Asset Pricing with Omitted Factors*

Stefano Giglio[†]
Yale School of Management
NBER and CEPR

Dacheng Xiu[‡]
Booth School of Business
University of Chicago

First version: November 2016
This version: May 7, 2018

Abstract

Standard estimators of risk premia are biased if the estimation model omits some priced factors. We propose a three-pass method to estimate the risk premia of observable factors in a linear asset pricing model, which is valid even when not all factors in the model are specified and observed. We show that the risk premium of a factor can be identified regardless of the rotation of the other control factors, as long as they together span the true factor space. Motivated by this rotation invariance result, our approach uses principal components of test assets to recover the factor space and additional cross-sectional and time-series regressions to obtain the risk premium of each observed factor. Our estimator is also equivalent to the average excess return of an appropriately-regularized mimicking portfolio maximally correlated with the observed factor. Our methodology also accounts for potential measurement error in the observed factors and detects when such factors are spurious or even useless. The methodology exploits the blessings of dimensionality, and we therefore apply it to a large panel of equity portfolios to estimate risk premia for several workhorse factors. The estimates are robust to the choice of test portfolios within equities as well as across many asset classes.

Keywords: Three-Pass Estimator, Regularized Mimicking Portfolio, Latent Factors, Omitted Factors, Measurement Error, Fama-MacBeth Regression, Principal Component Regression

---

*This paper was previously circulated as "Inference on Risk Premia in the Presence of Omitted Factors." We 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, Irina Zviadadze, and seminar and conference participants at the University of Chicago, Princeton University, Stanford University, University of Cambridge, Imperial College Business School, Duke University, Tinbergen Institute, 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, the 2017 Annual meeting of the American Finance Association, the HEC-McGill Winter Finance Workshop, the 2017 Asian Meeting of the Econometric Society, the 2017 China Finance Review International Conference, the 10th Annual SoFiE conference, the BI-SHoF Conference, and the European Finance Association 2017 Annual Meeting.

[†]Address: 165 Whitney Avenue, New Haven, CT 06520, USA. E-mail address: `stefano.giglio@yale.edu`.

[‡]Address: 5807 S Woodlawn Avenue, Chicago, IL 60637, USA. E-mail address: `dacheng.xiu@chicagobooth.edu`.

1 Introduction

One of the central predictions of asset pricing models is that some risk factors – for example, intermediary capital or aggregate liquidity – should 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 computed as the average excess return of the factor. This is for example the case for the CAPM, 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 constructing a tradable portfolio that isolates that risk, holding all other risks constant. While different estimators have been proposed to estimate risk premia (most prominently, two-step cross-sectional regressions like Fama-MacBeth 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, and some of these omitted risks are correlated with the factor of interest. This is a fundamental concern when testing asset pricing theories, because theoretical models are usually very stylized and cannot possibly explicitly account for *all* sources of risk in the economy.[1] While the possibility of omitted variable bias is known in the literature (see, for example, 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 paper to paper. Papers using the two-pass cross-sectional regression approach typically add arbitrarily chosen factors or characteristics as controls, like the Fama-French three factors; papers using the mimicking-portfolio approach usually select a small set of portfolios (for example, 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 for the omitted variable bias in linear factor models. We introduce a new three-pass methodology that exploits the large dimensionality of available test assets and 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 premise 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 model with p factors and we wish to determine the risk premium of one of them (call it g_t). We show that the risk premium of g_t is invariant to how all other $p-1$ factors are rotated; the only requirement needed for correctly recovering the risk premium of g_t is that the model used in the estimation includes factors that, together with g_t, span the same space as the true factors in the model, no matter how they are rotated.[2] Naturally,

[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] The invariance result we derive is distinct from similar results the literature has explored in the past (e.g., Roll and Ross (1980), Huberman et al. (1987), Cochrane (2009)). This literature has explored the conditions under which rotations

some other components of the model (for example, risk exposures with respect to all factors including g_t) are *not* invariant to the rotation, so they cannot be recovered unless all factors in the model are specified. Needless to say, this rotation invariance result does *not* hold in a standard regression setting for coefficient on any specific regressor.[3]

This invariance result implies that knowing the *identities* of all true p 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. This is because the factor space can be recovered from the test asset themselves. 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 rotation invariance result with principal component analysis (PCA) 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. 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 that the risk premium estimate for g_t is consistent, regardless of the rotation of the true factors that occurs when extracting PCs.

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 principal-component-augmented two-pass cross-sectional regression. Rather than selecting the control factors 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 would lead to a bias, or onto the entire set of test assets, which would be inefficient or even infeasible when the dimension of the space 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 the test assets onto which the factor of interest is projected. Our theoretical analysis, however, sheds

of a factor model retain the pricing ability of the original model. It has not, however, explored the invariance properties for individual factors within the model. Indeed, our invariance result for the risk premium of an individual factor g_t builds on the existing results to show that a particular invariance property holds not only for the pricing ability of the entire model, but for the risk premia of each individual factor as well. This additional step is crucial when trying to understand the economic importance of a specific factor g_t in the presence of omitted factors.

[3] For example, a regression of Y on two variables X and Z will yield a different coefficient for X than a regression on X and $(X + Z)$, despite the fact that the two variables X and $(X + Z)$ span the same space as X and Z.

light on the convergence of the two as the dimension 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 apply our methodology to a large set of 202 equity portfolios, sorted by different characteristics. We estimate and test the significance of the risk premia of tradable and non-tradable factors from a number of different models. 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 CAPM. 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 correcting for it and for exposure to unobservable factors, they command a risk premium of essentially zero. We do, however, find some empirical support for the consumption growth of stockholders from Malloy et al. (2009), as well as for factors related to financial frictions (like the liquidity factor of Pástor and Stambaugh (2003)).

We also show that our risk premia estimates remain similar when using 100 non-equity portfolios (options, bonds, currencies, commodities) in addition to, or instead of, equity portfolios. We show that once the unobservable factors that drive these different asset classes are accounted for, the risk premia for many factors are quite consistent with those estimated just using the cross-section of equities. This result suggests that indeed several common factors are priced in a consistent way across various asset classes. This consistency is hard to detect without properly controlling for the unobservable factors to which various groups of assets are exposed.

Our paper derives several important econometric properties of the 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. Our asymptotic theory allows for heteroscedasticity 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 any of the observable factors g_t is spurious or even useless (that is, 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.

1.1 Literature review

This paper sits at the confluence of several strands of literature, combining empirical asset pricing with high-dimensional factor analysis.

Using two-pass regressions to estimate asset pricing models dates back to Black et al. (1972) and

Fama and Macbeth (1973). Over the years, the econometric methodologies have been refined and extended; see for example Ferson and Harvey (1991), Shanken (1992), Jagannathan and Wang (1998), Welch (2008), and Lewellen et al. (2010). These papers, along with the majority of the literature, rely on large T and fixed n asymptotic analysis for statistical inference and only deal with models where all factors are specified and observable. Bai and Zhou (2015) and Gagliardini et al. (2016) extend the inferential theory to the large n and large T setting, which delivers better small-sample performance when n is large relative to T. Connor et al. (2012) use semiparametric methods to model time variation in the risk exposures as function of observable characteristics, again allowing for large n and T. Our asymptotic theory relies on a similar large n and large T analysis, yet we do not impose a fully specified model.

Our paper relates to the literature that has pointed out pitfalls in estimating and testing linear factor models. For instance, ignoring model misspecification and identification-failure leads to an overly positive assessment of the pricing performance of spurious (Kleibergen (2009)) or even useless factors (Kan and Zhang (1999a,b); Jagannathan and Wang (1998)), and biased risk premia estimates of true factors in the model. It is therefore more reliable to use inference methods that are robust to model misspecification (Shanken and Zhou (2007); Kan and Robotti (2008); Kleibergen (2009); Kan and Robotti (2009); Kan et al. (2013); Gospodinov et al. (2013); Kleibergen and Zhan (2014); Gospodinov et al. (2016); Bryzgalova (2015); Burnside (2016)). We study and correct the biases due to omitted variables and measurement error. Gagliardini et al. (2017) propose a diagnostic criterion to detect potentially omitted factors from the residuals of an observable factor model. Hou and Kimmel (2006) argue that in the case of omitted factors, the definition of risk premia can be ambiguous. Relying on a large number of test assets, our approach can provide consistent estimates of the risk premia without ambiguity, and detect spurious and useless factors. Lewellen et al. (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 the missing factors in the two-pass cross-sectional regressions.

Our paper is also related to the literature that advocates the use of mimicking portfolios in factor pricing models. Huberman et al. (1987) show that mimicking portfolios can be used in place of non-tradable factors in asset pricing models and provide three choices of mimicking portfolios, one of which is the maximally-correlated portfolio. Balduzzi and Robotti (2008) and more recently, Kleibergen and Zhan (2018), estimate and test asset pricing models using mimicking portfolios as the factors. In the empirical literature, the use of mimicking portfolios dates back at least to Breeden et al. (1989), who use this approach to test the CCAPM model. Lamont (2001) also advocates the use of mimicking portfolios to analyze other economic factors. Ang et al. (2006) and Adrian et al. (2014) construct aggregate volatility and intermediary leverage factor-mimicking portfolios, respectively. One particular advantage of mimicking portfolios is that such portfolios are available at higher frequencies or over longer time spans than the original economic risk factors.

The literature on factor models has expanded dramatically since the seminal paper by Ross (1976) on arbitrage pricing theory (APT). Chamberlain and Rothschild (1983) extend this framework to approximate factor models. Connor and Korajczyk (1986, 1988) and Lehmann and Modest (1988) tackle estimation and testing in the APT setting by extracting principal components of returns, without having to specify the factors explicitly. More recently, Kozak et al. (2017) show how few principal components capture a large fraction of the cross-section of expected returns, which we will also show in our data. Overall, one of the downsides of latent factor models is precisely the difficulty in interpreting the estimated risk premia. In our paper, we start from the same statistical intuition that we can use PCA to extract latent factors, but exploit it to estimate (interpretable) risk premia for the observable factors. Bai and Ng (2002) and Bai (2003) introduce asymptotic inferential theory on factor structures. In addition, Bai and Ng (2006) propose a test for whether a set of observable factors spans the space of factors present in a large panel of returns. In contrast, our paper exploits statistically the spanning of the latent factors in time series, and their ability to explain the cross-sectional variation of expected returns.

Section 2 discusses biases due to omitted variables and measurement error in the standard risk premia estimators. Section 3 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 4 provides the asymptotic theory on inference with our estimator, followed by an empirical study in Section 5. The appendix provides technical details and Monte Carlo simulations.

Throughout the paper, we use $(A : B)$ to denote the concatenation (by columns) of two matrices A and B. e_i is a vector with 1 in the ith entry and 0 elsewhere, whose dimension depends on the context. ι_k denotes a k-dimensional vector with all entries being 1. For any time series of vectors $\{a_t\}_{t=1}^{T}$, we denote $\bar{a} = \frac{1}{T}\sum_{t=1}^{T} a_t$. 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 $\mathbb{M}_A = \mathbb{I} - \mathbb{P}_A$.

2 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.[4] 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.

[4] As discussed in the introduction, the focus of this paper is on *nontradable* factors, for which the 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 is the risk premium – in which case, the methods we discuss here are still valid as an alternative estimator of the risk premium.

In what follows, we estimate the risk premium of a proxy for the first factor v_{1t}, denoted as g_t; its risk premium is therefore γ_1 in this simple setting. We begin with a review of the two estimators, then consider two potential sources of bias that can affect each estimator.

2.1 A Review

Two-pass regressions estimate the factor risk premia as follows. First, time series regressions of each test asset's excess return r_t onto the factors v_t 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 estimators 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.

2.2 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 potentially 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 estimated β, only part of it ($\widehat{\beta}_1$) 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.

Whereas in the two-pass approach the bias stems from the omission of some factors (v_{2t} in our example) in the regressions of returns onto the factors, in the mimicking-portfolio approach a related omitted-variable bias can arise from the omission of assets onto which g_t is projected.

To see the potential for omitted variable bias, it is useful to write down explicitly the formula for the mimicking-portfolio estimator. Consider the projection of g_t onto the excess returns of a chosen set of test assets, $\check{r}_t$.[5] This projection yields coefficients $w^g = \text{Var}(\check{r}_t)^{-1}\text{Cov}(\check{r}_t, g_t)$; these are the weights of the mimicking portfolio for g_t, whose excess return is then $r_t^g = (w^g)^\intercal \check{r}_t$. Therefore, we can write the expected excess return of the mimicking portfolio as: $\gamma_g^{\text{MP}} = (w^g)^\intercal \text{E}(\check{r}_t)$. Since the test assets $\check{r}_t$ follow the same pricing model, we can write $\check{r}_t = \breve{\beta}\gamma + \breve{\beta} v_t + \check{u}_t$. Substituting, we can write the formula for the mimicking-portfolio estimator of the risk premium of the first factor as: $\gamma_g^{\text{MP}} = \left\{(\breve{\beta}\Sigma^v\breve{\beta}^\intercal + \breve{\Sigma}^u)^{-1}(\breve{\beta}\Sigma^v e_1)\right\}^\intercal \breve{\beta}\gamma$, where e_1 is a column vector $(1 : 0)^\intercal$, 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.

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

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 $\gamma_g^{\mathrm{MP}} = \gamma_1$. There is one case in which the estimator will clearly be consistent: 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 $\gamma_g^{\mathrm{MP}} = \gamma_1$.

When these conditions are not satisfied, however, the mimicking-portfolio estimator will in general be biased, in particular 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 et al. (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.

2.3 Measurement Error Bias

Suppose now that the factor of interest may be observed with error; the econometrician can only observe $g_t = v_{1t} + z_t$, 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 these estimators. Consider first the two-pass regression approach. Independently of whether v_{2t} is observed or not, 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: $\gamma_g^{\mathrm{MP}} = \left\{(\breve{\beta}\Sigma^v\breve{\beta}^\intercal + \breve{\Sigma}^u)^{-1}(\breve{\beta}\Sigma^v e_1 + \breve{\Sigma}^{z,u})\right\}^\intercal \breve{\beta}\gamma \neq \gamma_1$, where $\breve{\Sigma}^{z,u} = \mathrm{Cov}(z_t, \breve{u}_t)$. Thus, measurement error z_t introduces a bias in the mimicking-portfolio estimator, unless idiosyncratic errors $\breve{u}_t$ in the spanning assets are fully diversified away, or are completely uncorrelated with idiosyncratic errors.

3 Methodology

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

3.1 Model Setup

We begin by introducing our baseline specification. Suppose v_t is a $p \times 1$ vector of factor innovations (i.e., mean-zero factors), and let r_t denote an $n \times 1$ vector of asset excess returns. The pricing model satisfies:

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

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

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. We discuss such extensions in greater detail in Section 5.6.4. Second, we impose weak assumptions on the structure of the errors. Most of our results hold for non-stationary processes with heteroscedasticity 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 T-bill rate. Later, we will examine a more general version of the model which allows the zero-beta rate to be different and to be estimated.

The objective of this paper is to estimate the risk premia of specific factors g_t without necessarily observing all true factors v_t. In the simple two-factor model of the previous section, 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.

More specifically, call g_t a set of d observable (tradable or nontradable) factors whose risk premia we want to estimate. g_t is related to the factors v_t as follows:

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

where η, the loading of g on v, is a $d \times p$ matrix, ξ is a $d \times 1$ constant, and z_t is a $d \times 1$ measurement-error vector. The risk premium of a factor g_t 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 $\gamma_g = \eta\gamma$.

We also allow for measurement error in g_t (captured by z_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 in our specification. For tradable factors, z_t can capture exposure to unpriced risks, or idiosyncratic risk that is not fully diversified. For this reason, we allow z_t to be correlated with the idiosyncratic risk u_t.

3.2 Rotation Invariance of Risk Premia

We now derive a simple rotation-invariance result that holds, generally, in linear asset pricing models, which is the key to the identification of $\eta\gamma$ when not all factors are observed.

Recall that the risk premium of g_t in our setup (equations (1) and (2)) is given by $\eta\gamma$. We now consider a rotation of the model where the entire model is expressed as a function of rotated factors $\hat{v}_t = Hv_t$ instead of the original factors v_t, with any full-rank $p \times p$ matrix H. To do so, rewrite the model as:

$$r_t = \beta H^{-1} H\gamma + \beta H^{-1} H v_t + u_t,$$
$$g_t = \xi + \eta H^{-1} H v_t + z_t.$$

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

$$r_t = \hat{\beta}\hat{\gamma} + \hat{\beta}\hat{v}_t + u_t, \tag{3}$$
$$g_t = \xi + \hat{\eta}\hat{v}_t + z_t. \tag{4}$$

We say that a parameter or quantity in the model is rotation-invariant if it is identical in the original model (equations (1) and (2)) or in any rotated model (equations (3) and (4)), for any invertible H. Some parameters of this model are clearly *not* rotation invariant. For example, risk exposures of assets to the factors are different between the two representations:

$$\hat{\beta} = \beta H^{-1} \neq \beta.$$

In other words, if one estimates risk exposures $\hat{\beta}$ from a rotation of the original model, one cannot recover the original β without knowing the transformation H.

The main result that we will use in this paper is that the *risk premium* of g_t, $\eta\gamma$, is rotation invariant. While neither η nor γ by itself is rotation invariant (because $\hat{\eta} \equiv \eta H^{-1} \neq \eta$ and $\hat{\gamma} \equiv H\gamma \neq \gamma$), their product is indeed independent of the rotation H:

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

This result guarantees that any consistent estimator of $\hat{\eta}\hat{\gamma}$, no matter how the underlying factors are rotated, will consistently estimate the risk premium γ_g.

3.3 The Three-Pass Estimator

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 (equations (1) and (2)) can then be written in matrix terms as

$$R = \beta\gamma + \beta V + U.$$

Writing ($\bar{R}$, $\bar{V}$, $\bar{G}$, $\bar{U}$, $\bar{Z}$) as the matrices of the demeaned variables, 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, ξ, does not have a meaningful interpretation or relevance for the purpose of estimating the risk premium, we only need the demeaned version of equation (2):

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

Our estimator only makes use of excess returns R and the factors of interest G. We assume that the true factors V are latent. 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 $n, T \to \infty$. While H itself cannot be recovered from the data, the invariance result guarantees that we can still consistently estimate γ_g.

The three-pass estimator. Given observable returns R and the factors of interest G, our estimator $\widehat{\gamma}_g$ of $\gamma_g \equiv \eta\gamma$ proceeds as follows:

(i) **PCA step**. Extract the PCs of returns, by conducting the PCA of the matrix $n^{-1}T^{-1}\bar{R}^\intercal \bar{R}$. Define the estimator for the factors and their loadings as:

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

where $\xi_1, \xi_2, \ldots, \xi_{\widehat{p}}$ are the eigenvectors corresponding to the largest $\widehat{p}$ eigenvalues of the matrix $n^{-1}T^{-1}\bar{R}^\intercal \bar{R}$. $\hat{p}$ is an estimator of the number of factors; we propose using the following estimator:

$$\widehat{p} = \arg\min_{1 \le j \le p_{\max}} \left(n^{-1}T^{-1}\lambda_j(\bar{R}^\intercal \bar{R}) + j \times \phi(n,T)\right) - 1,$$

where $p_{\max}$ is some upper bound of p and $\phi(n,T)$ is some penalty function.

(ii) **Cross-sectional regression step**. Run a cross-sectional ordinary least square (OLS) regression of average returns, $\bar{r}$, onto the estimated factor loadings $\widehat{\beta}$ to obtain the risk premia of the estimated latent factors:

$$\widehat{\gamma} = (\widehat{\beta}^\intercal \widehat{\beta})^{-1}\widehat{\beta}^\intercal \bar{r}.$$

(iii) **Time-series regression step**. Run a time-series regression of g_t onto the factors extracted from the PCA in step (i), and then obtain the estimator $\widehat{\eta}$ and the fitted value of the observable factor after removing measurement error, $\widehat{G}$:

$$\widehat{\eta} = \bar{G}\widehat{V}^\intercal(\widehat{V}\widehat{V}^\intercal)^{-1}, \quad \text{and} \quad \widehat{G} = \widehat{\eta}\widehat{V}.$$

The estimator of the risk premium for the observable factor g_t is then obtained by combining the estimates of the second and third steps:

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

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

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

The estimator can be easily extended to the case in which the zero-beta rate is allowed to be different from the observed risk-free rate. In that case, returns can be written as: $r_t = \gamma_0\iota_n + \beta\gamma + \beta v_t + u_t$, and step (ii) of the procedure can be modified to yield an estimate for γ_g together with the zero-beta rate γ_0. In compact forms, the estimators are given by

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

where $\mathbb{M}_{\widehat{\beta}} = \mathbb{I} - \widehat{\beta}(\widehat{\beta}^\intercal\widehat{\beta})^{-1}\widehat{\beta}^\intercal$ and $\mathbb{M}_{\iota_n} = \mathbb{I} - \iota_n(\iota_n^\intercal\iota_n)^{-1}\iota_n^\intercal$.

The first step of the three-pass procedure recovers the factors (up to a rotation: $\widehat{V} \to H\bar{V}$ for some unobserved invertible matrix H), by extracting the PCs of returns and selecting the first $\widehat{p}$ of them. The number of factors $\widehat{p}$ to use is itself estimated, and we will show in the next section that our proposed estimator is consistent for the true number of factors p. This estimator is based on a penalty function, similar to the one Bai and Ng (2002) propose. However, it takes on a different form, because we will work under weaker assumptions than Bai and Ng (2002). $p_{\max}$ is an economically reasonable upper bound for the number of factors, imposed only to improve the finite sample performance. It is not needed in asymptotic analysis.[6]

Note that we propose to extract PCs from the $T \times T$ matrix $n^{-1}T^{-1}\bar{R}^\intercal\bar{R}$, and normalize the estimated factors such that $\widehat{V}\widehat{V}^\intercal = \mathbb{I}_{\widehat{p}}$. Alternatively, one could consider extracting PCs from the $n \times n$ matrix $n^{-1}T^{-1}\bar{R}\bar{R}^\intercal$, and normalizing $\widehat{\beta}^\intercal\widehat{\beta} = \mathbb{I}_{\widehat{p}}$. The two ways of normalization yield identical risk premia estimates, though the former estimator is easier to analyze when also estimating the zero-beta rate.

Once the PCs are extracted in the first stage, the second stage estimates their risk premia. Given that the PCs capture a rotation of the true factors, their risk premia correspond to a rotation of the true risk premia ($H\gamma$). 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 our setting the covariance matrix of u_t only matters at the order of $O_p(n^{-1} + T^{-1})$, whereas the leading term of $\widehat{\gamma}_g$ is $O_p(n^{-1/2} + T^{-1/2})$.

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 the economic theory predicts. This step also removes the effect of measurement error, which the standard approaches cannot accomplish. Even though g_t can be multi-dimensional, the estimation for each observable factor is

[6] Other estimators for the number of factors could be applied instead, including but not limited to those proposed by Onatski (2010) and Ahn and Horenstein (2013). However, these alternative estimators require stronger assumptions than ours.

separate. Estimating the risk premium for one factor does not affect the estimation for the others at all, another important property of our estimator.

To sum up, our estimator uses PCA to recover factors v_t up to a rotation H; it estimates their risk premia $(H\gamma)$ in the second step; it estimates the loading of g_t onto the rotated factors (ηH^{-1}) in the third step; finally, it combines the two to produce a consistent estimator for $\eta H^{-1} H\gamma = \gamma_g$, where this last equality is a consequence of the invariance result we derived above.

3.4 Alternative Interpretations

In Section 2 we have discussed how, in general, the two-pass regressions and the mimicking-portfolio estimators tend to give different estimates, even when the model is correctly specified. As it turns out, our three-pass estimator can be interpreted *both* as an extension of the two-pass regressions and as an extension of the mimicking-portfolio estimator. In this section, we discuss how our estimator brings together these two different approaches.

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 equation (2), and recall that we are assuming no measurement error for now). 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.

Given this result, we can interpret our three-pass estimator as a factor-augmented cross-sectional regression estimator. Step (i) uses PCA to extract a rotation of the original factors v_t. Step (iii) removes measurement error from g_t and identifies $\widehat{\eta}$: 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 g_t together with $p-1$ PCs as controls. Risk premia for this model are estimated via cross-sectional regressions (step (ii)), 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 (ii) and (iii) of our procedure, it gives numerically identical results.

Mimicking-portfolio interpretation. Our three-pass procedure can be also interpreted as a mimicking-portfolio estimator, in which the principal components themselves are the portfolios on which g_t is projected. This represents an optimal 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: so $\breve{r}_t = w^\intercal r_t$. In Section 2.2 we showed that in general the mimicking-portfolio estimator is not consistent, unless the returns of the portfolios on which g_t is projected ($\breve{r}_t$) satisfy particular requirements. In turn, that means that w needs to be chosen carefully so that the mimicking-portfolio estimator that uses these $\breve{p}$ portfolios,

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

actually converges to $\gamma_g = \eta\gamma$.

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 of Appendix B.1 that the bias of the mimicking-portfolio estimator (equation 10) 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.

Intuitively, this choice of w is guaranteed to achieve asymptotically the two criteria highlighted in Section 2.2: 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 affect 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: $\widehat{\beta}(\widehat{\beta}^\intercal\widehat{\beta})^{-1}$, that is, a full-rank rotation of the estimated β. The resulting portfolio returns are exactly the PCs in step (i) of our procedure, i.e., $\widehat{V} = (\widehat{\beta}^\intercal\widehat{\beta})^{-1}\widehat{\beta}^\intercal\bar{R}$. In addition, these portfolios are (when n is large) free of idiosyncratic error. Step (iii) projects g_t onto these portfolios, thus identifying the weights of the mimicking portfolio, $\widehat{\eta}$. Our estimator of the risk premium of g_t is then obtained by multiplying the portfolio weights $\widehat{\eta}$ by the risk premia of these portfolios ($\widehat{\gamma}$) obtained in Step (ii).

Interestingly, Proposition 1 also suggests that 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 a subset $\breve{r}_t$, 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 principal component 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.

4 Asymptotic Theory

In this section, we present the large sample distribution of our estimator as $n, T \to \infty$. Our results hold under the same or even weaker assumptions compared to 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.

4.1 Determining the Number of Factors

Theorem 1. *Under Assumptions A.4 – A.7, and suppose that as* n, $T \to \infty$, $\phi(n,T) \to 0$, *and* $\phi(n,T)/(n^{-1/2} + T^{-1/2}) \to \infty$, *we have* $\widehat{p} \xrightarrow{p} p$.

By a simple conditioning argument, we can assume that $\widehat{p} = p$ when developing the limiting distributions of the estimators, see Bai (2003). In the remainder of the section, we assume $\widehat{p} = p$. Even though consistency cannot guarantee the recovery of the true number of factors in any finite sample, our derivation in Section 4.5 shows that as long as $p \le \widehat{p} \le K$ for some finite K, we can estimate the parameters Γ consistently.

A notable assumption behind is the so-called pervasive condition for a factor model, i.e., Assumption A.6. It requires the factors to be sufficiently strong that most assets have non-negligible 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. We defer a more detailed discussion of this to Section 5.6.3.

4.2 Limiting Distribution of the Risk Premia Estimator

We now present the main theorem of the paper – the asymptotic distribution of the estimator $\widehat{\gamma}_g$, which naturally needs more assumptions reported in detail in Appendix A.

Theorem 2. *Under Assumptions A.4 – A.11, and suppose* $\widehat{p} \xrightarrow{p} p$, *then as* n, $T \to \infty$, *we have*

$$\widehat{\gamma} - H\gamma = H\bar{v} + O_p(n^{-1} + T^{-1}), \quad \widehat{\eta} - \eta H^{-1} = T^{-1}\bar{Z}\bar{V}^{\intercal}H^{\intercal} + O_p(n^{-1} + T^{-1}),$$

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

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

where Φ is given by

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

Remarkably, the asymptotic covariance matrix 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 orders. Therefore, for the inference on the risk premium of g_t, 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.[7]

4.3 Allowing for Pricing Errors and Zero-beta Rate

Now we extend the above results to a more general setting, in which the zero-beta rate is unrestricted, and in which mispricing is allowed for in the model. This case represents the most general setting in which our estimator is consistent.

Suppose the cross-section of asset returns r_t follows

$$r_t = \alpha + \iota_n\gamma_0 + \beta\gamma + \beta v_t + u_t, \tag{12}$$

where the cross-sectional pricing error α is i.i.d., independent of β, u and v, with mean 0, standard deviation $\sigma^\alpha > 0$, and a finite fourth moment.

There is a large body of literature on testing the APT by exploring the deviation of α from 0, including Connor and Korajczyk (1988), Gibbons et al. (1989), MacKinlay and Richardson (1991), and more recently, Pesaran and Yamagata (2012) and Fan et al. (2015). This is, however, not the focus of this paper. Empirically, the pricing errors may exist for many reasons such as limits to arbitrage, transaction costs, market inefficiency, and so on, so that it is important to allow for a misspecified linear factor model. Gospodinov et al. (2014) and Kan et al. (2013) also consider this type of model misspecification in their two-pass cross-sectional regression setting.

In this case, we employ the alternative estimator (9).

