A Critique of the Use of $\boldsymbol{t}$-ratios in Model Selection

NAI-FU CHEN, RAYMOND KAN and CHU ZHANG*

Comments are welcome

First draft: April, 1997
This version: June, 1999

*Chen is from the University of California at Irvine and the Hong Kong University of Science and Technology, Kan is from the University of Toronto, Zhang is from the University of Alberta and the Hong Kong University of Science and Technology. We are grateful to Peter Bossaerts, Wayne Ferson, Robert Grauer, Burton Hollifield, Peter Klein, Yuming Li, Tom McCurdy, Kazumitsu Nawata, Zhenyu Wang, Guofu Zhou, seminar participants at Washington University in St. Louis, University of British Columbia, University of Toronto, participants at the 1998 UCI-UCLA-USC Finance Conference, the 1998 Western Finance Meetings and the 1998 Nippon Finance Association and Asia-Pacific Finance Association First Joint International Conference for many helpful comments and suggestions. Kan gratefully acknowledges financial support from the Social Sciences and Humanities Research Council of Canada.

A Critique of the Use of t-ratios in Model Selection

ABSTRACT

In this paper, we expose a subtle but serious problem of selecting models using t-ratios in multivariate regression methodologies. We illustrate the problem in the context of selecting empirical asset pricing models. Using a simple version of the widely used cross-sectional regression methodology, we show analytically that variables with the highest t-ratios may not be highly correlated with expected returns. Contrary to common belief, a high t-ratio may in fact be evidence of low explanatory power. The results in this study cast doubt on the economic significance of variables selected only on the basis of high t-ratios in multivariate regression methodologies and suggest that we should include other diagnostics in addition to t-ratios for model selection.

A central issue in finance is the determination of expected returns across different assets. This is the main focus of many well known theoretical asset pricing models including the Capital Asset Pricing Model (CAPM) of Sharpe (1964) and Lintner (1965), the Intertemporal Capital Asset Pricing Model of Merton (1973), the Arbitrage Pricing Theory of Ross (1976), and the Consumption Capital Asset Pricing Model of Breeden (1979). At the same time, there are many empirically motivated models that propose some firm-specific variables as explanations of the cross-sectional differences of expected returns. Some notable examples in this category are Basu (1977), Banz (1981), Bhandari (1988), Chan, Hamao, and Lakonishok (1991), and Fama and French (1992). A common feature in all these models is that the expected returns are linear in some firm-specific variables (they can be betas corresponding to some common factors or they can be some accounting ratios). In the face of so many competing models, one of the important tasks of empirical researchers is to find out which model does the best job in describing the cross-sectional differences of expected returns.

In practice, this question is often addressed using multivariate regression methodologies such as cross-sectional regressions (CSR), seemingly unrelated regressions (SUR), or generalized method of moments (GMM). A generally accepted practice is to test whether a variable is "priced" by itself (i.e., its slope coefficient has a "significant" $\boldsymbol{t}$-ratio) and whether it is "priced" when combined with other competing variables. For example, in the recent debate of the validity of the CAPM, the focal point is whether the CAPM betas and other competing variables are statistically significantly priced.

In univariate regression methodologies like the ordinary least squares (OLS) regressions, the square of the $\boldsymbol{t}$-ratio is an increasing function of the goodness-of-fit measure $\boldsymbol{R^2}$ for simple regressions (or partial $\boldsymbol{R^2}$ in the case of multiple regressions). Therefore, it is not surprising that researchers consider variables that have high $\boldsymbol{t}$-ratios (in absolute value) to be the ones that possess good explanatory power on the dependent variable. While there are still situations in which using $\boldsymbol{t}$-ratios to select variables in univariate regressions could be problematic, one would be tempted to favor explanatory variables with high t-ratios over those with low t-ratios, especially when the results are robust with respect to different samples.[1]

However, since we are interested in how well an asset pricing model explains the cross-sectional differences of expected returns, the regression methodologies used by most empirical asset pricing studies are multivariate regressions that involve data with both time series and cross-sectional dimensions. In this paper, we point out that the monotonic relation between the measure of goodness-of-fit (i.e., cross-sectional explanatory power of an asset pricing model on expected returns) and the square of the t-ratio does not necessarily hold in multivariate regressions. Therefore, the t-ratio in multivariate regression methodologies is, in general, not an indicator of how good a variable is in explaining the cross-sectional differences of expected returns and, therefore, should not be the only criterion for model selection.

To make our argument more concrete, we focus on the case of the CSR methodology in this paper because it is most commonly used for the purpose of model selection, even though our argument is also relevant to other methodologies such as SUR and GMM. We derive the properties of the t-ratios in the CSR methodology under potentially misspecified models and show analytically that variables with better explanatory power may not have higher t-ratios. A firm-specific variable that provides little explanatory power on the cross-sectional differences of expected returns can be found significantly priced by itself and in the presence of other competing firm-specific variables, whereas a firm-specific variable that provides much explanatory power on the expected returns may not be significantly priced by itself or with other variables. More surprisingly, in many situations, the high t-ratio attained by a candidate variable may in fact be an indication that such a variable is not very useful in explaining the cross-sectional differences of

[1] Well known problems with the use of t-ratios in univariate regressions include multicollinearity and data-mining. The multicollinearity problem is relatively easy to resolve. One just has to examine whether or not a variable has high t-ratio by itself. The data-mining problem can also be mitigated if one can do out-of-sample tests.

expected returns. As a result, variables that are chosen purely on the basis of their high t-ratios could very well be variables that are not very useful.

The rest of the paper is organized as follows. Section 1 provides analytical results of the t-ratios in the OLS CSR with a single firm-specific variable. Section 2 discusses the same problem but in the OLS CSR with multiple firm-specific variables. Section 3 discusses the corresponding results for the generalized least squares (GLS) CSR. Section 4 provides analytical results of the sample R^2. The final section concludes our findings and the Appendix contains proofs of all propositions.

1. The OLS Cross-Sectional Regressions with a Single Firm-Specific Variable

Since Fama and MacBeth (1973), the CSR methodology has been widely used in the accounting and finance literature. The CSR can be run by ordinary least squares (OLS), generalized least squares (GLS), or weighted least squares (WLS). Since the OLS version is easy to implement and it is feasible even with a large number of assets, it has been the most popular version of the CSR methodology. Despite the popularity of the CSR methodology, little is known about its properties under misspecified models. In this section, we first discuss the property of the t-ratio in the OLS CSR methodology when the model contains a single, constant firm-specific variable. Models with multiple firm-specific variables and the GLS CSR are discussed in subsequent sections.

Suppose we observe returns R_t on the N assets for time $t = 1, \ldots, T$ and they are independently distributed as $N(\mu, V)$, where μ is the expected returns on the N assets, and V is the variance-covariance matrix of the N assets. Suppose the model being tested is

$$\mu = \gamma_0 1_N + \gamma_1 x, \tag{1}$$

for some constants γ_0 and γ_1, where 1_N is the N-vector of ones and x is a firm-specific variable, not proportional to $\mathbf{1}_N$.[2] The OLS CSR is performed by first running an OLS regression of the return at time t, R_t, on $X = [\mathbf{1}_N,\ x]$ for every period to obtain estimate

$$\hat{\gamma}_t^{OLS} \equiv \begin{bmatrix} \hat{\gamma}_{0t}^{OLS} \\ \hat{\gamma}_{1t}^{OLS} \end{bmatrix} = (X'X)^{-1}(X'R_t), \quad t = 1, \ldots, T. \tag{2}$$

The sample averages of $\hat{\gamma}_{0t}^{OLS}$ and $\hat{\gamma}_{1t}^{OLS}$ are then reported as the point estimate of γ_0 and γ_1, respectively. In finding out whether or not a firm-specific variable x helps to explain the cross-sectional differences of expected returns, researchers often test the hypothesis $H_0 : \gamma_1 = 0$ using

$$t_{OLS} = \frac{\bar{\hat{\gamma}}_1^{OLS}}{s(\hat{\gamma}_1^{OLS})/\sqrt{T}}, \tag{3}$$

where $\bar{\hat{\gamma}}_1^{OLS}$ and $s(\hat{\gamma}_1^{OLS})$ are, respectively, the sample average and standard deviation of $\hat{\gamma}_{1t}^{OLS}$.

Since it is unlikely that any firm-specific variable x is totally uncorrelated with μ, *testing whether or not a given firm-specific variable is priced is somewhat meaningless.* Therefore, in many empirical studies, the t-ratio is used not only for testing whether or not a firm-specific variable is priced, but also for comparing different models, i.e., models with different xs. It is this use of the t-ratio that we will focus on. With different firm-specific variable xs, the model $\mu = \gamma_0 \mathbf{1}_N + \gamma_1 x$ cannot be all true, and it is even more likely that none of the models being compared is the true model. We call a model misspecified if there do not exist γ_0 and γ_1 such that (1) holds.

For a misspecified model, the first step toward understanding the property of the t-test is to identify the parameter that the OLS estimator in (2) estimates. Define

$$\begin{aligned} \gamma^{OLS} \equiv \begin{bmatrix} \gamma_0^{OLS} \\ \gamma_1^{OLS} \end{bmatrix} &= \text{argmin}_\gamma (\mu - X\gamma)'(\mu - X\gamma) \\ &= (X'X)^{-1}(X'\mu), \qquad (4) \\ \delta_{OLS}^2(x) &= \frac{T(x'M\mu)^2}{x'MVMx}, \qquad (5) \end{aligned}$$

[2]Although the problem that we discuss applies also to the case of time-varying x, μ and V, we present the simplest case of constant parameters in this paper to illustrate the main point.

where $\boldsymbol{M} = \boldsymbol{I_N} - \boldsymbol{1_N}(\boldsymbol{1}'_{\boldsymbol{N}}\boldsymbol{1_N})^{-1}\boldsymbol{1}'_{\boldsymbol{N}}$. The properties of the OLS estimate $\hat{\gamma}_t^{OLS}$ and the OLS t-test are summarized in the following proposition.

Proposition 1 *For a possibly misspecified model, (i) the OLS estimator $\hat{\gamma}_t^{OLS}$ is an unbiased estimator of γ^{OLS}, and (ii) $\mathbf{t_{OLS}}$ is distributed as a noncentral t-distribution with degrees of freedom $T - 1$ and the square of noncentrality parameter $\delta^2_{OLS}(x)$.*

The proposition shows that when the OLS CSR is used, the model is estimated at the parameter that minimizes the sum of squares of the pricing errors of the N assets. If $\gamma_1^{OLS} = 0$, then $\delta^2_{OLS}(x) = 0$, so the t-test is still valid for hypothesis $\gamma_1^{OLS} = 0$, whether or not the model is correctly specified. The question is how good the t-test is when it comes to model selection. To answer this question, we have to understand the magnitude of the t-ratio in various misspecified models. It turns out that the absolute expected value of the t-ratio as well as the probability of rejecting $H_0 : \gamma_1^{OLS} = 0$ in a two-tailed t-test are both positively related to $\delta^2_{OLS}(x)$ (see Johnson, Kotz, and Balakrishnan (1995, Ch. 31)). Therefore, variables with higher $\delta^2_{OLS}(x)$ will tend to have higher t-ratios and they are more likely to be accepted. The question then becomes whether or not a good model necessarily has a high $\delta^2_{OLS}(x)$.

The best model, of course, is a correctly specified model, i.e., an x such that it is perfectly correlated with μ. For such an x, it is easy to see that $\delta^2_{OLS}(x) = \delta^2_{OLS}(\mu)$. Will $\delta^2_{OLS}(x)$ reach maximum at $x = \mu$? The following proposition gives us the first hint that $\delta^2_{OLS}(x)$, and hence the t-test, is not good for model selection because there is no guarantee that the $\delta^2_{OLS}(x)$ of a misspecified model is less than the $\delta^2_{OLS}(\mu)$ of a correct model.

Proposition 2 *For the noncentrality parameters of the OLS t-ratios as functions of x,*

$$\delta^2_{OLS}(\mu) \leq \max_{x} \delta^2_{OLS}(x), \tag{6}$$

with the equality holding if and only if $M\mu = cMVM\mu$ for some constant $c \neq 0$.

In general, there is no reason why μ and V should be related by $M\mu = cMVM\mu$ and so one would not expect the equality between $\delta^2_{OLS}(\mu)$ and $\max_x \delta^2_{OLS}(x)$ to hold. In that case, a true model may lose to an inferior model if the t-ratio is used as the criterion of model selection.

To compare potentially misspecified models, we would like to use a measure, or an inverse measure, of model misspecification. To this end, we consider a population measure of the goodness-of-fit of a model under the OLS CSR,

$$\rho^2_{OLS}(x) = 1 - \frac{(\mu - X\gamma^{OLS})'(\mu - X\gamma^{OLS})}{\mu' M\mu} = \frac{(x'M\mu)^2}{(x'Mx)(\mu'M\mu)}. \quad (7)$$

$\rho^2_{OLS}(x)$ measures the percentage of the variation of μ explained by the firm-specific variable x. Under the OLS CSR, this is a natural choice because it is inversely related to the sum of squares of pricing errors of the model. We refer to $\rho^2_{OLS}(x)$ as the explanatory power of the model with firm-specific variable x. Although Proposition 2 suggests that the t-ratio in OLS CSR is not a proper measure of goodness-of-fit, we would like to find out if there is any linkage between $\delta^2_{OLS}(x)$ and $\rho^2_{OLS}(x)$.

Let $P\Lambda P'$ be the spectral decomposition of MVM, where $\Lambda = \text{Diag}(\lambda_1, \ldots, \lambda_{N-1})$ is a diagonal matrix of the nonzero eigenvalues of MVM listed in descending order, and P is an $N \times (N-1)$ orthonormal matrix with its columns equal to the eigenvectors of MVM associated with Λ. Suppose $\lambda_1 > \lambda_{N-1}$. It can be shown that the mapping from ρ^2_{OLS} to δ^2_{OLS} is not single valued. Instead, for firm-specific variables with the same explanatory power on the expected returns, ρ^2_{OLS}, there could be a wide range of statistical significance for their t-ratios, δ^2_{OLS}. We state this in the following proposition.

Proposition 3 *Suppose* $M\mu \neq cMVM\mu$ *for any constant* c*. Then, if* $\rho^2_{OLS}(x) = 0$*, then* $\delta^2_{OLS}(x) = 0$*; if* $\rho^2_{OLS}(x) = 1$*, then* $\delta^2_{OLS}(x) = T(\mu'M\mu)^2/(\mu'MVM\mu)$*; and for any given number* $\rho^2_0 \in (0,1)$*, the set* $\{\delta^2_{OLS}(x) : \rho^2_{OLS}(x) = \rho^2_0\}$ *contains more than one element; and* $\max_{x:\rho^2_{OLS}(x)=\rho^2_0} \delta^2_{OLS}(x)$ *is not a monotonically increasing* *function of* $\boldsymbol{\rho_0^2}$.

The range of $\boldsymbol{\delta^2_{OLS}}$ for a given value of $\boldsymbol{\rho^2_{OLS}}$ is determined by both $\boldsymbol{\mu}$ and $\boldsymbol{V}$. Analytical expressions of this range are given in the proof of Proposition 3. For real world data, this range is typically very wide. For illustration, we use 10 size-ranked, equally weighted portfolios of the combined NYSE-AMEX as an example. Since we do not know the true $\boldsymbol{\mu}$ and $\boldsymbol{V}$ of these 10 portfolios, we set $\boldsymbol{\mu}$ and $\boldsymbol{V}$ equal to their sample moments (i.e., $\bar{\boldsymbol{R}} = \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{R_t}$ and $\hat{\boldsymbol{V}} = \frac{1}{T-1}\sum_{t=1}^{T}(\boldsymbol{R_t} - \bar{\boldsymbol{R}})(\boldsymbol{R_t} - \bar{\boldsymbol{R}})'$, respectively) computed based on 330 monthly returns from the period July 1963 to December 1990 (the same period used in many recent studies).[3]

Table 1 about here

In Panel A of Table 1, we report the expected returns ($\boldsymbol{\mu}$) for these 10 size-ranked portfolios, measured in percentage per month. We also report the vector of firm-specific variable ($\boldsymbol{x}^*$) that maximizes $\boldsymbol{\delta^2_{OLS}(x)}$. In Panel B, we report the nonzero eigenvalues of $\boldsymbol{MVM}$. The nonzero eigenvalues of $\boldsymbol{MVM}$ are closely related to those of $\boldsymbol{V}$ as stated in the following lemma.

Lemma 1 *Let* $\hat{\boldsymbol{\lambda}}_1 \geq \hat{\boldsymbol{\lambda}}_2 \geq \cdots \geq \hat{\boldsymbol{\lambda}}_N > 0$ *be the eigenvalues of* $\boldsymbol{V}$. *Then,*

$$\hat{\boldsymbol{\lambda}}_1 \geq \boldsymbol{\lambda}_1 \geq \hat{\boldsymbol{\lambda}}_2 \geq \boldsymbol{\lambda}_2 \geq \cdots \geq \boldsymbol{\lambda}_{N-1} \geq \hat{\boldsymbol{\lambda}}_N. \tag{8}$$

When the returns on the $\boldsymbol{N}$ assets are heteroskedastic and correlated with each other, there will be a dispersion in the nonzero eigenvalues of $\boldsymbol{MVM}$ in general.[4]

[3]Monthly returns on the 10 size-ranked portfolios are constructed based the 100 size-beta-ranked portfolios data set used by Jagannathan and Wang (1996). We are grateful to Ravi Jagannathan and Zhenyu Wang for sharing the data set with us.

[4]Correlations between the returns of the assets do not necessarily give rise to dispersion of the nonzero eigenvalues of $\boldsymbol{MVM}$. For example, suppose returns are homoskedastic and the correlation between returns of any pair of assets is a constant, i.e., $\boldsymbol{V} = \boldsymbol{\sigma^2}[(\boldsymbol{1}-\boldsymbol{a})\boldsymbol{I_N} + \boldsymbol{a}\boldsymbol{1_N}\boldsymbol{1_N}']$, where $-\boldsymbol{1/N} < \boldsymbol{a} < \boldsymbol{1}$ is the common correlation coefficient. In this case, it is easy to show that the nonzero eigenvalues of $\boldsymbol{MVM}$ are all equal to $\boldsymbol{\sigma^2(1-a)}$.

It turns out that the extent to which the statistical significance of the t-ratio (δ^2_{OLS}) and the explanatory power of the model (ρ^2_{OLS}) are misaligned depends on, among other things, how small the ratio λ_{N-1}/λ_1 is. In Panel B, we can see that for the 10 size-ranked portfolios, we have $\lambda_{N-1}/\lambda_1 = 0.0173$. The following two examples help us gain some additional insight into why λ_{N-1}/λ_1 is often very small for the return data typically used in empirical tests of asset pricing models.

