Bayesian Fama-MacBeth Regressions

Svetlana Bryzgalova* Jiantao Huang† Christian Julliard‡

October 2024

Abstract

Commonly used frequentist estimation methods for linear factor models of asset returns are invalidated by weak and spurious factors. The problem is amplified by omitted variables and model misspecification, often calling for specialized non-standard estimation techniques. Conversely, the Bayesian analogue of the popular Fama and MacBeth (1973) two-pass regressions method provides reliable risk premia estimates for both tradable and nontradable factors, detects those weakly identified, delivers valid credible intervals for all objects of interest, and is intuitive, fast and simple to implement. In other words, weak and spurious factors are not a problem for the Bayesian estimation of Fama-MacBeth regressions.

Keywords: Cross-sectional asset pricing, factor models, weak factors, risk premia, Bayesian methods.

JEL Classification Codes: C11, G11, G12.

*Corresponding author; London Business School; sbryzgalova@london.edu. Address: Department of Finance, London Business School, Regents Park, London, NW1 4SA, UK.

†The University of Hong Kong, huangjt@hku.hk. Address: Faculty of Business and Economics, The University of Hong Kong, Pok Fu Lam, Hong Kong.

‡FMG, and SRC, London School of Economics, and CEPR; c.julliard@lse.ac.uk. Address: Department of Finance, London School of Economics, Houghton Street, London, WC2A 2AE, UK.

1 Introduction

Linear factor models are ubiquitous in asset pricing. The economics behind these models can be summarized in one sentence: The expected return of an asset equals its exposures to systematic risk factors multiplied by the factors' risk premia. A typical formulation of these models is described as follows:

$$\boldsymbol{R}_t = \boldsymbol{\beta_f}\boldsymbol{\lambda_f} + \boldsymbol{\beta_f}(\boldsymbol{f}_t - \boldsymbol{\mu_f}) + \boldsymbol{\epsilon}_t, \quad \boldsymbol{f}_t \perp \boldsymbol{\epsilon}_t, \quad t = 1, \ldots T, \tag{1}$$

where the returns of N test assets, in excess of the risk-free rate, are denoted by $\boldsymbol{R}_t = (R_{1t} \ldots R_{Nt})^\top$, $\boldsymbol{f}_t = (f_{1t} \ldots f_{Kt})^\top$ denotes K systematic (either tradable or nontradable) risk factors with unconditional expectations $\boldsymbol{\mu_f}$, $\boldsymbol{\beta_f}$ are the loadings on systematic risk, $\boldsymbol{\lambda_f}$ are the factors' risk premia, and $\boldsymbol{\epsilon}_t$ denotes the unpriced idiosyncratic errors.

Factors' risk premia, $\boldsymbol{\lambda_f}$, are often estimated via the Fama-MacBeth (FM) regression method (see Fama and MacBeth (1973)) due to its simplicity and the intuitive appeal of its hierarchical structure. In the FM procedure, the factor exposures of asset returns, $\boldsymbol{\beta_f} \in \mathbb{R}^{N\times K}$, are recovered from the following time-series regression:

$$\boldsymbol{R}_t = \boldsymbol{\mu_R} + \boldsymbol{\beta_f}(\boldsymbol{f}_t - \boldsymbol{\mu_f}) + \boldsymbol{\epsilon}_t, \tag{2}$$

The factors' risk premia, $\boldsymbol{\lambda_f} \in \mathbb{R}^K$, are then estimated from the cross-sectional regression:

$$\widehat{\boldsymbol{\mu}_R} = \widehat{\boldsymbol{\beta}}_{\boldsymbol{f}}\boldsymbol{\lambda_f} + \boldsymbol{\alpha}, \tag{3}$$

where $\widehat{\boldsymbol{\mu}_R}$ are the sample average returns of $\boldsymbol{R}_t$, $\widehat{\boldsymbol{\beta}}_{\boldsymbol{f}}$ denote the time-series estimates from equation (2), and $\boldsymbol{\alpha} \in \mathbb{R}^N$ is the vector of pricing errors. The ordinary or generalized least square (OLS or GLS) estimates of factors' risk premia, as well as their standard errors with the Shanken (1992) correction, are given by equations (12.11), (12.15), and (12.19) in Cochrane (2005).

In the presence of a spurious (or weak) factor, that is, such that $\boldsymbol{\beta}_j = \frac{\boldsymbol{C}}{\sqrt{T}}$, $\boldsymbol{C} \in \mathbb{R}^N$, risk premia are no longer identified. In particular, their estimates diverge (i.e., $\widehat{\boldsymbol{\lambda}}_j \not\to 0$ as $T \to \infty$), leading to inference problems for both the useless and the strong factors (see, e.g., Kan and Zhang (1999a,b) and Gospodinov, Kan, and Robotti (2017, 2019)), as well as highly inflated cross-sectional fit (see Kleibergen and Zhan (2015)).

Consequently, plenty of frequentist statistical methods have been proposed to address the weak identification of $\boldsymbol{\beta_f}$. Kleibergen (2009) presents several novel statistics that are robust to weak factors. Gospodinov, Kan, and Robotti (2014) derive robust standard errors for the risk premia estimates and show that the corresponding t-statistics are robust even when the model is misspecified and a useless factor is included. Bryzgalova (2015) introduces a penalized term in the FM regression to shrink the risk premium of the useless factor toward zero. Burnside (2016) suggests researchers use the tests of rank conditions of $\boldsymbol{\beta_f}$ for model reduction. Kleibergen and Zhan (2020) extend the GRS statistic in Gibbons, Ross, and Shanken (1989) to test the identification of risk premia. Last but not least, Anatolyev and Mikusheva (2022) adopt the idea of sample-splitting instrumental variable regressions, and propose a new estimator that is robust to the weak factors and the presence of strong unaccounted cross-sectional error dependence.

Different from the above solutions, we propose a hierarchical Bayesian analogue of the two-pass regressions in equations (2)–(3). Despite being an essential concern in the frequentist FM estimation, weak identification is of little consequence in this Bayesian setting. Furthermore, in the absence of identification and misspecification problems, our Bayesian method yields posteriors centred at the FM estimates and analogous inference. Our modelling ingredients are rather standard. In the time-series dimension, we assume that $\boldsymbol{f}_t$ and $\boldsymbol{R}_t$ follow a joint multivariate normal distribution, which implies a Normal-inverse-Wishart posterior distribution of the first two moments of the data under the flat prior. Conditional on the sampled time-series parameters, the posterior distribution of $\boldsymbol{f}_t$'s risk premia is a Dirac at the OLS or GLS estimates. As we show in the simulation studies, the Bayesian credible intervals provided by our framework deliver appropriate coverages of the true risk premia under the null hypothesis, regardless of strong or weak factors. Conversely, in the same simulations, the frequentist analogues always over-reject the null of zero risk premia for the weak factors.

The model in equation (1), as well as the two-pass regressions in equations (2)–(3), assume that $\boldsymbol{f}_t$ enter the stochastic discount factor and that no omitted factor can bias the risk premia estimates. Nevertheless, these assumptions, although often made in the existing empirical literature, are overly strong and fragile.[1] Since the Bayesian Fama MacBeth (BFM) OLS and GLS estimators inherit these assumptions, they are also under the same scrutiny.

To simultaneously address weak identification *and* omitted factor bias, we introduce another Bayesian estimator, denoted as BFM-OMIT. This estimator shares the same spirit of the three-pass regression of Giglio and Xiu (2021): Expected asset returns are explained by the exposures to the unobserved latent factors, and an observable factor is priced because it loads on the priced latent sources of risk. Instead of estimating the second step in equation (3), we regress the average returns on the eigenvectors of the covariance matrix of asset returns and next infer the risk premia of $\boldsymbol{f}_t$.

Unlike Giglio and Xiu (2021), we do not study a large cross-section (i.e., $N << T$).

[1]For instance, Dickerson, Julliard, and Mueller (2024) and Bryzgalova, Huang, and Julliard (2023) find that the stochastic discount factors for both corporate bonds and equities are dense in the space of observable factors, and that the popular low dimensional models in the literature are misspecified with very high probability.

Although the traditional latent factor selection procedures (e.g., Bai and Ng (2002)) cannot be applied in a small N setting, model selection and averaging can be conducted over the entire eigenspace of the covariance matrix of asset returns (e.g., using the spike-and-slab method of Bryzgalova, Huang, and Julliard (2023)). In the extreme case that all eigenvectors are essential in explaining asset returns, BFM-OMIT is equivalent to finding the mimicking portfolio by projecting the factor onto the entire original asset space.

The paper is organized as follows. Section 2 presents the Bayesian Fama-MacBeth methods. Section 3 studies the finite-sample performance of our estimators via Monte Carlo simulations. We conclude in Section 4 by showing several empirical examples.

