# A Skeptical Appraisal of Asset-Pricing Tests

**Jonathan Lewellen**
Dartmouth and NBER
jon.lewellen@dartmouth.edu

**Stefan Nagel**
Stanford and NBER
nagel_stefan@gsb.stanford.edu

**Jay Shanken**
Emory and NBER
jay_shanken@bus.emory.edu

This version: May 2008
First draft: May 2005

---

We are grateful to John Cochrane, Mathijs Cosemans, Wayne Ferson, Martin Lettau, Christopher Polk, Motohiro Yogo, an anonymous referee, and workshop participants at numerous universities and conferences for their helpful comments. We thank Ken French, Sydney Ludvigson, Stijn Van Nieuwerburgh, and Motohiro Yogo for providing data via their websites.

# A Skeptical Appraisal of Asset-Pricing Tests


## 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 $R^2$s and small pricing errors) in fact provides quite weak support for a model. We offer a number of suggestions for improving empirical tests and evidence that several proposed models don't work as well as originally advertised.

## 1. Introduction

The finance literature has proposed a wide variety of asset-pricing models in recent years, motivated, at least in part, by the well-known size and B/M effects in returns. The models suggest new risk factors that should be priced in expected returns, including labor income (Jagannathan and Wang, 1996; Heaton and Lucas, 2000; Jacobs and Wang, 2004), growth in GDP 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 proposed several new state variables to summarize macroeconomic conditions, including the consumption-to-wealth ratio (Lettau and Ludvigson, 2001), 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 sentiment reinforced by the fact that many of the new models have little in common economically with each other.

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 of the proposed models don't work as well as the original evidence suggested.

The heart of our critique is that the literature has often given itself a low hurdle to meet in claiming success: high cross-sectional $R^2$s (or low pricing errors) when average returns on the Fama-French 25 size-B/M portfolios are regressed on their factor loadings. This hurdle is low because the size-B/M portfolios are well-known to have a strong factor structure, i.e., Fama and French's (1993) three 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 $R^2$ is easy because almost any proposed factor is likely to produce betas that line up with expected returns – basically all that's 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 an SDF approach. Under the same conditions that give a high cross-sectional $R^2$, the true pricing errors in an SDF specification will be small or zero, 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 aren't important. Indeed, the covariance structure of size-B/M portfolios also implies that, even if we do find factors that don't explain any 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, while correlated with returns, are constructed to have zero true cross-sectional $R^2$s for the 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 $R^2$s reported in the literature aren't nearly as impressive as they might, on the surface, appear.

The obvious question is then: What can be done? How can we improve asset-pricing tests 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 expand the set of test assets to include other portfolios, sorted, for example, by industry, beta, or other characteristics. Second, since the problems are exacerbated by the fact that empirical tests often ignore theoretical restrictions on the cross-sectional slopes, another simple solution is to take the magnitude of the slopes seriously when theory provides appropriate guidance. For example, zero-beta rates should be close to the riskfree 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 GLS than for 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, as we explain in Section 2, it will suffer from the same problems as the OLS $R^2$ in some situations (as do other statistics like the Hansen-Jagannathan, 1997, distance).

Finally, since the problems are exacerbated by sampling issues, a fourth '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 doesn't work, or doesn't work better than existing theory, and then look for evidence to reject the null. (In event studies, for example, 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 don't 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 quite 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 tests and Section 4 applies the prescriptions to several recently proposed models. Section 5 concludes.

## 2. Interpreting asset-pricing tests

Our analysis uses the following notation: Let R be the vector of excess returns on N test assets (in excess of the riskfree rate) and F be a vector of K risk factors that perfectly explain expected returns on the assets, i.e., $\mu \equiv E[R]$ 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 the mean of F equals the cross-sectional risk premium on B, implying $\mu = B\ \mu_F$. Thus, our basic model is

$$R = B\ F + e, \tag{1}$$

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, e.g., 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, $\text{cov}(x, y) \equiv E[(x - \mu_x)(y - \mu_y)']$; i.e., the row dimension is determined by x and the column dimension is determined by y. We use $\iota$ 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 will serve as a benchmark but we won't be interested in it per se. Instead, we want to test a proposed model P consisting of J factors. The matrix of assets' factor loadings on P is denoted $C \equiv \text{cov}(R, P)\ \text{var}^{-1}(P)$, and we'll say that P 'explains the cross section of expected returns' if $\mu = C\ \gamma$ for some risk-premium vector $\gamma$. The premia would, ideally, 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:

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

where $\lambda$ 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 riskfree rate); (ii) $\lambda$ should be non-zero and may be restricted by theory; and (iii) $\alpha$ should be zero and the cross-sectional $R^2$ should be one. In practice, empirical tests often focus only 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 will appear 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 non-zero) correlation with F and explains little of the time-series variation in returns.*

Proof: The assumption that cov(e, P) = 0 implies cov(R, P) = B cov(F, P). Thus, stocks' loadings on P are linearly related to their loadings on F: $C \equiv \text{cov}(R, P)\,\text{var}^{-1}(P) = B\,\text{cov}(F, P)\,\text{var}^{-1}(P) = B\,Q$, where $Q \equiv \text{cov}(F, P)\,\text{var}^{-1}(P)$. It follows that $\mu = B\,\mu_F = C\,\lambda$, 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 doesn't seem to have much economic meaning in recent empirical applications. For example, in tests with size-B/M portfolios, we know that $R_M$, SMB, and HML (the model F in our notation) capture nearly all (more than 92%) of the time-series variation in returns, so the residual in R = B F + e is both small and largely idiosyncratic. In that setting, we don't 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 cov(e, P) wasn't close to zero. In turn, we're not at all surprised that many proposed models seem to 'explain' the cross-section of expected size and B/M returns about as well as $R_M$, SMB, and HML. The strong factor structure of size and B/M portfolios makes it likely that stocks' betas on almost any proposed factor will line up with their 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 pretty low hurdle to meet in claiming that a proposed model explains the size and B/M effects.

[1] This argument works cleanly if a proposed model has three factors. It should also apply when P has two factors since size-B/M portfolios all have multiple-regression market betas close to one (see Fama and French, 1993). In essence, the two-factor model of SMB and HML explains most of the cross-sectional variation of expected returns, so a proposed model really needs only two factors (as long as we ignore restrictions on the intercept).

It may be useful to point out that, in Observation 1, pricing errors for the proposed model are exactly zero, implying that 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 $R^2$; it also applies to formal statistics like Hansen's (1982) J-test, Shanken's (1985) $T^2$ statistic, and Hansen-Jagannathan's (1997) HJ-distance. We'll discuss 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 randomly chosen set of K assets perfectly explains the cross section of expected returns so long as the K assets aren't asked to price themselves (that is, the K assets aren't included on the left-hand side of the cross-sectional regression and the risk premia aren't 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 1 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 $\text{cov}(e, R_k) = \text{cov}(e, B_K F + e_K) = 0$. 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 APT assumption of a strict factor structure, any collection of K assets will work. Obtaining a high cross-sectional $R^2$ just isn't 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 K return factors aren't asked to price themselves, 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, $\lambda$ simplifies to $\mu_K/\text{cor}^2(R_K, F)$, which clearly doesn't equal $\mu_K$ unless $\text{cor}(R_K, F) = 1$ (the formula for $\lambda$ 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 $\lambda$ was taken seriously, i.e., $R_K$ would then price the cross section only if $\text{cor}(R_K, F) = 1$.

Observations 1 and 2 are rather special since, in order to get clean predictions, we've assumed 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 $\mu$ on C is expected to be J / K.*

Proof: By a 'generic sense,' we mean that we don't know anything about the proposed factors and, thus, 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). 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 = B Q, where Q is a K×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 $\mu$ on $\iota$ and C, the $R^2$ is $\mu'MC(C'MC)^{-1}C'M\mu / \mu'M\mu$. Substituting $\mu = B\,\mu_F$ and C = B Q, 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, then it should be easy to find other K-factor models that also explain the cross section of expected returns. The issue is a bit 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 aren't surprised at all if a proposed three-factor model explains the size and B/M effects about as well as the Fama-French 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 Fama-French factors,

[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 $E[q_i \mid Q'Q \text{ and } q_{j\neq i}] = 0$ 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 $E[Q(Q'Q)^{-1}Q'] = (1/K)\, E[tr(Q(Q'Q)^{-1}Q')\, I_K] = (1/K)\, E[tr(Q'Q(Q'Q)^{-1})]\, I_K = (J/K)\, I_K$.

since this requires that a single factor captures the pricing information in both SMB and HML. (We note again that size-B/M portfolios all have Fama-French 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.)

Figure 1, on the next page, illustrates these results using simulations with Fama and French'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 1963–2004 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, it focuses on explaining expected returns in population, not on sampling issues (which we consider later).

Each of the panels reports simulations using artificial factors to explain expected returns. The simulations in Panel A match the assumptions underlying 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 $P_i = w_i'F + \nu_i$, where $F = [R_M, \text{SMB}, \text{HML}]$ and $\nu_i$ 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 5,000 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 has the minimal property that it correlates at least somewhat with the common factors in returns (i.e., the FF factors) but, consistent with it being macroeconomic, is orthogonal to the idiosyncratic residuals of the size-B/M portfolios.

