# A Note on the GRS Test

Mark J. Kamstra* Ruoyao Shi†

This version: July 2021

**Abstract**

We clear up an ambiguity in Gibbons, Ross and Shanken (1989, GRS hereafter) by providing the correct formula of the GRS test statistic and proving its exact F distribution in the general multiple portfolio case. We generalize the Sharpe ratio based interpretation of the GRS test to the multiple portfolio case, which we argue paradoxically makes experts in asset pricing studies more susceptible to an incorrect formula. We theoretically and empirically illustrate the consequences of using the incorrect formula – over-rejecting and mis-ranking asset pricing models.

**Keywords**: GRS test, asset pricing, CAPM, multivariate test, portfolio efficiency, Sharpe ratio, over-rejection, model ranking

## 1 Introduction

In an influential paper, Gibbons, Ross and Shanken (1989, GRS hereafter) developed and analyzed a test for the ex ante mean-variance efficiency of portfolios. For the single portfolio case, they carefully developed the test statistic in a linear regression model, derived its finite-sample exact F distribution, investigated its power properties, and highlighted its significance in asset pricing theory by purveying an alternative

*Schulich School of Business, York University, mkamstra@yorku.ca. This research was funded in part by by the Social Sciences and Humanities Research Council of Canada.

†Department of Economics, University of California Riverside, ruoyao.shi@ucr.edu. The authors contributed equally and are in the alphabetical order. We are grateful for the helpful suggestions of Lisa Kramer, Cheng Chou, Diane Del Guercio, Jinyong Hahn, Zhipeng Liao, Lilian Ng, Raymond Kan and the research assistance of Mengnan (Cliff) Zhu.

interpretation involving the Sharpe ratio (Sharpe, 1966) – the excess return to a portfolio per unit of risk (or volatility, measured by standard deviation) – which is a key measure of portfolio efficiency. For the multiple portfolio case, however, GRS (1989, Section 7) were ambiguous on how the test statistic should be constructed.

In terms of asset pricing theory, the solution to the portfolio optimization problem that yields the Sharpe ratio[1] naturally requires the estimation of the variance-covariance matrix of the portfolio excess returns. As for statistics, the derivation of the finite-sample exact distribution of the GRS statistic exploits a projection matrix that, while resembling the variance-covariance matrix, is not. This equivocality of the "variance-covariance matrix" has apparently caused a confusion of the function of the GRS statistic, which is further exacerbated by the fact that the F distribution wrongfully conjures up a degrees of freedom adjustment (d.f. hereafter) that is improper in this case. This has led to the application of a very common incorrect formula that, paradoxically, is more likely to be used by financial economists, the experts in the field, than by someone who focuses only on the statistical aspect of the problem. We find that using the incorrect formula leads to: (i) a test statistic that does not follow the F distribution as prescribed and over-rejects the null hypothesis of portfolio efficiency; (ii) smaller models often being favored over larger ones when the statistic is used to rank asset pricing models. This error comes from mixing terms that fall out of portfolio optimization with a statistical object that comes from the finite-sample exact test derivation.

The purpose of this note is to clear up the ambiguity in the statement of the GRS test and in doing so, to highlight the error commonly made and its implications (both theoretical and empirical) for applied work. Section 2.1 investigates the statistical aspect of the GRS test by deriving the correct GRS test statistic along with its finite-sample exact distribution for the general multiple portfolio case, as well as exploring the consequences of using the incorrect formula. Section 2.2 highlights the interpretation of the GRS test using the Sharpe ratio and spells out the error often made by financial specialists in the context of the portfolio optimization problem. Section 3 concludes with some important final remarks. We will adopt the notation in GRS (1989, Section 7) whenever possible.

[1] Minimization of the volatility of a portfolio return subject to a portfolio mean return constraint, to be detailed in Section 2.2.

# 2 Testing the Efficiency of Portfolios of Assets

The problem is to test the mean-variance efficiency of $L$ portfolios, utilizing another $N$ assets (known as test assets). Let $\tilde{r}_{jt}$ denote the excess return on portfolio $j$ in period $t$, let $\tilde{r}_{pt} \equiv (\tilde{r}_{1t}, \ldots, \tilde{r}_{Lt})'$, and let $\tilde{r}_{it}$ denote the excess return on test asset $i$ in period $t$ ($j = 1, \ldots, L$, $i = 1, \ldots, N$ and $t = 1, \ldots, T$).

The mean-variance efficiency of the $L$ portfolios has two implications that are instructive for coming up with tests of this efficiency. For the first implication, consider the linear regression model (eq. (17) of GRS, 1989):

$$\tilde{r}_{it} = \delta_{i0} + \delta_i' \tilde{r}_{pt} + \tilde{\eta}_{it}, \quad \forall i = 1, \ldots, N, \tag{1}$$

where the $L$ portfolios serve as factors and $\tilde{\eta}_{it}$ denotes the disturbance term. If the $L$ portfolios are efficient, then the intercepts in model (1) will be zero; that is

$$H_0: \ \delta_{i0} = 0, \quad \forall i = 1, \ldots, N. \tag{2}$$

See for instance Sharpe (1964) and Litner (1965). A more interesting implication, however, is that if the $L$ portfolios are efficient, then the optimal portfolio consisting of the $L$ portfolios and the $N$ test assets will have the same Sharpe ratio as that consisting of the $L$ portfolios only (GRS, 1989).

These two implications are well known in the literature. In the following two subsections, we will elaborate on what they indicate for tests of portfolio efficiency for the general multiple portfolio case. This side-by-side comparison helps shed light on why the mistake is made so often.

## 2.1 GRS Test for Multiple Portfolios

To elaborate on the first implication, it helps to take a purely statistical approach to the linear regression model (1) and the hypothesis (2), abstracting any economic interpretation of them. We first provide the GRS test statistic for the general $L \geq 1$ case along with its finite-sample exact distribution.

