Finite Sample Analysis of Two-Pass Cross-Sectional Regressions

ROBERT CHEN and RAYMOND KAN*

First draft: September, 2003

*Both authors are from the Joseph L. Rotman School of Management at the University of Toronto. We thank seminar participants at University of Toronto for helpful discussions and comments. Kan gratefully acknowledges financial support from the Social Sciences and Humanities Research Council of Canada.
Finite Sample Analysis of Two-Pass Cross-Sectional Regressions

ABSTRACT

In this paper, we investigate finite sample properties of the two-pass cross-sectional regression (CSR) methodology, which is popular for estimation of risk premia and testing of beta pricing models. We find that the finite sample distribution of the estimated risk premium differs significantly from the asymptotic distribution. In particular, the risk premium estimates obtained from the second pass CSR of average returns on estimated betas can have serious bias even when the number of time series observations is reasonably large. In addition, the standard error of the estimated risk premium based on the asymptotic distribution overstates the actual standard error. A simple bias-adjustment on the estimated zero-beta rate and risk premium is proposed and the adjusted version is shown to have better finite sample properties than the unadjusted one.
In 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 estimation of risk premia and testing of beta asset pricing models. Although there are many variations of 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 a constant term and 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, 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 premium are biased. However, as the length of the times series used to estimate betas increases to infinity, Shanken (1992) shows that the estimation errors on the betas go to zero and hence the estimated zero-beta rate and risk premium from the second pass CSR are still consistent. The second problem is that the standard errors of the estimated zero-beta rate and risk premium that are due to Fama and MacBeth (1973) are inconsistent. Shanken (1992) proposes an asymptotically valid EIV adjustment to 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 literature. Aside from some simulation studies like Affleck-Graves and Bradfield (1993) and Shanken and Zhou (2000), we have very little understanding of the finite sample properties of the estimated zero-beta rate and risk premium from the second pass CSR. For example, while we know that there is a bias in the estimated risk premium, we do not know the magnitude of this bias, whether the bias is positive or negative, or what the parameters determining it are. While asymptotically the estimated zero-beta rate and risk premium from the second pass CSR are consistent, it is not clear if 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 many applications. 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 premium.

In this paper, we provide a finite sample analysis on the biases and variances of the estimated zero-beta rate and risk premium 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. With reasonable choice 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. Based on our analytical analysis of the bias, we propose a bias adjustment to the point estimates of the zero-beta rate and risk premium. Simulation results show that our bias-adjusted estimators perform better than the unadjusted ones in finite samples.

The rest of the paper is organized as follows. Section I provides an overview of the two-pass cross-sectional regression methodology and summarizes existing results. Section II presents a fast simulation approach that allows us to obtain the distribution and the moments of the estimated zero-beta rate and risk premium for a general multi-factor case. Section III presents analytical results on the finite sample bias and variance of the estimated zero-beta rate and risk premium for the single factor case. Section IV presents our bias-adjusted estimators of zero-beta rate and risk premium. Section V presents simulation results to examine the robustness of our analysis as well as to evaluate how well our bias-adjusted estimators of zero-beta rate and risk premium perform in finite samples. Section VI concludes the paper and the Appendix contains proofs of all propositions.

1 Under some assumptions, Kim (1995) presents an EIV adjustment when the number of test assets goes to infinity. As the number of test assets used in many studies is not very large, this EIV adjustment can also be inappropriate for those studies.
I. Overview of the Two-Pass Methodology

A. The 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

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

(1)

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

(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

$$
\mu_2 = 1_N \gamma_0 + \beta \gamma_1 = H \gamma,
$$

(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, $\gamma_0$ is usually called the zero-beta rate and $\gamma_1$ 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 $\gamma$ by first estimating $\beta$ using an OLS regression of regressing $R_t$ on a constant term and $f_t$

$$
R_t = \alpha + \beta f_t + \epsilon_t, \quad t = 1, \ldots, T,
$$

(4)

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

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

(5)

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

(6)

the OLS estimate of $\beta$ is given by

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

(7)
Equipped with the estimated beta 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 one is the OLS CSR. Under the OLS CSR, estimated $\gamma$ is given by

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

The variance of $\epsilon_t$ in (4) is given by

$$\Sigma = V_{22} - V_{21}V_{11}^{-1}V_{12}. \hspace{1cm} (9)$$

Suppose we know $\Sigma$, then a more efficient estimate of $\gamma$ can be obtained using a true GLS CSR in the second pass. The true GLS estimate of $\gamma$ is given by

$$\tilde{\gamma} = (\hat{H}'\Sigma^{-1}\hat{H})^{-1}(\hat{H}'\Sigma^{-1}\hat{\mu}_2). \hspace{1cm} (10)$$

In most applications, $\Sigma$ is unknown and needs to be estimated. In practice, one uses

$$\hat{\Sigma} = \hat{V}_{22} - \hat{V}_{21}\hat{V}_{11}^{-1}\hat{V}_{12} \hspace{1cm} (11)$$

to replace the $\Sigma$ 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 $\gamma$ from this estimated GLS CSR is given by

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

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

In the discussion above, we assume that the same estimated beta is used throughout the entire sample period, which allows to simply run a single CSR of $\hat{\mu}_2$ on $\hat{H}$ to estimate $\gamma$. 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 beta of the $N$ assets at time $t$, typically estimated from an earlier period. For example, the time $t$ estimate of OLS CSR estimate of $\gamma$ is given by

$$\tilde{\gamma}_t = (\hat{H}'_t\hat{H}_t)^{-1}(\hat{H}'_tR_t). \hspace{1cm} (13)$$
By repeating this OLS CSR period-by-period, we get a time series of \( \bar{\gamma}_t \) and the resulting estimate of \( \gamma \) is simply the time series average of \( \bar{\gamma}_t \). 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.

\textbf{B. Existing Results}

For finite sample inference, we make the assumption that \( Y_t \) is i.i.d. normal. Nevertheless, we will use simulation to examine the robustness of our results to departure 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 \( \gamma \) when conditional on the realizations of the factors. Conditional on \( \hat{\mu}_1 \), we have

\[ \hat{\mu}_2 \sim N(\alpha + \beta \hat{\mu}_1, \Sigma/T). \]  

When the asset pricing model is correct, \( \alpha \) is given by

\[ \alpha = \mu_2 - \beta \mu_1 = 1_N \gamma_0 + \beta \gamma_1 - \beta \mu_1, \]  

so conditional on \( \hat{\mu}_1 \), we have

\[ \hat{\mu}_2 \sim N(1_N \gamma_0 + \beta \gamma_1, \Sigma/T), \]  

where \( \gamma_1 = \gamma_1 - \mu_1 + \hat{\mu}_1 \) and it is called the \textit{ex post} risk premium by Shanken (1992).\(^2\)

Conditional on \( \hat{V}_{11} \), we have \( \hat{\beta} \) and \( \hat{\Sigma} \) independent of each other, with distributions given by

\[ \text{vec}(\hat{\beta}) \sim N(\text{vec}(\beta), \Sigma \otimes \hat{V}_{11}^{-1}/T), \]  

\[ T\hat{\Sigma} \sim W_N(T - K - 1, \Sigma). \]  

The above are standard results from the normality assumption. One can refer to Muirhead (1982, Theorems 1.5.2 and 3.2.10) for the proof of these results.\(^3\) Note that \( \hat{\beta} \) and \( \hat{\Sigma} \) are functions of elements of \( \hat{V} \), so they are also independent of \( \hat{\mu}_2 \).

\(^2\)Our analysis focuses on the case that the asset pricing model is correct because this is the case in which \( \gamma \) is well defined. Nevertheless, it is easy to generalize our results to the case of general \( \mu_2 \) in which the \( K \)-factor asset pricing model does not hold.

\(^3\)See also Lemma 1 of Shanken (1992).
While we know the distributions of $\hat{\mu}_2$, $\hat{\beta}$ and $\hat{\Sigma}$, the estimated $\gamma$ is a complicated function of these random variables, so obtaining the distribution or the moments of the estimated $\gamma$ in finite samples is nontrivial. In the existing literature, there are two methods researchers use in making statistical inference on $\gamma$. The first method ignores the fact that $\hat{\beta}$ is estimated with errors. If we treat $\hat{H}$ as $H$, we have

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

(19)

Using the fact that $V_{22} = \Sigma + \beta V_{11}\beta'$, we can write the variance of $\tilde{\gamma}$ and $\tilde{\gamma}$ as

$$\text{Var}[\tilde{\gamma}] = \frac{1}{T} (H'\Sigma^{-1}H)^{-1} H' \Sigma^{-1} [0 \ 0'] H',$$

(21)

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

$$\text{Var}[\hat{\gamma}] = \frac{1}{T} (H'\Sigma^{-1}H)^{-1} H' \Sigma^{-1} \left[H'\Sigma^{-1}V_{22}\Sigma^{-1}H\right](H'\Sigma^{-1}H)^{-1},$$

(22)

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

Note that if $\beta$ is known, we have

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

(24)

and $\tilde{\gamma}_t$ is i.i.d. normally distributed. Therefore, using a time-series of OLS CSR estimate $\tilde{\gamma}_t$ 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

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

(25)

where $s(a'\tilde{\gamma}_t)$ is the sample standard deviation of the time series $a'\tilde{\gamma}_t$ and $t_{T-1}$ is the central $t$-distribution with $T - 1$ degrees of freedom. This is the foundation for the 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, $\beta$ is estimated with errors, so the errors-in-variables problem is introduced in the second pass CSR when $\hat{\beta}$ instead of true $\beta$ is used. Shanken (1992) provides a nice asymptotic analysis of this problem. He shows that although $\beta$ is estimated with errors, the estimation error in $\hat{\beta}$ goes to zero as $T$ goes to infinity and the second pass CSR estimate of $\gamma$ is $T$-consistent. However, the usual standard error for the estimated $\gamma$ is inconsistent, and we need an adjustment to account for the estimation errors in $\hat{\beta}$. Shanken (1992) shows that

$$\sqrt{T}(\hat{\gamma} - \gamma) \overset{A}{\sim} N\left(0, \left(\begin{array}{c} \gamma_1 \end{array}\right)^\prime \left(\begin{array}{cc} \Sigma_1 & \gamma_1 \end{array}\right) \left(\begin{array}{c} \Sigma_1 \gamma_1 \gamma_1 & \Sigma \end{array}\right)^{-1} \left(\begin{array}{c} \gamma_1 \end{array}\right)^\prime \right),$$

(26)

$$\sqrt{T}(\tilde{\gamma} - \gamma) \overset{A}{\sim} N\left(0, \left(\begin{array}{c} \gamma_1 \end{array}\right)^\prime \left(\begin{array}{cc} \Sigma_1 & \gamma_1 \end{array}\right) \left(\begin{array}{c} \Sigma_1 \gamma_1 \gamma_1 & \Sigma \end{array}\right)^{-1} \left(\begin{array}{c} \gamma_1 \end{array}\right)^\prime \right),$$

(27)

$$\sqrt{T}(\tilde{\gamma} - \gamma) \overset{A}{\sim} N\left(0, \left(\begin{array}{c} \gamma_1 \end{array}\right)^\prime \left(\begin{array}{cc} \Sigma_1 & \gamma_1 \end{array}\right) \left(\begin{array}{c} \Sigma_1 \gamma_1 \gamma_1 & \Sigma \end{array}\right)^{-1} \left(\begin{array}{c} \gamma_1 \end{array}\right)^\prime \right),$$

(28)

For statistical inference, one replaces the terms $\beta$, $\Sigma$, $\gamma_1$ and $V_{11}$ in the asymptotic variance with their sample counterparts.

While the asymptotic results of Shanken (1992) are elegant, one may question their relevance for the applications that we typically encounter. In many studies, $\beta$ 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 $\gamma$ 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 $\mu$ and $V$ and simulate $T$ observations of $Y_t$ in order to draw one realization of estimated $\gamma$. Aside from being time consuming, the conclusion is specific to a given choice of $\mu$ and $V$, so one cannot make a general statement 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 $\gamma$. In the next section, we present a simplification of the problem which allows us to reduce the number of parameters as well as deliver a speedy simulation method that frees us from having to simulate $T$ observations of $Y_t$ for every draw of estimated $\gamma$. This analysis also provides us with some new results that are not available in literature.

\footnote{Jagannathan and Wang (1996) and Ahn and Gadarowski (1999) provide asymptotic results for more general cases.}
II. Finite Sample Analysis

A. Simulation Method

Before we discuss our simulation method, we first draw a distinction between conditional and unconditional distributions of $\hat{\gamma}$, $\tilde{\gamma}$ and $\hat{\gamma}$. By conditional distribution, what we meant is the distribution of the estimated $\gamma$ when conditional on the mean and the variance of the $K$ factors, i.e., conditional on $\hat{\mu}_1$ and $\hat{\Sigma}_1$. As the derivation of the unconditional distribution is based on the results from the conditional distribution, so we provide its analysis first.

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

A.1. OLS CSR

We turn our attention to the OLS case first. Although the OLS CSR estimate of $\gamma$ only involves $\hat{\mu}_2$ and $\hat{\beta}$, we still need the knowledge of $\Sigma$ to simulate $\hat{\mu}_2$ and $\hat{\beta}$ if we use (16) and (17). In order to reduce the number of parameters, we first apply the partitioned matrix inverse formula to $(\hat{H}'\hat{H})^{-1}$ and obtain

$$\begin{align*}
(\hat{H}'\hat{H})^{-1} \hat{H}' &= \begin{bmatrix}
\frac{1}{N}[1_N'(1_N'N\hat{\beta})'(\hat{\beta}'M\hat{\beta})^{-1}\hat{\beta}'M] \\
(\hat{\beta}'M\hat{\beta})^{-1}\hat{\beta}'M
\end{bmatrix},
\end{align*}$$

(29)

where $M = I_N - 1_N'(1_N'N)^{-1}1_N'$. With this expression, we can write the OLS CSR estimate of $\gamma_1$ and $\gamma_0$ as

$$\begin{align*}
\gamma_1 &= (\hat{\beta}'M\hat{\beta})^{-1}(\hat{\beta}'M\hat{\mu}_2), \\
\gamma_0 &= \frac{1}{N} \left(1_N'\hat{\mu}_2 - 1_N'\hat{\beta}\hat{\gamma}_1\right).
\end{align*}$$

(30) (31)

The conditional distributions in (16)–(18) only depend on the multivariate normality assumption of $\epsilon_t$ in (4), there is no need to make a joint normality assumption on $Y_t$. 
Let $PAP'$ 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 $P$ is an $N \times (N-1)$ matrix with its $i$th column equal to the eigenvector of $\Sigma^{\frac{1}{2}}M\Sigma^{\frac{1}{2}}$ associated with $\lambda_i$. Denote $\nu = \Sigma^{-\frac{1}{2}}1_N/(1_N'\Sigma^{-1}1_N)^{\frac{1}{2}}$, we note that $\Sigma^{\frac{1}{2}}M\Sigma^{\frac{1}{2}}\nu = 0_N$, so $\nu$ is orthogonal to $P$. Together, $P$ and $\nu$ form an orthonormal basis of $\mathbb{R}^N$ and we have

$$\nu\nu' + PP' = I_N. \tag{32}$$

Defining $Y$ and $Z$ as the following linear transformations of $\hat{\mu}_2$ and $\hat{\beta}$,

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} \sqrt{T}\nu'\Sigma^{-\frac{1}{2}}\hat{\mu}_2 \\ \sqrt{TP}\Sigma^{-\frac{1}{2}}\hat{\mu}_2 \end{bmatrix}, \tag{33}$$

$$Z = \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix} = \begin{bmatrix} \sqrt{T}\nu'\Sigma^{-\frac{1}{2}}\hat{\beta}V_{11}^{\frac{1}{2}} \\ \sqrt{TP}\Sigma^{-\frac{1}{2}}\hat{\beta}V_{11}^{\frac{1}{2}} \end{bmatrix}, \tag{34}$$

we now show that $\hat{\gamma}_0$ and $\hat{\gamma}_1$ can be written as functions of $Y$ and $Z$. Starting with $\hat{\gamma}_1$, we have

$$\hat{\gamma}_1 = (\hat{\beta}'M\hat{\beta})^{-1}(\hat{\beta}'M\hat{\mu}_2)$$

$$= (\hat{\beta}'\Sigma^{-\frac{1}{2}}\Sigma\Sigma^{\frac{1}{2}}\Sigma^{-\frac{1}{2}}\hat{\beta})^{-1}(\hat{\beta}'\Sigma^{-\frac{1}{2}}\Sigma\Sigma^{\frac{1}{2}}\Sigma^{-\frac{1}{2}}\hat{\mu}_2)$$

$$= (\hat{\beta}'\Sigma^{-\frac{1}{2}}PAP'\Sigma^{-\frac{1}{2}}\hat{\beta})^{-1}(\hat{\beta}'\Sigma^{-\frac{1}{2}}PAP'\Sigma^{-\frac{1}{2}}\hat{\mu}_2)$$

$$= \hat{V}_{11}^{-\frac{1}{2}}(Z_1'AZ_2)^{-1}(Z_2'AY_2). \tag{35}$$

As for $\hat{\gamma}_0$, we have

$$\hat{\gamma}_0 = \frac{1}{N} \left(1_N'\hat{\mu}_2 - 1_N'\hat{\beta}\hat{\gamma}_1\right)$$

$$= \frac{1}{N} \left[1_N'\Sigma^{\frac{1}{2}}(\nu\nu' + PP')\Sigma^{-\frac{1}{2}}\hat{\mu}_2 - 1_N'\Sigma^{\frac{1}{2}}(\nu\nu' + PP')\Sigma^{-\frac{1}{2}}\hat{\beta}\hat{\gamma}_1\right]$$

$$= \frac{1}{\sqrt{T}} \left[(1_N'\Sigma^{-1}1_N)^{-\frac{1}{2}}Y_1 + \xi'Y_2 - (1_N'\Sigma^{-1}1_N)^{-\frac{1}{2}}Z_1 + \xi'Z_2\right] (Z_2'AZ_2)^{-1}(Z_2'AY_2), \tag{36}$$

where $\xi = (P'\Sigma^{\frac{1}{2}}1_N)/N$. Under this approach, we need to simulate $Y$ and $Z$ instead of $\hat{\mu}_2$ and $\hat{\beta}$. Although the number of random variables remains the same, we need to know much less information in simulating $Y$ and $Z$. Conditional on $\hat{\mu}_1$ and $\hat{V}_{11}$, we have $Y_1$, $Y_2$, $Z_1$, and $Z_2$ being independent of each other and having the following normal distribution

$$Y \sim N\left(\begin{bmatrix} \sqrt{T}(\nu'\Sigma^{-\frac{1}{2}}1_N\hat{\gamma}_0 + \delta\hat{\gamma}_1) \\ \sqrt{T}\hat{\gamma}_1 \end{bmatrix}, I_N\right), \tag{37}$$

$$\text{vec}(Z) \sim N\left(\text{vec} \left(\begin{bmatrix} \sqrt{T}\delta\hat{V}_{11}^{\frac{1}{2}} \\ \sqrt{T}\hat{V}_{11}^{\frac{1}{2}} \end{bmatrix}\right), I_N \otimes I_K\right), \tag{38}$$

9
where we define $\delta = \nu' \Sigma^{-\frac{1}{2}} \beta$ and $\eta = P' \Sigma^{-\frac{1}{2}} \beta$. Therefore, conditional on $\hat{\mu}_1$ and $\hat{V}_{11}$, one only needs to know $\xi$, $1' \Sigma^{-1} 1_N$, $\gamma$, $\delta$, $\eta$ and $\Lambda$ to simulate $\hat{\gamma}$. In fact, multiplying $\Lambda$ by a constant would not change the distribution, so the distribution of $\hat{\gamma}$ only depends on $\lambda_i/\lambda_{N-1}$ 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 only $N(K + 2) + K - 1$ parameters.

### A.2. True GLS CSR

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

$$
\tilde{\gamma}_1 = (\hat{\beta}' \Sigma^{-\frac{1}{2}} (I_N - \nu \nu') \Sigma^{-\frac{1}{2}} \hat{\beta})^{-1} (\hat{\beta}' \Sigma^{-\frac{1}{2}} (I_N - \nu \nu') \Sigma^{-\frac{1}{2}} \hat{\mu}_2)
$$

$$
= (\hat{\beta}' \Sigma^{-\frac{1}{2}} P' \Sigma^{-\frac{1}{2}} \hat{\beta})^{-1} (\hat{\beta}' \Sigma^{-\frac{1}{2}} P' \Sigma^{-\frac{1}{2}} \hat{\mu}_2)
$$

$$
= \hat{V}_{11}^{\frac{1}{2}} (Z_2' Z_2)^{-1} (Z_2' Y_2),
$$

(39)

$$
\tilde{\gamma}_0 = \frac{1}{1' \Sigma^{-1} 1_N} (1' \Sigma^{-1} \hat{\mu}_2 - 1' \Sigma^{-1} \hat{\beta}_1)
$$

$$
= \frac{1}{(1' \Sigma^{-1} 1_N)^{\frac{1}{2}}} (\nu' \Sigma^{-\frac{1}{2}} \hat{\mu}_2 - \nu' \Sigma^{-\frac{1}{2}} \hat{\beta}_1)
$$

$$
= \frac{1}{T^{\frac{1}{2}} (1' \Sigma^{-1} 1_N)^{\frac{1}{2}}} [Y_1 - Z_1 (Z_2' Z_2)^{-1} (Z_2' Y_2)].
$$

(40)

Therefore, the distribution of $\tilde{\gamma}$ only depends on $(N + 1)K + 2$ parameters: $1' \Sigma^{-1} 1_N$, $\gamma$, $\delta$ and $\eta$. Comparing (39) and (40) with (35) and (36), we can see that the distribution of true GLS $\tilde{\gamma}$ can be obtained as a special case of the distribution of OLS $\hat{\gamma}$ by setting $\xi = 0_{N-1}$ and $\Lambda = I_{N-1}$.