2 Inference in Linear Factor Models

This section introduces our hierarchical Bayesian Fama-MacBeth (BFM) estimation method. Our ultimate goal is to estimate the risk premia of (tradable or nontradable) factors $\boldsymbol{f}$, which are defined as follows:

$$\boldsymbol{\lambda_f} = -\text{cov}(\boldsymbol{f}_t, M_t), \tag{4}$$

where M_t is the stochastic discount factor (SDF) that is normalized such that $\mathbb{E}[M_t] = 1$. The definition in equation (4) is natural for the tradable factors due to the fundamental asset pricing equation, i.e., $\mathbb{E}[M_t \boldsymbol{f}_t] = \mathbf{0}$ if $\boldsymbol{f}_t$ are tradable excess returns. For nontradable $\boldsymbol{f}$, we can interpret $-\text{cov}(\boldsymbol{f}_t, M_t)$ as the expected excess returns on the pseudo assets with stochastic growth rates $\boldsymbol{f}_t$.

In Subsection 2.1, we describe the baseline case without any omitted factor; that is, the observable factors $\boldsymbol{f}$ fully capture the sources of priced risk. However, the assumption of no omitted factor is strong (yet pervasive in the literature). Hence, in Subsection 2.2, we present a simple modification of the baseline approach that accounts for omitted variable bias in estimating factor risk premia.

2.1 Bayesian Fama-MacBeth without Omitted Factors

In this subsection, we assume that asset returns follow the data-generating process in equation (1)—the exposures to the observable factors $\boldsymbol{f}_t$ are sufficient to explain expected returns. Equivalently, this implies the following linear SDF:

$$M_t = 1 - \boldsymbol{\lambda}_{\boldsymbol{f}}^\top \boldsymbol{\Sigma}_{\boldsymbol{f}}^{-1} (\boldsymbol{f}_t - \boldsymbol{\mu}_{\boldsymbol{f}}), \tag{5}$$

which is consistent with the definition in equation (4). Hence, whenever we run the two-step FM regression in equations (2)–(3) to estimate $\boldsymbol{\lambda}_f$, we implicitly assume that $\boldsymbol{f}_t$ are the only relevant pricing factors in the linear SDF.

We first consider the time-series dimension of the estimation. Let $\boldsymbol{Y}_t = (\boldsymbol{R}_t^\top, \boldsymbol{f}_t^\top)^\top$, a $p \times 1$ vector ($p = N + K$).[2] We assume that $\{\boldsymbol{Y_t}\}_{t=1}^T$ follows an independent and identically distributed (iid) multivariate Gaussian distribution, that is, $\boldsymbol{Y_t} \overset{\text{iid}}{\sim} \mathcal{N}(\boldsymbol{\mu_Y}, \boldsymbol{\Sigma_Y})$, where $\boldsymbol{\mu_Y}$ and $\boldsymbol{\Sigma_Y}$ are, respectively, the unconditional vector of means and covariance matrix of $\boldsymbol{Y_t}$. This distributional assumption implies the following likelihood function for $\boldsymbol{Y_t}$:

$$p(\mathbf{Y} \mid \boldsymbol{\mu_Y}, \boldsymbol{\Sigma_Y}) \propto |\boldsymbol{\Sigma_Y}|^{-\frac{T}{2}} \exp\left\{-\frac{1}{2} tr\left[\boldsymbol{\Sigma_Y}^{-1} \sum_{t=1}^{T} (\boldsymbol{Y_t} - \boldsymbol{\mu_Y})(\boldsymbol{Y_t} - \boldsymbol{\mu_Y})^\top\right]\right\}, \tag{6}$$

where $\mathbf{Y} \equiv \{\boldsymbol{Y_t}\}_{t=1}^T$. For simplicity, we assign a diffuse prior to the parameters, $\pi(\boldsymbol{\mu_Y}, \boldsymbol{\Sigma_Y}) \propto |\boldsymbol{\Sigma_Y}|^{-\frac{p+1}{2}}$, yielding the following posterior distribution of $(\boldsymbol{\mu_Y}, \boldsymbol{\Sigma_Y})$:[3]

$$\boldsymbol{\mu_Y} \equiv \begin{pmatrix} \boldsymbol{\mu_R} \\ \boldsymbol{\mu_f} \end{pmatrix} \mid \boldsymbol{\Sigma_Y}, \mathbf{Y} \sim \mathcal{N}\left(\hat{\boldsymbol{\mu}}_{\boldsymbol{Y}},\ \boldsymbol{\Sigma_Y}/T\right), \tag{7}$$

$$\boldsymbol{\Sigma_Y} \equiv \begin{pmatrix} \boldsymbol{\Sigma_R} & \boldsymbol{\Sigma_{Rf}} \\ \boldsymbol{\Sigma_{Rf}^\top} & \boldsymbol{\Sigma_f} \end{pmatrix} \mid \mathbf{Y} \sim \mathcal{W}^{-1}\left(T-1,\ T\hat{\boldsymbol{\Sigma}}_{\boldsymbol{Y}}\right), \tag{8}$$

where $\hat{\boldsymbol{\mu}}_{\boldsymbol{Y}} \equiv \frac{1}{T}\sum_{t=1}^T \boldsymbol{Y_t}$, $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{Y}} = \frac{1}{T}\sum_{t=1}^T (\boldsymbol{Y_t} - \widehat{\boldsymbol{\mu}}_{\boldsymbol{Y}})(\boldsymbol{Y_t} - \widehat{\boldsymbol{\mu}}_{\boldsymbol{Y}})^\top$, and $\mathcal{W}^{-1}$ is the inverse-Wishart distribution (a multivariate generalization of the inverse-gamma distribution).

Remark 1. *The iid assumption of $\boldsymbol{Y}_t$ can be relaxed using the approach in Müller (2013). Specifically, replacing the posterior covariance matrix, $\boldsymbol{\Sigma_Y}/T$, in equation (7) with a Newey and West (1987)-type heteroskedasticity- and autocorrelation-consistent (HAC) sandwich estimator of the covariance matrix. Furthermore, since the cross-sectional layer (below) of the BFM estimator is conditional on the draws from the posterior distribution of the time series layer, Gaussianity of the latter could also be relaxed.*

The Normal-inverse-Wishart posterior in equations (7)–(8) implies that we can sample $(\boldsymbol{\mu_Y}, \boldsymbol{\Sigma_Y})$ sequentially: We first simulate the covariance matrix $\boldsymbol{\Sigma_Y}$ from the inverse-Wishart distribution conditional on the data, and next conditional on the data and the draw of $\boldsymbol{\Sigma_Y}$, we draw $\boldsymbol{\mu_Y}$ from a multivariate normal distribution. We can further infer the factor exposures of asset returns, $\boldsymbol{\beta_f} \in \mathbb{R}^{N\times K}$, using $\boldsymbol{\Sigma_Y}$: $\boldsymbol{\beta_f} = \boldsymbol{\Sigma_{Rf}}\boldsymbol{\Sigma_f}^{-1}$. Similarly, we can compute the covariance matrix of $\boldsymbol{\epsilon_t}$ in equation (2): $\boldsymbol{\Sigma_\epsilon} = \boldsymbol{\Sigma_R} - \boldsymbol{\Sigma_{Rf}}\boldsymbol{\Sigma_f}^{-1}\boldsymbol{\Sigma_{Rf}^\top}$. Note that the above posterior distribution is robust even if a weak factor is included in $\boldsymbol{f}$. However, in the presence of a weak factor, the covariance between returns and factors, $\boldsymbol{\Sigma_{Rf}}$ (so as $\boldsymbol{\beta_f}$),

[2] If some factors in $\boldsymbol{f}_t$ are tradable, we allow them to be test assets in $\boldsymbol{R}_t$. In this case, $\boldsymbol{Y}_t$ is the union of factors and returns, i.e., $\boldsymbol{Y}_t = \boldsymbol{f}_t \cup \boldsymbol{R}_t$.

[3] The full derivation of this step can be found in Appendix A1 of Bryzgalova, Huang, and Julliard (2023).

converges to a singular matrix that will be later used as the regressors in the estimation of factors' risk premia.

If the model is correctly specified, in the sense that all true factors are included, expected returns of the assets should be fully explained by their risk exposures, $\boldsymbol{\beta_f}$, and the prices of risk $\boldsymbol{\lambda_f}$, that is: $\boldsymbol{\mu_R} = \boldsymbol{\beta_f}\boldsymbol{\lambda_f}$ in population. Given $(\boldsymbol{\mu_R}, \boldsymbol{\beta_f})$, we have the rather standard least square estimate $(\boldsymbol{\beta_f}^\top\boldsymbol{\beta_f})^{-1}\boldsymbol{\beta_f}^\top\boldsymbol{\mu_R}$.[4] Therefore, we can define our first estimator.

