THE JOURNAL OF FINANCE • VOL. LIV, NO. 1 • FEBRUARY 1999

Two-Pass Tests of Asset Pricing Models with Useless Factors

RAYMOND KAN and CHU ZHANG*

ABSTRACT

In this paper we investigate the properties of the standard two-pass methodology of testing beta pricing models with misspecified factors. In a setting where a factor is useless, defined as being independent of all the asset returns, we provide theoretical results and simulation evidence that the second-pass cross-sectional regression tends to find the beta risk of the useless factor priced more often than it should. More surprisingly, this misspecification bias exacerbates when the number of time series observations increases. Possible ways of detecting useless factors are also examined.

WHEN TESTING ASSET PRICING MODELS relating risk premiums on assets to their betas, the primary question of interest is whether the beta risk of a particular factor is priced (i.e., whether the estimated risk premium associated with a given factor is significantly different from zero). Black, Jensen, and Scholes (1972) and Fama and MacBeth (1973) develop a two-pass methodology in which the beta of each asset with respect to a factor is estimated in a first-pass time series regression, and estimated betas are then used in second-pass cross-sectional regressions (CSRs) to estimate the risk premium of the factor. This two-pass methodology is very intuitive and has been widely used in the literature. The properties of the test statistics and goodness-of-fit measures under the two-pass methodology are usually developed under the assumptions that the asset pricing model is correctly specified and that the factors are correctly identified. Shanken (1992) provides an excellent discussion of this two-pass methodology, especially the large sample proper-

* Kan is at the University of Toronto, Zhang is at the Hong Kong University of Science and Technology and the University of Alberta. We thank K. C. Chan, Sean Cleary, Jin-Chuan Duan, Wayne Ferson, René Garcia, Mark Huson, Vijay Jog, Youngsoo Kim, George Kirikos, Peter Klein, Bob Korkie, Anthony Lynch, Vikas Mehrotra, Angelo Melino, Randall Morck, Sergei Sarkissian, Ken Shah, Tim Simin, Robert Stambaugh, René Stulz (the editor), Mike Vetsuypens, Zhenyu Wang, John Wei, Min-Teh Yu, Guofu Zhou, Xiaodong Zhu, seminar participants at the Hong Kong University of Science and Technology, National Central University (Taiwan), Queen's University, Southern Methodist University, University of Alberta, University of California at Irvine, University of Toronto, participants at the 1996 Northern Finance Meetings in Quebec City and the 1997 Western Finance Meetings in San Diego, and particularly Naifu Chen and Ravi Jagannathan for their helpful comments and discussions. We would also like to thank Ravi Jagannathan and Zhenyu Wang for sharing their data set with us, and Teresa Chan and Martin Forest for their research assistance. All remaining errors are ours.

ties of the two-pass CSR for the correctly specified model under the assumption that returns are conditionally homoskedastic. Jagannathan and Wang (1998) generalize Shanken's large sample results to the case of conditionally heteroskedastic returns. However, little is known about how these test statistics and goodness-of-fit measures behave if the model is misspecified.

In this paper, we study the properties of the two-pass CSR when the asset pricing model is misspecified. Misspecification of an asset pricing model can take various forms; here we focus on the extreme case in which the proposed factor is independent of all the asset returns used in the test. We call such a factor a *useless factor*. One might expect that when a useless factor is used in testing an asset pricing model, the hypothesis that its risk premium is zero would only be rejected with a low probability as indicated by the size of the test. We show that this view cannot be justified. Analytical and simulation results indicate that in a finite sample, the beta risk associated with a useless factor is found to be priced more often than the size of the test. A more surprising result is its large sample property. Since the problem arises because the betas are unobservable and estimated with errors, one might expect that as the number of the time series observations increases, the estimates of the betas will become more accurate and the above-mentioned problem will diminish. We show that this is not the case. In fact, as the number of time series observations goes to infinity, the probability of rejecting the null hypothesis that the risk premium of a useless factor equals zero goes to one.

The reason this problem arises is that the true betas of the assets with respect to the useless factor are all zeros, so the "true" risk premium with respect to the useless factor is in fact undefined. Therefore, as the estimated betas approach zero, the absolute value of the estimated risk premium needs to go to infinity, instead of zero, in order to "explain" the cross-sectional difference in the expected returns.

Although the misspecification bias in the case of a useless factor is due to the estimation errors of betas in the first-pass time series regression, traditional errors-in-variables (EIV) adjustments as suggested by Shanken (1992) and Kim (1995) cannot be used to correct for this bias. This is because such adjustments are derived under the assumption that the model tested is the correct one, and, therefore, they are not applicable to the case of misspecified models. We show that even with the EIV adjustments of Shanken (1992), the asymptotic probability of rejecting the null hypothesis that the risk premium of a useless factor equals zero is still much greater than the size of the test. Since the "true" risk premium of a useless factor is undefined, it is not possible to come up with an EIV adjustment that is appropriate for a useless factor, as well as for the true factor.

Our results present a significant complication regarding the interpretation of many empirical tests of asset pricing models. This investigation has particular relevance to the models of the Arbitrage Pricing Theory (APT) of Ross (1976) and the Intertemporal Capital Asset Pricing Model of Merton

(1973), in which factors and state variables are unidentified. These factors or state variables are chosen for empirical analysis based on economic intuition. We have no way of determining ex ante whether an asset pricing model to be tested is correct and whether the factors used are the correct ones. There is always a possibility that some proposed factors are in fact useless. Even for the well-known Capital Asset Pricing Model (CAPM) of Sharpe (1964), Lintner (1965), and Black (1972), in which the market portfolio is the sole factor, the problem still exists because the true market portfolio is unobservable (Roll (1977)).

Given the importance of the two-pass methodology in testing asset pricing models and the potential problem of misspecifying factors, a relevant question is how we can detect useless factors in the two-pass methodology. We suggest several tests that can serve as diagnostic tools and we also provide simulation results about their effectiveness.

The rest of the paper is organized as follows. Section I discusses the properties of the test statistics in the two-pass methodology under both correct specification and incorrect specifications. Section II provides simulation evidence that illustrates the magnitude of the bias caused by misspecification. The final section provides our conclusions and the Appendix contains proofs of all propositions.

I. Analytic Results on the Misspecification Bias

A. The Two-pass Regression under Correct Specification

For illustrative purposes the raw returns on N assets at time t, R_t, are assumed to be independently drawn across t from $N(\mu,V)$ where μ is its unconditional mean and V is its unconditional variance-covariance matrix. We also assume R_t are generated from the following one-factor model:[1]

$$R_t = \mu + \beta f_t + \varepsilon_t, \qquad t = 1,\dots,T, \tag{1}$$

where f_t is the realization of the common factor at time t, and $\beta = \mathrm{Cov}[R_t, f_t]/\mathrm{Var}[f_t]$ is a vector of the betas of the N assets with respect to the common factor f_t which is assumed to be constant over time. The term factor is used in a weak sense, so conditioned on $f = [f_1, f_2, \dots, f_T]$ we assume the error term ε_t to have mean zero but it can be a cross-sectionally correlated random vector (see Chamberlain and Rothschild (1983)).

Under an exact static one-factor beta pricing model, for some constants γ_0 and γ_1,

$$\mu = \gamma_0 1_N + \gamma_1 \beta, \tag{2}$$

[1] We thank Naifu Chen for suggesting this simple example to illustrate the problem. However, the subsequent results do not depend on the structure of one-factor models.

where 1_N is the N-vector of ones. When testing equation (2), the main interest is focused on the hypothesis $H_0: \gamma_1 = 0$. If a researcher can observe β, then an Ordinary Least Squares (OLS) CSR of R_t on β can be run for each period. By letting $X = [1_N, \beta]$ and $\gamma = [\gamma_0, \gamma_1]'$, the OLS estimate of γ at time t is

$$\hat{\gamma}_t^{OLS} = \begin{bmatrix} \hat{\gamma}_{0t}^{OLS} \\ \hat{\gamma}_{1t}^{OLS} \end{bmatrix} = (X'X)^{-1}(X'R_t). \tag{3}$$

Since $\mu = X\gamma$, under the assumption that returns are independently and identically distributed (i.i.d.) $N(\mu, V)$, it is easy to verify that

$$\hat{\gamma}_t^{OLS} \sim N(\gamma, (X'X)^{-1}(X'VX)(X'X)^{-1}). \tag{4}$$

In particular,

$$\hat{\gamma}_{1t}^{OLS} \sim N\left(\gamma_1, \frac{\beta' MVM\beta}{(\beta' M\beta)^2}\right), \tag{5}$$

where $M = I_N - (1_N 1_N')/(1_N' 1_N)$ and $\{\hat{\gamma}_{1t}^{OLS}\}$ is a sequence of i.i.d. unbiased estimators of γ_1. We can test $H_0: \gamma_1 = 0$ using a t-test on the time series of $\hat{\gamma}_{1t}^{OLS}$. The test statistic is given by

$$t_{OLS} = \frac{\bar{\hat{\gamma}}_1^{OLS}}{s(\hat{\gamma}_1^{OLS})/\sqrt{T}}, \tag{6}$$

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

Under the null hypothesis, t_{OLS} has a central t-distribution with $T - 1$ degrees of freedom. But when $\gamma_1 \neq 0$, t_{OLS} has a noncentral t-distribution with the square of its noncentrality parameter given by[2]

$$\delta_{OLS}^2(\beta) = \frac{T(\beta' M\mu)^2}{(\beta' MVM\, \beta)} = \frac{T(\mu' M\mu)^2}{(\mu' MVM\mu)} = \delta_{OLS}^2(\mu). \tag{7}$$

It is well known that if a random variable t_δ has a noncentral t-distribution with noncentrality parameter δ, then $P[|t_\delta| > d]$ for $d > 0$ is an increasing function of δ^2. Therefore, if μ is not constant across assets, then $\delta_{OLS}(\beta) \neq 0$ and the probability of rejecting the null hypothesis using a two-tailed t-test will be

[2] See, for example, Johnson, Kotz, and Balakrishnan (1995, Chap. 31). The square of the noncentrality parameter is invariant to rescaling of the factor, or rescaling of β, hence, it only depends on μ but not on γ_1.

higher than the size of the test obtained from a central t-distribution. In the parlance of asset pricing theory, one is likely to find β priced when using the t-test if $\gamma_1 \neq 0$.

The CSR can also be run by generalized least squares (GLS). The GLS estimate of γ at time t is

$$\hat{\gamma}_t^{GLS} = \begin{bmatrix} \hat{\gamma}_{0t}^{GLS} \\ \hat{\gamma}_{1t}^{GLS} \end{bmatrix} = (X'V^{-1}X)^{-1}(X'V^{-1}R_t). \tag{8}$$

It is easy to verify that

$$\hat{\gamma}_t^{GLS} \sim N(\gamma, (X'V^{-1}X)^{-1}), \tag{9}$$

and in particular, we have

$$\hat{\gamma}_{1t}^{GLS} \sim N\left(\gamma_1, \frac{1}{\tilde{\beta}'\widetilde{M}\tilde{\beta}}\right), \tag{10}$$

where $\widetilde{M} = I_N - (\tilde{1}_N\tilde{1}_N')/(\tilde{1}_N'\tilde{1}_N)$, $\tilde{1}_N = V^{-1/2}1_N$ and $\tilde{\beta} = V^{-1/2}\beta$. Therefore, $\{\hat{\gamma}_{1t}^{GLS}\}$ is also a sequence of i.i.d. unbiased estimators of γ_1 and similarly we can test $H_0\colon \gamma_1 = 0$ using the t-test given by

$$t_{GLS} = \frac{\bar{\hat{\gamma}}_1^{GLS}}{s(\hat{\gamma}_1^{GLS})/\sqrt{T}}, \tag{11}$$

where $\bar{\hat{\gamma}}_1^{GLS}$ and $s(\hat{\gamma}_1^{GLS})$ are the sample average and standard deviation of $\hat{\gamma}_{1t}^{GLS}$, respectively. Under the null hypothesis, t_{GLS} has a central t-distribution with $T - 1$ degrees of freedom. But when $\gamma_1 \neq 0$, t_{GLS} has a noncentral t-distribution with the square of its noncentrality parameter given by

$$\delta_{GLS}^2(\beta) = \frac{T(\tilde{\beta}'\widetilde{M}\tilde{\mu})^2}{(\tilde{\beta}'\widetilde{M}\tilde{\beta})} = T(\tilde{\mu}'\widetilde{M}\tilde{\mu}) = \delta_{GLS}^2(\mu), \tag{12}$$

where $\tilde{\mu} = V^{-1/2}\mu$. It is well known that $\delta_{GLS}^2(\beta) \geq \delta_{OLS}^2(\beta)$. Therefore, if β and V are observable, GLS CSR is more powerful than OLS CSR under the correct model.[3]

[3] Due to this inequality, Amihud, Christensen, and Mendelson (1992) suggest the CSR should be run using GLS instead of OLS.

