Finite Sample Analysis of Two-Pass Cross-Sectional Regressions

ROBERT CHEN
University of Toronto

RAYMOND KAN*
University of Toronto

First draft: September 2003
This version: April 2005

*We thank Karim Abadir, Ambrus Kecskés, Lynda Khalaf, Jan Magnus, Kevin Wang, seminar participants at the University of Toronto, and participants at the 2003 Northern Finance Meetings for helpful discussions and comments. Kan gratefully acknowledges financial support from the Social Sciences and Humanities Research Council of Canada. Corresponding author: Raymond Kan, Joseph L. Rotman School of Management, University of Toronto, 105 St. George Street, Toronto, Ontario, Canada M5S 3E6; Tel: (416)978-4291; Fax: (416)978-5433; email:kan@chass.utoronto.ca.

Finite Sample Analysis of Two-Pass Cross-Sectional Regressions

ABSTRACT

We investigate the finite sample properties of the two-pass cross-sectional regression (CSR) methodology, which is popular for estimating risk premia and testing beta pricing models. We find that the finite sample distributions of the estimated risk premia differ significantly from their asymptotic distributions. In particular, the risk premia estimates obtained from the second-pass CSR of average returns on estimated betas can be seriously biased even when the number of time series observations is reasonably large. In addition, the standard error of the estimated risk premia based on the asymptotic distribution overstates the actual standard error. We show that popular alternative estimators in the literature including the maximum likelihood estimator have no finite integral moments and therefore cannot be used to correct the bias. We propose a new bias adjustment of the estimated zero-beta rate and risk premia and we show that the adjusted version has a smaller bias than the unadjusted version.

In the empirical asset pricing literature, the popular two-pass cross-sectional regression (CSR) methodology developed by Black, Jensen, and Scholes (1972) and Fama and MacBeth (1973) is often used for estimating risk premia and testing beta asset pricing models. Although there are many variations on this two-pass methodology, its basis setup always involves two steps. In the first pass, the betas of the test assets are estimated using the usual ordinary least squares (OLS) time series regression of returns on some common factors. In the second pass, the returns on test assets are regressed on the estimated betas obtained from the first pass. By running this second-pass CSR on a period-by-period basis, we obtain a time series of the intercept and the slope coefficients. The average values of the intercept and the slope coefficients are then used as estimates of the zero-beta rate and the risk premia.

Since betas are estimated with errors in the first pass time series regression, the errors-in-variables (EIV) problem is introduced in the second-pass CSR. Measurement errors in the estimated betas cause two problems. The first problem is that the estimated zero-beta rate and risk premia are biased. However, as the length of the times series used to estimate betas increases to infinity, Shanken (1992) shows that the estimation errors of the betas approach to zero and hence the estimated zero-beta rate and risk premia from the second pass CSR are still consistent. The second problem is that the standard errors of the estimated zero-beta rate and risk premia that are due to Fama and MacBeth (1973) are inconsistent. Shanken (1992) proposes an asymptotically valid EIV adjustment of the standard errors.

Unlike the asymptotic results which are nicely presented by Shanken (1992), finite sample analysis of the two-pass CSR is largely unavailable in the literature. Aside from some simulation studies like Affleck-Graves and Bradfield (1993) and Shanken and Zhou (2000), there is very little understanding of the finite sample properties of the estimated zero-beta rate and risk premia from the second-pass CSR. For example, while we know that the estimated risk premium is biased, we do not know the magnitude, direction, or determining parameters of this bias. Although the estimated zero-beta rate and risk premia from the second-pass CSR are consistent, it is not clear that their finite sample biases are negligible. In many applications, betas of test assets are estimated using as few as 60 monthly observations, so it is reasonable to suspect that the asymptotic results may not be entirely relevant. In addition, it is also unclear how well the unadjusted standard errors and Shanken's EIV adjusted standard errors approximate the true standard errors of the estimated zero-beta rate and risk premia.

We provide a finite sample analysis of the biases and variances of the estimated zero-beta rate and risk premia from the second-pass CSR, for both the cases of ordinary least squares (OLS) and generalized least squares (GLS). For the single factor case, we provide analytical expressions of the finite sample bias and variance of the estimated zero-beta rate and risk premium. For the multi-factor case, we provide a simple and fast simulation approach in obtaining the biases and variances. For reasonable choices of parameters, we show that the finite sample bias of the risk premium can be more than 80% when betas are estimated using only 60 monthly observations. Even when betas are estimated using as many as 600 monthly observations, the bias can still be as high as 30% of the true value. In addition, we find that the unadjusted and EIV adjusted standard errors tend to overstate the true standard errors of the estimated zero-beta rate and risk premium. The biases in the point estimate of the risk premium coupled with the overstatement of its standard error can lead researchers to wrongly accept the null hypothesis of zero risk premium even when the risk premium is actually nonzero.

Attempts to correct the bias in the CSR estimators are available in the literature. These include the maximum likelihood estimator and the adjusted estimators of Litzenberger and Ramaswamy (1979) and Kim (1995). While these estimators are shown to be consistent when the number of assets goes to infinity, there is little understanding of their finite sample properties. In this paper, we also present a finite sample analysis of these alternative estimators. Surprisingly, we find that the first moments of these adjusted estimators do not exist and hence they cannot be used to correct the bias. Based on our analysis of the bias, we propose a new bias adjustment of the point estimates of the zero-beta rate and risk premium. Simulation results show that our bias-adjusted estimators perform better than the unadjusted estimators in finite samples.

The rest of the paper is organized as follows. Section 1 provides an overview of the two-pass CSR methodology and summarizes existing results and previous attempts to correct the biases in the CSR estimators. Section 2 presents a fast simulation approach that allows us to obtain the distribution and the moments of the estimated zero-beta rate and risk premia for the general multi-factor case. It also provides theoretical results on the existence of moments for the unadjusted and adjusted estimators. Section 3 presents analytical results of the finite sample bias and variance of the estimated zero-beta rate and risk premium for the single factor case. Section 4 presents our new bias-adjusted estimators of the risk premium. Section 5 presents simulation results to examine the robustness of our analysis as well as to evaluate how well our bias-adjusted estimators of risk premium perform in finite samples. Section 6 concludes the paper. The Appendix contains proofs of all propositions.

1. Overview of the Two-Pass Methodology

1.1 Two-Pass Cross-Sectional Regressions

Traditional asset pricing theories, such as those of Sharpe (1964), Lintner (1965), Black (1972), Merton (1973), Ross (1976) and Breeden (1979), relate the expected return on a financial asset to its covariances (or betas) with some systematic risk factors. Let $Y_t = [f_t',\ R_t']'$ be an $N+K$ vector where f_t is the realization of K systematic factors at time t and R_t is the return on N test assets at time t. Denote the mean and variance of Y_t as

$$\mu = E[Y_t] \equiv \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \tag{1}$$

$$V = \mathrm{Var}[Y_t] \equiv \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix}, \tag{2}$$

where V is assumed to be nonsingular. If the K-factor asset pricing model holds, the expected returns of the N assets are given by

$$\mu_2 = 1_N\gamma_0 + \beta\gamma_1 = H\gamma, \tag{3}$$

where 1_N is an N-vector of ones, $\beta = V_{21}V_{11}^{-1}$, $H = [1_N,\ \beta]$, and $\gamma = [\gamma_0,\ \gamma_1']'$. Under this setup, γ_0 is usually called the zero-beta rate and γ_1 is called the risk premia associated with the K factors.

Suppose we have T observations of Y_t. The popular two-pass CSR approach estimates γ by first estimating β using an OLS regression of R_t on a constant term and f_t

$$R_t = \alpha + \beta f_t + \epsilon_t, \quad t = 1, \ldots, T, \tag{4}$$

where ϵ_t is the regression residual at time t. Defining the sample mean and variance of Y_t as

$$\hat{\mu} = \frac{1}{T}\sum_{t=1}^{T} Y_t \equiv \begin{bmatrix} \hat{\mu}_1 \\ \hat{\mu}_2 \end{bmatrix}, \tag{5}$$

$$\hat{V} = \frac{1}{T}\sum_{t=1}^{T}(Y_t - \hat{\mu})(Y_t - \hat{\mu})' \equiv \begin{bmatrix} \hat{V}_{11} & \hat{V}_{12} \\ \hat{V}_{21} & \hat{V}_{22} \end{bmatrix}, \tag{6}$$

the OLS estimate of β is given by

$$\hat{\beta} = \hat{V}_{21}\hat{V}_{11}^{-1}. \tag{7}$$

Equipped with the estimated betas from the first pass, the second pass runs a CSR of $\hat{\mu}_2$ on $\hat{H} = [1_N,\ \hat{\beta}]$. The second-pass CSR can be run in various ways, the most popular being the OLS CSR. Under the OLS CSR, the estimated γ is given by

$$\tilde{\gamma} = (\hat{H}'\hat{H})^{-1}(\hat{H}'\hat{\mu}_2). \tag{8}$$

The variance of ϵ_t in (4) is given by

$$\Sigma = V_{22} - V_{21}V_{11}^{-1}V_{12}. \tag{9}$$

If Σ is known, then a more efficient estimate of γ can be obtained using a true GLS CSR in the second pass. The true GLS estimate of γ is given by

$$\check{\gamma} = (\hat{H}'\Sigma^{-1}\hat{H})^{-1}(\hat{H}'\Sigma^{-1}\hat{\mu}_2). \tag{10}$$

In most applications, Σ is unknown and needs to be estimated. In practice,

$$\hat{\Sigma} = \hat{V}_{22} - \hat{V}_{21}\hat{V}_{11}^{-1}\hat{V}_{12} \tag{11}$$

replaces the Σ in the true GLS CSR. When $T > N + K$, the inverse of $\hat{\Sigma}$ exists and it is possible to run the CSR using $\hat{\Sigma}^{-1}$ as the weighting matrix. The resulting estimate of γ from this estimated GLS CSR is given by

$$\hat{\gamma} = (\hat{H}'\hat{\Sigma}^{-1}\hat{H})^{-1}(\hat{H}'\hat{\Sigma}^{-1}\hat{\mu}_2). \tag{12}$$

There are other ways of running the second-pass CSR, but they are not as common as OLS and GLS, so we limit our discussion to the OLS and the true and estimated GLS cases. Nevertheless, our results can be easily generalized to other versions of the second-pass CSR.

In the discussion above, we assumed that the same $\hat{\beta}$ is used throughout the entire sample period, which allows us to simply run a single CSR of $\hat{\mu}_2$ on $\hat{H}$ to estimate γ. Although this approach is quite popular, there are also quite a lot of studies that allow $\hat{\beta}$ to change throughout the sample period. For example, the popular Fama-MacBeth (1973) methodology runs the OLS CSR on a period-by-period basis by regressing realized return R_t on $\hat{H}_t = [1_N,\ \hat{\beta}_t]$, where $\hat{\beta}_t$ is the estimated betas of the N assets at time t, typically estimated from an earlier period. For example, the time t estimate of the OLS CSR estimate of γ is given by

$$\tilde{\gamma}_t = (\hat{H}_t'\hat{H}_t)^{-1}(\hat{H}_t'R_t). \tag{13}$$

By repeating this OLS CSR period by period, we get a time series of $\tilde{\gamma}_t$ and the resulting estimate of γ is simply the time series average of $\tilde{\gamma}_t$. In this paper, we focus on the constant beta case, but some of our results are also applicable to the time-varying beta case with some slight modifications, and we will note the required modifications in the subsequent analysis.

1.2 Existing Results

For finite sample inference, we assume that Y_t is i.i.d. normal. Nevertheless, we will use simulation to examine the robustness of our results to departures from normality. Under the normality assumption, it is well known that $\hat{\mu}$ and $\hat{V}$ are independent, with $\hat{\mu} \sim N(\mu, V/T)$ and $T\hat{V} \sim W_{N+K}(T-1, V)$, where $W_{N+K}(T-1, V)$ is an $(N+K)$-dimensional central Wishart distribution with $T-1$ degrees of freedom and covariance matrix V.

In many situations, one is interested in the properties of the estimated γ conditional on the realizations of the factors. Conditional on $\hat{\mu}_1$,

$$\hat{\mu}_2 \sim N(\alpha + \beta\hat{\mu}_1, \Sigma/T). \tag{14}$$

When the asset pricing model is correct,

$$\alpha = \mu_2 - \beta\mu_1 = 1_N\gamma_0 + \beta\gamma_1 - \beta\mu_1, \tag{15}$$

so conditional on $\hat{\mu}_1$,

$$\hat{\mu}_2 \sim N(1_N\gamma_0 + \beta\bar{\gamma}_1, \Sigma/T), \tag{16}$$

where $\bar{\gamma}_1 = \gamma_1 - \mu_1 + \hat{\mu}_1$ and it is called the *ex post* risk premia by Shanken (1992).[1]

Conditional on $\hat{V}_{11}$, $\hat{\beta}$ and $\hat{\Sigma}$ are independent of each other, with distributions

$$\text{vec}(\hat{\beta}) \sim N(\text{vec}(\beta), (T\hat{V}_{11})^{-1} \otimes \Sigma), \tag{17}$$

$$T\hat{\Sigma} \sim W_N(T-K-1, \Sigma). \tag{18}$$

[1] Our analysis focuses on the case in which the asset pricing model is correct because this is the case in which γ is well-defined. Nevertheless, it is easy to generalize our results to the case of general μ_2 in which the K-factor asset pricing model does not hold.

The above are standard results from the normality assumption. One can refer to Muirhead (1982, Theorem 10.1.2) for the proof of these results.[2] Note that $\hat{\beta}$ and $\hat{\Sigma}$ are functions of elements of $\hat{V}$, so they are also independent of $\hat{\mu}_2$.

While the distributions of $\hat{\mu}_2$, $\hat{\beta}$, and $\hat{\Sigma}$ are known, the estimated γ is a complicated function of these random variables, so obtaining the distribution or the moments of the estimated γ in finite samples is nontrivial. In the existing literature, there are two methods to make inferences about γ. The first method ignores the fact that $\hat{\beta}$ is estimated with error. If we treat $\hat{H}$ as H, then

$$\tilde{\gamma} \sim N\left(\gamma, \frac{1}{T}(H'H)^{-1}(H'V_{22}H)(H'H)^{-1}\right), \tag{19}$$

$$\check{\gamma} \sim N\left(\gamma, \frac{1}{T}(H'\Sigma^{-1}H)^{-1}(H'\Sigma^{-1}V_{22}\Sigma^{-1}H)(H'\Sigma^{-1}H)^{-1}\right). \tag{20}$$

Using the fact that

$$V_{22} = \Sigma + \beta V_{11}\beta' = \Sigma + H \begin{bmatrix} 0 & 0_K' \\ 0_K & V_{11} \end{bmatrix} H', \tag{21}$$

where 0_K is a K-vector of zeros, we can write the variance of $\tilde{\gamma}$ and $\check{\gamma}$ as

$$\mathrm{Var}[\tilde{\gamma}] = \frac{1}{T}\left((H'H)^{-1}(H'\Sigma H)(H'H)^{-1} + \begin{bmatrix} 0 & 0_K' \\ 0_K & V_{11} \end{bmatrix}\right), \tag{22}$$

$$\mathrm{Var}[\check{\gamma}] = \frac{1}{T}\left((H'\Sigma^{-1}H)^{-1} + \begin{bmatrix} 0 & 0_K' \\ 0_K & V_{11} \end{bmatrix}\right). \tag{23}$$

As for the estimated GLS case, one often ignores the estimation error in $\hat{\Sigma}$ and relies on (20) to make statistical inferences.

Note that if β is known, then

$$\tilde{\gamma}_t = (H'H)^{-1}H'R_t \sim N(\gamma, (H'H)^{-1}H'V_{22}H(H'H)^{-1}), \tag{24}$$

and $\tilde{\gamma}_t$ is i.i.d. normal. Therefore, using a time-series of OLS CSR estimates $\tilde{\gamma}_t$ of length T, one can perform a t-test of $H_0 : a'\gamma = c$, where a is a constant $(K+1)$-vector and c is a constant scalar, using the test statistic

$$\frac{\frac{1}{T}\sum_{t=1}^{T} a'\tilde{\gamma}_t - c}{s(a'\tilde{\gamma}_t)/\sqrt{T}}, \tag{25}$$

where $s(a'\tilde{\gamma}_t)$ is the sample standard deviation of the time series $a'\tilde{\gamma}_t$. Under the null hypothesis, the test statistic has a t-distribution with $T-1$ degrees of freedom. This is the foundation of the

[2]See also Lemma 1 of Shanken (1992).

popular t-test that is used by Fama and MacBeth (1973) and many other studies. A similar t-test can also be performed using the GLS CSR estimates.

In reality, β is estimated with error, so the EIV problem is introduced in the second-pass CSR when $\hat{\beta}$ is used instead of the true β. Shanken (1992) provides a nice asymptotic analysis of this problem. He shows that although β is estimated with error, the estimation error in $\hat{\beta}$ goes to zero as T goes to infinity and the second-pass CSR estimate of γ is T-consistent. However, the usual standard error for the estimated γ is inconsistent, and we need an adjustment to account for the estimation errors in $\hat{\beta}$. Shanken (1992) shows that[3]

$$\sqrt{T}(\tilde{\gamma}-\gamma) \overset{A}{\sim} N\left(0_{K+1},(1+\gamma_1'V_{11}^{-1}\gamma_1)(H'H)^{-1}(H'\Sigma H)(H'H)^{-1}+\begin{bmatrix} 0 & 0_K' \\ 0_K & V_{11}\end{bmatrix}\right), \quad (26)$$

$$\sqrt{T}(\check{\gamma}-\gamma) \overset{A}{\sim} N\left(0_{K+1},(1+\gamma_1'V_{11}^{-1}\gamma_1)(H'\Sigma^{-1}H)^{-1}+\begin{bmatrix} 0 & 0_K' \\ 0_K & V_{11}\end{bmatrix}\right), \quad (27)$$

$$\sqrt{T}(\hat{\gamma}-\gamma) \overset{A}{\sim} N\left(0_{K+1},(1+\gamma_1'V_{11}^{-1}\gamma_1)(H'\Sigma^{-1}H)^{-1}+\begin{bmatrix} 0 & 0_K' \\ 0_K & V_{11}\end{bmatrix}\right). \quad (28)$$

For statistical inference, one replaces the terms β, Σ, γ_1, and V_{11} in the asymptotic variance with their sample counterparts.

While the asymptotic results of Shanken (1992) are elegant, their relevance for the applications that are typically encountered is questionable. In many studies, $\hat{\beta}$ is estimated using as few as 60 monthly observations and it can be very volatile, so there may be a serious finite sample bias in the estimated γ from the second-pass CSR. In addition, the asymptotic variance may also be inappropriate for finite sample analysis. To evaluate how relevant the asymptotic results are, one can perform a simulation experiment. However, a typical simulation study requires us to specify μ and V and simulate T observations of Y_t in order to draw one realization of the estimated γ. Aside from being time-consuming, the conclusion is specific to a given choice of μ and V, so one cannot generalize from such simulation studies. As a result, not much has been done to document the finite sample behavior of the second-pass CSR estimate of γ. In Section 2, we present a simplification of the problem which allows us to reduce the number of parameters. Such simplification also allows us to deliver a speedy simulation method that frees us from having to simulate T observations of Y_t for every draw of the estimated γ.

[3]Jagannathan and Wang (1998) and Ahn and Gadarowski (1999) provide asymptotic results for more general cases.

1.3 Adjusted Estimators

Since the estimators of γ in the two-pass CSR are subject to the EIV problem, they are biased in general. There have been some attempts in the literature to come up with adjusted estimators of γ to reduce this bias. The first attempt is due to Litzenberger and Ramaswamy (1979) who develop an adjustment for the weighted least squares CSR. Shanken (1992) later generalizes and improves this adjusted estimator. The idea behind this adjusted estimator is that while $\hat{H}$ is an unbiased estimator of H, $\hat{H}'A\hat{H}$ is in general not an unbiased estimator of $H'AH$ where A is an $N \times N$ matrix, so by subtracting the bias from $\hat{H}'A\hat{H}$ one hopes to improve the finite sample properties of the two-pass CSR estimators. Under this adjustment, the OLS, true GLS and estimated GLS CSR estimators of γ are, respectively,

$$\tilde{\gamma}^{LR} = \left(\hat{H}'\hat{H} - \begin{bmatrix} 0 & 0_K' \\ 0_K & \mathrm{tr}(\hat{\Sigma})\hat{V}_{11}^{-1}/(T-K-1) \end{bmatrix}\right)^{-1} (\hat{H}'\hat{\mu}_2), \tag{29}$$

$$\check{\gamma}^{LR} = \left(\hat{H}'\Sigma^{-1}\hat{H} - \begin{bmatrix} 0 & 0_K' \\ 0_K & N\hat{V}_{11}^{-1}/T \end{bmatrix}\right)^{-1} (\hat{H}'\Sigma^{-1}\hat{\mu}_2), \tag{30}$$

$$\hat{\gamma}^{LR} = \left(\hat{H}'\hat{\Sigma}^{-1}\hat{H} - \begin{bmatrix} 0 & 0_K' \\ 0_K & N\hat{V}_{11}^{-1}/(T-N-K-2) \end{bmatrix}\right)^{-1} (\hat{H}'\hat{\Sigma}^{-1}\hat{\mu}_2). \tag{31}$$

Under some conditions, Shanken (1992) shows that these adjusted estimators are N-consistent, i.e., when the number of assets goes to infinity, these adjusted estimators converge to $[\gamma_0,\ \bar{\gamma}_1']'$.

Another adjusted estimator that is N-consistent is due to Kim (1995), who develops an EIV-adjusted estimator of γ for the one-factor case based on the results of Fuller (1987). Here, we generalize his estimator to the K-factor case and provide a simple analytical expression for this estimator.[4] The idea behind Kim's adjusted estimator is to choose γ_0, $\bar{\gamma}_1$, and β to minimize a quadratic form $g'\Omega^{-1}g$ where

$$g(\beta,\gamma_0,\bar{\gamma}_1) = \begin{bmatrix} \hat{\mu}_2 - 1_N\gamma_0 - \beta\bar{\gamma}_1 \\ \mathrm{vec}(\hat{\beta}) - \mathrm{vec}(\beta) \end{bmatrix} = \begin{bmatrix} \hat{\mu}_2 - 1_N\gamma_0 - (\bar{\gamma}_1' \otimes I_N)b \\ \hat{b} - b \end{bmatrix}, \tag{32}$$

with $b = \mathrm{vec}(\beta)$, $\hat{b} = \mathrm{vec}(\hat{\beta})$, and

$$\Omega = \begin{bmatrix} \Sigma/T & \mathrm{O}_{N\times NK} \\ \mathrm{O}_{NK\times N} & \hat{V}_{11}^{-1} \otimes \Sigma/T \end{bmatrix} = \begin{bmatrix} 1 & 0_K' \\ 0_K & \hat{V}_{11}^{-1} \end{bmatrix} \otimes \Sigma/T. \tag{33}$$

[4] Kim's estimator is developed for Fama-MacBeth regressions which use rolling betas. Here we present the results only for the constant beta case. Generalizing the results to the rolling beta case is straightforward.

It makes sense to use Ω^{-1} as the weighting matrix because under the normality assumption, Ω is the variance of g conditional on $\hat{\mu}_1$ and $\hat{V}_{11}$. We call the estimator of γ from this procedure as "Kim's true GLS estimator" and denote it as $\breve{\gamma}^K$. However, when Σ is not known, we replace Σ by $\hat{\Sigma}$ in Ω and the resulting estimator is denoted as $\hat{\gamma}^K$.[5] For completeness, we also present the adjusted OLS estimator $\tilde{\gamma}^K$, in which we replace Σ by I_N.

In obtaining $\tilde{\gamma}^K$, $\breve{\gamma}^K$, and $\hat{\gamma}^K$, one needs to minimize $g'\Omega^{-1}g$. Although the objective function is nonlinear in the parameters, it turns out that a simple analytical solution is available. The following lemma provides this solution.

Lemma 1. *Let*

$$G = \begin{bmatrix} 1 & 0 & 0'_K \\ 0 & 0 & 0'_K \\ 0_K & 0_K & \hat{V}_{11}^{-1} \end{bmatrix} \tag{34}$$

and

$$\tilde{A} = [\hat{\mu}_2,\ \hat{H}]'[\hat{\mu}_2,\ \hat{H}], \tag{35}$$

$$\breve{A} = [\hat{\mu}_2,\ \hat{H}]'\Sigma^{-1}[\hat{\mu}_2,\ \hat{H}], \tag{36}$$

$$\hat{A} = [\hat{\mu}_2,\ \hat{H}]'\hat{\Sigma}^{-1}[\hat{\mu}_2,\ \hat{H}]. \tag{37}$$

Suppose $x = [x_1,\ x_2']'$ is the eigenvector associated with the largest eigenvalue of $\tilde{A}^{-1}G$, where x_1 is the first element of x. Kim's OLS estimator is given by $\tilde{\gamma}^K = -x_2/x_1$. $\breve{\gamma}^K$ and $\hat{\gamma}^K$ are obtained by replacing $\tilde{A}$ with $\breve{A}$ and $\hat{A}$, respectively.

Lemma 1 provides an analytical solution to Kim's estimator which greatly facilitates its use, especially for the multi-factor case. Of particular interest is $\hat{\gamma}^K$, which we prove in the Appendix that it is numerically identical to the maximum likelihood estimator of γ under the normality assumption, so Lemma 1 can also be used to analytically compute the maximum likelihood estimator of γ.[6] While both the adjusted estimators of Litzenberger and Ramaswamy (1979) and Kim (1995) are N-consistent, there is little understanding of their finite sample properties. In the next section, we provide a surprising theoretical result which suggests that none of these adjusted estimators have integral moments when N and T are both finite.

[5] In Kim's implementation, he further assumes Σ is diagonal to accommodate a large number of test assets.

[6] Shanken and Zhou (2000) also provide the analytical solution to the maximum likelihood estimation of γ but our expression here is slightly simpler than theirs.

2. Finite Sample Analysis

2.1 Simulation Method

Before we discuss our simulation method, we first distinguish between the conditional and unconditional distributions of $\tilde{\gamma}$, $\check{\gamma}$, and $\hat{\gamma}$. By conditional distribution, we mean the distribution of the estimated γ conditional on the sample mean and variance of the K factors, i.e., $\hat{\mu}_1$ and $\hat{V}_{11}$. As the derivation of the unconditional distribution is based on the results from the conditional distribution, we provide the conditional distribution analysis first.

The issue at hand is how to simulate $\tilde{\gamma}$, $\check{\gamma}$, and $\hat{\gamma}$ conditional on a given value of $\hat{\mu}_1$ and $\hat{V}_{11}$. From (16)–(18), the conditional distributions of $\hat{\mu}_2$, $\hat{\beta}$ and $\hat{\Sigma}$ are known, eliminating the need to directly simulate T observations of Y_t to obtain $\tilde{\gamma}$, $\check{\gamma}$, and $\hat{\gamma}$.[7] While this approach is substantially faster than the traditional simulation method of simulating a time series of Y_t, it requires specification of a large number of parameters, namely, γ, β, and Σ. We would like to simplify the problem by reducing the number of random variables that are needed to construct $\tilde{\gamma}$, $\check{\gamma}$, and $\hat{\gamma}$. This simplification also allows us to understand the essential parameters that determine the distributions of $\tilde{\gamma}$, $\check{\gamma}$, and $\hat{\gamma}$.