**Lemma 1** (Joint F test)**.** *Let $\tilde{r}_p \equiv [\tilde{r}_{p1}, \ldots, \tilde{r}_{pT}]'$, $\bar{r}_p \equiv T^{-1} \sum_{t=1}^{T} \tilde{r}_{pt}$, and let $\hat{\delta}_0$ be the least squares estimator of $\delta_0 \equiv (\delta_{10}, \ldots, \delta_{N0})'$; also let $\hat{\eta}_t \equiv (\hat{\eta}_{1t}, \ldots, \hat{\eta}_{Nt})'$ with $\hat{\eta}_{it}$ being the least squares residuals of model (1). We follow GRS (1989) to assume that the disturbance $\tilde{\eta}_t \equiv (\tilde{\eta}_{1t}, \ldots, \tilde{\eta}_{Nt})'$ has a joint normal distribution with mean zero and* *nonsingular variance-covariance matrix* $\Sigma$ *and is iid over* $t$.[2] *Define*

$$\breve{\Omega} \equiv \frac{1}{T}\sum_{t=1}^{T} \tilde{r}_{pt}\tilde{r}_{pt}', \tag{3}$$

$$\widehat{\Sigma} \equiv \frac{1}{T-L-1}\sum_{t=1}^{T} \hat{\eta}_t\hat{\eta}_t'. \tag{4}$$

*Then the statistic defined as*

$$\widetilde{W}^* \equiv \frac{T(T-N-L)}{N(T-L-1)}\left(1-\bar{r}_p'\breve{\Omega}^{-1}\bar{r}_p\right)\hat{\delta}_0'\widehat{\Sigma}^{-1}\hat{\delta}_0 \tag{5}$$

*follows the* $F_{N,T-N-L}$ *distribution under the* $H_0$.

From a purely statistical point of view, Lemma 1 is all that one needs to test the hypothesis (2), which is the usual joint F test of zero intercepts in a linear regression model, and its proof is standard (in Online Appendix A). Note that in this case, only $\Sigma$, the variance-covariance matrix of the **disturbances** $\tilde{\eta}_{it}$, needs to be estimated (by $\widehat{\Sigma}$ in eq. (4)). It is also obvious in the proof (eq. (A.3) and (A.5) in Online Appendix A) that $\breve{\Omega}$ naturally arises in the **projection** of $\tilde{r}_{it}$ onto the column space of $\tilde{r}_p$, the design matrix of model (1). The following theorem, however, gives the generalized GRS test statistic that is equivalent to $\widetilde{W}^*$ in Lemma 1, generalizes the original GRS test when $L = 1$, and is easier to interpret from the portfolio optimization point of view (eq. (11) below and the discussion that follows).

**Theorem 1** (Generalized GRS test)**.** *Suppose all the conditions of Lemma 1 are satisfied. Define*

$$\widetilde{\Omega} \equiv \frac{1}{T}\sum_{t=1}^{T}\left(\tilde{r}_{pt}-\bar{r}_p\right)\left(\tilde{r}_{pt}-\bar{r}_p\right)' = \frac{1}{T}\sum_{t=1}^{T}\tilde{r}_{pt}\tilde{r}_{pt}' - \bar{r}_p\bar{r}_p' \tag{6}$$

*and the generalized GRS test statistic*

$$\widetilde{W} \equiv \frac{T(T-N-L)}{N(T-L-1)}\left(1+\bar{r}_p'\widetilde{\Omega}^{-1}\bar{r}_p\right)^{-1}\hat{\delta}_0'\widehat{\Sigma}^{-1}\hat{\delta}_0. \tag{7}$$

[2] We acknowledge that it is difficult to take seriously the assumption of normality of returns – returns are bounded below by -100% due to limited liability in financial markets for publicly traded assets and returns are known to be heteroskedastic and dependent over time. Here we adopt the GRS (1989) setting for comparison purpose.

*Then $\widetilde{W} = \widetilde{W}^*$, and hence $\widetilde{W}$ follows the $F_{N,T-N-L}$ distribution under the $H_0$.*

**Remark 1** (Original GRS test when $L = 1$). *When $L = 1$, $\widetilde{\Omega}$ equals to $\frac{1}{T}\sum_{t=1}^{T} \tilde{r}_{pt}^2 - \bar{r}_p^2 = s_p^2$, the sample variance of $\tilde{r}_{pt}$* ***without*** *d.f. defined by GRS (1989, p.1124). It then immediately follows that $\widetilde{W}$ in eq. (7) is the same as the original GRS test statistic when $L = 1$ (GRS, 1989, p.1124).*

Although $\widetilde{\Omega}$ in eq. (6) is reminiscent of the maximum likelihood estimator (MLE hereafter) of the variance-covariance matrix $\Omega$ of the $L$ portfolio excess returns $\tilde{r}_{pt}$, it is a conceptually different object. The use of the Sherman-Morrison formula in the proof of Theorem 1 (in Online Appendix A) shows that $\widetilde{\Omega}$ is simply a result of rewriting $\breve{\Omega}$ (which in turn naturally arises in the projection of $\tilde{r}_{it}$ onto the column space of $\tilde{r}_p$), instead of an estimator of $\Omega$, so d.f. is not an admissible notion here.[3]

For the $L > 1$ case, GRS (1989, p.1146) gave a test statistic and its finite-sample exact F distribution under the $H_0$:

$$\widehat{W} \equiv \frac{T(T-N-L)}{N(T-L-1)} \left(1 + \bar{r}_p' \widehat{\Omega}^{-1} \bar{r}_p\right)^{-1} \hat{\delta}_0' \widehat{\Sigma}^{-1} \hat{\delta}_0 \sim F_{N,T-N-L}, \tag{8}$$

where they prescribed $\widehat{\Omega}$ to be the **sample** variance-covariance matrix of $\tilde{r}_{pt}$.[4] Since the sample variance-covariance matrix customarily entails d.f., i.e.,

$$\widehat{\Omega} \equiv \frac{1}{T-1} \sum_{t=1}^{T} \left(\tilde{r}_{pt} - \bar{r}_p\right) \left(\tilde{r}_{pt} - \bar{r}_p\right)' = \frac{T}{T-1} \widetilde{\Omega}, \tag{9}$$

the statistic $\widehat{W}$ suggested by GRS (1989) differs from our $\widetilde{W}$, and Theorem 1 therefore immediately implies that $\widehat{W}$ does not follow the $F_{N,T-N-L}$ distribution as prescribed.