### A.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 $\tilde{\gamma}$, and there is no need to simulate $\hat{\Sigma}$ from the central Wishart distribution in order to simulate $\hat{\gamma}$. To prepare for our derivation, we define

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

(41)
\[ \hat{A} = [\hat{\mu}_2, \hat{H}]\Sigma^{-1}[\hat{\mu}_2, \hat{H}]. \]  

(42)

The only difference between \( \hat{A} \) and \( \tilde{A} \) is that \( \hat{A} \) has \( \hat{\Sigma} \) in the middle and \( \tilde{A} \) has the true \( \Sigma \) in the middle. Using Theorem 3.2.11 of Muirhead (1982), conditional on \( \hat{\mu}_2 \) and \( \hat{\beta} \), we have

\[ \hat{A}^{-1} \sim W_{K+2}(T - N + 1, \tilde{A}^{-1}/T). \]  

(43)

Our first task is to express \( \tilde{\gamma} \) and \( \hat{\gamma} \) as elements of \( \tilde{A}^{-1} \) 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 \((i, j)\) block of \( \hat{A} \) and \( \hat{A}^{ij} \) the \((i, j)\) block of \( \hat{A}^{-1} \). \( \hat{A}_{ij} \) and \( \hat{A}^{ij} \) are similarly defined for \( \tilde{A} \) and \( \tilde{A}^{-1} \). From the partitioned matrix inverse formula, it is easy to verify that

\[ \tilde{\gamma} = \tilde{A}^{-1}_{22}\hat{A}_{21} = -\tilde{A}^{21}(\tilde{A}^{11})^{-1}, \]

\[ \hat{\gamma} = \hat{A}^{-1}_{22}\hat{A}_{21} = -\hat{A}^{21}(\hat{A}^{11})^{-1}. \]  

(44)

Conditional on \( \hat{\mu}_2 \) and \( \hat{\beta} \), from (43), we have

\[ \hat{A}^{11} \sim W_1(T - N + 1, \tilde{A}^{11}/T). \]  

(45)

It follows that

\[ U = \frac{T\hat{A}^{11}}{\tilde{A}^{11}} \sim \chi^2_{T-N+1}, \]  

(46)

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

\[ (\tilde{A}^{11})^{-1} = \frac{T(\hat{A}^{11})^{-1}}{U} = \frac{T(\tilde{A}^{11} - \tilde{A}^{12}\tilde{A}^{-1}_{22}\tilde{A}_{21})}{U}. \]  

(47)

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

\[ \hat{A}^{21} \sim N(\tilde{A}^{21}(\tilde{A}^{11})^{-1}\tilde{A}^{11}, \tilde{A}^{-1}_{22}\tilde{A}^{11}/T), \]  

(48)

and hence

\[ \hat{\gamma} = -\hat{A}^{21}(\hat{A}^{11})^{-1} \sim N(\hat{\gamma}, (\hat{A}^{11})^{-1}\hat{A}^{-1}_{22}/T). \]  

(49)

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 (47) to obtain the conditional distribution of \( \hat{\gamma} \) as

\[ \hat{\gamma} \sim N\left(\gamma, \frac{(\tilde{A}^{11} - \tilde{A}^{12}\tilde{A}^{-1}_{22}\tilde{A}_{21})\tilde{A}^{-1}_{22}}{U}\right). \]  

(50)
With some algebra, we can show that

\[
(\tilde{A}_{11} - \tilde{A}_{12}\tilde{A}_{22}^{-1}\tilde{A}_{21})\tilde{A}_{22}^{-1} = (Y_2' [I_{N-1} - Z_2(Z_2'Z_2)^{-1}Z_2'] Y_2)
\]

\[
\begin{bmatrix}
1+Z_1(Z_2'Z_2)^{-1}Z_1'T
\frac{T}{T^2(1'_N\Sigma^{-1}1_N)^2} \hat{V}_{11}^2
\frac{1}{T^2(1'_N\Sigma^{-1}1_N)^2} \hat{V}_{11}^2
\end{bmatrix}
\begin{bmatrix}
\frac{T}{T^2(1'_N\Sigma^{-1}1_N)^2} \hat{V}_{11}^2
\end{bmatrix}
\]

\]

So in order to simulate \( \hat{\gamma} \), we first simulate \( Y, Z, \) and \( U \), independent of each other. Then we compute \( \tilde{\gamma} \) and \( (\tilde{A}_{11} - \tilde{A}_{12}\tilde{A}_{22}^{-1}\tilde{A}_{21})\tilde{A}_{22}^{-1}/U \) and simulate \( \hat{\gamma} \) from the normal distribution in (50).

Under this approach, there is no need to simulate \( \hat{\Sigma} \) at all.

While providing a speedy simulation method, our analysis of the estimated GLS CSR also has an added benefit of allowing us to relate the mean and variance of \( \hat{\gamma} \) to those of \( \tilde{\gamma} \). The relations are given in the following lemma.

**Lemma 1** Assuming \( T > N + K \), conditional on \( \hat{\mu}_1 \) and \( \hat{V}_{11} \), the mean and variance of \( \hat{\gamma} \) from the estimated GLS and \( \tilde{\gamma} \) from the true GLS, when they exist, are related to each other by the following relationship

\[
E[\hat{\gamma}|\hat{\mu}_1, \hat{V}_{11}] = E[\tilde{\gamma}|\hat{\mu}_1, \hat{V}_{11}],
\]

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

Unconditionally, we have

\[
E[\hat{\gamma}] = E[\tilde{\gamma}],
\]

\[
\text{Var}[\hat{\gamma}] = \text{Var}[\tilde{\gamma}] + \frac{1}{T - N - 1} E[(\tilde{A}_{11} - \tilde{A}_{12}\tilde{A}_{22}^{-1}\tilde{A}_{21})\tilde{A}_{22}^{-1}].
\]

This lemma suggests that the expected value and hence the bias of \( \hat{\gamma} \) from the estimated GLS CSR is exactly the same as the one from the true GLS CSR. Therefore, if one is interested in finding out the bias of \( \hat{\gamma} \), one can simply use the corresponding results from the true GLS, which is easier to derive. However, the variance of \( \hat{\gamma} \) is larger than that of \( \tilde{\gamma} \), so using estimated \( \hat{\Sigma} \) instead of true \( \Sigma \) introduces additional volatility to the estimated \( \gamma \), especially when \( N \) is large relative to \( T \). Therefore, while \( \hat{\gamma} \) from the true GLS CSR is more efficient than the OLS CSR \( \tilde{\gamma} \), there is no guarantee that \( \hat{\gamma} \) from the estimated GLS CSR is more efficient than \( \hat{\gamma} \) in finite samples. This is particularly the case when \( N \) is large relative to \( T \).
A.4. Extensions

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

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

(56)

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

(57)

and they are independent of each other, so simulating the unconditional distribution of the estimated $\gamma$ 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 $\epsilon_t$ in (4) is i.i.d. normal when conditional on $\hat{\mu}_1$ and $\hat{V}_{11}$, we can still use our method to simulate the unconditional distribution of the estimated $\gamma$.

Our simulation method can also be extended to the situation when 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 we use the first $T$ periods to estimate $\beta$, but the second pass CSR is run using realized returns of period $t$, where $t > T$. The OLS CSR estimate of $\gamma$ at time $t$ is

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

(58)

Comparing (58) with (8), the only difference here is that we use $R_t$ instead of $\hat{\mu}_2$ as the dependent variable. As $R_t$ and $\hat{\mu}_2$ are both independent of $\hat{\beta}$, simulating $\hat{\gamma}_t$ requires very little modification in our simulation approach. Conditional on $f_t$ (instead of $\hat{\mu}_1$ as before), we have

$$R_t \sim N(1_N\gamma_0 + \beta\tilde{\gamma}_t, \Sigma),$$

(59)

where $\tilde{\gamma}_t = \gamma_1 - \mu_1 + f_t$ and is independent of $\hat{\beta}$. By changing the definition of $Y$ to

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} \nu'\Sigma^{-\frac{1}{2}}R_t \\ P'\Sigma^{-\frac{1}{2}}R_t \end{bmatrix} \sim N(\begin{bmatrix} \nu'\Sigma^{-\frac{1}{2}}1_N\gamma_0 + \delta\tilde{\gamma}_t \\ \eta\tilde{\gamma}_1 \end{bmatrix}, I_N),$$

(60)

we can then write

$$\tilde{\gamma}_t = (\hat{\beta}'M\hat{\beta})^{-1}(\hat{\beta}'MR_t) = \sqrt{T}V_{11}^{\frac{1}{2}}(Z_2'\Lambda Z_2)^{-1}(Z_2'\Lambda Y_2),$$

(61)

$$\gamma_0t = \frac{1}{N} \left(1_N' R_t - 1_N' \hat{\beta}\tilde{\gamma}_t \right)$$

$$= (1_N'\Sigma^{-1}1_N)^{-\frac{1}{2}}Y_1 + \xi'Y_2 - \left[(1_N'\Sigma^{-1}1_N)^{-\frac{1}{2}}Z_1 + \xi'Z_2 \right] (Z_2'\Lambda Z_2)^{-1}(Z_2'\Lambda Y_2).$$

(62)
Therefore, conditional on $f_t$ and $\hat{V}_{11}$, we can simulate $Y$ and $Z$ to generate $\hat{\gamma}_t$. For true GLS and estimated GLS, we can make the same modification to obtain the distribution of $\hat{\gamma}_t$ and $\hat{\gamma}_t$. Note that in our original setup, $T$ refers to the length of the time series used to estimate $\beta$ as well as 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 realized return in a different period. The Fama-MacBeth CSR can be repeated for many periods to obtain a time series of $\hat{\gamma}_t$, but the length of the time series of $\hat{\gamma}_t$ has no relation to the length of the beta estimation period ($T$).

B. Moments of CSR Estimates of Zero-Beta Rate and Risk Premium

B.1. Existence of Moments

Asymptotically, $\hat{\gamma}$, $\tilde{\gamma}$ and $\hat{\gamma}$ have a normal distribution according to (26)–(28), so all the moments of the estimated $\gamma$ exist in the asymptotic distribution. However, in finite samples, only a finite number of moments of $\hat{\gamma}$, $\tilde{\gamma}$ 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 $\gamma$ exist 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 $\gamma$ does not exist if $s \geq N - K$, so the unconditional $s$-th moment of the estimated $\gamma$ 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 we use just $N = 2$ assets to estimate the CAPM (i.e., $K = 1$), we can get an estimate of $\gamma$, but none of its moments exist (because its distribution has heavy tails). When we estimate the CAPM using $N = 3$ assets, the mean of the estimated $\gamma$ exists, but its variance does not exist. In general, we expect that when $K$ approaches $N$, the finite sample distribution of the estimated $\gamma$ becomes less and less like normal. Note that for the OLS and the true GLS case, 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 $\gamma$ 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 on \( \gamma \) can be problematic even for large \( T \).

In the previous subsection, we show that one can approximate the conditional distribution of the estimated \( \gamma \) by simulating \( Y_1, Y_2, Z_1 \) and \( Z_2 \). However, if one is only interested in the first and the second moment of the estimated \( \gamma \), then one only needs to simulate \( Z_2 \). The rest of this subsection discusses the conditional and unconditional mean and variance of the estimated \( \gamma \). For notational brevity, we use \( E^c \) and \( \text{Var}^c \) to denote conditional 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 \( \text{Var}^c[X] \equiv \text{Var}[X|\hat{\mu}_1, \hat{V}_{11}] \).

**B.2. Mean**

From Proposition 1, the conditional mean of \( \hat{\gamma}, \check{\gamma} \) and \( \hat{\gamma} \) exists when \( N > K + 1 \).\(^6\) From Lemma 1, we know that the conditional mean of \( \hat{\gamma} \) and \( \check{\gamma} \) are the same, so we only need to present the results for \( \check{\gamma} \) and \( \hat{\gamma} \) here. We start with the OLS CSR estimate of \( \gamma \). Conditional on \( \hat{\mu}_1 \) and \( \hat{V}_{11} \), \( Y_2 \) and \( Z_2 \) are independent, from (35) we have

\[
E^c[\hat{\gamma}_1] = \hat{V}_{11}^2 E^c[(Z_2'\Lambda Z_2)^{-1}Z_2']E^c[Y_2] = \sqrt{T}\hat{V}_{11}^2 E^c[(Z_2'\Lambda Z_2)^{-1}Z_2']\Lambda\eta\check{\gamma}_1. \tag{63}
\]

This expression suggests that conditional on \( \hat{\mu}_1 \) and \( \hat{V}_{11} \), the expected value of \( \check{\gamma}_1 \) depends on \( \eta, \Lambda \) and \( \check{\gamma}_1 \). In order to obtain the conditional mean of \( \check{\gamma}_1 \), we need to evaluate \( E^c[(Z_2'\Lambda Z_2)^{-1}Z_2'] \). 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 explicit expression for \( E^c[(Z_2'\Lambda Z_2)^{-1}Z_2'] \) is available, but we can draw a large number of \( Z_2 \) and approximate the expectation by using the average value of \( (Z_2'\Lambda Z_2)^{-1}Z_2' \). Note that in approximating \( E^c[\check{\gamma}_1] \), all we need to do is to simulate \( Z_2 \), there is no need to simulate \( Y_2 \). Besides saving computational time, our approach is more accurate because there is only one source of simulation error which is due to the use of the average value of \( (Z_2'\Lambda Z_2)^{-1}Z_2' \) to approximate \( E^c[(Z_2'\Lambda Z_2)^{-1}Z_2'] \). There is no need to worry about approximation errors coming from simulating \( Y_2 \).

Defining \( h = (1_N'\Sigma^{-1}/\beta)/(1_N'\Sigma^{-1}1_N) \), its conditional mean of \( \check{\gamma}_0 \) is given by

\[
E^c[\check{\gamma}_0] = \gamma_0 + h(\check{\gamma}_1 - E^c[\check{\gamma}_1]) + \xi'\eta\check{\gamma}_1 - \xi'\hat{\gamma}^c[Z_2(Z_2'\Lambda Z_2)^{-1}Z_2']\Lambda\eta\check{\gamma}_1 \tag{64}
\]

\(^6\)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.
using the fact that $Y_1$, $Y_2$ and $Z_1$ are independent of $Z_2$. Aside from $\eta$, $\Lambda$ and $\bar{\gamma}_1$, the conditional mean of $\bar{\gamma}_0$ also depends on $\gamma_0$, $\xi$ and $h$. In order to obtain $E^c[\bar{\gamma}_0]$, we can use the average value of $Z_2(Z_2'\Lambda Z_2)^{-1}Z_2'$ from the simulations to approximate $E^c[Z_2(Z_2'\Lambda Z_2)^{-1}Z_2']$.

Taking unconditional expectation on both sides of (63) and (64), we have

$$
E[\hat{\gamma}_1] = \sqrt{T} E[\hat{V}_{11}^{1/2}(Z_2'\Lambda Z_2)^{-1}Z_2']\Lambda\eta\bar{\gamma}_1, \tag{65}
$$

$$
E[\hat{\gamma}_0] = \gamma_0 + h(\gamma_1 - E[\hat{\gamma}_1]) + \xi'\eta\bar{\gamma}_1 - \xi'E[Z_2(Z_2'\Lambda Z_2)^{-1}Z_2']\Lambda\eta\bar{\gamma}_1. \tag{66}
$$

The only difference between obtaining the conditional and unconditional mean of $\hat{\gamma}$ is that in addition to $Z$, we also need to simulate $\hat{V}_{11}$ for the unconditional mean. In each simulation, we first simulate $\hat{V}_{11}$ using (57) and then $Z_2$ using (38), we can then approximate $E[\hat{V}_{11}^{1/2}(Z_2'\Lambda Z_2)^{-1}Z_2']$ using the average value of $\hat{V}_{11}^{1/2}(Z_2'\Lambda Z_2)^{-1}Z_2'$ from the simulations.

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

$$
E^c[\tilde{\gamma}_1] = \sqrt{T} E[\hat{V}_{11}^{1/2}(Z_2'Z_2)^{-1}Z_2']\eta\bar{\gamma}_1, \tag{67}
$$

$$
E^c[\tilde{\gamma}_0] = \gamma_0 + h(\gamma_1 - E^c[\tilde{\gamma}_1]). \tag{68}
$$

Therefore, the conditional mean of $\tilde{\gamma}_1$ only depends on $\eta$ and $\bar{\gamma}_1$, whereas the conditional mean of $\tilde{\gamma}_0$ depends on $\eta$, $\gamma_0$, $\bar{\gamma}_1$ and $(1_N'\Sigma^{-1}0)/(1_N'\Sigma^{-1}1_N)$. Unconditionally, we have

$$
E[\tilde{\gamma}_1] = \sqrt{T} E[\hat{V}_{11}^{1/2}(Z_2'Z_2)^{-1}Z_2']\eta\bar{\gamma}_1, \tag{69}
$$

$$
E[\tilde{\gamma}_0] = \gamma_0 + h(\gamma_1 - E[\tilde{\gamma}_1]). \tag{70}
$$

From the expressions above, we can see that the unconditional biases of both OLS and GLS estimates of $\gamma_0$ and $\gamma_1$ depend on the value of $\gamma_1$, but not the value of $\gamma_0$, so the actual value of the zero-beta rate is irrelevant in determining the bias of the estimated $\gamma$. 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 $\gamma$ are zero.

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

B.3. Variance

From Proposition 1, the conditional second moment $\bar{\gamma}$, $\bar{\gamma}$ and $\bar{\gamma}$ exists when $N > K + 2$. Starting with the OLS case, using the fact that $Y_2$ and $Z_2$ are independent and

$$E[\gamma_2 Y_2'] = I_{N-1} + T \eta \gamma_1' \eta',$$

we have the conditional second moment of $\hat{\gamma}_1$ given by

$$E^c[\hat{\gamma}_1 \hat{\gamma}_1'] = \hat{V}_{11}^{\frac{1}{2}} E^c[D_2 + TD_1 \bar{\gamma}_1 \gamma_1' D_1'] \hat{V}_{11}^{\frac{1}{2}},$$

where

$$D_1 = (Z_2' \Lambda Z_2)^{-1}(Z_2' \Lambda \eta),$$

$$D_2 = (Z_2' \Lambda Z_2)^{-1}(Z_2' \Lambda^2 Z_2)(Z_2' \Lambda Z_2)^{-1}.$$ (73)

With (72) and (63), we can then compute the conditional variance of $\hat{\gamma}_1$ as

$$\text{Var}^c[\hat{\gamma}_1] = E^c[\hat{\gamma}_1 \gamma_1'] - E^c[\hat{\gamma}_1] E^c[\hat{\gamma}_1]' .$$

After some straightforward but tedious algebra, it can be shown that the conditional variance of $\gamma_0$ is given by

$$\text{Var}^c[\gamma_0] = \frac{1 + E^c[\hat{\gamma}_1 \hat{V}_{11}^{\frac{1}{2}} \hat{\gamma}_1]}{T (1_N' \Sigma^{-1} 1_N)} + \frac{\xi' \xi - 2 E^c[D_3 (Z_2' \Lambda Z_2)^{-1} Z_2' \Lambda \xi] + \text{Var}^c[D_3 \hat{V}_{11}^{\frac{1}{2}} \hat{\gamma}_1]}{T}$$

$$+ \frac{E^c[\text{tr}(D_2 + TD_1 \bar{\gamma}_1 \gamma_1' D_1')] T (1_N' \Sigma^{-1} 1_N)}{T}$$

$$+ \frac{\xi' \xi + E^c[D_3 D_2 D_3' - 2 D_3 (Z_2' \Lambda Z_2)^{-1} Z_2' \Lambda \xi]}{T} .$$

(76)

where

$$D_3 = \sqrt{T \hat{V}_{11}^{\frac{1}{2}}} + \xi' Z_2 .$$

(77)

For the unconditional variances of $\bar{\gamma}_1$ and $\hat{\gamma}_0$, we use the fact that $\hat{\mu}_1$ and $\hat{V}_{11}$ are independent and

$$E[\gamma_1 \gamma_1'] = E[\gamma_1 \gamma_1'] = \frac{V_{11}}{T} + \gamma_1 \gamma_1'.$$
to obtain

$$
\text{Var}[\hat{\gamma}_1] = E[V_{11}^{\frac{1}{2}}(D_2 + D_1V_{11}D_1' + TD_1\hat{\gamma}_1\hat{\gamma}_1'D_1')V_{11}^{\frac{1}{2}}] - E[\hat{\gamma}_1]E[\hat{\gamma}_1]',
$$

(79)

$$
\text{Var}[\hat{\gamma}_0] = \text{Var}[D_3D_1\hat{\gamma}_1] + \frac{1 + E[\text{tr}(D_2 + D_1V_{11}D_1' + TD_1\hat{\gamma}_1\hat{\gamma}_1'D_1')]}{T(V_N^2\Sigma^{-1}1_N)} + \frac{\xi'\xi + E[aV_{11}a' + D_2D_2D_2' - 2D_3(Z_2\Lambda Z_2)^{-1}Z_2\Lambda \xi]}{T},
$$

(80)

where $a = h + \xi'\eta - D_3D_1$. For the true GLS case, the conditional variances of $\hat{\gamma}_1$ and $\hat{\gamma}_0$ are obtained by setting $\Lambda = I_{N-1}$ and $\xi = 0_{N-1}$ in the OLS case, and we have

$$
\text{Var}^c[\hat{\gamma}_1] = V_{11}^{\frac{1}{2}}E[C\tilde{D}_2 + T\tilde{D}_1\tilde{\gamma}_1\tilde{\gamma}_1'\tilde{D}_1']V_{11}^{\frac{1}{2}} - E[\hat{\gamma}_1]E[\hat{\gamma}_1]',
$$

(81)

$$
\text{Var}^c[\hat{\gamma}_0] = h\text{Var}^c[\hat{\gamma}_1]h' + \frac{1 + E[\text{tr}(D_2 + T\tilde{D}_1\tilde{\gamma}_1\tilde{\gamma}_1'D_1')]}{T(V_N^2\Sigma^{-1}1_N)},
$$

(82)

where

$$
\tilde{D}_1 = (Z_2'Z_2)^{-1}(Z_2'\eta),
$$

(83)

$$
\tilde{D}_2 = (Z_2'Z_2)^{-1}.
$$

(84)

The unconditional variance of $\hat{\gamma}_1$ and $\hat{\gamma}_0$ are obtained similarly. They are given by

$$
\text{Var}[\hat{\gamma}_1] = E[V_{11}^{\frac{1}{2}}(\tilde{D}_2 + \tilde{D}_1V_{11}\tilde{D}_1' + T\tilde{D}_1\tilde{\gamma}_1\tilde{\gamma}_1'\tilde{D}_1')\tilde{V}_{11}^{\frac{1}{2}}] - E[\hat{\gamma}_1]E[\hat{\gamma}_1]',
$$

(85)

and

$$
\text{Var}[\hat{\gamma}_0] = h\text{Var}[\hat{\gamma}_1]h' + \frac{1 + E[\text{tr}(\tilde{D}_2 + \tilde{D}_1V_{11}\tilde{D}_1' + T\tilde{D}_1\tilde{\gamma}_1\tilde{\gamma}_1'\tilde{D}_1')]}{T(V_N^2\Sigma^{-1}1_N)} + \frac{hV_{11}h' - 2hE[\sqrt{T}\tilde{V}_{11}^{\frac{1}{2}}\tilde{D}_1]V_{11}h}{T}.
$$

(86)

