Asset Pricing Tests with Mimicking Portfolios

December 7, 2014

Abstract

Mimicking portfolios for factors are often used in asset pricing studies. Current practice has generally ignored the impact of estimation errors on the weights of the mimicking portfolios. We show that such a practice can lead to gross understatement of the standard errors of the estimated risk premia associated with the mimicking portfolios, especially when the factors are not highly correlated with the returns on the test assets. In this paper, we present a methodology that properly takes into account the impact of the estimation errors of the mimicking portfolios on the standard error of estimated risk premia. In empirical applications, we report that the outcome of asset pricing tests can vary significantly, depending on whether the estimation errors on the weights of the mimicking portfolios are accounted for. Our findings thus cast doubt on existing empirical studies that use mimicking portfolios but ignore the estimation error problem.

JEL Classification: G12

Keywords: asset pricing; risk factors; mimicking portfolios; estimation error; risk premia; standard error.

1 Introduction

A large body of the asset pricing literature relies on the usage of mimicking portfolios, see, e.g., Breeden (1979) for the early discussion. These portfolios are constructed by mimicking the risk factors for asset returns, and are commonly used in asset pricing tests. Examples of the practical usage of mimicking portfolios can be found in, e.g., Chen *et al.* (1986), Breeden *et al.* (1989), Ferson and Harvey (1991), Pástor and Stambaugh (2003), Ang *et al.* (2006), Muir *et al.* (2013), etc.

Although risk factors and their mimicking portfolios co-exist in asset pricing studies, there are empirical as well as theoretical reasons that favor mimicking portfolios over factors. One of the reasons, for instance, concerns data availability, i.e., economic risk factors are usually only observable at a low frequency and/or within limited time periods, while the constructed mimicking portfolios for such factors could be available at higher frequencies and in extended time periods (see, e.g., Ang *et al.*, 2006). In addition, given mimicking portfolios are constructed with assets, their returns are more closely correlated with asset returns than risk factors in finite sample applications. This statistical quality is crucial for the inference on risk premia, since Kan and Zhang (1999), Kleibergen (2009) and Gospodinov *et al.* (2014) have warned that weak correlation of factors and returns induces spurious results that incorrectly favor weak or even useless factors in the Fama and MacBeth (1973) two-pass procedure.[1] In terms of testing, Huberman *et al.* (1987) show that mimicking portfolios can be used to test additional asset pricing restrictions that can not be tested with factors. See also Gibbons *et al.* (1989). Furthermore, mimicking portfolios could also be used to hedge against economic risks, see, e.g., Lamont (2001). Finally, since mimicking portfolios reflect the excess returns on zero-cost portfolios, an advantage of using mimicking portfolios rather than risk factors lies in that the estimation outcome can be straightforwardly interpreted from the investment perspective. See, e.g., Brennan *et al.* (1998) and Avramov and Chordia

[1] On the other hand, Shanken (1992), Balduzzi and Robotti (2008) and Chordia *et al.* (2013) suggest that the inference on risk premia could be improved with the usage of mimicking portfolios.

(2006).

Unlike the advantages listed above, the downside of using mimicking portfolios for asset pricing, however, is rarely discussed. Furthermore, the methods of constructing mimicking portfolios vary, and it is not clear yet which method provides the best finite sample performance. For example, Huberman *et al.* (1987) provide three different ways to generate mimicking portfolios. Lehmann and Modest (1988) propose a weighted least squares procedure as well as a minimum idiosyncratic risk procedure. In the empirical literature, Breeden *et al.* (1989), Lamont (2001), Vassalou (2003), Avramov and Chordia (2006), Kapadia (2011), Menkhoff *et al.* (2012) and Muir *et al.* (2013) construct mimicking portfolios by projecting factors to a set of portfolios which approximately span the space of returns. On the other hand, Pástor and Stambaugh (2003), Ang *et al.* (2006) and Chang *et al.* (2013) form factor mimicking portfolios by using the difference in return of the portfolio with highest correlation and the one with lowest correlation with factors.[2]

Our paper is motivated by the simple fact that the weights of mimicking portfolios are unknown and thus have to be estimated, no matter which construction method is adopted. Consequently, mimicking portfolios used in practice are contaminated by the estimation errors on their weights, which could potentially affect asset pricing tests that rely on mimicking portfolios. This downside of mimicking portfolios thus calls for the investigation of the estimation errors occurred during their construction, and the impact of these errors on subsequent tests.

In this paper, we derive the limiting behavior of the risk premium estimator using mimicking portfolios in the Fama and MacBeth (1973) two-pass procedure. In particular, we take the estimation errors on the weights of maximum correlation portfolios into account. We find that ignoring the errors could substantially understate the standard errors of risk premia, particularly when the risk factors are only weakly correlated with returns. In or-

[2] Mimicking portfolios that are generated by sorting on firm characteristics such as market equity and book-to-market ratio can be also considered as in this category if we assume the characteristics are correlated with their loadings of the undying factors as in Chan *et al.* (1998).

der to correct for this problem, we provide the asymptotic standard error of estimated risk premia that properly takes into the estimation errors of mimicking portfolios. Simulation evidence suggests that our asymptotic results provide a reasonable approximation in finite samples. We apply our methodology to analyze asset pricing models in Cochrane (1996), Li *et al.* (2006) and Muir *et al.* (2013). Interestingly, we find that once estimation errors of mimicking portfolios are accounted for, standard errors of risk premia increase around $50\% - 100\%$ in most cases. Our findings thus indicate that estimation errors associated with mimicking portfolios are important, and put into question the existing studies that ignore them.

The remainder of this paper is organized as follows. Section 2 presents our setup and analytical results. Section 3 contains the simulation outcome as well as the empirical application. Section 4 concludes. The proofs and technical details are included in the Appendix.

2 Setup and Analysis

2.1 Setup

Consider the vector Y_t, which consists of the $K \times 1$ vector f_t for risk factors and the $N \times 1$ vector R_t for asset returns:

$$Y_t = \begin{bmatrix} f_t \\ R_t \end{bmatrix} \tag{1}$$

where stationarity and ergodicity of Y_t are assumed, and $N \geq K + 1$.

The mean and variance of Y_t read:

$$\mu = \mathbb{E}[Y_t] = \begin{bmatrix} \mu_f \\ \mu_R \end{bmatrix}, \qquad V = \mathrm{Var}[Y_t] = \begin{bmatrix} V_f & V_{fR} \\ V_{Rf} & V_R \end{bmatrix}, \tag{2}$$

where V is assumed to be positive definite, and both μ and V can be consistently estimated by their sample counterparts based on T observations of Y_t, $t = 1, ..., T$:

$$\hat{\mu} = \begin{bmatrix} \hat{\mu}_f \\ \hat{\mu}_R \end{bmatrix} = \frac{1}{T}\sum_{t=1}^{T} Y_t, \qquad \hat{V} = \begin{bmatrix} \hat{V}_f & \hat{V}_{fR} \\ \hat{V}_{Rf} & \hat{V}_R \end{bmatrix} = \frac{1}{T}\sum_{t=1}^{T}(Y_t - \hat{\mu})(Y_t - \hat{\mu})'. \tag{3}$$

Let M be the $N \times K$ full rank matrix with the columns being the weights of the K mimicking portfolios. Huberman *et al.* (1987) show that M is factor mimicking if and only if

$$M = V_R^{-1} V_{Rf} L \tag{4}$$

where L is any nonsingular $K \times K$ matrix. The returns for the corresponding mimicking portfolios are then given by

$$g_t = M' R_t, \tag{5}$$

with mean and variance

$$\mu_g = M' \mu_R, \qquad V_g = M' V_R M, \tag{6}$$

respectively.

2.2 Mimicking Portfolios in CSR with Given Weights

Recently, Kan *et al.* (2013) provide the asymptotic distribution of the risk premium estimator in the two-pass cross-sectional regression (CSR, see, e.g., Fama and MacBeth 1973), when the linear factor model made of f_t and R_t is allowed to be potentially misspecified. We extend Kan *et al.* (2013)'s results for factors to mimicking portfolios. As a starting point, we consider the ideal scenario that the matrix of the weights of mimicking portfolios, denoted by M, is given.

In Theorem 1, we first list Kan *et al.* (2013)'s results (part 1 for factors f_t), then extend these results to mimicking portfolios (part 2 for mimicking portfolios g_t).

Theorem 1. *Let* $\hat{\beta} \equiv \hat{V}_{Rf}\hat{V}_f^{-1}$, $\hat{X} \equiv [1_N,\ \hat{\beta}]$. *Similarly for mimicking portfolios,* $\hat{\beta}_g \equiv \hat{V}_R M(M'\hat{V}_R M)^{-1}$, $\hat{X}_g \equiv [1_N,\ \hat{\beta}_g]$. W *is a positive definite weighting matrix.*

1. *For factors (see Kan et al. 2013):*

   (i) If W is known, the asymptotic distribution of $\hat{\gamma} = (\hat{X}'W\hat{X})^{-1}\hat{X}'W\hat{\mu}_R$ is given by $\sqrt{T}(\hat{\gamma} - \gamma) \overset{A}{\sim} N(0, V_{\hat{\gamma}})$, where $V_{\hat{\gamma}} = \sum_{j=-\infty}^{\infty} \mathbb{E}[h_t h'_{t+j}]$, with

$$h_t = (\gamma_t - \gamma) - (\phi_t - \phi)w_t + Hz_t. \tag{7}$$

   (ii) For the feasible GLS case with $W = V_R^{-1}$, which is estimated by $\hat{V}_R^{-1}$, the asymptotic distribution of $\hat{\gamma} = (\hat{X}'\hat{V}_R^{-1}\hat{X})^{-1}\hat{X}'\hat{V}_R^{-1}\hat{\mu}_R$ is given by $\sqrt{T}(\hat{\gamma} - \gamma) \overset{A}{\sim} N(0, V_{\hat{\gamma}})$, where $V_{\hat{\gamma}} = \sum_{j=-\infty}^{\infty} \mathbb{E}[h_t h'_{t+j}]$, with

$$h_t = (\gamma_t - \gamma) - (\phi_t - \phi)w_t + Hz_t - (\gamma_t - \gamma)u_t. \tag{8}$$

   For both (i) and (ii), $\gamma_t = [\gamma_{0,t},\ \gamma'_{1,t}]' = AR_t$, $\gamma = [\gamma_0,\ \gamma'_1]' = A\mu_R$, $A = HX'W$, $H = (X'WX)^{-1}$, $X = [1_N,\ \beta]$, $\beta = V_{Rf}V_f^{-1}$, $\phi_t = [\gamma_{0,t},\ (\gamma_{1,t} - f_t)']'$, $\phi = [\gamma_0,\ (\gamma_1 - \mu_f)']'$, $w_t = \gamma'_1 V_f^{-1}(f_t - \mu_f)$, $u_t = e'W(R_t - \mu_R)$, $e = \mu_R - X\gamma$, $z_t = [0,\ u_t(f_t - \mu_f)'V_f^{-1}]'$.

2. *For mimicking portfolios with given weights:*

   (iii) If W is known, the asymptotic distribution of $\hat{\gamma}_g = (\hat{X}'_g W\hat{X}_g)^{-1}\hat{X}'_g W\hat{\mu}_R$ is given by $\sqrt{T}(\hat{\gamma}_g - \gamma_g) \overset{A}{\sim} N(0, V_{\hat{\gamma}_g})$, where $V_{\hat{\gamma}_g} = \sum_{j=-\infty}^{\infty} \mathbb{E}[h_{g,t} h'_{g,t+j}]$ with

$$h_{g,t} = (\gamma_{g,t} - \gamma_g) - (\phi_{g,t} - \phi_g)w_{g,t} + H_g z_{g,t} \tag{9}$$

   (iv) For the feasible GLS case with $W = V_R^{-1}$, which is estimated by $\hat{V}_R^{-1}$, the asymptotic distribution of $\hat{\gamma}_g = (\hat{X}'_g\hat{V}_R^{-1}\hat{X}_g)^{-1}\hat{X}'_g\hat{V}_R^{-1}\hat{\mu}_R$ is given by $\sqrt{T}(\hat{\gamma}_g - \gamma_g) \overset{A}{\sim} N(0, V_{\hat{\gamma}_g})$, where $V_{\hat{\gamma}_g} = \sum_{j=-\infty}^{\infty} \mathbb{E}[h_{g,t} h'_{g,t+j}]$ with

$$h_{g,t} = (\gamma_{g,t} - \gamma_g) - (\phi_{g,t} - \phi_g)w_{g,t} + H_g z_{g,t} - (\gamma_{g,t} - \gamma_g)u_{g,t} \tag{10}$$

For both (iii) and (iv), $\gamma_{g,t} = [\gamma_{g0,t}, \gamma'_{g1,t}]' = A_g R_t$, $\gamma_g = [\gamma_{g0},\ \gamma'_{g1}]' = A_g\mu_R$, $A_g = H_g X'_g W$, $H_g = (X'_g W X_g)^{-1}$, $X_g = [1_N,\ \beta_g]$, $\beta_g = V_R M(M'V_R M)^{-1}$, $\phi_{g,t} = [\gamma_{g0,t},\ (\gamma_{g1,t} - g_t)']'$, $\phi_g = [\gamma_{g0}, (\gamma_{g1} - \mu_g)']'$, $w_{g,t} = \gamma'_{g1} V_g^{-1}(g_t - \mu_g)$, $u_{g,t} = e'_g W(R_t - \mu_R)$, $e_g = \mu_R - X_g\gamma_g$, $z_{g,t} = [0,\ u_{g,t}(g_t - \mu_g)'V_g^{-1}]'$.

It is easy to check that (i)(ii) in Theorem 1 coincide with (iii)(iv), if we replace each object for factors in (i)(ii) with its counterpart in (iii)(iv) resulting from mimicking portfolios. As a result, in the ideal scenario that the weights of mimicking portfolios are given, mimicking portfolios can be treated in the same manner as factors in CSR.