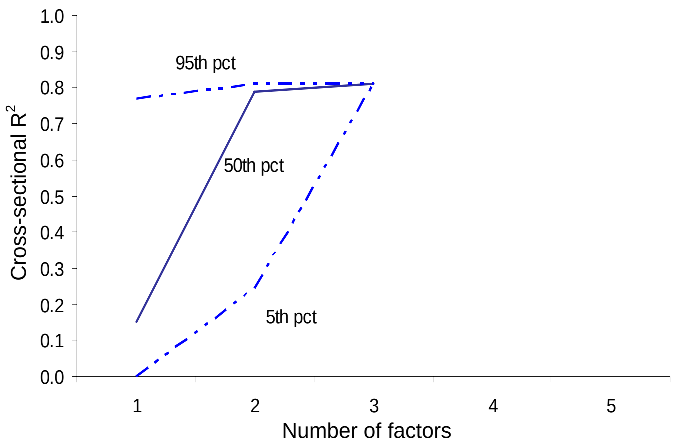
The figure suggests that it can be easy to find factors that help explain expected returns on the size-B/M portfolios. Taken individually, half of our artificial factors produce an OLS $R^2$ greater than 0.15 and 23% produce an $R^2$ greater than 0.50 (the latter isn't reported in the figure). Using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

In Panel B, we relax the assumption that our artificial factors are completely uncorrelated with the

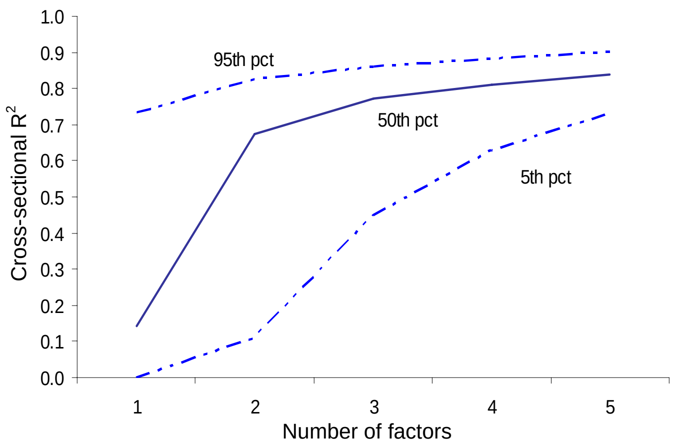
**Figure 1. Population $R^2$s for artificial factors.**

This figure explores how easy it is to find factors that explain, in population, the cross section of expected returns on Fama and French's 25 size-B/M portfolios. We randomly generate artificial factors, as described in each panel, and estimate the population $R^2$ when size-B/M portfolios' expected returns are regressed on their factor loadings. The quarterly average returns, variances, and covariances of the portfolios and Fama-French factors ($R_M$, SMB, and HML) from 1963–2004 are treated as population parameters in the simulations. The plots show the 5th, 50th, and 95th percentiles of simulated distributions based on 5,000 draws of up to 5 factors at a time.

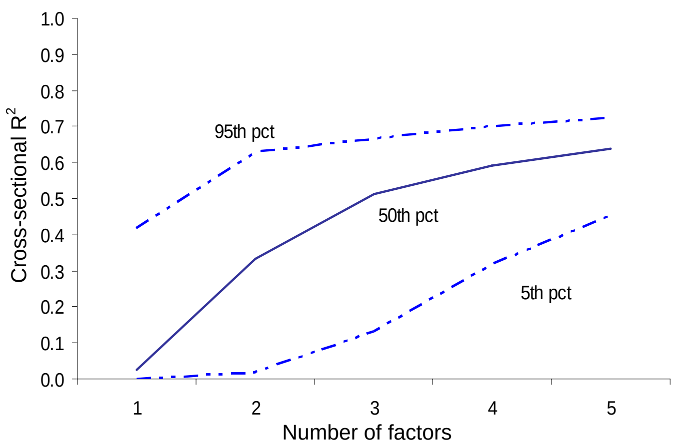
*Panel A: Random draws of macro factors.*
Macro factors, correlated with $R_M$, SMB, and HML but not with the residuals of size-B/M portfolios, are constructed by randomly drawing a 3×1 vector of weights from a standard normal distribution that defines a factor $P = F'w + \nu$, where $F = [R_M, \text{SMB}, \text{HML}]$ and $\nu$ is orthogonal to returns.




*Panel B: Random draws of factor portfolios.*
Zero-investment factors, formed from the size-B/M portfolios, are generated by randomly drawing a 25×1 vector of weights from a standard normal distribution.




*Panel C: Random draws of zero-mean factor portfolios.*
Zero-investment factors, formed from the size-B/M portfolios, are generated by randomly drawing a 25×1 vector of weights from a standard normal distribution, keeping only factors with roughly zero expected returns.



residuals of the 25 size-B/M portfolios. Specifically, we generate factors that are simply random combinations of the size-B/M portfolios: a 25×1 vector of weights is drawn from a N[0, 1] distribution, and the weights are then shifted and re-scaled to have a mean of zero and to have one dollar long and one dollar short. We repeat the simulations 5,000 times, with up to five factors in a model. 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 don't change), and these simulations explore how easy it is to stumble across mimicking portfolios that explain the cross section of expected returns. The simulations also show how special, in a sense, the Fama–French 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; 10% of the two-factor models, and 30% of the three-factor models, actually explain more cross-sectional variation than the Fama-French factors.[3]

Finally, Panel C repeats the simulations in Panel B with a small twist: we use only artificial factors, i.e., random 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 $R^2$s 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. With one, three, and five factors, the median $R^2$s in Panel C are 0.03, 0.52, and 0.64, while the 75th percentiles are 0.11, 0.60, and 0.68, respectively (the latter aren't shown in the figure). 24% 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 $R^2$s would be zero.

[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. But our simulations do indicate that the particular factors they constructed aren't unique in their ability to explain cross-section variation in expected returns on the size-B/M portfolios.

The results above 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.*

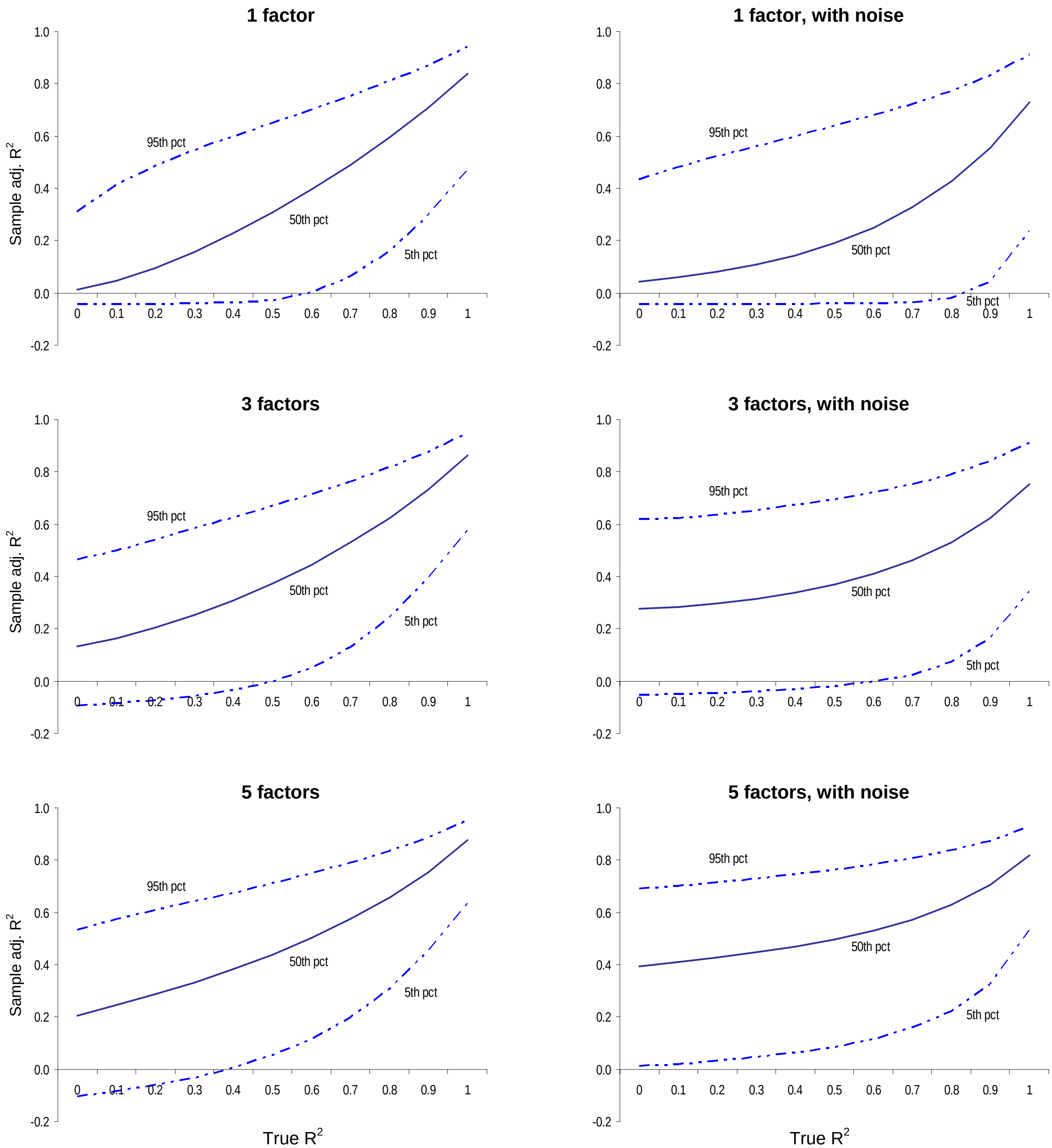
Observation 4 is intentionally informal and, in lieu of a proof, we offer simulations using Fama and French's 25 size-B/M portfolios to illustrate the point. The simulations differ from those in Figure 1 because, rather than study the population $R^2$ for artificial factors, we now focus on sampling variation in estimated $R^2$s *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 have 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 adj. $R^2$ for the artificial data by regressing average returns on estimated factor loadings. The second step is repeated 4,000 times to construct a sampling distribution of the adj. $R^2$. In addition, to make sure the particular matrix of loadings generated in step 1 isn't important, we repeat that step 10 times, giving us a total sample of 40,000 adj. $R^2$s corresponding to an assumed true $R^2$.

Figure 2 shows results for models with 1, 3, and 5 factors. The left-hand column plots the distribution of the sample adjusted $R^2$ (5th, 50th, and 95th percentiles) corresponding to true $R^2$s of 0.0 to 1.0 for models in which the factors are portfolio returns, as 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 3/4 of their total 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.