Although GRS (1989) were careful to point out that $\widetilde{\Omega}$ (without d.f.) should be used for the $L = 1$ case and Shanken (1986) clearly stated the same for the $L > 1$ case in a separate paper, their ambiguity and choice of words for the $L > 1$ case in GRS (1989) made readers particularly susceptible to incorrectly using $\widehat{\Omega}$ (with d.f.) instead, which leads to a test statistic $\widehat{W}$ that does not follow the prescribed $F_{N,T-N-L}$ distribution. For example, popular packages in R and Stata use $\widehat{\Omega}$ when computing the GRS test statistic.

[3] The equivalence between $\widetilde{W}^*$ and $\widetilde{W}$ is obvious for the $L = 1$ case, but requires the Sherman-Morrison formula to show for the $L > 1$ case.

[4] And everything else in eq. (8) is as defined in Lemma 1.

We discover two major consequences of using the incorrect GRS test statistic in empirical asset pricing studies. First, the incorrect GRS test statistic $\widehat{W}$ over-rejects the null hypothesis of portfolio efficiency when gauged against the $F_{N,T-N-L}$ distribution. Our simulation experiments (detailed in Online Appendix B and results reported in Table B.1) show that the over-rejection can be large if the number of test assets ($N$) is large even when the sample period ($T$) is long – we see substantial over-rejection of the null hypothesis with 25 test assets and five factors for $T = 400$ calibrated to typical monthly data. The following remark provides the justification for our empirical findings.

**Remark 2** (Over-rejection)**.** *Take the ratio between $\widehat{W}$ in eq. (8) and $\widetilde{W}$ in eq. (7), then by the relationship between $\widehat{\Omega}$ and $\widetilde{\Omega}$ in eq. (9) we get the ratio*

$$\frac{\widehat{W}}{\widetilde{W}} = \frac{1 + \bar{r}_p' \widetilde{\Omega}^{-1} \bar{r}_p}{1 + \frac{T-1}{T} \bar{r}_p' \widetilde{\Omega}^{-1} \bar{r}_p}, \tag{10}$$

*which measures how much the incorrect formula inflates the GRS test statistic. Define a function $g(x) = \frac{1+x}{1+\frac{T-1}{T}x}$. Since the first order derivative of this function is $g'(x) = \frac{1/T}{\left(1+\frac{T-1}{T}x\right)^2} > 0$, we know that $g(x)$ is a monotonically increasing function of $x$. This, combined with the facts that $g(0) = 1$ and $\bar{r}_p' \widetilde{\Omega}^{-1} \bar{r}_p \geq 0$, implies that $\widehat{W}/\widetilde{W} \geq 1$. As a result, when $\widehat{W}$ is gauged against the $F_{N,T-N-L}$, the distribution of $\widetilde{W}$, it will over-reject the null hypothesis.*

In addition, the significance of the GRS test statistic in recent financial studies, as Fama and French (2015) advocate, resides in the ranking of competing asset pricing models rather than testing them, since, as Fama and French (2015) remark, all models are merely approximations of the asset pricing mechanism and will be rejected by a sufficiently powerful test given enough data. Models with smaller GRS test statistics are favored as they are regarded as fitting the data better. Using the data borrowed from Fama and French (2015, 2016), we show that the incorrect GRS test statistic $\widehat{W}$ can easily flip the ranking between the single factor CAPM and four/five-factor Fama-French models (detailed in Online Appendix C and results reported in Table C.1). The next remark provides some intuition for our empirical findings.

**Remark 3** (Model ranking)**.** *Some back-of-the-envelop calculation suggests that $\bar{r}_p' \widetilde{\Omega}^{-1} \bar{r}_p$ tends to be larger for models with more factors. To see this, let $\mu_{\tilde{r}_p}$ denote the mean vector of $\tilde{r}_{pt}$, then by the central limit theorem, we have $\sqrt{T}(\bar{r}_p - \mu_{\tilde{r}_p}) \xrightarrow{d.} \mathcal{N}(0, \Omega)$;*

*and by the law of large numbers, we have* $\widetilde{\Omega} \xrightarrow{p.} \Omega$. *These two results imply that* $T(\bar{r}_p - \mu_{\tilde{r}_p})'\widetilde{\Omega}^{-1}(\bar{r}_p - \mu_{\tilde{r}_p}) \xrightarrow{d.} \chi^2_L$. *Note that* $E(\chi^2_L) = L$, *so this in turn implies that for fixed* $T$, *the mean of* $\bar{r}_p'\widetilde{\Omega}^{-1}\bar{r}_p$ *is approximately* $E(\bar{r}_p'\widetilde{\Omega}^{-1}\bar{r}_p) \approx \frac{L}{T} + 2E(\bar{r}_p'\Omega^{-1}\mu_{\tilde{r}_p}) - \mu_{\tilde{r}_p}'\Omega^{-1}\mu_{\tilde{r}_p} = \frac{L}{T} + \mu_{\tilde{r}_p}'\Omega^{-1}\mu_{\tilde{r}_p}$, *where* $\mu_{\tilde{r}_p}'\Omega^{-1}\mu_{\tilde{r}_p}$ *is expected to increase with* $L$ *since the dimensions of both* $\mu_{\tilde{r}_p}$ *and* $\Omega$ *increase with* $L$. *As a result, the random variable* $\bar{r}_p'\widetilde{\Omega}^{-1}\bar{r}_p$ *tends to increase with* $L$ ***on average***.[5] *Combined with Remark 2, this means that the ratio in eq. (10) tends to be larger for larger models; that is, smaller models tend to be disproportionally favored if the incorrect GRS test statistic* $\widehat{W}$ *is used to rank models, compared to the ranking based on the correct GRS statistic* $\widetilde{W}$.