2.3 Mimicking Portfolios in CSR with Estimated Weights

In practice, the weights of mimicking portfolios are unknown and must be estimated. Empirical researchers thus have to work with

$$\hat{g}_t = \hat{M}' R_t \tag{11}$$

where $\hat{M}$ is an estimator of M. Consequently, mimicking portfolios used in practice are contaminated by the estimation error on their weights. In many situations, $\hat{M}$ is a consistent estimator of M, so using $\hat{g}_t$ instead of g_t does not impact the consistency of the estimated risk premia. Nevertheless, the estimation error of $\hat{M}$ has a first order impact on the asymptotic variance of the estimated risk premia and its impact cannot be ignored, especially when the factors are poorly mimicked by the returns on the assets.

In order to explicitly analyze the asymptotic variance of the risk premium estimator under estimated mimicking portfolios, we focus on the case of maximum correlation portfolios, which corresponds to $M = V_R^{-1} V_{Rf}$, i.e., $L = I_K$, so:[3]

$$g_t = V_{fR} V_R^{-1} R_t, \qquad \hat{g}_t = \hat{V}_{fR}\hat{V}_R^{-1} R_t. \tag{12}$$

[3] Other types of mimicking portfolios, such as those suggested by Huberman *et al.* (1987), can be considered as linear transformations of maximum correlation portfolios, so can be analyzed in a similar manner.

These mimicking portfolios are obtained by projecting f_t on R_t and a constant term:

$$f_t = \mu_f - V_{fR}V_R^{-1}\mu_R + V_{fR}V_R^{-1}R_t + \eta_t = \mu_f - \mu_g + g_t + \eta_t \tag{13}$$

where $\eta_t = (f_t - \mu_f) - (g_t - \mu_g)$ is uncorrelated with R_t, and we denote its variance by $V_\eta = V_f - V_{fR}V_R^{-1}V_{Rf}$.

In Theorem 2, we provide the asymptotic distribution of the risk premium estimator, when the betas with respect to $\hat{g}_t = \hat{V}_{fR}\hat{V}_R^{-1}R_t$ are used in CSR.

Theorem 2. *Let* $\hat{\beta}_{\hat{g}} \equiv \hat{V}_{Rf}(\hat{V}_{fR}\hat{V}_R^{-1}\hat{V}_{Rf})^{-1}$, $\hat{X}_{\hat{g}} \equiv [1_N,\ \hat{\beta}_{\hat{g}}]$, *$W$ is a positive definite weighting matrix. If W is known, the asymptotic distribution of* $\hat{\gamma}_{\hat{g}} = (\hat{X}_{\hat{g}}'W\hat{X}_{\hat{g}})^{-1}\hat{X}_{\hat{g}}'W\hat{\mu}_R$ *is given by* $\sqrt{T}(\hat{\gamma}_{\hat{g}} - \gamma_g) \overset{A}{\sim} N(0, V_{\hat{\gamma}_{\hat{g}}})$, *where* $V_{\hat{\gamma}_{\hat{g}}} = \sum_{j=-\infty}^{\infty}\mathbb{E}[h_{\hat{g},t}h_{\hat{g},t+j}']$ *with*

$$h_{\hat{g},t} = \begin{bmatrix} 1 & 0_K' \\ 0_K & V_gV_f^{-1} \end{bmatrix} h_t + \begin{bmatrix} 0 \\ \left[V_\eta V_f^{-1}(f_t-\mu_f)(f_t-\mu_f)' - \eta_t\eta_t'\right]V_f^{-1}\gamma_1 \end{bmatrix} \tag{14}$$

and h_t is provided in (7), $V_g = V_{fR}V_R^{-1}V_{Rf}$. *In addition, when W is unknown, (14) similarly holds for the asymptotic distribution of* $\hat{\gamma}_{\hat{g}} = (\hat{X}_{\hat{g}}'\hat{V}_R^{-1}\hat{X}_{\hat{g}})^{-1}\hat{X}_{\hat{g}}'\hat{V}_R^{-1}\hat{\mu}_R$ *with h_t provided in (8).*

As in Theorem 2, $h_{\hat{g},t}$ can be written as the scaled h_t in (i)(ii) of Theorem 1 plus an extra term that depends on η_t, the error in the construction of mimicking portfolios by the time series regression approach. In particular, Theorem 2 coincides with (i)(ii) of Theorem 1 when η_t equals zero.

When η_t is nonzero but sufficiently small (i.e., $\eta_t \approx 0_K$, so $V_\eta \approx 0_{K\times K}$, $V_g \approx V_f$), (14) suggests that the results for factors in Theorem 1 and those for mimicking portfolios in Theorem 2 are almost identical. From this perspective, it is not completely unreasonable to treat mimicking portfolios in the same manner as factors in CSR, assuming that the error in mimicking portfolio construction is small. This treatment, however, is expected to be problematic when the error is substantial, as shown below.

2.4 Consequence of Ignoring the Estimation Error on the Weights

Consider the following scenario: with the data of f_t and R_t, a researcher constructs mimicking portfolios $\hat{g}_t = \hat{V}_{fR}\hat{V}_R^{-1}R_t$. After construction, the researcher treats the constructed $\hat{g}_t$ in the same manner as g_t. By doing so, the researcher ignores the estimation error and effectively turns to $h_{g,t}$ in (iii)(iv) of Theorem 1 to derive standard errors of risk premia, instead of $h_{\hat{g},t}$ in Theorem 2.

To highlight the difference between $h_{g,t}$ and $h_{\hat{g},t}$, we rewrite $h_{\hat{g},t}$ as follows.

Corollary 2.1. *Consider mimicking portfolios under $L = I_K$. If W is known, $h_{g,t}$ in (9) and $h_{\hat{g},t}$ in (14) satisfy:*

$$h_{\hat{g},t} = h_{g,t} + \delta_t \tag{15}$$

where

$$\delta_t = -(\phi_{g,t} - \phi_g)\gamma_{g1}' V_g^{-1}\eta_t + H_g \begin{bmatrix} 0 \\ u_{g,t}V_g^{-1}\eta_t \end{bmatrix} + \begin{bmatrix} 0 \\ \eta_t \end{bmatrix} \gamma_{g1}' V_g^{-1}(g_t - \mu_g).$$

In addition, (15) also holds for the $h_{g,t}$ and $h_{\hat{g},t}$ corresponding to feasible GLS.

In the theorem below, we continue to compare the variances of the risk premium estimator corresponding to $h_{\hat{g},t}$ and $h_{g,t}$, under the joint elliptical distribution assumption of Y_t.

Theorem 3. *If Y_t is i.i.d. and follows the elliptical distribution, then Corollary 2.1 implies:*

$$\mathbb{E}[h_{\hat{g},t}h_{\hat{g},t}'] = \mathbb{E}[h_{g,t}h_{g,t}'] + \mathbb{E}[\delta_t\delta_t'] \tag{16}$$

where

$$
\begin{aligned}
&\mathbb{E}[\delta_t\delta_t'] \\
=&(1+\kappa)\left(A_g V_R A_g' - \begin{bmatrix} 0 & 0_K' \\ 0_K & V_g \end{bmatrix}\right)(\gamma_{g1}' V_g^{-1} V_\eta V_g^{-1}\gamma_{g1}) \\
&+(1+\kappa)H_g \begin{bmatrix} 0 & 0_K' \\ 0_K & V_g^{-1}V_\eta V_g^{-1} \end{bmatrix} H_g(e'WV_RWe) + (1+\kappa)\begin{bmatrix} 0 & 0_K' \\ 0_K & V_\eta \end{bmatrix}(\gamma_{g1}' V_g^{-1}\gamma_{g1})
\end{aligned}
$$

with κ the kurtosis parameter of the elliptical distribution.[4]

Theorem 3 indicates that ignoring estimation error of mimicking portfolios in CSR could effectively under-estimate the variance of the risk premium estimator, since $\mathbb{E}[h_{g,t}h_{g,t}'] \leq \mathbb{E}[h_{\hat{g},t}h_{\hat{g},t}']$.

In addition, the cost of ignoring estimation error is reflected by $\mathbb{E}[\delta_t\delta_t']$, which increases in V_η. For factors that are minorly correlated with returns, V_η tends to be large. Consequently, the variance of the risk premium estimator is more likely to be severely under-estimated for mimicking portfolios on such factors.

Corollary 3.1. *In the special case that $K = 1$ (i.e., there is only one factor considered for asset pricing), under the i.i.d. and joint elliptical distribution of Y_t, variances of $\hat{\gamma}_{\hat{g}}$ in Theorem 2 and $\hat{\gamma}$ in Theorem 1 satisfy:*

$$
V_{\hat{\gamma}_{\hat{g}}} = \begin{bmatrix} 1 & 0_K' \\ 0_K & V_g V_f^{-1} \end{bmatrix} V_{\hat{\gamma}} \begin{bmatrix} 1 & 0_K' \\ 0_K & V_g V_f^{-1} \end{bmatrix}' \tag{17}
$$

When $K = 1$, Corollary 3.1 suggests that the variance of the risk premium estimator under mimicking portfolios equals the scaled version of its counterpart under original factors.

[4]The expression of $\mathbb{E}[h_{g,t}h_{g,t}']$ results from replacing each object in Lemma 1, 2 of Kan *et al.* (2013) with its counterpart from mimicking portfolios.

2.5 Further Discussion

In this part, we briefly discuss several issues that are related to our analysis above.

2.5.1 Beta, Risk Premia, Pricing Error for Factor and Mimicking Portfolios

Firstly, we relate the beta, risk premia and pricing error for original risk factors denoted by f_t to their counterparts resulting from mimicking portfolios denoted by g_t:

$$\beta_g = \beta V_f L V_g^{-1}, \ \gamma_g = \begin{bmatrix} 1 & 0_K' \\ 0_K & V_g L^{-1} V_f^{-1} \end{bmatrix} \gamma, \ e_g = e \tag{18}$$

where β, γ, e and β_g, γ_g, e_g have been used in Theorem 1.

A subtle point in (18) is that pricing errors that use original risk factors and mimicking portfolios are identical. From this perspective, there is little advantage of using the pricing error under mimicking portfolios in specification tests.

In addition, when $L = I_K$, $\gamma_g = \begin{bmatrix} 1 & 0_K' \\ 0_K & V_g V_f^{-1} \end{bmatrix} \gamma$. This result, together with Corollary 3.1 above, suggests that there does not appear to be advantage of using mimicking portfolios over factors under $K = 1$, since the t-ratios in the two setup are identical if estimation error of mimicking portfolios is correctly accounted for.

2.5.2 Factors and Mimicking Portfolios in CSR

Secondly, we explore the scenario that factors and mimicking portfolios are simultaneously used in CSR, which is not uncommon in practice. The asymptotic distribution of the risk premium estimator in this scenario is thus of practical importance. For brevity, we provide the result in the appendix.

2.5.3 An Alternative Scheme

Finally, we consider an alternative scheme for forming mimicking portfolios that are often used in the empirical literature. As will be shown below, the alternative scheme nests our setup above as a special case.

Let R_t be the returns on N test assets at time t, r_t be the returns on L benchmark assets at time t, and f_t be the realizations of K factors at time t, where we assume $N \geq K+1$ and $L \geq K$. Instead of projecting f_t on R_t to form mimicking portfolios, researchers often project f_t on r_t to obtain the mimicking portfolios. Let

$$Y_t = \begin{bmatrix} f_t \\ r_t \\ R_t \end{bmatrix}, \tag{19}$$

and the mean and covariance matrix of Y_t is defined as

$$\mu = \mathbb{E}[Y_t] = \begin{bmatrix} \mu_f \\ \mu_r \\ \mu_R \end{bmatrix}, \qquad V = \mathrm{Var}[Y_t] = \begin{bmatrix} V_f & V_{fr} & V_{fR} \\ V_{rf} & V_r & V_{rR} \\ V_{Rf} & V_{Rr} & V_R \end{bmatrix}. \tag{20}$$

The returns on the mimicking portfolios obtained by projecting f_t on r_t are given by

$$\tilde{g}_t = V_{fr}V_r^{-1}r_t, \tag{21}$$

with mean and variance

$$\mu_{\tilde{g}} = V_{fr}V_r^{-1}\mu_r, \qquad V_{\tilde{g}} = V_{fr}V_r^{-1}V_{rf}, \tag{22}$$

respectively. Note that $\tilde{g}_t$ coincides with g_t in (12), if $r_t = R_t$.

It follows that the beta of R_t with respect to $\tilde{g}_t$ is

$$\beta_{\tilde{g}} = \mathrm{Cov}[R_t, \tilde{g}_t^{'}]\mathrm{Var}[\tilde{g}_t]^{-1} = V_{Rr}V_r^{-1}V_{rf}(V_{fr}V_r^{-1}V_{rf})^{-1}. \tag{23}$$

Let $X_{\tilde{g}} = [1_N,\ \beta_{\tilde{g}}]$. The pseudo parameters for risk premia in the two-pass CSR (with weighting matrix W) are then given by

$$\gamma_{\tilde{g}} = \begin{bmatrix} \gamma_{\tilde{g}0} \\ \gamma_{\tilde{g}1} \end{bmatrix} = (X_{\tilde{g}}^{'}WX_{\tilde{g}})^{-1}X_{\tilde{g}}^{'}W\mu_R. \tag{24}$$

We provide the asymptotic distribution of the estimator for $\gamma_{\tilde{g}}$, when the feasible version of $\tilde{g}_t$, $\hat{\tilde{g}}_t = \hat{V}_{fr}\hat{V}_r^{-1}r_t$, is used in the two-pass CSR.