[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 $\mu_s$ 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 $\mu_s$, and to make $\mathrm{cov}(g_i, g_j) = 0$ for $i \neq j$). The factor portfolios in the simulations have covariance with returns (a 25×1 vector) given by the $g_j$.

**Figure 2. Sample distribution of the cross-sectional adj. $R^2$.**

This figure shows the sample distribution of the cross-sectional adj. $R^2$ (average returns regressed on estimated factor loadings) for Fama and French's 25 size-B/M portfolios from 1963–2004 (quarterly returns). The plots use one to five randomly generated factors that together have the true $R^2$ reported on the x-axis. In the left-hand panels, the factors are combinations of the size-B/M portfolios (the weights are randomly drawn to produce the given $R^2$, as described in the text). In the right-hand panels, noise is added to the factors equal to 3/4 of a factor's total variance, to simulate factors that are not perfectly spanned by returns. The plots are based on 40,000 bootstrap simulations (10 sets of random factors; 4,000 simulations with each).



The figure shows that a sample $R^2$ needs to be quite high to be statistically significant, especially for models with several factors. Focusing on the right-hand column, the 95th percentile of the sampling distribution using one factor is 44%, using three factors is 62%, and using five factors is 69% – 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 still wouldn't be terribly surprising to find fairly high $R^2$s in sample. Further, with multiple factors, the ability of the sample $R^2$ to discriminate between good and bad models is quite small, since the distribution of the sample $R^2$ is similar across a wide range of true $R^2$s. 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 5% 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 $R^2$ doesn't seem to be a very useful metric for distinguishing among models.

*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 the cross-sectional $R^2$ in simple CAPM tests isn't 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 on assets with covariance structure like that of the size-B/M portfolios – is quite different. The closest overlap comes from our simulations in Panel C of Figure 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 'useless' factors, defined as factors that are independent of asset returns. They show that regressions with useless factors can be misleading because the usual 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 $R^2$s and apply to factors that are correlated with asset returns.

Ferson, Sarkissian, and Simin (1999) also critique asset-pricing tests, emphasizing how difficult it may 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, unrelated to risk. Our analysis is different since 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, Reisman, 1992; Shanken, 1992a). Most closely, Nawalkha (1997) derives results like Observations 1 and 2 above, 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 (like 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. How can we improve empirical 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 $R^2$ or small pricing errors often has little economic meaning and, in our view, should not be taken as providing 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 tests often focus on size-B/M portfolios due to the importance of the size and value anomalies. 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 the size-B/M portfolios. One simple solution, then, is to include portfolios that don't correlate as strongly with SMB and HML. Reasonable choices include industry-, beta-, volatility-, or loading-sorted portfolios (the last being loadings on the proposed factor; an alternative would be to use individual stocks, though errors-in-variables problems could make this impractical, or statistically based portfolios like those described by Ahn, Conrad, and Dittmar, 2007). 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 don't 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 regressions, or in risk, on the right-hand side. One is necessary, not both.

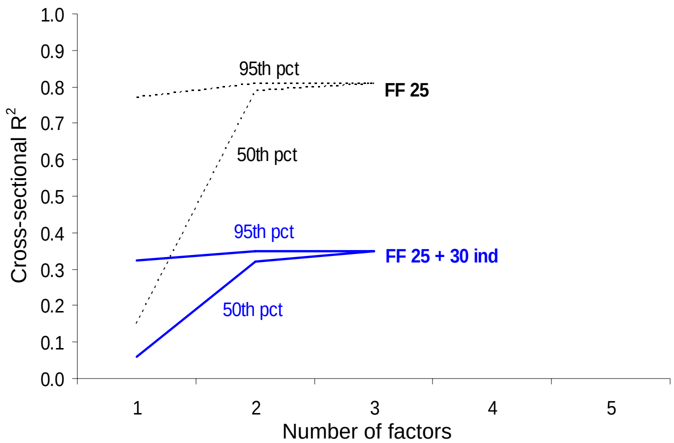
Second, we acknowledge the concern that no model is perfect and can be expected to explain all patterns in the data. This truism makes it tempting to view a model as successful if it explains even one or two anomalies, like the size and B/M effects. The problem with this view, however, is that tests with size-B/M portfolios alone just don't provide a sufficient test of a model, for all of the reasons discussed in Section 2. It doesn't seem practical to judge a model as 'successful' if it only works 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 don't provide a meaningful way to distinguish among the theories (though some of our suggestions below can help).

Figure 3 illustrates the benefits of using additional portfolios as test assets. We replicate the simulations in Figure 1 but, rather than use size-B/M portfolios alone, we augment them with Fama and French's 30 industry portfolios. As before, we explore how well artificial factors explain, in population, the cross section of expected returns (average returns and covariances from 1963–2004 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 $P_i = w_i'F + \nu_i$, where $F = [R_M, SMB, HML]$ and $\nu_i$ 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. And 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 cross-sectional $R^2$ (in population). We refer the reader to Section 2 for a more detailed description of the simulations.

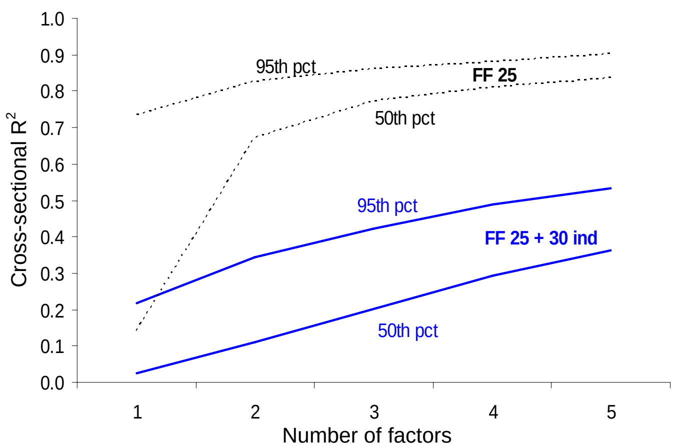
**Figure 3. Population $R^2$s for artificial factors: Size-B/M and industry portfolios.**

This figure compares how easy it is to find factors that explain, in population, the cross section of expected returns on Fama and French's 25 size-B/M portfolios (dotted lines) vs. 55 portfolios consisting of the size-B/M portfolios and Fama and French's 30 industry portfolios (solid lines). We randomly generate artificial factors, as described in each panel, and estimate the population $R^2$ when portfolios' expected returns are regressed on their factor loadings. The quarterly average returns, variances, and covariances of the portfolios and Fama-French factors ($R_M$, SMB, and HML) from 1963–2004 are treated as population parameters in the simulations. The plots show the 5th, 50th, and 95th percentiles of simulated distributions based on 5,000 draws of up to 5 factors at a time.

*Panel A: Random draws of macro factors.*
Macro factors, correlated with $R_M$, SMB, and HML but not with the residuals of size-B/M-industry portfolios, are constructed by randomly drawing a 3×1 vector of weights from a standard normal distribution that defines a factor $P = F'w + \nu$, where $F = [R_M, \text{SMB}, \text{HML}]$ and $\nu$ is orthogonal to returns.




*Panel B: Random draws of factor portfolios.*
Zero-investment factors, formed from the size-B/M-industry portfolios, are generated by randomly drawing a 55×1 vector of weights from a standard normal distribution.




*Panel C: Random draws of zero-mean factor portfolios.*
Zero-investment factors, formed from the size-B/M portfolios, are generated by randomly drawing a 55×1 vector of weights from a standard normal distribution, but only factors with roughly zero expected returns are kept.

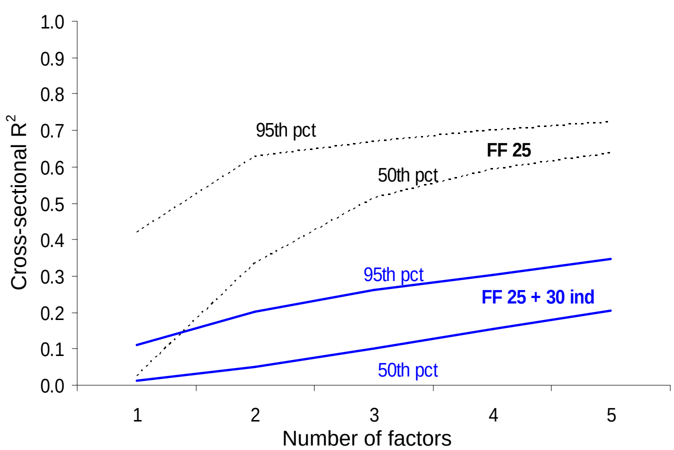


Figure 3 shows that it is much 'harder,' using artificial factors, to explain expected returns on the 55 portfolios than on the 25 size-B/M portfolios (the median and 95th percentiles for the latter are repeated from Fig. 1 for comparison). For example, for models with three factors, the median $R^2$ for the full set of 55 portfolios is 35% in Panel A, 21% in Panel B, and 11% in Panel C, compared with median $R^2$s for the 25 size-B/M portfolios of 81%, 78%, and 52%, respectively. The difference between the 25 size-B/M portfolios and the full set of 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.

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

The literature sometimes emphasizes a model's high cross-sectional $R^2$ but doesn't 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 the zero-beta rate should equal the riskfree rate. The standard retort is that Brennan's (1971) model relaxes this constraint if borrowing and lending rates differ, but this argument isn't convincing in our view: (riskless) borrowing and lending rates just aren't 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's unreasonable to ask a consumption-based model to explain it. But it isn't clear why we should accept a model that can't explain the level of expected returns.