Theorem 3. *Under Assumptions A.2, A.4 – A.14, and suppose $\widehat{p} \xrightarrow{p} p$, then as n, $T \to \infty$, we have*

$$n^{1/2}\left(\widehat{\gamma}_0 - \gamma_0\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \left(1 - \beta_0^\intercal(\Sigma^\beta)^{-1}\beta_0\right)^{-1}(\sigma^\alpha)^2\right),$$

$$\left(T^{-1}\Phi + n^{-1}\Upsilon\right)^{-1/2}\left(\widetilde{\gamma}_g - \eta\gamma\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \mathbb{I}_d\right),$$

[7] Indeed, we can show that our estimator is asymptotically equivalent to the infeasible GLS.

where the asymptotic covariance matrices Φ *is given by* (11)*, and* Υ *is defined by*

$$\Upsilon = (\sigma^\alpha)^2 \eta \left(\Sigma^\beta - \beta_0\beta_0^\intercal\right)^{-1} \eta^\intercal.$$

Unlike the CLT in Theorem 2, the result of Theorem 3 does not impose any restrictions on the relative rates of n and T. Also, the above analysis assumes that the factor loading β is uncorrelated with the pricing error α. In fact, if they are correlated, then our estimator would instead converge to the "pseudo-true" parameter $\eta\left(\gamma + \text{plim}_{n\to\infty}(\beta^\intercal \mathbb{M}_{\iota_n}\beta)^{-1}\beta^\intercal\alpha\right)$, which is difficult to interpret, see, e.g., Kan et al. (2013).

4.4 Goodness-of-Fit Measures

To measure the goodness-of-fit in the cross-section of expected returns, we define the usual (population) cross-sectional R^2 for the latent factors in (12):

$$\mathrm{R}_v^2 = \frac{\gamma^\intercal(\Sigma^\beta - \beta_0\beta_0^\intercal)\gamma}{(\sigma^\alpha)^2 + \gamma^\intercal(\Sigma^\beta - \beta_0\beta_0^\intercal)\gamma}.$$

To measure the signal-to-noise ratio of each observable factor, we define the time-series R^2 for each observable factor g $(1\times T)$, for the time-series regression of g_t on the latent factors:

$$\mathrm{R}_g^2 = \frac{\eta\Sigma^v\eta^\intercal}{\eta\Sigma^v\eta^\intercal + \Sigma^z}, \quad \text{where } \eta \text{ is a } 1\times p \text{ vector.}$$

To calculate these measures in sample, we use

$$\widehat{\mathrm{R}}_v^2 = \frac{\bar{r}^\intercal\mathbb{M}_{\iota_n}\widehat{\beta}(\widehat{\beta}^\intercal\mathbb{M}_{\iota_n}\widehat{\beta})^{-1}\widehat{\beta}^\intercal\mathbb{M}_{\iota_n}\bar{r}}{\bar{r}^\intercal\mathbb{M}_{\iota_n}\bar{r}} \quad \text{and} \quad \widehat{\mathrm{R}}_g^2 = \frac{\widehat{\eta}\widehat{V}\widehat{V}^\intercal\widehat{\eta}^\intercal}{\bar{G}\bar{G}^\intercal}, \quad \text{respectively,}$$

where $\bar{G} = g - \bar{g}$ is a $1\times T$ vector. We can consistently estimate the cross-sectional R^2 for the latent factors as well as the time-series R^2 for each observable factor.

Theorem 4. *Under Assumptions A.2, and A.4 – A.14, and suppose* $\widehat{p} \xrightarrow{p} p$*, then as* $n, T\to\infty$*, we have*

$$\widehat{\mathrm{R}}_v^2 \xrightarrow{p} \mathrm{R}_v^2 \quad \textit{and} \quad \widehat{\mathrm{R}}_g^2 \xrightarrow{p} \mathrm{R}_g^2.$$

4.5 Robustness to the Choice of p

Although $\widehat{p}$ is a consistent estimator of p, it is possible that in a finite sample $\widehat{p}\neq p$. In fact, without a consistent estimator of $\widehat{p}$, as long as our choice, denoted by $\breve{p}$, is greater than or equal to p, the estimators based on $\breve{p}$, denoted by $\breve{\gamma}_0$ and $\breve{\gamma}_g$, are consistent, as the next theorem shows.

Theorem 5. *Suppose Assumptions A.2, and A.4 – A.14 hold. In addition, assume that* u_t *is i.i.d.* $\mathcal{N}\left(0,(\sigma^u)^2\mathbb{I}_n\right)$*, independent of* z_t *and* v_t*. If* $\breve{p}\geq p$ *and* $\breve{p}\leq K$ *as* $n/T\to c\in(0,\infty)$*, then* $\breve{\gamma}_0$ *and* $\breve{\gamma}_g$ *are* *consistent estimators of γ_0 and $\eta\gamma$, and it holds that*

$$\breve{\gamma}_0 - \widehat{\gamma}_0 = O_p(n^{-1/2}), \quad \breve{\gamma}_g - \widetilde{\gamma}_g = O_p(n^{-1/2}).$$

The above theorem establishes the desired robustness to the inclusion of "noise" factors.[8] While we cannot establish its asymptotic distribution, simulation exercises suggest that the differences between the asymptotic variances of $\breve{\gamma}_g$ and $\widetilde{\gamma}_g$ are tiny. This is also the case for our empirical study.

4.6 Asymptotic Variances Estimation

We develop consistent estimators of the asymptotic covariances in Theorem 8. The case for estimator (8) in Theorem 2 is simpler. We can estimate them for inference on risk premia using:

$$\widehat{\Phi} = \left(\widetilde{\gamma}^\intercal(\widehat{\Sigma}^v)^{-1} \otimes \mathbb{I}_d\right)\widehat{\Pi}_{11}\left((\widehat{\Sigma}^v)^{-1}\widetilde{\gamma} \otimes \mathbb{I}_d\right) + \left(\widetilde{\gamma}^\intercal(\widehat{\Sigma}^v)^{-1} \otimes \mathbb{I}_d\right)\widehat{\Pi}_{12}\widehat{\eta}^\intercal + \widehat{\eta}\widehat{\Pi}_{21}\left((\widehat{\Sigma}^v)^{-1}\widetilde{\gamma} \otimes \mathbb{I}_d\right) + \widehat{\eta}\widehat{\Pi}_{22}\widehat{\eta}^\intercal,$$
$$\widehat{\Upsilon} = \widehat{\sigma^\alpha}^2\widehat{\eta}\left(\widehat{\Sigma}^\beta - \widehat{\beta}_0\widehat{\beta}_0^\intercal\right)^{-1}\widehat{\eta}^\intercal,$$

where $\widehat{\Pi}_{11}, \widehat{\Pi}_{12}, \widehat{\Pi}_{22}$, are the HAC-type estimators of Newey and West (1987), defined as:

$$\widehat{\Pi}_{11} = \frac{1}{T}\sum_{t=1}^{T}\text{vec}(\widehat{z}_t\widehat{v}_t^\intercal)\text{vec}(\widehat{z}_t\widehat{v}_t^\intercal)^\intercal + \frac{1}{T}\sum_{m=1}^{q}\sum_{t=m+1}^{T}\left(1 - \frac{m}{q+1}\right)\left(\text{vec}(\widehat{z}_{t-m}\widehat{v}_{t-m}^\intercal)\text{vec}(\widehat{z}_t\widehat{v}_t^\intercal)^\intercal + \text{vec}(\widehat{z}_t\widehat{v}_t^\intercal)\text{vec}(\widehat{z}_{t-m}\widehat{v}_{t-m}^\intercal)^\intercal\right),$$
$$\widehat{\Pi}_{12} = \frac{1}{T}\sum_{t=1}^{T}\text{vec}(\widehat{z}_t\widehat{v}_t^\intercal)\widehat{v}_t^\intercal + \frac{1}{T}\sum_{m=1}^{q}\sum_{t=m+1}^{T}\left(1 - \frac{m}{q+1}\right)\left(\text{vec}(\widehat{z}_{t-m}\widehat{v}_{t-m}^\intercal)\widehat{v}_t^\intercal + \text{vec}(\widehat{z}_t\widehat{v}_t^\intercal)\widehat{v}_{t-m}^\intercal\right),$$
$$\widehat{\Pi}_{22} = \frac{1}{T}\sum_{t=1}^{T}\widehat{v}_t\widehat{v}_t^\intercal + \frac{1}{T}\sum_{m=1}^{q}\sum_{t=m+1}^{T}\left(1 - \frac{m}{q+1}\right)\left(\widehat{v}_{t-m}\widehat{v}_t^\intercal + \widehat{v}_t\widehat{v}_{t-m}^\intercal\right),$$

and

$$\widehat{Z} = \bar{G} - \widehat{\eta}\widehat{V}, \quad \widehat{\Sigma}^\beta = n^{-1}\widehat{\beta}^\intercal\widehat{\beta}, \quad \widehat{\Sigma}^v = T^{-1}\widehat{V}\widehat{V}^\intercal, \quad \widehat{\beta}_0 = n^{-1}\widehat{\beta}^\intercal\iota_n, \quad \widehat{\sigma^\alpha}^2 = n^{-1}\left\|\bar{r} - (\iota_n : \widehat{\beta})\widetilde{\Gamma}\right\|_{\text{F}}^2,$$
$$\widetilde{\gamma} = \left(\widehat{\beta}^\intercal\mathbb{M}_{\iota_n}\widehat{\beta}\right)^{-1}\widehat{\beta}^\intercal\mathbb{M}_{\iota_n}\bar{r}, \quad \widetilde{\Gamma} = (\widehat{\gamma}_0 : \widetilde{\gamma}^\intercal)^\intercal,$$

with $q \to \infty$, $q(T^{-1/4} + n^{-1/4}) \to 0$, as $n, T \to \infty$.

To prove the validity of these estimators, we need additional assumptions, because the estimands are more complicated than the parameters we estimate.

Theorem 6. *Under Assumptions A.2, and A.4 – A.16, and suppose that $\widehat{p} \xrightarrow{p} p$, then as n, $T \to \infty$, $n^{-3}T \to 0$, $q(T^{-1/4} + n^{-1/4}) \to 0$, $\widehat{\Phi} \xrightarrow{p} \Phi$ and $\widehat{\Upsilon} \xrightarrow{p} \Upsilon$.*

[8] To prove this result, we need much stronger assumptions on u_t. This is because the proof relies on the use of random matrix theory to analyze the eigenvalues and eigenvectors of large sample covariance matrices. The i.i.d. assumption is typically imposed in most scenarios.

4.7 Testing the Strength of an Observed Factor

As discussed in the introduction, a recent literature has explored the potential biases associated with the presence of weak factors (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 equation (2)). If either $\eta = 0$ (g_t is not a priced factor) or the factor is very noisy (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 proxy g_t has zero or low exposures 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. The R^2 of that regression 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 $\mathbb{H}_0 : \eta = 0$ vs $\mathbb{H}_1 : \eta \neq 0$, and construct a Wald Test. Our test statistic is given by

$$\widehat{W} = T\widehat{\eta}\left(\widehat{\Sigma}_v^{-1}\widehat{\Pi}_{11}\widehat{\Sigma}_v^{-1}\right)^{-1}\widehat{\eta}^{\intercal},$$

where $\widehat{\Pi}_{11}$ and $\widehat{\Sigma}_v$ are constructed in Section 4.6.

The next theorem establishes the desired size control and the consistency of the test.

Theorem 7. *Suppose $d = 1$ and $\widehat{p} \xrightarrow{p} p$. Under Assumptions A.2, and A.4 – A.16, and as n, $T \to \infty$, $n^{-2}T \to 0$, $q(T^{-1/4} + n^{-1/4}) \to 0$, we have*

$$\lim_{n,T\to\infty} \mathrm{P}\left(\widehat{W} > \chi^2_{\widehat{p}}(1-\alpha_0)|\mathbb{H}_0\right) = \alpha_0, \quad \textit{and} \quad \lim_{n,T\to\infty} \mathrm{P}\left(\widehat{W} > \chi^2_{\widehat{p}}(1-\alpha_0)|\mathbb{H}_1\right) = 1,$$

where $\chi^2_{\widehat{p}}(1-\alpha_0)$ is the $(1-\alpha_0)$-quantile of the chi-squared distribution with $\widehat{p}$ degree of freedom.

Our assumptions that the latent factors are pervasive, while observable factors can potentially be weak, are not in conflict with existing empirical evidence. It is known from the literature (e.g., Bernanke and Kuttner (2005) and 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 well known to be weak in some cases, like for example industrial production (see for example Gospodinov et al. (2014) and Bryzgalova (2015)).

5 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.

5.1 Data

We conduct our empirical analysis on a large set of standard portfolios of U.S. equities, testing several asset pricing models that have focused on risk premia in equity markets. We target U.S. equities because of their better data quality and because they are available for a long time period. However, our methodology could be applied to any country or asset class.

We include in our analysis 202 portfolios: 25 portfolios sorted by size and book-to-market ratio, 17 industry portfolios, 25 portfolios sorted by operating profitability and investment, 25 portfolios sorted by size and variance, 35 portfolios sorted by size and net issuance, 25 portfolios sorted by size and accruals, 25 portfolios sorted by size and beta, and 25 portfolio sorted by size and momentum. This set of portfolios captures a vast cross section of anomalies and exposures to different factors; at the same time, they are easily available on Kenneth French's website, and therefore represent a natural starting point to illustrate our methodology.[9]

Although some of these portfolio returns have been available since 1926, we conduct most of our analysis on the period from July of 1963 to December of 2015 (630 months), for which all of the returns are available. 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 (McLean and Pontiff (2015); Harvey et al. (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). Here is a list of models and corresponding observable factors g_t we include in our analysis:[10]

1. Capital Asset Pricing Model (*CAPM*): the value-weighted market return, constructed from the Center for Research in Security Prices (CRSP) for all stocks listed on the NYSE, AMEX, or NASDAQ.

2. Fama-French three factors (*FF3*): in addition to the market return, the model includes SMB (size) and HML (value).

3. Carhart's four-factor model (*FF4*) that adds a momentum factor (MOM) to $FF3$.

4. Fama-French five-factor model (*FF5*), from Fama and French (2015). The model adds to $FF3$ RMW (operating profitability) and CMA (investment).[11]

5. Betting-against-beta factor (*BAB*) from Frazzini and Pedersen (2014).

[9]See the description of all portfolio construction on Kenneth French's website: `http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`.

[10]Factor time series for models 1-4 are obtained from Kenneth French's website; for models 5-6, from AQR's website; for model 7, from the Federal Reserve Bank of St. Louis; for model 8, from Sydney Ludvigson's website; for model 9, from Lubos Pastor's website; for model 10, from Bryan Kelly's website; for model 11, from the various sources indicated in Novy-Marx (2014); for model 12, from Toby Moskowitz's website.

[11]We have also explored the four-factor model of Hou et al. (2015), which includes the market return, ME (size), IA (investment), ROE (profitability). Results are qualitatively and quantitatively similar to the FF5.

6. Quality-minus-junk factor (QMJ) from Asness et al. (2013).

7. Industrial production growth (IP). Industrial production is a macroeconomic factor available for the entire sample period at the monthly frequency. We use AR(1) innovations as the factor.

8. The first three principal components of 279 macro-finance variables constructed by Ludvigson and Ng (2009) (LN), also available at the monthly frequency. We estimate a VAR(1) with those three principal components, and use innovations as factors.

9. The liquidity factor from Pástor and Stambaugh (2003).

10. Two intermediary capital factors, one from He et al. (2016) and one from Adrian et al. (2014).

11. 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.[12]

12. Two consumption-based factors from Malloy et al. (2009). They include both an aggregate consumption series and a stockholder's consumption series.

5.2 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. Figure 1 (left panel) reports the first eight eigenvalues of the covariance matrix of returns for our panel of 202 portfolios. As typical for large panels, the first eigenvalue tends to be much larger than the others, so on the right panel we plot the eigenvalues excluding the first one. We observe a noticeable decrease in the eigenvalues up to four factors, suggesting $\breve{p} = 4$. This is also the number suggested by our estimator. As discussed in Section 4, our analysis is consistent as long as the number of factors $\breve{p}$ is at least as large as the true dimension p; to show the robustness of our results, we report the estimates separately using four, five, and six factors. The analysis is robust to using more factors.

The model with four PCs has a cross-sectional R^2 of 65%, indicating that it accounts for a significant fraction of the cross-sectional variation in expected returns for the 202 test portfolios, but leaving some unexplained variation. This number is comparable with the 73% cross-sectional R^2 one obtains using the FF3 model on the cross-section of 25 portfolios sorted by size and book-to-market, yet, we obtain it for a cross-section eight times as large, and using a model with just one more factor. We report in Figure 2 the actual and predicted expected excess returns for the model. Each panel of the figure highlights one of the eight test-asset groups that comprise our total of 202 portfolios. The fit is better for some groups of assets than others, but overall the factor model performs relatively well. These results change little when using returns in excess of T-bill rate instead of estimating the zero-beta rate.

[12] 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.

5.3 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 (T-bill) rate. In this case, we can compare our three-pass procedure with both Fama-MacBeth cross-sectional regressions and with the mimicking portfolio approach.

For each factor (or group of factors) g_t that we consider, Table 1 reports risk premia estimated using these different methodologies. The first column 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, that is however possible only for tradable factors.

The rest of the table considers three implementations of the two-pass cross-sectional estimator, three 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 results) and controlling for the Fama-French three factors (Market, SMB, HML; third set of results).

Next, using the mimicking-portfolio approach, we project the factor g_t onto the market portfolio alone (first set of results), onto the Fama-French 3 factors (second set of results), and onto all 202 test assets (third set of results). Note that the latter version of the approach (that projects onto all the available assets) is rarely applied in the literature, as it is inefficient or even infeasible when $n > T$; in our case, however, we have $n = 202$ and $T = 630$, so this projection is feasible and we therefore report it for comparison.[13]

Finally, we report our three-pass estimator at the end, using four principal components. We explore robustness with respect to the number of factors in the next table, for reasons of space.

To help with the interpretation of the table, we first examine one example in detail. Consider the profitability factor proposed by Fama and French (2015), RMW. The time-series average excess return is 25bp per month. Estimating the RMW risk premium in a two-pass cross-sectional regression with no controls yields an estimate of -16bp. Adding the market gives -4bp, and further adding SMB and HML gives 32bp. The results clearly depend on which controls are used in the estimation. Similarly, consider the three implementations of the mimicking-portfolio approach. When we project RMW 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 and significant risk premia estimates; when we project it on all 202 portfolios, we obtain a positive and significant estimate. Therefore, the results also vary dramatically with the choice of portfolios on which the factor is projected. Finally, our estimator provides a statistically significant 15bp estimate for the risk premium of this factor.

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

[13] If the factor is tradable and is itself included in the set of assets on which it is projected, the mimicking portfolio approach will yield an estimate equal to the average time-series excess return of the factor. We consider here the case in which the factor is not added to the set of test assets for the projection, to show how the estimator depends on the choice of projection portfolios. This sensitivity to the choice of assets is crucial for the case of nontradable factors, in which case the factor itself cannot be added to the space of returns on which the projection is applied.

Two-pass cross-sectional estimator. Both potential sources of bias described in this paper for two-pass cross-sectional regressions are visible in the table. First, for most factors, there are significant differences in the estimates obtained using different control factors (namely: no controls, market alone, FF3). This shows the potential for quantitatively meaningful biases that could arise if the wrong set of controls is specified. At the same time, it highlights how the standard procedure of arbitrarily selecting controls in these regressions can influence the resulting risk premia estimates.

The second potential source of bias is due to measurement error. As discussed above, 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). This appears to be the reason for the often extreme risk premia estimates obtained using the two-pass regression, for what appear to be weak factors (for example, the Novy-Marx (2014) factors). As described in Section 4.7, our three-pass procedure is immune to the problem of weak observable factors – in fact, the next table will report the results of our test to detect such factors.

Mimicking-portfolio estimator. The mimicking-portfolio estimator is similarly sensitive to the choice of portfolios on which factors are projected. It is not uncommon to see opposite signs across the different set of results for this estimator. For example, MOM is estimated to have a negative and significant risk premium when projected onto the market or onto the FF3 portfolios, but a positive risk premium when projected on the entire set of 202 portfolios. The first macro factor from Ludvigson and Ng (2009), instead, appears statistically significant when projected onto the FF3 portfolios, but not the market alone or all portfolios. These results highlight a quantitatively meaningful bias that could arise when important portfolios are omitted from the projection.

Three-pass estimator. The last column of the table reports the results using our three-pass estimator, using four principal components. For the case of tradable factors, the estimator produces results that are mostly close to the average excess returns of the factors. In all cases, the *sign* of the estimated risk premium is the same as the average return of the factor, which does not hold generally for the other estimators. For example, it estimates a market risk premium of around 50bp (exactly in line with the average market excess return) and a momentum risk premium of 77bp.

The three-pass procedure finds several nontradable factor risk premia economically and statistically significant: the liquidity factor of Pástor and Stambaugh (2003), both intermediary factors of He et al. (2016) and Adrian et al. (2014), the first macro PC from Ludvigson and Ng (2009), and also stockholders' consumption growth from Malloy et al. (2009). Nonetheless, several other nontradable factors do not appear to have statistically significant risk premia, for example the Novy-Marx (2014) factors or IP growth.

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.

Allowing for a unconstrained zero-beta rate. Table 2 shows the results produced by our more general estimator (9), that allows the zero-beta rate to be different from the T-bill rate.[14] We do not report the mimicking portfolio results here because that estimation approach does not allow for an unconstrained zero-beta rate. Instead, we report in this table our three-pass results for different numbers of PCs, from 4 to 6. Also, we do not report the estimates for the zero-beta rate for reasons of space. For the three-pass procedure, they are consistently around 0.55% per month (close to the average T-bill return of 0.4% per month).[15] For the two-pass regressions, they are in the vast majority of cases significantly above 1% per month.

The results with the unconstrained zero-beta rate are mostly similar to those presented in the previous table, but there are some additional noteworthy results. First, the table shows that risk premia estimates using the three-pass method are very robust to the number of PCs used (from 4 to 6). For example, for the liquidity factor we estimate a risk premium of 26bp with 4 and 5 factors, and 25bp with 6 factors. Similar results hold for all factors, both tradable and nontradable.

Second, the table also reports the R^2 of the time-series regression of each observed factor g_t onto the $\breve{p}$ latent factors; we refer to this as R_g^2. 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^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 (for IP, for example, it is below 1%), indicating that these factors are dominated by noise. We highlight this point in Figure 3, which shows the time series of cumulated innovations in the original and cleaned (i.e., fitted) factors, for a few of them. The figures provide a graphical representation of the extent to which the PCs of returns capture the variation in each factor. While for many of the tradable factors the original and cleaned factors track each other closely, for others the cleaned factor displays much lower variation than the original factor: the difference is the measurement error that our procedure has eliminated.

Finally, the last column of the table also reports the p-value for the test of the null that each factor g_t is weak, described in Section 4.7.[16] 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.

The zero-beta rate and the sign of the market risk premium. A well-known result in the empirical asset pricing literature is that market risk premia are estimated to be negative in standard two-pass regressions with an unrestricted zero-beta rate. This happens in our dataset as well. The two-pass estimators of the market risk premium yield a zero-beta rate estimate of around 125bp per month (not reported in the table) and a market risk premium of -20bp to -57bp per month, depending on the specification (first row of Table 2).

[14] The inference based on Theorem 3 in this case is also robust to the presence of pricing errors (alphas) that satisfy certain conditions.

[15] Recall that in the case of the three-pass procedure, the estimate of the zero-beta rate, obtained at step (ii), does not depend on the factor g_t.

[16] To be more conservative, we use $\breve{p} = 6$ for this test, corresponding to the rightmost set of results in the table, but results are similar for all values of $\breve{p}$.

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 the three-pass procedure lowers the estimated zero-beta rate (to 55bp per month) and produces a positive and significant estimate of the market risk premium (37bp), as the right side of the first row shows.

We can further investigate the relationship between market beta and expected returns after controlling for the explanatory power of the omitted factors using a residual regression approach.[17] Figure 4 plots the expected return vs. market beta, after partialing out the component explained by the other factors. The left panel uses the standard two-pass estimator for the market risk premium in the FF3 model, whereas the right panel uses our three-pass methodology. The solid red line in each graph corresponds to the slope estimated from the historical average return of the market, whereas the dashed line corresponds to the fitted line, using the cross-sectional estimate of the slope. Given that the market is a tradable factor, if the model is correctly specified, the two lines should overlap.

This is clearly not the case for the two-pass cross-sectional regression (left panel). However, the right panel of the figure shows that indeed, once the omitted factors are accounted for, there is a clear positive relationship between market beta and expected returns, and the slope is close to the average excess return of the market portfolio. Overall, the fact that the market risk premium significantly changes sign depending on whether we control for omitted factors serves as a strong warning that omitting factors could have important effects on our statistical and economic conclusions about the pricing of risks.

5.4 Observable and Unobservable Factors

The core of our estimation methodology is the link between the observable factors g_t and the unobservable factors v_t, through Equation (2). In particular, η represents the loadings of g_t onto the p PCs, and therefore reveals the exposures of the observable factors to the latent factors extracted by PCA.

In Table 3 we decompose the variance of g_t explained by the set of PC factors into the parts attributable to exposure to each individual PC factor (which is possible because PC factors are orthogonal to each other). Each row of the table sums up to 100%. This allows us to highlight which latent factors are most responsible for the variation of each observable factor. Note that the PC factors are ordered by their eigenvalues (largest to smallest).

The first row shows that the market return loads mostly onto the first PC (i.e., on the factor with the largest eigenvalue). This is expected because the market represents the largest source of common variation across assets. The other observable factors show interesting variation in their exposure to the PCs. For example, SMB loads on both the first and second factors, HML mostly on the third one, and MOM mostly on the fourth one. RMW loads substantially on at least four factors (including the sixth

[17] Recall that the estimate of the market risk premium using cross-sectional methods is the slope of a multivariate regression of average returns onto the betas of returns with respect to the market and the control factors. It is well known that the slope of any multivariate regression with respect to a specific variable (in this case, the market beta) can be also obtained by first separately regressing the outcome variable (average returns) and the regressor of interest (market beta) on the remaining regressors (the betas with respect to the other factors), and in a second stage regressing the residuals of the two regressions against each other. In this way, we can first partial out the component of the cross-section of expected returns and of market betas explained by the control factors, and then study the univariate relation between the residuals.

one), and CMA loads mostly on the same factor as HML. However, CMA and HML are still strongly distinguished by a differential exposure to the other factors.

Macro factors load onto the PCs in nontrivial ways. IP is mostly exposed to the fourth and sixth factors. The consumption factor loads mostly on the first two latent factors, whereas the stockholder consumption factor loads most strongly on the fifth one. The Novy-Marx (2014) factors and the Ludvigson and Ng (2009) factors are exposed to many risk sources.

Finally, both the liquidity factor of Pástor and Stambaugh (2003) and the intermediary factor of He et al. (2016) are strongly exposed to the first latent factor.

A caveat about these results is that the PCs cannot be interpreted as "fundamental" factors, but rather as an unknown rotation of them. So this decomposition cannot be viewed as linking the observable factors to any *interpretable* latent factors. Rather, it simply shows similarities and differences in the source of common variation different observable factors are related to.

5.5 Risk Premia across Asset Classes

The main analysis presented in the previous sections has focused on a large cross-section of equity portfolios, for which a long time series is available. In this section we explore risk premia for the same factors discussed above as we look instead at non-equity portfolios.

We obtain from Asaf Manela's website the time series of non-equity portfolio returns used in He et al. (2016), which in turn collects portfolio data from various sources. The data includes ten maturity-sorted government bond portfolios, ten corporate bond portfolios sorted on yield spread, six sovereign bond portfolios, 18 S&P 500 option portfolios sorted on moneyness and maturity, six currency portfolios sorted on interest rate differentials, six currency portfolios sorted on currency momentum, 24 commodity futures returns, and 20 CDS portfolios sorted by spread, for a total of 100 non-equity portfolios. Due to data availability for the non-equity portfolios, the sample covers the period 1970-2012. In addition, since not all portfolios are available for the entire time period, we estimate the variance and pairwise covariances separately with all available data. While the resulting covariance matrix is not necessarily positive-definite, it leads to consistent estimates for PCs.

Table 4 reports the results of the risk premia estimation with our three-pass procedure using equity and non-equity assets. The left panel of the table shows the results using the 202 equity portfolios (as in our main analysis) over this sample period. The results are qualitatively and quantitatively similar to our baseline results that use a longer sample. The middle panel of the table uses as test assets the 100 non-equity portfolios together with some equity portfolios (the Fama-French 25 portfolios), whereas the right panel uses *only* the non-equity portfolios. We use five principal components in the first two panels and six in the third (consistent with our estimator for p in this sample), achieving a cross-sectional R^2 of 58%, 57%, and 54% respectively.

For many (but not all) factors, the estimates of risk premia over these different groups of test portfolios are surprisingly stable. For example, the market risk premium is estimated to be positive and large in the cross-section of non-equity portfolios; similarly, the liquidity factor displays a risk premium of around 20bp per month in all of these samples, and the same goes for the intermediary capital factors (significance is a bit reduced in this table relative to the previous tables, because the sample is shorter).

Finally, IP is estimated to have a zero risk premium in every case. So while the risk premia associated with some factors (like equity momentum or profitability) do not seem to be stable across markets, the analysis does uncover factors that seem to display consistent risk premia in all markets we study.

These results suggest the existence of common risk factors across different markets. While some degree of segmentation surely exists across these markets, these results indicate that at least some aggregate risk factors may be pervasive across many markets and their risk premia are consistent across them. Unlike FF3 or FF5 for equity markets, there are no well-established models or factors that serve as a good benchmark across markets. In this case, key to correctly uncovering the risk premium of these factors is to properly control for the non-observable factors specific to each market, something that our approach can achieve.

5.6 Robustness and Extensions

In this section we explore the robustness of our main results to the choice of test portfolios, sample period, and choice of the method to recover the factor space, particularly as it relates to the potential weakness of factors. We also discuss how to handle time variation in risk premia and risk exposures.

5.6.1 Robustness to the Choice of Test Portfolios

Our main empirical results are obtained using a large set of 202 portfolios, and our methodology is specifically designed to be used with as many assets as possible, so that all relevant dimensions of risk will be expressed in the cross section. It is natural, however, to wonder to what extent the results are affected by the particular selection of test assets.

