Asset Pricing with Omitted Factors

Stefano Giglio

Yale University, National Bureau of Economic Research, and Center for Economic and Policy Research

Dacheng Xiu

University of Chicago

Standard estimators of risk premia in linear asset pricing models are biased if some priced factors are omitted. We propose a three-pass method to estimate the risk premium of an observable factor, which is valid even when not all factors in the model are specified or observed. The risk premium of the observable factor can be identified regardless of the rotation of the other control factors if together they span the true factor space. Our approach uses principal components of test asset returns to recover the factor space and additional regressions to obtain the risk premium of the observed factor.

I. Introduction

One of the central predictions of asset pricing models is that some risk factors—for example, intermediary capital or aggregate liquidity—should

This paper was previously circulated as "Inference on Risk Premia in the Presence of Omitted Factors." We are grateful to the editor, Harald Uhlig, and three anonymous referees for many insightful comments. We also benefited tremendously from discussions with Jushan Bai, Svetlana Bryzgalova, John Cochrane, George Constantinides, Gene Fama, Patrick Gagliardini, René Garcia, Valentin Haddad, Christian Hansen, Lars Hansen, Kris Jacobs, Raymond Kan, Frank Kleibergen, Serhiy Kozak, Nan Li, Toby Moskowitz, Stavros Panageas, Shri Santosh, Ivo Welch, and Irina Zviadadze as well as seminar and conference participants at the University of Chicago, Princeton University, Stanford University, University of Cambridge, Imperial College Business School, Duke University, Emory, Tinbergen

Electronically published May 10, 2021
[*Journal of Political Economy*, 2021, vol. 129, no. 7]

command a risk premium: investors should be compensated for their exposure to those factors, holding constant their exposure to all other sources of risk.

Sometimes, this prediction is easy to test in the data: when the factor predicted by theory is itself a portfolio (what we refer to as a *tradable* factor), the risk premium can be directly computed as the average excess return of the factor. For example, this is the case for the capital asset pricing model, where the theory-predicted factor is the market portfolio.

Most theoretical models, however, predict that investors are concerned about *nontradable* risks: risks that are not themselves portfolios, like consumption, inflation, liquidity, and so on. Estimating the risk premium of a nontradable factor requires the construction of a tradable portfolio that isolates that risk, holding all other risks constant. While different estimators have been proposed for risk premia (most prominently, two-pass cross-sectional regressions, like Fama-MacBeth regressions and mimicking-portfolio projections), they are all affected by one common potential issue: omitted variable bias.

Omitted variable bias arises in standard risk premia estimators whenever the model used in the estimation does not fully account for all priced sources of risk in the economy. This is a fundamental concern when testing asset pricing theories, because theoretical models are usually very stylized and cannot possibly explicitly account for all the risks that are at play in the economy.[1] While the possibility of omitted variable bias is known in the literature (see, e.g., Burmeister and McElroy 1988; Jagannathan and Wang 1998), no systematic solution has been proposed so far; rather, this problem is typically addressed in ad hoc ways that differ from case to case. Studies using the two-pass cross-sectional regression approach typically add somewhat arbitrarily chosen factors or characteristics as controls, like the Fama-French three factors;[2] studies using the mimicking-portfolio

Institute, Boston College, Singapore Management University, Federal Reserve Bank of Dallas, Xiamen University, Tsinghua University, Hong Kong University of Science and Technology, Durham Business School, University of Liverpool Management School, University of Virginia–McIntire, University of Minnesota, Luxembourg School of Finance, University of British Columbia, Nankai University, Fudan University, University of Miami, the 2017 Annual Meeting of the American Finance Association, the 2017 National Bureau of Economic Research–National Science Foundation Time Series Conference, the Hautes Études Commerciales–McGill Winter Finance Workshop, the 2017 Asian Meeting of the Econometric Society, the 2017 China Finance Review International Conference, the 10th Annual Society for Financial Econometrics conference, the BI Norwegian Business School–Swedish House of Finance (BI-SHoF) Conference, and the European Finance Association 2017 Annual Meeting.

[1] A symptom of this omission is the fact that the pricing ability of the models is often poor when tested using only the factors explicitly predicted by the theory. This suggests that other factors may be present in the data that are not accounted for by the model.

[2] For example, in an early contribution to this literature, Burmeister and McElroy (1988) propose choosing additional portfolios (like the market) to complement the set of observable factors and represent unobservable factors. See a summary of this approach and references to other related papers in Elton et al. (2014, 165–67).

approach usually select a small set of portfolios (e.g., portfolios sorted by size and book to market) on which to project the factor of interest. There is, however, no theoretical guarantee that the controls or the spanning portfolios are adequate to correct the omitted variable bias.

In this paper we propose a general solution to the omitted variable problem in linear asset pricing models. We introduce a new three-pass methodology that exploits (1) the large dimensionality of available test assets and (2) a rotation invariance result to correctly recover the risk premium of any observable factor, even when not all true risk factors are observed and included in the model.

The starting point of our procedure is a simple but general rotation invariance result that holds for risk premia in linear factor models. Suppose that returns follow a linear factor model with p factors, and we wish to determine the risk premium of one of them, g_t. What we describe as a rotation invariance result is the observation that the risk premium of g_t is invariant to how the $p - 1$ control factors are rotated—any $p - 1$ linear combinations of the original p factors can serve as control factors as long as the rotated factor model spans the same risks as the original model.[3]

This result implies that knowing the identities of all true factors is not necessary to estimate the risk premium of one of them (g_t). As long as the entire factor space can be recovered, the risk premium of g_t can be identified even when the other factors are neither observed nor known. A natural way to recover the factor space in this scenario is to extract principal components (PCs) of the test asset returns. Our methodology therefore combines the principal component analysis (PCA) with two-pass cross-sectional regressions to provide consistent estimates of the risk premium for any observed factor.

Our methodology proceeds in three steps. First, we use PCA to extract factors and their loadings from a large panel of test asset returns, thus recovering the factor space (i.e., some unknown rotation of the p factors). Second, we run a cross-sectional regression using only the PCs (without the factor of interest g_t) to find their risk premia. Third, we estimate a time series regression of g_t onto the PCs that uncovers the relation between g_t and the latent factors and, in addition, removes potential measurement error from g_t. The risk premium of g_t is then estimated as the product of the loadings of g_t on the PCs (estimated in the third step) and their risk premia (estimated in the second step). The invariance result discussed above is what guarantees the identification of the risk premium

[3] The invariance result we derive is distinct from similar results that the literature has explored in the past (e.g., Roll and Ross 1980; Huberman, Kandel, and Stambaugh 1987; Cochrane 2009). This literature has explored the conditions under which rotations of a factor model retain the pricing ability of the original model. It has not, however, explored the invariance properties for risk premia of individual factors within the model.

of g_t, regardless of the rotation of the true factors that occurs when extracting PCs.

A caveat worth mentioning is that one premise of our approach is that all priced risks in the economy should be reflected in asset returns and that these risk factors are pervasive, so that relevant and priced factors can be recovered from the PCA step. This assumption—which is also behind the arbitrage pricing theory of Ross (1976) and the approximate factor model of Chamberlain and Rothschild (1983)—might not hold exactly in practice, for example, if markets are less integrated or if large subsets of test assets are not exposed to certain priced risk factors (i.e., the factors are weak). That said, recent empirical work in the asset pricing literature (e.g., Kelly, Pruitt, and Su 2019) provides evidence in support of this assumption.[4]

Our three-pass procedure can be interpreted in light of the two standard methods for risk premium estimation. First, it can be viewed as a PC-augmented two-pass cross-sectional regression. Rather than selecting the control factors for g_t arbitrarily, the PCs of the test asset returns are used as controls; these stand in for the omitted factors and, thanks to the rotation invariance result, fully correct the omitted variable bias. Second, our procedure can be interpreted as a regularized version of the mimicking-portfolio approach. The factor g_t is projected onto the PCs of returns (the PCs are themselves portfolios) rather than onto an arbitrarily chosen set of portfolios, which could lead to a bias, or onto the entire set of test assets, which would be inefficient or even infeasible when the number of test assets is larger than the sample size.

The fact that our procedure can be interpreted equivalently as an extension of both methods is particularly surprising because in standard settings (when the number of test assets is fixed), the two estimators differ even in large samples because the risk premium of a factor (in population) is not the same as the expected excess return of its mimicking portfolio, unless the factor itself is tradable. The former is a constant parameter that does not depend on the test assets, whereas the latter depends on the test assets onto which the factor of interest is projected. Our theoretical analysis, however, sheds light on the convergence of the two methods as the number of test assets increases. Our three-pass procedure reveals the numerical equivalence in this scenario between the extensions of the two procedures, as long as PCA is used to span the entire factor space and avoid the curse of dimensionality.

We then extend the theory of our estimator in two directions. First, we explore the possibility that using mixed-frequency data may help the efficiency of our estimator. Specifically, while the factors of interest (especially

[4] We report an in-depth discussion and analysis of this issue in the online appendix.

the nontradable ones that are the focus of the paper) are typically available only at low frequencies, the returns are often available at high frequency. Perhaps surprisingly, we find that the use of mixed-frequency data cannot improve the asymptotic efficiency of our procedure. As we show in the paper, the reason is that to a first order, the limit to efficiency is not the frequency of the data but the total span of the sample; intuitively, this is because the first-order estimation error for the risk premium of a nontradable factor comes from the expected excess return of its mimicking portfolio, and it is well known that high-frequency data do not help better pin down expected returns.

The second direction we explore focuses on the mimicking-portfolio interpretation of our estimator. To the extent that one views our estimator as a regularized mimicking-portfolio estimator on the basis of a large set of test assets, it is interesting to ask whether other regularization techniques might yield a similar result. We explore here in detail the ridge regression approach: that is, a regularized projection (using ridge) of g_t onto all test assets. We derive new theoretical results showing that the ridge estimator is indeed consistent, but in our setup the baseline PCA approach is more efficient; intuitively, if there is a finite number of strong factors underlying asset returns, PCA can efficiently separate signal (the factors) from noise (idiosyncratic error), whereas a ridge regression relies on both signal and noise and is therefore less efficient. Together, the exploration of these theoretical extensions confirms the good theoretical properties of our baseline three-pass estimator.

We apply our methodology to a large set of 647 portfolios that include equities sorted along many characteristics, bonds, and currencies. We estimate and test the significance of the risk premia of tradable and nontradable factors from a number of models proposed in the literature. We show that the conclusions about the magnitude and significance of the risk premia often depend dramatically on whether we account for omitted factors (using our estimator) or ignore them (using standard methods). In contrast with the existing literature, we find a risk premium of the market portfolio that is positive, significant, and close to the time series average of market excess returns, even when we allow for an unrestricted zero-beta rate following the Black (1972) version of the capital asset pricing model. We also decompose the variance of each observed factor into the components due to exposures to the latent factors as well as the component due to measurement error. We find that several macroeconomic factors are dominated by noise, and after we correct for it and for exposure to unobservable factors, they command a risk premium of essentially zero. We do, however, find empirical support for the consumption growth of stockholders from Malloy, Moskowitz, and Vissing-Jorgensen (2009) as well as for factors related to financial frictions (like the liquidity factor of Pástor and Stambaugh 2003).

Our paper derives several important econometric properties of the three-pass estimator. We establish the consistency and derive the asymptotic distribution when both the number of test portfolios n and the number of observations T are large.[5] Our asymptotic theory allows for heteroskedasticity and correlation across both the time series and the cross-sectional dimensions while explicitly accounting for the propagation of estimation errors through the multiple estimation steps.

Moreover, the increasing dimensionality simplifies the asymptotic variance of the risk premium estimates, for which we also provide an estimator. In addition, we construct a consistent estimator for the number of latent factors while also showing that even without it, the risk premium estimates remain consistent. Finally, a notable advantage of our procedure is that inference remains valid even when the observable factor g_t is spurious or even useless (i.e., totally uncorrelated with asset returns). In the paper, we also provide a test of the null that the observed factor g_t is weak. Our methodology therefore provides a novel approach to inference in the presence of weak observable factors.

This paper sits at the confluence of several large strands of literature combining empirical asset pricing with high-dimensional factor analysis. Using two-pass regressions to estimate asset pricing models dates back to Jensen, Black, and Scholes (1972) and Fama and Macbeth (1973). Over the years, the econometric methodologies have been refined and extended, and some recent papers have also explored the large-n, large-T setting, though not in the context of solving the omitted factor problem (e.g., Connor, Hagmann, and Linton 2012; Bai and Zhou 2015; Fan, Liao, and Yao 2015; Gagliardini, Ossola, and Scaillet 2016, 2019). Our paper also relates to the large literature that has explored pitfalls in estimating and testing linear factor models, like model misspecification and the presence of weak factors (e.g., Kan and Zhang 1999a, 1999b; Jagannathan and Wang 1998), and that has proposed methods that are more robust to misspecification (e.g., Kleibergen 2009; Gospodinov, Kan, and Robotti 2013; Bryzgalova 2015). Given the mimicking-portfolio interpretation of our estimator, our paper naturally builds on a vast literature on theoretical and empirical analysis using mimicking portfolios, dating back to Huberman, Kandel, and Stambaugh (1987) and Breeden, Gibbons, and Litzenberger (1989) and, more recently, Balduzzi and Robotti (2008) and Ang et al. (2006), for example. Finally, we build on the literature that has explored the use of PCs in asset pricing (e.g., Chamberlain and Rothschild 1983; Connor and Korajczyk 1986, 1988; and, recently, Kozak, Nagel, and Santosh 2018).

[5] Lewellen, Nagel, and Shanken (2010) highlight the danger of focusing on a small cross section of assets with a strongly low-dimensional factor structure and suggest increasing the number of assets used to test the model. We point to an additional reason to use a large number of assets: to control properly for omitted factors.

The paper is organized as follows. Section II discusses biases due to omitted variables and measurement error in the standard risk premia estimators. Section III introduces our three-pass estimation procedure and discusses how it can be interpreted as an extension of both the cross-sectional regression approach and the mimicking-portfolio approach. Section IV provides the asymptotic theory on inference with our estimator, followed by an empirical study in section V. Appendixes A and B provide technical details. The online appendix contains additional theoretical results, Monte Carlo simulations, additional empirical analysis, and supplementary mathematical proofs.

Throughout the paper, we use $(A:B)$ to denote the concatenation (by columns) of two matrices A and B. The term e_i is a vector with 1 in the ith entry and 0 elsewhere, whose dimension depends on the context. The term ι_k denotes a k-dimensional vector with all entries being 1, and $\mathbb{I}_d$ denotes the $d \times d$ identity matrix. For any time series of vectors $\{a_t\}_{t=1}^T$, we denote $\bar{a} = (1/T)\Sigma_{t=1}^T a_t$. In addition, we write $\bar{a}_t = a_t - \bar{a}$. We use the capital letter A to denote the matrix $(a_1 : a_2 : \ldots : a_T)$ and write $\bar{A} = A - \bar{a}\iota_T^\intercal$ correspondingly. We denote $\mathbb{P}_A = A(A^\intercal A)^{-1}A^\intercal$ and $\mathbb{M}_A = \mathbb{I}_d - \mathbb{P}_A$ for some $d \times T$ matrix A. We use $a \vee b$ to denote the maximum of a and b, and we use $a \wedge b$ as their minimum for any scalars a and b.

II. Biases in Standard Risk Premia Estimators

In this section we illustrate how the standard risk premia estimators—the two-pass regression approach (like Fama-MacBeth) and the mimicking-portfolio approach—suffer from potential biases induced by omitted factors and measurement error. For illustration purposes, we show these results in a simple two-factor model, but all the results easily extend to more general specifications.

Suppose that $v_t = (v_{1t}, v_{2t})^\intercal$ is a vector of two potentially correlated factors. We assume that both have been demeaned, so we interpret v_{1t} and v_{2t} as factor innovations.[6] Assuming that the risk-free rate is observed, we express the model in terms of excess returns:

$$r_t = \beta\gamma + \beta v_t + u_t,$$

where u_t is idiosyncratic risk, $\beta = (\beta_1 : \beta_2)$ is a matrix of risk exposures, and $\gamma = (\gamma_1, \gamma_2)^\intercal$ is the vector of risk premia for the two factors. In what follows, we are interested in estimating the risk premium of a proxy for

[6] As discussed in the introduction, the focus of this paper is on nontradable factors, whose means have no direct relevance for the factors' risk premia. This is why we write the model directly in terms of factor innovations. Of course, if the factors are instead tradable, the mean of the factor itself—minus the risk-free rate—is the risk premium, in which case the methods we discuss here are still valid as an alternative estimator of the risk premium.

the first factor v_{1t}, denoted as g_t; its risk premium is therefore γ_1 in this simple setting.

We begin with a brief review of the two standard estimators of γ. Two-pass regressions estimate the factor risk premia as follows. First, time series regressions of each test asset's excess return onto the factors estimate the assets' risk exposures, β_1 and β_2. Second, a cross-sectional regression of average returns onto the estimated β_1 and β_2 yields the risk premia estimates of γ_1 and γ_2.