A related restriction, mentioned earlier, is that the risk premium for any factor portfolio should be the portfolio's expected excess return. For example, the cross-sectional price of market-beta risk should be the market equity premium; the price of interest-rate risk, captured by movements in long-term Tbond returns, should be the expected Tbond return over the riskfree rate. In practice, this type of restriction could be tested in cross-sectional regressions or, better yet, 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 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 $\beta_t$ and $\gamma_t$ are the conditional beta and equity premium, respectively, and $\beta$ and $\gamma$ are their unconditional means. The cross-sectional slope on $\varphi_i = cov(\beta_{it}, \gamma_t)$, in the unconditional regression, should clearly be one but that constraint is often ignored 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 (i) the statistical properties of feasible GLS and (ii) 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 (i.e., if $\mu$ and B are expected returns and loadings for the test assets, OLS regresses $\mu$ on $[\iota\ B]$ while GLS regresses $V^{-1/2}\mu$ on $V^{-1/2}[\iota\ B]$, where $V = var(R)$). We believe these concerns are misplaced, or at least overstated, and that GLS actually has a number of advantages over OLS.

The statistical concerns with GLS are real but not prohibitive. The main issue is that, since 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 when the number of assets is large relative to the length of the time series (Gibbons, Ross, and Shanken, 1989, provide examples in a closely related context; see also Shanken and Zhou, 2007). But we see little reason this problem can't be overcome using simulation methods or, in special cases, the finite-sample results of Shanken (1985) or Gibbons et al.

The second 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 little connection, in general, 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 may not be 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 the cost of capital 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.

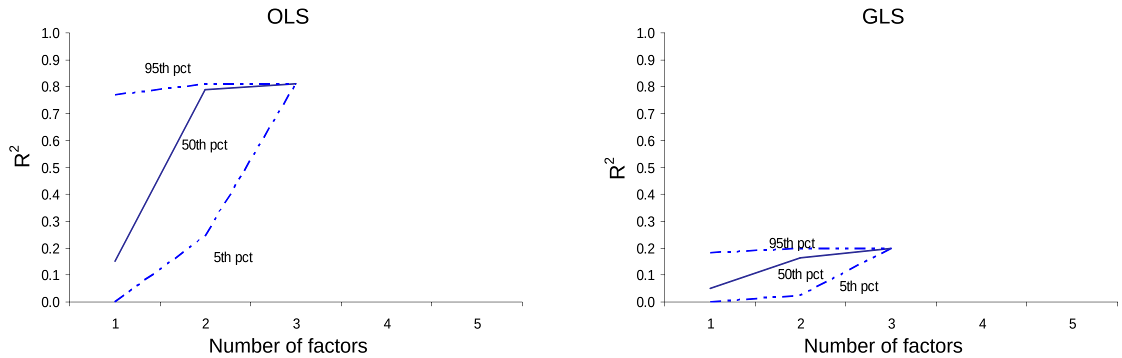
The same ideas apply to models with non-return factors. In this case, Appendix A shows that a GLS regression is equivalent to using maximally-correlated mimicking portfolios in place of the actual factors *and* imposing the constraint that the risk premia on the portfolios equal their excess returns (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 GLS $R^2$ is also closely linked to formal asset-pricing tests like Shanken's (1985) cross-sectional test of linearity and the Hansen-Jagannathan (1987) distance (see the appendix and 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 the factor mimicking portfolios are far from mean-variance efficient, while the GLS $R^2$ is high only if a model can explain the high Sharpe ratios available on the test portfolios.

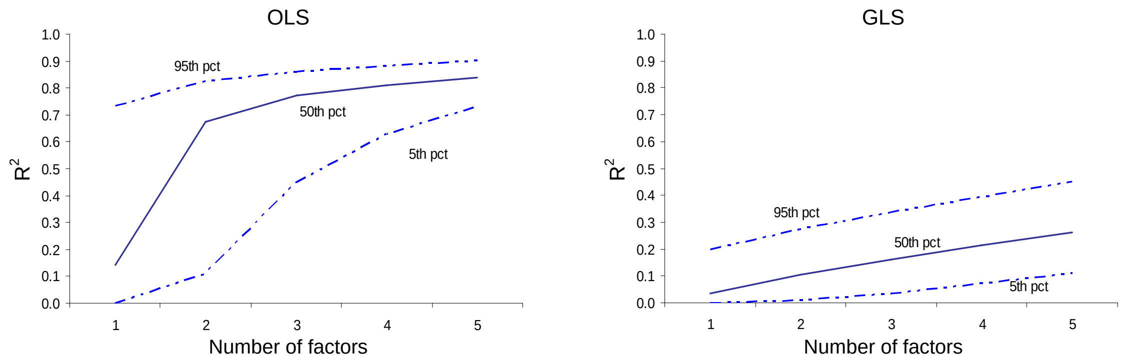
The implicit restrictions imposed by GLS aren't a full solution to the problems discussed in Section 2. Indeed, Observations 1 and 2 apply equally to OLS and GLS regressions: both $R^2$s are one given the stated assumptions. But our simulations with artificial factors, which relax the strong assumptions of the formal propositions, suggest that, in practice, finding a high population GLS $R^2$s is much less likely than finding a high OLS $R^2$. Figure 4, on the next page, illustrates this result. The figure shows GLS $R^2$s for the same simulations as Figure 1, using artificial factors to explain expected returns on Fama-French's size-B/M portfolios (treating their sample moments 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 $R^2$s are dramatically lower than OLS $R^2$s. 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 $R^2$s are appropriately zero, since the risk premia on the factors match their zero expected returns, while the OLS $R^2$ are often 50% or more in models with multiple factors.

## Figure 4. Population OLS and GLS $R^2$s for artificial factors.

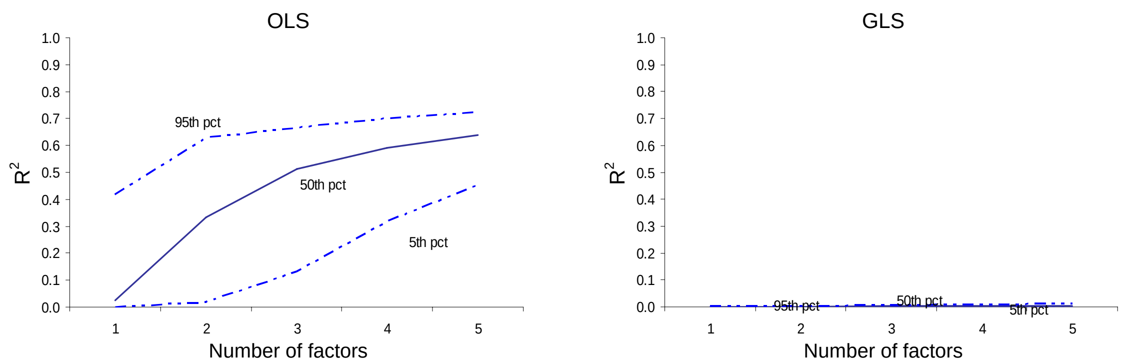
This figure explores how easy it is to find factors that explain, in population, the cross section of expected returns on Fama and French's 25 size-B/M portfolios. We randomly generate artificial factors, as described in each panel, and estimate the population OLS and GLS $R^2$s when size-B/M portfolios' expected returns are regressed on their factor loadings. The quarterly average returns, variances, and covariances of the portfolios and Fama-French factors ($R_M$, SMB, and HML) from 1963–2004 are treated as population parameters in the simulations. The plots are based on 5,000 draws of up to 5 factors at a time.




***Panel A: Random draws of macro factors***. Macro factors, correlated with $R_M$, SMB, and HML but not with the residuals of the 25 size-B/M portfolios, are constructed by randomly drawing a 3×1 vector of weights from a standard normal distribution that defines a factor $P = F'w + \nu$, where $F = [R_M, \text{SMB}, \text{HML}]$ and $\nu$ is orthogonal to returns.




***Panel B: Random draws of factor portfolios***. Zero-investment factors, formed from the size-B/M portfolios, are generated by randomly drawing a 25×1 vector of weights from a standard normal distribution.




***Panel C: Random draws of zero-mean factor portfolios***. Zero-investment factors, formed from the size-B/M portfolios, are generated by randomly drawing a 25×1 vector of weights from a standard normal distribution, but only factors with roughly zero expected returns are kept.

The advantage of GLS over OLS regressions in the simulations seems to come from two sources. The first, mentioned 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, the Fama-French factors explain much less cross-sectional variation in the size-B/M portfolios' expected returns in a GLS regression than in an OLS regression: 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 the factor as one of the test assets on the left-hand side of the cross-sectional regression.

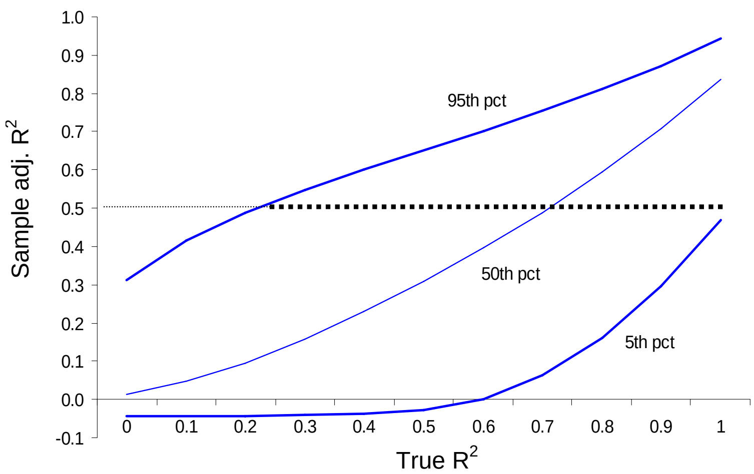
Prescription 4 builds on Prescription 2, in particular, the idea that the cross-sectional price of risk for a factor portfolio should be the factor's expected excess return. One simple way to build this restriction into a cross-sectional regression is to ask the factor to price itself; that is, to test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

Prescription 4 is most important when the cross-sectional regression is estimated with GLS rather than OLS. As mentioned above, when a factor portfolio is included as a left-hand side asset, GLS forces the regression to price the asset perfectly: the estimated slope on the factor's loading 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 test portfolio, 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 – indeed, it does nothing to address the concern that it may be easy to find factors that produce high population $R^2$s – 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 without regard to its sampling properties.