Unlike the mean, the variance of $\hat{\gamma}$ from the estimated GLS is not the same as the variance of $\hat{\gamma}$ from the true GLS. From Lemma 1, we have

$$
\text{Var}^c[\hat{\gamma}_1] = \text{Var}^c[\hat{\gamma}_1] + \frac{V_{11}^{\frac{1}{2}}E[C(Y_2'[I_{N-1} - Z_2(Z_2'Z_2)^{-1}Z_2']Y_2]\tilde{D}_2]\tilde{V}_{11}^{\frac{1}{2}}}{T - N - 1}
$$

$$
= \text{Var}^c[\hat{\gamma}_1] + \frac{V_{11}^{\frac{1}{2}}E[C(Y_2'[I_{N-1} - Z_2(Z_2'Z_2)^{-1}Z_2']Y_2]\tilde{D}_2[Z_2]\tilde{V}_{11}^{\frac{1}{2}}}{T - N - 1}
$$

$$
= \text{Var}^c[\hat{\gamma}_1] + \Delta^c,
$$

(87)
where
\[
\Delta^c = \frac{E^c[((N - K - 1) + \hat{T}\gamma_1\hat{C}\gamma_1)\hat{V}_{11}\hat{D}_2\hat{V}_{11}]}{T - N - 1},
\]
with
\[
\hat{C} = \eta[I_{N-1} - Z_2(Z_2'Z_2)^{-1}Z_2']\eta.
\]

The last equality follows because, conditional on \(\mu_1, \hat{V}_{11}\) and \(Z_2\), we have \(Y_2'[I_{N-1} - Z_2(Z_2'Z_2)^{-1}Z_2']Y_2 \sim \chi^2_{N-K-1}(T\gamma_1'\hat{C}\gamma_1)\) and its expected value is \(N - K - 1 + T\gamma_1'\hat{C}\gamma_1\). Similarly, the conditional variance of \(\gamma_0\) is given by

\[
\text{Var}^c[\gamma_0] = \text{Var}^c[\gamma_0] + \frac{E^c[(Y_2'[I_{N-1} - Z_2(Z_2'Z_2)^{-1}Z_2']Y_2)(1 + Z_1\hat{D}_2Z_1')]}{T(T - N - 1)(1_N'\Sigma^{-1}1_N)}
\]

\[
= \frac{E^c[((N - K - 1) + T\gamma_1'\hat{C}\gamma_1)(1 + T\delta\hat{V}_{11}\hat{D}_2\hat{V}_{11}'h' + \text{tr}(\hat{D}_2))]}{T(T - N - 1)(1_N'\Sigma^{-1}1_N)}
\]

\[
= \text{Var}^c[\gamma_0] + h\Delta^c h' + \frac{E^c[((N - K - 1) + T\gamma_1'\hat{C}\gamma_1)(1 + \text{tr}(\hat{D}_2))]}{T(T - N - 1)(1_N'\Sigma^{-1}1_N)},
\]

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

Unconditionally, from Lemma 1 we have
\[
\text{Var}[\gamma_1] = \text{Var}[\gamma_1] + \Delta,
\]
where \(\Delta = E[\Delta^c]\) and
\[
\text{Var}[\gamma_0] = \text{Var}[\gamma_0] + h\Delta h' + \frac{E[(((N - K - 1) + T\gamma_1'\hat{C}\gamma_1 + \text{tr}(\hat{C}V_{11}))(1 + \text{tr}(\hat{D}_2))]}{T(T - N - 1)(1_N'\Sigma^{-1}1_N)}.
\]

Note that similar to 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 \(\hat{V}_{11}\) (for the unconditional variance). In order to approximate those expectations, we just need to simulate \(Z_2\) and \(\hat{V}_{11}\) and use the average values of those functions.

It is interesting to note that the conditional and unconditional variances of both the OLS and the GLS estimators of \(\gamma_0\) and \(\gamma_1\) depend on the value of \(\gamma_1\), but not \(\gamma_0\), so the actual value of the zero-beta rate is irrelevant in determining the variance of the estimated \(\gamma\). The reason that \(\gamma_1\) plays a role in determining the variance of the estimated \(\gamma\) but not \(\gamma_0\) is that there are measurement errors in the betas, but not in the vector of ones. In fact, even as \(T \to \infty\), we can see from the
expressions of the asymptotic variance of the estimated $\gamma$ in (26)–(28) that the measurement errors in betas still have an effect on the variance of the estimated $\gamma$, and the effect is nicely summarized by the term $\gamma_1V_{11}^{-1}\gamma_1$.

III. Analytical Results for the Single Factor Case

A. Finite Sample Distribution of Estimated Risk Premium

In the previous section, we suggest that one can simulate the conditional distribution of $\hat{\gamma}$, $\check{\gamma}$ and $\hat{\gamma}$ by simulating some normal random variables $Y$ and $Z$ (and also a chi-squared random variable $U$ for $\hat{\gamma}$). For conditional first and second moments of the estimated $\gamma$, we can approximate them by simulating only $Z_2$ (an $(N-1) \times K$ normal random variable). For unconditional first and second moments, we also need to simulate $\hat{V}_{11}$. 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 $\gamma$ without doing a simulation. For $K > 1$, it is a formidable task and we are unable to accomplish it. Nevertheless, we can do it for $K = 1$ and we present the analytical results for the single factor case in this section. While the case of a single factor is a special case, it is 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 important for us to understand the behavior of the estimated $\gamma$ 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 $\hat{\gamma}_1$ of the OLS CSR, we need to know how to evaluate the conditional distribution of $(Z_2^T\Lambda Z_2)^{-1}Z_2^TY_2$. When $K = 1$, $Z_2$ is just an $(N-1)$-vector of normal random variables. Let $X = [Y_2', Z_2']'$, conditional on $\hat{\mu}_1$ and $\hat{V}_{11}$, $X \sim N(\mu_X, I_{2N-2})$, we have

$$
\mu_X = \begin{bmatrix}
\sqrt{T} \eta \gamma_1 \\
\sqrt{T} \eta \gamma_1^2
\end{bmatrix}.
$$

Defining

$$
A = \begin{bmatrix}
O_{(N-1) \times (N-1)} & \Lambda/2 \\
\Lambda/2 & O_{(N-1) \times (N-1)}
\end{bmatrix},
$$

$$
B = \begin{bmatrix}
O_{(N-1) \times (N-1)} & O_{(N-1) \times (N-1)} \\
O_{(N-1) \times (N-1)} & \Lambda
\end{bmatrix},
$$

20
we can easily verify that \((X'AX)/(X'BX) = (Z_2'\Lambda Y_2)/(Z_2'\Lambda Z_2)\) and from (35), we can write
\[
\hat{\gamma}_1 = \hat{V}_{11}^{1/2} \left( \frac{X'AX}{X'BX} \right),
\]
and
\[
P[\hat{\gamma}_1 > c | \hat{\mu}_1, \hat{V}_{11}] = P \left[ \frac{X'AX}{X'BX} > \frac{c}{\hat{V}_{11}^{1/2}} \hat{\mu}_1, \hat{V}_{11} \right] = P \left[ X' \left( A - \frac{c}{\hat{V}_{11}^{1/2}} B \right) X > 0 \bigg| \hat{\mu}_1, \hat{V}_{11} \right].
\]
Let \(QDQ'\) be the eigenvalue decomposition of \(A - c\hat{V}_{11}^{-1/2} B\), where \(D\) is an \(n\)-dimensional \((n \leq 2N-2)\) diagonal matrix of the nonzero eigenvalues of \(A - c\hat{V}_{11}^{-1/2} B\) and \(Q\) is an \((2N-2) \times n\) matrix with its columns being the corresponding eigenvectors. Defining \(x = Q'X\), we can then write
\[
P[\hat{\gamma}_1 > c | \hat{\mu}_1, \hat{V}_{11}] = P[x'Dx > 0 | \hat{\mu}_1, \hat{V}_{11}].
\]
Conditional on \(\hat{\mu}_1\) and \(\hat{V}_{11}\), \(x \sim N(Q'\mu_X, I_n)\), so \(x'Dx\) is just a linear combination of \(n\) independent \(\chi_1^2\) random variables, and the probability can be easily evaluated using numerical procedures from Farebrother (1984, 1990).\(^7\) Similarly, to obtain the conditional distribution of \(\hat{\gamma}_1\) from the true GLS, we can simply set \(\Lambda = I_{N-1}\) in the above procedure.

B. Mean of Estimated Zero-Beta Rate and Risk Premium

It turns out that for the single factor case, the conditional and unconditional mean of the estimated \(\gamma\) can be written as 1-dimensional integrals. Starting with the case of the OLS CSR, we present the conditional mean of \(\hat{\gamma}\) in the following proposition.

**Proposition 2:** For the single factor case, the conditional means of the second pass OLS CSR estimators of \(\gamma_1\) and \(\gamma_0\) exist for \(N \geq 3\) and they are given by
\[
E^c[\hat{\gamma}_1] = \frac{T\hat{V}_{11}}{2} \int_0^1 \left[ g_1 \left( \prod_{i=1}^{N-1} a_i \right)^{1/2} e^{-\frac{TV_{11}}{2} \sum_{i=1}^{N-1} \eta_i^2 (a_iy-1) y^{N-3} } dy \right] \hat{\gamma}_1,
\]
\[
E^c[\hat{\gamma}_0] = \gamma_0 + h(\hat{\gamma}_1 - E^c[\hat{\gamma}_1]) + (\xi'\eta - c) \hat{\gamma}_1,
\]
with \(h = (1_N' \Sigma^{-1} \beta)/(1_N' \Sigma^{-1} 1_N)\) and
\[
c = E^c \left[ \frac{Z_2' \xi \eta' \Lambda Z_2}{Z_2' \Lambda Z_2} \right] = \frac{1}{2} \int_0^1 (g_2 + T\hat{V}_{11} g_1 g_3 y) \left( \prod_{i=1}^{N-1} a_i \right)^{1/2} e^{-\frac{TV_{11}}{2} \sum_{i=1}^{N-1} \eta_i^2 (a_iy-1) y^{N-3} } dy,
\]

\(^7\)Forchini (2001) provides an explicit expression for the cumulative density function of \(X'AX/(X'BX)\).
where

\[ a_i = \frac{1}{\lambda_i^* - (\lambda_i^* - 1) y}. \]  

(102)

\( \eta_i \) is the \( i \)-th element of \( \eta = P'\Sigma^{-\frac{1}{2}} \beta \), \( \xi_i \) is the \( i \)-th element of \( \xi = P'\Sigma^{\frac{1}{2}} \eta / N \), \( \lambda_i^* = \lambda_i / \lambda_{N-1} \),

\[ g_1 = \sum_{i=1}^{N-1} a_i \lambda_i^* \eta_i^2, \quad g_2 = \sum_{i=1}^{N-1} a_i \lambda_i^* \eta_i \xi_i, \quad \text{and} \quad g_3 = \sum_{i=1}^{N-1} a_i \eta_i \xi_i. \]

Although (99) and (101) look complex, they are only 1-dimensional integrals and can be easily evaluated with the knowledge of \( \eta_i^2, \xi_i \) and \( \lambda_i^* \).

For the true and the estimated GLS CSR, the conditional mean of the estimated \( \gamma \) is obtained by substituting \( \lambda_i^* = 1 \) and \( \xi_i = 0 \) in (99) and (101). The resulting expressions are very simple and are summarized in the following proposition.

**Proposition 3:** For the single factor case, the conditional means of the second pass GLS CSR estimators of \( \gamma \) exist for \( N \geq 3 \) and they are given by

\[ E^c[\hat{\gamma}_1] = E^c[\tilde{\gamma}_1] = \left( \frac{T \hat{V}_{11} \eta}{2} \right) \int_0^1 e^{-\frac{y}{2} (y-1)} y^{\frac{N-3}{2}} dy \bar{\gamma}_1, \]  

(103)

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

(104)

In fact, we do not even need to do numerical integration to obtain the conditional expected value of the estimated \( \gamma \) from the second pass GLS CSR. The following lemma presents a simplification formula.

**Lemma 2** Suppose \( b \) is a positive scalar and \( n \) is a nonnegative integer, we have

\[ b \int_0^1 e^{b(y-1)} y^2 dy = \sum_{r=0}^{\frac{n}{2}} \frac{\left( \frac{n}{2} - r + 1 \right)_r}{(-b)^r} - \frac{(\frac{n}{2})! e^{-b}}{(-b)^{\frac{n}{2}}} \]  

(105)

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

\[ b \int_0^1 e^{b(y-1)} y^2 dy = \sum_{r=0}^{\frac{n-1}{2}} \frac{\left( \frac{n}{2} - r + 1 \right)_r}{(-b)^r} - \frac{(\frac{n}{2})! e^{-b}}{(-b)^{\frac{n-1}{2}}} \frac{D(\sqrt{b})}{\sqrt{b}} \]  

(106)

for odd \( n \), where

\[ D(x) = e^{-x^2} \int_0^x e^t dt \]  

(107)

is the Dawson’s integral, which is readily available from many mathematical programs.
If we define
\[ \tilde{\kappa} = \frac{\theta}{2} \int_{0}^{1} e^{\frac{\eta \theta}{2}} y^{\frac{N-3}{2}} dy, \]
where \( \theta = T \hat{V}_{11} \eta' \eta \), then we have \( E[\tilde{\gamma}_1] = \tilde{\kappa} \tilde{\gamma}_1 \). As \( \tilde{\kappa} \) depends on \( \hat{V}_{11} \) but not \( \hat{\mu}_1 \), we also have
\[ E[\tilde{\gamma}_1 | \hat{V}_{11}] = \tilde{\kappa} \hat{\gamma}_1, \]
and we can think of \( \tilde{\kappa} - 1 \) as the percentage bias of \( \tilde{\gamma}_1 \) when conditional on \( \hat{V}_{11} \). Note that \( \tilde{\kappa} \) is only a function of \( \theta \) and \( N \), so these two parameters jointly determine the conditional percentage bias of \( \tilde{\gamma}_1 \). The following lemma gives some properties of \( \tilde{\kappa} \).

**Lemma 3** Conditional on \( \hat{V}_{11} \), \( \tilde{\kappa} \) is an increasing function of \( \theta \) and a decreasing function of \( N \), and we have \( 0 < \tilde{\kappa} < 1 \). As \( \theta \) approaches infinity, \( \tilde{\kappa} \) approaches the limit of one.

Lemma 3 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 \( T \hat{V}_{11} \eta' \eta \). In order to understand what \( T \hat{V}_{11} \eta' \eta \) represents, we write the GLS CSR of \( \hat{\mu}_2 \) on \( 1_N \) and \( \beta \) as an OLS CSR of
\[ \Sigma^{-\frac{1}{2}} \hat{\mu}_2 = \Sigma^{-\frac{1}{2}} 1_N \gamma_0 + \Sigma^{-\frac{1}{2}} \beta \gamma_1 + e, \]
where \( e \) is an \( N \)-vector of error terms. Premultiplying both sides by \( P' \) and noting that \( P' \Sigma^{-\frac{1}{2}} 1_N = 0_{N-1} \), we have
\[ P' \Sigma^{-\frac{1}{2}} \hat{\mu}_2 = P' \Sigma^{-\frac{1}{2}} \beta \gamma_1 + \varepsilon, \]
where \( \varepsilon = P' e \). Let \( y = P' \Sigma^{-\frac{1}{2}} \hat{\mu}_2 \), we can think of the true GLS CSR estimates \( \gamma_1 \) by running the following OLS regression of
\[ y_i = \gamma_1 \eta_i + \varepsilon_i, \quad i = 1, \ldots, N - 1. \]
Of course, we do not use true \( \beta \) but estimated \( \beta \) in the CSR. Define \( \eta_i = P' \Sigma^{-\frac{1}{2}} \beta = \eta_i + n_i \), where \( n_i \) is the measurement error of \( \eta_i \). Therefore, the regression that we run is actually
\[ y_i = \gamma_1 \eta_i^* + \varepsilon_i^*, \quad i = 1, \ldots, N - 1. \]

The resulting estimate of \( \gamma_1 \) from this OLS regression is
\[ \tilde{\gamma}_1 = \frac{\sum_{i=1}^{N-1} \eta_i^* y_i}{\sum_{i=1}^{N-1} (\eta_i^*)^2}. \]
Note that this is the classical problem of EIV and the bias depends on the ratio of \( \sum_{i=1}^{N-1} \eta_i^2/(N - 1) = \eta^2/(N - 1) \) (the signal) to \( \text{Var}[n_i] \) (the noise). Note that conditional on \( \hat{V}_{11} \),

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

(115)

so \( n_i \)'s are independent of each other and their variance is \( (T\hat{V}_{11})^{-1} \). With this analysis, we now understand that \( T\hat{V}_{11} \eta^2/(N - 1) \) is a measure of signal-to-noise ratio in the estimated betas. This explains why the percentage bias of \( \tilde{\gamma}_1 \) is a decreasing function of \( \theta \) for a fixed \( N \). This result is largely consistent with the traditional analysis of errors-in-variables in the regression framework, which suggests that when the independent variable is measured with errors, 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 the traditional EIV analysis provides only asymptotic results when \( N \to \infty \) and we provide exact finite sample results here. Lemma 3 suggests that there are two ways to reduce the bias of \( \tilde{\gamma}_1 \): one is to increase the length of the time series, another is to use test assets that have a wide dispersion in \( \beta \). As for the effect of the number of test assets, we know from Lemma 3 that \( \tilde{\kappa} \) is a decreasing function of \( N \) for a fixed value of \( T\hat{V}_{11} \eta^2/N \). However, the effect of increasing \( N \) on the bias is not clear because \( \eta^2/N \) also typically increases with \( N \). If we reasonably assume that \( \eta^2/N \) is a constant for different choices of \( N \), then we can find out whether \( \tilde{\kappa} \) is an increasing function of \( N \). In Figure 1, we plot \( \tilde{\kappa} \) as a function of \( T\hat{V}_{11} \eta^2/(N - 1) \) for \( N = 5, 10, 25, \) and 100 over the range of \( 0 \leq T\hat{V}_{11} \eta^2/(N - 1) \leq 10 \) (which covers the range of \( T\hat{V}_{11} \eta^2/(N - 1) \) that we encounter in typical applications). As we see in Figure 1, if \( \eta^2/N \) is constant across different choices of \( N \), then the bias is an increasing function of \( N \), but the difference between different choices of \( N \) is quite small.\(^8\) This also suggests that while \( \tilde{\kappa} \) is a function of \( T\hat{V}_{11} \eta^2 \) and \( N \), the bias of the GLS estimate of \( \gamma_1 \) is mostly determined by the signal-to-noise ratio \( T\hat{V}_{11} \eta^2/(N - 1) \).

Figure 1 about here

8In practice, there is another advantage of using smaller \( N \) in the two-pass CSR. It is that when we use a few well diversified portfolios instead of a large number of individual stocks as test assets, the variance of the residuals of the test assets \( \Sigma \) is typically smaller and hence we can get a larger \( \eta^2/N \), which further reduces the bias.
Similar to the case of \( \tilde{\kappa} \), we have \( E[\tilde{\gamma}_1|\hat{V}_{11}] = \tilde{\kappa}\gamma_1 \) and we can interpret \( \tilde{\kappa} - 1 \) as the percentage bias of \( \tilde{\gamma}_1 \) when conditional on \( \hat{V}_{11} \). Unlike the case of \( \tilde{\kappa} \) which depends on only \( N \) and \( TV_{11}\eta'\eta, \tilde{\kappa} \) depends on \( \lambda_i^*, \eta_2^2, N \) and \( TV_{11} \), so the individual elements of \( \Lambda \) and \( \eta \) 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 we cannot claim that \( \tilde{\gamma}_1 \) is biased toward zero. In fact, for some choices of \( \lambda_i^* \) and \( \eta_i \), we can have \( \tilde{\kappa} > 1 \). The intuition that EIV causes the slope coefficient to be biased toward zero does not apply here because unlike the GLS CSR, the measurement errors of the independent variable (\( \hat{\beta} \)) are in general not uncorrelated with each other in the OLS CSR. In general, depending on the values of \( \lambda_i \) and \( \eta_i \), the bias of \( \tilde{\gamma}_1 \) can be more or less than that of \( \tilde{\gamma}_1 \). The following lemma compares \( \tilde{\kappa} \) and \( \tilde{\kappa} \) for two extreme cases.

**Lemma 4** If \( \eta_1 \neq 0 \) and \( \eta_2 = \cdots = \eta_{N-1} = 0 \), we have \( \tilde{\kappa} \geq \tilde{\kappa} \). If \( \eta_1 = \cdots = \eta_{N-2} = 0 \) and \( \eta_{N-1} \neq 0 \), we have \( \tilde{\kappa} \leq \tilde{\kappa} \). The equalities hold if and only if \( \lambda_1 = \lambda_{N-1} \).

Heuristically, Lemma 4 suggests that when \( \eta_1^2 \) is large and \( \eta_2^2 \) to \( \eta_{N-1}^2 \) are small (i.e., when \( \Sigma^{-\frac{1}{2}}\beta \) is close to proportional to the eigenvector associated with the largest eigenvalue of \( \Sigma M \Sigma \)), it is possible to find that the bias of the GLS CSR estimate of \( \gamma_1 \) is more severe than that of the OLS CSR. On the contrary, if \( \eta_{N-1}^2 \) is large, but \( \eta_1^2 \) to \( \eta_{N-2}^2 \) are small (i.e., when \( \Sigma^{-\frac{1}{2}}\beta \) is close to proportional to the eigenvector associated with the smallest nonzero eigenvalue of \( \Sigma M \Sigma \)), then \( 0 < \tilde{\kappa} \leq \tilde{\kappa} < 1 \) and the bias of the OLS CSR estimate of \( \gamma_1 \) is more severe than that of the GLS CSR. Since there is no theoretical relation between \( \beta \) and \( \Sigma \), it is not clear whether we should expect the OLS or the GLS CSR estimate of \( \gamma_1 \) to have more bias. It is also important to note that the bias of the OLS CSR estimate of \( \gamma_1 \) 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 \( \tilde{\gamma} \) and \( \hat{\gamma} \), which are invariant to portfolio repacking of the original \( N \) test assets.\[10\]

---

\[9\] Since \( \eta = P'\Sigma^{-\frac{1}{2}}\beta \), so \( \eta_i = P'_i\Sigma^{-\frac{1}{2}}\beta \), where \( P_i \) is the eigenvector associated with the \( i \)-th largest eigenvalue of \( \Sigma M \Sigma \).

\[10\] See Kandel and Stambaugh (1995) for a discussion of the invariance property of the GLS CSR. The analysis of Kandel and Stambaugh (1995) is based on the population measures of \( \mu_2 \) and \( \beta \), but it can be easily generalized to show that \( \tilde{\gamma}_1 \) and \( \gamma_1 \) are invariant to portfolio repackaging.
So far we have discussed the conditional mean of ˇγ, ˜γ and ˆγ. The unconditional mean can be obtained by using the fact that $T \hat{V}_{11}/V_{11} \sim \chi^2_{T-1}$ and is independent of ˆµ1. In order to facilitate our presentation of the unconditional results, we define two 1-dimensional integrals. The first integral is defined as

$$\varphi_{m,n}(g) = \int_0^1 g(y) \left( \prod_{i=1}^{N-1} a_i \right)^{\frac{1}{2}} \frac{y^{m+n}}{\left[ 1 + V_{11} \sum_{i=1}^{N-1} \eta_i^2 (1 - a_i y) \right]^{\frac{n}{2}}} \ dy,$$  \hspace{1cm} (117)

where $a_i$ is defined in (102), $g$ is a function of $y$, and $m > -2$. The second integral is defined as

$$\phi_m = \int_0^1 \frac{y^m}{\left[ 1 + V_{11} \eta \eta (1 - y) \right]^{\frac{m+n}{2}}} \ dy.$$  \hspace{1cm} (118)

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

$$\phi_m = 2F_1 \left( \frac{m+2}{2}, \frac{T+1}{2}, \frac{m+4}{2}, \frac{V_{11} \eta \eta}{1 + V_{11} \eta \eta} \right),$$  \hspace{1cm} (119)

where

$$2F_1(a, b, c, x) = \sum_{r=1}^{\infty} \frac{(a)_r (b)_r}{(c)_r} \frac{x^r}{r!}$$  \hspace{1cm} (120)

is the hypergeometric function, which is readily available in many mathematical programs. Using these two 1-dimensional integrals, the following proposition provides the expressions for the unconditional mean of γ0 and γ1 from the second pass CSR.