Definition 1 (Bayesian Fama-MacBeth (BFM))**.** *Conditional on $\boldsymbol{\mu_R}$, $\boldsymbol{\beta_f}$, and the data $\boldsymbol{Y} = \{\boldsymbol{Y_t}\}_{t=1}^T$, under the null hypothesis of a correctly specified expected return-beta representation ($\boldsymbol{\mu_R} = \boldsymbol{\beta_f}\boldsymbol{\lambda_f}$), the posterior distribution of $\boldsymbol{\lambda_f}$ is a Dirac at $(\boldsymbol{\beta_f}^\top\boldsymbol{\beta_f})^{-1}\boldsymbol{\beta_f}^\top\boldsymbol{\mu_R}$. Therefore, conditional on $\boldsymbol{Y} = \{\boldsymbol{Y_t}\}_{t=1}^T$, we first sample $\boldsymbol{\mu}_{Y,(j)}$ and $\boldsymbol{\Sigma}_{Y,(j)}$ from the Normal-inverse-Wishart distribution in equations (7)–(8), next compute $\boldsymbol{\beta}_{f,(j)} = \boldsymbol{\Sigma_{Rf,(j)}}\boldsymbol{\Sigma}_{f,(j)}^{-1}$, and finally compute $\boldsymbol{\lambda_{f,(j)}} = (\boldsymbol{\beta_{f,(j)}^\top}\boldsymbol{\beta_{f,(j)}})^{-1}\boldsymbol{\beta_{f,(j)}^\top}\boldsymbol{\mu_{R,(j)}}$.*

The BFM estimator defined above accounts for both uncertainty about expected returns (via the sampling of $\boldsymbol{\mu_R}$) and uncertainty about factor loadings (via the sampling of $\boldsymbol{\beta_f}$). Although the BFM estimator seems analogous to the frequentist FM approach, its hierarchical structure of posterior sampling enables us to detect weak and spurious factors in finite samples. Similar to the frequentist case, the near singularity of $(\boldsymbol{\beta_{f,(j)}^\top}\boldsymbol{\beta_{f,(j)}})^{-1}$ for weak factors causes the draws of $\boldsymbol{\lambda_{f,(j)}}$ to diverge. Nevertheless, $\boldsymbol{\beta_{f,(j)}^\top}\boldsymbol{\mu_{R,(j)}}$ tends to switch sign across draws because the posterior distribution of $\boldsymbol{\beta_f}$ for a weak factor is centred at around zero. Hence, the BFM estimator of $\boldsymbol{\lambda_f}$ puts substantial probability mass on both values above and below zero, and the resulting posterior credible intervals tend to include zero with high probability, making weak factors easily detectable.

The cross-sectional step of the FM regression is often performed via GLS rather than least squares. In our setting, under the null of the model, this leads to $\hat{\boldsymbol{\lambda}} = (\boldsymbol{\beta_f}^\top\boldsymbol{\Sigma_\epsilon}^{-1}\boldsymbol{\beta_f})^{-1}\boldsymbol{\beta_f}^\top\boldsymbol{\Sigma_\epsilon}^{-1}\boldsymbol{\mu_R}$. Therefore, we define the following GLS-type estimator.

Definition 2 (Bayesian Fama-MacBeth GLS (BFM-GLS))**.** *Conditional on $\boldsymbol{\mu_R}$, $\boldsymbol{\beta_f}$, $\boldsymbol{\Sigma_\epsilon}$, and $\boldsymbol{Y} = \{\boldsymbol{Y_t}\}_{t=1}^T$, under the null hypothesis of a correctly specified expected return-beta representation ($\boldsymbol{\mu_R} = \boldsymbol{\beta_f}\boldsymbol{\lambda_f}$), the posterior distribution of $\boldsymbol{\lambda_f}$ is a Dirac at $(\boldsymbol{\beta_f}^\top\boldsymbol{\Sigma_\epsilon}^{-1}\boldsymbol{\beta_f})^{-1}\boldsymbol{\beta_f}^\top\boldsymbol{\Sigma_\epsilon}^{-1}\boldsymbol{\mu_R}$. Therefore, conditional on $\boldsymbol{Y} = \{\boldsymbol{Y_t}\}_{t=1}^T$, we first sample $\boldsymbol{\mu_{Y,(j)}}$ and $\boldsymbol{\Sigma_{Y,(j)}}$ from the Normal-inverse-Wishart (7)–(8), compute $\boldsymbol{\beta_{f,(j)}} = \boldsymbol{\Sigma_{Rf,(j)}}\boldsymbol{\Sigma_{f,(j)}^{-1}}$, $\boldsymbol{\Sigma_{\epsilon,(j)}} = \boldsymbol{\Sigma_{R,(j)}} - \boldsymbol{\Sigma_{Rf,(j)}}\boldsymbol{\Sigma_{f,(j)}^{-1}}\boldsymbol{\Sigma_{Rf,(j)}^\top}$, and finally obtain $\boldsymbol{\lambda_{f,(j)}} = (\boldsymbol{\beta_{f,(j)}^\top}\boldsymbol{\Sigma_{\epsilon,(j)}^{-1}}\boldsymbol{\beta_{f,(j)}})^{-1}\boldsymbol{\beta_{f,(j)}^\top}\boldsymbol{\Sigma_{\epsilon,(j)}^{-1}}\boldsymbol{\mu_{R,(j)}}$.*

Our Bayesian framework can also quantify the posterior uncertainty about the cross-sectional fit of the model, that is, the cross-sectional R^2. Conditional on the posterior draws

[4]We can further include a common intercept in the cross-sectional regression. That is, $\boldsymbol{\mu_R} = \boldsymbol{\beta}\boldsymbol{\lambda}$, where $\boldsymbol{\beta} = (\boldsymbol{1}_N, \boldsymbol{\beta_f})$ and $\boldsymbol{\lambda} = (\lambda_c, \boldsymbol{\lambda_f^\top})^\top$. The least square estimate of $\boldsymbol{\lambda}$ is then $(\boldsymbol{\beta}^\top\boldsymbol{\beta})^{-1}\boldsymbol{\beta}^\top\boldsymbol{\mu_R}$.

of the parameters, we can easily obtain the posterior distribution of R^2, defined as

$$R^2_{ols} = 1 - \frac{(\boldsymbol{\mu_R} - \boldsymbol{\beta_f}\boldsymbol{\lambda_f})^\top(\boldsymbol{\mu_R} - \boldsymbol{\beta_f}\boldsymbol{\lambda_f})}{(\boldsymbol{\mu_R} - \bar{\mu}_R\mathbf{1_N})^\top(\boldsymbol{\mu_R} - \bar{\mu}_R\mathbf{1_N})}, \quad R^2_{gls} = 1 - \frac{(\boldsymbol{\mu_R} - \boldsymbol{\beta_f}\boldsymbol{\lambda_{f,gls}})^\top\boldsymbol{\Sigma_\epsilon}^{-1}(\boldsymbol{\mu_R} - \boldsymbol{\beta_f}\boldsymbol{\lambda_{f,gls}})}{(\boldsymbol{\mu_R} - \bar{\mu}_R\mathbf{1_N})^\top\boldsymbol{\Sigma_\epsilon}^{-1}(\boldsymbol{\mu_R} - \bar{\mu}_R\mathbf{1_N})}, \quad (9)$$

where $\bar{\mu}_R = \frac{1}{N}\sum_i^N \mu_{Ri}$. That is, for each posterior draw of $(\boldsymbol{\mu_R},\ \boldsymbol{\beta_f},\ \boldsymbol{\lambda_f},\ \boldsymbol{\Sigma_\epsilon})$, we can construct the corresponding draw for the R^2 from equation (9), hence tracing out its posterior distribution. We can think of equation (9) as the population R^2, where $\boldsymbol{\mu_R}$, $\boldsymbol{\beta_f}$, $\boldsymbol{\lambda_f}$, and $\boldsymbol{\Sigma_\epsilon}$ are unknown. After observing the data, we infer the posterior distribution of $\boldsymbol{\mu_R}$, $\boldsymbol{\beta_f}$, $\boldsymbol{\lambda_f}$, and $\boldsymbol{\Sigma_\epsilon}$, and from these we can recover the distribution of the R^2. Furthermore, we can estimate the posterior distribution of other test statistics, such as the cross-sectional T^2 statistic of Shanken (1985) and the F-statistic of Gibbons, Ross, and Shanken (1989).

The assumption that factor loadings $\boldsymbol{\beta_f}$ can entirely explain expected returns is likely too good to be true. Nevertheless, we can allow for the presence of pricing errors in the cross-sectional dimension: $\boldsymbol{\mu_R} = \boldsymbol{\beta_f}\boldsymbol{\lambda_f} + \boldsymbol{\alpha}$, where $\boldsymbol{\alpha}$ is a $N \times 1$ vector of pricing alphas. The BFM estimator in Definition 1 (the BFM-GLS estimator in Definition 2) remains consistent under the canonical assumption $\boldsymbol{\beta}_f^\top\boldsymbol{\alpha} = \mathbf{0}$ ($\boldsymbol{\beta}_f^\top\boldsymbol{\Sigma_\epsilon}^{-1}\boldsymbol{\alpha} = \mathbf{0}$).

2.2 Accounting for Omitted Factors

The BFM estimators in Subsection 2.1 rely on the tenuous assumption that a predetermined set of observable factors $\boldsymbol{f}$ subsumes all the pricing information in the SDF. Empirically, researchers are often concerned with omitting other pricing factors, which may lead to biased estimates of factors' risk premia. Past literature (see, e.g., Burmeister and McElroy (1988)) has acknowledged the omitted variable bias, but Giglio and Xiu (2021) is the first to formally propose a solution to estimate risk premia in the presence of omitted factors. We now show how to handle omitted factors within the BFM framework.