2.1.1 OLS CSR

We turn our attention to the OLS case first. Apply the partitioned matrix inverse formula to $(\hat{H}'\hat{H})^{-1}$ and obtain

$$(\hat{H}'\hat{H})^{-1}\hat{H}' = \begin{bmatrix} \frac{1}{N}[1_N' - (1_N'\hat{\beta})(\hat{\beta}'M\hat{\beta})^{-1}\hat{\beta}'M] \\ (\hat{\beta}'M\hat{\beta})^{-1}\hat{\beta}'M \end{bmatrix} \equiv \begin{bmatrix} A_0 \\ A_1 \end{bmatrix}, \tag{38}$$

where $M = I_N - 1_N(1_N'1_N)^{-1}1_N'$. With this expression, write the OLS CSR estimate of γ_0 and γ_1 as $\tilde{\gamma}_0 = A_0\hat{\mu}_2$, and $\tilde{\gamma}_1 = A_1\hat{\mu}_2$. Using (16) and the fact that $\hat{\mu}_2$ is independent of $\hat{\beta}$, we can first simulate $\hat{\beta}$ and then simulate $\tilde{\gamma} = (\hat{H}'\hat{H})^{-1}\hat{H}'\hat{\mu}_2$ as

$$\begin{bmatrix} \tilde{\gamma}_0 \\ \tilde{\gamma}_1 \end{bmatrix} \sim N\left(\begin{bmatrix} \gamma_0 + A_0\beta\bar{\gamma}_1 \\ A_1\beta\bar{\gamma}_1 \end{bmatrix}, \frac{1}{T}\begin{bmatrix} A_0\Sigma A_0' & A_0\Sigma A_1' \\ A_1\Sigma A_0' & A_1\Sigma A_1' \end{bmatrix} \right). \tag{39}$$

Although the OLS CSR estimate of γ only involves $\hat{\mu}_2$ and $\hat{\beta}$, we still need to know Σ to simulate $\hat{\mu}_2$ and $\hat{\beta}$ if we use (16) and (17). To reduce the number of parameters required for the simulation,

[7] The conditional distributions in (16)–(18) only depend on the multivariate normality assumption of ϵ_t in (4). There is no need to assume that Y_t is jointly normal.

we perform a transformation on $\hat{\beta}$. Let $P\Lambda P'$ be the eigenvalue decomposition of $\Sigma^{\frac{1}{2}}M\Sigma^{\frac{1}{2}}$, where $\Lambda = \text{Diag}(\lambda_1, \ldots, \lambda_{N-1})$ with $\lambda_1 \geq \cdots \geq \lambda_{N-1} > 0$ being the $N-1$ nonzero eigenvalues of $\Sigma^{\frac{1}{2}}M\Sigma^{\frac{1}{2}}$, and P be an $N \times (N-1)$ matrix with its ith column equal to the eigenvector of $\Sigma^{\frac{1}{2}}M\Sigma^{\frac{1}{2}}$ associated with λ_i. Denote $\nu = \Sigma^{-\frac{1}{2}}1_N/(1_N'\Sigma^{-1}1_N)^{\frac{1}{2}}$. Then $\Sigma^{\frac{1}{2}}M\Sigma^{\frac{1}{2}}\nu = 0_N$, so ν is orthogonal to P and $Q = [\nu,\ P]$ is an orthonormal basis of $\mathbb{R}^N$, i.e., $QQ' = Q'Q = I_N$. We define Z as the following linear transformation of $\hat{\beta}$,

$$Z \equiv \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix} = \sqrt{T}Q'\Sigma^{-\frac{1}{2}}\hat{\beta}\hat{V}_{11}^{\frac{1}{2}} = \begin{bmatrix} \sqrt{T}\nu'\Sigma^{-\frac{1}{2}}\hat{\beta}\hat{V}_{11}^{\frac{1}{2}} \\ \sqrt{T}P'\Sigma^{-\frac{1}{2}}\hat{\beta}\hat{V}_{11}^{\frac{1}{2}} \end{bmatrix} \tag{40}$$

and its mean is

$$\mu_Z = \sqrt{T}Q'\Sigma^{-\frac{1}{2}}\beta\hat{V}_{11}^{\frac{1}{2}} = \begin{bmatrix} \sqrt{T(1_N'\Sigma^{-1}1_N)}h\hat{V}_{11}^{\frac{1}{2}} \\ \eta\hat{V}_{11}^{\frac{1}{2}} \end{bmatrix}, \tag{41}$$

where we define $h = (1_N'\Sigma^{-1}\beta)/(1_N'\Sigma^{-1}1_N)$ and $\eta = \sqrt{T}P'\Sigma^{-1}\beta$. We consider Z as a normalized version of $\hat{\beta}$ because conditional on $\hat{V}_{11}$, the distribution of Z is given by

$$\text{vec}(Z) \sim N\left(\text{vec}(\mu_Z), I_{NK}\right), \tag{42}$$

so all the elements of Z are independent of each other. Finally, defining

$$\delta \equiv \begin{bmatrix} \delta_1 \\ \delta_2 \end{bmatrix} = \frac{1}{\sqrt{T}N}Q'\Sigma^{\frac{1}{2}}1_N = \begin{bmatrix} \frac{1}{\sqrt{T(1_N'\Sigma^{-1}1_N)}} \\ \frac{1}{\sqrt{T}N}P'\Sigma^{\frac{1}{2}}1_N \end{bmatrix}, \tag{43}$$

we can write A_0 and A_1 as

$$\begin{aligned} A_0 &= \frac{1}{N}1_N'[I_N - \hat{\beta}(\hat{\beta}'M\hat{\beta})^{-1}\hat{\beta}'M] \\ &= \frac{1}{N}1_N'\Sigma^{\frac{1}{2}}QQ'\Sigma^{-\frac{1}{2}}[I_N - \hat{\beta}(\hat{\beta}'\Sigma^{-\frac{1}{2}}P\Lambda P'\Sigma^{-\frac{1}{2}}\hat{\beta})^{-1}\hat{\beta}'\Sigma^{-\frac{1}{2}}P\Lambda P'\Sigma^{-\frac{1}{2}}] \\ &= \sqrt{T}\delta'[Q' - Z(Z_2'\Lambda Z_2)^{-1}Z_2'\Lambda P']\Sigma^{-\frac{1}{2}}, \end{aligned} \tag{44}$$

$$\begin{aligned} A_1 &= (\hat{\beta}'M\hat{\beta})^{-1}\hat{\beta}'M \\ &= (\hat{\beta}'\Sigma^{-\frac{1}{2}}P\Lambda P'\Sigma^{-\frac{1}{2}}\hat{\beta})^{-1}\hat{\beta}'\Sigma^{-\frac{1}{2}}P\Lambda P'\Sigma^{-\frac{1}{2}} \\ &= \sqrt{T}\hat{V}_{11}^{\frac{1}{2}}(Z_2'\Lambda Z_2)^{-1}Z_2'\Lambda P'\Sigma^{-\frac{1}{2}}. \end{aligned} \tag{45}$$

Therefore, we can first simulate Z and then simulate $\tilde{\gamma}$ using the following normal distribution

$$\begin{bmatrix} \tilde{\gamma}_0 \\ \tilde{\gamma}_1 \end{bmatrix} \sim N\left(\begin{bmatrix} \gamma_0 + (h + \delta_2'\eta - \delta'ZD_1)\bar{\gamma}_1 \\ \hat{V}_{11}^{\frac{1}{2}}D_1\bar{\gamma}_1 \end{bmatrix}, \begin{bmatrix} \delta'\delta - 2\delta'ZD_2 + \delta'ZD_3Z'\delta & (D_2' - \delta'ZD_3)\hat{V}_{11}^{\frac{1}{2}} \\ \hat{V}_{11}^{\frac{1}{2}}(D_2 - D_3Z'\delta) & \hat{V}_{11}^{\frac{1}{2}}D_3\hat{V}_{11}^{\frac{1}{2}} \end{bmatrix}\right), \tag{46}$$

where $D_1 = (Z_2'\Lambda Z_2)^{-1}Z_2'\Lambda\eta$, $D_2 = (Z_2'\Lambda Z_2)^{-1}Z_2'\Lambda\delta_2$, and $D_3 = (Z_2'\Lambda Z_2)^{-1}(Z_2'\Lambda^2 Z_2)(Z_2'\Lambda Z_2)^{-1}$. Under this approach, we need to simulate Z instead of $\hat{\beta}$. Although the number of random variables remains the same, we need to know much less to simulate Z. Conditional on $\hat{\mu}_1$ and $\hat{V}_{11}$, we only need to know h, η, δ, γ_0, $\bar{\gamma}_1$, and Λ to simulate $\tilde{\gamma}$. In fact, multiplying Λ by a constant would not change the distribution, so the distribution of $\tilde{\gamma}$ only depends on λ_i/λ_{N-1} for $i = 1, \ldots, N-2$. Aside from N and T, the above analysis suggests that the conditional distribution of $\tilde{\gamma}$ can be written as a function of $N(K+2)+K-1$ parameters.

2.1.2 True GLS CSR

While the true GLS CSR estimate of γ depends on $\hat{\beta}$, $\hat{\mu}_2$, and Σ, it turns out that the distribution of $\check{\gamma}$ is easier to simulate than $\tilde{\gamma}$. Using the partitioned matrix inverse formula on $(\hat{H}'\Sigma^{-1}\hat{H})^{-1}$ and our earlier definition of Z and η, we can write

$$\begin{aligned}(\hat{H}'\Sigma^{-1}\hat{H})^{-1}\hat{H}'\Sigma^{-\frac{1}{2}} &= \left[\begin{array}{c}\frac{1}{1_N'\Sigma^{-1}1_N}[1_N'\Sigma^{-\frac{1}{2}} - 1_N'\Sigma^{-1}\hat{\beta}(\hat{\beta}'\Sigma^{-\frac{1}{2}}PP'\Sigma^{-\frac{1}{2}}\hat{\beta})^{-1}\hat{\beta}'\Sigma^{-\frac{1}{2}}PP'] \\ (\hat{\beta}'\Sigma^{-\frac{1}{2}}PP'\Sigma^{-\frac{1}{2}}\hat{\beta})^{-1}\hat{\beta}'\Sigma^{-\frac{1}{2}}PP'\end{array}\right] \\ &= \left[\begin{array}{c}\frac{1}{\sqrt{1_N'\Sigma^{-1}1_N}}[\nu' - Z_1(Z_2'Z_2)^{-1}Z_2'P'] \\ \sqrt{T}\hat{V}_{11}^{\frac{1}{2}}(Z_2'Z_2)^{-1}Z_2'P'\end{array}\right]. \end{aligned} \tag{47}$$

Therefore, in order to simulate $\check{\gamma} = (\hat{H}'\Sigma^{-1}\hat{H})^{-1}\hat{H}'\Sigma^{-\frac{1}{2}}\hat{\mu}_2$, we can first simulate Z and then simulate $\check{\gamma}$ using the following normal distribution

$$\left[\begin{array}{c}\check{\gamma}_0 \\ \check{\gamma}_1\end{array}\right] \sim N\left(\left[\begin{array}{c}\gamma_0 + h\bar{\gamma}_1 - \frac{Z_1(Z_2'Z_2)^{-1}Z_2'\eta}{\sqrt{T(1_N'\Sigma^{-1}1_N)}}\bar{\gamma}_1 \\ \hat{V}_{11}^{\frac{1}{2}}(Z_2'Z_2)^{-1}(Z_2'\eta)\bar{\gamma}_1\end{array}\right], \left[\begin{array}{cc}\frac{1+Z_1(Z_2'Z_2)^{-1}Z_1'}{T(1_N'\Sigma^{-1}1_N)} & -\frac{Z_1(Z_2'Z_2)^{-1}\hat{V}_{11}^{\frac{1}{2}}}{\sqrt{T(1_N'\Sigma^{-1}1_N)}} \\ -\frac{\hat{V}_{11}^{\frac{1}{2}}(Z_2'Z_2)^{-1}Z_1'}{\sqrt{T(1_N'\Sigma^{-1}1_N)}} & \hat{V}_{11}^{\frac{1}{2}}(Z_2'Z_2)^{-1}\hat{V}_{11}^{\frac{1}{2}}\end{array}\right]\right). \tag{48}$$

Comparing (48) with (46), the distribution of the true GLS $\check{\gamma}$ can be obtained as a special case of the distribution of OLS $\tilde{\gamma}$ by setting $\delta_2 = 0_{N-1}$ and $\Lambda = I_{N-1}$. While we can simulate $\check{\gamma}$ by simulating Z, we present here a much more efficient method for simulating $\check{\gamma}$. Let $\Theta = \hat{V}_{11}^{\frac{1}{2}}\eta'\eta\hat{V}_{11}^{\frac{1}{2}}$ and $[Q_1,\ Q_2]$ be an orthonormal basis of $\mathbb{R}^{N-1}$ with $Q_1 = \eta\hat{V}_{11}^{\frac{1}{2}}\Theta^{-\frac{1}{2}}$. Defining $Z_3 = Q_1'Z_2$ and $Z_4 = Q_2'Z_2$, it is easy to verify that Z_3 and Z_4 are independent of each other and that

$$\operatorname{vec}(Z_3) \sim N\left(\operatorname{vec}(\Theta^{\frac{1}{2}}), I_{K^2}\right), \tag{49}$$

$$\operatorname{vec}(Z_4) \sim N\left(0_{K(N-K-1)}, I_{K(N-K-1)}\right). \tag{50}$$

Writing $W_4 = Z_4'Z_4$, $W_4 \sim W_K(N-K-1, I_K)$ and it is independent of Z_3. Since $Q_1Q_1' + Q_2Q_2' = I_{N-1}$,

$$Z_2'Z_2 = Z_2'[Q_1,\ Q_2][Q_1,\ Q_2]'Z_2 = Z_3'Z_3 + Z_4'Z_4 = Z_3'Z_3 + W_4. \tag{51}$$

From (48), conditional on $\hat{\mu}_1$, $\hat{V}_{11}$, and Z_2, $\check{\gamma}_1$ is normally distributed and its mean and variance are given by

$$E[\check{\gamma}_1|\hat{\mu}_1, \hat{V}_{11}, Z_2] = \hat{V}_{11}^{\frac{1}{2}}(Z_2'Z_2)^{-1}(Z_2'\eta)\bar{\gamma}_1 = \hat{V}_{11}^{\frac{1}{2}}(Z_3'Z_3 + W_4)^{-1}Z_3'\Theta^{\frac{1}{2}}\hat{V}_{11}^{-\frac{1}{2}}\bar{\gamma}_1, \tag{52}$$

$$\mathrm{Var}[\check{\gamma}_1|\hat{\mu}_1, \hat{V}_{11}, Z_2] = \hat{V}_{11}^{\frac{1}{2}}(Z_2'Z_2)^{-1}\hat{V}_{11}^{\frac{1}{2}} = \hat{V}_{11}^{\frac{1}{2}}(Z_3'Z_3 + W_4)^{-1}\hat{V}_{11}^{\frac{1}{2}}, \tag{53}$$

which depend on Z_2 only through Z_3 and W_4. Therefore, conditional on $\hat{\mu}_1$, $\hat{V}_{11}$, Z_3, and W_4,

$$\check{\gamma}_1 \sim N\left(\hat{V}_{11}^{\frac{1}{2}}(Z_3'Z_3 + W_4)^{-1}Z_3'\Theta^{\frac{1}{2}}\hat{V}_{11}^{-\frac{1}{2}}\bar{\gamma}_1, \hat{V}_{11}^{\frac{1}{2}}(Z_3'Z_3 + W_4)^{-1}\hat{V}_{11}^{\frac{1}{2}}\right). \tag{54}$$

From (48), conditional on $\hat{\mu}_1$, $\hat{V}_{11}$, Z and $\check{\gamma}_1$, $\check{\gamma}_0$ is normal and its mean and variance are given by

$$E[\check{\gamma}_0|\hat{\mu}_1, \hat{V}_{11}, \check{\gamma}_1, Z] = \gamma_0 + h\bar{\gamma}_1 - \frac{Z_1}{\sqrt{T(1_N'\Sigma^{-1}1_N)}}\hat{V}_{11}^{-\frac{1}{2}}\check{\gamma}_1, \tag{55}$$

$$\mathrm{Var}[\check{\gamma}_0|\hat{\mu}_1, \hat{V}_{11}, \check{\gamma}_1, Z] = \frac{1}{T(1_N'\Sigma^{-1}1_N)}. \tag{56}$$

Note that the conditional mean of $\check{\gamma}_0$ only depends on $\hat{\mu}_1$, $\hat{V}_{11}$, $\check{\gamma}_1$ and Z_1, and the conditional variance of $\check{\gamma}_0$ is a constant. Therefore, conditional on only $\hat{\mu}_1$, $\hat{V}_{11}$ and $\check{\gamma}_1$,

$$\check{\gamma}_0 \sim N\left(\gamma_0 + h(\bar{\gamma}_1 - \check{\gamma}_1), (1 + \check{\gamma}_1'\hat{V}_{11}^{-1}\check{\gamma}_1)/(T1_N'\Sigma^{-1}1_N)\right). \tag{57}$$

In summary, we suggest the following steps to simulate the conditional distribution of $\check{\gamma}$:

1. Simulate Z_3 using $\mathrm{vec}(Z_3) \sim N(\mathrm{vec}(\Theta^{\frac{1}{2}}), I_{K^2})$ and $W_4 \sim W_K(N-K-1, I_K)$, independently of each other.

2. Simulate $\check{\gamma}_1$ using (54).

3. Simulate $\check{\gamma}_0$ using (57).

Note that this approach requires us to simulate Z_3, W_4, and a $(K+1)$-dimensional normal distribution to generate a realization of $\check{\gamma}$. Since Z_3 and W_4 are K-dimensional random matrices, the computation time only depends on K but not N. As the number of factors K is typically small but the number of test assets N can be large, our approach provides an extremely efficient method of simulating the conditional distribution of $\check{\gamma}$.

This simulation method also allows us to understand the essential parameters that determine the conditional distribution of $\check{\gamma}_1$ and $\check{\gamma}_0$. Conditional on $\hat{\mu}_1$ and $\hat{V}_{11}$, from (54) and (57), the distribution of $\check{\gamma}_1$ only depends on Θ and $\bar{\gamma}_1$ but the distribution of $\check{\gamma}_0$ also depends on γ_0, $1_N'\Sigma^{-1}1_N$, and $1_N'\Sigma^{-1}\beta$. As Θ is symmetric, the conditional distribution of $\check{\gamma}$ only depends on $(K+1)(K+4)/2$ parameters, which are far fewer than in the OLS case.

2.1.3 Estimated GLS CSR

At first sight, the estimated GLS CSR appears to be much more complicated than the true GLS CSR because one needs to simulate $\hat{\Sigma}$ in order to compute $\hat{\gamma}$. However, the analysis provided here shows that simulating $\hat{\gamma}$ does not require much more effort than simulating $\check{\gamma}$, and there is no need to simulate $\hat{\Sigma}$. To prepare for our derivation, we define $\check{A}$ and $\hat{A}$ as in (36) and (37), respectively. The only difference between $\hat{A}$ and $\check{A}$ is that $\hat{A}$ has $\hat{\Sigma}$ in the middle but $\check{A}$ has the true Σ in the middle. Using Theorem 3.2.11 of Muirhead (1982), conditional on $\hat{\mu}_2$ and $\hat{\beta}$,

$$\hat{A}^{-1} \sim W_{K+2}(T-N+1, \check{A}^{-1}/T). \tag{58}$$

Our first task is to express $\check{\gamma}$ and $\hat{\gamma}$ as elements of $\check{A}^{-1}$ and $\hat{A}^{-1}$. Partition $\hat{A}$ into 2 by 2 blocks with dimension 1 and $K+1$, respectively. Denote $\hat{A}_{ij}$ as the (i,j)th block of $\hat{A}$ and $\hat{A}^{ij}$ the (i,j)th block of $\hat{A}^{-1}$. $\check{A}_{ij}$ and $\check{A}^{ij}$ are similarly defined for $\check{A}$ and $\check{A}^{-1}$, respectively. From the partitioned matrix inverse formula, it is easy to verify that

$$\begin{aligned} \check{\gamma} &= \check{A}_{22}^{-1}\check{A}_{21} = -\check{A}^{21}(\check{A}^{11})^{-1}, \\ \hat{\gamma} &= \hat{A}_{22}^{-1}\hat{A}_{21} = -\hat{A}^{21}(\hat{A}^{11})^{-1}. \end{aligned} \tag{59}$$

Conditional on $\hat{\mu}_2$ and $\hat{\beta}$, from (58),

$$\hat{A}^{11} \sim W_1(T-N+1, \check{A}^{11}/T). \tag{60}$$

It follows that

$$U = \frac{T\hat{A}^{11}}{\check{A}^{11}} \sim \chi^2_{T-N+1}, \tag{61}$$

and U is independent of $\hat{\mu}_2$ and $\hat{\beta}$, and hence independent of $\check{A}^{11}$. Therefore, we have

$$(\hat{A}^{11})^{-1} = \frac{T(\check{A}^{11})^{-1}}{U} = \frac{T(\check{A}_{11} - \check{A}_{12}\check{A}_{22}^{-1}\check{A}_{21})}{U}. \tag{62}$$

Conditional on $\hat{\mu}_2$, $\hat{\beta}$, and $\hat{A}^{11}$, from Theorem 3.2.10 of Muirhead (1982),

$$\hat{A}^{21} \sim N(\check{A}^{21}(\check{A}^{11})^{-1}\hat{A}^{11}, \check{A}_{22}^{-1}\hat{A}^{11}/T), \tag{63}$$

and hence

$$\hat{\gamma} = -\hat{A}^{21}(\hat{A}^{11})^{-1} \sim N(\check{\gamma}, (\hat{A}^{11})^{-1}\check{A}_{22}^{-1}/T). \tag{64}$$

Conditional on $\hat{\mu}_2$, $\hat{\beta}$, and $\hat{A}_{11}$, or equivalently conditional on $\hat{\mu}_2$, $\hat{\beta}$, and U, we can now use (62) to obtain the conditional distribution of $\hat{\gamma}$ as

$$\hat{\gamma} \sim N\left(\check{\gamma}, \frac{(\check{A}_{11} - \check{A}_{12}\check{A}_{22}^{-1}\check{A}_{21})\check{A}_{22}^{-1}}{U}\right). \tag{65}$$

With some algebra,

$$\begin{aligned} & (\check{A}_{11} - \check{A}_{12}\check{A}_{22}^{-1}\check{A}_{21})\check{A}_{22}^{-1} \\ = \;& T\hat{\mu}_2'\Sigma^{-\frac{1}{2}}P[I_{N-1} - Z_2(Z_2'Z_2)^{-1}Z_2']P'\Sigma^{-\frac{1}{2}}\hat{\mu}_2 \begin{bmatrix} \frac{1+Z_1(Z_2'Z_2)^{-1}Z_1'}{T(1_N'\Sigma^{-1}1_N)} & -\frac{Z_1(Z_2'Z_2)^{-1}\hat{V}_{11}^{\frac{1}{2}}}{\sqrt{T(1_N'\Sigma^{-1}1_N)}} \\ -\frac{\hat{V}_{11}^{\frac{1}{2}}(Z_2'Z_2)^{-1}Z_1'}{\sqrt{T(1_N'\Sigma^{-1}1_N)}} & \hat{V}_{11}^{\frac{1}{2}}(Z_2'Z_2)^{-1}\hat{V}_{11}^{\frac{1}{2}} \end{bmatrix}. \end{aligned} \tag{66}$$

Conditional on Z_2, $U_1 \equiv T\hat{\mu}_2'\Sigma^{-\frac{1}{2}}P[I_{N-1} - Z_2(Z_2'Z_2)^{-1}Z_2']P'\Sigma^{-\frac{1}{2}}\hat{\mu}_2 \sim \chi^2_{N-K-1}(\omega)$, with its non-centrality parameter ω given by

$$\begin{aligned} \omega &= \bar{\gamma}_1'\eta'[I_{N-1} - Z_2(Z_2'Z_2)^{-1}Z_2']\eta\bar{\gamma}_1 \\ &= \bar{\gamma}_1'\hat{V}_{11}^{-\frac{1}{2}}\Theta^{\frac{1}{2}}[I_K - Z_3(Z_3'Z_3 + W_4)^{-1}Z_3']\Theta^{\frac{1}{2}}\hat{V}_{11}^{-\frac{1}{2}}\bar{\gamma}_1 \\ &= \bar{\gamma}_1'\hat{V}_{11}^{-\frac{1}{2}}\Theta^{\frac{1}{2}}(I_K + Z_3W_4^{-1}Z_3')^{-1}\Theta^{\frac{1}{2}}\hat{V}_{11}^{-\frac{1}{2}}\bar{\gamma}_1, \end{aligned} \tag{67}$$

which only depends on Z_2 through Z_3 and W_4.

Note that conditional on $\hat{\mu}_1$, $\hat{V}_{11}$ and Z, U_1 is independent of $\check{\gamma}$. This is because from (47), $\check{\gamma}$ only depends on $\hat{\mu}_2$ only through $\nu'\Sigma^{-\frac{1}{2}}\hat{\mu}_2$ and $Z_2'P'\Sigma^{-\frac{1}{2}}\hat{\mu}_2$, and both are independent of U_1. Therefore, conditional on Z, U and U_1 and using the distribution result in (48), we have

$$\begin{bmatrix} \hat{\gamma}_0 \\ \hat{\gamma}_1 \end{bmatrix} \sim N\left(\begin{bmatrix} \gamma_0 + h\bar{\gamma}_1 - \frac{Z_1(Z_2'Z_2)^{-1}Z_2'\eta}{\sqrt{T(1_N'\Sigma^{-1}1_N)}}\bar{\gamma}_1 \\ \hat{V}_{11}^{\frac{1}{2}}(Z_2'Z_2)^{-1}(Z_2'\eta)\bar{\gamma}_1 \end{bmatrix}, \left(1 + \frac{U_1}{U}\right)\begin{bmatrix} \frac{1+Z_1(Z_2'Z_2)^{-1}Z_1'}{T(1_N'\Sigma^{-1}1_N)} & -\frac{Z_1(Z_2'Z_2)^{-1}\hat{V}_{11}^{\frac{1}{2}}}{\sqrt{T(1_N'\Sigma^{-1}1_N)}} \\ -\frac{\hat{V}_{11}^{\frac{1}{2}}(Z_2'Z_2)^{-1}Z_1'}{\sqrt{T(1_N'\Sigma^{-1}1_N)}} & \hat{V}_{11}^{\frac{1}{2}}(Z_2'Z_2)^{-1}\hat{V}_{11}^{\frac{1}{2}} \end{bmatrix}\right). \tag{68}$$

Note that the conditional distribution of $\hat{\gamma}$ from the estimated GLS only differs from the conditional distribution of $\check{\gamma}$ from the true GLS in that the variance of $\hat{\gamma}$ is larger than the variance of $\check{\gamma}$ by a factor of $1 + \frac{U_1}{U}$. With these results and following the same analysis for simulating $\check{\gamma}$, we can simulate the conditional distribution of $\hat{\gamma}$ using the following steps:

1. Simulate Z_3 using $\text{vec}(Z_3) \sim N(\text{vec}(\Theta^{\frac{1}{2}}), I_{K^2})$ and $W_4 \sim W_K(N-K-1, I_K)$, independently of each other.

2. Simulate $U \sim \chi^2_{T-N+1}$ and $U_1 \sim \chi^2_{N-K-1}(\omega)$ independently of each other, where ω is given by (67).

3. Simulate $\hat{\gamma}_1$ using

$$\hat{\gamma}_1 \sim N\left(\hat{V}_{11}^{\frac{1}{2}}(Z_3'Z_3 + W_4)^{-1}Z_3'\Theta^{\frac{1}{2}}\hat{V}_{11}^{-\frac{1}{2}}\bar{\gamma}_1, \left(1 + \frac{U_1}{U}\right)\hat{V}_{11}^{\frac{1}{2}}(Z_3'Z_3 + W_4)^{-1}\hat{V}_{11}^{\frac{1}{2}}\right). \tag{69}$$

4. Simulate $\hat{\gamma}_0$ using

$$\hat{\gamma}_0 \sim N\left(\gamma_0 + h(\bar{\gamma}_1 - \hat{\gamma}_1), \frac{1 + \hat{\gamma}_1'\hat{V}_{11}^{-1}\hat{\gamma}_1 + \frac{U_1}{U}}{T(1_N'\Sigma^{-1}1_N)}\right). \tag{70}$$

Note that in comparison to the simulation of $\check{\gamma}$, we only need to generate two additional chi-squared random variables U and U_1 to simulate $\hat{\gamma}$. Therefore, our approach of simulating $\hat{\gamma}$ is very efficient. In addition, we do not need to know any additional parameters (like Σ) to simulate $\hat{\gamma}$, which suggests that the conditional distribution of $\hat{\gamma}$ depends on the same $(K+1)(K+4)/2$ parameters that determine the conditional distribution of $\check{\gamma}$.

Besides providing a speedy simulation method, our analysis of the estimated GLS CSR also has the added benefit of relating the mean and variance of $\hat{\gamma}$ to those of $\check{\gamma}$, as given in the following lemma.