Example 1. Suppose returns follow a one-factor structure with homoskedastic and independent idiosyncratic risk, $V = \beta\beta' + \sigma^2 I_N$, where $\beta \neq k1_N$ for any scalar k. In this case, $\hat{\lambda}_1 = \sigma^2 + \beta'\beta$ and $\hat{\lambda}_2 = \cdots = \hat{\lambda}_N = \sigma^2$, whereas $\lambda_1 = \sigma^2 + \beta' M\beta$ and $\lambda_2 = \cdots = \lambda_{N-1} = \sigma^2$. When $\beta' M\beta$ is relatively large compared with σ^2, then V and MVM has one large eigenvalue and the rest of the eigenvalues are small. In this case, the ratio λ_{N-1}/λ_1 can be very small. In empirical tests of asset pricing models, well diversified portfolios are often used as test assets to improve the power of the tests and, in many cases, the variance-covariance matrix of the returns of the portfolios often has one large eigenvalue relative to others.

Example 2. Suppose returns follow a K-factor ($K < N$) structure with homoskedastic and independent idiosyncratic risk, $V = BB' + \sigma^2 I_N$, where B is $N \times K$ and it has full column rank. In addition, we assume 1_N is not in the span of B. In this case, $\hat{\lambda}_i = \sigma^2 + \hat{d}_i$ for $i = 1, \ldots, K$ and $\hat{\lambda}_{K+1} = \cdots = \hat{\lambda}_N = \sigma^2$, where $\hat{d}_1 \geq \cdots \geq \hat{d}_K$ are the nonzero eigenvalues of BB'. For the nonzero eigenvalues of MVM, we have $\lambda_i = \sigma^2 + d_i$ for $i = 1, \ldots, K$, and $\lambda_{K+1} = \cdots = \lambda_{N-1} = \sigma^2$, where $d_1 \geq \cdots \geq d_K$ are the nonzero eigenvalues of $MBB'M$. Therefore, even if returns follow a K-factor model, but if just one linear combination of factors can explain a lot of the variance of the returns of the assets or if the portfolios are well diversified, then the ratio λ_{N-1}/λ_1 is still very small.

With this in mind, we then investigate the magnitude of the problem of using OLS t-ratios for the 10 size-ranked portfolios case. If a researcher proposes a firm-specific variable $\boldsymbol{x} = \boldsymbol{\mu}$ (or some linear combinations of $\mathbf{1_N}$ and $\boldsymbol{\mu}$), such a firm-specific variable should be the best one in explaining the cross-sectional differences of the expected returns of the 10 portfolios. However, Panel C of Table 1 shows that in the OLS CSR, we can only reject the null hypothesis that our perfect firm-specific variable ($\boldsymbol{x} = \boldsymbol{\mu}$) is priced at the 5% level in less than half of the times when the number of time series observations is 330.[5] On the other hand, we find $\boldsymbol{x}^*$ ($= \arg\max_{\boldsymbol{x}} \boldsymbol{\delta^2_{OLS}(x)}$)[6] to be significantly priced at the 5% level with a probability of 0.965 (with 330 time series observations), despite the fact that $\boldsymbol{x}^*$ has very low correlation with $\boldsymbol{\mu}$ and its $\boldsymbol{\rho^2_{OLS}(x)}$ is only 0.114. Thus, the poor model is more significantly priced than the true model. The magnitude of the problem that we have shown in this example is not unusual. For the case when the parameters are determined by the 100 size-beta ranked portfolios, the problem is even more severe. These results suggest that the OLS $\boldsymbol{t}$-ratio is a poor indicator of how good a model is.

While our analysis focuses on the noncentrality parameter, it should be noted that by replacing $\boldsymbol{\mu}$ with $\boldsymbol{\bar{R}}$ and $\boldsymbol{V}$ with $\boldsymbol{\hat{V}}$, Propositions 2 and 3 also apply to the sample OLS $\boldsymbol{t}$-ratio. This is because $\boldsymbol{t^2_{OLS}}$ is just a sample version of $\boldsymbol{\delta^2_{OLS}}$. A case of special interest is when $\boldsymbol{N > T}$. One of the reasons for the popularity of the OLS CSR is because the OLS CSR is feasible even when $\boldsymbol{N > T}$. However, when $\boldsymbol{N > T}$, $\boldsymbol{\hat{V}}$ is singular and one can always find poor firm-specific variables that have arbitrary high sample OLS $\boldsymbol{t}$-ratios.[7] Thus, variables with high OLS $\boldsymbol{t}$-ratios, whether *ex ante* or *ex*

[5] Affleck-Graves and Bradfield (1993) and Chan and Lakonishok (1993) suggest that with the typical number of time series observations, the two-pass OLS CSR methodology does not possess enough power to determine if the risk premium of market betas is significantly different from zero because the returns are too noisy.

[6] Any linear combinations of $\mathbf{1_N}$ and $\boldsymbol{x}^*$ will do exactly the same job. The one that we report is chosen such that it has the same cross-sectional mean and standard deviation as $\boldsymbol{\mu}$.

[7] To see that, we rewrite

$$\begin{aligned} t^2_{OLS} &= \frac{T(x'M\bar{R})^2}{x'M\hat{V}Mx} = \frac{T(T-1)(x'M\bar{R})^2}{x'M\left[\sum_{t=1}^T (R_t-\bar{R})(R_t-\bar{R})'\right]Mx} \\ &= \frac{T(T-1)(x'M\bar{R})^2}{\sum_{t=1}^T \left[x'M(R_t-\bar{R})\right]^2}. \end{aligned}$$

post, do not have to be good variables in explaining the expected returns, and one should be extremely cautious in choosing variables based on their OLS t-ratios.

Let us now return to the 10 size-ranked portfolios case and examine the general relation between ρ^2_{OLS} and δ^2_{OLS}. Figure 1 presents the lower bound and upper bound of δ^2_{OLS} for $0 \leq \rho^2_{OLS} \leq 1$ (330 time series observations). The lower bound of δ^2_{OLS} is monotonic in ρ^2_{OLS} but the upper bound of δ^2_{OLS} is not. The upper bound of δ^2_{OLS} goes up very quickly with ρ^2_{OLS}. Even with $\rho^2_{OLS} = 0.05$, there are firm-specific variables that will be on average more significantly priced than the "true" model that provides perfect explanatory power. As shown in Table 1, the maximum δ^2_{OLS} is reached when $\rho^2_{OLS} = 0.114$. While such a variable is not very powerful in explaining the cross-sectional differences of expected returns, it is, in fact, the most significantly priced among all the firm-specific variables. Figure 1 shows that the problem we report in Table 1 is very widespread and it is not just a special case. In Figure 2, we present a plot similar to Figure 1 but with μ and V determined by the sample estimates of 100 size-beta-ranked portfolios. The pattern is similar, i.e., there are many misspecified models that are more significantly priced than the true model and the one that is most significantly priced has a low ρ^2_{OLS} of 0.272.

While Figures 1 and 2 show a clear picture that the OLS t-ratio is an ambiguous indicator of how good a firm-specific variable is, there is a stronger message behind these figures. It suggests that firm-specific variables with very high expected value of OLS t-ratios are in fact very bad in explaining the cross-sectional differences of expected returns. To see this more clearly, we exchange the x-axis and y-axis in Figure 1 and replot it in Figure 3. Instead of labelling the x-axis as δ^2_{OLS}, we transform it into its

If $N > T$, we can choose x to be uncorrelated with $R_t - \bar{R}$ for $t = 1, \ldots, T$, and the denominator of the t-ratio is zero. To the extent that such an x is correlated with $\bar{R}$ (i.e., $\bar{R}$ is not in the span of $[1_N, R_1 - \bar{R}, \ldots, R_T - \bar{R}]$), its OLS t-ratio is infinity. If $\bar{R}$ is not proportional to μ and $N > T+1$, we can even choose x to be uncorrelated with μ and $R_t - \bar{R}$ for $t = 1, \ldots, T$. Such firm-specific variables will have infinite sample OLS t-ratios but zero explanatory power on expected returns, i.e., $\rho^2_{OLS} = 0$. Since both t_{OLS} and ρ^2_{OLS} are continuous in x, the sample OLS t-ratios can be arbitrarily large for those firm-specific variables that are close enough to the x which has $t_{OLS} = \infty$ and $\rho^2_{OLS} = 0$.

corresponding absolute expected value of OLS $\boldsymbol{t}$-ratio. With the change of axes and rescaling, Figure 3 shows the lower bound and upper bound of $\boldsymbol{\rho^2_{OLS}}$ for a firm-specific variable which has a particular absolute expected value of OLS $\boldsymbol{t}$-ratio. It shows, for example, a firm-specific variable that has an expected $\boldsymbol{t}$-ratio of 3.5 in the OLS CSR, cannot explain more than 20% of the cross-sectional variation of the expected returns on the 10 size-ranked portfolios. These results suggest that if a researcher should do pre-test screening by picking variables based on high OLS $\boldsymbol{t}$-ratios in the univariate OLS CSR, it would likely give him a bunch of variables that are not very useful.

Note that the shapes of the graphs in Figures 1–3 and the relative ranking of $\boldsymbol{\delta^2_{OLS}}$ of firm-specific variables do not depend on $\boldsymbol{T}$. If an inferior firm-specific variable has a higher $\boldsymbol{\delta^2_{OLS}}$ than a superior firm-specific variable, then we can expect the inferior variable to have a higher OLS $\boldsymbol{t}$-ratio than the superior variable for every sample period, regardless of its length. Therefore, looking at subperiod results or out-of-sample period results is not going to help with this problem. *The inferior variable is indeed priced consistently in a statistical sense.* The problem is that the measure of statistical significance of the $\boldsymbol{t}$-ratio ($\boldsymbol{\delta^2_{OLS}}$) does not always coincide with the explanatory power of the variable ($\boldsymbol{\rho^2_{OLS}}$).

The misalignment between the statistical significance of the $\boldsymbol{t}$-ratio and the measure of goodness-of-fit in the OLS CSR comes from the fact that the measure of goodness-of-fit depends only on $\boldsymbol{x}$ and $\boldsymbol{\mu}$, but not on $\boldsymbol{V}$. However, in assessing the statistical significance of the estimated risk premium associated with $\boldsymbol{x}$, we look at $\boldsymbol{x}$, $\boldsymbol{\mu}$, as well as $\boldsymbol{V}$. That $\boldsymbol{V}$ plays a role in determining the statistical significance of a test statistic is common in all statistical tests. Therefore, using other methodologies such as Seemingly Unrelated Regressions or the Generalized Methods of Moments will face the same problem.

The fact that there is a misalignment between the measure of goodness-of-fit and the statistical significance of the $\boldsymbol{t}$-ratio of a firm-specific variable does not explain why poor firm-specific variables can have high OLS $\boldsymbol{t}$-ratios. To gain more intuition about this problem, we consider the OLS CSR estimate of the risk premium at time t with the firm-specific variable scaled such that $x'Mx = 1$. Now, since $\hat{\gamma}_{1t}^{OLS} = \frac{x'MR_t}{x'Mx} = x'MR_t$ when $x'Mx = 1$, its expected value and variance are given by

$$E[\hat{\gamma}_{1t}^{OLS}] = x'M\mu = (\mu'M\mu)^{\frac{1}{2}}\rho_{OLS}(x), \tag{9}$$

$$\mathrm{Var}[\hat{\gamma}_{1t}^{OLS}] = x'MVMx = \sum_{i=1}^{N-1}\lambda_i(p_i'x)^2, \tag{10}$$

where p_i is the eigenvector of MVM associated with the eigenvalue λ_i.

From these two expressions, we can see that if we choose x close to μ, then $\rho_{OLS}^2(x)$ is very high and naturally the absolute value of the numerator of its OLS t-ratio is also going to be very high. However, the denominator of the OLS t-ratio is determined by the variance of $\hat{\gamma}_{1t}^{OLS}$. If p_1 is also highly correlated with μ,[8] $\mathrm{Var}[\hat{\gamma}_{1t}^{OLS}]$ will be large (i.e., λ_1), and the resulting OLS t-ratio can be easily insignificant. On the other hand, if we pick x close to p_{N-1}, $\mathrm{Var}[\hat{\gamma}_{1t}^{OLS}]$ will be small (i.e., λ_{N-1}) and the resulting OLS t-ratio can be very high even though x has very little correlation with μ.

As an illustration, consider the case when returns follow a one-factor structure as discussed in Example 1 of the last section (i.e., $V = \beta\beta' + \sigma^2 I_N$). If expected returns follow an approximate one-factor pricing model, i.e., $\mu \approx \gamma_0 1_N + \gamma_1\beta$ for some constants γ_0 and $\gamma_1 \neq 0$, and if we choose x (normalized to $x'Mx = 1$) to be proportional to β, then such a firm-specific variable obviously explains a lot of the cross-sectional differences of the expected returns. However, V is also mostly determined by β (p_1 is $M\beta/(\beta'M\beta)^{\frac{1}{2}}$), so the risk premium for the firm-specific variable proportional to β has the largest possible variance ($\lambda_1 = \beta'M\beta + \sigma^2$) among all the firm-specific variables with $x'Mx = 1$.[9] As a result, the OLS t-ratio of the firm-specific variable proportional to β may not be very high.

[8] Since p_i is orthonormal and orthogonal to 1_N, $p_i'x$ is the correlation coefficient of x with p_i.

[9] When $V = \beta\beta' + \sigma^2 I_N$, we have $\mathrm{Var}[\hat{\gamma}_{1t}^{OLS}] = x'M[\beta\beta' + \sigma^2 I_N]Mx = (x'M\beta)^2 + \sigma^2$ for firm-specific variables normalized to $x'Mx = 1$. From this expression, it is obvious that if we choose x to be proportional to β (i.e., $x = \beta/(\beta'M\beta)^{\frac{1}{2}}$), the variance of its estimated risk premium will be the largest.

On the other hand, as long as $\boldsymbol{\mu}$ is not completely explained by $\boldsymbol{\beta}$, we can find a firm-specific variable $\boldsymbol{x}$ that is proportional to the part of $\boldsymbol{\mu}$ that is uncorrelated with $\boldsymbol{\beta}$. Such a firm-specific variable has low but still nonzero correlation with $\boldsymbol{\mu}$. However, since it does not account for much of the structure of $\boldsymbol{V}$, the variance of $\hat{\gamma}_{1t}^{OLS}$ is only equal to σ^2, the unsystematic risk of the test assets. For test assets that are well diversified portfolios, σ^2 is small and hence the OLS t-ratios of such poor variables can be large.

It is quite ironic that when the CAPM is almost true and when we use well diversified portfolios as the test assets, we are still susceptible to the problems of finding anomalies to the CAPM and finding the betas not to be statistically significantly priced. More generally, if returns of the test assets are driven by a small number of factors and their expected returns are approximately linear in the betas with respect to those factors, we can find variables with little explanatory power on $\boldsymbol{\mu}$ to be statistically significantly priced.

We now turn our attention to $\rho^2_{OLS}(\boldsymbol{x}^*)$ where $\boldsymbol{x}^* = \arg\max_{\boldsymbol{x}} \delta^2_{OLS}(\boldsymbol{x})$. This measure tells us whether the statistically most significant models in the OLS CSR are good or bad models. If $\rho^2_{OLS}(\boldsymbol{x}^*)$ is high, then although statistical significance of the t-ratio does not align exactly with the measure of goodness-of-fit in the OLS CSR, we can still assume that models with high absolute value of t-ratios are good models. Since this measure depends on $\boldsymbol{\mu}$ and $\boldsymbol{V}$, it is difficult to make a general statement about its magnitude without knowing $\boldsymbol{\mu}$ and $\boldsymbol{V}$. The following lemma gives the worst case scenario for a given $\boldsymbol{V}$.

Lemma 2 *For the firm-specific variable that is most significantly priced in the OLS cross-sectional regression,* $\boldsymbol{x}^* = \arg\max_{\boldsymbol{x}} \delta^2_{OLS}(\boldsymbol{x})$, *we have*

$$\rho^2_{OLS}(\boldsymbol{x}^*) \geq \frac{4\left(\frac{\lambda_{N-1}}{\lambda_1}\right)}{\left[1 + \left(\frac{\lambda_{N-1}}{\lambda_1}\right)\right]^2}, \tag{11}$$

with equality holds when $\eta_1^2 = \frac{\lambda_1}{\lambda_1 + \lambda_{N-1}}$, $\eta_i = 0$ *for* $1 < i < N-1$, *and* $\eta^2_{N-1} =$

$\frac{\boldsymbol{\lambda_{N-1}}}{\boldsymbol{\lambda_1+\lambda_{N-1}}}$, *where* $\boldsymbol{\lambda_1 \geq \cdots \geq \lambda_{N-1} > 0}$ *are the nonzero eigenvalues of* $\boldsymbol{MVM}$ *and* $\boldsymbol{\eta_i}$ *is the correlation coefficient between* $\boldsymbol{\mu}$ *and the eigenvector of* $\boldsymbol{MVM}$ *associated with* $\boldsymbol{\lambda_i}$*.*

From Lemma 2, we can see that a necessary condition for the statistically most significant model in the OLS CSR to be a bad model is $\boldsymbol{\lambda_1 \gg \lambda_{N-1}}$, which is a common property of $\boldsymbol{V}$ found in real world data. However, $\boldsymbol{\lambda_1 \gg \lambda_{N-1}}$ is only a necessary condition for $\boldsymbol{\rho^2_{OLS}(x^*)}$ to be low. For example, when $\boldsymbol{\mu}$ is perfectly correlated with one of the columns of $\boldsymbol{P}$, then even though $\boldsymbol{\lambda_1 \gg \lambda_{N-1}}$, we still have $\boldsymbol{\rho^2_{OLS}(x^*) = 1}$. In order for $\boldsymbol{\rho^2_{OLS}(x^*)}$ to be very small, we also need $\boldsymbol{\eta_1^2}$ to be very close but not equal to one, i.e., $\boldsymbol{\mu}$ to be close but not exactly proportional to the eigenvector associated with the largest eigenvalue of $\boldsymbol{MVM}$. In Panel B of Table 1, we can see that $\boldsymbol{\lambda_{N-1}/\lambda_1 = 0.0173}$ and $\boldsymbol{\eta_1 = 0.9235}$ for the 10 size-ranked portfolios, and hence $\boldsymbol{\rho^2_{OLS}(x^*)}$ is very small. This is a case in point that if the returns of the test assets are driven by a small number of common factors and if the expected returns of the test assets are approximately linear in the betas with respect to these factors, then $\boldsymbol{\rho^2_{OLS}(x^*)}$ will be small.