We follow Giglio and Xiu (2021) and estimate the risk premium of each factor separately.[5] In particular, we consider an univariate f_t, and assume the data-generating process:

$$f_t = \mu_f + \boldsymbol{\eta}_f^\top\boldsymbol{v}_t + w_t, \quad \boldsymbol{R}_t = \boldsymbol{\beta_v}\boldsymbol{\lambda_v} + \boldsymbol{\beta_v}\boldsymbol{v}_t + \boldsymbol{\epsilon}_t, \quad \boldsymbol{\Sigma_v} = \boldsymbol{I}_P, \quad \boldsymbol{v}_t \perp w_t \perp \boldsymbol{\epsilon}_t, \qquad (10)$$

where the vector $\boldsymbol{v}_t$ denotes P latent systematic factors that are normalized to be uncorrelated and have mean zero. Moreover, expected asset returns are fully explained by the exposures to $\boldsymbol{v}_t$, and f_t loads on $\boldsymbol{v}_t$ and possibly contains an unspanned component w_t (e.g., measurement error).

The data-generating process in equation (10) is equivalent to the following linear SDF:

$$M_t = 1 - \boldsymbol{\lambda_v}^\top\boldsymbol{\Sigma_v}^{-1}\boldsymbol{v}_t = 1 - \boldsymbol{\lambda_v}^\top\boldsymbol{v}_t, \qquad (11)$$

[5] In their language, "estimating the risk premium for one factor does not affect the estimation for the others at all, another important property of our estimator."

which implies that $\lambda_f = \boldsymbol{\lambda}_{\boldsymbol{v}}^\top \boldsymbol{\eta}_f$, using the definition of risk premia in equation (4).

Remark 2. *For expositional simplicity, we follow the canonical assumption that the latent factors are strong, in that the eigenvalues of $\boldsymbol{\beta}_{\boldsymbol{v}}^\top \boldsymbol{\beta}_{\boldsymbol{v}}$ coincide with the largest P eigenvalues of $\boldsymbol{\Sigma}_{\boldsymbol{R}}$. Yet, canonical latent factor selection procedures (e.g., Bai and Ng (2002)), which require both N and T to be sufficiently large, may not be appropriate with small N relative to T. In this case, model selection (and even Bayesian model averaging of the SDF) can be conducted over the space of latent factors using the spike-and-slab method of Bryzgalova, Huang, and Julliard (2023) to determine which columns in $\boldsymbol{\beta}_{\boldsymbol{v}}$ explain expected returns $\boldsymbol{\mu}_{\boldsymbol{R}}$. This differs from traditional latent factor selection in that it leverages cross-sectional pricing information, instead of merely fitting the time-series variation of returns.*

To show how to control for omitted factors, we first rewrite the SDF in equation (11) as the sum of two orthogonal components:

$$M_t = 1 - (\boldsymbol{\eta}_f^\top \boldsymbol{\lambda}_{\boldsymbol{v}})^\top (\boldsymbol{\eta}_f^\top \boldsymbol{\eta}_f)^{-1} \boldsymbol{\eta}_f^\top \boldsymbol{v}_t - (\boldsymbol{H}_1^\top \boldsymbol{\lambda}_{\boldsymbol{v}})^\top (\boldsymbol{H}_1^\top \boldsymbol{H}_1)^{-1} \boldsymbol{H}_1^\top \boldsymbol{v}_t, \tag{12}$$

where $\boldsymbol{H}_1^\top \boldsymbol{v}_t$ are the omitted latent factors, and $\boldsymbol{H}_1$ is chosen such that $\boldsymbol{H}_1^\top \boldsymbol{v}_t$ are orthogonal to the priced component of the observable factor, $\boldsymbol{\eta}_f^\top \boldsymbol{v}_t$.

Two major differences between the SDFs in equations (5) and (12) are noteworthy. First, only the spanned component of f_t, denoted as $\boldsymbol{\eta}_f^\top \boldsymbol{v}_t$, enters the SDF in equation (12), whereas the unspanned element w_t is eliminated. Second, $\boldsymbol{H}_1^\top \boldsymbol{v}_t$ are the priced components omitted by f_t. As Giglio and Xiu (2021) show, including the omitted latent factors as control is key to recovering the risk premium of the observable factor.

We now discuss the choice of $\boldsymbol{H}_1$ from the theoretical perspective. In particular, we aim to rotate $\boldsymbol{v}_t$ to $(\boldsymbol{\eta}_f^\top \boldsymbol{v}_t, \boldsymbol{H}_1 \boldsymbol{v}_t)$ such that the SDF in equation (12) contains the same pricing information as that in equation (11). Using the SDF in equation (12), we can derive the beta-pricing representation as follows:

$$\boldsymbol{\mu}_{\boldsymbol{R}} = -\mathrm{cov}(\boldsymbol{R}_t, M_t) = \boldsymbol{\beta}_{\boldsymbol{v}} \boldsymbol{\eta}_f (\boldsymbol{\eta}_f^\top \boldsymbol{\eta}_f)^{-1} \boldsymbol{\eta}_f^\top \boldsymbol{\lambda}_{\boldsymbol{v}} + \boldsymbol{\beta}_{\boldsymbol{v}} \boldsymbol{H}_1 (\boldsymbol{H}_1^\top \boldsymbol{H}_1)^{-1} \boldsymbol{H}_1^\top \boldsymbol{\lambda}_{\boldsymbol{v}} = \boldsymbol{\beta}_{\boldsymbol{v}} \boldsymbol{\lambda}_{\boldsymbol{v}},$$

which implies that $\boldsymbol{\eta}_f (\boldsymbol{\eta}_f^\top \boldsymbol{\eta}_f)^{-1} \boldsymbol{\eta}_f^\top + \boldsymbol{H}_1 (\boldsymbol{H}_1^\top \boldsymbol{H}_1)^{-1} \boldsymbol{H}_1^\top = \boldsymbol{I}_{\boldsymbol{P}}$. Note that the rank of $\boldsymbol{I}_{\boldsymbol{P}} - \boldsymbol{\eta}_f (\boldsymbol{\eta}_f^\top \boldsymbol{\eta}_f)^{-1} \boldsymbol{\eta}_f^\top$ is $P - 1$ when $\boldsymbol{\eta}_f$ is not a zero vector. Hence, a natural choice of $\boldsymbol{H}_1$ is a $P \times (P-1)$ matrix collecting the first $(P-1)$ columns of the unitary matrix from the singular value decomposition (SVD) of $\boldsymbol{I}_P - \boldsymbol{\eta}_f (\boldsymbol{\eta}_f^\top \boldsymbol{\eta}_f)^{-1} \boldsymbol{\eta}_f^\top$: That is, $\boldsymbol{H}_1^\top \boldsymbol{H}_1 = \boldsymbol{I}_{P-1}$, $\boldsymbol{H}_1^\top \boldsymbol{\eta}_{\boldsymbol{f}} = \boldsymbol{0}$, and $\boldsymbol{H}_1 \boldsymbol{H}_1^\top = \boldsymbol{I}_{\boldsymbol{P}} - \boldsymbol{\eta}_f (\boldsymbol{\eta}_f^\top \boldsymbol{\eta}_f)^{-1} \boldsymbol{\eta}_f^\top$.

BFM does not directly estimate $\boldsymbol{\eta}_f$. Instead, conditional on the time-series step (again given by the posterior distributions in equations (7)–(8)), we obtain the estimates of $\boldsymbol{\Sigma}_{\boldsymbol{R}}$ and $\boldsymbol{\beta}_{\boldsymbol{f}}$ $(= \boldsymbol{\Sigma}_{\boldsymbol{R}\boldsymbol{f}} / \sigma_f^2)$. We then identify $\boldsymbol{\beta}_{\boldsymbol{v}}$ as the eigenvectors of $\boldsymbol{\Sigma}_{\boldsymbol{R}}$ corresponding to the

P largest eigenvalues,[6] and $\boldsymbol{\eta}_f$ equals $(\boldsymbol{\beta}_{\boldsymbol{v}}^\top \boldsymbol{\beta}_{\boldsymbol{v}})^{-1}\boldsymbol{\beta}_{\boldsymbol{v}}^\top \boldsymbol{\Sigma}_{\boldsymbol{Rf}}$, which is further used to identify $\boldsymbol{H}_1$. Based on $\boldsymbol{\mu}_{\boldsymbol{R}} = \boldsymbol{\beta}_{\tilde{\boldsymbol{v}}}\boldsymbol{\lambda}_{\tilde{\boldsymbol{v}}}$, where $\boldsymbol{\beta}_{\tilde{\boldsymbol{v}}} = (\boldsymbol{\beta}_{\boldsymbol{f}}, \boldsymbol{\beta}_{\boldsymbol{v}}\boldsymbol{H}_1)$, the risk premia estimates, $\boldsymbol{\lambda}_{\tilde{\boldsymbol{v}}}$, are $(\boldsymbol{\beta}_{\tilde{\boldsymbol{v}}}^\top \boldsymbol{\beta}_{\tilde{\boldsymbol{v}}})^{-1}\boldsymbol{\beta}_{\tilde{\boldsymbol{v}}}^\top \boldsymbol{\mu}_{\boldsymbol{R}}$, and λ_f is the first element of $\boldsymbol{\lambda}_{\tilde{\boldsymbol{v}}}$. Therefore, the BFM estimators with omitted factors are equivalent to those in Definitions 1 and 2 by replacing $\boldsymbol{\beta}_{\boldsymbol{f}}$ with $\boldsymbol{\beta}_{\tilde{\boldsymbol{v}}}$, yielding the risk premium estimate of f_t: $\lambda_f = \sigma_f^2(\boldsymbol{\eta}_f^\top \boldsymbol{\eta}_f)^{-1}\boldsymbol{\eta}_f^\top \boldsymbol{\lambda}_{\boldsymbol{v}}$.