Theorem 4. *Let $\hat{\beta}_{\hat{\tilde{g}}} = \hat{V}_{Rr}\hat{V}_r^{-1}\hat{V}_{rf}(\hat{V}_{fr}\hat{V}_r^{-1}\hat{V}_{rf})^{-1}$, $\hat{X}_{\hat{\tilde{g}}} = [1_N,\ \hat{\beta}_{\hat{\tilde{g}}}]$, W is a positive definite weighting matrix.*

(i) If W is known, the asymptotic distribution of $\hat{\gamma}_{\hat{\tilde{g}}} = (\hat{X}_{\hat{\tilde{g}}}^{'}W\hat{X}_{\hat{\tilde{g}}})^{-1}\hat{X}_{\hat{\tilde{g}}}^{'}W\hat{\mu}_R$ is given by $\sqrt{T}(\hat{\gamma}_{\hat{\tilde{g}}} - \gamma_{\tilde{g}}) \overset{A}{\sim} N(0, V_{\hat{\gamma}_{\hat{\tilde{g}}}})$, where $V_{\hat{\gamma}_{\hat{\tilde{g}}}} = \sum_{j=-\infty}^{\infty}\mathbb{E}[h_{\hat{\tilde{g}},t}h_{\hat{\tilde{g}},t+j}^{'}]$ with

$$h_{\hat{\tilde{g}},t} = (\gamma_{\tilde{g},t} - \gamma_{\tilde{g}}) - (\phi_{\tilde{g},t} - \phi_{\tilde{g}})w_{\tilde{g},t} + H_{\tilde{g}}z_{\tilde{g},t} + \tilde{\delta}_t \tag{25}$$

(ii) For the feasible GLS case with $W = V_R^{-1}$, which is estimated by $\hat{V}_R^{-1}$, the asymptotic distribution of $\hat{\gamma}_{\hat{\tilde{g}}} = (\hat{X}_{\hat{\tilde{g}}}^{'}\hat{V}_R^{-1}\hat{X}_{\hat{\tilde{g}}})^{-1}\hat{X}_{\hat{\tilde{g}}}^{'}\hat{V}_R^{-1}\hat{\mu}_R$ is given by $\sqrt{T}(\hat{\gamma}_{\hat{\tilde{g}}} - \gamma_{\tilde{g}}) \overset{A}{\sim} N(0, V_{\hat{\gamma}_{\hat{\tilde{g}}}})$, where $V_{\hat{\gamma}_{\hat{\tilde{g}}}} = \sum_{j=-\infty}^{\infty}\mathbb{E}[h_{\hat{\tilde{g}},t}h_{\hat{\tilde{g}},t+j}^{'}]$ with

$$h_{\hat{\tilde{g}},t} = (\gamma_{\tilde{g},t} - \gamma_{\tilde{g}}) - (\phi_{\tilde{g},t} - \phi_{\tilde{g}})w_{\tilde{g},t} + H_{\tilde{g}}z_{\tilde{g},t} - (\gamma_{\tilde{g},t} - \gamma_{\tilde{g}})u_{\tilde{g},t} + \tilde{\delta}_t \tag{26}$$

For both (i) and (ii), $\gamma_{\tilde{g},t} = [\gamma_{\tilde{g}0,t}, \gamma_{\tilde{g}1,t}^{'}]' = A_{\tilde{g}}R_t$, $\gamma_{\tilde{g}} = [\gamma_{\tilde{g}0},\ \gamma_{\tilde{g}1}^{'}]' = A_{\tilde{g}}\mu_R$, $A_{\tilde{g}} = H_{\tilde{g}}X_{\tilde{g}}^{'}W$, $H_{\tilde{g}} = (X_{\tilde{g}}^{'}WX_{\tilde{g}})^{-1}$, $X_{\tilde{g}} = [1_N,\ \beta_{\tilde{g}}]$, $\beta_{\tilde{g}} = V_{Rr}V_r^{-1}V_{rf}(V_{fr}V_r^{-1}V_{rf})^{-1}$, $\phi_{\tilde{g},t} = [\gamma_{\tilde{g}0,t},\ (\gamma_{\tilde{g}1,t} - \tilde{g}_t)']'$, $\phi_{\tilde{g}} = [\gamma_{\tilde{g}0}, (\gamma_{\tilde{g}1} - \mu_{\tilde{g}})']'$, $w_{\tilde{g},t} = \gamma_{\tilde{g}1}^{'}V_{\tilde{g}}^{-1}(\tilde{g}_t - \mu_{\tilde{g}})$, $u_{\tilde{g},t} = e_{\tilde{g}}^{'}W(R_t - \mu_R)$, $e_{\tilde{g}} = \mu_R - X_{\tilde{g}}\gamma_{\tilde{g}}$,

$$z_{\tilde{g},t} = [0,\ u_{\tilde{g},t}(\tilde{g}_t - \mu_{\tilde{g}})'V_{\tilde{g}}^{-1}]',\ and\ \tilde{\delta}_t = -\left\{A_{\tilde{g}}V_{Rr}V_r^{-1}(r_t - \mu_r) - \begin{bmatrix} 0 \\ \tilde{g}_t - \mu_{\tilde{g}} \end{bmatrix}\right\} r_{\tilde{g}1}^{'}V_{\tilde{g}}^{-1}\tilde{\eta}_t + H_{\tilde{g}}\begin{bmatrix} 0 \\ V_{\tilde{g}}^{-1}\tilde{\eta}_t \end{bmatrix} e_{\tilde{g}}^{'}WV_{Rr}V_r^{-1}(r_t - \mu_r) + \begin{bmatrix} 0 \\ \tilde{\eta}_t \end{bmatrix} w_{\tilde{g},t},\ with\ \tilde{\eta}_t = (f_t - \mu_f) - (\tilde{g}_t - \mu_{\tilde{g}}).$$

It is easy to verify that when $r_t = R_t$ (i.e., benchmark assets are also used as test assets), $\tilde{\delta}_t$ reduces to δ_t in Corollary 2.1, and Theorem 4 and Theorem 2 coincide. Put differently, Theorem 4 nests Theorem 2, by allowing test assets and benchmark assets to be different.

Furthermore, when $r_t = R_t$ and $\tilde{\eta}_t = 0_K$ (i.e., estimation error of mimicking portfolios is negligible), $\tilde{\delta}_t$ reduces to zero, and Theorem 4 coincides with (iii)(iv) of Theorem 1 under maximum correlation portfolios.

3 Simulation and Application

3.1 Simulation Evidence

To investigate the finite sample performance of the asymptotic results in Section 2, we conduct a simulation experiment.

3.1.1 D.G.P.

In the data generation process (D.G.P.), $Y_t = (f_t', R_t')'$, $t = 1, ..., T$, is independently drawn from a multivariate normal distribution, whose mean and variance are calibrated from real data sets as follows.

For returns, we set $N = 25$, and employ the commonly used 25 size and book-to-market sorted portfolios for the calibration of mean and variance. For factors, we set $K = 3$ and consider two three-factor models for comparison. One model (denoted by *M-S* in Table 1 and 2) consists of three statistically strong factors whose mean and variances are calibrated from Fama and French (1993) factors. In contrast, the other three-factor model (denoted by M-W) consists of factors that are only weakly correlated with returns, and we calibrate the mean and variance for these factors from the data set of Lettau and Ludvigson (2001), where consumption wealth ratio, consumption growth and the interaction of these two serve as the three factors. The covariance of factors and returns, which reflects whether factors and returns are strongly or weakly correlated, is thus calibrated from the data for factors and returns described above, using the same time period as in Lettau and Ludvigson (2001).

Note that in the above D.G.P., the model specification error is not necessarily equal to zero. In order to explore the case that this error is zero so there is no model misspecification, we also consider the setup that $\mu_R = \hat{X}\hat{\gamma}$ in D.G.P., where $\hat{X}$ and $\hat{\gamma}$ result from estimation using the data of Lettau and Ludvigson (2001).[5]

After the data is generated, we regress the simulated factors on the simulated returns to derive the weights of mimicking portfolios. By replacing factors with constructed mimicking portfolios in the two models above, we compute risk premia and the associated standard errors by both OLS ($W = I$) and GLS ($W = V_R^{-1}$) in the Fama and MacBeth (1973) two-pass procedure. The number of Monte Carlo replications we use is 2000, and the reported standard errors in Table 1 for OLS and Table 2 for GLS result from these replications. In both tables, we consider $T = 500$, which is close to the typical sample size in practice, as well as $T = 2000$ to facilitate comparison. Finally, the two cases depending on model misspecification exists or not are both presented.

3.1.2 Results

We present the OLS outcome in Table 1 and the GLS result in Table 2. In each table, we report three types of standard errors of risk premia. In the column of $var(\hat{\gamma}_{\hat{g}})^{1/2}$, we report the standard error of the risk premium estimator using constructed mimicking portfolios: we compute the estimator 2000 times and report the sample standard error. Since these reported numbers result from a large number of replications, they can be considered as real

[5]See also Kan *et al.* (2013) for the similar treatment.

standard errors of risk premia. Under $\widehat{var}(\hat{\gamma}_{\hat{g}})^{1/2}$, we report two more standard errors by the asymptotic theory in Section 2: in particular, in the column of $h_{g,t}$, the estimation error of mimicking portfolios is ignored; under $h_{\hat{g},t}$, the estimation error is accounted for. Note that these notations are also used in Section 2. The reported numbers under $h_{g,t}$ and $h_{\hat{g},t}$ result from the average of 2000 replications.

When factors are only weakly correlated with returns (Model M-W), Table 1 shows that ignoring the estimation error in mimicking portfolios (columns of $h_{g,t}$) could severely underestimate the standard errors of risk premia by OLS. In contrast, the standard error that accounts for estimation error in mimicking portfolios (columns of $h_{\hat{g},t}$) provides good approximations to actual standard errors.

Nevertheless, when factors are strongly correlated with returns (Model M-S), ignoring the estimation error in mimicking portfolios does not appear to severely underestimate the standard errors of risk premia, as standard errors corresponding to $h_{g,t}$ and $h_{\hat{g},t}$ are close to each other in Table 1.

The above findings remain unaffected, no matter whether model misspecification exists or not. In particular, when there is no model misspecification but factors are minorly correlated with returns, Table 1 still suggests that taking account of estimating error of mimicking portfolios can cause substantial differences in standard errors of risk premia.

Overall, the findings in Table 1 are consistent with our analytical results in Section 2. These findings also remain qualitatively similar, when GLS is used as the estimation method for risk premia in Table 2.

3.2 Empirical Application

To illustrate how the adjustment for estimation errors on the weights of mimicking portfolios affects the outcome of asset pricing tests in CSR, we consider three empirical examples, which are adopted from Cochrane (1996), Li *et al.* (2006) and Muir *et al.* (2013) respectively.

The factors considered in our empirical application include the residential investment growth $\triangle I_{Res}$ and nonresidential investment growth $\triangle I_{Nres}$ in Cochrane (1996), the investment growth rate in the financial cooperations $Finan$, the nonfinancial corporate business $Nfinco$, and the household sector $Hholds$ in Li *et al.* (2006), the funding liquidity Lev in Muir *et al.* (2013).

To construct mimicking portfolios, we regress the factors above on the 25 Fama-French size and book-to-market sorted portfolios using the time period 1973Q1-2009Q4, during which both factors and returns have quarterly data available.

With the constructed mimicking portfolios, we estimate risk premia by the two-pass procedure of Fama and MacBeth (1973) and compute the associated standard errors. The OLS and GLS results are reported in Table 3 and 4, respectively. In particular, we compute four standard errors corresponding to the methods in Fama and MacBeth (1973), Shanken (1992), Jagannathan and Wang (1996) and Kan *et al.* (2013). Finally, we also compute two standard errors according to our proposed approach that accounts for errors-in-weights (EIW): for EIW (c) , we assume the model is correctly specified; in contrast, for EIW (m), we allow for model misspecification.

As shown by Table 3, once we account for the estimation error of mimicking portfolios, the outcome of t-test in CSR for risk premia can be substantially altered. For example, in both Cochrane (1996) and Li *et al.* (2006), we find the evidence that risk premium that was significant at 5% when ignoring estimation error becomes insignificant, after this error is accounted for. Similarly in Muir *et al.* (2013), significance at 1% reduces to 5%, if we take estimation error of mimicking portfolios into consideration.

Overall, it is commonplace in Tables 3 and 4 that the standard error associated with the estimate of risk premium increases by a big margin (i.e., in most cases, around 50% $\sim$ 100%), once estimation error of mimicking portfolios as well as model misspecification are accounted for. Therefore, estimation errors on the weights of the mimicking portfolios are of first order importance and should not be ignored. For improved statistical inferences, we recommend researchers to adopt our new standard error to compute the standard errors for estimated risk premia associated with mimicking portfolios.

4 Conclusion

This paper targets a common practice in asset pricing studies: using mimicking portfolios constructed from original risk factors for asset pricing tests without taking into account the estimation errors on their weights. We suggest a methodology for computing the standard error of the estimated risk premia, which takes into account the estimation errors of the mimicking portfolios as well as potential model misspecification. Our empirical evidence shows that the suggested adjustment on standard error is statistically relevant (in most of the cases that we study, the increase in the standard error is more than 50%). In many cases, the estimated risk premium associated with the mimicking portfolios become statistically insignificant once we take into account the estimation error associated with mimicking portfolios. Consequently, our findings cast doubt on the reliability of existing empirical asset pricing studies that rely on the use of mimicking portfolios but fail to take into account of the impact of the estimation errors on the mimicking portfolios.

Appendix

Proofs as well as some additional results are contained in the appendix. Omitted proofs for the results in the paper can be derived straightforwardly.

A: Preliminary

In order to simplify the derivation in the appendix, the result and notation below will be used.