**Remark 4** (Model ranking using p-value)**.** *Beyond this sensitivity of model ranking to the d.f. calculation, we note that the use of the raw GRS statistic is subject to a familiar critique akin to that of the use of* $R^2$ *for model selection. In this context, we suggest using the p-value of the statistic* $\widetilde{W}$ *from the F distribution.*

## 2.2 Interpretation of the GRS Test Using the Sharpe Ratio

To elaborate on the second implication, that if the $L$ portfolios are efficient then the optimal portfolio consisting of the $L$ portfolios and the $N$ test assets will have the same Sharpe ratio as that consisting of the $L$ portfolios only, we first consider a general portfolio optimization problem that yields the Sharpe ratio. Let $\tilde{r}$ denote a vector of excess returns of $K$ assets ($K \geq 1$), and let $\mu_{\tilde{r}}$ and $\Omega_{\tilde{r}}$ be their ex ante mean vector and variance-covariance matrix, respectively. Let $m$ be the target mean excess return and $\omega$ be a vector of $K$ asset weights. The optimal portfolio weights $\omega^*$ solve

$$\min_{\omega} \omega'\Omega_{\tilde{r}}\omega, \quad \text{subject to} \quad \omega'\mu_{\tilde{r}} = m.$$

Recall that the Sharpe ratio captures the mean excess return to a portfolio per unit of volatility (standard deviation), so the squared Sharpe ratio of the optimal portfolio composed of these $K$ assets is

$$\theta^{*2} \equiv \left(\frac{m}{\sqrt{\omega^{*\prime}\Omega_{\tilde{r}}\omega^*}}\right)^2 = \mu_{\tilde{r}}'\Omega_{\tilde{r}}^{-1}\mu_{\tilde{r}},$$

[5] We need to point out that the argument here is based on an approximation, as $E(\bar{r}_p'\widetilde{\Omega}^{-1}\bar{r}_p)$ is a non-linear function of $\bar{r}_p$ and $\widetilde{\Omega}$. Moreover, $\bar{r}_p'\widetilde{\Omega}^{-1}\bar{r}_p$ may deviate from its mean for a particular sample. Therefore, it is entirely possible that the incorrect formula of the GRS test favors larger models in some cases.

in which the variance-covariance matrix $\Omega_{\tilde{r}}$ of the $K$ assets plays a central role.

Applying this general result separately to the Sharpe Ratio of the $L$ portfolios and the $N$ test assets, and to the Sharpe ratio of the $L$ portfolios alone (detailed in Online Appendix D), we get the following $W$, involving the squared Sharpe ratio $\theta^{*2}_{N+L}$ of both the $L$ portfolios and the $N$ test assets and the squared Sharpe ratio $\theta^{*2}_{p}$ of the $L$ portfolios only,

$$W \equiv \left(\frac{\sqrt{1+\theta^{*2}_{N+L}}}{\sqrt{1+\theta^{*2}_{p}}}\right)^2 - 1 = \left(1+\mu'_{\tilde{r}_p}\Omega^{-1}\mu_{\tilde{r}_p}\right)^{-1}\delta_0'\Sigma^{-1}\delta_0. \tag{11}$$

Under the null hypothesis of portfolio efficiency, $W$ is zero, which generalizes eq. (7) in GRS (1989) to the $L > 1$ case.

Note that in the derivation of eq. (11) (in Online Appendix D), both $\Omega$ and $\Sigma$ are understood to be variance-covariance matrices (of the portfolio excess returns $\tilde{r}_{pt}$ and of the disturbance term $\tilde{\eta}_t$, respectively). Based on eq. (11), tests for the portfolio efficiency can be constructed as an ex post (sample) analog of $W$, by replacing each of its components with its estimator. It is natural to replace $\mu_{\tilde{r}_p}$ with $\bar{r}_p$, and $\delta_0$ with $\hat{\delta}_0$. When it comes to the variance-covariance matrices $\Omega$ and $\Sigma$, however, two simple estimators are equally commonly used – the unbiased estimators ($\widehat{\Omega}$ and $\widehat{\Sigma}$) and the MLE ($\widetilde{\Omega}$ and $\widetilde{\Sigma}$).[6] As far as the optimal portfolio theory is concerned, any combination of them, including both $\widetilde{W}$ and $\widehat{W}$, suffices to test whether the two Sharpe ratios are the same (i.e., whether $W = 0$), provided that the correct finite-sample exact distribution is used.[7] The tricky part is that since both $\Omega$ and $\Sigma$ are interpreted as variance-covariance matrices in the optimal portfolio theory, the **exact** $F_{N,T-N-L}$ distribution strongly conjures up a d.f. adjustment for **both** of them, which results in $\widehat{W}$ that does not follow the $F_{N,T-N-L}$ distribution, while in fact only the combination of $\widetilde{\Omega}$ (without d.f.) and $\widehat{\Sigma}$ (with d.f.) leads to $\widetilde{W}$, which follows the convenient $F_{N,T-N-L}$ distribution. This perhaps explains why financial economists, the experts in the field, are more likely to use the incorrect statistic $\widehat{W}$ than someone who focuses on the statistical problem rather than the portfolio optimization problem.

[6] $\widetilde{\Sigma}$ is not formally defined in this paper.

[7] Without using the Sherman-Morrison formula, the correspondence between $\widetilde{W}^*$ and $W$, on the other hand, is not straightforward.

# 3 Concluding Remarks

The GRS test statistic, as a sample analog of the solution to a portfolio optimization problem and used in a finite-sample F test, should from the perspective of a financial economist involve $\widehat{\Omega}$, the d.f. adjusted estimator of the variance-covariance matrix $\Omega$ of the portfolio excess returns. As a consequence, an applied researcher in finance is more likely to use the statistic $\widehat{W}$ which incorrectly implements a d.f. adjustment. Certainly in practice this appears to be the case.

Paradoxically, this error is clearly visible when turning a blind eye to the economic interpretation of the GRS test and taking a purely statistical approach. Even though $\widetilde{\Omega}$ is numerically the same as the usual MLE (without d.f.) of $\Omega$, $\widetilde{\Omega}$ naturally arises in the projection onto the column space of $\tilde{r}_p$, the design matrix of model (1), not as an estimator of the variance-covariance matrix $\Omega$. Hence a d.f. adjustment should not be applied.

Finally, if $T$ is large enough such that the difference between $\widehat{W}$ and $\widetilde{W}$ becomes negligible, then the $\chi^2_N$ distribution is adequate if all that is required is a joint test on the intercepts.[8] The $F_{N,T-N-L}$ distribution is, however, inherently pertinent to **exact** tests where one should make a point of computing the degrees of freedom correctly. Further, among all the potential constructions of the GRS test statistic, $\widetilde{W}$ is the only form with a known small-sample distribution, and the p-value of the F test is suited to the purposes of ranking models. For this reason we recommend the exact F test construction with its attendant $F_{N,T-N-L}$ distribution.

# References


Fama, E. F. and French, K. R., 2015. A Five-Factor Asset Pricing Model. *Journal of Financial Economics* 116(1), 1-22.