**Proposition 4:** For the single factor case, the unconditional means of the second pass OLS CSR estimators of γ1 and γ0 exist for $N \geq 3$ and they are given by $E[\tilde{\gamma}_1] = \tilde{\kappa}_u \gamma_1$ where

$$\tilde{\kappa}_u = \frac{(T-1)V_{11}}{2} \varphi_{N-3,T+1}(g_1),$$  \hspace{1cm} (121)

and

$$E[\bar{\gamma}_0] = \gamma_0 + h(\gamma_1 - E[\tilde{\gamma}_1]) + (\xi^\prime \eta - c_u) \gamma_1,$$  \hspace{1cm} (122)

with $h = (1_N^\prime \Sigma^{-1} \beta)/(1_N^\prime \Sigma^{-1} 1_N)$ and

$$c_u = E \left[ \frac{Z_2^\prime \xi \eta \Lambda Z_2}{Z_2^\prime A Z_2} \right] = \frac{1}{2} \varphi_{N-3,T-1}(g_2) + \frac{(T-1)V_{11}}{2} \varphi_{N-1,T+1}(g_1 g_3),$$  \hspace{1cm} (123)

where $g_1$ to $g_3$ are defined in Proposition 2. In addition, the unconditional means of the second pass GLS CSR estimators of γ1 and γ0 exist for $N \geq 3$ and they are given by $E[\hat{\gamma}_1] = E[\bar{\gamma}_1] = \hat{\kappa}_u \gamma_1$ where

$$\hat{\kappa}_u = \frac{(T-1)V_{11} \eta \eta \varphi_{N-3}}{2},$$  \hspace{1cm} (124)
From Proposition 4, we can see that the unconditional percentage bias of \( \tilde{\gamma}_1 \) depends on \( V_{11} \eta', \eta, N, \) and \( T. \) As \( 0 < \hat{\kappa} < 1, \) \( \tilde{\kappa}_u = E[\tilde{\kappa}] \) is also bounded by 0 and 1. For the unconditional percentage bias of \( \hat{\gamma}_1, \) it depends on \( V_{11}, \eta_i^2, \lambda_i^*, N, \) and \( T. \) Just as in the conditional case, \( \tilde{\kappa}_u \) can be greater than or less than \( \tilde{\kappa}_u \) depending on the value of the parameters \( \lambda_i^* \) and \( \eta_i. \) Without knowing the value of these parameters, it is not clear whether the GLS or the OLS CSR estimate of \( \gamma_1 \) has more bias unconditionally.

### C. Variance of Estimated Zero-Beta Rate and Risk Premium

Similar to the conditional and unconditional mean, the conditional and unconditional variance of the estimated \( \gamma \) can also be written as 1-dimensional integrals for the single factor case. As conditional variance is less relevant for the purpose of statistical inference, we only present the results on unconditional variance.\(^{11}\) Nevertheless, the results on conditional variance of the estimated \( \gamma \) can be easily obtained from the proof in the Appendix.

Starting with the case of the OLS CSR, the unconditional variance of \( \hat{\gamma} \) is presented in the following proposition.

**Proposition 5:** For the single factor case, the unconditional variances of the second pass OLS CSR estimators of \( \gamma_1 \) and \( \gamma_0 \) exist for \( N \geq 4. \) The unconditional variance of \( \gamma_1 \) is given by

\[
\text{Var}[\gamma_1] = \frac{(T - 1)(T + 1)V_{11}^2}{4T} \varphi^d_{N-3,T+3}(g_4 + ag_1^2) + \frac{(T - 1)V_{11}}{4T} \varphi^d_{N-5,T+1}(g_5 + ag_6) - \tilde{\kappa}_u^2 \gamma_1, \tag{126}
\]

where \( \varphi_{m,n}^d(g) = \varphi_{m,n}(g) - \varphi_{m+2,n}(g), \) \( g_4 = \sum_{i=1}^{N-1} a_i^2 \lambda_i^2 \eta_i^2, \) \( g_5 = \sum_{i=1}^{N-1} a_i \lambda_i^2, \) \( g_6 = \sum_{i=1}^{N-1} a_i^2 \lambda_i^2 \eta_i^2, \) \( a = T \gamma_1^2 + V_{11}, \) and \( \tilde{\kappa}_u \) is defined in (121). For the unconditional variance of \( \gamma_0, \) it is given by

\[
\text{Var}[\gamma_0] = h^2 E[\gamma_1^2] + \frac{V_{11}}{T}[h + \xi'(\hat{\gamma}_1 + c_u)]^2
+ \frac{\xi' \xi - 2(hQ_1 + Q_2) + 2hQ_4 + Q_5 - a(h\tilde{\kappa}_u + c_u)^2}{T} + \frac{1 + Q_3}{T(1'N \Sigma^{-1}1_N)}, \tag{127}
\]

\(^{11}\)When conditional on \( \hat{\mu}_1, \) the returns and hence the estimated \( \gamma_1 \) are much less volatile than when unconditional on \( \hat{\mu}_1. \) However, for statistical inference purposes, we are more interested in finding out the value of \( \gamma_1 \) rather than that of \( \bar{\gamma}_1 = \gamma_1 - \mu_1 + \hat{\mu}_1. \) As a result, when making inference on \( \gamma_1, \) we need to take into account the sampling variability of \( \hat{\mu}_1 \) by using the unconditional variance of the estimated \( \gamma_1. \)
where \( h = (1'N\Sigma^{-1}\beta)/(1'N\Sigma^{-1}1_N) \), \( c_u \) is defined in (123), and \( Q_1 \) to \( Q_5 \) are the following 1-dimensional integrals

\[
Q_1 = \frac{(T-1)V_{11}}{2} \varphi_{N-3,T+1}(g_2),
\]

\[
Q_2 = \frac{1}{2} \varphi_{N-3,T-1} \left( \sum_{i=1}^{N-1} a_i \lambda_i^2 \xi_i^2 \right) + \frac{(T-1)V_{11}}{2} \varphi_{N-1,T+1}(g_2g_3),
\]

\[
Q_3 = \frac{(T-1)V_{11}}{4} \varphi_{N-3,T+1}^d (g_4 + ag_2^2) + \frac{1}{4} \varphi_{N-5,T-1}^d (g_5 + ag_6),
\]

\[
Q_4 = \frac{(T-1)(T+1)V_{11}^2}{4} \varphi_{N-3,T+1}^d ((g_4 + ag_1^2)g_3)
+ \frac{(T-1)V_{11}}{4} \varphi_{N-1,T+1}^d ((g_5 + ag_6)g_3 + 2g_7 + 2ag_1g_2),
\]

\[
Q_5 = \frac{(T-1)(T+1)V_{11}^2}{4} \varphi_{N-1,T+3}^d ((g_4 + ag_1^2)g_3^2)
+ \frac{(T-1)V_{11}}{4} \varphi_{N-1,T+1}^d ((g_5 + ag_6)g_3^2 + (g_4 + ag_1^2)g_8 + 4(g_7 + ag_1g_2)g_3)
+ \frac{1}{4} \varphi_{N-3,T-1}^d (g_5 + ag_6)g_8 + 2 \sum_{i=1}^{N-1} a_i^2 \lambda_i^{2*} \xi_i^2 + 2ag_2^2,
\]

where \( g_1 \) to \( g_3 \) are defined in Proposition 2, \( g_7 = \sum_{i=1}^{N-1} a_i^2 \lambda_i^{2*} \eta_i \xi_i \) and \( g_8 = \sum_{i=1}^{N-1} a_i \xi_i^2 \).

For the true GLS CSR, the variance of \( \hat{\gamma} \) can be obtained by setting \( \lambda_i^* = 1 \) and \( \xi_i = 0 \) in the expressions for the OLS case. After some simplifications, the results are given in the following proposition.

**Proposition 6:** For the single factor case, the unconditional variances of the second pass true GLS CSR estimators of \( \gamma_1 \) and \( \gamma_0 \) exist for \( N \geq 4 \). The unconditional variance of \( \hat{\gamma}_1 \) is given by

\[
\text{Var}[\hat{\gamma}_1] = \frac{(T-1)V_{11}}{4T} \left[ (N-2)\alpha \eta^' \phi_{N-3} + [2 - (N-4)\alpha \eta^' \phi_{N-5}] - \hat{\kappa}_u \hat{\gamma}_1 \right],
\]

where \( a = T\gamma_1^2 + V_{11} \) and \( \hat{\kappa}_u \) is defined in (124). The unconditional variance of \( \hat{\gamma}_0 \) is given by

\[
\text{Var}[\hat{\gamma}_0] = h^2 \text{Var}[\hat{\gamma}_1] + \frac{V_{11}}{T} h^2 (1 - 2\hat{\kappa}_u) + \frac{1 + \hat{Q}_3}{T(1'N\Sigma^{-1}1_N)},
\]

where \( h = (1'N\Sigma^{-1}\beta)/(1'N\Sigma^{-1}1_N) \) and

\[
\hat{Q}_3 = \frac{N-1 + \alpha \eta^' \eta}{(N-1)(N-3)} - \frac{(T-1)V_{11} \alpha \eta^' \eta}{4} \left[ \frac{(N-2)\alpha \eta^' \eta}{N-1} \phi_{N-1} + \frac{2 - (N-4)\alpha \eta^' \eta}{N-3} \phi_{N-3} \right].
\]
Finally, our next proposition presents the unconditional variance of the estimated $\gamma$ from the estimated GLS CSR.

**Proposition 7:** For the single factor case, the unconditional variances of the second pass estimators of $\gamma_1$ and $\gamma_0$ from the estimated GLS CSR exist for $N \geq 4$. The unconditional variance of $\hat{\gamma}_1$ is given by

$$
\text{Var}[\hat{\gamma}_1] = \text{Var}[\tilde{\gamma}_1] + \Delta,
$$

where

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

and $a = T\gamma_1^2 + V_{11}$. The unconditional variance of $\hat{\gamma}_0$ is given by

$$
\text{Var}[\hat{\gamma}_0] = \text{Var}[\tilde{\gamma}_0] + h^2\Delta + \frac{(N - 2)d}{(T - N - 1)T(1'N\Sigma^{-1}1N)},
$$

where $h = (1'N\Sigma^{-1}\beta)/(1'N\Sigma^{-1}1N)$ and

$$
d = \frac{(N - 2)(N - 1 + a\eta'\eta)}{(N - 1)(N - 3)} - \frac{(T - 1)V_{11}\eta'\eta}{4} \left( \frac{a\eta'\eta}{N - 1} \phi_{N-1} + \frac{a\eta'\eta + 2}{N - 3} \phi_{N-3} \right).
$$

**IV. Bias-adjusted Estimators of Zero-Beta Rate and Risk Premium**

Since the second pass CSR estimators of $\gamma_0$ and $\gamma_1$ are biased in finite samples, we would like to correct for this bias. In this section, we present a bias-adjusted version of the two-pass CSR estimators of zero-beta rate and risk premium. We focus our discussion on the 1-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 simulated 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 we know the value of $\tilde{\kappa}$ in (108) and $\hat{\kappa}$ in (116), then we can construct the following adjusted estimated GLS and OLS estimators of $\gamma_1$

$$
\tilde{\gamma}_1^a = \frac{\hat{\gamma}_1}{\hat{\kappa}},
$$

$$
\hat{\gamma}_1^a = \frac{\hat{\gamma}_1}{\hat{\kappa}}.
$$
As \( E[\hat{\gamma}_1|\hat{V}_{11}] = \hat{\kappa}_1 \) and \( E[\hat{\gamma}_1|\hat{V}_{11}] = \kappa_1 \), the adjusted estimators are unbiased when conditional on \( \hat{V}_{11} \), and hence also unconditionally unbiased. For the zero-beta rate, the adjusted GLS and OLS estimators are given by
\[
\hat{\gamma}_0^a = \hat{\gamma}_0 - \frac{h(1 - \hat{\kappa})}{\hat{\kappa}} \gamma_1, \tag{142}
\]
\[
\hat{\gamma}_0^a = \hat{\gamma}_0 - \frac{h(1 - \hat{\kappa}) + \xi'\eta - c}{\hat{\kappa}} \gamma_1, \tag{143}
\]
where \( h = (1_N'\Sigma^{-1}\beta)/(1_N'\Sigma^{-1}1_N) \) and \( c \) is defined in (101). In order to construct the unbiased estimators of \( \gamma_0 \), we need to know \( \hat{\kappa} \) and \( h \) for the GLS case. For the OLS case, we also need to know \( \xi'\eta \) and \( c \). Similar to the adjusted estimators for \( \gamma_1 \), \( \hat{\gamma}_0^a \) and \( \hat{\gamma}_0^a \) are conditionally and unconditionally unbiased.

In practice, \( \hat{\kappa} \) and \( \hat{\kappa} \) are in general unknown, so we need to estimate them. Note that when we use the estimated \( \hat{\kappa} \) and \( \hat{\kappa} \) instead of the true \( \hat{\kappa} \) and \( \hat{\kappa} \), the adjusted estimators are no longer unbiased. Nevertheless, it is reasonable to expect that by making the adjustment, we can reduce the bias in finite samples. We start with the problem of estimating \( \hat{\kappa} \). From (108), we can see that \( \hat{\kappa} \) is only determined by \( \theta = T\hat{V}_{11}\eta'\eta \) and \( N \), and the only quantity that is unknown is \( \eta'\eta \). A sensible approach to estimate \( \hat{\kappa} \) is to replace \( \eta'\eta \) in (108) by \( \hat{\eta}'\hat{\eta} \), where \( \hat{\eta} = \hat{P}'\hat{\Sigma}^{-\frac{1}{2}}\hat{\beta} \), with \( \hat{\beta} \) and \( \hat{\Sigma} \) being the sample estimates of \( \beta \) and \( \Sigma \), and \( \hat{P}'\hat{\Lambda}\hat{P}' \) is the eigenvalue decomposition of \( \hat{\Sigma}^{\frac{1}{2}}M\hat{\Sigma}^{\frac{1}{2}} \), where \( \hat{\Lambda} = \text{Diag}(\hat{\lambda}_1, \ldots, \hat{\lambda}_{N-1}) \) is 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. However, when \( N \) is large relative to \( T \), such a procedure could lead to a severely upward biased estimate of \( \eta'\eta \). The following lemma gives the conditional distribution of \( \hat{V}_{11}\hat{\eta}'\hat{\eta} \) and its expectation.

**Lemma 5** Conditional on \( \hat{V}_{11} \), \( \hat{V}_{11}\hat{\eta}'\hat{\eta} \) is distributed as
\[
\hat{V}_{11}\hat{\eta}'\hat{\eta} \sim \frac{(N-1)}{(T-N)} F_{N-1,T-N}(T\hat{V}_{11}\hat{\eta}'\hat{\eta}), \tag{144}
\]
where \( F_{N-1,T-N}(T\hat{V}_{11}\hat{\eta}'\hat{\eta}) \) is a noncentral \( F \)-distribution with \( N - 1 \) and \( T - N \) degrees of freedom and a noncentrality parameter of \( T\hat{V}_{11}\hat{\eta}'\hat{\eta} \), and its conditional expected value is
\[
E[\hat{V}_{11}\hat{\eta}'\hat{\eta}|\hat{V}_{11}] = \left( \frac{T}{T-N-2} \right) \hat{V}_{11}\hat{\eta}'\hat{\eta} + \frac{N-1}{T-N-2}. \tag{145}
\]
Lemma 5 shows that \( \hat{V}_{11}\hat{\eta}'\hat{\eta} \) tends to overestimate \( \hat{V}_{11}\hat{\eta}'\hat{\eta} \), and the overstatement is particularly severe when \( N \) is relatively large to \( T \). As \( \hat{\kappa} \) is an increasing function of \( T\hat{V}_{11}\hat{\eta}'\hat{\eta} \), using the sample
estimate of $\hat{\eta}'\hat{\eta}$ instead of the true $\eta'\eta$ will tend to make the resulting estimated $\tilde{\kappa}$ to be too high. 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 $\theta = T\hat{V}_{11}\eta'\eta$ instead of the sample one. From Lemma 5, the unbiased estimator of $\theta$ is given by

$$\hat{\theta}_u = (T - N - 2)\hat{V}_{11}\hat{\eta}'\hat{\eta} - (N - 1).$$  \hfill (146)

However, there are two problems with this estimator. First, $\hat{\theta}_u$ can be negative with positive probability, but it is impossible to compute $\tilde{\kappa}$ when $\hat{\theta}_u$ is negative. Second, when $\hat{\theta}_u$ is close to zero, the estimated $\tilde{\kappa}$ is also close to zero and the adjusted estimator $\hat{\gamma}_1^a = \hat{\gamma}_1/\tilde{\kappa}$ is extremely large. This would have the effect of making the distribution of the adjusted estimator have a long right tail, especially when $T$ is small. In addition, while $\hat{\theta}_u$ is an unbiased estimator of $\theta$, $1/\tilde{\kappa}$ is not a linear function of $\theta$, so the resulting adjusted estimator of $\gamma_1$ can still be severely biased.

From Lemma 5, we have $X = (T - N)\hat{V}_{11}\hat{\eta}'\hat{\eta}/(N - 1) \sim F_{N-1,T-N}(\theta)$, so the problem of estimating $\theta$ 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 $\theta$ under quadratic loss function. However, 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 $1/\tilde{\kappa}$ which is a nonlinear function of $\theta$. If an estimator of $\theta$ takes a value that is very close to zero, then the implied estimator of $1/\tilde{\kappa}$ will take a very large value. In order for the implied estimator of $1/\tilde{\kappa}$ to be well behaved, we would like to prevent our estimator of $\theta$ from taking 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

$$L(\theta, \hat{\theta}) = \frac{\hat{\theta}}{\theta} - \log \left( \frac{\hat{\theta}}{\theta} \right) - 1.$$  \hfill (147)

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 $\theta$ that have a high probability of taking small values. By adopting an estimator of $\theta$ 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 for estimating the noncentrality parameter of a noncentral chi-squared distribution, so we need to extend their analysis to the case of noncentral $F$-distribution. The following lemma presents the Bayes estimator of $\theta$ under the Stein’s loss function.

**Lemma 6** Under the Stein’s loss function (147), the Bayes estimator of $\theta$ for the class of improper priors $\pi(\theta) = \theta^b$ where $b > 0$ is given by

$$\hat{\theta} = 2b \frac{2F_1 \left( 1 + b, \frac{T-1}{2}, \frac{N-1}{2}, \frac{z}{1+z} \right)}{2F_1 \left( b, \frac{T-1}{2}, \frac{N-1}{2}, \frac{z}{1+z} \right)}, \quad (148)$$

where $2F_1$ is the hypergeometric function and $z = \hat{V}_{11} \hat{\eta}' \hat{\eta}$.

In Figure 2, we plot $\hat{\theta}$ and $\hat{\theta}_u$ 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, $\hat{\theta} = 2b = 1$. As $z$ gets larger, $\hat{\theta}$ 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 $\hat{\theta}$ as an estimator of $\theta$, we 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 $\hat{\theta}$ 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 even make $\hat{\theta}$ to be greater than $(T - N - 2)z$ because in this situation we know $(T - N - 2)z$ is unusually small. As the minimum of $\hat{\theta}$ is $2b$, we can choose a value of $b$ such that $2b$ is considered to be an absolute lower bound of $\theta$.

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 to have a $\theta$ that is less than one.\(^{12}\)

\(^{12}\)If $\theta$ is indeed less than one, then our adjusted estimator will not offer sufficient adjustment. Nevertheless, some adjustment is still better than no adjustment at all.

32
We now turn to the problem of estimating $\hat{\kappa}$. The value of $\hat{\kappa}$ depends on $\theta_i = T\hat{V}_{11}\hat{\eta}_i^2$ and $\lambda_i$. For $\lambda_i$, we can simply use the sample estimate $\hat{\lambda}_i$. As for $\theta_i$, we can use its sample estimator $T\hat{V}_{11}\hat{\eta}_i^2$, but just as in the estimation of $\theta$, the sample estimator tends to overstate $\theta_i$. Analogous to the estimation of $\theta$, we use the following estimator for $\theta_i$:

$$
\hat{\theta}_i = 2b_2F_1\left(1 + b, \frac{T-N-1}{2}, \frac{1}{2}, \frac{z_i}{1 + z_i} \right),
$$

(149)

where $b > 0$ and $z_i = \hat{V}_{11}\hat{\eta}_i^2$. In our empirical work, we choose $b = 0.5$.

In order to obtain the adjusted GLS estimator of $\gamma_0$ in (142), we need to know the value of $h = (1\Sigma^{-1}\beta)/(1\Sigma^{-1}1_N)$ in addition to $\hat{\kappa}$. To estimate $h$, we simply use its sample estimate $\hat{h} = (1\hat{\Sigma}^{-1}\hat{\beta})/(1\hat{\Sigma}^{-1}1_N)$. As for the adjusted OLS estimator of $\gamma_0$ in (143), we also need to estimate $c$ in (101), which in turn requires us to estimate $\xi = P\hat{\Sigma}^{1/2}1_N/N$. For our applications, we use its sample estimate $\hat{\xi} = \hat{P}\hat{\Sigma}^{1/2}1_N/N$.