The BFM estimator with the additional control $(\boldsymbol{\beta}_{\boldsymbol{v}}\boldsymbol{H}_1)$ is robust to the presence of spurious and omitted factors. Nevertheless, f_t's risk premium based on these FM regressions has an unsatisfactory property, as highlighted in the Remark below.

Remark 3. *The BFM estimate, $\lambda_f = \sigma_f^2(\boldsymbol{\eta}_f^\top \boldsymbol{\eta}_f)^{-1}\boldsymbol{\eta}_f^\top \boldsymbol{\lambda}_{\boldsymbol{v}}$, differs from the definition in equation (4), where $\lambda_f = \boldsymbol{\eta}_f^\top \boldsymbol{\lambda}_{\boldsymbol{v}}$, by a scale factor capturing the proportion of variation in f_t explained by the spanned component $\boldsymbol{\eta}_f^\top \boldsymbol{v}_t$ (i.e., $\boldsymbol{\eta}_f^\top \boldsymbol{\eta}_f/\sigma_f^2$). The existence of the term $\sigma_f^2(\boldsymbol{\eta}_f^\top \boldsymbol{\eta}_f)^{-1} = 1 + \sigma_w^2(\boldsymbol{\eta}_f^\top \boldsymbol{\eta}_f)^{-1}$ is economically unsound: As the variance of the unspanned component w_t increases, the FM estimates of risk premia are mechanically inflated. This is inconsistent with canonical portfolio choice theory, which conjectures that unspanned risk should not be compensated. This inconsistency always exists unless f_t is fully spanned by the sources of risk in the SDF (equivalently, $\sigma_w^2 = 0$). Moreover, the scale factor captures exactly the source of weak identification: $(\boldsymbol{\eta}_f^\top \boldsymbol{\eta}_f)^{-1}$ tends to be singular for a weak factor.*

To resolve the above inconsistency, we leverage the definition of risk premium in equation (4), $-\text{cov}(M_t, f_t) = \boldsymbol{\eta}_f^\top \boldsymbol{\lambda}_{\boldsymbol{v}}$, where the unspanned component no longer determines the risk compensation.[7] That is, instead of directly estimating λ_f, we revise the second (cross-sectional) step of the canonical FM estimation to be over the space of latent factors: We regress $\boldsymbol{\mu}_{\boldsymbol{R}}$ on $\boldsymbol{\beta}_{\boldsymbol{v}}$ to obtain the risk premia of latent factors $\boldsymbol{v}_t$, $\boldsymbol{\lambda}_{\boldsymbol{v}} = (\boldsymbol{\beta}_{\boldsymbol{v}}^\top \boldsymbol{\beta}_{\boldsymbol{v}})^{-1}\boldsymbol{\beta}_{\boldsymbol{v}}^\top \boldsymbol{\mu}_{\boldsymbol{R}}$. Since the time-series step in equations (7)–(8) determines $\boldsymbol{\eta}_f$, which equals $(\boldsymbol{\beta}_{\boldsymbol{v}}^\top \boldsymbol{\beta}_{\boldsymbol{v}})^{-1}\boldsymbol{\beta}_{\boldsymbol{v}}^\top \boldsymbol{\Sigma}_{\boldsymbol{Rf}}$, the natural estimate of f_t's risk premium is then $\lambda_f = \boldsymbol{\lambda}_{\boldsymbol{v}}^\top (\boldsymbol{\beta}_{\boldsymbol{v}}^\top \boldsymbol{\beta}_{\boldsymbol{v}})^{-1}\boldsymbol{\beta}_{\boldsymbol{v}}^\top \boldsymbol{\Sigma}_{\boldsymbol{Rf}}$. This modified version of the BFM estimator is summarized in Definition 3.

Definition 3 (Bayesian Fama-MacBeth with omitted factors (BFM-OMIT))**.** *Conditional on the j-th posterior draw of the first two moments of the data, including $\boldsymbol{\mu}_{\boldsymbol{R},(\boldsymbol{j})}$, $\boldsymbol{\beta}_{\boldsymbol{f},(\boldsymbol{j})}$, $\boldsymbol{\Sigma}_{\boldsymbol{R},(\boldsymbol{j})}$, $\boldsymbol{\Sigma}_{\boldsymbol{Rf},(\boldsymbol{j})}$, and the data $\boldsymbol{Y} = \{\boldsymbol{Y}_{\boldsymbol{t}}\}_{t=1}^T$, we perform the eigendecomposition of $\boldsymbol{\Sigma}_{\boldsymbol{R},(\boldsymbol{j})}$, $\boldsymbol{\Sigma}_{\boldsymbol{R},(\boldsymbol{j})} = \boldsymbol{U}_{(j)}\boldsymbol{\Lambda}_{(j)}\boldsymbol{U}_{(j)}^\top$, and identify $\boldsymbol{\beta}_{\boldsymbol{v},(\boldsymbol{j})}$ as the first P columns of $\boldsymbol{U}_{(j)}\boldsymbol{\Lambda}_{(j)}^{\frac{1}{2}}$. Under the null hypothesis of a correctly specified expected return-beta representation ($\boldsymbol{\mu}_{\boldsymbol{R}} = \boldsymbol{\beta}_{\boldsymbol{v}}\boldsymbol{\lambda}_{\boldsymbol{v}}$), the*

[6] Consider the eigendecomposition of $\boldsymbol{\Sigma}_{\boldsymbol{R}}$: $\boldsymbol{\Sigma}_{\boldsymbol{R}} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^\top$. $\boldsymbol{\beta}_{\boldsymbol{v}}$ collect the first P columns of $\boldsymbol{U}\boldsymbol{\Lambda}^{\frac{1}{2}}$.

[7] This definition of risk premium satisfies the rotation invariance property, as highlighted in Giglio and Xiu (2021). For any nonsingular matrix $\boldsymbol{A} \in \mathbb{R}^{P\times P}$, let $\boldsymbol{z}_t = \boldsymbol{A}\boldsymbol{v}_t$ be the rotated latent factors, and $\boldsymbol{\beta}_{\boldsymbol{z}} = \boldsymbol{\beta}_{\boldsymbol{v}}\boldsymbol{A}^{-1}$. This rotation implies that the loadings of f_t on $\boldsymbol{z}_t$ is equal to $\text{cov}(f_t, \boldsymbol{z}_t)\boldsymbol{\Sigma}_{\boldsymbol{z}}^{-1} = \boldsymbol{\eta}_{\boldsymbol{f}}^\top \boldsymbol{A}^\top(\boldsymbol{A}\boldsymbol{A}^\top)^{-1} = \boldsymbol{\eta}_{\boldsymbol{f}}^\top \boldsymbol{A}^{-1}$. Moreover, the risk premium of $\boldsymbol{z}_t$ is $\boldsymbol{\lambda}_{\boldsymbol{z}} = (\boldsymbol{\beta}_{\boldsymbol{z}}^\top \boldsymbol{\beta}_{\boldsymbol{z}})^{-1}\boldsymbol{\beta}_{\boldsymbol{z}}^\top \boldsymbol{\mu}_{\boldsymbol{R}} = \boldsymbol{A}\boldsymbol{\lambda}_{\boldsymbol{v}}$. Hence, the risk premia of f_t is $\boldsymbol{\eta}_{\boldsymbol{f}}^\top \boldsymbol{A}^{-1}\boldsymbol{\lambda}_{\boldsymbol{z}} = \boldsymbol{\eta}_{\boldsymbol{f}}^\top \boldsymbol{\lambda}_{\boldsymbol{v}}$. In short, the risk premium of f_t is invariant to any nonsingular rotation of $\boldsymbol{v}_t$.

posterior distribution of $\boldsymbol{\lambda}_{\boldsymbol{v},(\boldsymbol{j})}$ *is a Dirac at* $(\boldsymbol{\beta}_{\boldsymbol{v},(\boldsymbol{j})}^\top \boldsymbol{\beta}_{\boldsymbol{v},(\boldsymbol{j})})^{-1} \boldsymbol{\beta}_{\boldsymbol{v},(\boldsymbol{j})}^\top \boldsymbol{\mu}_{\boldsymbol{R},(\boldsymbol{j})}$, *and the risk premium of* f_t *is* $\lambda_{f,(j)} = \boldsymbol{\lambda}_{\boldsymbol{v},(\boldsymbol{j})}^\top (\boldsymbol{\beta}_{\boldsymbol{v},(\boldsymbol{j})}^\top \boldsymbol{\beta}_{\boldsymbol{v},(\boldsymbol{j})})^{-1} \boldsymbol{\beta}_{\boldsymbol{v},(\boldsymbol{j})}^\top \boldsymbol{\Sigma}_{\boldsymbol{Rf},(\boldsymbol{j})}$.