Fama, E. F. and French, K. R., 2016. Dissecting Anomalies With A Five-Factor Model. *The Review of Financial Studies* 29(1), 69-103.

Gibbons, M., Ross, S., and Shanken, J., 1989. A Test of the Efficiency of A Given Portfolio. *Econometrica* 57(5), 1121-1152.

Lintner, J., 1965. The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets. *The Review of Economics and Statistics* 47(1), 13-37.


[8] Both of the properly re-scaled test statistics $N\widehat{W}$ and $N\widetilde{W}$ have a limit $\chi^2_N$ distribution as $T \to \infty$.

Shanken, J., 1986. Testing Portfolio Efficiency when the Zero-Beta Rate is Unknown: A Note. *The Journal of Finance* 41(1), 269-276.

Sharpe, W., 1964. Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk. *Journal of Finance* 19, 425-42.

Sharpe, W. F., 1966. Mutual fund performance. *The Journal of Business* 39(1), 119-138.

# (NOT FOR PUBLICATION) ONLINE APPENDICES FOR "A NOTE ON THE GRS TEST"

Mark J. Kamstra[9] Ruoyao Shi[10]

This version: July 2021

## A Proofs

To prove Theorem 1 in the paper, we use the following lemmas.

**Lemma 2.** *If a random vector $Y$ and a random matrix $W$ satisfy: (i) $Y \sim \mathcal{N}_d(\mu, \Sigma)$, the $d$ dimensional normal distribution; (ii) $W \sim \mathcal{W}_d(f, \Sigma)$, the $d \times d$ dimensional Wishart distribution; and (iii) $Y \perp W$. Then given the Hotelling's T-squared defined as $T^2 \equiv f(Y-\mu)'W^{-1}(Y-\mu)$, we have $F \equiv \frac{f-d+1}{fd}T^2 \sim F_{d,f-d-1}$.*

**Lemma 3** (Sherman-Morrison formula)**.** *Suppose $A$ is an invertible $L \times L$ matrix and $u$ and $v$ are $L \times 1$ vectors. If $A + uv'$ is invertible, then $(A+uv')^{-1} = A^{-1} - \frac{A^{-1}uv'A^{-1}}{1+v'A^{-1}u}$.*

Lemma 2 is a standard result in multivariate statistics (see, e.g., Anderson, 2003, Theorem 5.2.2), and Lemma 3 is a standard result in linear algebra (see, e.g., Bartlett, 1951, pp. 107).

**Proof of Lemma 1.** The proof proceeds in three steps.

**Step 1.** In this step, we will show that under the null hypothesis (2),

$$\sqrt{T(1-\bar{r}_p'\breve{\Omega}^{-1}\bar{r}_p)}\hat{\delta}_0 \sim \mathcal{N}_N(0, \Sigma), \tag{A.1}$$

where $\breve{\Omega}$ is defined in eq. (3).

Let $\ell_T$ denote a $T \times 1$ vector with every element being one, and let $I_T$ denote the $T \times T$ identity matrix. Define $P_{p,T} = \tilde{r}_p\left(\tilde{r}_p'\tilde{r}_p\right)^{-1}\tilde{r}_p'$ as the $T \times T$ projection matrix (onto the column space of $\tilde{r}_p$) and its $T \times T$ complement matrix $Q_{p,T} = I_T - P_{p,T}$. It is a standard result (e.g., Hayashi, 2000, pp. 18-19) that the OLS estimator of $\delta_{i0}$ satisfies $\hat{\delta}_{i0} - \delta_{i0} = (\ell_T'Q_{p,T}\ell_T)^{-1}\ell_T'Q_{p,T}\tilde{\eta}_i$, where $\tilde{\eta}_i \equiv (\tilde{\eta}_{i1}, \ldots, \tilde{\eta}_{iT})'$ for $\forall i = 1, \ldots, N$. Since $\tilde{\eta}_{it}$ has normal distribution, and let $\sigma_{ii}^2$ denote the $(i,i)$ entry of $\Sigma$, then it is a standard

[9]Schulich School of Business, York University, mkamstra@yorku.ca.
[10]Department of Economics, University of California Riverside, ruoyao.shi@ucr.edu.

result (e.g., Hayashi, 2000, Section 1.3) that $\sqrt{\ell_T' Q_{p,T} \ell_T} \left(\hat{\delta}_{i0} - \delta_{i0}\right) \sim \mathcal{N}_1(0, \sigma_{ii}^2)$. It then only takes some algebra to show that

$$\sqrt{\ell_T' Q_{p,T} \ell_T} \left(\hat{\delta}_0 - \delta_0\right) \sim \mathcal{N}_N(0, \Sigma). \tag{A.2}$$

Now let's take a closer look at $\ell_T' Q_{p,T} \ell_T$:

$$\begin{aligned}
\ell_T' Q_{p,T} \ell_T &= \ell_T' \ell_T - \ell_T' \tilde{r}_p \left(\tilde{r}_p' \tilde{r}_p\right)^{-1} \tilde{r}_p' \ell_T \\
&= T - \left(\sum_{t=1}^T \tilde{r}_{pt}'\right) \left(\sum_{t=1}^T \tilde{r}_{pt} \tilde{r}_{pt}'\right)^{-1} \left(\sum_{t=1}^T \tilde{r}_{pt}\right) \\
&= T - T \left(\frac{1}{T} \sum_{t=1}^T \tilde{r}_{pt}'\right) \left(\frac{1}{T} \sum_{t=1}^T \tilde{r}_{pt} \tilde{r}_{pt}'\right)^{-1} \left(\frac{1}{T} \sum_{t=1}^T \tilde{r}_{pt}\right) \\
&= T(1 - \bar{r}_p' \breve{\Omega}^{-1} \bar{r}_p).
\end{aligned} \tag{A.3}$$

Recall that $\delta_0 = 0$ under the null hypothesis (2), so eq. (A.2) and (A.3) together imply (A.1), the claim of Step 1.