In practice, betas are not observable and have to be estimated. The popular two-pass CSR methodology involves first estimating betas in a first-pass time series OLS regression of R_t on f_t. The estimated betas, $\hat{\beta}$, are then used to run the CSR of equations (3) or (8) by replacing X with $\hat{X} = [1_N, \hat{\beta}]$. When estimated betas instead of true betas are used in the second-pass CSR, the estimators $\bar{\hat{\gamma}}^{OLS}$ and $\bar{\hat{\gamma}}^{GLS}$ are biased but as the estimation period is lengthened, the estimation errors of $\hat{\beta}$ diminish and $\bar{\hat{\gamma}}^{OLS}$ and $\bar{\hat{\gamma}}^{GLS}$ are still consistent. Nevertheless, as discussed in Shanken (1992) and Jagannathan and Wang (1998), EIV adjustments are still required to obtain asymptotically correct standard errors of $\bar{\hat{\gamma}}^{OLS}$ and $\bar{\hat{\gamma}}^{GLS}$, where $\gamma_1 \neq 0$. For the case $\gamma_1 = 0$, such an EIV adjustment is not required and the t-tests of equations (6) and (11) are valid asymptotically.

When the same estimated betas, $\hat{\beta}$, are used to run CSR every period, the estimated risk premium ($\bar{\hat{\gamma}}_1^{OLS}$ or $\bar{\hat{\gamma}}_1^{GLS}$) described above is numerically equivalent to that in the single CSR of $\bar{R}$ on $\hat{\beta}$, where $\bar{R}$ is the time series average of R_t. However, it is important to distinguish the conventional t-ratio of the slope coefficient in this single CSR from the t-ratio in the two-pass CSR. As recognized by Black et al. (1972) and Miller and Scholes (1972), error terms in this single CSR are heteroskedastic and cross-sectionally (positively) correlated and hence the conventional t-ratio in this single OLS CSR tends to overstate the actual significance of the estimated risk premium. Therefore, researchers do not use this conventional t-ratio in the single OLS CSR to test $H_0 : \gamma_1 = 0$.

From the single CSR of $\bar{R}$ on estimated β, R^2s are often reported as a measure of the goodness-of-fit of the model. For OLS CSR, the sample R^2_{OLS} is given by

$$R^2_{OLS}(\hat{\beta}) = \frac{(\bar{R}'M\hat{\beta})^2}{(\bar{R}'M\bar{R})(\hat{\beta}'M\hat{\beta})}. \tag{13}$$

For GLS CSR, the sample R^2_{GLS} is given by

$$R^2_{GLS}(\hat{\beta}) = \frac{(\tilde{\bar{R}}'\widetilde{M}\tilde{\hat{\beta}})^2}{(\tilde{\bar{R}}'\widetilde{M}\tilde{\bar{R}})(\tilde{\hat{\beta}}'\widetilde{M}\tilde{\hat{\beta}})}, \tag{14}$$

where $\tilde{\hat{\beta}} = V^{-1/2}\hat{\beta}$, and $\tilde{\bar{R}} = V^{-1/2}\bar{R}$. Under the correct specification, as $T \to \infty$, $\hat{\beta} \to \beta$, and $\bar{R} \to \mu$. Therefore, both the sample R^2_{OLS} and R^2_{GLS} tend to one as T tends to infinity. However, for us to use R^2_{OLS} and R^2_{GLS} as measures of goodness-of-fit, we need to understand their properties under incorrect specifications. It turns out that when the average returns are cross-sectionally correlated, the sample R^2_{OLS} in the single CSR, as an increasing function of the square of the conventional OLS t-ratio, also tends to overstate the goodness-of-fit.

B. Misspecification Bias under Incorrect Specifications

Although misspecification may take various forms, we consider the extreme case where the chosen factor, g_t, is a useless factor in the sense that $\{R_1, R_2, \ldots, R_T\}$ are conditionally independent of $g = \{g_1, g_2, \ldots, g_T\}$. Without knowing g is useless, the researcher estimates the betas of the N assets with respect to the useless factor in the first-pass time series regression. Conditioned on g and under the assumption that $R_t \sim N(\mu, V)$, the estimated betas of the N assets, b, have a distribution given by

$$b \sim N(0, V/s_{gg}), \tag{15}$$

where $s_{gg} = \sum_{t=1}^{T}(g_t - \bar{g})^2$ and $\bar{g} = \sum_{t=1}^{T} g_t/T$. Since the distributions of the t-ratio and sample R^2 do not depend on s_{gg}, the unconditional distributions of the t-ratio and R^2 are identical to the ones conditioned on g. Therefore, we do not need to place any restrictions on the joint distribution of g. g_t could be correlated over time and it could even have time-varying distributions. Without knowing that b is the estimated betas with respect to a useless factor, the risk premium γ_1 is estimated in the second-pass CSR of R_t on b for each period and the hypothesis $H_0\colon \gamma_1 = 0$ is tested using the t-test described in the last subsection. Although the mean of b is zero for all assets, realizations are nonzero and may provide some explanatory power for expected returns. In this subsection, we address the following questions for a fixed number of time series observations, T:

1. For a given realization of b, what is the probability of rejecting $H_0\colon \gamma_1 = 0$ using the t-test? And what is the probability of rejecting $H_0\colon \gamma_1 = 0$, unconditioned on the realization of b, using the t-test to test if a useless factor is priced?
2. How do the goodness-of-fit measures, $R^2_{OLS}(b)$ and $R^2_{GLS}(b)$, of the CSR of $\bar{R}$ on b behave? Will the R^2s be inflated?

Running the CSR of R_t on $[1_N, b]$ using OLS and GLS, we obtain the OLS and GLS estimates of γ_1 for each period as

$$\hat{\gamma}_{1t}^{OLS}(b) = \frac{b'MR_t}{b'Mb}, \tag{16}$$

$$\hat{\gamma}_{1t}^{GLS}(b) = \frac{\tilde{b}'\widetilde{M}\tilde{R}_t}{\tilde{b}'\widetilde{M}\tilde{b}}, \tag{17}$$

where $\tilde{R}_t = V^{-1/2}R_t$ and $\tilde{b} = V^{-1/2}b$, and we can compute the OLS and GLS t-ratios in equations (6) and (11) to test $H_0\colon \gamma_1 = 0$. Researchers often treat $\hat{\gamma}_{1t}^{OLS}(b)$ and $\hat{\gamma}_{1t}^{GLS}(b)$ as i.i.d. normal conditioned on b, and compute the p-values of the OLS and GLS t-ratios based on a central t-distribution with $T - 1$ degrees of freedom. However, when we use the same period to estimate b as well as to run the OLS CSR, then $\hat{\gamma}_{1t}^{OLS}(b)$ and $\hat{\gamma}_{1t}^{GLS}(b)$ are no longer i.i.d. conditioned on b. We are able to derive the exact distribution of the t-ratios but due to its length and complexity, we choose to present the approximate distribution of the t-ratios by ignoring the dependence of $\hat{\gamma}_{1t}^{OLS}(b)$ and $\hat{\gamma}_{1t}^{GLS}(b)$ for simplicity.[4]

PROPOSITION 1: *If $\mu \neq k1_N$ for any scalar k, then, conditioned on b, the OLS and GLS t-ratios of testing $H_0\colon \gamma_1 = 0$ have an approximate noncentral t-distribution with noncentrality parameters*

$$\delta_{OLS}(b) = \frac{\sqrt{T}(b'M\mu)}{(b'MVMb)^{1/2}}, \tag{18}$$

$$\delta_{GLS}(b) = \frac{\sqrt{T}(\tilde{b}'\widetilde{M}\tilde{\mu})}{(\tilde{b}'\widetilde{M}\tilde{b})^{1/2}}, \tag{19}$$

and $T - 1$ degrees of freedom. Except for a set of b with a zero measure, $\delta_{OLS}(b) \neq 0$ and $\delta_{GLS}(b) \neq 0$. Unconditioned on b, the expected value and the variance of the OLS and GLS t-ratios are given by

$$E[t_{OLS}(b)] = 0, \tag{20}$$

$$\mathrm{Var}[t_{OLS}(b)] = \left(\frac{T-1}{T-3}\right) + \left(\frac{T-1}{T-3}\right)E[\delta_{OLS}^2(b)], \tag{21}$$

$$E[t_{GLS}(b)] = 0, \tag{22}$$

$$\mathrm{Var}[t_{GLS}(b)] = \left(\frac{T-1}{T-3}\right) + \left(\frac{T-1}{T-3}\right)E[\delta_{GLS}^2(b)]. \tag{23}$$

Therefore, conditioned on b, the probability of rejecting the null hypothesis using a two-tailed t-test will be higher than the size of the test obtained from a central t-distribution. In other words, one is likely to find b priced when using the t-test. Although conditioned on almost every b, the t-ratio does not have a mean equal to zero, the unconditional mean of the t-ratio is equal to zero and therefore we do not expect the estimated risk premium to take a particular sign. However, the unconditional variance of the t-ratio is higher than the variance of a central t-distribution, so we expect the two-tailed t-test to overreject the null hypothesis unconditionally. It should be emphasized that the results in Proposition 1 are just approximations. The

[4] The derivation of the exact distribution of the t-ratio is available upon request. The approximate distribution is very close to the exact distribution when T is reasonably large.

results based on the exact distribution of the t-ratios suggest it is possible that we can underreject the null hypothesis unconditionally when the factor is useless. This could happen when T is small or when μ is close to $k1_N$.

We now have the answer to the first question: In most of the cases, the null hypothesis $H_0 : \gamma_1 = 0$ will be rejected with a higher probability than the size of the test, due to the misspecification bias in the t-test. This indicates the serious problem in using this t-test to determine whether a beta risk is priced under incorrect specifications.

Note that since b is a random variable, both $\delta^2_{OLS}(b)$ and $\delta^2_{GLS}(b)$ are random variables. Though the misspecification bias applies to both the OLS and GLS t-tests, it is desirable to compare $E[\delta^2_{OLS}(b)]$ and $E[\delta^2_{GLS}(b)]$ because they reveal whether the OLS t-test or the GLS t-test is more susceptible to misspecification bias when T is finite. However, whether $E[\delta^2_{OLS}(b) - \delta^2_{GLS}(b)]$ is positive or negative depends on both μ and V, so one cannot make a general statement about relative superiority of OLS or GLS in detecting useless factors. The following proposition gives an analytical expression for $E[\delta^2_{OLS}(b)]$ and $E[\delta^2_{GLS}(b)]$. By eigenvalue decomposition, we have $V^{1/2}MV^{1/2} = H\Lambda H'$, where H is an $N \times (N-1)$ orthonormal matrix and $\Lambda = \mathrm{Diag}(\lambda_1, \ldots, \lambda_{N-1})$ where $0 < \lambda_1 \leq \cdots \leq \lambda_{N-1}$ are the $N-1$ nonzero eigenvalues of $V^{1/2}MV^{1/2}$.