Three recent studies suggest problems of using the OLS CSR in evaluating the CAPM. Roll and Ross (1994) show that the closeness of a market index to the efficient frontier does not always tell us what the cross-sectional covariance is between expected returns and betas (with respect to the inefficient market index). For example, a market portfolio that is very close to the efficient frontier can produce zero covariance between expected returns and betas. Kandel and Stambaugh (1995) show that a market portfolio that has very minor inefficiency (as measured by how close it is to the frontier) could produce a $\boldsymbol{\rho^2_{OLS}}$ that is almost indistinguishable from zero but yet for a market portfolio far away from the frontier, the risk premium and $\boldsymbol{\rho^2_{OLS}}$ of its betas could be very large. Grauer (1999) constructs examples to show that whether expected returns are correlated with betas tells us very little about the validity of the CAPM, and the difference between the intercept in the OLS CSR and riskless rate has little to do with how close the market portfolio is to the efficient frontier. The common point in these papers is that ρ^2_{OLS} is not always an appropriate measure of how close a market portfolio is to the mean-variance efficient frontier. The point in our paper is fundamentally different. What we suggest is that even if ρ^2_{OLS} is an appropriate measure to compare models, the t-ratios in the OLS CSR do not provide reliable information about ρ^2_{OLS}. While the results of this study can also be specialized to the case of testing the CAPM, our main concern is on the inference derived from t-ratios in the OLS CSR with general firm-specific variables. In our case, the incorrect inferences from using OLS t-ratios arise not from the wrong choice of measure of goodness-of-fit, but from the misalignment of statistical significance of the t-ratio and the measure of goodness-of-fit in model comparison.

2. The OLS Cross-Sectional Regressions with Multiple Firm-Specific Variables

The previous section suggests serious problems of using the OLS t-ratio as a device for pre-test screening and for selecting models with a single firm-specific variable. We now turn to the case of models with multiple firm-specific variables. The most common practice is to put a number of firm-specific variables in a multiple CSR and choose the ones with significant t-ratios and drop the ones with insignificant t-ratios. *One of the myths in using the CSR methodology is that when two or more firm-specific variables are included in the model, the good variables will drive out the bad variables.* In other words, superior firm-specific variables will be more significantly priced than the inferior firm-specific variables in a multiple CSR. In this section, we show that such a belief cannot be justified.

To simplify our argument, we assume the researcher has two firm-specific variables, x_1 and x_2, and he would like to use them to explain μ.[10] Denote $X = [1_N,\ x_1,\ x_2]$.

[10] The analysis of the case with more than two firm-specific variables can be easily generalized from

The $\boldsymbol{\gamma}$ that minimizes the sums of squares of the pricing errors $(\boldsymbol{\mu} - \boldsymbol{X\gamma})'(\boldsymbol{\mu} - \boldsymbol{X\gamma})$ is given by

$$\boldsymbol{\gamma}^{OLS} \equiv \begin{bmatrix} \gamma_0^{OLS} \\ \gamma_1^{OLS} \\ \gamma_2^{OLS} \end{bmatrix} = (\boldsymbol{X}'\boldsymbol{X})^{-1}(\boldsymbol{X}'\boldsymbol{\mu}). \tag{12}$$

If $(\boldsymbol{\mu} - \boldsymbol{X\gamma})'(\boldsymbol{\mu} - \boldsymbol{X\gamma})$ is the relevant measure of model misspecification, the marginal explanatory power of $\boldsymbol{x_1}$ and $\boldsymbol{x_2}$ can be measured by the partial coefficients of determination:

$$\rho^2_{OLS}(\boldsymbol{x_1}|\boldsymbol{x_2}) = 1 - \frac{(\boldsymbol{\mu} - \boldsymbol{X}\gamma^{OLS})'(\boldsymbol{\mu} - \boldsymbol{X}\gamma^{OLS})}{\boldsymbol{\mu}'\boldsymbol{M_2}\boldsymbol{\mu}} = \frac{(\boldsymbol{x}_1'\boldsymbol{M_2}\boldsymbol{\mu})^2}{(\boldsymbol{x}_1'\boldsymbol{M_2}\boldsymbol{x_1})(\boldsymbol{\mu}'\boldsymbol{M_2}\boldsymbol{\mu})}, \tag{13}$$

$$\rho^2_{OLS}(\boldsymbol{x_2}|\boldsymbol{x_1}) = 1 - \frac{(\boldsymbol{\mu} - \boldsymbol{X}\gamma^{OLS})'(\boldsymbol{\mu} - \boldsymbol{X}\gamma^{OLS})}{\boldsymbol{\mu}'\boldsymbol{M_1}\boldsymbol{\mu}} = \frac{(\boldsymbol{x}_2'\boldsymbol{M_1}\boldsymbol{\mu})^2}{(\boldsymbol{x}_2'\boldsymbol{M_1}\boldsymbol{x_2})(\boldsymbol{\mu}'\boldsymbol{M_1}\boldsymbol{\mu})}, \tag{14}$$

where

$$\boldsymbol{M_1} = \boldsymbol{I_N} - \boldsymbol{X_1}(\boldsymbol{X}_1'\boldsymbol{X_1})^{-1}\boldsymbol{X}_1', \tag{15}$$

$$\boldsymbol{M_2} = \boldsymbol{I_N} - \boldsymbol{X_2}(\boldsymbol{X}_2'\boldsymbol{X_2})^{-1}\boldsymbol{X}_2' \tag{16}$$

are the projection matrices onto the space orthogonal to $\boldsymbol{X_1} = [\boldsymbol{1_N},\ \boldsymbol{x_1}]$ and $\boldsymbol{X_2} = [\boldsymbol{1_N},\ \boldsymbol{x_2}]$, respectively. The partial coefficient of determination is a measure of the fraction of the cross-sectional variance of the expected returns that is explained by one firm-specific variable, conditioned on that the other firm-specific variable and the constant term are included in the model.[11]

To estimate $\boldsymbol{\gamma}^{OLS}$, we can run a multiple OLS CSR of $\boldsymbol{R_t}$ on $\boldsymbol{X}$ every period. The OLS CSR estimate of $\boldsymbol{\gamma}^{OLS}$ at time $\boldsymbol{t}$ is

$$\hat{\gamma}_t^{OLS} = \begin{bmatrix} \hat{\gamma}_{0t}^{OLS} \\ \hat{\gamma}_{1t}^{OLS} \\ \hat{\gamma}_{2t}^{OLS} \end{bmatrix} = (\boldsymbol{X}'\boldsymbol{X})^{-1}(\boldsymbol{X}'\boldsymbol{R_t}). \tag{17}$$

the results here.

[11] From the identity

$$\frac{1 - \rho^2_{OLS}(\boldsymbol{x_1})}{1 - \rho^2_{OLS}(\boldsymbol{x_2})} = \frac{1 - \rho^2_{OLS}(\boldsymbol{x_1}|\boldsymbol{x_2})}{1 - \rho^2_{OLS}(\boldsymbol{x_2}|\boldsymbol{x_1})},$$

we can see that the ranking of the partial coefficients of determination gives us the same information about the ranking of simple coefficients of determination. If $\rho^2_{OLS}(\boldsymbol{x_1}|\boldsymbol{x_2}) > \rho^2_{OLS}(\boldsymbol{x_2}|\boldsymbol{x_1})$, it tells us that $\rho^2_{OLS}(\boldsymbol{x_1}) > \rho^2_{OLS}(\boldsymbol{x_2})$ and $\boldsymbol{x_1}$ alone is also a better firm-specific variable than $\boldsymbol{x_2}$ alone.

The OLS t-ratios for testing $H_0 : \gamma_1^{OLS} = 0$ and $H_0 : \gamma_2^{OLS} = 0$ are computed using the time series of $\hat{\gamma}_{1t}^{OLS}$ and $\hat{\gamma}_{2t}^{OLS}$ just like in the case of a single firm-specific variable. The OLS t-ratios of x_1 and x_2 have noncentral t-distribution with $T-1$ degrees of freedom and their noncentrality parameters are given by

$$\delta_{OLS}^2(x_1|x_2) = \frac{T(x_1' M_2 \mu)^2}{x_1' M_2 V M_2 x_1}, \tag{18}$$

$$\delta_{OLS}^2(x_2|x_1) = \frac{T(x_2' M_1 \mu)^2}{x_2' M_1 V M_1 x_2}. \tag{19}$$

In running the multiple OLS CSR with both firm-specific variables x_1 and x_2, the statistical significance of the t-ratios of the two variables are represented by $\delta_{OLS}^2(x_1|x_2)$ and $\delta_{OLS}^2(x_2|x_1)$.

It is important to realize that whether a firm-specific variable is priced or not in a multiple OLS CSR depends on what other variables are included in the model. If one of the variables, say x_2, fully explains the cross-sectional variance of the expected returns (i.e., $\mu = \gamma_0 1_N + \gamma_2 x_2$ for some constants γ_0 and γ_2), then $x_1' M_2 \mu = 0$ and any other variable x_1 would not be priced. Therefore, if we have the correct model, it will subsume the explanatory power of any other imperfect firm-specific variables, and the traditional t-test is well justified in this case.[12] However, as discussed in the last section, $x_2 = \mu$ may not be significantly priced by itself, and hence it may not even be included in the multiple CSR to compete with other firm-specific variables. When x_2 does not fully explain μ, then almost any variable x_1 will be priced, and therefore testing $H_0 : \gamma_1^{OLS} = 0$ in the multiple OLS CSR is just as meaningless as in the simple OLS CSR case.

While the marginal explanatory power of the variables can be judged by their partial

[12] For a given imperfect firm-specific variable x_1, it does not take the true model to subsume the explanatory power of x_1 in a multiple CSR. In fact, there exist many imperfect firm-specific variables x_2 that can completely subsume the explanatory power of x_1. For example, if $\rho_{OLS}^2(x_1) = \rho_0^2$ where $0 < \rho_0^2 < 1$, then by including in the multiple OLS CSR another firm-specific variable

$$x_2 = M\mu - \rho_0^2 M_1 \mu + \rho_0 (1-\rho_0^2)^{\frac{1}{2}} (\mu' M_1 \mu)^{\frac{1}{2}} u, \tag{20}$$

where $u \in [1_N,\ x_1,\ \mu]^{\perp}$ and $u'u = 1$, we have $x_1' M_2 \mu = 0$ and x_1 will not be priced.

coefficients of determination, the ranking of the statistical significance of $\boldsymbol{t}$-ratios in the multiple OLS CSR does not convey any reliable information about the ranking of the partial coefficients of determination. Just like in the case of single firm-specific variable, the expressions for statistical significance (18) and (19) are in general not proportional to the expressions for marginal explanatory power (13) and (14). Hence, comparing $\boldsymbol{t}$-ratios in a multiple OLS CSR is also misleading. In general, there is no guarantee that superior variables are priced more significantly than inferior variables in the multiple OLS CSR. The following proposition is a direct extension of Proposition 2 to the multiple firm-specific variables case.

Proposition 4 *Conditioned on that $\boldsymbol{x_2}$ is included in a multiple cross-sectional regression of regressing returns on both $\boldsymbol{x_1}$ and $\boldsymbol{x_2}$ (and a constant term), the noncentrality parameters of the $\boldsymbol{t}$-ratios of the estimated risk premium of $\boldsymbol{x_1}$ satisfy*

$$\boldsymbol{\delta^2_{OLS}(\mu|x_2) \leq \max_{x_1} \delta^2_{OLS}(x_1|x_2)}, \tag{21}$$

with the equality holding if and only if $\boldsymbol{M_2\mu = cM_2VM_2\mu}$ for some constant $\boldsymbol{c \neq 0}$.

By the iterated projection theorem, it is easy to show that

$$\boldsymbol{\max_{x_1} \delta^2_{OLS}(x_1|x_2) \leq \delta^2_{OLS}(x^*)}, \tag{22}$$

where $\boldsymbol{x^* = \arg\max \delta^2_{OLS}(x)}$, with the equality holding if $\boldsymbol{x_2}$ is uncorrelated with $\boldsymbol{x^*}$. Therefore, the statistically most significant firm-specific variable in a multiple OLS CSR cannot be more significantly priced than the $\boldsymbol{x^*}$ in a simple OLS CSR. This result by no means suggests that the problems of using OLS $\boldsymbol{t}$-ratios are mitigated in the multiple OLS CSR. For example, it is possible that $\boldsymbol{\rho^2_{OLS}(x_1^*) < \rho^2_{OLS}(x^*)}$ for $\boldsymbol{x_1^* = \arg\max \delta^2_{OLS}(x_1|x_2)}$, so the statistically most significant firm-specific variable in the multiple OLS CSR can be even worse than the one in the simple OLS CSR. Therefore, unless one of the variables perfectly explains $\boldsymbol{\mu}$, the problems of using $\boldsymbol{t}$-ratios in selecting variables still exist in the multiple OLS CSR. In addition, it is also likely that

$\boldsymbol{\delta^2_{OLS}(\mu|x_2) < \delta^2_{OLS}(\mu)}$, so the true model can be less significantly priced in the multiple OLS CSR than in the simple OLS CSR,[13] which makes it even more difficult to find support of the true model in the multiple OLS CSR.

Table 2 about here

As an illustration, in the 10 size ranked portfolios case that we discussed earlier, we find that $\boldsymbol{x^* = \arg\max \delta^2_{OLS}(x)}$ is the most significantly priced firm-specific variable in the simple OLS CSR. In Table 2, we put such a variable against some competition in a multiple OLS CSR. Besides $\boldsymbol{x^*}$, we also include another variable $\boldsymbol{x_1}$ (reported in Panel A) in the OLS CSR. The variable $\boldsymbol{x_1}$ is chosen to be proportional to the residuals from regressing $\boldsymbol{\mu}$ on $\boldsymbol{x^*}$, so the constant term $\boldsymbol{1_N}$ together with $\boldsymbol{x_1}$ and $\boldsymbol{x^*}$ fully explain $\boldsymbol{\mu}$. Although $\boldsymbol{\rho^2_{OLS}(x_1) = 0.886}$ and $\boldsymbol{x_1}$ explains a lot more of the cross-sectional differences of the expected returns than $\boldsymbol{x^*}$, we still find $\boldsymbol{x^*}$ to be more significantly priced than $\boldsymbol{x_1}$. For example, in the multiple OLS CSR, we find $\boldsymbol{x_1}$ to be priced at the 5% level with a probability of only 0.415 but $\boldsymbol{x^*}$ is found to be priced at the 5% level with a probability of 0.965. Therefore, it does not appear that the good variable could come close to driving out the bad variable, even in a multiple OLS CSR.

It should be noted that we choose $\boldsymbol{x_1}$ to be uncorrelated with $\boldsymbol{x^*}$ in our example; multicollinearity is thus not a reason for the low statistical significance of $\boldsymbol{x_1}$. On the other hand, if we allow $\boldsymbol{x_1}$ to be correlated with $\boldsymbol{x^*}$, then the statistical significance of $\boldsymbol{x^*}$ will be reduced. In this case, it is important to realize that it is not because $\boldsymbol{x_1}$ is close to $\boldsymbol{\mu}$ that drives out $\boldsymbol{x^*}$; it is only because $\boldsymbol{x_1}$ is close to $\boldsymbol{x^*}$.

With a simple modification of Proposition 3, we can also derive the lower bound and upper bound of $\boldsymbol{\delta^2_{OLS}(x_1|x^*)}$ as a function of $\boldsymbol{\rho^2_{OLS}(x_1|x^*)}$. In Figure 4, we present a plot of the lower bound and upper bound of $\boldsymbol{\delta^2_{OLS}(x_1|x^*)}$ as a function of $\boldsymbol{\rho^2_{OLS}(x_1|x^*)}$

[13] The inequality $\boldsymbol{\delta^2_{OLS}(\mu|x_2) \leq \delta^2_{OLS}(\mu)}$ does not always hold. For some choice of $\boldsymbol{x_2}$, it is possible that the true model can be more significantly priced in a multiple OLS CSR than in a simple OLS CSR.

for the 10 size-ranked portfolios case.[14] It shows that once $\boldsymbol{x}^*$ is included in the model, no other firm-specific variables can have $\delta^2_{OLS}(x_1|x^*)$ greater than 3.640, so it is virtually impossible for a researcher to drive out x^* by putting other firm-specific variables in the model, unless the other variables are correlated with x^*. Ironically, the one that puts up the best competition against x^* is a variable with only $\rho^2_{OLS}(x_1|x^*) = 0.189$. Choosing variables based on the t-ratios in the multiple OLS CSR does not help us drive out bad variables obtained from the simple OLS CSR. Instead, this practice often proposes more bad variables to be included in the model.

3. The GLS Cross-Sectional Regressions

For models with a single firm-specific variable, the GLS CSR estimate of the model (with the true variance-covariance matrix V known) at time t is

$$\hat{\gamma}_t^{GLS} \equiv \begin{bmatrix} \hat{\gamma}_{0t}^{GLS} \\ \hat{\gamma}_{1t}^{GLS} \end{bmatrix} = (X'V^{-1}X)^{-1}(X'V^{-1}R_t). \tag{23}$$

It is easy to verify that it is an unbiased estimator of $\gamma^{GLS} \equiv (X'V^{-1}X)^{-1}X'V^{-1}\mu$. The t-test of $H_0 : \gamma_1^{OLS}$ is given by

$$t_{GLS} = \frac{\bar{\hat{\gamma}}_1^{GLS}}{s(\hat{\gamma}_1^{GLS})/\sqrt{T}}, \tag{24}$$

where $\bar{\hat{\gamma}}_1^{GLS}$ and $s(\hat{\gamma}_1^{GLS})$ are the sample average and standard deviation of $\hat{\gamma}_{1t}^{GLS}$. Under the assumption that R_ts are independently distributed as $N(\mu, V)$, t_{GLS} has a noncentral t-distribution with $T-1$ degrees of freedom and the square of its noncentrality parameter is given by

$$\delta^2_{GLS}(x) = \frac{T(\tilde{x}'\tilde{M}\tilde{\mu})^2}{\tilde{x}'\tilde{M}\tilde{x}}, \tag{25}$$

where $\tilde{x} = V^{-\frac{1}{2}}x$, $\tilde{\mu} = V^{-\frac{1}{2}}\mu$, and $\tilde{M} = I_N - \tilde{1}_N(\tilde{1}'_N\tilde{1}_N)^{-1}\tilde{1}'_N$ with $\tilde{1}_N = V^{-\frac{1}{2}}1_N$.

[14] Note that the lower bound and upper bound of $\delta^2(x_1|x^*)$ is the same whether x_1 is correlated with x^* or not.