To investigate this question systematically, we perform the following robustness exercise. From the 202 test portfolios we use in our empirical exercise, we randomly select (without replacement) half of them, and we re-estimate the risk premia of all observable factors in this subsample. We repeat this exercise 10,000 times, thus obtaining a distribution of risk premia estimates across subsamples of 101 portfolios each, randomly selected.

Figure 5 shows the results for several factors. Note that all panels of the figure report the same range of risk premia (x axis, between -20bp and 100bp), so that the histograms are easily comparable across panels. The results are quite heterogeneous across factors. In the top left panel, we see that the risk premium for the market return is clearly positive in the vast majority of cases (it is below zero only in a small set of subsamples). At the same time, its exact magnitude varies across subsamples. The top right panel shows that instead the risk premia of SMB and HML are much more precisely estimated using our three-pass regression method, and similarly for MOM (middle left panel).

The last three panels show interesting results for non-tradable factors. Confirming the results of Table 2 and Figure 3, IP is a useless factor, with a risk premium of effectively zero across all subsamples. On the contrary, liquidity and intermediary capital factors all appear positively priced across subsamples.

Overall, our subsample results show that the conclusions of our empirical analysis are robust to the selection of the test assets.

5.6.2 Robustness to the Choice of Time Period

A potential concern when working with PCs is the stability of the estimated loadings and factors over time. The extent to which our risk premia estimates are consistent across time periods is an empirical question that we explore in this section.

Similarly to the robustness with respect to the test assets, we perform our robustness check with respect to the sample period by resampling half of the time periods randomly without replacement, and looking at the variability of the risk premia estimates. Simple resampling in the time series is possible in our context because of the low serial correlation of returns and factor innovations over time.

Figure 6 shows the results. Both quantitatively and qualitatively, the results are very similar to the ones in the previous section (where we randomly resampled the cross-section as opposed to the time series). The results show that all of the main conclusions of our main analysis hold when looking across subsamples.

While the stability across subsamples may seem surprising, it is useful to note that our risk premia estimator is not *only* based on PCA. Instead, a key step is the projection of the factor of interest g_t onto the extracted PCs. So any rotation that makes the extracted factors differ across subsamples will be entirely offset by a corresponding rotation of the loading of g_t onto those factors – resulting in stable risk premia estimates for the observable factors.

5.6.3 Robustness to the Presence of Weak Latent Factors

For the invariance result to hold, it is essential that the control factors, together with g_t, span the entire factor space. The first step of our three-pass procedure involves using PCA to recover the factor space. As we discuss in the paper, our procedure works even if the *observable* factor g_t is weak (in fact, we propose a test for whether g_t is weak); however, PCA will not necessarily recover the entire factor space if the underlying *latent* factors are weak. In this section we summarize our main theoretical and empirical arguments for using PCA in practice, and propose an additional robustness test to mitigate the concern that the presence of weak factors may distort our results.

In theory, weak latent factors – unobservable factors for which the dispersion of risk exposures is small in the cross-section – can affect our estimator because they have low eigenvalues, and PCA might fail to separate them from noise. However, for weak factors to bias our estimates of risk premia for observable factors, they also need to have themselves high risk premia, which allows them to explain a significant portion of the cross-section of average returns. But large risk premia for factors with low eigenvalues imply high Sharpe ratios. A first theoretical argument in favor of focusing on the PCs with largest eigenvalues are *good-deal bounds*, which impose a theoretical upper bound on the potential bias from weak factors (Kozak et al. (2017) make precisely this argument to support using PCA in this context).

A second, empirical, argument is that we can easily add additional PCs with lower and lower eigenvalues, and verify that the risk premia estimates are stable. For example, in Tables 1 and 2, we explore the robustness of the results using 4, 5 and 6 factors. These robustness exercises are theoretically motivated by Theorem 5, which guarantees that adding "too many" PCs does not affect the consistency of our estimator.

A third way to verify that weak latent factors are not driving our empirical results is the comparison of the risk premia estimated for tradable factors using our three-pass procedure with those obtained as time-series average of the portfolios' excess returns. As discussed in previous sections, the two should be the same if the factor model is correctly specified. Biases due to the presence of weak latent factors should produce significant differences between the estimates using cross-sectional methods and the time-series averages.

Finally, we propose here an additional robustness test with respect to the possibility of weak factors, based on changing the objective function when extracting the statistical factors from the panel of returns. Recall that the first step towards PCA is to calculate eigenvalues of the covariance matrix of returns, which equal the variances of the corresponding PCs, and that the constructed factors are eigenvectors associated with the largest few eigenvalues.

Since weak factors are factors with low eigenvalues, which however explain the cross-section of returns, we can modify the objective function to account for the contribution to the cross-sectional variation. That is, rather than finding factors that best explain the time-series comovement of stock returns, we find factors that strike a balance between explaining the time-series comovement of stock returns and the cross-sectional variation of expected returns. This alternative objective function was first proposed by Connor and Korajczyk (1986). It is a convenient reference point because it puts equal weight on the two components of the objective function – the time-series and the cross-sectional variation.

As shown in Bai and Ng (2002), our PCA formula given in (7) is the solution to the following optimization problem:

$$\min_{\beta,\bar{V}} n^{-1}T^{-1}\left\|\bar{R}-\beta\bar{V}\right\|_{\mathrm{F}}^{2}, \quad \text{subject to} \quad T^{-1}\bar{V}\bar{V}^{\intercal}=\mathbb{I}_{\widehat{p}},$$

where $\|\cdot\|_{\mathrm{F}}$ is the Frobenius norm of a matrix. By our rotation invariance result, it would give the same risk premia estimates if we were to use an alternative normalization $n^{-1}\beta^{\intercal}\beta=\mathbb{I}_n$. Connor and Korajczyk (1986) suggest another optimization problem (henceforce CK):

$$\min_{\beta,\bar{V},\gamma} n^{-1}T^{-1}\left\|\bar{R}-\beta\bar{V}\right\|_{\mathrm{F}}^{2}+wn^{-1}\left\|\bar{r}-\beta\gamma\right\|_{\mathrm{F}}^{2}, \quad \text{subject to} \quad n^{-1}\beta^{\intercal}\beta=\mathbb{I}_n,$$

where they choose $w=1$. The solution turns out to be

$$\tilde{\beta}=n^{1/2}(\tilde{\varsigma}_1:\tilde{\varsigma}_2:\ldots:\tilde{\varsigma}_{\widehat{p}}), \quad \text{and} \quad \tilde{V}=(\tilde{\beta}^{\intercal}\tilde{\beta})^{-1}\tilde{\beta}^{\intercal}\bar{R},$$

where $\tilde{\varsigma}_1,\tilde{\varsigma}_2,\ldots,\tilde{\varsigma}_{\widehat{p}}$ are the eigenvectors associated with the largest eigenvalues of the matrix $n^{-1}T^{-1}\bar{R}\bar{R}^{\intercal}+wn^{-1}\bar{r}\bar{r}^{\intercal}$. Note that starting from CK's formulation, setting $w=0$ (thus focusing entirely on time-series comovement) is equivalent to PCA.

The CK approach can be used instead of the standard PCA in step (i) of our three-step procedure. Since the second term of the objective function is the cross-sectional R^2, it may help select latent factors that have large risk premia but are weak. We can then continue steps (ii) and (iii) as in Section 3 using the estimated latent factors together with g_t to estimate risk premia. Note that the CK approach does not allow for an unrestricted zero-beta rate or pricing errors.

In Table 5, we show how the risk premia estimates differ between the baseline PCA (left panel, $w = 0$) and the CK approach (right panel, $w = 1$). The fact that there is virtually no difference in the risk premia estimates suggests that weak factors are either not present in this dataset we consider, or if they are, they have small enough risk premia that ignoring them has little consequence for our estimates.

Taken together, these considerations lead us to conclude that for the purposes of estimating risk premia, using PCA to recover the factor space represents a simple yet robust solution.

5.6.4 Individual Assets vs. Portfolios

In this paper, we recommend using characteristic-sorted portfolios instead of individual stocks. The main advantage of using portfolios is that their risk exposures are more stable over time, as discussed at length in the asset pricing literature. This is particularly important in our setting, because we assume the betas of the test assets are constant.

To see this intuition more formally, call $\tilde{r}_t$ is the vector of time-t returns for m individual stocks, and c_t a $m \times n$ matrix of characteristics (or their functions) observed at time t for the m stocks. The typical procedure to construct characteristic-sorted portfolios in asset pricing categorizes stocks at each time $t-1$ into groups based on one or more observed characteristics, and then obtains the portfolio return at time t using equal or market-value weights for stocks in each group.

The sorting procedure can be represented mathematically by constructing the matrix c_{t-1} stacking side-by-side the n dummy variables corresponding to each characteristic-sorted group. For example, to construct 10 size-based portfolios, c_{t-1} would be an $m \times 10$ matrix containing 10 dummy variables, each indicating the size group to which each stock belongs at time $t-1$. The n characteristic-sorted portfolio returns from $t-1$ to t are simply the coefficients of a cross-sectional regression of $\tilde{r}_t$ onto c_{t-1}, since c_{t-1} contains only dummies.

More generally, given any matrix c_{t-1} (that could include dummies or continuous variables), the characteristics-sorted portfolio returns at time t are:

$$r_t = (c_{t-1}^{\intercal} c_{t-1})^{-1} c_{t-1}^{\intercal} \tilde{r}_t, \tag{13}$$

where the term $(c_{t-1}^{\intercal} c_{t-1})^{-1} c_{t-1}$ therefore represents the time-$(t-1)$ portfolio weights.

Using this expression that links r_t and $\tilde{r}_t$, it is immediate to find that if individual factor exposures are linear functions of c_{t-1} (e.g., Rosenberg (1974)), then the sorted portfolios have constant factor exposures. Specifically, extending our setup (1) to include time-varying factor exposures for individual asset returns, we have:

$$\tilde{r}_t = \beta_{t-1}\gamma_{t-1} + \beta_{t-1}\tilde{v}_t + \tilde{u}_t,$$

where $\tilde{r}_t$ and $\tilde{u}_t$ are $m \times 1$ vectors, β_{t-1} is an $m \times n$ matrix of time varying exposures, following

$$\beta_{t-1} = c_{t-1}\beta + \varepsilon_{t-1}, \tag{14}$$

for some $n \times p$ matrix β, $m \times n$ matrix of observable characteristics c_{t-1}, and some $n \times p$ matrix of residuals ε_{t-1}.

Prior to applying our three-pass estimation procedure, we construct characteristics-sorted portfolios:

$$r_t = (c_{t-1}^\intercal c_{t-1})^{-1} c_{t-1}^\intercal \tilde{r}_t = \beta\gamma + \beta v_t + u_t,$$

where

$$\gamma = \mathrm{E}(\gamma_{t-1}), \quad v_t = \tilde{v}_t + \gamma_{t-1} - \mathrm{E}(\gamma_{t-1}), \quad u_t = (c_{t-1}^\intercal c_{t-1})^{-1} c_{t-1}^\intercal \left(\tilde{u}_t + \varepsilon_{t-1}(\gamma_{t-1} + \tilde{v}_t)\right).$$

Therefore, our methodology to estimate risk premia can be applied even if individual stock risk exposures are time-varying, as long as characteristic-sorted portfolios that have constant factor exposures are used as test assets. Also, we can interpret the estimated risk premia as estimates of their time-series average.

In this paper, we take the portfolio-formation step as given, and use characteristic-sorted portfolios that have been proposed in the literature. In contrast, Kelly et al. (2017) construct such portfolios using characteristics and individual stocks for a model specification test. Their results show that PCs based on such portfolios explain more cross-sectional variations than those based on individual stocks, which is consistent with the formal result shown above that characteristic-sorted portfolios will have constant betas if the characteristics are chosen appropriately.

6 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 relies 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 enough factors to 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 principal components of returns as controls in the two-pass regressions, completing the factor space. It can also be thought 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, that occurs as $n \to \infty$.

Equally important to what we can recover is what we cannot recover if some factors are omitted:

how the pricing kernel loads onto the observed factors, as well as the set of true risk exposures to each factor. These can only be pinned down under much stronger assumptions – by identifying all the factors that drive the pricing kernel, and explicitly specifying how they enter the pricing kernel. Instead, a notable property of factor risk premia is precisely that they can be recovered even without specifying the identities of all factors, and this is what we focus on in this paper.

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 space of the omitted factors. It also explicitly takes into account the possibility of measurement error in any observed factor.

The application of the methodology to workhorse factor models using equity test assets yields several compelling results. Contrary to most existing estimates, we find that the risk premium estimate associated with market risk exposure is positive and significant even when the zero-beta rate is left unrestricted, and close to the time-series average excess return of the market portfolio. This confirms that our methodology correctly recovers the risk premium of the market (and similar results hold for other tradable factors), thus mitigating misspecification concerns. The most interesting results appear for non-tradable factors. Many standard macroeconomic factors appear insignificant, whereas non-tradable factors related to various market frictions (like liquidity and intermediary leverage) appear strongly significant even when considered as part of richer linear pricing models that include additional factors. Similar results hold when looking across asset classes; the stability of the risk premia estimates for several factors across markets suggests the presence of pervasive aggregate risks that can be detected once factors specific to the various asset classes are properly accounted for – which in this case is achieved using the three-pass methodology we propose.

References

Adrian, T., E. Etula, and T. Muir (2014). Financial intermediaries and the cross-section of asset returns. *The Journal of Finance 69*(6), 2557–2596.

Ahn, S. C. and A. R. Horenstein (2013). Eigenvalue ratio test for the number of factors. *Econometrica 81*, 1203–1227.

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

Asness, C. S., A. Frazzini, and L. H. Pedersen (2013). Quality Minus Junk. Technical report, AQR.

Bai, J. (2003). Inferential Theory for Factor Models of Large Dimensions. *Econometrica 71*(1), 135–171.

Bai, J. and Y. Liao (2013). Statistical inferences using large estimated covariances for panel data and factor models. Technical report, Columbia University.

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

Bai, J. and S. Ng (2006). Evaluating latent and observed factors in macroeconomics and finance. *Journal of Econometrics 131*(1), 507–537.

Bai, J. and G. Zhou (2015). Fama–MacBeth two-pass regressions: Improving risk premia estimates. *Finance Research Letters 15*, 31–40.

Bai, Z. and J. W. Silverstein (2009). *Spectral Analysis of Large Dimensional Random Matrices*. Springer.

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

Bernanke, B. S. and K. Kuttner (2005). What explains the stock market's reaction to federal reserve policy. *The Journal of Finance 60*, 1221–1257.

Black, F. (1972). Capital market equilibrium with restricted borrowing. *The Journal of Business 45*(3), 444–455.

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

Breeden, D. T., M. R. Gibbons, and R. H. Litzenberger (1989, June). Empirical Test of the Consumption-Oriented CAPM. *The Journal of Finance 44*(2), 231–262.

Bryzgalova, S. (2015). Spurious Factors in Linear Asset Pricing Models. Technical report, Stanford University.

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

Chamberlain, G. and M. Rothschild (1983). Arbitrage, factor structure, and mean-variance analysis on large asset markets. *Econometrica 51*, 1281–1304.

Cochrane, J. H. (2009). *Asset Pricing:(Revised Edition)*. Princeton university press.

Connor, G., M. Hagmann, and O. Linton (2012). Efficient semiparametric estimation of the fama-french model and extensions. *Econometrica 80*(2), 713–754.

Connor, G. and R. A. Korajczyk (1986). Performance measurement with the arbitrage pricing theory: A new framework for analysis. *Journal of Financial Economics 15*(3), 373–394.

Connor, G. and R. A. Korajczyk (1988). Risk and return in an equilibrium APT: Application of a new test methodology. *Journal of Financial Economics 21*(2), 255–289.

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

Fama, E. F. and K. R. French (2015). A five-factor asset pricing model. *Journal of Financial Economics 116*(1), 1–22.

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

Fan, J., Y. Liao, and M. Mincheva (2011). High-dimensional covariance matrix estimation in approximate factor models. *Annals of Statistics 39*(6), 3320–3356.

Fan, J., Y. Liao, and M. Mincheva (2013). Large covariance estimation by thresholding principal orthogonal complements. *Journal of the Royal Statistical Society, B 75*, 603–680.

Fan, J., Y. Liao, and J. Yao (2015). Power enhancement in high-dimensional cross-sectional tests. *Econometrica 83*(4), 1497–1541.

Ferson, W. E. and C. R. Harvey (1991). The variation of economic risk premiums. *Journal of Political Economy 99*(2), 385–415.

Frazzini, A. and L. H. Pedersen (2014). Betting against beta. *Journal of Financial Economics 111*(1), 1–25.

Gagliardini, P., E. Ossola, and O. Scaillet (2016). Time-varying risk premium in large cross-sectional equity datasets. *Econometrica 84*(3), 985–1046.

Gagliardini, P., E. Ossola, and O. Scaillet (2017). A diagnostic criterion for approximate factor structure. Technical report, Swiss Finance Institute.

Gibbons, M., S. A. Ross, and J. Shanken (1989). A test of the efficiency of a given portfolio. *Econometrica 57*(5), 1121–1152.

Gospodinov, N., R. Kan, and C. Robotti (2013). Chi-squared tests for evaluation and comparison of asset pricing models. *Journal of Econometrics 173*(1), 108–125.

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

Gospodinov, N., R. Kan, and C. Robotti (2016). Spurious inference in reduced-rank asset-pricing models. Technical report, Imperial College London.

Harvey, C. R., Y. Liu, and H. Zhu (2016). ...and the Cross-Section of Expected Returns. *The Review of Financial Studies 29*(1), 5–68.

He, Z., B. Kelly, and A. Manela (2016). Intermediary asset pricing: New evidence from many asset classes. Technical report, National Bureau of Economic Research.

Horn, R. A. and C. R. Johnson (2013). *Matrix Analysis* (Second ed.). Cambridge University Press.

Hou, K. and R. Kimmel (2006). On the estimation of risk premia in linear factor models. Technical report, Working Paper, Ohio State University.

Hou, K., C. Xue, and L. Zhang (2015). Digesting anomalies: An investment approach. *Review of Financial Studies, Forthcoming 28*(3), 650–705.

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

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

Kan, R. and C. Robotti (2008). Specification tests of asset pricing models using excess returns. *Journal of Empirical Finance 15*(5), 816–838.

Kan, R. and C. Robotti (2009). Model comparison using the hansen-jagannathan distance. *Review of Financial Studies 22*(9), 3449–3490.

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

Kan, R. and C. Zhang (1999a). GMM tests of stochastic discount factor models with useless factors. *Journal of Financial Economics 54*(1), 103–127.

Kan, R. and C. Zhang (1999b). Two-Pass Tests of Asset Pricing Models. *The Journal of Finance LIV*(1), 203–235.

Kelly, B., S. Pruitt, and Y. Su (2017). Characteristics are risk exposures. Technical report, University of Chicago.

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

Kleibergen, F. and Z. Zhan (2014). Mimicking portfolios of macroeconomic factors. Technical report, Brown University Working Paper.

Kleibergen, F. and Z. Zhan (2018). Identification-robust inference on risk premia of mimicking portfolios of non-traded factors. *Journal of Financial Econometrics 16*(2), 155–190.

Kozak, S., S. Nagel, and S. Santosh (2017). Interpreting factor models. *Journal of Finance, forthcoming*.

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

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

Lewellen, J., S. Nagel, and J. Shanken (2010). A skeptical appraisal of asset pricing tests. *Journal of Financial Economics 96*(2), 175–194.

Lucca, D. O. and E. Moench (2015). The pre-FOMC announcement drift. *The Journal of Finance 70*(1), 329–371.

Ludvigson, S. C. and S. Ng (2009). A factor analysis of bond risk premia. Technical report, National Bureau of Economic Research.

MacKinlay, A. C. and M. P. Richardson (1991). Using Generalized Method of Moments to Test Mean-Variance Efficiency. *The Journal of Finance 46*(2), 511–527.

Malloy, C. J., T. J. Moskowitz, and A. VISSING-JØRGENSEN (2009). Long-run stockholder consumption risk and asset returns. *The Journal of Finance 64*(6), 2427–2479.

McLean, R. D. and J. Pontiff (2015). Does Academic Research Destroy Stock Return Predictability? *The Journal of Finance LXXI*(1), 1–48.

Moon, H. R. and M. Weidner (2015). Linear regression for panel with unknown number of factors as interactive fixed effects. *Econometrica 83*(4), 1543–1579.

Newey, W. K. and K. D. West (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica 55*, 703–708.

Novy-Marx, R. (2014). Predicting anomaly performance with politics, the weather, global warming, sunspots, and the stars. *Journal of Financial Economics 112*(2), 137–146.

Onatski, A. (2010). Determining the number of factors from empirical distribution of eigenvalues. *Review of Economics and Statistics 92*, 1004–1016.

Onatski, A. (2012). Asymptotics of the principal components estimator of large factor models with weakly influential factors. *Journal of Econometrics 168*, 244–258.

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

Pesaran, M. and T. Yamagata (2012). Testing CAPM with a large number of assets. Technical report, University of South California.

Roll, R. and S. A. Ross (1980). An empirical investigation of the arbitrage pricing theory. *The Journal of Finance 35*(5), 1073–1103.

Rosenberg, B. (1974). Extra-market components of covariance in security returns. *Journal of Financial and Quantitative Analysis 9*(2), 263–274.

Ross, S. A. (1976). The Arbitrage Theory of Capital Asset Pricing. *Journal of Economics Theory 13*, 341–360.

Shanken, J. (1992). On the Estimation of Beta Pricing Models. *The Review of Financial Studies 5*(1), 1–33.

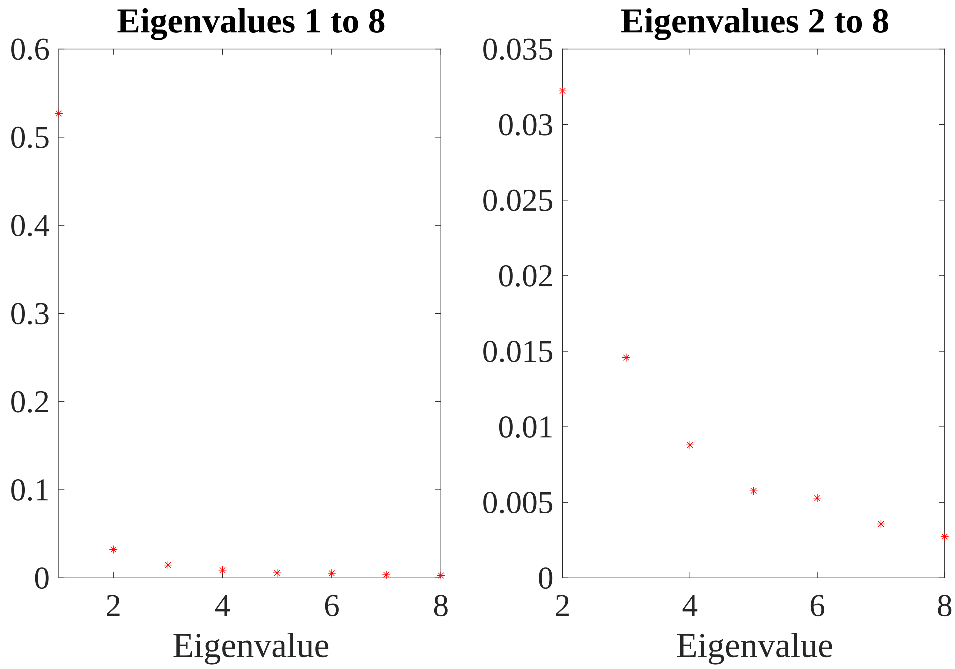
Shanken, J. and G. Zhou (2007). Estimating and testing beta pricing models: Alternative methods and their performance in simulations. *Journal of Financial Economics 84*(1), 40–86.

Welch, I. (2008). The link between fama-french time-series tests and fama-macbeth cross-sectional tests. Technical report, UCLA.

White, H. (2000). *Asymptotic Theory for Econometricians: Revised Edition.* Emerald Group Publishing Limited.

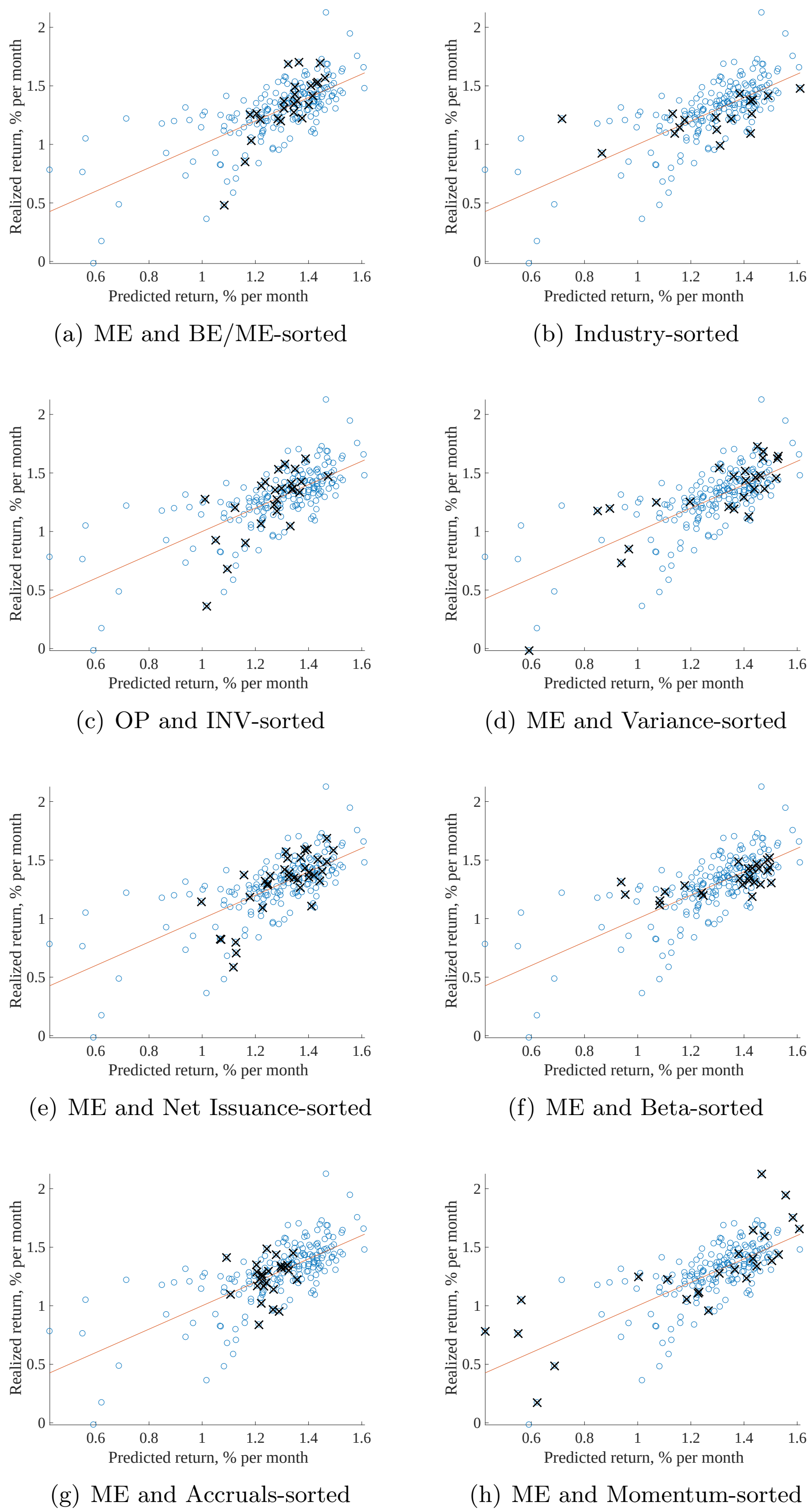
7 Figures and Tables

Figure 1: First Eight Eigenvalues of the Covariance Matrix of 202 Equity Portfolios



Note: The left panel reports the first eight eigenvalues of the covariance matrix of our 202 test portfolios. The right panel zooms in to the eigenvalues two through eight.