PROPOSITION 2

$$E[\delta^2_{OLS}(b)] = E\left[\frac{T(Z'\Lambda\eta)^2}{Z'\Lambda^2 Z}\right] \leq \delta^2_{OLS}(\beta), \tag{24}$$

$$E[\delta^2_{GLS}(b)] = \frac{\delta^2_{GLS}(\beta)}{N-1} = \frac{T(\tilde{\mu}'\widetilde{M}\tilde{\mu})}{N-1}, \tag{25}$$

where $\eta = H'V^{-1/2}\mu$ *and* $Z \sim N(0, I_{N-1})$. *The exact analytical expression of* $E[\delta^2_{OLS}(b)]$ *is given in the Appendix.*

Proposition 2 suggests that for the GLS t-test when b is the estimated beta with respect to a useless factor, the unconditional expectation of the square of the noncentrality parameter is $1/(N-1)$ of that for the true beta. For the OLS t-test, even though b is the estimated beta with respect to a useless factor, the unconditional expectation of the square of the noncentrality parameter can be as high as that for the true beta. The equality can be attained when the returns have an exact one-factor structure without noise.[5] This is because when returns follow an exact one-factor structure without noise, we have

$$R_t = \mu + \beta f_t, \tag{26}$$

[5] We thank Naifu Chen for pointing this out to us.

and the estimated betas of the returns with respect to a useless factor g_t are given by

$$b = \frac{\sum_{t=1}^{T} R_t (g_t - \bar{g})}{\sum_{t=1}^{T} (g_t - \bar{g})^2}$$

$$= \frac{\sum_{t=1}^{T} \beta f_t (g_t - \bar{g})}{\sum_{t=1}^{T} (g_t - \bar{g})^2}$$

$$= c_{fg}\, \beta, \tag{27}$$

where $\bar{g} = \sum_{t=1}^{T} g_t / T$ and c_{fg} is the slope coefficient of regressing f_t on g_t. Although $E[c_{fg}] = 0$, its realization is not equal to zero with probability 1. Therefore, b is always a linear function of β and it is virtually impossible to distinguish betas estimated with respect to a useless factor from the betas with respect to a true factor.

Proposition 2 also suggests that, generally, the misspecification bias does not necessarily diminish as $N \to \infty$. For example, when $V = I_N$ (i.e., when there is no difference between OLS and GLS), we have

$$E[\delta^2_{GLS}(b)] = E[\delta^2_{OLS}(b)] = \frac{T(\mu' M \mu)}{N-1} = T\left[\frac{\sum_{i=1}^{N} (\mu_i - \bar{\mu})^2}{N-1}\right], \tag{28}$$

where $\bar{\mu} = (\mu' 1_N)/N$. The term in the last brackets is the cross-sectional variance of the expected returns of the N assets. To the extent that test assets are randomly drawn from a universe of firms, the cross-sectional variance does not decrease as the number of assets increases. Therefore, using more test assets may not reduce the misspecification bias.

For two independent random variables, it is well known that if one of these two variables has a spherical distribution, then the sample R^2 in the OLS regression of one variable on the other is distributed as a Beta$(\frac{1}{2}, (N-2)/2)$ where N is the number of observations, and the sample OLS R^2 has an expected value of $1/(N-1)$. However, in the OLS CSR of $\bar{R}$ on b, the distribution of R^2_{OLS} is difficult to derive because the N observations of b are correlated and the N observations of $\bar{R}$ are correlated and have different means. Although, under the normality assumption, $\bar{R}$ is independent of b, the OLS CSR R^2 does not have a Beta distribution except in some special cases, and its expected value is equal to (proof available upon request)

$$E[R^2_{OLS}(b)] = \sum_{i=1}^{N-1} E\left[\frac{\lambda_i Y_i^2}{\left(\sum_{j=1}^{N-1} \lambda_j Y_j^2\right)}\right] E\left[\frac{\lambda_i Z_i^2}{\left(\sum_{j=1}^{N-1} \lambda_j Z_j^2\right)}\right], \tag{29}$$

where $Y \sim N(\sqrt{T}\eta, I_{N-1})$ and $Z \sim N(0, I_{N-1})$ with η and λ_i as defined in Proposition 2. From this expression, we can see that when λ_{N-1} is very large compared with λ_1 to λ_{N-2} (for example, when returns are close to having an exact one-factor structure), $E[R^2_{OLS}(b)] \approx 1$ even if b is estimated with respect to a useless factor.

PROPOSITION 3: *In the cross-sectional regression of $\bar{R}$ on b, $R^2_{GLS}(b)$ is distributed as* Beta$(\frac{1}{2}, (N-2)/2)$. *$R^2_{OLS}(b)$ will be distributed as* Beta$(\frac{1}{2}, (N-2)/2)$ *if $MVM = cM$ for some constant $c > 0$.*

Therefore, a partial answer to the second question is that $R^2_{GLS}(b)$ of the CSR behaves like that of two independent variables with i.i.d. observations. Since its distribution is known and independent of T, R^2_{GLS} can be easily used to test whether the proposed factor is useless. However, since returns are cross-sectionally correlated, $R^2_{OLS}(b)$ of the CSR will typically not behave like the one between two independent variables with i.i.d. observations. Even though the factor is useless, the expected value of $R^2_{OLS}(b)$ can be much higher than $1/(N-1)$. This result should not be confused with those in Roll and Ross (1994), Grauer (1994), and Kandel and Stambaugh (1995). Their studies suggest that when betas are computed based on an inefficient portfolio, R^2_{OLS} and the risk premium could assume almost any value, having no relationship as to how close the inefficient portfolio is to the efficient frontier. Our results differ from theirs in two ways. First, we do not require the factor to be a portfolio of the N assets. Second, we deal with issues of sampling distribution, whereas they deal mainly with issues of population moments. If we can observe the population moments of useless factors and the returns of the assets (whose betas are all zero), we would not have any misspecification bias because the population R^2 should always be equal to zero. Although the problem we discuss here is related to theirs (i.e., misspecification), the results we obtain cannot be easily foreshadowed from their studies, which do not study the sampling distribution of R^2 or t-ratio.

C. Large Sample Properties

In this subsection, we discuss the properties of t-tests and R^2s when the number of time series observations, T, increases, assuming the first-pass time series regression and the second-pass CSR are performed using returns of the same period.[6] Since the bias problem with a useless factor arises from the errors of the estimated betas, and the variances of such estimated betas go down with T, one might expect that the misspecification bias in the t-test will diminish as T increases, and the t-test will work at least asymptotically. Unfortunately, this is not the case. In fact, the following proposition suggests that when T increases, the bias is even larger for the t-tests.

PROPOSITION 4: *Suppose $\mu \neq k1_N$ for any scalar k. As $T \to \infty$, $|t_{OLS}(b)| \to \infty$ and $|t_{GLS}(b)| \to \infty$ with probability one. As a result, the probability of rejecting the null hypothesis, $H_0\colon \gamma_1 = 0$, tends to one.*

This proposition illustrates the seriousness of misspecification bias in the two-pass methodology. The bias problem cannot be alleviated by increasing the number of time series observations. Intuitively, the reason for this result is that although the estimated betas become more accurate and tend to zero stochastically, the estimate of γ_1 does not. Since the expected returns, μ, are not constant across assets, the intercept term of the CSR cannot fully explain the variation in μ and it leaves something for the estimated betas to explain. As the estimates b tend to zero, the calculated OLS slope coefficient needs to go to infinity to explain the variation in μ,[7] so the numerator of the OLS t-ratio tends to infinity. However, the time series estimate of the standard error remains finite. That is, the denominator of the OLS t-ratio remains finite. As a result, when $T \to \infty$, $t_{OLS}(b)$ explodes. The same is true for $t_{GLS}(b)$.

One might suspect that this result comes from multicollinearity because, in the limit, the regressors of the CSR are an N-vector of ones and an N-vector of zeros. That this is not true can be shown in two ways. First, if one runs a CSR using excess returns without the intercept term, there will not be multicollinearity, but it can be easily shown that the problem remains. Second, suppose it is true that $\mu = k1_N$ for some scalar k. Then the multicollinearity is present, but it can be shown that the t-ratio has an asymptotic standard normal distribution and it will not explode. These two points make it clear that the problem comes from misspecification, rather than multicollinearity.

The following proposition describes the limit of R^2_{OLS} and R^2_{GLS} as T increases.

PROPOSITION 5: *Suppose $\mu \neq k1_N$ for any scalar k. Let $Z \sim N(0, I_{N-1})$. Denote*

$$h = \frac{(Z'\Lambda\eta)^2}{Z'\Lambda Z}, \tag{30}$$

[6] It is easy to show that Proposition 4 holds as long as the number of time series observations in performing the CSR goes to infinity. Therefore, the asymptotic results of Proposition 4 hold even when the betas of the useless factor are estimated using a different period from the one in which the CSR is performed.

[7] From the expression, $\bar{\hat{\gamma}}_1^{OLS}(b) = b'M\bar{R}/b'Mb$, it is easy to see that $|\bar{\hat{\gamma}}_1^{OLS}(b)| \to \infty$ as $b \to 0$, because $M\bar{R} \to M\mu \neq 0$ if $\mu \neq k1_N$ for any k, so the denominator tends to zero at a faster rate than the numerator.

where η and Λ are defined in Proposition 2. As $T \to \infty$,

$$R^2_{OLS} \xrightarrow{D} \frac{h}{\mu' M \mu}, \tag{31}$$

$$R^2_{GLS} \xrightarrow{D} \text{Beta}\left(\frac{1}{2}, \frac{N-1}{2}\right). \tag{32}$$

This proposition says that both R^2_{OLS} and R^2_{GLS} converge in distribution to some random variables. They do not converge to zero, even though the betas of a useless factor all converge to zeros. The result for R^2_{GLS} is anticipated from Proposition 3, which states that the distribution of R^2_{GLS} is $\text{Beta}(\frac{1}{2},(N-2)/2)$ and does not depend on T.

Since the t-tests are not reliable, and from Proposition 4 even more unreliable for larger T, the fact that R^2_{GLS} has a Beta distribution for the case of useless factors provides us with a diagnostic test to detect useless factors. In theory, one can also use R^2_{OLS} to detect useless factors if one knows the distribution of sample R^2_{OLS}. However, the distribution of sample R^2_{OLS} for the useless factor case depends on T, μ, and V. Since information on μ and V is generally not available to the researcher, it is difficult to assess the distribution of sample R^2_{OLS} even through simulations. Although sample R^2_{GLS} is superior to sample R^2_{OLS} in testing whether the factor is useless, one should not interpret our results as suggesting that R^2_{GLS} is superior to R^2_{OLS} in every situation. If R^2 is not used for detecting useless factors but is meant to provide a measure of goodness-of-fit to compare models, then to the extent that the test assets are economically meaningful, Jagannathan and Wang (1996) argue that R^2_{OLS} could be a better metric than R^2_{GLS}.

D. EIV Adjustment

Since the misspecification bias is partly due to the fact that we estimate betas with errors in the first-pass time series regression, one may think that we can correct the problem using the EIV adjustment proposed by Shanken (1992). Instead of computing OLS and GLS t-ratios as in equations (6) and (11), Shanken (1992) suggests that when the betas are estimated with errors, we should compute the EIV-adjusted OLS and GLS t-ratios as:

$$t^*_{OLS} = \frac{\bar{\hat{\gamma}}_1^{OLS}}{\left[\dfrac{s^2(\hat{\gamma}_1^{OLS})}{T} + \left(\dfrac{\bar{\hat{\gamma}}_1^{OLS}}{\hat{\sigma}_g}\right)^2 \left(\dfrac{s^2(\hat{\gamma}_1^{OLS})}{T} - \dfrac{\hat{\sigma}_g^2}{T}\right)\right]^{1/2}}, \tag{33}$$

$$t^*_{GLS} = \frac{\bar{\hat{\gamma}}_1^{GLS}}{\left[\dfrac{s^2(\hat{\gamma}_1^{GLS})}{T} + \left(\dfrac{\bar{\hat{\gamma}}_1^{GLS}}{\hat{\sigma}_g}\right)^2 \left(\dfrac{s^2(\hat{\gamma}_1^{GLS})}{T} - \dfrac{\hat{\sigma}_g^2}{T}\right)\right]^{1/2}}, \tag{34}$$

where $\hat{\sigma}_g^2 = s_{gg}/T$. The second term in the denominator is designed to take into account the measurement errors in the estimated betas. Shanken (1992) states that such an adjustment is not needed when testing $H_0\colon \gamma_1 = 0$. But this is true only if the model is correctly specified under the null hypothesis, i.e., $\mu = \gamma_0 1_N$ and expected returns are constant across assets. Here, we are interested in the properties of adjusted t-tests in the case of useless factors.