For GLS CSR with true $\boldsymbol{V}$, it is easy to show that $\boldsymbol{\delta^2_{GLS}(x)}$ is an increasing and linear function of $\boldsymbol{\rho^2_{GLS}(x)}$, defined as

$$\boldsymbol{\rho^2_{GLS}(x) = 1 - \frac{(\mu - X\gamma^{GLS})' V^{-1} (\mu - X\gamma^{GLS})}{\tilde{\mu}' \tilde{M} \tilde{\mu}} = \frac{(\tilde{x}' \tilde{M} \tilde{\mu})^2}{(\tilde{x}' \tilde{M} \tilde{x})(\tilde{\mu}' \tilde{M} \tilde{\mu})}}. \tag{26}$$

Therefore, if one can justify using $\boldsymbol{\rho^2_{GLS}(x)}$ as a measure of the explanatory power of the model, then the misalignment of the statistical significance of the $\boldsymbol{t}$-ratio and the explanatory power of the model does not occur in GLS CSR. This fact also carries over to the multiple firm-specific variables case.[15] Coupled with the fact that under the true model, the GLS $\boldsymbol{t}$-test (with $\boldsymbol{V}$ known) is more powerful than the OLS $\boldsymbol{t}$-test, this seems to suggest that we should use the GLS CSR instead of OLS CSR and the $\boldsymbol{t}$-ratios in the GLS CSR can be used for model selection.

However, it is important to note the following caveats. First, while $\boldsymbol{\delta^2_{GLS}(x)}$ and $\boldsymbol{\rho^2_{GLS}(x)}$ are positively related, $\boldsymbol{\rho^2_{GLS}(x)}$ as a measure of explanatory power of the model lacks justification. Unlike parameter estimation where efficiency is gained through the use of $\boldsymbol{V^{-1}}$, weighting pricing errors by $\boldsymbol{V^{-1}}$ is not natural. Second, when a model is misspecified, the GLS $\boldsymbol{t}$-test may not be always more powerful than the OLS $\boldsymbol{t}$-test. Third, the $\boldsymbol{V}$ matrix is typically unknown. The properties of the GLS with known $\boldsymbol{V}$ do not necessarily carry over to the GLS with estimated $\boldsymbol{V}$ in a finite sample. Consequently, we do not advocate the use of the $\boldsymbol{t}$-ratio in GLS CSR for model selection. In fact, when the number of assets $\boldsymbol{N}$ is greater than the number of time series observations $\boldsymbol{T}$, the estimated GLS is not feasible anyway.

As an illustration, we report in Panel C of Table 1 the results of running the GLS CSR (with the true $\boldsymbol{V}$) with a single firm-specific variable. For true GLS CSR, we find that the true model, $\boldsymbol{\mu}$, performs just as well as $\boldsymbol{x^*}$ in the case of OLS CSR, and it will be found to be priced very often. However, the $\boldsymbol{x^*}$ which is statistically most significant for the simple OLS CSR still fares very well in the simple GLS CSR. It is found to be priced almost as often as the true model. In Panel B of Table 2, we also report the

[15] Details are available upon request.

results of running a multiple GLS CSR with both $\boldsymbol{x_1}$ and $\boldsymbol{x}^*$. $\boldsymbol{x}^*$, in this case, is just as significantly priced in the multiple GLS CSR as in the multiple OLS CSR, and $\boldsymbol{x_1}$ is only marginally more significant. Therefore, without having the true model, it is very difficult to detect that $\boldsymbol{x}^*$ is a poor variable (in terms of $\boldsymbol{\rho^2_{OLS}}$) because in both simple and multiple OLS and GLS CSR, such a firm-specific variable $\boldsymbol{x}^*$ is likely to be found significantly priced.

4. The Sample Coefficient of Determination

Since the OLS $\boldsymbol{t}$-ratio has such undesirable properties for the purpose of model comparisons, a natural question to ask is whether there are other sample test statistics that would allow us to do a better job in model comparisons. In particular, we would like to infer how much the model can explain the expected returns of the test assets. If the appropriate measure of goodness-of-fit is $\boldsymbol{\rho^2_{OLS}}$, then a natural candidate is its sample counterpart, i.e., the sample coefficient of determination.[16]

In the OLS CSR of $\boldsymbol{R_t}$ on $\boldsymbol{X} = [\boldsymbol{1_N},\ \boldsymbol{x}]$, we can compute the sample coefficient of determination on a period-by-period basis, measured as[17]

$$\boldsymbol{R^2_{OLS,t}(x) = \frac{(x'MR_t)^2}{(x'Mx)(R_t'MR_t)}}, \tag{27}$$

and the average of this time series of $\boldsymbol{R^2_{OLS,t}}$ can be reported as a sample measure of goodness-of-fit, denoted as

$$\boldsymbol{R^2_{OLS}(x) = \frac{1}{T}\sum_{t=1}^{T} R^2_{OLS,t}(x)}. \tag{28}$$

Although many studies using the OLS CSR do not report $\boldsymbol{R^2_{OLS}}$, some authors do report this average $\boldsymbol{R^2_{OLS}}$ along with the $\boldsymbol{t}$-ratios in the OLS CSR.

[16]To avoid possible confusion, it should be noted that Kandel and Stambaugh (1995) use $\boldsymbol{R^2_{OLS}}$ to denote the coefficient of determination based on true expected returns. In our notation, coefficient of determination based on true expected returns is denoted as $\boldsymbol{\rho^2_{OLS}}$.

[17]The analysis is the same for the case of multiple firm-specific variables. The only change is we need to replace $\boldsymbol{x}$ with the fitted returns $\boldsymbol{\hat{R}_t^{OLS} \equiv X\gamma_t^{OLS}}$ in the definition of $\boldsymbol{R^2_{OLS,t}(x)}$.

If $\boldsymbol{x}$ is constant over time, it is probably more logical to report another sample measure of goodness-of-fit, measured by the OLS coefficient of determination between $\boldsymbol{x}$ and the average returns, $\bar{\boldsymbol{R}}$, as

$$\bar{R}^2_{OLS}(x) = \frac{(x'M\bar{R})^2}{(x'Mx)(\bar{R}'M\bar{R})}. \tag{29}$$

Few authors report this measure because it is only feasible if the firm-specific variable does not change over time.

Similar to δ^2_{OLS}, the following proposition suggests that the mappings from ρ^2_{OLS} to both $E[R^2_{OLS}]$ and $E[\bar{R}^2_{OLS}]$ are in general not single valued.

Proposition 5 *Denote* $H = E\left[\frac{vv'}{v'v}\right]$ *where* $v \sim N(P'\mu, \Lambda)$ *with* P *and* Λ *defined before Proposition 3. Suppose* $M\mu \neq cPHP'\mu$ *for any constant* c*. Then, for any given number* $\rho_0^2 \in [0, 1)$*, the sets* $\{E[R^2_{OLS}(x)] : \rho^2_{OLS}(x) = \rho_0^2\}$ *and* $\{E[\bar{R}^2_{OLS}(x)] : \rho^2_{OLS}(x) = \rho_0^2\}$ *contain more than one element.*

In Figure 5, we plot the lower bound and upper bound of $E[R^2_{OLS}]$ for $0 \leq \rho^2_{OLS} \leq 1$ for the case of 10 size-ranked portfolios. A number of interesting observations can be made about Figure 5. First, although the lower bound and upper bound of $E[R^2_{OLS}]$ are mostly increasing in ρ^2_{OLS}, the maximal value of $E[R^2_{OLS}]$ is 0.5597 and it is attained by a firm-specific variable with $\rho^2_{OLS} = 0.925$. Second, even when $\rho^2_{OLS} = 0$, the $E[R^2_{OLS}]$ is not equal to zero.[18] Similarly, for the correctly specified model with $\rho^2_{OLS} = 1$, we only have $E[R^2_{OLS}] = 0.521$ and it is nowhere near its population value. Note that increasing the length of the time series only makes the realized value of R^2_{OLS} closer to its expected value but it will not change its expected value. Third, there is a wide range between the lower bound and upper bound of $E[R^2_{OLS}]$. For example, a firm-specific variable with $\rho^2_{OLS} = 0$ can have $E[R^2_{OLS}] = 0.1449$ but, for some firm-specific variable with $\rho^2_{OLS} = 0.456$, its $E[R^2_{OLS}]$ can be as low as 0.1446. Therefore, like the OLS t-ratio, using R^2_{OLS} to select models could also be seriously questionable.

[18] We rule out the case of $x = k1_N$ for some scalar k. When $x = k1_N$, then $R^2_{OLS} = R^2_{GLS} = 0$.

While for the 10 size-ranked portfolios case, the firm-specific variable with the highest $E[R^2_{OLS}]$ is a good model, this is not always the case for other test assets. In Figure 6, we do a similar plot for the 100 size-beta-ranked portfolios. As we can see, for some firm-specific variable with $\rho^2_{OLS} = 0$, it can have $E[R^2_{OLS}]$ as high as 0.232 whereas for the correctly specified model, its $E[R^2_{OLS}]$ is only 0.0822. In addition, the maximum $E[R^2_{OLS}]$ is attained for a variable with only $\rho^2_{OLS} = 0.078$. Therefore, picking models based on R^2_{OLS} will also give us very bad models in the case of 100 size-beta-ranked portfolios.

On the other hand, although the mapping between ρ^2_{OLS} and $E[\bar{R}^2_{OLS}]$ is not single valued when T is finite, $\bar{R}^2_{OLS}$ has a more desirable asymptotic property than $R^2_{OLS}(x)$ because as $T \to \infty$, $\bar{R} \to \mu$ and $\bar{R}^2_{OLS}(x) \to \rho^2_{OLS}(x)$. In Figures 7 and 8, we present the lower bound and upper bound of $E[\bar{R}^2_{OLS}]$ for $0 \leq \rho^2_{OLS} \leq 1$ for the cases of 10 size-ranked portfolios and 100 size-beta ranked portfolios, respectively. When $T = 330$, although there is still a range of $E[\bar{R}^2_{OLS}]$ for a given value of ρ^2_{OLS}, the range is much narrower than the ones for $E[R^2_{OLS}]$ and hence good models are more likely to have higher $E[\bar{R}^2_{OLS}]$. Therefore, for models with constant firm-specific variables, it is advisable to use $\bar{R}^2_{OLS}$, instead of R^2_{OLS}, to rank models.

For the case of GLS CSR of R_t on X, we can compute the sample coefficient of determination on a period-by-period basis, measured as

$$R^2_{GLS,t}(x) = \frac{(\tilde{x}'\tilde{M}\tilde{R}_t)^2}{(\tilde{x}'\tilde{M}\tilde{x})(\tilde{R}_t'\tilde{M}\tilde{R}_t)}, \tag{30}$$

where $\tilde{R}_t = V^{-\frac{1}{2}}R_t$. The average of this time series of $R^2_{GLS,t}$ can be reported as a sample measure of goodness-of-fit for the GLS CSR, denoted as

$$R^2_{GLS}(x) = \frac{1}{T}\sum_{t=1}^{T} R^2_{GLS,t}(x). \tag{31}$$

When x is constant over time, we can also compute the GLS coefficient of determination between x and $\bar{R}$ as

$$\bar{R}^2_{GLS}(x) = \frac{(\tilde{x}'\tilde{M}\tilde{\bar{R}})^2}{(\tilde{x}'\tilde{M}\tilde{x})(\tilde{\bar{R}}'\tilde{M}\tilde{\bar{R}})}, \tag{32}$$

where $\tilde{\bar{\boldsymbol{R}}} = \boldsymbol{V}^{-\frac{1}{2}}\bar{\boldsymbol{R}}$.

Similar to the case of $\boldsymbol{\delta^2_{GLS}}$, the following proposition shows that both $\boldsymbol{E}[\boldsymbol{R^2_{GLS}}]$ and $\boldsymbol{E}[\bar{\boldsymbol{R}}^2_{\boldsymbol{GLS}}]$ are linear functions of $\boldsymbol{\rho^2_{GLS}}$.

Proposition 6 *For a firm-specific variable* $\boldsymbol{x}$*, we have*

$$\boldsymbol{E}[\boldsymbol{R^2_{GLS}(x)}] = \boldsymbol{h} + [\boldsymbol{1} - (\boldsymbol{N}-\boldsymbol{1})\boldsymbol{h}]\boldsymbol{\rho^2_{GLS}(x)}, \tag{33}$$

$$\boldsymbol{E}[\bar{\boldsymbol{R}}^2_{\boldsymbol{GLS}}\boldsymbol{(x)}] = \boldsymbol{k} + [\boldsymbol{1} - (\boldsymbol{N}-\boldsymbol{1})\boldsymbol{k}]\boldsymbol{\rho^2_{GLS}(x)}, \tag{34}$$

where

$$\boldsymbol{h} = \frac{\boldsymbol{e^{-a}}}{\boldsymbol{2}}\int_{\boldsymbol{0}}^{\boldsymbol{1}} \boldsymbol{e^{au}u^{\frac{N-3}{2}}}\mathbf{d}\boldsymbol{u}, \tag{35}$$

$$\boldsymbol{k} = \frac{\boldsymbol{e^{-aT}}}{\boldsymbol{2}}\int_{\boldsymbol{0}}^{\boldsymbol{1}} \boldsymbol{e^{aTu}u^{\frac{N-3}{2}}}\mathbf{d}\boldsymbol{u}, \tag{36}$$

with $\boldsymbol{a} = (\tilde{\boldsymbol{\mu}}'\tilde{\boldsymbol{M}}\tilde{\boldsymbol{\mu}})/\boldsymbol{2}$.

Therefore, to the extent that we can justify using $\boldsymbol{\rho^2_{GLS}}$ as a measure of goodness-of-fit, we can also use either $\boldsymbol{R^2_{GLS}}$ or $\bar{\boldsymbol{R}}^2_{\boldsymbol{GLS}}$ to rank models. In practice, since $\boldsymbol{h} \gg \boldsymbol{k}$ when $\boldsymbol{T}$ is large, it is a lot easier to distinguish good models from bad models by using $\bar{\boldsymbol{R}}^2_{\boldsymbol{GLS}}$ than by using $\boldsymbol{R^2_{GLS}}$.

5. Concluding Remarks

Since asset pricing models are at best approximations, testing whether or not an asset pricing model is literally true is not very interesting by itself. For most practical purposes, the most interesting part of empirical analysis of asset pricing models rests in the evaluation and comparison of models. In this paper, we advocate that such a task can only be accomplished by making it clear about what we mean by a better model. When the purpose of an asset pricing model is to explain the cross-sectional differences of expected returns, the goodness of a should depend only on how many of the variations in the expected returns of the test assets are explained by the explanatory variables in the model. Test statistics invariably depend also on the variance-covariance matrix of the returns on the test assets, and hence statistical significance of a variable does not always tell us about its explanatory power. This misalignment is not surprising since test statistics are designed for the purpose of testing a model but not for the purpose of comparison and ranking of models.

Instead of merely pointing out that statistical significance does not always imply good explanatory power, we establish a linkage between the two in the sense that the statistical significance of the $\boldsymbol{t}$-ratio of a variable allows us to determine the lower bound and upper bound for its explanatory power (and vice versa). This linkage allows us to understand better when misalignment of statistical significance and explanatory power occurs. For the case of the commonly used OLS CSR, we show that the $\boldsymbol{t}$-ratios are grossly inappropriate for model selection because (i) good models or even the true model do not always have high expected value of OLS $\boldsymbol{t}$-ratios and (ii) in many situations, a high OLS $\boldsymbol{t}$-ratio attained by a candidate variable often suggests that the variable does not explain the expected returns of the test assets very well. For the case of true GLS CSR, the $\boldsymbol{t}$-ratio is more desirable because the true model will be found priced more often than the wrong models. However, for the purpose of model comparison, there is still a misalignment of statistical significance with explanatory power in general. Based on these results, we suggest that model selection based exclusively on $\boldsymbol{t}$-ratios can lead to acceptance of many poor variables. Therefore, it is extremely important to include other diagnostics, such as the sample counterparts of the measure of goodness-of-fit, in model comparison.

Appendix

Proof of Proposition 1: (i) From (2) and the assumption that $\boldsymbol{E}[\boldsymbol{R_t}] = \boldsymbol{\mu}$, it is easy to verify that

$$\boldsymbol{E}[\hat{\gamma}_t^{OLS}] = (\boldsymbol{X}'\boldsymbol{X})^{-1}(\boldsymbol{X}'\boldsymbol{\mu}) = \gamma^{OLS}. \tag{A1}$$

Thus, $\hat{\gamma}_t^{OLS}$ is an unbiased estimator of γ^{OLS}.

(ii) It is easy to verify that

$$\hat{\gamma}_{1t}^{OLS} = \frac{\boldsymbol{x}'\boldsymbol{M}\boldsymbol{R_t}}{\boldsymbol{x}'\boldsymbol{M}\boldsymbol{x}}. \tag{A2}$$

Therefore, from the independence and normality assumption of $\boldsymbol{R_t}$, we have

$$\hat{\gamma}_{1t}^{OLS} \overset{\text{i.i.d.}}{\sim} \boldsymbol{N}\left(\gamma_1^{OLS}, \frac{\boldsymbol{x}'\boldsymbol{M}\boldsymbol{V}\boldsymbol{M}\boldsymbol{x}}{(\boldsymbol{x}'\boldsymbol{M}\boldsymbol{x})^2}\right). \tag{A3}$$

Using results in Johnson, Kotz and Balakrishnan (1995, Ch. 31), $\boldsymbol{t_{OLS}}$ has a noncentral $\boldsymbol{t}$-distribution with its square of noncentrality parameter equal to $\boldsymbol{\delta^2_{OLS}(x)}$.

Proof of Proposition 2: We begin by citing a well known result here for later use. By the Cauchy-Schwarz inequality, for a fixed vector $\boldsymbol{a}$ and a square matrix $\boldsymbol{B}$,

$$(\boldsymbol{z}'\boldsymbol{a})^2 = [(\boldsymbol{z}'\boldsymbol{B}^{\frac{1}{2}})(\boldsymbol{B}^{-\frac{1}{2}}\boldsymbol{a})]^2 \leq (\boldsymbol{z}'\boldsymbol{B}\boldsymbol{z})(\boldsymbol{a}'\boldsymbol{B}^{-1}\boldsymbol{a})$$

with equality holds if and only if $\boldsymbol{z}$ is proportional to $\boldsymbol{B}^{-1}\boldsymbol{a}$. Therefore,

$$\max_{\boldsymbol{z}} \frac{(\boldsymbol{z}'\boldsymbol{a})^2}{\boldsymbol{z}'\boldsymbol{B}\boldsymbol{z}} = \boldsymbol{a}'\boldsymbol{B}^{-1}\boldsymbol{a}, \tag{A4}$$

and the maximum is attained if and only if $\boldsymbol{c}\boldsymbol{z} = \boldsymbol{B}^{-1}\boldsymbol{a}$ for some constant $\boldsymbol{c} \neq \boldsymbol{0}$.