The mimicking-portfolio approach instead estimates the risk premium of g_t by projecting that factor onto a set of tradable asset returns, therefore constructing a tradable portfolio that is maximally correlated with g_t (which is why it is also referred to as the maximally correlated mimicking portfolio). The risk premium of g_t is then estimated as the average excess return of its mimicking portfolio.

A. *Omitted Variable Bias*

Consider first estimating the risk premium of $g_t = v_{1t}$ using a two-pass cross-sectional regression that omits v_{2t}. It is easy to see that this omission can induce a bias in each of the two steps of the procedure. The time series step yields a biased estimate of β_1, as long as the omitted factor v_{2t} is correlated with v_{1t} (a standard omitted variable bias problem). The magnitude of this bias depends on the time series correlation of the factors. In the cross-sectional step of the procedure, a second omitted variable bias occurs: rather than regressing average returns onto the entire matrix of risk exposures with respect to both factors, β, only part of it ($\hat{\beta}_1$) would be used, since the factor v_{2t} is omitted. The magnitude of this second bias depends on the cross-sectional correlation of risk exposures, β_1 and β_2. Eventually, both biases (omission of v_{2t} in the first step and omission of β_2 in the second step) affect the estimated risk premium for g_t using the two-pass regression approach.

In the mimicking-portfolio estimator, a related omitted variable bias can instead arise from the omission of assets onto which g_t is projected. To see the potential for omitted variable bias in the mimicking-portfolio approach, it is useful to write down explicitly the formula for the estimator. Consider the projection of $g_t = v_{1t}$ onto the excess returns of a chosen set of test assets, $\breve{r}_t$.[7] This projection yields coefficients $w^g = \mathrm{Var}(\breve{r}_t)^{-1}\mathrm{Cov}(\breve{r}_t, g_t)$; these are the weights of the mimicking portfolio for g_t, whose excess return is then $r_t^g = (w^g)^\top \breve{r}_t$. Therefore, we can write the expected excess return of the mimicking portfolio as $\gamma_g^{\mathrm{MP}} = (w^g)^\top \mathrm{E}(\breve{r}_t)$. Since the test assets $\breve{r}_t$ follow the same pricing model as the

[7] We deliberately use $\breve{r}_t$ instead of r_t, which we reserve for the universe of available test assets. The choice of assets for projection could be the entire universe of test assets r_t or some portfolios of r_t.

universe r_t, we can write $\breve{r}_t = \breve{\beta}\gamma + \breve{\beta}v_t + \breve{u}_t$. Substituting, we can write the formula for the mimicking-portfolio estimator of the risk premium of the first factor as $\gamma_g^{\text{MP}} = \{(\breve{\beta}\Sigma^v\breve{\beta}^\top + \breve{\Sigma}^u)^{-1}(\breve{\beta}\Sigma^v e_1)\}^\top\breve{\beta}\gamma$, where e_1 is a column vector $(1, 0)^\top$, Σ^v is the covariance matrix of the factors, and $\breve{\Sigma}^u$ is the covariance matrix of the idiosyncratic risk of the assets used in the projection.

The formula above shows that, in general, not all choices of the assets on which to project g_t will result in a consistent estimator of γ_1; that is, it is not guaranteed that $\gamma_g^{\text{MP}} = \gamma_1$. There is one case in which these two population quantities are identical: if the assets are chosen to be p portfolios that (1) are well diversified (so that $\breve{\Sigma}^u \approx 0$) and (2) fully span the true factors v_t, so that $\breve{\beta}$ is invertible and $v_t = \breve{\beta}^{-1}\breve{r}_t$; if both conditions hold, we indeed have $\gamma_g^{\text{MP}} = \gamma_1$.

When these conditions are not satisfied, however, the mimicking-portfolio estimator will in general be biased. In particular, the mimicking-portfolio estimator will be biased if the set of assets used in the projection omits some portfolios that help span all risk factors in v_t. The existing literature that has used the mimicking-portfolio approach has typically ignored this bias. For example, when constructing a mimicking portfolio for consumption growth, Malloy, Moskowitz, and Vissing-Jorgensen (2009) project it onto four portfolios sorted by size and book to market. But naturally there are other risks in the economy in addition to size and value that may be correlated to consumption growth and that may not be captured by those four portfolios. In that case, the estimator may be affected by omitted variable bias.

B. Measurement Error Bias

We now consider measurement error in g_t because this is often plausible in practice. For nontradable factors, which are the primary focus of this paper, there are often many choices the researcher needs to make to construct the empirical counterpart of a theory-predicted factor. For example, there are many ways to construct an aggregate liquidity factor in practice. The construction of the empirical factor is likely to introduce some measurement error, which we allow for in our specification. For tradable factors, measurement error can capture exposure to unpriced risks or idiosyncratic risk that is not fully diversified.

Suppose that the econometrician can observe only $g_t = v_{1t} + z_t$, where z_t is measurement error orthogonal to the factors but potentially correlated with u_t.

Measurement error in g_t adds another source of bias to the standard estimators. Consider first the two-pass regression approach. Independently of whether v_{2t} is observed, measurement error in g_t will induce an attenuation bias in the estimated β_1 in the time series regression (since the regressor g_t is measured with error). In turn, this first-stage bias affects the second-step estimate, leading to a biased estimate of γ_1.

Measurement error affects the mimicking portfolio as well. In the presence of measurement error z_t, the formula for γ_g^{MP} has an additional term: $\gamma_g^{\mathrm{MP}} = \{(\breve{\beta}\Sigma^v\breve{\beta}^\top + \breve{\Sigma}^u)^{-1}(\breve{\beta}\Sigma^v e_1 + \breve{\Sigma}^{z,u})\}^\top\breve{\beta}\gamma \neq \gamma_1$, where $\breve{\Sigma}^{z,u} = \mathrm{Cov}(z_t, \breve{u}_t)$. Thus, measurement error z_t can introduce a bias in the mimicking-portfolio estimator, unless idiosyncratic errors $\breve{u}_t$ in the spanning assets are uncorrelated with z_t.

III. Methodology

In this section we present our three-pass risk premia estimator, which tackles both the omitted variable and the measurement error biases in estimating risk premia in linear factor models.

Consider a general linear factor model with p factors:

$$r_t = \beta\gamma + \beta v_t + u_t, \quad \mathrm{E}(v_t) = \mathrm{E}(u_t) = 0, \quad \text{and} \quad \mathrm{Cov}(u_t, v_t) = 0, \tag{1}$$

where v_t are innovations of the p factors (i.e., mean zero factors), r_t are excess returns on n assets, u_t are idiosyncratic errors, β are factor loadings, and γ is the vector of risk premia for the p factors.

The objective of this paper is to estimate the risk premia of one or more factors g_t without necessarily observing all true factors v_t. In the simple two-factor model of section II, we assumed that g_t was a proxy of the first factor v_{1t}. Here we introduce a more general specification for g_t that nests this case and also allows for measurement error.

Call g_t a set of d observable (tradable or nontradable) factors whose risk premia we aim to estimate. The term g_t is related to the factors v_t as follows:

$$g_t = \delta + \eta v_t + z_t, \quad \mathrm{E}(z_t) = 0, \quad \text{and} \quad \mathrm{Cov}(z_t, v_t) = 0, \tag{2}$$

where η captures the relation between g_t and the unobservable factors v_t, and z_t is measurement error in g_t. The risk premium of g_t—the objective of our analysis—is defined as the expected excess return of a portfolio with beta of 1 with respect to g_t and beta of 0 with respect to all other factors (including the unobservable ones), and in this model it corresponds to $\gamma_g = \eta\gamma$.[8]

From equations (1) and (2), it is clear that neither η (the loading of g_t on the unobservable fundamental factors) nor γ (the risk premia of the unobservable factors) can be identified if the factors v_t are not observed. A fundamental identification question then arises: under what conditions

[8] This specification nests the case where g_t is the first factor by choosing η to be the vector $(1, 0, 0, \ldots, 0)$ and setting δ and z_t to zero.

can we identify the product $\eta\gamma$ even if we cannot separately identify η and γ?

The answer—upon which we build our methodology—lies in a simple but powerful property of risk premia in linear factor models, which we denote *rotation invariance*. It states that the product $\eta\gamma$ can be identified even if one observes only an arbitrary full-rank rotation of the factors, that is, if one just observes $\hat{v}_t = Hv_t$, with H any full-rank $p \times p$ matrix, but observes neither v_t nor H.

To see that, rewrite the model as

$$r_t = \beta H^{-1} H\gamma + \beta H^{-1} Hv_t + u_t,$$

$$g_t = \delta + \eta H^{-1} Hv_t + z_t.$$

Defining $\hat{\eta} \coloneqq \eta H^{-1}$, $\hat{\gamma} \coloneqq H\gamma$, and $\hat{\beta} \coloneqq \beta H^{-1}$, we can write the model entirely in terms of the rotated factors $\hat{v}_t$:

$$r_t = \hat{\beta}\hat{\gamma} + \hat{\beta}\hat{v}_t + u_t, \tag{3}$$

$$g_t = \delta + \hat{\eta}\hat{v}_t + z_t. \tag{4}$$

As long as $\hat{v}_t$ is observed, we can then identify $\hat{\eta} = \eta H^{-1}$ (it is the vector of regression coefficients of g_t on $\hat{v}_t$) as well as $\hat{\gamma} = H\gamma$ (they are the risk premia of $\hat{v}_t$, which can be obtained, e.g., via standard cross-sectional regressions). While we clearly cannot recover η or γ separately because we do not know H, we can still identify the risk premium of g_t because

$$\hat{\eta}\hat{\gamma} = \eta H^{-1} H\gamma = \eta\gamma = \gamma_g.$$

It is obvious that this so-called invariance property does not hold for the other quantities in the model, like η, γ, or β—it is a property that is specific to γ_g.

The methodology we propose in this paper combines this insight with a well-known result from the econometric literature on latent factors (e.g., Bai 2003): that if the factors in v_t are sufficiently strong, PCA consistently recovers a rotation of the factor space; that is, $\hat{v}_t = Hv_t$ for some unobservable matrix H.

Following these arguments, our three-pass procedure (1) estimates the rotated factors $\hat{v}_t$ via PCA; (2) estimates via a two-pass cross-sectional regression $\hat{\gamma} = H\gamma$, that is, the risk premia of $\hat{v}_t$; and (3) estimates $\hat{\eta} = \eta H^{-1}$ via a time series regression of g_t onto the estimated $\hat{v}_t$. The risk premia of g_t, $\eta\gamma$, can then be estimated by taking the product of the estimates of $\hat{\eta}$ and $\hat{\gamma}$ at steps 2 and 3.

Having presented the broad ideas behind our methodology, section III.A delves deeper and more formally into our three-pass estimator; it discusses in detail the role each step plays and addresses various implementation concerns. After that, section III.B discusses different interpretations of our estimator, relating it specifically to the existing methodologies.

A. *The Three-Pass Estimator*

Our analysis is presented in the context of the model of equations (1) and (2). Before we formally discuss the estimator, it is useful to add a few notes on the model.

First, the model assumes constant loadings and risk premia. These assumptions are restrictive for individual stocks but applicable to characteristic-sorted portfolios, which we will use in our empirical study. Our analysis is still applicable to certain conditional models that allow for time-varying risk premia and risk exposures by taking a stand on appropriate conditioning information, (e.g., characteristics or state variables) at the cost of greater statistical complexity.[9] Second, we impose weak assumptions on the structure of the errors. Most of our results hold for nonstationary processes with heteroskedasticity and dependence in both the time series and the cross-sectional dimensions. For ease of presentation, we defer the technical details to appendix A. Third, this baseline model imposes that the zero-beta rate is equal to the observed treasury bill rate. Online appendix I.2 examines a more general version of the model that allows the zero-beta rate to be different and to be estimated.

We now present our three-pass estimator. We start by writing the model in matrix form for notational convenience. We denote R as the $n \times T$ matrix of excess returns, V the $p \times T$ matrix of factors, G the $d \times T$ matrix of observable factors, U the $n \times T$ matrix of idiosyncratic errors, and Z the $d \times T$ matrix of measurement error. Our model (eqq. [1], [2]) can then be written in matrix terms as

$$R = \beta\gamma\iota_T^{\top} + \beta V + U.$$

Writing $(\bar{R}, \bar{V}, \bar{G}, \bar{U}, \bar{Z})$ as the matrices of the demeaned variables, we find that this equation then becomes

$$\bar{R} = \beta\bar{V} + \bar{U}. \tag{5}$$

Next, we write the equation for g_t in matrix form. Given that for nontradable factors (like inflation or liquidity) the mean of g_t, δ, does not have a meaningful interpretation or relevance for the purpose of estimating the risk premium, we need only the demeaned version of equation (2):

$$\bar{G} = \eta\bar{V} + \bar{Z}. \tag{6}$$

[9] We discuss such extensions in greater detail in online app. III.9.

Our estimator makes use of only excess returns R and the factors of interest G. We do not require the true factors V to be known or observable. As reported above, the procedure exploits an important result from Bai and Ng (2002) and Bai (2003) that guarantees that by applying PCA to the panel of observed return innovations $\bar{R}$, we can recover β and $\bar{V}$ up to some invertible matrix H as long as $n, T \to \infty$.

The three-pass estimator.—Given observable returns R and the factors of interest G, we can write the three steps of our estimator for $\gamma_g = \eta\gamma$ as follows:

1. *PCA step.*—Extract the PCs of returns by conducting the PCA of the matrix $n^{-1}T^{-1}\bar{R}^{\top}\bar{R}$. Define the estimator for the factors and their loadings as

$$\hat{V} = T^{1/2}(\xi_1 : \xi_2 : \dots : \xi_{\hat{p}})^{\top} \quad \text{and} \quad \hat{\beta} = T^{-1}\bar{R}\hat{V}^{\top}, \tag{7}$$

   where $\xi_1, \xi_2, \dots, \xi_{\hat{p}}$ are the normalized eigenvectors (of length 1) corresponding to the largest $\hat{p}$ eigenvalues of the matrix $n^{-1}T^{-1}\bar{R}^{\top}\bar{R}$ and $\hat{p}$ is some consistent estimator of the number of factors.[10]
2. *Cross-sectional regression step.*—Run a cross-sectional ordinary least squares (OLS) regression of average returns, $\bar{R}$, onto the estimated factor loadings, $\hat{\beta}$, to obtain the risk premia of the estimated latent factors:

$$\hat{\gamma} = (\hat{\beta}^{\top}\hat{\beta})^{-1}\hat{\beta}^{\top}\bar{r}.$$

3. *Time series regression step.*—Run a time series regression of g_t onto the extracted factors from step 1 and then obtain the estimator, $\hat{\eta}$, and the fitted value of the observable factor, $\hat{G}$:

$$\hat{\eta} = \bar{G}\hat{V}^{\top}(\hat{V}\hat{V}^{\top})^{-1} \quad \text{and} \quad \hat{G} = \hat{\eta}\hat{V}.$$

As indicated before, the estimator of the risk premium for the observable factor g_t is obtained by combining the estimates of the second and third steps:

$$\hat{\gamma}_g = \hat{\eta}\hat{\gamma}.$$

Our three-pass estimator also has a more compact form:

$$\hat{\gamma}_g = \bar{G}\hat{V}^{\top}(\hat{V}\hat{V}^{\top})^{-1}(\hat{\beta}^{\top}\hat{\beta})^{-1}\hat{\beta}^{\top}\bar{r}. \tag{8}$$

The estimator can be easily extended to the case in which the zero-beta rate is allowed to differ from the observed risk-free rate. In that case,

[10] There are various estimators of p available in the existing literature, which work under similar but different assumptions. We propose one such estimator in online app. I.1 and prove its consistency under assumptions of this paper. In what follows, we directly assume the existence of a consistent estimator $\hat{p}$.

returns can be written as $r_t = \gamma_0\iota_n + \beta\gamma + \beta v_t + u_t$, and step 2 of the procedure can be modified to yield an estimate for γ_g together with the zero-beta rate γ_0. In compact form, the estimators are given by

$$\hat{\gamma}_0 = (\iota_n^\top \mathbb{M}_{\hat{\beta}}\iota_n)^{-1}\iota_n^\top \mathbb{M}_{\hat{\beta}}\bar{r}, \quad \tilde{\gamma}_g = \bar{G}\hat{V}^\top(\hat{V}\hat{V}^\top)^{-1}\left(\hat{\beta}^\top \mathbb{M}_{\iota_n}\hat{\beta}\right)^{-1}\hat{\beta}^\top \mathbb{M}_{\iota_n}\bar{r}, \tag{9}$$

where $\mathbb{M}_{\hat{\beta}} = \mathbb{I}_n - \hat{\beta}(\hat{\beta}^\top\hat{\beta})^{-1}\hat{\beta}^\top$ and $\mathbb{M}_{\iota_n} = \mathbb{I}_n - \iota_n(\iota_n^\top\iota_n)^{-1}\iota_n^\top$. We discuss this extension in online appendix I.2.