**Step 2.** In this step, we will show that $\hat{\delta}_0 \perp \widehat{\Sigma}$ and

$$(T - L - 1)\widehat{\Sigma} \sim \mathcal{W}_N(T - L - 1, \Sigma). \tag{A.4}$$

Let $X = [\ell_T, \tilde{r}_p]$ denote the $T \times (L + 1)$ design matrix of eq. (1). Define the $T \times T$ projection matrix $P = X\left(X'X\right)^{-1} X'$ and its complement $Q = I_T - P$. Let $\tilde{\eta} = [\tilde{\eta}_1, \ldots, \tilde{\eta}_N]$ denote the $T \times N$ matrix of all disturbances in eq. (1). Then by the standard results of the OLS estimators with normal disturbances (e.g. Hayashi, 2000, Section 1.3), we have $\hat{\delta}_0 \perp \widehat{\Sigma}$ and $(T - L - 1)\widehat{\Sigma} = \sum_{t=1}^T \hat{\eta}_t \hat{\eta}_t' = \tilde{\eta}' Q \tilde{\eta} = \tilde{\eta}' U D U' \tilde{\eta}$, where the last equality holds by the singular value decomposition of $Q$, in which $U$ is a $T \times T$ unitary matrix, and $D$ is a $T \times T$ diagonal matrix with $T - L - 1$ diagonal entries being ones and the rest being zeros. Since we assume that the rows of $\tilde{\eta}$ are mutually independent and follow $\mathcal{N}_N(0, \Sigma)$ distribution, the rows of $U'\tilde{\eta}$ are also mutually independent and follow $\mathcal{N}_N(0, \Sigma)$ distribution. This further implies that $\tilde{\eta}' U D U' \tilde{\eta}$ has the same distribution as such sum $S = \sum_{j=1}^{T-L-1} \xi_j \xi_j'$, where $\xi_j$ are mutually independent and $\xi_j \sim \mathcal{N}_N(0, \Sigma)$ ($j = 1, \ldots, T - L - 1$). By construction, the distribution of $S$ is the Wishart distribution $\mathcal{W}_N(T - L - 1, \Sigma)$. This proves the claim of Step 2.

**Step 3.** In this step, we apply Lemma 2 to the results of Steps 1 and 2. After some simple algebra, we get $\widetilde{W}^* \sim F_{N, T-N-L}$ with $\widetilde{W}^*$ defined in eq. (5). This completes the proof of Lemma 1.

**Proof of Theorem 1.** Based on Lemma 1, we only need to show that $\widetilde{W}^*$ defined in eq. (5) equals $\widetilde{W}$ in eq. (7). By comparing eq. (6) and (3), we see that $\widetilde{\Omega} = \breve{\Omega} - \bar{r}_p \bar{r}_p'$, so it suffices to show that

$$1 - \bar{r}_p' \breve{\Omega}^{-1} \bar{r}_p = \left(1 + \bar{r}_p' \widetilde{\Omega}^{-1} \bar{r}_p\right)^{-1} = \left[1 + \bar{r}_p' \left(\breve{\Omega} - \bar{r}_p \bar{r}_p'\right)^{-1} \bar{r}_p\right]^{-1}. \tag{A.5}$$

Applying Lemma 3 with $A = \breve{\Omega}$, $u = \bar{r}_p$ and $v = -\bar{r}_p$, we get $\left(\breve{\Omega} - \bar{r}_p \bar{r}_p'\right)^{-1} = \breve{\Omega}^{-1} + \frac{\breve{\Omega}^{-1} \bar{r}_p \bar{r}_p' \breve{\Omega}^{-1}}{1 - \bar{r}_p' \breve{\Omega}^{-1} \bar{r}_p}$, which implies that $1 + \bar{r}_p' \left(\breve{\Omega} - \bar{r}_p \bar{r}_p'\right)^{-1} \bar{r}_p = 1 + \bar{r}_p \breve{\Omega}^{-1} \bar{r}_p + \frac{\left(\bar{r}_p' \breve{\Omega}^{-1} \bar{r}_p\right)^2}{1 - \bar{r}_p' \breve{\Omega}^{-1} \bar{r}_p} = \left(1 - \bar{r}_p' \breve{\Omega}^{-1} \bar{r}_p\right)^{-1}$, which further immediately implies eq. (A.5). This completes the proof of Theorem 1.

# B Simulation Results

Following the literature, we generate portfolio excess returns $\tilde{r}_{pt}$ as normal, independent and identically distributed, calibrated to monthly U.S. stock returns. The excess return for test asset $i$ and time $t$ is generated based on model (1), which is rewritten here:

$$\tilde{r}_{it} = \delta_{i0} + \sum_{j=1}^{L} \delta_{ij} \tilde{r}_{jt} + \tilde{\eta}_{it}, \tag{B.1}$$

where $\tilde{\eta}_{it} \sim iid\ \ normal$ across $t$ with mean 0 and volatility $\sigma_{ii}$, and $\tilde{r}_{jt} \sim iid\ \ normal$ across $t$ with mean $(\mu_j/L)$, volatility $\sigma_j$ and $E\left[\tilde{\eta}_{it}\tilde{r}_{jt}\right] = 0$. We set $\mu_j = 0.01$, $\sigma_j = 0.02$, $\sigma_{ii} = 0.08$, $\delta_{ij} = 1$, $\forall\ i, j$ and we explore only the case of $\delta_{i0} = 0$, $\forall\ i$.

We explore size properties of the correct and incorrect formulas of the GRS statistic for numbers of portfolios (L) from 1 to 5, test assets (N) from 1 to 25, and sample sizes (T) from 36 (months) to 400. This spans typical applications of the GRS test. Our simulations show that the performance of the incorrect GRS formula generally suffers deterioration as the number of firms and factors increases, as one might expect. We present a subset of simulation results in Table B.1, where the over-rejection due to the degree of freedom adjustment for the sample variance-covariance matrix of the portfolio excess returns is displayed.

The correct formula of the GRS test generally presents no evidence of incorrect size, as our simulation setting is one in which it should have correct finite-sample exact size. The incorrect formula of the GRS test shows over-rejection even at 400 months for $(N, L) = (5, 4)$ and at 36 months months the over-rejection can be somewhat large, for instance rejecting 10.6% of the time at the 10% level for $(N, L) = (5, 4)$.

Bootstrap simulations using non-normal data, available on request, show a smaller deviation of the correct and incorrect GRS test size performance.