Since $s^2(\hat{\gamma}_1^{OLS}) > \hat{\sigma}_g^2$ and $s^2(\hat{\gamma}_1^{GLS}) > \hat{\sigma}_g^2$, we have $|t_{OLS}^*| < |t_{OLS}|$ and $|t_{GLS}^*| < |t_{GLS}|$ in any sample, so the rejection rate of the null hypothesis using the adjusted t-ratio will always be lower than the rejection rate of using the unadjusted t-ratio. Therefore, the adjusted t-ratio can help to reduce the misspecification bias even though it is incorrect for the case of useless factor.

Unlike the unadjusted t-ratios, the limits of adjusted t-ratios as the sample size T goes to infinity are finite (though unbounded) random variables. The properties of the asymptotic distributions are given in the following proposition. For convenience, they are stated in terms of t_{OLS}^{*2} and t_{GLS}^{*2}. Let $F_{\lim t_{OLS}^{*2}}(u)$ be the limiting distribution function of t_{OLS}^{*2} and let $F_{\chi_k^2}(u)$ be the distribution function of a χ_k^2 variable.

Proposition 6: *As T goes to infinity,*

$$t_{OLS}^{*2} \xrightarrow{D} \frac{(Z'\Lambda Z)^2}{Z'\Lambda^2 Z} \tag{35}$$

$$t_{GLS}^{*2} \xrightarrow{D} \chi_{N-1}^2, \tag{36}$$

where $Z \sim N(0, I_{N-1})$ and Λ is defined in Proposition 2, and

$$F_{\chi_1^2}(u) > F_{\lim t_{OLS}^{*2}}(u) \geq F_{\chi_{N-1}^2}(u) \quad \text{for } u > 0. \tag{37}$$

Under the correctly specified model, t_{OLS}^{*2} and t_{GLS}^{*2} both should have limiting distribution of χ_1^2, so the acceptance/rejection decision is based on the distribution of χ_1^2. However, when the factor is useless, t_{OLS}^{*2} and t_{GLS}^{*2} no longer have limiting distribution of χ_1^2. When $N > 2$, overrejection occurs asymptotically if we use the EIV-adjusted GLS t-ratio. For the case of OLS, the overrejection also occurs asymptotically, but the severity depends on the relative magnitude of the nonzero eigenvalues of $V^{1/2}MV^{1/2}$. If the nonzero eigenvalues are equal, the limiting distribution of the t_{OLS}^{*2} is χ_{N-1}^2 and the overrejection is severe. If the nonzero eigenvalues are unequal, the limiting distribution of t_{OLS}^{*2} is closer to that of χ_1^2 and the overrejection is less severe.

E. Discussion

To understand why the misspecification bias occurs, we examine the null and the alternative hypotheses of the test. When the asset pricing model is correctly specified and the factor is correctly identified, the alternative hypothesis says that the expected returns are a linear function of the betas with respect to the true factor, but the null says the expected returns are constant across all assets. Such a null should be rejected in favor of the alternative when the true beta risk is priced. However, when the asset pricing model is incorrectly specified, then the alternative hypothesis, which says that the beta risk of a useless factor is priced, and the null, which says that the expected returns are constant across assets, are both wrong. As a result, there will be a good chance for the null hypothesis to be rejected in favor of the alternative. Rejection simply means that the alternative (that the useless factor is priced) is better than the null (that expected returns are constant across assets), which is not a very interesting benchmark. As the number of time series observations increases, the problem gets even worse, as we explain in the previous subsection.

The two-pass methodology has its advantage when the model is correctly specified, but our analysis indicates that the methodology is inadequate in the case of useless factors. A natural question is how one can detect a useless factor. We offer some suggestions.

The first suggestion is to test whether the betas of the assets with respect to a particular factor are significantly different from zero in the first-pass time series regression before we run the second-pass CSR.[8] If we cannot reject the hypothesis that the betas are jointly equal to zero, then we should be concerned about whether the factor is useless. Chen, Roll, and Ross (1986) and Ferson and Harvey (1993) performed such a test in their studies, but unfortunately this procedure has been largely ignored by many researchers.

The second suggestion pertains to the use of R^2_{OLS} and R^2_{GLS} in detecting useless factors. Since R^2_{OLS} can be highly inflated for useless factors and its distribution is in general unknown, simulations are required to find out its distribution in order to determine whether the factor is useless. From Proposition 3, R^2_{GLS} is not inflated much by the misspecification and its distribution for a useless factor is known. Therefore, in the CSR of the returns on the betas with respect to a common factor, it is more convenient to use R^2_{GLS} to check if the factor is useless. However, as a goodness-of-fit measure to compare models, Jagannathan and Wang (1996) suggests that R^2_{GLS} is inferior to R^2_{OLS} because the latter measure applies to the transformed returns and betas.

The third suggestion is to use Shanken's EIV-adjusted t-ratio for testing the null hypothesis in the case of conditional homoskedastic returns. Such an adjustment is far from being perfect, but it helps reduce the overrejection rate as compared with the unadjusted t-ratio. More generally, one can use the EIV-adjusted t-ratio of Jagannathan and Wang (1997), which allows for conditional heteroskedasticity in the returns.

[8] In theory, a variable could be a legitimate factor even though it has very low (but not zero) correlations with returns, because if we add pure measurement errors to a true factor, the betas with respect to this new factor still explain the cross-sectional differences of expected returns perfectly. Empirically, a very noisy factor is not very useful because there will be large errors in the estimated betas.

If there are two or more independent sets of samples available, it is always beneficial to perform the test on all the samples separately and draw inference upon the joint results. But for a given sample, a fourth way to detect useless factors is to utilize the fact that the t-ratios of a useless factor have an unconditional mean equal to zero. The difference between a true factor and a useless factor is that the beta estimates are stable for a true factor, but unstable for a useless factor. Thus, we can split the whole sample period into several subperiods, estimate betas of the assets for each subperiod, perform CSR for each subperiod using estimated betas from the respective subperiod, and reject the hypothesis $\gamma_1 = 0$ if the hypothesis is rejected in all subperiods in the same direction.[9] One empirical question is the trade-off between detecting useless factors and maintaining the power of the test under the correct model. In the next section, we illustrate this with simulation and report the performance of such a test.

II. Simulation Results

A. The Data

To evaluate the magnitude of the misspecification bias discussed in the previous section, we rely on simulation evidence. We use both actual and simulated returns. For actual returns, we choose two sets of portfolios which are commonly used in the empirical literature. The first set is 10 size-ranked equally weighted portfolios of the combined NYSE and AMEX stocks, sorted by market value at the end of June in each year. The second set of portfolios is 100 size-and-beta-ranked portfolios that are obtained by ranking the stocks within each size portfolio by their value-weighted betas estimated using 24 to 60 months of past return data and subdividing them into 10 beta portfolios. The portfolios are equally weighted and are rebalanced on a monthly basis. The monthly return series for both sets covers from July 1963 to December 1990.[10]

While actual portfolio returns are relevant in evaluating the impact of misspecification bias in actual empirical studies, actual returns alone are not sufficient for us to gauge its impact. The problem is that we do not know the data generating process of actual returns, and because we have only one realization of actual returns, the analysis is bound to be a conditional analysis and may not be generalizable to other realizations of returns. Moreover, actual returns could be nonnormal, conditionally heteroskedastic, and serially correlated. All these features may bias the t-test, thus we cannot attribute the overrejection of $H_0: \gamma_1 = 0$ for useless factors entirely to the problem we

[9] That is, all the γ_1s are significantly positive, or all are significantly negative, with the size appropriately adjusted. In practice, the subperiod joint test has an additional advantage when the betas are time varying.

[10] Monthly returns on the 100 size-beta portfolios are kindly provided to us by Jagannathan and Wang. Monthly returns on the 10 size portfolios are constructed based on these 100 size-beta portfolios.

discuss in the previous section. For this reason, we use simulated returns to demonstrate the magnitude of the misspecification bias due to useless factors only. To this end, we simulate i.i.d. returns from $N(\mu,V)$ where μ and V are set equal to the average and estimated variance-covariance matrix of the actual returns. To facilitate comparison with the results of using actual portfolio returns, we generate two sets of parameters for the simulated returns. One set corresponds to the 10 size portfolios and the other corresponds to the 100 size-beta portfolios. If we observe a bias of similar magnitude in both actual returns and simulated returns, we then have more confidence that the observed magnitude of the bias in the actual returns is not driven by the other violations of the assumptions for the t-test and that the bias is not unique to a particular realization of returns.

B. Simulation Results for Fixed T

In this subsection, we report simulation results with a fixed number of time series observations: $T = 330$, which corresponds to the number of time series observations used by Fama and French (1992) and Jagannathan and Wang (1996). The purpose of the simulations is to find out the magnitude of the misspecification bias that we can expect to observe in real world data. The number of replications in all our simulations is 10,000.

First, we take the actual returns as given, and for each simulation a useless factor is generated as an independent $N(0,1)$ variate. (The mean and standard deviation of the useless factor are irrelevant.[11]) Using the two-pass procedure, the betas with respect to this useless factor are estimated in the first-pass time series regression and γ_1 is estimated in the second-pass CSR. The purpose of this exercise is to determine how often the betas of useless factors will be found priced. Table I reports the results of this experiment. The left half of Table I reports the simulation results of CSR using only the betas of the useless factor. In the case of 10 size portfolios, rejection rates are quite different between OLS and GLS t-tests. For OLS CSR, the two-tailed t-test slightly underrejects the null hypothesis for significance levels at 1 percent and 5 percent, but grossly overrejects the null at the 10 percent level. For GLS CSR, the two-tailed t-test overrejects the null at all three significance levels but the rejection rates are not as bad as OLS CSR at the 10 percent level. One possible explanation for the difference between OLS and GLS results is that actual returns are fraught with problems of conditional heteroskedasticity, nonnormality, and a time-varying factor structure and these problems have different impacts on the OLS t-test and GLS t-test. Whatever the reason for the overrejection, the simulation results indicate

[11] If R_t is i.i.d. $N(\mu,V)$, then the distribution and time series properties of the useless factor are irrelevant. However, when we condition on actual returns, the choice of the distribution and time series properties of the useless factor could matter. We generate the useless factor as an independent $N(0,1)$ variate here just to illustrate the typical magnitude of the misspecification bias. Conditioned on the actual data, the misspecification bias could be higher or lower than what we report if the useless factor is not normally distributed or autocorrelated.

Table I

Probability of Rejecting $H_0: \gamma_1 = 0$ and Empirical Distribution of R^2 Using Actual Returns and Estimated Betas of a Random Factor

The table presents the probability of rejecting $H_0: \gamma_1 = 0$ in 10,000 simulations using the two-tailed t-test at various significance levels. N is the number of assets. For $N = 10$, the assets are 10 size portfolios and for $N = 100$, the assets are 100 size-beta portfolios. In both cases, the returns are equally weighted monthly returns constructed using the combined NYSE-AMEX monthly file over the period July 1963 to December 1990 ($T = 330$). In each simulation, a useless factor of 330 observations is randomly drawn from $N(0,1)$ and the beta with respect to this factor, b, is estimated for each asset. Monthly returns, R_t are regressed on b with and without β^{vw}, where β^{vw} is the beta of the returns with respect to the value-weighted NYSE-AMEX market index:

$$R_t = \gamma_0 1_N + \gamma_1 b + \varepsilon_t,$$

$$R_t = \gamma_0 1_N + \gamma_1 b + \gamma_2 \beta^{vw} + \varepsilon_t,$$

using ordinary least squares (OLS) and generalized least squares (GLS). A two-tailed t-test is performed to test $H_0: \gamma_1 = 0$ using the time series of the estimated slope coefficients. The table also presents the empirical distribution of the OLS and GLS R^2 between the average returns and the fitted expected returns in the 10,000 simulations.