The first step of the three-pass procedure recovers the factors v (up to a rotation: $H\bar{V}$ for some unobserved invertible matrix H) by extracting the PCs of returns and selecting the first $\hat{p}$ of them. We propose to extract PCs from the $T \times T$ matrix, $n^{-1}T^{-1}\bar{R}^\top\bar{R}$, and normalize the estimated factors such that $\hat{V}\hat{V}^\top = \mathbb{I}_{\hat{p}}$. Alternatively, one could consider extracting PCs from the $n \times n$ matrix $n^{-1}T^{-1}\bar{R}\bar{R}^\top$, and the normalization takes a different form: $\hat{\beta}^\top\hat{\beta} = \mathbb{I}_{\hat{p}}$. The two ways of normalization yield numerically identical risk premia estimates.

Once the PCs are extracted in the first stage, the second stage estimates their risk premia. The estimation of risk premia in the second step can be done in different ways. We suggest using an OLS regression for its simplicity. Either a generalized least squares (GLS) regression or a weighted least squares (WLS) regression is possible, but either of the two would require estimating a large number of parameters (e.g., the covariance matrix of u_t in GLS or its diagonal elements in WLS). As it turns out, these estimators will not improve the asymptotic efficiency of the OLS to the first order. This is different from the standard large T and fixed n case because in the large n setting, the covariance matrix of u_t matters only at the order of $O_p(n^{-1} + T^{-1})$, whereas the leading term of $\hat{\gamma}_g$ is of $O_p(T^{-1/2})$.

The third step is a new addition to the standard two-pass procedure. It is critical because it translates the uninterpretable risk premia of latent factors to those of factors that the economic theory predicts. This step also removes the effect of measurement error, which the standard approaches cannot accomplish.[11] Even though g_t can be multidimensional, the estimation for each observable factor is separate. Estimating the risk premium for one factor does not affect the estimation for the others at all, another important property of our estimator.

B. *Interpretations of the Three-Pass Estimator*

In this section, we discuss various interpretations of our estimator. In particular, we show that our estimator can be interpreted both as an extension

[11] The third step resembles similar regressions considered also in Bai and Ng (2006) to test whether observable factors are spanned by latent ones rather than in combination with the first two steps of our procedure (which allow us to estimate the risk premium of the factors).

of the two-pass regressions and as an extension of the mimicking-portfolio estimator. As discussed above—and as we also review formally below—in general, the two-pass regression and the mimicking-portfolio estimator tend to give different estimates, even when the model is correctly specified. We show below that our estimator instead represents the *convergence* of these two approaches: a case in which both estimators give exactly the same answer. Finally, we introduce a third interpretation based on the stochastic discount factor.

1. Two-Pass Regression Interpretation

The two-pass interpretation of our results derives directly from the rotation invariance of risk premia. To begin, suppose that we know the entire model, equations (1) and (2). Also, suppose for simplicity that g_t is only one factor ($d = 1$; the results extend to any d), and there is no measurement error.

We now construct a specific rotation of this model in which the factor g_t appears as the first of the p factors, together with $p - 1$ additional control factors. To do so, construct a matrix H in which the first row is η and the remaining $p - 1$ rows are arbitrary (with the only condition that the resulting H is full rank). The factors of the rotated model are Hv_t; since η is the first row of H, the first factor in this rotation is ηv_t, which is just g_t (see eq. [2]). Similarly, the risk premia of the rotated factors are $H\gamma$, and the risk premium of the first factor, g_t, is $\eta\gamma$, again because the first row of H is η.

Consider now applying a two-pass cross-sectional regression in this particular rotation, assuming that all the rotated factors Hv_t are observed. Given that the model is correctly specified, the two-pass regression will recover all the risk premia $H\gamma$; therefore, it will also recover $\eta\gamma$ as the risk premium for g_t. But this result holds for any matrix H where the first row is η, independently of the other rows of H. This implies that a two-pass estimation of a model where g_t appears with $p - 1$ arbitrary linear combinations of v_t will deliver the correct estimate for the risk premium of g_t independently of how the remaining $p - 1$ controls are rotated. The only requirement is that H is invertible: that is, that g_t together with the controls spans the same space as the original factors v_t.

We can then interpret our three-pass estimator as a factor-augmented cross-sectional regression estimator. Step 1 uses PCA to extract a rotation of the original factors v_t. Step 3 removes measurement error from g_t and identifies $\hat{\eta}$: this tells us how to rotate the estimated model so that g_t appears as the first factor. We can then construct a rotated model with $\hat{g}_t$ together with $p - 1$ PCs as controls. Risk premia for this model are estimated via cross-sectional regressions (step 2) that will then deliver a risk premium of $\eta\gamma$ for g_t. While this cross-sectional regression interpretation of the estimator inverts the ordering of steps 2 and 3 of our procedure, it gives numerically identical results.

2. Mimicking-Portfolio Interpretation

Our three-pass procedure can also be interpreted as a mimicking-portfolio estimator, in which the PCs themselves are the portfolios on which g_t is projected. This is an ideal choice of portfolios that ensures that the estimator is consistent.

Suppose that, out of the universe of test assets, we construct $\breve{p}$ portfolios on which we project g_t. We refer to w as the $n \times \breve{p}$ matrix of portfolio weights that are used to construct these portfolios of r_t: so the returns of the portfolios we will use in the estimation will be $\breve{r}_t = w^\top r_t$. In section II.A, we showed that in general the mimicking-portfolio estimator is not consistent. That said, it can be consistent if the returns of the portfolios on which g_t is projected ($\breve{r}_t$) satisfy certain requirements. In turn, this means that w needs to be chosen carefully so that the mimicking-portfolio estimator that uses these $\breve{p}$ portfolios,

$$\gamma_g^{\mathrm{MP}} = \eta \Sigma^v \breve{\beta}^\top (\breve{\Sigma}^r)^{-1} \breve{\beta} \gamma + \breve{\Sigma}^{z,u} (\breve{\Sigma}^r)^{-1} \breve{\beta} \gamma, \tag{10}$$

actually converges to $\gamma_g = \eta\gamma$. This formula also shows clearly that unless the returns of the assets $\breve{r}_t$ are chosen appropriately, the mimicking-portfolio approach will in general yield a different estimate than the two-step cross-sectional estimator.

We now derive a novel property of mimicking-portfolio estimators (not studied in the existing literature, to the best of our knowledge) that helps us choose w appropriately when the number of test assets n is large. In particular, we prove in proposition 1 that the difference between the mimicking-portfolio based risk premia γ_g^{MP} and $\eta\gamma$ disappears as $n \to \infty$, as long as the portfolios on which to project g_t are constructed by choosing w equal to β or some full-rank rotation of it.

PROPOSITION 1. Suppose that assumptions A1–A3 hold. The risk premium of the mimicking portfolio that is maximally correlated with g_t, γ_g^{MP}, satisfies $\gamma_g^{\mathrm{MP}} - \eta\gamma = o_p(1)$ as $n \to \infty$.

Intuitively, this choice of w is guaranteed to achieve asymptotically the two criteria highlighted in section II.A: these portfolios manage to average out idiosyncratic errors while maintaining their exposure to the factors. The second part is important. Many portfolios can average out idiosyncratic errors, but they might also average out exposures to certain factors, in which case the omitted variable bias discussed in this paper would bias the estimates.

Our three-pass method corresponds exactly to a mimicking-portfolio estimator where the portfolios onto which g_t is projected are constructed using a particular choice for w: $\hat{\beta}(\hat{\beta}^\top\hat{\beta})^{-1}$, that is, a full-rank rotation of the estimated β. The resulting portfolio returns are exactly the PCs in step 1 of our procedure, that is, $\hat{V} = (\hat{\beta}^\top\hat{\beta})^{-1}\hat{\beta}^\top\bar{R}$. In addition, these portfolios are (when n is large) free of idiosyncratic error. Step 3 projects g_t onto these portfolios, thus identifying the weights of the mimicking portfolio, $\hat{\eta}$. Our estimator of the risk premium of g_t is then obtained by multiplying the portfolio weights $\hat{\eta}$ by the risk premia of these portfolios ($\hat{\gamma}$) obtained in step 2.

Interestingly, as we prove in appendix A (see discussion following assumption A3), another valid choice of w would be the identity matrix. Therefore, the mimicking-portfolio estimator would also be unbiased if the factor is projected onto the entire universe of potential test assets r_t as opposed to an appropriately chosen subset $\check{r}_t$, again as long as $n \to \infty$. Intuitively, when g_t is projected onto a larger and larger set of test assets, the mimicking portfolio will diversify the idiosyncratic errors while at the same time spanning the factor space, thus reducing the bias. However, as $n \to \infty$, the mimicking-portfolio estimator becomes increasingly inefficient, as the number of right-hand-side regressors increases; when n is larger than T, it actually becomes infeasible. Our three-pass procedure can therefore be interpreted as a regularized mimicking-portfolio estimator that exploits the benefits in terms of bias reduction that occur when $n \to \infty$ but preserves feasibility and efficiency via PC regressions.

To sum up, in standard cases with fixed n, the two-pass cross-sectional regression and mimicking-portfolio approaches tend to give different answers about the risk premium of a factor g_t. Our three-pass estimator represents the convergence of these two approaches that occurs when PCs are used to span the space of a large number of test assets.

3. Stochastic Discount Factor Interpretation

We conclude with a third interpretation of the procedure in terms of the stochastic discount factor. As discussed in Cochrane (2009), in linear factor models the risk premium of any factor g_t is simply the negative of the univariate covariance of g_t with the stochastic discount factor m_t. Formally, in our setting, $m_t = 1 - \gamma^\top\Sigma_v^{-1}v_t$, so that $-\text{Cov}(g_t, m_t) = \eta\gamma$. The first two steps of the procedure effectively recover the stochastic discount factor m_t in the linear factor model (which is invariant to the rotation of the factors, as discussed in the existing literature; see, e.g., Roll and Ross 1980; Huberman, Kandel, and Stambaugh 1987). The requirement of spanning the factor space is what allows to estimate the stochastic discount factor consistently. Step 3 effectively computes the univariate covariance between g_t and the stochastic discount factor m_t estimated in the first two stages. The invariance result is at play here because step 3 involves only a univariate covariance with m_t, which itself is invariant to the rotation of the factor space.

IV. Asymptotic Theory

In this section, we present the large sample distribution of our estimator as n, $T \to \infty$. For clarity of presentation, we leave the technical details of all assumptions to appendix A. Our results hold under similar or weaker assumptions compared with those in Bai (2003). This is because our goals are different. Our main target is $\eta\gamma$ instead of the asymptotic distributions of factors and their loadings.

One notable assumption is the so-called pervasive condition for a factor model, that is, assumption A6. It requires the factors to be sufficiently strong that most assets have nonnegligible exposures. This is a key identification condition, which dictates that the eigenvalues corresponding to the factor components of the return covariance matrix grow rapidly at a rate n, so that as n increases, they can be separated from the idiosyncratic component whose eigenvalues grow at a lower rate. The pervasiveness assumption precludes weak but priced latent factors—though, as will be clear later, it still allows for weak observable factors.[12]

A. *Limiting Distribution of the Risk Premia Estimator*

We now present the main theorem of the paper—the asymptotic distribution of the estimator $\hat{\gamma}_g$.

THEOREM 1. Under assumptions A1, A2, and A4–A11, and if we suppose that $\hat{p} \xrightarrow{p} p$, then as n, $T \to \infty$, we have

$$\begin{aligned} \hat{\gamma} - H\gamma &= H\bar{v} + O_p(n^{-1} + T^{-1}), \\ \hat{\eta} - \eta H^{-1} &= T^{-1}\bar{Z}\bar{V}^{\intercal}H^{\intercal} + O_p(n^{-1} + T^{-1}), \end{aligned}$$

for some matrix H that is invertible with probability approaching 1. Moreover, if $T^{1/2}n^{-1} \to 0$,

$$T^{1/2}\left(\hat{\gamma}_g - \eta\gamma\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Phi),$$

where the asymptotic covariance matrix is given by

$$\begin{aligned} \Phi = {} & \left(\gamma^{\intercal}(\Sigma^{v})^{-1} \otimes \mathbb{I}_d\right)\Pi_{11}\left((\Sigma^{v})^{-1}\gamma \otimes \mathbb{I}_d\right) + \left(\gamma^{\intercal}(\Sigma^{v})^{-1} \otimes \mathbb{I}_d\right)\Pi_{12}\eta^{\intercal} \\ & + \eta\Pi_{21}\left((\Sigma^{v})^{-1}\gamma \otimes \mathbb{I}_d\right) + \eta\Pi_{22}\eta^{\intercal}, \end{aligned} \tag{11}$$

and Π_{11}, Π_{12}, and Π_{22} are defined in assumption A11.

Remarkably, Φ does not depend on the covariance matrix of the residual u_t or the estimation error of β. Their impact on the asymptotic variance is of higher order because of the blessings of dimensionality ($n \to \infty$).

[12] We defer a more detailed discussion of weak factors to online app. III.4.

This is in sharp contrast to the classical fixed-n asymptotic results in Cochrane (2009), in which both β and Σ^u enter the asymptotic variance of the risk premia estimator and in which the time series estimation error of $\hat{\beta}$ contributes to an extra Shanken adjustment term in the asymptotic variance (Shanken 1992). Consequently, to conduct inference on $\eta\gamma$, there is no need to estimate the large covariance matrix of u_t. This also implies that the usual GLS or WLS estimator would not improve the efficiency of the OLS estimator to the first order.[13]

To illustrate the intuition behind this result, we compare the asymptotic variance expressions in a special case, where there is no measurement error and all factors are known and observable (i.e., $\eta = \mathbb{I}_p$). The large-T fixed-n analysis yields

$$\mathrm{Avar}_{\mathrm{OLS}}(\hat{\gamma}) = (\beta^\intercal\beta)^{-1}\beta^\intercal\Sigma^u\beta(\beta^\intercal\beta)^{-1}\underbrace{\left(1+\gamma^\intercal(\Sigma^v)^{-1}\gamma\right)}_{\text{Shanken adjustment for }\hat{\beta}} + \Sigma^v,$$

$$\mathrm{Avar}_{\mathrm{GLS}}(\hat{\gamma}) = (\beta^\intercal(\Sigma^u)^{-1}\beta)^{-1}\underbrace{\left(1+\gamma^\intercal(\Sigma^v)^{-1}\gamma\right)}_{\text{Shanken adjustment for }\hat{\beta}} + \Sigma^v.$$

Under the assumptions of our paper, we can show that $\|(\beta^\intercal\beta)^{-1}\| = O(n^{-1})$, $\|\beta^\intercal\Sigma^u\beta\| = O(n)$, and $\|(\beta^\intercal(\Sigma^u)^{-1}\beta)^{-1}\| = O(n^{-1})$. So it is easy to see that the leading order of the asymptotic variances of both OLS and GLS is Σ^v as n increases.[14] In this special case, our equation (11) simplifies to $\Phi = \Pi_{22}$, which, under the stationarity of v_t, is identical to Σ^v. Also, if the factors are themselves tradable portfolios, the asymptotic variance of their sample average returns is equal to Σ^v. In fact, Σ^v is the minimal variance that an estimator could achieve in this setting. Our estimator achieves this bound without the knowledge of factor identities.

In the general setting we consider (i.e., $\eta \neq \mathbb{I}_p$, $z_t \neq 0$), we can decompose the estimation error of $\hat{\gamma}_g$ into its two components: the error due to estimating $\hat{\gamma}$ and the error due to estimating $\hat{\eta}$. The dominating error term in the former arises from the time series average of the factor innovation v_t, whereas in the latter the dominating error term arises from the time series regression of g_t on v_t. A direct application of the delta method on these two leading terms yields the desired central limit result in theorem 1. The remaining error terms are of the order $O_p(n^{-1} + T^{-1})$ and are due to the

[13] Indeed, to formally prove this, we can show our estimator is asymptotically equivalent to the infeasible GLS estimator.

[14] Without loss of generality, we can rewrite model (1) in terms of observable factors f_t instead of their innovations v_t; i.e., $r_t = \beta\gamma + \beta(f_t - f) + u_t$. So our informal analysis of this special case echos Gagliardini, Ossola, and Scaillet (2016), who formally show that the estimation error of $\gamma - f$ is $O_p(n^{-1/2}T^{-1/2})$. Obviously, the estimation error of f is $O_p(T^{-1/2})$, which dominates the estimation error of $\gamma = \gamma - f + f$.

ignorance of the true factors, the error in variable bias in $\hat{\beta}$, and the error in the cross-sectional regression (cf. n. 13). Under the condition that $T^{1/2}n^{-1} = o(1)$, these errors are negligible with respect to $O_p(T^{-1/2})$, namely, the convergence rate of the achieved central limit result. In particular, this means that the asymptotic variance (11) is identical to that of the case in which all factors are observable, so that asymptotically our estimator behaves as if factors were fully observable. In a finite sample, however, the asymptotic efficiency loss due to not observing the factors could be large when n is relatively small. Also, $\hat{p}$ is likely not identical to the true number of factors, which further affects the finite sample performance.