We now simplify the formulation of λ_f to gain economic insights. Consider the eigen-decomposition of $\boldsymbol{\Sigma}_{\boldsymbol{R}}$: $\boldsymbol{\Sigma}_{\boldsymbol{R}} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^\top$. Since $\boldsymbol{\Sigma}_{\boldsymbol{R}}$ is drawn from its posterior distribution in equation (8), the estimation uncertainty of the entire eigenspace of $\boldsymbol{\Sigma}_{\boldsymbol{R}}$ is fully accounted. We let $\boldsymbol{U}_P$ collect the first P columns of $\boldsymbol{U}$, and $\boldsymbol{\Lambda}_P$ collect the first P rows and columns of $\boldsymbol{\Lambda}$. We can then show that $\lambda_f = \boldsymbol{\Sigma}_{\boldsymbol{Rf}}^\top \boldsymbol{U}_P \boldsymbol{\Lambda}_P^{-1} \boldsymbol{U}_P^\top \boldsymbol{\mu}_{\boldsymbol{R}}$. Note that $\boldsymbol{U}_P^\top \boldsymbol{\mu}_{\boldsymbol{R}}$ are the expected returns of the first P PCs of $\boldsymbol{R}_t$, i.e., $\boldsymbol{U}_P^\top \boldsymbol{R}_t$, with a covariance matrix $\boldsymbol{\Lambda}_P$, and $\boldsymbol{\Sigma}_{\boldsymbol{Rf}}^\top \boldsymbol{U}_P \boldsymbol{\Lambda}_P^{-1}$ is the loadings of f_t on the PCs $\boldsymbol{U}_P^\top \boldsymbol{R}_t$. Therefore, λ_f can be interpreted as the expected return of the factor-mimicking portfolio that consists of PCs. A special case is $P = N$, which implies that $\lambda_f = \boldsymbol{\Sigma}_{\boldsymbol{Rf}}^\top \boldsymbol{\Sigma}_{\boldsymbol{R}}^{-1} \boldsymbol{\mu}_{\boldsymbol{R}}$. In this case, we project f_t onto the original asset space with a dimension N. Although $\boldsymbol{\Sigma}_{\boldsymbol{Rf}}^\top \boldsymbol{\Sigma}_{\boldsymbol{R}}^{-1} \boldsymbol{\mu}_{\boldsymbol{R}}$ is theoretically a more robust estimator of λ_f, the estimate of $\boldsymbol{\Sigma}_{\boldsymbol{R}}^{-1}$ tends to be rather noisy due to highly correlated asset returns. Employing PCs can be seen as an empirical shortcut to reduce the dimensionality and to approximate the priced source of risk in the nontraded factor.

Finally, we can relax the assumption of zero pricing errors in equation (10): $\boldsymbol{\mu}_{\boldsymbol{R}} = \boldsymbol{\beta}_{\boldsymbol{v}} \boldsymbol{\lambda}_{\boldsymbol{v}} + \boldsymbol{\alpha}$, where $\boldsymbol{\alpha} \perp \boldsymbol{\beta}_{\boldsymbol{v}}$. The latter orthogonality assumption is theoretically not as restrictive as the one in Section 2.1 because we can always include a relatively large number of latent factors and select the priced ones (as described in Remark 2) to ensure $\boldsymbol{\alpha}$ and $\boldsymbol{\lambda}_{\boldsymbol{v}}$ to be as orthogonal as possible.

3 Simulation

We now investigate the performance of our BFM estimators via Monte Carlo simulations. We consider both strong and weak factors, and allow for potential misspecification in the linear factor model.

The cross-section of excess returns comprises the 25 Fama-French portfolios sorted by size and value, plus 12 industry portfolios. Factors and test asset returns are simulated jointly from normal distributions, as follows:

$$f_{t,useless} \overset{\text{iid}}{\sim} \mathcal{N}(0, (1\%)^2), \qquad \begin{pmatrix} \boldsymbol{R_t} \\ f_{t,hkm} \end{pmatrix} \overset{\text{iid}}{\sim} \mathcal{N}\left(\begin{bmatrix} \bar{\boldsymbol{\mu}}_{\boldsymbol{R}} \\ \bar{f}_{hkm} \end{bmatrix}, \begin{bmatrix} \hat{\boldsymbol{\Sigma}}_{\boldsymbol{R}} & \widehat{\boldsymbol{C}}_{\boldsymbol{hkm}} \\ \widehat{\boldsymbol{C}}_{\boldsymbol{hkm}}^\top & \widehat{\sigma}_{hkm}^2 \end{bmatrix} \right), \tag{13}$$

where the nontraded intermediary factor from He, Kelly, and Manela (2017) (HKM factor) is the strong factor, the mean vector and covariance matrix of $(\boldsymbol{R_t}^\top, f_{t,hkm})$ are calibrated as their sample estimates, and the useless factor is simulated from an independent standard normal distribution. All the model parameters are estimated on monthly data from January

1970 to December 2019.

The fundamental element in simulations is the vector of expected asset returns, $\bar{\boldsymbol{\mu}}_{\boldsymbol{R}}$. We consider two settings that are mapped into the frameworks in Sections 2.1 and 2.2.

First, we assume that the HKM factor is in the linear SDF, but it does not fully explain the expected returns, with pricing errors orthogonal to the factor loadings, as follows:

$$\bar{\boldsymbol{\mu}}_{\boldsymbol{R}} = \frac{\widehat{\boldsymbol{C}}_{\boldsymbol{hkm}}}{\widehat{\sigma}^2_{hkm}} \widehat{\lambda}_{hkm} + \widehat{\boldsymbol{\alpha}}_{hkm}, \quad \widehat{\boldsymbol{\alpha}}_{hkm} \perp \widehat{\boldsymbol{C}}_{\boldsymbol{hkm}}, \tag{14}$$

where $\widehat{\lambda}_{hkm}$ is estimated via the two-step FM procedure, and pricing errors $\widehat{\boldsymbol{\alpha}}_{hkm}$ are calibrated such that they are orthogonal to factor exposures and that the pseudo-true cross-sectional adjusted R^2_{ols} is about 50.2% (and $R^2_{gls} = 37.0\%$).

To illustrate the properties of the frequentist and Bayesian approaches in the simulation based on equation (14), we consider three estimation setups: (a) the model includes only a strong factor (HKM); (b) the model includes only a useless factor; and (c) the model includes both strong and useless factors, with the following sample sizes: T = 200, 600, 1,000, and 20,000. We compare the performance of the OLS/GLS standard frequentist and Bayesian Fama-MacBeth estimators (FM and BFM, correspondingly) with the focus on risk premia recovery, testing, and identification of strong and useless factors.

Panel A of Table 1 reports the simulation results where the data are drawn from equations (13) and (14), and the strong and weak factors are simultaneously included in the regressions. We compare the performance of the frequentist Fama-MacBeth tests, which are constructed using standard t-statistics adjusted for Shanken correction, with our BFM estimators described in Definitions 1 and 2. In the BFM and BFM-GLS, we rely on the quantiles of the posterior distribution to form credible intervals for parameters. The last two columns also report the quantiles of the posterior means of R^2 across the simulations. We observe that the cross-sectional fit is rather noisy, with wide 90% confidence intervals across simulations, even in a relatively large sample of 1,000 months. Moreover, as pointed out by the previous literature, the canonical frequentist Fama-MacBeth often over-rejects the null hypothesis of zero risk premia for useless factors: Its OLS/GLS t-statistic would be above a 10%-critical value in more than 60% of the simulations. In contrast, our BFM estimators tend to deliver the proper coverages of the true values, and reject the null of no risk premia with frequency asymptotically approaching the size of the tests.[8]

In the second simulation setup, expected returns are explained by their exposures to

[8]Table A1 in the appendix reports the estimation results in which strong and weak factors are individually estimated by the frequentist and Bayesian Fama-MacBeth regressions. The simulation results are similar to those in Panel A of Table 1. Note that a common intercept is not included in the cross-sectional step; hence, the adjusted R^2 can be significantly negative in some simulations.