1. Define $\psi = \begin{bmatrix} \mu \\ vec(V) \end{bmatrix}$, $\hat{\psi} = \begin{bmatrix} \hat{\mu} \\ vec(\hat{V}) \end{bmatrix}$, then by the conventional GMM results under just-identification:

$$\sqrt{T}(\hat{\psi} - \psi) \overset{A}{\sim} N(0_{(N+K)\times(N+K+1)}, \sum_{j=-\infty}^{\infty} \mathbb{E}(\zeta_t \zeta'_{t+j}))$$

where

$$\zeta_t = \begin{bmatrix} Y_t - \mu \\ vec((Y_t - \mu)(Y_t - \mu)' - V) \end{bmatrix}$$

2. Define the matrix C and its inverse C^{-1}:

$$C = \begin{bmatrix} 1 & 0'_K \\ 0_K & V_g L^{-1} V_f^{-1} \end{bmatrix}, \; C^{-1} = \begin{bmatrix} 1 & 0'_K \\ 0_K & V_f L V_g^{-1} \end{bmatrix}$$

In particular, when $L = I_K$:

$$C = \begin{bmatrix} 1 & 0'_K \\ 0_K & V_g V_f^{-1} \end{bmatrix}, \; C^{-1} = \begin{bmatrix} 1 & 0'_K \\ 0_K & V_f V_g^{-1} \end{bmatrix}$$

B: Proof for Theorem 1

Proof. We consider two cases, depending on W is given or not.

W is known

Since $\hat{\gamma}_g = (\hat{X}'_g W \hat{X}_g)^{-1} \hat{X}'_g W \hat{\mu}_R$ is a smooth function of $\hat{\mu}$ and $\hat{V}$, we only need to derive $\frac{\partial \gamma_g}{\partial \psi'}$ and then apply the Delta method for $h_{g,t}$. That is,

$$h_{g,t} = \frac{\partial \gamma_g}{\partial \psi'} \zeta_t$$

with

$$\frac{\partial \gamma_g}{\partial \psi'} = \begin{pmatrix} \frac{\partial \gamma_g}{\partial \mu'_f} & \frac{\partial \gamma_g}{\partial \mu'_R} & \frac{\partial \gamma_g}{\partial vec(V)'} \end{pmatrix}$$

where $\frac{\partial \gamma_g}{\partial \mu'_f} = 0_{(K+1)\times K}$, $\frac{\partial \gamma_g}{\partial \mu'_R} = A_g = (\hat{X}'_g W \hat{X}_g)^{-1} \hat{X}'_g W$. So only $\frac{\partial \gamma_g}{\partial vec(V)'}$ needs to be derived.

Given $\gamma_g = (X_g' W X_g)^{-1} X_g' W \mu_R = H_g X_g' W \mu_R$, we use the product rule:

$$\frac{\partial \gamma_g}{\partial vec(V)'} = (\mu_R' W X_g \otimes I_{K+1}) \frac{\partial vec(H_g)}{\partial vec(V)'} + (\mu_R' W \otimes H_g) \frac{\partial vec(X_g')}{\partial vec(V)'}$$

where

$$\begin{aligned}
(\mu_R' W X_g \otimes I_{K+1}) \frac{\partial vec(H_g)}{\partial vec(V)'} &= -(\mu_R' W X_g \otimes I_{K+1})(H_g \otimes H_g) \frac{\partial vec(H_g^{-1})}{\partial vec(V)'} \\
&= -(\gamma_g' \otimes H_g) \left[(X_g' W \otimes I_{K+1}) \frac{\partial vec(X_g')}{\partial vec(V)'} + (I_{K+1} \otimes X_g' W) \frac{\partial vec(X_g)}{\partial vec(V)'} \right] \\
&= -(H_g \otimes \gamma_g' X_g' W) \frac{\partial vec(X_g)}{\partial vec(V)'} - (\gamma_g' \otimes H_g X_g' W) \frac{\partial vec(X_g)}{\partial vec(V)'} \\
(\mu_R' W \otimes H_g) \frac{\partial vec(X_g')}{\partial vec(V)'} &= (H_g \otimes \mu_R' W) \frac{\partial vec(X_g)}{\partial vec(V)'}
\end{aligned}$$

Combining the pieces above, with $e_g = \mu_R - X_g \gamma_g$:

$$\frac{\partial \gamma_g}{\partial vec(V)'} = (H_g \otimes e_g' W) \frac{\partial vec(X_g)}{\partial vec(V)'} - (\gamma_g' \otimes H_g X_g' W) \frac{\partial vec(X_g)}{\partial vec(V)'}$$

Note that

$$\begin{aligned}
\frac{\partial vec(X_g)}{\partial vec(V)'} &= \frac{\partial vec(X_g)}{\partial vec(\beta_g)'} \frac{\partial vec(\beta_g)}{\partial vec(V)'} \\
&= ([0_K, I_K]' \otimes I_N) \frac{\partial vec(\beta_g)}{\partial vec(V)'}
\end{aligned}$$

Hence what remains to be derived is $\frac{\partial vec(\beta_g)}{\partial vec(V)'}$.

In order to derive $\frac{\partial vec(\beta_g)}{\partial vec(V)'}$, and notice that $\beta_g = V_R M (M' V_R M)^{-1}$, and

$$V_R = [0_{N\times K}, I_N] V [0_{N\times K}, I_N]', \;\; \frac{\partial vec(V_R)}{\partial vec(V)'} = [0_{N\times K}, I_N] \otimes [0_{N\times K}, I_N]$$

Given these results, we apply the product rule to $\frac{\partial vec(\beta_g)}{\partial vec(V)'}$:

$$\frac{\partial vec(\beta_g)}{\partial vec(V)'} = [(M' V_R M)^{-1} M' \otimes I_N] \frac{\partial vec(V_R)}{\partial vec(V)'} + (I_K \otimes V_R M) \frac{\partial vec((M' V_R M)^{-1})}{\partial vec(V)'}$$

where

$$\begin{aligned}
[(M' V_R M)^{-1} M' \otimes I_N] \frac{\partial vec(V_R)}{\partial vec(V)'} &= [(M' V_R M)^{-1} M' \otimes I_N]([0_{N\times K}, I_N] \otimes [0_{N\times K}, I_N]) \\
&= [0_{K\times K}, (M' V_R M)^{-1} M'] \otimes [0_{N\times K}, I_N]
\end{aligned}$$

and

$$
\begin{aligned}
&(I_K \otimes V_R M)\frac{\partial vec((M'V_R M)^{-1})}{\partial vec(V)'} \\
=& -(I_K \otimes V_R M)[(M'V_R M)^{-1} \otimes (M'V_R M)^{-1}]\frac{\partial vec(M'V_R M)}{\partial vec(V)'} \\
=& -[(M'V_R M)^{-1} \otimes \beta_g]\frac{\partial vec(M'V_R M)}{\partial vec(V)'} \\
=& -[(M'V_R M)^{-1} \otimes \beta_g](M' \otimes M')\frac{\partial vec(V_R)}{\partial vec(V)'} \\
=& -[0_{K\times K}, (M'V_R M)^{-1}M'] \otimes [0_{N\times K}, \beta_g M']
\end{aligned}
$$

Combining these pieces, we get:

$$
\frac{\partial vec(\beta_g)}{\partial vec(V)'} = [0_{K\times K}, (M'V_R M)^{-1}M'] \otimes [0_{N\times K}, I_N - \beta_g M']
$$

and

$$
\begin{aligned}
\frac{\partial vec(X_g)}{\partial vec(V)'} &= \frac{\partial vec(X_g)}{\partial vec(\beta_g)'}\frac{\partial vec(\beta_g)}{\partial vec(V)'} \\
&= ([0_K, I_K]' \otimes I_N)([0_{K\times K}, (M'V_R M)^{-1}M'] \otimes [0_{N\times K}, I_N - \beta_g M']) \\
&= [0_K, I_K]'[0_{K\times K}, (M'V_R M)^{-1}M'] \otimes [0_{N\times K}, I_N - \beta_g M']
\end{aligned}
$$

Recall that:

$$
\frac{\partial \gamma_g}{\partial vec(V)'} = (H_g \otimes e_g' W)\frac{\partial vec(X_g)}{\partial vec(V)'} - (\gamma_g' \otimes A_g)\frac{\partial vec(X_g)}{\partial vec(V)'}
$$

where $A_g = H_g X_g' W$.

For $(H_g \otimes e_g' W)\frac{\partial vec(X_g)}{\partial vec(V)'}$, it can be simplified by using $\beta_g' W e_g = 0$:

$$
(H_g \otimes e_g' W)\frac{\partial vec(X_g)}{\partial vec(V)'} = \left[0_{(K+1)\times K}, H_g[0_N, M(M'V_R M)^{-1}]'\right] \otimes [0_K', e_g' W]
$$

For $(\gamma_g' \otimes A_g)\frac{\partial vec(X_g)}{\partial vec(V)'}$:

$$
(\gamma_g' \otimes A_g)\frac{\partial vec(X_g)}{\partial vec(V)'} = \left[0_K', \gamma_g'[0_N, M(M'V_R M)^{-1}]'\right] \otimes [0_{(K+1)\times K}, A_g - A_g\beta_g M']
$$

Combining these pieces, $\frac{\partial \gamma_g}{\partial vec(V)'}$ thus has two terms:

$$
\begin{aligned}
\frac{\partial \gamma_g}{\partial vec(V)'} =& \left[0_{(K+1)\times K}, H_g[0_N, M(M'V_R M)^{-1}]'\right] \otimes [0_K', e_g' W] \\
&- \left[0_K', \gamma_g'[0_N, M(M'V_R M)^{-1}]'\right] \otimes [0_{(K+1)\times K}, A_g - A_g\beta_g M']
\end{aligned}
$$

With $\frac{\partial \gamma_g}{\partial vec(V)'}$ above, we derive $h_{g,t}$ as follows.

$$
\begin{aligned}
h_{g,t} &= \frac{\partial \gamma_g}{\partial \psi'}\zeta_t \\
&= \begin{pmatrix} 0_{(K+1)\times K} & A_g & \frac{\partial \gamma_g}{\partial vec(V)'} \end{pmatrix} \begin{pmatrix} Y_t - \mu \\ vec((Y_t-\mu)(Y_t-\mu)' - V) \end{pmatrix} \\
&= A_g(R_t - \mu_R) + \frac{\partial \gamma_g}{\partial vec(V)'} vec((Y_t-\mu)(Y_t-\mu)' - V)
\end{aligned}
$$

By plugging in $\frac{\partial \gamma_g}{\partial vec(V)'}$, we end up with that $h_{g,t}$ consists of the terms below.

The term $A_g(R_t - \mu_R)$ can be rewritten as:

$$
A_g(R_t - \mu_R) = \gamma_{g,t} - \gamma_g
$$

For the pieces caused by the two terms in $\frac{\partial \gamma_g}{\partial vec(V)'}$, we report them one by one.

First of all, $\left[0_{(K+1)\times K}, H_g[0_N, M(M'V_RM)^{-1}]'\right] \otimes [0'_K, e'_gW]vec((Y_t-\mu)(Y_t-\mu)' - V)$ can be rewritten as:

$$
\begin{aligned}
&\left[0_{(K+1)\times K}, H_g[0_N, M(M'V_RM)^{-1}]'\right] \otimes [0'_K, e'_gW]vec((Y_t-\mu)(Y_t-\mu)' - V) \\
=&vec([0'_K, e'_gW]\left((Y_t-\mu)(Y_t-\mu)' - V\right) \begin{bmatrix} 0_{K\times(K+1)} \\ (H_g[0_N, M(M'V_RM)^{-1}]')' \end{bmatrix}) \\
=&vec(e'_gW((R_t-\mu_R)(R_t-\mu_R)' - V_R)[0_N, M(M'V_RM)^{-1}]H'_g) \\
=&H_g[0_N, M(M'V_RM)^{-1}]'((R_t-\mu_R)(R_t-\mu_R)' - V_R)We_g \\
=&H_g[0_N, M(M'V_RM)^{-1}]'(R_t-\mu_R)u_{g,t} \\
=&H_gz_{g,t}
\end{aligned}
$$

where $u_{g,t} = e'_gW(R_t - \mu_R)$, $z_{g,t} = [0, u_{g,t}(g_t - \mu_g)'(M'V_RM)^{-1}]'$.

Secondly, $\left[0'_K, \gamma'_g[0_N, M(M'V_RM)^{-1}]'\right] \otimes [0_{(K+1)\times K}, A_g - A_g\beta_gM']vec((Y_t-\mu)(Y_t-\mu)' - V)$ can be rewritten as:

$$
\begin{aligned}
&\left[0'_K, \gamma'_g[0_N, M(M'V_RM)^{-1}]'\right] \otimes [0_{(K+1)\times K}, A_g - A_g\beta_gM']vec((Y_t-\mu)(Y_t-\mu)' - V) \\
=&[0_{(K+1)\times K}, A_g - A_g\beta_gM]((Y_t-\mu)(Y_t-\mu)' - V) \begin{pmatrix} 0_K \\ (\gamma'_g[0_N, M(M'V_RM)^{-1}]')' \end{pmatrix} \\
=&(A_g - A_g\beta_gM')[(R_t-\mu_R)(R_t-\mu_R)' - V_R]M(M'V_RM)^{-1}\gamma_{g1} \\
=&(\gamma_{g,t} - \gamma_g)(g_t-\mu_g)'(M'V_RM)^{-1}\gamma_{g1} - A_g\beta_g(g_t - \mu_g)w_{g,t} \\
=&(\gamma_{g,t} - \gamma_g)w_{g,t} - A_g\beta_g(g_t - \mu_g)w_{g,t} \\
=&(\phi_{g,t} - \phi_g)w_{g,t}
\end{aligned}
$$

where $w_{g,t} = (g_t - \mu_g)'(M'V_RM)^{-1}\gamma_{g1}$, $\phi_{g,t} = [\gamma_{g0,t}, (\gamma_{g1,t} - g_t)']'$, $\phi_g = [\gamma_{g0}, (\gamma_{g1} - \mu_g)']'$.