Lemma 2. *Assuming $T > N + K$, conditional on $\hat{\mu}_1$ and $\hat{V}_{11}$, the means and variances of $\hat{\gamma}$ from the estimated GLS and $\check{\gamma}$ from true GLS, when they exist, are related to each other by the following relationship*

$$E[\hat{\gamma}|\hat{\mu}_1, \hat{V}_{11}] = E[\check{\gamma}|\hat{\mu}_1, \hat{V}_{11}], \tag{71}$$

$$\text{Var}[\hat{\gamma}|\hat{\mu}_1, \hat{V}_{11}] = \text{Var}[\check{\gamma}|\hat{\mu}_1, \hat{V}_{11}] + \frac{1}{T-N-1}E[(\check{A}_{11} - \check{A}_{12}\check{A}_{22}^{-1}\check{A}_{21})\check{A}_{22}^{-1}|\hat{\mu}_1, \hat{V}_{11}]. \tag{72}$$

Unconditionally,

$$E[\hat{\gamma}] = E[\check{\gamma}], \tag{73}$$

$$\mathrm{Var}[\hat{\gamma}] = \mathrm{Var}[\check{\gamma}] + \frac{1}{T-N-1} E[(\check{A}_{11} - \check{A}_{12}\check{A}_{22}^{-1}\check{A}_{21})\check{A}_{22}^{-1}]. \tag{74}$$

This lemma suggests that the expected value, and hence the bias of $\hat{\gamma}$ from the estimated GLS CSR, is exactly the same as its true GLS CSR counterpart. Therefore, to find the bias of $\hat{\gamma}$, one can simply use the corresponding results from true GLS, which is easier to derive. However, the variance of $\hat{\gamma}$ is larger than that of $\check{\gamma}$, so using the estimated $\hat{\Sigma}$ instead of the true Σ introduces additional volatility into the estimated γ, especially when N is large relative to T. Therefore, while $\check{\gamma}$ from the true GLS CSR is more efficient than $\tilde{\gamma}$ from the OLS CSR, there is no guarantee that $\hat{\gamma}$ from the estimated GLS CSR is more efficient than $\tilde{\gamma}$ in finite samples, particularly when N is large relative to T.

2.1.4 Extensions

So far we have focused our discussions on simulating the conditional distributions of $\tilde{\gamma}$, $\check{\gamma}$, and $\hat{\gamma}$. To simulate their unconditional distributions, we only need to simulate $\hat{\mu}_1$ and $\hat{V}_{11}$ before drawing Y, Z, and U. Under the normality assumption,

$$\hat{\mu}_1 \sim N(\mu_1, V_{11}/T), \tag{75}$$

$$T\hat{V}_{11} \sim W_K(T-1, V_{11}), \tag{76}$$

and they are independent of each other, so simulating the unconditional distribution of the estimated γ is relatively easy. In fact, our simulation approach can be used even when f_t is not normally distributed. As long as one can simulate $\hat{\mu}_1$ and $\hat{V}_{11}$, and ϵ_t in (4) is i.i.d. normal when conditional on $\hat{\mu}_1$ and $\hat{V}_{11}$, our method can still be used to simulate the unconditional distribution of the estimated γ.

Our simulation method can also be extended to the situation in which the beta used in the second-pass CSR is estimated from a period that is different from that of the realized return, which is often the case in the Fama-MacBeth regression. Suppose the first T periods are used to estimate β, but the second-pass CSR is run using realized returns of period t, where $t > T$. The OLS CSR estimate of γ at time t is

$$\tilde{\gamma}_t = (\hat{H}'\hat{H})^{-1}(\hat{H}'R_t). \tag{77}$$

Comparing (77) to (8), the only difference is that R_t is used instead of $\hat{\mu}_2$ as the dependent variable. As R_t and $\hat{\mu}_2$ are both independent of $\hat{\beta}$, simulating $\tilde{\gamma}_t$ requires very little modification of our simulation approach. Conditional on f_t (instead of $\hat{\mu}_1$ as before),

$$R_t \sim N(1_N\gamma_0 + \beta\bar{\gamma}_{1t}, \Sigma), \tag{78}$$

where $\bar{\gamma}_{1t} = \gamma_1 - \mu_1 + f_t$ and is independent of $\hat{\beta}$. Compared with (16), the only change that we need to make in our simulation method is to replace $\bar{\gamma}_1$ by $\bar{\gamma}_{1t}$ and multiply the variance term by T in (46). For true GLS and estimated GLS, we can make the same modification to obtain the distribution of $\check{\gamma}_t$ and $\hat{\gamma}_t$. Note that in our original setup, T refers to the length of the time series used to estimate β as well as to the length of the times series used to compute the average return $\hat{\mu}_2$. Under the setting for the Fama-MacBeth regression that we discuss here, T is only used to denote the length of the beta estimation period. The dependent variable here is no longer the average return over the beta estimation period but the realized return in a different period. The Fama-MacBeth CSR can be repeated for many periods to obtain a time series of $\tilde{\gamma}_t$, but the length of the time series of $\tilde{\gamma}_t$ has no relation to the length of the beta estimation period (T).

2.2 Moments of CSR Estimates of Zero-Beta Rate and Risk Premia

2.2.1 Existence of Moments

Asymptotically, $\tilde{\gamma}$, $\check{\gamma}$, and $\hat{\gamma}$ have a normal distribution according to (26)–(28), so all the moments of the estimated γ exist in the asymptotic distribution. However, in finite samples, only a finite number of the moments of $\tilde{\gamma}$, $\check{\gamma}$, and $\hat{\gamma}$ exist. The following proposition presents this result, which appears to be largely unknown in the finance literature.

Proposition 1. *Conditional on $\hat{\mu}_1$ and $\hat{V}_{11}$, the s-th moment of the second-pass OLS and true GLS CSR estimators of γ exists if and only if $s < N - K$. For the estimated GLS CSR, the s-th moment of $\hat{\gamma}$ exists if and only if $s < \min[N - K, T - N + 1]$.*

Proposition 1 suggests that the conditional s-th moment of the estimated γ does not exist if $s \geq N - K$, so the unconditional s-th moment of the estimated γ also does not exist if $s \geq N - K$. Proposition 1 provides the first clue that the asymptotic distribution can be problematic for finite sample inference. For example, when just $N = 2$ assets are used to estimate the CAPM (i.e.,

$K = 1$), we can estimate γ, but none of its moments exist (because its distribution has heavy tails). When the CAPM is estimated using $N = 3$ assets, the mean of the estimated γ exists, but its variance does not. In general, we expect that when K approaches N, the finite sample distribution of the estimated γ becomes less and less normal. Note that for the OLS and the true GLS cases, Proposition 1 suggests that the existence of moments depends on N and K but not on T. If $s \geq N - K$ and the s-th moment of the estimated γ does not exist, then having a longer time series does not help. Therefore, the traditional practice of using the normal or the t-distribution with $T - 1$ degrees of freedom to make inference about γ can be problematic even for large T.

Contrary to the CSR estimators which have at least some finite moments, the following proposition suggests that the adjusted estimators of Litzenberger and Ramaswamy (1979) and Kim (1995) (which includes the maximum likelihood estimator as a special case) actually have no integral moments in finite samples.

Proposition 2. *Conditional on $\hat{\mu}_1$ and $\hat{V}_{11}$, the s-th moment of the adjusted estimators $\tilde{\gamma}^{LR}$, $\check{\gamma}^{LR}$, $\hat{\gamma}^{LR}$, $\tilde{\gamma}^{K}$, $\check{\gamma}^{K}$, and $\hat{\gamma}^{K}$ does not exist for $s \geq 1$ when N and T are finite.*

The result in Proposition 2 is actually quite general and it is not limited to the normality case. It suggests that the mean and variance of these adjusted estimators do not exist, so if one uses the mean squared error criterion to compare estimators, these adjusted estimators are definitely inferior to the unadjusted CSR estimators. This result is quite ironic because these adjusted estimators were developed to reduce the bias of the CSR estimators but their means do not exist.

Intuitively, the adjusted estimators do not have integral moments because their distribution has heavy tails. The heavy tails arise because these adjusted estimators can all be written as a ratio of two functions of $\hat{\mu}_2$ and $\hat{\beta}$, and there is a set of points of $\hat{\mu}_2$ and $\hat{\beta}$ such that the numerator is nonzero but the denominator is zero. When there is a sufficient probability in the neighborhood of this set of points, the tail is so fat that the mean of the adjusted estimators fails to exist. The consequence of the nonexistence of moments is that occasionally there will be some extreme outliers from these adjusted estimators that will render these adjusted estimators unreliable.[8]

[8]While the nonexistence of moments of these adjusted estimators was never proven before in the literature, researchers had noted the outlier problems associated with these adjusted estimators. For example, Banz (1981) suggests a serious problem in applying Litzenberger and Ramaswamy estimator is that the unbiased matrix estimate is not guaranteed to be positive definite for a given sample, and Shanken and Weinstein (1990) encounter this problem

Shanken (1992) shows that the GLS CSR estimator $\hat{\gamma}$ has the same asymptotic distribution as the maximum likelihood estimator (which is the same as $\hat{\gamma}^K$), yet Propositions 1 and 2 suggest that their finite sample properties are vastly different, with the latter having no integral moments. While these results are somewhat surprising and unknown in the finance literature, they can be anticipated to some extent from various studies of simultaneous equations models in the econometrics literature. For the estimation of simultaneous equations models, there have been extensive studies on the relative merits of limited information maximum likelihood (LIML) method and two stage least squares (2SLS). Although these two methods are asymptotically equivalent, they have very different finite sample properties. It was shown in the econometrics literature that the distribution of the LIML estimator has tails similar to those of a multivariate Cauchy distribution, which has no integral moments, whereas the 2SLS estimator has tails similar to those of a multivariate t-distribution, which has some finite integral moments.[9] In our context, Kim's adjusted estimator $\hat{\gamma}^K$ is analogous to the LIML estimator and the two-pass GLS CSR estimator $\hat{\gamma}$ is analogous to the 2SLS estimator. As a result, it is not entirely surprising that some of the integral moments of $\hat{\gamma}$ exist but no integral moments of $\hat{\gamma}^K$ exist.

It should be emphasized that outlier behavior is only one characteristic of the finite sample distribution of an estimator, so one should not dismiss the usefulness of the adjusted estimators of Litzenberger and Ramaswamy (1979) and Kim (1995) simply because they do not have integral moments. While having no finite first moment, the adjusted estimators of γ may actually approach their asymptotic distributions more rapidly than the unadjusted ones. Nevertheless, one needs to be careful in interpreting the simulation results of these adjusted estimators. It is because the adjusted estimators have no integral moments, so the sample mean and variance from the simulations cannot be used to approximate the true mean and variance.

In the previous subsection, we showed that one can approximate the conditional distribution of the estimated γ by simulating first Z and then a $(K+1)$-dimensional normal random variable (plus two chi-squared random variables for the estimated GLS). However, if one is only interested in the first and the second moments of the estimated γ, then one only needs to simulate Z_2. The rest of this subsection discusses the conditional and unconditional mean and variance of the estimated γ.

in a multi-factor context. Amsler and Schmidt (1985) report that the maximum likelihood estimator was "occasionally off by spectacularly large amounts" in simulations.

[9] See Anderson (1982) and Phillips (1983) for a review of this literature. Phillips (1984, 1985) provide the exact density function of the LIML estimators.

For notational brevity, we use E^c and Var^c to denote mean and variance when conditional on $\hat{\mu}_1$ and $\hat{V}_{11}$, i.e., $E^c[X] \equiv E[X|\hat{\mu}_1, \hat{V}_{11}]$ and $\text{Var}^c[X] \equiv \text{Var}[X|\hat{\mu}_1, \hat{V}_{11}]$.

2.2.2 Mean

From Proposition 1, the conditional mean of $\tilde{\gamma}$, $\check{\gamma}$, and $\hat{\gamma}$ exists when $N > K+1$.[10] From Lemma 2, the conditional mean of $\hat{\gamma}$ and $\check{\gamma}$ are the same, so we only need to present the results for $\tilde{\gamma}$ and $\check{\gamma}$ here. We start with the OLS CSR estimate of γ. Conditional on $\hat{\mu}_1$ and $\hat{V}_{11}$, Y_2 and Z_2 are independent, so from (46),

$$E^c[\tilde{\gamma}_1] = \hat{V}_{11}^{\frac{1}{2}} E^c[(Z_2'\Lambda Z_2)^{-1} Z_2'\Lambda\eta]\bar{\gamma}_1. \tag{79}$$

This expression suggests that conditional on $\hat{\mu}_1$ and $\hat{V}_{11}$, the expected value of $\tilde{\gamma}_1$ depends on η, Λ and $\bar{\gamma}_1$. In order to obtain the conditional mean of $\tilde{\gamma}_1$, we need to evaluate $E^c[(Z_2'\Lambda Z_2)^{-1} Z_2'\Lambda\eta]$. For the case of $K = 1$, this expectation can be evaluated directly and we present the results in the next section. For $K > 1$, no simple expression for $E^c[(Z_2'\Lambda Z_2)^{-1} Z_2'\Lambda\eta]$ is available, but a large number of Z_2 can be drawn to approximate the expectation by using the average value of $(Z_2'\Lambda Z_2)^{-1} Z_2'\Lambda\eta$.

From (46), the conditional mean of $\tilde{\gamma}_0$ is

$$E^c[\tilde{\gamma}_0] = \gamma_0 + h(\bar{\gamma}_1 - E^c[\tilde{\gamma}_1]) + \delta_2'\eta\bar{\gamma}_1 - E^c[\delta_2' Z_2 (Z_2'\Lambda Z_2)^{-1} Z_2'\Lambda\eta]\bar{\gamma}_1. \tag{80}$$

Aside from η, Λ, and $\bar{\gamma}_1$, the conditional mean of $\tilde{\gamma}_0$ also depends on γ_0, δ_2, and h. In order to obtain $E^c[\tilde{\gamma}_0]$, we can use the average value of $\delta_2' Z_2 (Z_2'\Lambda Z_2)^{-1} Z_2'\Lambda\eta$ from the simulations to approximate $E^c[\delta_2' Z_2 (Z_2'\Lambda Z_2)^{-1} Z_2'\Lambda\eta]$.

Taking the unconditional expectation of both sides of (79) and (80),

$$E[\tilde{\gamma}_1] = E[\hat{V}_{11}^{\frac{1}{2}} (Z_2'\Lambda Z_2)^{-1} Z_2'\Lambda\eta]\gamma_1, \tag{81}$$

$$E[\tilde{\gamma}_0] = \gamma_0 + h(\gamma_1 - E[\tilde{\gamma}_1]) + \delta_2'\eta\gamma_1 - E[\delta_2' Z_2 (Z_2'\Lambda Z_2)^{-1} Z_2'\Lambda\eta]\gamma_1. \tag{82}$$

The only difference between obtaining the conditional and unconditional mean of $\tilde{\gamma}$ is that in addition to Z_2, we also need to simulate $\hat{V}_{11}$ for the unconditional mean. In each simulation, we first simulate $\hat{V}_{11}$ using (76) and then Z_2 using (42). We can then approximate $E[\hat{V}_{11}^{\frac{1}{2}} (Z_2'\Lambda Z_2)^{-1} Z_2'\Lambda\eta]$ using the average value of $\hat{V}_{11}^{\frac{1}{2}} (Z_2'\Lambda Z_2)^{-1} Z_2'\Lambda\eta$ from the simulations.

[10] For the GLS CSR, we also need $T > N$ for the mean of $\hat{\gamma}$ to exist, but this condition is automatically satisfied because we need $T > N + K$ for the estimated GLS to be feasible.

The mean of $\check{\gamma}$ and $\check{\gamma}$ is easy to obtain. We simply set $\delta_2 = 0_N$ and $\Lambda = I_{N-1}$ in the expressions for the OLS case. For the conditional mean,

$$E^c[\hat{\gamma}_1] = E^c[\check{\gamma}_1] = \hat{V}_{11}^{\frac{1}{2}} E^c[(Z_2'Z_2)^{-1}Z_2'\eta]\bar{\gamma}_1 = \hat{V}_{11}^{\frac{1}{2}} E^c[(Z_3'Z_3 + W_4)^{-1}Z_3']\Theta^{\frac{1}{2}}\hat{V}_{11}^{-\frac{1}{2}}\bar{\gamma}_1, \tag{83}$$

$$E^c[\hat{\gamma}_0] = E^c[\check{\gamma}_0] = \gamma_0 + h(\bar{\gamma}_1 - E^c[\check{\gamma}_1]). \tag{84}$$

Therefore, the conditional mean of $\check{\gamma}_1$ and $\hat{\gamma}_1$ only depends on Θ and $\bar{\gamma}_1$, whereas the conditional mean of $\check{\gamma}_0$ and $\hat{\gamma}_0$ depends on Θ, γ_0, $\bar{\gamma}_1$ and $(1_N'\Sigma^{-1}\beta)/(1_N'\Sigma^{-1}1_N)$. In addition, we only need to simulate Z_3 and W_4 to approximate $E^c[\check{\gamma}_1]$, which is much faster than simulating Z_2 when N is large. Unconditionally,

$$E[\hat{\gamma}_1] = E[\check{\gamma}_1] = E[\hat{V}_{11}^{\frac{1}{2}}(Z_3'Z_3 + W_4)^{-1}Z_3'\Theta^{\frac{1}{2}}\hat{V}_{11}^{-\frac{1}{2}}]\gamma_1, \tag{85}$$

$$E[\hat{\gamma}_0] = E[\check{\gamma}_0] = \gamma_0 + h(\gamma_1 - E[\check{\gamma}_1]). \tag{86}$$

From the expressions above, the unconditional biases of both OLS and GLS estimates of γ_0 and γ_1 depend on the value of γ_1 but not the value of γ_0, so the actual value of the zero-beta rate is irrelevant in determining the bias of the estimated γ. For the special case that $\gamma_1 = 0_K$ (i.e., expected returns are constant across assets), the unconditional biases for both OLS and GLS estimates of γ are zero.

The expressions for the unconditional mean of $\hat{\gamma}$ and $\check{\gamma}$ that we derived above assume that the dependent variable in the second-pass CSR is $\hat{\mu}_2$. However, the unconditional mean remains the same if R_t is used as the dependent variable, where $t > T$ (i.e., returns of the test assets fall outside of the beta estimation period). This is because R_t (with $t > T$) and $\hat{\mu}_2$ are both independent of $\hat{\beta}$ and have the same unconditional mean $\mu_2 = 1_N\gamma_0 + \beta\gamma_1$. Therefore, the expectation of the estimated γ that we derive here can also be used for the case of the Fama-MacBeth regression.

2.2.3 Variance

From Proposition 1, the conditional second moment of $\tilde{\gamma}$, $\check{\gamma}$, and $\hat{\gamma}$ exists when $N > K+2$. Starting with the OLS case, we use (46) to obtain the conditional second moment of $\tilde{\gamma}_1$ as

$$E^c[\tilde{\gamma}_1\tilde{\gamma}_1'] = \hat{V}_{11}^{\frac{1}{2}} E^c[D_3 + D_1\bar{\gamma}_1\bar{\gamma}_1'D_1']\hat{V}_{11}^{\frac{1}{2}}, \tag{87}$$

where D_1 and D_3 are defined right after (46). With (87) and (79), the conditional variance of $\tilde{\gamma}_1$ is

$$\text{Var}^c[\tilde{\gamma}_1] = E^c[\tilde{\gamma}_1\tilde{\gamma}_1'] - E^c[\tilde{\gamma}_1]E^c[\tilde{\gamma}_1]'. \tag{88}$$

After some straightforward but tedious algebra, the conditional variance of $\tilde{\gamma}_0$ is

$$\begin{aligned}\mathrm{Var}^c[\tilde{\gamma}_0] &= \mathrm{Var}^c[D_4\hat{V}_{11}^{-\frac{1}{2}}\tilde{\gamma}_1] + \frac{E^c[\tilde{\gamma}_1'\hat{V}_{11}^{-1}\tilde{\gamma}_1]}{T(1_N'\Sigma^{-1}1_N)} + \delta'\delta - 2E^c[D_4D_2] \\ &= \mathrm{Var}^c[D_4D_1\bar{\gamma}_1] + E^c[D_4D_3D_4'] + \frac{E^c[\mathrm{tr}(D_3 + D_1\bar{\gamma}_1\bar{\gamma}_1'D_1')]}{T(1_N'\Sigma^{-1}1_N)} + \delta'\delta - 2E^c[D_4D_2], \quad (89)\end{aligned}$$

where D_2 is defined right after (46) and $D_4 = h\hat{V}_{11}^{\frac{1}{2}} + \delta_2'Z_2$. For the unconditional variances of $\tilde{\gamma}_1$ and $\tilde{\gamma}_0$, use the fact that $\hat{\mu}_1$ and $\hat{V}_{11}$ are independent and that

$$E[\bar{\gamma}_1\bar{\gamma}_1'|\hat{V}_{11}] = E[\bar{\gamma}_1\bar{\gamma}_1'] = \gamma_1\gamma_1' + \frac{V_{11}}{T} \quad (90)$$

to obtain

$$\mathrm{Var}[\tilde{\gamma}_1] = E\left[\hat{V}_{11}^{\frac{1}{2}}(D_3 + D_1\left(\gamma_1\gamma_1' + \frac{V_{11}}{T}\right)D_1')\hat{V}_{11}^{\frac{1}{2}}\right] - E[\tilde{\gamma}_1]E[\tilde{\gamma}_1]', \quad (91)$$

$$\begin{aligned}\mathrm{Var}[\tilde{\gamma}_0] &= \mathrm{Var}[D_4D_1\gamma_1] + E[dV_{11}d'/T + D_4D_3D_4'] \\ &\quad + \frac{E[\mathrm{tr}(D_3 + D_1\left(\gamma_1\gamma_1' + \frac{V_{11}}{T}\right)D_1')]}{T(1_N'\Sigma^{-1}1_N)} + \delta'\delta - 2E[D_4D_2], \quad (92)\end{aligned}$$

where $d = h + \delta_2'\eta - D_4D_1$. For the true GLS case, the conditional variances of $\check{\gamma}_1$ and $\check{\gamma}_0$ are obtained by setting $\delta_2 = 0_{N-1}$ and $\Lambda = I_{N-1}$ in the OLS case, and

$$\mathrm{Var}^c[\check{\gamma}_1] = \hat{V}_{11}^{\frac{1}{2}}E^c[D_5 + D_6\bar{\gamma}_1\bar{\gamma}_1'D_6']\hat{V}_{11}^{\frac{1}{2}} - E^c[\check{\gamma}_1]E^c[\check{\gamma}_1]', \quad (93)$$

$$\mathrm{Var}^c[\check{\gamma}_0] = h\mathrm{Var}^c[\check{\gamma}_1]h' + \frac{1 + E^c[\mathrm{tr}(D_5 + D_6\bar{\gamma}_1\bar{\gamma}_1'D_6')]}{T(1_N'\Sigma^{-1}1_N)}, \quad (94)$$

where $D_5 = (Z_2'Z_2)^{-1} = (Z_3'Z_3 + W_4)^{-1}$ and $D_6 = (Z_2'Z_2)^{-1}(Z_2'\eta) = (Z_3'Z_3 + W_4)^{-1}Z_3'\Theta^{\frac{1}{2}}\hat{V}_{11}^{-\frac{1}{2}}$. The unconditional variances of $\check{\gamma}_1$ and $\check{\gamma}_0$ are obtained similarly to be

$$\mathrm{Var}[\check{\gamma}_1] = E\left[\hat{V}_{11}^{\frac{1}{2}}(D_5 + D_6\left(\gamma_1\gamma_1' + \frac{V_{11}}{T}\right)D_6')\hat{V}_{11}^{\frac{1}{2}}\right] - E[\check{\gamma}_1]E[\check{\gamma}_1]', \quad (95)$$

$$\begin{aligned}\mathrm{Var}[\check{\gamma}_0] &= h\mathrm{Var}[\check{\gamma}_1]h' + \frac{1 + E[\mathrm{tr}(D_5 + D_6\left(\gamma_1\gamma_1' + \frac{V_{11}}{T}\right)D_6')]}{T(1_N'\Sigma^{-1}1_N)} \\ &\quad + \frac{hV_{11}h' - 2hE[\hat{V}_{11}^{\frac{1}{2}}D_6]V_{11}h'}{T}. \quad (96)\end{aligned}$$

Unlike the mean of $\hat{\gamma}$, the variance of $\hat{\gamma}$ from the estimated GLS is not the same as the variance of $\check{\gamma}$ from the true GLS. From Lemma 2 and (66),

$$\mathrm{Var}^c[\hat{\gamma}_1] = \mathrm{Var}^c[\check{\gamma}_1] + \frac{\hat{V}_{11}^{\frac{1}{2}}E^c[U_1D_5]\hat{V}_{11}^{\frac{1}{2}}}{T-N-1} = \mathrm{Var}^c[\check{\gamma}_1] + \Delta^c, \quad (97)$$

where

$$\Delta^c = \frac{E^c[(N-K-1+\omega)\hat{V}_{11}^{\frac{1}{2}} D_5 \hat{V}_{11}^{\frac{1}{2}}]}{T-N-1}, \tag{98}$$

with ω defined in (67). The last equality follows because, conditional on $\hat{\mu}_1$, $\hat{V}_{11}$ and Z_2, $U_1 \sim \chi^2_{N-K-1}(\omega)$ and its expected value is $N-K-1+\omega$. Similarly, the conditional variance of $\hat{\gamma}_0$ is given by

$$\begin{aligned}
\mathrm{Var}^c[\hat{\gamma}_0] &= \mathrm{Var}^c[\check{\gamma}_0] + \frac{E^c[U_1(1+Z_1 D_5 Z_1')]}{T(T-N-1)(1_N'\Sigma^{-1}1_N)} \\
&= \mathrm{Var}^c[\check{\gamma}_0] + \frac{E^c[(N-K-1+\omega)(1+T(1_N'\Sigma^{-1}1_N)hD_5h' + \mathrm{tr}(D_5))]}{T(T-N-1)(1_N'\Sigma^{-1}1_N)} \\
&= \mathrm{Var}^c[\check{\gamma}_0] + h\Delta^c h' + \frac{E^c[(N-K-1+\omega)(1+\mathrm{tr}(D_5))]}{T(T-N-1)(1_N'\Sigma^{-1}1_N)},
\end{aligned} \tag{99}$$

with the second equality following from the identity $E[x'Ax] = \mu_x' A \mu_x + \mathrm{tr}(AV_x)$, where x is a vector of random variables with mean μ_x and covariance matrix V_x.

Unconditionally, from Lemma 2,

$$\mathrm{Var}[\hat{\gamma}_1] = \mathrm{Var}[\check{\gamma}_1] + \Delta, \tag{100}$$

where $\Delta = E[\Delta^c]$ and

$$\mathrm{Var}[\hat{\gamma}_0] = \mathrm{Var}[\check{\gamma}_0] + h\Delta h' + \frac{E[((N-K-1) + \gamma_1' C \gamma_1 + \mathrm{tr}(CV_{11}/T))(1+\mathrm{tr}(D_5))]}{T(T-N-1)(1_N'\Sigma^{-1}1_N)}, \tag{101}$$

with $C = \eta'[I_{N-1} - Z_2(Z_2'Z_2)^{-1}Z_2']\eta = \hat{V}_{11}^{-\frac{1}{2}}\Theta^{\frac{1}{2}}(I_K + Z_3 W_4^{-1} Z_3')^{-1}\Theta^{\frac{1}{2}}\hat{V}_{11}^{-\frac{1}{2}}$. Note that as in the case of the mean, the expressions for the variance given in this subsection are all written as functions of expectations of some functions of Z_2 and $\hat{V}_{11}$. In order to approximate these expectations, we just need to simulate Z_2 and $\hat{V}_{11}$ and use the average simulated values of these functions to approximate their expectations. For the GLS case, we do not even need to simulate Z_2, since simply simulating Z_3 and W_4 will suffice.

3. Analytical Results for the Single Factor Case

3.1 Finite Sample Distribution of Estimated Risk Premia

In the previous section, we suggested that one can simulate the conditional distribution of $\tilde{\gamma}$ by simulating some normal random variables Z_1 and Z_2. The conditional first and second moments of $\tilde{\gamma}$ can be approximated by simulating only Z_2 (an $(N-1)\times K$ normal random variable). For unconditional first and second moments, $\hat{V}_{11}$ must also be simulated. For the GLS case, $\check{\gamma}$ and $\hat{\gamma}$ can be simulated without simulating Z_2. Only simulation of some K-dimensional random vectors and matrices is required to simulate $\check{\gamma}$ and $\hat{\gamma}$. While this approach is much faster than the traditional approach of simulating data on the returns and the factors, we would ideally like to evaluate the moments of the estimated γ without doing a simulation. In this section, we present the analytical results for the single factor case,[11] a special case albeit one of great importance in the finance literature. The capital asset pricing model (CAPM) and the consumption capital asset pricing model (CCAPM) are both single factor models, making it worthwhile to understand the behavior of the estimated γ for the single factor case.