Table 1: Tests of risk premia in a misspecified model with both useless and strong factors

	T	λ_{strong} 10%	5%	1%	$\lambda_{useless}$ 10%	5%	1%	R^2_{adj} 5th	95th
Panel A: Simulations based on equations (13) and (14)									
FM-OLS	200	0.086	0.036	0.005	0.066	0.018	0.000	-29.4%	48.5%
	600	0.095	0.047	0.006	0.100	0.036	0.001	-10.2%	56.1%
	1,000	0.088	0.038	0.008	0.126	0.046	0.002	3.7%	58.9%
	20,000	0.098	0.046	0.010	0.634	0.513	0.217	42.6%	60.3%
BFM-OLS	200	0.048	0.016	0.002	0.001	0.001	0.000	-13.9%	24.3%
	600	0.083	0.033	0.007	0.007	0.002	0.000	-5.0%	37.7%
	1,000	0.078	0.033	0.005	0.008	0.003	0.001	3.1%	43.5%
	20,000	0.065	0.031	0.003	0.078	0.035	0.008	43.0%	56.5%
FM-GLS	200	0.102	0.050	0.006	0.194	0.116	0.036	-31.3%	31.3%
	600	0.105	0.051	0.011	0.246	0.160	0.062	-11.7%	38.6%
	1,000	0.094	0.049	0.009	0.291	0.207	0.091	-2.8%	40.5%
	20,000	0.092	0.048	0.013	0.678	0.620	0.521	23.9%	39.4%
BFM-GLS	200	0.060	0.018	0.001	0.020	0.005	0.001	-0.7%	25.3%
	600	0.089	0.041	0.007	0.028	0.011	0.001	2.2%	31.9%
	1,000	0.085	0.038	0.009	0.035	0.009	0.000	7.6%	35.0%
	20,000	0.082	0.039	0.007	0.084	0.044	0.011	30.7%	40.8%
Panel B: Simulations based on equations (13) and (15)									
FM-OLS	200	0.197	0.128	0.030	0.124	0.049	0.002	-37.8%	31.1%
	600	0.398	0.278	0.108	0.233	0.128	0.018	-27.4%	26.5%
	1,000	0.558	0.428	0.220	0.323	0.186	0.033	-21.8%	24.9%
	20,000	0.963	0.950	0.894	0.795	0.727	0.446	-6.1%	19.2%
FM-GLS	200	0.110	0.054	0.009	0.322	0.232	0.116	-34.7%	14.0%
	600	0.120	0.058	0.015	0.460	0.378	0.240	-23.1%	10.7%
	1,000	0.107	0.059	0.012	0.510	0.436	0.296	-20.3%	9.6%
	20,000	0.178	0.120	0.044	0.758	0.712	0.631	-13.9%	5.1%
BFM-OMIT ($P = 5$)	200	0.052	0.017	0.001	0.004	0.000	0.000	26.5%	59.2%
	600	0.092	0.041	0.008	0.041	0.013	0.001	21.2%	58.6%
	1,000	0.087	0.039	0.005	0.053	0.020	0.001	22.0%	57.4%
	20,000	0.094	0.049	0.008	0.097	0.041	0.008	38.4%	48.9%
BFM-OMIT ($P = 3$)	200	0.057	0.020	0.001	0.007	0.001	0.000	12.6%	49.2%
	600	0.093	0.048	0.009	0.062	0.022	0.003	9.4%	50.4%
	1,000	0.097	0.049	0.011	0.066	0.025	0.004	10.5%	48.9%
	20,000	0.466	0.329	0.137	0.096	0.044	0.007	25.0%	37.1%
BFM-OMIT ($P = 6$)	200	0.053	0.020	0.001	0.005	0.000	0.000	31.6%	64.1%
	600	0.090	0.040	0.008	0.043	0.014	0.001	24.6%	62.1%
	1,000	0.089	0.039	0.005	0.054	0.021	0.001	25.7%	60.2%
	20,000	0.094	0.049	0.008	0.093	0.043	0.008	36.8%	47.7%

Frequency of rejecting the null hypothesis $H_0 : \lambda_i = \lambda_i^*$ for pseudo-true values of λ_{strong} and $\lambda^*_{useless} \equiv 0$ in a misspecified model with a strong and a useless factor. Last two columns: 5th and 95th percentiles of cross-sectional R^2_{adj} across 2,000 simulations, evaluated at the point estimates for FM and at the posterior mean for BFM. In Panel A, the data are simulated based on equations (13) and (14), with the true value of R^2_{adj} equal to 50.2% (37.0%) for OLS (GLS) estimation. In this simulation setup, we compare the frequentist FM methods with our BFM estimators in Definitions 1 and 2. In Panel B, we simulate the data from equations (13) and (15), with the true R^2_{adj} equal to 44.0%. We estimate the risk premia based on the BFM-OMIT method in Definition 3. As a comparison, we also show the OLS/GLS frequentist FM estimates.

latent factors. The HKM factor, instead, does not enter the SDF and is priced because it loads on these priced latent factors. In this case, we assume the formulation of $\bar{\boldsymbol{\mu}}_{\boldsymbol{R}}$ as follows:

$$\bar{\boldsymbol{\mu}}_{\boldsymbol{R}} = \widehat{\boldsymbol{\beta}}_{\boldsymbol{v}} \widehat{\boldsymbol{\lambda}}_{\boldsymbol{v}} + \widehat{\boldsymbol{\alpha}}_{\boldsymbol{v}}, \quad \widehat{\boldsymbol{\alpha}}_{\boldsymbol{v}} \perp \widehat{\boldsymbol{\beta}}_{\boldsymbol{v}}, \tag{15}$$

where $\widehat{\boldsymbol{\beta}}_{\boldsymbol{v}}$ are calibrated as the eigenvectors of $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{R}}$ corresponding to the largest five eigenvalues, and $\widehat{\boldsymbol{\lambda}}_{\boldsymbol{v}}$ and $\widehat{\boldsymbol{\alpha}}_{\boldsymbol{v}}$ are the sample estimates from the two-step regression. Unlike the simulations based on equation (14), we estimate the risk premia of strong and weak factors individually in this simulation setting.

In Panel B of Table 1, we investigate the performance of our BFM-OMIT estimator in Definition 3. Not surprisingly, the traditional frequentist FM estimates are biased—which can be seen from the over-rejection of the null hypothesis—due to the omitted variable bias, measurement error, and weak identification of the simulated factors. In contrast, our BFM-OMIT estimator, using the correct number of latent components (Panel with $P = 5$), provides appropriate posterior coverages for the true risk premia, no matter whether a factor is strong or weak. Next, we take a more conservative strategy to include one additional, unpriced latent factor into the cross-sectional step (Panel with $P = 6$)—similar simulation results are obtained. Nevertheless, the correctness of the BFM-OMIT approach requires us to include a sufficient amount of latent factors in the cross-sectional step. For instance, when the number of latent components is three ($P = 3$), the risk premia estimates of the strong factor are significantly biased asymptotically.

4 Empirical Examples

This section applies the BFM methods to estimate the risk premia of real durable and nondurable consumption growth, as well as the nontraded HKM intermediary factor.[9] As a comparison, we report the frequentist FM estimates. We use the same cross-section of 37 equity portfolios as in the simulation studies. Table 2 reports the risk premia estimates (the λ_f columns) and the cross-sectional fit (the R^2_{adj} columns). For the frequentist FM regressions, we report the t-statistics within the parentheses, whereas the 90% posterior credible intervals (CIs) are shown in the square brackets for the BFM estimates.

Our BFM estimators in Definitions 1 and 2 are robust to the weak identification. Take durable consumption growth, for example. Note that durable consumption growth has a weak correlation with asset returns and, hence, is very likely to be a weak factor. Although the frequentist FM-OLS estimate of its risk premium is significant at the 10% significance

[9]Data on durable and nondurable consumption growth are from Table 7.1 of the Bureau of Economic Research. The HKM intermediary factor is downloaded from the authors' websites.

Table 2: Risk premia estimates of three examples

Frequentist FM				BFM				BFM-OMIT			
FM-OLS		FM-GLS		BFM-OLS		BFM-GLS		$P = 5$		$P = 10$	
λ_f	R^2_{adj}	λ_f	R^2_{adj}	λ_f	R^2_{adj}	λ_f	R^2_{adj}	λ_f	R^2_{adj}	λ_f	R^2_{adj}
Panel A. Quarterly durable consumption growth: 1963Q3–2019Q4											
1.468	-418%	-0.105	-40%	1.111	-514%	-0.072	-29%	0.000	37%	0.017	61%
(1.892)		(-0.487)		[-0.264, 2.184]		[-0.383, 0.236]		[-0.060, 0.061]		[-0.057, 0.090]	
Panel B. Quarterly nondurable consumption growth: 1963Q3–2019Q4											
1.266	-39%	0.375	-35%	1.318	-137%	0.235	-27%	0.034	37%	0.045	61%
(2.141)		(1.716)		[0.455, 2.637]		[-0.099, 0.575]		[-0.015, 0.086]		[-0.022, 0.113]	
Panel C. HKM intermediary factor: 1970M1–2019M12											
0.188	-48%	0.119	-46%	0.188	-35%	0.118	-34%	0.112	23%	0.099	50%
(3.341)		(2.514)		[0.092, 0.284]		[0.035, 0.201]		[0.057, 0.167]		[0.040, 0.158]	

This table reports the risk premia estimates of three economic factors (real durable and nondurable consumption growth, and the HKM intermediary factor) in the cross-section of 25 Fama-French size- and value-sorted portfolios plus 12 industry portfolios. The estimations are based on the traditional frequentist FM regression, our BFM approach in Definitions 1 and 2, and the BFM-OMIT method in Definition 3. We estimate their risk premia individually, excluding the common intercept in the cross-sectional step. We report both the risk premia estimates (λ_f) and the cross-sectional fit (adjusted R^2 defined in equation (9)). In addition to the point estimates, we tabulate the t-statistics of λ_f (see the numbers within parentheses) in the frequentist FM estimation and the 90% posterior credible intervals (see the numbers with brackets) of λ_f in the BFM and BFM-OMIT methods.

level, the 90% CIs given by BFM-OLS contain the zero risk premium; hence, in the Bayesian method, we cannot reject the null hypothesis of zero risk premium. This observation echoes our previous simulation studies, which show that weak identification is inconsequential in our Bayesian methods.