As a result, $h_{g,t}$ consists of three terms:

$$
h_{g,t} =(\gamma_{g,t} - \gamma_g) - (\phi_{g,t} - \phi_g)w_{g,t} + H_gz_{g,t}
$$

W is unknown

Since $\hat{\gamma}_g = (\hat{X}_g'\hat{V}_R^{-1}\hat{X}_g)^{-1}\hat{X}_g'\hat{V}_R^{-1}\hat{\mu}_R$ is also a smooth function of $\hat{\mu}$ and $\hat{V}$, we only need to derive $\frac{\partial \gamma_g}{\partial \psi'}$. Given $\gamma_g = (X_g'V_R^{-1}X_g)^{-1}X_g'V_R^{-1}\mu_R = H_g X_g'V_R^{-1}\mu_R$, we use the product rule:

$$\frac{\partial \gamma_g}{\partial vec(V)'} = (\mu_R'V_R^{-1}X_g \otimes I_{K+1})\frac{\partial vec(H_g)}{\partial vec(V)'} + (\mu_R'V_R^{-1} \otimes H_g)\frac{\partial vec(X_g')}{\partial vec(V)'} + (\mu_R' \otimes H_g X_g')\frac{\partial vec(V_R^{-1})}{\partial vec(V)'}$$

where

$$\begin{aligned}
&(\mu_R'V_R^{-1}X_g \otimes I_{K+1})\frac{\partial vec(H_g)}{\partial vec(V)'} \\
=& -(\mu_R'V_R^{-1}X_g \otimes I_{K+1})(H_g \otimes H_g)\frac{\partial vec(H_g^{-1})}{\partial vec(V)'} \\
=& -(\gamma_g' \otimes H_g)\left[(X_g'V_R^{-1} \otimes I_{K+1})\frac{\partial vec(X_g')}{\partial vec(V)'} + (X_g' \otimes X_g')\frac{\partial vec(V_R^{-1})}{\partial vec(V)'} + (I_{K+1} \otimes X_g'V_R^{-1})\frac{\partial vec(X_g)}{\partial vec(V)'}\right] \\
=& -(H_g \otimes \gamma_g'X_g'V_R^{-1})\frac{\partial vec(X_g)}{\partial vec(V)'} + (\gamma_g'X_g'V_R^{-1} \otimes H_g X_g'V_R^{-1})\frac{\partial vec(V_R)}{\partial vec(V)'} - (\gamma_g' \otimes H_g X_g'V_R^{-1})\frac{\partial vec(X_g)}{\partial vec(V)'}
\end{aligned}$$

and

$$\begin{aligned}
(\mu_R'V_R^{-1} \otimes H_g)\frac{\partial vec(X_g')}{\partial vec(V)'} &= (H_g \otimes \mu_R'V_R^{-1})\frac{\partial vec(X_g)}{\partial vec(V)'} \\
(\mu_R' \otimes H_g X_g')\frac{\partial vec(V_R^{-1})}{\partial vec(V)'} &= -(\mu_R'V_R^{-1} \otimes H_g X_g'V_R^{-1})\frac{\partial vec(V_R)}{\partial vec(V)'}
\end{aligned}$$

Combining the pieces above, with $e_g = \mu_R - X_g\gamma_g$:

$$\frac{\partial \gamma_g}{\partial vec(V)'} = (H_g \otimes e_g'V_R^{-1})\frac{\partial vec(X_g)}{\partial vec(V)'} - (\gamma_g' \otimes H_g X_g'V_R^{-1})\frac{\partial vec(X_g)}{\partial vec(V)'} - (e_g'V_R^{-1} \otimes H_g X_g'V_R^{-1})\frac{\partial vec(V_R)}{\partial vec(V)'}$$

Compared to the case W is given, the difference is that we have the third term.

$$-(e_g'V_R^{-1} \otimes A_g)\frac{\partial vec(V_R)}{\partial vec(V)'} = -[0_K', e_g'V_R^{-1}] \otimes [0_{(K+1)\times K}, A_g]$$

This further induces the extra term $-(\gamma_{g,t} - \gamma_g)u_{g,t}$ for $h_{g,t}$:

$$\begin{aligned}
&-[0_K', e_g'V_R^{-1}] \otimes [0_{(K+1)\times K}, A_g]vec((Y_t - \mu)(Y_t - \mu)' - V) \\
=& -[0_{(K+1)\times K}, A_g]((Y_t - \mu)(Y_t - \mu)' - V)\begin{pmatrix} 0_K \\ V_R^{-1}e_g \end{pmatrix} \\
=& -A_g[(R_t - \mu_R)(R_t - \mu_R)' - V_R]V_R^{-1}e_g \\
=& -(\gamma_{g,t} - \gamma_g)u_{g,t}
\end{aligned}$$

□

C: Proof for Theorem 2

Proof. We can show that (under $L = I_K$, see (18))

$$\gamma_g = \begin{bmatrix} \gamma_0 \\ V_g V_f^{-1} \gamma_1 \end{bmatrix} = C\gamma, \qquad \hat{\gamma}_g = \begin{bmatrix} \hat{\gamma}_0 \\ \hat{V}_g \hat{V}_f^{-1} \hat{\gamma}_1 \end{bmatrix}.$$

We already know the explicit expression of

$$h_t \equiv \frac{\partial \gamma}{\partial \varphi'} \zeta_t.$$

We are interested in obtaining

$$\begin{aligned} h_{\hat{g},t} &= \frac{\partial \gamma_g}{\partial \varphi'} \zeta_t \\ &= C \frac{\partial \gamma}{\partial \varphi'} \zeta_t + \begin{bmatrix} 0'_{(N+K)+(N+K)^2} \\ (\gamma_1' \otimes I_K) \frac{\partial \mathrm{vec}(V_g V_f^{-1})}{\partial \varphi'} \end{bmatrix} \zeta_t \\ &= C h_t + \begin{bmatrix} 0 \\ (\gamma_1' \otimes I_K) \frac{\partial \mathrm{vec}(V_g V_f^{-1})}{\partial \mathrm{vec}(V)'} \mathrm{vec}((Y_t - \mu)(Y_t - \mu)') \end{bmatrix}. \end{aligned}$$

where

$$\begin{aligned} (\gamma_1' \otimes I_K) \frac{\partial \mathrm{vec}(V_g V_f^{-1})}{\partial \mathrm{vec}(V)'} &= (\gamma_1' V_f^{-1} \otimes I_K) \frac{\partial \mathrm{vec}(V_g)}{\partial \mathrm{vec}(V)'} - (\gamma_1' V_f^{-1} \otimes V_g V_f^{-1}) \frac{\partial \mathrm{vec}(V_f)}{\partial \mathrm{vec}(V)'} \\ &= (\gamma_1' V_f^{-1} \otimes I_K) \big([0_{K\times K},\ V_{fR} V_R^{-1}] \otimes [I_K,\ 0_{K\times N}] \\ &\quad + [I_K,\ -V_{fR} V_R^{-1}] \otimes [0_{K\times K},\ V_{fR} V_R^{-1}] \big) \\ &\quad - (\gamma_1' V_f^{-1} \otimes V_g V_f^{-1}) ([I_K,\ 0_{K\times N}] \otimes [I_K,\ 0_{K\times N}]). \end{aligned}$$

It follows that

$$h_{\hat{g},t} = C h_t + \begin{bmatrix} 0 \\ (f_t - \mu_f)(g_t - \mu_g)' V_f^{-1} \gamma_1 + (g_t - \mu_g) \eta_t' V_f^{-1} \gamma_1 - V_g V_f^{-1} (f_t - \mu_f)(f_t - \mu_f)' V_f^{-1} \gamma_1 \end{bmatrix}$$

After simplification, we can also write this as

$$h_{\hat{g},t} = C h_t + \begin{bmatrix} 0 \\ [V_\eta V_f^{-1} (f_t - \mu_f)(f_t - \mu_f)' - \eta_t \eta_t'] V_f^{-1} \gamma_1 \end{bmatrix}.$$

□

D: Proof for Corollary 2.1

Proof. For convenience, define:

$$\triangle_t \equiv \begin{bmatrix} 0 \\ \left[V_\eta V_f^{-1}(f_t-\mu_f)(f_t-\mu_f)' - \eta_t\eta_t'\right] V_f^{-1}\gamma_1 \end{bmatrix}$$

With $\triangle_t$ above, h_t and $h_{\hat{g},t}$ are related as follows (W is given):

$$\begin{aligned} h_{\hat{g},t} &= Ch_t + \triangle_t \\ &= C(\gamma_t-\gamma) - C(\phi_t-\phi)w_t + CHz_t + \triangle_t \\ &= (\gamma_{g,t}-\gamma_g) - (\phi_{g,t}-\phi_g)w_t - \begin{bmatrix} 0 \\ g_t-\mu_g \end{bmatrix} w_t + C\begin{bmatrix} 0 \\ f_t-\mu_f \end{bmatrix} w_t + H_g C^{-1'} z_t + \triangle_t \\ &= h_{g,t} - (\phi_{g,t}-\phi_g)\eta_t' V_f^{-1}\gamma_1 + H_g \begin{bmatrix} 0 \\ u_{g,t}V_g^{-1}\eta_t \end{bmatrix} + \triangle_t - \begin{bmatrix} 0 \\ g_t-\mu_g \end{bmatrix} w_t + C\begin{bmatrix} 0 \\ f_t-\mu_f \end{bmatrix} w_t \end{aligned}$$

Plug in the expression of $\triangle_t$ to get

$$\triangle_t - \begin{bmatrix} 0 \\ g_t-\mu_g \end{bmatrix} w_t + C\begin{bmatrix} 0 \\ f_t-\mu_f \end{bmatrix} w_t = \begin{bmatrix} 0 \\ \eta_t \end{bmatrix} \gamma_{g1}' V_g^{-1}(g_t-\mu_g)$$

As a result:

$$h_{\hat{g},t} = h_{g,t} - (\phi_{g,t}-\phi_g)\eta_t' V_f^{-1}\gamma_1 + H_g \begin{bmatrix} 0 \\ u_{g,t}V_g^{-1}\eta_t \end{bmatrix} + \begin{bmatrix} 0 \\ \eta_t \end{bmatrix} \gamma_{g1}' V_g^{-1}(g_t-\mu_g)$$

and note that $\gamma_{g1}'V_g^{-1} = \gamma_1'V_f^{-1}$, so we may also write:

$$h_{\hat{g},t} = h_{g,t} - (\phi_{g,t}-\phi_g)\eta_t' V_g^{-1}\gamma_{g1} + H_g \begin{bmatrix} 0 \\ u_{g,t}V_g^{-1}\eta_t \end{bmatrix} + \begin{bmatrix} 0 \\ \eta_t \end{bmatrix} \gamma_{g1}' V_g^{-1}(g_t-\mu_g)$$

The proof for the feasible GLS case is similar and thus omitted. □

E: Proof for Theorem 3

Proof. We first show $\mathbb{E}(h_{g,t}\delta_t') = 0_{(K+1)\times(K+1)}$, where:

$$h_{g,t} = (\gamma_{g,t}-\gamma_g) - (\phi_{g,t}-\phi_g)w_{g,t} + H_g z_{g,t}$$

or

$$h_{g,t} = (\gamma_{g,t}-\gamma_g) - (\phi_{g,t}-\phi_g)w_{g,t} + H_g z_{g,t} - (\gamma_{g,t}-\gamma_g)u_{g,t}$$

and

$$\delta_t = -(\phi_{g,t}-\phi_g)\gamma_{g1}'V_g^{-1}\eta_t + H_g \begin{bmatrix} 0 \\ u_{g,t}V_g^{-1}\eta_t \end{bmatrix} + \begin{bmatrix} 0 \\ \eta_t \end{bmatrix} \gamma_{g1}' V_g^{-1}(g_t-\mu_g)$$

Note that $\gamma_{g,t}-\gamma_g$, $\phi_{g,t}-\phi_g$, $w_{g,t}$, $u_{g,t}$, $\gamma_{g1}'V_g^{-1}\eta_t$, η_t and $\gamma_{g1}'V_g^{-1}(g_t-\mu_g)$ are all jointly elliptically distributed with zero expectations.

Below we discuss the interaction of the terms in $h_{g,t}$ with δ_t, one by one.

$\underline{\gamma_{g,t} - \gamma_g}$: The interaction of $\gamma_{g,t} - \gamma_g$ and δ_t is made of third moments, thus has zero expectation.

$\underline{(\phi_{g,t} - \phi_g)w_{g,t}}$: Both $\phi_{g,t} - \phi_g$ and $w_{g,t}$ are uncorrelated with η_t. Thus the interaction of $(\phi_{g,t} - \phi_g)w_{g,t}$ and δ_t has zero expectation.

$\underline{H_g z_{g,t}}$: $z_{g,t}$ contains $u_{g,t}$ and $g_t - \mu_g$, both are uncorrelated with η_t. Thus the interaction of $H_g z_{g,t}$ and δ_t has zero expectation.

$\underline{(\gamma_{g,t} - \gamma_g)u_{g,t}}$: Both $\gamma_{g,t} - \gamma_g$ and $u_{g,t}$ contain $R_t - \mu_R$, which is uncorrelated with η_t. Thus the interaction of $(\gamma_{g,t} - \gamma_g)u_{g,t}$ and δ_t has zero expectation.

With the four pieces above, we have $\mathbb{E}(h_{g,t}\delta_t') = 0_{(K+1)\times(K+1)}$.