As $\hat{\theta}$ is a nonlinear function of $(T - N - 2)\hat{V}_{11}\hat{\eta}_i\hat{\eta}_j$, it is nontrivial to obtain the finite sample distribution of our adjusted estimators. We therefore 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, we should remark that both $\hat{\kappa}$ and $\hat{\kappa}$ 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.

V. Simulation Experiment

A. Design of Experiment

We perform simulation experiments to examine the robustness of our analytical results to departure from normality, as well as the finite sample properties of our adjusted second pass CSR estimators of zero-beta rate and risk premium. In choosing parameters for our simulation experiments, we

\[\text{\footnotesize{\cite{13}}\text{Although we obtain the distribution of } \hat{V}_{11}\hat{\eta}_i\hat{\eta}_j \text{ in Lemma 5, the distributions of the individual elements } \hat{V}_{11}\hat{\eta}_i^2 \text{ are unknown. When } \hat{\Sigma} \text{ is a good approximation of } \Sigma, \text{ we have } \sqrt{T}\hat{V}_{11}\hat{\eta}_i^2 \approx \sqrt{T}\hat{V}_{11}\hat{\eta}_i^2(\Sigma^{-1/2}\hat{\beta}) \sim N(\sqrt{T}\hat{V}_{11}\hat{\eta}_i^2, 1), \text{ so } T\hat{V}_{11}\hat{\eta}_i^2 \text{ is approximately distributed as } \chi^2(T\hat{V}_{11}\hat{\eta}_i^2).} \text{ To account for the variation of } \Sigma, \text{ we assume } \hat{V}_{11}\hat{\eta}_i^2 \text{ is approximately distributed as } \chi^2(T\hat{V}_{11}\hat{\eta}_i^2)/\chi^2_{T-N}, \text{ which in turn implies } \hat{V}_{11}\hat{\eta}_i\hat{\eta}_j \sim \chi^2_{T-N}(T\hat{V}_{11}\hat{\eta}_i^2)/\chi^2_{T-N} \text{ and it is consistent with the results in Lemma 5.}}\]
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 market portfolio of the NYSE. 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 variations 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 of the three sets of portfolios over the period 1941/2–2002/12. The sample estimates of $\beta$ and $\Sigma$ from this period is 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 $\beta$ and $\Sigma$ over the period 1959/2–2002/12. In Table I, we report a summary of the parameters for our three sets of portfolios under two different assumptions of the factors. For the factor with high explanatory power, the parameters are reported in Panel A. We find that the betas of the $N = 10$ assets case have lower cross-sectional variations than the other two cases. It 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. Table I also reports the signal-to-noise ratio $V_{11} \eta' \eta / (N - 1)$ (multiplied by 100) for our three sets of test assets. It can be seen that the signal-to-noise ratio is highest for the $N = 100$ assets case, so we can expect that the GLS CSR estimator of the risk premium to have the least bias for this case. However, the value of signal-to-noise ratio is chosen based on the sample estimates $\hat{\eta}' \hat{\eta}$. From Lemma 5, we know the sample estimate $\hat{\eta}' \hat{\eta}$ tends to overestimate the true $\eta' \eta$, especially when $N$ is large. Therefore, it is entirely possible that the higher signal-to-noise ratio for the $N = 100$ 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 $N = 100$ assets case as an upper bound of what we can expect from real world data when the
factor resembles the return on a market portfolio.

Panel B reports the parameters for the case that the factor has low explanatory power. When the factor is chosen to mimic the growth rate of per capita consumption, we find that the cross-sectional variance of the betas across the portfolios tends to be higher than that of the case when 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. It 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 I reports the three largest and three smallest (standardized) nonzero eigenvalues of the matrix $\Sigma_1^2 M \Sigma_2^2$. If $\Sigma_1^2 M \Sigma_2^2$ is proportional to the identity matrix, then we should have the standardized eigenvalues all equal to one. Instead, we find that $\lambda_1$ is much higher than $\lambda_{N-1}$ in all the cases. This suggests that the OLS and GLS CSR estimators of risk premium can have very different properties. Also reported in Table I are the absolute value of $\eta_i = p_i' \Sigma^{-\frac{1}{2}} \beta$ corresponding to the three largest and three smallest eigenvalues, where $p_i$ is the eigenvector of $\Sigma_1^2 M \Sigma_2^2$ associated with $\lambda_i$. Although there is no theoretical relation between $\Sigma$ and $\beta$, we typically find in the data that $\eta_1^2$ to $\eta_3^2$ are much larger than $\eta_{N-3}^2$ to $\eta_{N-1}^2$. From Lemma 4, we should expect that with our choice of $\lambda_i^2$ and $\eta_i^2$, the GLS CSR estimator of $\gamma_1$ has more bias than the OLS CSR estimator of $\gamma_1$.

B. Biases in Estimated Zero-Beta Rate and Risk Premium

With our chosen parameters, we can compute the expected value of the CSR estimators of $\gamma_0$ and $\gamma_1$ using the formulae in Proposition 4. In Table II, we report the unconditional biases of OLS and GLS CSR estimators of $\gamma_0$ and $\gamma_1$, both as a percentage of the value of the true $\gamma_1$. As the percentage bias is independent of the choice of the actual values of $\gamma_0$ and $\gamma_1$, the numbers in Table I are applicable for all choices of $\gamma_0$ and $\gamma_1$, 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 not need to distinguish these two versions of GLS here. Panel A reports the results for the factor with high explanatory power. We see that 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 $N = 10$ assets case, the biases for the GLS CSR estimators of $\gamma_0$ and $\gamma_1$ are 75.3% and −75.8%, respectively. As for the OLS CSR, the bias is smaller but it is still at a high level of 59.9% and −56.8% for $\hat{\gamma}_0$ and $\hat{\gamma}_1$. 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 pattern also holds the $N = 25$ and $N = 100$ assets cases. However, as the signal-to-noise ratio is higher for the $N = 25$ and $N = 100$ assets case, the bias is of smaller magnitude. Nevertheless, the biases are still rather significant, especially when $T$ is small.

Panel B reports the results for the factor with low explanatory power. As the consumption betas are estimated with a lot of noise, we find that the percentage biases of the CSR estimators of $\gamma_0$ and $\gamma_1$ are huge. When $T = 60$ months, the bias of the GLS estimator of $\gamma_1$ is more than −80% for all three sets of test assets. Even when we use $T = 600$ months to estimate the consumption betas, the bias of the GLS estimator of risk premium is still more than −30%. Given the huge bias of the estimated risk premium, it would be quite difficult for researchers to find the consumption betas to be priced even if the consumption CAPM is exactly correct. As in the case of Panel A, we find that the OLS estimators to have smaller biases than the GLS ones but they 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 $\hat{\gamma}_1$ goes down from −61.3% to −1.2% as $T$ increases from 60 months to 360 months. However, when $T$ goes up to 480 months, the bias of $\hat{\gamma}_1$ turns positive and increases to 2.4%, and this bias actually increases further to 3.6% when $T$ goes up to 600 months. Although not shown in the table, the bias of $\hat{\gamma}_1$ will eventually go to 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, and there is also no guarantee that the bias of $\hat{\gamma}_1$ is negative.
C. Comparison of Asymptotic and Finite Sample Standard Deviation

For statistical inference, we need to know the standard deviation of the CSR estimators of $\gamma_0$ and $\gamma_1$. Traditionally, asymptotic results are used for this purpose. With the results in Propositions 5–7, we now know the finite sample standard deviation of these estimators. In Table III, we report the asymptotic and finite sample standard deviation of the OLS and estimated GLS CSR estimators of $\gamma_0$ and $\gamma_1$. Panel A contains the results for the factor with high explanatory power and Panel B contains the results for the factor with low explanatory power. In computing the asymptotic and finite sample standard deviation of the CSR estimators, we need to make an assumption on the value of $\gamma_1$. As the high explanatory power factor case is chosen to mimic the value-weighted return of the market, we assume $\gamma_1$ is 0.6% per month. By choosing this value, we have the CAPM in mind which suggests 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 $\gamma_1$ 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))

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

where $\rho$ is the coefficient of relative risk aversion, and $c_t$ is the growth rate of aggregate consumption.

Using our monthly data on 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. Then by assuming $\rho = 5$, we obtain our value of 0.028%/month for $\gamma_1$.

The asymptotic standard deviations in Table III 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 separately report them. As for the finite sample standard deviations, they are based on formulae in Propositions 5 and 7. 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. Together with the bias as documented in Table II,
the overstatement of the standard error of the estimated risk premium can lead to wrong acceptance of the null hypothesis $H_0 : \gamma_1 = 0$ even though the true $\gamma_1$ is nonzero. As a result, it is hard to find evidence that a factor is priced when the beta estimation period is short.

Table III also allows us to compare the merits of the estimators from the OLS and the estimated GLS. It is well known that the true GLS CSR estimator is more efficient than the OLS CSR estimator. However, as we need to estimate $\Sigma$, it is not clear that the advantage of the true GLS CSR carries over to the estimated GLS CSR. As a result, many researchers opt to use the simpler OLS CSR. Table III shows that when $N = 10$ and $N = 25$, the estimated GLS continues to dominate 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 we find OLS to be superior is when $T = 120$. This is a case that $T$ is close to $N$, so $\hat{\Sigma}$ is very volatile which leads to added volatility to the estimated GLS estimator. Other than this case, we find that the estimated GLS largely dominates OLS in terms of estimation efficiency. Therefore, unless $T$ is very close to $N$, it is advisable to choose the estimated GLS CSR over the OLS CSR if one is concerned with estimation efficiency.

D. Nonnormal Distributions

While the analytical results in this paper are derived under the assumption of multivariate normality, we have good reasons 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 multivariate normality assumption of the residuals, it is rather robust to departure from normality of the residuals. To examine if our finite sample results on the CSR estimators of zero-beta rate and risk premium are robust to departure from normality assumption, we consider a case that 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 what we often find in the data. As we cannot obtain finite sample distribution of the CSR estimators under the $t$-distribution assumption, we rely on simulation. In order to make easy comparison with our results under normality, we simulate the factor and the returns of the test assets using exactly the
same $\mu$ and $V$ as in the normality case. In Table IV, we present the percentage bias of the OLS and estimated GLS estimator of $\gamma_0$ and $\gamma_1$ under our alternative distribution assumption using exactly the same format as in Table II. The results are based on 100,000 simulations. By comparing the numbers in Table II and Table IV, we can see that the percentage bias of the CSR estimators under the two distribution assumptions are fairly close to each other, with the only exceptions in the GLS case when $N = 100$ and $T$ is small.

In Table V, we report the finite sample standard deviations of the OLS and estimated GLS CSR estimators of $\gamma_0$ and $\gamma_1$ in the 100,000 simulations under the $t$-distribution assumption. By comparing the numbers in Table III and Table V, we again find that the analytical results for the normality case is a very good approximation for the $t$-distribution case, even when $T$ is small. The only noticeable difference again comes from the GLS case when $N = 100$ 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 size of $T$. In view of the simulation evidence here, we consider our analytical finite sample results are good approximations even when the factor and the returns are not multivariate normally distributed.

E. Simulation Results on Bias-adjusted Estimators

We now turn our attention 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 II.B, 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.$^{14}$ Using the same parameters as before, we simulate the bias-adjusted estimators of $\gamma_0$ and $\gamma_1$ under the OLS and the estimated GLS CSR for 100,000 times. In Table VI, we report the percentage biases of the adjusted estimators of $\gamma_0$

$^{14}$Details are available upon request.
and $\gamma_1$. By comparing the numbers in Table II and Table VI, we can see a dramatic reduction of biases for the bias-adjusted estimators as compared with the unadjusted estimators. When $T$ is small, the bias-adjusted estimators do not offer sufficient bias adjustment. This is because the adjustment factor that we use are based on estimated parameters but not the true ones, and when $T$ is small, the estimated parameters are less reliable. It should be noted that there is one case in Panel A ($N = 100$ and $T = 120$) where the bias-adjusted OLS estimator actually offers too big an adjustment, leading to even more bias than the unadjusted estimator. As a whole, while our bias-adjusted estimators do not totally eliminate the bias, we consider they are quite effective in reducing the bias in the unadjusted estimators.

The reduction of bias has its cost, it comes at the expense of increasing the volatility of the estimator. This is the case even if we know the true adjustment factor. For example, the GLS bias-adjusted estimator of $\gamma_1$ is given by $\hat{\gamma}_1^a = \hat{\gamma}_1 / \tilde{\kappa}$, so we have $\text{Var}[\hat{\gamma}_1^a] = \text{Var}[\hat{\gamma}_1] / \tilde{\kappa}^2$. Since $\tilde{\kappa} < 1$, so we must have $\text{Var}[\hat{\gamma}_1^a] > \text{Var}[\hat{\gamma}_1]$. In addition, using the estimated $\tilde{\kappa}$ instead of the true $\kappa$ adds yet another source of variability to $\hat{\gamma}_1^a$, so we can expect our bias-adjusted estimators are more volatile than the unadjusted ones. In Table VII, we report the finite sample standard deviations of the bias-adjusted estimators in the 100,000 simulations. By comparing the numbers in Table III and Table VII, we can see that when $T$ is small and the bias of the unadjusted estimator is large, we can see a big increase in the volatility of the bias-adjusted estimator as compared with the unadjusted one. The trade-off between the bias and volatility of the adjusted and unadjusted estimators depends on the applications, so it is difficult to make a general statement as to which estimator is more desirable. However, in many empirical asset pricing studies, the value of the risk premium is of central importance, so if an estimator of risk premium is heavily biased, it would be misleading to the researchers. In fact, it will be even more misleading if such an estimator is less volatile because it can give the researcher the wrong impression that the risk premium estimate is accurate.
VI. Conclusion

Due to its easy implementation, the two-pass CSR methodology has been the most popular methodology used by empirical researchers for estimating risk premium associated with systematic factors. Despite its simplicity, the finite sample properties of the estimated risk premium from this two-pass approach are complicated due to the errors-in-variables problem associated with the use of estimated betas in the second pass CSR. Traditionally, researchers have either ignored the errors-in-variables 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 premium.

In this paper, we provide an analysis of the finite sample bias and variance of the estimators of zero-beta rate and risk premium from the second pass CSR. Under the normality assumption, we give explicit expressions on 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 premium. For the single factor case, we find that the GLS CSR estimator of risk premium on average underestimates the true risk premium. For reasonable choice 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. For reasonable choice of parameters, 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 departure from the normality assumption.

In order to correct for the finite sample bias of the estimated zero-beta rate and risk premium, we suggest a simple bias adjustment to the second pass CSR estimators of zero-beta rate and risk premium. Using simulations, we find that our adjusted version of the second pass CSR estimators of zero-beta rate and 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. For future research,

\footnote{A set of Matlab programs for calculating the adjusted estimators is available upon request.}
it is of interest to compare the finite sample performance of the CSR estimator of risk premium with other estimators of risk premium (like the ones from the maximum likelihood method and the generalized method of moments). Such an analysis will enhance our understanding of relative merits of various estimators in finite samples. It is also of interest to find out 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 of the CAPM when using our adjusted estimators of zero-beta rate and risk premium.
Appendix

Proof of Lemma 1: Conditional on $\hat{\mu}_1$, $\hat{V}_{11}$, $\hat{\mu}_2$, $\hat{\beta}$ and $U$, from (50), we have

$$E[\hat{\gamma}|\hat{\mu}_1, \hat{V}_{11}, \hat{\mu}_2, \hat{\beta}, U] = \hat{\gamma}. \quad (A1)$$

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

$$E[\hat{\gamma}|\hat{\mu}_1, \hat{V}_{11}] = E[\hat{\gamma}|\hat{\mu}_1, \hat{V}_{11}]. \quad (A2)$$

By taking expectation on both sides with respect to $\hat{\mu}_1$ and $\hat{V}_{11}$, we get

$$E[\hat{\gamma}] = E[\hat{\gamma}]. \quad (A3)$$

Similarly, conditional on $\hat{\mu}_1$, $\hat{V}_{11}$, $\hat{\mu}_2$, $\hat{\beta}$ and $U$, from (50), we have

$$\text{Var}[\hat{\gamma}|\hat{\mu}_1, \hat{V}_{11}, \hat{\beta}, \hat{\mu}_2, U] = \frac{(\tilde{A}_{11} - \tilde{A}_{12}^{-1}\tilde{A}_{22})\tilde{A}_{22}^{-1}}{U}. \quad (A4)$$

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

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

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

$$= \text{Var}[\hat{\gamma}|\hat{\mu}_1, \hat{V}_{11}] + \frac{1}{T - N - 1} E\left[ (\tilde{A}_{11} - \tilde{A}_{12}\tilde{A}_{22}^{-1}\tilde{A}_{21})\tilde{A}_{22}^{-1}|\hat{\mu}_1, \hat{V}_{11}\right], \quad (A5)$$

where the last equality follows from the fact that $U$ is independent of $Y$ and $Z$ and $E[1/U] = 1/(T - N - 1)$. Finally, the unconditional variance of $\hat{\gamma}$ is given by

$$\text{Var}[\hat{\gamma}] = \text{Var}[E[\hat{\gamma}|\hat{\mu}_1, \hat{V}_{11}]] + E[\text{Var}[\hat{\gamma}|\hat{\mu}_1, \hat{V}_{11}]]$$

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

$$= \text{Var}[\hat{\gamma}] + \frac{1}{T - N - 1} E[ (\tilde{A}_{11} - \tilde{A}_{12}\tilde{A}_{22}^{-1}\tilde{A}_{21})\tilde{A}_{22}^{-1}], \quad (A6)$$

where the second equality follows from (A2) and (A5). This completes the proof. \textit{Q.E.D.}

Proof of Proposition 1: We first prove the necessary and sufficient condition for the existence of the conditional $s$-th moment of $\hat{\gamma}_1$. 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 $C$ be a $p$-vector of normal random variables. Suppose $x_i = [C_i, A_i]' \sim N(\mu_i, s_iI_{q+1})$ where $A_i$ is the $i$-th row of $A$, and $x_i$ is independent across $i$. Then, the $s$-th moment of $(A'A)^{-1}A'C$ exists if and only if $s < p - q + 1$.

Writing $A = \Lambda^{\frac{1}{2}}Z_2$, $C = \Lambda^{\frac{1}{2}}Y_2$, conditional on $\hat{\mu}_1$ and $\hat{V}_{11}$, $A$ and $C$ are normally distributed, and $[C_i, A_i]'$ has a variance $\lambda_iI_N$, 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 $\gamma_1 = \hat{V}_{11}^{\frac{1}{2}}(Z_2'\Lambda Z_2)^{-1}(Z_2'\Lambda Y_2)$ exists) if and only if $s < N - K$.

For $\gamma_0$, from (36), we have

$$\gamma_0 = \frac{1}{\sqrt{T}} \left[ (1_N'\Sigma^{-1}1_N)^{-\frac{1}{2}}Y_1 + \xi'Y_2 - \left[(1_N'\Sigma^{-1}1_N)^{-\frac{1}{2}}Z_1 + \xi'Z_2 \right](Z_2'\Lambda Z_2)^{-1}(Z_2'\Lambda Y_2) \right]. \quad (A7)$$

Conditional on $\hat{\mu}_1$ and $\hat{V}_{11}$, the term $(1_N'\Sigma^{-1}1_N)^{-\frac{1}{2}}Y_1 + \xi'Y_2$ 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 independent of $(Z_2'\Lambda Z_2)^{-1}(Z_2'\Lambda Y_2)$, whose finite $s$-th moment exists if and only if $s < N - K$. Finally, the last term $\xi'Z_2(Z_2'\Lambda Z_2)^{-1}(Z_2'\Lambda Y_2)$ has all of its moments exist because $Y_2$ is normally distributed (with all moments existing) and it is independent of $\xi'Z_2(Z_2'\Lambda Z_2)^{-1}Z_2'$, which also has all moments exist. This is because for any $(N-1)$-vector $c$, from the Cauchy-Schwarz inequality, we have

$$(\xi'Z_2(Z_2'\Lambda Z_2)^{-1}Z_2'c)^2 \leq (\xi'Z_2(Z_2'\Lambda Z_2)^{-1}Z_2'\xi)(c'Z_2(Z_2'\Lambda Z_2)^{-1}Z_2'c) < K^2(\xi'\Lambda^{-1}\xi)(c'\Lambda^{-1}c), \quad (A8)$$

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

$$\frac{\xi'Z_2(Z_2'\Lambda Z_2)^{-1}Z_2'\xi}{\xi'\Lambda^{-1}\xi} = \frac{x'[\Lambda^{\frac{1}{2}}Z_2(Z_2'\Lambda Z_2)^{-1}Z_2'\Lambda^{\frac{1}{2}}]x}{x'x} \leq d_{\text{max}} < \text{tr}(\Lambda^{\frac{1}{2}}Z_2(Z_2'\Lambda Z_2)^{-1}Z_2'\Lambda^{\frac{1}{2}}) = K. \quad (A9)$$

Therefore, $|\xi'Z_2(Z_2'\Lambda Z_2)^{-1}Z_2'c|$ is bounded from above and all of its moments exist. Since conditional on $\hat{\mu}_1$ and $\hat{V}_{11}$, all the terms of $\gamma_0$ have all moments exist 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 $\gamma_0$ exists if and only if $s < N - K$.