In fact, when $K=1$, we can even evaluate the exact conditional distribution of the estimated risk premium for the OLS and the true GLS CSR. In order to obtain the conditional distribution of $\tilde{\gamma}_1$ of the OLS CSR, we need to evaluate the conditional distribution of $(Z_2'\Lambda Z_2)^{-1}Z_2'\Lambda Y_2$, where $Y_2=\sqrt{T}P'\Sigma^{-\frac{1}{2}}\hat{\mu}_2$. When $K=1$, Z_2 is just an $(N-1)$-vector of normal random variables. Let $X=[Y_2',\ Z_2']'$. Then conditional on $\hat{\mu}_1$ and $\hat{V}_{11}$, $X\sim N(\mu_X, I_{2N-2})$, and

$$\mu_X=\begin{bmatrix}\eta\bar{\gamma}_1\\ \eta\hat{V}_{11}^{\frac{1}{2}}\end{bmatrix}. \tag{102}$$

Defining

$$A = \begin{bmatrix}\mathrm{O}_{(N-1)\times(N-1)} & \Lambda/2\\ \Lambda/2 & \mathrm{O}_{(N-1)\times(N-1)}\end{bmatrix}, \tag{103}$$

$$B = \begin{bmatrix}\mathrm{O}_{(N-1)\times(N-1)} & \mathrm{O}_{(N-1)\times(N-1)}\\ \mathrm{O}_{(N-1)\times(N-1)} & \Lambda\end{bmatrix}, \tag{104}$$

it is easy to see that $(X'AX)/(X'BX)=(Z_2'\Lambda Y_2)/(Z_2'\Lambda Z_2)$, and from (45), we can write

$$\tilde{\gamma}_1=A_1\hat{\mu}_2=\hat{V}_{11}^{\frac{1}{2}}\left(\frac{X'AX}{X'BX}\right), \tag{105}$$

and

$$P[\tilde{\gamma}_1>c|\hat{\mu}_1,\hat{V}_{11}]=P\left[\frac{X'AX}{X'BX}>\frac{c}{\hat{V}_{11}^{\frac{1}{2}}}\middle|\hat{\mu}_1,\hat{V}_{11}\right]=P\left[X'\left(A-\frac{c}{\hat{V}_{11}^{\frac{1}{2}}}B\right)X>0\middle|\hat{\mu}_1,\hat{V}_{11}\right]. \tag{106}$$

[11] Although we can obtain analytical expressions for the moments of the estimated γ for the multi-factor case, the expressions involve zonal polynomials and they are difficult to evaluate. Therefore, we recommend using the simulation approach to obtain the moments of the estimated γ for the multi-factor case.

Let GDG' be the eigenvalue decomposition of $A - c\hat{V}_{11}^{-\frac{1}{2}}B$, where D is an n-dimensional ($n \leq 2N-2$) diagonal matrix of the nonzero eigenvalues of $A - c\hat{V}_{11}^{-\frac{1}{2}}B$ and G is an $(2N-2)\times n$ matrix with its columns being the corresponding eigenvectors. Defining $x = G'X$, we can then write

$$P[\tilde{\gamma}_1 > c|\hat{\mu}_1, \hat{V}_{11}] = P[x'Dx > 0|\hat{\mu}_1, \hat{V}_{11}]. \tag{107}$$

Conditional on $\hat{\mu}_1$ and $\hat{V}_{11}$, $x \sim N(G'\mu_X, I_n)$, so $x'Dx$ is just a linear combination of n independent χ^2_1 random variables, and the probability can be evaluated using the numerical procedure suggested by Imhof (1961).[12] Similarly, to obtain the conditional distribution of $\check{\gamma}_1$ from the true GLS, simply set $\Lambda = I_{N-1}$ in the above procedure.

3.2 Mean of Estimated Risk Premium

It turns out that for the single factor case, the conditional and unconditional means of the estimated γ can be written as 1-dimensional integrals. In order to conserve space, we present only the results for the estimated risk premium in the paper. The results for the zero-beta rate are less important and they are reported in a separate technical appendix, which is available upon request. Starting with the case of the OLS CSR, we present the conditional mean of $\tilde{\gamma}_1$ in the following proposition.

Proposition 3. *For the single factor case, the conditional mean of the second-pass OLS CSR estimators of γ_1 exists for $N \geq 3$ and it is given by*

$$E^c[\tilde{\gamma}_1] = \left[\frac{\hat{V}_{11}}{2}\int_0^1 g_1 \left(\prod_{i=1}^{N-1} a_i\right)^{\frac{1}{2}} e^{\frac{\hat{V}_{11}}{2}\sum_{i=1}^{N-1}\eta_i^2(a_i y - 1)} y^{\frac{N-3}{2}} \mathrm{d}y\right]\bar{\gamma}_1, \tag{108}$$

where

$$a_i = \frac{1}{\lambda_i^* - (\lambda_i^* - 1)y}, \tag{109}$$

η_i is the i-th element of $\eta = \sqrt{T}P'\Sigma^{-\frac{1}{2}}\beta$, $\lambda_i^* = \lambda_i/\lambda_{N-1}$, and $g_1 = \sum_{i=1}^{N-1} a_i\lambda_i^*\eta_i^2$.

Although (108) looks complex, it is only a 1-dimensional integral and can be easily evaluated when η and Λ are known.

[12] See Davies (1980), Farebrother (1990), Ansley, Kohn, and Shively (1992), and Lu and King (2002) for improvements to and implementation issues with this numerical method. Forchini (2001) provides an explicit expression for the cumulative density function of $X'AX/(X'BX)$.

For the true and the estimated GLS CSR, the conditional mean of the estimated γ_1 is obtained by substituting $\lambda_i^* = 1$ in the OLS case. The resulting expressions are very simple and are summarized in the following proposition.

Proposition 4. *For the single factor case, the conditional mean of the second pass GLS CSR estimators of γ_1 exists for $N \geq 3$ and it is*

$$E^c[\hat{\gamma}_1] = E^c[\check{\gamma}_1] = \check{\kappa}\bar{\gamma}_1, \tag{110}$$

where

$$\check{\kappa} = \frac{\theta}{2}\int_0^1 e^{\frac{\theta(y-1)}{2}} y^{\frac{N-3}{2}} \mathrm{d}y \tag{111}$$

and $\theta = \hat{V}_{11}\eta'\eta$.

In fact, numerical integration is not needed to obtain the conditional expected value of the estimated γ_1 from the second-pass GLS CSR. The following lemma presents a simplification formula.

Lemma 3. *Suppose b is a positive scalar and n is a nonnegative integer. Then*

$$b\int_0^1 e^{b(y-1)} y^{\frac{n}{2}} \mathrm{d}y = \sum_{r=0}^{\frac{n}{2}} \frac{\left(\frac{n}{2} - r + 1\right)_r}{(-b)^r} - \frac{\left(\frac{n}{2}\right)! e^{-b}}{(-b)^{\frac{n}{2}}} \tag{112}$$

for even n, where $(a)_r = a(a+1)\cdots(a+r-1)$ with $(a)_0 \equiv 1$, and

$$b\int_0^1 e^{b(y-1)} y^{\frac{n}{2}} \mathrm{d}y = \sum_{r=0}^{\frac{n-1}{2}} \frac{\left(\frac{n}{2} - r + 1\right)_r}{(-b)^r} - \frac{\left(\frac{3}{2}\right)_{(n-1)/2}}{(-b)^{\frac{n-1}{2}}} \frac{\mathcal{D}(\sqrt{b})}{\sqrt{b}} \tag{113}$$

for odd n, where

$$\mathcal{D}(x) = e^{-x^2}\int_0^x e^{t^2} \mathrm{d}t \tag{114}$$

is Dawson's integral, which is readily available in many mathematical programs.

As $\tilde{\kappa}$ depends on $\hat{V}_{11}$ but not $\hat{\mu}_1$, we also have

$$E[\check{\gamma}_1 | \hat{V}_{11}] = \check{\kappa}\gamma_1, \tag{115}$$

and we can think of $\check{\kappa} - 1$ as the percentage bias of $\check{\gamma}_1$ conditional on $\hat{V}_{11}$. Note that $\check{\kappa}$ is only a function of θ and N, so these two parameters jointly determine the conditional percentage bias of $\check{\gamma}_1$. The following lemma describes some properties of $\check{\kappa}$.

Lemma 4. *Conditional on $\hat{V}_{11}$, $\breve{\kappa}$ is an increasing function of θ and a decreasing function of N, and $0 < \breve{\kappa} < 1$. As θ approaches infinity, $\breve{\kappa}$ approaches one.*

Lemma 4 suggests that $\tilde{\gamma}_1$ (and also $\hat{\gamma}_1$) are biased toward zero, and the magnitude of the percentage bias is an increasing function of N and a decreasing function of θ. To understand what θ represents, write the GLS CSR of $\hat{\mu}_2$ on 1_N and β as an OLS CSR of

$$\Sigma^{-\frac{1}{2}}\hat{\mu}_2 = \Sigma^{-\frac{1}{2}}1_N\gamma_0 + \Sigma^{-\frac{1}{2}}\beta\gamma_1 + e, \tag{116}$$

where e is an N-vector of error terms. Premultiplying both sides by P' and noting that $P'\Sigma^{-\frac{1}{2}}1_N = 0_{N-1}$,

$$P'\Sigma^{-\frac{1}{2}}\hat{\mu}_2 = P'\Sigma^{-\frac{1}{2}}\beta\gamma_1 + \varepsilon, \tag{117}$$

where $\varepsilon = P'e$. Let $y = P'\Sigma^{-\frac{1}{2}}\hat{\mu}_2$ and $x = P'\Sigma^{-\frac{1}{2}}\beta$. Then we can think of the true GLS CSR estimate of γ_1 as being obtained by running the following OLS regression

$$y_i = \gamma_1 x_i + \varepsilon_i, \qquad i = 1, \ldots, N-1 \tag{118}$$

when β is known. Of course, we do not use the true β but the estimated β in the CSR. Define $x_i^* = P'\Sigma^{-\frac{1}{2}}\hat{\beta} = x_i + n_i$, where n_i is the measurement error of x_i^*. Therefore, the regression that is run is actually

$$y_i = \gamma_1 x_i^* + \varepsilon_i^* = \gamma_1(x_i + n_i) + \varepsilon_i^*, \qquad i = 1, \ldots, N-1. \tag{119}$$

The resulting estimate of γ_1 from this OLS regression is

$$\tilde{\gamma}_1 = \frac{\sum_{i=1}^{N-1} x_i^* y_i}{\sum_{i=1}^{N-1}(x_i^*)^2}. \tag{120}$$

Note that this is the classical EIV problem and the bias depends on the ratio of $\sum_{i=1}^{N-1} x_i^2/(N-1) = \eta'\eta/[T(N-1)]$ (the signal) to $\mathrm{Var}[n_i]$ (the noise). Note that conditional on $\hat{V}_{11}$,

$$n = P'\Sigma^{-\frac{1}{2}}\hat{\beta} - P'\Sigma^{-\frac{1}{2}}\beta \sim N(0_{N-1}, (T\hat{V}_{11})^{-1}I_{N-1}), \tag{121}$$

so the n_i's are independent of each other and their variance is $(T\hat{V}_{11})^{-1}$. With this analysis, $\hat{V}_{11}\eta'\eta/(N-1) = \theta/(N-1)$ is a measure of signal-to-noise ratio in the estimated betas, which explains why the percentage bias of $\tilde{\gamma}_1$ is a decreasing function of θ for a fixed N. This result is largely consistent with the traditional EIV analysis in the regression framework, which suggests that when the independent variable is measured with error, the estimated slope coefficient in the regression is biased toward zero and the bias depends on the signal-to-noise ratio of the independent variable. The only difference is that the traditional EIV analysis provides only asymptotic results when $N \to \infty$ and we provide exact finite sample results here. Lemma 4 suggests that there are two ways to reduce the bias of $\breve{\gamma}_1$: one is to increase the length of the time series, the other is to use test assets that have a wide dispersion in β. As for the effect of the number of test assets, from Lemma 4, $\breve{\kappa}$ is a decreasing function of N for a fixed value of $\hat{V}_{11}\eta'\eta$. However, the effect of increasing N on the bias is not clear because $\eta'\eta$ also typically increases with N. If we reasonably assume that $\eta'\eta/(N-1)$ is a constant for different choices of N, then we can find out whether $\breve{\kappa}$ is an increasing function of N. In Figure 1, we plot $\breve{\kappa}$ as a function of $\hat{V}_{11}\eta'\eta/(N-1)$ for $N = 5$, 10, 25, and 100 over the range $0 \le \hat{V}_{11}\eta'\eta/(N-1) \le 10$ (which covers the range of $\hat{V}_{11}\eta'\eta/(N-1)$ that we encounter in typical applications). As Figure 1 shows, if $\eta'\eta/(N-1)$ is constant across different choices of N, then the bias is an increasing function of N, but the difference in bias between different choices of N is quite small.[13] This also suggests that while $\breve{\kappa}$ is a function of $\hat{V}_{11}\eta'\eta$ and N, the bias of the GLS estimate of γ_1 is mostly determined by the signal-to-noise ratio $\hat{V}_{11}\eta'\eta/(N-1)$.

Figure 1 about here

The analysis of the bias of $\tilde{\gamma}_1$, however, is far more complicated. Denote $\tilde{\kappa}$ as

$$\tilde{\kappa} = \frac{\hat{V}_{11}}{2}\int_0^1 \left(\sum_{i=1}^{N-1} a_i\lambda_i^*\eta_i^2\right)\left(\prod_{i=1}^{N-1} a_i\right)^{\frac{1}{2}} e^{\frac{\hat{V}_{11}}{2}\sum_{i=1}^{N-1}\eta_i^2(a_i y-1)} y^{\frac{N-3}{2}} \mathrm{d}y. \tag{122}$$

Similar to the case of $\breve{\kappa}$, $E[\tilde{\gamma}_1|\hat{V}_{11}] = \tilde{\kappa}\gamma_1$ and we can interpret $\tilde{\kappa}-1$ as the percentage bias of $\tilde{\gamma}_1$ conditional on $\hat{V}_{11}$. Unlike the case of $\breve{\kappa}$, which depends only on N and $\hat{V}_{11}\eta'\eta$, the case of $\tilde{\kappa}$ depends on λ_i^*, η_i^2, N, and $\hat{V}_{11}$, so the individual elements of Λ and η are important in determining the bias of $\tilde{\gamma}_1$. It is important to note that $\tilde{\kappa}_1$ is not bounded above by one, and $\tilde{\gamma}_1$ is not necessarily biased toward zero. In fact, for some choices of λ_i and η_i, $\tilde{\kappa} > 1$. The intuition that EIV cause the slope coefficient to be biased toward zero does not apply here because unlike in the case of GLS CSR, the measurement errors of the independent variable ($\hat{\beta}$) are in general correlated with each other

[13] In practice, there is another advantage of using smaller N in the two-pass CSR. This advantage is that when a few well diversified portfolios are used instead of a large number of individual stocks as test assets, the variance of the residuals of the test assets (Σ) is typically smaller and hence $\eta'\eta/(N-1)$ is larger, which further reduces the bias.

in the OLS CSR. In general, depending on the values of λ_i and η_i, the bias of $\tilde{\gamma}_1$ can be more or less than that of $\check{\gamma}_1$. The following lemma compares $\tilde{\kappa}$ and $\check{\kappa}$ for two extreme cases.

Lemma 5. *If $\eta_1 \neq 0$ and $\eta_2 = \cdots = \eta_{N-1} = 0$, then $\tilde{\kappa} \geq \check{\kappa}$. If $\eta_1 = \cdots = \eta_{N-2} = 0$ and $\eta_{N-1} \neq 0$, then $\tilde{\kappa} \leq \check{\kappa}$. The equalities hold if and only if $\lambda_1 = \lambda_{N-1}$.*

Heuristically, Lemma 5 suggests that when η_1^2 is large and η_2^2 to η_{N-1}^2 are small (i.e., when $\Sigma^{-\frac{1}{2}}\beta$ is close to being proportional to the eigenvector associated with the largest eigenvalue of $\Sigma^{\frac{1}{2}}M\Sigma^{\frac{1}{2}}$),[14] the bias of the GLS CSR estimate of γ_1 is more severe than the bias of the OLS CSR estimate of γ_1. On the contrary, if η_{N-1}^2 is large but η_1^2 to η_{N-2}^2 are small (i.e., when $\Sigma^{-\frac{1}{2}}\beta$ is close to being proportional to the eigenvector associated with the smallest nonzero eigenvalue of $\Sigma^{\frac{1}{2}}M\Sigma^{\frac{1}{2}}$), then $0 < \tilde{\kappa} \leq \check{\kappa} < 1$ and the bias of the OLS CSR estimate of γ_1 is more severe than the bias of the GLS CSR estimate of γ_1. Since there is no theoretical relation between β and Σ, it is not clear whether we should expect the OLS or the GLS CSR estimate of γ_1 to have a larger bias. It is also important to note that the bias of the OLS CSR estimate of γ_1 is not invariant to repackaging of the original N test assets. If we construct N new portfolios from the N original test assets, the bias of the resulting new estimate of $\tilde{\gamma}_1$ is in general different from that of the old estimate obtained using the original N assets. This is not the case for $\check{\gamma}$ and $\hat{\gamma}$, which are invariant to portfolio repackaging of the original N test assets.[15]

So far we have discussed the conditional means of $\tilde{\gamma}_1$, $\check{\gamma}_1$, and $\hat{\gamma}_1$. The unconditional means can be obtained by using the facts that $T\hat{V}_{11}/V_{11} \sim \chi^2_{T-1}$ and is independent of $\hat{\mu}_1$. In order to facilitate our presentation of the unconditional results, we define two 1-dimensional integrals. The first integral is

$$\varphi_{m,n}(g) = \int_0^1 g(y)\left(\prod_{i=1}^{N-1} a_i\right)^{\frac{1}{2}} \frac{y^{\frac{m}{2}}}{\left[1 + \frac{V_{11}}{T}\sum_{i=1}^{N-1}\eta_i^2(1-a_i y)\right]^{\frac{n}{2}}}\mathrm{d}y, \tag{123}$$

where a_i is defined as in (109), g is a function of y, and $m > -2$. The second integral is defined as

$$\phi_m = \int_0^1 \frac{y^{\frac{m}{2}}}{\left[1 + \frac{V_{11}}{T}\eta'\eta(1-y)\right]^{\frac{T+1}{2}}}\mathrm{d}y. \tag{124}$$

[14]Since $\eta = \sqrt{T}P'\Sigma^{-\frac{1}{2}}\beta$, we have $\eta_i = \sqrt{T}P_i'\Sigma^{-\frac{1}{2}}\beta$, where P_i is the eigenvector associated with the i-th largest eigenvalue of $\Sigma^{\frac{1}{2}}M\Sigma^{\frac{1}{2}}$.

[15]See Kandel and Stambaugh (1995) for a discussion of the invariance property of the GLS CSR estimator. Their analysis is based on the population measures of μ_2 and β, but it can be easily generalized to show that $\check{\gamma}_1$ and $\hat{\gamma}_1$ are invariant to portfolio repackaging.

In fact, ϕ_m is a special case of $\varphi_{m,n}(g)$, with $g(y) = 1$, $a_i = 1$ and $n = T + 1$. Note that ϕ_m can also be written as

$$\phi_m = {}_2F_1\left(\frac{m+2}{2}, \frac{T+1}{2}, \frac{m+4}{2}, \frac{V_{11}\eta'\eta}{T + V_{11}\eta'\eta}\right), \tag{125}$$

where

$$ {}_2F_1(a, b, c, x) = \sum_{r=1}^{\infty} \frac{(a)_r (b)_r}{(c)_r} \frac{x^r}{r!} \tag{126}$$

is the hypergeometric function, which is readily available in many mathematical programs. Using these two 1-dimensional integrals, the following proposition provides the expression for the unconditional mean of the estimated γ_1 from the second-pass CSR.

Proposition 5. *For the single factor case, the unconditional mean of the second pass OLS CSR estimator of* γ_1 *exists for* $N \geq 3$ *and it is* $E[\tilde{\gamma}_1] = \tilde{\kappa}_u \gamma_1$ *where*

$$\tilde{\kappa}_u = \frac{(T-1)V_{11}}{2T} \varphi_{N-3,T+1}(g_1), \tag{127}$$

with g_1 *defined as in Proposition 3. In addition, the unconditional mean of the second-pass GLS CSR estimators of* γ_1 *exists for* $N \geq 3$ *and it is* $E[\hat{\gamma}_1] = E[\check{\gamma}_1] = \check{\kappa}_u \gamma_1$ *where*

$$\check{\kappa}_u = \frac{(T-1)V_{11}\eta'\eta}{2T} \phi_{N-3}. \tag{128}$$

From Proposition 5, the unconditional percentage bias of GLS $\check{\gamma}_1$ depends on $V_{11}\eta'\eta$, N, and T. As $0 < \check{\kappa} < 1$, $\check{\kappa}_u = E[\check{\kappa}]$ is also bounded by 0 and 1. The unconditional percentage bias of OLS $\tilde{\gamma}_1$ depends on V_{11}, η_i^2, λ_i^*, N, and T. Just as in the conditional case, $\tilde{\kappa}_u$ can be greater than or less than $\check{\kappa}_u$ depending on the value of the parameters λ_i^* and η_i. Without knowing the value of these parameters, it is not clear whether the GLS or the OLS CSR estimate of γ_1 has more bias unconditionally.

3.3 Variance of Estimated Risk Premium

Similarly to the conditional and unconditional mean, the conditional and unconditional variance of the estimated γ_1 can also be written as 1-dimensional integrals for the single factor case. To conserve space, we only present the results for the unconditional variance here but the expressions for the conditional variance of the estimated γ_1 can be easily obtained from the proof in the Appendix.

Starting with the case of the OLS CSR, the unconditional variance of $\tilde{\gamma}_1$ is presented in the following proposition.

Proposition 6. *For the single factor case, the unconditional variance of the second-pass OLS CSR estimator of γ_1 exists for $N \geq 4$. The unconditional variance of $\tilde{\gamma}_1$ is*

$$\mathrm{Var}[\tilde{\gamma}_1] = \frac{(T-1)(T+1)V_{11}^2}{4T^2}\varphi^d_{N-3,T+3}(g_2 + ag_1^2) + \frac{(T-1)V_{11}}{4T}\varphi^d_{N-5,T+1}(g_3 + ag_4) - \tilde{\kappa}_u^2\gamma_1^2, \quad (129)$$

where $\varphi^d_{m,n}(g) = \varphi_{m,n}(g) - \varphi_{m+2,n}(g)$, $g_1 = \sum_{i=1}^{N-1} a_i\lambda_i^*\eta_i^2$, $g_2 = \sum_{i=1}^{N-1} a_i^2\lambda_i^{*2}\eta_i^2$, $g_3 = \sum_{i=1}^{N-1} a_i\lambda_i^{*2}$, $g_4 = \sum_{i=1}^{N-1} a_i\lambda_i^{*2}\eta_i^2$, $a = \gamma_1^2 + \frac{V_{11}}{T}$ with a_i, η_i and λ_i^* defined in Proposition 3, and $\tilde{\kappa}_u$ is defined as in (127).

For the true GLS CSR, the variance of $\check{\gamma}$ can be obtained by setting $\lambda_i^* = 1$ in the expression for the OLS case. After some simplification, the results are given in the following proposition.

Proposition 7. *For the single factor case, the unconditional variance of the second-pass true GLS CSR estimator of γ_1 exists for $N \geq 4$. The unconditional variance of $\check{\gamma}_1$ is*

$$\mathrm{Var}[\check{\gamma}_1] = \frac{(T-1)V_{11}}{4T}\left[(N-2)a\eta'\eta\phi_{N-3} + [2-(N-4)a\eta'\eta]\phi_{N-5}\right] - \check{\kappa}_u^2\gamma_1^2, \quad (130)$$

where $a = \gamma_1^2 + \frac{V_{11}}{T}$ and $\check{\kappa}_u$ is defined as in (128).

Finally, the next proposition presents the unconditional variance of the estimated γ_1 from the estimated GLS CSR.

Proposition 8. *For the single factor case, the unconditional variance of the second-pass estimator of γ_1 from the estimated GLS CSR exists for $N \geq 4$. The unconditional variance of $\hat{\gamma}_1$ is*

$$\mathrm{Var}[\hat{\gamma}_1] = \mathrm{Var}[\check{\gamma}_1] + \Delta, \quad (131)$$

where

$$\Delta = \frac{(N-2)(T-1)V_{11}}{4T(T-N-1)}[(a\eta'\eta + 2)\phi_{N-5} - a\eta'\eta\phi_{N-3}], \quad (132)$$

and $a = \gamma_1^2 + \frac{V_{11}}{T}$.

4. Bias-adjusted Estimator of Risk Premium

Since the second-pass CSR estimators of γ_1 are biased in finite samples, we would like to correct this bias. In this section, we present a bias-adjusted version of the two-pass CSR estimators of the risk premium. The bias-adjusted version of the zero-beta rate is presented in a separate technical appendix. We focus our discussion on the single factor case because we have an analytical solution of the finite sample bias. Our method, however, can be extended to the multi-factor case provided that the simulation method is used to approximate the finite sample bias. For the GLS CSR, we only focus on the estimated GLS case as it is typically more relevant than the true GLS case.

If the value of $\breve{\kappa}$ in (111) and $\tilde{\kappa}$ in (122) are known, then we can construct the following adjusted OLS estimators and estimated GLS estimators of γ_1

$$\tilde{\gamma}_1^a = \tilde{\gamma}_1/\tilde{\kappa}, \tag{133}$$

$$\hat{\gamma}_1^a = \hat{\gamma}_1/\breve{\kappa}. \tag{134}$$

As $E[\tilde{\gamma}_1|\hat{V}_{11}] = \tilde{\kappa}\gamma_1$ and $E[\hat{\gamma}_1|\hat{V}_{11}] = \breve{\kappa}\gamma_1$, the adjusted estimators are unbiased conditional on $\hat{V}_{11}$, and hence are also unconditionally unbiased.

In practice, $\tilde{\kappa}$ and $\breve{\kappa}$ are in general unknown, so they must be estimated. In addition, the evaluation of $\tilde{\kappa}$ involves numerical integration, which makes the adjusted estimators somewhat difficult to use. We therefore propose the following approximation of $\tilde{\kappa}$ and $\breve{\kappa}$.[16]

Lemma 6. *Using a Taylor series expansion, $\tilde{\kappa}$ and $\breve{\kappa}$ can be approximated by*

$$\tilde{\kappa}^a = \frac{1}{(\mathrm{tr}(\Lambda)+\theta_1)^3}\left(\theta_1[(\mathrm{tr}(\Lambda)+\theta_1)^2+2\theta_2]+2[\mathrm{tr}(\Lambda^2)\theta_1-\mathrm{tr}(\Lambda)\theta_2]\right), \tag{135}$$

$$\breve{\kappa}^a = \frac{\theta(b^2+2\theta)}{b^3}, \tag{136}$$

where $\theta_1 = \hat{V}_{11}\eta'\Lambda\eta$, $\theta_2 = \hat{V}_{11}\eta'\Lambda^2\eta$, $\theta = \hat{V}_{11}\eta'\eta$, and $b = N-1+\theta$.

To evaluate $\tilde{\kappa}^a$, η and Λ must be estimated. It is natural to estimate them with their sample estimates $\hat{\eta} = \sqrt{T}\hat{P}'\hat{\Sigma}^{-\frac{1}{2}}\hat{\beta}$ and $\hat{\Lambda}$, where $\hat{P}\hat{\Lambda}\hat{P}'$ is the eigenvalue decomposition of $\hat{\Sigma}^{\frac{1}{2}}M\hat{\Sigma}^{\frac{1}{2}}$, with $\hat{\Lambda} = \mathrm{Diag}(\hat{\lambda}_1,\ldots,\hat{\lambda}_{N-1})$ being a diagonal matrix of the $N-1$ nonzero eigenvalues of $\hat{\Sigma}^{\frac{1}{2}}M\hat{\Sigma}^{\frac{1}{2}}$, and the columns of $\hat{P}$ are the corresponding eigenvectors.

We now turn to the problem of estimating $\breve{\kappa}^a$. From (136), $\breve{\kappa}^a$ is only determined by $\theta = \hat{V}_{11}\eta'\eta$ and N, and the only unknown quantity is $\eta'\eta$. A sensible approach to estimate $\breve{\kappa}^a$ is to replace $\eta'\eta$

[16] Paolella (2003) evaluates various schemes for approximating the mean of the ratio of quadratic forms of normal random variables and find that the approximation based on a Taylor series expansion is the most accurate. In our application, the adjusted estimator of γ_1 based on the approximate formulas produce almost identical results to those based on the exact formulas.

in (136) by $\hat{\eta}'\hat{\eta}$. However, when N is large relative to T, such a procedure could lead to a severely upwardly-biased estimate of $\eta'\eta$. The following lemma gives the conditional distribution of $\hat{V}_{11}\hat{\eta}'\hat{\eta}$ and its expectation.