Figure 2: Predicted and Realized Average Excess Returns in a Six-Factor Model



(a) ME and BE/ME-sorted

(b) Industry-sorted

(c) OP and INV-sorted

(d) ME and Variance-sorted

(e) ME and Net Issuance-sorted

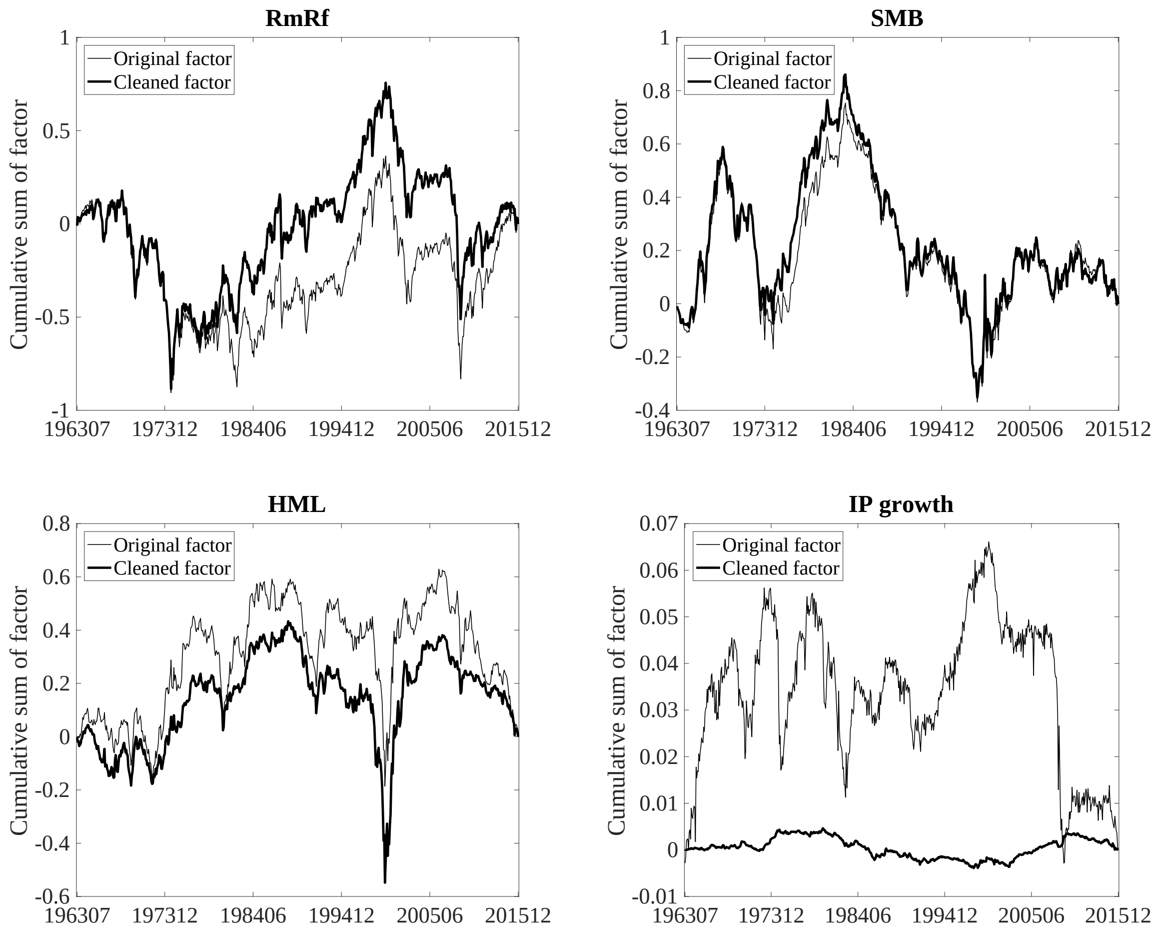
(f) ME and Beta-sorted

(g) ME and Accruals-sorted

(h) ME and Momentum-sorted

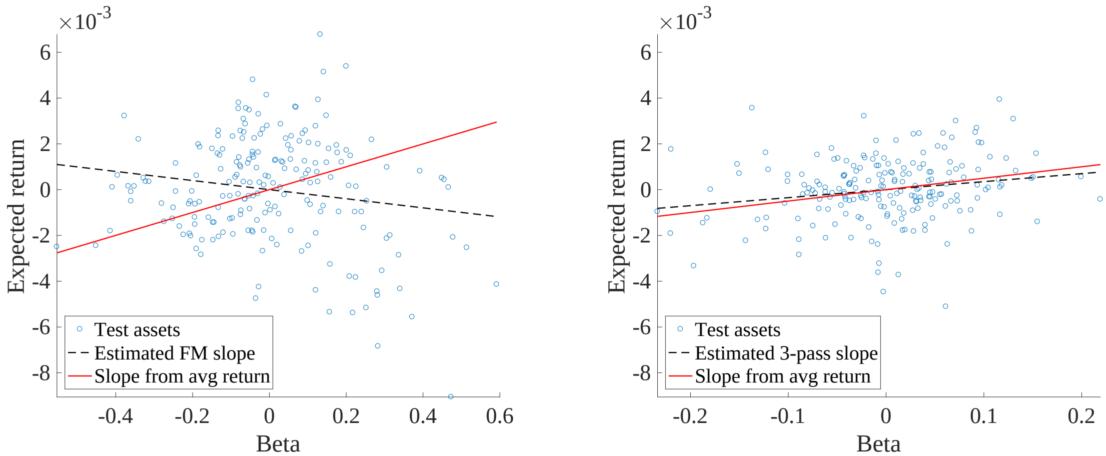
Note: This figure reports the predicted average excess returns of the 202 test portfolios against the realized average excess returns. Each panel highlights a different set of test assets. The solid line is the 45-degree line.

Figure 3: Cumulative Factor Time Series with and without Measurement Error



Note: This figure reports the time series of cumulative factor innovations for RmRf, SMB, HML, and IP (thin line) together with the time series obtained from removing measurement error from the factor (thick line).

Figure 4: Market Beta and Expected Return

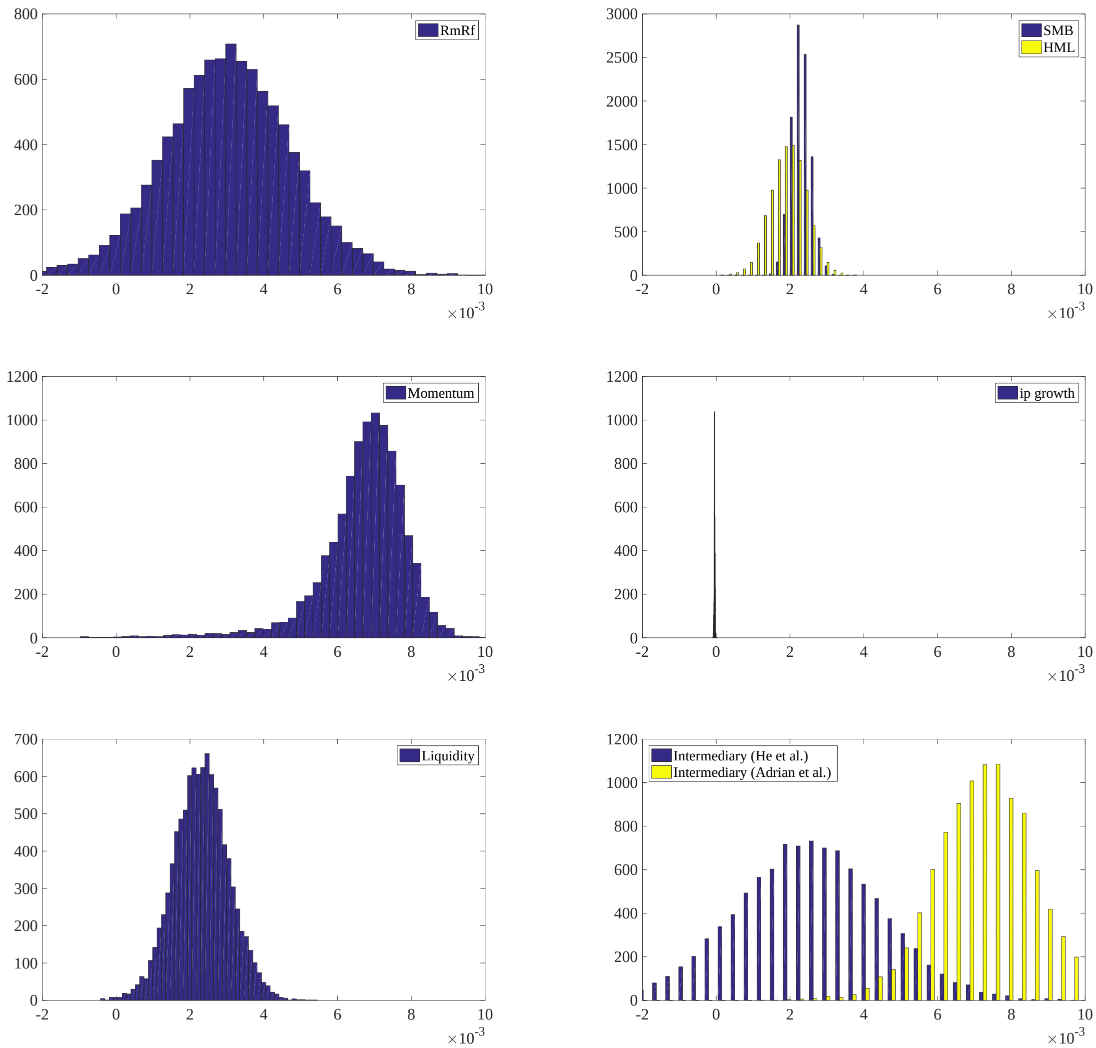


(a) Fama-MacBeth

(b) Three-pass estimator

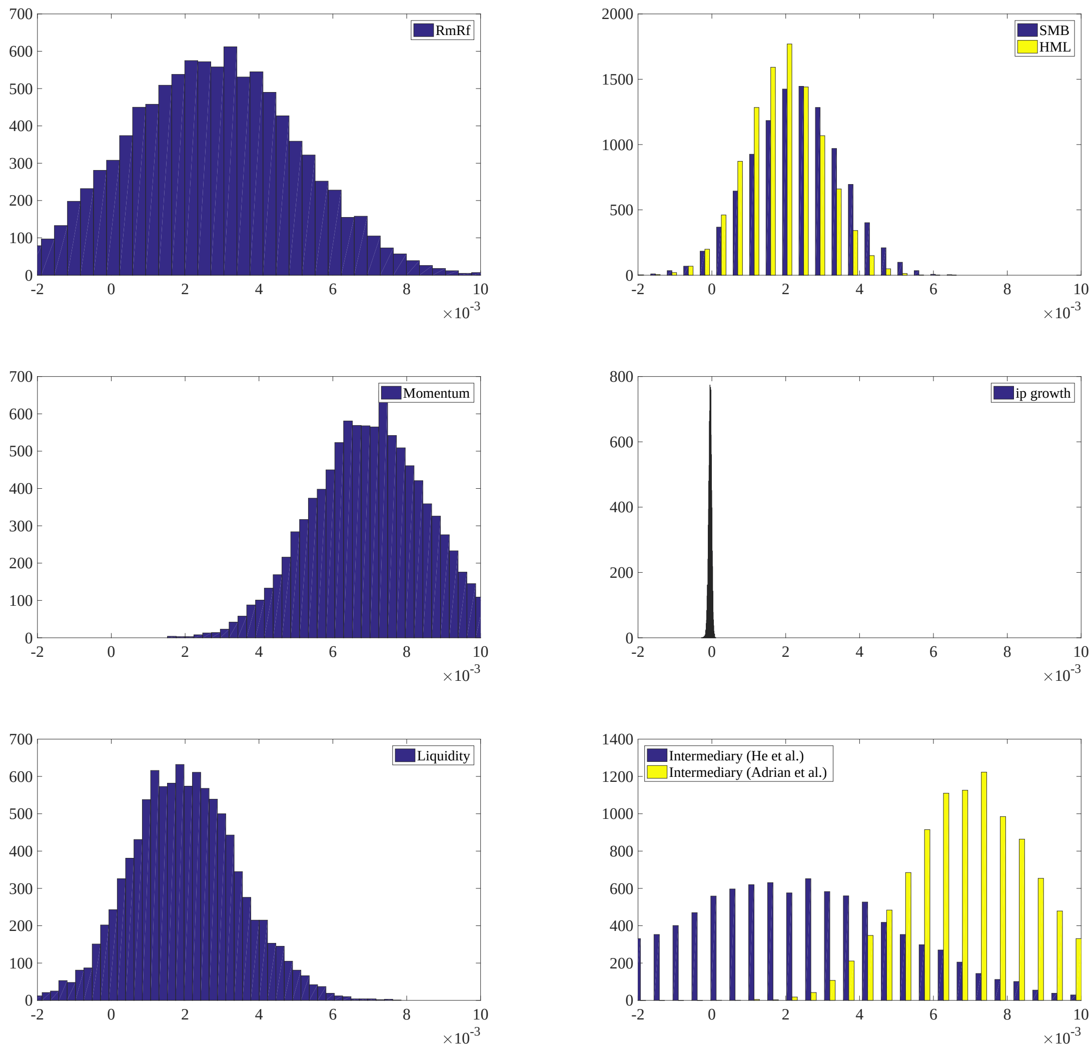
Note: This figure plots expected returns against market beta after partialing out the components explained by the other factors (using residual regression approach). The left panel uses standard two-pass regression with the Fama-French three-factor model. The right panel uses our three-pass regression approach. In each graph, the solid red line corresponds to the market risk premium estimate obtained from the time-series average return of the market portfolio; the dashed line is the two-pass regression slope. If the model is correctly specified, the two lines should coincide.

Figure 5: Robustness to the Set of Test Portfolios: Resampling Exercise



Note: This figure reports the histograms of risk premia estimated using the three-pass estimator across subsamples of the set of 202 test portfolios. We generate 10,000 subsamples by randomly drawing (without replacement) half of the portfolios from the baseline set of 202 portfolios. In each sample we estimate the risk premium of each factor using the three-pass estimator, setting $\check{p} = 6$. The histogram reports the frequency of the risk premia estimates across samples. All figures report the same range for the risk premia, between -20bp and 100bp per month.

Figure 6: Robustness to the Time Period: Resampling Exercise



Note: This figure reports the histograms of risk premia estimated using the three-pass estimator across subsamples of the time period. We generate 10,000 subsamples by randomly drawing (without replacement) half of the available time periods (using all of the portfolios available in the selected periods). In each sample we estimate the risk premium of each factor using the three-pass estimator, setting $\breve{p} = 6$. The histogram reports the frequency of the risk premia estimates across samples. All figures report the same range for the risk premia, between -20bp and 100bp per month.

Table 1: Three-Pass Regression: Empirical Results with Restricted Zero-beta Rate

Factors	Avg. Ret.	two-pass regression no controls		two-pass regression w/ R_m		two-pass regression w/ FF3		Mimicking-portf. w/ R_m		Mimicking-portf. w/ FF3		Mimicking-portf. w/ all assets		three-pass regression	
		γ	stderr	γ	stderr	γ	stderr	γ	stderr	γ	stderr	γ	stderr		
Market	0.50	0.59^{***}	(0.19)	0.59^{***}	(0.19)	0.49^{***}	(0.18)	0.50^{***}	(0.18)	0.50^{***}	(0.18)	0.52^{***}	(0.18)	0.51^{***}	(0.18)
SMB	0.23	0.63^{***}	(0.20)	0.16	(0.13)	0.13	(0.13)	0.11^{***}	(0.04)	0.23^{*}	(0.12)	0.25^{**}	(0.12)	0.24^{*}	(0.13)
HML	0.34	-1.02^{***}	(0.37)	0.43^{***}	(0.14)	0.46^{***}	(0.13)	-0.10^{**}	(0.04)	0.34^{***}	(0.11)	0.34^{***}	(0.11)	0.21^{**}	(0.11)
Momentum	0.71	-1.78^{***}	(0.65)	0.69^{***}	(0.18)	0.90^{***}	(0.17)	-0.06^{*}	(0.04)	-0.20^{***}	(0.07)	0.71^{***}	(0.17)	0.77^{***}	(0.17)
RMW	0.25	−0.16	(0.14)	−0.04	(0.14)	0.32^{***}	(0.10)	-0.05^{**}	(0.02)	-0.09^{**}	(0.04)	0.29^{***}	(0.09)	0.15^{**}	(0.06)
CMA	0.30	-0.47^{***}	(0.18)	0.38^{***}	(0.11)	0.49^{***}	(0.10)	-0.09^{***}	(0.03)	0.11^{*}	(0.06)	0.27^{***}	(0.08)	0.13^{*}	(0.08)
BAB	0.84	−0.67	(0.51)	1.11^{***}	(0.19)	1.39^{***}	(0.19)	−0.03	(0.03)	0.16^{**}	(0.06)	0.65^{***}	(0.14)	0.59^{***}	(0.10)
QMJ	0.35	-0.45^{***}	(0.15)	−0.02	(0.13)	0.37^{***}	(0.11)	-0.14^{***}	(0.05)	-0.26^{***}	(0.07)	0.22^{**}	(0.10)	0.05	(0.08)
Liquidity		2.10^{***}	(0.65)	4.54^{***}	(1.09)	2.16^{***}	(0.64)	0.21^{**}	(0.08)	0.24^{**}	(0.10)	0.23	(0.29)	0.33^{***}	(0.11)
Interm. (He)		0.73^{*}	(0.43)	−0.18	(0.57)	0.10	(0.51)	0.52^{**}	(0.23)	0.74^{***}	(0.24)	0.59^{*}	(0.33)	0.50^{**}	(0.25)
Interm. (Adrian)		1.67^{***}	(0.24)	1.68^{***}	(0.24)	1.79^{***}	(0.24)	0.12^{**}	(0.06)	0.65^{***}	(0.14)	0.95^{***}	(0.19)	0.89^{***}	(0.15)
NY temp.		-450.78^{*}	(252.27)	382.27^{**}	(162.32)	−12.56	(125.49)	−5.83	(6.35)	1.54	(9.52)	43.84	(57.48)	−5.15	(10.34)
Global temp.		−5.47	(4.80)	−3.52	(4.91)	0.72	(2.83)	−0.02	(0.10)	0.00	(0.15)	−0.03	(0.99)	−0.10	(0.16)
El Niño		73.86^{***}	(16.70)	14.22	(10.64)	-23.01^{***}	(7.04)	0.44	(0.35)	0.70	(0.48)	−1.83	(3.10)	0.32	(0.58)
Sunspots		170.38	(727.85)	2690.51^{***}	(530.06)	2044.61^{***}	(484.89)	−21.70	(20.28)	−12.79	(25.12)	−210.62	(160.69)	16.72	(27.44)
IP growth		-1.22^{***}	(0.35)	-0.29^{***}	(0.05)	-0.30^{***}	(0.05)	−0.00	(0.00)	−0.00	(0.00)	−0.02	(0.01)	−0.01	(0.00)
Macro PC 1		87.09^{***}	(21.11)	73.83^{***}	(19.86)	19.22	(14.35)	0.88	(0.55)	2.17^{**}	(1.09)	4.56	(5.22)	2.33^{*}	(1.23)
Macro PC 2		−17.96	(16.98)	−15.25	(19.49)	-20.86^{*}	(10.70)	-0.79^{*}	(0.48)	-1.31^{*}	(0.77)	3.30	(3.66)	−1.57	(0.99)
Macro PC 3		20.21	(14.04)	25.86^{**}	(10.50)	6.65	(9.50)	−0.58	(0.40)	−0.13	(0.76)	−4.21	(3.47)	−0.15	(1.04)
Cons. growth		0.44^{***}	(0.16)	0.10	(0.10)	0.07	(0.07)	0.00	(0.00)	0.01	(0.01)	0.01	(0.03)	0.01	(0.01)
Stockholder cons.		7.36^{***}	(2.40)	4.11^{***}	(1.32)	2.58^{***}	(0.72)	0.05	(0.04)	0.03	(0.06)	−0.13	(0.34)	0.12^{***}	(0.05)

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 T-bill rate: "Avg. Ret.", 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, respectively; three versions of the mimicking-portfolio estimator, projecting factors onto the market portfolio, the Fama-French three factors, and onto all 202 test assets, respectively; and the three-pass estimator we propose in this paper.

Table 2: Three-Pass Regression: Empirical Results with Unrestricted Zero-beta Rate

		two-pass regression no controls		two-pass regression w/ Rm		two-pass regression w/ FF3		three-pass regression $\check{p}=4$			three-pass regression $\check{p}=5$			three-pass regression $\check{p}=6$			p-value for g weak null
Factors	Avg. Ret.	γ	stderr	γ	stderr	γ	stderr	γ	stderr	R_g^2	γ	stderr	R_g^2	γ	stderr	R_g^2	
Market	0.50	−0.20	(0.28)	−0.20	(0.28)	−0.57**	(0.25)	0.37*	(0.20)	98.18	0.37*	(0.21)	98.93	0.35	(0.22)	99.08	0.00
SMB	0.23	0.11	(0.13)	0.19	(0.13)	0.17	(0.13)	0.23*	(0.13)	93.90	0.23*	(0.13)	94.88	0.23*	(0.13)	97.19	0.00
HML	0.34	0.23	(0.16)	0.27**	(0.13)	0.23*	(0.13)	0.21*	(0.11)	66.86	0.21*	(0.11)	67.90	0.20*	(0.11)	75.37	0.00
Momentum	0.71	0.64***	(0.19)	0.64***	(0.18)	0.81***	(0.17)	0.75***	(0.18)	91.18	0.75***	(0.18)	91.52	0.74***	(0.18)	92.19	0.00
RMW	0.25	−0.13	(0.14)	−0.09	(0.14)	0.30***	(0.10)	0.13**	(0.06)	33.93	0.13**	(0.07)	37.42	0.13*	(0.07)	45.81	0.00
CMA	0.30	0.28***	(0.10)	0.34***	(0.11)	0.37***	(0.09)	0.14*	(0.08)	44.58	0.14*	(0.08)	45.68	0.13	(0.08)	55.38	0.00
BAB	0.84	0.55**	(0.24)	1.29***	(0.28)	1.27***	(0.23)	0.55***	(0.11)	45.40	0.55***	(0.11)	45.85	0.54***	(0.12)	49.00	0.00
QMJ	0.35	0.03	(0.13)	−0.02	(0.13)	0.39***	(0.11)	0.07	(0.08)	59.63	0.07	(0.08)	63.13	0.07	(0.09)	70.43	0.00
Liquidity		0.02	(0.97)	3.58***	(1.17)	0.98	(0.64)	0.26**	(0.12)	11.99	0.26**	(0.12)	12.02	0.25**	(0.12)	12.02	0.00
Interm. (He)		0.02	(0.64)	−0.77	(0.54)	−0.22	(0.51)	0.30	(0.27)	60.10	0.29	(0.29)	62.08	0.28	(0.30)	62.13	0.00
Interm. (Adrian)		1.25***	(0.32)	1.30***	(0.33)	1.32***	(0.25)	0.79***	(0.16)	49.28	0.78***	(0.17)	49.28	0.77***	(0.17)	51.58	0.00
NY temp.		316.11	(209.09)	160.17	(156.72)	−31.38	(126.14)	−1.34	(10.61)	0.50	−2.13	(11.53)	0.50	−1.53	(10.54)	0.51	0.87
Global temp.		2.65	(4.92)	3.41	(4.66)	−3.55	(2.75)	−0.07	(0.17)	1.74	−0.08	(0.18)	1.85	−0.04	(0.18)	1.87	0.03
El Niño		−6.38	(9.46)	0.29	(9.75)	−1.92	(5.91)	0.01	(0.59)	0.97	0.09	(0.65)	1.01	0.01	(0.60)	1.01	0.66
Sunspots		1867.70***	(668.06)	1285.31***	(448.63)	1074.20***	(409.74)	25.12	(27.74)	0.71	23.17	(30.41)	0.73	20.54	(27.33)	0.87	0.57
IP growth		−0.16***	(0.05)	−0.16***	(0.05)	−0.15***	(0.04)	−0.00	(0.00)	0.30	−0.00	(0.00)	0.34	−0.00	(0.00)	0.52	0.80
Macro PC 1		32.65*	(18.36)	47.71**	(19.85)	−21.33	(13.20)	2.04	(1.25)	1.40	1.99	(1.25)	1.63	1.90	(1.24)	1.64	0.10
Macro PC 2		−21.14	(16.23)	−37.06**	(17.38)	20.84**	(10.23)	−1.29	(0.99)	1.98	−1.27	(1.00)	2.00	−1.21	(0.99)	2.02	0.06
Macro PC 3		29.25**	(12.42)	10.11	(10.75)	−13.86	(9.56)	−0.23	(1.06)	1.94	−0.25	(1.06)	3.29	−0.15	(1.12)	4.42	0.00
Cons. growth		0.11	(0.10)	0.28***	(0.10)	0.12*	(0.07)	0.01	(0.01)	3.10	0.01	(0.01)	3.12	0.01	(0.01)	3.60	0.00
Stockholder cons.		0.58	(1.40)	−0.23	(1.09)	0.45	(0.62)	0.04	(0.05)	1.33	0.07	(0.06)	2.64	0.06	(0.06)	2.66	0.22

Note: For each factor, the table reports the risk premia estimates using different methods, with unrestricted zero-beta rate (not reported in the table for reasons of space): "Avg. Ret.", 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, respectively; and three versions of the three-pass estimator we propose in this paper, using 4, 5, and 6 principal components, respectively. The three-pass results report, in addition to the estimate of the risk premium and the standard error, the R^2 of the time-series regression of the factor onto the PCs. The table does not report results with the mimicking-portfolio approach because it does not allow for unrestricted zero-beta rate. The last column in the table reports the p-value of a Wald test for the null that the observable factor g_t is weak, using $\check{p}=6$.

Table 3: Loading of Observable Factors onto Latent Factors (% of Variation Explained)

Factors	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6
Market	91.0	6.3	1.7	0.1	0.8	0.2
SMB	31.0	64.0	0.6	0.9	1.0	2.4
HML	7.0	1.3	75.5	4.9	1.4	9.9
MOM	3.1	0.3	2.0	93.5	0.4	0.7
RMW	17.2	37.4	15.4	4.1	7.6	18.3
CMA	19.8	0.0	60.6	0.1	2.0	17.5
BAB	0.9	3.6	72.7	15.4	0.9	6.4
QMJ	57.3	15.9	2.4	9.0	5.0	10.4
Liquidity	95.0	2.8	1.8	0.2	0.2	0.0
Interm. (He et al)	81.2	12.5	0.1	2.9	3.2	0.1
Interm. (Adrian et al)	20.3	6.8	52.0	16.4	0.0	4.5
NY temp.	21.7	8.6	46.5	19.7	1.1	2.4
Global temp.	0.5	85.9	5.8	0.9	5.4	1.4
El Niño	58.8	2.0	9.0	25.8	4.1	0.3
Sunspots	41.9	2.4	2.4	34.8	2.5	16.0
IP growth	11.9	3.5	2.3	39.4	8.2	34.7
Macro PC 1	58.0	2.9	24.4	0.0	13.9	0.8
Macro PC 2	81.9	13.0	2.3	0.6	1.0	1.2
Macro PC 3	23.7	1.0	16.3	3.0	30.4	25.6
Cons. growth	26.5	54.4	5.0	0.1	0.5	13.5
Stockholder cons.	17.9	23.2	2.6	6.5	49.2	0.6

Note: The table reports the decomposition of the variance of the observable factors g_t explained by the six latent factors. Each row adds up to 100%.

Table 4: Risk Premia across Asset Classes

Factors	Avg ret	202 equity γ	stderr	R^2_g	FF25 + 100 non-equity γ	stderr	R^2_g	100 non-equity γ	stderr	R^2_g
Market	0.47	0.26	(0.24)	98.85	0.73^{***}	(0.23)	89.84	0.68^{***}	(0.22)	52.37
SMB	0.18	0.17	(0.14)	94.71	0.22^{*}	(0.12)	31.13	0.15^{**}	(0.06)	2.08
HML	0.41	0.25^{*}	(0.13)	74.33	0.16^{**}	(0.07)	10.43	0.07	(0.07)	4.44
MOM	0.67	0.59^{***}	(0.20)	79.32	−0.30	(0.18)	6.49	−0.27	(0.18)	6.55
RMW	0.29	0.15^{*}	(0.08)	48.06	-0.17^{***}	(0.05)	11.27	-0.16^{***}	(0.05)	7.10
CMA	0.40	0.17^{*}	(0.09)	49.03	−0.04	(0.06)	14.17	−0.10	(0.06)	9.35
BAB	0.87	0.55^{***}	(0.12)	43.90	0.22^{**}	(0.09)	7.05	0.16	(0.12)	9.68
QMJ	0.38	0.09	(0.10)	68.82	-0.44^{***}	(0.11)	45.15	-0.41^{***}	(0.11)	26.95
Liquidity		0.21	(0.13)	11.76	0.24	(0.18)	11.45	0.21	(0.18)	5.88
Interm. (He et al)		0.25	(0.28)	62.21	1.03^{***}	(0.30)	52.14	0.94^{***}	(0.31)	34.95
Interm. (Adrian et al)		0.69^{***}	(0.16)	46.42	0.50^{***}	(0.11)	14.91	0.35^{***}	(0.09)	6.93
NY temp.		0.11	(10.81)	0.63	6.84	(22.41)	1.26	−2.39	(23.40)	1.91
Global temp.		−0.04	(0.17)	1.92	−0.18	(0.37)	0.58	−0.19	(0.42)	0.59
El Niño		0.02	(0.64)	1.48	1.75	(1.18)	1.14	1.79	(1.21)	0.95
Sunspots		26.37	(29.80)	0.90	−10.38	(46.08)	0.86	−6.54	(48.91)	0.88
IP growth		−0.00	(0.00)	0.79	0.00	(0.01)	0.96	−0.01	(0.01)	3.43
Macro PC 1		1.66	(1.27)	1.57	0.02	(2.82)	2.24	0.84	(2.69)	1.92
Macro PC 2		−0.81	(0.90)	1.49	−1.95	(1.69)	2.64	−1.70	(1.79)	2.32
Macro PC 3		−0.55	(1.14)	4.96	-3.80^{**}	(1.55)	7.36	-3.68^{**}	(1.87)	7.09
Cons. growth		0.00	(0.01)	3.30	0.03	(0.03)	3.69	0.04^{*}	(0.03)	2.31
Stockholder cons.		0.05	(0.06)	2.46	0.12	(0.20)	2.50	0.13	(0.23)	3.11

Note: The table reports the results of risk premia estimation for various models using our three-pass procedure. The left side of the panel uses 202 equity portfolios as test assets. The center panel uses the 25 Fama-French portfolios plus 100 non-equity assets. The right panel uses only the 100 non-equity assets. Sample period covers 1970-2012. The number of factors $\check{p}$ used is 5 for the left and middle panel and 6 for the right panel.

