well-specified SDF. Hansen and Jagannathan (1997) show that the distance between w and the set of true SDFs, $D \equiv \min_m E[(w-m)^2]$, is the same as $\alpha'H^{-1}\alpha$, where $H \equiv E[(1+R)(1+R)']$ is the second-moment matrix of gross returns. To get a confidence interval for D, Appendix B shows that $D = \theta_z^2/(1+r_f)^2$, where θ_z^2 is the model's unexplained squared Sharpe ratio, as defined earlier (see also Kan and Zhou, 2004). Thus, like the GRS F statistic, the estimate of D is small-sample F up to a constant of proportionality (assuming that P consists of return factors). A confidence interval can then be obtained using the approach described above.

4. Empirical examples

The prescriptions above are relatively straightforward to implement and, while not a complete solution to the problems discussed in Section 2, should help to improve the power and rigor of empirical tests. As an illustration, we report tests for several models that have been proposed recently in the literature. Cross-sectional tests in the original studies focus on FF's size-B/M portfolios, precisely the scenario for which our concerns are greatest. Our goal here is neither to disparage the papers—indeed, we believe they provide economically important insights—nor to provide a full review of the often extensive empirical tests in each paper, but only to show that our prescriptions can dramatically change how well a model seems to work.

We investigate models for which data are readily available. The models include: (i) Lettau and Ludvigson's (LL, 2001) conditional consumption CAPM (CCAPM), in which the conditioning variable is the aggregate consumption-to-wealth ratio CAY (available on Ludvigson's website); (ii) Lustig and Van Nieuwerburgh's (LVN, 2004) conditional CCAPM, in which the conditioning variable is the housing collateral ratio MYMO (we consider only their linear model with separable preferences; MYMO is available on Van Nieuwerburgh's website;); (iii) Santos and Veronesi's (SV, 2006) conditional CAPM, in which the

---

[5] Appendix A explains these results in detail. The variance Σ_a is derived under the null that $\alpha=0$. Also, Shanken (1985) analyzes GLS, not OLS, cross-sectional regressions. Our appendix shows that the OLS-based test described here is equivalent to his GLS-based test.

Table 1
Empirical tests of asset pricing models, 1963–2004.

The table reports slopes, Shanken t-statistics (in parentheses), and other statistics from cross-sectional regressions of average excess returns on estimated factor loadings for eight asset pricing models. Returns are quarterly, in percent. The test assets are Fama and French's 25 size-B/M portfolios used alone or together with their 30 industry portfolios. The OLS R^2 is an adjusted R^2. The cross-sectional T^2 statistic tests whether pricing errors in the cross-sectional regression are all zero, with simulated p-values in brackets; it is proportional to the distance, q, that a model's mimicking portfolios are from the minimum-variance boundary, measured as the difference between the maximum generalized squared Sharpe ratio and that attainable from the mimicking portfolios. The sample estimate of q is reported in the final column. Ninety-five percent confidence intervals for the true R^2s and q are reported in brackets next to the sample values. The models are estimated from 1963 to 2004 except Yogo's, for which we have factor data through 2001. The variables are: cay=Lettau and Ludvigson's (2001) consumption-to-wealth ratio; Δc=log consumption growth; my=Lustig and Van Nieuwerburgh's (2004) housing-collateral ratio based on mortgage data; R_M=CRSP value-weighted excess return; s^w=labor income to consumption ratio; ΔI_{HH}, ΔI_{Corp}, ΔI_{Ncorp}=log investment growth for households, nonfinancial corporations, and the noncorporate sector, respectively; Δc_{Ndur}, Δc_{Dur}=Yogo's (2006) log consumption growth for nondurables and durables, respectively; SMB, HML=Fama and French's (1993) size and B/M factors.

Model/assets	Variables				OLS R^2	GLS R^2	T^2	q
LL (2001)	**Const**	**cay**	**Δc**	**cay × Δc**				
FF25	3.33 (2.28)	−0.81 (−1.25)	0.25 (0.84)	0.00 (0.42)	0.58 [0.30, 1.00]	0.05 [0.00, 0.50]	33.9 [p=0.08]	0.44 [0.00, 0.72]
FF25+30 ind.	2.50 (3.29)	−0.48 (−1.23)	0.09 (0.53)	−0.00 (−1.10)	0.00 [0.00, 0.35]	0.01 [0.00, 0.20]	193.8 [p=0.00]	1.31 [0.18, 1.08]
LVN (2004)	**Const**	**my**	**Δc**	**my × Δc**				
FF25	3.58 (2.22)	4.23 (0.76)	0.02 (0.04)	0.10 (1.57)	0.57 [0.35, 1.00]	0.02 [0.00, 0.35]	20.8 [p=0.57]	0.45 [0.00, 0.48]
FF25+30 ind.	2.78 (3.51)	0.37 (0.13)	−0.02 (−0.09)	0.03 (1.40)	0.09 [0.00, 1.00]	0.00 [0.00][a]	157.1 [p=0.04]	1.32 [0.00, 0.96]
SV (2006)	**Const**	**R_M**	**$s^w × R_M$**					
FF25	3.07 (1.96)	−0.95 (−0.58)	−0.21 (−2.06)		0.27 [0.00, 1.00]	0.02 [0.00, 0.40]	26.0 [p=0.63]	0.46 [0.00, 0.30]
FF25+30 ind.	2.57 (2.77)	−0.49 (−0.44)	−0.09 (−1.99)		0.08 [0.00, 1.00]	0.02 [0.00, 0.40]	160.8 [p=0.07]	1.31 [0.00, 0.72]
LVX (2006)	**Const**	**ΔI_{HH}**	**ΔI_{Corp}**	**ΔI_{Ncorp}**				
FF25	2.47 (2.13)	−0.80 (−0.39)	−2.65 (−1.03)	−8.59 (−1.96)	0.80 [0.75, 1.00]	0.26 [0.05, 1.00]	25.2 [p=0.29]	0.34 [0.00, 0.48]
FF25+30 ind.	2.22 (3.14)	0.20 (0.19)	−0.93 (−0.58)	−5.11 (−2.32)	0.42 [0.20, 1.00]	0.04 [0.00, 0.55]	141.2 [p=0.11]	1.27 [0.00, 0.84]
Yogo (2006)	**Const**	**Δc_{Ndur}**	**Δc_{Dur}**	**R_M**				
FF25	1.98 (1.36)	0.28 (1.00)	0.67 (2.33)	0.48 (0.29)	0.18 [0.00, 1.00]	0.04 [0.00, 0.55]	24.9 [p=0.69]	0.46 [0.00, 0.30]
FF25+30 ind.	1.95 (2.27)	0.18 (1.01)	0.19 (1.52)	0.12 (0.11)	0.02 [0.00, 0.60]	0.05 [0.00, 1.00]	159.3 [p=0.06]	1.24 [0.00, 0.78]
CAPM	**Const**	**R_M**						
FF25	2.90 (3.18)	−0.44 (−0.39)			−0.03 [0.00, 0.55]	0.01 [0.00, 0.25]	77.5 [p=0.00]	0.46 [0.12, 0.48]
FF25+30 ind.	2.03 (2.57)	0.10 (0.09)			−0.02 [0.00, 0.35]	0.00 [0.00, 0.05]	225.2 [p=0.00]	1.34 [0.18, 0.96]
Cons. CAPM	**Const**	**Δc**						
FF25	1.70 (2.47)	0.24 (0.89)			0.05 [0.00, 1.00]	0.01 [0.00, 0.50]	60.6 [p=0.01]	0.46 [0.06, 0.66]
FF25+30 ind.	2.07 (3.51)	0.03 (0.15)			−0.02 [0.00, 0.65]	0.00 [0.00][a]	224.5 [p=0.00]	1.34 [0.18, 1.02]
Fama-French	**Const**	**R_M**	**SMB**	**HML**				
FF25	2.99 (2.33)	−1.42 (−0.98)	0.80 (1.70)	1.44 (3.11)	0.78 [0.60, 1.00]	0.19 [0.05, 0.65]	56.1 [p=0.00]	0.37 [0.06, 0.42]
FF25+30 ind.	2.21 (2.14)	−0.49 (−0.41)	0.60 (1.24)	0.87 (1.80)	0.31 [0.00, 0.90]	0.06 [0.05, 0.35]	200.4 [p=0.00]	1.24 [0.12, 0.90]