The proof for $\gamma$ is the same as the proof for $\gamma_0$, except for setting $\Lambda = I_{N-1}$. As for $\gamma$, from the estimated GLS, from (50) and (51), we have

$$\hat{\gamma} = \hat{\gamma} + U^{-\frac{1}{2}}(\hat{A}_{11} - \hat{A}_{12}\hat{A}_{22}^{-1}\hat{A}_{21})^{\frac{1}{2}}\hat{A}_{22}^{-\frac{1}{2}}Z_3, \quad (A10)$$
where $Z_3 \sim N(0_{K+1}, I_{K+1})$ and it is independent of $Y_2, Z_1, Z_2$ and $U$. For the first term, the conditional $s$-th moment of $\tilde{\gamma}$ exists if and only if $s < N - K$. As for the second term, $Z_3$ has all of its moments exist and it 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 $(\tilde{A}_{11} - \tilde{A}_{12} \tilde{A}_{22}^{-1} \tilde{A}_{21})^{\frac{1}{2}}$. As $U \sim \chi^2_{T-N+1}$, we have $U^{-\frac{1}{2}}$ exist if and only if $s < T - N + 1$ (see, for example, Johnson, Kotz, and Balakrishnan (1995, Chapter 27)). As $\tilde{A}_{11} - \tilde{A}_{12} \tilde{A}_{22}^{-1} \tilde{A}_{21} = (Y_2'[I_{N-1} - Z_2(Z_2'Z_2)^{-1}Z_2']Y_2)/T \leq Y_2'Y_2/T$ and $Y_2'Y_2$ has a noncentral chi-squared distribution with all of its moments existing, so all the moments of the term $\tilde{A}_{11} - \tilde{A}_{12} \tilde{A}_{22}^{-1} \tilde{A}_{21}$ also exist. Finally, from (51), we see that the $s$-th moment of $\tilde{A}_{22}^{-\frac{1}{2}}$ exists if and only if the $s$-th moment of $(Z_2'Z_2)^{-\frac{1}{2}}$ exists. From 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, we have shown that the conditional $s$-th moment of $\tilde{\gamma}$ exists if and only if $s < \min[N - K, T - N + 1]$. This completes the proof. $Q.E.D.$

**Proof of Propositions 2:** 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

$$Q = X'AX\over X'BX,$$

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

$$E[Q] = \int_0^\infty \frac{\mu'_X(I_n + 2tB)^{-1}A(I_n + 2tB)^{-1}\mu_X + \text{tr}((I_n + 2tB)^{-1}A)}{|I_n + 2tB|^{\frac{1}{2}}} \times \exp \left(\frac{\mu'_X[(I_n + 2tB)^{-1} - I_n]\mu_X}{2}\right) \, dt$$

$$E[Q^2] = \int_0^\infty \frac{t}{|I_n + 2tB|^{\frac{3}{2}}} \exp \left(\frac{\mu'_X[(I_n + 2tB)^{-1} - I_n]\mu_X}{2}\right) \times$$

$$\left(\left[\mu_X(I_n + 2tB)^{-1}A(I_n + 2tB)^{-1}\mu_X + \text{tr}((I_n + 2tB)^{-1}A)^2 + 2\text{tr}(((I_n + 2tB)^{-1}A)^2) \right] + 4\mu_X(I_n + 2tB)^{-1}A(I_n + 2tB)^{-1}A(I_n + 2tB)^{-1}A\mu_X \right) \, dt.$$ (A13)

Let $X = [Y_2', Z_2']'$ and $A$ and $B$ are defined in (94) and (95), we have $Q = (Z_2'\Lambda Y_2)/(Z_2'\Lambda Z_2)$ and $X \sim N(\mu_X, I_{2(N-1)})$ when conditional on $\hat{\mu}_1$ and $\hat{V}_{11}$, where $\mu_X$ is defined in (93). Substituting $A$, $B$ and $\mu_X$ in (A12) and (A13), we have

$$E^C \begin{bmatrix} \hat{V}_{11}^{-\frac{1}{2}}Z_2'\Lambda Y_2 \\ Z_2'\Lambda Z_2 \end{bmatrix} = \left(\int_0^\infty T\hat{V}_{11}^{-\frac{1}{2}}(\Lambda^{-1} + 2tI_{N-1})^{-1}\eta \exp \left(\frac{T\hat{V}_{11}^{-\frac{1}{2}}[(I_{N-1} + 2t\Lambda)^{-1} - I_{N-1}]\eta}{2}\right) \, dt\right) \tilde{\gamma}_1(A14)$$
\[ E^c \left[ \tilde{V}_{11} \left( \frac{Z_2^2 \Lambda Y_2}{Z_2^2 \Lambda Z_2} \right)^2 \right] \]

\[ = \int_0^\infty \frac{t}{|I_{N-1} + 2t\Lambda|^2} \exp \left( \frac{T\tilde{V}_{11}\eta'[\left(I_{N-1} + 2t\Lambda\right)^{-1} - I_{N-1}]\eta}{2} \right) \]

\[ \times \left[ \frac{\gamma_1^2}{\gamma_1^2} \left( T^2 \tilde{V}_{11}^2 \left[ \eta' \left( \Lambda^{-1} + 2tI_{N-1} \right)^{-1} \right]^2 + T\tilde{V}_{11}\eta' \Lambda \left(I_{N-1} + 2t\Lambda\right)^{-1} \Lambda \eta \right) \right. \]

\[ + \tilde{V}_{11} \left( \text{tr}(\Lambda^2 (I_{N-1} + 2t\Lambda)^{-1}) + T\tilde{V}_{11}\Lambda (I_{N-1} + 2t\Lambda)^{-2} \Lambda \eta \right) \] \text{dt}. \hspace{1cm} (A15)

Using a change of variable \( y = 1/(1 + 2t\Lambda_{N-1}) \) in (A14) and the fact that \( 1/(1 + 2t\lambda_i) = a_iy \), we obtain (99). For (100), the expression follows from (64) and the expression of \( c \) is obtained from (A12) by substituting \( X = Z_2, \mu_X = \sqrt{T}\tilde{V}_{11}\eta, \Lambda = (\xi\eta'\Lambda + \Lambda\eta\xi')/2 \) and \( B = \Lambda \).

\textbf{Proof of Lemma 2:} Repeated use of integration by parts gives us

\[ b \int_0^1 y^{\frac{n}{2}} e^{b(y-1)} \, dy = 1 - \frac{n}{b} + \left( \frac{n}{2} \right) \left( \frac{n}{2} - 1 \right) \frac{b^2}{b^2} - \cdots + (-1)^{\frac{n}{2}} \frac{(\frac{n}{2})!}{b^\frac{n}{2}} (1 - e^{-b}) \]

\[ = \sum_{r=0}^{\frac{n}{2}} \frac{(-1)^{\frac{n}{2}}}{(-b)^r} \frac{(\frac{n}{2} - r + 1)_r}{(\frac{n}{2} - r)_r} \frac{D(\sqrt{b})}{\sqrt{b}} \] \hspace{1cm} (A16)

for even \( n \), and

\[ b \int_0^1 y^{\frac{n}{2}} e^{b(y-1)} \, dy = 1 - \frac{n}{b} + \left( \frac{n}{2} \right) \left( \frac{n}{2} - 1 \right) \frac{b^2}{b^2} - \cdots + (-1)^{\frac{n-1}{2}} \frac{(\frac{n}{2})!}{b^{\frac{n+1}{2}}} \frac{D(\sqrt{b})}{\sqrt{b}} \]

\[ = \sum_{r=0}^{\frac{n-1}{2}} \frac{(-1)^{\frac{n}{2}}}{(-b)^r} \frac{(\frac{n}{2} - r + 1)_r}{(\frac{n}{2} - r)_r} \frac{D(\sqrt{b})}{\sqrt{b}} \] \hspace{1cm} (A17)

for odd \( n \). This completes the proof. \hspace{1cm} Q.E.D.

\textbf{Proof of Lemma 3:} As we must have \( N \geq 3 \) for the first moment of \( \tilde{\gamma}_1 \) to exist, so \( y^{\frac{N-3}{2}} \) is a nonincreasing function of \( N \) for \( 0 \leq y \leq 1 \), and \( \tilde{\kappa} \) is also a nonincreasing function of \( N \) for a fixed \( \theta \). It then follows that the upper bound of \( \tilde{\kappa} \) is

\[ \tilde{\kappa} = \frac{\theta}{2} e^{-\frac{\theta}{2}} \int_0^1 y^{\frac{N-3}{2}} e^{\frac{\theta}{2} y} \, dy \leq \frac{\theta}{2} e^{-\frac{\theta}{2}} \int_0^1 e^{\frac{\theta}{2} y} \, dy = 1 - e^{-\frac{\theta}{2}} < 1. \] \hspace{1cm} (A18)

In order to show that \( \tilde{\kappa} \) is an increasing function of \( \theta \), we use integration by parts to obtain

\[ \frac{\partial \tilde{\kappa}}{\partial \theta} = \frac{1}{2} - \frac{N - 3 + \theta}{4} \int_0^1 y^{\frac{N-3}{2}} e^{\frac{\theta}{2} (y-1)} \, dy. \] \hspace{1cm} (A19)

For \( 0 \leq y \leq 1 \), we have \( f(y) = 1 - ye^{c(1-y)} \geq 0 \) for \( c \geq 0 \). This is 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 \( f(1) = 0 \), we have \( f(y) \geq 0 \) for
$0 \leq y \leq 1$. Putting $c = \frac{N-3}{2}$ in $f(y)$, we have $y^{\frac{N-3}{2}} \leq e^{\left(\frac{N-3}{2}\right)(y-1)}$ for $0 \leq y \leq 1$ and hence

\[
\frac{\partial \kappa}{\partial \theta} \geq \frac{1}{2} \frac{N-3+\theta}{4} \int_0^1 e^{\left(\frac{N-3+\theta}{2}\right)(y-1)} dy = \frac{1}{2} e^{-\left(\frac{N-3+\theta}{2}\right)} > 0. \tag{A20}
\]

Using the inequality $e^{\left(\frac{y-1}{2}\right)} \geq y^c$ for $0 < y < 1$, we obtain a lower bound for $\kappa$ as

\[
\kappa = \frac{\theta}{2} \int_0^1 y^{\frac{N-3}{2}} e^\frac{\theta}{2}(y-1) dy \geq \frac{\theta}{2} \int_0^1 y^{\frac{N-3}{2}} y\frac{\theta}{2} dy = \frac{\theta}{N-1+\theta} > 0. \tag{A21}
\]

As $\theta \to \infty$, we have both the lower bound (A21) and the upper bound (A18) of $\kappa$ approach one. Therefore, $\lim_{\theta \to \infty} \kappa = 1$. This completes the proof. Q.E.D.

**Proof of Lemma 4:** When $\eta_1 = \cdots = \eta_{N-2} = 0$, we can use the fact that $\lambda_{N-1} = 1$ and $a_{N-1} = 1$ to write

\[
\kappa = \frac{T\hat{V}_{11}\eta_2^{N-1}}{2} \int_0^1 \left( \prod_{i=1}^{N-1} a_i \right)^{\frac{1}{N}} e^{\frac{\tau V_{11}\eta_2^{N-1}}{2} y^{\frac{N-3}{2}} dy} \leq \frac{T\hat{V}_{11}\eta_2^{N-1}}{2} \int_0^1 e^{\frac{\tau V_{11}\eta_2^{N-1}}{2} y^{\frac{N-3}{2}} dy} = \kappa, \tag{A22}
\]

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

When $\eta_2 = \cdots = \eta_{N-1} = 0$, we use a change of variable $y = 1/(1+2t\lambda_1)$ in (A14) instead of $y = 1/(1+2\theta\lambda_{N-1})$ to get

\[
\kappa = \frac{T\hat{V}_{11}2}{} \int_0^1 \left( \sum_{i=1}^{N-1} b_i \frac{\lambda_i}{\lambda_1} \right) \left( \prod_{i=1}^{N-1} b_i \right) \frac{1}{2} e^{\frac{\tau V_{11}\eta_2^{N-1}}{2} y^{\frac{N-3}{2}} dy}, \tag{A23}
\]

where $b_i = \lambda_1 / [\lambda_i - (\lambda_i - \lambda_1)]$. Putting $\eta_2 = \cdots = \eta_{N-1} = 0$ in (A23), we have

\[
\kappa = \frac{T\hat{V}_{11}\eta_2^2}{2} \int_0^1 \left( \prod_{i=1}^{N-1} b_i \right) \frac{1}{2} e^{\frac{\tau V_{11}\eta_2^2}{2} y^{\frac{N-3}{2}} dy} \geq \frac{T\hat{V}_{11}\eta_2^2}{2} \int_0^1 e^{\frac{\tau V_{11}\eta_2^2}{2} y^{\frac{N-3}{2}} dy} = \kappa, \tag{A24}
\]

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 $\lambda_1 = \lambda_{N-1}$. This completes the proof. Q.E.D.

**Proof of Proposition 4:** Replacing $T\hat{V}_{11}$ by $V_{11}v$ in the integral for the conditional mean, where $v \sim \chi_{2N-1}^2$. Using the fact that the density function of $v$ is

\[
f(v) = \frac{1}{\Gamma\left(\frac{2N-2}{2}\right)} \left( \frac{1}{2} \right)^{\frac{2N-2}{2}} v^{\frac{2N-2}{2} - 1} e^{-v}, \tag{A25}
\]

47
we can integrate the conditional mean using the density function of \( v \). After some simplification and using the identity
\[
\int_0^{\infty} v^a e^{-av} dv = \frac{\Gamma(n + 1)}{a^{n+1}}
\]
for \( a > 0 \), we obtain the unconditional mean. This completes the proof. \( Q.E.D. \)

**Proof of Proposition 5:** With the expression of \( \hat{\gamma}_1 \) in (35) and using a change of variable of \( y = 1/(1 + 2t\lambda_{N-1}) \) in (A15), we have
\[
E^c[\hat{\gamma}_1^2] = E\left[ \hat{V}_{11} \left( \frac{Z'_2aY_2}{Z'_2\Lambda Z_2} \right)^2 \right] = \frac{\hat{V}_{11}}{4} \int_0^1 \left[ T\hat{\gamma}_1^2(T\hat{V}_{11}g_2^2y + g_6) + (T\hat{V}_{11}g_4y + g_5) \right] \left( \prod_{i=1}^{N-1} a_i \right)^{1/2} \times e^{\frac{T\hat{V}_{11}}{2} \sum_{i=1}^{N-1} \eta^2_i(a_iy - 1) y^2} (1 - y) dy.
\]

(A27)

Since \( \hat{\gamma}_1 \) is independent of \( \hat{V}_{11} \) and \( E[\hat{\gamma}_1^2] = \hat{\gamma}_1^2 + \frac{\hat{V}_{11}}{4} = \frac{\gamma}{q} \), we have
\[
E[\hat{\gamma}_1^2|\hat{V}_{11}] = \frac{\hat{V}_{11}}{4} \int_0^1 \left[ T\hat{V}_{11}(g_4 + ag_1^2)y + (g_5 + ag_6) \right] \left( \prod_{i=1}^{N-1} a_i \right)^{1/2} \times e^{\frac{T\hat{V}_{11}}{2} \sum_{i=1}^{N-1} \eta^2_i(a_iy - 1) y^2} (1 - y) dy.
\]

(A28)

Then, by integrating the above expression with respect to the density function of \( \hat{V}_{11} \) as in the proof of Proposition 4 and using the fact that \( E[\hat{\gamma}_1] = \hat{\kappa}_u \gamma_1 \), we obtain (126).

For the proof of \( \hat{\gamma}_0 \), we first define \( D_4 = (Z'_2aY_2)/(Z'_2\Lambda Z_2) \) and \( D_5 = (Z'_2a\xi)/(Z'_2\Lambda Z_2) \). We then define the constants \( Q_1 \) to \( Q_5 \) as \( Q_1 = E[\sqrt{T}\hat{V}_{11}^2 D_5], Q_2 = E[D_5(\xi'Z_2)], Q_3 = E[D_4^2], Q_4 = E[D_4^2(\xi'Z_2)] \) and \( Q_5 = E[D_4^2(\xi'Z_2)^2] \). We now show that \( Q_1 \) to \( Q_5 \) are given by the expressions (128)–(132). The proof of \( Q_1 \) is similar to the proof of \( E[\hat{\gamma}_1] \), the proof of \( Q_2 \) is similar to the proof of \( c_u \) in Proposition 4, and the proof of \( Q_3 \) is similar to the proof of \( E[\hat{\gamma}_1^2] \), so we do not repeat them here. It remains to prove the expressions for \( Q_4 \) and \( Q_5 \). Theorem 5 of Magnus (1990) provides the expression for \( E[(X'AX)^2(a'X)/(X'BX)^2] \) and \( E[(X'AX)^2(X'CX)/(X'BX)^2] \), where \( a \) is a vector, \( B \) is a positive semidefinite matrix, and \( A \) and \( C \) are symmetric matrices. With the appropriate choice of \( A \) and \( B \) as in Proposition 2, we can show that
\[
E^c \left[ \sqrt{T}\hat{V}_{11}^2 \left( \frac{Z'_2aY_2}{Z'_2\Lambda Z_2} \right)^2 (\xi'Z_2) \right]
\]
48
\[ \int_0^\infty \frac{t}{|I_{N-1} + 2t\Lambda|^{\frac{3}{2}}} \exp\left(\frac{T\hat{V}_{11}y'(I_{N-1} + 2t\Lambda)^{-1} - I_{N-1}|\eta}{2}\right) \]
\[ \times \left\{ \left( T\hat{V}_{11}\hat{\gamma}_1\eta'\Lambda(I_{N-1} + 2t\Lambda)^{-1}|\eta \right)^2 + \text{tr} \left( (I_{N-1} + 2t\Lambda)^{-1}\Lambda^2 \right) + T\hat{\gamma}_1^2\eta'\Lambda^2(I_{N-1} + 2t\Lambda)^{-1}|\eta \right. \\
\[ + T\hat{V}_{11}\eta'\Lambda^2(I_{N-1} + 2t\Lambda)^{-2}|\eta \right) T\hat{V}_{11}\xi'(I_{N-1} + 2t\Lambda)^{-1}|\eta + 2T\hat{V}_{11}\eta'\Lambda^2(I_{N-1} + 2t\Lambda)^{-2}|\eta \]
\[ + 2T\hat{\gamma}_1^2\left( T\hat{V}_{11}\eta'\Lambda(I_{N-1} + 2t\Lambda)^{-1}|\eta \right) (\eta'\Lambda(I_{N-1} + 2t\Lambda)^{-1}|\eta) \right\} dt. \quad (A29) \]
\[ E^c \left[ \left( \begin{array}{c} Z_2\Lambda Y_2 \\ Z_2\Lambda Z_2 \end{array} \right)^2 \right] \]
\[ = \int_0^\infty \frac{t}{|I_{N-1} + 2t\Lambda|^{\frac{3}{2}}} \exp\left(\frac{T\hat{V}_{11}y'(I_{N-1} + 2t\Lambda)^{-1} - I_{N-1}|\eta}{2}\right) \]
\[ \times \left\{ \left( T\hat{V}_{11}\hat{\gamma}_1\eta'\Lambda(I_{N-1} + 2t\Lambda)^{-1}|\eta \right)^2 + \text{tr} \left( (I_{N-1} + 2t\Lambda)^{-1}\Lambda^2 \right) + T\hat{\gamma}_1^2\eta'\Lambda^2(I_{N-1} + 2t\Lambda)^{-1}|\eta \right. \\
\[ + T\hat{V}_{11}\eta'\Lambda^2(I_{N-1} + 2t\Lambda)^{-2}|\eta \right) T\hat{V}_{11}\xi'(I_{N-1} + 2t\Lambda)^{-1}|\eta + 2T\hat{V}_{11}\eta'\Lambda^2(I_{N-1} + 2t\Lambda)^{-2}|\eta \]
\[ + 2T\hat{\gamma}_1^2\left( T\hat{V}_{11}\eta'\Lambda(I_{N-1} + 2t\Lambda)^{-1}|\eta \right) (\eta'\Lambda(I_{N-1} + 2t\Lambda)^{-1}|\eta) \right\} dt. \quad (A30) \]

With a change of variable \( y = 1/(1 + 2t\lambda_{N-1}) \) and taking unconditional expectation, we obtain the expressions for \( Q_4 \) and \( Q_5 \).

From the first expression of (76), we have

\[ \text{Var}^c[\hat{\gamma}_0] = \frac{\xi'\xi - 2(hE^c[\sqrt{T}\hat{V}_{11}D_5] + E^c[D_5(\xi'Z_2)])}{T} + \frac{1 + E^c[D_4^2]}{T(I_N^2\Sigma^{-1}1_N)}. \quad (A31) \]

Using the formula for unconditional variance, we have

\[ \text{Var}[\hat{\gamma}_0] = E[\text{Var}^c[\hat{\gamma}_0]] + \text{Var}[E^c[\hat{\gamma}_0]] \\
= \frac{1 + Q_5}{T(I_N^2\Sigma^{-1}1_N)} + \frac{\xi'\xi - 2(hQ_1 + Q_2) + E^c[h\sqrt{T}\hat{V}_{11}D_4 + D_4(\xi'Z_2)]}{T} \]
\[ + \text{Var}[h + \xi'\eta - (h\bar{\kappa} + c)\hat{\gamma}_1], \quad (A32) \]

where \( \bar{\kappa} \) and \( c \) are defined in Proposition 2. Using the following simplifications

\[ E[\text{Var}^c[h\sqrt{T}\hat{V}_{11}D_4 + D_4(\xi'Z_2)]] \]
We can arrive at the expression in (127). This completes the proof. \(Q.E.D.\)

**Proof of Proposition 6:** By setting \(\lambda_i^* = 1\) in the expression of \(\text{Var}[\tilde{\gamma}_1]\), we get

\[
\text{Var}[\tilde{\gamma}_1] = \frac{(T - 1)V_{11}}{4T} \left[ (T + 1)V_{11} \phi_{N-3,T+3}(\eta'\eta + a(\eta'\eta)^2) + \varphi_{N-5,T+1}(N - 1 + a\eta'\eta) \right] - \tilde{\kappa}_u^2 \tilde{\gamma}_1^2. \tag{A35}
\]

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

\[
\text{Var}[\tilde{\gamma}_1] = \frac{(T - 1)V_{11}}{4T} \left[ (T + 1)V_{11} \eta'N(1 + a\eta'\eta)(\phi_{N-3,T+3} - \phi_{N-1,T+3}) \right. \\
+ (N - 1 + a\eta'\eta)(\phi_{N-5} - \phi_{N-3}) \left. - \tilde{\kappa}_u^2 \tilde{\gamma}_1^2 \right]. \tag{A36}
\]

Using integration by parts, we have

\[
\phi_{N-3,T+3} = \frac{2 - (N - 3)\phi_{N-5}}{(T + 1)V_{11}\eta'}, \tag{A37}
\]

\[
\phi_{N-1,T+3} = \frac{2 - (N - 1)\phi_{N-3}}{(T + 1)V_{11}\eta'}. \tag{A38}
\]

Substituting these two expressions in (A36), we get (133). Similarly, by setting \(\lambda_i^* = 1\) and \(\xi_i = 0\) in the expression of \(\text{Var}[\tilde{\gamma}_0]\), we get

\[
\text{Var}[\tilde{\gamma}_0] = h^2 E[\tilde{\gamma}_1^2] + \frac{V_{11}}{T} h^2 (1 - \tilde{\kappa}_u)^2 - \frac{h^2 \tilde{\kappa}_u^2}{T} + \frac{1 + \hat{Q}_3}{T(1'N\Sigma^{-1}1_N)}
\]

\[
= h^2 \text{Var}[\tilde{\gamma}_1] + \frac{V_{11}}{T} h^2 (1 - 2\tilde{\kappa}_u) + \frac{1 + \hat{Q}_3}{T(1'N\Sigma^{-1}1_N)}, \tag{A39}
\]

where

\[
\hat{Q}_3 = \frac{(T - 1)V_{11}}{4} \varphi_{N-3,T+1}(\eta'\eta + a(\eta'\eta)^2) + \frac{1}{4} \varphi_{N-5,T-1}(N - 1 + a\eta'\eta)
\]

\[
= \frac{(T - 1)V_{11}\eta'(1 + a\eta'\eta)}{4} (\phi_{N-3} - \phi_{N-1}) + \frac{(N - 1 + a\eta'\eta)}{4} (\phi_{N-5,T-1} - \phi_{N-3,T-1}). \tag{A40}
\]
Using integration by parts, we have
\[
\phi_{N-5,T-1} = \frac{2 - (T - 1)V_{11}\eta'\eta\phi_{N-3}}{N - 3}, \quad (A41)
\]
\[
\phi_{N-3,T-1} = \frac{2 - (T - 1)V_{11}\eta'\eta\phi_{N-1}}{N - 1}. \quad (A42)
\]
Substituting these two expressions in (A40), we get the expression of \( \tilde{Q}_3 \) in Proposition 6. This completes the proof. \( Q.E.D. \)

**Proof of Proposition 7:** From (88) and (91), for \( K = 1 \), we have
\[
\Delta = \frac{E[((N - 2) + T\gamma_1^2\tilde{C})\tilde{V}_{11}\tilde{D}_2]}{T - N - 1} - \frac{(N - 2)E[\tilde{V}_{11}\tilde{C}\tilde{D}_2]}{T_N - 1} = \frac{1}{T - N - 1} \left\{ [(N - 2) + a\eta'\eta]E \left[ \frac{\tilde{V}_{11}}{Z_2'Z_2} \right] - aE \left[ \tilde{V}_{11} \left( \frac{Z_2'\eta}{Z_2'Z_2} \right)^2 \right] \right\}. \quad (A43)
\]
Using a similar method as in the proof of Proposition 2, we can show that
\[
E \left[ \frac{\tilde{V}_{11}}{Z_2'Z_2} \right] = \frac{(T - 1)V_{11}}{2T} \phi_{N-5}, \quad (A44)
\]
\[
E \left[ \tilde{V}_{11} \left( \frac{Z_2'\eta}{Z_2'Z_2} \right)^2 \right] = \frac{(T - 1)V_{11}\eta'\eta}{4T} [(N - 2)\phi_{N-3} - (N - 4)\phi_{N-5}]. \quad (A45)
\]
Substituting these two expressions in (A43), we obtain the expression of \( \Delta \) in Proposition 7.