Since $\boldsymbol{M}\boldsymbol{V}\boldsymbol{M} = \boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}'$ where $\boldsymbol{P}$ and $\boldsymbol{\Lambda}$ are defined in the main text before Proposition 3, we have $\boldsymbol{P}\boldsymbol{P}' = \boldsymbol{M}$[19] and

$$\delta^2_{OLS}(x) = \frac{\boldsymbol{T}(\boldsymbol{x}'\boldsymbol{M}\boldsymbol{\mu})^2}{\boldsymbol{x}'\boldsymbol{M}\boldsymbol{V}\boldsymbol{M}\boldsymbol{x}} = \frac{\boldsymbol{T}(\boldsymbol{x}'\boldsymbol{P}\boldsymbol{P}'\boldsymbol{\mu})^2}{\boldsymbol{x}'\boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}'\boldsymbol{x}} = \frac{\boldsymbol{T}(\boldsymbol{z}'\boldsymbol{a})^2}{\boldsymbol{z}'\boldsymbol{\Lambda}\boldsymbol{z}} \tag{A5}$$

[19]Since $\boldsymbol{1}'_{\boldsymbol{N}}\boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}'\boldsymbol{1_N} = \boldsymbol{1}'_{\boldsymbol{N}}\boldsymbol{M}\boldsymbol{V}\boldsymbol{M}\boldsymbol{1_N} = \boldsymbol{0}$, $\boldsymbol{P}'\boldsymbol{1_N}$ must be a zero vector, $\boldsymbol{0_{N-1}}$, because $\boldsymbol{\Lambda}$ is positive definite. Therefore, the columns of $\boldsymbol{P}$ are orthogonal to $\boldsymbol{1_N}$ and $\boldsymbol{P}$ is a basis for $[\boldsymbol{1_N}]^{\perp}$. The projection matrix onto $[\boldsymbol{1_N}]^{\perp}$ is $\boldsymbol{M} = \boldsymbol{P}(\boldsymbol{P}'\boldsymbol{P})^{-1}\boldsymbol{P}' = \boldsymbol{P}\boldsymbol{P}'$.

by writing $\boldsymbol{z} = \boldsymbol{P}'\boldsymbol{x}$ and $\boldsymbol{a} = \boldsymbol{P}'\boldsymbol{\mu}$. Invoking our earlier results, $\boldsymbol{\mu}$ is one of the $\boldsymbol{x}$s which maximize $\delta^2_{OLS}(\boldsymbol{x})$ if and only if $\boldsymbol{cP}'\boldsymbol{\mu} = \boldsymbol{\Lambda}^{-1}\boldsymbol{P}'\boldsymbol{\mu}$, or $\boldsymbol{cP\Lambda P}'\boldsymbol{\mu} = \boldsymbol{PP}'\boldsymbol{\mu}$, or $\boldsymbol{cMVM\mu} = \boldsymbol{M\mu}$ for some constant $\boldsymbol{c} \neq \boldsymbol{0}$. This completes the proof. *Q.E.D.*

Proof of Proposition 3: The cases $\boldsymbol{\rho}^2_{OLS} = \boldsymbol{0}$ and $\boldsymbol{1}$ are easy to prove. For the case $\boldsymbol{\rho}^2_{\boldsymbol{0}} \in (\boldsymbol{0}, \boldsymbol{1})$, we will provide a constructive proof for calculating the maximal and minimal $\delta^2_{OLS}(\boldsymbol{x})$ for $\boldsymbol{\rho}^2_{OLS}(\boldsymbol{x}) = \boldsymbol{\rho}^2_{\boldsymbol{0}}$, after we introduce some notations. When $\boldsymbol{\mu} \neq \boldsymbol{k}\boldsymbol{1}_N$ for any scalar $\boldsymbol{k}$, $\boldsymbol{\mu}'\boldsymbol{M\mu} \neq \boldsymbol{0}$ and we can define

$$\boldsymbol{\eta} = \frac{\boldsymbol{P}'\boldsymbol{\mu}}{(\boldsymbol{\mu}'\boldsymbol{M\mu})^{\frac{1}{2}}} \equiv (\eta_1, \ldots, \eta_{N-1})', \tag{A6}$$

where $\boldsymbol{P}$ is defined in the main text before Proposition 3. Since $\boldsymbol{\eta}'\boldsymbol{\eta} = \boldsymbol{1}$, η_i^2 measures the fraction of the cross-sectional variance of $\boldsymbol{\mu}$ that is explained by the $\boldsymbol{i}$-th column of $\boldsymbol{P}$.[20] Denote the set of $\boldsymbol{i}$ such that $\eta_i \neq \boldsymbol{0}$ as $\boldsymbol{S}$, i.e.,

$$\boldsymbol{S} = \{i : \eta_i \neq 0,\ 1 \leq i \leq N - 1\}. \tag{A7}$$

For each nonzero eigenvalue of $\boldsymbol{MVM}$, $\boldsymbol{\lambda}_i$, we denote the set of $\boldsymbol{j}$ such that $\boldsymbol{\lambda}_j = \boldsymbol{\lambda}_i$ as $\boldsymbol{S}_i$, i.e.,

$$\boldsymbol{S}_i = \{j : \lambda_j = \lambda_i,\ 1 \leq j \leq N - 1\}. \tag{A8}$$

That $\boldsymbol{M\mu} \neq \boldsymbol{cMVM\mu}$ is equivalent to $\boldsymbol{\eta} \neq \boldsymbol{c\Lambda\eta}$ where $\boldsymbol{\Lambda}$ is defined in the proof of Proposition 2. It follows that there are at least two elements $\boldsymbol{i}$ and $\boldsymbol{j}$ from $\boldsymbol{S}$ such that $\boldsymbol{\lambda}_i \neq \boldsymbol{\lambda}_j$. For $\boldsymbol{0} < \boldsymbol{\rho}^2_{\boldsymbol{0}} < \boldsymbol{1}$, we define two sets[21]

$$\begin{aligned}
\Phi_1^\lambda &= \{\phi : f_\lambda(\phi) = \rho_0^2\}, \\
\Phi_2^\lambda &= \{\phi : \phi = \lambda_i \text{ for some } i \text{ if } f_\lambda(\lambda_i) > \rho_0^2, \text{ and} \\
&\qquad \eta_i(\lambda_i - \lambda_j) = 0 \text{ for at least one } j \neq i\},
\end{aligned}$$

where

$$f_\lambda(\phi) = \begin{cases} \frac{\left[\sum_{i\in S} \eta_i^2(\lambda_i - \phi)^{-1}\right]^2}{\sum_{i\in S} \eta_i^2(\lambda_i-\phi)^{-2}}, & \text{if } \phi \neq \lambda_i,\ i \in S; \\ \sum_{j\in S_i} \eta_j^2, & \text{if } \phi = \lambda_i,\ i \in S. \end{cases} \tag{A9}$$

[20]This is because the columns of $\boldsymbol{P}$ are orthogonal to each other and to $\boldsymbol{1}_N$ (i.e., their mean is 0), and their norm is 1, elements of $\boldsymbol{\eta}$ are the correlation coefficients between $\boldsymbol{\mu}$ and the columns of $\boldsymbol{P}$.

[21]In the usual case that $\eta_i \neq \boldsymbol{0}$ for all $\boldsymbol{i}$ and $\boldsymbol{\lambda}_i$s are distinct, Φ_2^λ is empty.

Note both $\boldsymbol{\Phi_1^\lambda}$ and $\boldsymbol{\Phi_2^\lambda}$ depend on $\boldsymbol{\rho_0^2}$.[22] We begin by listing some properties of $\boldsymbol{f_\lambda}$.

(i) $\boldsymbol{f_\lambda(\phi)}$ is continuous on $\boldsymbol{(-\infty, \infty)}$.

(ii) $\boldsymbol{0 \leq f_\lambda(\phi) \leq 1}$; $\boldsymbol{\lim_{\phi \to -\infty} f_\lambda(\phi) = \lim_{\phi \to \infty} f_\lambda(\phi) = 1}$.

(iii) For $\boldsymbol{0 < \rho_0^2 < 1}$, $\boldsymbol{\Phi_1^\lambda}$ contains at least two elements.

These properties can be easily verified as follows:

(i) The continuity needs to be verified only at $\boldsymbol{\phi = \lambda_i}$, $\boldsymbol{i \in S}$. By L'Hôpital's rule, $\boldsymbol{\lim_{\phi \to \lambda_i} f_\lambda(\phi) = \sum_{j \in S_i} \eta_j^2}$, so $\boldsymbol{f_\lambda(\phi)}$ is continuous.

(ii) Let

$$\boldsymbol{g_\lambda(\phi) = \sum_{i \in S} \frac{\eta_i^2}{\lambda_i - \phi}, \qquad \phi \neq \lambda_i,\ i \in S,} \tag{A10}$$

$$\boldsymbol{h_\lambda(\phi) = \sum_{i \in S} \frac{\eta_i^2}{(\lambda_i - \phi)^2}, \qquad \phi \neq \lambda_i,\ i \in S.} \tag{A11}$$

Then by the Cauchy-Schwarz inequality and $\boldsymbol{\sum_{i \in S} \eta_i^2 = \eta' \eta = 1}$,

$$\boldsymbol{g_\lambda(\phi)^2 = \left(\sum_{i \in S} \eta_i [\eta_i (\lambda_i - \phi)^{-1}] \right)^2 \leq \left(\sum_{i \in S} \eta_i^2 \right) \left(\sum_{i \in S} \eta_i^2 (\lambda_i - \phi)^{-2} \right) = h_\lambda(\phi).}$$

Therefore $\boldsymbol{0 \leq f_\lambda(\phi) = g_\lambda(\phi)^2 / h_\lambda(\phi) \leq 1}$. The proof for the limit as $\boldsymbol{\phi \to \pm\infty}$ is elementary.

(iii) It is easy to check that for all $\boldsymbol{i \in S}$, $\boldsymbol{\lim_{\phi \to \lambda_i^-} g_\lambda(\phi) = \infty}$, and $\boldsymbol{\lim_{\phi \to \lambda_i^+} g_\lambda(\phi) = -\infty}$. Since $\boldsymbol{S}$ contains at least two elements for which the eigenvalues are distinct, take $\boldsymbol{i}$ and $\boldsymbol{j}$ from $\boldsymbol{S}$ such that $\boldsymbol{\lambda_i < \lambda_j}$ are closest to each other. Since $\boldsymbol{g_\lambda(\phi)}$ is continuous on any closed interval contained in $\boldsymbol{(\lambda_i, \lambda_j)}$, there exists a $\boldsymbol{\phi'}$ in between such that $\boldsymbol{g_\lambda(\phi') = 0}$. Since $\boldsymbol{h_\lambda(\phi') > 0}$, we also have $\boldsymbol{f_\lambda(\phi') = 0}$. From (i) and (ii), there exist at least two $\boldsymbol{\phi}$'s, one less than $\boldsymbol{\phi'}$ and the other greater than $\boldsymbol{\phi'}$, such that $\boldsymbol{f_\lambda(\phi) = \rho_0^2}$ for $\boldsymbol{0 < \rho_0^2 < 1}$.

[22] By multiplying the numerator and denominator of $\boldsymbol{f_\lambda(\phi)}$ by $\boldsymbol{\prod_{i \in S} (\lambda_i - \phi)^2}$, we can see that the solutions to $\boldsymbol{f_\lambda(\phi) = \rho_0^2}$ are in fact solutions to a polynomial, which are easy to obtain numerically.

Now we are ready to prove the rest of Proposition 3. For $\mathbf{0} < \boldsymbol{\rho}_0^2 < \mathbf{1}$, the solution to the minimal and maximal $\boldsymbol{\delta}_{OLS}^2$ given $\boldsymbol{\rho}_{OLS}^2(\boldsymbol{x}) = \boldsymbol{\rho}_0^2$ are given by

$$\min_{x:\rho_{OLS}^2(x)=\rho_0^2} \delta_{OLS}^2(x) = \frac{T(\mu' M \mu)\rho_0^2}{\phi^{**} + \frac{\rho_0^2}{\sum_{i\in S}\eta_i^2(\lambda_i-\phi^{**})^{-1}}}, \tag{A12}$$

$$\max_{x:\rho_{OLS}^2(x)=\rho_0^2} \delta_{OLS}^2(x) = \frac{T(\mu' M \mu)\rho_0^2}{\phi^{*} + \frac{\rho_0^2}{\sum_{i\in S}\eta_i^2(\lambda_i-\phi^{*})^{-1}}}, \tag{A13}$$

where $\boldsymbol{\phi}^* = \min \boldsymbol{\Phi}_1^{\lambda} \cup \boldsymbol{\Phi}_2^{\lambda}$ and $\boldsymbol{\phi}^{**} = \max \boldsymbol{\Phi}_1^{\lambda} \cup \boldsymbol{\Phi}_2^{\lambda}$. The proof for maximal $\boldsymbol{\delta}_{OLS}^2$ is given here. The proof for minimal is similar. First, the scale of $\boldsymbol{x}$ does not change $\boldsymbol{\delta}_{OLS}^2$ and $\boldsymbol{\rho}_{OLS}^2$, so we can normalize $\boldsymbol{x}$ such that $\boldsymbol{x'Mx} = \mathbf{1}$. Let $\boldsymbol{y} = \boldsymbol{P'x}$. Then

$$\rho_{OLS}^2 = (\eta' y)^2, \tag{A14}$$

$$\delta_{OLS}^2 = \frac{T(\mu' M \mu)(\eta' y)^2}{y'\Lambda y}. \tag{A15}$$

Therefore, the maximization of $\boldsymbol{\delta}_{OLS}^2$ for a given $\boldsymbol{\rho}_{OLS}^2 = \boldsymbol{\rho}_0^2$ is the solution to the following problem.

$$\min_{y}\ y'\Lambda y, \qquad \text{s.t. } \eta' y = \rho_0, \quad y'y = 1.$$

The Lagrange function of the problem is,

$$\mathcal{L}(y, \tau, \phi) = \frac{1}{2} y'\Lambda y - \tau(\eta' y - \rho_0) - \frac{1}{2}\phi(y'y - 1). \tag{A16}$$

The first order condition is

$$(\Lambda - \phi I_{N-1})y = \tau\eta. \tag{A17}$$

Let us consider $\boldsymbol{\phi} \neq \boldsymbol{\lambda}_i$, $\boldsymbol{i} = \mathbf{1}, \ldots, \boldsymbol{N} - \mathbf{1}$. Then, $\boldsymbol{y} = \boldsymbol{\tau}(\boldsymbol{\Lambda} - \boldsymbol{\phi I}_{N-1})^{-1}\boldsymbol{\eta}$. From the linear constraint,

$$\tau = \frac{\rho_0}{\eta'(\Lambda - \phi I_{N-1})^{-1}\eta} = \frac{\rho_0}{\sum_{i\in S}\eta_i^2(\lambda_i - \phi)^{-1}}. \tag{A18}$$

Combined with the quadratic constraint,

$$\left[\eta'(\Lambda - \phi I_{N-1})^{-1}\eta\right]^2 = \rho_0^2\eta'(\Lambda - \phi I_{N-1})^{-2}\eta, \tag{A19}$$

which can be rewritten as $f_\lambda(\phi) = \rho_0^2$.

There are possibly other solutions to the first order condition for which $\phi = \lambda_i$ for some i. There are two cases to consider. (a) If $S_i \cap S$ is not empty, then there is at least one $j \in S_i$ such that $\eta_j \neq 0$. For the first order condition to hold, τ must be zero, $y_j = 0$ for all $j \notin S_i$, $\sum_{j \in S_i} y_j \eta_j = \rho_0$, and $\sum_{j \in S_i} y_j^2 = 1$. From the Cauchy-Schwarz inequality, $f_\lambda(\lambda_i) = \sum_{j \in S_i} \eta_j^2 \geq (\sum_{j \in S_i} y_j \eta_j)^2 / \sum_{j \in S_i} y_j^2 = \rho_0^2$. If the equality holds, then $\lambda_i \in \Phi_1^\lambda$. Otherwise, $\lambda_i \in \Phi_2^\lambda$.[23] (b) If $S_i \cap S$ is empty, then $\eta_j = 0$ for all $j \in S_i$. The first order condition requires $y_j = \tau \eta_j / (\lambda_j - \lambda_i)$ for $j \notin S_i$. The linear constraint implies

$$\tau = \frac{\rho_0}{\sum_{j \notin S_i} \eta_j^2 (\lambda_j - \lambda_i)^{-1}} = \frac{\rho_0}{\sum_{j \in S} \eta_j^2 (\lambda_j - \lambda_i)^{-1}}.$$

Summing up y_j^2 for $j \notin S_i$, we have

$$\sum_{j \notin S_i} y_j^2 = \tau^2 \sum_{j \notin S_i} \frac{\eta_j^2}{(\lambda_j - \lambda_i)^2} = \frac{\rho_0^2 \sum_{j \in S} \eta_j^2 (\lambda_j - \lambda_i)^{-2}}{\left[\sum_{j \in S} \eta_j^2 (\lambda_j - \lambda_i)^{-1}\right]^2} = \frac{\rho_0^2}{f_\lambda(\lambda_i)}.$$

If $\sum_{j \notin S_i} y_j^2 = 1$, we have $f_\lambda(\lambda_i) = \rho_0^2$ and $\lambda_i \in \Phi_1^\lambda$. If $\sum_{j \notin S_i} y_j^2 < 1$, then $\lambda_i \in \Phi_2^\lambda$. On the other hand, it is easy to see that all $\phi \in \Phi_1^\lambda \cup \Phi_2^\lambda$ with corresponding τs and ys are indeed solutions to the first order condition.

Now for $\phi \in \Phi_1^\lambda \cup \Phi_2^\lambda$, premultiplying y' to the first order condition gives,[24]

$$y' \Lambda y = \phi + \tau \rho_0 = \phi + \frac{\rho_0^2}{\sum_{i \in S} \eta_i^2 (\lambda_i - \phi)^{-1}}. \tag{A20}$$

We now show that the minimum of $y' \Lambda y$ is attained at the minimum in $\Phi_1^\lambda \cup \Phi_2^\lambda$. By

[23] In the case when λ_i has multiplicity of one and $i \in S$, we can only have $f_\lambda(\lambda_i) = \rho_0^2$ but not $f_\lambda(\lambda_i) > \rho_0^2$, and λ_i cannot be in Φ_2^λ. This requirement is reflected in the definition of Φ_2^λ.