Lemma 7. *Conditional on* $\hat{V}_{11}$*,*

$$\hat{\theta} = \hat{V}_{11}\hat{\eta}'\hat{\eta} \sim \frac{T(N-1)}{(T-N)} F_{N-1,T-N}(\theta), \tag{137}$$

where $F_{N-1,T-N}(\theta)$ *is a noncentral* F*-distribution with* $N-1$ *and* $T-N$ *degrees of freedom and a noncentrality parameter* θ*, and*

$$E[\hat{\theta}|\hat{V}_{11}] = \left(\frac{T}{T-N-2}\right)\theta + \frac{T(N-1)}{T-N-2}. \tag{138}$$

Lemma 7 shows that $\hat{\theta}$ tends to overestimate θ, and the overstatement is particularly severe when N is relatively large to T. As $\breve{\kappa}$ is an increasing function of θ, using the sample estimate of $\hat{\eta}'\hat{\eta}$ instead of the true $\eta'\eta$ will tend to overestimate $\breve{\kappa}^a$. This, in turn, implies that on average, the adjusted estimator $\hat{\gamma}_1^a$ will still be biased toward zero.

To account for this problem, one may like to use the unbiased estimator of θ instead of the sample estimator. From Lemma 7, the unbiased estimator of θ is given by

$$\hat{\theta}_u = \left(\frac{T-N-2}{T}\right)\hat{\theta} - (N-1). \tag{139}$$

However, this estimator has three problems. First, $\hat{\theta}_u$ can be negative with positive probability, which is unreasonable as θ can only be positive. Second, when $\hat{\theta}_u$ is close to zero, the estimated $\breve{\kappa}^a$ is also close to zero and the adjusted estimator $\hat{\gamma}_1^a = \hat{\gamma}_1/\breve{\kappa}^a$ is extremely large, causing the distribution of the adjusted estimator to have fat tails, especially when T is small. Third, while $\hat{\theta}_u$ is an unbiased estimator of θ, $1/\breve{\kappa}^a$ is not a linear function of θ, so the resulting adjusted estimator of γ_1 can still be severely biased.

From Lemma 7, we have $X = (T-N)\hat{\theta}/[T(N-1)] \sim F_{N-1,T-N}(\theta)$, so the problem of estimating θ using X is equivalent to the problem of estimating the noncentrality parameter of a noncentral F-distribution using a single observation. This problem has been studied by a number of researchers in statistics and various attempts were made to improve upon the unbiased estimator. For example, Rukhin (1993) and Kubokawa, Robert, and Saleh (1993) both propose estimators that are superior to the unbiased estimator of θ under the quadratic loss function. However, the quadratic loss function on the noncentrality parameter is not entirely appropriate for our application here. Our objective is to come up with a good estimator of $1/\breve{\kappa}^a$ that is a nonlinear function of θ. If an estimator of θ takes a value that is very close to zero, then the implied estimator of $1/\breve{\kappa}^a$ will be very large. In order for the implied estimator of $1/\breve{\kappa}^a$ to be well-behaved, our estimator of θ must not take very small values. For this purpose, a more sensible loss function on $\hat{\theta}$ is a Stein's type loss function which takes the following form

$$L(\theta, \hat{\theta}) = \frac{\hat{\theta}}{\theta} - \log\left(\frac{\hat{\theta}}{\theta}\right) - 1. \tag{140}$$

Note that the term $-\log(\hat{\theta}/\theta)$ takes a large value when $\hat{\theta}/\theta$ is small. Therefore, this loss function heavily penalizes estimators of θ that have a high probability of taking small values. By adopting an estimator of θ that minimizes this loss function, the event of making a large adjustment to $\hat{\gamma}_1$ is less likely to occur.

Using this Stein's type loss function, Fourdrinier, Philippe, and Robert (2000) provide a Bayes estimator under the improper prior $\pi(\theta) = \theta^b$, where $b > 0$. However, their estimator is designed to estimate the noncentrality parameter of a noncentral chi-squared distribution, so we need to extend their analysis to the case of the noncentral F-distribution. The following lemma presents the Bayes estimator of θ under Stein's loss function.

Lemma 8. *Under Stein's loss function (140), the Bayes estimator of θ for the class of improper priors $\pi(\theta) = \theta^b$ where $b > 0$ is*

$$\hat{\theta}_a = 2b \frac{{}_2F_1\left(1+b, \frac{T-1}{2}, \frac{N-1}{2}, \frac{z}{1+z}\right)}{{}_2F_1\left(b, \frac{T-1}{2}, \frac{N-1}{2}, \frac{z}{1+z}\right)}, \tag{141}$$

where ${}_2F_1$ is the hypergeometric function and $z = \hat{V}_{11}\hat{\eta}'\hat{\eta}/T$.

In Figure 2, we plot $\hat{\theta}$, $\hat{\theta}_u$ and $\hat{\theta}_a$ as a function of z for $N = 10$, $T = 100$, and $b = 0.5$. It can be seen that $\hat{\theta}$ is an increasing and convex function of z. When z is equal to zero, $\hat{\theta}_a = 2b = 1$. As z gets larger, $\hat{\theta}_a$ becomes more like a linear function of z and behaves almost like the unbiased estimator $\hat{\theta}_u = (T-N-2)z-(N-1)$. To understand why it makes sense to use $\hat{\theta}_a$ as an estimator of θ, note that $(T-N-2)z$ behaves almost like a $\chi^2_{N-1}(\theta)$ random variable and it has an expected value of

$\theta + N - 1$. When $(T - N - 2)z$ is very large, it is more likely that part of its large value is due to the upward bias of $N - 1$, so we adjust the estimator $\hat{\theta}_a$ downward by making it less than $(T - N - 2)z$. However, when $(T - N - 2)z$ is small, we should not subtract $N - 1$ from $(T - N - 2)z$ because even if $\theta = 0$, a small $(T - N - 2)z$ (say less than $N - 1$) indicates that $(T - N - 2)z$ is in fact less than its expected value. Therefore, we should subtract a smaller amount from $(T - N - 2)z$, causing $\hat{\theta}$ to be a nonlinear function of z. When z is very small, we want $\hat{\theta}_a$ to be even greater than $(T - N - 2)z$ because in this situation $(T - N - 2)z$ is unusually small. As the minimum of $\hat{\theta}_a$ is $2b$, we can choose a value of b such that $2b$ is an absolute lower bound of θ. For our empirical work, we choose $b = 0.5$ because we believe that in most empirical applications, it would be very unlikely to find test assets that have a θ that is less than one.

Figure 2 about here

Note that when estimated $\tilde{\kappa}^a$ and $\breve{\kappa}^a$ are used instead of $\tilde{\kappa}$ and $\breve{\kappa}$ in constructing the adjusted estimators, the adjusted estimators are no longer unbiased. Nevertheless, it is reasonable to expect that by making these adjustments, the bias of the estimated γ_1 is reduced in finite samples. Since it is difficult to obtain the finite sample distribution of our adjusted estimators, we rely on simulation experiments to examine the performance of the adjusted estimators. While the adjusted and unadjusted estimators can have very different properties in finite samples, both $\tilde{\kappa}$ and $\breve{\kappa}$ converge to one at a rate of $1/T$, so our adjusted estimators have exactly the same asymptotic distributions as the unadjusted ones, and the asymptotic results of Shanken (1992) are also applicable to our adjusted estimators.

However, the use of asymptotic standard error can cause serious size problem for statistical inference on γ_1, especially for the case of estimated GLS. This is because the expression for the asymptotic variance of $\hat{\gamma}_1$ completely ignores the finite sample variability of $\hat{\Sigma}$. Ideally, we would like to compute the t-ratio as the estimated risk premium divided by its finite sample standard deviation. In addition, similar to the bias adjustment, we would like to use the finite sample standard deviation of the estimated risk premium that is conditional on $\hat{V}_{11}$. However, the expressions in Propositions 6 to 8 are for unconditional variance and they require numerical integration, which may be difficult to use. We therefore propose the following approximation of the finite sample varaince of the estimated risk premium when conditional on $\hat{V}_{11}$.

Lemma 9. *Using a Taylor series expansion,* $\mathrm{Var}[\tilde{\gamma}_1|\hat{V}_{11}]$, $\mathrm{Var}[\breve{\gamma}_1|\hat{V}_{11}]$, *and* $\mathrm{Var}[\hat{\gamma}_1|\hat{V}_{11}]$ *can be approximated by*

$$\begin{aligned}
\mathrm{Var}[\tilde{\gamma}_1|\hat{V}_{11}] &\approx \frac{[(\theta_2 + 2\mathrm{tr}(\Lambda^2))\theta_1^2 - 2\theta_2 \mathrm{tr}(\Lambda)\theta_1 + (\mathrm{tr}(\Lambda))^2\theta_2]\gamma_1^2}{(\mathrm{tr}(\Lambda)+\theta_1)^4} \\
&\quad + \frac{(\theta_1^2+\theta_2)\frac{V_{11}}{T} + (\mathrm{tr}(\Lambda^2)+\theta_2)\hat{V}_{11}}{(\mathrm{tr}(\Lambda)+\theta_1)^2}, \qquad (142) \\
\mathrm{Var}[\breve{\gamma}_1|\hat{V}_{11}] &\approx \frac{[\theta^2+(N-1)^2]\theta\gamma_1^2}{b^4} + \frac{\theta(1+\theta)\frac{V_{11}}{T} + b\hat{V}_{11}}{b^2}, \qquad (143) \\
\mathrm{Var}[\hat{\gamma}_1|\hat{V}_{11}] &\approx \frac{(T-N-2)\mathrm{Var}[\breve{\gamma}_1|\hat{V}_{11}] + [(N-1)\hat{V}_{11} + a\theta]c - (\breve{\kappa}^a\gamma_1)^2}{T-N-1}, \qquad (144)
\end{aligned}$$

where θ_1, θ_2, θ, b, *and* $\breve{\kappa}^a$ *are defined in Lemma 6,* $a = \gamma_1^2 + \frac{V_{11}}{T}$ *and* $c = (b^2 + 2b + 2\theta)/b^3$.

Similar to the bias adjustment, computing the approximate variance requires the knowledge of some unknown parameters. In practice, we estimate θ by our adjusted estimator of $\hat{\theta}_a$ and the other parameters by their sample counterparts.

5. Simulation Experiment

5.1 Experimental Design

We perform simulation experiments to examine the robustness of our analytical results to departures from normality and the finite sample properties of our adjusted second-pass CSR estimators of the zero-beta rate and the risk premium. In choosing parameters for our simulation experiments, we attempt to cover a wide range of possible test assets and factors that are used in empirical studies. For the number of test assets, we consider three cases, $N = 10$, 25, and 100. For the 10 assets case, the parameters of the assets are chosen to mimic the 10 size-ranked portfolios of the NYSE. For the 25 assets case, the parameters are chosen to mimic the 25 size and book-to-market-ranked portfolios of the combined NYSE-AMEX-NASDAQ. For the 100 assets case, the parameters are chosen to mimic the 100 size and beta-ranked portfolios of the NYSE. For the parameters of the factor, we consider two cases. In the first case, the parameters of the factor are chosen to mimic the behavior of the value-weighted NYSE market portfolio. In the second case, the parameters are chosen to mimic the behavior of the growth rate of per capita consumption in nondurables. The main difference between these two factors is that the value-weighted market return explains a substantial portion of the time series variation of returns of well-diversified stock portfolios, whereas the growth rate of consumption has low explanatory power on the returns of stock portfolios.

We collect monthly returns for the three sets of portfolios over the period 1941/2–2002/12. The sample estimates of β and Σ from this period are used to determine the parameters for our simulations. For the growth rate of per capita consumption in nondurables, we only have monthly data starting from 1959/2, so the parameters for the low explanatory factor case are determined using sample estimates of β and Σ over the period 1959/2–2002/12. In Table 1, we report a summary of the parameters for our three sets of portfolios under the two different factor assumptions. For the factor with high explanatory power, the parameters are reported in Panel A. The betas in the 10 assets case have lower cross-sectional variations than in the other two cases. This is because over the sample period 1941/1–2002/12, the estimated betas of the ten size ranked portfolios range from only 0.96 to 1.15, and they are not all that different from each other. Table 1 also reports the signal-to-noise ratio $V_{11}\eta'\eta/(N-1)$ (for $T=100$) for our three sets of test assets. The signal-to-noise ratio is highest for the 100 assets case, so we can expect that the GLS CSR estimator of the risk premium has the least bias in this case. However, the value of signal-to-noise ratio is chosen based on the sample estimates $\hat{\eta}'\hat{\eta}$. From Lemma 6, the sample estimate $\hat{\eta}'\hat{\eta}$ tends to overestimate the true $\eta'\eta$, especially when N is large, so it is entirely possible that the higher signal-to-noise ratio for the 100 assets case is due to this bias. We do not attempt to make an adjustment here. Instead, we think of the signal-to-noise ratio for the 100 assets case as an upper bound on what we can expect from real world data when the factor resembles the return on a market portfolio.

Table 1 about here

Panel B reports the parameters for the case in which the factor has low explanatory power. When the factor is chosen to mimic the growth rate of per capita consumption, the cross-sectional variations of the betas across the portfolios tend to be higher than in the case in which the factor has high explanatory power. However, as the factor has lower explanatory power, the variance of the residuals from the regression of returns on factors is also higher. This implies that the consumption betas of the portfolios are estimated with a lot of noise, and, as a result, the signal-to-noise ratios in Panel B are much lower than the corresponding ones in Panel A.

For each case, Table 1 reports the three largest and three smallest (standardized) nonzero eigenvalues of the matrix $\Sigma^{\frac{1}{2}}M\Sigma^{\frac{1}{2}}$. If $\Sigma^{\frac{1}{2}}M\Sigma^{\frac{1}{2}}$ is proportional to the identity matrix, then the standardized eigenvalues should all equal one. Instead, λ_1 is much higher than λ_{N-1} in all cases. This suggests that the OLS and GLS CSR estimators of the risk premium can have very different properties. Table 1 also reports the absolute values of $\eta_i^* = p_i'\Sigma^{-\frac{1}{2}}\beta = \eta_i/\sqrt{T}$ corresponding to the three largest and three smallest eigenvalues, where p_i is the eigenvector of $\Sigma^{\frac{1}{2}}M\Sigma^{\frac{1}{2}}$ associated with λ_i. Although there is no theoretical relation between Σ and β, we typically find in the data that η_1^2 to η_3^2 are much larger than η_{N-3}^2 to η_{N-1}^2. From Lemma 5, it is expected that with our choice of λ_i^2 and η_i^2 the GLS CSR estimator of γ_1 has more bias than the OLS CSR estimator of γ_1.

5.2 Biases in Estimated Zero-Beta Rate and Risk Premium

With our chosen parameters, we can compute the expected value of the CSR estimators of γ_1 using the formulas in Proposition 5. Table 2 reports the unconditional biases of the OLS and GLS CSR estimators of γ_1 as a percentage of the value of the true γ_1. As the percentage bias is independent of the choice of values of γ_0 and γ_1, the numbers in Table 1 are applicable to all choices of γ_0 and γ_1, as long as $\gamma_1 \neq 0$. Also note that the estimators from the true GLS and the estimated GLS have the same bias, so we do not need to distinguish between these two versions of GLS here. Panel A reports the results for the factor with high explanatory power. When the length of the beta estimation period is $T = 60$ months, the betas of the portfolios are estimated with a lot of noise and there is a severe bias in the estimated zero-beta rate and risk premium. For the 10 assets case, the biases for the GLS CSR estimators of γ_1 is -75.8%. As for the OLS CSR, the bias is smaller but it is still high at -56.8%. When T increases, the biases for both the OLS and GLS CSR estimators tend to be lower. However, even for T as high as 600 months, the GLS CSR estimator of risk premium still shows a -21% bias. Similar patterns also hold for the 25 and 100 assets cases. However, as the signal-to-noise ratio is higher for the 25 and 100 assets case, the biases are of smaller magnitude but they are still rather significant, especially when T is small.

Table 2 about here

Panel B reports the results for the factor with low explanatory power. As the consumption betas are estimated with a lot of noise, the percentage biases of the CSR estimators of γ_1 are huge. When $T = 60$ months, the bias of the GLS estimator of γ_1 is more than -80% for all three sets of test assets. Even when $T = 600$ months is used to estimate the consumption betas, the bias of the GLS estimator of the risk premium is still more than -30%. Given the huge bias of the estimated risk premium, it would be quite difficult to find the consumption betas to be priced even if the consumption CAPM is exactly correct. As in the case of Panel A, the OLS estimators have smaller biases than the GLS estimators but the biases are still significant. As T increases, the bias of the OLS estimator does not always exhibit the same monotonic pattern as in the GLS case. For example, when $N = 10$, the bias of $\tilde{\gamma}_1$ falls from -61.3% to -1.2% as T increases from 60 months to 360 months. However, when T rises to 480 months, the bias of $\tilde{\gamma}_1$ turns positive and increases to 2.4%, and further increases to 3.6% when T goes up to 600 months. Although not shown in the table, the bias of $\tilde{\gamma}_1$ will eventually approach zero as T increases further. However, for intermediate values of T, there is no guarantee that a longer beta estimation period will reduce the bias of the OLS estimator, nor is there any guarantee that the bias of $\tilde{\gamma}_1$ will be negative.

5.3 Comparison of Asymptotic and Finite Sample Standard Deviation

For statistical inference on risk premium, we need to know the standard deviation of the CSR estimators of γ_1. Traditionally, asymptotic results are used for this purpose. With the results in Propositions 6–8, we now know the finite sample standard deviation of these estimators. Table 3 reports the asymptotic and finite sample standard deviation of the OLS and estimated GLS CSR estimators of γ_1. Panels A and B contain the results for the factor with high and low explanatory power, respectively. In computing the asymptotic and finite sample standard deviation of the CSR estimators, we need to make an assumption about the value of γ_1. As the high explanatory power factor case is chosen to mimic the value-weighted return of the market, we assume γ_1 is 0.6% per month. By choosing this value, we have the CAPM in mind which suggests that the risk premium of the market beta should be the expected excess return on the market portfolio. For the low explanatory power case, we choose γ_1 to be 0.028% per month. In choosing this value, we have the consumption CAPM in mind, which suggests that under a utility function with constant relative risk aversion, the risk premium for the consumption beta should be (see Breeden, Gibbons, and Litzenberger (1989))

$$\gamma_1 = \frac{\rho \mathrm{Var}[c_t]}{1 - \rho E[c_t]}, \tag{145}$$

where ρ is the coefficient of relative risk aversion, and c_t is the growth rate of aggregate consumption. Using our monthly data on the growth rate of per capita consumption, we estimate the mean and standard deviation of c_t to be 0.105%/month and 0.947%/month, respectively. Then, by assuming $\rho = 5$, $\gamma_1 = 0.028\%$/month.

Table 3 about here

The asymptotic standard deviations in Table 3 are computed based on (26) and (28) using the true parameters (after dividing the asymptotic variance by T). These are EIV-adjusted standard errors from Shanken (1992). The unadjusted ones are very close to the EIV-adjusted ones, so we do not report them separately. As for the finite sample standard deviations, they are based on formulas in Propositions 6 and 8. By comparing the asymptotic and finite sample standard deviations, we find that the asymptotic standard deviation tends to overstate the finite sample standard deviation, especially when T is small. This shows that while the asymptotic variance formula attempts to take into account of the effect of EIV, it treats the estimated risk premium as consistent and completely ignores its finite sample bias. When T is small, the finite sample distribution of the estimated risk premium can be heavily biased toward zero, leading to a substantially smaller finite sample standard deviation as compared with the one that is predicted by the asymptotic standard deviation formula.

The results in Table 3 allow us to compare the merits of the estimators from OLS and the estimated GLS. It is well known that the true GLS CSR estimator is more efficient than the OLS CSR estimator. However, since Σ has to be estimated, it is not clear that the advantage of the true GLS CSR estimator carries over to the estimated GLS CSR estimator. As a result, many researchers opt to use the simpler OLS CSR estimator. Table 3 shows that when $N = 10$ and $N = 25$, the estimated GLS continues to dominate the OLS in terms of estimation efficiency, even for T as short as 60 months, and the improvement is often very big. When $N = 100$, the only case that the OLS estimator is superior is when $T = 120$. In this case, T is close to N, so $\hat{\Sigma}$ is very volatile, which leads to added volatility to the estimated GLS estimator. Other than this case, the estimated GLS estimator largely dominates OLS the estimator in terms of estimation efficiency. Therefore, unless T is very close to N, it is advisable to choose the estimated GLS CSR estimator over the OLS CSR estimator if one is concerned about estimation efficiency.

5.4 Nonnormal Distributions

While the analytical results in this paper are derived under the assumption of multivariate normality, we have good reason to believe that they work fairly well even though the factors and returns are not normally distributed. For example, the work of MacKinlay (1985) and Zhou (1993) shows that although the F-test of Gibbons, Ross, and Shanken (1989) for testing the mean-variance efficiency of a given portfolio relies on the multivariate normality of the residuals, it is rather robust to departures from normality. To examine if our finite sample results on the CSR estimators of the zero-beta rate and the risk premium are robust to departures from normality, we consider the case in which the factor has a t-distribution with five degrees of freedom, and the residuals of the test assets have a multivariate t-distribution with five degrees of freedom. Under this alternative distribution assumption, the factor and the returns have fat tails, which is often the case in the data. As we cannot obtain the finite sample distribution of the CSR estimators under the t-distribution assumption, we rely on simulation. In order to easily compare with our results under normality and nonnormality, we simulate the factor and the returns of the test assets using exactly the same μ and V as in the normality case. Table 4 presents the percentage bias of the OLS and estimated GLS estimators of γ_1 under our alternative distribution assumption using exactly the same format as in Table 2. The results are based on 100,000 simulations. By comparing Tables 2 and 4, the percentage biases of the CSR estimators under the two distribution assumptions are fairly close to each other, with the only exceptions being the GLS case with $N = 100$ and T is small.

Table 4 about here

Table 5 reports the finite sample standard deviations of the OLS and estimated GLS CSR estimators of γ_1 in the 100,000 simulations under the t-distribution assumption. By comparing Tables 3 and 5, the analytical results for the normality case again prove to be a very good approximation to the t-distribution case, even when T is small. The only noticeable difference again comes from the GLS case with $N = 100$ and T is small. This robustness result is not surprising because while $\hat{\beta}$ is not exactly normal and $\hat{\Sigma}$ is not exactly Wishart when the residuals are not multivariate normally distributed, such approximations are in fact quite good even for moderate sizes of T. In view of the simulation evidence here, we conclude that our analytical finite sample results can still be good approximations even when the factor and the returns are not multivariate normally distributed.

Table 5 about here

5.5 Simulation Results on Bias-adjusted Estimators

We now turn to the bias-adjusted estimators. To evaluate their performance, we rely on simulation. However, since the bias-adjusted estimators only depend on $\hat{\mu}_2$, $\hat{\beta}$, and $\hat{\Sigma}$, so there is no need to simulate the returns and the factors. In fact, using the same approach as in Section 2.2, we only need to simulate $\hat{\beta}$ and $\hat{\Sigma}$ or a normalized version of them in order to approximate the mean and variance of the adjusted estimators.[17] Using the same parameters as before, we simulate the bias-adjusted estimators of γ_1 under the OLS and the estimated GLS CSR for 100,000 times. Table 6 reports the percentage biases of the adjusted estimators of γ_1. Comparing Tables 2 and 6, there is a dramatic reduction of biases for the bias-adjusted estimators as compared to the unadjusted estimators. When T is small, the bias-adjusted estimators do not offer a sufficient bias adjustment. This is because the adjustment factor used is based on estimated rather than true parameters, and the estimated parameters are less reliable when T is small. It should be noted that there are a few cases ($N = 10$ and $T > 240$) where the bias-adjusted OLS estimator actually over-adjusts, leading to even more bias than the unadjusted estimator. As a whole, however, our bias-adjusted estimators are quite effective in reducing the bias in the unadjusted estimators.

Table 6 about here

The reduction of bias has its cost, namely, the increase in the volatility of the estimator, which occurs even if the true adjustment factor is known. For example, the GLS bias-adjusted estimator of γ_1 is given by $\hat{\gamma}_1^a = \hat{\gamma}_1/\breve{\kappa}$, so $\text{Var}[\hat{\gamma}_1^a] = \text{Var}[\hat{\gamma}_1]/\breve{\kappa}^2 > \text{Var}[\hat{\gamma}_1]$ since $\breve{\kappa} < 1$. In addition, using the estimated $\breve{\kappa}$ instead of the true $\breve{\kappa}$ adds yet another source of variability to $\hat{\gamma}_1^a$, so our bias-adjusted estimators are more volatile than the unadjusted estimators. Table 7 reports the finite sample standard deviations of the bias-adjusted estimators in the 100,000 simulations. By comparing Tables 3 and 7, we note that when T is small and the bias of the unadjusted estimator is large, there is a big increase in the volatility of the bias-adjusted estimator as compared to the unadjusted estimator, so in terms of root mean squared error, our adjusted estimators can be worse than the unadjusted estimators. However, in many empirical asset pricing studies, the value of the risk

[17] Details are available upon request.

premium is of central importance, so a heavily biased estimator of the risk premium is seriously misleading. In fact, it will be even more misleading if such an estimator is less volatile because without taking into account of its bias, one is more likely to make a wrong statistical inference by believing that the risk premium estimate is in fact very accurate.

Table 7 about here

6. Conclusion

Due to its easy implementation, the two-pass CSR methodology has been used extensively to estimate risk premia associated with systematic factors. Despite the methodology's simplicity, the finite sample properties of the estimated risk premiua from this two-pass approach are complicated by the EIV problem associated with the use of estimated betas in the second-pass CSR. Traditionally, researchers have either ignored the EIV problem or relied on the asymptotic results of Shanken (1992). Neither approach addresses the issue of correcting the finite sample bias of the estimated zero-beta rate and risk premia.

We analyze the finite sample bias and variance of the estimators of the zero-beta rate and risk premia from the second-pass CSR. Under the normality assumption, we give explicit expressions for the finite sample bias and variance of the estimated zero-beta rate and risk premium for the single factor case. For the multi-factor case, we offer an efficient simulation approach to obtain the finite sample bias and variance of the estimated zero-beta rate and risk premia. For the single factor case, we find that the GLS CSR estimator of the risk premium on average underestimates the true risk premium. For reasonable choices of parameters, this understatement is very severe, especially when the beta estimation period is short. For the OLS CSR, the estimated risk premium can overestimate or underestimate the true risk premium. In many cases, we find that the bias of the risk premium from the OLS CSR is still negative and it tends to be smaller than that from the GLS CSR. While our analytical results are derived under the normality assumption, simulation evidence suggests that they are fairly robust to departures from normality.

While the unadjusted CSR estimators can have serious biases, the popular adjusted estimators due to Litzenberger and Ramaswamy (1979) and Kim (1995) as well as the maximum likelihood estimator do not fare any better. In fact, under fairly general conditions, we show that these adjusted estimators do not even have finite first moments. As a result, they cannot be used to correct the bias of the unadjusted estimators. We suggest a simple bias adjustment to the second-pass CSR estimators of the zero-beta rate and the risk premium to correct their finite sample biases.[18] Using simulations, we find that our adjusted version of the second-pass CSR estimators of the risk premium can significantly reduce the bias of the unadjusted estimators. Since the value of the risk premium is of central importance in many finance applications, researchers should be cautious in relying on the unadjusted estimated risk premium from the second-pass CSR for making inferences, especially when the beta estimation period is short. It is up to future research to ascertain whether the failure of the CAPM as documented by some recent empirical studies can be partly explained by the EIV problem and whether there will be more support for the CAPM using our adjusted estimators of the risk premium.

[18] A set of Matlab programs for calculating the adjusted estimators is available upon request.