Recall that δ_t contains three parts

$$\delta_t = -(\phi_{g,t} - \phi_g)\gamma_{g1}' V_g^{-1}\eta_t + H_g \begin{bmatrix} 0 \\ u_{g,t}V_g^{-1}\eta_t \end{bmatrix} + \begin{bmatrix} 0 \\ \eta_t \end{bmatrix} \gamma_{g1}' V_g^{-1}(g_t - \mu_g)$$

By the property of the elliptical distribution, the interaction terms of the three parts in δ_t all have zero expectations, so $\mathbb{E}(\delta_t\delta_t')$ reduces to the sum of three variances below:

$$(1+\kappa)\left(A_g V_R A_g' - \begin{bmatrix} 0 & 0_K' \\ 0_K & V_g \end{bmatrix}\right) \cdot (\gamma_{g1}' V_g^{-1} V_\eta V_g^{-1}\gamma_{g1})$$

$$(1+\kappa)H_g \begin{bmatrix} 0 & 0_K' \\ 0_K & V_g^{-1}V_\eta V_g^{-1} \end{bmatrix} H_g \cdot (e'WV_RWe)$$

$$(1+\kappa)\begin{bmatrix} 0 & 0_K' \\ 0_K & V_\eta \end{bmatrix} \cdot (\gamma_{g1}' V_g^{-1}\gamma_{g1})$$

which correspond to the three parts in δ_t, respectively. □

F: Derivation of Equation (18)

Proof. β_g corresponding to g_t now reads:

$$\begin{aligned} \beta_g &= cov(R_t, g_t)var(g_t)^{-1} \\ &= V_R M(M'V_R M)^{-1} \\ &= \beta V_f L V_g^{-1} \end{aligned}$$

which implies that $X_g = [1_N,\ \beta_g] = [1_N,\ \beta]C^{-1} = XC^{-1}$.

The risk premium γ_g corresponding to β_g is thus:

$$\begin{aligned} \gamma_g &= (X_g' W X_g)^{-1} X_g' W \mu_R \\ &= C\gamma \\ &= \begin{bmatrix} \gamma_0 \\ V_g L^{-1} V_f^{-1}\gamma_1 \end{bmatrix} \end{aligned}$$

In addition, the pricing errors are given by $e_g = \mu_R - X_g\gamma_g = \mu_R - XC^{-1}C\gamma = e$. □

G: Proof for Theorem 4

Proof. Given $\gamma_{\tilde{g}} = (X_{\tilde{g}}' V_R^{-1} X_{\tilde{g}})^{-1} X_{\tilde{g}}' V_R^{-1} \mu_R = H_{\tilde{g}} X_{\tilde{g}}' V_R^{-1} \mu_R$, we use the product rule:

$$\frac{\partial \gamma_{\tilde{g}}}{\partial vec(V)'} = (\mu_R' V_R^{-1} X_{\tilde{g}} \otimes I_{K+1}) \frac{\partial vec(H_{\tilde{g}})}{\partial vec(V)'} + (\mu_R' V_R^{-1} \otimes H_{\tilde{g}}) \frac{\partial vec(X_{\tilde{g}}')}{\partial vec(V)'} + (\mu_R' \otimes H_{\tilde{g}} X_{\tilde{g}}') \frac{\partial vec(V_R^{-1})}{\partial vec(V)'}$$

where

$$\begin{aligned}
&(\mu_R' V_R^{-1} X_{\tilde{g}} \otimes I_{K+1}) \frac{\partial vec(H_{\tilde{g}})}{\partial vec(V)'} \\
=& -(\mu_R' V_R^{-1} X_{\tilde{g}} \otimes I_{K+1})(H_{\tilde{g}} \otimes H_{\tilde{g}}) \frac{\partial vec(H_{\tilde{g}}^{-1})}{\partial vec(V)'} \\
=& -(\gamma_{\tilde{g}}' \otimes H_{\tilde{g}}) \left[(X_{\tilde{g}}' V_R^{-1} \otimes I_{K+1}) \frac{\partial vec(X_{\tilde{g}}')}{\partial vec(V)'} + (X_{\tilde{g}}' \otimes X_{\tilde{g}}') \frac{\partial vec(V_R^{-1})}{\partial vec(V)'} + (I_{K+1} \otimes X_{\tilde{g}}' V_R^{-1}) \frac{\partial vec(X_{\tilde{g}})}{\partial vec(V)'} \right] \\
=& -(H_{\tilde{g}} \otimes \gamma_{\tilde{g}}' X_{\tilde{g}}' V_R^{-1}) \frac{\partial vec(X_{\tilde{g}})}{\partial vec(V)'} + (\gamma_{\tilde{g}}' X_{\tilde{g}}' V_R^{-1} \otimes H_{\tilde{g}} X_{\tilde{g}}' V_R^{-1}) \frac{\partial vec(V_R)}{\partial vec(V)'} - (\gamma_{\tilde{g}}' \otimes H_{\tilde{g}} X_{\tilde{g}}' V_R^{-1}) \frac{\partial vec(X_{\tilde{g}})}{\partial vec(V)'}
\end{aligned}$$

and

$$\begin{aligned}
(\mu_R' V_R^{-1} \otimes H_{\tilde{g}}) \frac{\partial vec(X_{\tilde{g}}')}{\partial vec(V)'} &= (H_{\tilde{g}} \otimes \mu_R' V_R^{-1}) \frac{\partial vec(X_{\tilde{g}})}{\partial vec(V)'} \\
(\mu_R' \otimes H_{\tilde{g}} X_{\tilde{g}}') \frac{\partial vec(V_R^{-1})}{\partial vec(V)'} &= -(\mu_R' V_R^{-1} \otimes H_{\tilde{g}} X_{\tilde{g}}' V_R^{-1}) \frac{\partial vec(V_R)}{\partial vec(V)'}
\end{aligned}$$

Combining the pieces above, with $e_{\tilde{g}} = \mu_R - X_{\tilde{g}} \gamma_{\tilde{g}}$:

$$\frac{\partial \gamma_{\tilde{g}}}{\partial vec(V)'} = (H_{\tilde{g}} \otimes e_{\tilde{g}}' V_R^{-1}) \frac{\partial vec(X_{\tilde{g}})}{\partial vec(V)'} - (\gamma_{\tilde{g}}' \otimes H_{\tilde{g}} X_{\tilde{g}}' V_R^{-1}) \frac{\partial vec(X_{\tilde{g}})}{\partial vec(V)'} - (e_{\tilde{g}}' V_R^{-1} \otimes H_{\tilde{g}} X_{\tilde{g}}' V_R^{-1}) \frac{\partial vec(V_R)}{\partial vec(V)'}$$

Compared to the case $W = V_R^{-1}$ is given, the difference is that we have the third term.

$$-(e_{\tilde{g}}' V_R^{-1} \otimes A_{\tilde{g}}) \frac{\partial vec(V_R)}{\partial vec(V)'} = -[0_{K+L}', e_{\tilde{g}}' V_R^{-1}] \otimes [0_{(K+1)\times(K+L)}, A_{\tilde{g}}]$$

This further induces the extra term $-(\gamma_{\tilde{g},t} - \gamma_{\tilde{g}}) u_{\tilde{g},t}$ for $h_{\hat{g},t}$:

$$\begin{aligned}
&-[0_{K+L}', e_{\tilde{g}}' V_R^{-1}] \otimes [0_{(K+1)\times(K+L)}, A_{\tilde{g}}] vec((Y_t - \mu)(Y_t - \mu)' - V) \\
=& -[0_{(K+1)\times(K+L)}, A_{\tilde{g}}]((Y_t - \mu)(Y_t - \mu)' - V) \begin{pmatrix} 0_{K+L} \\ V_R^{-1} e_{\tilde{g}} \end{pmatrix} \\
=& -A_{\tilde{g}}[(R_t - \mu_R)(R_t - \mu_R)' - V_R] V_R^{-1} e_{\tilde{g}} \\
=& -(\gamma_{\tilde{g},t} - \gamma_{\tilde{g}}) u_{\tilde{g},t}
\end{aligned}$$

For $\frac{\partial vec(X_{\tilde{g}})}{\partial vec(V)'}$, we derive $\frac{\partial vec(\beta_{\tilde{g}})}{\partial vec(V)'}$ first:

$$\begin{aligned}\frac{\partial vec(\beta_{\tilde{g}})}{\partial vec(V)'} =& [0_{K\times K}, V_{\tilde{g}}^{-1}V_{fr}V_r^{-1}, 0_{K\times N}] \otimes [-\beta_{\tilde{g}}, 0_{N\times L}, I_N] \\ &+ [V_{\tilde{g}}^{-1}, -V_{\tilde{g}}^{-1}V_{fr}V_r^{-1}, 0_{K\times N}] \otimes [0_{N\times K}, V_{Rr}V_r^{-1} - \beta_{\tilde{g}}V_{fr}V_r^{-1}, 0_{N\times N}]\end{aligned}$$

$$\begin{aligned}\frac{\partial vec(X_{\tilde{g}})}{\partial vec(V)'} &= \frac{\partial vec(X_{\tilde{g}})}{\partial vec(\beta_{\tilde{g}})'}\frac{\partial vec(\beta_{\tilde{g}})}{\partial vec(V)'} \\ &= ([0_K, I_K]' \otimes I_N)\frac{\partial vec(\beta_{\tilde{g}})}{\partial vec(V)'} \\ &= [0_K, I_K]'[0_{K\times K}, V_{\tilde{g}}^{-1}V_{fr}V_r^{-1}, 0_{K\times N}] \otimes [-\beta_{\tilde{g}}, 0_{N\times L}, I_N] \\ &+ [0_K, I_K]'[V_{\tilde{g}}^{-1}, -V_{\tilde{g}}^{-1}V_{fr}V_r^{-1}, 0_{K\times N}] \otimes [0_{N\times K}, V_{Rr}V_r^{-1} - \beta_{\tilde{g}}V_{fr}V_r^{-1}, 0_{N\times N}]\end{aligned}$$

Consequently:

$$\begin{aligned}(H_{\tilde{g}} \otimes e_{\tilde{g}}'W)\frac{\partial vec(X_{\tilde{g}})}{\partial vec(V)'} &= H_{\tilde{g}}[0_K, I_K]'[0_{K\times K}, V_{\tilde{g}}^{-1}V_{fr}V_r^{-1}, 0_{K\times N}] \otimes [0_K', 0_L', e_{\tilde{g}}'W] \\ &+ H_{\tilde{g}}[0_K, I_K]'[V_{\tilde{g}}^{-1}, -V_{\tilde{g}}^{-1}V_{fr}V_r^{-1}, 0_{K\times N}] \otimes [0_K', e_{\tilde{g}}'WV_{Rr}V_r^{-1}, 0_N']\end{aligned}$$

and

$$\begin{aligned}&H_{\tilde{g}}[0_K, I_K]'[0_{K\times K}, V_{\tilde{g}}^{-1}V_{fr}V_r^{-1}, 0_{K\times N}] \otimes [0_K', 0_L', e_{\tilde{g}}'W]vec((Y_t-\mu)(Y_t-\mu)'-V) \\ =& vec(e_{\tilde{g}}'W(R_t-\mu_R)(r_t-\mu_r)' - V_{Rr})V_r^{-1}V_{rf}V_{\tilde{g}}^{-1}[0_K, I_K]H_{\tilde{g}}) \\ =& H_{\tilde{g}}[0_K, I_K]'V_{\tilde{g}}^{-1}(\tilde{g}_t - \mu_{\tilde{g}})u_{\tilde{g},t} \\ =& H_{\tilde{g}}z_{\tilde{g},t}\end{aligned}$$

$$\begin{aligned}&H_{\tilde{g}}[0_K, I_K]'[V_{\tilde{g}}^{-1}, -V_{\tilde{g}}^{-1}V_{fr}V_r^{-1}, 0_{K\times N}] \otimes [0_K', e_{\tilde{g}}'WV_{Rr}V_r^{-1}, 0_N']vec((Y_t-\mu)(Y_t-\mu)'-V) \\ =& vec(e_{\tilde{g}}'WV_{Rr}V_r^{-1}((r_t-\mu_r)(f_t-\mu_f)' - V_{rf})V_{\tilde{g}}^{-1}[0_K, I_K]H_{\tilde{g}} \\ &- e_{\tilde{g}}'WV_{Rr}V_r^{-1}((r_t-\mu_r)(r_t-\mu_r)' - V_r)V_r^{-1}V_{rf}V_{\tilde{g}}^{-1}[0_K, I_K]H_{\tilde{g}}) \\ =& vec(e_{\tilde{g}}'WV_{Rr}V_r^{-1}(r_t-\mu_r)\tilde{\eta}_t'V_{\tilde{g}}^{-1}[0_K, I_K]H_{\tilde{g}}) \\ =& H_{\tilde{g}}\begin{bmatrix} 0 \\ V_{\tilde{g}}^{-1}\tilde{\eta}_t \end{bmatrix}(r_t-\mu_r)'V_r^{-1}V_{rR}We_{\tilde{g}}\end{aligned}$$