B. *Mixed-Frequency Data*

In this section we explore the possibility of using mixed-frequency data to improve the performance of our estimator. More specifically, even though the factor g_t may be available only at a low frequency (e.g., quarterly or monthly for macro factors), returns of test assets are available at a higher frequency (e.g., daily). Let $\Delta = m^{-1}$ be the sampling frequency of high-frequency returns. Let $a^h_{t+k\Delta}$ denote the high-frequency return from $t + (k-1)\Delta$ to $t + k\Delta$ for $a = r, u$, and v. We can recycle the old notation a_t and use it as the low-frequency cumulative return from $t-1$ to t, that is, $a_t := \sum_{k=1}^{m} a^h_{t-1+k\Delta}$. We shall make assumptions on the high-frequency dynamics of $a^h_{t+k\Delta}$, such that the corresponding low-frequency cumulative return a_t satisfies the assumptions in appendix A and that we can evaluate the efficiency gain, if any, achievable with high-frequency data. In particular, we assume that the high-frequency test asset returns follow a linear factor model:

$$r^h_{t+k\Delta} = \beta\gamma\Delta + \beta v^h_{t+k\Delta} + u^h_{t+k\Delta} \tag{12}$$

for $1 \le k \le m$ and $1 \le t \le T$.

We then revise our procedure to make use of high-frequency returns. Specifically, in the first two steps of the three-pass procedure, we conduct PCA and cross-sectional regressions using $r^h_{t+k\Delta}$ and obtain $\hat{V}^h$, $\hat{\beta}^h$, and $\hat{\gamma}^h$, respectively, where the superscript h emphasizes the use of high-frequency data. Note that $\hat{V}^h$ is a $d \times (mT)$ matrix of the estimated high-frequency factors. For the third and last step, we instead regress g_t onto the low-frequency cumulative returns of the high-frequency factors:

$$\hat{\eta}^l = \bar{G}\hat{V}^{l\top}(\hat{V}^l\hat{V}^{l\top})^{-1},$$

where $\hat{V}^l = \hat{V}^h(\iota_m \otimes \mathbb{I}_T)$. Consequently, the new risk premia estimator is

$$\hat{\gamma}^m_g = m \times \hat{\eta}^l\hat{\gamma}^h. \tag{13}$$

The multiplier m ensures that the risk premia estimates are in the same unit as those based on low-frequency data.

The next theorem shows that this estimator is in fact asymptotically equivalent to our benchmark, $\hat{\gamma}_g$, which uses only low-frequency data (cumulative returns).

THEOREM 2. Suppose that the high-frequency returns $r^h_{t+k\Delta}$ satisfies (12), with its components β, $v^h_{t+k\Delta}$, and $u^h_{t+k\Delta}$ satisfying the same conditions as those given by assumptions A4–A7 and A9, except that the sample size is replaced by mT and the variance Σ^v is replaced by $\Sigma^v\Delta$.[15] In addition, assumptions A2, A8, A10, and A11 hold for the low-frequency cumulative returns, v_t and u_t. Suppose $\hat{p}\xrightarrow{p}p$; then as n, $T\to\infty$, with probably approaching 1, there exists some invertible matrix H^h, such that

$$\hat{\gamma}^h - H^h\gamma\Delta = m^{-1}H^h\bar{v} + O_p(n^{-1} + T^{-1}),$$
$$\hat{\eta}^l - \eta(H^h)^{-1} = \bar{Z}\bar{V}^\intercal(\bar{V}\bar{V}^\intercal)^{-1}(H^h)^{-1} + O_p(n^{-1} + T^{-1}).$$

Moreover, we have $\hat{\gamma}^m_g = \hat{\gamma}_g + O_p(n^{-1} + T^{-1})$.

Intuitively, g_t is available only at a lower frequency, so that the efficiency gain, if any, could arise only from part of the estimator that uses high-frequency data: $\hat{\gamma}^h$. But recall that the sample average return, $\bar{v}^h = (mT)^{-1}\sum_{t=1}^T\sum_{k=1}^m v^h_{t-1+k}$, is the leading term of $\hat{\gamma}^h - H^h\gamma\Delta$. And the sample average remains the same regardless of the sampling frequency of the observed returns: $\bar{v}^h = m^{-1}\bar{v}$. As a result, there is no efficiency gain to the first order when using (13) with high-frequency test asset returns.

C. *Using the Ridge Estimator*

In section III.B, we pointed out that our three-pass estimator can be regarded as the average excess return of a regularized mimicking portfolio, which uses PCs as basis assets onto which the factor of interest is projected. Effectively, we construct this mimicking portfolio using the PC regression (see, e.g., Friedman, Hastie, and Tibshirani 2009). In this section, we consider an alternative ridge regression approach to the construction of mimicking portfolios.

Ridge regression is a shrinkage method originally motivated to improve estimation and prediction in a linear regression problem (see Hoerl and Kennard 2000). Compared with the best linear unbiased OLS estimator, the ridge estimator is biased, but its variance may be smaller; thus, it potentially achieves a better trade-off in terms of mean squared error.

As discussed in section III.B, one approach to mimicking-portfolio construction is to regress g_t onto the entire set of test asset returns. However,

[15] The scaling factor Δ ensures that the variance of $v_t = \Sigma_{k=1}^m v^h_{t-1+k\Delta}$ remains Σ_v. Note that we consider only the case of a fixed Δ, so we do not keep track of Δ on the right-hand-side bounds in assumptions A4, A7, and A9.

this estimator is rather inefficient when n is comparable to T and becomes infeasible when n is larger than T. Using the ridge regression for estimating the mimicking-portfolio weights solves this issue. Specifically, the portfolio weights are given by $\hat{w}_g^{\mu} = (\bar{R}\bar{R}^{\top} + \mu\mathbb{I}_n)^{-1}\bar{R}\bar{G}^{\top}$, where μ is a tuning parameter. With $\mu > 0$, the matrix $\bar{R}\bar{R}^{\top} + \mu\mathbb{I}_n$ is always invertible. Moreover, via singular value decomposition of $\bar{R}$, that is, $\bar{R} = \varsigma D\xi^{\top}$, the ridge estimator can be written as $\hat{w}_g^{\mu} = \varsigma(DD^{\top} + \mu\mathbb{I}_n)^{-1}D\xi^{\top}\bar{G}^{\top}$, where ς is an $n \times n$ matrix of left singular vectors, D is an $n \times T$ diagonal matrix of singular values of $\bar{R}$, and ξ is a $T \times T$ matrix of right singular vectors. Clearly, the ridge estimator tends to result in more stable portfolio weights by shrinking the singular values of $\bar{R}$. The larger μ is, the larger the shrinkage effect is.

The ridge regression version of our risk premia estimator will then be

$$\hat{\gamma}_g^{\mu} = (\hat{w}_g^{\mu})^{\top}\bar{r} = \bar{G}\bar{R}^{\top}(\bar{R}\bar{R}^{\top} + \mu\mathbb{I}_n)^{-1}\bar{r} = \bar{G}\xi D(DD^{\top} + \mu\mathbb{I}_n)^{-1}\varsigma^{\top}\bar{r}. \tag{14}$$

To compare this estimator with our PCA-based three-pass procedure, we can rewrite (8) as

$$\hat{\gamma}_g = \bar{G}\hat{V}^{\top}(\hat{V}\hat{V}^{\top})^{-1}(\hat{\beta}^{\top}\hat{\beta})^{-1}\hat{\beta}^{\top}\bar{r} = \bar{G}\xi_{1:\hat{p}}D_{1:\hat{p}}^{-1}\varsigma_{1:\hat{p}}^{\top}\bar{r},$$

where $\xi_{1:\hat{p}}$ and $\varsigma_{1:\hat{p}}$ are submatrices of ξ and ς, respectively, including only their first $\hat{p}$ columns, and $D_{1:\hat{p}}$ is the $\hat{p} \times \hat{p}$ submatrix on the top left corner of D. Clearly, the difference between $\hat{\gamma}_g^{\mu}$ and $\hat{\gamma}_g$ is that the latter depends on only the first $\hat{p}$ singular values and singular vectors, whereas the former relies on all of them.

With an appropriate choice of μ, the next theorem establishes the consistency of the ridge estimator.

Theorem 3. Suppose that assumptions A1, A2, and A4–A11 hold, $\hat{p}\xrightarrow{p}p$, and μ is chosen such that $\mu n^{-1}T^{-1} \to 0$ and $\mu^{-1}(nT^{1/4} + n^{1/4}T)$ 0; then we have $\hat{\gamma}_g^{\mu} - \hat{\gamma}_g = o_p(1)$.

This theorem shows that using ridge regression in place of PCA in our analysis yields an alternative consistent estimator of risk premia. However, note that in general when the true number of factors is finite, the ridge regression is not as efficient as the three-pass estimator. Intuitively, the reason is that the ridge estimator puts weight on all assets without trying to distinguish latent factors from noise; since each of the spanning portfolios loads on the factors but also contains noise (idiosyncratic or unpriced risk), the ridge estimator will reflect some of that noise. Instead, the PCA analysis explicitly separates the latent factors from the noise and is asymptotically as efficient as if factors were observable (under the assumption that all latent factors are pervasive). In our empirical analysis, we therefore focus on the PCA-based three-pass estimator, but we also report for robustness the results obtained using the ridge regression approach.[16]

[16] The comparison between these estimators in the case of an infinite number of factors or in the case of weak factors is beyond the scope of the paper, and we leave it for future work.

D. Testing the Strength of an Observed Factor

As discussed in the introduction, a recent literature has explored the issues that arise when making inference on risk premia in the presence of weak factors (i.e., factors that are only weakly reflected in the cross section of test assets). Our methodology is in fact robust to the case in which observable factors g_t are weak. In particular, whether g_t is strong or weak can be captured by the signal-to-noise ratio of its relationship with the underlying factors v_t (from eq. [2]). If either $\eta = 0$ (g_t is not a pervasive factor) or $\eta \to 0$ (measurement error z_t dominates the g_t variation), then g_t will be weak, and returns exposures to g_t will be small.

Our procedure estimates equation (2) in the third pass and is therefore able to detect whether an observable factor g_t has zero or low exposure to the fundamental factors (η is small) or whether it is noisy (z_t is large) and corrects for it when estimating the risk premium. To measure the signal-to-noise ratio of each observable factor, we define the time series R^2 for each observable factor $g(1 \times T)$ in the time series regression of g_t on the latent factors, $R_g^2 = \eta\Sigma^v\eta^\top/(\Sigma^v\eta^\top + \Sigma^z)$, as well as its estimator, $\widehat{R}_g^2 = \hat{\eta}\hat{V}\hat{V}^\top\hat{\eta}^\top/\bar{G}\bar{G}^\top$, which we show to be consistent in theorem I.4 of the online appendix. This R_g^2 reveals how noisy g is, which, as we report in our empirical analysis, varies substantially across factor proxies. In this section, we provide a Wald test for the null hypothesis that a factor g is weak.

Without loss of generality, it is sufficient to consider the $d = 1$ case. To do so, we formulate the hypotheses $\mathbb{H}_0 : \eta = 0$ versus $\mathbb{H}_1 : \eta \neq 0$ and construct a Wald test. Our test statistic is given by

$$\hat{W} = T\hat{\eta}\left((\hat{\Sigma}^v)^{-1}\hat{\Pi}_{11}(\hat{\Sigma}^v)^{-1}\right)^{-1}\hat{\eta}^\top,$$

where $\hat{\Pi}_{11}$ and $\hat{\Sigma}^v$ are constructed in section IV.E.

The next theorem establishes the desired size control and consistency of the test as well as its convergence property under a sequence of local alternatives $\mathbb{H}_T : \eta = \eta_0 T^{-1/2}$.

THEOREM 4. Suppose $d = 1$, $\hat{p} \xrightarrow{p} p$, and $\hat{\Pi}_{11} \xrightarrow{p} H\Pi_{11}H^\top$ for the same H matrix in theorem 1. Under assumptions A2 and A4–A11, as n, $T \to \infty$, $T^{1/2}n^{-1} \to 0$, we have

$$\lim_{n,T\to\infty} \mathrm{P}\left(\hat{W} > \chi_{\hat{p}}^2(1-\alpha_0)|\mathbb{H}_0\right) = \alpha_0 \text{ and } \lim_{n,T\to\infty} \mathrm{P}\left(\hat{W} > \chi_{\hat{p}}^2(1-\alpha_0)|\mathbb{H}_1\right) = 1,$$

where $\chi_{\hat{p}}^2(1-\alpha_0)$ is the $(1-\alpha_0)$ quantile of the χ^2 distribution with $\hat{p}$ degree of freedom. Moreover, $\hat{W}$ follows a noncentral χ^2 distribution with $\hat{p}$ degrees of freedom and noncentrality parameter $\eta_0\Sigma^v\Pi_{11}^{-1}\Sigma^v\eta_0^\top$ under the sequence of local alternative hypotheses $\mathbb{H}_T$.

Note that our assumption that the latent factors are pervasive while observable factors can potentially be weak is not in conflict with existing empirical evidence. It is known from the literature (e.g., Bernanke and

Kuttner 2005; Lucca and Moench 2015) that the stock market and the bond market strongly react to Federal Reserve and government policies and that macroeconomic risks affect equity premia; fundamental macroeconomic shocks seem to be pervasive. At the same time, we do not observe all fundamental economic shocks directly and have instead to rely on observable proxies; these are known to be weak in some cases, like in the case of industrial production (see Gospodinov, Kan, and Robotti 2014; Bryzgalova 2015).

E. Asymptotic Variances Estimation

We develop consistent estimators of the asymptotic covariances in theorem 1:

$$\begin{aligned}\hat{\Phi} = {} & \left(\hat{\gamma}^{\top}(\hat{\Sigma}^{v})^{-1} \otimes \mathbb{I}_d\right)\hat{\Pi}_{11}\left((\hat{\Sigma}^{v})^{-1}\hat{\gamma} \otimes \mathbb{I}_d\right) + \left(\hat{\gamma}^{\top}(\hat{\Sigma}^{v})^{-1} \otimes \mathbb{I}_d\right)\hat{\Pi}_{12}\hat{\eta}^{\top} \\ & + \hat{\eta}\hat{\Pi}_{21}\left((\hat{\Sigma}^{v})^{-1}\hat{\gamma} \otimes \mathbb{I}_d\right) + \hat{\eta}\hat{\Pi}_{22}\hat{\eta}^{\top},\end{aligned}$$

where $\hat{\Pi}_{11}$, $\hat{\Pi}_{12}$, and $\hat{\Pi}_{22}$ are the heteroskedasticity and autocorrelation consistent type estimators of Newey and West (1987), defined as

$$\begin{aligned}\hat{\Pi}_{11} = {} & \frac{1}{T}\sum_{t=1}^{T}\mathrm{vec}(\hat{z}_t\hat{v}_t^{\top})\mathrm{vec}(\hat{z}_t\hat{v}_t^{\top})^{\top} \\ & + \frac{1}{T}\sum_{m=1}^{q}\sum_{t=m+1}^{T}\left(1 - \frac{m}{q+1}\right)\left(\mathrm{vec}(\hat{z}_{t-m}\hat{v}_{t-m}^{\top})\mathrm{vec}(\hat{z}_t\hat{v}_t^{\top})^{\top} + \mathrm{vec}(\hat{z}_t\hat{v}_t^{\top})\mathrm{vec}(\hat{z}_{t-m}\hat{v}_{t-m}^{\top})^{\top}\right), \\ \hat{\Pi}_{12} = {} & \frac{1}{T}\sum_{t=1}^{T}\mathrm{vec}(\hat{z}_t\hat{v}_t^{\top})\hat{v}_t^{\top} + \frac{1}{T}\sum_{m=1}^{q}\sum_{t=m+1}^{T}\left(1 - \frac{m}{q+1}\right)\left(\mathrm{vec}(\hat{z}_{t-m}\hat{v}_{t-m}^{\top})\hat{v}_t^{\top} + \mathrm{vec}(\hat{z}_t\hat{v}_t^{\top})\hat{v}_{t-m}^{\top}\right), \\ \hat{\Pi}_{22} = {} & \frac{1}{T}\sum_{t=1}^{T}\hat{v}_t\hat{v}_t^{\top} + \frac{1}{T}\sum_{m=1}^{q}\sum_{t=m+1}^{T}\left(1 - \frac{m}{q+1}\right)\left(\hat{v}_{t-m}\hat{v}_t^{\top} + \hat{v}_t\hat{v}_{t-m}^{\top}\right),\end{aligned}$$

where $\hat{Z} = \bar{G} - \hat{\eta}\hat{V}$, $\hat{\Sigma}^{\beta} = n^{-1}\hat{\beta}^{\top}\hat{\beta}$, $\hat{\Sigma}^{v} = T^{-1}\hat{V}\hat{V}^{\top}$, and q is the usual lag parameter in Newey-West type of estimators. Theorem I.6 in the online appendix establishes the desired consistency of these estimators.