[a] The model's GLS R^2 falls below the 2.5th percentile of the sampling distribution for all values of the true GLS R^2, i.e., the estimated GLS R^2 is unusually small given any true R^2.

conditioning variable is the labor income-to-consumption ratio s^w; (iv) Li, Vassalou, and Xing's (LVX, 2006) investment model, in which the factors are investment growth rates for households (ΔI_{HH}), nonfinancial corporations (ΔI_{Corp}), and the noncorporate sector (ΔI_{Ncorp}) (we consider only this version of their model); and (v) Yogo's (2006) durable-consumption CAPM, in which the factors are the market return, R_M, and the growth rates in durable and nondurable consumption, Δc_{Dur} and Δc_{Ndur} (we consider only his linear model; the consumption series are available on Yogo's website). For comparison, we also report results for three benchmark models: the unconditional CAPM, the unconditional CCAPM, and FF's three-factor model.

Table 1 reports cross-sectional regressions for all eight models. The tests use quarterly excess returns (in percent) from 1963 to 2004 and highlight our suggestions in Section 3. Specifically, we compare results using FF's 25 size-B/M portfolios alone (FF25) with results for the expanded set of 55 portfolios that includes their 30 industry portfolios (FF25+30 ind). Our choice of industry portfolios is based on the notion that they should provide, in any reasonable sense, a fair test of the models (in contrast to, say, momentum portfolios whose returns seem to be anomalous relative to any of the models). We report OLS regressions supplemented by the GLS R^2, the cross-sectional T^2 statistic, and the sample estimate of the statistic q described earlier, equal to the difference between the maximum generalized squared Sharpe ratio and that attainable from a model's mimicking portfolios (q is zero if the model fully explains the cross section of expected returns).

Confidence intervals for the true values of q and the cross-sectional R^2s are obtained using the approach described in Section 3. For the R^2, we simulate the distribution of the sample R^2 for true R^2s between 0.0 and 1.0 and invert plots like Fig. 5. The simulations are similar to those in Figs. 2 and 5, with the actual factors

for each model now used in place of artificial factors.[6] We also use simulations to get a confidence interval for q, instead of relying on asymptotic theory, because the length of the time series in our tests (168 quarters) is small relative to the number of test assets (25 or 55). The confidence interval for q is based on the T^2 statistic because q determines the noncentrality parameter of T^2's (asymptotic) distribution. Thus, we simulate the distribution of the T^2 statistic for various values of q and replicate the tests in Fig. 6, with q playing the same role as θ_z^2 in the GRS F test. The p-value we report for the T^2 statistic also comes from these simulations, with $q=0$.

Table 1 shows four key results. First, adding industry portfolios dramatically changes the performance of the models, in terms of slope estimates, cross-sectional R^2s, and T^2 statistics. Compared to regressions with only size-B/M portfolios, the slopes estimated using all 55 portfolios are almost always closer to zero and the cross-sectional R^2s drop substantially. The adjusted OLS R^2 drops from 58% to 0% for LL's model, 57% to 9% for LVN's model, 27% to 8% for SV's model, 80% to 42% for LVX's model, and 18% to 3% for Yogo's model. In addition, for these five models, the T^2 statistics are insignificant in tests with size-B/M portfolios but reject, or nearly reject, the models using the expanded set of 55 portfolios. The performance of FF's three-factor model is similar to the other five—it has an R^2 of 78% for the size-B/M portfolios and 31% for all 55 portfolios—while the simple and consumption CAPMs have small adjusted R^2s for both sets of test assets.

The second key result is that the sample OLS R^2 is often uninformative about a model's true population performance. Our simulations show that, across the five main models in Table 1, a 95% confidence interval for the true R^2 has an average width greater than 0.70 using either set of test assets. In regressions with size-B/M portfolios, we cannot reject that all models work perfectly, but we also cannot reject that the true R^2s are quite small, with an average lower bound for the confidence intervals of 0.28. (LVX's model is an outlier, with a lower bound of 0.75.) In tests with all 55 portfolios, four of the five confidence intervals include 0.00 and the fifth includes 0.20—that is, using just the sample R^2, we cannot reject that the models have essentially no explanatory power. Two of the confidence intervals cover the entire range of R^2s from 0.00 to 1.00. The table suggests that sampling variation in the R^2 is just too large to use it as a reliable metric of performance.