Appendix

Proof of Lemma 1. Since the proofs of $\tilde{\gamma}^K$ and $\hat{\gamma}^K$ are almost identical to that of $\check{\gamma}^K$, we only provide the proof of $\check{\gamma}^K$ here. The objective is to choose γ_0, $\bar{\gamma}_1$, and b to minimize

$$f(\gamma_0, \bar{\gamma}_1, b) = g'\Omega^{-1}g. \tag{A1}$$

Conditional on γ_0 and $\bar{\gamma}_1$, g is linear in b. Writing $g = y - Xb$, where

$$X = \begin{bmatrix} \bar{\gamma}_1' \otimes I_N \\ I_{NK} \end{bmatrix} = \begin{bmatrix} \bar{\gamma}_1' \\ I_K \end{bmatrix} \otimes I_N, \tag{A2}$$

$$y = \begin{bmatrix} \hat{\mu}_2 - 1_N\gamma_0 \\ \hat{b} \end{bmatrix}, \tag{A3}$$

it is easy to show that the b that minimizes (A1) is $b^* = (X'\Omega^{-1}X)^{-1}(X'\Omega^{-1}y)$. Conditional on b^*, write f as a function of γ_0 and $\bar{\gamma}_1$ alone:

$$f(\gamma_0, \bar{\gamma}_1) = y'[\Omega^{-1} - \Omega^{-1}X(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}]y. \tag{A4}$$

Since $X'\Omega^{-1}X = (\bar{\gamma}_1\bar{\gamma}_1' + \hat{V}_{11}) \otimes T\Sigma^{-1}$,

$$(X'\Omega^{-1}X)^{-1} = \left(\hat{V}_{11}^{-1} - \frac{\hat{V}_{11}^{-1}\bar{\gamma}_1\bar{\gamma}_1'\hat{V}_{11}^{-1}}{1 + \bar{\gamma}_1'\hat{V}_{11}^{-1}\bar{\gamma}_1}\right) \otimes (\Sigma/T) \tag{A5}$$

and

$$\Omega^{-1} - \Omega^{-1}X(X'\Omega^{-1}X)^{-1}X'\Omega^{-1} = \frac{[1,\ -\bar{\gamma}_1']'[1,\ -\bar{\gamma}_1'] \otimes T\Sigma^{-1}}{1 + \bar{\gamma}_1'\hat{V}_{11}^{-1}\bar{\gamma}_1}. \tag{A6}$$

Using the fact that

$$\begin{aligned} ([1,\ -\bar{\gamma}_1'] \otimes \Sigma^{-\frac{1}{2}})y &= ([1,\ -\bar{\gamma}_1'] \otimes \Sigma^{-\frac{1}{2}})\mathrm{vec}([\hat{\mu}_2 - 1_N\gamma_0,\ \hat{\beta}]) \\ &= \mathrm{vec}\left(\Sigma^{-\frac{1}{2}}[\hat{\mu}_2 - 1_N\gamma_0,\ \hat{\beta}][1,\ -\bar{\gamma}_1']'\right) \\ &= \Sigma^{-\frac{1}{2}}(\hat{\mu}_2 - 1_N\gamma_0 - \hat{\beta}\gamma_1), \end{aligned} \tag{A7}$$

write

$$f(\gamma_0, \bar{\gamma}_1) = \frac{T(\hat{\mu}_2 - 1_N\gamma_0 - \hat{\beta}\gamma_1)'\Sigma^{-1}(\hat{\mu}_2 - 1_N\gamma_0 - \hat{\beta}\gamma_1)}{1 + \bar{\gamma}_1'\hat{V}_{11}^{-1}\bar{\gamma}_1} = \frac{Tx'\check{A}x}{x'Gx}, \tag{A8}$$

where $x = [1,\ -\gamma_0,\ -\bar{\gamma}_1']'$. Choosing x to minimize $f(\gamma_0, \bar{\gamma}_1)$ is the same as choosing x to maximize $(x'Gx)/(x'Ax)$. By the Rayleigh-Ritz theorem,

$$\max_x \frac{x'Gx}{x'\check{A}x} = \lambda_1, \tag{A9}$$

where λ_1 is the largest eigenvalue of $\breve{A}^{-\frac{1}{2}}G\breve{A}^{-\frac{1}{2}}$. As $\breve{A}^{-\frac{1}{2}}G\breve{A}^{-\frac{1}{2}}$ and $\breve{A}^{-1}G$ share the same set of eigenvalues, the maximum is attained when x is proportional to the eigenvector of $\breve{A}^{-1}G$ associated with λ_1, because with this choice of x, $\breve{A}^{-1}Gx = \lambda_1 x$ and hence $x'Gx = x'\breve{A}\breve{A}^{-1}Gx = \lambda_1 x'\breve{A}x$, so the maximum of $x'Gx/(x'\breve{A}x)$ is attained. This completes the proof. *Q.E.D.*

Proof that Kim's estimated GLS estimator of γ is numerically identical to the maximum likelihood estimator of γ. From Theorem 3 of Shanken (1992), we know that under the joint normality assumption of returns and the factors, the maximum likelihood estimator for γ_0 and γ_1 is the one that minimizes the function

$$f(\hat{\gamma}_0, \hat{\gamma}_1) = \frac{e'\hat{\Sigma}^{-1}e}{1 + \hat{\gamma}_1'\hat{V}_{11}^{-1}\hat{\gamma}_1}, \tag{A10}$$

where $e = \hat{\alpha} - 1_N\hat{\gamma}_0 - \hat{\beta}(\hat{\gamma}_1 - \hat{\mu}_1) = \hat{\mu}_2 - 1_N\hat{\gamma}_0 - \hat{\beta}\hat{\gamma}_1$. Using the notations in the proof of Lemma 1, the objective function can be written as

$$f(\hat{\gamma}_0, \hat{\gamma}_1) = \min_x \frac{x'\hat{A}x}{x'Gx}, \tag{A11}$$

where $x = [1,\ -\hat{\gamma}_0,\ -\hat{\gamma}_1']'$, and this objective function is identical to the one for the Kim's estimated GLS estimator. This completes the proof. *Q.E.D.*

Proof of Lemma 2. Conditional on $\hat{\mu}_1$, $\hat{V}_{11}$, $\hat{\mu}_2$, $\hat{\beta}$, and U, from (65),

$$E[\hat{\gamma}|\hat{\mu}_1, \hat{V}_{11}, \hat{\mu}_2, \hat{\beta}, U] = \breve{\gamma}. \tag{A12}$$

Taking the expectation of both sides with respect to $\hat{\mu}_2$, $\hat{\beta}$, and U, the law of iterated expectations gives

$$E[\hat{\gamma}|\hat{\mu}_1, \hat{V}_{11}] = E[\breve{\gamma}|\hat{\mu}_1, \hat{V}_{11}]. \tag{A13}$$

Taking the expectation of both sides with respect to $\hat{\mu}_1$ and $\hat{V}_{11}$,

$$E[\hat{\gamma}] = E[\breve{\gamma}]. \tag{A14}$$

Similarly, conditional on $\hat{\mu}_1$, $\hat{V}_{11}$, $\hat{\mu}_2$, $\hat{\beta}$, and U, from (65),

$$\mathrm{Var}[\hat{\gamma}|\hat{\mu}_1, \hat{V}_{11}, \hat{\beta}, \hat{\mu}_2, U] = \frac{(\breve{A}_{11} - \breve{A}_{12}\breve{A}_{22}^{-1}\breve{A}_{21})\breve{A}_{22}^{-1}}{U}. \tag{A15}$$

Conditional on $\hat{\mu}_1$ and $\hat{V}_{11}$,

$$\mathrm{Var}[\hat{\gamma}|\hat{\mu}_1, \hat{V}_{11}] = \mathrm{Var}[E[\hat{\gamma}|\hat{\mu}_1, \hat{V}_{11}, \hat{\mu}_2, \hat{\beta}, U]|\hat{\mu}_1, \hat{V}_{11}] + E[\mathrm{Var}[\hat{\gamma}|\hat{\mu}_1, \hat{V}_{11}, \hat{\beta}, \hat{\mu}_2, U]|\hat{\mu}_1, \hat{V}_{11}]$$

$$= \text{Var}[\breve{\gamma}|\hat{\mu}_1, \hat{V}_{11}] + E\left[\frac{1}{U}(\breve{A}_{11} - \breve{A}_{12}\breve{A}_{22}^{-1}\breve{A}_{21})\breve{A}_{22}^{-1}\right| \hat{\mu}_1, \hat{V}_{11}\bigg]$$

$$= \text{Var}[\breve{\gamma}|\hat{\mu}_1, \hat{V}_{11}] + \frac{1}{T-N-1}E\left[(\breve{A}_{11} - \breve{A}_{12}\breve{A}_{22}^{-1}\breve{A}_{21})\breve{A}_{22}^{-1}\right| \hat{\mu}_1, \hat{V}_{11}\bigg], \quad \text{(A16)}$$

where the last equality follows because U is independent of $\hat{\mu}_2$ and $\hat{\beta}$ and $E[1/U] = 1/(T-N-1)$. Finally, the unconditional variance of $\hat{\gamma}$ is

$$\begin{aligned} \text{Var}[\hat{\gamma}] &= \text{Var}[E[\hat{\gamma}|\hat{\mu}_1, \hat{V}_{11}]] + E[\text{Var}[\hat{\gamma}|\hat{\mu}_1, \hat{V}_{11}]] \\ &= \text{Var}[E[\breve{\gamma}|\hat{\mu}_1, \hat{V}_{11}]] + E[\text{Var}[\breve{\gamma}|\hat{\mu}_1, \hat{V}_{11}]] + \frac{1}{T-N-1}E[(\breve{A}_{11} - \breve{A}_{12}\breve{A}_{22}^{-1}\breve{A}_{21})\breve{A}_{22}^{-1}] \\ &= \text{Var}[\breve{\gamma}] + \frac{1}{T-N-1}E[(\breve{A}_{11} - \breve{A}_{12}\breve{A}_{22}^{-1}\breve{A}_{21})\breve{A}_{22}^{-1}], \end{aligned} \quad \text{(A17)}$$

where the second equality follows from (A13) and (A16). This completes the proof. *Q.E.D.*

Proof of Proposition 1. We first prove the necessary and sufficient conditions for the existence of the conditional s-th moment of $\tilde{\gamma}_1$. Theorem 1 of Kinal (1980) establishes the following lemma.

Kinal's Lemma. Let A be a $p \times q$ matrix of normal random variables, where $p > q$, and let C be a p-vector of normal random variables. Suppose $x_i = [C_i,\ A_i]' \sim N(\mu_i, s_i I_{q+1})$ where A_i is the i-th row of A and x_i are independent across i. Then, the s-th moment of $(A'A)^{-1}A'C$ exists if and only if $s < p - q + 1$.

Writing $A = \Lambda^{\frac{1}{2}} Z_2$, $C = \Lambda^{\frac{1}{2}} Y_2$ where $Y_2 = \sqrt{T} P' \Sigma^{-\frac{1}{2}} \hat{\mu}_2$, conditional on $\hat{\mu}_1$ and $\hat{V}_{11}$, A and C are normally distributed, and $[C_i,\ A_i]'$ has variance $\lambda_i I_N$, so A and C satisfy the conditions of Kinal's lemma. This implies that the conditional s-th moment of $(Z_2'\Lambda Z_2)^{-1}(Z_2'\Lambda Y_2)$ exists (which in turn implies that the conditional s-th moment of $\tilde{\gamma}_1 = \hat{V}_{11}^{\frac{1}{2}}(Z_2'\Lambda Z_2)^{-1}(Z_2'\Lambda Y_2)$ exists) if and only if $s < N - K$.

For $\tilde{\gamma}_0$, from (44), we define $Y_1 = \sqrt{T}\nu'\Sigma^{-\frac{1}{2}}\hat{\mu}_2$ and we can write

$$\tilde{\gamma}_0 = \delta_1 Y_1 + \delta_2' Y_2 - (\delta_1 Z_1 + \delta_2' Z_2)(Z_2'\Lambda Z_2)^{-1}(Z_2'\Lambda Y_2) = A_0 \hat{\mu}_2. \quad \text{(A18)}$$

Conditional on $\hat{\mu}_1$ and $\hat{V}_{11}$, the term $\delta_1 Y_1 + \delta_2' Y_2$ has a normal distribution and all of its moments exist. The s-th moment of the term $Z_1(Z_2'\Lambda Z_2)^{-1}(Z_2'\Lambda Y_2)$ exists if and only if $s < N - K$. This is because Z_1 is normally distributed (with all moments existing) and is independent of $(Z_2'\Lambda Z_2)^{-1}(Z_2'\Lambda Y_2)$, the finite s-th moment of which exists if and only if $s < N - K$. Finally, all of the moments of the last term $\delta_2' Z_2(Z_2'\Lambda Z_2)^{-1}(Z_2'\Lambda Y_2)$ exist because Y_2 is normally distributed (with all moments existing) and is independent of $\delta_2' Z_2(Z_2'\Lambda Z_2)^{-1}Z_2'$ (all moments of which exist). This is because for any $(N-1)$-vector c, from the Cauchy-Schwarz inequality,

$$(\delta_2' Z_2(Z_2'\Lambda Z_2)^{-1}Z_2'c)^2 \leq (\delta_2' Z_2(Z_2'\Lambda Z_2)^{-1}Z_2'\delta_2)(c' Z_2(Z_2'\Lambda Z_2)^{-1}Z_2'c) < K^2(\delta_2'\Lambda^{-1}\delta_2)(c'\Lambda^{-1}c), \tag{A19}$$

The second inequality follows because, by writing $x = \Lambda^{-\frac{1}{2}}Z_2$ and $d_{\max}$ as the largest eigenvalue of $\Lambda^{\frac{1}{2}}Z_2(Z_2'\Lambda Z_2)^{-1}Z_2'\Lambda^{\frac{1}{2}}$,

$$\frac{\delta_2' Z_2(Z_2'\Lambda Z_2)^{-1}Z_2'\delta_2}{\delta_2'\Lambda^{-1}\delta_2} = \frac{x'[\Lambda^{\frac{1}{2}}Z_2(Z_2'\Lambda Z_2)^{-1}Z_2'\Lambda^{\frac{1}{2}}]x}{x'x} \leq d_{\max} < \mathrm{tr}(\Lambda^{\frac{1}{2}}Z_2(Z_2'\Lambda Z_2)^{-1}Z_2'\Lambda^{\frac{1}{2}}) = K. \tag{A20}$$

Therefore, $|\delta_2' Z_2(Z_2'\Lambda Z_2)^{-1}Z_2'c|$ is bounded from above and all of its moments exist. Since, conditional on $\hat{\mu}_1$ and $\hat{V}_{11}$, all moments exist for all the terms of $\tilde{\gamma}_0$ except for one term which has a finite s-th moment if and only $s < N-K$, it follows that the conditional s-th moment of $\tilde{\gamma}_0$ exists if and only if $s < N-K$.

The proof for $\check{\gamma}$ is the same as the proof for $\tilde{\gamma}$ except for setting $\Lambda = I_{N-1}$. As for $\hat{\gamma}$ from the estimated GLS, from (65) and (66),

$$\hat{\gamma} = \check{\gamma} + U^{-\frac{1}{2}}(\check{A}_{11} - \check{A}_{12}\check{A}_{22}^{-1}\check{A}_{21})^{\frac{1}{2}}\check{A}_{22}^{-\frac{1}{2}}Y_3, \tag{A21}$$

where $Y_3 \sim N(0_{K+1}, I_{K+1})$ and is independent of Y_2, Z_1, Z_2, and U. For the first term, the conditional s-th moment of $\check{\gamma}$ exists if and only if $s < N-K$. As for the second term, all of the moments of Y_3 exist and Y_3 is independent of Y_2, Z_2, and U, so we only need to consider the existence of moments for $U^{-\frac{1}{2}}$ and $(\check{A}_{11} - \check{A}_{12}\check{A}_{22}^{-1}\check{A}_{21})^{\frac{1}{2}}\check{A}_{22}^{-\frac{1}{2}}$. Since $U \sim \chi^2_{T-N+1}$, then $U^{-\frac{s}{2}}$ exist if and only if $s < T-N+1$ (see, for example, Johnson, Kotz, and Balakrishnan (1995, Chapter 27)). As $\check{A}_{11} - \check{A}_{12}\check{A}_{22}^{-1}\check{A}_{21} = (Y_2'[I_{N-1} - Z_2(Z_2'Z_2)^{-1}Z_2']Y_2)/T \leq Y_2'Y_2/T$ and $Y_2'Y_2$ has a noncentral chi-squared distribution with all of its moments existing, so all the moments of the term $\check{A}_{11} - \check{A}_{12}\check{A}_{22}^{-1}\check{A}_{21}$ also exist. Finally, from (66), the s-th moment of $\check{A}_{22}^{-\frac{1}{2}}$ exists if and only if the s-th moment of $(Z_2'Z_2)^{-\frac{1}{2}}$ exists. The proof of Theorem 1 in Kinal (1980) (see also Magnus (1990)) establishes that the s-th moment of $(Z_2'Z_2)^{-\frac{1}{2}}$ exists if and only if $s < N-K$. Combining these results, the conditional s-th moment of $\hat{\gamma}$ exists if and only if $s < \min[N-K, T-N+1]$. This completes the proof. *Q.E.D.*

Proof of Proposition 2. We only prove the nonexistence of moments for $\check{\gamma}^{LR}$ and $\check{\gamma}^{K}$ here. The proofs for the other cases are similar. We first present a general theorem on the existence of moments that is due to Sargan (1976).[19]

Sargan's Theorem. If $\hat{\theta} = \psi(p)/\phi(p)$, where p is a random vector and $\hat{\theta}$ is a scalar function of p, and there exists a p_0 in the domain of definition of p such that

1. $\psi(p)$ is continuous at p_0 with $\psi(p_0) \neq 0$,

2. $\phi(p)$ has continuous derivatives at p_0, denoted by ϕ_p, for which $\phi_p'\phi_p > 0$ and $\phi(p_0) = 0$,

3. p has a continuous density function f with $f(p_0) > 0$,

then $E[|\hat{\theta}|^s] = \infty$ for $s \geq 1$.

In order to apply this theorem to prove that $\check{\gamma}^{LR}$ and $\check{\gamma}^{K}$ have no integral moments, we just need to show that elements of these two estimators can be written as ratios of two functions of $\hat{\mu}_2$ and $\hat{\beta}$, and that there is some choice of $\hat{\mu}_2$ and $\hat{\beta}$ such that the denominator is zero but the numerator is not. In addition, we need to show that ϕ_p has at least one nonzero element. Condition 3 in Sargan's Theorem is automatically satisfied because $\hat{\mu}_2$ and $\hat{\beta}$ are normally distributed conditional on $\hat{\mu}_1$ and $\hat{V}_{11}$.

For $\check{\gamma}^{LR}$, we define $\omega = \Sigma^{-\frac{1}{2}}1_N$, $Y = \Sigma^{-\frac{1}{2}}\hat{\mu}_2$, $Z = \Sigma^{-\frac{1}{2}}\hat{\beta}$ and $p = [Y,\ \mathrm{vec}(Z)']'$. Write

$$\check{\gamma}^{LR} = W^{-1}\begin{bmatrix} \omega'Y \\ Z'Y \end{bmatrix} = \frac{\mathrm{adj}(W)}{|W|}\begin{bmatrix} \omega' \\ Z' \end{bmatrix} Y, \tag{A22}$$

where adj(W) stands for the adjoint of the W matrix, which is given by

$$W(p) = \begin{bmatrix} \omega'\omega & \omega'Z \\ Z'\omega & Z'Z - N\hat{V}_{11}^{-1}/T \end{bmatrix}. \tag{A23}$$

Under this expression, $|W(p)|$ is the function $\phi(p)$ in Sargan's Theorem. Let e_i be a K-vector of all zeros except for its i-th element, which is equal to one. When $\hat{V}_{11}$ is positive definite, the matrix $N\hat{V}_{11}^{-1}/T + I_K - e_ie_i'$ is also positive definite, and let $N\hat{V}_{11}^{-1}/T + I_K - e_ie_i' = P\Lambda P'$, where Λ is a diagonal matrix which contains the eigenvalues, and the columns of P are the corresponding eigenvectors. Let $Z^* = Q\Lambda^{\frac{1}{2}}P'$, where Q is an $N \times K$ matrix with $Q'Q = I_K$ and the columns of Q are orthogonal to ω. Note that the ith column of Z^* cannot be a zero vector (otherwise,

[19] See also Theorem 3.9.1 of Phillips (1983).

the i-th diagonal element of $N\hat{V}_{11}^{-1}/T$ would be zero, contradicting the fact that $\hat{V}_{11}$ is positive definite), so without loss of generality, assume its (j,i)-th element is nonzero (i.e., $Z_{ji}^* \neq 0$). Now, let $Z_0 = Z^* + \omega e_i'$ and $Y_0 = h_j$ where h_j is an N-vector of all zeros except for its j-th element, which is equal to one. With this choice of $p_0 = [Y_0,\ \mathrm{vec}(Z_0)']'$,

$$W(p_0) = \begin{bmatrix} \omega'\omega & (\omega'\omega)e_i' \\ (\omega'\omega)e_i & I_K + (\omega'\omega - 1)e_i e_i' \end{bmatrix} \tag{A24}$$

and we can easily verify that

$$\mathrm{adj}(W(p_0)) \begin{bmatrix} \omega' \\ Z_0' \end{bmatrix} Y_0 = (\omega'\omega) \begin{bmatrix} -Z_{ji}^* \\ Z_{ji}^* e_i \end{bmatrix}, \tag{A25}$$

so the first and the $(i+1)$-th elements are nonzero. However, for $p = p_0$, the function in the denominator of $\check{\gamma}^{LR}$ is zero because

$$|W(p_0)| = (\omega'\omega)|I_K + (\omega'\omega - 1)e_i e_i' - (\omega'\omega)e_i e_i'| = 0. \tag{A26}$$

Since

$$\frac{\partial |W(p)|}{\partial Z} = (\omega'\omega)\frac{\partial |Z'CZ - N\hat{V}_{11}^{-1}/T|}{\partial Z} = 2(\omega'\omega)CZ\mathrm{adj}(Z'CZ - N\hat{V}_{11}^{-1}/T), \tag{A27}$$

where $C = I_N - \omega(\omega'\omega)^{-1}\omega'$, use $CZ_0 = Z^*$ and $Z_0'CZ_0 - N\hat{V}_{11}^{-1}/T = I_K - e_i e_i'$ to obtain

$$\left.\frac{\partial |W(p)|}{\partial Z}\right|_{p=p_0} = 2(\omega'\omega)Z^*\mathrm{adj}(I_K - e_i e_i') = 2(\omega'\omega)Z^* e_i e_i', \tag{A28}$$

which is a nonzero matrix as the i-th column of Z^* is a nonzero vector. Therefore, the derivative condition is satisfied. Note that in the proof above, we have shown that the first and the $(i+1)$-th elements of $\check{\gamma}^{LR}$ have no integral moments. However, the choice of i is entirely arbitrary, so by varying i, all the elements of $\check{\gamma}^{LR}$ can be shown to have no integral moments.

Turning our attention to $\check{\gamma}^{K}$, we need to find the eigenvector associated with the largest eigenvalue of $\check{A}^{-1}G$. This involves solving the equation $\check{A}^{-1}Gx = \lambda x$, or equivalently $(G - \lambda\check{A})x = 0_{K+2}$. Using the same notation as before,

$$(G - \lambda\check{A})x = \begin{bmatrix} 1 - \lambda Y'Y & -\lambda Y'\omega & -\lambda Y'Z \\ -\lambda\omega'Y & -\lambda\omega'\omega & -\lambda\omega'Z \\ -\lambda Z'Y & -\lambda Z'\omega & \hat{V}_{11}^{-1} - \lambda Z'Z \end{bmatrix} x = 0_{K+2}. \tag{A29}$$

Let $Q = [q_1, \; \cdots \; q_{K+1}]$ be an $N \times (K+1)$ matrix such that $Q'Q = I_{K+1}$ and the columns of Q are orthogonal to ω. Let $Y_0 = q_{K+1}$ and $Z_0 = [q_1, \; \cdots, \; q_{i-1}, \; \frac{1}{2}(q_i + \omega/(\omega'\omega)^{\frac{1}{2}}), \; q_{i+1}, \; \cdots, \; q_K]\hat{V}_{11}^{-\frac{1}{2}}$. When $p = p_0$, the problem becomes

$$\begin{bmatrix} 1-\lambda & 0 & 0_K' \\ 0 & -\lambda\omega'\omega & -\frac{\lambda}{2}(\omega'\omega)^{\frac{1}{2}} e_i' \hat{V}_{11}^{-\frac{1}{2}} \\ 0_K & -\frac{\lambda}{2}(\omega'\omega)^{\frac{1}{2}} \hat{V}_{11}^{-\frac{1}{2}} e_i & \hat{V}_{11}^{-\frac{1}{2}}[(1-\lambda)I_K + \frac{1}{2}\lambda e_i e_i']\hat{V}_{11}^{-\frac{1}{2}} \end{bmatrix} x = 0_{K+2}. \tag{A30}$$

Solving this problem, $\lambda_1 = 4$, and the associated eigenvector is $x = [x_1, \; x_2, \; x_3']'$, where $x_1 = 0$ and

$$x_2 = \frac{1}{[1 + 4(\omega'\omega)e_i'\hat{V}_{11}e_i]^{\frac{1}{2}}} \neq 0, \tag{A31}$$

$$x_3 = -2(\omega'\omega)^{\frac{1}{2}} \hat{V}_{11}^{\frac{1}{2}} e_i x_2. \tag{A32}$$

In particular, the i-th element of x_3 is $e_i'x_3 = -2(\omega'\omega)^{\frac{1}{2}}(e_i'\hat{V}_{11}^{\frac{1}{2}}e_i)x_2 \neq 0$. Since $\check{\gamma}_0^K = -x_2/x_1$ and $\check{\gamma}_1^K = -x_3/x_1$, with our choice of p_0, the numerators for $\check{\gamma}_0^K$ and the i-th element of $\check{\gamma}_1^K$ are nonzero but their denominators are zero. As the choice of i is arbitrary, it remains to verify the derivative condition. From (A29),

$$x_1 = -\frac{1}{Y'Y - d}(Y'\omega x_2 + Y'Zx_3). \tag{A33}$$

Differentiating x_1 with respect to $\text{vec}(Z)$,

$$\begin{aligned} \frac{\partial x_1}{\partial \text{vec}(Z)} &= \frac{-\lambda}{(1-\lambda Y'Y)^2}\left[(1-\lambda Y'Y)\frac{\partial(Y'\omega x_2 + Y'Zx_3)}{\partial \text{vec}(Z)} - \delta\right] \\ &= \frac{-\lambda}{(1-\lambda Y'Y)^2}\left[(1-\lambda Y'Y)\left(Y'\omega\frac{\partial x_2}{\partial \text{vec}(Z)} + Y'Z\frac{\partial x_3}{\partial \text{vec}(Z)} + x_3 \otimes Y\right) - \delta\right], \end{aligned} \tag{A34}$$

where the last term is obtained using the identity $Y'Zx_3 = (x_3 \otimes Y)'\text{vex}(Z)$ and $\partial(a'y)/\partial y = a$, and

$$\delta = (Y'\omega x_2 + Y'Zx_3)\frac{\partial(Y'Y - d)}{\partial \text{vec}(Z)}. \tag{A35}$$

Evaluating (A34) at $p = p_0$, we have $\delta = 0_{N(K+1)}$ and hence

$$\left.\frac{\partial x_1}{\partial \text{vec}(Z)}\right|_{p=p_0} = \frac{4}{3}(x_3 \otimes q_{K+1}). \tag{A36}$$

Since both x_3 and q_{K+1} are nonzero vectors, $\frac{4}{3}(x_3 \otimes q_{K+1})$ is also a nonzero vector, and the derivative condition is satisfied. This completes the proof. Note that the normality assumption on $\hat{\mu}_2$ and $\hat{\beta}$ is not needed in the proof. All that is needed is that conditional on $\hat{\mu}_1$ and $\hat{V}_{11}$, the joint distribution of $\hat{\mu}_2$ and $\hat{\beta}$ is absolutely continuous with nonzero density. *Q.E.D.*

Proof of Propositions 3. Using a lemma from Sawa (1972), Hoque (1985, Theorems 1 and 2) and Magnus (1986, Theorem 6) show that, for a ratio of quadratic forms of normal random variables

$$Q = \frac{X'AX}{X'BX}, \tag{A37}$$

where $X \sim N(\mu_X, I_n)$, A is a symmetric matrix and B is a positive semidefinite matrix, we have

$$\begin{aligned} E[Q] &= \int_0^\infty \frac{\mu_X'(I_n+2tB)^{-1}A(I_n+2tB)^{-1}\mu_X + \mathrm{tr}((I_n+2tB)^{-1}A)}{|I_n+2tB|^{\frac{1}{2}}} \\ &\quad \times \exp\left(\frac{\mu_X'[(I_n+2tB)^{-1}-I_n]\mu_X}{2}\right)\mathrm{d}t \qquad \text{(A38)} \\ E[Q^2] &= \int_0^\infty \frac{t}{|I_n+2tB|^{\frac{1}{2}}}\exp\left(\frac{\mu_X'[(I_n+2tB)^{-1}-I_n]\mu_X}{2}\right)\times \\ &\quad \Big(\left[\mu_X'(I_n+2tB)^{-1}A(I_n+2tB)^{-1}\mu_X + \mathrm{tr}((I_n+2tB)^{-1}A)\right]^2 + 2\mathrm{tr}(((I_n+2tB)^{-1}A)^2) \\ &\quad + 4\mu_X'(I_n+2tB)^{-1}A(I_n+2tB)^{-1}A(I_n+2tB)^{-1}\mu_X\Big)\,\mathrm{d}t. \qquad \text{(A39)} \end{aligned}$$