V. Empirical Analysis

In this section we apply our three-pass methodology to the data. We estimate the risk premia of several factors, both traded and not traded, and show how our results differ from those obtained using standard two-pass cross-sectional regressions and mimicking portfolios.

A. Data

We conduct our empirical analysis on a large set of 647 portfolios that include US equities sorted by a large number of characteristics as well as treasury bonds, corporate bonds, and currencies (for a detailed description of the test portfolios and the data sources for the empirical analysis, see online app. III). Note that our methodology is designed to work specifically in large-n environments, so there is no cost in adding test portfolios. Because of limited data availability for the nonequity portfolios, the sample covers the period 1976–2010. We perform the analysis at the monthly frequency and work with factors that are available at the monthly frequency.

Although the asset pricing literature has proposed an extremely large number of factors (Harvey, Liu, and Zhu 2016; McLean and Pontiff 2016), we focus here on a few representative ones. Recall that the observable factors g_t in the three-pass methodology can be either an individual factor or groups of factors. We consider here both cases to illustrate the methodology; importantly, the risk premia estimates for any factors using our three-pass methodology do not depend on whether other factors are included in g_t (though this does matter for the two-pass cross-sectional estimator).

The factors we consider include both tradable and nontradable factors. The tradable factors are market (in excess of the risk-free rate), size (SMB), value (HML), profitability (RMW), investment (CMA), momentum (MOM), betting against beta (BAB; from Frazzini and Pedersen 2014), and quality minus junk (QMJ; from Asness, Frazzini, and Pedersen 2013). The nontradable factors are AR(1) innovations in industrial production growth (IP), VAR(1) innovations in the first three PCs of 279 macrofinance variables from Ludvigson and Ng (2010), the liquidity factor of Pástor and Stambaugh (2003), two intermediary capital factors (one from He, Kelly, and Manela [2017] and one from Adrian, Etula, and Muir [2014]), four factors from Novy-Marx (2014; high monthly temperature in Manhattan, global land surface temperature anomaly, quasiperiodic Pacific Ocean temperature anomaly [El Niño], and the number of sunspots), and two consumption-based factors from Malloy, Moskowitz, and Vissing-Jorgensen (2009), which include both an aggregate consumption series and a stockholder's consumption series.[17]

B. Factors from the Large Panel of Returns

The first step for estimating the observable factor risk premia is to determine the dimension of the latent factor model, p. In the online appendix, figure III.3 (left panel) reports the first 20 eigenvalues of the covariance matrix of returns for our panel of 647 portfolios. As typical for large panels,

[17] These time series have been proposed by Novy-Marx (2014) as examples of variables that appear to predict returns in standard predictive regressions but whose economic link to the stock market seems weak. We use AR(1) innovations in these series as factors and test whether our procedure identifies the weak link to the economy and reveals the series as weak or unpriced in the cross section of returns.

the first eigenvalue tends to be much larger than the others. In the right panel, we zoom in on the eigenvalues 5–20. We observe a noticeable decrease in the eigenvalues after the seventh one, suggesting $\hat{p} = 7$. This is also the number suggested by our estimator given by online appendix I.1. The discussion therein (theorem I.2) also suggests that our estimator is consistent as long as the number of factors we use (e.g., $\breve{p}$) is at least as large as the true dimension p. Indeed, the additional analysis in online appendix III.2 shows the robustness of our empirical results with respect to the choice of $\breve{p}$.[18]

The model with seven PCs has a cross-sectional R^2 of 59%, indicating that it accounts for a significant fraction of the cross-sectional variation in expected returns for the 647 test portfolios but leaving some unexplained variation. This number is comparable with the 73% cross-sectional R^2 that one obtains using the Fama-French three-factor model on the cross section of 25 portfolios sorted by size and book to market, yet we obtain it for a cross section 26 times as large and using a model with just four more factors. Using 10 and 13 factors (as we do in the online appendix) raises this cross-sectional R^2 to 66% and 68%, respectively.

C. Risk Premia Estimates

We now present the estimates for the risk premia of observable factors using excess returns under the assumption that the zero-beta rate is equal to the observed risk-free (treasury bill) rate. For each factor (or group of factors) g_t that we consider, table B1 reports risk premia estimated using different methodologies.[19] Column 1 reports the time series average excess return of the factor when the factor is tradable. This represents a model-free estimator of the factor risk premium; however, this is possible only for tradable factors.

Columns 3–14 of table B1 consider three implementations of the two-pass cross-sectional estimator, two implementations of the mimicking-portfolio estimator, and our three-pass estimator. For each set of results, we report the risk premium estimate and its standard error.

Using the two-pass cross-sectional regression, we estimate the risk premium of each observable factor g_t without any additional control factors (first set of results), controlling for the market return (second set of

[18] Figure III.4 in the online appendix further shows that with seven factors eliminated, the correlation matrix of the residuals appears more sparse, which is consistent with the implications of a factor model.

[19] As a side note, the asymptotic variance of risk premia estimators based on the two-pass cross-sectional regression with observable factors, v_t, simplifies to Σ_v in the case of large n and large T regime, as we have pointed out in sec. IV.A. We thereby use $\hat{\Pi}_{22}$ above (with $\hat{v}_t$ replaced by v_t since they are observable) throughout simulations and empirical studies for their inference.

results), and controlling for the Fama-French three factors (market, SMB, HML; third set of results).[20] Next, using the mimicking-portfolio approach, we project the factor g_t onto the market portfolio alone (first set of results) and onto the Fama-French three factors (second set of results). In theory, one could also project g_t onto the entire universe of 647 test assets. Note that the latter version of the mimicking-portfolio approach (that projects onto all the available assets) is rarely applied in the literature, as it is at best inefficient because n is often large relative to T. In our case, it is actually infeasible, since the number of test assets is greater than the number of time periods used in the estimation ($n > T$)—so the mimicking portfolio cannot be computed. Finally, we report our three-pass estimator in columns 13 and 14 of table B1 using $\breve{p} = 7$, as suggested by our estimator.

To help with the interpretation of table B1, we first examine one example in detail. Consider the momentum factor (MOM). The time series average excess return is 69 basis points (bps) per month, with a 24-bp standard error. Since this factor is traded, we would expect any consistent estimator to recover a risk premium close to 69 bps. Estimating the momentum risk premium in a two-pass cross-sectional regression with no controls yields a strongly significant estimate of −201 bps per month. Adding the market as a control in the cross-sectional regression gives 20 bps but is statistically insignificant, and further adding SMB and HML gives 71 bps. Clearly, the results strongly depend on which factors are used as controls in the estimation. Similarly, consider the two implementations of the mimicking-portfolio approach. When we project MOM onto the market alone or onto the Fama-French three factors (the latter being a typical choice of portfolios for the projection in the empirical literature), we obtain negative risk premia estimates, small in the first case (−5 bps) and larger and significant in the second case (−21 bps). Not only are the mimicking-portfolio estimates at odds with the observed average excess return of the momentum portfolio (69 bps per month), but also they vary significantly with the choice of portfolios on which the factor is projected. Finally, our estimator provides a statistically significant 49-bp estimate for the risk premium of this factor.

We now summarize the main patterns of results obtained using different estimators in table B1.

1. Two-Pass Cross-Sectional Estimator

Two-pass cross-sectional estimators are subject to two potential biases: omitted controls and measurement error. Both are clearly visible in table B1.

[20] When the factor of interest is also in the set of controls—e.g., when we estimate the risk premium for the market in cols. 5–12, in which controls are R_m or Fama-French three factors—we exclude it from the set of controls.

First, for most factors, there are extreme differences in the estimates obtained using different control factors (i.e., no controls, market, Fama-French three factors), like in the case of MOM described above. This shows how sensitive this estimator is to the control factors and illustrates the potential for quantitatively meaningful biases that could arise if the wrong set of controls is specified.

The second potential source of bias is due to measurement error. As discussed in section II.B, measurement error induces a bias in risk premia estimates. In addition, two-pass cross-sectional regressions have well-known biases due to the presence of weak factors in the model (factors that are dominated by noise). Looking at table B1, measurement error appears to be the reason for the often extreme risk premia estimates obtained using the two-pass regression for what appear to be weak or noisy factors. For example, three of the four pure-noise factors of Novy-Marx (2014) are estimated to have huge magnitudes and statistical significance by the cross-sectional estimator. As described in section IV.D, our three-pass procedure is immune to the problem of weak or noisy observable factors, and it estimates that the Novy-Marx factor risk premia are statistically indistinguishable from zero.

2. Mimicking-Portfolio Estimator

The mimicking-portfolio estimator is similarly sensitive to the choice of portfolios on which factors are projected. In several cases, in fact, the estimator yields opposite signs across the different sets of projection portfolios. For example, BAB and CMA are estimated to have a negative risk premium when projected onto the market and a positive one when projected onto the Fama-French three-factor model portfolios. In addition, in many cases, the mimicking-portfolio estimator yields estimates of risk premia that have the wrong sign relative to the average excess return for traded factors, which suggests that the estimator is biased (omitted variable bias). Finally, as discussed above, the mimicking-portfolio approach becomes infeasible when g_t is projected on all available assets (since $n > T$), highlighting the need to select a subset of the assets for the estimator to be feasible, which in turn can induce a bias if the set of projection assets chosen by the econometrician does not span the right space (see sec. II.A). Overall, table B1 shows that, like the two-step cross-sectional estimator, the mimicking-portfolio estimator is subject to a quantitatively meaningful bias that can arise if important portfolios are omitted from the projection.

3. Three-Pass Estimator

Columns 13 and 14 of table B1 reports the results using our three-pass estimator, using seven PCs. For the case of tradable factors, the estimator produces results that are generally close (economically and statistically) to the average excess returns of the factors, contrary to the alternative estimators discussed above.

Among the nontradable factors, our estimator finds that several of them carry economically and statistically significant risk premia: the liquidity factor of Pástor and Stambaugh (2003), both intermediary factors of He, Kelly, and Manela (2017) and Adrian, Etula, and Muir (2014), the first macro PC from Ludvigson and Ng (2010), and also stockholders' consumption growth from Malloy, Moskowitz, and Vissing-Jorgensen (2009). Instead, several other nontradable factors do not appear to have statistically significant risk premia, for example, the Novy-Marx (2014) factors. IP growth appears marginally statistically different from zero but extremely small in magnitude (−1 bps).

To conclude, for the tradable factors we study, the three-pass estimator produces results that are broadly consistent with the time series average returns of those factors; for the nontradable factors, they produce estimates that have economically reasonable magnitudes. The results are often noticeably different from those produced by the other estimators, which vary substantially across implementations.

4. Measurement Error in Factors and a Test for Weak Factors

Column 15 of table B1 reports the R^2 of the time series regression of each observed factor g_t onto the $\hat{p}$ latent factors; we refer to this as R_g^2. The R_g^2 will be lower than 100% when measurement error is present in the factor g_t. In the data, we find great heterogeneity among factors in terms of their measurement error. For some of them (like the market or SMB), this R_g^2 is extremely high, suggesting that the factor is measured essentially without error. For many other factors, and especially so for nontradable factors, the R_g^2 is much lower (e.g., for IP, it is around 2%), indicating that these factors are dominated by noise.[21] Online appendix III.8 explores this result in greater detail, showing how our procedure can be used to de-noise the factors.

Finally, column 16 of table B1 reports the p-value for the test of the null that each factor g_t is weak, described in section IV.D. A rejection of the null indicates that g_t is a strong factor for the cross section of test portfolios. For several—but not all—of the nontradable factors, we fail to reject that the factor is weak.

[21] Of course, the low R^2 could instead be due to the test assets not spanning the factor or the selected number of PCs being too small. Recall, however, that our empirical results are obtained using a large cross section that includes the most important dimensions of risk in the equity market, treasury and corporate bonds, and currency markets. Online app. III.2 further illustrates the robustness with respect to using more PCs.

5. The Zero-Beta Rate and the Sign of the Market Risk Premium

Our estimator can also be applied in the case the zero-beta rate is estimated (instead of being set equal to the treasury bill rate). Results (reported in table III.9 of the online appendix) are broadly similar to the baseline case in which the zero-beta rate is restricted, though of course the exact estimates of the risk premia are different. There is, however, one interesting result that is worth remarking and that applies specifically to the case in which the zero-beta rate is estimated. A well-known fact in the empirical asset pricing literature is that in standard factor models (like the Fama-French three-factor model), the market risk premium is estimated to be negative in the cross section of equity portfolios when using Fama-MacBeth regressions with an unrestricted zero-beta rate. We confirm this pattern in our data as well (we obtain a zero-beta rate estimate of 129 bps per month and a market risk premium of -24 bps using the Fama-MacBeth estimator). If this puzzling result is due to the omission of important controls (measurement error bias is unlikely because of the large R_g^2), we expect our three-pass estimator to correct for it. Indeed, our estimator yields a positive estimate of the market risk premium in all cases.

D. Additional Empirical Results and Robustness Tests

The online appendix presents several additional empirical results and robustness tests. Among them, it reports the results obtained when estimating the zero-beta rate, when using a greater number of factors as controls (specifically, 10 and 13), and when using only equity portfolios to perform the estimation. The online appendix also shows robustness with respect to the choice of test portfolios and time periods.

Finally, the online appendix also explores alternatives to standard PCA for reducing the dimensionality of the returns space. First, we show that the results are similar when changing the penalty function used to estimate the factors in a way that emphasizes not only the covariation among returns but also the ability of the factors to explain the cross section of risk premia (similar to Connor and Korajczyk [1986] and Lettau and Pelger [2018]). Second, we show that using the ridge regression (see sec. IV.C) instead of PCA yields similar results.

VI. Conclusion

We propose a three-pass methodology to estimate the risk premium of observable factors in a linear asset pricing model that is consistent even when not all factors in the model are specified and observed. The methodology builds on a simple invariance result that states that to correct the omitted variable problem in cases where not all factors are observed, it is sufficient to control for arbitrarily rotated factors that span the entire factor space. In this case, the risk premia for observable factors are consistently estimated even though the risk exposures cannot be identified. We propose to employ PCA to recover the factor space and effectively use the PCs as controls in the cross-sectional regressions together with the observable factors.

Our three-pass procedure can be viewed as an extension of both the standard two-pass cross-sectional regression approach and the mimicking-portfolio estimator of risk premia. In particular, it can be thought of as a factor-augmented two-pass cross-sectional estimator, where the model adds PCs of returns as controls in the two-pass regressions, completing the factor space. It can also be thought of as a regularized mimicking-portfolio estimator, in which the factor of interest is projected onto the PCs of returns (themselves portfolios). As we discuss in the paper, our method represents the convergence of the two methods, which occurs as $n \to \infty$.

The main advantage of our methodology is that it provides a systematic way to tackle the concern that the model predicted by theory is misspecified because of omitted factors. Rather than relying on arbitrarily chosen control factors or computing risk premia only on subsets of the test assets, our methodology utilizes the large dimension of testing assets available to span the factor space. It also explicitly takes into account the possibility of measurement error in any observed factor.

An application of our estimator to several nontradable factors yields interesting empirical results. In our main application, we show that while many standard macroeconomic factors (e.g., industrial production and PCs of macroeconomic series) appear to have insignificant risk premia, nontradable factors related to various market frictions (like liquidity and intermediary leverage) have in fact robustly strong and significant risk premia when considered as part of richer linear pricing models that include additional factors extracted from the cross section of returns. Our methodology can therefore help discriminate which macroeconomic (or other nontradable) factors are priced by investors.

Appendix A

Assumptions and Technical Details

We need more notation. We use $\lambda_j(A)$, $\lambda_{\min}(A)$, and $\lambda_{\max}(A)$ to denote the jth, the minimum, and the maximum eigenvalues of a matrix A. By convention, $\lambda_1(A) = \lambda_{\max}(A)$. In addition, we use $\|A\|_1$, $\|A\|_\infty$, $\|A\|$, and $\|A\|_F$ to denote the $\mathbb{L}_1$ norm, the $\mathbb{L}_\infty$ norm, the operator norm (or $\mathbb{L}_2$ norm), and the Frobenius norm of a matrix $A = (a_{ij})$, that is, $\max_j \Sigma_i |a_{ij}|$, $\max_i \Sigma_j |a_{ij}|$, $(\lambda_{\max}(A^\intercal A))^{1/2}$, and $(\mathrm{Tr}(A^\intercal A))^{1/2}$, respectively. We also use $\| A \|_{\mathrm{MAX}} = \max_{i,j} |a_{ij}|$ to denote the $\mathbb{L}_\infty$

norm of A on the vector space. The term K is a generic constant that may change from line to line.