Nevertheless, BFM-OLS and BFM-GLS are built upon the assumption that the single factor itself is the only relevant priced source of risk in the SDF, which is, ex-post, almost impossible due to the extremely negative cross-sectional R^2.[10]

To account for the model misspecification, we repeat the same exercise using the BFM-OMIT approach described in Definition 3. Unlike the models containing only univariate observable factors, the five (ten) latent factors explain 23–37% (50–61%) of the cross-sectional variation. Also, risk premia estimates based on BFM-OMIT tend to be smaller than those of frequentist FM or BFM. For instance, nondurable consumption growth commands a significantly positive risk premium (about 1.3) in both FM and BFM OLS estimations, whereas the BFM-OMIT estimates (0.034–0.045) are tiny and not statistically significant.

Different from the two consumption growth measures, the HKM factor is consistently priced in the cross-section. Although the factor itself cannot explain the cross-sectional spreads of average returns, it loads on the priced latent factors and, therefore, commands a significantly positive risk premium. This observation echoes the discussion in Chapter 13.4

[10]Note that the (adjusted) R^2s are negative for most entries because a common intercept is not included in the cross-sectional step of the two-step estimation.

of Cochrane (2005): A factor being priced (that is, the factor-mimicking portfolio carrying a significant risk premium) is distinct from a factor entering the SDF (that is, a factor capable of pricing assets given the other factors). As a final caveat to the cross-sectional step in the two-pass regressions, researchers should not regress mean asset returns ($\boldsymbol{\mu_R}$) on the beta loadings on the observable factor ($\boldsymbol{\beta_f}$) to obtain the risk premium estimate, unless they have strong beliefs that the priced risks in the SDF completely span the tested factor.

References

Anatolyev, S., and A. Mikusheva (2022): "Factor models with many assets: strong factors, weak factors, and the two-pass procedure," *Journal of Econometrics*, 229(1), 103–126.

Bai, J., and S. Ng (2002): "Determining the number of factors in approximate factor models," *Econometrica*, 70(1), 191–221.

Bryzgalova, S. (2015): "Spurious Factors in Linear Asset Pricing Models," working paper.

Bryzgalova, S., J. Huang, and C. Julliard (2023): "Bayesian solutions for the factor zoo: We just ran two quadrillion models," *Journal of Finance*, 78(1), 487–557.

Burmeister, E., and M. B. McElroy (1988): "Joint estimation of factor sensitivities and risk premia for the arbitrage pricing theory," *Journal of Finance*, 43(3), 721–733.

Burnside, C. (2016): "Identification and inference in linear stochastic discount factor models with excess returns," *Journal of Financial Econometrics*, 14(2), 295–330.

Cochrane, J. H. (2005): *Asset Pricing: Revised Edition*. Princeton University Press, Princeton, New Jersey.

Dickerson, A., C. Julliard, and P. Mueller (2024): "The Corporate Bond Factor Zoo," London School of Economics Manuscript, available at SSRN: https://ssrn.com/abstract=4589786.

Fama, E. F., and J. D. MacBeth (1973): "Risk, Return, and Equilibrium: Empirical Tests," *Journal of Political Economy*, 81(3), 607–636.

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

Giglio, S., and D. Xiu (2021): "Asset pricing with omitted factors," *Journal of Political Economy*, 129(7), 1947–1990.

Gospodinov, N., R. Kan, and C. Robotti (2014): "Misspecification-Robust Inference in Linear Asset-Pricing Models with Irrelevant Risk Factors," *Review of Financial Studies*, 27(7), 2139–2170.

——— (2017): "Spurious inference in reduced-rank asset-pricing models," *Econometrica*, 85(5), 1613–1628.

——— (2019): "Too Good to Be True? Fallacies in Evaluating Risk Factor Models," *Journal of Financial Economics*, 132(2), 451–471.

He, Z., B. Kelly, and A. Manela (2017): "Intermediary asset pricing: New evidence from many asset classes," *Journal of Financial Economics*, 126(1), 1–35.

Kan, R., and C. Zhang (1999a): "GMM Tests of Stochastic Discount Factor Models with Useless Factors," *Journal of Financial Economics*, 54(1), 103–127.

——— (1999b): "Two-Pass Tests of Asset Pricing Models with Useless Factors," *Journal of Finance*, 54(1), 203–235.

Kleibergen, F. (2009): "Tests of Risk Premia in Linear Factor Models," *Journal of Econometrics*, 149(2), 149–173.

Kleibergen, F., and Z. Zhan (2015): "Unexplained Factors and Their Effects on Second Pass R-Squareds," *Journal of Econometrics*, 189(1), 101–116.

——— (2020): "Robust inference for consumption-based asset pricing," *Journal of Finance*, 75(1), 507–550.

Müller, U. K. (2013): "Risk of Bayesian inference in misspecified models, and the sandwich covariance matrix," *Econometrica*, 81(5), 1805–1849.

Newey, W. K., and K. D. West (1987): "A Simple, Positive Semidefinite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703–08.

Shanken, J. (1985): "Multivariate tests of the zero-beta CAPM," *Journal of Financial Economics*, 14(3), 327–348.

——— (1992): "On the Estimation of Beta-Pricing Models," *Review of Financial Studies*, 5, 1–33.

Appendix

Table A1: Tests of risk premia in a misspecified model with either strong or useless factors

		Only a strong factor					Only a useless factor				
		λ_{strong}			R^2_{adj}		$\lambda_{useless}$			R^2_{adj}	
	T	10%	5%	1%	5th	95th	10%	5%	1%	5th	95th
Panel A: OLS											
	200	0.096	0.040	0.006	-35.3%	40.6%	0.046	0.010	0.000	-2069.4%	-2.7%
FM	600	0.104	0.050	0.009	-13.8%	51.6%	0.151	0.051	0.002	-2658.3%	-43.1%
	1,000	0.096	0.043	0.009	0.1%	54.9%	0.203	0.078	0.004	-2901.4%	-67.9%
	20,000	0.101	0.056	0.011	41.9%	55.4%	0.311	0.166	0.041	-2998.3%	-129.4%
	200	0.060	0.020	0.003	-23.0%	17.3%	0.023	0.004	0.001	-909.8%	-80.2%
BFM	600	0.091	0.042	0.008	-12.6%	33.0%	0.083	0.042	0.005	-1393.2%	-179.2%
	1,000	0.090	0.042	0.008	-3.4%	39.6%	0.092	0.041	0.007	-1567.4%	-270.8%
	20,000	0.107	0.054	0.013	40.8%	54.0%	0.109	0.052	0.012	-1868.7%	-604.1%
Panel B: GLS											
	200	0.104	0.054	0.007	-16.3%	33.3%	0.072	0.039	0.012	-530.4%	2.3%
FM	600	0.110	0.054	0.010	-1.9%	40.9%	0.052	0.028	0.009	-666.7%	-1.6%
	1,000	0.096	0.048	0.011	5.9%	42.2%	0.038	0.018	0.005	-733.9%	-10.3%
	20,000	0.100	0.048	0.012	29.5%	39.6%	0.032	0.021	0.013	-855.9%	-25.8%
	200	0.063	0.024	0.002	-1.4%	24.2%	0.023	0.006	0.000	-5.4%	17.7%
BFM	600	0.091	0.043	0.008	2.2%	31.7%	0.035	0.011	0.002	-4.2%	20.2%
	1,000	0.086	0.041	0.008	7.6%	35.1%	0.041	0.016	0.001	-1.3%	21.6%
	20,000	0.100	0.044	0.013	30.8%	40.4%	0.078	0.042	0.010	12.7%	21.0%

Frequency of rejecting the null hypothesis $H_0 : \lambda_i = \lambda_i^*$ for pseudo-true values of λ_i^* in a misspecified model with either a strong or a useless factor. Last two columns: 5th and 95th percentiles of cross-sectional R^2_{adj} across 2,000 simulations, evaluated at the point estimates for FM and at the posterior mean for BFM. The data are simulated based on equations (13) and (14), and the strong and weak factors are individually included in the regressions. The true value of the cross-sectional R^2_{adj} is 50.2% (37.0%) for the OLS (GLS) estimation.