	Without β^{vw}				With β^{vw}			
	$N = 10$		$N = 100$		$N = 10$		$N = 100$	
Panel A: Probability of Rejecting $H_0: \gamma_1 = 0$								
Significance Level	OLS	GLS	OLS	GLS	OLS	GLS	OLS	GLS
0.01	0.001	0.029	0.033	0.031	0.008	0.038	0.082	0.036
0.05	0.045	0.127	0.142	0.104	0.090	0.143	0.413	0.108
0.10	0.692	0.206	0.232	0.168	0.294	0.238	0.595	0.176
Panel B: Distribution of R^2								
Percentile	OLS	GLS	OLS	GLS	OLS	GLS	OLS	GLS
0.010	0.058	0.002	0.002	0.000	68.059	0.047	1.349	0.664
0.025	0.366	0.012	0.012	0.001	68.093	0.059	1.365	0.665
0.050	1.433	0.051	0.045	0.005	68.205	0.102	1.425	0.668
0.100	5.088	0.207	0.182	0.018	68.704	0.280	1.635	0.682
0.200	17.308	0.840	0.727	0.067	70.469	1.033	2.521	0.734
0.300	31.457	1.991	1.591	0.152	73.012	2.288	3.921	0.823
0.400	46.251	3.641	2.842	0.282	75.977	4.109	5.995	0.952
0.500	57.768	5.971	4.557	0.457	78.810	6.699	8.637	1.134
0.600	67.457	9.052	6.921	0.716	81.452	10.149	11.871	1.400
0.700	75.449	13.384	9.989	1.070	84.256	15.065	15.975	1.781
0.800	81.584	19.480	13.982	1.626	86.919	21.877	21.171	2.329
0.900	86.914	30.547	20.887	2.735	89.993	32.655	28.276	3.464
0.950	89.992	40.264	26.111	3.871	92.042	43.547	33.981	4.658
0.975	92.075	48.438	31.128	5.042	93.555	52.817	38.349	5.809
0.990	93.979	59.573	35.981	6.664	95.057	62.065	43.282	7.424
Average	51.547	11.178	7.813	1.002	79.080	12.259	12.119	1.698

that the t-tests are not reliable when the model is misspecified. As for the case of 100 size-beta portfolios, the overrejection rates are in general lower than the 10 size portfolios case but they are still very significant.

The bias of the t-test on the risk premium of the useless factor exists because the model is misspecified. Some may think that including more factors may help to alleviate this problem. This is not necessarily the case if the additional factors included are also not the correct factors. To illustrate this, we add the betas of the portfolios with respect to the value-weighted NYSE-AMEX market index in the CSR along with the betas of the useless factor. The betas of both the useless factor and the value-weighted NYSE-AMEX market index are simple regression betas, so the set of value-weighted NYSE-AMEX betas stays the same on every simulation.[12] The right half of Table I reports the simulation results. Compared with the results without the value-weighted NYSE-AMEX betas, we find that in most cases the overrejection rates of $H_0 : \gamma_1 = 0$ for the useless factor are substantially increased. Therefore, including more factors in the model does not always help to exclude useless factors. It could even exacerbate the misspecification bias if the other factors included in the model do not nest the true data generating process.

The sample R^2s for OLS regressions of $\bar{R}$ on b appear to be very high. The sample R^2_{GLS}, by Proposition 3, should follow a Beta distribution with mean $1/(N-1)$. Although the GLS is run with an estimated V matrix, the distribution of sample R^2_{GLS} in Table I is still quite similar to a Beta distribution.

Though the results in Table I show that the bias in the t-test is quite severe, one could attribute the bias to many possible sources. In Table II, we repeat the same experiment except that in every replication we simulate both the returns from $N(\mu, V)$ and the useless factor from $N(0,1)$. By simulating also the returns in every replication, we can isolate the magnitude of the misspecification bias from other sources of bias in the t-test.

For GLS CSR, we report both the true GLS (using the actual V) and the estimated GLS (using the estimated $\hat{V}$). For the t-ratios, we can see that the rejection rates are mostly higher than the ones reported in Table I, indicating that the misspecification bias is one of the main reasons causing overrejection of $H_0 : \gamma_1 = 0$. In the 10 assets case, we find that the estimated GLS has roughly the same properties as the true GLS since with 330 observations, the variance-covariance matrix of the returns on the 10 assets can be estimated quite accurately. However, in the case of 100 assets, the variance-covariance matrix of their returns cannot be estimated very accurately with only 330 observations and it introduces another source of bias in the GLS CSR t-ratio. The effect of this bias is to further increase the rejection rate. Therefore, although true GLS is less likely to reject $H_0 : \gamma_1 = 0$ for the useless factors given our choice of parameters, the estimated GLS turns out to be worse than the OLS in the case of 100 assets because the estimation errors of the variance-covariance matrix of the returns further contaminate the t-ratios.

[12] We have also performed simulations using multiple regression betas and the results are qualitatively similar.

Table II

Probability of Rejecting $H_0: \gamma_1 = 0$ and Empirical Distribution of R^2 Using Simulated Returns and Estimated Betas of a Random Factor

The table presents the probability of rejecting $H_0: \gamma_1 = 0$ in 10,000 simulations using the two-tailed t-test at various significance levels. In each simulation, 330 observations of a useless factor are randomly drawn from $N(0,1)$ and 330 observations of returns on N assets are independently drawn from $N(\mu, V)$, where μ and V are chosen based on the sample estimates over the period July 1963 to December 1990. The parameters are estimated from 10 size portfolios for $N = 10$, and from 100 size-beta portfolios for $N = 100$. The simulated monthly returns, R_t, are regressed on their betas (b) estimated with respect to the useless factor:

$$R_t = \gamma_0 1_N + \gamma_1 b + \varepsilon_t,$$

using ordinary least squares (OLS), true generalized least squares (GLS) (using true V) and estimated GLS (using $\hat{V}$ estimated from the simulated returns). A two-tailed t-test is performed to test $H_0: \gamma_1 = 0$ using the time series of the estimated slope coefficients. The table also presents the empirical distribution of OLS, true GLS, and estimated GLS R^2 between the average returns and the fitted expected returns in the 10,000 simulations.

	$N = 10$			N = 100		
Panel A: Probability of Rejecting $H_0: \gamma_1 = 0$						
Significance level	OLS	GLS(V)	GLS($\hat{V}$)	OLS	GLS(V)	GLS($\hat{V}$)
0.01	0.188	0.107	0.114	0.111	0.061	0.172
0.05	0.386	0.231	0.242	0.236	0.154	0.297
0.10	0.502	0.318	0.332	0.320	0.238	0.384
Panel B: Distribution of R^2						
Percentile	OLS	GLS(V)	GLS($\hat{V}$)	OLS	GLS(V)	GLS($\hat{V}$)
0.010	0.033	0.002	0.003	0.002	0.000	0.000
0.025	0.168	0.011	0.012	0.012	0.001	0.001
0.050	0.722	0.047	0.048	0.045	0.004	0.004
0.100	2.864	0.211	0.214	0.160	0.016	0.016
0.200	10.957	0.898	0.883	0.634	0.064	0.065
0.300	22.090	2.054	1.987	1.426	0.157	0.152
0.400	34.139	3.735	3.670	2.594	0.284	0.283
0.500	44.963	6.055	6.030	4.233	0.474	0.465
0.600	56.651	9.295	9.200	6.454	0.736	0.728
0.700	66.416	13.628	13.695	9.593	1.108	1.108
0.800	75.161	20.064	19.985	13.863	1.680	1.690
0.900	83.745	30.129	30.518	21.045	2.765	2.758
0.950	88.624	39.908	40.099	28.023	3.941	3.881
0.975	91.477	48.553	48.804	33.729	5.045	5.120
0.990	93.796	58.373	58.458	39.392	6.536	6.355
Average	44.333	11.281	11.273	7.846	1.013	1.011

For the R^2, we observe the same pattern as in Table I. The R^2_{OLS} is highly inflated, but R^2_{GLS} (for both true GLS and estimated GLS) behaves quite like a Beta distribution with a mean approximately equal to $1/(N - 1)$.

In summary, our simulation experiment shows that the misspecification bias could lead to overrejection of the null hypothesis using the central

t-distribution. The magnitude of this bias is significant for the typical return data used in empirical research and should not be ignored. The results suggest that one should not jump to the conclusion that a factor is "priced" whenever one finds its estimated risk premium has a significant t-ratio. Furthermore, we find that the R^2_{OLS} is inflated by a large amount and renders it inappropriate to detect useless factors. On the other hand, R^2_{GLS} is not subject to this problem and can serve as a useful measure to detect useless factors.

C. The Rejection Rate as T Increases

In Table III, we report simulation results for different numbers of time series observations used in OLS and GLS CSR in the cases of 10 assets and 100 assets. The parameters of the two sets of returns are chosen in exactly the same way as in Table II, but we increase the length of time series observations of the simulated returns from $T = 120$ to $T = 1200$ by an increment of 120. By looking at time series of different lengths, we can better understand the magnitude of misspecification biases for different samples. Table III shows that the unconditional means of the computed OLS and GLS t-ratios are very close to zero, but their variances go up roughly linearly with T. It is obvious that they can get much higher than that of a central t-distribution. As a result, the rejection rates using the central t-distribution are much higher than the one suggested by the size of the test. Even for $T = 120$, we find that the rejection rates are often more than twice the size of the test. When $T = 1200$, we find that a useless factor is priced at the 10 percent level with a probability of more than 0.5, and can be as high as 0.892 (for OLS CSR when $N = 10$). This experiment shows how fast the rejection rate increases and how the misspecification bias becomes more severe as the length of time series increases.

In Table IV, we report the simulation results of using the EIV-adjusted t-ratios instead of using the unadjusted t-ratios. Similar to the unadjusted t-ratios, we find the rejection rate of the null hypothesis $H_0: \gamma_1 = 0$ to be an increasing function of the length of time series observations. The unconditional means of the EIV-adjusted OLS and GLS t-ratios are very close to zero, and their variances go up with T. However, unlike the unadjusted t-ratios, the variances of the adjusted t-ratios do not explode but instead converge to some limits. The rejection rates of using the EIV-adjusted t-ratios are in general less than the numbers in Table III for the unadjusted t-ratios. However, for $N = 100$, the rejection rates of using the EIV-adjusted t-ratio are still very close to the ones using the unadjusted t-ratio. For $T = 360$, the EIV-adjusted t-ratio rejects the null hypothesis with a probability of more than twice the size of the test. The only case where the EIV-adjusted t-ratio does not create significant overrejection of the null hypothesis is the OLS case for $N = 10$. Therefore, although EIV-adjusted t-ratios are better behaved than the unadjusted ones, we still find useless factors to be priced when T or N are large.

Table III

Probability of Rejecting $H_0: \gamma_1 = 0$ and Unconditional Mean and Variance of OLS and GLS t-ratios of the Risk Premium Associated with the Betas of a Random Factor for Different Lengths of Time Series

The table presents the probability of rejecting $H_0: \gamma_1 = 0$ in 10,000 simulations using the two-tailed t-test at various significance levels and the mean and variance of the t-ratios for different lengths of time series (T). In each simulation, T observations of a useless factor are randomly drawn from $N(0,1)$ and T observations of returns on N assets are independently drawn from $N(\mu, V)$, where μ and V are chosen based on the sample estimates over the period July 1963 to December 1990. The parameters are estimated from 10 size portfolios for $N = 10$, and from 100 size-beta portfolios for $N = 100$. The simulated monthly returns, R_t, are regressed on their betas (b) estimated with respect to the useless factor:

$$R_t = \gamma_0 1_N + \gamma_1 b + \varepsilon_t,$$

using ordinary least squares (OLS) and estimated generalized least squares (GLS). A two-tailed t-test is performed to test $H_0: \gamma_1 = 0$ using the time series of the estimated slope coefficients. The table also presents the mean and the variance of the OLS and GLS t-ratios in the 10,000 simulations.