The third key result is that none of the models provides much improvement over the simple or consumption CAPM when performance is measured by the GLS R^2 or q. This is true even for tests with size-B/M portfolios, for which OLS R^2s are quite high, and is consistent with our view that the GLS R^2 provides a more rigorous hurdle than the OLS R^2. The average GLS R^2 is only 0.08 across the five models using size-B/M portfolios and 0.02 using the full set of 55 portfolios (compared with GLS R^2s of 0.00–0.02 for the simple and consumption CAPMs). Just as important, confidence intervals for the true GLS R^2 typically rule out values close to one. Across the five models, the average upper bound for the true GLS R^2 is 0.56 for the size-B/M portfolios and 0.43 for all 55 portfolios (all but one of the confidence intervals include 0.00).

The distance q is closely related to the GLS R^2 and, not surprisingly, suggests similar conclusions. It can be interpreted as the maximum generalized squared Sharpe ratio (defined relative to the optimal zero-beta rate) on a portfolio that is uncorrelated with the factors, equal to zero if the model is well specified. For the size-B/M portfolios, the sample q is 0.46 for the simple and consumption CAPMs, dropping to 0.44 for LL's model, 0.45 for LVN's model, 0.46 for SV's model, 0.34 for LVX's model, and 0.46 for Yogo's model. Adding the 30 industry portfolios, the CAPM has a sample q of 1.34, compared with 1.31, 1.32, 1.31, 1.27, and 1.24 for the other models. Confidence intervals for the true q are generally quite wide, so even when we cannot reject that q is zero, we also cannot reject that q is large. Again, this is true even for the size-B/M portfolios, for which the models seem to perform well if we narrowly focus on the T^2 statistic's generally large p-values under the null.

Finally, in the spirit of taking seriously the cross-sectional parameters (Prescription 2), the table shows that none of the models explains the level of expected returns: the estimated intercepts are all substantially greater than zero. The regressions use excess quarterly returns, so the intercepts can be interpreted as the estimated quarterly zero-beta rates over and above the risk-free rate. Annualized, the zero-beta rates range from 7.8% to 14.3% *above the risk-free rate*. These estimates cannot reasonably be attributed to differences in lending versus borrowing costs.

Most of the other parameters in Table 1 are not pinned down precisely by theory, making it difficult to test restrictions on the slopes. The key exception is that both factors in SV's conditional CAPM are portfolio returns— $s^w R_M$ can be interpreted as a dynamic portfolio, provided the riskfree asset is traded—implying that the risk premia should equal the factors' expected excess returns (Lewellen and Nagel, 2006, provide an alternative interpretation of this constraint and discuss restrictions on the slopes in conditional consumption CAPM models like those of LL and LVN). The point estimates for SV's model, using either set of portfolios, are far from the factors' average excess returns during the sample (1.53% for R_M and -0.01% for $s^w R_M$), suggesting that unrestricted regressions might significantly overstate the model's explanatory power. More formally, if we impose the restrictions (i.e., we require the model to price R_M and $s^w R_M$), the OLS and GLS

---

[6] The only other difference is that, to simulate data for different true cross-sectional R^2s, we keep the true factor loadings the same in all simulations, equal to the historical estimates, and change the vector of true expected returns to give the right R^2. Specifically, expected returns in the simulations equal $\mu = h(C\lambda) + \varepsilon$, where C is the estimated matrix of factor loadings for a model, λ is the estimated vector of cross-sectional slopes, h is a scalar constant, and ε is randomly drawn from a N[0, σ_ε^2] distribution. The constants h and σ_ε are chosen to give the right cross-sectional R^2 and to maintain the historical cross-sectional dispersion in average returns. In principle, the simulations should consider how the distribution of the sample R^2 changes as a function of all unknown parameters that affect its distribution, not just as a function of the true R^2. Our approach of varying just the true R^2, given estimates of the other parameters, provides an approximate confidence interval.

R^2s become negative for both sets of portfolios, even with an intercept in the regression. The T^2 statistic jumps from 26.0 to 89.2 for the size-B/M portfolios and from 160.8 to 233.7 for the full set of 55 portfolios (the p-value drops from 0.63 to 0.00 in the first case and from 0.07 to 0.00 in the second). The confidence interval for q, closely related to the T^2 statistic, extends from 0.16 to 0.72 for the size-B/M portfolios and from 0.16 to 1.12 for the full set of portfolios. The R^2s are lower, of course, if we force the zero-beta rate to be the risk-free rate, and the T^2 statistics increase further, to 96.9 and 240.1 (again with p-values of 0.00).[7]

In sum, despite the seemingly impressive ability of the models to explain cross-sectional variation in average returns on size-B/M portfolios, none of the models performs very well once we expand the set of test portfolios, consider the GLS R^2 and confidence intervals for the true R^2s and q statistics, or ask the models to price the risk-free asset and, in the case of SV's model, the factor portfolios.

5. Conclusion

Our basic conclusion is that asset pricing models should not be judged by their success in explaining average returns on size-B/M portfolios (or, more generally, on portfolios for which a couple of factors are known to explain most of the time-series and cross-sectional variation in returns). High cross-sectional explanatory power for size-B/M portfolios, in terms of high R^2 or small pricing errors, is simply not a sufficiently high hurdle by which to evaluate a model. In addition, the sample cross-sectional R^2 and other common test statistics do not appear to be very informative about the true (population) performance of a model, at least in our tests with size, B/M, and industry portfolios.

The problems we highlight are not just sampling issues, though the sample properties of test statistics do make them worse. In population, if returns have a covariance structure like that of size-B/M portfolios, true expected returns will line up with true factor loadings so long as a proposed factor is correlated with returns only through the variation captured by the two or three common components in returns. The problems are also not solved by using an SDF approach, since SDF tests are very similar to traditional cross-sectional regressions.

The paper offers four main suggestions for improving empirical tests. First, because the problems are tied to the strong covariance structure of size-B/M portfolios, one suggestion is to include other portfolios in the tests, for example, portfolios sorted by industry or factor loadings.