In addition:

$$\begin{aligned}&-(\gamma_{\tilde{g}}' \otimes A_{\tilde{g}})\frac{\partial vec(X_{\tilde{g}})}{\partial vec(V)'} \\ =& [0_K', r_{\tilde{g}1}'V_{\tilde{g}}^{-1}V_{fr}V_r^{-1}, 0_N'] \otimes [A_{\tilde{g}}\beta_{\tilde{g}}, 0_{(K+1)\times L}, -A_{\tilde{g}}] \\ &+ [r_{\tilde{g}1}'V_{\tilde{g}}^{-1}, -r_{\tilde{g}1}'V_{\tilde{g}}^{-1}V_{fr}V_r^{-1}, 0_N'] \otimes [0_{(K+1)\times K}, A_{\tilde{g}}\beta_{\tilde{g}}V_{fr}V_r^{-1} - A_{\tilde{g}}V_{Rr}V_r^{-1}, 0_{(K+1)\times N}]\end{aligned}$$

and

$$
\begin{aligned}
&[0'_K, r'_{\tilde{g}1}V_{\tilde{g}}^{-1}V_{fr}V_r^{-1}, 0'_N] \otimes [A_{\tilde{g}}\beta_{\tilde{g}}, 0_{(K+1)\times L}, -A_{\tilde{g}}]vec((Y_t-\mu)(Y_t-\mu)'-V) \\
=& -A_{\tilde{g}}((R_t-\mu_R)(r_t-\mu_r)'-V_{Rr})V_r^{-1}V_{rf}V_{\tilde{g}}^{-1}r_{\tilde{g}1} \\
&+A_{\tilde{g}}\beta_{\tilde{g}}((f_t-\mu_f)(r_t-\mu_r)'-V_{fr})V_r^{-1}V_{rf}V_{\tilde{g}}^{-1}r_{\tilde{g}1} \\
=& -(\gamma_{\tilde{g},t}-\gamma_{\tilde{g}})w_{\tilde{g},t}+A_{\tilde{g}}\beta_{\tilde{g}}(f_t-\mu_f)w_{\tilde{g},t}
\end{aligned}
$$

$$
\begin{aligned}
&[r'_{\tilde{g}1}V_{\tilde{g}}^{-1}, -r'_{\tilde{g}1}V_{\tilde{g}}^{-1}V_{fr}V_r^{-1}, 0'_N] \otimes [0_{(K+1)\times K}, A_{\tilde{g}}\beta_{\tilde{g}}V_{fr}V_r^{-1}-A_{\tilde{g}}V_{Rr}V_r^{-1}, 0_{(K+1)\times N}] \\
&vec((Y_t-\mu)(Y_t-\mu)'-V) \\
=&(A_{\tilde{g}}\beta_{\tilde{g}}V_{fr}V_r^{-1}-A_{\tilde{g}}V_{Rr}V_r^{-1})((r_t-\mu_r)(f_t-\mu_f)'-V_{rf})V_{\tilde{g}}^{-1}r_{\tilde{g}1} \\
&-(A_{\tilde{g}}\beta_{\tilde{g}}V_{fr}V_r^{-1}-A_{\tilde{g}}V_{Rr}V_r^{-1})((r_t-\mu_r)(r_t-\mu_r)'-V_r)V_r^{-1}V_{rf}V_{\tilde{g}}^{-1}r_{\tilde{g}1} \\
=&A_{\tilde{g}}\beta_{\tilde{g}}(\tilde{g}_t-\mu_{\tilde{g}})\tilde{\eta}'_tV_{\tilde{g}}^{-1}r_{\tilde{g}1}-A_{\tilde{g}}V_{Rr}V_r^{-1}(r_t-\mu_r)\tilde{\eta}'_tV_{\tilde{g}}^{-1}r_{\tilde{g}1}
\end{aligned}
$$

We derive $h_{\hat{\tilde{g}},t}$ as follows.

$$
\begin{aligned}
h_{\hat{\tilde{g}},t} &= \frac{\partial\gamma_{\tilde{g}}}{\partial\psi'}\zeta_t \\
&= \begin{pmatrix} 0_{(K+1)\times(K+L)} & A_{\tilde{g}} & \frac{\partial\gamma_{\tilde{g}}}{\partial vec(V)'} \end{pmatrix}\begin{pmatrix} Y_t-\mu \\ vec((Y_t-\mu)(Y_t-\mu)'-V) \end{pmatrix} \\
&= A_{\tilde{g}}(R_t-\mu_R)+\frac{\partial\gamma_{\tilde{g}}}{\partial vec(V)'}vec((Y_t-\mu)(Y_t-\mu)'-V)
\end{aligned}
$$

By plugging in the pieces for $\frac{\partial\gamma_{\tilde{g}}}{\partial vec(V)'}$ above (i.e., $(H_{\tilde{g}}\otimes e'_{\tilde{g}}V_R^{-1})\frac{\partial vec(X_{\tilde{g}})}{\partial vec(V)'}$, $-(\gamma'_{\tilde{g}}\otimes A_{\tilde{g}})\frac{\partial vec(X_{\tilde{g}})}{\partial vec(V)'}$, $-(e'_{\tilde{g}}V_R^{-1}\otimes A_{\tilde{g}})\frac{\partial vec(V_R)}{\partial vec(V)'}$), we end up with

$$
\begin{aligned}
&h_{\hat{\tilde{g}},t} \\
=&(\gamma_{\tilde{g},t}-\gamma_{\tilde{g}})+H_{\tilde{g}}z_{\tilde{g},t}+H_{\tilde{g}}\begin{bmatrix} 0 \\ V_{\tilde{g}}^{-1}\tilde{\eta}_t \end{bmatrix}(r_t-\mu_r)'V_r^{-1}V_{rR}We_{\tilde{g}}-(\gamma_{\tilde{g},t}-\gamma_{\tilde{g}})u_{\tilde{g},t}-(\gamma_{\tilde{g},t}-\gamma_{\tilde{g}})w_{\tilde{g},t} \\
&+A_{\tilde{g}}\beta_{\tilde{g}}(f_t-\mu_f)w_{\tilde{g},t}+A_{\tilde{g}}\beta_{\tilde{g}}(\tilde{g}_t-\mu_{\tilde{g}})\tilde{\eta}'_tV_{\tilde{g}}^{-1}r_{\tilde{g}1}-A_{\tilde{g}}V_{Rr}V_r^{-1}(r_t-\mu_r)\tilde{\eta}'_tV_{\tilde{g}}^{-1}r_{\tilde{g}1} \\
=&(\gamma_{\tilde{g},t}-\gamma_{\tilde{g}})-(\phi_{\tilde{g},t}-\phi_{\tilde{g}})w_{\tilde{g},t}+H_{\tilde{g}}z_{\tilde{g},t}-(\gamma_{\tilde{g},t}-\gamma_{\tilde{g}})u_{\tilde{g},t}+\tilde{\delta}_t
\end{aligned}
$$

where $\tilde{\delta}_t$ is defined as

$$
H_{\tilde{g}}\begin{bmatrix} 0 \\ V_{\tilde{g}}^{-1}\tilde{\eta}_t \end{bmatrix}(r_t-\mu_r)'V_r^{-1}V_{rR}We_{\tilde{g}}+\begin{bmatrix} 0 \\ \tilde{\eta}_t \end{bmatrix}w_{\tilde{g},t}+\{\begin{bmatrix} 0 \\ \tilde{g}_t-\mu_{\tilde{g}} \end{bmatrix}-A_{\tilde{g}}V_{Rr}V_r^{-1}(r_t-\mu_r)\}r'_{\tilde{g}1}V_{\tilde{g}}^{-1}\tilde{\eta}_t.
$$

□

H: Factors and Mimicking Portfolios in CSR

Consider

$$f_t^c = [f_{1t}',\ g_{2t}']'$$

where f_{1t} is the $K_1 \times 1$ vector of factors with $K_1 \leq K$, g_{2t} is the $(K - K_1) \times 1$ vector of mimicking portfolios, with $f_t = [f_{1t}',\ f_{2t}']'$, $g_t = [g_{1t}',\ g_{2t}']'$.

Note that if $K_1 = K$, then f_t^c reduces to f_t; in contrast, if $K_1 = 0$, then f_t^c reduces to g_t. In general, f_t^c consists of factors as well as mimicking portfolios.

As in the main text, we consider mimicking portfolios resulting from the time series regression approach, i.e., $g_{2t} = V_{f_2R}V_R^{-1}R_t$, where V_{f_2R} denotes the covariance of f_{2t} and R_t. In addition, since estimation error of mimicking portfolios is not necessarily negligible, we focus on the feasible version of f_t^c, denote by

$$\hat{f}_t^c = [f_{1t}',\ (\hat{V}_{f_2R}\hat{V}_R^{-1}R_t)']'$$

The theorem below provides the asymptotic distribution of the risk premium estimator, when $\hat{f}_t^c$ is used in CSR.

Theorem 5. *Let $\hat{\beta}^c \equiv \hat{V}_{Rf}\hat{V}_{f^c}^{-1}$, $\hat{X}^c \equiv [1_N,\ \hat{\beta}^c]$, W is a positive definite weighting matrix. If W is known, the asymptotic distribution of $\hat{\gamma}^c = (\hat{X}^{c\prime}W\hat{X}^c)^{-1}\hat{X}^{c\prime}W\hat{\mu}_R$ is given by:*

$$\sqrt{T}(\hat{\gamma}^c - \gamma^c) \overset{A}{\sim} N(0, V_{\hat{\gamma}^c})$$

where $V_{\hat{\gamma}^c} = \sum_{j=-\infty}^{\infty} \mathbb{E}[h_t^c h_{t+j}^{c\prime}]$ with

$$h_t^c = \begin{bmatrix} 1 & 0_K' \\ 0_K & V_{f^c}V_f^{-1} \end{bmatrix} h_t + \begin{bmatrix} 0 \\ [(V_f - V_{f^c})V_f^{-1}(f_t - \mu_f)(f_t - \mu_f)' + S\eta_t\eta_t'S - \eta_t\eta_t']V_f^{-1}\gamma_1 \end{bmatrix}$$

and h_t is provided in (7), $V_{f^c} = \begin{bmatrix} V_{f_1} & V_{f_1R}V_R^{-1}V_{Rf_2} \\ V_{f_2R}V_R^{-1}V_{Rf_1} & V_{f_2R}V_R^{-1}V_{Rf_2} \end{bmatrix}$, $\gamma^c = \begin{bmatrix} 1 & 0_K' \\ 0_K & V_{f^c}V_f^{-1} \end{bmatrix}\gamma$, $S = \begin{bmatrix} I_{K_1} & 0_{K_1\times(K-K_1)} \\ 0_{(K-K_1)\times K_1} & 0_{(K-K_1)\times(K-K_1)} \end{bmatrix}$. Similarly, the asymptotic distribution result above holds for $\hat{\gamma}^c = (\hat{X}^{c\prime}\hat{V}_R^{-1}\hat{X}^c)^{-1}\hat{X}^{c\prime}\hat{V}_R^{-1}\hat{\mu}_R$, with h_t in (8).

When $\hat{f}_t^c$ only contains factors (i.e., $K_1 = K$), it is easy to see that h_t^c reduces to h_t, so Theorem 5 coincides with Theorem 1. In contrast, when $\hat{f}_t^c$ only contains mimicking portfolios constructed by the time series regression approach (i.e., $K_1 = 0$), Theorem 5 coincides with Theorem 2.

If traded factors denoted by the $K_1 \times 1$ vector g_{1t} are simultaneously used with the $(K - K_1) \times 1$ vector of mimicking portfolios $\hat{g}_{2,t} = \hat{V}_{f_2R}\hat{V}_R^{-1}R_t$ in CSR, the asymptotic distribution of the risk premium estimator is similarly provided by Theorem 5, if we replace the objects corresponding to f_{1t} with the counterparts resulting from g_{1t}.

References

Ang, A., Hodrick, R. J., Xing, Y. and Zhang, X. (2006). The cross-section of volatility and expected returns. *Journal of Finance*, **61** (1), 259–299.

Avramov, D. and Chordia, T. (2006). Asset pricing models and financial market anomalies. *Review of Financial Studies*, **19** (3), 1001–1040.

Balduzzi, P. and Robotti, C. (2008). Mimicking portfolios, economic risk premia, and tests of multi-beta models. *Journal of Business and Economic Statistics*, **26** (3), 354–368.

Breeden, D. T. (1979). An intertemporal asset pricing model with stochastic consumption and investment opportunities. *Journal of Financial Economics*, **7** (3), 265–296.

—, Gibbons, M. R. and Litzenberger, R. H. (1989). Empirical tests of the consumption-oriented capm. *Journal of Finance*, **44** (2), 231–262.

Brennan, M. J., Chordia, T. and Subrahmanyam, A. (1998). Alternative factor specifications, security characteristics, and the cross-section of expected stock returns. *Journal of Financial Economics*, **49** (3), 345–373.

Chan, L. K., Karceski, J. and Lakonishok, J. (1998). The risk and return from factors. *Journal of Financial and Quantitative Analysis*, **33** (02), 159–188.

Chang, B. Y., Christoffersen, P. and Jacobs, K. (2013). Market skewness risk and the cross section of stock returns. *Journal of Financial Economics*, **107** (1), 46–68.

Chen, N.-F., Roll, R. and Ross, S. A. (1986). Economic forces and the stock market. *Journal of Business*, pp. 383–403.

Chordia, T., Goyal, A. and Shanken, J. (2013). Cross-sectional asset pricing with individual stocks: Betas versus characteristics, working paper.

Cochrane, J. (1996). A cross-sectional test of an investment-based asset pricing model. *Journal of Political Economy*, **104** (3), 572–621.

Fama, E. and French, K. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, **33** (1), 3–56.

— and MacBeth, J. (1973). Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy*, **81** (3), 607–636.

Ferson, W. E. and Harvey, C. R. (1991). The variation of economic risk premiums. *Journal of Political Economy*, pp. 385–415.

Gibbons, M. R., Ross, S. A. and Shanken, J. (1989). A test of the efficiency of a given portfolio. *Econometrica*, pp. 1121–1152.

Gospodinov, N., Kan, R. and Robotti, C. (2014). Misspecification-robust inference in linear asset-pricing models with irrelevant risk factors. *Review of Financial Studies*, p. hht135.

Huberman, G., Kandel, S. and Stambaugh, R. F. (1987). Mimicking portfolios and exact arbitrage pricing. *Journal of Finance*, **42** (1), 1–9.

Jagannathan, R. and Wang, Z. (1996). The conditional capm and the cross-section of expected returns. *Journal of Finance*, pp. 3–53.

— and — (1998). An asymptotic theory for estimating beta-pricing models using cross-sectional regression. *Journal of Finance*, **53** (4), 1285–1309.

Kan, R., Robotti, C. and Shanken, J. (2013). Pricing model performance and the two-pass cross-sectional regression methodology. *Journal of Finance*, **68** (6), 2617–2649.

— and Zhang, C. (1999). Two-pass tests of asset pricing models with useless factors. *Journal of Finance*, pp. 203–235.