Let $X = [Y_2',\ Z_2']'$ and let A and B be defined as in (103) and (104). Then $Q = (Z_2'\Lambda Y_2)/(Z_2'\Lambda Z_2)$ and $X \sim N(\mu_X, I_{2(N-1)})$ when conditional on $\hat{\mu}_1$ and $\hat{V}_{11}$, where μ_X is defined as in (102). Substituting A, B, and μ_X into (A38) and (A39),

$$\begin{aligned} & E^c\left[\hat{V}_{11}^{\frac{1}{2}}\frac{Z_2'\Lambda Y_2}{Z_2'\Lambda Z_2}\right] \\ =\ & \left(\int_0^\infty \frac{\hat{V}_{11}\eta'(\Lambda^{-1}+2tI_{N-1})^{-1}\eta}{|I_{N-1}+2t\Lambda|^{\frac{1}{2}}}\exp\left(\frac{\hat{V}_{11}\eta'[(I_{N-1}+2t\Lambda)^{-1}-I_{N-1}]\eta}{2}\right)\mathrm{d}t.\right)\bar{\gamma}_1, \qquad \text{(A40)} \\ & E^c\left[\hat{V}_{11}\left(\frac{Z_2'\Lambda Y_2}{Z_2'\Lambda Z_2}\right)^2\right] \\ =\ & \int_0^\infty \frac{t}{|I_{N-1}+2t\Lambda|^{\frac{1}{2}}}\exp\left(\frac{\hat{V}_{11}\eta'[(I_{N-1}+2t\Lambda)^{-1}-I_{N-1}]\eta}{2}\right) \\ & \times\Big[\bar{\gamma}_1^2\left(\hat{V}_{11}^2\left[\eta'(\Lambda^{-1}+2tI_{N-1})^{-1}\eta\right]^2 + \hat{V}_{11}\eta'\Lambda(I_{N-1}+2t\Lambda)^{-1}\Lambda\eta\right) \\ & + \hat{V}_{11}\left(\mathrm{tr}(\Lambda^2(I_{N-1}+2t\Lambda)^{-1}) + \hat{V}_{11}\eta'\Lambda(I_{N-1}+2t\Lambda)^{-2}\Lambda\eta\right)\Big]\,\mathrm{d}t. \qquad \text{(A41)} \end{aligned}$$

Using a change of variables $y = 1/(1+2t\lambda_{N-1})$ in (A40) and the fact that $1/(1+2t\lambda_i) = a_i y$, we obtain (108). This completes the proof. *Q.E.D.*

Proof of Lemma 3. Repeated use of integration by parts gives

$$\begin{aligned} b\int_0^1 y^{\frac{n}{2}} e^{b(y-1)}\mathrm{d}y &= 1-\frac{\frac{n}{2}}{b}+\frac{\left(\frac{n}{2}\right)\left(\frac{n}{2}-1\right)}{b^2}-\cdots+(-1)^{\frac{n}{2}}\frac{\left(\frac{n}{2}\right)!}{b^{\frac{n}{2}}}(1-e^{-b}) \\ &= \sum_{r=0}^{\frac{n}{2}}\frac{\left(\frac{n}{2}-r+1\right)_r}{(-b)^r}-\frac{\left(\frac{n}{2}\right)!e^{-b}}{(-b)^{\frac{n}{2}}} \end{aligned} \tag{A42}$$

for even n, and

$$\begin{aligned} b\int_0^1 y^{\frac{n}{2}} e^{b(y-1)}\mathrm{d}y &= 1-\frac{\frac{n}{2}}{b}+\frac{\left(\frac{n}{2}\right)\left(\frac{n}{2}-1\right)}{b^2}-\cdots+(-1)^{\frac{n-1}{2}}\frac{\left(\frac{n}{2}\right)\cdots\left(\frac{3}{2}\right)}{b^{\frac{n-1}{2}}}\left(1-\frac{\mathcal{D}(\sqrt{b})}{\sqrt{b}}\right) \\ &= \sum_{r=0}^{\frac{n-1}{2}}\frac{\left(\frac{n}{2}-r+1\right)_r}{(-b)^r}-\frac{\left(\frac{3}{2}\right)_{(n-1)/2}}{(-b)^{\frac{n-1}{2}}}\frac{\mathcal{D}(\sqrt{b})}{\sqrt{b}} \end{aligned} \tag{A43}$$

for odd n. This completes the proof. *Q.E.D.*

Proof of Lemma 4. As $N \geq 3$ for the first moment of $\check{\gamma}_1$ to exist, so $y^{\frac{N-3}{2}}$ is a nonincreasing function of N for $0 \leq y \leq 1$, and $\check{\kappa}$ is also a nonincreasing function of N for a fixed θ. It then follows that the upper bound of $\check{\kappa}$ is

$$\check{\kappa}=\frac{\theta}{2}e^{-\frac{\theta}{2}}\int_0^1 y^{\frac{N-3}{2}}e^{\frac{\theta}{2}y}\mathrm{d}y \leq \frac{\theta}{2}e^{-\frac{\theta}{2}}\int_0^1 e^{\frac{\theta}{2}y}\mathrm{d}y = 1-e^{-\frac{\theta}{2}} < 1. \tag{A44}$$

In order to show that $\check{\kappa}$ is an increasing function of θ, integration by parts gives

$$\frac{\partial\check{\kappa}}{\partial\theta}=\frac{1}{2}-\frac{N-3+\theta}{4}\int_0^1 y^{\frac{N-3}{2}}e^{\frac{\theta}{2}(y-1)}\mathrm{d}y. \tag{A45}$$

For $0 \leq y \leq 1$, $f(y) = 1 - y^c e^{c(1-y)} \geq 0$ for $c \geq 0$, because for $0 \leq y \leq 1$, $f'(y) = ce^{c(1-y)}y^{c-1}(y-1) \leq 0$ and $f(y)$ is nonincreasing. Since $f(1) = 0$, $f(y) \geq 0$ for $0 \leq y \leq 1$. Putting $c = \frac{N-3}{2}$ into $f(y)$, $y^{\frac{N-3}{2}} \leq e^{\left(\frac{N-3}{2}\right)(y-1)}$ for $0 \leq y \leq 1$ and hence

$$\frac{\partial\check{\kappa}}{\partial\theta} \geq \frac{1}{2}-\frac{N-3+\theta}{4}\int_0^1 e^{\left(\frac{N-3+\theta}{2}\right)(y-1)}\mathrm{d}y = \frac{1}{2}e^{-\left(\frac{N-3+\theta}{2}\right)} > 0. \tag{A46}$$

Using the inequality $e^{c(y-1)} \geq y^c$ for $0 < y < 1$, a lower bound for $\check{\kappa}$ is

$$\check{\kappa}=\frac{\theta}{2}\int_0^1 y^{\frac{N-3}{2}}e^{\frac{\theta}{2}(y-1)}\mathrm{d}y \geq \frac{\theta}{2}\int_0^1 y^{\frac{N-3}{2}}y^{\frac{\theta}{2}}\mathrm{d}y = \frac{\theta}{N-1+\theta} > 0. \tag{A47}$$

As $\theta \to \infty$, both the lower bound (A47) and the upper bound (A44) of $\check{\kappa}$ approach one. Therefore, $\lim_{\theta\to\infty}\check{\kappa} = 1$. This completes the proof. *Q.E.D.*

Proof of Lemma 5. When $\eta_1 = \cdots = \eta_{N-2} = 0$, use the fact that $\lambda_{N-1}^* = 1$ and $a_{N-1} = 1$ to write

$$\tilde{\kappa} = \frac{\hat{V}_{11}\eta_{N-1}^2}{2} \int_0^1 \left(\prod_{i=1}^{N-1} a_i \right)^{\frac{1}{2}} e^{\frac{\hat{V}_{11}\eta_{N-1}^2}{2}(y-1)} y^{\frac{N-3}{2}} \mathrm{d}y \leq \frac{\hat{V}_{11}\eta'\eta}{2} \int_0^1 e^{\frac{\hat{V}_{11}\eta'\eta}{2}(y-1)} y^{\frac{N-3}{2}} \mathrm{d}y = \check{\kappa}, \quad \text{(A48)}$$

where the inequality follows from the fact that $0 \leq a_i \leq 1$.

When $\eta_2 = \cdots = \eta_{N-1} = 0$, use a change of variables $y = 1/(1 + 2t\lambda_1)$ in (A40) instead of $y = 1/(1 + 2t\lambda_{N-1})$ to get

$$\tilde{\kappa} = \frac{\hat{V}_{11}}{2} \int_0^1 \left(\sum_{i=1}^{N-1} b_i \left(\frac{\lambda_i}{\lambda_1} \right) \eta_i^2 \right) \left(\prod_{i=1}^{N-1} b_i \right)^{\frac{1}{2}} e^{\frac{\hat{V}_{11}}{2} \sum_{i=1}^{N-1} \eta_i^2 (b_i y - 1)} y^{\frac{N-3}{2}} \mathrm{d}y, \quad \text{(A49)}$$

where $b_i = \lambda_1/[\lambda_i - (\lambda_i - \lambda_1)y]$. Putting $\eta_2 = \cdots = \eta_{N-1} = 0$ into (A49),

$$\tilde{\kappa} = \frac{\hat{V}_{11}\eta_1^2}{2} \int_0^1 \left(\prod_{i=1}^{N-1} b_i \right)^{\frac{1}{2}} e^{\frac{\hat{V}_{11}\eta_1^2}{2}(y-1)} y^{\frac{N-3}{2}} \mathrm{d}y \geq \frac{\hat{V}_{11}\eta'\eta}{2} \int_0^1 e^{\frac{\hat{V}_{11}\eta'\eta}{2}(y-1)} y^{\frac{N-3}{2}} \mathrm{d}y = \check{\kappa}, \quad \text{(A50)}$$

where the inequality follows from the fact that $b_i \geq 1$. Note that the inequalities are equalities if and only if $a_i = 1$ and $b_i = 1$ for $1 \leq i \leq N-1$, which hold if and only if $\lambda_1 = \lambda_{N-1}$. This completes the proof. *Q.E.D.*

Proof of Proposition 5. Replacing $T\hat{V}_{11}$ by $V_{11}v$ in the integral for the conditional mean, where $v \sim \chi_{T-1}^2$, and using the fact that the density function of v is

$$f(v) = \frac{1}{\Gamma\left(\frac{T-1}{2}\right)} \left(\frac{1}{2} \right)^{\frac{T-1}{2}} v^{\frac{T-3}{2}} e^{-\frac{v}{2}}, \quad \text{(A51)}$$

we can integrate the conditional mean using the density function of v. After some simplification and using the identity

$$\int_0^\infty v^n e^{-av} \mathrm{d}v = \frac{\Gamma(n+1)}{a^{n+1}} \quad \text{(A52)}$$

for $a > 0$, we obtain the unconditional mean. This completes the proof. *Q.E.D.*

Proof of Proposition 6. With the expression of $\tilde{\gamma}_1$ defined by (45) and using a change of variables of $y = 1/(1 + 2t\lambda_{N-1})$ in (A41),

$$\begin{aligned} E^c[\tilde{\gamma}_1^2] &= E^c\left[\hat{V}_{11} \left(\frac{Z_2'\Lambda Y_2}{Z_2'\Lambda Z_2} \right)^2 \right] \\ &= \frac{\hat{V}_{11}}{4} \int_0^1 \left[\bar{\gamma}_1^2 (\hat{V}_{11} g_1^2 y + g_4) + (\hat{V}_{11} g_2 y + g_3) \right] \left(\prod_{i=1}^{N-1} a_i \right)^{\frac{1}{2}} \\ &\quad \times e^{\frac{\hat{V}_{11}}{2} \sum_{i=1}^{N-1} \eta_i^2 (a_i y - 1)} y^{\frac{N-5}{2}} (1-y) \mathrm{d}y. \end{aligned} \quad \text{(A53)}$$

Since $\bar{\gamma}_1$ is independent of $\hat{V}_{11}$ and $E[\bar{\gamma}_1^2] = \gamma_1^2 + \frac{V_{11}}{T} = a$,

$$E[\tilde{\gamma}_1^2|\hat{V}_{11}] = \frac{\hat{V}_{11}}{4}\int_0^1 \left[\hat{V}_{11}(g_2 + ag_1^2)y + (g_3 + ag_4)\right]\left(\prod_{i=1}^{N-1} a_i\right)^{\frac{1}{2}} \times e^{\frac{\hat{V}_{11}}{2}\sum_{i=1}^{N-1}\eta_i^2(a_i y-1)} y^{\frac{N-5}{2}}(1-y)\mathrm{d}y. \tag{A54}$$

Then, by integrating the above expression with respect to the density function of $\hat{V}_{11}$ as in the proof of Proposition 5 and using the fact that $E[\tilde{\gamma}_1] = \tilde{\kappa}_u\gamma_1$, we obtain (129). This completes the proof. *Q.E.D.*

Proof of Proposition 7. By setting $\lambda_i^* = 1$ in the expression of $\mathrm{Var}[\check{\gamma}_1]$,

$$\mathrm{Var}[\check{\gamma}_1] = \frac{(T-1)V_{11}}{4T}\left[\frac{(T+1)V_{11}}{T}\varphi^d_{N-3,T+3}(\eta'\eta + a(\eta'\eta)^2) + \varphi^d_{N-5,T+1}(N-1+a\eta'\eta)\right] - \check{\kappa}_u^2\gamma_1^2. \tag{A55}$$

When $\lambda_i^* = 1$, denote $\phi_{m,n} = \varphi_{m,n}(1)$. Note that when $n = T+1$, we simply write $\phi_{m,T+1}$ as ϕ_m. Using these notations,

$$\mathrm{Var}[\check{\gamma}_1] = \frac{(T-1)V_{11}}{4T}\left[\frac{(T+1)V_{11}\eta'\eta}{T}(1+a\eta'\eta)(\phi_{N-3,T+3} - \phi_{N-1,T+3}) + (N-1+a\eta'\eta)(\phi_{N-5} - \phi_{N-3})\right] - \tilde{\kappa}_u^2\gamma_1^2. \tag{A56}$$

Using integration by parts,

$$\phi_{N-3,T+3} = \frac{2-(N-3)\phi_{N-5}}{(T+1)V_{11}\eta'\eta/T}, \tag{A57}$$

$$\phi_{N-1,T+3} = \frac{2-(N-1)\phi_{N-3}}{(T+1)V_{11}\eta'\eta/T}. \tag{A58}$$

Substituting these two expressions into (A56), we get (130). This completes the proof. *Q.E.D.*

Proof of Proposition 8. From (98) and (100) and denoting $C = \eta'[I_{N-1} - Z_2(Z_2'Z_2)^{-1}Z_2']\eta$, for $K = 1$,

$$\begin{aligned}\Delta &= \frac{E[((N-2)+\bar{\gamma}_1^2 C)\hat{V}_{11}D_5]}{T-N-1} \\ &= \frac{(N-2)E[\hat{V}_{11}D_5] + aE[\hat{V}_{11}CD_5]}{T-N-1} \\ &= \frac{1}{T-N-1}\left\{[(N-2)+a\eta'\eta]E\left[\frac{\hat{V}_{11}}{Z_2'Z_2}\right] - aE\left[\hat{V}_{11}\left(\frac{Z_2'\eta}{Z_2'Z_2}\right)^2\right]\right\}. \end{aligned} \tag{A59}$$

Using a similar method as in the proof of Proposition 3,

$$E\left[\frac{\hat{V}_{11}}{Z_2'Z_2}\right] = \frac{(T-1)V_{11}}{2T}\phi_{N-5}, \quad \text{(A60)}$$

$$E\left[\hat{V}_{11}\left(\frac{Z_2'\eta}{Z_2'Z_2}\right)^2\right] = \frac{(T-1)V_{11}\eta'\eta}{4T}\left[(N-2)\phi_{N-3}-(N-4)\phi_{N-5}\right]. \quad \text{(A61)}$$

Substituting these two expressions in (A59), we obtain the expression for Δ in Proposition 8. This completes the proof. *Q.E.D.*

Proof of Lemma 6. Applying a Taylor series expansion to x/y at (μ_x, μ_y) and dropping all the terms of order higher than two,

$$E\left[\frac{x}{y}\right] \approx \frac{\mu_x}{\mu_y}\left(1-\frac{\mathrm{Cov}[x,y]}{\mu_x\mu_y}+\frac{\mathrm{Var}[y]}{\mu_y^2}\right), \quad \text{(A62)}$$

where $\mu_x = E[x]$ and $\mu_y = E[y]$. From (105), $E^c[\tilde{\gamma}_1] = \hat{V}_{11}^{\frac{1}{2}}E^c[(X'AX)/(X'BX)]$. Let $x = X'AX$, $y = X'BX$ and using the results from Graybill (1983, p.367), we have

$$\mu_x = \hat{V}_{11}^{\frac{1}{2}}(\eta'\Lambda\eta)\bar{\gamma}_1, \quad \text{(A63)}$$

$$\mu_y = \mathrm{tr}(\Lambda)+\hat{V}_{11}(\eta'\Lambda\eta), \quad \text{(A64)}$$

$$\mathrm{Var}[y] = 2\mathrm{tr}(\Lambda^2)+4\hat{V}_{11}(\eta'\Lambda^2\eta), \quad \text{(A65)}$$

$$\mathrm{Cov}[x,y] = 2\bar{\gamma}_1(\eta'\Lambda^2\eta)\hat{V}_{11}^{\frac{1}{2}}. \quad \text{(A66)}$$

Therefore,

$$\begin{aligned}\tilde{\kappa} &= \frac{\hat{V}_{11}^{\frac{1}{2}}E^c[\frac{X'AX}{X'BX}]}{\bar{\gamma}_1} \\ &\approx \frac{\theta_1^3+2\mathrm{tr}(\Lambda)\theta_1^2+\left((\mathrm{tr}(\Lambda))^2+2\mathrm{tr}(\Lambda^2)+2\theta_2\right)\theta_1-2\mathrm{tr}(\Lambda)\theta_2}{(\mathrm{tr}(\Lambda)+\theta_1)^3} = \tilde{\kappa}^a, \quad \text{(A67)}\end{aligned}$$

where $\theta_1 = \hat{V}_{11}(\eta'\Lambda\eta)$ and $\theta_2 = \hat{V}_{11}(\eta'\Lambda^2\eta)$.

For $\breve{\kappa}^a$, let $\Lambda = I_{N-1}$. Then $\theta_1 = \theta_2 = \theta = \hat{V}_{11}\eta'\eta$ and (A67) becomes

$$\breve{\kappa} \approx \frac{\theta[(N-1+\theta)^2+2\theta]}{(N-1+\theta)^3} = \breve{\kappa}^a. \quad \text{(A68)}$$

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

Proof of Lemma 7. Let

$$\hat{A} = [1_N,\ \hat{\beta}]'\hat{\Sigma}^{-1}[1_N,\ \hat{\beta}], \tag{A69}$$

$$\check{A} = [1_N,\ \hat{\beta}]'\Sigma^{-1}[1_N,\ \hat{\beta}]. \tag{A70}$$

From Theorem 3.2.11 of Muirhead (1982), conditional on $\hat{\beta}$,

$$\hat{A}^{-1} \sim W_2(T-N, \check{A}^{-1}/T). \tag{A71}$$

Note that the $(2,2)$ elements of $\hat{A}^{-1}$ and $\check{A}^{-1}$ are given by $T/\hat{\eta}'\hat{\eta}$ and $T\hat{V}_{11}/Z_2'Z_2$, respectively, where

$$Z_2 = \sqrt{T}P'\Sigma^{-\frac{1}{2}}\hat{\beta}\hat{V}_{11}^{\frac{1}{2}} \sim N(\eta\hat{V}_{11}^{\frac{1}{2}}, I_{N-1}). \tag{A72}$$

Therefore, conditional on $\hat{\beta}$, from (A71),

$$\frac{T}{\hat{\eta}'\hat{\eta}} \sim W_1\left(T-N, \frac{\hat{V}_{11}}{Z_2'Z_2}\right) \tag{A73}$$

and hence

$$U_2 = \frac{TZ_2'Z_2}{\hat{V}_{11}\hat{\eta}'\hat{\eta}} \sim \chi^2_{T-N}. \tag{A74}$$

Since the distribution of U_2 is independent of $\hat{\beta}$ and hence independent of $Z_2'Z_2$,

$$\hat{V}_{11}\hat{\eta}'\hat{\eta} \sim \frac{TZ_2'Z_2}{U_2}. \tag{A75}$$

Finally, conditional on $\hat{V}_{11}$, $Z_2'Z_2 \sim \chi^2_{N-1}(\hat{V}_{11}\eta'\eta)$. Using the definition of the noncentral F-distribution, we then obtain the distribution of $\hat{V}_{11}\hat{\eta}'\hat{\eta}$. The expectation of $\hat{V}_{11}\hat{\eta}'\hat{\eta}$ is simply obtained from the expected value of a noncentral F-distribution, which is available from Johnson, Kotz, and Balakrishnan (1995, Ch. 30). This completes the proof. *Q.E.D.*

Proof of Lemma 8. The Bayes estimator with respect to the loss function in (140) is the estimator that minimizes the posterior risk

$$\int_0^\infty \left[\frac{\hat{\theta}}{\theta} - \log\left(\frac{\hat{\theta}}{\theta}\right)\right] f(\theta|z)\mathrm{d}\theta, \tag{A76}$$

where $f(\theta|z)$ is the posterior density of θ and $z = \hat{V}_{11}\hat{\eta}'\hat{\eta}/T$. Taking the derivative of (A76) with respect to $\hat{\theta}$,

$$\hat{\theta} = \frac{1}{\int_0^\infty \theta^{-1}f(\theta|z)\mathrm{d}\theta}. \tag{A77}$$

As the posterior of θ is

$$f(\theta|z) = \frac{f(z|\theta)\pi(\theta)}{\int_0^\infty f(z|\theta)\pi(\theta)\mathrm{d}\theta} = \frac{\theta^b f(z|\theta)}{\int_0^\infty \theta^b f(z|\theta)\mathrm{d}\theta}, \tag{A78}$$

substituting (A78) in (A77),

$$\hat{\theta} = \frac{\int_0^\infty \theta^b f(\theta|z)\mathrm{d}\theta}{\int_0^\infty \theta^{b-1} f(\theta|z)\mathrm{d}\theta}. \tag{A79}$$

From Johnson, Kotz, and Balakrishnan (1995, Ch. 30, p.484), the density function of z is

$$f(z|\theta) = \frac{e^{-\frac{\theta}{2}} z^{\frac{N-3}{2}}}{B\left(\frac{N-1}{2}, \frac{T-N}{2}\right)(1+z)^{\frac{T-1}{2}}} \sum_{r=0}^{\infty} \left[\frac{z}{2(1+z)}\right]^r \frac{\left(\frac{T-1}{2}\right)_r}{\left(\frac{N-1}{2}\right)_r} \frac{\theta^r}{r!}, \tag{A80}$$

where $B(a,b)$ is the beta function. Using (A52),

$$\begin{aligned}\int_0^\infty \theta^d f(z|\theta)\mathrm{d}\theta &= \frac{z^{\frac{N-3}{2}} 2^{d+1}}{B\left(\frac{N-1}{2}, \frac{T-N}{2}\right)(1+z)^{\frac{T-1}{2}}} \sum_{r=0}^{\infty} \Gamma(d+r+1) \frac{\left(\frac{T-1}{2}\right)_r}{\left(\frac{N-1}{2}\right)_r} \frac{\left(\frac{z}{1+z}\right)^r}{r!} \\ &= \frac{z^{\frac{N-3}{2}} 2^{d+1}\Gamma(d+1)}{B\left(\frac{N-1}{2}, \frac{T-N}{2}\right)(1+z)^{\frac{T-1}{2}}} {}_2F_1\left(d+1, \frac{T-1}{2}, \frac{N-1}{2}, \frac{z}{1+z}\right). \end{aligned} \tag{A81}$$

Using this identity in both the numerator and denominator of (A79), we obtain our Bayes estimator in (141). This completes the proof. *Q.E.D.*

Proof of Lemma 9. As in the proof for Lemma 6, we apply a second order Taylor series expansion to $\mathrm{Var}[x/y]$

$$\mathrm{Var}\left[\frac{x}{y}\right] \approx \frac{\mathrm{Var}[x]}{\mu_y^2} - \frac{2\mu_x \mathrm{Cov}[x,y]}{\mu_y^3} + \frac{\mu_x^2 \mathrm{Var}[y]}{\mu_y^4}. \tag{A82}$$

To derive an approximation for $\mathrm{Var}[\tilde{\gamma}_1|\hat{V}_{11}] = \hat{V}_{11}\mathrm{Var}[(X'AX/X'BX)|\hat{V}_{11}]$, we write $x = X'AX$ and $y = X'BX$. Conditional on $\hat{V}_{11}$ only,

$$X \equiv \begin{bmatrix} Y_2 \\ Z_2 \end{bmatrix} \sim N\left(\begin{bmatrix} \eta\gamma_1 \\ \eta\hat{V}_{11}^{\frac{1}{2}} \end{bmatrix}, \begin{bmatrix} I_{N-1} + V_{11}\eta\eta'/T & O_{(N-1)\times(N-1)} \\ O_{(N-1)\times(N-1)} & I_{N-1} \end{bmatrix}\right). \tag{A83}$$

From Graybill (1983, p.367)

$$\mu_x = \hat{V}_{11}^{\frac{1}{2}}(\eta'\Lambda\eta)\gamma_1, \tag{A84}$$

$$\mathrm{Var}[x] = \mathrm{tr}(\Lambda^2) + \hat{V}_{11}V_{11}(\eta'\Lambda\eta)^2/T + [V_{11}/T + (\hat{V}_{11} + \gamma_1^2)](\eta'\Lambda^2\eta), \tag{A85}$$

$$\mathrm{Cov}[x,y] = 2\gamma_1(\eta'\Lambda^2\eta)\hat{V}_{11}^{\frac{1}{2}}, \tag{A86}$$

while μ_y and Var[y] are the same as (A64) and (A65), respectively. Substituting these expressions into (A82), we have (142).

Let $\Lambda = I_{N-1}$ in (142), we obtain (143) as $\mathrm{Var}[\check{\gamma}_1|\hat{V}_{11}]$. To find $\mathrm{Var}[\hat{\gamma}_1|\hat{V}_{11}]$, we use the formula $\mathrm{Var}[X] = E[\mathrm{Var}[X|Y]] + \mathrm{Var}[E[X|Y]]$. From (97) and (98),

$$\begin{aligned}\mathrm{Var}[\hat{\gamma}_1|\hat{V}_{11}] &= E[\mathrm{Var}^c[\hat{\gamma}_1]|\hat{V}_{11}] + \mathrm{Var}[E^c[\hat{\gamma}_1]|\hat{V}_{11}] \\ &= E[\mathrm{Var}^c[\check{\gamma}_1]|\hat{V}_{11}] + E[\Delta^c|\hat{V}_{11}] + \mathrm{Var}[E^c[\check{\gamma}_1]|\hat{V}_{11}] \\ &= \mathrm{Var}[\check{\gamma}_1|\hat{V}_{11}] + E[\Delta^c|\hat{V}_{11}], \end{aligned} \tag{A87}$$

where

$$\begin{aligned} & E[\Delta^c|\hat{V}_{11}] \\ = \;& \frac{1}{T-N-1}\left\{\left[(N-1)\hat{V}_{11} + \left(\gamma_1^2 + \frac{V_{11}}{T}\right)\theta\right] E\left[\left.\frac{1}{Z_2'Z_2}\right|\hat{V}_{11}\right] - E\left[\check{\gamma}_1^2|\hat{V}_{11}\right]\right\} \\ = \;& \frac{1}{T-N-1}\left\{\left[(N-1)\hat{V}_{11} + a\theta\right] E\left[\left.\frac{1}{Z_2'Z_2}\right|\hat{V}_{11}\right] - \left(\mathrm{Var}\left[\check{\gamma}_1|\hat{V}_{11}\right] + E[\check{\gamma}_1|\hat{V}_{11}]^2\right)\right\}. \end{aligned} \tag{A88}$$

To approximate $E[\frac{1}{Z_2'Z_2}|\hat{V}_{11}]$, we plug in $x = 1$ and $y = Z_2'Z_2$ in (A62) to obtain the expression for c. Combining all the terms in (A87) and (A88), we have (144). This completes the proof. *Q.E.D.*

References

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

Ahn, S., and C. Gadarowski, 1999, "Two-pass Cross-sectional Regression of Factor Pricing Model: Minimum Distance Approach," working paper, Arizona State University.

Amsler, C. E., and P. Schmidt, 1985, "A Monte Carlo Investigation of the Accuracy of Multivariate CAPM Tests," *Journal of Financial Economics*, 14, 359–375.