**Figure 5. Sample distribution of the cross-sectional adj. $R^2$.**
This figure repeats the '1 factor' panel of Fig. 2. It shows the sample distribution of the cross-sectional adj. $R^2$, 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 from 1963–2004 (quarterly returns). The simulated factor is a combination of the size-B/M portfolios (the weights are randomly drawn to produce the given $R^2$, as described in Section 2). The plot is based on 40,000 bootstrap simulations (10 sets of simulated factors; 4,000 simulations with each).




The distribution of the sample $R^2$ can be derived analytically in special cases but we're not aware of a general formula that incorporates first-stage estimation error in factor loadings. An alternative is to use simulations like those in Figure 2, one panel of which is repeated in Figure 5. The simulations indicate that the sample $R^2$ (OLS) 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' Figure 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 dark dotted line in the graph. 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.

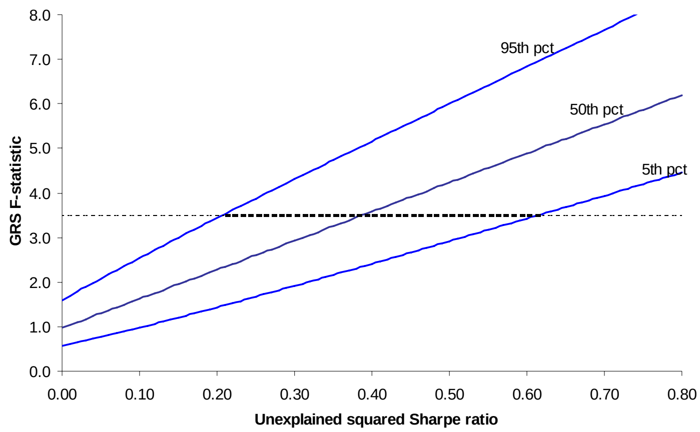
Prescription 6 has the same goal as Prescription 5: to provide a better measure of how well a model performs. Again, Prescription 6 doesn't address our concern that it's easy to find factors that produce small population pricing errors for size-B/M portfolios. But confidence intervals should at least make clear when a test has low power – we may not reject that a model works perfectly, but we also won't reject that the pricing errors are quite large. And confidence intervals can also reveal the opposite, that a model is rejected because the pricing errors are precisely estimated, 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 like Figure 4, 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 $\chi^2$) 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 time-series intercepts are zero when excess returns are regressed on a set of factor portfolios, $R = \alpha + B\, R_P + e$. (The F-test is valid only if the factors are portfolio returns.) Let $\hat{a}$ be the OLS estimate of $\alpha$ 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 $\chi^2$ 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 $\alpha'\Sigma^{-1}\alpha$ is the model's unexplained squared Sharpe ratio, the difference between the population squared Sharpe ratio of the tangency portfolio ($\theta_\tau^2$) and that attainable from P ($\theta_P^2$). Thus, an exact confidence interval for $\theta_z^2 \equiv \alpha'\Sigma^{-1}\alpha$ can be found by inverting a graph like Fig. 5 showing the sample distribution of F as a function of $\theta_z^2$ (i.e., finding the set of $\theta_z^2$ for which F is between the chosen percentiles of an F-distribution with non-centrality parameter $c^{-1}\,\theta_z^2$).

Fig. 6 illustrates the confidence-interval approach for testing the unconditional CAPM. The test uses quarterly excess returns on Fama and French's 25 size-B/M portfolios from 1963–2004 and our market

**Figure 6. Sample distribution of the GRS F-statistic and confidence interval for $\theta_z^2$.**
This figure provides a test of the CAPM using quarterly excess returns on Fama and French's 25 size-B/M portfolios from 1963–2004. The sample distribution of the GRS F-statistic, for a given value of the true unexplained squared Sharpe ratio, $\theta_z^2$, can be found by slicing the graph along the x-axis (fixing $\theta_z^2$ then scanning up to find percentiles of the sample distribution). A confidence interval for $\theta_z^2$, given the sample F-statistic, is found by slicing along the y-axis (fixing F then scanning across). $\theta_z^2$ is the difference between the squared Sharpe ratio of the tangency portfolio and that of the CRSP value-weighted index. $F = c^{-1}\,\hat{a}'\,\Sigma_{OLS}^{-1}\,\hat{a}\,(T–N–1)/[N(T–2)]$; it has an F-distribution with noncentrality parameter $c^{-1}\theta_z^2$ and degrees of freedom N and T–N–1, where N=25, T=168, and $c \approx 1/T$. The sample F-statistic is 3.49 and the corresponding 90% confidence interval for $\theta_z^2$ is depicted by the dark dotted line.




proxy is the CRSP value-weighted index. The size and B/M effects are quite strong during this sample (the average absolute quarterly alpha is 0.96% across the 25 portfolios), and the GRS F-statistic strongly rejects the CAPM: F = 3.49 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 $\theta_z^2$ is [0.21, 0.61]. Interpreted differently, following MacKinlay (1995), there exists a portfolio z that is uncorrelated with the market and, 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 (residuals) in the cross-sectional regression, $\mu = z\,\iota + B\,\lambda + \alpha$. (The $T^2$ test can be used with non-return factors and doesn't restrict the zero-beta rate to be $r_f$, 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 $x = [\iota\ b]$ 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 $y = I - x(x'x)^{-1}x'$ 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 $\Sigma_a$ based on consistent estimates of $\lambda$, $\Sigma_F$, y, and $\Sigma$ (the pseudoinverse is required because $\Sigma_a$ is singular). Appendix A shows that $T^2$ is asymptotically $\chi^2$ 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 what we'll refer to as a generalized Sharpe ratio, using the zero-beta rate, $r_f + z$, in place of the riskfree 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; it 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 $\chi^2$ distribution or a simulated small-sample distribution.[6]

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 $\theta_z^2$ 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 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 focused on Fama and French's size-B/M portfolios, precisely the

---

[5] Appendix A explains these results in detail. The variance $\Sigma_a$ assumes that returns are IID over time and that $\alpha = 0$. Also, Shanken analyzes GLS, not OLS, cross-sectional regressions. The appendix shows that the OLS-based test described here is equivalent to his GLS-based test.

[6] A third possibility, applying the logic of Shanken (1985), would be to replace the asymptotic $\chi^2$ distribution with a finite-sample F distribution. Details are available on request.

scenario for which our concerns are greatest. Our goal here is not to disparage the papers – indeed, we believe the studies 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 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 ($\Delta I_{HH}$), non-financial corporations ($\Delta I_{Corp}$), and the non-corporate sector ($\Delta I_{Ncorp}$) (we consider only this version of their model); and (v) Yogo's (2006) durable–consumption CAPM, in which the factors are the growth in durable and non-durable consumption, $\Delta c_{Dur}$ and $\Delta c_{Non}$, and the market return ($R_M$) (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 consumption CAPM, and Fama and French's (FF 1993) three-factor model.

Table 1 reports cross-sectional regressions for the eight models. The tests use quarterly excess returns (in %) from 1963–2004 and highlight our suggestions in the previous section. Specifically, we compare results using Fama and French's 25 size-B/M portfolios alone ('FF25' in the table) 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 above, 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 $R^2$s are obtained using the approach described in Section 3. For the $R^2$, we simulate the distribution of the sample $R^2$ for true $R^2$s between 0.0 to 1.0 and invert plots like Figure 5; the simulations are similar to those in Figures 2 and 5, with the actual factors for each model now used in place of artificial factors.[7] We also use simulations to get a confidence interval for q, rather than rely 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 since q determines the non-centrality parameter of $T^2$'s (asymptotic) distribution. Thus, we simulate the distribution of the $T^2$ statistic for various values of q and invert a plot like Figure 6, with q playing the same role as $\theta_z^2$ 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 $R^2$s, and $T^2$ statistics. Compared to regressions using only size-B/M portfolios, the slope estimates are almost always closer to zero and the cross-sectional $R^2$s often drop substantially. The adj. OLS $R^2$ drops from 58% to 0% for LL's model, from 57% to 9% for LVN's model, from 27% to 8% for SV's model, from 80% to 42% for LVX's model, and from 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 adj. $R^2$s for both sets of test assets.

The second key result is that the sample OLS $R^2$ is often very 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 of 0.72 for the size-B/M portfolios and 0.75 for the expanded set of 55 portfolios. For regressions with size-B/M portfolios, we cannot reject that all models work perfectly, as expected, but neither can we reject that the true $R^2$s are quite small, with an average lower bound for the confidence intervals of 0.28. (Li, Vassalou, and Xing's model is an outlier, with a lower bound of 0.75.) For regressions 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 can't reject that the models have essentially no explanatory power. Two of the confidence intervals cover the entire range of $R^2$s 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

[7] The only other difference is that, to simulate data for different true cross-sectional $R^2$s, 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, $\lambda$ is the estimated vector of cross-sectional slopes, h is a scalar constant, and $\varepsilon$ is randomly drawn from a MVN[0, $\sigma_\varepsilon^2$ I] distribution; h and $\sigma_\varepsilon$ are chosen to give the right cross-sectional $R^2$ and to maintain the historical cross-sectional dispersion in expected returns.

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 either the GLS $R^2$ or q. This is true even for tests with size-B/M portfolios, for which OLS $R^2$s 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 $R^2$s 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 true $R^2$s close to one: across the five models, the average upper bound for the true GLS $R^2$ is 0.56 for the 25 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, compared with 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 simple and consumption CAPM q's are both 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 can't reject that q is zero, we also cannot reject that q is very 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 p-value 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 (in %), so the intercepts can be interpreted as the estimated quarterly zero-beta rates over and above the riskfree rate. Annualized, the zero-beta rates range from 7.8% to 14.3% *above the riskfree rate*. These estimates cannot reasonably be attributed to differences in lending vs. borrowing costs.