	OLS					GLS				
	t-ratio		Prob. of Rejecting $H_0: \gamma_1 = 0$			t-ratio		Prob. of Rejecting $H_0: \gamma_1 = 0$		
T	Mean	Var.	1%	5%	10%	Mean	Var.	1%	5%	10%
Panel A: $N = 10$										
120	0.031	1.998	0.055	0.166	0.258	0.003	1.686	0.044	0.128	0.202
240	−0.004	3.031	0.121	0.289	0.405	−0.001	2.286	0.083	0.191	0.277
360	0.004	4.047	0.202	0.405	0.525	−0.003	2.799	0.126	0.247	0.336
480	−0.002	5.066	0.285	0.504	0.622	0.008	3.333	0.161	0.290	0.378
600	0.021	6.142	0.374	0.600	0.707	0.007	3.993	0.204	0.343	0.427
720	0.002	7.083	0.449	0.672	0.767	0.012	4.539	0.237	0.379	0.461
840	0.015	8.041	0.516	0.729	0.810	0.018	5.077	0.270	0.402	0.484
960	0.034	9.126	0.587	0.774	0.841	0.025	5.708	0.300	0.432	0.513
1080	0.041	10.098	0.642	0.810	0.868	0.017	6.233	0.322	0.455	0.534
1200	0.031	11.165	0.696	0.841	0.892	0.021	6.842	0.343	0.482	0.559
Panel B: $N = 100$										
120	0.017	1.487	0.032	0.102	0.176	−0.072	9.817	0.390	0.506	0.579
240	0.017	2.164	0.077	0.184	0.265	−0.022	3.569	0.169	0.297	0.378
360	−0.003	2.788	0.120	0.254	0.338	−0.044	3.535	0.171	0.302	0.383
480	−0.022	3.392	0.169	0.303	0.389	−0.041	3.886	0.189	0.325	0.410
600	−0.034	3.974	0.209	0.343	0.428	−0.048	4.361	0.217	0.344	0.431
720	−0.027	4.654	0.248	0.385	0.469	−0.065	4.870	0.243	0.376	0.454
840	−0.025	5.277	0.280	0.414	0.502	−0.081	5.325	0.266	0.398	0.482
960	−0.014	5.832	0.301	0.442	0.517	−0.072	5.774	0.285	0.422	0.501
1080	−0.020	6.489	0.331	0.466	0.544	−0.074	6.298	0.309	0.438	0.514
1200	−0.014	7.178	0.356	0.488	0.563	−0.068	6.853	0.326	0.458	0.535

Table IV

Probability of Rejecting $H_0 : \gamma_1 = 0$ and Unconditional Mean and Variance of OLS and GLS EIV-adjusted *t*-ratios of the Risk Premium Associated with the Betas of a Random Factor for Different Lengths of Time Series

The table presents the probability of rejecting $H_0 : \gamma_1 = 0$ in 10,000 simulations using the two-tailed errors-in-variables (EIV) adjusted t-test at various significance levels and the mean and variance of the t-ratios for different lengths of time series (T) as well as for the limiting distribution. In each simulation, T observations of a useless factor are randomly drawn from $N(0,1)$ and T observations of returns on N assets are independently drawn from $N(\mu, V)$, where μ and V are chosen based on the sample estimates over the period July 1963 to December 1990. The parameters are estimated from 10 size portfolios for $N = 10$, and from 100 size-beta portfolios for $N = 100$. The simulated monthly returns, R_t, are regressed on their betas (b) estimated with respect to the useless factor:

$$R_t = \gamma_0 1_N + \gamma_1 b + \varepsilon_t,$$

using ordinary least squares (OLS) and estimated generalized least squares (GLS). A two-tailed EIV-adjusted t-test is performed to test $H_0 : \gamma_1 = 0$ using the time series of the estimated slope coefficients.

	OLS					GLS				
	EIV-adjusted t-ratio		Prob. of Rejecting $H_0 : \gamma_1 = 0$			EIV-adjusted t-ratio		Prob. of Rejecting $H_0 : \gamma_1 = 0$		
T	Mean	Var.	1%	5%	10%	Mean	Var.	1%	5%	10%
Panel A: $N = 10$										
120	0.017	0.696	0.000	0.002	0.012	0.002	1.144	0.002	0.044	0.115
240	0.000	0.895	0.000	0.004	0.024	0.004	1.403	0.006	0.073	0.167
360	0.003	1.037	0.000	0.006	0.034	−0.002	1.608	0.009	0.102	0.216
480	−0.003	1.156	0.000	0.010	0.049	0.003	1.798	0.018	0.130	0.253
600	0.011	1.255	0.000	0.013	0.061	0.002	2.014	0.024	0.164	0.302
720	0.003	1.335	0.000	0.016	0.072	0.006	2.183	0.035	0.194	0.334
840	0.008	1.409	0.000	0.020	0.086	0.010	2.331	0.044	0.222	0.365
960	0.017	1.476	0.001	0.026	0.104	0.013	2.484	0.052	0.246	0.393
1080	0.015	1.514	0.001	0.031	0.108	0.010	2.604	0.060	0.264	0.415
1200	0.017	1.554	0.001	0.034	0.120	0.010	2.734	0.069	0.282	0.438
Limit	0.000	2.335	0.039	0.153	0.286	0.000	9.000	0.675	0.922	0.975
Panel B: $N = 100$										
120	0.014	1.140	0.005	0.050	0.116	−0.069	9.309	0.388	0.506	0.578
240	0.010	1.553	0.021	0.103	0.199	−0.020	3.362	0.159	0.292	0.374
360	−0.003	1.894	0.036	0.157	0.268	−0.043	3.289	0.160	0.295	0.378
480	−0.018	2.193	0.058	0.202	0.317	−0.040	3.568	0.176	0.318	0.404
600	−0.023	2.455	0.081	0.235	0.353	−0.046	3.945	0.202	0.337	0.425
720	−0.019	2.743	0.101	0.275	0.390	−0.061	4.354	0.229	0.370	0.450
840	−0.015	2.989	0.121	0.304	0.423	−0.077	4.710	0.252	0.390	0.476
960	−0.009	3.153	0.140	0.324	0.444	−0.068	5.051	0.271	0.415	0.496
1080	−0.009	3.378	0.157	0.350	0.469	−0.069	5.442	0.295	0.429	0.507
1200	−0.008	3.576	0.178	0.370	0.488	−0.066	5.840	0.310	0.449	0.529
Limit	0.000	14.029	0.881	1.000	1.000	0.000	99.000	1.000	1.000	1.000

D. Subperiod Joint Test

As we have suggested, one way to mitigate the misspecification bias caused by a useless factor is to run CSR for different subperiods and reject the hypothesis $H_0 : \gamma_1 = 0$ if and only if the t-ratios have the same sign and are significant in all subperiods. Let α be the significance level for such a test. Denote $t_n(\alpha)$ as the upper 100α percentage points of the central t-distribution with n degrees of freedom. For an entire period of T observations, let t_1 and t_2 be the t-ratios for two subperiods of $T/2$ observations. The subperiod joint test is to reject $H_0 : \gamma_1 = 0$ at a significance level of α if and only if

$$t_1 > t_{\frac{T}{2}-1}\left(\sqrt{\frac{\alpha}{2}}\right) \quad \text{and} \quad t_2 > t_{\frac{T}{2}-1}\left(\sqrt{\frac{\alpha}{2}}\right) \tag{38}$$

or

$$t_1 < -t_{\frac{T}{2}-1}\left(\sqrt{\frac{\alpha}{2}}\right) \quad \text{and} \quad t_2 < -t_{\frac{T}{2}-1}\left(\sqrt{\frac{\alpha}{2}}\right). \tag{39}$$

Now, under the assumption that returns are uncorrelated over time, such a test has a significance level α if the null hypothesis is correct. To see how such a test behaves for a useless factor, we again rely on simulation with both returns and useless factors simulated.

Table V reports the results of the rejection rates for both the true beta and the beta of a useless factor using the t-test over the entire period and the joint t-test over two subperiods. The true beta used in simulation is simply the vector μ for generating returns. For the true beta, the rejection rates from the joint t-test over two subperiods are still quite high as compared with the rejection rates from the t-test over the entire period. They are only slightly smaller for the OLS and the true GLS, indicating the small loss of the power of the joint t-test. For estimated GLS with $N = 10$, the power is even higher. This means that the subperiod joint test still maintains relatively high power in rejecting the null hypothesis under the correctly specified model. For the beta of a useless factor, although the subperiod joint test still overrejects the null hypothesis, the rejection rates are substantially reduced compared to those of the full period test. On average, the rejection rates are reduced by more than half. We conclude that the subperiod joint test is fairly effective in detecting useless factors without sacrificing too much the ability to reject the null hypothesis under the correctly specified model.

III. Concluding Remarks

In this paper, we argue that there is a problem of misspecification bias in the two-pass methodology of testing beta pricing models when the factor is misspecified. This problem renders the t-test inadequate. Simulation evi-

Table V

Probability of Rejecting $H_0: \gamma_1 = 0$ Using the t-test over the Entire Period and the Joint t-test over Two Subperiods

The table presents the probability of rejecting $H_0: \gamma_1 = 0$ in 10,000 simulations using a two-tailed t-test over the entire period and a joint t-test over two subperiods at various significance levels. In each simulation, 330 observations of a useless factor are randomly drawn from $N(0,1)$ and 330 observations of returns on N assets are independently drawn from $N(\mu, V)$, where μ and V are chosen based on the sample estimates over the period July 1963 to December 1990. The parameters are estimated from 10 size portfolios for $N = 10$, and from 100 size-beta portfolios for $N = 100$. The simulated monthly returns, R_t, are regressed on their betas (b) estimated with respect to the useless factor:

$$R_t = \gamma_0 1_N + \gamma_1 b + \varepsilon_t,$$

using ordinary least squares (OLS), true generalized least squares (GLS) (using true V) and estimated GLS (using $\hat{V}$ estimated from the simulated returns) with the entire period and two subperiods. Let α be the significance level for the test. Denote $t_n(\alpha)$ as the upper 100α percentage points of the central t-distribution with n degrees of freedom. Let t_1 and t_2 be the t-ratios of two subperiods. The hypothesis $H_0: \gamma_1 = 0$ is rejected by the joint t-test if and only if

$$t_1 > t_{\frac{T}{2}-1}\left(\sqrt{\frac{\alpha}{2}}\right) \quad \text{and} \quad t_2 > t_{\frac{T}{2}-1}\left(\sqrt{\frac{\alpha}{2}}\right)$$

or

$$t_1 < -t_{\frac{T}{2}-1}\left(\sqrt{\frac{\alpha}{2}}\right) \quad \text{and} \quad t_2 < -t_{\frac{T}{2}-1}\left(\sqrt{\frac{\alpha}{2}}\right)$$

where T is the number of time series observations.

	Probability of Rejecting $H_0: \gamma_1 = 0$							
	$N = 10$				$N = 100$			
	True β		Useless b		True β		Useless b	
Significance Level	Full	Joint	Full	Joint	Full	Joint	Full	Joint
OLS								
0.01	0.256	0.206	0.188	0.074	0.949	0.873	0.111	0.038
0.05	0.487	0.414	0.386	0.165	0.988	0.957	0.236	0.103
0.10	0.615	0.531	0.502	0.234	0.996	0.975	0.320	0.162
GLS(V)								
0.01	0.886	0.795	0.107	0.035	1.000	1.000	0.061	0.021
0.05	0.966	0.910	0.231	0.102	1.000	1.000	0.154	0.074
0.10	0.982	0.949	0.318	0.166	1.000	1.000	0.238	0.129
GLS($\hat{V}$)								
0.01	0.596	0.778	0.114	0.038	1.000	1.000	0.172	0.046
0.05	0.815	0.906	0.242	0.108	1.000	1.000	0.297	0.121
0.10	0.897	0.945	0.332	0.169	1.000	1.000	0.384	0.182

dence suggests that the t-test rejects the zero risk premium for a useless factor with a probability more than twice the size of the test for a typical length of time series used in empirical studies. The problem is exacerbated when the number of time series observations increases. This type of misspecification bias may provide misleading results.

Since the two-pass methodology does have some merits that other testing methodologies do not possess,[13] and some of the factors used in empirical studies could be useless, the relevant question here is how one can detect the problem. The diagnostics we suggest are as follows.