Table 5: Alternative Objective Function for Latent Factor Estimation

Factors	PCA ($w = 0$)			CK ($w = 1$)		
	$\breve{p} = 4$	$\breve{p} = 5$	$\breve{p} = 6$	$\breve{p} = 4$	$\breve{p} = 5$	$\breve{p} = 6$
Market	0.51	0.52	0.52	0.51	0.52	0.52
SMB	0.24	0.23	0.23	0.24	0.24	0.23
HML	0.21	0.21	0.22	0.22	0.22	0.22
MOM	0.77	0.76	0.76	0.81	0.80	0.81
RMW	0.15	0.15	0.15	0.15	0.16	0.15
CMA	0.13	0.13	0.13	0.14	0.14	0.14
BAB	0.59	0.60	0.60	0.61	0.62	0.62
QMJ	0.05	0.06	0.05	0.06	0.07	0.06
Liquidity	0.33	0.33	0.33	0.34	0.34	0.34
Interm. (He et al)	0.50	0.51	0.51	0.49	0.51	0.51
Interm. (Adrian et al)	0.89	0.89	0.90	0.92	0.92	0.93
NY temp.	-5.15	-5.66	-5.70	-5.07	-5.53	-5.58
Global temp.	-0.10	-0.05	-0.05	-0.09	-0.05	-0.05
El Niño	0.32	0.39	0.39	0.31	0.38	0.38
Sunspots	16.72	13.74	13.29	18.31	15.16	14.72
IP growth	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01
Macro PC 1	2.33	2.38	2.38	2.37	2.42	2.42
Macro PC 2	-1.57	-1.58	-1.58	-1.59	-1.60	-1.60
Macro PC 3	-0.15	-0.07	-0.06	-0.13	-0.05	-0.05
Cons. growth	0.01	0.01	0.01	0.01	0.01	0.01
Stockholder cons.	0.12	0.17	0.17	0.13	0.18	0.18

Note: The table reports the results of risk premia estimation for various models using our three-pass procedure, using different penalty functions to extract the latent factors. We report results for $\breve{p} = 4, 5, 6$. The left panel uses standard PCA as in our baseline case. The right panel extracts factors using the Connor and Korajczyk (1986, 1988) approach, which gives weight to the cross-sectional R^2. Following Connor and Korajczyk (1986, 1988), the zero-beta rate is imposed to be equal to the risk-free rate. Sample period covers 1970-2012.

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 $\|A\|_{\mathrm{F}}$ 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, $\max_j \sum_i |a_{ij}|$, $\max_i \sum_j |a_{ij}|$, $\sqrt{\lambda_{\max}(A^\intercal A)}$, and $\sqrt{\mathrm{Tr}(A^\intercal A)}$, 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.

Let $(\mathrm{P}, \Omega, \mathcal{F})$ be the probability space. K is a generic constant that may change from line to line. We say a sequence of centered multivariate random variables $\{y_t\}_{t\geq 1}$ satisfy the exponential-type tail condition, if there exist some constants a and b, such that $\mathrm{P}\left(|y_{it}| > y\right) \leq \exp\{-(y/b)^a\}$, for all i and t. We say a sequence of random variables satisfy the strong mixing condition if the mixing coefficients satisfy $\alpha_m \leq \exp(-Km^c)$, for $m = 1, 2, \ldots$, and some constants $c > 0$ and $K > 0$.

For clarity, we restate the assumptions on the dynamics of returns and factors, i.e., (1) and (2), introduced in the main text:

Assumption A.1. *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:*

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

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

Assumption A.2. *There is an observable $d \times 1$ vector, g_t, of factor proxies, which satisfies:*

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

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

Next, we impose some restrictive assumptions, which are only used in Proposition 1 and Section 3.4 to illustrate the intuition of our result and the connection between the two-pass cross-sectionals regression and the factor mimicking portfolios. Our asymptotic analysis below does not rely on this assumption.

Assumption A.3. *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 $\breve{u}_t := w^\intercal r_t$, i.e., $\breve{\beta}^{z,u}$, satisfy the following conditions: $\lambda_{\min}^{-1}(\Sigma^v) = O_p(1)$, $\left\|\breve{\beta}^{z,u}\breve{\beta}\right\|_{\mathrm{MAX}} = O_p(1)$, $\lambda_{\min}^{-1}(\breve{\beta}^\intercal\breve{\beta}) = O_p(1)$, $\lambda_{\max}(\breve{\Sigma}^u) = O_p(s_n n^{-1})$, where $\breve{\beta} := w^\intercal \beta$, $s_n = o_p(n)$.*

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 $\breve{\beta}^\intercal\breve{\beta}$ 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 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 A.3 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: $w = \widehat{\beta}(\widehat{\beta}^\intercal\widehat{\beta})^{-1}$.

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 A.3.

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 A.4. *There exists a positive constant K, such that for all n and T,*

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

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

Assumption A.4 is similar to part of Assumption C in Bai (2003), which imposes restrictions on the cross-sectional dependence and heteroskedasticity of u_t.

Assumption A.5. *The factor innovation V satisfies:*

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

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 A.5 imposes rather weak conditions on the time series behavior of the factors. It certainly holds if factors are stationary and satisfy the exponential-type tail condition and the strong mixing condition, see, Fan et al. (2013).

Assumption A.6. *The factor loadings matrix* β *satisfies*

$$\left\| n^{-1}\beta^\intercal\beta - \Sigma^\beta \right\| = o_p(1), \quad as \quad n \to \infty,$$

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 A.7. *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*:*

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

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

This assumption can be shown from the 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 A.8. *The residual innovation* Z *satisfies:*

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

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

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

Similar to Assumption A.5, Assumption A.8 holds if z_t is stationary, and satisfies the exponential-type tail condition and some strong mixing condition. It is more general than the i.i.d. assumption on z_t, which also applies to non-tradable factor proxies in the empirical applications.

Assumption A.9. *For all* n *and* T*, and* $i, j \leq p$*,* $l \leq d$*, the following moment conditions hold:*

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

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

Assumption A.9 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 A.10. *For all n and T, and $l \leq d$, $j \leq p$, the following moment conditions hold:*

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

Similar to Assumption A.9, Assumption A.10 restricts the dependence between the idiosyncratic component u_t and the projection residual z_t. If z_t, u_t, and β are independent, (i) - (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 A.1 and A.2. It is thereby reasonable to allow for dependence between z_t and u_t. For non-tradable factors, z_ts can also be correlated with u_t in general.

Assumption A.11. *As $T \to \infty$, the following joint central limit theorem holds:*

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

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

$$\Pi_{11} = \lim_{T\to\infty}\frac{1}{T}\mathrm{E}\left(\mathrm{vec}(ZV^\intercal)\mathrm{vec}(ZV^\intercal)^\intercal\right),$$
$$\Pi_{12} = \lim_{T\to\infty}\frac{1}{T}\mathrm{E}\left(\mathrm{vec}(ZV^\intercal)\iota_T^\intercal V^\intercal\right),$$
$$\Pi_{22} = \lim_{T\to\infty}\frac{1}{T}\mathrm{E}\left(V\iota_T\iota_T^\intercal V^\intercal\right).$$

Assumption A.11 describes the joint asymptotic distribution of $ZV^\intercal$ 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, (e.g., Theorem 5.20 of White (2000)). Needless to say, it is stronger than Assumption A.5, which is sufficient for identification and consistency.

Assumption A.12 restates the assumptions on the dynamics of (12) in the main text.

Assumption A.12. *Suppose the cross-section of asset returns r_t follows*

$$r_t = \alpha + \iota_n\gamma_0 + \beta\gamma + \beta v_t + u_t,$$

where the cross-sectional pricing error α is i.i.d., independent of β, u and v, with mean 0, standard deviation $\sigma^\alpha > 0$, and a finite fourth moment.

Assumption A.13. *There exists a $p \times 1$ vector β_0, such that $\left\|n^{-1}\beta^\intercal\iota_n - \beta_0\right\|_{\mathrm{MAX}} = o_p(1)$. Moreover, the matrix*

$$\begin{pmatrix} 1 & \beta_0^\intercal \\ \beta_0 & \Sigma^\beta \end{pmatrix} \quad \textit{is of full rank.}$$

This assumption is imposed only in the case when we allow for and estimate the zero-beta rate. The rank condition ensures that in the limit the factor loadings and ι_n are not perfectly correlated in the cross section, and in particular, that the zero-beta rate γ_0 is identifiable.

Assumption A.14. *Define, for any $i, i' \leq n$, $t, t' \leq T$,*

$$\mathrm{E}(u_{it} u_{i't}) = \sigma_{ii',t}, \quad \textit{and} \quad \mathrm{E}(u_{it} u_{i't'}) = \sigma_{ii',tt'}.$$

The following moment conditions hold, for all n and T, and $i, j \leq p$, $l \leq d$,

$$(i) \quad \max_{1 \leq t \leq T} |\sigma_{ii',t}| \leq |\sigma_{ii'}|, \quad \textit{for some } \sigma_{ii'}. \quad \textit{In addition, } n^{-1} \sum_{i=1}^{n} \sum_{i'=1}^{n} |\sigma_{ii'}| \leq K.$$

$$(ii) \quad n^{-1} T^{-1} \sum_{i=1}^{n} \sum_{i'=1}^{n} \sum_{t=1}^{T} \sum_{t'=1}^{T} |\sigma_{ii',tt'}| \leq K.$$

$$(iii) \quad \mathrm{E}\left(\sum_{t=1}^{T} \sum_{k=1}^{n} v_{jt} u_{kt} \right)^2 \leq KnT.$$

Assumption A.14 imposes restrictions on the time series dependence of u_t. Stationarity of u_t is not required. Eigenvalues of the residual covariance matrices $\mathrm{E}(u_t u_t^\intercal)$ are not necessarily bounded. Assumption A.14 is similar to part of Assumption C in Bai (2003). We only need it when we allow for a zero-beta rate.

Assumption A.15. *The following conditions hold:*

$$(i) \quad \sum_{t'=1}^{T} |\gamma_{n,tt'}| \leq K, \quad \textit{for all } t.$$

$$(ii) \quad \sum_{i'=1}^{n} |\sigma_{ii'}| \leq K, \quad \textit{for all } i.$$

This assumption is identical to Assumption E in Bai (2003). It restricts the eigenvalues of $\mathrm{E}(u_t u_t^\intercal)$ and $\mathrm{E}(u_t^\intercal u_t)$ to be bounded as the dimension increases, because the $\mathbb{L}_\infty$-norm is stronger than the operator norm for symmetric matrices. We need this to bound the estimation error of factors uniformly over t, which in turn leads to the consistency of the asymptotic variance estimation.

For the same reason, we also need Assumption A.16, which Fan et al. (2011) and Fan et al. (2015) also adopt:

Assumption A.16. *The sequence of $\{u_t, v_t, z_t\}_{t \geq 1}$ is jointly strong mixing, and satisfies the exponential-type tail condition. Moreover, for all $t', t \leq T$,*

$$\mathrm{E}\left(u_t^\intercal u_{t'} - \mathrm{E} u_t^\intercal u_{t'}\right)^4 \leq Kn^2, \quad \mathrm{E}\, \|\beta^\intercal u_t\|^4 \leq Kn^2.$$

B Additional Results

B.1 Mimicking Portfolios

Proposition 1. *Suppose Assumptions A.1 - A.3 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 $n \to \infty$.*

B.2 Limiting Distribution of the Denoised Factors

As discussed above, our framework allows for measurement error in the observable factor proxies g. Theorem 4 in the main text indicates that we can separate the error from the factors using the extracted PCs. Moreover, we can conduct inference on $\widehat{g}_t$, provided an additional assumption:

Assumption B.17. *For each t, as $n \to \infty$,*

$$n^{-1/2}\beta^\intercal u_t \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \Omega_t\right),$$

where, writing $\beta = (\beta_1 : \beta_2 : \ldots : \beta_n)^\intercal$,

$$\Omega_t = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n}\sum_{i'=1}^{n}\beta_i\beta_{i'}^\intercal \mathrm{E}(u_{it}u_{i't}). \tag{B.1}$$

Assumption B.17 is identical to Assumption F3 in Bai (2003), which is used to describe the asymptotic distribution of the estimated factors at each point in time.

Theorem 8. *Under Assumptions A.2, and A.4 –A.12, A.15, A.16, and B.17, and suppose that $\widehat{p} \xrightarrow{p} p$, then as n, $T \to \infty$, we have*

$$\Psi_t^{-1/2}\left(\widehat{g}_t - \eta v_t\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \mathbb{I}_d),$$

where $\Psi_t = T^{-1}\Psi_{1t} + n^{-1}\Psi_{2t}$,

$$\begin{aligned}
\Psi_{1t} =& \Big\{ \left(v_t^\intercal \left(\Sigma^v\right)^{-1} \otimes \mathbb{I}_d\right) \Pi_{11} \left(\left(\Sigma^v\right)^{-1} v_t \otimes \mathbb{I}_d\right) - \left(v_t^\intercal \left(\Sigma^v\right)^{-1} \otimes \mathbb{I}_d\right) \Pi_{12}\eta^\intercal \\
&- \eta\Pi_{12}^\intercal \left(\left(\Sigma^v\right)^{-1} v_t \otimes \mathbb{I}_d\right) + \eta\Pi_{22}\eta^\intercal \Big\}, \quad \textit{and} \\
\Psi_{2t} =& \eta\left(\Sigma^\beta\right)^{-1}\Omega_t\left(\Sigma^\beta\right)^{-1}\eta^\intercal.
\end{aligned}$$

In Bai (2003), the latent factors can be estimated at the $n^{-1/2}$-rate, provided that $n^{1/2}T^{-1} \to 0$. In our setting, the estimation error consists of the errors in estimating $\widehat{\eta}$ and $\widehat{v}_t$. Because $\widehat{\eta}$ is estimated up to a $T^{-1/2}$-rate error which dominates T^{-1} terms, the convergence rate of $\widehat{g}_t$ does not rely on any relationship between n and T.

B.3 Asymptotic Covariance Matrix

To estimate the asymptotic covariance matrices Ψ_{1t} and Ψ_{2t} in Theorem 8, we can simply replace v_t, Σ^v, Π_{11}, Π_{12}, Π_{22}, η, Σ^β by their sample analogues, $\widehat{v}_t$, $\widehat{\Sigma}^v$, $\widehat{\Pi}_{11}$, $\widehat{\Pi}_{12}$, $\widehat{\Pi}_{22}$, $\widehat{\eta}$, $\widehat{\Sigma}^\beta$, in the $\widehat{\Psi}_{1t}$ and $\widehat{\Psi}_{2t}$ constructions. With respect to Ω_t, we need an additional assumption:

Assumption B.18. *The innovation u_{it} is stationary, and its covariance matrix Σ^u is sparse, i.e., there exists some $h \in [0, 1/2)$, with $\omega_T = (\log n)^{1/2} T^{-1/2} + n^{-1/2}$, such that*

$$s_n = \max_{1 \le i \le n} \sum_{i'=1}^{n} |\Sigma^u_{ii'}|^h, \quad \textit{where} \quad s_n = o_p\left(\left(\omega_T^{1-h} + n^{-1} + T^{-1}\right)^{-1}\right).$$

Given this assumption, (B.1) and its estimator can be rewritten as

$$\Omega = \lim_{n \to \infty} \frac{1}{n} \beta^\intercal \Sigma^u \beta, \quad \text{and} \quad \widehat{\Omega}_t = \widehat{\Omega} = \frac{1}{n} \widehat{\beta}^\intercal \widehat{\Sigma}^u \widehat{\beta}, \tag{B.2}$$

where, for $1 \le i, i' \le n$,

$$\widehat{\Sigma}^u_{ii'} = \begin{cases} \widetilde{\Sigma}^u_{ii}, & i = i' \\ s_{ii'}(\Sigma^u_{ii'}), & i \neq i' \end{cases}, \quad \widetilde{\Sigma}^u = \frac{1}{T} \sum_{t=1}^{T} \widehat{u}_t \widehat{u}_t^\intercal,$$

and $s_{ii'}(z) : \mathbb{R} \to \mathbb{R}$ is a general thresholding function with an entry dependent threshold $\tau_{ii'}$ such that (i) $s_{ii'}(z) = 0$ if $|z| < \tau_{ii'}$; (ii) $|s_{ii'}(z) - z| \le \tau_{ii'}$; and (iii) $|s_{ii'}(z) - z| \le a\tau^2_{ii'}$, if $|z| > b\tau_{ii'}$, with some $a > 0$ and $b > 1$. $\tau_{ii'}$ can be chosen as:

$$\tau_{ii'} = c(\widehat{\Sigma}_{ii} \widehat{\Sigma}_{i'i'})^{1/2} \omega_T, \quad \text{for some constant } c > 0.$$

Bai and Liao (2013) adopt a similar estimator of Σ^u for efficient estimation of factor models.

With estimators of their components constructed, our estimators for Ψ_{1t} and Ψ_{2t} are defined as:

$$\begin{aligned}
\widehat{\Psi}_{1t} = & T^{-1} \Big\{ \left(\widehat{v}_t^\intercal (\widehat{\Sigma}^v)^{-1} \otimes \mathbb{I}_d\right) \widehat{\Pi}_{11} \left((\widehat{\Sigma}^v)^{-1} \widehat{v}_t \otimes \mathbb{I}_d\right) - \left(\widehat{v}_t^\intercal (\widehat{\Sigma}^v)^{-1} \otimes \mathbb{I}_d\right) \widehat{\Pi}_{12} \widehat{\eta}^\intercal - \widehat{\eta} \widehat{\Pi}_{12}^\intercal \left((\widehat{\Sigma}^v)^{-1} \widehat{v}_t \otimes \mathbb{I}_d\right) \\
& + \widehat{\eta} \widehat{\Pi}_{22} \widehat{\eta}^\intercal \Big\}, \\
\widehat{\Psi}_{2t} = & n^{-1} \widehat{\eta} \left(\widehat{\Sigma}^\beta\right)^{-1} \widehat{\Omega}_t \left(\widehat{\Sigma}^\beta\right)^{-1} \widehat{\eta}^\intercal,
\end{aligned}$$

where $\widehat{\Omega}_t$ is given by (B.2). The next theorem establishes the desired consistency of $\widehat{\Psi}_{1t}$ and $\widehat{\Psi}_{2t}$:

Theorem 9. *Under Assumptions A.2, A.3 – A.16, B.17, B.18, we have*

$$\widehat{\Psi}_{1t} - \Psi_{1t} \xrightarrow{p} 0, \quad \textit{and} \quad \widehat{\Psi}_{2t} - \Psi_{2t} \xrightarrow{p} 0.$$

C Simulations

In this section, we study the finite sample performance of our inference procedure using Monte Carlo simulations. We consider a five-factor data-generating process, where the latent factors are calibrated to match the de-noised five Fama-French factors (RmRf, SMB, HML, RMW, CMA, see Fama and French (2015)) from our empirical study in the next section. Suppose that we do not observe all five factors, but instead some noisy version of the three Fama-French factors (RmRf, SMB, HML, see Fama and French (1993)), plus a potentially spurious macro factor calibrated to industrial production growth (IP) in our empirical study. Our simulations, therefore, include both the issue of omitted factors and that of a spurious factor. We calibrate the parameters γ_0, γ, η, Σ^v, Σ^z, Σ^u, $(\sigma^\alpha)^2$, β_0, and Σ^β to exactly match their counterparts in the data (in our estimation of the Fama-French five-factor model). We then generate the realizations of v_t, z_t, u_t, α, and β from a multivariate normal using the calibrated means and covariances.

We report in Tables C1, C2, and C3 the bias and the root-mean-square error of the estimates using standard two-pass regressions and our three-pass approach. We choose different numbers of factors to estimate the model, $\breve{p} = 4$, 5, and 6, whereas the true value is 5. The five rows in each panel provide the results for the zero-beta rate, RmRf, SMB, HML, and IP, respectively. Throughout these tables, we find that the three-pass estimators with $\breve{p} = 5$ or 6 outperform the other estimators, in particular when n and T are large. By comparison, the two-pass estimates have substantial biases. Moreover, the biases for the market factor premium are substantial and negative even when n and T are large. The three-pass estimator with $\breve{p} = 4$ has an obvious bias, compared to the cases with $\breve{p} = 5$ and 6, because an omitted-factor problem still affects it (4 factors do not span the entire factor space).

We then plot in Figure C1 the histograms of the standardized risk premia estimates using Fama-MacBeth standard errors for the two-pass estimator (right column) and the estimated asymptotic standard errors for the three-pass method with $\breve{p} = 5$ (left column).[18] The histograms on the right deviate substantially from the standard normal distribution, whereas those on the left match the normal distribution very well, which verifies our central limit results despite a small sample size $T = 240$ and a moderate dimension $n = 200$. There exist some small higher order biases for γ_0, which would disappear with a larger n and T in simulations not included here.

Next, we report in Table C4 the estimated number of factors. We choose $\phi(n, T) = K(\log n + \log T)(n^{-1/2} + T^{-1/2})$, where $K = 0.5 \times \widehat{\lambda}$, $\widehat{\lambda}$ is the median of the first $p_{\max}$ eigenvalues of $n^{-1}T^{-1}\bar{R}^\intercal \bar{R}$. The median eigenvalue helps adjust the magnitude of the penalty function for better finite sample accuracy. Although the estimator is consistent, it cannot give the true number of factors without error, in particular when n or T is small, potentially due to the ad-hoc choice of tuning parameters.[19] In the empirical study, we apply this estimator of p and select slightly more factors to ensure the robustness of the estimates, as suggested by Theorem 5.

[18] We have also implemented the standard errors of the two-pass estimators using the formula given by Bai and Zhou (2015), which provides desirable performance when both n and T are large. However, we do not find substantial differences compared to the Fama-MacBeth method, so we omit those histograms.

[19] The eigenvalue ratio-based test by Ahn and Horenstein (2013) does not work well in our simulation setting because the first eigenvalue dominates the rest by a wide margin, so that their test often suggests 1 factor.

Then we evaluate the size and power properties of the proposed test in Section 4.7. To check the size control, we create a purely noisy factor with $\eta = 0$ and variance calibrated to be the average variance of the four factors we consider. The left panel of Figure C2 plots the histogram of the test statistic under the null against the density of a χ^2-distribution with 5 degrees of freedom. The distributions match reasonably well given $T = 240$. To evaluate the power, we plot on the right panel of Figure C2 the average rejection probabilities against the signal-to-noise strength measured by R_g^2 for a sequence of factors. These factors only load on the market factor, and share the same total variance calibrated to be the average variance as above, with different R_g^2s ranging from 0 to 10%. As expected, we observe the rejection probability elevates to 1 as R_g^2 increases.

Finally, we compare the performance of these estimators with the mimicking portfolio estimators under more restrictive dynamics in which γ_0 is known and $\alpha = 0$. So we estimate the model using (8) and excess returns. We consider two sets of mimicking portfolios: one set (MP3) uses three portfolios as spanning assets to project factors, where portfolio weights are exactly proportional to the market, SMB, and HML beta. Using three base assets clearly leads to an omitted variable problem because these three assets cannot span the space of five factors. The second set of mimicking portfolios (MP) uses all assets as basis assets for projection. There is no omitted variable bias in this case as we prove in Proposition 1, but these estimators are not as efficient as the three-pass estimators. They become infeasible when $n > T$. Figure C3 verifies these statements. Indeed, the deviation from normality is clearly visible for all estimators but ours. MP3 and two-pass estimates show visible biases whereas MP estimates display distortion due to the curse of dimensionality (n is of a similar scale to T). Tables C5 - C7 further illustrate that the RMSEs of the mimicking portfolio estimators are often larger than those of the three-pass estimators, due to large biases of MP3 and large variances of MP.

Overall, the three-pass estimators outperform the two-pass and mimicking portfolio estimators by a large margin in many cases. The MP estimator using all assets ranks the second, despite being infeasible when n is greater than T. The biases in the two-pass and MP3 are substantial, yet they are unfortunately the most common choices in the empirical literature.

Table C1: Simulation Results for $n = 50$

			Two-Pass Estimator		Three-Pass Estimators					
					$\breve{p} = 4$		$\breve{p} = 5$		$\breve{p} = 6$	
T	Param	True	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
120	γ_0	0.546	0.092	0.201	-0.002	0.177	0.012	0.170	0.021	0.168
	RmRf	0.372	-0.097	0.440	-0.009	0.420	-0.010	0.418	-0.015	0.419
	SMB	0.229	-0.092	0.311	-0.040	0.268	-0.042	0.269	-0.042	0.270
	HML	0.209	0.198	0.397	0.008	0.207	-0.009	0.209	-0.012	0.211
	IP	-0.003	-0.008	0.106	0.001	0.009	0.001	0.010	0.001	0.010
240	γ_0	0.546	0.186	0.227	0.102	0.161	0.047	0.130	0.041	0.127
	RmRf	0.372	-0.186	0.361	-0.074	0.308	-0.038	0.300	-0.033	0.299
	SMB	0.229	-0.084	0.233	-0.066	0.204	-0.023	0.195	-0.020	0.195
	HML	0.209	0.089	0.284	-0.065	0.158	-0.044	0.153	-0.042	0.154
	IP	-0.003	-0.031	0.141	0.000	0.006	0.000	0.007	0.000	0.007
480	γ_0	0.546	0.144	0.175	0.043	0.104	0.037	0.101	0.035	0.101
	RmRf	0.372	-0.118	0.248	-0.045	0.216	-0.040	0.215	-0.037	0.214
	SMB	0.229	-0.090	0.184	-0.021	0.143	-0.018	0.143	-0.018	0.143
	HML	0.209	0.080	0.227	-0.030	0.117	-0.027	0.116	-0.026	0.116
	IP	-0.003	-0.036	0.174	0.001	0.004	0.001	0.004	0.001	0.005

Note: In this table, we report the bias (Column "Bias") and the root-mean-square error (Column "RMSE") of the zero-beta rate and risk premia estimates using two-pass and three-pass estimators with $\breve{p} = 4$, 5, and 6, for $n = 50$, and $T = 120$, 240, and 480, respectively. The true data-generating process has five factors, and the parameters are calibrated based on the de-noised five Fama-French factors (RmRf, SMB, HML, RMW, and CMA). The true zero-beta rate is 0.546, and the true risk premia of four noisy yet observed factors (RmRf, SMB, HML, and IP) are provided in the "True" column. All numbers are in percentages.

Table C2: Simulation Results for $n = 100$

			Two-Pass Estimator		Three-Pass Estimators					
					$\breve{p} = 4$		$\breve{p} = 5$		$\breve{p} = 6$	
T	Param	True	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
120	γ_0	0.546	0.215	0.255	0.051	0.120	0.037	0.109	0.034	0.107
	RmRf	0.372	-0.184	0.457	-0.020	0.404	-0.019	0.405	-0.017	0.405
	SMB	0.229	-0.088	0.309	-0.033	0.272	-0.019	0.272	-0.017	0.272
	HML	0.209	0.011	0.330	-0.055	0.213	-0.038	0.211	-0.036	0.212
	IP	-0.003	-0.015	0.102	0.000	0.010	0.000	0.010	0.000	0.011
240	γ_0	0.546	0.219	0.242	0.010	0.080	0.012	0.080	0.015	0.080
	RmRf	0.372	-0.198	0.363	0.004	0.292	-0.008	0.293	-0.011	0.293
	SMB	0.229	-0.088	0.229	-0.016	0.193	-0.009	0.192	-0.008	0.192
	HML	0.209	0.112	0.278	-0.035	0.157	-0.018	0.157	-0.014	0.157
	IP	-0.003	-0.029	0.137	0.000	0.007	0.000	0.007	0.000	0.007
480	γ_0	0.546	0.191	0.207	0.014	0.068	0.011	0.067	0.008	0.067
	RmRf	0.372	-0.155	0.263	-0.006	0.204	-0.012	0.205	-0.010	0.205
	SMB	0.229	-0.167	0.230	-0.022	0.142	-0.014	0.141	-0.013	0.141
	HML	0.209	0.085	0.222	-0.019	0.112	-0.012	0.112	-0.010	0.112
	IP	-0.003	-0.056	0.193	0.000	0.005	0.000	0.005	0.000	0.005

Note: In this table, we report the bias (Column "Bias") and the root-mean-square error (Column "RMSE") of the zero-beta rate and risk premia estimates using two-pass and three-pass estimators with $\breve{p} = 4$, 5, and 6, for $n = 100$, and $T = 120$, 240, and 480, respectively. The true data-generating process has five factors, and the parameters are calibrated based on the de-noised five Fama-French factors (RmRf, SMB, HML, RMW, and CMA). The true zero-beta rate is 0.546, and the true risk premia of four noisy yet observed factors (RmRf, SMB, HML, and IP) are provided in the "True" column. All numbers are in percentages.

Table C3: Simulation Results for $n = 200$

T	Param	True	Two-Pass Estimator Bias	Two-Pass Estimator RMSE	Three-Pass Estimators $\breve{p}=4$ Bias	$\breve{p}=4$ RMSE	$\breve{p}=5$ Bias	$\breve{p}=5$ RMSE	$\breve{p}=6$ Bias	$\breve{p}=6$ RMSE
120	γ_0	0.546	0.203	0.227	0.037	0.085	0.028	0.069	0.028	0.069
	RmRf	0.372	-0.181	0.458	-0.022	0.409	-0.016	0.407	-0.016	0.407
	SMB	0.229	-0.078	0.298	-0.010	0.268	-0.004	0.269	-0.003	0.269
	HML	0.209	0.119	0.343	-0.017	0.214	-0.017	0.216	-0.016	0.216
	IP	-0.003	-0.013	0.097	0.000	0.010	0.000	0.011	0.000	0.011
240	γ_0	0.546	0.167	0.181	0.016	0.053	0.011	0.049	0.010	0.049
	RmRf	0.372	-0.154	0.338	0.006	0.291	-0.004	0.292	-0.003	0.292
	SMB	0.229	-0.102	0.237	-0.025	0.198	-0.004	0.198	-0.004	0.198
	HML	0.209	0.135	0.283	-0.017	0.154	-0.013	0.154	-0.013	0.154
	IP	-0.003	-0.033	0.142	0.000	0.007	0.000	0.007	0.000	0.007
480	γ_0	0.546	0.178	0.189	0.011	0.045	0.007	0.044	0.006	0.043
	RmRf	0.372	-0.148	0.255	0.007	0.201	-0.008	0.202	-0.008	0.202
	SMB	0.229	-0.146	0.213	-0.031	0.143	-0.008	0.140	-0.008	0.140
	HML	0.209	0.196	0.286	-0.017	0.114	-0.002	0.113	-0.002	0.114
	IP	-0.003	-0.076	0.203	0.000	0.005	0.000	0.005	0.000	0.005

Note: In this table, we report the bias (Column "Bias") and the root-mean-square error (Column "RMSE") of the zero-beta rate and risk premia estimates using two-pass and three-pass estimators with $\breve{p} = 4$, 5, and 6, for $n = 200$, and $T = 120$, 240, and 480, respectively. The true data-generating process has five factors, and the parameters are calibrated based on the de-noised five Fama-French factors (RmRf, SMB, HML, RMW, and CMA). The true zero-beta rate is 0.546, and the true risk premia of four noisy yet observed factors (RmRf, SMB, HML, and IP) are provided in the "True" column. All numbers are in percentages.