Most of the other parameters in Table 1 aren't 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 – implying that the risk premia should equal the factors' expected excess returns (in excess of the zero-beta rate if the intercept is freely estimated; see

Lewellen and Nagel, 2006, for an alternative interpretation of this constraint and a discussion of the predicted slopes in conditional CCAPM 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 may 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 $R^2$s 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 $R^2$s are, of course, lower if we also force the zero-beta rate to be the riskfree rate, and the $T^2$ statistics increase further, to 96.9 and 240.1 (again with p-values of 0.00).[8]

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 $R^2$s and q statistics, or ask the models to price the riskfree asset and, in the case of SV's model, the factor portfolios.

## 5. Conclusion

The main point of the paper is easily summarized: Asset-pricing models cannot 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 $R^2$, as well as more formal test statistics based on the weighted sum of squared pricing errors, are not 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, i.e., they are not solved by getting standard errors

[8] An alternative to imposing the slope constraints directly is just to include the two factors as additional test assets, as suggested in Prescription 4. The OLS results in this case are fairly similar to those in Table 1: the $R^2$s change by less than 0.01, while the $T^2$ statistics increase from 26.0 to 32.5 (p-value of 0.58) for the size-B/M portfolios and from 160.8 to 208.2 (p-value of 0.01) for all 55 portfolios. GLS regressions provide stronger evidence against the model because they implicitly force the regression to price the two factors, as discussed earlier, so they're equivalent to the tests just discussed that impose the slope constraints directly.

right (but sampling issues 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 key suggestions for improving empirical tests. First, since the problems are tied to the strong covariance structure of size–B/M portfolios, one simple suggestion is to expand the set of test assets to include portfolios sorted in other ways, for example, by industry or factor loadings. Second, since the problems are exacerbated by the fact that empirical tests often ignore theoretical restrictions on the cross-sectional intercept and slopes, another suggestion is to take their magnitudes seriously when theory provides appropriate guidance. Third, since the problems we discuss appear to be less severe for GLS regressions, another suggestion is to report the GLS $R^2$ in addition to, or instead of, the OLS $R^2$. Last, since the problems are exacerbated by sampling issues, our fourth suggestion is to report confidence intervals for cross-sectional $R^2$s 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 don't provide a perfect solution.

The paper also 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 under both the null and alternative hypotheses, offering an economic interpretation of the non-centrality parameter; and (iv) we describe a way to obtain confidence intervals for the GRS F-statistic, cross-sectional $T^2$ statistic, and Hansen-Jagannathan 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 under the null and alternatives, provides an economic interpretation of the non-centrality parameter, and discusses the connection between the $T^2$-statistic and the GLS $R^2$.

Let $R_t$ be the $N \times 1$ vector of excess returns and $F_t$ be the $K \times 1$ vector of factors in period t. Both are assumed, in this appendix, to be IID over time. The matrix of factor loadings is estimated in the first-pass time-series regression, $R_t = c + B\,F_t + e_t$, and the relation between expected returns and B is estimated in the second-pass cross-sectional regression, $\mu = z\,\iota + B\,\gamma + \alpha = X\,\lambda + \alpha$, where $\mu \equiv E[R_t]$, $\lambda' \equiv [z\ \ \gamma']$, $X \equiv [\iota\ \ B]$, and $\alpha$ is the vector of the true pricing errors. More precisely, the parameters in the cross-sectional equation depend on whether we are interested in an OLS or GLS regression. For OLS, the population slope is $\lambda = (X'X)^{-1}X'\mu$ and the pricing errors are $\alpha \equiv [I - X(X'X)^{-1}X']\,\mu \equiv y\,\mu$. For GLS, the slope is $\lambda^* = (X'V^{-1}X)^{-1}X'V^{-1}\mu$ and the pricing errors are $\alpha^* \equiv [I - X(X'V^{-1}X)^{-1}X'V^{-1}]\,\mu \equiv y^*\,\mu$, where $V \equiv \text{var}(R_t)$. In practice, of course, the cross-sectional regression is estimated with average returns substituted for $\mu$ and estimates of B substituted for the true loadings.

We begin with a few population results that will be 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 will use V and $\Sigma$ 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 $q^* = \alpha^{*\prime}\,[y^*\,\Sigma\,y^{*\prime}]^+\,\alpha^*$, where a superscript '+' denotes a pseudoinverse, will be 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 $\Sigma$ is replaced by V. Together with Result 1, this result will imply that the $T^2$ statistics, from either OLS or GLS, are the same regardless of which covariance matrix we use.*

Proof: $y\,X = y^*\,X = 0$ implies that $y\,B = y^*\,B = 0$, since $X = [\iota\;\; B]$. Thus, $y\,V = y\,(B\,\Sigma_F\,B' + \Sigma) = y\,\Sigma$ and $y^*\,V = y^*\,(B\,\Sigma_F\,B' + \Sigma) = y^*\,\Sigma$. The result follows immediately. □

**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 $q^* = \alpha^{*\prime}\,[y^*\,\Sigma\,y^{*\prime}]^+\,\alpha^*$. Using the definition of a pseudoinverse, it is easy to confirm that $[y\,\Sigma\,y]^+ = \Sigma^{-1}y^*$, implying that $q = \alpha'\Sigma^{-1}y^*\alpha$. Further, $[y^*\,\Sigma\,y^{*\prime}]^+ = \Sigma^{-1}(y^*)^+$, implying that $q^* = \alpha^{*\prime}\,\Sigma^{-1}(y^*)^+\,\alpha^*$.[9] Moreover, $\alpha^* = y^*\alpha$, from which it follows that $q^* = \alpha'y^{*\prime}\Sigma^{-1}(y^*)^+y^*\alpha = \alpha'(\Sigma^{-1}y^*)'(y^*)^+y^*\alpha = \alpha'\Sigma^{-1}y^*\alpha = q$, where the second-to-last equality uses the fact that $\Sigma^{-1}y^*$ is symmetric and that $y^*(y^*)^+y^* = y^*$. □

**Result 4**. *The OLS and GLS quadratics equal $q^* = \alpha^{*\prime}\,\Sigma^{1}\,\alpha^* = \alpha^{*\prime}\,V^{1}\,\alpha^*$. 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, $q^* = \alpha'\,\Sigma^{-1}\,y^*y^*\,\alpha = \alpha'\,y^{*\prime}\,\Sigma^{-1}\,y^*\,\alpha = \alpha^{*\prime}\,\Sigma^{-1}\,\alpha^*$. □

Mimicking portfolios, $R_P$, for the factors are defined as the K portfolios maximally correlated with F. The $N \times K$ matrix of mimicking-portfolio weights are slopes in the regression $F = k + w_P'\,R + s$, where $\text{cov}(R, s) = 0$ (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, $w_P = V^{-1}\,\text{cov}(R, F) = V^{-1}\,B\,\Sigma_F$ and stocks' loadings on the mimicking portfolios are $C = \text{cov}(R, R_P)\,\Sigma_P^{-1} = V\,w_P\,\Sigma_P^{-1} = B\,\Sigma_F\,\Sigma_P^{-1}$.

**Result 5**. *The cross-sectional regression (OLS or GLS) of $\mu$ on B is equivalent to the cross-sectional regression of $\mu$ on C, with or without an intercept, in the sense that the intercept, $R^2$, 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., $q^* = \alpha^{*\prime}\,V^{-1}\,\alpha^*$, where V is invariant to the set of factors. □

[9] More precisely, $\Sigma^{-1}(y^*)^+$ is a generalized inverse of $y^*\Sigma y^{*\prime}$, 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.

**Result 6**. *The GLS regression of $\mu$ on B or $\mu$ 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 $\alpha_P^* = w_P' \alpha^* = \Sigma_F B' V^{-1} \alpha^* = \Sigma_F B' V^{-1} y^* \mu$. 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 $\gamma_P^*$ are the GLS slopes on C. Therefore, $\alpha_P^* = w_P' \alpha^* = \mu_P - z^* \iota - \gamma_P^* = 0$, where $\mu_P = w_P' \mu$ and $C_P = w_P' C = I_K$. Solving for $\gamma_P^*$ proves the second half of the result. □

**Result 7**. *Pricing errors in a GLS cross-sectional regression of $\mu$ on C are identical to the intercepts in a time-series regression of $R - z^*\iota$ on a constant and $R_P - z^*\iota$. It follows that the quadratics q and q* equal $\theta_\tau^2(z^*) - \theta_P^2(z^*)$, where $\theta_i(z^*)$ is a generalized Sharpe ratio with respect to $r_f + z^*$, defined as $(\mu_i - z^*) / \sigma_i$, and $\sigma_i$ is asset i's standard deviation, $\tau$ is the 'tangency' portfolio with respect to $r_f + z^*$, and $\theta_P$ 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 $\alpha^*$ since $\gamma_P^* = \mu_P - z^*\iota$. The intepretation of the quadratics then follows immediately from the well-known intepretation of $\alpha_{TS}' \Sigma^{-1} \alpha_{TS}$ (Jobson and Korkie, 1982; Gibbons, Ross, 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 $\mu_{gmv}$ is the expected return on the global minimum variance portfolio (note that Q depends only on assets 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 $\mu_{gmv}$ (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 first claim follows directly from the definition of the GLS $R^2$, i.e., GLS $R^2 = 1 - \alpha^{*\prime} V^{-1} \alpha^* / (\mu - z_{nf}\, \iota)' V^{-1} (\mu - z_{nf}\, \iota)$, where $z_{nf}$ is the GLS intercept when $\mu$ is regressed only a constant. $z_{nf}$ is the same as $\mu_{gmv}$, and $\alpha^{*\prime} V^{-1} \alpha^*$ is the same as q* (see Result 4). The second claim, which we state without further proof, is a multifactor generalization of the results of Kandel and Stambaugh (1995) with mimicking portfolios substituted for non-return factors (see Result 5). The key fact is that $q^* = Q - (\mu_P -$

$\mu_{gmv}$ ι)′ $\Sigma_P{*}^{-1}$ ($\mu_P$ – $\mu_{gmv}$ ι), where $\Sigma_P$* 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 $\mu_P = \mu_{gmv}$ ι. □

Together, Results 1–8 describe key properties of GLS cross-sectional regressions, establish the equality between 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, α = 0 and α* = 0, 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 (defined precisely below). Let r be average returns, b be the sample (first-pass, time-series regression) estimate of B, x = [ι b] be the corresponding estimate of X, v and S be the usual estimates of V and Σ, $\hat{y} = I - x(x'x)^{-1}x'$ be the sample estimate of y, and $\hat{y}^* = I - x(x'v^{-1}x)^{-1}x'v^{-1}$ be the sample estimate of y*. The estimated OLS cross-sectional regression is $r = x\hat{\lambda} + \hat{a}$, where $\hat{\lambda} = (x'x)^{-1}x'r$ and $\hat{a} = \hat{y}\,r$ is the sample OLS pricing error. 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, since $r = (1/T)\,\Sigma_t\,R_t$, the estimated slope and pricing errors can be interpreted as time-series averages of period-by-period Fama-MacBeth cross-sectional estimates. We will focus on OLS regressions but, as a consequence of the sample analog of Result 3 above, we show that the $T^2$ statistic is equivalent from OLS and GLS.