Second, because the problems are exacerbated by the fact that empirical tests often neglect theoretical restrictions on the cross-sectional intercept and slopes, another suggestion is to take their magnitudes seriously when theory provides appropriate guidance. Third, because the problems appear to be less severe for GLS regressions, a partial solution is to report the GLS R^2 in addition to, or instead of, the OLS R^2. Last, because the problems are exacerbated by sampling issues, our fourth suggestion is to report confidence intervals for cross-sectional R^2s and other test statistics using the techniques described in the paper. Together, these prescriptions should help to improve the power and informativeness of empirical tests, though they clearly do not provide a perfect solution.

The paper contributes to the cross-sectional asset pricing literature in a number of additional ways: (i) we provide a novel interpretation of the GLS R^2 in terms of the relative mean-variance efficiency of factor-mimicking portfolios, building on the work of Kandel and Stambaugh (1995); (ii) we show that the cross-sectional T^2 statistic based on OLS regressions is equivalent to that from GLS regressions (identical in sample except for the Shanken-correction terms), and we show that both are a transformation of the GLS R^2; (iii) we derive the asymptotic properties of the cross-sectional T^2 statistic when the true pricing errors differ from zero and provide an economic interpretation of the noncentrality parameter; and (iv) we describe a way to obtain confidence intervals for measures of model misspecification based on the GRS F statistic, cross-sectional T^2 statistic, and HJ distance, in addition to confidence intervals for the cross-sectional R^2. These results are helpful for understanding cross-sectional asset pricing tests.

Appendix A

This appendix derives the asymptotic distribution of the cross-sectional T^2 statistic, provides an economic interpretation of the distribution's noncentrality parameter, and discusses the connection between the T^2 statistic and the GLS R^2.

Let R_t be the vector of excess returns on N test assets and F_t be a vector of K factors in period t. Both are assumed, in this appendix, to be IID over time. The assets' factor loadings are estimated in the first-pass time-series regression, $R_t=c+BF_t+e_t$, and the relation between expected returns and B is estimated in the second-pass cross-sectional regression, $E[R]=z\iota+B\gamma+\alpha$, or more compactly, $\mu=X\lambda+\alpha$, where $\mu\equiv E[R_t]$, $X\equiv[\iota\ B]$, $\lambda'\equiv[z\ \gamma']$, and α is the vector of true pricing errors. To be precise, λ and α depend on whether we are considering OLS or GLS regressions: The true OLS slope is $\lambda=(X'X)^{-1}X'\mu$ and pricing errors are $\alpha=y\mu$, where $y\equiv I-X(X'X)^{-1}X'$; the true GLS slope is $\lambda^*=(X'V^{-1}X)^{-1}X'V^{-1}\mu$ and pricing errors are $\alpha^*=y^*\mu$, where $y^*\equiv I-X(X'V^{-1}X)^{-1}X'V^{-1}$ and $V\equiv\text{var}(R_t)$. In practice, of course, the cross-sectional regression is estimated with average returns substituted for μ and estimates of B substituted for the true loadings.

We begin with a few population results that are useful for interpreting empirical tests. We omit the time subscript until we turn to sample statistics.

Result 1. *The cross-sectional slope and pricing errors in a GLS regression are the same if V is replaced by $\Sigma \equiv var(e)$. Thus, we use V and Σ interchangeably in the GLS results below depending on which is more convenient for the issue at hand.*

Proof. See Shanken (1985). The result follows from $(X'V^{-1}X)^{-1}X'V^{-1} = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}$. □

Recall that $\alpha = y\mu$ and that $\alpha^* = y^*\mu$. The quadratics $q = \alpha'[y\Sigma y]^+\alpha$ and $q^* = \alpha^{*'}[y^*\Sigma y^{*'}]^+\alpha^*$, where a superscript $+$ denotes a pseudoinverse, are important for interpreting the cross-sectional T^2 test. The analysis below uses the facts, easily confirmed, that y and $\Sigma^{-1}y^*$ are symmetric, y and y^* are idempotent ($y = yy$ and $y^* = y^*y^*$), $y = yy^*$, $y^* = y^*y$, and $yX = y^*X = 0$.

Result 2. *The quadratics q and q^* are unchanged if Σ is replaced by V. Together with Result 1, this result implies that the T^2 statistic, from either OLS or GLS, is the same regardless of which covariance matrix is used.*

Proof. $yX = y^*X = 0$ implies that $yB = y^*B = 0$. Result 2 then follows from the fact $V = B\Sigma_F B' + \Sigma$, so $yV = y\Sigma$ and $y^*V = y^*\Sigma$. □

Result 3. *The OLS and GLS quadratics are identical, i.e., $q = q^*$.*

Proof. The quadratics are defined as $q = \alpha'[y\Sigma y]^+\alpha$ and $q^* = \alpha^{*'}[y^*\Sigma y^{*'}]^+\alpha^*$. Using the definition of a pseudo-inverse, it is straightforward to show that $[y\Sigma y]^+ = \Sigma^{-1}y^*$ and $[y^*\Sigma y^{*'}]^+ = \Sigma^{-1}(y^*)^+$, implying that $q = \alpha'\Sigma^{-1}y^*\alpha$ and $q^* = \alpha^{*'}\Sigma^{-1}(y^*)^+\alpha^*$.[8] In addition, $\alpha^* = y^*\alpha$ and $\Sigma^{-1}y^* = y^{*'}\Sigma^{-1}$, from which it follows that $q^* = \alpha'y^{*'}\Sigma^{-1}(y^*)^+y^*\alpha = \alpha'\Sigma^{-1}y^*(y^*)^+y^*\alpha = \alpha'\Sigma^{-1}y^*\alpha = q$. □

Result 4. *The OLS and GLS quadratics simplify to $q^* = \alpha^{*'}\Sigma^{-1}\alpha^* = \alpha^{*'}V^{-1}\alpha^*$. This result implies that our cross-sectional T^2 statistic matches that of Shanken (1985).*

Proof. Result 3 shows that $q^* = \alpha'\Sigma^{-1}y^*\alpha$. Recall that y^* is idempotent, $\Sigma^{-1}y^*$ is symmetric, and $\alpha^* = y^*\alpha$. Therefore, $q^* = \alpha'\Sigma^{-1}y^*y^*\alpha = \alpha'y^{*'}\Sigma^{-1}y^*\alpha = \alpha^{*'}\Sigma^{-1}\alpha^*$. □

Mimicking portfolios for the factors are the K portfolios, R_P, maximally correlated with F. The weights defining R_P can be interpreted as slopes in the regression $F = k + w_P' R + s$, where $cov(R, s) = 0$ (we ignore the constraint that $w_P' \iota = \iota$ for simplicity; the weights can be scaled up or down to make the constraint hold without changing the substance of any results). Thus, $w_P = V^{-1}cov(R, F) = V^{-1}B\Sigma_F$ and stocks' loadings on the mimicking portfolios are $C = cov(R, R_P)\Sigma_P^{-1} = Vw_P\Sigma_P^{-1} = B\Sigma_F\Sigma_P^{-1}$.