For clarity, we restate the assumptions on the dynamics of returns and factors, that is, (1) and (2):

Assumption A1. Suppose that f_t is a $p \times 1$ vector of asset pricing factors and that r_t denotes an $n \times 1$ vector of excess returns of the testing assets. The pricing model satisfies

$$r_t = \beta\gamma + \beta v_t + u_t, \quad f_t = f + v_t, \quad \mathrm{E}(v_t) = \mathrm{E}(u_t) = 0, \quad \text{and} \quad \mathrm{Cov}(u_t, v_t) = 0,$$

where v_t is a $p \times 1$ vector of innovations of f_t, u_t is an $n \times 1$ vector of idiosyncratic components, β is an $n \times p$ factor loading matrix, and γ is the $p \times 1$ risk premia vector.

Assumption A2. There is an observable $d \times 1$ vector, g_t, of factor proxies that satisfies

$$g_t = \delta + \eta v_t + z_t, \quad \mathrm{E}(z_t) = 0, \quad \text{and} \quad \mathrm{Cov}(z_t, v_t) = 0,$$

where η, the loading of g on v, is a $d \times p$ matrix, δ is a $d \times 1$ constant, and z_t is a $d \times 1$ measurement-error vector.

Next, we impose one restrictive assumption, which is used only in proposition 1 and section III.B to illustrate the intuition of our result and the connection between the two-pass cross-sectional regression and the factor mimicking portfolios. Our asymptotic analysis below does not rely on this assumption.

Assumption A3. Suppose that v_t, z_t, and u_t in (1) are stationary time series independent of β, respectively, and that the weights of the spanning portfolios, $\breve{r}_t$, are given by the $n \times \breve{p}$ matrix w, with $\breve{p} \geq p$. The covariance matrices of v_t and u_t (i.e., Σ^v and Σ^u) and the loading of z_t on $\breve{u}_t := w^\top r_t$ (i.e., $\breve{\beta}^{z,u}$) satisfy the following conditions: $\lambda_{\min}^{-1}(\Sigma^v) = O_p(1)$, $\| \breve{\beta}^{z,u}\breve{\beta} \|_{\mathrm{MAX}} = O_p(1)$, $\lambda_{\min}^{-1}(\breve{\beta}^\top\breve{\beta}) = O_p(1)$, and $\lambda_{\max}(\breve{\Sigma}^u) = O_p(s_n n^{-1})$, where $\breve{\beta} := w^\top\beta$, $s_n = o_p(n)$.

The condition on $\lambda_{\min}(\Sigma^v)$ requires the set of factors in (1) to have a full rank covariance matrix; the second condition on $\breve{\beta}^{z,u}\breve{\beta}$ restricts the exposure of g_t to the idiosyncratic errors $\breve{u}_t$; the condition on $\breve{\beta}^\top\breve{\beta}$ resembles the usual pervasiveness assumption that guarantees nontrivial exposure of spanning portfolios to factors; the last restriction on $\breve{\Sigma}^u$ ensures that the idiosyncratic errors of spanning portfolios are diversifiable. These conditions turn out sufficient for the difference between the risk premium of g_t and that of its factor-mimicking portfolios to diminish, as shown in proposition 1.

There are two notable choices of w that are relevant for our study. The first case sets $w = n^{-1/2}\mathbb{I}_n$; that is, g_t is projected onto the entire set of test assets r_t. In this case, the conditions in assumption A3 are similar to the identification conditions of the approximate factor models. In particular, the last condition is more general than the bounded eigenvalue assumption introduced in Chamberlain and Rothschild (1983). The second choice sets $w = n^{-1}\beta H$ for any invertible matrix H. That is, the base portfolios are constructed using weights proportional to the exposure of the test assets. Because β is unknown, this case is not feasible. However, it is precisely what motivates the (feasible) construction of the three-pass estimator: $w = \hat{\beta}(\hat{\beta}^\top\hat{\beta})^{-1}$.

The following assumptions are more general, which we rely on to derive the asymptotic results in the paper. These high-level assumptions can be justified using stronger and more primitive conditions, such as those in assumption A3.

We proceed with the idiosyncratic component u_t and define, for any t, $t' \leq T$:

$$\gamma_{n,tt'} = \mathrm{E}\left(n^{-1}\sum_{i=1}^{n} u_{it}u_{it'}\right).$$

ASSUMPTION A4. There exists a positive constant K, such that for all n and T,

$$\text{(i)}\ T^{-1}\sum_{t=1}^{T}\sum_{t'=1}^{T}|\gamma_{n,tt'}| \leq K, \quad \max_{1\leq t\leq T}\gamma_{n,tt} \leq K,$$

$$\text{(ii)}\ T^{-2}\sum_{s=1}^{T}\sum_{t=1}^{T}\mathrm{E}\left(\sum_{j=1}^{n}(u_{js}u_{jt} - \mathrm{E}(u_{js}u_{jt}))\right)^2 \leq Kn.$$

Assumption A4 is similar to part of assumption C in Bai (2003), which imposes restrictions on the cross-sectional dependence and heteroskedasticity of u_t.

ASSUMPTION A5. The factor innovation V satisfies

$$\| \bar{v} \|_{\mathrm{MAX}} = O_p(T^{-1/2}), \quad \| T^{-1}VV^{\top} - \Sigma^{v} \|_{\mathrm{MAX}} = O_p(T^{-1/2}),$$

where Σ^v is a $p \times p$ positive-definite matrix and $0 < K_1 < \lambda_{\min}(\Sigma^v) \leq \lambda_{\max}(\Sigma^v) < K_2 < \infty$.

Assumption A5 imposes rather weak conditions on the time series behavior of the factors. It holds if factors are stationary, strong mixing, and satisfy some moment conditions.

ASSUMPTION A6. The factor loadings matrix β satisfies

$$\| n^{-1}\beta^{\top}\beta - \Sigma^{\beta} \| = o_p(1)$$

as $n \to \infty$, where Σ^β is a $p \times p$ positive-definite matrix and $0 < K_1 < \lambda_{\min}(\Sigma^\beta) \leq \lambda_{\max}(\Sigma^\beta) < K_2 < \infty$.

This is the key identifying assumption that imposes all factors to be pervasive and hence excludes weaker ones. Onatski (2012) develops the inference methodology in a framework that allows for weak factors using a Pitman-drift-like asymptotic device.

ASSUMPTION A7. The factor loadings matrix β and the idiosyncratic error u_t satisfy the following moment conditions for all $1 \leq j \leq p$ and for all n and T:

$$\text{(i)}\ \mathrm{E}\sum_{t=1}^{T}\left(\sum_{i=1}^{n}\beta_{ij}u_{it}\right)^2 \leq KnT,$$

$$\text{(ii)}\ \mathrm{E}\left(\sum_{t=1}^{T}\sum_{i=1}^{n}\beta_{ij}u_{it}\right)^2 \leq KnT.$$

The above assumption can be derived from a stronger cross-sectional independence assumption between β and u_t as well as some moment conditions on β, which are imposed by Bai (2003).

Assumption A8. The residual innovation Z satisfies

$$\| \bar{z} \|_{\mathrm{MAX}} = O_p(T^{-1/2}), \quad \| T^{-1} ZZ^{\top} - \Sigma^z \|_{\mathrm{MAX}} = O_p(T^{-1/2}),$$

where Σ^z is positive definite and $0 < K_1 < \lambda_{\min}(\Sigma^z) \le \lambda_{\max}(\Sigma^z) < K_2 < \infty$. In addition,

$$\| ZV^{\top} \|_{\mathrm{MAX}} = O_p(T^{1/2}).$$

Similar to assumption A5, assumption A8 holds if z_t is stationary, strong mixing, and satisfies some moment condition. It is more general than the independently and identically distributed assumption on z_t, which also applies to nontradable factor proxies in the empirical applications.

Assumption A9. For all n and T and $i, j \le p$, $l \le d$, the following moment conditions hold:

$$\text{(i) } \mathrm{E}\sum_{k=1}^{n}\left(\sum_{t=1}^{T} v_{jt} u_{kt}\right)^2 \le KnT,$$

$$\text{(ii) } \mathrm{E}\left(\sum_{t=1}^{T}\sum_{k=1}^{n} v_{it} u_{kt} \beta_{kj}\right)^2 \le KnT.$$

Assumption A9 resembles assumption D in Bai (2003). The variables in each summation have zero means, so that the required rate can be justified under more primitive assumptions. In fact, it holds trivially if v_t and u_t are independent.

Assumption A10. For all n and T and $l \le d$, $j \le p$, the following moment conditions hold:

$$\text{(i) } \mathrm{E}\sum_{k=1}^{n}\left(\sum_{t=1}^{T} z_{lt} u_{kt}\right)^2 \le KnT,$$

$$\text{(ii) } \mathrm{E}\left(\sum_{t=1}^{T}\sum_{k=1}^{n} z_{lt} u_{kt} \beta_{kj}\right)^2 \le KnT.$$

Similar to assumption A9, assumption A10 restricts the dependence between the idiosyncratic component u_t and the projection residual z_t. If z_t, u_t, and β are independent, conditions i and ii are easy to verify. For a tradable portfolio factor in g_t, we can interpret its corresponding z_t as certain undiversified idiosyncratic risk, since z_t is a portfolio of u_t, as implied from assumptions A1 and A2. It is thereby reasonable to allow for dependence between z_t and u_t. For nontradable factors, z_t's can also be correlated with u_t in general.

Assumption A11. As $T \to \infty$, the following joint central limit theorem holds:

$$T^{1/2}\begin{pmatrix} T^{-1}\mathrm{vec}(ZV^{\top}) \\ \bar{v} \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Pi_{11} & \Pi_{12} \\ \Pi_{12}^{\top} & \Pi_{22} \end{pmatrix}\right),$$

where Π_{11}, Π_{12}, and Π_{22} are $dp \times dp$, $dp \times p$, and $p \times p$ matrices, respectively, defined as

$$\Pi_{11} = \lim_{T\to\infty}\frac{1}{T}E\left(\mathrm{vec}(ZV^\top)\mathrm{vec}(ZV^\top)^\top\right),$$

$$\Pi_{12} = \lim_{T\to\infty}\frac{1}{T}E(\mathrm{vec}(ZV^\top)\iota_T^\top V^\top),$$

$$\Pi_{22} = \lim_{T\to\infty}\frac{1}{T}E(V\iota_T\iota_T^\top V^\top).$$

Assumption A11 describes the joint asymptotic distribution of $ZV^\top$ and $V\iota_T$. Because the dimensions of these random processes are finite, this assumption is a fairly standard result of some central limit theorem for mixing processes, for example, theorem 5.20 of White (2000). Needless to say, it is stronger than assumption A5, which is sufficient for identification and consistency.

Appendix B

Mathematical Proofs

We provide here the proofs of all theorems in the main text and leave proofs of the technical lemmas used below to the online appendix.

B1. Proofs of Main Theorems

B1.1. Proof of Proposition 1

Given the weights of the mimicking portfolios, and by assumption A1, these portfolio returns (i.e., $\breve{r}_t = w^\top r_t$) satisfy the following factor model:

$$\breve{r}_t = \breve{\beta}\gamma + \breve{\beta}v_t + \breve{u}_t, \tag{B1}$$

where $\breve{\beta} = w^\top\beta$ and $\breve{u}_t = w^\top u_t$. Using assumption A2, conditioning on β, the bias of the factor mimicking-portfolio estimator is

$$\gamma_g^{\mathrm{MP}} - \eta\gamma = \eta\Sigma^v\breve{\beta}^\top(\breve{\Sigma}^r)^{-1}\breve{\beta}\gamma - \eta\gamma + \breve{\Sigma}^{z,u}(\breve{\Sigma}^r)^{-1}\breve{\beta}\gamma,$$

where $\breve{\Sigma}^{z,u} = \Sigma^{z,u}w$ and $\breve{\Sigma}^r = w^\top\Sigma^r w$.

Because of (B1), we have $\breve{\Sigma}^r = \breve{\beta}\Sigma^v\breve{\beta}^\top + \breve{\Sigma}^u$, so that by Woodbury matrix identity

$$(\breve{\Sigma}^r)^{-1} = (\breve{\Sigma}^u)^{-1} - (\breve{\Sigma}^u)^{-1}\breve{\beta}\left(\breve{\beta}^\top(\breve{\Sigma}^u)^{-1}\breve{\beta} + (\Sigma^v)^{-1}\right)^{-1}\breve{\beta}^\top(\breve{\Sigma}^u)^{-1},$$

$$\breve{\beta}^\top(\breve{\Sigma}^r)^{-1}\breve{\beta} = \breve{\beta}^\top(\breve{\Sigma}^u)^{-1}\breve{\beta} - \breve{\beta}^\top(\breve{\Sigma}^u)^{-1}\breve{\beta}\left(\breve{\beta}^\top(\breve{\Sigma}^u)^{-1}\breve{\beta} + (\Sigma^v)^{-1}\right)^{-1}\breve{\beta}^\top(\breve{\Sigma}^u)^{-1}\breve{\beta}.$$

This further implies that

$$\left(\breve{\beta}^\top(\breve{\Sigma}^r)^{-1}\breve{\beta}\right)^{-1} = \left(\breve{\beta}^\top(\breve{\Sigma}^u)^{-1}\breve{\beta}\right)^{-1} + \Sigma^v,$$

where we use the fact that $\breve{p} \geq p$.

Using this equation and by direct calculations, we obtain

$$\begin{aligned}\eta\Sigma^{v}\breve{\beta}^{\top}(\breve{\Sigma}^{r})^{-1}\breve{\beta}\gamma - \eta\gamma &= \eta\left(\left(\mathbb{I}_p + \left(\breve{\beta}^{\top}(\breve{\Sigma}^{u})^{-1}\breve{\beta}\right)^{-1}(\Sigma^{v})^{-1}\right)^{-1} - \mathbb{I}_p\right)\gamma \\ &= -\eta\left(\mathbb{I}_p + A\right)^{-1}A\gamma,\end{aligned}$$

where $A = (\breve{\beta}^{\top}(\breve{\Sigma}^{u})^{-1}\breve{\beta})^{-1}(\Sigma^{v})^{-1}$. We can then show that the maximum eigenvalue of A goes to zero as $n \to \infty$, so the bias disappears asymptotically.

In fact, under assumption A3, $\lambda_{\max}(\Sigma^{\smile u}) = O_p(s_n n^{-1})$, then

$$\lambda_{\max}(A) = \lambda_{\min}^{-1}(\Sigma^{v}\breve{\beta}^{\top}(\breve{\Sigma}^{u})^{-1}\breve{\beta}) \leq \lambda_{\min}^{-1}(\Sigma^{v})\lambda_{\min}^{-1}(\breve{\beta}^{\top}\breve{\beta})\lambda_{\max}(\breve{\Sigma}^{u}) = O_p(s_n n^{-1}).$$

By Weyl's inequality, for any fixed $\epsilon > 0$, with probability approaching 1

$$\lambda_{\min}(\mathbb{I}_p + A) \geq \lambda_{\min}(\mathbb{I}_p) + \lambda_{\min}(A) > 1 - \epsilon.$$

It then follows that

$$\|\eta\Sigma^{v}\breve{\beta}^{\top}(\breve{\Sigma}^{r})^{-1}\breve{\beta}\gamma - \eta\gamma\| \leq \|\eta\|\,\|\gamma\|\,\lambda_{\max}(A)\lambda_{\min}^{-1}(\mathbb{I}_p + A) = O_p(s_n n^{-1}).$$

On the other hand, we have

$$\begin{aligned}\breve{\Sigma}^{z,u}(\breve{\Sigma}^{r})^{-1}\breve{\beta} &= \breve{\Sigma}^{z,u}(\breve{\Sigma}^{u})^{-1}\breve{\beta} - \breve{\Sigma}^{z,u}(\breve{\Sigma}^{u})^{-1}\breve{\beta}\left(\breve{\beta}^{\top}(\breve{\Sigma}^{u})^{-1}\breve{\beta} + (\Sigma^{v})^{-1}\right)^{-1}\breve{\beta}^{\top}(\breve{\Sigma}^{u})^{-1}\breve{\beta} \\ &= \breve{\Sigma}^{z,u}(\breve{\Sigma}^{u})^{-1}\breve{\beta}\left(\mathbb{I}_p + A\right)^{-1}A.\end{aligned}$$

Using a similar analysis as above, we have

$$\|\breve{\Sigma}^{z,u}(\breve{\Sigma}^{r})^{-1}\breve{\beta}\gamma\| \leq K\|\breve{\beta}^{z,u}\breve{\beta}\|_{\mathrm{MAX}}\|\gamma\|\,\lambda_{\max}(A)\lambda_{\min}^{-1}(\mathbb{I}_p + A) = O_p(s_n n^{-1}),$$

which concludes the proof. QED

B1.2. Proof of Theorem 1

By a simple conditioning argument, we can assume that $\hat{p} = p$ when developing the limiting distributions of the estimators (see Bai 2003). In the remainder of the proofs, we assume that $\hat{p} = p$. That said, the consistency of $\hat{p}$ with respect to p cannot guarantee the recovery of the true number of factors in any finite sample. We leave the discussion on this issue to online appendix I.1.