For \( \text{Var}[\gamma_0] \), we set \( K = 1 \) in (92) and get
\[
\text{Var}[\gamma_0] = \text{Var}[\gamma_0] + h^2\Delta + \frac{E[((N - 2) + T\gamma_1^2\tilde{C} + V_{11}\tilde{C})\phi_{N-3}]}{T(T - N - 1)(V_N'\Sigma^{-1}1_N)} = \text{Var}[\gamma_0] + h^2\Delta + \frac{(N - 2)E[\tilde{D}_2] + a(E[\tilde{C}] + E[\tilde{C}\tilde{D}_2])}{T(T - N - 1)(V_N'\Sigma^{-1}1_N)}. \quad (A46)
\]
Using a similar method as in the proof of Proposition 2, we can show that
\[
E[\tilde{D}_2] = E \left[ \frac{1}{Z_2'Z_2} \right] = \frac{1}{N - 3} - \frac{(T - 1)V_{11}\eta'\eta}{2(N - 3)} \phi_{N-3}, \quad (A47)
\]
\[
E[\tilde{C}] = \eta'\eta - E \left[ \frac{Z_2'\eta Z_2'Z_2}{Z_2'Z_2} \right] = \frac{(N - 2)\eta'\eta}{N - 1} \left[ 1 - \frac{(T - 1)V_{11}\eta'\eta}{2} \phi_{N-1} \right], \quad (A48)
\]
\[
E[\tilde{C}\tilde{D}_2] = \eta'\eta E[\tilde{D}_2] - E \left[ \left( \frac{Z_2'\eta}{Z_2'Z_2} \right)^2 \right] = (N - 2)\eta'\eta \left[ \frac{1}{(N - 1)(N - 3)} + \frac{(T - 1)V_{11}\eta'\eta}{4} \left( \frac{\phi_{N-1}}{N - 1} - \frac{\phi_{N-3}}{N - 3} \right) \right]. \quad (A49)
\]
Substituting these expressions in (A46) and after some simplifications, we obtain the expression of $\text{Var}[\hat{\gamma}_0]$ in Proposition 7. This completes the proof.  \[Q.E.D.\]

Proof of Lemma 5: Let
\begin{align*}
\hat{A} &= [1_N, \hat{\beta}'\hat{\Sigma}^{-1}[1_N, \hat{\beta}], \quad (A50) \\
\bar{A} &= [1_N, \hat{\beta}'\Sigma^{-1}[1_N, \hat{\beta}], \quad (A51)
\end{align*}
From Theorem 3.2.11 of Muirhead (1982), conditional on $\hat{\beta}$, we have
\[\hat{A}^{-1} \sim W_2(T - N, \bar{A}^{-1}/T). \quad (A52)\]
Note that the $(2, 2)$ element of $\hat{A}^{-1}$ and $\bar{A}^{-1}$ are given by $1/\hat{\eta}'\hat{\eta}$ and $1/\bar{\eta}'\bar{\eta}$, respectively, where
\[\bar{\eta} = P'\Sigma^{-1/2}\hat{\beta} \sim N(\eta, I_{N-1}). \quad (A53)\]
Therefore, conditional on $\hat{\beta}$, from (A52), we have
\[\frac{1}{\bar{\eta}'\bar{\eta}} \sim W_1\left(T - N, \frac{1}{T\bar{\eta}'\bar{\eta}}\right), \quad (A54)\]
and hence
\[V = \frac{T\bar{\eta}'\hat{\eta}}{\bar{\eta}'\bar{\eta}} \sim \chi^2_{T - N}. \quad (A55)\]
Since the distribution of $V$ is independent of $\hat{\beta}$ and hence independent of $\bar{\eta}'\bar{\eta}$, we have
\[\hat{V}_{11}\bar{\eta}'\hat{\eta} \sim \frac{TV_{11}\bar{\eta}'\hat{\eta}}{V}. \quad (A56)\]
Finally, conditional on $\hat{V}_{11}$, we have $T\hat{V}_{11}\bar{\eta}'\hat{\eta} \sim \chi^2_{N-1}(TV_{11}\bar{\eta}'\hat{\eta})$. Using the definition of the noncentral $F$-distribution, we then obtain the distribution of $\hat{V}_{11}\bar{\eta}'\hat{\eta}$. The expectation of $\hat{V}_{11}\bar{\eta}'\hat{\eta}$ 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 6: The Bayes estimator with respect to the loss function (147) is the estimator that minimizes the posterior risk
\[\int_0^\infty \left[ \frac{\hat{\theta}}{\theta} - \log\left( \frac{\theta}{\hat{\theta}} \right) \right] f(\theta|z)\,d\theta, \quad (A57)\]
where \( f(\theta | z) \) is the posterior density of \( \theta \) and \( z = \hat{V}_1 \hat{\eta}' \hat{\eta} \). Taking derivative of (A57) with respect to \( \hat{\theta} \), we have

\[
\hat{\theta} = \frac{1}{\int_0^\infty \theta^{-1} f(\theta | z) d\theta}. \tag{A58}
\]

As the posterior of \( \theta \) is given by

\[
f(\theta | z) = \frac{f(z | \theta) \pi(\theta)}{\int_0^\infty f(z | \theta) \pi(\theta) d\theta} = \frac{\theta^b f(z | \theta)}{\int_0^\infty \theta^b f(z | \theta) d\theta}, \tag{A59}
\]

substituting (A59) in (A58), we obtain

\[
\hat{\theta} = \frac{\int_0^\infty \theta^b f(\theta | z) d\theta}{\int_0^\infty \theta^b f(\theta | z) d\theta}. \tag{A60}
\]

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

\[
f(z | \theta) = \frac{e^{-\frac{\theta}{2} z} \frac{N-3}{2}}{B \left( \frac{N-1}{2}, \frac{T-N}{2} \right) (1 + z)^\frac{1}{2}} \sum_{r=0}^\infty \frac{\left( \frac{z}{2(1+z)} \right)^r}{\left( \frac{N-1}{2} \right)_r} \frac{(\frac{T-1}{2})_r}{r!}, \tag{A61}
\]

where \( B(a, b) \) is the beta function. Using (A26), we obtain the following identity

\[
\int_0^\infty \theta^d f(z | \theta) d\theta = \frac{z \frac{N-3}{2} 2^{d+1}}{B \left( \frac{N-1}{2}, \frac{T-N}{2} \right) (1 + z)^\frac{1}{2}} \sum_{r=0}^\infty \frac{\Gamma(d + r + 1)}{r!} \left( \frac{T-1}{2} \right)_r \frac{(\frac{z}{1+z})^r}{\left( \frac{N-1}{2} \right)_r} = \frac{z \frac{N-3}{2} 2^{d+1} \Gamma(d + 1)}{B \left( \frac{N-1}{2}, \frac{T-N}{2} \right) (1 + z)^\frac{1}{2}} _2 F_1 \left( d + 1, \frac{T-1}{2}, \frac{N-1}{2}, \frac{z}{1+z} \right). \tag{A62}
\]

Using this identity in both the numerator and denominator of (A60), we obtain our Bayes estimator in (148). This completes the proof.

\( Q.E.D. \)
References


MacKinlay, A. Craig, 1985, Analysis of multivariate financial tests, Ph.D. dissertation, Graduate School of Business, University of Chicago.


The figure plots $\tilde{\kappa} = E[\tilde{\gamma}_1|\hat{V}_{11}] / \gamma_1$ as a function of $TV_{11}{\eta}'\eta/(N-1)$ for different values of $N$, where $\tilde{\gamma}_1$ is the second pass GLS CSR estimate of risk premium, $N$ is the number of test assets, $T$ is the number of time series observations used for estimation of $\beta$, $V_{11}$ is the realized variance of the factor, and $\eta'{\eta}/(N-1)$ is a measure of dispersion of $\beta$ across the test assets.
Figure 2

Representation of Two Estimators of $T\hat{V}_{11}'\eta$ for Different Values of $\hat{V}_{11}'\eta$

The figure plots two estimators of $\hat{\theta} = T\hat{V}_{11}'\eta$ as a function of $z = \hat{V}_{11}'\eta$ when $N = 10$ and $T = 100$. The dotted line is for the estimator $\hat{\theta}_u = (T - N - 2)\hat{V}_{11}'\eta - (N - 1)$, which is an unbiased estimator of $\theta$. The solid line is for the estimator $\hat{\theta}$, which is defined as $\hat{\theta} = 2b_2F_1(1+b, (T - 1)/2, (N - 1)/2, (N - 1)/2, z/(1 + z))$, where $2F_1$ is the hypergeometric function and $b$ is set to 0.5. $\hat{\theta}$ is the Bayes estimator of $\theta$ under Stein’s loss function and a prior of $\pi(\theta) = \theta^b$. 

58
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 a high explanatory power on the returns of the test assets whereas the factor in Panel B is chosen to have a low explanatory power on the returns of the test assets. For each panel, the table presents the standard deviation (in percentage) of the factor \( V_{i1} \) and three sets of parameters corresponding to three different cases of number of test assets \( N \). For each case, the table presents the cross-sectional standard deviation of beta \( \sigma_\beta \), the average standard deviation (in percentage) of the regression residuals for the test assets \( \bar{\sigma}_\epsilon \), GLS cross-sectional variance of normalized beta \( V_{11} \eta' \eta / (N - 1) \). In addition, the table also presents the three largest and three smallest normalized eigenvalues \( (\lambda_i / \bar{\lambda}) \) of the matrix \( \Sigma_{12} \Sigma_{12}' \) and the corresponding absolute values of \( \eta_i = p_i' \Sigma^{-1/2} \beta \), where \( p_i \) is the eigenvector associated with \( \lambda_i \).

### Panel A: Factor with high explanatory power \( (V_{i1} = 4.092) \)

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \sigma_\beta )</td>
<td>0.049</td>
<td>0.109</td>
<td>0.245</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \bar{\sigma}_\epsilon )</td>
<td>2.362</td>
<td>2.934</td>
<td>3.816</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( 100V_{11}\eta' \eta / (N - 1) )</td>
<td>0.520</td>
<td>1.728</td>
<td>2.071</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( i )</td>
<td>( \lambda_i / \bar{\lambda} )</td>
<td>(</td>
<td>\eta_i</td>
<td>)</td>
<td>( \lambda_i / \bar{\lambda} )</td>
</tr>
<tr>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
</tr>
<tr>
<td>1</td>
<td>6.551</td>
<td>2.759</td>
<td>9.749</td>
<td>5.530</td>
<td>28.128</td>
</tr>
<tr>
<td>2</td>
<td>1.000</td>
<td>2.418</td>
<td>4.712</td>
<td>2.250</td>
<td>6.159</td>
</tr>
<tr>
<td>3</td>
<td>0.431</td>
<td>3.013</td>
<td>1.686</td>
<td>0.613</td>
<td>3.799</td>
</tr>
<tr>
<td>( N - 3 )</td>
<td>0.154</td>
<td>0.206</td>
<td>0.177</td>
<td>1.411</td>
<td>0.135</td>
</tr>
<tr>
<td>( N - 2 )</td>
<td>0.130</td>
<td>0.202</td>
<td>0.158</td>
<td>2.175</td>
<td>0.125</td>
</tr>
<tr>
<td>( N - 1 )</td>
<td>0.112</td>
<td>0.137</td>
<td>0.157</td>
<td>2.841</td>
<td>0.122</td>
</tr>
</tbody>
</table>

### Panel B: Factor with low explanatory power \( (V_{i1} = 0.747) \)

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \sigma_\beta )</td>
<td>0.255</td>
<td>0.336</td>
<td>0.481</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \bar{\sigma}_\epsilon )</td>
<td>5.088</td>
<td>5.428</td>
<td>6.079</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( 100V_{11}\eta' \eta / (N - 1) )</td>
<td>0.227</td>
<td>0.309</td>
<td>0.351</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( i )</td>
<td>( \lambda_i / \bar{\lambda} )</td>
<td>(</td>
<td>\eta_i</td>
<td>)</td>
<td>( \lambda_i / \bar{\lambda} )</td>
</tr>
<tr>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
</tr>
<tr>
<td>1</td>
<td>6.356</td>
<td>16.606</td>
<td>11.554</td>
<td>17.630</td>
<td>25.705</td>
</tr>
<tr>
<td>2</td>
<td>1.012</td>
<td>2.900</td>
<td>4.236</td>
<td>4.903</td>
<td>10.583</td>
</tr>
<tr>
<td>3</td>
<td>0.488</td>
<td>6.990</td>
<td>1.449</td>
<td>10.277</td>
<td>4.369</td>
</tr>
<tr>
<td>( N - 3 )</td>
<td>0.165</td>
<td>2.330</td>
<td>0.152</td>
<td>9.710</td>
<td>0.117</td>
</tr>
<tr>
<td>( N - 2 )</td>
<td>0.140</td>
<td>0.868</td>
<td>0.136</td>
<td>10.246</td>
<td>0.110</td>
</tr>
<tr>
<td>( N - 1 )</td>
<td>0.127</td>
<td>4.552</td>
<td>0.116</td>
<td>1.875</td>
<td>0.098</td>
</tr>
</tbody>
</table>
Table II
Unconditional Percentage Biases of Zero-Beta Rate and Risk Premium Estimators from the Second Pass Cross-Sectional Regressions

The table presents the unconditional biases of OLS and GLS CSR estimators of zero-beta rate ($\gamma_0$) and risk premium ($\gamma_1$) as a percentage of the true value of the risk premium for different length of beta estimation period ($T$) and number of test assets ($N$) when the factors and returns are multivariate normally distributed. The factor in Panel A is chosen to have a high explanatory power on the returns of the test assets whereas the factor in Panel B is chosen to have a low explanatory power on the returns of the test assets. The true GLS and the estimated GLS CSR estimators of $\gamma_0$ and $\gamma_1$ have the same bias, except that the estimated GLS is infeasible for $T = 60$ and $N = 100$.

### Panel A: Factor with high explanatory power

<table>
<thead>
<tr>
<th>$N$ = 10</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$T$</td>
<td>$\hat{\gamma}_0$</td>
<td>$\hat{\gamma}_1$</td>
<td>$\hat{\gamma}_0$</td>
<td>$\hat{\gamma}_1$</td>
<td>$\hat{\gamma}_0$</td>
<td>$\hat{\gamma}_1$</td>
</tr>
<tr>
<td>60</td>
<td>59.9</td>
<td>-56.8</td>
<td>75.3</td>
<td>-75.8</td>
<td>32.8</td>
<td>-29.9</td>
</tr>
<tr>
<td>120</td>
<td>35.4</td>
<td>-33.2</td>
<td>59.5</td>
<td>-59.9</td>
<td>16.5</td>
<td>-14.8</td>
</tr>
<tr>
<td>240</td>
<td>13.6</td>
<td>-12.3</td>
<td>41.2</td>
<td>-41.5</td>
<td>7.7</td>
<td>-6.7</td>
</tr>
<tr>
<td>360</td>
<td>5.6</td>
<td>-4.7</td>
<td>31.2</td>
<td>-31.4</td>
<td>4.9</td>
<td>-4.3</td>
</tr>
<tr>
<td>480</td>
<td>2.2</td>
<td>-1.6</td>
<td>25.1</td>
<td>-25.2</td>
<td>3.5</td>
<td>-3.1</td>
</tr>
<tr>
<td>600</td>
<td>0.7</td>
<td>-0.2</td>
<td>20.9</td>
<td>-21.0</td>
<td>2.8</td>
<td>-2.4</td>
</tr>
</tbody>
</table>

### Panel B: Factor with low explanatory power

<table>
<thead>
<tr>
<th>$N$ = 10</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$T$</td>
<td>$\hat{\gamma}_0$</td>
<td>$\hat{\gamma}_1$</td>
<td>$\hat{\gamma}_0$</td>
<td>$\hat{\gamma}_1$</td>
<td>$\hat{\gamma}_0$</td>
<td>$\hat{\gamma}_1$</td>
</tr>
<tr>
<td>60</td>
<td>60.8</td>
<td>-61.3</td>
<td>60.9</td>
<td>-88.0</td>
<td>70.1</td>
<td>-65.1</td>
</tr>
<tr>
<td>120</td>
<td>38.6</td>
<td>-36.4</td>
<td>54.0</td>
<td>-78.0</td>
<td>50.0</td>
<td>-43.7</td>
</tr>
<tr>
<td>240</td>
<td>15.1</td>
<td>-11.1</td>
<td>43.7</td>
<td>-63.1</td>
<td>28.4</td>
<td>-22.0</td>
</tr>
<tr>
<td>360</td>
<td>5.4</td>
<td>-1.2</td>
<td>36.4</td>
<td>-52.6</td>
<td>18.2</td>
<td>-12.7</td>
</tr>
<tr>
<td>480</td>
<td>1.3</td>
<td>2.4</td>
<td>31.1</td>
<td>-44.9</td>
<td>12.8</td>
<td>-8.1</td>
</tr>
<tr>
<td>600</td>
<td>-0.4</td>
<td>3.6</td>
<td>27.0</td>
<td>-39.0</td>
<td>9.7</td>
<td>-5.7</td>
</tr>
</tbody>
</table>

60
Table III  
Asymptotic and Finite Sample Standard Deviation of Zero-Beta Rate and Risk Premium Estimators from the Second Pass Cross-Sectional Regressions

The table presents the unconditional standard deviations (in percentage per month) of OLS and estimated GLS CSR estimators of zero-beta rate ($\gamma_0$) and risk premium ($\gamma_1$) for different length of sample period ($T$) and number of test assets ($N$) when the factors and returns are multivariate normally distributed. The factor in Panel A is chosen to have a high explanatory power on the returns of the test assets whereas the factor in Panel B is chosen to have a low explanatory power on the returns of the test assets. The value of $\gamma_1$ 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 one is based on the asymptotic variance formula from Shanken (1992), and the second one is based on the exact finite sample standard deviation of the estimators.