1. The hypothesis that all the betas with respect to a factor are zero should be tested before the second-pass CSR is run.
2. In the second-pass CSR, the OLS R^2 can be used as a measure of goodness of fit, but to test whether the factor is useless, simulations are needed to find its distribution. The GLS R^2 can be used to detect useless factors because its distribution is readily available, but GLS R^2 is inappropriate as a goodness-of-fit measure because it applies to transformed data (as reasoned in Jagannathan and Wang (1996)).
3. Shanken's EIV adjustment can be used to reduce the overrejection rates for useless factors when the returns are conditionally homoskedastic. But in the presence of conditional heteroskedasticity in returns, the EIV adjustment developed in Jagannathan and Wang (1997) should be used.
4. As a trade-off between detecting useless factors and maintaining the power of the test, a subperiod joint test can be performed. A more effective way, when possible, is to use another independent sample to examine the significance of the risk premium associated with a proposed factor.

These suggested diagnostic methods are not perfect, and they should be combined, contrasted, and used with care.[14]

As opposed to the traditional treatment, which assumes the proposed model is the correct model, we assume the proposed factor is useless in this paper. Both assumptions are extreme cases, and therefore both are unlikely to be true. In practice, probably all models suffer from some sort of misspecification and probably all proposed factors are not strictly useless. Between the

[13] The methodologies developed by Gibbons (1982) and Gibbons, Ross, and Shanken (1989) are one-pass regressions, but they can only be applied to the cases where only asset returns are used as factors.

[14] In an earlier version of this paper, we question whether the growth rate of labor income used in Jagannathan and Wang (1996) is a useless factor. Although we cannot reject the hypothesis that the labor betas are jointly equal to zero, simulation evidence of various test statistics as well as evidence from Japan (Jagannathan, Kubota, and Takehara (1997)) suggest that it is inappropriate to claim that the growth rate of labor income is a useless factor simply based on the insignificance of the labor betas. Moreover, the labor beta of the value-weighted market portfolio is negative and statistically significant, which further indicates that the growth rate of labor income is unlikely to be a useless factor.

true factor and useless factors, there are many misspecified models. Existing empirical asset pricing models probably all fall into this category of misspecified models. Testing whether these models are right or wrong is not very interesting by itself; a more interesting and challenging question is how we compare the performance of these models. Hansen and Jagannathan (1997) address this question from a certain perspective, and we hope future research will continue to address this important question.

Appendix

Proof of Proposition 1: Conditioned on b, and assuming $\hat{\gamma}_{1t}^{OLS}$ and $\hat{\gamma}_{1t}^{GLS}$ are i.i.d. normal, the OLS and GLS t-ratios for testing $H_0: \gamma_1 = 0$ are given by

$$t_{OLS}(b) = \frac{\bar{\hat{\gamma}}_1^{OLS}(b)}{s(\hat{\gamma}_1^{OLS}(b))/\sqrt{T}} \tag{A1}$$

and

$$t_{GLS}(b) = \frac{\bar{\hat{\gamma}}_1^{GLS}(b)}{s(\hat{\gamma}_1^{GLS}(b))/\sqrt{T}}. \tag{A2}$$

Using equations (16) and (17), it is easy to verify their noncentrality parameters. Note that if $\mu = k1_N$ for some scalar k, then $\delta_{OLS}(b) = \delta_{GLS}(b) = 0$ for every realization of b and both the OLS and GLS t-tests are properly specified. If $\mu \neq k1_N$ for any scalar k, then $\delta_{OLS}(b) = 0$ if and only if $b'M\mu = 0$ and $\delta_{GLS}(b) = 0$ if and only if $\tilde{b}'\widetilde{M}\tilde{\mu} = 0$. If b has a continuous distribution, then both sets of b have measure zero.

Unconditionally, $t_{OLS}(b)$ has a compound noncentral t-distribution that depends on the distribution of $\delta_{OLS}(b)$. Its expected value and variance are given by

$$E[t_{OLS}(b)] = \left(\frac{T-1}{2}\right)^{1/2} \frac{\Gamma\left(\frac{T-2}{2}\right)}{\Gamma\left(\frac{T-1}{2}\right)} E[\delta_{OLS}(b)], \tag{A3}$$

$$\begin{aligned}\mathrm{Var}[t_{OLS}(b)] = {} & \left(\frac{T-1}{T-3}\right) + \left[\left(\frac{T-1}{T-3}\right) - \left(\frac{T-1}{2}\right)\frac{\Gamma\left(\frac{T-2}{2}\right)^2}{\Gamma\left(\frac{T-1}{2}\right)^2}\right] E[\delta_{OLS}^2(b)] \\ & + \left[\left(\frac{T-1}{2}\right)\frac{\Gamma\left(\frac{T-2}{2}\right)^2}{\Gamma\left(\frac{T-1}{2}\right)^2}\right] \mathrm{Var}[\delta_{OLS}(b)].\end{aligned} \tag{A4}$$

Since $\delta_{OLS}(b)$ is an odd function of b and the normal density function of b is an even function of b, it follows that $E[\delta_{OLS}(b)] = 0$ and $\mathrm{Var}[\delta_{OLS}(b)] = E[\delta^2_{OLS}(b)]$, and we have

$$E[t_{OLS}(b)] = 0, \tag{A5}$$

$$\mathrm{Var}[t_{OLS}(b)] = \left(\frac{T-1}{T-3}\right) + \left(\frac{T-1}{T-3}\right)E[\delta^2_{OLS}(b)]. \tag{A6}$$

Therefore, although the t-ratio has an unconditional mean of zero, its variance is higher than that of the central t-distribution. The unconditional expected value and variance of $t_{GLS}(b)$ are similarly obtained by replacing $\delta_{OLS}(b)$ by $\delta_{GLS}(b)$ in (A3) and (A4). Q.E.D.

Proof of Proposition 2: For $E[\delta^2_{OLS}(b)]$, define $Z = \sqrt{s_{gg}}H'V^{-1/2}b \sim N(0, I_{N-1})$ and we can write

$$E[\delta^2_{OLS}(b)] = E\left[\frac{T(Z'\Lambda\eta)^2}{Z'\Lambda^2 Z}\right] = T\sum_{i=1}^{N-1} E\left[\frac{\lambda_i^2 Z_i^2}{\sum_{j=1}^{N-1}\lambda_j^2 Z_j^2}\right]\eta_i^2. \tag{A7}$$

The off-diagonal elements do not matter because, by symmetry, $E[Z_i Z_j/(Z'\Lambda^2 Z)]$ vanishes when $i \neq j$. For analytical expression of $E[\delta^2_{OLS}]$, we apply the results of Sawa (1978) and obtain

$$E[\delta^2_{OLS}(b)] = T\int_0^\infty \frac{1}{\prod_{j=1}^{N-1}(1+2t\lambda_j^2)^{1/2}} \sum_{i=1}^{N-1}\frac{\lambda_i^2\eta_i^2}{(1+2t\lambda_i^2)}\,\mathrm{d}t. \tag{A8}$$

The numerical integration of this expression can be facilitated by a change of variable with $u = 1/(1+2t\lambda_1^2)$ and the integral can be evaluated over u from 0 to 1.

Similarly, we can write

$$\delta^2_{OLS}(\beta) = \frac{T(\mu'M\mu)^2}{\mu'MVM\mu} = \frac{T(\eta'\Lambda\,\eta)^2}{\eta'\Lambda^2\,\eta} = \frac{T\left(\sum_{i=1}^{N-1}\lambda_i\eta_i^2\right)^2}{\sum_{i=1}^{N-1}\lambda_i^2\eta_i^2}. \tag{A9}$$

To prove the inequality, it suffices to show the following

$$
\begin{aligned}
&\left(\sum_{i=1}^{N-1} E\left[\frac{\lambda_i^2 Z_i^2}{\sum_{j=1}^{N-1} \lambda_j^2 Z_j^2}\right]\eta_i^2\right)\left(\sum_{i=1}^{N-1} \lambda_i^2 \eta_i^2\right)\\
&\quad \leq \left(\sum_{i=1}^{N-1} E\left[\frac{\lambda_i^2 Z_i^2}{\lambda_i^2 Z_i^2 + \lambda_{N-1}^2 Z_{N-1}^2}\right]\eta_i^2\right)\lambda_{N-1}\left(\sum_{i=1}^{N-1} \lambda_i \eta_i^2\right)\\
&\quad = \left(\sum_{i=1}^{N-1} \lambda_{N-1} E\left[\frac{Z_i^2}{Z_i^2 + \frac{\lambda_{N-1}^2}{\lambda_i^2} Z_{N-1}^2}\right]\eta_i^2\right)\left(\sum_{i=1}^{N-1} \lambda_i \eta_i^2\right)\\
&\quad = \left(\sum_{i=1}^{N-1} \left[\frac{\lambda_{N-1}}{1 + \frac{\lambda_{N-1}}{\lambda_i}}\right]\eta_i^2\right)\left(\sum_{i=1}^{N-1} \lambda_i \eta_i^2\right)\\
&\quad \leq \left(\sum_{i=1}^{N-1} \lambda_i \eta_i^2\right)^2. \qquad \text{(A10)}
\end{aligned}
$$

The last equality follows because for $c \geq 0$,

$$
E\left[\frac{Z_i^2}{Z_i^2 + c^2 Z_{N-1}^2}\right] = \frac{1}{1+c}. \qquad \text{(A11)}
$$

For $E[\delta_{GLS}^2]$, since $\widetilde{M}$ is idempotent, there exists an $N \times (N-1)$ orthonormal matrix Q (the columns of Q are simply the $N-1$ eigenvectors of $\widetilde{M}$ associated with the $N-1$ eigenvalues of 1) such that $Q'Q = I_{N-1}$ and $QQ' = \widetilde{M}$. Define $a = Q'\tilde{\mu}$ and $Z = \sqrt{s_{gg}}Q'\tilde{b}$, then we have $Z \sim N(0, I_{N-1})$ and

$$
\begin{aligned}
E[\delta_{GLS}^2(b)] &= T(\tilde{\mu}'\widetilde{M}\tilde{\mu})E\left[\frac{(\tilde{b}'\widetilde{M}\tilde{\mu})^2}{(\tilde{\mu}'\widetilde{M}\tilde{\mu})(\tilde{b}'\widetilde{M}\tilde{b})}\right]\\
&= \delta_{GLS}^2(\beta)E\left[\frac{(a'Z)^2}{(a'a)(Z'Z)}\right]\\
&= \delta_{GLS}^2(\beta)E\left[\frac{Z'BZ}{Z'Z}\right] \qquad \text{(A12)}
\end{aligned}
$$

by writing $B = (aa')/(a'a)$. It is easy to verify that B is an $(N-1) \times (N-1)$ symmetric idempotent matrix of rank 1. Since Z has a spherical distribution, using Theorem 1.5.7 in Muirhead (1982), $(Z'BZ)/(Z'Z)$ is distributed as $\text{Beta}(\frac{1}{2}, (N-2)/2)$ and its expected value is $1/(N-1)$. Q.E.D.

Proof of Proposition 3: That $R^2_{GLS}(b)$ follows a Beta$(\frac{1}{2},(N-2)/2)$ distribution when $b \sim N(0,V/s_{gg})$ follows directly from the proof of Muirhead (1982), Theorem 5.1.1, which states that in order for the squared sample correlation between two variables to follow the Beta distribution, only the observations of one variable need to be spherical. Define $U = Q'\tilde{\bar{R}}$ and $Z = \sqrt{s_{gg}}Q'\tilde{b}$, where Q is defined in the proof of Proposition 2, then we have

$$R^2_{GLS}(b) = \frac{(\tilde{\bar{R}}'\widetilde{M}\tilde{b})^2}{(\tilde{\bar{R}}'\widetilde{M}\tilde{\bar{R}})(\tilde{b}'\widetilde{M}\tilde{b})} = \frac{(U'Z)^2}{(U'U)(Z'Z)}. \tag{A13}$$

Since $Z \sim N(0,I_{N-1})$ and it has a spherical distribution, the proof of Theorem 5.1.1 in Muirhead (1982) goes through.