Table B.1: Null Rejection Rates

| Test Assets ($N$) | Portfolios ($L$) | Periods ($T$) | $\widehat{W}$ (Incorrect GRS) 1% | 5% | 10% | $\widetilde{W}$ (Correct GRS) 1% | 5% | 10% |
|---|---|---|---|---|---|---|---|---|
| 1 | 3 | 36 | **0.0107** | **0.0519** | **0.1027** | **0.0101** | 0.0500 | 0.1000 |
| | | 60 | **0.0104** | **0.0509** | **0.1016** | 0.0100 | 0.0498 | 0.0999 |
| | | 100 | **0.0103** | **0.0509** | **0.1011** | 0.0100 | **0.0502** | **0.1001** |
| | | 400 | 0.0100 | **0.0501** | **0.1001** | 0.0100 | 0.0499 | 0.0998 |
| 5 | 4 | 36 | **0.0111** | **0.0540** | **0.1064** | **0.0101** | **0.0502** | **0.1002** |
| | | 60 | **0.0107** | **0.0524** | **0.1040** | 0.0100 | 0.0499 | 0.1000 |
| | | 100 | **0.0104** | **0.0515** | **0.1024** | 0.0100 | 0.0500 | **0.1001** |
| | | 400 | 0.0100 | **0.0504** | **0.1007** | 0.0099 | 0.0500 | 0.1000 |
| 25 | 5 | 36 | **0.0106** | **0.0527** | **0.1050** | 0.0100 | 0.0499 | 0.0999 |
| | | 60 | **0.0109** | **0.0538** | **0.1064** | 0.0100 | 0.0500 | **0.1001** |
| | | 100 | **0.0108** | **0.0529** | **0.1046** | **0.0101** | **0.0502** | 0.1000 |
| | | 400 | **0.0102** | **0.0508** | **0.1012** | 0.0099 | 0.0500 | 0.0999 |

Notes: (1) Bold-faced numbers are rejection rates larger than the nominal values.
(2) The results are based on 5,000,000 simulations.
(3) The models are: $\tilde{r}_{it} = \delta_{i0} + \sum_{j=1}^{L} \delta_{ij}\tilde{r}_{jt} + \tilde{\eta}_{it}$, where $\tilde{r}_{jt} \sim iid\,\mathcal{N}\left(\mu_j/L, \sigma_j^2\right)$, $\tilde{\eta}_{it} \sim iid\,\mathcal{N}\left(0, \sigma_{ii}^2\right)$ and $E\left[\tilde{\eta}_{it}\tilde{r}_{jt}\right] = 0$. For $\forall j$, $\mu_j = 0.01$, $\sigma_j = 0.02$, $\sigma_{ii} = 0.08$, and $\delta_{ij} = 1$.

# C Empirical Results

We now turn to some empirical examples, borrowing from Fama and French (2015, 2016). The models they report GRS statistics for include the CAPM, the Fama-French 3-factor model, several variations of a 4-factor model, and their 5-factor model. The test assets they explore include five by five sortings based on market capitalization and various anomaly variables including operating profitability, return volatility, residual volatility, accruals and so on. We use data retrieved from the French data library, and we consider five year periods (Table C.1 Panel A) and twenty year periods (Table C.1 Panel B) drawn from the time period 1963-2019.[11].

Our results with test assets can be found in Table C.1. What we find is that the rankings of the models are sensitive to miscalculation of the GRS statistic. Consider the case of the test assets comprised of the 25 accruals-ME (market capitalization) portfolios and a five year estimation horizon. The simple CAPM and a five factor model flip rankings if we use the incorrectly calculated GRS statistic, in spite of an average absolute alpha for the CAPM that is twice as high as for the five factor model. Similarly, if we look at the five by five sorting by CAPM beta and market capitalization, the simple CAPM model is misranked above a four factor model when the incorrectly calculated GRS statistic is employed. If we expand the set of test assets

[11]We thank Ken French for making this valuable resource freely available.

Table C.1: Ranking Between Small And Large Models

| Test Assets<br>Small Model<br>Large Model | \|Average\|<br>Annualized Alpha<br>$\overline{\lvert\hat{\delta}_0\rvert}$ | Correct GRS<br>$\widetilde{W}$ | <br>Rank | Incorrect GRS<br>$\widehat{W}$ | <br>Rank |
|---|---|---|---|---|---|
| **Panel A:** | | | | | |
| 5x5 Accruals × ME | | | | | |
| Mkt | 3.02 | 0.8148 | **2** | 0.8150 | **1** |
| Mkt SMB HML RMW CMA | 1.44 | 0.8144 | **1** | 0.8174 | **2** |
| 5x5 Beta × ME | | | | | |
| Mkt | 3.73 | 2.4026 | **2** | 2.4026 | **1** |
| Mkt SMB HML RMW | 3.46 | 2.3951 | **1** | 2.4038 | **2** |
| **Panel B:** | | | | | |
| 5x5 ME × UMD | | | | | |
| Mkt | 3.99 | 2.1984 | **2** | 2.2078 | **1** |
| Mkt SMB HML RMW CMA | 3.36 | 2.1816 | **1** | 2.2293 | **2** |
| 5x5 ME × OP | | | | | |
| Mkt | 1.74 | 2.5132 | **2** | 2.5242 | **1** |
| Mkt SMB HML RMW CMA | 1.34 | 2.4907 | **1** | 2.5456 | **2** |
| 5x5 Variance × ME | | | | | |
| Mkt | 4.96 | 4.5547 | **2** | 4.5744 | **1** |
| Mkt SMB HML CMA | 2.72 | 4.5483 | **1** | 4.6284 | **2** |
| 2x4x4 ME × MEBE × OP | | | | | |
| Mkt | 4.70 | 2.8994 | **2** | 2.8996 | **1** |
| Mkt SMB HML CMA | 2.79 | 2.8993 | **1** | 2.9005 | **2** |

Notes: (1) In all cases, the incorrect GRS statistic inflates the large model more to flip the ranks.
(2) For Panel A the sample periods cover January 2002 to December 2006 and January 2007 to December 2011, respectively.
(3) For Panel B the sample periods are 20 year windows selected from 1963 to 2019.
(4) For detailed description of the factor and test asset construction see Fama and French (2015, 2016).

(results available on request), we see that very few models maintain their ranking if we use the incorrectly calculated GRS statistic.