Our analysis below uses the following facts:

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

Combining these facts, the pricing error in period t is $\hat{a}_t \equiv \hat{y}\,R_t = \hat{y}\,(B - b)\,\gamma + \hat{y}\,\alpha + \hat{y}\,(B - b)\,UF_t + \hat{y}\,e_t$ and the time-series average is $\hat{a} = \hat{y}\,(B - b)\,\gamma + \hat{y}\,\alpha + \hat{y}\,(B - b)\,\overline{UF} + \hat{y}\,\bar{e}_t$, where an upper bar denotes a time-series average. Asymptotically, $b \to B$, $\hat{y} \to y$, $\overline{UF} \to 0$, and $\bar{e}_t \to 0$, where → 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, $\hat{y}\,(B - b)\,\overline{UF}$, 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 â 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 precisely, $d = \hat{y}\,(B - b)\,\gamma + \hat{y}\,\bar{e}_t$. The asymptotic distribution is the same substituting y for ŷ, and the two terms have asymptotic mean of zero and are uncorrelated with each other under the standard assumptions of OLS regressions (i.e., in a regression, estimation error in the slopes is uncorrelated with the mean of the residuals). The asymptotic covariance is, therefore, $\Sigma_d = \text{var}[y\,(B - b)\,\gamma] + \text{var}[y\,\bar{e}_t]$. Let $\text{vec}(b - B)$ 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 $\Sigma_d$, the first term becomes $\text{var}[y\,(B - b)\,\gamma] = \gamma'\Sigma_F^{-1}\gamma\,(y\,\Sigma\,y) / T$ and the second term becomes $\text{var}[y\,\bar{e}_t] = y\,\Sigma\,y / T$. 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 $\Sigma_d$ substituting the statistics ŷ, S, $\hat{\gamma}$, and $S_F$ for the population parameters y, Σ, γ, and $\Sigma_F$. 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 $\Sigma_F$ are used for both). Moreover, Result 4 implies that $T^2 = \hat{a}^{*\prime}\,S^{-1}\,\hat{a}^*\,[T / (1 + \hat{\gamma}'S_F^{-1}\hat{\gamma})]$.

**Result 10**. *The cross-sectional* $T^2$ *statistic is asymptotically* $\chi^2$ *with degrees of freedom* $N - K - 1$ *and non-centrality parameter* $n = k\,q^*$, *where* $k = T / (1 + \gamma'\Sigma_F^{-1}\gamma)$.[10] *Equivalently, from Result 7, the non-centrality parameter is* $n = k\,[\theta_\tau^2(z^*) - \theta_P^2(z^*)]$.

Proof: The cross-sectional pricing errors are $\hat{a} = d + \hat{y}\,\alpha = \hat{y}\,(d + \alpha)$, where the first equality follows from

[10] We use the terminology of a limiting distribution somewhat informally here (notice that, as the result is stated, the non-centrality 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 $\chi^2$ with non-centrality parameter $kq^* = \delta^{*\prime}\Sigma^{-1}\delta^* / (1 + \gamma'\Sigma_F^{-1}\gamma)$, where $k = T / (1 + \gamma'\Sigma_F^{-1}\gamma)$ and $q^* = \delta^{*\prime}\Sigma^{-1}\delta^* / T$.

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 facts from the proof of Result 3, it is straightforward to show that $\hat{y} [\hat{y} S \hat{y}]^+ \hat{y} = [\hat{y} S \hat{y}]^+$, which implies 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 $\Sigma_d^+/T$, so the $T^2$ statistic has the same asymptotic distribution as $(d+\alpha)' \Sigma_d^+ (d+\alpha)$. [$S_d^+/T \rightarrow \Sigma_d^+/T$ since $S_d^+/T = S^{-1}\hat{y}^* / (1 + \hat{\gamma}' S_F^{-1} \hat{\gamma})$, 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 $\Sigma_d$ has rank N–K–1. This implies that, asymptotically, $d'\Sigma_d^+ d$ is central $\chi^2$ while $(d+\alpha)' \Sigma_d^+ (d+\alpha)$ is noncentral $\chi^2$ with non-centrality parameter $\alpha' \Sigma_d^+ \alpha$, both with degrees of freedom N–K–1. Result 10 then follows from observing that the non-centrality parameter equals $q^* T / (1+\gamma'\Sigma_F^{-1}\gamma)$. □

**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 $w = g_0 + g_1' R_P$. 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), then the m* which 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 $R_\tau$ 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, $\theta_x/\theta_\tau$. Thus, $s_1$ gives the combination of $R_P$ that has the maximum squared Sharpe ratio, denoted $\theta_P^2$, 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 $\theta_z^2$ can be interpreted as the proposed model's unexplained squared Sharpe ratio.

The analysis above is cast in terms of population parameters but equivalent results go through in sample. Thus, the estimated HJ-distance, $\hat{D}$, 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: $\hat{D} = F\ c\ /\ (1 + r_f)^2\ [N\ (T–K–1)\ /\ (T–N–K)]$, where $c = (1+s_P^2)/T$. It follows immediately that, up to a constant of proportionality, $\hat{D}$ has an F distribution with non-centrality parameter $c^{-1}\theta_z^2 = c^{-1}(1 + r_f)^2 D$.

**References**

Ait-Sahalia, Yacine, Jonathan Parker, and Motohiro Yogo, 2004. Luxury goods and the equity premium. *Journal of Finance* 59, 2959-3004.

Acharya, Viral and Lasse Pedersen, 2005. Asset pricing with liquidity risk. *Journal of Financial Economics* 77 (2), 375-410.

Ahn, Dong-Hyun, Jennifer Conrad, and Robert Dittmar, 2007. Basis assets. *Review of Financial Studies*, forthcoming.

Bansal, Ravi, Robert Dittmar, and Christian Lundblad, 2005. Consumption, dividends, and the cross-section of equity returns. *Journal of Finance* 60, 1639-1672.

Black, Fisher, Michael C. Jensen, and Myron Scholes, 1972. The capital asset pricing model: Some empirical findings, in: *Studies in the theory of capital markets*, ed. by Michael C. Jensen. New York: Praeger.

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

Brennan, Michael, Ashley Wang, and Yihong Xia, 2004. Estimation and test of a simple model of intertemporal asset pricing. *Journal of Finance* 59, 1743-1775.

Campbell, John and Tuomo Vuolteenaho, 2004. Bad beta, good beta. *American Economic Review* 94, 1249-1275.

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

Cochrane, John, 2001. *Asset Pricing*. Princeton University Press, Princeton, NJ.

Daniel, Kent and Sheridan Titman, 2005. Testing factor-model explanations of market anomalies. Working paper (Northwestern University, Evanston, IL).

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

Ferson, Wayne, Sergei Sarkissian, and Timothy Simin, 1999. The alpha factor asset pricing model: A parable. *Journal of Financial Markets* 2, 49-68.

Gibbons, Michael, Stephen Ross, and Jay Shanken, 1989. A test of the efficiency of a given portfolio. *Econometrica* 57, 1121-1152.

Hansen, Lars, 1982. Large Sample Properties of Generalized Method of Moments Estimators. *Econometrica* 50, 1029-1054.

Hansen, Lars, John Heaton, and Nan Li, 2006. Consumption strikes back? Measuring long-run risk. Working paper (University of Chicago, Chicago, IL).

Hansen, Lars and Ravi Jagannathan, 1997. Assessing Specification Errors in Stochastic Discount Factor

Models. *Journal of Finance* 52, 557-590.

Heaton, John and Deborah Lucas, 2000. Asset pricing and portfolio choice: The role of entrepreneurial risk. *Journal of Finance* 55 (3), 1163-1198.

Jacobs, Kris and Kevin Wang, 2004. Idiosyncratic consumption risk and the cross section of asset returns. *Journal of Finance* 59, 2211-2252.

Jagannathan, Ravi and Zhenyu Wang, 1996. The conditional CAPM and the cross-section of stock returns. *Journal of Finance* 51, 3-53.

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

Kan, Raymond and Chou Zhang, 1999. Two-pass tests of asset pricing models with useless factors. *Journal of Finance* 54, 203-235.

Kan, Raymond and Guofu Zhou, 2004. Hansen-Jagannathan distance: Geometry and exact distribution. Working paper (University of Toronto and Washington University in St. Louis).

Kandel, Shmuel and Robert Stambaugh, 1987. On correlations and inferences about mean-variance efficiency. *Journal of Financial Economics* 18, 61-90.

Kandel, Shmuel and Robert Stambaugh, 1995. Portfolio inefficiency and the cross-section of expected returns. *Journal of Finance* 50, 157-184.