Let $\hat{\Lambda}$ be the $p \times p$ diagonal matrix of the p largest eigenvalues of $n^{-1}T^{-1}\bar{R}^{\top}\bar{R}$. We define a $p \times p$ matrix:

$$H = n^{-1}T^{-1}\hat{\Lambda}^{-1}\hat{V}\bar{V}^{\top}\beta^{\top}\beta. \tag{B2}$$

We have the following decomposition:

$$\begin{aligned}\hat{\gamma} - H\gamma &= \left(\hat{\beta}^{\top}\hat{\beta}\right)^{-1}\hat{\beta}^{\top}\left((\beta - \hat{\beta}H)\gamma + \beta\bar{v} + \bar{u}\right) \\ &= H\bar{v} + n\left(\hat{\beta}^{\top}\hat{\beta}\right)^{-1}n^{-1}\left(H^{-\top}\beta^{\top}\bar{u} + H^{-\top}\beta^{\top}(\beta - \hat{\beta}H)\gamma\right. \\ &\quad \left. + (\hat{\beta} - \beta H^{-1})^{\top}\bar{u} + H^{-\top}\beta^{\top}(\beta - \hat{\beta}H)\bar{v} + (\hat{\beta}^{\top} - H^{-\top}\beta^{\top})(\beta - \hat{\beta}H)(\gamma + \bar{v})\right).\end{aligned}$$

On the one hand, by lemma 4b and 4e, we have

$$\begin{aligned}
& n^{-1}\,\|\,\hat\beta^\top\hat\beta - H^{-\top}\beta^\top\beta H^{-1}\,\|_{\mathrm{MAX}} \\
\le & \|\,n^{-1}\big(\hat\beta^\top - H^{-\top}\beta^\top\big)\big(\hat\beta - \beta H^{-1}\big)\|_{\mathrm{MAX}} + n^{-1}\|\,H^{-\top}\beta^\top\big(\hat\beta - \beta H^{-1}\big)\|_{\mathrm{MAX}} \\
& + n^{-1}\|\,\big(\hat\beta - \beta H^{-1}\big)^\top\beta H^{-1}\,\|_{\mathrm{MAX}} \\
= & \; O_p(n^{-1} + T^{-1}).
\end{aligned} \tag{B3}$$

Therefore, by assumption A6 and lemma 4a, 4c, and 4d, we have

$$\hat\gamma - H\gamma = H\bar v + O_p(n^{-1} + T^{-1}). \tag{B4}$$

On the other hand, we note that

$$\hat\eta - \eta H^{-1} = \eta H^{-1}\big(H\bar V - \hat V\big)\hat V^\top(\hat V\hat V^\top)^{-1} + \bar Z\hat V^\top(\hat V\hat V^\top)^{-1},$$

and by lemma 5a and 5b, we have

$$\hat\eta - \eta H^{-1} = T^{-1}\bar Z\bar V^\top H^\top + O_p(n^{-1} + T^{-1}). \tag{B5}$$

Moreover, by lemma 5c, it follows that

$$\|\,\hat\eta - \eta H^{-1}\,\| = O_p(n^{-1} + T^{-1/2}). \tag{B6}$$

Combining (B4), (B5), and lemma 2, we obtain

$$\hat\gamma_g - \eta\gamma = \eta\bar v + T^{-1}\bar Z\bar V^\top(\Sigma^v)^{-1}\gamma + O_p(n^{-1} + T^{-1}).$$

Since

$$\begin{aligned}
\mathrm{vec}\big(T^{-1}\bar Z\bar V^\top(\Sigma^v)^{-1}\gamma\big) &= \big(\gamma^\top(\Sigma^v)^{-1}\otimes\mathbb{I}_d\big)\big(\mathrm{vec}(T^{-1}ZV^\top) + \mathrm{vec}(\bar z\bar v^\top)\big) \\
&= \big(\gamma^\top(\Sigma^v)^{-1}\otimes\mathbb{I}_d\big)\mathrm{vec}(T^{-1}ZV^\top) + O_p(T^{-1}),
\end{aligned}$$

it follows from assumption A11 that

$$\begin{aligned}
& T^{1/2}\begin{pmatrix} T^{-1}\bar Z\bar V^\top(\Sigma^v)^{-1}\gamma \\ \eta\bar v \end{pmatrix} \\
\xrightarrow{\mathcal{L}} & \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \big(\gamma^\top(\Sigma^v)^{-1}\otimes\mathbb{I}_d\big)\Pi_{11}\big((\Sigma^v)^{-1}\gamma\otimes\mathbb{I}_d\big) & \big(\gamma^\top(\Sigma^v)^{-1}\otimes\mathbb{I}_d\big)\Pi_{12}\eta^\top \\ \cdot & \eta\Pi_{22}\eta^\top \end{pmatrix}\right).
\end{aligned}$$

Therefore, by the delta method and imposing $T^{1/2}n^{-1}\to 0$, we obtain

$$T^{1/2}\big(T^{-1}\bar Z\bar V^\top(\Sigma^v)^{-1}\gamma + \eta\bar v\big)\xrightarrow{\mathcal{L}}\mathcal{N}(0,\Phi),$$

where Φ is given in the main text. This concludes the proof. QED

B1.3. Proof of Theorem 2

Following the proof of theorem 1, we can define

$$H^h = m^{-1}T^{-1}(\hat\Lambda^h)^{-1}\hat V^h\bar V^{h\top}\beta^\top\beta, \tag{B7}$$

where $\hat{\Lambda}^h$ is the $p \times p$ diagonal matrix of the p largest eigenvalues of $n^{-1}m^{-1}T^{-1}\bar{R}^{h\top}\bar{R}^h$. By assumptions A4, A5, A6, A7, and A9, we can establish the high-frequency version of (B4):

$$\hat{\gamma}^h - H^h\gamma\Delta = H^h\bar{v}^h + O_p(n^{-1} + T^{-1}),$$

where $\bar{v}^h = (mT)^{-1}\Sigma_{t=1}^{T}\Sigma_{k=1}^{m}v_{t-1+k\Delta}^h = (mT)^{-1}\Sigma_{t=1}^{T}v_t = m^{-1}\bar{v}$. This leads to the first result of the theorem.

On the other hand, we note that by assumption A2,

$$\hat{\eta}^l - \eta(H^h)^{-1} = \eta(H^h)^{-1}\left(H^h\bar{V} - \hat{V}^l\right)\hat{V}^{l\top}(\hat{V}^l\hat{V}^{l\top})^{-1} + \bar{Z}\hat{V}^{l\top}(\hat{V}^l\hat{V}^{l\top})^{-1}.$$

By a similar proof of lemma 1, we can establish

$$\|\hat{V}^h - H^h\bar{V}^h\|_{\mathrm{F}} = O_p(n^{-1/2}T^{1/2} + 1),$$

and hence by $\bar{V} = \bar{V}^h(\iota_m \otimes \mathbb{I}_T)$ and $\hat{V}^l = \hat{V}^h(\iota_m \otimes \mathbb{I}_T)$, we obtain

$$\|\hat{V}^l - H^h\bar{V}\|_{\mathrm{F}} \leq K\|\hat{V}^h - H^h\bar{V}^h\|\|\iota_m \otimes \mathbb{I}_T\| = O_p(n^{-1/2}T^{1/2} + 1). \tag{B8}$$

Additionally, using the following decomposition,

$$-(\hat{V}^l - H^h\bar{V})\bar{V}^\top = m^{-1}n^{-1}T^{-1}(\hat{\Lambda}^h)^{-1}\left(\hat{V}^h\bar{U}^{h\top}\beta\bar{V}\bar{V}^\top + \hat{V}^h\bar{V}^{h\top}\beta^\top\bar{U}\bar{V}^\top + \hat{V}^h\bar{U}^{h\top}\bar{U}\bar{V}^\top\right),$$

and by a similar proof of lemma 3a, we can show that

$$\|\left(H^h\bar{V} - \hat{V}^l\right)\bar{V}^\top\|_{\mathrm{MAX}} = O_p(n^{-1}T + 1). \tag{B9}$$

Combining (B8) and (B9), we have

$$\begin{aligned}\|(H^h\bar{V} - \hat{V}^l)\hat{V}^{l\top}\|_{\mathrm{MAX}} &\leq K\|\hat{V}^l - H^h\bar{V}\|_{\mathrm{F}}^2 + K\|\left(H^h\bar{V} - \hat{V}^l\right)\bar{V}^\top\|_{\mathrm{MAX}}\|H^h\|_{\mathrm{MAX}}\\ &= O_p(n^{-1}T + 1).\end{aligned}$$

Moreover, by (B8), we have

$$\begin{aligned}T^{-1}\|H^h\bar{V}\bar{V}^\top H^{h\top} - \hat{V}^l\hat{V}^{l\top}\| &\leq T^{-1}\|\hat{V}^l - H^h\bar{V}\|\|\hat{V}^l\| + T^{-1}\|H^h\|\|\bar{V}\|\|\hat{V}^l\\ &\quad - H^h\bar{V}\|\\ &= O_p(n^{-1/2} + T^{-1/2}),\end{aligned}$$

which, combined with Weyl's inequalities, implies that

$$\left|\lambda_{\min}(T^{-1}H^h\bar{V}\bar{V}^\top H^{h\top}) - \lambda_{\min}(T^{-1}\hat{V}^l\hat{V}^{l\top})\right| = o_p(1).$$

Therefore, we have $\lambda_{\min}^{-1}(T^{-1}\hat{V}^l\hat{V}^{l\top}) = O_p(1)$. Now using the above inequalities and a similar result of lemma 5b, we obtain

$$\begin{aligned}&\|\hat{\eta}^l - \eta(H^h)^{-1} - \bar{Z}\bar{V}^\top(\bar{V}\bar{V}^\top)^{-1}(H^h)^{-1}\|_{\mathrm{MAX}}\\ \leq &\|\eta\|\|(H^h)^{-1}\|\|(H^h\bar{V} - \hat{V}^l)\hat{V}^{l\top}\|_{\mathrm{MAX}}\|(\hat{V}^l\hat{V}^{l\top})^{-1}\|\\ &+\|\bar{Z}\bar{V}^\top H^{h\top}\|\|(H^h\bar{V}\bar{V}^\top H^{h\top})^{-1} - (\hat{V}^l\hat{V}^{l\top})^{-1}\|\\ &+\|\bar{Z}(H^h\bar{V} - \hat{V}^l)^\top\|\|(\hat{V}^l\hat{V}^{l\top})^{-1}\|\\ = &O_p(n^{-1} + T^{-1}).\end{aligned}$$

This establishes the second result of the theorem.

Finally, combining these two results, we have

$$m \times \hat{\eta}^l \hat{\gamma}^h - \eta\gamma = \eta\bar{v} + T^{-1}\bar{Z}\bar{V}^\top (T^{-1}\bar{V}\bar{V}^\top)^{-1}\gamma + O_p(n^{-1} + T^{-1}),$$

which concludes the proof. QED

B1.4. Proof of Theorem 3

Consider the singular value decomposition of $\bar{R}$:

$$\bar{R} = \sum_{i=1}^{\min(n,T-1)} \lambda_i(\bar{R}^\top\bar{R})^{1/2}\varsigma_i\xi_i^\top, \tag{B10}$$

where ς_i and ξ_i are $n \times 1$ and $T \times 1$ singular vectors, respectively. We use ς and ξ to denote the $n \times n$ and $T \times T$ matrices of singular vectors.

By simple calculations, we can rewrite the ridge estimator as

$$\hat{\gamma}_g^\mu = \sum_{i=1}^{\min(n,T-1)} \lambda_i(\bar{R}^\top\bar{R})^{1/2}(\lambda_i(\bar{R}^\top\bar{R}) + \mu)^{-1}\bar{G}\xi_i\varsigma_i^\top\bar{r}.$$

Notice that by (7), we have

$$\hat{V} = T^{1/2}(\xi_1 : \xi_2 : \ldots : \xi_p)^\top, \quad \hat{\beta} = n^{1/2}(\varsigma_1 : \varsigma_2 : \ldots\varsigma_p)\hat{\Lambda}^{1/2}.$$

So it follows that

$$\hat{\gamma}_g^\mu = \hat{\gamma}_g + \mathcal{R}_1 + \mathcal{R}_2,$$

where, writing E as a $p \times p$ diagonal matrix with $-\mu\lambda_i(\bar{R}^\top\bar{R})^{-1}(\lambda_i(\bar{R}^\top\bar{R}) + \mu)^{-1}$ on the (i, i)th entry,

$$\mathcal{R}_1 = \bar{G}\hat{V}^\top E\hat{\beta}^\top\bar{r}, \quad \mathcal{R}_2 = \sum_{i=p+1}^{\min(n,T-1)} \frac{\lambda_i(\bar{R}^\top\bar{R})^{1/2}}{\lambda_i(\bar{R}^\top\bar{R}) + \mu}\bar{G}\xi_i\varsigma_i^\top\bar{r}.$$

By the definitions of $\hat{\gamma}$ and $\hat{\eta}$, (B3), (B6), and (B4), we have

$$\| \bar{G}\hat{V}^\top \| = T\| \hat{\eta} \| = O_p(T), \quad \| \hat{\beta}^\top\bar{r} \| \leq \| \hat{\beta}^\top\hat{\beta} \| \| \hat{\gamma} \| = O_p(n).$$

Moreover, by (V.39), as $\mu n^{-1}T^{-1} \to 0$,

$$\| E \| \leq \mu\| \Lambda^{-1} \|^2 n^{-2}T^{-2} = O_p(\mu n^{-2}T^{-2}),$$

which leads to $\| \mathcal{R}_1 \| = O_p(\mu n^{-1}T^{-1}) = o_p(1)$.

On the other hand, writing $\xi_{-(1:p)} = (\xi_{p+1} : \xi_{p+2} : \ldots : \xi_{\min(n,T-1)})$ and $\varsigma_{-(1:p)} = (\varsigma_{p+1} : \varsigma_{p+2} : \ldots : \varsigma_{\min(n,T-1)})$, we have

$$\| \mathcal{R}_2 \| \leq \max_{p+1\leq i\leq\min(n,T-1)} \frac{\lambda_i(\bar{R}^\top\bar{R})^{1/2}}{\lambda_i(\bar{R}^\top\bar{R}) + \mu} \| \bar{G}\xi_{-(1:p)} \| \| \varsigma_{-(1:p)}^\top\bar{r} \| t.$$

Note also that by lemmas 1 and 3b, equation (V.27), and the fact that $\| \bar{Z} \| = O_p(T^{1/2})$, $\| \xi \| = 1$, and $\| \varsigma \| = 1$, we have

$$
\begin{aligned}
\| \bar{G}\xi_{-(1:p)} \| &\leq \| \eta \bar{V}\xi_{-(1:p)} \| + \| \bar{Z}\xi_{-(1:p)} \| \\
&\leq \| \eta \| \| \bar{V} - H^{-1}\hat{V} \| \| \xi \| + \| \bar{Z} \| \| \xi \| = O_p(T^{1/2}), \\
\| \varsigma^{\top}_{-(1:p)} \bar{r} \| &\leq \| \varsigma^{\top}_{-(1:p)} \beta(\gamma + \bar{v}) \| + \| \varsigma^{\top}_{-(1:p)} \bar{u} \| \\
&\leq \| \varsigma \| \| \hat{\beta}H - \beta \| \| \gamma + \bar{v} \| + \| \varsigma \| \| \bar{u} \| = O_p(1 + n^{1/2}T^{-1/2}).
\end{aligned}
$$

By (V.41), we have

$$
\begin{aligned}
\max_{p+1\leq i\leq \min(n,T-1)} \frac{\lambda_i(\bar{R}^{\top}\bar{R})^{1/2}}{\lambda_i(\bar{R}^{\top}\bar{R}) + \mu} &\leq \mu^{-1} \max_{p+1\leq i\leq \min(n,T-1)} \lambda_i(\bar{R}^{\top}\bar{R})^{1/2} \\
&= O_p(\mu^{-1}\sqrt{nT^{1/2} + n^{1/2}T}).
\end{aligned}
$$

It thereby follows that

$$
\| \mathcal{R}_2 \| = O_p(\mu^{-1}(nT^{1/4} + n^{1/4}T)) = o_p(1),
$$

which concludes the proof. QED

B1.5. Proof of Theorem 4

By (B5), we have

$$
\hat{\eta} - \eta H^{-1} = T^{-1}\bar{Z}\bar{V}^{\top}H^{\top} + O_p(n^{-1} + T^{-1}).
$$