It is not difficult to find cases using a fifteen or twenty year window for which rankings of models flip. It is also common to find a model with a large average alpha incorrectly ranked above another which a much smaller alpha. See for instance the case of a five by five sorting by market capitalization and momentum (UMD), five by five sorting by market capitalization and operating profit (OP), five by five sorting by market capitalization and total return variance, two by four by four sorting by market capitalization, book-to-market (BEME) and operating profits, all cases for which the CAPM is ranked incorrectly as superior to a four or five factor model.

The important insights to take away from these results include that the error in calculating the GRS statistic is unlikely to overturn studies using several decades of monthly return data, but can have a material impact on empirical results when twenty or fewer years of data are used, which is not uncommon in empirical asset pricing studies. For example, Barillas and Shanken (2018) do not use the GRS but do model comparisons on a little less than fifteen years of monthly data. Sha and Gao (2019) use 144 months of data and exploit six metrics to evaluation factor model performance, including the GRS statistic. Baek and Bilson (2015) consider 234 months of data in a subsample estimation. Chiah, Chai, Zhong and Li (2016) use 23 years of data when comparing models using the GRS. One takeaway from these papers is that many situations involving specialized data (like Sha and Gao 2019 and their exploration of mutual fund returns in China) or sub-sample robustness checks (like Baek and Bilson 2015) necessarily are constrained to shorter samples than fifty or even twenty years, so that the bias from an incorrectly calculated GRS statistic becomes large.

# D Derivation of Eq. (11)

The optimal portfolio weights $\omega^*$ solve

$$\min_{\omega} \omega' \Omega_{\tilde{r}} \omega, \quad \text{subject to} \quad \omega' \mu_{\tilde{r}} = m.$$

The first order conditions for this problems are $\omega^* = \varphi \Omega_{\tilde{r}}^{-1} \mu_{\tilde{r}}$ and $\varphi = m/(\mu_{\tilde{r}}' \Omega_{\tilde{r}}^{-1} \mu_{\tilde{r}})$, where $\varphi$ is the Lagrange multiplier. Recall that the Sharpe ratio captures the mean excess return to a portfolio per unit of volatility (standard deviation), so the squared Sharpe ratio of the optimal portfolio composed of these $K$ assets is

$$\theta^{*2} \equiv \left( \frac{m}{\sqrt{\omega^{*\prime} \Omega_{\tilde{r}} \omega^*}} \right)^2 = \mu_{\tilde{r}}' \Omega_{\tilde{r}}^{-1} \mu_{\tilde{r}},$$

in which the variance-covariance matrix $\Omega_{\tilde{r}}$ plays a central role.

Applying this general result, we know that when the constituent assets are the $L$

portfolios, the squared Sharpe ratio is

$$\theta_p^{*2} = \mu'_{\tilde{r}_p} \Omega^{-1} \mu_{\tilde{r}_p}. \tag{D.1}$$

When the constituent assets include both the $N$ test assets and the $L$ portfolios, the squared Sharpe ratio is $\theta_{N+L}^{*2} = \mu'_{\tilde{r}_{N+L}} \Omega_{\tilde{r}_{N+L}}^{-1} \mu_{\tilde{r}_{N+L}}$, where $\mu_{\tilde{r}_{N+L}} \equiv \left(\mu'_{\tilde{r}_N}, \mu'_{\tilde{r}_p}\right)'$,

$$\Omega_{\tilde{r}_{N+L}} \equiv \begin{bmatrix} \delta\Omega\delta' + \Sigma & \delta\Omega \\ \Omega\delta' & \Omega \end{bmatrix}, \tag{D.2}$$

and $\delta \equiv [\delta_1, \ldots, \delta_N]'$ with $\delta_i$ being the slope coefficient in model (1). Eq. (D.2) holds because we can rewrite the variance-covariance matrix of the $N$ test assets and their covariance matrix with the $L$ portfolios using $\Omega$, $\Sigma$ and $\delta$ (in the same way as $\hat{V}$ on p.1143 and eq. (24) in GRS, 1989). Applying the inverse formula for a block matrix and noticing the relationship between $\mu_{\tilde{r}_N}$ and $\mu_{\tilde{r}_p}$ implied by model (1), we get

$$\theta_{N+L}^{*2} = \theta_p^{*2} + \delta_0' \Sigma^{-1} \delta_0, \tag{D.3}$$

which is essentially the same as eq. (22) and (23) in MacKinlay (1995). This, together with eq. (D.1) and simple algebra, further implies eq. (11).

# E Software Packages

SAS and R packages to implement our generalized GRS test can be found at the authors' websites: http://markkamstra.com/data.html (SAS) and https://ruoyaoshi.github.io/ (R). A Stata package `grsftest` coded by Mengnan (Cliff) Zhu can be found at https://ideas.repec.org/c/boc/bocode/s458828.html.

# References


Anderson, T.W., 2003. An Introduction to Multivariate Statistical Analysis, Third Edition. *Wiley-Interscience.*

Baek, S. and Bilson, J.F., 2015. Size and value risk in financial firms. *Journal of Banking and Finance* 55, 295-326.

Barillas, F. and Shanken, J., 2018. Comparing asset pricing models. *The Journal of Finance* 73(2), 715-754.

Bartlett, M. S., 1951. An Inverse Matrix Adjustment Arising in Discriminant Analysis. *The Annals of Mathematical Statistics* 22(1), 107-111.

Chiah, M., Chai, D., Zhong, A. and Li, S., 2016. A Better Model? An empirical investigation of the Fama-French five-factor model in Australia. *International*

*Review of Finance* 16(4), 595-638.

Fama, E. F. and French, K.R., 2015. A Five-Factor Asset Pricing Model. *Journal of Financial Economics* 116(1), 1-22.

Fama, E. F. and French, K.R., 2016. Dissecting Anomalies With A Five-Factor Model. *The Review of Financial Studies* 29(1), 69-103.

Gibbons, M., Ross, S., and Shanken, J., 1989. A Test of the Efficiency of A Given Portfolio. *Econometrica* 57(5), 1121-1152.

Hayashi, F., 2000. Econometrics. *Princeton: Princeton University Press.*

MacKinlay, A. G., 1995. Multifactor models do not explain deviations from the CAPM. *Journal of Financial Economics* 38(1), 3-28.

Sha, Y. and Gao, R., 2019. Which is the best: A comparison of asset pricing factor models in Chinese mutual fund industry. *Economic Modelling* 83, 8-16.