Anderson, T. W., 1982, "Some Recent Developments on the Distribution of Single Equation Estimators," in W. Hildenbrand (ed.), *Advances in Econometrics*, Cambridge University Press, Cambridge.

Ansley, C. F., R. Kohn, and T. S. Shively, 1992, "Computing p-values of the Generalized Durbin-Watson and Other Invariant Test Statistics," *Journal of Econometrics*, 54, 277-300.

Banz, R., 1981, "The Relations between Returns and Market Values of Common Stocks," *Journal of Financial Economics*," 9, 3–18.

Black, F., 1972, "Capital Market Equilibrium with Restricted Borrowing," *Journal of Business*, 45, 444–454.

Black, F., M. C. Jensen, and M. Scholes, 1972, "The Capital Asset Pricing Model: Some Empirical Findings," in M. C. Jensen (ed.), *Studies in the Theory of Capital Markets*, Praeger, New York.

Breeden, D. T., 1979, "An Intertemporal Asset Pricing Model with Stochastic Consumption and Investment Opportunities," *Journal of Financial Economics*, 7, 265–296.

Breeden, D. T., M. Gibbons, and R. H. Litzenberger, 1989, "Empirical Tests of the Consumption-oriented CAPM," *Journal of Finance*, 44, 231–262.

Davies, R. B., 1980, "The Distribution of a Linear Combination of χ^2 Random Variables," *Applied Statistics*, 29, 323–333.

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

Farebrother, R. W., 1990, "The Distribution of a Quadratic Form in Normal Variables," *Applied Statistics*, 39, 294–309.

Forchini, G., 2001, "The Distribution of a Ratio of Quadratic Forms in Noncentral Normal Variables," working paper, University of York.

Fourdrinier D., A. Philippe, and C. P. Robert, 2000, "Estimation of a Non-centrality Parameter under Stein-type-like Losses," *Journal of Statistical Planning and Inference*, 87, 43–54.

Fuller, W. A., 1987, *Measurement Error Models*, Wiley, New York.

Gibbons, M. R., S. A. Ross, and J. Shanken, 1989, "A Test of the Efficiency of a Given Portfolio," *Econometrica*, 57, 1121–1152.

Graybill, F. A., 1983, *Matrices with Applications in Statistics* (2nd ed.), Wadsworth and Brooks/Cole, Pacific Grove, C. A.

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

Imhof, J. P., 1961, "Computing the Distribution of Quadratic Forms in Normal Variables," *Biometrika*, 48, 419–426.

Jagannathan, R., and Z. Wang, 1998, "An Asymptotic Theory for Estimating Beta-Pricing Models Using Cross-sectional Regression," *Journal of Finance*, 53, 1285–1309.

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

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

Kim, D., 1995, "The Errors in the Variables Problem in the Cross-section of Expected Stock Returns," *Journal of Finance*, 50, 1605–1634.

Kinal, T. W., 1980, "The Existence of Moments of k-class Estimators," *Econometrica*, 48, 241–249.

Kubokawa, T., C. P. Robert, and A. K. Md. E. Saleh, 1993, "Estimation of Noncentrality Parameters," *Canadian Journal of Statistics*, 21, 45–57.

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

Litzenberger, R. H., and K. Ramaswamy, 1979, "The Effect of Personal Taxes and Dividends on Capital Asset Prices," *Journal of Financial Economics*, 7, 163–195.

Lu, Z., and M. L. King, 2002, "Improving the Numerical Technique for Computing the Accumulated Distribution of a Quadratic Form in Normal Variables," *Econometric Reviews*, 21, 149–165.

MacKinlay, A. C., 1985, "Analysis of Multivariate Financial Tests," Ph.D. dissertation, Graduate School of Business, University of Chicago.

Magnus, J. R., 1986, "The Exact Moments of a Ratio of Quadratic Forms in Normal Variables," *Annales D'économie et de Statistique*, 4, 95–109.

Magnus, J. R., 1990, "On Certain Moments Relating to Ratios of Quadratic Forms in Normal Variables: Further Results," *Sankhyā: The Indian Journal of Statistics*, 52, Series B, Part 1, 1–13.

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

Muirhead, R. J., 1982, *Aspects of Multivariate Statistical Theory*, Wiley, New York.

Paolella, M. S., 2003, "Computing Moments of Ratios of Quadratic Forms in Normal Variables," *Computational Statistics and Data Analysis*, 42, 313-331.

Phillips, P. C. B., 1983, "Exact Small Sample Theory in the Simultaneous Equations Model," in M. D. Intriligator and Z. Griliches(eds.), *Handbooks of Econometrics*, North-Holland, Amsterdam, 449–516.

Phillips, P. C. B., 1984, "The Exact Distribution of LIML: I," *International Economic Review*, 25, 249–261.

Phillips, P. C. B., 1985, "The Exact Distribution of LIML: II," *International Economic Review*, 26, 21–36.

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

Rukhin, Andrew L., 1993, "Estimation of the Noncentrality Parameter of an F-distribution," *Journal of Statistical Planning and Inference*, 35, 201–211.

Sargan, J. D., 1976, "Existence of the Moments of Estimated Reduced Form Coefficients," Discussion Paper No. A6, London School of Economics.

Sawa, T., 1972, "Finite Sample Properties of the k-class Estimators," *Econometrica*, 40, 653–680.

Shanken, J., 1992, "On the Estimation of Beta-pricing Models," *Review of Financial Studies*, 5, 1–33.

Shanken, J., and M. Weinstein, 1990, "Macroeconomic Variables and Asset Pricing: Further Results," working paper, University of Rochester.

Shanken, J., and G. Zhou, 2000, "Asset Pricing Tests in the Two-pass Cross-sectional Regression Model," working paper, Washington University in St. Louis.

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

Zhou, G., 1993, "Asset Pricing Test under Alternative Distributions," *Journal of Finance*, 48, 1925–1942.

Table 1
Summary of Parameters in Simulation Experiments

Panel A: Factor with high explanatory power ($V_{11}^{\frac{1}{2}} = 4.092$)

	$N = 10$		$N = 25$		$N = 100$	
σ_β	0.049		0.109		0.245	
$\bar{\sigma}_\epsilon$	2.362		2.934		3.816	
$V_{11}\eta'\eta/(N-1)$	0.520		1.728		2.071	
i	$\lambda_i/\bar{\lambda}$	$\|\eta_i^*\|$	$\lambda_i/\bar{\lambda}$	$\|\eta_i^*\|$	$\lambda_i/\bar{\lambda}$	$\|\eta_i^*\|$
1	6.551	2.759	9.749	5.530	28.128	5.601
2	1.000	2.418	4.712	2.250	6.159	20.768
3	0.431	3.013	1.686	0.613	3.799	9.348
$N-3$	0.154	0.206	0.177	1.411	0.135	3.566
$N-2$	0.130	0.202	0.158	2.175	0.125	1.244
$N-1$	0.112	0.137	0.157	2.841	0.122	1.315

Panel B: Factor with low explanatory power ($V_{11}^{\frac{1}{2}} = 0.747$)

	$N = 10$		$N = 25$		$N = 100$	
σ_β	0.255		0.336		0.481	
$\bar{\sigma}_\epsilon$	5.088		5.428		6.079	
$V_{11}\eta'\eta/(N-1)$	0.227		0.309		0.351	
i	$\lambda_i/\bar{\lambda}$	$\|\eta_i^*\|$	$\lambda_i/\bar{\lambda}$	$\|\eta_i^*\|$	$\lambda_i/\bar{\lambda}$	$\|\eta_i^*\|$
1	6.356	16.606	11.554	17.630	25.705	20.859
2	1.012	2.900	4.236	4.903	10.583	12.235
3	0.488	6.990	1.449	10.277	4.369	2.712
$N-3$	0.165	2.330	0.152	9.710	0.117	6.179
$N-2$	0.140	0.868	0.136	10.246	0.110	1.797
$N-1$	0.127	4.552	0.116	1.875	0.098	14.304

The table presents a summary of the parameters of the test assets and factors used in our simulation experiments. The factor in Panel A is chosen to have high explanatory power on the returns of the test assets whereas the factor in Panel B is chosen to have low explanatory power on the returns of the test assets. For each panel, the table presents the standard deviation (in percentages) of the factor ($V_{11}^{\frac{1}{2}}$) and three sets of parameters, each corresponding to different numbers of test assets (N). For each case, the table presents the cross-sectional standard deviation of beta (σ_β), the average standard deviation (in percentages) of the regression residuals for the test assets ($\bar{\sigma}_\epsilon$), and the GLS cross-sectional variance of the normalized beta ($V_{11}\eta'\eta/(N-1)$) for $T = 100$. In addition, the table also presents the three largest and three smallest normalized eigenvalues ($\lambda_i/\bar{\lambda}$) of the matrix $\Sigma^{\frac{1}{2}}M\Sigma^{\frac{1}{2}}$ and the corresponding absolute values of $\eta_i^* = p_i'\Sigma^{-\frac{1}{2}}\beta$, where p_i is the eigenvector associated with λ_i.

Table 2
Unconditional Percentage Biases of Risk Premium Estimators from Second-Pass Cross-Sectional Regressions

Panel A: Factor with high explanatory power

	$N = 10$		$N = 25$		$N = 100$	
T	OLS	GLS	OLS	GLS	OLS	GLS
60	−56.8	−75.8	−29.9	−48.9	−16.6	−45.2
120	−33.2	−59.9	−14.8	−31.7	−8.6	−28.8
240	−12.3	−41.5	−6.7	−18.5	−4.4	−16.7
360	−4.7	−31.4	−4.3	−13.1	−2.9	−11.7
480	−1.6	−25.2	−3.1	−10.1	−2.2	−9.0
600	−0.2	−21.0	−2.4	−8.2	−1.8	−7.4

Panel B: Factor with low explanatory power

	$N = 10$		$N = 25$		$N = 100$	
T	OLS	GLS	OLS	GLS	OLS	GLS
60	−61.3	−88.0	−65.1	−84.5	−66.3	−82.9
120	−36.4	−78.0	−43.7	−72.7	−47.2	−70.5
240	−11.1	−63.1	−22.0	−56.7	−28.3	−54.2
360	−1.2	−52.6	−12.7	−46.3	−19.6	−44.0
480	2.4	−44.9	−8.1	−39.1	−14.8	−37.0
600	3.6	−39.0	−5.7	−33.8	−11.8	−32.0

The table presents the unconditional biases of the OLS and GLS CSR estimators of the risk premium (γ_1) as a percentage of the true value of the risk premium for different lengths of beta estimation period (T) and for different numbers of test assets (N) when the factors and returns are multivariate normally distributed. The factor in Panel A is chosen to have high explanatory power on the returns of the test assets whereas the factor in Panel B is chosen to have low explanatory power on the returns of the test assets. The true GLS and the estimated GLS CSR estimators of γ_1 have the same bias, except that the estimated GLS is infeasible for $T = 60$ and $N = 100$.

Table 3
Asymptotic and Finite Sample Standard Deviation of Risk Premium Estimators from Second-Pass Cross-Sectional Regressions

Panel A: Factor with high explanatory power

	$N = 10$				$N = 25$				$N = 100$			
	OLS		GLS		OLS		GLS		OLS		GLS	
T	Asy.	Exact	Asy.	Exact	Asy.	Exact	Asy.	Exact	Asy.	Exact	Asy.	Exact
60	3.952	2.696	2.525	1.445	1.693	1.269	0.983	0.828	0.807	0.692	n/a	n/a
120	2.794	2.319	1.785	1.245	1.197	1.030	0.695	0.603	0.570	0.527	0.457	0.614
240	1.976	1.846	1.262	1.025	0.846	0.788	0.492	0.451	0.403	0.387	0.323	0.313
360	1.613	1.567	1.031	0.891	0.691	0.660	0.401	0.377	0.329	0.321	0.264	0.254
480	1.397	1.382	0.893	0.798	0.598	0.579	0.348	0.331	0.285	0.279	0.229	0.221
600	1.250	1.248	0.798	0.729	0.535	0.521	0.311	0.299	0.255	0.251	0.204	0.198

Panel B: Factor with low explanatory power

	$N = 10$				$N = 25$				$N = 100$			
	OLS		GLS		OLS		GLS		OLS		GLS	
T	Asy.	Exact	Asy.	Exact	Asy.	Exact	Asy.	Exact	Asy.	Exact	Asy.	Exact
60	0.768	0.517	0.682	0.283	0.653	0.338	0.367	0.188	0.456	0.194	n/a	n/a
120	0.543	0.472	0.482	0.258	0.462	0.313	0.260	0.152	0.322	0.186	0.134	0.159
240	0.384	0.398	0.341	0.228	0.327	0.272	0.184	0.128	0.228	0.167	0.095	0.076
360	0.314	0.341	0.278	0.207	0.267	0.241	0.150	0.114	0.186	0.151	0.078	0.063
480	0.272	0.299	0.241	0.191	0.231	0.217	0.130	0.104	0.161	0.138	0.067	0.056
600	0.243	0.267	0.216	0.178	0.207	0.198	0.116	0.096	0.144	0.127	0.060	0.051

The table presents the unconditional standard deviations (in percentages per month) of the OLS and estimated GLS CSR estimators of the the risk premium (γ_1) for different lengths of sample period (T) and for different numbers of test assets (N) when the factors and returns are multivariate normally distributed. The factor in Panel A is chosen to have high explanatory power on the returns of the test assets whereas the factor in Panel B is chosen to have low explanatory power on the returns of the test assets. The value of γ_1 is assumed to be 0.6% per month for Panel A and 0.028% per month for Panel B. For each panel, two measures of standard deviations are presented, the first based on the asymptotic variance formula from Shanken (1992) and the second based on the exact finite sample standard deviation of the estimators.

Table 4
Unconditional Percentage Biases of Risk Premium Estimators from Second-Pass Cross-Sectional Regressions under Nonnormality

Panel A: Factor with high explanatory power

	$N = 10$		$N = 25$		$N = 100$	
T	OLS	GLS	OLS	GLS	OLS	GLS
60	−52.5	−73.0	−29.7	−43.0	−17.1	n/a
120	−33.1	−57.7	−15.0	−28.6	−8.8	−21.3
240	−12.5	−40.3	−6.9	−17.2	−4.3	−13.2
360	−4.3	−30.7	−4.3	−12.5	−2.7	−9.6
480	−1.7	−24.7	−2.9	−9.7	−2.0	−7.7
600	0.1	−20.5	−2.2	−7.9	−1.6	−6.4

Panel B: Factor with low explanatory power

	$N = 10$		$N = 25$		$N = 100$	
T	OLS	GLS	OLS	GLS	OLS	GLS
60	−60.3	−86.1	−65.5	−81.7	−67.5	n/a
120	−37.5	−74.2	−43.9	−69.9	−47.7	−61.7
240	−11.7	−61.6	−23.5	−54.8	−28.7	−45.8
360	1.0	−50.2	−12.1	−45.2	−19.3	−37.5
480	3.9	−40.5	−7.3	−38.4	−15.0	−32.8
600	5.5	−34.4	−4.9	−32.7	−11.6	−28.6

The table presents the unconditional biases of the OLS and estimated GLS CSR estimators of the risk premium (γ_1) as a percentage of the true value of the risk premium for different lengths of the beta estimation period (T) and for different numbers of test assets (N). The factor in Panel A is chosen to have high explanatory power on the returns of the test assets whereas the factor in Panel B is chosen to have low explanatory power on the returns of the test assets. The returns on the test assets have the same mean, variance and betas as those used in Table 2 except that the factors here have a t-distribution with five degrees of freedom, and the residuals of the test assets have a multivariate t-distribution with five degrees of freedom. The results in the table are based on 100,000 simulations.

Table 5
Unconditional Standard Deviation of Risk Premium Estimators from Second-Pass Cross-Sectional Regressions under Nonnormality

Panel A: Factor with high explanatory power

	$N = 10$		$N = 25$		$N = 100$	
T	OLS	GLS	OLS	GLS	OLS	GLS
60	2.697	1.417	1.277	0.787	0.697	n/a
120	2.311	1.220	1.027	0.584	0.528	0.554
240	1.839	1.011	0.787	0.440	0.386	0.302
360	1.559	0.884	0.659	0.370	0.320	0.247
480	1.377	0.792	0.576	0.326	0.279	0.216
600	1.246	0.724	0.519	0.296	0.251	0.194

Panel B: Factor with low explanatory power

	$N = 10$		$N = 25$		$N = 100$	
T	OLS	GLS	OLS	GLS	OLS	GLS
60	0.525	0.281	0.345	0.184	0.205	n/a
120	0.473	0.255	0.313	0.148	0.189	0.148
240	0.394	0.226	0.270	0.125	0.167	0.072
360	0.337	0.206	0.239	0.112	0.150	0.060
480	0.297	0.190	0.215	0.102	0.137	0.054
600	0.266	0.177	0.196	0.095	0.127	0.049

The table presents the unconditional standard deviations (in percentages per month) of the OLS and estimated GLS CSR estimators of the risk premium (γ_1) for different lengths of sample period (T) and for different numbers of test assets (N). The factor in Panel A is chosen to have high explanatory power on the returns of the test assets whereas the factor in Panel B is chosen to have low explanatory power on the returns of the test assets. The value of γ_1 is assumed to be 0.6% per month for Panel A and 0.028% per month for Panel B. The returns on the test assets have the same mean, variance and betas as those used in Table 3 except that the factors here have a t-distribution with five degrees of freedom, and the residuals of the test assets have a multivariate t-distribution with five degrees of freedom. The results in the table are based on 100,000 simulations.

Table 6
Unconditional Percentage Biases of Bias-Adjusted Estimators of Risk Premium from Second-Pass Cross-Sectional Regressions

Panel A: Factor with high explanatory power

	$N = 10$		$N = 25$		$N = 100$	
T	OLS	GLS	OLS	GLS	OLS	GLS
60	−23.5	−23.1	−4.7	17.2	n/a	n/a
120	3.7	10.4	1.4	5.7	−0.5	2.1
240	13.5	21.5	1.2	1.0	−0.1	0.1
360	11.1	15.5	0.7	0.4	0.0	0.0
480	8.2	9.8	0.5	0.3	0.0	0.0
600	6.1	6.1	0.3	0.2	0.0	0.0

Panel B: Factor with low explanatory power

	$N = 10$		$N = 25$		$N = 100$	
T	OLS	GLS	OLS	GLS	OLS	GLS
60	−33.7	−58.1	−41.9	−30.5	n/a	n/a
120	2.2	−26.8	−13.4	11.1	−20.6	6.8
240	24.6	7.0	4.6	25.0	−4.8	11.0
360	23.8	19.6	7.1	17.3	−0.9	3.1
480	18.1	22.4	6.4	10.1	0.1	1.5
600	13.1	21.1	5.1	6.0	0.4	0.9

The table presents the unconditional biases of the bias-adjusted estimators of the risk premium (γ_1) from the OLS and estimated GLS CSR, as a percentage of the true value of the risk premium for different lengths of beta estimation period (T) and for different numbers of test assets (N) when the factors and returns are multivariate normally distributed. The factor in Panel A is chosen to have high explanatory power on the returns of the test assets whereas the factor in Panel B is chosen to have low explanatory power on the returns of the test assets. The results in the table are based on 100,000 simulations.

Table 7
Unconditional Standard Deviation of Bias-Adjusted Estimators of Risk Premium from Second-Pass Cross-Sectional Regressions

Panel A: Factor with high explanatory power

	$N = 10$		$N = 25$		$N = 100$	
T	OLS	GLS	OLS	GLS	OLS	GLS
60	7.757	5.988	1.777	2.266	n/a	n/a
120	5.154	4.475	1.240	0.995	0.575	0.907
240	2.821	2.701	0.860	0.566	0.405	0.377
360	1.976	1.797	0.697	0.439	0.330	0.288
480	1.583	1.321	0.602	0.371	0.286	0.243
600	1.356	1.054	0.537	0.327	0.256	0.214

Panel B: Factor with low explanatory power

	$N = 10$		$N = 25$		$N = 100$	
T	OLS	GLS	OLS	GLS	OLS	GLS
60	1.743	1.287	0.665	1.052	n/a	n/a
120	1.386	1.122	0.547	0.778	0.282	0.676
240	0.959	0.866	0.393	0.456	0.223	0.200
360	0.615	0.679	0.309	0.291	0.186	0.119
480	0.446	0.543	0.258	0.207	0.162	0.091
600	0.348	0.441	0.225	0.163	0.145	0.076

The table presents the unconditional standard deviations (in percentages per month) of the bias-adjusted estimators of the risk premium (γ_1) from the OLS and estimated GLS CSR for different lengths of sample period (T) and for different numbers of test assets (N) when the factors and returns are multivariate normally distributed. The factor in Panel A is chosen to have high explanatory power on the returns of the test assets whereas the factor in Panel B is chosen to have low explanatory power on the returns of the test assets. The value of γ_1 is assumed to be 0.6% per month for Panel A and 0.028% per month for Panel B. The results in the table are based on 100,000 simulations.

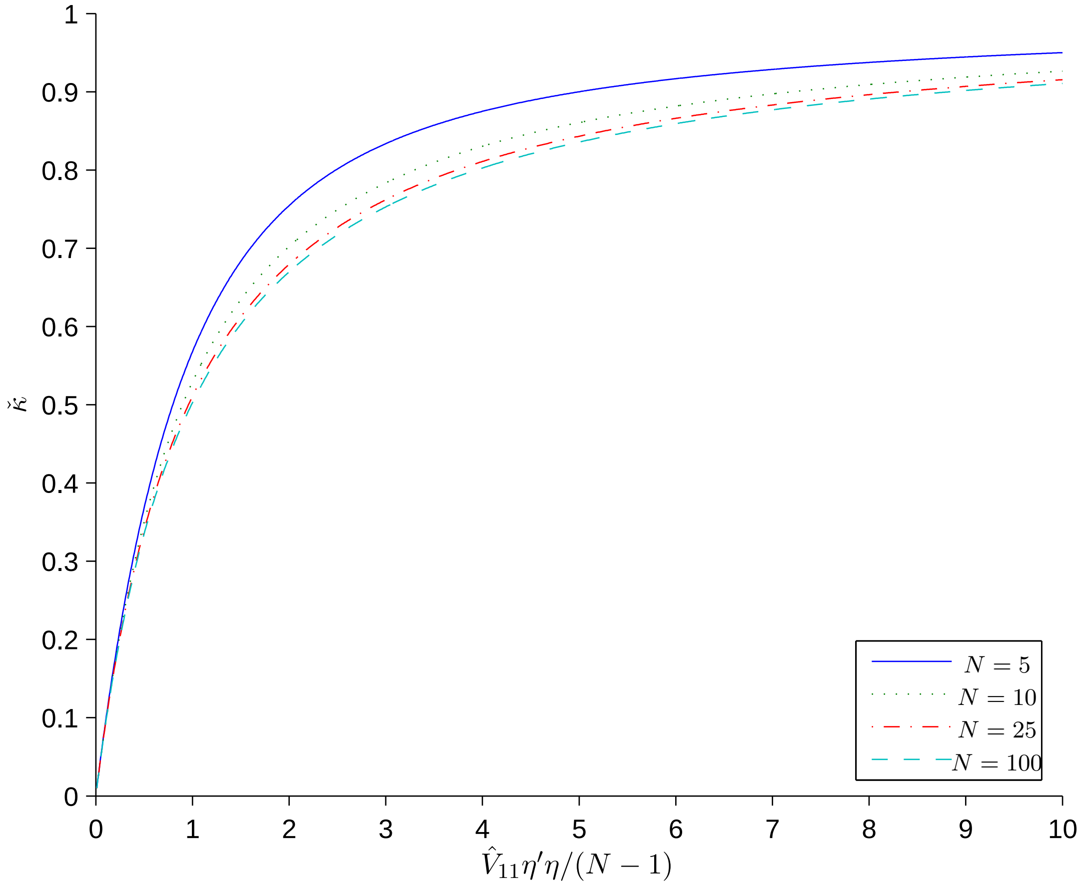

Figure 1
Conditional expected value of the second pass GLS cross-sectional regression estimate of the risk premium for the single factor case. The figure plots $\breve{\kappa} = E[\breve{\gamma}_1|\hat{V}_{11}]/\gamma_1$ as a function of $\hat{V}_{11}\eta'\eta/(N-1)$ for different values of N, where $\breve{\gamma}_1$ is the second-pass GLS CSR estimate of the risk premium, N is the number of test assets, $\hat{V}_{11}$ is the realized variance of the factor, and $\eta'\eta/(N-1)$ is a measure of the dispersion of β across the test assets.

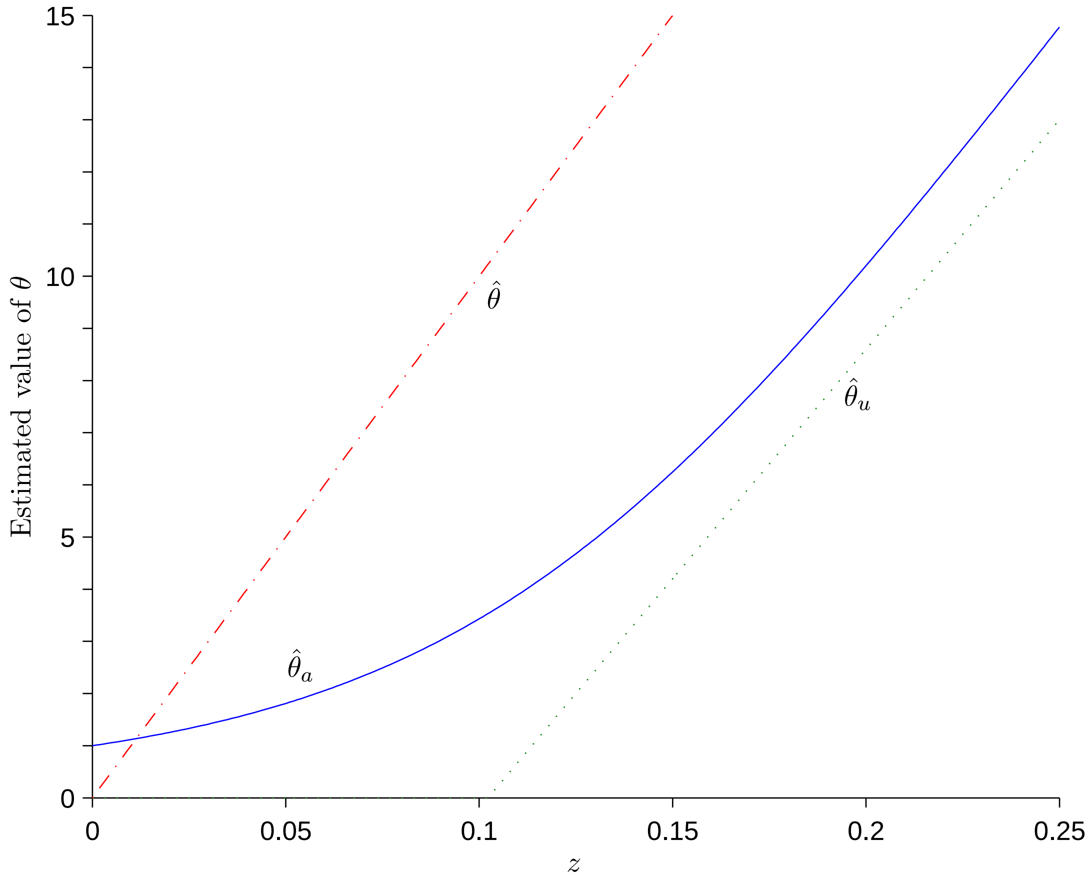

Figure 2
Representation of three estimators of $\theta = \hat{V}_{11}\eta'\eta$ for different values of $\hat{V}_{11}\hat{\eta}'\hat{\eta}/T$. The figure plots three estimators of $\theta = \hat{V}_{11}\eta'\eta$ as a function of $z = \hat{V}_{11}\hat{\eta}'\hat{\eta}/T$ when $N = 10$ and $T = 100$. The dashed-dotted line is for the sample estimator $\hat{\theta} = \hat{V}_{11}\hat{\eta}'\hat{\eta}$, which is an upward biased estimator of θ. The dotted line is for the estimator $\hat{\theta}_u = (T-N-2)z-(N-1)$, which is an unbiased estimator of θ. The solid line is for the estimator $\hat{\theta}_a = 2b_2F_1(1+b,(T-1)/2,(N-1)/2,z/(1+z))/_2F_1(b,(T-1)/2,(N-1)/2,z/(1+z))$, where ${}_2F_1$ is the hypergeometric function and b is set to 0.5. $\hat{\theta}_a$ is the Bayes estimator of θ under Stein's loss function and a prior of $\pi(\theta) = \theta^b$.