Table C4: Simulation Results for the Number of Factors

T	$n=50$ Median	$n=50$ Stderr	$n=100$ Median	$n=100$ Stderr	$n=200$ Median	$n=200$ Stderr
120	3	0.35	4	0.62	5	0.08
240	4	0.52	5	0.53	5	0.20
480	3	0.31	5	0.25	5	0.39

Note: In this table, we report the median (Column "Median") and the standard error (Column "Stderr") of the estimates for the number of factors. The true number of factors in the data generating process is five.

Table C5: Simulation Results for $n = 50$

T	Param	True	Two-Pass Estimator Bias	Two-Pass Estimator RMSE	Mimicking Portfolios using three assets Bias	Mimicking Portfolios using three assets RMSE	Mimicking Portfolios using all assets Bias	Mimicking Portfolios using all assets RMSE
	RmRf	0.372	-0.001	0.419	0.015	0.398	0.003	0.406
120	SMB	0.229	-0.067	0.304	0.162	0.302	-0.002	0.282
	HML	0.209	0.148	0.370	-0.308	0.322	-0.018	0.254
	IP	-0.003	-0.008	0.091	0.002	0.008	0.000	0.027
	RmRf	0.372	-0.172	0.345	-0.019	0.275	-0.012	0.282
240	SMB	0.229	-0.052	0.224	0.119	0.214	0.005	0.191
	HML	0.209	0.282	0.379	-0.273	0.279	-0.007	0.159
	IP	-0.003	-0.011	0.123	0.002	0.006	0.001	0.013
	RmRf	0.372	-0.220	0.309	-0.041	0.206	-0.019	0.205
480	SMB	0.229	-0.091	0.178	0.071	0.151	-0.010	0.139
	HML	0.209	0.307	0.366	-0.262	0.266	0.004	0.110
	IP	-0.003	-0.020	0.140	0.003	0.004	0.000	0.007

Three-Pass Estimators

T	Param	True	$\breve{p} = 4$ Bias	$\breve{p} = 4$ RMSE	$\breve{p} = 5$ Bias	$\breve{p} = 5$ RMSE	$\breve{p} = 6$ Bias	$\breve{p} = 6$ RMSE
	RmRf	0.372	0.019	0.401	0.018	0.402	0.015	0.402
120	SMB	0.229	-0.018	0.261	-0.014	0.262	-0.008	0.264
	HML	0.209	-0.056	0.195	-0.053	0.198	-0.050	0.200
	IP	-0.003	0.001	0.008	0.001	0.009	0.001	0.010
	RmRf	0.372	-0.079	0.288	-0.056	0.284	-0.043	0.281
240	SMB	0.229	0.029	0.184	0.030	0.186	0.022	0.186
	HML	0.209	0.044	0.142	0.013	0.139	0.004	0.139
	IP	-0.003	0.001	0.006	0.001	0.006	0.001	0.007
	RmRf	0.372	-0.088	0.221	-0.068	0.215	-0.043	0.208
480	SMB	0.229	-0.016	0.137	-0.020	0.138	-0.025	0.139
	HML	0.209	0.090	0.133	0.070	0.125	0.044	0.111
	IP	-0.003	0.001	0.004	0.001	0.004	0.001	0.004

Note: In this table, we report the bias (Column "Bias") and the root-mean-square error (Column "RMSE") of the risk premia estimates using mimicking portfolios with three assets or all assets, two-pass and three-pass estimators with $\breve{p} = 4$, 5, and 6, for $n = 50$, and $T = 120$, 240, and 480, respectively. The true data-generating process has five factors, and the parameters are calibrated based on the de-noised five Fama-French factors (RmRf, SMB, HML, RMW, and CMA). In this setting, there is no α. The true zero-beta rate is 0, and the true risk premia of four noisy yet observed factors (RmRf, SMB, HML, and IP) are provided in the "True" column. All numbers are in percentages.

Table C6: Simulation Results for $n = 100$

T	Param	True	Two-Pass Estimator Bias	Two-Pass Estimator RMSE	Mimicking Portfolios using three assets Bias	Mimicking Portfolios using three assets RMSE	Mimicking Portfolios using all assets Bias	Mimicking Portfolios using all assets RMSE
	RmRf	0.372	-0.180	0.443	0.087	0.395	0.008	0.410
120	SMB	0.229	-0.165	0.338	0.046	0.267	-0.001	0.324
	HML	0.209	0.086	0.333	-0.374	0.387	-0.013	0.427
	IP	-0.003	-0.011	0.090	0.002	0.009	0.000	0.074
	RmRf	0.372	-0.215	0.372	0.036	0.286	0.000	0.288
240	SMB	0.229	-0.158	0.266	0.099	0.213	-0.004	0.201
	HML	0.209	0.232	0.340	-0.341	0.347	-0.001	0.180
	IP	-0.003	-0.028	0.144	0.002	0.006	0.000	0.020
	RmRf	0.372	-0.240	0.321	-0.009	0.199	-0.004	0.203
480	SMB	0.229	-0.087	0.174	0.127	0.182	-0.002	0.139
	HML	0.209	0.183	0.269	-0.307	0.310	-0.005	0.116
	IP	-0.003	-0.062	0.180	0.002	0.004	0.000	0.009

Three-Pass Estimators

T	Param	True	$\breve{p} = 4$ Bias	$\breve{p} = 4$ RMSE	$\breve{p} = 5$ Bias	$\breve{p} = 5$ RMSE	$\breve{p} = 6$ Bias	$\breve{p} = 6$ RMSE
	RmRf	0.372	0.017	0.393	0.013	0.393	0.013	0.393
120	SMB	0.229	-0.028	0.270	-0.018	0.270	-0.017	0.270
	HML	0.209	-0.010	0.210	-0.013	0.212	-0.014	0.213
	IP	-0.003	0.000	0.010	0.000	0.010	0.000	0.011
	RmRf	0.372	0.001	0.286	0.001	0.286	0.001	0.286
240	SMB	0.229	-0.015	0.193	-0.013	0.193	-0.012	0.194
	HML	0.209	-0.002	0.147	-0.003	0.148	-0.001	0.149
	IP	-0.003	0.000	0.007	0.000	0.007	0.000	0.007
	RmRf	0.372	-0.010	0.202	-0.005	0.202	-0.005	0.202
480	SMB	0.229	0.001	0.137	-0.003	0.137	-0.003	0.137
	HML	0.209	-0.004	0.107	-0.008	0.107	-0.008	0.107
	IP	-0.003	0.000	0.005	0.000	0.005	0.000	0.005

Note: In this table, we report the bias (Column "Bias") and the root-mean-square error (Column "RMSE") of the risk premia estimates using mimicking portfolios with three assets or all assets, two-pass and three-pass estimators with $\breve{p} = 4$, 5, and 6, for $n = 100$, and $T = 120$, 240, and 480, respectively. The true data-generating process has five factors, and the parameters are calibrated based on the de-noised five Fama-French factors (RmRf, SMB, HML, RMW, and CMA). In this setting, there is no α. The true zero-beta rate is 0, and the true risk premia of four noisy yet observed factors (RmRf, SMB, HML, and IP) are provided in the "True" column. All numbers are in percentages.

Table C7: Simulation Results for $n = 200$

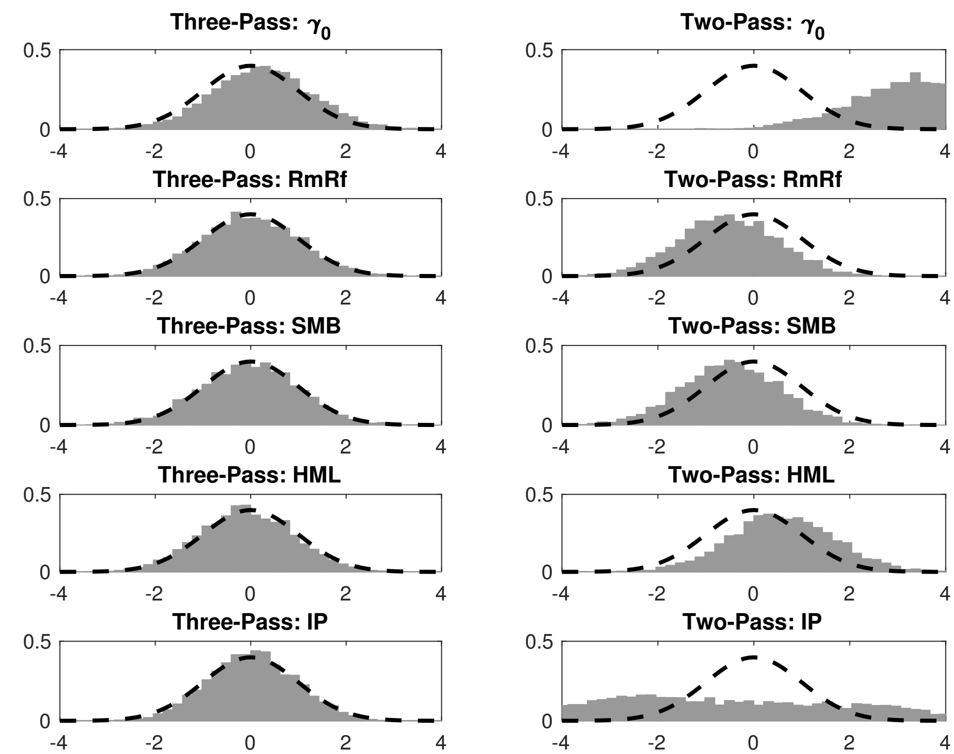
T	Param	True	Two-Pass Estimator Bias	Two-Pass Estimator RMSE	Mimicking Portfolios using three assets Bias	Mimicking Portfolios using three assets RMSE	Mimicking Portfolios using all assets Bias	Mimicking Portfolios using all assets RMSE
120	RmRf	0.372	-0.201	0.462	0.064	0.402	NA	NA
	SMB	0.229	-0.064	0.297	0.064	0.267	NA	NA
	HML	0.209	0.031	0.324	-0.422	0.437	NA	NA
	IP	-0.003	-0.017	0.110	0.002	0.008	NA	NA
240	RmRf	0.372	-0.179	0.344	-0.018	0.281	-0.003	0.294
	SMB	0.229	-0.092	0.223	0.121	0.220	-0.002	0.222
	HML	0.209	0.247	0.347	-0.316	0.323	0.004	0.298
	IP	-0.003	-0.027	0.134	0.002	0.006	0.000	0.050
480	RmRf	0.372	-0.174	0.273	0.027	0.200	-0.001	0.202
	SMB	0.229	-0.070	0.164	0.115	0.175	-0.001	0.140
	HML	0.209	0.155	0.248	-0.374	0.377	0.001	0.129
	IP	-0.003	-0.056	0.183	0.002	0.004	0.000	0.014

Three-Pass Estimators

T	Param	True	$\breve{p} = 4$ Bias	$\breve{p} = 4$ RMSE	$\breve{p} = 5$ Bias	$\breve{p} = 5$ RMSE	$\breve{p} = 6$ Bias	$\breve{p} = 6$ RMSE
120	RmRf	0.372	0.010	0.406	-0.001	0.407	-0.001	0.407
	SMB	0.229	-0.022	0.270	-0.007	0.269	-0.007	0.270
	HML	0.209	-0.018	0.212	-0.019	0.214	-0.018	0.214
	IP	-0.003	0.000	0.010	0.000	0.011	0.000	0.011
240	RmRf	0.372	-0.035	0.285	-0.003	0.284	-0.003	0.284
	SMB	0.229	0.003	0.188	-0.009	0.189	-0.007	0.189
	HML	0.209	0.042	0.155	0.000	0.150	0.000	0.151
	IP	-0.003	0.001	0.007	0.000	0.007	0.000	0.007
480	RmRf	0.372	-0.012	0.201	-0.001	0.201	-0.001	0.201
	SMB	0.229	0.004	0.137	-0.002	0.137	-0.002	0.137
	HML	0.209	0.010	0.110	0.000	0.109	0.000	0.109
	IP	-0.003	0.000	0.005	0.000	0.005	0.000	0.005

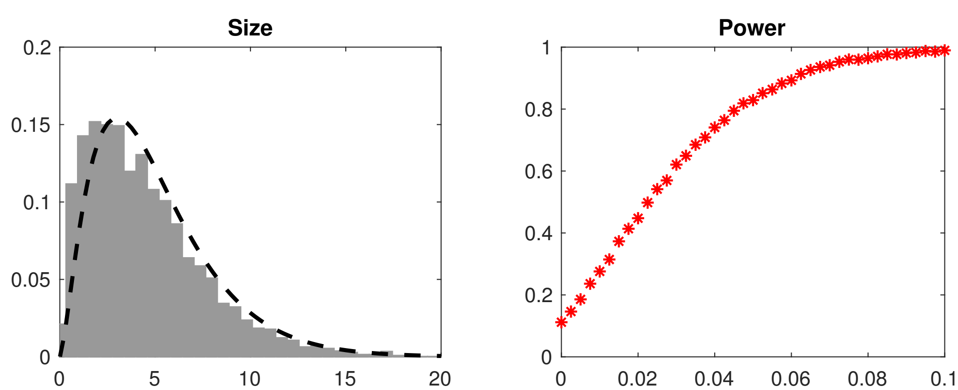
Note: In this table, we report the bias (Column "Bias") and the root-mean-square error (Column "RMSE") of the risk premia estimates using mimicking portfolios with three assets or all assets, two-pass and three-pass estimators with $\breve{p} = 4$, 5, and 6, for $n = 200$, and $T = 120$, 240, and 480, respectively. The true data-generating process has five factors, and the parameters are calibrated based on the de-noised five Fama-French factors (RmRf, SMB, HML, RMW, and CMA). In this setting, there is no α. The true zero-beta rate is 0, and the true risk premia of four noisy yet observed factors (RmRf, SMB, HML, and IP) are provided in the "True" column. All numbers are in percentages. NA means the estimates are "infeasible."

Figure C1: Histograms of the Standardized Estimates in Simulations



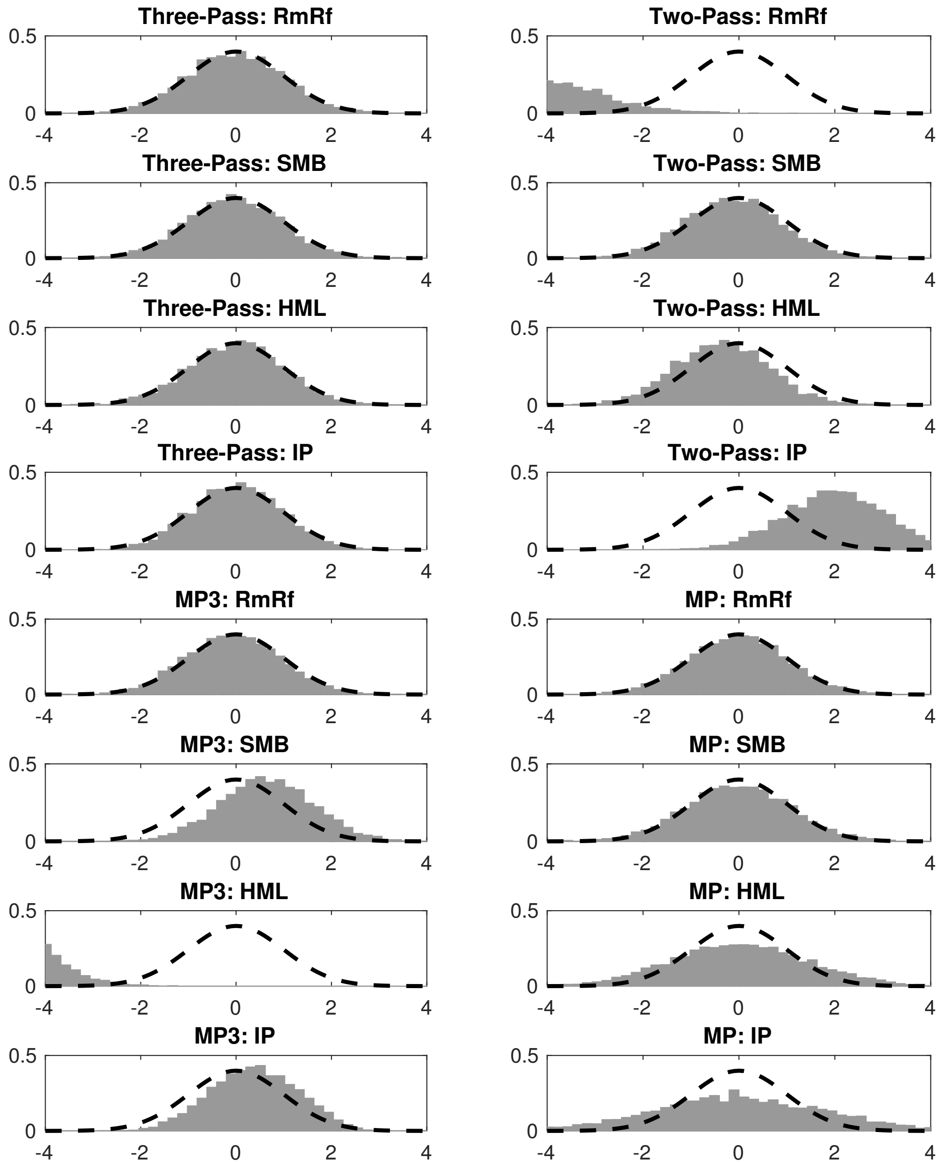
Note: The right panels provide the histograms of the standardized two-pass risk premia estimates using the Fama-MacBeth approach for standard error estimation, whereas the left panels provide the histograms of the standardized three-pass estimates using asymptotic standard errors. We simulate the models with $n = 200$ and $T = 240$.

Figure C2: Size and Power of the Test Statistic



Note: The left panel provides the histogram of the standardized test statistic under the null hypothesis $\eta = 0$ along with the density of the chi-squared distribution with 5 degrees of freedom, whereas the right panel plots the rejection probability (y-axis) against R_g^2 (x-axis). We fix $n = 200$ and $T = 240$.

Figure C3: Histograms of the Standardized Estimates in Simulations



Note: The left panels provide the histograms of the standardized mimicking portfolio estimators for four parameters, using three (bottom left) or all assets (bottom right), respectively, as well as those of the two-pass estimates (top right), and three-pass estimates (top left). We simulate the models with $n = 200$ and $T = 240$.

D Mathematical Proofs

D.1 Proofs of Main Theorems

Proof of Proposition 1. Given the weights of the mimicking portfolios, and by Assumptions A.1 and A.2, these portfolio returns, i.e., $\breve{r}_t = w^\intercal r_t$, satisfy the following factor model:

$$\breve{r}_t = \breve{\beta}\gamma + \breve{\beta}v_t + \breve{u}_t, \tag{D.3}$$

where $\breve{\beta} = w^\intercal \beta$, $\breve{u}_t = w^\intercal u_t$. Conditioning on β, the bias of the factor mimicking-portfolio estimator is:

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

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

Because of (D.3), we have $\breve{\Sigma}^r = \breve{\beta}\Sigma^v\breve{\beta}^\intercal + \breve{\Sigma}^u$, so that by Woodbury matrix identity

$$\begin{aligned}
(\breve{\Sigma}^r)^{-1} =& (\breve{\Sigma}^u)^{-1} - (\breve{\Sigma}^u)^{-1}\breve{\beta}\left(\breve{\beta}^\intercal(\breve{\Sigma}^u)^{-1}\breve{\beta} + (\Sigma^v)^{-1}\right)^{-1}\breve{\beta}^\intercal(\breve{\Sigma}^u)^{-1}, \\
\breve{\beta}^\intercal(\breve{\Sigma}^r)^{-1}\breve{\beta} =& \breve{\beta}^\intercal(\breve{\Sigma}^u)^{-1}\breve{\beta} - \breve{\beta}^\intercal(\breve{\Sigma}^u)^{-1}\breve{\beta}\left(\breve{\beta}^\intercal(\breve{\Sigma}^u)^{-1}\breve{\beta} + (\Sigma^v)^{-1}\right)^{-1}\breve{\beta}^\intercal(\breve{\Sigma}^u)^{-1}\breve{\beta}.
\end{aligned}$$

This further implies that

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

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

Using this equation and by direct calculations, we obtain that

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

where $A = \left(\breve{\beta}^\intercal(\breve{\Sigma}^u)^{-1}\breve{\beta}\right)^{-1}(\Sigma^v)^{-1}$. 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 A.3, $\lambda_{\max}(\breve{\Sigma}^u) = O_p(s_n n^{-1})$, then

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

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

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

On the other hand, we have

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

Using similar analysis as above, we have

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

which concludes the proof. □

Proof of Theorem 1. We take two steps to prove it.
Step 1: Since

$$\bar{R}^\intercal\bar{R} - \bar{V}^\intercal\beta^\intercal\beta\bar{V} = \bar{U}^\intercal\beta\bar{V} + \bar{V}^\intercal\beta^\intercal\bar{U} + \bar{U}^\intercal\bar{U},$$

then by Weyl's inequality, we have, for $1 \leq j \leq p$,

$$\left|\lambda_j(\bar{R}^\intercal\bar{R}) - \lambda_j(\bar{V}^\intercal\beta^\intercal\beta\bar{V})\right| \leq \left\|\bar{U}^\intercal\bar{U}\right\| + \left\|\bar{U}^\intercal\beta\bar{V}\right\| + \left\|\bar{V}^\intercal\beta^\intercal\bar{U}\right\|.$$

We analyze the terms on the right-hand side one by one.
(i) To begin with, write $\Gamma^u = (\gamma_{n,tt'})$. Note that

$$\left\|\bar{U}^\intercal\bar{U} - n\Gamma^u\right\| \leq \left\|U^\intercal U - n\Gamma^u\right\|_{\mathrm{F}} + 2\left\|\iota_T\bar{u}^\intercal U\right\|_{\mathrm{F}} + \left\|\iota_T\bar{u}^\intercal\bar{u}\iota_T^\intercal\right\|_{\mathrm{F}}.$$

By Assumption A.4(ii),

$$\mathrm{E}\left\|U^\intercal U - n\Gamma^u\right\|_{\mathrm{F}}^2 = \sum_{s=1}^T\sum_{t=1}^T \mathrm{E}\left(\sum_{j=1}^n(u_{js}u_{jt} - \mathrm{E}(u_{js}u_{jt}))\right)^2 \leq KnT^2, \tag{D.4}$$

and by Assumption A.4(i),

$$\mathrm{E}\left\|\bar{u}\right\|_{\mathrm{F}}^2 = T^{-2}\mathrm{E}\sum_{i=1}^n\sum_{t=1}^T\sum_{t'=1}^T u_{it}u_{it'} \leq nT^{-2}\sum_{t=1}^T\sum_{t'=1}^T|\gamma_{n,tt'}| \leq KnT^{-1}, \tag{D.5}$$

$$\mathrm{E}\left\|U\right\|_{\mathrm{F}}^2 = \sum_{i=1}^n\sum_{t=1}^T \mathrm{E}u_{it}^2 \leq n\sum_{t=1}^T\gamma_{n,tt} \leq KnT, \tag{D.6}$$

it follows that

$$\left\|\iota_T\bar{u}^\intercal U\right\|_{\mathrm{F}} \leq \left\|\iota_T\right\|_{\mathrm{F}}\left\|\bar{u}^\intercal\right\|_{\mathrm{F}}\left\|U\right\|_{\mathrm{F}} = O_p(nT^{1/2}), \quad \left\|\iota_T\bar{u}^\intercal\bar{u}\iota_T^\intercal\right\|_{\mathrm{F}} \leq \left\|\iota_T\right\|_{\mathrm{F}}^2\left\|\bar{u}^\intercal\right\|_{\mathrm{F}}^2 = O_p(n),$$

and hence that

$$\left\|\bar{U}^\intercal \bar{U} - n\Gamma^u\right\| = O_p(n^{1/2}T) + O_p(nT^{1/2}). \tag{D.7}$$

Next, writing $\rho_{n,st} = \gamma_{n,st}/\sqrt{\gamma_{n,ss}\gamma_{n,tt}}$, by Assumption A.4(i) and the fact that $|\rho_{n,st}| \leq 1$,

$$\begin{aligned}\|\Gamma^u\|_{\mathrm{F}}^2 =&\sum_{s=1}^{T}\sum_{t=1}^{T}\gamma_{n,st}^2 = \sum_{s=1}^{T}\sum_{t=1}^{T}\gamma_{n,ss}\gamma_{n,tt}\rho_{n,st}^2\\ \leq& K\sum_{s=1}^{T}\sum_{t=1}^{T}|\gamma_{n,ss}\gamma_{n,tt}|^{1/2}|\rho_{n,st}| \leq K\sum_{s=1}^{T}\sum_{t=1}^{T}|\gamma_{n,st}| \leq KT,\end{aligned} \tag{D.8}$$

so we have $n\|\Gamma^u\| = O_p(nT^{1/2})$. Therefore, we obtain

$$\left\|\bar{U}^\intercal \bar{U}\right\| \leq \left\|\bar{U}^\intercal \bar{U} - n\Gamma^u\right\| + n\|\Gamma^u\| = O_p(nT^{1/2}) + O_p(n^{1/2}T). \tag{D.9}$$

(ii) By Assumption A.7, we have

$$\mathrm{E}\|U^\intercal\beta\|_{\mathrm{F}}^2 = \mathrm{E}\sum_{j=1}^{p}\sum_{t=1}^{T}\left(\sum_{i=1}^{n}\beta_{ij}u_{it}\right)^2 \leq KnT, \tag{D.10}$$

$$\mathrm{E}\|\bar{u}^\intercal\beta\|_{\mathrm{F}}^2 = \mathrm{E}\sum_{k=1}^{p}\left(\sum_{i=1}^{n}\bar{u}_i\beta_{ik}\right)^2 \leq KnT^{-1}, \tag{D.11}$$

it follows that

$$\left\|\bar{U}^\intercal\beta\right\|_{\mathrm{F}} \leq \|U^\intercal\beta\|_{\mathrm{F}} + \|\iota_T\|_{\mathrm{F}}\|\bar{u}^\intercal\beta\|_{\mathrm{F}} = O_p(n^{1/2}T^{1/2}). \tag{D.12}$$

Also, by Assumption A.5,

$$T^{-1}\left\|\bar{V}\bar{V}^\intercal\right\|_{\mathrm{MAX}} \leq \left\|T^{-1}VV^\intercal - \Sigma^v\right\|_{\mathrm{MAX}} + \|\Sigma^v\|_{\mathrm{MAX}} + \|\bar{v}\bar{v}^\intercal\|_{\mathrm{MAX}} \leq K, \tag{D.13}$$

we have

$$\left\|\bar{V}\right\| \leq \left\|\bar{V}\bar{V}^\intercal\right\|^{1/2} \leq K\left\|\bar{V}\bar{V}^\intercal\right\|_{\mathrm{MAX}}^{1/2} = O_p(T^{1/2}). \tag{D.14}$$

Therefore, we have

$$\left\|\bar{V}^\intercal\beta^\intercal\bar{U}\right\| = \left\|\bar{U}^\intercal\beta\bar{V}\right\| \leq \left\|\bar{U}^\intercal\beta\right\|_{\mathrm{F}}\left\|\bar{V}\right\| = O_p(n^{1/2}T).$$

Combining (i) and (ii), we have for $1 \leq j \leq p$,

$$n^{-1}T^{-1}\left|\lambda_j(\bar{R}^\intercal\bar{R}) - \lambda_j(\bar{V}^\intercal\beta^\intercal\beta\bar{V})\right| = O_p(n^{-1/2} + T^{-1/2}) = o_p(1). \tag{D.15}$$

(iii) Moreover, by Assumption A.6, (D.14), and Weyl's inequality again,

$$\left|n^{-1}T^{-1}\lambda_j(\bar{V}^\intercal\beta^\intercal\beta\bar{V}) - T^{-1}\lambda_j(\bar{V}^\intercal\Sigma^\beta\bar{V})\right| \leq \left\|n^{-1}\beta^\intercal\beta - \Sigma^\beta\right\| T^{-1}\left\|\bar{V}^\intercal\right\|\left\|\bar{V}\right\| = o_p(1),$$

and combined with Assumption A.5, and the fact that $\|\bar{v}\| \leq K\|\bar{v}\|_{\mathrm{MAX}} = O_p(T^{-1/2})$,

$$\begin{aligned}&\left|T^{-1}\lambda_j(\bar{V}^\intercal\Sigma^\beta\bar{V}) - \lambda_j\left(\left(\Sigma^\beta\right)^{1/2}\Sigma^v\left(\Sigma^\beta\right)^{1/2}\right)\right|\\ \leq &\left\|T^{-1}\bar{V}\bar{V}^\intercal - \Sigma^v\right\|\left\|\Sigma^\beta\right\| \leq \left(\left\|T^{-1}VV^\intercal - \Sigma^v\right\| + \|\bar{v}\bar{v}^\intercal\|\right)\left\|\Sigma^\beta\right\| = o_p(1),\end{aligned}$$

where we also use the fact that the non-zero eigenvalues of $\bar{V}^\intercal\Sigma^\beta\bar{V}$ are identical to the non-zero eigenvalues of $\left(\Sigma^\beta\right)^{1/2}\bar{V}\bar{V}^\intercal\left(\Sigma^\beta\right)^{1/2}$. Therefore, for $1 \leq j \leq p$,

$$\left|n^{-1}T^{-1}\lambda_j(\bar{R}^\intercal\bar{R}) - \lambda_j\left(\left(\Sigma^\beta\right)^{1/2}\Sigma^v\left(\Sigma^\beta\right)^{1/2}\right)\right| = o_p(1). \tag{D.16}$$