Therefore, when $n^{-2}T = o(1)$, we can rewrite

$$
\widehat{W} = T\left(\eta H^{-1} + T^{-1}\bar{Z}\bar{V}^{\top}H^{\top}\right)\left((\hat{\Sigma}^{v})^{-1}\hat{\Pi}_{11}(\hat{\Sigma}^{v})^{-1}\right)^{-1}\left(\eta H^{-1} + T^{-1}\bar{Z}\bar{V}^{\top}H^{\top}\right)^{\top} + o_p(1).
$$

Under $\mathbb{H}_0:\eta = 0$, we have

$$
\widehat{W} = \left(T^{-1/2}\bar{Z}\bar{V}^{\top}\right)\left(H^{-1}(\hat{\Sigma}^{v})^{-1}\hat{\Pi}_{11}(\hat{\Sigma}^{v})^{-1}H^{-\top}\right)^{-1}\left(T^{-1/2}\bar{Z}\bar{V}^{\top}\right)^{\top} + o_p(1).
$$

By assumption A11, to show $\hat{W}\xrightarrow{\mathcal{L}}\chi^2_p$ under $\mathbb{H}_0$, it is sufficient to establish that

$$
H^{-1}(\hat{\Sigma}^{v})^{-1}\hat{\Pi}_{11}(\hat{\Sigma}^{v})^{-1}H^{-\top}\xrightarrow{p}\Pi_{11}.
$$

This holds by assumption as well as the fact that $\hat{\Sigma}^{v} = \mathbb{I}_d$, which leads to the first claim. The second claim is straightforward because H is invertible with probability approaching 1, $\| H \| = O_p(1)$ by lemma 2, and $T^{-1}\bar{Z}\bar{V}^{\top} = O_p(T^{-1/2})$ by assumption A8, which is dominated by ηH^{-1} under $\mathbb{H}_1$. Finally, under $\mathbb{H}_T:\eta = \eta_0 T^{-1/2}$, we have

$$
\hat{W} = \left(\eta_0 H^{-1} + T^{-1/2}\bar{Z}\bar{V}^{\top}H^{\top}\right)\left((\hat{\Sigma}^{v})^{-1}\hat{\Pi}_{11}(\hat{\Sigma}^{v})^{-1}\right)^{-1}\left(\eta_0 H^{-1} + T^{-1/2}\bar{Z}\bar{V}^{\top}H^{\top}\right)^{\top} + o_p(1).
$$

By lemma 2 and the same derivation as above, we have

$$
H^{\top}(\hat{\Sigma}^{v})^{-1}\hat{\Pi}_{11}(\hat{\Sigma}^{v})^{-1}H = H^{\top}H\Pi_{11}H^{\top}H + o_p(1) = (\Sigma^{v})^{-1}\Pi_{11}(\Sigma^{v})^{-1} + o_p(1),
$$

so it follows that

$$
\hat{W} = \left(\eta_0\Sigma^{v} + T^{-1/2}\bar{Z}\bar{V}^{\top}\right)\Pi_{11}^{-1}\left(\eta_0\Sigma^{v} + T^{-1/2}\bar{Z}\bar{V}^{\top}\right)^{\top} + o_p(1),
$$

which yields the desired result. QED

B2. *Empirical Results*

TABLE B1
THREE-PASS REGRESSION: EMPIRICAL RESULTS

	Average Return		Two-Pass Regression: No Controls		Two-Pass Regression: With R_m		Two-Pass Regression: With FF3		Mimicking Portfolio: With R_m		Mimicking Portfolio: With FF3		Three-Pass Regression			
Factors	γ (1)	SE (2)	γ (3)	SE (4)	γ (5)	SE (6)	γ (7)	SE (8)	γ (9)	SE (10)	γ (11)	SE (12)	γ (13)	SE (14)	R_g^2 (15)	p (g Weak) (16)
Market	.51**	.23	.56**	.23	.56**	.23	.51**	.22	.51**	.22	.51**	.22	.51**	.23	99.57	.00
SMB	.25	.15	.82**	.34	.07	.16	.10	.16	.08**	.04	.25	.16	.20	.16	97.24	.00
HML	.35**	.17	−.85**	.38	.30*	.18	.35**	.17	−.13**	.06	.35**	.15	.20	.15	83.03	.00
MOM	.69***	.24	−2.01**	.88	.20	.26	.71***	.24	−.05	.05	−.21*	.11	.49**	.23	89.82	.00
RMW	.38***	.13	.04	.16	−.00	.17	.27**	.13	−.07**	.03	−.09	.06	.22*	.11	71.48	.00
CMA	.32***	.11	−.59**	.24	.34**	.14	.42***	.12	−.10**	.05	.12	.08	.14	.10	59.03	.00
BAB	.94***	.22	−1.59*	.85	1.10***	.29	1.21***	.27	−.06	.05	.23**	.11	.57***	.15	47.43	.00
QMJ	.44***	.14	−.50**	.21	.01	.16	.25*	.14	−.15**	.07	−.29***	.09	.06	.13	84.29	.00
Liquidity			2.26**	.90	3.44***	1.09	.57	.68	.21*	.11	.32**	.14	.37**	.16	12.11	.00
Intermediary (He)			1.01**	.45	.19	.49	.43	.45	.57**	.25	.78***	.27	.60**	.31	69.05	.00
Intermediary (Adrian)			1.37***	.30	1.52***	.28	1.58***	.27	.10*	.06	.61***	.15	.72***	.16	51.99	.00
New York temperature			−319.01	255.73	125.89	152.76	−277.96**	124.08	−2.35	5.42	10.71	10.94	−.69	13.90	.76	.84
Global temperature			−6.65	4.85	−5.29	4.92	−3.33	2.07	−.01	.09	.11	.17	.05	.21	2.21	.09
El Niño			56.85***	17.42	19.23*	11.08	−15.34**	7.11	.39	.33	.94	.59	.41	.82	1.58	.43
Sunspots			−409.37	937.73	1,637.60***	467.40	882.89**	405.40	−19.30	19.49	−4.33	30.42	4.01	35.63	.86	.72
IP growth			−.36**	.14	−.27***	.07	−.14***	.05	−.00	.00	−.01	.01	−.01*	.00	2.25	.21
Macro PC 1			84.90***	24.76	87.26***	20.95	39.96***	13.57	1.22	.75	2.49*	1.43	3.26**	1.58	2.34	.29
Macro PC 2			9.35	15.93	9.28	16.34	23.91***	8.97	−.91	.59	−2.05**	1.03	−.88	1.27	4.05	.09
Macro PC 3			−5.94	14.30	−6.70	12.11	−31.24***	9.74	−.99	.64	−.61	1.21	−1.25	1.51	6.60	.01
Consumption growth			.26*	.16	−.03	.11	.07	.05	−.00	.00	−.00	.01	.00	.01	4.07	.07
Stockholder consumption			6.26***	2.14	2.48**	1.20	1.08*	.58	.05	.04	.03	.06	.17**	.08	2.50	.32

NOTE.—For each factor, the table reports the risk premia estimates using different methods, with the restriction that the zero-beta rate is equal to the observed treasury bill rate: the time series average return of the factor, available when the factor is tradable; three versions of the two-pass cross-sectional regression using no control factors in the model, using the market, and using the Fama-French three factors (FF3), respectively; two versions of the mimicking-portfolio estimator, projecting factors onto the market portfolio and the FF3 (given that we have more portfolios than observations, it is not feasible to use the mimicking-portfolio approach with all test portfolios); the three-pass estimator we propose in this paper, using $\check{p} = 7$ latent factors; the R^2 of the projection of g_t onto the latent factors; and the p-value of the test that factor g_t is weak. SE = standard error; He = He, Kelly, and Manela (2017); Adrian = Adrian, Etula, and Muir (2014).

* $p < .10$.
** $p < .05$.
*** $p < .01$.

References

Adrian, Tobias, Erkko Etula, and Tyler Muir. 2014. "Financial Intermediaries and the Cross-Section of Asset Returns." *J. Finance* 69 (6): 2557–96.

Ang, Andrew, Robert Hodrick, Yuhang Xing, and Xiaoyan Zhang. 2006. "The Cross-Section of Volatility and Expected Returns." *J. Finance* 61:259–99.

Asness, Clifford S., Andrea Frazzini, and Lasse Heje Pedersen. 2013. "Quality Minus Junk." Technical report.

Bai, Jushan. 2003. "Inferential Theory for Factor Models of Large Dimensions." *Econometrica* 71 (1): 135–71.

Bai, Jushan, and Serena Ng. 2002. "Determining the Number of Factors in Approximate Factor Models." *Econometrica* 70:191–221.

———. 2006. "Evaluating Latent and Observed Factors in Macroeconomics and Finance." *J. Econometrics* 131 (1): 507–37.

Bai, Jushan, and Guofu Zhou. 2015. "Fama-Macbeth Two-Pass Regressions: Improving Risk Premia Estimates." *Finance Res. Letters* 15:31–40.

Balduzzi, Pierluigi, and Cesare Robotti. 2008. "Mimicking Portfolios, Economic Risk Premia, and Tests of Multi-Beta Models." *J. Business and Econ. Statis.* 26:354–68.

Bernanke, Ben S., and Kenneth Kuttner. 2005. "What Explains the Stock Market's Reaction to Federal Reserve Policy?" *J. Finance* 60:1221–57.

Black, Fischer. 1972. "Capital Market Equilibrium with Restricted Borrowing." *J. Bus.* 45 (3): 444–55. http://www.jstor.org/stable/2351499.

Breeden, Douglas T., Michael R. Gibbons, and Robert H. Litzenberger. 1989. "Empirical Test of the Consumption-Oriented CAPM." *J. Finance* 44 (2): 231–62.

Bryzgalova, Svetlana. 2015. "Spurious Factors in Linear Asset Pricing Models." Tech. report, Stanford Univ.

Burmeister, Edwin, and Marjorie B. McElroy. 1988. "Joint Estimation of Factor Sensitivities and Risk Premia for the Arbitrage Pricing Theory." *J. Finance* 43 (3): 721–33.

Chamberlain, Gary, and Michael Rothschild. 1983. "Arbitrage, Factor Structure, and Mean-Variance Analysis on Large Asset Markets." *Econometrica* 51:1281–304.

Cochrane, John H. 2009. *Asset Pricing*. Princeton, NJ: Princeton Univ. Press.

Connor, Gregory, Matthias Hagmann, and Oliver Linton. 2012. "Efficient Semiparametric Estimation of the Fama-French Model and Extensions." *Econometrica* 80 (2): 713–54. https://doi.org/10.3982/ECTA7432.

Connor, Gregory, and Robert A. Korajczyk. 1986. "Performance Measurement with the Arbitrage Pricing Theory: A New Framework for Analysis." *J. Financial Econ.* 15 (3): 373–94.

———. 1988. "Risk and Return in an Equilibrium APT: Application of a New Test Methodology." *J. Financial Econ.* 21 (2): 255–89.

Elton, Edwin J., Martin J. Gruber, Stephen J. Brown, and William N. Goetzmann. 2014. *Modern Portfolio Theory and Investment Analysis.* Hoboken, NJ: Wiley.

Fama, Eugene F., and James D. Macbeth. 1973. "Risk, Return, and Equilibrium: Empirical Tests." *J.P.E.* 81 (3): 607–36.

Fan, Jianqing, Yuan Liao, and Jiawei Yao. 2015. "Power Enhancement in High-Dimensional Cross-Sectional Tests." *Econometrica* 83 (4): 1497–541.

Frazzini, Andrea, and Lasse Heje Pedersen. 2014. "Betting against Beta." *J. Financial Econ.* 111 (1): 1–25. https://doi.org/10.1016/j.jfineco.2013.10.005.

Friedman, Jerome, Trevor Hastie, and Robert Tibshirani. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.* 2nd ed. New York: Springer.

Gagliardini, Patrick, Elisa Ossola, and Olivier Scaillet. 2016. "Time-Varying Risk Premium in Large Cross-Sectional Equity Datasets." *Econometrica* 84 (3): 985–1046.

———. 2019. "A Diagnostic Criterion for Approximate Factor Structure." *J. Econometrics* 212 (2): 503–21.
Gospodinov, Nikolay, Raymond Kan, and Cesare Robotti. 2013. "Chi-Squared Tests for Evaluation and Comparison of Asset Pricing Models." *J. Econometrics* 173 (1): 108–25. https://doi.org/10.1016/j.jeconom.2012.11.002.
———. 2014. "Misspecification-Robust Inference in Linear Asset-Pricing Models with Irrelevant Risk Factors." *Rev. Financial Studies* 27 (7): 2139–70.
Harvey, Campbell R., Yan Liu, and Heqing Zhu. 2016. ". . . and the Cross-Section of Expected Returns." *Rev. Financial Studies* 29 (1): 5–68.
He, Zhiguo, Bryan Kelly, and Asaf Manela. 2017. "Intermediary Asset Pricing: New Evidence from Many Asset Classes." *J. Financial Econ.* 126 (1): 1–35.
Hoerl, Arthur E., and Robert W. Kennard. 2000. "Ridge Regression: Biased Estimation for Nonorthogonal Problems." *Technometrics* 42 (1): 80–86.
Huberman, Gur, Shmuel Kandel, and Robert F. Stambaugh. 1987. "Mimicking Portfolios and Exact Arbitrage Pricing." *J. Finance* 42 (1): 1–9.
Jagannathan, Ravi, and Zhenyu Wang. 1998. "An Asymptotic Theory for Estimating Beta-Pricing Models Using Cross-Sectional Regression." *J. Finance* 53 (4): 1285–309.
Jensen, Michael C., Fischer Black, and Myron S. Scholes. 1972. "The Capital Asset Pricing Model: Some Empirical Tests." In *Studies in the Theory of Capital Markets*, edited by Michael C. Jensen. Westport, CT: Praeger.
Kan, Raymond, and Chu Zhang. 1999a. "GMM Tests of Stochastic Discount Factor Models with Useless Factors." *J. Financial Econ.* 54 (1): 103–27.
———. 1999b. "Two-Pass Tests of Asset Pricing Models." *J. Finance* 54 (1): 203–35.
Kelly, Bryan, Seth Pruitt, and Yinan Su. 2019. "Characteristics Are Covariances: A Unified Model of Risk and Return." *J. Financial Econ.* 134 (3): 501–24.
Kleibergen, Frank. 2009. "Tests of Risk Premia in Linear Factor Models." *J. Econometrics* 149 (2): 149–73. https://doi.org/10.1016/j.jeconom.2009.01.013.
Kozak, Serhiy, Stefan Nagel, and Shrihari Santosh. 2018. "Interpreting Factor Models." *J. Finance* 73 (3): 1183–223.
Lettau, Martin, and Markus Pelger. 2018. "Factors That Fit the Time Series and Cross-Section of Stock Returns." Working Paper no. 24858, NBER, Cambridge, MA.
Lewellen, Jonathan, Stefan Nagel, and Jay Shanken. 2010. "A Skeptical Appraisal of Asset Pricing Tests." *J. Financial Econ.* 96 (2): 175–94. https://doi.org/10.1016/j.jfineco.2009.09.001.
Lucca, David O., and Emanuel Moench. 2015. "The Pre-FOMC Announcement Drift." *J. Finance* 70 (1): 329–71.
Ludvigson, Sydney C., and Serena Ng. 2010. "A Factor Analysis of Bond Risk Premia." In *Handbook of Empirical Economics and Finance*, vol. 1, edited by Aman Ulah and David E. A. Giles, 313–72. Boca Raton, FL: Chapman & Hall.
Malloy, Christopher J., Tobias J. Moskowitz, and Annette Vissing-Jorgensen. 2009. "Long-Run Stockholder Consumption Risk and Asset Returns." *J. Finance* 64 (6): 2427–79.
McLean, R. David, and Jeffrey Pontiff. 2016. "Does Academic Research Destroy Stock Return Predictability?" *J. Finance* 71 (1): 1–48.
Newey, Whitney K., and Kenneth D. West. 1987. "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix." *Econometrica* 55:703–8.
Novy-Marx, Robert. 2014. "Predicting Anomaly Performance with Politics, the Weather, Global Warming, Sunspots, and the Stars." *J. Financial Econ.* 112 (2): 137–46.
Onatski, Alexei. 2012. "Asymptotics of the Principal Components Estimator of Large Factor Models with Weakly Influential Factors." *J. Econometrics* 168:244–58.

Pástor, Luboš, and Robert F. Stambaugh. 2003. "Liquidity Risk and Expected Stock Returns." *J.P.E.* 111 (3): 642–85.

Roll, Richard, and Stephen A. Ross. 1980. "An Empirical Investigation of the Arbitrage Pricing Theory." *J. Finance* 35 (5): 1073–103.

Ross, Stephen A. 1976. "The Arbitrage Theory of Capital Asset Pricing." *J. Econ. Theory* 13:341–60.

Shanken, Jay. 1992. "On the Estimation of Beta Pricing Models." *Rev. Financial Studies* 5 (1): 1–33.

White, Halbert. 2000. *Asymptotic Theory for Econometricians.* Bingley: Emerald Group.