Result 5. *The cross-sectional regression (OLS or GLS) of μ on B is equivalent to the cross-sectional regression of μ on C, with or without an intercept, in the sense that the intercept,*

---

[8] More precisely, $\Sigma^{-1}(y^*)^+$ is a generalized inverse of $y^*\Sigma y^{*'}$, though not necessarily the pseudoinverse (the pseudoinverse of A is such that $AA^+A = A$, $A^+AA^+ = A^+$, and A^+A and AA^+ are symmetric; a generalized inverse ignores the two symmetry conditions). It can be shown that using $\Sigma^{-1}(y^*)^+$ in the quadratic q^* is equivalent to using the pseudoinverse.

R^2, *pricing errors, and quadratics q and q^* are the same in both.*

Proof. The first three claims, that the intercept, R^2, and pricing errors are the same, follow directly from the fact that C is a nonsingular transformation of B. The final claim, that the quadratics are the same regardless of whether we use F or R_P, follows from the fact that the pricing errors are the same and the quadratics can be based on V, i.e., $q^* = \alpha^{*'}V^{-1}\alpha^*$, where V is invariant to the set of factors. □

Result 6. *The GLS regression of μ on B or μ on C, with or without an intercept, prices the mimicking portfolios perfectly, i.e., $\alpha_P^* = w_P' \alpha^* = 0$. It follows that the slopes on C equal $\mu_P - z^*\iota$, the expected return on the mimicking portfolio in excess of the GLS zero-beta rate (for this last result, we assume that w_P is scaled to make $w_P' \iota = \iota$).*

Proof. From the discussion prior to Result 5, $w_P = V^{-1}B\Sigma_F$, implying that $\alpha_P^* = \Sigma_F B'V^{-1}\alpha^* = \Sigma_F B'V^{-1}y^*\mu$. Further, $X'V^{-1}y^* = 0$, from which it follows that $B'V^{-1}y^* = 0$ and, hence, $\alpha_P^* = 0$. This proves the first half of the result. Also, by definition, $\alpha^* = \mu - z^*\iota - C\gamma_P^*$, where γ_P^* are the GLS slopes on C. Therefore, $\alpha_P^* = w_P' \alpha^* = \mu_P - z^*\iota - \gamma_P^* = 0$, where $w_P' \mu$ and $C_P = w_P' C = I_K$. Solving for γ_P^* proves the second half of the result. □

Result 7. *Pricing errors in a GLS cross-sectional regression of μ on C are identical to the intercepts in a time-series regression of $R - z^*\iota$ on $R_P - z^*\iota$. It follows that q and q^* equal $\theta_\tau^2(z^*) - \theta_P^2(z^*)$, where $\theta(z^*)$ is an asset's generalized Sharpe ratio with respect to $r_f + z^*$ (the asset's expected return in excess of $r_f + z^*$ divided by its standard deviation), τ is the tangency portfolio with respect to $r_f + z^*$, and θ_P is the maximum squared generalized Sharpe ratio attainable from R_P.*

Proof. Intercepts in the time-series regression are $\alpha_{TS} = \mu - z^*\iota - C(\mu_P - z^*\iota)$. From Result 6, these equal α^* since $\gamma'_P = \mu_P - z^*\iota$. The interpretation of the quadratics then follows immediately from the well-known interpretation of $\alpha'_{TS}\Sigma^{-1}\alpha_{TS}$ (Jobson and Korkie, 1982; Gibbons, Ross, and Shanken, 1989), with the only change that the Sharpe ratios need to be defined relative to $r_f + z^*$. □

Result 8. *The GLS R^2 equals $1 - q/Q = 1 - q^*/Q$, where $Q = (\mu - \mu_{gmv}\iota)'V^{-1}(\mu - \mu_{gmv}\iota)$ and μ_{gmv} is the expected return on the global minimum variance portfolio (Q depends only on asset returns, not the factors being tested). Further, the GLS R^2 is zero if and only if the factors' mimicking portfolios all have expected returns equal to μ_{gmv} (i.e., they lie exactly in the middle of mean-variance space), and the GLS R^2 is one if and only if some combination of the mimicking portfolios lies on the minimum-variance boundary.*

Proof. The GLS R^2 is defined as $1 - \alpha^{*'}V^{-1}\alpha^*/(\mu - z_{nf}\iota)'V^{-1}(\mu - z_{nf}\iota)$, where z_{nf} is the GLS intercept when μ is regressed on a constant. The first claim in Result 8 follows from observing that $\alpha^{*'}V^{-1}\alpha^*$ is the same as q^* (see Result 4) and z_{nf} is the same as μ_{gmv}. The second claim, which we state without further proof, is a multifactor generalization of results in Kandel and Stambaugh (1995), with mimicking portfolios substituted for non-return factors. The key fact is that $q^* = Q - (\mu_P - \mu_{gmv}\iota)'\Sigma_P^{*-1}(\mu_P - \mu_{gmv}\iota)$, where Σ_P^*

is the residual covariance matrix when R_P is regressed on the R_{gmv}. Thus, q^* is zero (the GLS R^2 is one) only if some combination of R_P lies on the minimum-variance boundary, and q^* equals Q (the GLS R^2 is zero) only if $\mu_P = \mu_{gmv}\iota$.　□

Together, Results 1–8 describe key properties of GLS cross-sectional regressions, prove the equality of the OLS and GLS quadratics q and q^*, and establish the connections among the location of R_P in mean-variance space, the GLS R^2, and the quadratics. All of the results have exact parallels in sample, redefining population moments as sample statistics.