Lettau, Martin and Sydney Ludvigson, 2001. Resurrecting the (C)CAPM: A cross-sectional test when risk premia are time-varying. *Journal of Political Economy* 109, 1238-1287.

Lewellen, Jonathan and Stefan Nagel, 2006. The conditional CAPM does not explain asset-pricing anomalies. *Journal of Financial Economics* 82, 289-314.

Li, Qing, Maria Vassalou, and Yuhang Xing, 2006. Sector investment growth rates and the cross section of equity returns. *Journal of Business* 89, 1637-1665.

Lustig, Hanno and Stijn Van Nieuwerburgh, 2004. Housing collateral, consumption insurance, and risk premia. *Journal of Finance* 60 (3), 1167-1221.

Mehra, Rajnish and Edward Prescott, 1985. The equity premium: A puzzle. *Journal of Monetary Economics* 15, 145-161.

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

Pastor, Lubos and Robert Stambaugh, 2003. Liquidity risk and expected stock returns. *Journal of Political Economy* 111, 642-685.

Parker, Jonathan and Christian Julliard, 2005. Consumption risk and the cross section of expected returns. *Journal of Political Economy* 113, 185-222.

Petkova, Ralitsa, 2006. Do the Fama-French factors proxy for innovations in predictive variables?

*Journal of Finance* 61, 581-612.

Reisman, Haim, 1992. Reference variables, factor structure, and the approximate multibeta representation. *Journal of Finance* 47, 1303-1314.

Roll, Richard and Stephen Ross, 1994. On the cross-sectional relation between expected returns and betas. *Journal of Finance* 49, 101-121.

Santos, Tano and Pietro Veronesi, 2006. Labor income and predictable stock returns. *Review of Financial Studies* 19, 1-44.

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

Shanken, Jay, 1987. Multivariate proxies and asset pricing relations: Living with the Roll critique. *Journal of Financial Economics* 18, 91-110.

Shanken, Jay, 1990. Intertemporal asset pricing: An empirical investigation. *Journal of Econometrics* 45, 99-120.

Shanken, Jay, 1992a. On the current state of the arbitrage pricing theory, *Journal of Finance* 47, 1569-1574.

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

Shanken, Jay and Goufu Zhou, 2007. Estimating and testing beta pricing models: Alternative methods and their performance in simulations. *Journal of Financial Economics* 84, 40-86.

Stock, James, 1991. Confidence intervals for the largest autoregressive root in U.S. macroeconomic time series. *Journal of Monetary Economics* 28, 435-459.

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

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

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

The table reports slopes, Shanken (1992b) t-statistics (in parentheses), and $R^2$s from cross-sectional regressions of average excess returns on factor loadings for eight models proposed in the literature. Returns are quarterly (%). 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 residuals (pricing errors) in the cross-sectional regression are all zero, as described in the text, with simulated p-values in brackets. $T^2$ 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. 95% confidence intervals for the true $R^2$s and q are reported in brackets below the sample values. The models are estimated from 1963–2004 except Yogo's, for which we have factor data through 2001.

| Model / assets | Variables | | | | OLS $R^2$ | GLS $R^2$ | $T^2$ | q |
|---|---|---|---|---|---|---|---|---|
| **LL (2001)** | **const.** | **cay** | **Δc** | **cay×Δc** | | | | |
| FF25 | 3.33 | -0.81 | 0.25 | 0.00 | 0.58 | 0.05 | 33.9 | 0.44 |
| | (2.28) | (-1.25) | (0.84) | (0.42) | [0.30, 1.00] | [0.00, 0.50] | [p=0.08] | [0.00, 0.72] |
| FF25 + 30 ind. | 2.50 | -0.48 | 0.09 | -0.00 | 0.00 | 0.01 | 193.8 | 1.31 |
| | (3.29) | (-1.23) | (0.53) | (-1.10) | [0.00, 0.35] | [0.00, 0.20] | [p=0.00] | [0.18, 1.08] |
| **LVN (2004)** | **const.** | **my** | **Δc** | **my×Δc** | | | | |
| FF25 | 3.58 | 4.23 | 0.02 | 0.10 | 0.57 | 0.02 | 20.8 | 0.45 |
| | (2.22) | (0.76) | (0.04) | (1.57) | [0.35, 1.00] | [0.00, 0.35] | [p=0.57] | [0.00, 0.48] |
| FF25 + 30 ind. | 2.78 | 0.37 | -0.02 | 0.03 | 0.09 | 0.00 | 157.1 | 1.32 |
| | (3.51) | (0.13) | (-0.09) | (1.40) | [0.00, 1.00] | [0.00]* | [p=0.04] | [0.00, 0.96] |
| **SV (2006)** | **const.** | $\mathbf{R_M}$ | $\mathbf{s^w \times R_M}$ | | | | | |
| FF25 | 3.07 | -0.95 | -0.21 | | 0.27 | 0.02 | 26.0 | 0.46 |
| | (1.96) | (-0.58) | (-2.06) | | [0.00, 1.00] | [0.00, 0.40] | [p=0.63] | [0.00, 0.30] |
| FF25 + 30 ind. | 2.57 | -0.49 | -0.09 | | 0.08 | 0.02 | 160.8 | 1.31 |
| | (2.77) | (-0.44) | (-1.99) | | [0.00, 1.00] | [0.00, 0.40] | [p=0.07] | [0.00, 0.72] |
| **LVX (2006)** | **const.** | $\mathbf{\Delta I_{HH}}$ | $\mathbf{\Delta I_{Corp}}$ | $\mathbf{\Delta I_{Ncorp}}$ | | | | |
| FF25 | 2.47 | -0.80 | -2.65 | -8.59 | 0.80 | 0.26 | 25.2 | 0.34 |
| | (2.13) | (-0.39) | (-1.03) | (-1.96) | [0.75, 1.00] | [0.05, 1.00] | [p=0.29] | [0.00, 0.48] |
| FF25 + 30 ind. | 2.22 | 0.20 | -0.93 | -5.11 | 0.42 | 0.04 | 141.2 | 1.27 |
| | (3.14) | (0.19) | (-0.58) | (-2.32) | [0.20, 1.00] | [0.00, 0.55] | [p=0.11] | [0.00, 0.84] |
| **Yogo (2006)** | **const.** | $\mathbf{\Delta c_{Ndur}}$ | $\mathbf{\Delta c_{Dur}}$ | $\mathbf{R_M}$ | | | | |
| FF25 | 1.98 | 0.28 | 0.67 | 0.48 | 0.18 | 0.04 | 24.9 | 0.46 |
| | (1.36) | (1.00) | (2.33) | (0.29) | [0.00, 1.00] | [0.00, 0.55] | [p=0.69] | [0.00, 0.30] |
| FF25 + 30 ind. | 1.95 | 0.18 | 0.19 | 0.12 | 0.02 | 0.05 | 159.3 | 1.24 |
| | (2.27) | (1.01) | (1.52) | (0.11) | [0.00, 0.60] | [0.00, 1.00] | [p=0.06] | [0.00, 0.78] |
| **CAPM** | **const.** | $\mathbf{R_M}$ | | | | | | |
| FF25 | 2.90 | -0.44 | | | -0.03 | 0.01 | 77.5 | 0.46 |
| | (3.18) | (-0.39) | | | [0.00, 0.55] | [0.00, 0.25] | [p=0.00] | [0.12, 0.48] |
| FF25 + 30 ind. | 2.03 | 0.10 | | | -0.02 | 0.00 | 225.2 | 1.34 |
| | (2.57) | (0.09) | | | [0.00, 0.35] | [0.00, 0.05] | [p=0.00] | [0.18, 0.96] |

Table 1 continues on next page (variables are defined at the end of the table)

**Table 1, continued.**

| Model / assets | Variables | | | | OLS $R^2$ | GLS $R^2$ | $T^2$ | q |
|---|---|---|---|---|---|---|---|---|
| **Cons. CAPM** | **const.** | **$\Delta c$** | | | | | | |
| FF25 | 1.70 | 0.24 | | | 0.05 | 0.01 | 60.6 | 0.46 |
| | (2.47) | (0.89) | | | [0.00, 1.00] | [0.00, 0.50] | [p=0.01] | [0.06, 0.66] |
| FF25 + 30 ind. | 2.07 | 0.03 | | | -0.02 | 0.00 | 224.5 | 1.34 |
| | (3.51) | (0.15) | | | [0.00, 0.65] | [0.00]* | [p=0.00] | [0.18, 1.02] |
| **Fama–French** | **const.** | **$R_M$** | **SMB** | **HML** | | | | |
| FF25 | 2.99 | -1.42 | 0.80 | 1.44 | 0.78 | 0.19 | 56.1 | 0.37 |
| | (2.33) | (-0.98) | (1.70) | (3.11) | [0.60, 1.00] | [0.05, 0.65] | [p=0.00] | [0.06, 0.42] |
| FF25 + 30 ind. | 2.21 | -0.49 | 0.60 | 0.87 | 0.31 | 0.06 | 200.4 | 1.24 |
| | (2.14) | (-0.41) | (1.24) | (1.80) | [0.00, 0.90] | [0.05, 0.35] | [p=0.00] | [0.12, 0.90] |

* 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 $R^2$.

Variables:
$R_M$ = CRSP value-weighted excess return
$\Delta c$ = log consumption growth
cay = Lettau and Ludvigson's (2001) consumption-to-wealth ratio (de-meaned)
my = Lustig and Van Nieuwerburgh's (2004) housing-collateral ratio based on mortgage data (de-meaned)
$s^w$ = labor income to consumption ratio (de-meaned)
$\Delta I_{HH}$, $\Delta I_{Corp}$, $\Delta I_{Ncorp}$ = log investment growth for households, non-financial corporations, and the non-corporate sector
$\Delta c_{Ndur}$, $\Delta c_{Dur}$ = Yogo's (2006) log consumption growth for non-durables and durables
SMB, HML = Fama and French's (1993) size and B/M factors