### Panel A: Factor with high explanatory power

<table>
<thead>
<tr>
<th>$N = 10$</th>
<th>$N = 25$</th>
<th>$N = 100$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>OLS</strong></td>
<td><strong>GLS</strong></td>
<td><strong>OLS</strong></td>
</tr>
<tr>
<td>$T$</td>
<td>$\hat{\gamma}_0$</td>
<td>$\hat{\gamma}_1$</td>
</tr>
<tr>
<td>60</td>
<td>3.956</td>
<td>3.952</td>
</tr>
<tr>
<td>120</td>
<td>2.797</td>
<td>2.794</td>
</tr>
<tr>
<td>240</td>
<td>1.978</td>
<td>1.976</td>
</tr>
<tr>
<td>360</td>
<td>1.615</td>
<td>1.613</td>
</tr>
<tr>
<td>480</td>
<td>1.398</td>
<td>1.397</td>
</tr>
<tr>
<td>600</td>
<td>1.251</td>
<td>1.250</td>
</tr>
</tbody>
</table>

### Finite Sample Standard Deviation

<table>
<thead>
<tr>
<th>$N = 10$</th>
<th>$N = 25$</th>
<th>$N = 100$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>OLS</strong></td>
<td><strong>GLS</strong></td>
<td><strong>OLS</strong></td>
</tr>
<tr>
<td>$T$</td>
<td>$\hat{\gamma}_0$</td>
<td>$\hat{\gamma}_1$</td>
</tr>
<tr>
<td>60</td>
<td>2.702</td>
<td>2.696</td>
</tr>
<tr>
<td>120</td>
<td>2.307</td>
<td>2.319</td>
</tr>
<tr>
<td>240</td>
<td>1.831</td>
<td>1.846</td>
</tr>
<tr>
<td>360</td>
<td>1.556</td>
<td>1.567</td>
</tr>
<tr>
<td>480</td>
<td>1.373</td>
<td>1.382</td>
</tr>
<tr>
<td>600</td>
<td>1.241</td>
<td>1.248</td>
</tr>
</tbody>
</table>

61
Table III
Asymptotic and Finite Sample Standard Deviation of Zero-Beta Rate and Risk Premium Estimators from the Second Pass Cross-Sectional Regressions (Continued)

Panel B: Factor with low explanatory power

<table>
<thead>
<tr>
<th>N</th>
<th>(\hat{\gamma}_0)</th>
<th>(\hat{\gamma}_1)</th>
<th>(\hat{\gamma}_0)</th>
<th>(\hat{\gamma}_1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>60</td>
<td>0.817</td>
<td>0.768</td>
<td>0.705</td>
<td>0.682</td>
</tr>
<tr>
<td>120</td>
<td>0.578</td>
<td>0.543</td>
<td>0.499</td>
<td>0.482</td>
</tr>
<tr>
<td>240</td>
<td>0.409</td>
<td>0.384</td>
<td>0.353</td>
<td>0.341</td>
</tr>
<tr>
<td>360</td>
<td>0.334</td>
<td>0.314</td>
<td>0.288</td>
<td>0.278</td>
</tr>
<tr>
<td>480</td>
<td>0.289</td>
<td>0.272</td>
<td>0.249</td>
<td>0.241</td>
</tr>
<tr>
<td>600</td>
<td>0.258</td>
<td>0.243</td>
<td>0.223</td>
<td>0.216</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>N</th>
<th>(\hat{\gamma}_0)</th>
<th>(\hat{\gamma}_1)</th>
<th>(\hat{\gamma}_0)</th>
<th>(\hat{\gamma}_1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>60</td>
<td>0.776</td>
<td>0.517</td>
<td>0.638</td>
<td>0.283</td>
</tr>
<tr>
<td>120</td>
<td>0.574</td>
<td>0.472</td>
<td>0.447</td>
<td>0.258</td>
</tr>
<tr>
<td>240</td>
<td>0.425</td>
<td>0.398</td>
<td>0.323</td>
<td>0.228</td>
</tr>
<tr>
<td>360</td>
<td>0.353</td>
<td>0.341</td>
<td>0.268</td>
<td>0.207</td>
</tr>
<tr>
<td>480</td>
<td>0.306</td>
<td>0.299</td>
<td>0.235</td>
<td>0.191</td>
</tr>
<tr>
<td>600</td>
<td>0.273</td>
<td>0.267</td>
<td>0.212</td>
<td>0.178</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>N</th>
<th>(\hat{\gamma}_0)</th>
<th>(\hat{\gamma}_1)</th>
<th>(\hat{\gamma}_0)</th>
<th>(\hat{\gamma}_1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>60</td>
<td>0.681</td>
<td>0.653</td>
<td>0.459</td>
<td>0.367</td>
</tr>
<tr>
<td>120</td>
<td>0.481</td>
<td>0.462</td>
<td>0.324</td>
<td>0.260</td>
</tr>
<tr>
<td>240</td>
<td>0.340</td>
<td>0.327</td>
<td>0.229</td>
<td>0.184</td>
</tr>
<tr>
<td>360</td>
<td>0.278</td>
<td>0.267</td>
<td>0.187</td>
<td>0.150</td>
</tr>
<tr>
<td>480</td>
<td>0.241</td>
<td>0.231</td>
<td>0.162</td>
<td>0.130</td>
</tr>
<tr>
<td>600</td>
<td>0.215</td>
<td>0.207</td>
<td>0.145</td>
<td>0.116</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>N</th>
<th>(\hat{\gamma}_0)</th>
<th>(\hat{\gamma}_1)</th>
<th>(\hat{\gamma}_0)</th>
<th>(\hat{\gamma}_1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>60</td>
<td>0.632</td>
<td>0.338</td>
<td>0.580</td>
<td>0.188</td>
</tr>
<tr>
<td>120</td>
<td>0.445</td>
<td>0.313</td>
<td>0.355</td>
<td>0.152</td>
</tr>
<tr>
<td>240</td>
<td>0.319</td>
<td>0.272</td>
<td>0.238</td>
<td>0.128</td>
</tr>
<tr>
<td>360</td>
<td>0.265</td>
<td>0.241</td>
<td>0.191</td>
<td>0.114</td>
</tr>
<tr>
<td>480</td>
<td>0.232</td>
<td>0.217</td>
<td>0.165</td>
<td>0.104</td>
</tr>
<tr>
<td>600</td>
<td>0.209</td>
<td>0.198</td>
<td>0.147</td>
<td>0.096</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>N</th>
<th>(\hat{\gamma}_0)</th>
<th>(\hat{\gamma}_1)</th>
<th>(\hat{\gamma}_0)</th>
<th>(\hat{\gamma}_1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>60</td>
<td>0.563</td>
<td>0.194</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>120</td>
<td>0.376</td>
<td>0.186</td>
<td>0.562</td>
<td>0.159</td>
</tr>
<tr>
<td>240</td>
<td>0.252</td>
<td>0.167</td>
<td>0.209</td>
<td>0.076</td>
</tr>
<tr>
<td>360</td>
<td>0.201</td>
<td>0.151</td>
<td>0.153</td>
<td>0.063</td>
</tr>
<tr>
<td>480</td>
<td>0.173</td>
<td>0.138</td>
<td>0.127</td>
<td>0.056</td>
</tr>
<tr>
<td>600</td>
<td>0.154</td>
<td>0.127</td>
<td>0.111</td>
<td>0.051</td>
</tr>
</tbody>
</table>
Table IV
Unconditional Percentage Biases of Zero-Beta Rate and Risk Premium Estimators from the Second Pass Cross-Sectional Regressions under Nonnormality

The table presents the unconditional biases of OLS and estimated GLS CSR estimators of zero-beta rate ($\gamma_0$) and risk premium ($\gamma_1$) as a percentage of the true value of the risk premium for different length of beta estimation period (T) and number of test assets (N). The factor in Panel A is chosen to have a high explanatory power on the returns of the test assets whereas the factor in Panel B is chosen to have a 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 II except that the factor here 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. The results in the table are based on 100,000 simulations.

Panel A: Factor with high explanatory power

<table>
<thead>
<tr>
<th>T</th>
<th>$\gamma_0$</th>
<th>$\gamma_1$</th>
<th>$\gamma_0$</th>
<th>$\gamma_1$</th>
<th>$\gamma_0$</th>
<th>$\gamma_1$</th>
<th>$\gamma_0$</th>
<th>$\gamma_1$</th>
<th>$\gamma_0$</th>
<th>$\gamma_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>60</td>
<td>55.6</td>
<td>-52.5</td>
<td>72.6</td>
<td>-73.0</td>
<td>33.4</td>
<td>-29.7</td>
<td>44.1</td>
<td>-43.0</td>
<td>18.9</td>
<td>-17.1</td>
</tr>
<tr>
<td>120</td>
<td>35.4</td>
<td>-33.1</td>
<td>57.4</td>
<td>-57.7</td>
<td>17.4</td>
<td>-15.0</td>
<td>29.3</td>
<td>-28.6</td>
<td>9.7</td>
<td>-8.8</td>
</tr>
<tr>
<td>240</td>
<td>14.0</td>
<td>-12.5</td>
<td>40.2</td>
<td>-40.3</td>
<td>8.0</td>
<td>-6.9</td>
<td>17.5</td>
<td>-17.2</td>
<td>4.9</td>
<td>-4.3</td>
</tr>
<tr>
<td>360</td>
<td>5.4</td>
<td>-4.3</td>
<td>30.7</td>
<td>-30.7</td>
<td>4.9</td>
<td>-4.3</td>
<td>12.5</td>
<td>-12.5</td>
<td>3.3</td>
<td>-2.7</td>
</tr>
<tr>
<td>480</td>
<td>2.6</td>
<td>-1.7</td>
<td>24.7</td>
<td>-24.7</td>
<td>3.3</td>
<td>-2.9</td>
<td>9.7</td>
<td>-9.7</td>
<td>2.5</td>
<td>-2.0</td>
</tr>
<tr>
<td>600</td>
<td>0.6</td>
<td>0.1</td>
<td>20.5</td>
<td>-20.5</td>
<td>2.6</td>
<td>-2.2</td>
<td>7.9</td>
<td>-7.9</td>
<td>1.9</td>
<td>-1.6</td>
</tr>
</tbody>
</table>

Panel B: Factor with low explanatory power

<table>
<thead>
<tr>
<th>T</th>
<th>$\gamma_0$</th>
<th>$\gamma_1$</th>
<th>$\gamma_0$</th>
<th>$\gamma_1$</th>
<th>$\gamma_0$</th>
<th>$\gamma_1$</th>
<th>$\gamma_0$</th>
<th>$\gamma_1$</th>
<th>$\gamma_0$</th>
<th>$\gamma_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>60</td>
<td>52.7</td>
<td>-60.3</td>
<td>61.2</td>
<td>-86.1</td>
<td>67.9</td>
<td>-65.5</td>
<td>34.7</td>
<td>-81.7</td>
<td>66.8</td>
<td>-67.5</td>
</tr>
<tr>
<td>120</td>
<td>32.0</td>
<td>-37.5</td>
<td>47.8</td>
<td>-74.2</td>
<td>52.4</td>
<td>-43.9</td>
<td>26.1</td>
<td>-69.9</td>
<td>49.5</td>
<td>-47.7</td>
</tr>
<tr>
<td>240</td>
<td>17.9</td>
<td>-11.7</td>
<td>43.8</td>
<td>-61.6</td>
<td>30.2</td>
<td>-23.5</td>
<td>21.3</td>
<td>-54.8</td>
<td>35.1</td>
<td>-28.7</td>
</tr>
<tr>
<td>360</td>
<td>5.2</td>
<td>1.0</td>
<td>35.0</td>
<td>-50.2</td>
<td>19.8</td>
<td>-12.1</td>
<td>17.7</td>
<td>-45.2</td>
<td>27.5</td>
<td>-19.3</td>
</tr>
<tr>
<td>480</td>
<td>1.1</td>
<td>3.9</td>
<td>27.5</td>
<td>-40.5</td>
<td>15.6</td>
<td>-7.3</td>
<td>16.7</td>
<td>-38.4</td>
<td>21.7</td>
<td>-15.0</td>
</tr>
<tr>
<td>600</td>
<td>-1.8</td>
<td>5.5</td>
<td>23.4</td>
<td>-34.4</td>
<td>12.6</td>
<td>-4.9</td>
<td>14.2</td>
<td>-32.7</td>
<td>18.6</td>
<td>-11.6</td>
</tr>
</tbody>
</table>
Table V
Unconditional Standard Deviation of Zero-Beta Rate and Risk Premium Estimators from the Second Pass Cross-Sectional Regressions under Nonnormality

The table presents the unconditional standard deviations (in percentage per month) of OLS and estimated GLS CSR estimators of zero-beta rate ($\gamma_0$) and risk premium ($\gamma_1$) for different length of sample period ($T$) and number of test assets ($N$). The factor in Panel A is chosen to have a high explanatory power on the returns of the test assets whereas the factor in Panel B is chosen to have a low explanatory power on the returns of the test assets. The value of $\gamma_1$ 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 III except that the factor here 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. The results in the table are based on 100,000 simulations.

Panel A: Factor with high explanatory power

<table>
<thead>
<tr>
<th>$T$</th>
<th>OLS $\hat{\gamma}_0$</th>
<th>OLS $\hat{\gamma}_1$</th>
<th>OLS $\hat{\gamma}_0$</th>
<th>OLS $\hat{\gamma}_1$</th>
<th>OLS $\hat{\gamma}_0$</th>
<th>OLS $\hat{\gamma}_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>60</td>
<td>2.705</td>
<td>2.697</td>
<td>1.450</td>
<td>1.417</td>
<td>1.211</td>
<td>0.763</td>
</tr>
<tr>
<td>120</td>
<td>2.299</td>
<td>2.311</td>
<td>1.220</td>
<td>1.220</td>
<td>0.958</td>
<td>0.532</td>
</tr>
<tr>
<td>240</td>
<td>1.825</td>
<td>1.839</td>
<td>0.997</td>
<td>1.011</td>
<td>0.733</td>
<td>0.386</td>
</tr>
<tr>
<td>360</td>
<td>1.548</td>
<td>1.559</td>
<td>0.868</td>
<td>0.884</td>
<td>0.615</td>
<td>0.320</td>
</tr>
<tr>
<td>480</td>
<td>1.368</td>
<td>1.377</td>
<td>0.774</td>
<td>0.792</td>
<td>0.539</td>
<td>0.281</td>
</tr>
<tr>
<td>600</td>
<td>1.240</td>
<td>1.246</td>
<td>0.707</td>
<td>0.724</td>
<td>0.485</td>
<td>0.254</td>
</tr>
</tbody>
</table>

Panel B: Factor with low explanatory power

<table>
<thead>
<tr>
<th>$T$</th>
<th>OLS $\hat{\gamma}_0$</th>
<th>OLS $\hat{\gamma}_1$</th>
<th>OLS $\hat{\gamma}_0$</th>
<th>OLS $\hat{\gamma}_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>60</td>
<td>0.776</td>
<td>0.525</td>
<td>0.601</td>
<td>0.281</td>
</tr>
<tr>
<td>120</td>
<td>0.573</td>
<td>0.473</td>
<td>0.430</td>
<td>0.255</td>
</tr>
<tr>
<td>240</td>
<td>0.424</td>
<td>0.394</td>
<td>0.317</td>
<td>0.226</td>
</tr>
<tr>
<td>360</td>
<td>0.351</td>
<td>0.337</td>
<td>0.265</td>
<td>0.206</td>
</tr>
<tr>
<td>480</td>
<td>0.307</td>
<td>0.297</td>
<td>0.233</td>
<td>0.190</td>
</tr>
<tr>
<td>600</td>
<td>0.274</td>
<td>0.266</td>
<td>0.211</td>
<td>0.177</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$N$</th>
<th>OLS $\hat{\gamma}_0$</th>
<th>OLS $\hat{\gamma}_1$</th>
<th>OLS $\hat{\gamma}_0$</th>
<th>OLS $\hat{\gamma}_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0.631</td>
<td>0.345</td>
<td>0.510</td>
<td>0.184</td>
</tr>
<tr>
<td>20</td>
<td>0.447</td>
<td>0.313</td>
<td>0.328</td>
<td>0.148</td>
</tr>
<tr>
<td>40</td>
<td>0.320</td>
<td>0.270</td>
<td>0.227</td>
<td>0.125</td>
</tr>
<tr>
<td>80</td>
<td>0.264</td>
<td>0.239</td>
<td>0.185</td>
<td>0.112</td>
</tr>
<tr>
<td>160</td>
<td>0.232</td>
<td>0.215</td>
<td>0.160</td>
<td>0.102</td>
</tr>
<tr>
<td>320</td>
<td>0.209</td>
<td>0.196</td>
<td>0.143</td>
<td>0.095</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$N$</th>
<th>OLS $\hat{\gamma}_0$</th>
<th>OLS $\hat{\gamma}_1$</th>
<th>OLS $\hat{\gamma}_0$</th>
<th>OLS $\hat{\gamma}_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0.563</td>
<td>0.205</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>20</td>
<td>0.378</td>
<td>0.189</td>
<td>0.452</td>
<td>0.148</td>
</tr>
<tr>
<td>40</td>
<td>0.253</td>
<td>0.167</td>
<td>0.182</td>
<td>0.072</td>
</tr>
<tr>
<td>80</td>
<td>0.202</td>
<td>0.150</td>
<td>0.138</td>
<td>0.060</td>
</tr>
<tr>
<td>160</td>
<td>0.174</td>
<td>0.137</td>
<td>0.117</td>
<td>0.054</td>
</tr>
<tr>
<td>320</td>
<td>0.155</td>
<td>0.127</td>
<td>0.103</td>
<td>0.049</td>
</tr>
</tbody>
</table>
Table VI
Unconditional Percentage Biases of Bias-adjusted Estimators of Zero-Beta Rate and Risk Premium from the Second Pass Cross-Sectional Regressions

The table presents the unconditional biases of bias-adjusted estimators of zero-beta rate ($\gamma_0$) and risk premium ($\gamma_1$) from the OLS and estimated GLS CSR, as a percentage of the true value of the risk premium for different length of beta estimation period ($T$) and number of test assets ($N$) when the factors and returns are multivariate normally distributed. The factor in Panel A is chosen to have a high explanatory power on the returns of the test assets whereas the factor in Panel B is chosen to have a low explanatory power on the returns of the test assets. The results in the table are based on 100,000 simulations.

Panel A: Factor with high explanatory power

<table>
<thead>
<tr>
<th>$T$</th>
<th>OLS</th>
<th>GLS</th>
<th>OLS</th>
<th>GLS</th>
<th>OLS</th>
<th>GLS</th>
<th>OLS</th>
<th>GLS</th>
<th>OLS</th>
<th>GLS</th>
<th>OLS</th>
<th>GLS</th>
</tr>
</thead>
<tbody>
<tr>
<td>60</td>
<td>46.5</td>
<td>-44.6</td>
<td>22.9</td>
<td>-23.1</td>
<td>8.7</td>
<td>-7.8</td>
<td>-17.3</td>
<td>17.2</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>120</td>
<td>19.4</td>
<td>-18.2</td>
<td>-10.4</td>
<td>10.5</td>
<td>0.5</td>
<td>-0.2</td>
<td>-5.7</td>
<td>5.7</td>
<td>-23.6</td>
<td>21.4</td>
<td>-2.3</td>
<td>2.3</td>
</tr>
<tr>
<td>240</td>
<td>-0.1</td>
<td>0.5</td>
<td>-21.2</td>
<td>21.4</td>
<td>-0.8</td>
<td>0.9</td>
<td>-1.0</td>
<td>1.0</td>
<td>-2.6</td>
<td>2.4</td>
<td>-0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>360</td>
<td>-4.1</td>
<td>4.2</td>
<td>-15.2</td>
<td>15.3</td>
<td>-0.6</td>
<td>0.6</td>
<td>-0.4</td>
<td>0.4</td>
<td>-1.0</td>
<td>0.9</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>480</td>
<td>-4.4</td>
<td>4.4</td>
<td>-9.4</td>
<td>9.5</td>
<td>-0.4</td>
<td>0.4</td>
<td>-0.2</td>
<td>0.2</td>
<td>-0.5</td>
<td>0.5</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>600</td>
<td>-3.8</td>
<td>3.8</td>
<td>-5.8</td>
<td>5.8</td>
<td>-0.3</td>
<td>0.3</td>
<td>-0.2</td>
<td>0.2</td>
<td>-0.3</td>
<td>0.3</td>
<td>0.0</td>
<td>0.0</td>
</tr>
</tbody>
</table>

Panel B: Factor with low explanatory power

<table>
<thead>
<tr>
<th>$T$</th>
<th>OLS</th>
<th>GLS</th>
<th>OLS</th>
<th>GLS</th>
<th>OLS</th>
<th>GLS</th>
<th>OLS</th>
<th>GLS</th>
<th>OLS</th>
<th>GLS</th>
<th>OLS</th>
<th>GLS</th>
</tr>
</thead>
<tbody>
<tr>
<td>60</td>
<td>46.7</td>
<td>-50.9</td>
<td>39.9</td>
<td>-57.6</td>
<td>50.8</td>
<td>-50.9</td>
<td>12.6</td>
<td>-30.0</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>120</td>
<td>22.0</td>
<td>-22.4</td>
<td>18.3</td>
<td>-26.2</td>
<td>29.0</td>
<td>-26.3</td>
<td>-4.5</td>
<td>11.3</td>
<td>13.1</td>
<td>-11.1</td>
<td>-1.7</td>
<td>8.9</td>
</tr>
<tr>
<td>240</td>
<td>-0.4</td>
<td>2.7</td>
<td>-4.9</td>
<td>7.2</td>
<td>8.6</td>
<td>-5.1</td>
<td>-10.2</td>
<td>25.2</td>
<td>7.9</td>
<td>-4.2</td>
<td>-1.7</td>
<td>11.7</td>
</tr>
<tr>
<td>360</td>
<td>-6.3</td>
<td>8.7</td>
<td>-13.6</td>
<td>19.7</td>
<td>1.7</td>
<td>1.1</td>
<td>-7.1</td>
<td>17.3</td>
<td>3.7</td>
<td>-1.0</td>
<td>-0.5</td>
<td>3.1</td>
</tr>
<tr>
<td>480</td>
<td>-6.9</td>
<td>8.7</td>
<td>-15.5</td>
<td>22.4</td>
<td>-0.7</td>
<td>2.6</td>
<td>-4.1</td>
<td>10.1</td>
<td>1.8</td>
<td>0.1</td>
<td>-0.2</td>
<td>1.5</td>
</tr>
<tr>
<td>600</td>
<td>-4.5</td>
<td>7.1</td>
<td>-14.5</td>
<td>21.0</td>
<td>-1.3</td>
<td>2.7</td>
<td>-2.4</td>
<td>6.0</td>
<td>0.9</td>
<td>0.5</td>
<td>-0.1</td>
<td>0.9</td>
</tr>
</tbody>
</table>
Table VII
Unconditional Standard Deviation of Bias-adjusted Estimators of Zero-Beta Rate and Risk Premium from the Second Pass Cross-Sectional Regressions

The table presents the unconditional standard deviations (in percentages per month) of bias-adjusted estimators of zero-beta rate ($\gamma_0$) and risk premium ($\gamma_1$) from the OLS and estimated GLS CSR for different length of sample period ($T$) and number of test assets ($N$) when the factors and returns are multivariate normally distributed. The factor in Panel A is chosen to have a high explanatory power on the returns of the test assets whereas the factor in Panel B is chosen to have a low explanatory power on the returns of the test assets. The value of $\gamma_1$ 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.

<table>
<thead>
<tr>
<th>N = 10</th>
<th>N = 25</th>
<th>N = 100</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OLS</td>
<td>GLS</td>
</tr>
<tr>
<td></td>
<td>$\hat{\gamma}_0$</td>
<td>$\hat{\gamma}_1$</td>
</tr>
<tr>
<td>60</td>
<td>3.724</td>
<td>3.698</td>
</tr>
<tr>
<td>120</td>
<td>3.018</td>
<td>3.010</td>
</tr>
<tr>
<td>240</td>
<td>2.192</td>
<td>2.192</td>
</tr>
<tr>
<td>360</td>
<td>1.757</td>
<td>1.758</td>
</tr>
<tr>
<td>480</td>
<td>1.491</td>
<td>1.492</td>
</tr>
<tr>
<td>600</td>
<td>1.313</td>
<td>1.313</td>
</tr>
</tbody>
</table>

Panel B: Factor with low explanatory power

<table>
<thead>
<tr>
<th>N = 10</th>
<th>N = 25</th>
<th>N = 100</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OLS</td>
<td>GLS</td>
</tr>
<tr>
<td></td>
<td>$\hat{\gamma}_0$</td>
<td>$\hat{\gamma}_1$</td>
</tr>
<tr>
<td>60</td>
<td>0.949</td>
<td>0.723</td>
</tr>
<tr>
<td>120</td>
<td>0.702</td>
<td>0.633</td>
</tr>
<tr>
<td>240</td>
<td>0.503</td>
<td>0.497</td>
</tr>
<tr>
<td>360</td>
<td>0.401</td>
<td>0.403</td>
</tr>
<tr>
<td>480</td>
<td>0.336</td>
<td>0.335</td>
</tr>
<tr>
<td>600</td>
<td>0.291</td>
<td>0.288</td>
</tr>
</tbody>
</table>