[24] If $\phi = \lambda_j$, $j \in S$, then $\tau = 0$ but (A20) still holds since $\lim_{\phi \to \lambda_j} \frac{\rho_0^2}{\sum_{i \in S} \eta_i^2 (\lambda_i - \phi)^{-1}} = 0$.

(iii) we can choose ϕ_1 and ϕ_2, with $\phi_1 < \phi_2$, either from Φ_1^λ or from Φ_2^λ. By definition,

$$\rho_0^4 \leq f_\lambda(\phi_1) f_\lambda(\phi_2) = \frac{\left[\sum_{i\in S} \frac{\eta_i^2}{(\lambda_i-\phi_1)}\right]^2 \left[\sum_{i\in S} \frac{\eta_i^2}{(\lambda_i-\phi_2)}\right]^2}{\left[\sum_{i\in S} \frac{\eta_i^2}{(\lambda_i-\phi_1)^2}\right] \left[\sum_{i\in S} \frac{\eta_i^2}{(\lambda_i-\phi_2)^2}\right]}$$
$$< \frac{\left[\sum_{i\in S} \frac{\eta_i^2}{(\lambda_i-\phi_1)}\right]^2 \left[\sum_{i\in S} \frac{\eta_i^2}{(\lambda_i-\phi_2)}\right]^2}{\left[\sum_{i\in S} \frac{\eta_i^2}{(\lambda_i-\phi_1)(\lambda_i-\phi_2)}\right]^2}, \tag{A21}$$

with the last inequality being the Cauchy-Schwarz inequality. Note that the Cauchy-Schwarz inequality is strict because there are at least two distinct λ_is, for which $i \in S$. Therefore,

$$\begin{aligned} & y(\phi_2)'\Lambda y(\phi_2) - y(\phi_1)'\Lambda y(\phi_1) \\ &= (\phi_2 - \phi_1)\left[1 - \frac{\rho_0^2 \sum_{i\in S} \frac{\eta_i^2}{(\lambda_i-\phi_1)(\lambda_i-\phi_2)}}{\sum_{i\in S} \frac{\eta_i^2}{(\lambda_i-\phi_1)} \sum_{i\in S} \frac{\eta_i^2}{(\lambda_i-\phi_2)}}\right] > 0. \end{aligned}$$

This inequality also shows that the set of $\{\delta_{OLS}^2(x) : \rho_{OLS}^2(x) = \rho_0^2\}$ for a given $\rho_0^2 \in (0, 1)$ is more than a singleton.

Finally, that $\max_{x:\rho_{OLS}^2(x)=\rho_0^2} \delta_{OLS}^2(x)$ is not a monotonically increasing function of ρ_0^2 follows from Proposition 2 and the fact that $\max_{x:\rho_{OLS}^2(x)=\rho_0^2} \delta_{OLS}^2(x)$ is continuous in ρ_0^2. This completes the proof. *Q.E.D.*

Proof of Lemma 1: Let $H = (P, \frac{1_N}{\sqrt{N}})$ where P is defined in the main text before Proposition 3. H is orthonormal and hence the eigenvalues of $H'VH$ are the same as the eigenvalues of V. But since

$$H'VH = \begin{pmatrix} P'VP & \frac{P'V1_N}{\sqrt{N}} \\ \frac{1_N'VP}{\sqrt{N}} & \frac{1_N'V1_N}{N} \end{pmatrix} = \begin{pmatrix} \Lambda & \frac{P'V1_N}{\sqrt{N}} \\ \frac{1_N'VP}{\sqrt{N}} & \frac{1_N'V1_N}{N} \end{pmatrix},$$

then by the Sturmian interlacing inequalities (see for example, Horn and Johnson (1990, p.185)), (8) holds. *Q.E.D.*

Proof of Lemma 2: Define P and Λ as in the proof of Proposition 2. By the Kantorovich's inequality (see for example, Horn and Johnson (1990, p.444)), for any $(N-1)$-vector

$\boldsymbol{y}$, we have

$$(\boldsymbol{y}'\boldsymbol{y})^2 \geq \frac{4\boldsymbol{\lambda_1\lambda_{N-1}}}{(\boldsymbol{\lambda_1+\lambda_{N-1}})^2}(\boldsymbol{y}'\boldsymbol{\Lambda y})(\boldsymbol{y}'\boldsymbol{\Lambda}^{-1}\boldsymbol{y}), \tag{A22}$$

with the equality holds when $\boldsymbol{y} = [\boldsymbol{k}, \boldsymbol{0}, \ldots, \boldsymbol{0}, \boldsymbol{k}]'$ for some constant $\boldsymbol{k} \neq \boldsymbol{0}$.[25] From the proof of Proposition 2, we have $\boldsymbol{cP}'\boldsymbol{x}^* = \boldsymbol{\Lambda}^{-1}\boldsymbol{P}'\boldsymbol{\mu}$. Then by writing $\boldsymbol{y} = \boldsymbol{\Lambda}^{-\frac{1}{2}}\boldsymbol{P}'\boldsymbol{\mu}$, we have

$$\begin{aligned}\rho^2_{OLS}(\boldsymbol{x}^*) &= \frac{(\boldsymbol{x}^{*\prime}\boldsymbol{M\mu})^2}{(\boldsymbol{x}^{*\prime}\boldsymbol{Mx}^*)(\boldsymbol{\mu}'\boldsymbol{M\mu})} = \frac{(\boldsymbol{\mu}'\boldsymbol{P}'\boldsymbol{\Lambda}^{-1}\boldsymbol{P}'\boldsymbol{\mu})^2}{(\boldsymbol{\mu}'\boldsymbol{P\Lambda}^{-2}\boldsymbol{P}'\boldsymbol{\mu})(\boldsymbol{\mu}'\boldsymbol{PP}'\boldsymbol{\mu})} \\ &= \frac{(\boldsymbol{y}'\boldsymbol{y})^2}{(\boldsymbol{y}'\boldsymbol{\Lambda}^{-1}\boldsymbol{y})(\boldsymbol{y}'\boldsymbol{\Lambda y})} \geq \frac{4\boldsymbol{\lambda_1\lambda_{N-1}}}{(\boldsymbol{\lambda_1+\lambda_{N-1}})^2},\end{aligned}$$

with equality holds when $\boldsymbol{y} = [\boldsymbol{k}, \boldsymbol{0}, \ldots, \boldsymbol{0}, \boldsymbol{k}]'$ for some constant $\boldsymbol{k} \neq \boldsymbol{0}$. Since $\boldsymbol{y}'\boldsymbol{\Lambda y} = \boldsymbol{\mu}'\boldsymbol{M\mu}$, this implies $\boldsymbol{k}^2 = \frac{\boldsymbol{\mu}'\boldsymbol{M\mu}}{\boldsymbol{\lambda_1+\lambda_{N-1}}}$ and hence

$$\boldsymbol{\eta}_i^2 = \frac{\boldsymbol{\lambda_i y_i^2}}{\boldsymbol{\mu}'\boldsymbol{M\mu}} = \begin{cases} \frac{\boldsymbol{\lambda_i}}{\boldsymbol{\lambda_1+\lambda_{N-1}}} & \text{if } \boldsymbol{i=1} \text{ or } \boldsymbol{N-1}; \\ \boldsymbol{0} & \text{otherwise.} \end{cases}$$

where $\boldsymbol{\eta_i}$ is defined in (A6). *Q.E.D.*

Proof of Proposition 4: The proof is almost identical to the proof of Proposition 2, so we only sketch out the beginning of it. Define $\boldsymbol{\Lambda_2}$ as a diagonal matrix with its diagonal elements equal to the $\boldsymbol{N-2}$ nonzero eigenvalues of $\boldsymbol{M_2VM_2}$ and $\boldsymbol{P_2}$ an $\boldsymbol{N \times (N-2)}$ orthonormal matrix with its columns equal to the eigenvectors of $\boldsymbol{M_2VM_2}$ associated with $\boldsymbol{\Lambda_2}$. We have $\boldsymbol{M_2VM_2} = \boldsymbol{P_2\Lambda_2P_2'}$ and $\boldsymbol{M_2} = \boldsymbol{P_2P_2'}$. The rest of the proof follows exactly like the proof of Proposition 2 by replacing $\boldsymbol{P}$ with $\boldsymbol{P_2}$, $\boldsymbol{Q}$ with $\boldsymbol{Q_2}$, and $\boldsymbol{\Lambda}$ with $\boldsymbol{\Lambda_2}$. *Q.E.D.*

Proof of Proposition 5: Since the proof of $\boldsymbol{E}[\boldsymbol{\bar{R}^2_{OLS}}]$ is almost identical to that of $\boldsymbol{E}[\boldsymbol{R^2_{OLS}}]$, we will only prove the case of $\boldsymbol{E}[\boldsymbol{R^2_{OLS}}]$ here. To obtain the analytical expression of $\boldsymbol{H}$, we define $\boldsymbol{\nu} = \boldsymbol{\Lambda}^{-\frac{1}{2}}\boldsymbol{P}'\boldsymbol{\mu} \equiv (\boldsymbol{\nu_1}, \ldots, \boldsymbol{\nu_{N-1}})'$ and $\boldsymbol{z} = \boldsymbol{\Lambda}^{-\frac{1}{2}}\boldsymbol{v} \sim \boldsymbol{N}(\boldsymbol{\nu}, \boldsymbol{I_{N-1}})$, we can write

$$\boldsymbol{H} = \boldsymbol{E}\left[\frac{\boldsymbol{vv}'}{\boldsymbol{v}'\boldsymbol{v}}\right] = \boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{E}\left[\frac{\boldsymbol{zz}'}{\boldsymbol{z}'\boldsymbol{\Lambda z}}\right]\boldsymbol{\Lambda}^{\frac{1}{2}}. \tag{A23}$$

[25]When there are multiple smallest or largest eigenvalues of $\boldsymbol{\Lambda}$, the equality can hold for other values of $\boldsymbol{y}$.

Using results of Sawa (1978) and Hoque (1985), the (i,j)th element of H is given by

$$\sqrt{\lambda_i\lambda_j}\int_0^\infty \frac{\exp\left(\frac{\nu'\left[(I_{N-1}+2t\Lambda)^{-1}-I_{N-1}\right]\nu}{2}\right)}{|I_{N-1}+2t\Lambda|^{\frac{1}{2}}}\left[\frac{\delta_{ij}}{1+2t\lambda_i}+\frac{\nu_i\nu_j}{(1+2t\lambda_i)(1+2t\lambda_j)}\right]\mathrm{d}t, \tag{A24}$$

where δ_{ij} is the Kronecker delta function which equals one when $i = j$ and zero otherwise. It is easy to compute H numerically to any desirable precision.[26] Let $\xi_1 \geq \cdots \geq \xi_{N-1} > 0$ be the eigenvalues of H and denote $\Xi = \mathrm{Diag}(\xi_1,\ldots,\xi_{N-1})$. Also let U be an orthonormal matrix with its columns equal to the eigenvectors of H so that $H = U\Xi U'$. When $\mu \neq k1_N$ for any scalar k, $\mu' M\mu \neq 0$ and we can define η as

$$\eta = \frac{U'P'\mu}{(\mu' M\mu)^{\frac{1}{2}}} \equiv (\eta_1,\ldots,\eta_{N-1})', \tag{A25}$$

S and S_i as in (A7) and (A8). Also define f_ξ similar to f_λ as in (A9), we have

$$f_\xi(\phi) = \begin{cases} \frac{\left[\sum_{i\in S}\eta_i^2(\xi_i-\phi)^{-1}\right]^2}{\sum_{i\in S}\eta_i^2(\xi_i-\phi)^{-2}}, & \text{if } \phi \neq \xi_i,\ i \in S; \\ \sum_{j\in S_i}\eta_j^2, & \text{if } \phi = \xi_i,\ i\in S. \end{cases} \tag{A26}$$

Since the scale of x does not change $E[R^2_{OLS}]$ and ρ^2_{OLS}, we can normalize x such that $x'Mx = 1$. Defining $v = P'R_t$, we have

$$\begin{aligned} E[R^2_{OLS}(x)] &= E[R^2_{OLS,t}(x)] = E\left[\frac{(x'MR_t)^2}{(x'Mx)(R_t'MR_t)}\right] \\ &= (x'P)E\left[\frac{vv'}{v'v}\right](P'x) = (x'P)H(P'x) = x'PU\Xi U'P'x. \end{aligned}$$

Let $y = U'P'x$. Then $\rho^2_{OLS} = (\eta' y)^2$, and $E[R^2_{OLS}] = y'\Xi y$. Therefore, the maximization of $E[R^2_{OLS}]$ for a given $\rho^2_{OLS} = \rho^2_0$ is the solution to the following problem.

$$\max_y\ y'\Xi y, \qquad \text{s.t. } \eta' y = \rho_0,\ y'y = 1.$$

It is easy to see that $M\mu \neq cPHP'\mu$ for any constant c implies there exist at least two elements i and j from S such that $\xi_i \neq \xi_j$. With Φ_1^ξ and Φ_2^ξ defined similarly to

[26] To facilitate numerical integration, we can use a change of variable of $u = 1/(1+2t\lambda_1)$ and the integral can be evaluated over u from 0 to 1.

Φ_1^λ and Φ_2^λ in the proof of Proposition 3 and following the proof there, we have when $x \neq k1_N$ for any scalar k,

$$\min_{x:\rho^2_{OLS}(x)=\rho_0^2} E[R^2_{OLS}(x)] = \phi^* + \frac{\rho_0^2}{\sum_{i\in S}\eta_i^2(\xi_i-\phi^*)^{-1}}, \tag{A27}$$

$$\max_{x:\rho^2_{OLS}(x)=\rho_0^2} E[R^2_{OLS}(x)] = \phi^{**} + \frac{\rho_0^2}{\sum_{i\in S}\eta_i^2(\xi_i-\phi^{**})^{-1}}, \tag{A28}$$

where $\phi^* = \min \Phi_1^\xi \cup \Phi_2^\xi$ and $\phi^{**} = \max \Phi_1^\xi \cup \Phi_2^\xi$. *Q.E.D.*

Proof of Proposition 6: Since the proof of $E[\bar{R}^2_{GLS}]$ is almost identical to that of $E[R^2_{GLS}]$, we will only prove the case of $E[R^2_{GLS}]$ here. We first define $z = Q'\tilde{R}_t \sim N(Q'\tilde{\mu}, I_{N-1})$ where $Q = V^{\frac{1}{2}}P\Lambda^{-\frac{1}{2}}$, we have $QQ' = \tilde{M}$ and

$$\begin{aligned} E[R^2_{GLS}(x)] = E[R^2_{GLS,t}(x)] &= E\left[\frac{(\tilde{x}'\tilde{M}\tilde{R}_t)^2}{(\tilde{x}'\tilde{M}\tilde{x})(\tilde{R}_t'\tilde{M}\tilde{R}_t)}\right] \\ &= \frac{(\tilde{x}'Q)}{(\tilde{x}'\tilde{M}\tilde{x})^{\frac{1}{2}}} E\left[\frac{zz'}{z'z}\right] \frac{(Q'\tilde{x})}{(\tilde{x}'\tilde{M}\tilde{x})^{\frac{1}{2}}}. \end{aligned} \tag{A29}$$

Define $\nu = Q'\tilde{\mu}$, then from (A24), the (i,j)th element of $E\left[\frac{zz'}{z'z}\right]$ is equal to

$$e^{-\frac{\nu'\nu}{2}} \int_0^\infty \frac{\exp\left(\frac{\nu'\nu}{2(1+2t)}\right)}{(1+2t)^{\frac{N-1}{2}}} \left[\frac{\delta_{ij}}{1+2t} + \frac{\nu_i\nu_j}{(1+2t)^2}\right] \mathrm{d}t. \tag{A30}$$

With a change of variable of $u = 1/(1+2t)$ and define $a = (\nu'\nu)/2 = (\tilde{\mu}'\tilde{M}\tilde{\mu})/2$, we have

$$\begin{aligned} E\left[\frac{zz'}{z'z}\right] &= \left(\frac{e^{-a}}{2}\int_0^1 e^{au}u^{\frac{N-3}{2}}\mathrm{d}u\right) I_{N-1} + \left(\frac{e^{-a}}{2}\int_0^1 e^{au}u^{\frac{N-1}{2}}\mathrm{d}u\right)\nu\nu' \\ &= hI_{N-1} + \left[\frac{1-(N-1)h}{2a}\right]\nu\nu', \end{aligned} \tag{A31}$$

with the last equality obtained from integration by parts. Therefore,

$$\begin{aligned} E[R^2_{GLS}(x)] &= \frac{h(\tilde{x}'\tilde{M}\tilde{x})}{\tilde{x}'\tilde{M}\tilde{x}} + \left[\frac{1-(N-1)h}{\tilde{\mu}'\tilde{M}\tilde{\mu}}\right]\frac{(\tilde{x}'\tilde{M}\tilde{\mu})^2}{\tilde{x}'\tilde{M}\tilde{x}} \\ &= h + [1-(N-1)h]\rho^2_{GLS}(x). \end{aligned} \tag{A32}$$

This completes the proof. *Q.E.D.*

References

Affleck-Graves, J. F., and D. J. Bradfield, 1993, "An Examination of the Power of Univariate Tests of the CAPM: A Simulation Approach," *Journal of Economics and Business*, 45, 17–33.

Banz, R. W., 1981, "The Relationship Between Return and Market Value of Common Stocks," *Journal of Financial Economics*, 9, 3–18.

Basu, S., 1977, "The Investment Performance of Common Stocks in Relation to Their Price-Earnings Ratios: A Test of the Efficient Market Hypothesis," *Journal of Finance*, 33, 663–682.

Bhandari, L. C., 1988, "Debt/Equity Ratio and Expected Common Stock Returns," *Journal of Finance*, 43, 507–528.

Breeden, D. T., 1979, "An Intertemporal Asset Pricing Model with Stochastic Consumption and Investment Opportunities," *Journal of Financial Economics*, 17, 265–296.

Chan, L. K. C., Y. Hamao, and J. Lakonishok, 1991, "Fundamentals and Stock Returns in Japan," *Journal of Finance*, 46, 1739–1764.

Chan, L. K. C., and J. Lakonishok, 1993, "Are the Reports of Beta's Death Premature?" *Journal of Portfolio Management*, 19, 51–62.

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

Fama, E. F., and J. D. MacBeth, 1973, "Risk, Return and Equilibrium: Empirical Tests," *Journal of Political Economy*, 81, 607–636.

Grauer, R. R., 1999, "On the Cross-Sectional Relation Between Expected Returns, Betas and Size," *Journal of Finance*, 54, 773–789.