That $R^2_{OLS}(b)$ follows a Beta$(\frac{1}{2},(N-2)/2)$ distribution when $MVM = cM$ for some constant $c > 0$ can be shown as follows. Define H and Λ as in Proposition 2, then we premultiply and postmultiply $MVM = cM$ by $V^{1/2}$, and we have $\Lambda = cI_{N-1}$ and $V^{1/2}MV^{1/2} = cHH'$ or $H' = (1/c)H'V^{1/2}MV^{1/2}$. Since $H'\tilde{1}_N = 0$, we have $\widetilde{M} = HH'$. Therefore, $V^{-1/2}\widetilde{M}V^{-1/2} = V^{-1/2}HH'V^{-1/2} = (1/c)M$, and hence GLS R^2 and OLS R^2 are the same. Q.E.D.

Proof of Proposition 4: Since the case of GLS is almost identical to the case of OLS, we will only prove the case of OLS here. In the literature of probability theory (see, e.g., Amemiya (1985) or Davidson (1994)), the notation $z_T = O_p(T^a)$ for a sequence of random variables z_T means that, for any $\varepsilon > 0$, there exists an M_ε such that $P[|z_T/T^a| < M_\varepsilon] > 1 - \varepsilon$; that is, z_T is at most of order T^a. There is also a notation $z_T = o_p(T^a)$ that means plim $z_T/T^a = 0$; that is, z_T is of an order less than T^a. In the following we will use $z_T = O_p(T^a)$ in a narrower sense, that it is $O_p(T^a)$ but not $o_p(T^a)$; in other words, z_T is exactly of order T^a.

From large sample theory of regression analysis, $b \to 0$ with probability one and, according to the central limit theorem, the rate of convergence is $T^{-1/2}$. That is, $b = O_p(T^{-1/2})$.

As $T \to \infty$, if $\mu \neq k1_N$ for any scalar k, $M\bar{R} \to M\mu \neq 0$. Hence,

$$\bar{\hat{\gamma}}_1^{OLS} = \frac{1}{T}\sum_{t=1}^{T}\hat{\gamma}_{1t}^{OLS} = \frac{1}{T}\sum_{t=1}^{T}\frac{b'MR_t}{b'Mb} = \frac{b'M\bar{R}}{b'Mb} = O_p(T^{1/2}). \tag{A14}$$

That is, $\bar{\hat{\gamma}}_1^{OLS} \to \infty$ in probability at an order $\sqrt{T}$. Let

$$\hat{V} = \frac{1}{T-1}\sum_{t=1}^{T}(R_t - \bar{R})(R_t - \bar{R})'. \tag{A15}$$

Then $\hat{V} \to V$ with probability one, and

$$s^2(\hat{\gamma}_1^{OLS}) = \frac{1}{T-1}\sum_{t=1}^{T}(\hat{\gamma}_{1t}^{OLS} - \bar{\hat{\gamma}}_1^{OLS})^2 = \frac{b'M\hat{V}Mb}{(b'Mb)^2} = O_p(T). \tag{A16}$$

Therefore, both $s^2(\hat{\gamma}_1^{OLS})/T$ and $s(\hat{\gamma}_1^{OLS})/\sqrt{T}$ are $O_p(1)$. As a result,

$$t_{OLS}(b) = \frac{\bar{\hat{\gamma}}_1^{OLS}}{s(\hat{\gamma}_1^{OLS})/\sqrt{T}} = O_p(T^{1/2}). \tag{A17}$$

By definition, for any $M > 0$,

$$P[|t_{OLS}(b)| > M] \to 1, \qquad \text{as } T \to \infty. \tag{A18}$$

The probability of the hypothesis $H_0\colon \gamma_1 = 0$ to be rejected using $t_{OLS}(b)$ tends to one. Q.E.D.

Proof of Proposition 5: Since $H'V^{-1/2}\bar{R} \xrightarrow{P} \eta$ and $Z = \sqrt{s_{gg}}H'V^{-1/2}b \sim N(0, I_{N-1})$, by using the Cramer–Slutsky theorem (see Amemiya (1985, p. 89) or Davidson (1994, p. 355)),

$$R_{OLS}^2 = \frac{(\bar{R}'Mb)^2}{(\bar{R}'M\bar{R})(b'Mb)} \xrightarrow{D} \frac{h}{(\mu'M\mu)}. \tag{A19}$$

For $h = (Z'\Lambda\eta)^2/(Z'\Lambda Z)$ defined in (30), its distribution is bounded. To see this, note the lowest value for h is 0 when the realization of b is orthogonal to $M\mu$. The highest value for h is $\mu'M\mu$, using the result $R_{OLS}^2 \leq 1$, which happens when the realization of b is a linear function of 1_N and μ. Therefore h has a continuous distribution over $[0, \mu'M\mu]$ and $\text{Var}[h] > 0$. The proof for R_{GLS}^2 follows directly from Proposition 3. Q.E.D.

Proof of Proposition 6: t_{OLS}^{*2} can be written as

$$t_{OLS}^{*2} = \frac{(\bar{\hat{\gamma}}_1^{OLS})^2}{\dfrac{s^2(\hat{\gamma}_1^{OLS})}{T} + \left(\dfrac{\bar{\hat{\gamma}}_1^{OLS}}{\hat{\sigma}_g}\right)^2\left[\dfrac{s^2(\hat{\gamma}_1^{OLS})}{T} - \dfrac{\hat{\sigma}_g^2}{T}\right]}. \tag{A20}$$

Since $s^2(\hat{\gamma}_1^{OLS})/T = O_p(1)$ and $\bar{\hat{\gamma}}_1^{OLS} = O_p(T^{1/2})$, it follows that

$$t_{OLS}^{*2} - \frac{s_{gg}}{s^2(\hat{\gamma}_1^{OLS})} \xrightarrow{P} 0. \tag{A21}$$

From the Cramer–Slutsky theorem, the limiting distribution of t_{OLS}^{*2} is the same as that of

$$\frac{s_{gg}}{s^2(\hat{\gamma}_1^{OLS})} = \frac{s_{gg}(b'Mb)^2}{b'M\hat{V}Mb}, \tag{A22}$$

which, in turn, has the same limiting distribution as $s_{gg}(b'Mb)^2/(b'MVMb)$ since $\hat{V} \to V$. Define $Z = \sqrt{s_{gg}}H'V^{-1/2}b \sim N(0,I_{N-1})$, where H is defined in Proposition 2, we have

$$\frac{s_{gg}(b'Mb)^2}{b'MVMb} = \frac{(Z'\Lambda Z)^2}{Z'\Lambda^2 Z} = \frac{\left(\sum_{i=1}^{N-1} \lambda_i Z_i^2\right)^2}{\sum_{i=1}^{N-1} \lambda_i^2 Z_i^2}. \tag{A23}$$

To show the first inequality, we note that

$$\left(\sum_{i=1}^{N-1} \lambda_i Z_i^2\right)^2 > \lambda_{N-1} Z_{N-1}^2 \left(\sum_{i=1}^{N-1} \lambda_i Z_i^2\right) \geq Z_{N-1}^2 \left(\sum_{i=1}^{N-1} \lambda_i^2 Z_i^2\right). \tag{A24}$$

Therefore $\lim t_{OLS}^{*2} > Z_{N-1}^2$, or $F_{\chi_1^2}(u) > F_{\lim t_{OLS}^{*2}}(u)$. To show the other inequality, note from the Cauchy–Schwarz inequality,

$$\left[\sum_{i=1}^{N-1} (\lambda_i Z_i) Z_i\right]^2 \leq \sum_{i=1}^{N-1} (\lambda_i Z_i)^2 \sum_{i=1}^{N-1} Z_i^2. \tag{A25}$$

So that $\lim t_{OLS}^{*2} \leq \sum_{i=1}^{N-1} Z_i^2$, or $F_{\lim t_{OLS}^{*2}}(u) \geq F_{\chi_{N-1}^2}(u)$.

For GLS, the limiting distribution of t_{GLS}^{*2} is simply $s_{gg}(\tilde{b}'\widetilde{M}\tilde{b})$. Define $Z = \sqrt{s_{gg}}Q'\tilde{b}$, where Q is defined in the proof of Proposition 2, then we have $Z \sim N(0,I_{N-1})$ and

$$s_{gg}(\tilde{b}'\widetilde{M}\tilde{b}) = Z'Z, \tag{A26}$$

which is a χ_{N-1}^2 random variable. Q.E.D.

REFERENCES

Amemiya, Takeshi, 1985, *Advanced Econometrics* (Harvard University Press, Boston, Massachusetts).

Amihud, Yakov, Bent Jesper Christensen, and Haim Mendelson, 1992, Further evidence on the risk-return relationship, Working paper, New York University.

Black, Fischer, 1972, Capital market equilibrium with restricted borrowing, *Journal of Business* 45, 444–454.

Black, Fischer, Michael C. Jensen, and Myron Scholes, 1972, The capital asset pricing model: Some empirical findings; in Michael C. Jensen, ed.: *Studies in the Theory of Capital Markets* (Praeger, New York).

Chamberlain, Gary, and Michael Rothschild, 1983, Arbitrage, factor structure and mean-variance analysis on large asset markets, *Econometrica* 51, 1281–1304.

Chen, Naifu, Richard R. Roll, and Stephen A. Ross, 1986, Economic forces and the stock market, *Journal of Business* 59, 383–403.

Davidson, James, 1994, *Stochastic Limit Theory: An Introduction to Econometricians* (Oxford University Press, New York).

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

Fama, Eugene F., and James D. MacBeth, 1973, Risk, return and equilibrium: Empirical tests, *Journal of Political Economy* 81, 607–636.

Ferson, Wayne E., and Campbell Harvey, 1993, The risk and predictability of international equity returns, *Review of Financial Studies* 6, 527–566.

Gibbons, Michael R., 1982, Multivariate tests of financial models: A new approach, *Journal of Financial Economics* 10, 3–28.

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

Grauer, Robert R., 1994, Tests of the capital asset pricing model based on the cross-section of expected returns, Working paper, Simon Fraser University.

Hansen, Lars Peter, and Ravi Jagannathan, 1997, Assessing specification errors in stochastic discount factor models, *Journal of Finance* 52, 557–590.

Jagannathan, Ravi, Keiichi Kubota, and Hitoshi Takehara, 1997, Relationship between labor-income risk and average return: Empirical evidence from the Japanese stock market, Discussion paper 117, Institute for Empirical Macroeconomics, Federal Reserve Bank of Minneapolis.

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

Jagannathan, Ravi, and Zhenyu Wang, 1998, An asymptotic theory for estimating beta-pricing models using cross-sectional regression, *Journal of Finance* 53, 1285–1309.

Johnson, Norman L., Samuel Kotz, and N. Balakrishnan, 1995, *Continuous Univariate Distributions, Vol. 2* (Wiley, New York).

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

Kim, Dongcheol, 1995, The errors in the variables problem in the cross-section of expected stock returns, *Journal of Finance* 50, 1605–1634.

Lintner, John, 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.

Merton, Robert C., 1973, An intertemporal capital asset pricing model, *Econometrica* 41, 867–887.

Miller, Merton H., and Myron Scholes, 1972, Rates of return in relation to risk: A reexamination of some recent findings; in Michael C. Jensen, ed.: *Studies in the Theory of Capital Markets* (Praeger, New York).

Muirhead, Robb J., 1982, *Aspects of Multivariate Statistical Theory* (Wiley, New York).

Roll, Richard, 1977, A critique of the asset pricing theory's tests; Part I: On past and potential testability of theory, *Journal of Financial Economics* 4, 129–176.

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

Ross, Stephen A., 1976, The arbitrage theory of capital asset pricing, *Journal of Economic Theory* 13, 341–360.

Sawa, Takamitsu, 1978, The exact moments of the least squares estimator for the autoregressive model, *Journal of Econometrics* 8, 159–172.

Shanken, Jay, 1992, On the estimation of beta-pricing models, *Review of Financial Studies* 5, 1–33.

Sharpe, William F., 1964, Capital asset prices: A theory of market equilibrium under conditions of risk, *Journal of Finance* 19, 425–442.