Kapadia, N. (2011). Tracking down distress risk. *Journal of Financial Economics*, **102** (1), 167–182.

Kleibergen, F. (2009). Tests of risk premia in linear factor models. *Journal of Econometrics*, **149** (2), 149–173.

Lamont, O. A. (2001). Economic tracking portfolios. *Journal of Econometrics*, **105** (1), 161–184.

Lehmann, B. N. and Modest, D. M. (1988). The empirical foundations of the arbitrage pricing theory. *Journal of Financial Economics*, **21** (2), 213–254.

Lettau, M. and Ludvigson, S. (2001). Resurrecting the (C) CAPM: A cross-sectional test when risk premia are time-varying. *Journal of Political Economy*, **109** (6), 1238–1287.

Li, Q., Vassalou, M. and Xing, Y. (2006). Sector investment growth rates and the cross section of equity returns. *Journal of Business*, **79** (3), 1637–1665.

Menkhoff, L., Sarno, L., Schmeling, M. and Schrimpf, A. (2012). Carry trades and global foreign exchange volatility. *Journal of Finance*, **67** (2), 681–718.

Muir, T., Adrian, T. and Etula, E. (2013). Financial intermediaries and the cross-section of asset returns. *Journal of Finance*, forthcoming.

Pástor, L. and Stambaugh, R. (2003). Liquidity risk and expected stock returns. *Journal of Political Economy*, **111** (3), 642–685.

Shanken, J. (1992). On the estimation of beta-pricing models. *Review of Financial Studies*, pp. 1–33.

Vassalou, M. (2003). News related to future gdp growth as a risk factor in equity returns. *Journal of Financial Economics*, **68** (1), 47–73.

Table 1: Standard Errors of Risk Premia - OLS

Panel A: Correct specification

	Model M-W					
	$T = 500$			$T = 2000$		
	$var(\hat{\gamma}_{\hat{g}})^{1/2}$	$\widehat{var}(\hat{\gamma}_{\hat{g}})^{1/2}$		$var(\hat{\gamma}_{\hat{g}})^{1/2}$	$\widehat{var}(\hat{\gamma}_{\hat{g}})^{1/2}$	
		$h_{g,t}$	$h_{\hat{g},t}$		$h_{g,t}$	$h_{\hat{g},t}$
Factor 1	0.0674	0.0495	0.0631	0.0313	0.0248	0.0316
2	0.0222	0.0151	0.0200	0.0103	0.0075	0.0100
3	0.0043	0.0031	0.0039	0.0020	0.0016	0.0020
	Model M-S					
	$T = 500$			$T = 2000$		
	$var(\hat{\gamma}_{\hat{g}})^{1/2}$	$\widehat{var}(\hat{\gamma}_{\hat{g}})^{1/2}$		$var(\hat{\gamma}_{\hat{g}})^{1/2}$	$\widehat{var}(\hat{\gamma}_{\hat{g}})^{1/2}$	
		$h_{g,t}$	$h_{\hat{g},t}$		$h_{g,t}$	$h_{\hat{g},t}$
Factor 1	0.8483	0.9001	0.9015	0.4391	0.4502	0.4509
2	0.2666	0.2627	0.2629	0.1283	0.1314	0.1315
3	0.2387	0.2355	0.2359	0.1199	0.1177	0.1179

Panel B: Mis-specification

	Model M-W					
	$T = 500$			$T = 2000$		
	$var(\hat{\gamma}_{\hat{g}})^{1/2}$	$\widehat{var}(\hat{\gamma}_{\hat{g}})^{1/2}$		$var(\hat{\gamma}_{\hat{g}})^{1/2}$	$\widehat{var}(\hat{\gamma}_{\hat{g}})^{1/2}$	
		$h_{g,t}$	$h_{\hat{g},t}$		$h_{g,t}$	$h_{\hat{g},t}$
Factor 1	0.0712	0.0511	0.0675	0.0338	0.0255	0.0337
2	0.0262	0.0165	0.0239	0.0123	0.0083	0.0119
3	0.0046	0.0032	0.0044	0.0022	0.0016	0.0022
	Model M-S					
	$T = 500$			$T = 2000$		
	$var(\hat{\gamma}_{\hat{g}})^{1/2}$	$\widehat{var}(\hat{\gamma}_{\hat{g}})^{1/2}$		$var(\hat{\gamma}_{\hat{g}})^{1/2}$	$\widehat{var}(\hat{\gamma}_{\hat{g}})^{1/2}$	
		$h_{g,t}$	$h_{\hat{g},t}$		$h_{g,t}$	$h_{\hat{g},t}$
Factor 1	1.0994	1.1830	1.1882	0.5835	0.5909	0.5935
2	0.2669	0.2629	0.2631	0.1283	0.1315	0.1316
3	0.2400	0.2366	0.2371	0.1201	0.1183	0.1185

Note: Model M-W uses mimicking portfolios from three weak factors, while Model M-S uses mimicking portfolios from three strong factors. $var(\hat{\gamma}_{\hat{g}})^{1/2}$ stands for the standard error of the risk premium estimator obtained by Monte Carlo replications; $\widehat{var}(\hat{\gamma}_{\hat{g}})^{1/2}$ stands for the standard error of the risk premium estimator by the asymptotic theory: under $h_{g,t}$, the estimation error of mimicking portfolios is ignored; under $h_{\hat{g},t}$, the estimation error is accounted for. The reported numbers result from 2000 replications.

Table 2: Standard Errors of Risk Premia - GLS

Panel A: Correct specification

Model M-W						
	$T=500$			$T=2000$		
	$var(\hat{\gamma}_{\hat{g}})^{1/2}$	$\widehat{var}(\hat{\gamma}_{\hat{g}})^{1/2}$		$var(\hat{\gamma}_{\hat{g}})^{1/2}$	$\widehat{var}(\hat{\gamma}_{\hat{g}})^{1/2}$	
		$h_{g,t}$	$h_{\hat{g},t}$		$h_{g,t}$	$h_{\hat{g},t}$
Factor 1	0.0343	0.0315	0.0330	0.0168	0.0158	0.0165
2	0.0122	0.0104	0.0111	0.0057	0.0052	0.0055
3	0.0021	0.0020	0.0020	0.0010	0.0010	0.0010

Model M-S						
	$T=500$			$T=2000$		
	$var(\hat{\gamma}_{\hat{g}})^{1/2}$	$\widehat{var}(\hat{\gamma}_{\hat{g}})^{1/2}$		$var(\hat{\gamma}_{\hat{g}})^{1/2}$	$\widehat{var}(\hat{\gamma}_{\hat{g}})^{1/2}$	
		$h_{g,t}$	$h_{\hat{g},t}$		$h_{g,t}$	$h_{\hat{g},t}$
Factor 1	0.6183	0.6377	0.6380	0.3182	0.3188	0.3189
2	0.2588	0.2597	0.2598	0.1267	0.1299	0.1300
3	0.2296	0.2307	0.2309	0.1178	0.1153	0.1154

Panel B: Mis-specification

Model M-W						
	$T=500$			$T=2000$		
	$var(\hat{\gamma}_{\hat{g}})^{1/2}$	$\widehat{var}(\hat{\gamma}_{\hat{g}})^{1/2}$		$var(\hat{\gamma}_{\hat{g}})^{1/2}$	$\widehat{var}(\hat{\gamma}_{\hat{g}})^{1/2}$	
		$h_{g,t}$	$h_{\hat{g},t}$		$h_{g,t}$	$h_{\hat{g},t}$
Factor 1	0.0520	0.0396	0.0552	0.0269	0.0198	0.0276
2	0.0184	0.0126	0.0184	0.0092	0.0063	0.0092
3	0.0030	0.0024	0.0032	0.0016	0.0012	0.0016

Model M-S						
	$T=500$			$T=2000$		
	$var(\hat{\gamma}_{\hat{g}})^{1/2}$	$\widehat{var}(\hat{\gamma}_{\hat{g}})^{1/2}$		$var(\hat{\gamma}_{\hat{g}})^{1/2}$	$\widehat{var}(\hat{\gamma}_{\hat{g}})^{1/2}$	
		$h_{g,t}$	$h_{\hat{g},t}$		$h_{g,t}$	$h_{\hat{g},t}$
Factor 1	0.8880	0.9643	0.9678	0.4783	0.4815	0.4832
2	0.2992	0.3051	0.3056	0.1486	0.1525	0.1528
3	0.2722	0.2717	0.2726	0.1344	0.1358	0.1363

Note: Model M-W uses mimicking portfolios from three weak factors, while Model M-S uses mimicking portfolios from three strong factors. $var(\hat{\gamma}_{\hat{g}})^{1/2}$ stands for the standard error of the risk premium estimator obtained by Monte Carlo replications; $\widehat{var}(\hat{\gamma}_{\hat{g}})^{1/2}$ stands for the standard error of the risk premium estimator by the asymptotic theory: under $h_{g,t}$, the estimation error of mimicking portfolios is ignored; under $h_{\hat{g},t}$, the estimation error is accounted for. The reported numbers result from 2000 replications.

Table 3: Estimation of Risk Premia in CSR with Mimicking Portfolios - OLS

Model	Const.	$\triangle I_{Nres}$	$\triangle I_{Res}$	Const.	$Hholds$	$Nfinco$	$Finan$	Const.	Lev
risk premium	0.8155	0.0029	0.0091	0.7459	0.0094	0.0077	0.0004	0.6856	3.0366
standard error									
FM	0.7348	0.0023	0.0036	0.7596	0.0032	0.0053	0.0025	1.0106	0.9758
Shanken	0.7975	0.0025	0.0038	0.8522	0.0035	0.0057	0.0028	1.1129	1.0452
JW	0.8411	0.0028	0.0040	0.8904	0.0038	0.0064	0.0031	1.1995	1.0067
KRS	0.8828	0.0023	0.0044	0.9552	0.0045	0.0060	0.0044	1.2135	1.0426
EIW(c)	1.1237	0.0028	0.0052	1.2294	0.0048	0.0070	0.0034	1.7071	1.3196
EIW(m)	1.1243	0.0035	0.0056	1.3399	0.0067	0.0098	0.0085	1.7160	1.3097
t-ratio									
FM	1.11	1.27	2.53	0.98	2.96	1.46	0.15	0.68	3.11
Shanken	1.02	1.19	2.39	0.88	2.71	1.34	0.14	0.62	2.91
JW	0.97	1.04	2.26	0.84	2.50	1.20	0.12	0.57	3.02
KRS	0.92	1.26	2.09	0.78	2.09	1.28	0.09	0.56	2.91
EIW(c)	0.73	1.04	1.74	0.61	1.95	1.10	0.11	0.40	2.30
EIW(m)	0.73	0.85	1.62	0.56	1.41	0.78	0.04	0.40	2.32

Note: The test assets are the 25 Fama-French size and book-to-market sorted portfolios and the sample period covers 1973Q1-2009Q4. The three models are adopted from Cochrane (1996), Li *et al.* (2006), and Muir *et al.* (2013), respectively. Estimates of risk premium are calculated by OLS in the Fama and MacBeth (1973) two-pass procedure using mimicking portfolios. Six types of standard errors (and thus six t-ratios) of risk premia are provided: FM-Fama and MacBeth (1973), Shanken-Shanken (1992), JW-Jagannathan and Wang (1998), KRS-Kan *et al.* (2013) and our proposed EIW (errors-in-weights) standard error. EIW(c) assumes correct model specification, while EIW (m) allows for model misspecification.

Table 4: Estimation of Risk Premia in CSR with Mimicking Portfolios - GLS

Model	Const.	$\triangle I_{Nres}$	$\triangle I_{Res}$	Const.	$Hholds$	$Nfinco$	$Finan$	Const.	Lev
risk premium	1.7012	−0.0017	0.0029	1.7563	0.0025	0.0018	−0.0032	1.9723	0.8374
standard error									
FM	0.5338	0.0009	0.0020	0.5299	0.0017	0.0027	0.0014	0.5225	0.5875
Shanken	0.5486	0.0009	0.0020	0.5469	0.0017	0.0027	0.0014	0.5267	0.5883
JW	0.5498	0.0009	0.0020	0.5325	0.0017	0.0027	0.0014	0.5244	0.5832
KRS	0.6113	0.0009	0.0021	0.5990	0.0017	0.0027	0.0014	0.5933	0.5960
EIW(c)	0.6013	0.0011	0.0023	0.5514	0.0019	0.0027	0.0015	0.5413	0.5786
EIW(m)	0.7304	0.0018	0.0038	0.6646	0.0032	0.0045	0.0026	0.6842	1.1483
t-ratio									
FM	3.19	−1.79	1.46	3.31	1.47	0.66	−2.38	3.77	1.43
Shanken	3.10	−1.79	1.46	3.21	1.46	0.66	−2.37	3.74	1.42
JW	3.09	−1.79	1.47	3.30	1.47	0.66	−2.39	3.76	1.44
KRS	2.78	−1.80	1.44	2.93	1.44	0.66	−2.39	3.32	1.41
EIW(c)	2.83	−1.55	1.27	3.19	1.29	0.67	−2.18	3.64	1.45
EIW(m)	2.33	−0.95	0.77	2.64	0.78	0.40	−1.27	2.88	0.73

Note: The test assets are the 25 Fama-French size and book-to-market sorted portfolios and the sample period covers 1973Q1-2009Q4. The three models are adopted from Cochrane (1996), Li *et al.* (2006), and Muir *et al.* (2013), respectively. Estimates of risk premium are calculated by GLS in the Fama and MacBeth (1973) two-pass procedure using mimicking portfolios. Six types of standard errors (and thus six t-ratios) of risk premia are provided: FM-Fama and MacBeth (1973), Shanken-Shanken (1992), JW-Jagannathan and Wang (1998), KRS-Kan *et al.* (2013) and our proposed EIW (errors-in-weights) standard error. EIW(c) assumes correct model specification, while EIW (m) allows for model misspecification.