Hoque, A., 1985, "The Exact Moments of Forecast Error in the General Dynamic Model," *Sankhyā: The Indian Journal of Statistics*, 47, Series B, Part 1, 128–143.

Horn, R. A., and C. R. Johnson, 1990, *Matrix Analysis*, Cambridge, New York.

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

Johnson, N. L., S. Kotz, and N. Balakrishnan, 1995, *Continuous Univariate Distributions, Volume 2* (Wiley, New York).

Kandel, S., and R. F. Stambaugh, 1995, "Portfolio Inefficiency and the Cross-Section of Expected Returns," *Journal of Finance*, 50, 157–184.

Lintner, J., 1965, "The Valuation of Risky Assets and the Selection of Risky Investments in the Portfolios and Capital Budgets," *Review of Economics and Statistics*, 47, 13–37.

McCloskey, D. N., and S. T. Ziliak, 1996, "The Standard Error of Regressions," *Journal of Economic Literature*, 34, 97–114.

Merton, R. C., 1973, "An Intertemporal Capital Asset Pricing Model," *Econometrica*, 41, 867–887.

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

Ross, S. A., 1976, "The Arbitrage Theory of Capital Asset Pricing," *Journal of Economic Theory*, 13, 341–360.

Sawa, T., 1978, "The Exact Moments of the Least Squares Estimator for the Autoregressive Model," *Journal of Econometrics*, 8, 159–172.

Sharpe, W. F., 1964, "Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk," *Journal of Finance*, 19, 425–442.

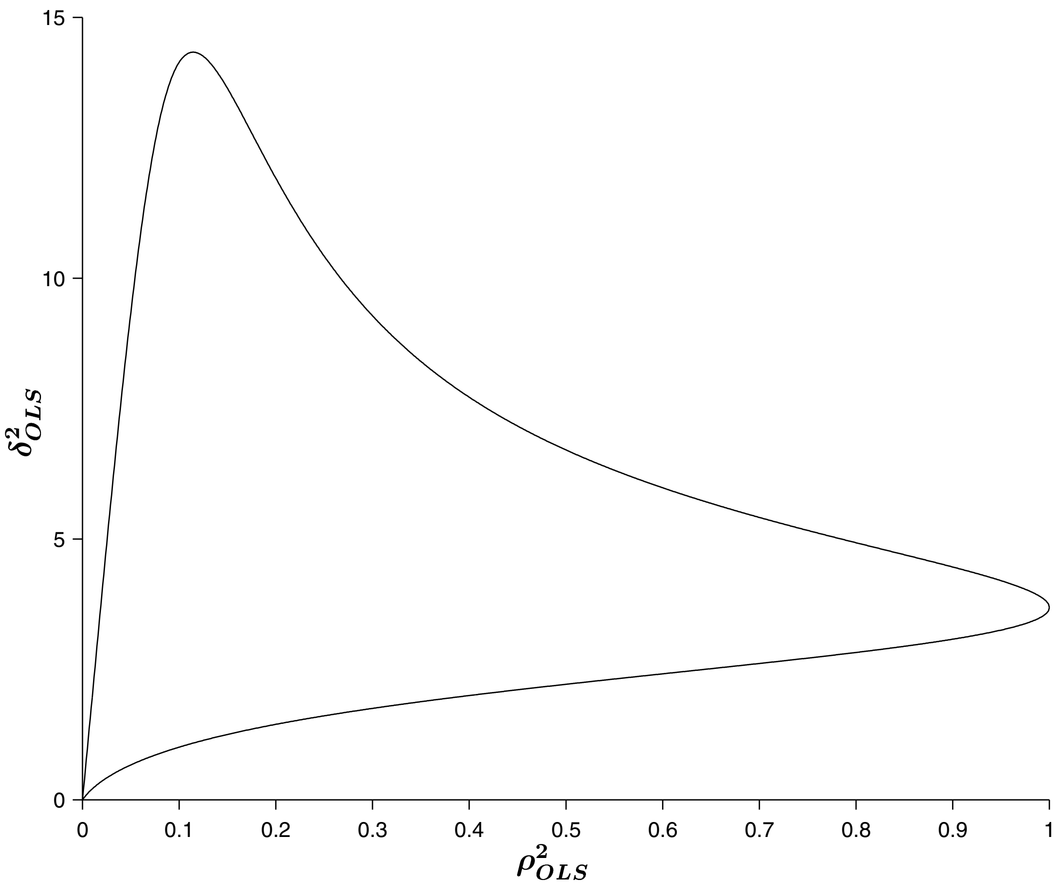

Figure 1
Lower Bound and Upper Bound of δ^2_{OLS} as a Function of ρ^2_{OLS} for 10 Size Ranked Portfolios when $T = 330$

The figure presents the lower bound and upper bound of the square of noncentrality parameter ($\delta^2_{OLS}(x)$) of the t-ratio of the slope coefficient obtained in OLS cross-sectional regression of regressing returns on 10 test assets on firm-specific variables with different explanatory power on the expected returns using 330 observations. The explanatory power of a firm-specific variable x is measured as $\rho^2_{OLS}(x) = \frac{(x'M\mu)^2}{(x'Mx)(\mu'M\mu)}$. The parameters of the 10 test assets are given in Table 1. The maximum $\delta^2_{OLS}(x)$ is reached when $\rho^2_{OLS}(x) = 0.114$.

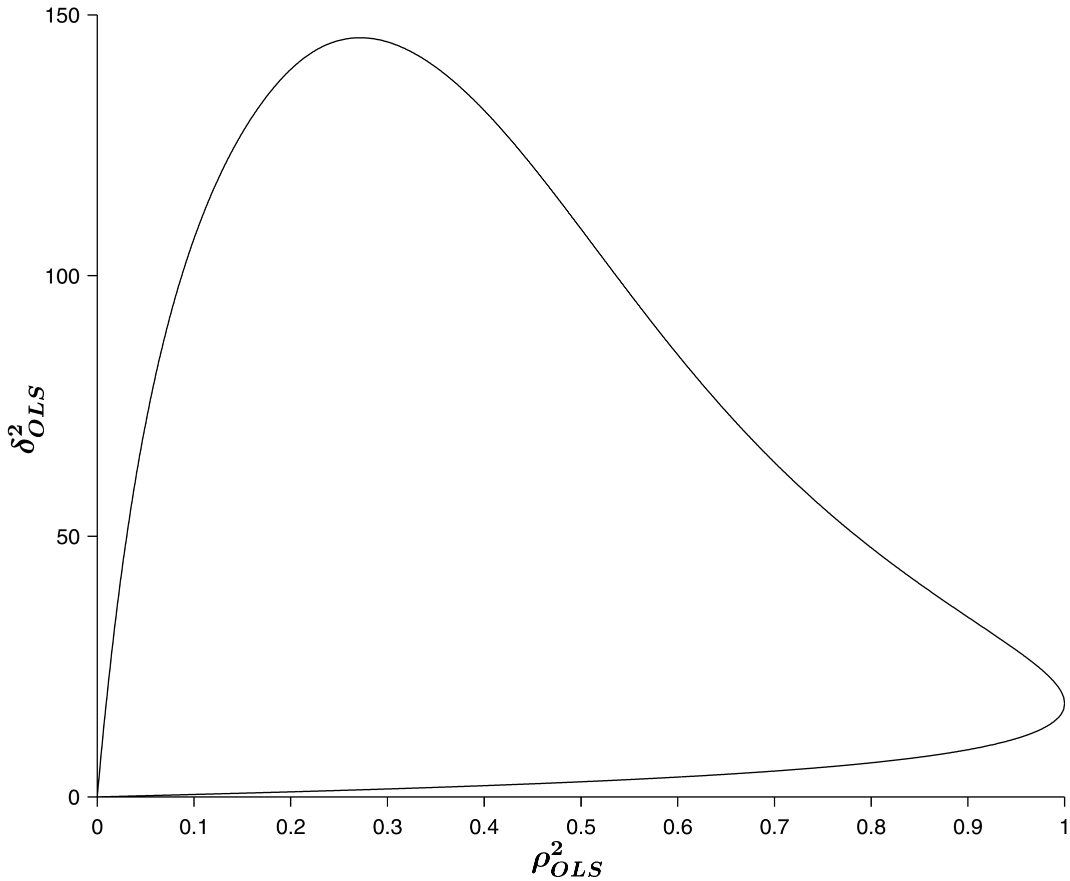

Figure 2
Lower Bound and Upper Bound of δ^2_{OLS} as a Function of ρ^2_{OLS} for 100 Size-Beta Ranked Portfolios when $T = 330$

The figure presents the lower bound and upper bound of the square of noncentrality parameter ($\delta^2_{OLS}(x)$) of the t-ratio of the slope coefficient obtained in OLS cross-sectional regression of regressing returns on 100 size-beta ranked portfolios on firm-specific variables with different explanatory power on the expected returns using 330 observations. The explanatory power of a firm-specific variable x is measured as $\rho^2_{OLS}(x) = \frac{(x'M\mu)^2}{(x'Mx)(\mu'M\mu)}$. The parameters of the 100 size-beta ranked portfolios are determined based on the sample estimates of the data used by Jagannathan and Wang (1996). The maximum $\delta^2_{OLS}(x)$ is reached when $\rho^2_{OLS}(x) = 0.272$.

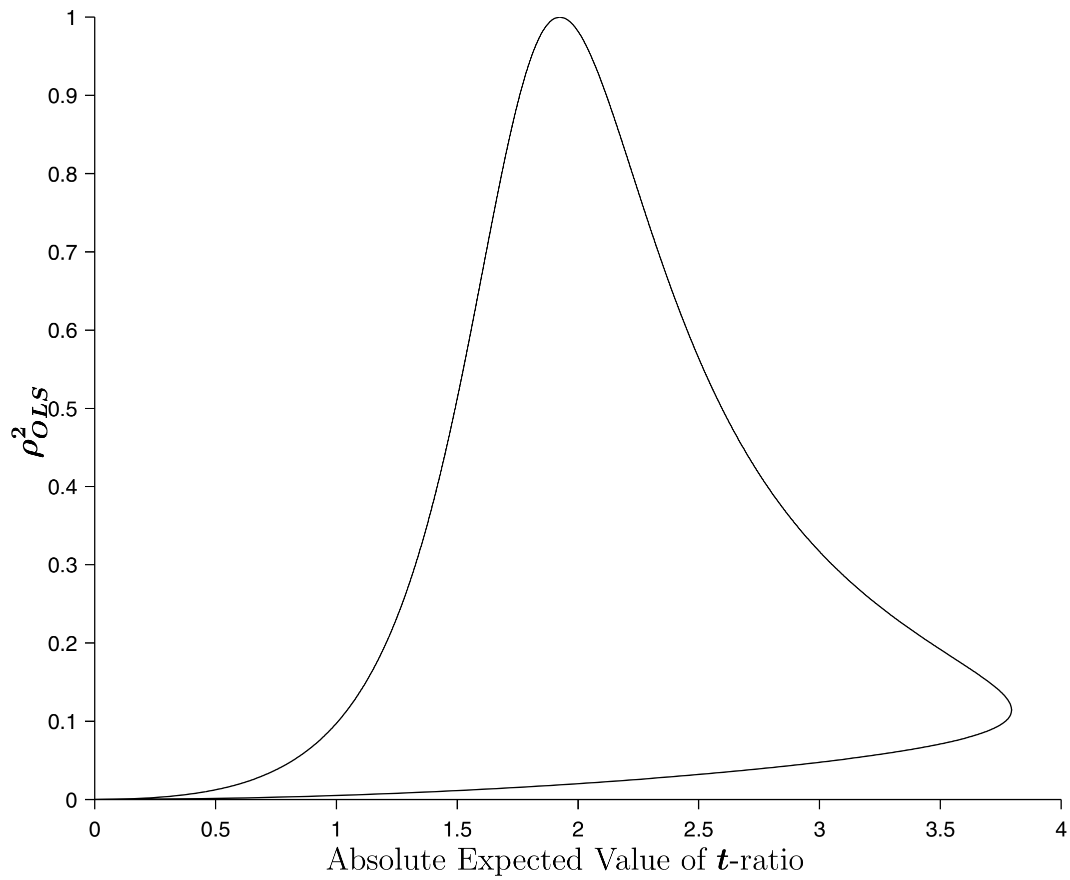

Figure 3
Lower Bound and Upper Bound of ρ^2_{OLS} for Firm-specific Variables with Different Absolute Values of Expected t-ratio for 10 Size Ranked Portfolios when $T = 330$
The figure presents the lower bound and upper bound of the explanatory power of a firm-specific variable $\boldsymbol{x}$ on the expected return when the variable has a given absolute value of expected $\boldsymbol{t}$-ratio of the slope coefficient in the OLS cross-sectional regression of regressing returns on 10 test assets on the firm-specific variable using 330 observations. The explanatory power of a firm-specific variable $\boldsymbol{x}$ is measured as $\rho^2_{OLS}(x) = \frac{(x'M\mu)^2}{(x'Mx)(\mu'M\mu)}$. The parameters of the 10 test assets are given in Table 1.

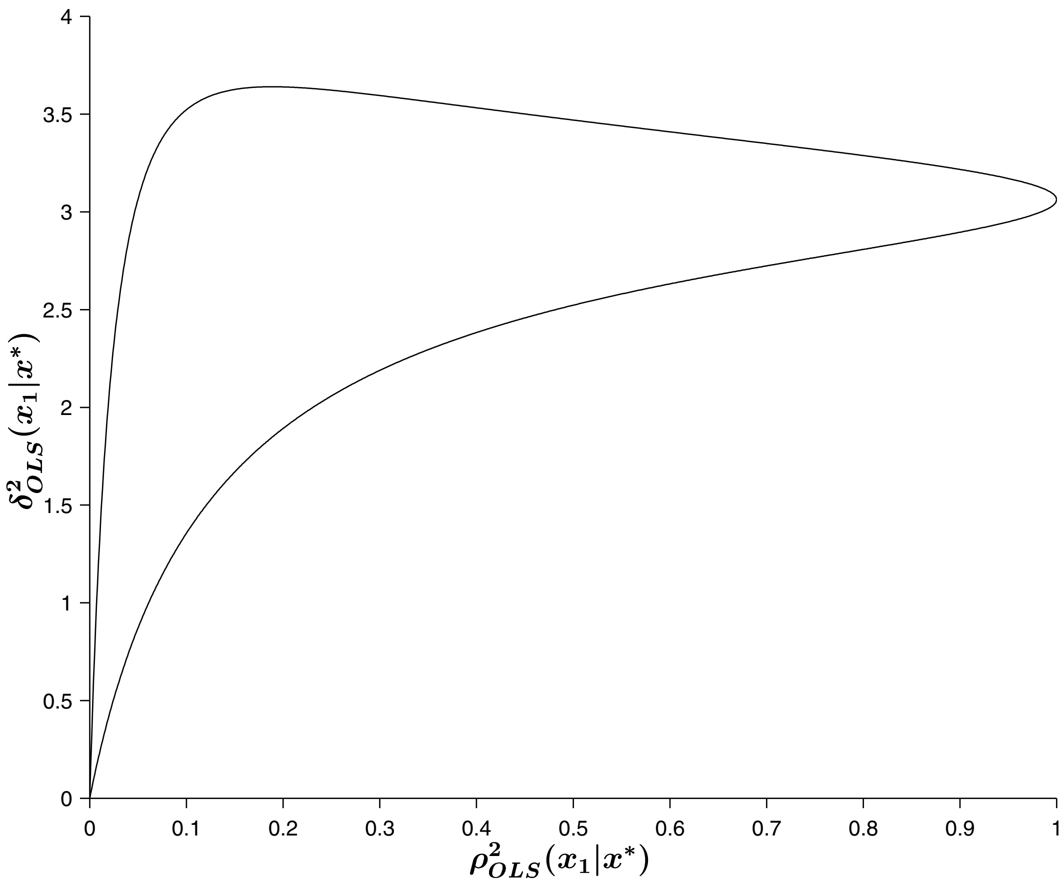

Figure 4
Lower Bound and Upper Bound of $\delta^2_{OLS}(x_1|x^*)$ as a Function of $\rho^2_{OLS}(x_1|x^*)$ for 10 Size Ranked Portfolios when $T = 330$
The figure presents the lower bound and upper bound of the square of noncentrality parameter ($\boldsymbol{\delta^2_{OLS}(x_1|x^*)}$) of the $\boldsymbol{t}$-ratio of the slope coefficient of a firm-specific variable with different marginal explanatory power on the expected returns using 330 observations. The $\boldsymbol{t}$-ratio of the firm-specific variable $\boldsymbol{x_1}$ is obtained in a multiple OLS cross-sectional regression of regressing returns on 10 test assets on this firm-specific variable together with another firm-specific variable $\boldsymbol{x^*}$ as described in Table 1. The marginal explanatory power of a firm-specific variable $\boldsymbol{x_1}$ is measured as $\boldsymbol{\rho^2_{OLS}(x_1|x^*) = \frac{(x_1' M_2 \mu)^2}{(x_1' M_2 x_1)(\mu' M_2 \mu)}}$ where $\boldsymbol{M_2}$ is the projection matrix onto the space orthogonal to $[\boldsymbol{1_N},\ \boldsymbol{x^*}]$. The parameters of the 10 test assets are given in Table 1. The maximum $\boldsymbol{\delta^2_{OLS}(x_1|x^*) = 3.640}$ is reached when $\boldsymbol{\rho^2_{OLS}(x_1|x^*) = 0.189}$.