Our final results consider the asymptotic properties of the cross-sectional T^2 statistic under the null that pricing errors are zero and generic alternatives that they are not. The T^2 statistic is, roughly speaking, the sample analog of the quadratics q and q^* based on the traditional two-pass methodology. Let r be the vector of average returns, b be the sample estimate of B, and v and S be the usual estimates of V and Σ. The corresponding estimates of X, y, and y^* are $x \equiv [\iota\ b]$, $\hat{y} \equiv I - x(x'x)^{-1}x'$, and $\hat{y}^* \equiv I - x(x'v^{-1}x)^{-1}x'v^{-1}$. Therefore, the estimated OLS cross-sectional regression is $r = x\hat{\lambda} + \hat{a}$, where $\hat{\lambda} = (x'x)^{-1}x'r$ and $\hat{a} = \hat{y}r$, and the estimated GLS regression is $r = x\hat{\lambda}^* + \hat{a}^*$, where $\hat{\lambda}^* = (x'v^{-1}x)^{-1}x'v^{-1}r$ and $\hat{a}^* = \hat{y}^*r$. Equivalently, by substituting $r = (1/T)\Sigma_t R_t$ into the equations, the slope and pricing errors can be interpreted as time-series averages of period-by-period Fama-MacBeth estimates. We focus on OLS regressions in what follows but, as a consequence of the sample analog of Result 3 above, we show that the T^2 statistics from OLS and GLS are equivalent.

Our analysis below uses the following facts:

(1) $R_t = \mu + B\mathrm{UF}_t + e_t$, where $\mathrm{UF}_t = F_t - \mu_F$.
(2) $\mu = X\lambda + \alpha = x\lambda + (X-x)\lambda + \alpha = x\lambda + (B-b)\gamma + \alpha$.
(3) $B\mathrm{UF}_t = b\,\mathrm{UF}_t + (B-b)\mathrm{UF}_t$.
(4) $\hat{y}x = 0$ and $\hat{y}b = 0$.

Combining these facts, the pricing error in period t is $\hat{a}_t \equiv \hat{y}R_t = \hat{y}(B-b)\gamma + \hat{y}\alpha + \hat{y}(B-b)\mathrm{UF}_t + \hat{y}e_t$ and the time-series average is $\hat{a} = \hat{y}(B-b)\gamma + \hat{y}\alpha + \hat{y}(B-b)\overline{\mathrm{UF}} + \hat{y}\bar{e}_t$, where an upper bar denotes a time-series average. Asymptotically, $b \to B$, $\hat{y} \to y$, $\overline{\mathrm{UF}} \to 0$, and $\bar{e}_t \to 0$, where $\to$ denotes convergence in probability. These observations, together with $y\alpha = \alpha$, imply that $\hat{a}$ is a consistent estimator of α. Also, the second-to-last term, $\hat{y}(B-b)\overline{\mathrm{UF}}$, converges to zero at a faster rate than the other terms and, for our purposes, can be dropped: $\hat{a} = \hat{y}(B-b)\gamma + \hat{y}\alpha + \hat{y}\bar{e}_t$.

Result 9. *Define $d \equiv \hat{a} - \hat{y}\alpha$. Asymptotically, $T^{1/2}d$ converges in distribution to $N(0, T\Sigma_d)$, where $\Sigma_d = y\Sigma y(1+\gamma'\Sigma_F^{-1}\gamma)/T$.*

Proof. This result follows from observing that d is the same as $\hat{a}$ when $\alpha = 0$, the scenario considered by Shanken (1985, 1992b), and the term $(1+\gamma'\Sigma_F^{-1}\gamma)$ is the Shanken correction for estimation error in b. More formally, $d = \hat{y}(B-b)\gamma + \hat{y}\bar{e}_t$. The asymptotic distribution is the same substituting y for $\hat{y}$, and the two terms have asymptotic mean of zero and are uncorrelated with each other under the standard assumptions of OLS regres-

sions (i.e., in a regression, estimation error in the slopes is uncorrelated with the mean of the residuals). The asymptotic covariance is, therefore, $\Sigma_d = \mathrm{var}[y(B-b)\gamma] + \mathrm{var}[y\bar{e}_t]$. Let $\mathrm{vec}(b-B)$ be the $NK \times 1$ vector version of $b-B$, stacking the loadings for asset 1, then asset 2, etc., which has asymptotic variance $\Sigma \otimes \Sigma_F^{-1}/T$ from standard regression results. Rearranging and simplifying the formula for Σ_d, the first term becomes $\mathrm{var}[y(B-b)\gamma] = \gamma'\Sigma_F^{-1}\gamma(y\Sigma y)/T$ and the second term becomes $\mathrm{var}[y\bar{e}_t] = y\Sigma y/T$. Summing these gives the covariance matrix.　□

A corollary of Result 9 is that, under the null that $\alpha = 0$, $T^{1/2}\hat{a}$ also converges in distribution to $N(0, T\Sigma_d)$. To test whether $\alpha = 0$, the cross-sectional T^2 statistic is then naturally defined as $T^2 \equiv \hat{a}'S_d^+\hat{a}$, where S_d is the sample estimate of Σ_d substituting the statistics $\hat{y}$, S, $\hat{\gamma}$, and S_F for the population parameters y, Σ, γ, and Σ_F. Thus, $T^2 = \hat{a}[\hat{y}S\hat{y}]^+\hat{a}[T/(1+\hat{\gamma}'S_F^{-1}\hat{\gamma})]$. The key quadratic here, $\hat{q} = \hat{a}'[\hat{y}S\hat{y}]^+\hat{a}$, is the sample counterpart of q defined earlier. Result 3 implies that this OLS-based T^2 statistic is identical to a GLS-based T^2 statistic defined using $\hat{q}^*$, the sample equivalent of the GLS quadratic q^* (the T^2 statistics are identical assuming the same Shanken-correction term, $\hat{\gamma}'S_F^{-1}\hat{\gamma}$, is used for both; they are asymptotically equivalent under the null as long as consistent estimates of γ and Σ_F are used for both). Moreover, Result 4 implies that $T^2 = \hat{a}^{*'}S^{-1}\hat{a}^*[T/(1+\hat{\gamma}'S_F^{-1}\hat{\gamma})]$.