Step 2: By Assumptions A.5 and A.6, there exists $0 < K_1, K_2 < \infty$, such that

$$K_1 < \lambda_{\min}(\Sigma^v)\lambda_{\min}(\Sigma^\beta) \leq \lambda_{\min}(\Sigma^v\Sigma^\beta) \leq \lambda_{\max}(\Sigma^v\Sigma^\beta) \leq \lambda_{\max}(\Sigma^v)\lambda_{\max}(\Sigma^\beta) < K_2.$$

Therefore the eigenvalues of $(\Sigma^\beta)^{1/2}\Sigma^v(\Sigma^\beta)^{1/2}$ are bounded away from 0 and ∞, we have by (D.16), for $1 \leq j \leq p$,

$$K_1 < n^{-1}T^{-1}\lambda_j(\bar{R}^\intercal\bar{R}) < K_2. \tag{D.17}$$

On the other hand, we can write

$$\bar{R}\bar{R}^\intercal = \tilde{\beta}\bar{V}\bar{V}^\intercal\tilde{\beta}^\intercal + \bar{U}\left(\mathbb{I}_T - \bar{V}^\intercal(\bar{V}\bar{V}^\intercal)^{-1}\bar{V}\right)\bar{U}^\intercal, \tag{D.18}$$

where $\tilde{\beta} = \beta + U\bar{V}^\intercal(\bar{V}\bar{V}^\intercal)^{-1}$. By (4.3.2a) of Theorem 4.3.1 and (4.3.14) of Corollary 4.3.12 in Horn and Johnson (2013), for $p+1 \leq j \leq n$, we have

$$\lambda_j(\bar{R}\bar{R}^\intercal) \leq \lambda_{j-p}\left(\bar{U}(\mathbb{I}_T - \bar{V}^\intercal(\bar{V}\bar{V}^\intercal)^{-1}\bar{V})\bar{U}^\intercal\right) + \lambda_{p+1}(\tilde{\beta}\bar{V}\bar{V}^\intercal\tilde{\beta}) \leq \lambda_{j-p}(\bar{U}\bar{U}^\intercal) \leq \lambda_1(\bar{U}\bar{U}^\intercal).$$

Moreover, by (D.9), we have

$$\lambda_1(\bar{U}\bar{U}^\intercal) = \left\|\bar{U}^\intercal\bar{U}\right\| = O_p(nT^{1/2}) + O_p(n^{1/2}T),$$

hence for $p+1 \leq j \leq n$, there exists some $K > 0$, such that

$$n^{-1}T^{-1}\lambda_j(\bar{R}^\intercal\bar{R}) \leq K(n^{-1/2} + T^{-1/2}). \tag{D.19}$$

Now we define, for $1 \le j \le n$,

$$f(j) = n^{-1}T^{-1}\lambda_j(\bar{R}^\intercal \bar{R}) + j \times \phi(n,T).$$

(D.17) and (D.19) together imply that for $1 \le j \le p$,

$$\begin{aligned} f(j) - f(p+1) =& n^{-1}T^{-1}\left(\lambda_j(\bar{R}^\intercal \bar{R}) - \lambda_{p+1}(\bar{R}^\intercal \bar{R})\right) + (j-p-1)\phi(n,T) \\ >& \lambda_j\left(\left(\Sigma^\beta\right)^{1/2}\Sigma^v\left(\Sigma^\beta\right)^{1/2}\right) + o_p(1) > K, \end{aligned}$$

for some $K > 0$; and for $p+1 < j \le n$, we have

$$\mathrm{P}(f(j) < f(p+1)) = \mathrm{P}\left((j-p-1)\phi(n,T) < n^{-1}T^{-1}\left(\lambda_{p+1}(\bar{R}^\intercal \bar{R}) - \lambda_j(\bar{R}^\intercal \bar{R})\right)\right) \to 0.$$

Therefore, $p+1 = \arg\min_{1 \le j \le p} f(j)$ holds with probability approaching 1, and hence $\widehat{p} \xrightarrow{p} p$. □

Proof of Theorem 2. Let $\widehat{\Lambda}$ be the $p \times p$ diagonal matrix of the p largest eigenvalues of $n^{-1}T^{-1}\bar{R}^\intercal \bar{R}$. We define a $p \times p$ matrix:

$$H = n^{-1}T^{-1}\widehat{\Lambda}^{-1}\widehat{V}\bar{V}^\intercal \beta^\intercal \beta. \tag{D.20}$$

We have the following decomposition:

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

On the one hand, by Lemmas 4(b) and 4(e) we have

$$\begin{aligned} & n^{-1}\left\|\widehat{\beta}^\intercal\widehat{\beta} - H^{-\intercal}\beta^\intercal\beta H^{-1}\right\|_{\mathrm{MAX}} \\ \le& \left\|n^{-1}\left(\widehat{\beta}^\intercal - H^{-\intercal}\beta^\intercal\right)\left(\widehat{\beta} - \beta H^{-1}\right)\right\|_{\mathrm{MAX}} + n^{-1}\left\|H^{-\intercal}\beta^\intercal\left(\widehat{\beta} - \beta H^{-1}\right)\right\|_{\mathrm{MAX}} \\ & + n^{-1}\left\|\left(\widehat{\beta} - \beta H^{-1}\right)^\intercal\beta H^{-1}\right\|_{\mathrm{MAX}} \\ =& O_p(n^{-1} + T^{-1}). \end{aligned} \tag{D.21}$$

Therefore, by Assumption A.6 and Lemma 4(a), (c), and (d), we have

$$\widehat{\gamma} - H\gamma = H\bar{v} + O_p(n^{-1} + T^{-1}). \tag{D.22}$$

On the other hand, we note that

$$\widehat{\eta} - \eta H^{-1} = \eta H^{-1}\left(H\bar{V} - \widehat{V}\right)\widehat{V}^\intercal(\widehat{V}\widehat{V}^\intercal)^{-1} + \bar{Z}\widehat{V}^\intercal(\widehat{V}\widehat{V}^\intercal)^{-1},$$

and by Lemma 5(a) and (b), we have

$$\widehat{\eta} - \eta H^{-1} = T^{-1}\bar{Z}\bar{V}^{\intercal}H^{\intercal} + O_p(n^{-1} + T^{-1}). \tag{D.23}$$

Moreover, by Lemma 5(c), it follows that

$$\left\|\widehat{\eta} - \eta H^{-1}\right\| = O_p(n^{-1} + T^{-1/2}). \tag{D.24}$$

Combining (D.22), (D.23), and Lemma 2, we obtain

$$\widehat{\gamma}_g - \eta\gamma = \eta\bar{v} + T^{-1}\bar{Z}\bar{V}^{\intercal}(\Sigma^v)^{-1}\gamma + O_p(n^{-1} + T^{-1}).$$

Since

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

it follows from Assumption A.11 that

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

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

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

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

Proof of Theorem 3. We summarize the parameters of interest in $\Gamma = (\gamma_0 : (\eta\gamma)^{\intercal})^{\intercal}$, and denote

$$\widetilde{\Gamma} := (\widetilde{\gamma}_0, \widetilde{\gamma}^{\intercal})^{\intercal} = \left((\iota_n : \widehat{\beta})^{\intercal}(\iota_n : \widehat{\beta})\right)^{-1}(\iota_n : \widehat{\beta})^{\intercal}\bar{r}, \quad \widehat{\Gamma} := \begin{pmatrix}\widehat{\gamma}_0\\ \widehat{\gamma}_g\end{pmatrix} := \begin{pmatrix}1 & 0\\ 0 & \widehat{\eta}\end{pmatrix}\widetilde{\Gamma} = \begin{pmatrix}\widetilde{\gamma}_0\\ \widehat{\eta}\widetilde{\gamma}\end{pmatrix}.$$

Because $\widehat{\beta}$ and $\widehat{\eta}$ only rely on $\bar{R}$ and $\bar{G}$, which do not depend on $\gamma_0\iota_n$ and α, we can recycle the estimates derived in Lemmas 1 – 5, despite that the DGP is given by Assumption A.12 instead of Assumption A.1.

We use the following decomposition:

$$\widetilde{\Gamma} - \begin{pmatrix}\gamma_0\\ H\gamma\end{pmatrix}$$

$$= \left((\iota_n : \widehat{\beta})^\intercal (\iota_n : \widehat{\beta}) \right)^{-1} (\iota_n : \widehat{\beta})^\intercal \left(\left(\beta - \widehat{\beta} H \right) \gamma + \beta \bar{v} + \alpha + \bar{u} \right)$$

$$= \begin{pmatrix} 0 \\ H\bar{v} \end{pmatrix} + \left\{ \frac{1}{n} \begin{pmatrix} \iota_n^\intercal \iota_n & \iota_n^\intercal \widehat{\beta} \\ \widehat{\beta}^\intercal \iota_n & \widehat{\beta}^\intercal \widehat{\beta} \end{pmatrix} \right\}^{-1} \left\{ \frac{1}{n} \begin{pmatrix} \iota_n^\intercal \alpha \\ H^{-\intercal} \beta^\intercal \alpha \end{pmatrix} + \frac{1}{n} \begin{pmatrix} \iota_n^\intercal \bar{u} + \iota_n^\intercal (\beta - \widehat{\beta} H) \gamma \\ H^{-\intercal} \beta^\intercal \bar{u} + H^{-\intercal} \beta^\intercal (\beta - \widehat{\beta} H) \gamma \end{pmatrix} \right.$$

$$\left. + \frac{1}{n} \begin{pmatrix} \iota_n^\intercal (\beta - \widehat{\beta} H) \bar{v} \\ (\widehat{\beta} - \beta H^{-1})^\intercal (\alpha + \bar{u}) + H^{-\intercal} \beta^\intercal (\beta - \widehat{\beta} H) \bar{v} + (\widehat{\beta}^\intercal - H^{-\intercal} \beta^\intercal)(\beta - \widehat{\beta} H)(\gamma + \bar{v}) \end{pmatrix} \right\}. \tag{D.25}$$

By Lemma 6(b), we have

$$n^{-1} \left\| \iota_n^\intercal (\widehat{\beta} - \beta H^{-1}) \right\|_{\mathrm{MAX}} = O_p(n^{-1} + T^{-1}).$$

Therefore, we have

$$\frac{1}{n} \begin{pmatrix} \iota_n^\intercal \iota_n & \iota_n^\intercal \widehat{\beta} \\ \widehat{\beta}^\intercal \iota_n & \widehat{\beta}^\intercal \widehat{\beta} \end{pmatrix} = \frac{1}{n} \begin{pmatrix} \iota_n^\intercal \iota_n & \iota_n^\intercal \beta H^{-1} \\ H^{-\intercal} \beta^\intercal \iota_n & H^{-\intercal} \beta^\intercal \beta H^{-1} \end{pmatrix} + O_p(n^{-1} + T^{-1}). \tag{D.26}$$

Using this, and by Lemmas 2, 4, 5, 6, and 7, we have

$$\widehat{\Gamma} - \begin{pmatrix} \gamma_0 \\ \eta \gamma \end{pmatrix} = \begin{pmatrix} 0 \\ T^{-1} \bar{Z} \bar{V}^\intercal (\Sigma^v)^{-1} \gamma + \eta \bar{v} \end{pmatrix}$$

$$+ \begin{pmatrix} 1 & 0 \\ 0 & \eta \end{pmatrix} \left\{ \frac{1}{n} \begin{pmatrix} \iota_n^\intercal \iota_n & \iota_n^\intercal \beta \\ \beta^\intercal \iota_n & \beta^\intercal \beta \end{pmatrix} + o_p(1) \right\}^{-1} \times \left\{ \frac{1}{n} \begin{pmatrix} \iota_n^\intercal \alpha \\ \beta^\intercal \alpha \end{pmatrix} + o_p(n^{-1/2} + T^{-1/2}) \right\}.$$

Moreover, by Cramér-Wold theorem and Lyapunov's central limit theorem, we can obtain

$$n^{-1/2} \begin{pmatrix} \iota_n^\intercal \alpha \\ \beta^\intercal \alpha \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N} \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \beta_0^\intercal \\ \beta_0 & \Sigma^\beta \end{pmatrix} (\sigma^\alpha)^2 \right), \tag{D.27}$$

where we use $\left\| n^{-1} \beta^\intercal \iota_n - \beta_0 \right\|_{\mathrm{MAX}} = o(1)$ and $\left\| n^{-1} \beta^\intercal \beta - \Sigma^\beta \right\|_{\mathrm{MAX}} = o(1)$. Also, Assumptions A.6 and A.13 ensure that $(1 - \beta_0^\intercal (\Sigma^\beta)^{-1} \beta_0)$ and $(\Sigma^\beta - \beta_0 \beta_0^\intercal)$ are invertible. Therefore, by the Delta method, we have

$$n^{1/2} \left(\widehat{\gamma}_0 - \gamma_0 \right) \xrightarrow{\mathcal{L}} \mathcal{N} \left(0, \left(1 - \beta_0^\intercal (\Sigma^\beta)^{-1} \beta_0 \right)^{-1} (\sigma^\alpha)^2 \right),$$

Similarly, we have

$$n^{1/2} \begin{pmatrix} 0 & \eta \end{pmatrix} \left\{ \frac{1}{n} \begin{pmatrix} \iota_n^\intercal \iota_n & \iota_n^\intercal \beta \\ \beta^\intercal \iota_n & \beta^\intercal \beta \end{pmatrix} + o_p(1) \right\}^{-1} \times \frac{1}{n} \begin{pmatrix} \iota_n^\intercal \alpha \\ \beta^\intercal \alpha \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N} \left(0, \Upsilon \right),$$

where

$$\Upsilon = (\sigma^\alpha)^2 \eta \left(\Sigma^\beta - \beta_0 \beta_0^\intercal \right)^{-1} \eta^\intercal.$$

By the same asymptotic independence argument as in the proof of Theorem 3 in Bai (2003), we establish the desired result:

$$\left(T^{-1}\Phi + n^{-1}\Upsilon\right)^{-1/2}(\widetilde{\gamma}_g - \eta\gamma) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \mathbb{I}_d).$$

□

Proof of Theorem 4. By Assumptions A.5, A.6, and A.13, Lemma 4, (D.11), and (D.27), we have

$$\begin{aligned}
&n^{-1}\iota_n^\intercal \bar{r} = \gamma_0 + \beta_0^\intercal\gamma + O_p(n^{-1/2} + T^{-1/2}),\\
&n^{-1}\bar{r}^\intercal\bar{r} = \gamma^\intercal\Sigma^\beta\gamma + \gamma_0^2 + (\sigma^\alpha)^2 + \gamma^\intercal\beta_0\gamma_0 + \beta_0^\intercal\gamma\gamma_0 + O_p(n^{-1/2} + T^{-1/2}),
\end{aligned}$$

it then follows that

$$n^{-1}\bar{r}^\intercal\mathbb{M}_{\iota_n}\bar{r} = n^{-1}\bar{r}^\intercal\bar{r} - (n^{-1}\iota_n^\intercal\bar{r})^2 = \gamma^\intercal(\Sigma^\beta - \beta_0\beta_0^\intercal)\gamma + (\sigma^\alpha)^2 + o_p(1).$$

On the other hand, by Assumption A.5, Lemma 3, (D.5), we have

$$\begin{aligned}
n^{-1}\left\|H^\intercal\widehat{\beta}^\intercal\mathbb{M}_{\iota_n}\bar{r} - \beta^\intercal\mathbb{M}_{\iota_n}\bar{r}\right\|_{\text{MAX}} =&\left\|(H^\intercal\widehat{\beta}^\intercal - \beta^\intercal)\mathbb{M}_{\iota_n}(\alpha + \beta\gamma + \beta\bar{v} + \bar{u})\right\|_{\text{MAX}}\\
\leq& n^{-1}\left\|H^\intercal\widehat{\beta}^\intercal - \beta^\intercal\right\|_{\text{F}}\|\alpha + \beta\gamma + \beta\bar{v} + \bar{u}\|_{\text{F}} = O_p(n^{-1/2} + T^{-1/2}).
\end{aligned}$$

Similarly, we have

$$\begin{aligned}
&n^{-1}\beta^\intercal\mathbb{M}_{\iota_n}\bar{r} = \left(\Sigma^\beta - \beta_0\beta_0^\intercal\right)\gamma + o_p(1),\\
&n^{-1}\beta^\intercal\mathbb{M}_{\iota_n}\beta = \Sigma^\beta - \beta_0\beta_0^\intercal + o_p(1),
\end{aligned}$$

therefore, we obtain

$$(n^{-1}\beta^\intercal\mathbb{M}_{\iota_n}\bar{r})^\intercal\left(n^{-1}\beta^\intercal\mathbb{M}_{\iota_n}\beta\right)^{-1}\left(n^{-1}\beta^\intercal\mathbb{M}_{\iota_n}\bar{r}\right) = \gamma^\intercal\left(\Sigma^\beta - \beta_0\beta_0^\intercal\right)\gamma + o_p(1),$$

which establishes $\widehat{\mathrm{R}}_v^2 \xrightarrow{p} \mathrm{R}_v^2$.

By Lemma 2, (D.48), (D.24) and the fact that $\|\eta\|_{\text{MAX}} \leq K$, we have

$$\begin{aligned}
&\left\|T^{-1}\widehat{\eta}\widehat{V}\widehat{V}^\intercal\eta^\intercal - \eta\Sigma^v\eta^\intercal\right\|_{\text{MAX}}\\
\leq&\left\|(\widehat{\eta} - \eta H^{-1})(\widehat{\eta} - \eta H^{-1})^\intercal\right\|_{\text{MAX}} + \left\|(\widehat{\eta} - \eta H^{-1})H^{-\intercal}\eta^\intercal\right\|_{\text{MAX}} + \left\|\eta H^{-1}(\widehat{\eta} - \eta H^{-1})^\intercal\right\|_{\text{MAX}}\\
&+ \left\|\eta(H^{-1}H^{-\intercal} - \Sigma^v)\eta^\intercal\right\|_{\text{MAX}}\\
=&O_p(n^{-1/2} + T^{-1/2}).
\end{aligned}$$

Also, by Assumptions A.5, A.8, and A.11, we have

$$T^{-1}\bar{G}\bar{G}^\intercal = T^{-1}(\eta\bar{V} + \bar{Z})(\eta\bar{V} + \bar{Z})^\intercal \xrightarrow{p} \eta\Sigma^v\eta^\intercal + \Sigma^z,$$

hence it follows that $\widehat{\mathrm{R}}_g^2 \xrightarrow{p} \mathrm{R}_g^2$. $\square$

Proof of Theorem 5. We denote the estimators of $\bar{V}$ and β based on $\breve{p}$ as $\breve{V}$ and $\breve{\beta}$ respectively. Without loss of generality, we can assume $\widehat{p} = p$.

Consider the singular value decomposition of $n^{-1/2}T^{-1/2}\bar{R}$ by scaling (D.63), we have

$$n^{-1/2}T^{-1/2}\varsigma_{p+1:\breve{p}}^{\intercal}\bar{R} = \Lambda_{p+1:\breve{p}}^{1/2}\xi_{p+1:\breve{p}}^{\intercal} \quad \text{and} \quad n^{-1/2}T^{-1/2}\bar{R}\xi_{p+1:\breve{p}} = \varsigma_{p+1:\breve{p}}\Lambda_{p+1:\breve{p}}^{1/2}, \tag{D.28}$$

where $\Lambda_{p+1:\breve{p}}$ is a $(\breve{p}-p)\times(\breve{p}-p)$ diagonal matrix with the ith entry on the diagonal being $n^{-1}T^{-1}\lambda_i(\bar{R}^{\intercal}\bar{R})$, $\xi_{p+1:\breve{p}} = (\xi_{p+1} : \xi_{p+2} : \cdots : \xi_{\breve{p}})$ is $T \times (\breve{p}-p)$, and $\varsigma_{p+1:\breve{p}} = (\varsigma_{p+1} : \varsigma_{p+2} : \ldots : \varsigma_{\breve{p}})$ is $n \times (\breve{p}-p)$. It is also easy to observe that

$$\breve{V}^{\intercal} = \left(\widehat{V}^{\intercal} : T^{1/2}\xi_{p+1:\breve{p}}\right), \quad \breve{\beta} = \left(\widehat{\beta} : n^{1/2}\varsigma_{p+1:\breve{p}}\Lambda_{p+1:\breve{p}}^{1/2}\right), \quad \widehat{V}\xi_{p+1:\breve{p}} = 0, \quad \text{and} \quad \varsigma_{p+1:\breve{p}}^{\intercal}\widehat{\beta} = 0.$$

Using the formula of block-diagonal matrix inversion, we can decompose

$$\begin{aligned}
\breve{\Gamma} - \widehat{\Gamma} = &\begin{pmatrix} 1 & \\ & \widehat{\eta} \end{pmatrix} \left\{ \frac{1}{n} \begin{pmatrix} \iota_n^{\intercal}\iota_n & \iota_n^{\intercal}\widehat{\beta} \\ \widehat{\beta}^{\intercal}\iota_n & \widehat{\beta}^{\intercal}\widehat{\beta} \end{pmatrix} \right\}^{-1} \\
&\times \begin{pmatrix} -n^{-1}\iota_n^{\intercal}\varsigma_{p+1:\breve{p}}\Lambda_{p+1:\breve{p}}^{1/2}\Delta^{-1}\Lambda_{p+1:\breve{p}}^{1/2}\varsigma_{p+1:\breve{p}}^{\intercal}\left(\mathbb{I}_n - \iota_n(\iota_n^{\intercal}\mathbb{M}_{\widehat{\beta}}\iota_n)^{-1}\iota_n^{\intercal}\mathbb{M}_{\widehat{\beta}}\right)\bar{r} \\ 0 \end{pmatrix} \\
&+ \begin{pmatrix} 0 \\ T^{-1/2}n^{-1/2}\bar{G}\xi_{p+1:\breve{p}}\Delta^{-1}\Lambda_{p+1:\breve{p}}^{1/2}\varsigma_{p+1:\breve{p}}^{\intercal}\left(\mathbb{I}_n - \iota_n(\iota_n^{\intercal}\mathbb{M}_{\widehat{\beta}}\iota_n)^{-1}\iota_n^{\intercal}\mathbb{M}_{\widehat{\beta}}\right)\bar{r} \end{pmatrix},
\end{aligned}$$

where $\Delta = \Lambda_{p+1:\breve{p}}^{1/2}\left(\mathbb{I}_{\breve{p}-p} - \varsigma_{p+1:\breve{p}}^{\intercal}\iota_n(\iota_n^{\intercal}\mathbb{M}_{\widehat{\beta}}\iota_n)^{-1}\iota_n^{\intercal}\varsigma_{p+1:\breve{p}}\right)\Lambda_{p+1:\breve{p}}^{1/2}$.

We analyze the right-hand side terms in the following. First, since $\mathbb{M}_{\widehat{\beta}}\widehat{\beta} = 0$ and $\varsigma_{p+1:\breve{p}}^{\intercal}\widehat{\beta} = 0$, we have

$$\varsigma_{p+1:\breve{p}}^{\intercal}\left(\mathbb{I}_n - \iota_n(\iota_n^{\intercal}\mathbb{M}_{\widehat{\beta}}\iota_n)^{-1}\iota_n^{\intercal}\mathbb{M}_{\widehat{\beta}}\right)\bar{r} = \varsigma_{p+1:\breve{p}}^{\intercal}\left(\mathbb{I}_n - \iota_n(\iota_n^{\intercal}\mathbb{M}_{\widehat{\beta}}\iota_n)^{-1}\iota_n^{\intercal}\mathbb{M}_{\widehat{\beta}}\right)(\alpha + (\beta - \widehat{\beta}H)(\bar{v} + \gamma) + \bar{u}).$$

Since α is independent of $\varsigma_{p+1:\breve{p}}$, we have

$$\mathrm{E}\left\|\varsigma_{p+1:\breve{p}}^{\intercal}\alpha\right\|_{\mathrm{F}}^2 = \sum_{j=1}^{\breve{p}-p}\mathrm{E}\left(\sum_{i=1}^{n}\varsigma_{p+1:\breve{p},ij}\alpha_i\right)^2 = \mathrm{E}\left\|\varsigma_{p+1:\breve{p}}\right\|_{\mathrm{F}}^2(\sigma^{\alpha})^2 \leq K,$$

hence $\left\|\varsigma_{p+1:\breve{p}}^{\intercal}\alpha\right\|_{\mathrm{F}} = O_p(1)$. By Lemma 3(b) and (D.5), we have

$$\left\|\varsigma_{p+1:\breve{p}}^{\intercal}(\beta - \widehat{\beta}H)\right\| = O_p(1 + n^{1/2}T^{-1/2}), \quad \left\|\varsigma_{p+1:\breve{p}}^{\intercal}\bar{u}\right\| = O_p(n^{1/2}T^{-1/2}).$$

On the other hand, since by Lemma 7, (D.26), and (D.27), we have

$$\left\|\iota_n^{\intercal}\mathbb{M}_{\widehat{\beta}}\alpha\right\| = O_p(n^{1/2}), \quad (\iota_n^{\intercal}\mathbb{M}_{\widehat{\beta}}\iota_n)^{-1} = O_p(n^{-1}).$$

By Lemmas 4 and 6, we have

$$\left\|\iota_n^\intercal \mathbb{M}_{\widehat{\beta}}\bar{u}\right\| = O_p(1+nT^{-1}), \quad \left\|\iota_n^\intercal \mathbb{M}_{\widehat{\beta}}(\beta - \widehat{\beta}H)\right\| = O_p(1+nT^{-1}).$$

Combined with Lemma 8, it implies that

$$\left\|\varsigma_{p+1:\breve{p}}^\intercal \iota_n (\iota_n^\intercal \mathbb{M}_{\widehat{\beta}}\iota_n)^{-1}\left(\iota_n^\intercal \mathbb{M}_{\widehat{\beta}}(\alpha + (\beta - \widehat{\beta}H)(\bar{v}+\gamma) + \bar{u})\right)\right\| = o_p(1+n^{1/2}T^{-1}).$$

With the above estimates, we obtain

$$\begin{aligned}
&\left\|\varsigma_{p+1:\breve{p}}^\intercal \left(\mathbb{I}_n - \iota_n (\iota_n^\intercal \mathbb{M}_{\widehat{\beta}}\iota_n)^{-1}\iota_n^\intercal \mathbb{M}_{\widehat{\beta}}\right)\bar{r}\right\| \\
\leq &\left\|\varsigma_{p+1:\breve{p}}^\intercal (\alpha + (\beta - \widehat{\beta}H)(\bar{v}+\gamma) + \bar{u})\right\| + \left\|\varsigma_{p+1:\breve{p}}^\intercal \iota_n (\iota_n^\intercal \mathbb{M}_{\widehat{\beta}}\iota_n)^{-1}\left(\iota_n^\intercal \mathbb{M}_{\widehat{\beta}}(\alpha + (\beta - \widehat{\beta}H)(\bar{v}+\gamma) + \bar{u})\right)\right\| \\
= &O_p(1+n^{1/2}T^{-1/2}). \qquad \text{(D.29)}
\end{aligned}$$

Next, by Sherman-Morrison-Woodbury formula, we have

$$\begin{aligned}
\left\|\Lambda_{p+1:\breve{p}}^{1/2}\Delta^{-1}\Lambda_{p+1:\breve{p}}^{1/2}\right\| &= \left\|\mathbb{I}_{\breve{p}-p} + \varsigma_{p+1:\breve{p}}^\intercal \iota_n \iota_n^\intercal \varsigma_{p+1:\breve{p}}\left(\iota_n^\intercal \mathbb{M}_{\widehat{\beta}}\iota_n - \iota_n^\intercal \varsigma_{p+1:\breve{p}}\varsigma_{p+1:\breve{p}}^\intercal \iota_n\right)^{-1}\right\| \\
&= \left\|\mathbb{I}_{\breve{p}-p} + n^{-1}\varsigma_{p+1:\breve{p}}^\intercal \iota_n \iota_n^\intercal \varsigma_{p+1:\breve{p}}\left(1 - n^{-1}\iota_n^\intercal \varsigma_{1:\breve{p}}\varsigma_{1:\breve{p}}^\intercal \iota_n\right)^{-1}\right\|,
\end{aligned}$$

where we use the fact that

$$n^{-1}\iota_n^\intercal \mathbb{M}_{\widehat{\beta}}\iota_n = 1 - n^{-1}\iota_n^\intercal \widehat{\beta}(\widehat{\beta}^\intercal \widehat{\beta})^{-1}\widehat{\beta}^\intercal \iota_n = 1 - n^{-1}\iota_n^\intercal \varsigma_{1:p}\varsigma_{1:p}^\intercal \iota_n.$$

By Lemma 8, for any $\breve{p} \leq K$, we have

$$\left(1 - n^{-1}\iota_n^\intercal \varsigma_{1:\breve{p}}\varsigma_{1:\breve{p}}^\intercal \iota_n\right)^{-1} = \left(1 - n^{-1}\iota_n^\intercal \varsigma_{1:p}\varsigma_{1:p}^\intercal \iota_n + o_p(1)\right)^{-1} = \left(1 - \beta_0^\intercal (\Sigma^\beta)^{-1}\beta_0\right)^{-1} + o_p(1).$$

Also, $\left\|n^{-1}\varsigma_{p+1:\breve{p}}^\intercal \iota_n \iota_n^\intercal \varsigma_{p+1:\breve{p}}\right\| = o_p(1)$, which in turn leads to

$$\left\|\Lambda_{p+1:\breve{p}}^{1/2}\Delta^{-1}\Lambda_{p+1:\breve{p}}^{1/2}\right\| = O_p(1). \qquad \text{(D.30)}$$