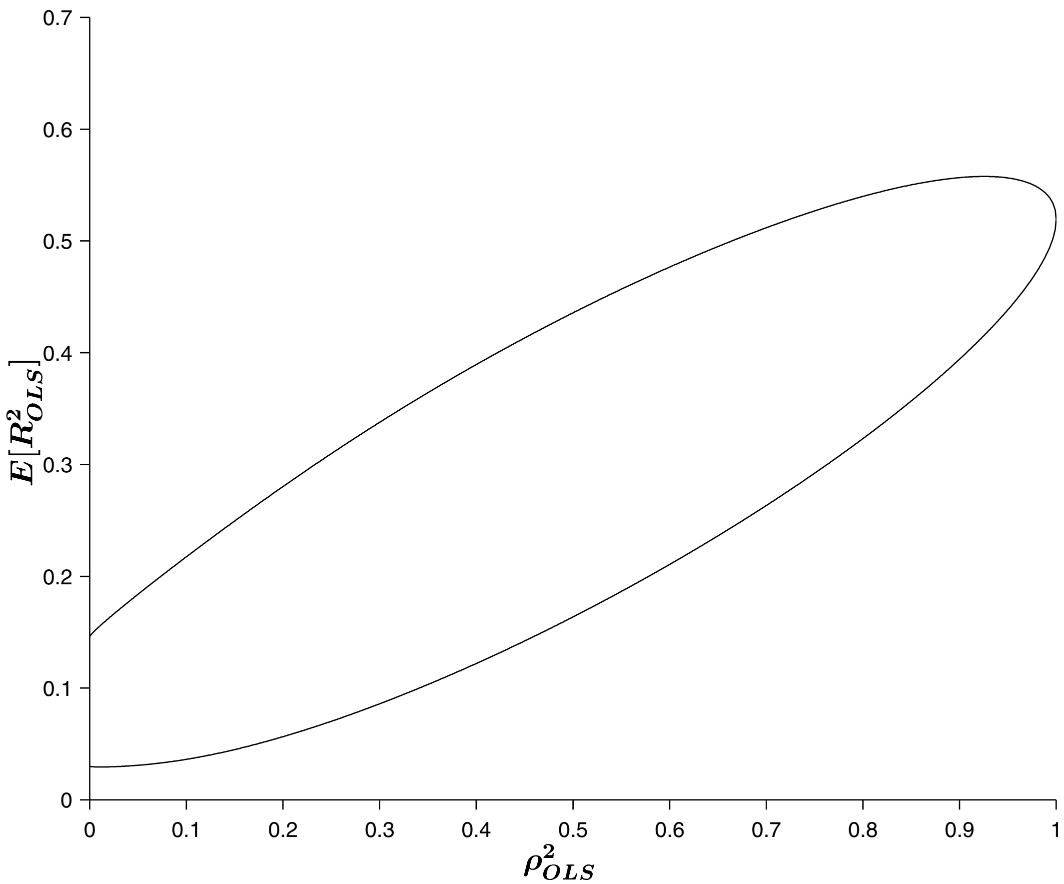

Figure 5
Lower Bound and Upper Bound of $E[R^2_{OLS}]$ as a Function of ρ^2_{OLS} for 10 Size Ranked Portfolios
The figure presents the lower bound and upper bound of the expected value of the average sample coefficient of determination in an OLS cross-sectional regression of regressing returns on 10 assets on firm-specific variables with different explanatory power on the expected returns. The explanatory power of a firm-specific variable $\boldsymbol{x}$ is measured as $\rho^2_{OLS}(x) = \frac{(x'M\mu)^2}{(x'Mx)(\mu'M\mu)}$. The parameters of the 10 test assets are given in Table 1. The maximum $E[R^2_{OLS}(x)]$ is reached when $\rho^2_{OLS}(x) = 0.925$.

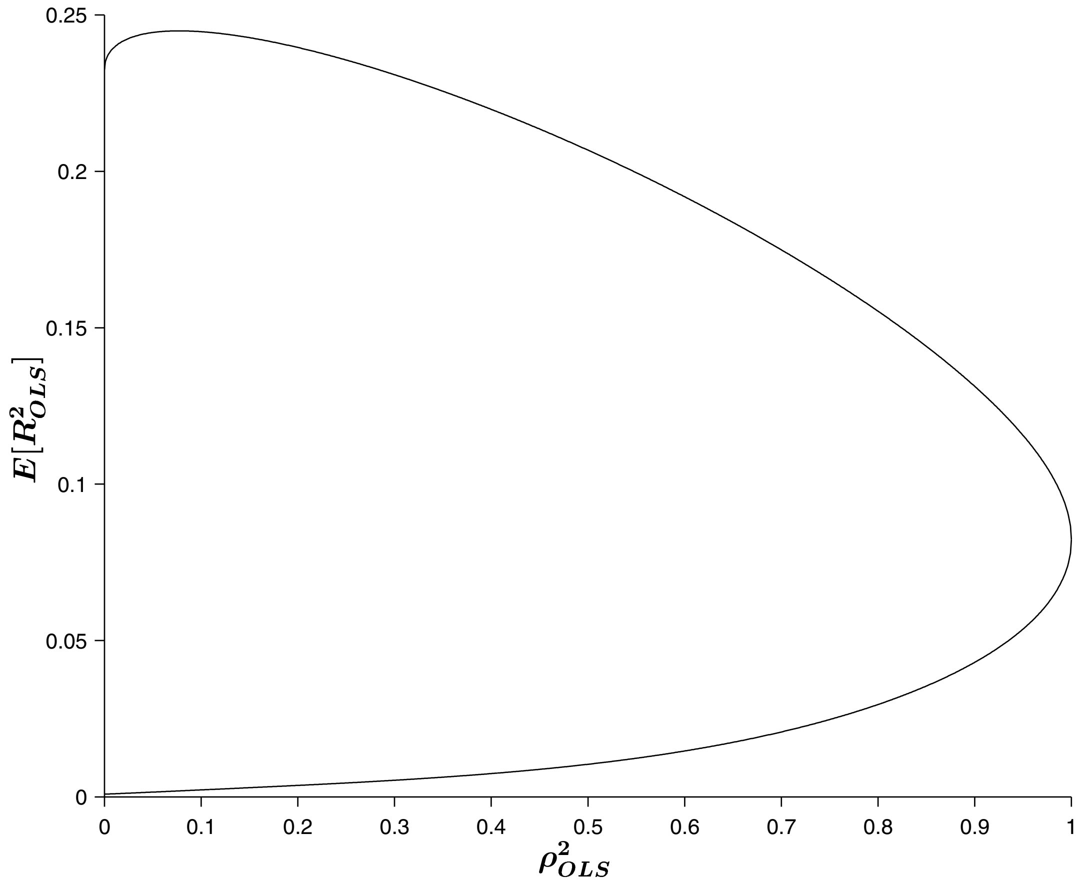

Figure 6

Lower Bound and Upper Bound of $E[R^2_{OLS}]$ as a Function of ρ^2_{OLS} for 100 Size-Beta Ranked Portfolios

The figure presents the lower bound and upper bound of the expected value of the average sample coefficient of determination in an OLS cross-sectional regression of regressing returns on 100 size-beta ranked portfolios on firm-specific variables with different explanatory power on the expected returns. The explanatory power of a firm-specific variable x is measured as $\rho^2_{OLS}(x) = \frac{(x'M\mu)^2}{(x'Mx)(\mu'M\mu)}$. The parameters of the 100 size-beta ranked portfolios are determined based on the sample estimates of the data used by Jagannathan and Wang (1996). The maximum $E[R^2_{OLS}(x)]$ is reached when $\rho^2_{OLS}(x) = 0.078$.

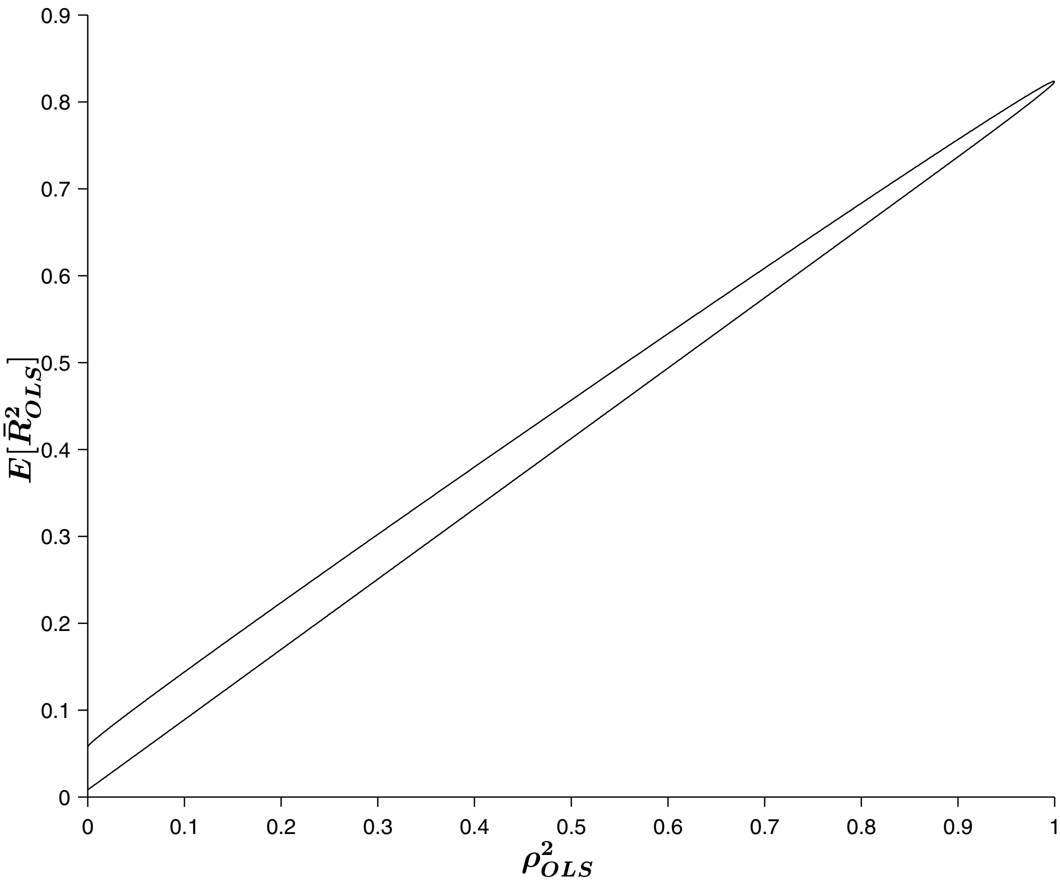

Figure 7
Lower Bound and Upper Bound of $E[\bar{R}^2_{OLS}]$ as a Function of ρ^2_{OLS} for 10 Size Ranked Portfolios when $T = 330$
The figure presents the lower bound and upper bound of the expected value of the sample coefficient of determination in an OLS cross-sectional regression of regressing average returns on 10 assets on firm-specific variables with different explanatory power on the expected returns. The average returns are computed using 330 observations and the explanatory power of a firm-specific variable $\boldsymbol{x}$ is measured as $\boldsymbol{\rho^2_{OLS}(x) = \frac{(x'M\mu)^2}{(x'Mx)(\mu'M\mu)}}$. The parameters of the 10 test assets are given in Table 1.

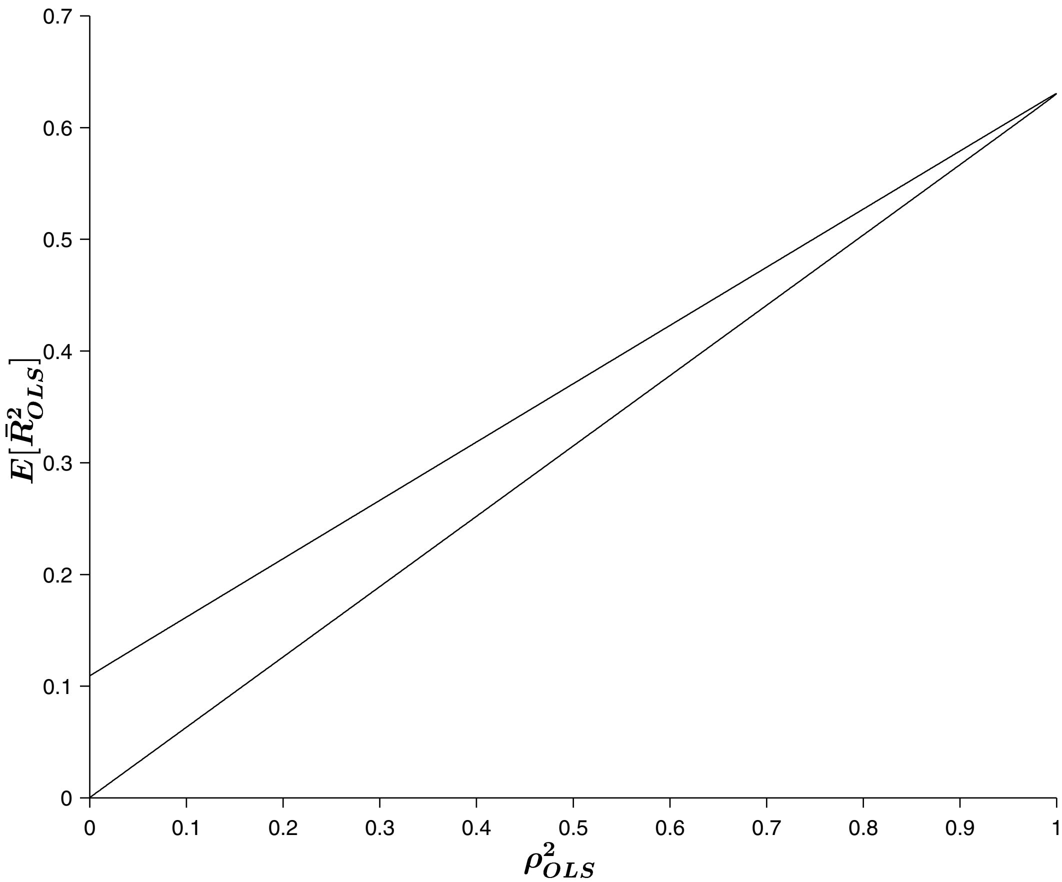

Figure 8
Lower Bound and Upper Bound of $E[\bar{R}^2_{OLS}]$ as a Function of ρ^2_{OLS} for 100 Size-Beta Ranked Portfolios when $T = 330$
The figure presents the lower bound and upper bound of the expected value of the sample coefficient of determination in an OLS cross-sectional regression of regressing average returns on 100 size-beta ranked portfolios on firm-specific variables with different explanatory power on the expected returns. The average returns are computed using 330 observations and the explanatory power of a firm-specific variable $\boldsymbol{x}$ is measured as $\rho^2_{OLS}(x) = \frac{(x'M\mu)^2}{(x'Mx)(\mu'M\mu)}$. The parameters of the 100 size-beta ranked portfolios are determined based on the sample estimates of the data used by Jagannathan and Wang (1996).

Table 1
Distribution of t-ratios and Rejection Rates of $H_0 : \gamma_1 = 0$ for Two Choices of Firm-specific Variables in a Simple Cross-sectional Regression

Panel A: Firm-specific variables for 10 size portfolios

	1	2	3	4	5	6	7	8	9	10
μ	1.454	1.215	1.200	1.239	1.158	1.088	1.017	1.090	0.944	0.872
x^*	1.241	0.939	1.058	1.282	1.206	1.078	0.957	1.459	0.989	1.069

$$\rho^2_{OLS}(x^*) = 0.114$$

Panel B: Nonzero eigenvalues (λ) of MVM and correlation coefficients of corresponding eigenvectors with expected returns (η)

	1	2	3	4	5	6	7	8	9
λ	23.862	2.819	0.913	0.816	0.698	0.590	0.519	0.456	0.415
η	0.961	0.022	−0.105	0.122	−0.049	−0.014	0.115	0.132	−0.128

Panel C: Characteristics of t-ratios for $T = 330$

Distribution of t-ratios

	OLS		GLS	
	μ	x^*	μ	x^*
δ^2	3.689	14.334	14.334	10.694
|Mean|	1.925	3.795	3.795	3.278
Variance	1.012	1.028	1.028	1.023

Probability of rejecting $H_0 : \gamma_1 = 0$

Significance	OLS		GLS	
Level	μ	x^*	μ	x^*
0.01	0.253	0.883	0.883	0.751
0.05	0.482	0.965	0.965	0.903
0.10	0.607	0.984	0.984	0.947

Panel A of the table presents two different choices of firm-specific variables used in explaining the cross-sectional differences of expected returns of 10 portfolios. The 10 portfolio returns are assumed to be independently distributed as $N(\mu, V)$, where μ and V (not reported) are set equal to the average and the estimated variance-covariance matrix of the equally weighted monthly returns on 10 size portfolios of combined NYSE-AMEX stocks over the period July 1963 to December 1990. The first choice of firm-specific variable is μ and the second choice is x^* which maximizes $\delta^2_{OLS}(x)$. Panel B presents the nonzero eigenvalues (λ) of MVM and the correlation coefficients (η) of the corresponding eigenvectors with μ. Panel C reports the characteristics of the t-ratios in OLS CSR and GLS CSR using μ or x^* as the explanatory variable on 330 monthly returns that are generated independently from $N(\mu, V)$. The probabilities of rejecting $H_0 : \gamma_1 = 0$ using the two-tailed t-test at various significance levels are also reported.

Table 2
Distribution of t-ratios and Rejection Rates of $H_0 : \gamma_1 = 0$ and $H_0 : \gamma_2 = 0$ for Two Firm-specific Variables in a Multiple Cross-sectional Regression

Panel A: Firm-specific variables for 10 size portfolios

	1	2	3	4	5	6	7	8	9	10
μ	1.454	1.215	1.200	1.239	1.158	1.088	1.017	1.090	0.944	0.872
x_1	1.434	1.288	1.230	1.191	1.131	1.103	1.072	0.969	0.982	0.877
x^*	1.241	0.939	1.058	1.282	1.206	1.078	0.957	1.459	0.989	1.069

$\rho^2_{OLS}(x_1) = 0.886 \quad \rho^2_{OLS}(x^*) = 0.114$

Panel B: Characteristics of t-ratios for $T = 330$

Distribution of t-ratios

	OLS		GLS	
	x_1	x^*	x_1	x^*
δ^2	3.063	14.334	3.640	14.334
|Mean|	1.754	3.795	1.912	3.795
Variance	1.011	1.028	1.012	1.028

Probability of rejecting $H_0 : \gamma_1 = 0$ and $H_0 : \gamma_2 = 0$

Significance	OLS		GLS	
Level	x_1	x^*	x_1	x^*
0.01	0.202	0.883	0.249	0.883
0.05	0.415	0.965	0.477	0.965
0.10	0.541	0.984	0.602	0.984

Panel A of the table presents the expected returns of 10 portfolios and two different firm-specific variables used in explaining the cross-sectional differences of expected returns of the 10 portfolios. The 10 portfolio returns are assumed to be independently distributed as $N(\mu, V)$, where μ and V (not reported) are set equal to the average and the estimated variance-covariance matrix of the equally weighted monthly returns on 10 size portfolios of combined NYSE-AMEX stocks over the period July 1963 to December 1990. The first firm-specific variable is x_1 and the second one is x^*. The two firm-specific variables are uncorrelated with each other and they together fully explain the cross-sectional differences of μ. Panel B reports the characteristics of the t-ratios in multiple OLS CSR and GLS CSR using both x_1 and x^* as the explanatory variables on 330 monthly returns that are generated independently from $N(\mu, V)$. The probabilities of rejecting $H_0 : \gamma_1 = 0$ and $H_0 : \gamma_2 = 0$ using the two-tailed t-test at various significance levels are also reported.