Result 10. *The cross-sectional T^2 statistic is asymptotically χ^2 with degrees of freedom $N-K-1$ and noncentrality parameter $q^*T/(1+\gamma'\Sigma_F^{-1}\gamma)$. Alternatively, from Result 7, the noncentrality parameter can be written as $[\theta_\tau^2(z^*) - \theta_P^2(z^*)]T/(1+\gamma'\Sigma_F^{-1}\gamma)$.*[9]

Proof. The cross-sectional pricing errors are $\hat{a} = d + \hat{y}\alpha = \hat{y}(d+\alpha)$, where the first equality follows from the definition of d and the second follows from the fact that $\hat{y}d = d$. The T^2 statistic, therefore, becomes $T^2 = (d+\alpha)'\hat{y}S_d^+\hat{y}(d+\alpha)$. Using the definition S_d^+ and facts from the proof of Result 3, it is straightforward to show that $\hat{y}S_d^+\hat{y} = S_d^+$ and, thus, $T^2 = (d+\alpha)'S_d^+(d+\alpha)$. S_d^+/T is a consistent estimate of Σ_d^+/T, so the T^2 statistic has the same asymptotic distribution as $(d+\alpha)'\Sigma_d^+(d+\alpha)$. [$S_d^+/T \to \Sigma_d^+/T$ because $S_d^+/T = S^{-1}\hat{y}'/(1+\hat{\gamma}'S_F^{-1}\hat{\gamma})$, which clearly converges to $\Sigma_d^+/T = \Sigma^{-1}y^*/(1+\gamma'\Sigma_F^{-1}\gamma)$.] From Result 9, $T^{1/2}d$ converges in distribution to $N(0, T\Sigma_d)$, where Σ_d has rank $N-K-1$. This implies that, asymptotically, $d'\Sigma_d^+d$ is central χ^2, while $(d+\alpha)'\Sigma_d^+(d+\alpha)$ is noncentral χ^2 with noncentrality parameter $\alpha'\Sigma_d^+\alpha$, both with degrees of freedom $N-K-1$. Result 10 then follows from observing that the noncentrality parameter equals $q^*T/(1+\gamma'\Sigma_F^{-1}\gamma)$.　□

---

[9] We use the terminology of a limiting distribution somewhat informally here because, as the result is stated, the noncentrality parameter goes to infinity as T gets large unless α and q^* are zero. The asymptotic result can be stated more formally by considering pricing errors that go to zero as T gets large: Suppose that $\alpha^* = T^{-1/2}\delta^*$, for some fixed vector δ^*. For this sequence of α^*, the T^2 statistic converges in distribution to a χ^2 with noncentrality parameter $\delta^{*'}\Sigma^{-1}\delta^*/(1+\gamma'\Sigma_F^{-1}\gamma)$.

Appendix B

This appendix derives the small-sample distribution of the HJ distance when returns are multivariate normal and the factors in the proposed model are portfolio returns (or have been replaced by maximally correlated mimicking portfolios). R is defined, for the purposes of this appendix, to be the $N+1$ vector of total rates of return on the test assets, including the riskless asset.

Let $w = g_0 + g'_1 R_P$ be a proposed SDF. The HJ distance is defined as $D \equiv \min_m E[(m-w)^2]$, where m represents any well-specified SDF, i.e., any variable for which $E[m(1+R)]=1$. Hansen and Jagannathan (1997) show that, if w is linear in asset returns (or is the projection of a non-return w onto the space of asset returns), the m^* that solves the minimization problem is linear in the return on the tangency portfolio, i.e., $m^* = v_0 + v_1 R_\tau$ for some constants v_0 and v_1, and $D = E[(m^*-w)^2]$.

The constants g_0 and g_1 are generally unknown and chosen to minimize $D = E[(m^*-g_0-g'_1 R_P)^2]$. This problem is simply a standard least-squares projection problem, so D turns out to be nothing more than the residual variance when m^* is regressed on a constant and R_P. Equivalently, D is v_1^2 times the residual variance when R_τ is regressed on a constant and R_P: $D = v_1^2 \mathrm{var}(\varepsilon)$, where ε is from the regression $R_\tau = s_0 + s'_1 R_P + \varepsilon$. Kandel and Stambaugh (1987) and Shanken (1987) show that the correlation between any portfolio and the tangency portfolio equals the ratio of their Sharpe ratios, θ_x/θ_τ. Thus, s_1 gives the combination of R_P that has the maximum squared Sharpe ratio, denoted θ_P^2, from which it follows that $D = v_1^2(1-\theta_P^2/\theta_\tau^2)\sigma_\tau^2$. The constant v_1 equals $-\theta_\tau/[\sigma_\tau(1+r_f)]$ (see, e.g., Cochrane, 2001), implying that the HJ distance is $D = (\theta_\tau^2-\theta_P^2)/(1+r_f)^2 = \theta_z^2/(1+r_f)^2$, where θ_z^2 can be interpreted as the proposed model's unexplained squared Sharpe ratio.

The analysis here is cast in terms of population parameters but equivalent results go through in sample. Thus, the estimated HJ distance, $\tilde{D}$, is proportional to the difference between the sample squared Sharpe ratios of the ex post tangency portfolio and the portfolios in R_P. Following the discussion in Section 4, the sample HJ distance is therefore proportional to the GRS F statistic: $\hat{D} = Fc/(1+r_f)^2[N(T-K-1)/(T-N-K)]$, where $c = (1+s_P^2)/T$. It follows immediately that, up to a constant of proportionality, $\hat{D}$ has an F distribution with noncentrality parameter $c^{-1}\theta_z^2 = c^{-1}(1+r_f)^2 D$.

References

Acharya, V., Pedersen, L., 2005. Asset pricing with liquidity risk. Journal of Financial Economics 77, 375–410.

Ahn, D.-H., Conrad, J., Dittmar, R., 2009. Basis assets. Review of Financial Studies 22, 5133–5174.

Ait-Sahalia, Y., Parker, J., Yogo, M., 2004. Luxury goods and the equity premium. Journal of Finance 59, 2959–3004.

Bansal, R., Dittmar, R., Lundblad, C., 2005. Consumption, dividends, and the cross section of equity returns. Journal of Finance 60, 1639–1672.

Black, F., Jensen, M.C., Scholes, M., 1972. The capital asset pricing model: some empirical findings. In: Jensen, M.C. (Ed.), Studies in the Theory of Capital Markets. Praeger, New York, pp. 79–121.

Brennan, M., 1971. Capital market equilibrium with divergent borrowing and lending rates. Journal of Financial and Quantitative Analysis 6, 1197–1205.

Brennan, M., Wang, A., Xia, Y., 2004. Estimation and test of a simple model of intertemporal asset pricing. Journal of Finance 59, 1743–1775.

Campbell, J., Cochrane, J., 1999. By force of habit: a consumption-based explanation of aggregate stock market behavior. Journal of Political Economy 107, 105–251.

Campbell, J., Vuolteenaho, T., 2004. Bad beta, good beta. American Economic Review 94, 1249–1275.

Cochrane, J., 1996. A cross-sectional test of an investment-based asset pricing model. Journal of Political Economy 104, 572–621.

Cochrane, J., 2001. Asset Pricing. Princeton University Press, Princeton, NJ.

Daniel, K., Titman, S., 2005. Testing factor-model explanations of market anomalies. Unpublished working paper. Northwestern University, Evanston, IL.

Duffee, G., 2005. Time variation in the covariance between stock returns and consumption growth. Journal of Finance 60, 1673–1712.

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

Ferson, W., Sarkissian, S., Simin, T., 1999. The alpha factor asset pricing model: a parable. Journal of Financial Markets 2, 49–68.

Gibbons, M., Ross, S., Shanken, J., 1989. A test of the efficiency of a given portfolio. Econometrica 57, 1121–1152.

Hansen, L., 1982. Large sample properties of generalized method of moments estimators. Econometrica 50, 1029–1054.

Hansen, L., Heaton, J., Li, N., 2006. Consumption strikes back? Measuring long-run risk. Unpublished working paper. University of Chicago, Chicago, IL.

Hansen, L., Jagannathan, R., 1997. Assessing specification errors in stochastic discount factor models. Journal of Finance 52, 557–590.

Heaton, J., Lucas, D., 2000. Asset pricing and portfolio choice: the role of entrepreneurial risk. Journal of Finance 55, 1163–1198.

Jacobs, K., Wang, K., 2004. Idiosyncratic consumption risk and the cross section of asset returns. Journal of Finance 59, 2211–2252.

Jagannathan, R., Kubota, K., Takehara, H., 1998. Relationship between labor-income risk and average return: empirical evidence from the Japanese stock market. Journal of Business 71, 319–347.

Jagannathan, R., Wang, Z., 1996. The conditional CAPM and the cross section of stock returns. Journal of Finance 51, 3–53.

Jobson, J.D., Korkie, B., 1982. Potential performance and tests of portfolio efficiency. Journal of Financial Economics 10, 433–466.

Kan, R., Zhang, C., 1999. Two-pass tests of asset pricing models with useless factors. Journal of Finance 54, 203–235.

Kan, R., Zhou, G., 2004. Hansen-Jagannathan distance: geometry and exact distribution. Unpublished working paper. University of Toronto, Toronto, Canada.

Kandel, S., Stambaugh, R., 1987. On correlations and inferences about mean-variance efficiency. Journal of Financial Economics 18, 61–90.

Kandel, S., Stambaugh, R., 1995. Portfolio inefficiency and the cross section of expected returns. Journal of Finance 50, 157–184.

Lettau, M., Ludvigson, S., 2001. Resurrecting the (C)CAPM: a cross-sectional test when risk premia are time-varying. Journal of Political Economy 109, 1238–1287.

Lewellen, J., Nagel, S., 2006. The conditional CAPM does not explain asset pricing anomalies. Journal of Financial Economics 82, 289–314.

Li, Q., Vassalou, M., Xing, Y., 2006. Sector investment growth rates and the cross section of equity returns. Journal of Business 89, 1637–1665.

Lustig, H., Van Nieuwerburgh, S., 2004. Housing collateral, consumption insurance, and risk premia. Journal of Finance 60, 1167–1221.

MacKinlay, A.C., 1995. Multifactor models do not explain deviations from the CAPM. Journal of Financial Economics 38, 3–28.

Mehra, R., Prescott, E., 1985. The equity premium: a puzzle. Journal of Monetary Economics 15, 145–161.

Nawalkha, S., 1997. A multibeta representation theorem for linear asset pricing theories. Journal of Financial Economics 46, 357–381.

Parker, J., Julliard, C., 2005. Consumption risk and the cross section of expected returns. Journal of Political Economy 113, 185–222.

Pastor, L., Stambaugh, R., 2003. Liquidity risk and expected stock returns. Journal of Political Economy 111, 642–685.

Petkova, R., 2006. Do the Fama-French factors proxy for innovations in predictive variables?. Journal of Finance 61, 581–612.

Reisman, H., 1992. Reference variables, factor structure, and the approximate multibeta representation. Journal of Finance 47, 1303–1314.

Roll, R., Ross, S., 1994. On the cross-sectional relation between expected returns and betas. Journal of Finance 49, 101–121.

Santos, T., Veronesi, P., 2006. Labor income and predictable stock returns. Review of Financial Studies 19, 1–44.

Shanken, J., 1985. Multivariate tests of the zero-beta CAPM. Journal of Financial Economics 14, 327–348.

Shanken, J., 1987. Multivariate proxies and asset pricing relations: living with the Roll critique. Journal of Financial Economics 18, 91–110.

Shanken, J., 1990. Intertemporal asset pricing: an empirical investigation. Journal of Econometrics 45, 99–120.

Shanken, J., 1992a. On the current state of the arbitrage pricing theory. Journal of Finance 47, 1569–1574.

Shanken, J., 1992b. On the estimation of beta-pricing models. Review of Financial Studies 5, 1–34.

Shanken, J., Zhou, G., 2007. Estimating and testing beta pricing models: alternative methods and their performance in simulations. Journal of Financial Economics 84, 40–86.

Stock, J., 1991. Confidence intervals for the largest autoregressive root in US macroeconomic time series. Journal of Monetary Economics 28, 435–459.

Vassalou, M., 2003. News related to future GDP growth as a risk factor in equity returns. Journal of Financial Economics 68, 47–73.

Yogo, M., 2006. A consumption-based explanation of expected stock returns. Journal of Finance 61, 539–580.