Combining (D.29), (D.30), and using Lemma 8 again, we have

$$\left\|n^{-1}\iota_n^\intercal \varsigma_{p+1:\breve{p}}\Lambda_{p+1:\breve{p}}^{1/2}\Delta^{-1}\Lambda_{p+1:\breve{p}}^{1/2}\varsigma_{p+1:\breve{p}}^\intercal \left(\mathbb{I}_n - \iota_n(\iota_n^\intercal \mathbb{M}_{\widehat{\beta}}\iota_n)^{-1}\iota_n^\intercal \mathbb{M}_{\widehat{\beta}}\right)\bar{r}\right\| = o_p(n^{-1/2}+T^{-1/2}).$$

Next, by Lemmas 1 and 2,

$$\left\|\eta \bar{V}\xi_{p+1:\breve{p}}\right\| = \left\|\eta(\bar{V} - H^{-1}\widehat{V})\xi_{p+1:\breve{p}}\right\| = O_p(1+n^{-1/2}T^{1/2}).$$

By the independence of $\bar{Z}$ and $\xi_{p+1:\breve{p}}$, we also have $\left\|\bar{Z}\xi_{p+1:\breve{p}}\right\| = O_p(1)$. On the other hand, it follows from (D.18) that

$$\bar{R}\bar{R}^\intercal + \bar{U}\bar{V}^\intercal(\bar{V}\bar{V}^\intercal)^{-1}\bar{V}\bar{U}^\intercal = \bar{U}\bar{U}^\intercal + \tilde{\beta}\bar{V}\bar{V}^\intercal\tilde{\beta}^\intercal,$$

where $\tilde{\beta} = \beta + U\bar{V}^\intercal(\bar{V}\bar{V}^\intercal)^{-1}$. By (4.3.2a) and (4.3.2b) of Theorem 4.3.1 in Horn and Johnson (2013), for $p+1 \leq j \leq \breve{p}$,

$$\lambda_{j+p}(\bar{U}\bar{U}^\intercal) + \lambda_{n-1}(\tilde{\beta}\bar{V}\bar{V}^\intercal\tilde{\beta}^\intercal) \leq \lambda_{j+p}(\bar{R}\bar{R}^\intercal + \bar{U}\bar{V}^\intercal(\bar{V}\bar{V}^\intercal)^{-1}\bar{V}\bar{U}^\intercal) \leq \lambda_j(\bar{R}\bar{R}^\intercal) + \lambda_{p+1}(\bar{U}\bar{V}^\intercal(\bar{V}\bar{V}^\intercal)^{-1}\bar{V}\bar{U}^\intercal).$$

Since $\operatorname{rank}(\tilde{\beta}\bar{V}\bar{V}^\intercal\tilde{\beta}^\intercal) \leq p$ and $\operatorname{rank}(\bar{U}\bar{V}^\intercal(\bar{V}\bar{V}^\intercal)^{-1}\bar{V}\bar{U}^\intercal) \leq p$, we obtain,

$$\lambda_{\breve{p}+p}(\bar{U}\bar{U}^\intercal) \leq \lambda_{j+p}(\bar{U}\bar{U}^\intercal) \leq \lambda_j(\bar{R}\bar{R}^\intercal) \leq \lambda_{j-p}(\bar{U}\bar{U}^\intercal) \leq \lambda_1(\bar{U}\bar{U}^\intercal).$$

Since the empirical distribution of eigenvalues of $T^{-1}\bar{U}\bar{U}^\intercal$ follows the Marčenko-Pastur law (see, e.g., Theorem 3.6 in Bai and Silverstein (2009)), as $n/T \to c \in (0, \infty)$, and this limiting law has a bounded support, then there exist upper and lower bounds for finitely many largest eigenvalues

$$0 < K' \leq T^{-1}\lambda_j(\bar{R}\bar{R}^\intercal) \leq K, \quad p+1 \leq j \leq \breve{p},$$

so that $\left\|\Lambda_{p+1:\breve{p}}^{-1/2}\right\| = O_p(n^{1/2})$. We therefore obtain that

$$\begin{aligned}
&\left\|T^{-1/2}n^{-1/2}\bar{G}\xi_{p+1:\breve{p}}\Delta^{-1}\Lambda_{p+1:\breve{p}}^{1/2}\varsigma_{p+1:\breve{p}}^\intercal\left(\mathbb{I}_n - \iota_n(\iota_n^\intercal\mathbb{M}_{\widehat{\beta}}\iota_n)^{-1}\iota_n^\intercal\mathbb{M}_{\widehat{\beta}}\right)\bar{r}\right\| \\
\leq &\left\|T^{-1/2}n^{-1/2}\bar{G}\xi_{p+1:\breve{p}}\right\|\left\|\Lambda_{p+1:\breve{p}}^{-1/2}\right\|\left\|\Lambda_{p+1:\breve{p}}^{1/2}\Delta^{-1}\Lambda_{p+1:\breve{p}}^{1/2}\right\|\left\|\varsigma_{p+1:\breve{p}}^\intercal\left(\mathbb{I}_n - \iota_n(\iota_n^\intercal\mathbb{M}_{\widehat{\beta}}\iota_n)^{-1}\iota_n^\intercal\mathbb{M}_{\widehat{\beta}}\right)\bar{r}\right\| \\
= &O_p(n^{-1/2}),
\end{aligned}$$

which concludes the proof. □

Proof of Theorem 6. Again, we assume $\widehat{p} = p$. To prove the consistency of $\widehat{\Phi}$, without loss of generality, we focus on the case of Π_{12}, and show that

$$(\widetilde{\gamma}^\intercal \otimes \mathbb{I}_d)\,\widehat{\Pi}_{12}\widehat{\eta}^\intercal \xrightarrow{p} \left(\gamma^\intercal\,(\Sigma^v)^{-1} \otimes \mathbb{I}_d\right)\Pi_{12}\eta^\intercal. \tag{D.31}$$

The proof for the other two terms in $\widehat{\Phi}$ is similar and hence is omitted.

Note that by (D.56), Lemma 2, Lemma 3(a), and Assumption A.5, we have

$$\begin{aligned}
&\left\|T^{-1}H^{-1}\widehat{V}\widehat{V}^\intercal H^{-\intercal} - \Sigma^v\right\|_{\text{MAX}} \\
= &\left\|T^{-1}H^{-1}(\widehat{V} - H\bar{V})\widehat{V}^\intercal H^{-\intercal} + T^{-1}\bar{V}(\widehat{V}^\intercal - \bar{V}^\intercal H^\intercal)H^{-\intercal} + T^{-1}VV^\intercal - \Sigma^v - \bar{v}\bar{v}^\intercal\right\|_{\text{MAX}} \\
= &O_p(n^{-1} + T^{-1/2}).
\end{aligned}$$

By (D.24), Lemma 2, and the proof of Theorem 2, we have

$$\|\widehat{\eta}H - \eta\|_{\mathrm{MAX}} = O_p(n^{-1} + T^{-1/2}), \quad \left\|H^{-1}\widetilde{\gamma} - \gamma\right\|_{\mathrm{MAX}} = O_p(n^{-1/2} + T^{-1/2}). \tag{D.32}$$

Therefore, to prove (D.31), we only need to show that

$$\widetilde{\Pi}_{12} := (H^{-1} \otimes \mathbb{I}_d)\widehat{\Pi}_{12}H^{-\intercal} \xrightarrow{p} \Pi_{12}, \tag{D.33}$$

with which, and by the continuous mapping theorem, we have

$$\begin{aligned}\left(\widetilde{\gamma}^{\intercal}\left(\widehat{\Sigma}^{v}\right)^{-1} \otimes \mathbb{I}_d\right)\widehat{\Pi}_{12}\widehat{\eta}^{\intercal} &= \left((H^{-1}\widetilde{\gamma})^{\intercal}\left(H^{-1}\widehat{\Sigma}^{v}H^{-\intercal}\right)^{-1} \otimes \mathbb{I}_d\right)(H^{-1} \otimes \mathbb{I}_d)\widehat{\Pi}_{12}H^{-\intercal}(\widehat{\eta}H)^{\intercal} \\ &\xrightarrow{p} \left(\gamma^{\intercal}\left(\Sigma^{v}\right)^{-1} \otimes \mathbb{I}_d\right)\Pi_{12}\eta^{\intercal}.\end{aligned}$$

Writing $\widetilde{V} = H^{-1}\widehat{V}$, we have

$$\widetilde{\Pi}_{12,(i-1)d+j,i'} = \mathrm{vec}(e_j e_i^{\intercal})^{\intercal}(H^{-1} \otimes \mathbb{I}_d)\widehat{\Pi}_{12}H^{-\intercal}e_{i'} = \mathrm{vec}(e_j e_i^{\intercal}H^{-1})^{\intercal}\widehat{\Pi}_{12}H^{-\intercal}e_{i'} = T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T}\widehat{z}_{jt}\widetilde{v}_{it}Q_{ts}\widetilde{v}_{i's},$$

where $Q_{st} = \left(1 - \frac{|s-t|}{q+1}\right)1_{|s-t|\leq q}$.

In fact, to show (D.33), by Lemma 2 we only need to prove for any fixed $1 \leq i, i' \leq p$, and $1 \leq j, j' \leq d$,

$$\widetilde{\Pi}_{12,(i-1)d+j,i'} - T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T} z_{jt}v_{it}Q_{ts}v_{i's} \xrightarrow{p} 0, \tag{D.34}$$

since by the identical proof of Theorem 2 in Newey and West (1987), we have

$$T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T} z_{jt}v_{it}Q_{ts}v_{i's} - \Pi_{12,(i-1)d+j,i'} \xrightarrow{p} 0.$$

Note that

$$\begin{aligned}&\text{the left-hand side of (D.34)} \\ =&T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T}\Big\{(\widehat{z}_{jt} - z_{jt})(\widetilde{v}_{it} - v_{it})Q_{ts}(\widetilde{v}_{i's} - v_{i's}) + (\widehat{z}_{jt} - z_{jt})(\widetilde{v}_{it} - v_{it})Q_{ts}v_{i's} \\ &+ (\widehat{z}_{jt} - z_{jt})v_{it}Q_{ts}\widetilde{v}_{i's} + z_{jt}(\widetilde{v}_{it} - v_{it})Q_{ts}\widetilde{v}_{i's} + z_{jt}\widetilde{v}_{it}Q_{ts}(\widetilde{v}_{i's} - v_{i's})\Big\}.\end{aligned}$$

We analyze these terms one by one. Since we have

$$\widehat{Z} - \bar{Z} = \eta\bar{V} - \widehat{\eta}\widehat{V} = (\eta H^{-1} - \widehat{\eta})H\bar{V} - (\widehat{\eta} - \eta H^{-1})(\widehat{V} - H\bar{V}) - \eta H^{-1}(\widehat{V} - H\bar{V}), \tag{D.35}$$

it follows from (D.24), (D.44), and Lemmas 1 and 2 that

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

Moreover, by Lemma 9, Assumption A.16, (D.35), and (D.32), we have

$$
\begin{aligned}
& \left\|\widehat{Z}-\bar{Z}\right\|_{\mathrm{MAX}} \\
\leq & \left\|\eta H^{-1}-\widehat{\eta}\right\|_{\mathrm{MAX}}\|H\|\left\|\bar{V}\right\|_{\mathrm{MAX}}+\left\|\widehat{\eta}-\eta H^{-1}\right\|_{\mathrm{MAX}}\left\|\widehat{V}-H \bar{V}\right\|_{\mathrm{MAX}}+\left\|\eta H^{-1}\right\|\left\|\widehat{V}-H \bar{V}\right\|_{\mathrm{MAX}} \\
= & O_p((\log T)^{1/a} T^{-1/2}+n^{-1/2} T^{1/4}).
\end{aligned}
$$

By Cauchy-Schwartz inequality, Lemmas 1, 9, and using the fact that $|Q_{ts}| \leq 1_{|t-s| \leq q}$ and $\left\|\bar{v} \iota_T^{\intercal}\right\|_{\mathrm{F}}=\|\bar{v}\|_{\mathrm{F}}\left\|\iota_T^{\intercal}\right\|_{\mathrm{F}} \leq K T^{1/2}\|\bar{v}\|_{\mathrm{MAX}}=O_p(1)$, we have

$$
\begin{aligned}
& \left|T^{-1} \sum_{t=1}^T \sum_{s=1}^T(\widehat{z}_{jt}-z_{jt})(\widetilde{v}_{it}-v_{it}) Q_{ts}(\widetilde{v}_{i's}-v_{i's})\right| \\
\leq & K q T^{-1}\left(\left\|\widetilde{V}-\bar{V}\right\|_{\mathrm{MAX}}+\left\|\bar{v} \iota_T^{\intercal}\right\|_{\mathrm{MAX}}\right)\left(\left\|\widetilde{V}-\bar{V}\right\|_{\mathrm{F}}+\left\|\bar{v} \iota_T^{\intercal}\right\|_{\mathrm{F}}\right)\left(\left\|\widehat{Z}-\bar{Z}\right\|_{\mathrm{F}}+\left\|\bar{z} \iota_T^{\intercal}\right\|_{\mathrm{F}}\right) \\
= & O_p\left(q(T^{-1}+n^{-1})(T^{1/4} n^{-1/2}+T^{-1})\right).
\end{aligned}
$$

Similarly, because of $\left\|\widetilde{V}\right\|_{\mathrm{F}} \leq O_p(T^{1/2})$ implied by (D.45), $\|Z\|_{\mathrm{MAX}}=O_p((\log T)^{1/a})$ by Assumption A.16 and Lemma 2, and by Assumptions A.5 and A.8, we have

$$
\begin{aligned}
& \left|T^{-1} \sum_{t=1}^T \sum_{s=1}^T(\widehat{z}_{jt}-z_{jt})(\widetilde{v}_{it}-v_{it}) Q_{ts} v_{i's}\right| \\
\leq & K q T^{-1}\|V\|_{\mathrm{MAX}}\left(\left\|\widetilde{V}-\bar{V}\right\|_{\mathrm{F}}+\left\|\bar{v} \iota_T^{\intercal}\right\|_{\mathrm{F}}\right)\left(\left\|\widehat{Z}-\bar{Z}\right\|_{\mathrm{F}}+\left\|\bar{z} \iota_T^{\intercal}\right\|_{\mathrm{F}}\right)=O_p\left(q(\log T)^{1/a}(n^{-1}+T^{-1})\right), \\
& \left|T^{-1} \sum_{t=1}^T \sum_{s=1}^T(\widehat{z}_{jt}-z_{jt}) v_{it} Q_{ts} \widetilde{v}_{i's}\right| \\
\leq & K q T^{-1}\|V\|_{\mathrm{MAX}}\left\|\widetilde{V}\right\|_{\mathrm{F}}\left(\left\|\widehat{Z}-\bar{Z}\right\|_{\mathrm{F}}+\left\|\bar{z} \iota_T^{\intercal}\right\|_{\mathrm{F}}\right)=O_p\left(q(\log T)^{1/a}\left(n^{-1/2}+T^{-1/2}\right)\right), \\
& \left|T^{-1} \sum_{t=1}^T \sum_{s=1}^T z_{jt}(\widetilde{v}_{it}-v_{it}) Q_{ts} \widetilde{v}_{i's}\right| \\
\leq & K q T^{-1}\|Z\|_{\mathrm{MAX}}\left\|\widetilde{V}\right\|_{\mathrm{F}}\left(\left\|H^{-1} \widehat{V}-\bar{V}\right\|_{\mathrm{F}}+\left\|\bar{v} \iota_T^{\intercal}\right\|_{\mathrm{F}}\right)=O_p\left(q(\log T)^{1/a}\left(n^{-1/2}+T^{-1/2}\right)\right), \\
& \left|T^{-1} \sum_{t=1}^T \sum_{s=1}^T z_{jt} \widetilde{v}_{it} Q_{ts}(\widetilde{v}_{i's}-v_{i's})\right| \\
\leq & K q T^{-1}\|Z\|_{\mathrm{MAX}}\left\|\widetilde{V}\right\|_{\mathrm{F}}\left(\left\|H^{-1} \widehat{V}-\bar{V}\right\|_{\mathrm{F}}+\left\|\bar{v} \iota_T^{\intercal}\right\|_{\mathrm{F}}\right)=O_p\left(q(\log T)^{1/a}\left(n^{-1/2}+T^{-1/2}\right)\right).
\end{aligned}
$$

All the above terms converge to 0, as $T, n \to \infty$, with $qT^{-1/4} + qn^{-1/4} \to 0$ and $n^{-3}T \to 0$, which establishes (D.34).

Finally, to show the consistency of $\widehat{\Upsilon}$, we first note

$$\left\| H^\intercal \left(\widehat{\Sigma}^\beta - \widehat{\beta}_0 \widehat{\beta}_0^\intercal \right) H - \left(\Sigma^\beta - \beta_0 \beta_0^\intercal \right) \right\|_{\text{MAX}} \le \left\| H^\intercal \widehat{\Sigma}^\beta H - \Sigma^\beta \right\|_{\text{MAX}} + \left\| H^\intercal \widehat{\beta}_0 \widehat{\beta}_0^\intercal H - \beta_0 \beta_0^\intercal \right\|_{\text{MAX}}.$$

By Lemmas 2, 4(b), (e), and Assumption A.6,

$$\begin{aligned}
&\left\| H^\intercal \widehat{\Sigma}^\beta H - \Sigma^\beta \right\|_{\text{MAX}} \\
\le& \left\| n^{-1} H^\intercal \widehat{\beta}^\intercal \widehat{\beta} H - n^{-1} \beta^\intercal \beta \right\|_{\text{MAX}} + \left\| n^{-1} \beta^\intercal \beta - \Sigma^\beta \right\|_{\text{MAX}} \\
\le& \left\| n^{-1} \left(H^\intercal \widehat{\beta}^\intercal - \beta^\intercal \right) (\widehat{\beta} H - \beta) + n^{-1} \left(H^\intercal \widehat{\beta}^\intercal - \beta^\intercal \right) \beta - n^{-1} \beta^\intercal (\beta - \widehat{\beta} H) \right\|_{\text{MAX}} + o_p(1) \\
=& o_p(1). \qquad\qquad (\text{D.36}) \\
&\left\| H^\intercal \widehat{\beta}_0 \widehat{\beta}_0^\intercal H - \beta_0 \beta_0^\intercal \right\|_{\text{MAX}} \\
\le& \left\| \left(H^\intercal \widehat{\beta}_0 - \beta_0 \right) \left(\widehat{\beta}_0^\intercal H - \beta_0^\intercal \right) + \beta_0 \left(\widehat{\beta}_0^\intercal H - \beta_0^\intercal \right) + \left(H^\intercal \widehat{\beta}_0 - \beta_0 \right) \beta_0^\intercal \right\|_{\text{MAX}} \\
=& o_p(1),
\end{aligned}$$

where we also use Lemma 6(b):

$$\left\| H^\intercal \widehat{\beta}_0 - \beta_0 \right\|_{\text{MAX}} = n^{-1} \left\| \left(H^\intercal \widehat{\beta}^\intercal - \beta^\intercal \right) \iota_n \right\|_{\text{MAX}} = o_p(1).$$

Next, by Lemma 3(b) and (D.32), we have

$$\begin{aligned}
\widehat{\sigma^\alpha}^2 - (\sigma^\alpha)^2 =& n^{-1} \left\| \bar{r} - \iota_n \widetilde{\gamma}_0 - \widehat{\beta} \widetilde{\gamma} \right\|_{\text{F}}^2 - (\sigma^\alpha)^2 \\
=& n^{-1} \left\| \iota_n (\gamma_0 - \widetilde{\gamma}_0) + \beta \gamma - \widehat{\beta} \widetilde{\gamma} + \beta \bar{v} + \bar{u} \right\|_{\text{F}}^2 + n^{-1} \|\alpha\|_{\text{F}}^2 - (\sigma^\alpha)^2 \\
\le& n^{-1} \|\iota_n\|_{\text{F}}^2 \|\gamma_0 - \widetilde{\gamma}_0\|_{\text{F}}^2 + n^{-1} \|\beta\|_{\text{F}}^2 \|\bar{v}\|_{\text{F}}^2 + n^{-1} \|\bar{u}\|_{\text{F}}^2 + n^{-1} \left\| (\widehat{\beta} H - \beta) \gamma \right\|_{\text{F}}^2 \\
&+ n^{-1} \left\| (\widehat{\beta} H - \beta)(H^{-1} \widetilde{\gamma} - \gamma) \right\|_{\text{F}}^2 + n^{-1} \left\| \beta (H^{-1} \widetilde{\gamma} - \gamma) \right\|_{\text{F}}^2 + o_p(1).
\end{aligned}$$

Therefore, by (D.24) and the continuous mapping theorem,

$$\widehat{\sigma^\alpha}^2 \widehat{\eta} H H^{-1} \left(\widehat{\Sigma}^\beta - \widehat{\beta}_0 \widehat{\beta}_0^\intercal \right)^{-1} H^{-\intercal} H^\intercal \widehat{\eta}^\intercal \xrightarrow{p} \Upsilon,$$

which concludes the proof. □

Proof of Theorem 7. By the conditioning argument, we assume $\widehat{p} = p$. By (D.23), we have

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

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

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

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

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

By Assumption A.11, to show $\widehat{W} \xrightarrow{\mathcal{L}} \chi_p^2$ under $\mathbb{H}_0$, it is sufficient to establish that

$$H^{-1}\widehat{\Sigma}_v^{-1}\widehat{\Pi}_{11}\widehat{\Sigma}_v^{-1}H^{-\intercal} \xrightarrow{p} \Pi_{11}.$$

By the same argument as in the proof of (D.33) and the fact that $\widehat{\Sigma}_v = \mathbb{I}_d$, we have

$$H^{-1}\widehat{\Pi}_{11}H^{-\intercal} \xrightarrow{p} \Pi_{11},$$

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 $T^{-1}\bar{Z}\bar{V}^\intercal = O_p(T^{-1/2})$ by Assumption A.8. □

Proof of Theorem 8. For any $1 \le t \le T$, we have

$$\widehat{g}_t - \eta v_t = (\widehat{\eta} - \eta H^{-1})(\widehat{v}_t - H\bar{v}_t) + (\widehat{\eta} - \eta H^{-1})H\bar{v}_t + \eta H^{-1}(\widehat{v}_t - H\bar{v}_t) - \eta\bar{v} \tag{D.37}$$

By (D.43), we have

$$\begin{aligned}\widehat{v}_t - H\bar{v}_t =& n^{-1}T^{-1}\widehat{\Lambda}^{-1}(\widehat{V} - H\bar{V})\left(\bar{U}^\intercal\beta\bar{v}_t + \bar{U}^\intercal\bar{u}_t\right) + n^{-1}T^{-1}\widehat{\Lambda}^{-1}\left(H\bar{V}\bar{U}^\intercal\beta\bar{v}_t + H\bar{V}\bar{U}^\intercal\bar{u}_t\right)\\ &+ n^{-1}T^{-1}\widehat{\Lambda}^{-1}\widehat{V}\bar{V}^\intercal\beta^\intercal\bar{u}_t.\end{aligned} \tag{D.38}$$

By Assumption B.17, we have $\|\beta^\intercal u_t\| = O_p(n^{1/2})$, so that using (D.11),

$$\|\beta^\intercal\bar{u}_t\|_{\mathrm{F}} \le \|\beta^\intercal u_t\|_{\mathrm{F}} + \|\beta^\intercal\bar{u}\|_{\mathrm{F}} = O_p(n^{1/2}). \tag{D.39}$$

By Assumption A.4(i), Assumptions A.15 and A.16, using the fact that $|\rho_{n,st}| \le 1$, we have

$$\begin{aligned}\mathrm{E}\,\|U^\intercal u_t\|_{\mathrm{F}}^2 =& \mathrm{E}\sum_{s=1}^{T}\left(n\gamma_{n,st} + \sum_{k=1}^{n}\left(u_{ks}u_{kt} - \mathrm{E}(u_{ks}u_{kt})\right)\right)^2\\ &\le Kn^2\sum_{s=1}^{T}\gamma_{n,st}^2 + KnT \le n^2\sum_{s=1}^{T}|\gamma_{n,st}| + KnT = Kn^2 + KnT,\\ \mathrm{E}\,\|u_t\|_{\mathrm{F}}^2 \le& \sum_{k=1}^{n}\mathrm{E}u_{kt}^2 \le \sum_{k=1}^{n}|\sigma_{kk'}| \le K.\end{aligned} \tag{D.40}$$

Then from (D.5) and (D.57), it follows that

$$\left\|\bar{U}^\intercal \bar{u}_t\right\|_{\mathrm{F}} \leq \left\|\bar{U}^\intercal \bar{u}\right\|_{\mathrm{F}} + \|U^\intercal u_t\|_{\mathrm{F}} + \|\iota_T\|_{\mathrm{F}} \|\bar{u}^\intercal\|_{\mathrm{F}} \|u_t\|_{\mathrm{F}} = O_p(n + n^{1/2}T^{1/2}).$$

The above estimates, along with (D.12), Lemma 1, and $\|\bar{v}_t\| = O_p(1)$, lead to

$$\begin{aligned}&\left\|n^{-1}T^{-1}\widehat{\Lambda}^{-1}(\widehat{V} - H\bar{V})\left(\bar{U}^\intercal \beta \bar{v}_t + \bar{U}^\intercal \bar{u}_t\right)\right\|_{\mathrm{MAX}}\\ \leq & n^{-1}T^{-1}\left\|\widehat{\Lambda}^{-1}\right\|_{\mathrm{MAX}}\left\|\widehat{V} - H\bar{V}\right\|_{\mathrm{F}}\left(\left\|\bar{U}^\intercal\beta\right\|_{\mathrm{F}}\|\bar{v}_t\| + \left\|\bar{U}^\intercal \bar{u}_t\right\|_{\mathrm{F}}\right) = O_p(n^{-1} + T^{-1}).\end{aligned}$$

Moreover, it follows from (D.5), (D.50), and (D.55) that

$$\begin{aligned}&\left\|n^{-1}T^{-1}\widehat{\Lambda}^{-1}\left(H\bar{V}\bar{U}^\intercal\beta\bar{v}_t + H\bar{V}\bar{U}^\intercal\bar{u}_t\right)\right\|_{\mathrm{MAX}}\\ \leq & Kn^{-1}T^{-1}\left\|\widehat{\Lambda}^{-1}\right\|_{\mathrm{MAX}}\|H\|\left(\left\|\bar{V}\bar{U}^\intercal\beta\right\|_{\mathrm{MAX}}\|\bar{v}_t\| + \left\|\bar{V}\bar{U}^\intercal\right\|_{\mathrm{F}}\left(\|u_t\|_{\mathrm{F}} + \|u\|_{\mathrm{F}}\right)\right)\\ = & O_p(n^{-1/2}T^{-1/2} + T^{-1}).\end{aligned}$$

We thereby focus on the remaining term, which by Lemma 1, (D.14) and (D.39), satisfies

$$n^{-1}T^{-1}\left\|\widehat{\Lambda}^{-1}\widehat{V}\bar{V}^\intercal\beta^\intercal\bar{u}_t\right\|_{\mathrm{MAX}} \leq Kn^{-1}T^{-1}\left\|\widehat{\Lambda}^{-1}\right\|_{\mathrm{MAX}}\left\|\widehat{V}\right\|_{\mathrm{F}}\left\|\bar{V}^\intercal\right\|_{\mathrm{F}}\|\beta^\intercal\bar{u}_t\|_{\mathrm{MAX}} = O_p(n^{-1/2}).$$

Therefore, we have

$$\|\widehat{v}_t - H\bar{v}_t\|_{\mathrm{MAX}} = O_p(n^{-1/2} + T^{-1}). \tag{D.41}$$

Then by (D.37), (D.38), and (D.23), we have

$$\left\|\widehat{g}_t - \eta v_t - \left(T^{-1}\bar{Z}\bar{V}^\intercal H^\intercal H v_t + n^{-1}T^{-1}\eta H^{-1}\widehat{\Lambda}^{-1}\widehat{V}\bar{V}^\intercal\beta^\intercal u_t - \eta\bar{v}\right)\right\|_{\mathrm{MAX}} = o_p(n^{-1/2} + T^{-1/2}).$$

Next, we note that by Assumption A.11 and Lemma 2,

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

By (D.20) and Assumptions A.6 and B.17, we have

$$n^{-1/2}T^{-1}\eta H^{-1}\widehat{\Lambda}^{-1}\widehat{V}\bar{V}^\intercal\beta^\intercal u_t = n^{1/2}\eta(\beta^\intercal\beta)^{-1}\beta^\intercal u_t \xrightarrow{\mathcal{L}}\mathcal{N}\left(0, \eta\left(\Sigma^\beta\right)^{-1}\Omega_t\left(\Sigma^\beta\right)^{-1}\eta^\intercal\right).$$

The desired result follows from the same asymptotic independence argument as in Bai (2003). □

Proof of Theorem 9. For $\widehat{\Psi}_{1t}$, we can follow exactly the same proof as that of Theorem 6, since, similar to (D.32) for $\widetilde{\gamma}$, we have the same estimate for $\widehat{v}_t$ by (D.41).

As to $\widehat{\Psi}_{2t}$, similarly, we only need to show

$$\left\|H^\intercal \widehat{\Omega} H - \Omega\right\|_{\mathrm{MAX}} = o_p(1).$$

Then by the continuous mapping theorem, along with (D.32) and (D.36), we have

$$\widehat{\Psi}_{2t} = \widehat{\eta} H \left(H^\intercal \widehat{\Sigma}^\beta H\right)^{-1} H^\intercal \widehat{\Omega}_t H \left(H^\intercal \widehat{\Sigma}^\beta H\right)^{-1} H^\intercal \widehat{\eta}^\intercal \xrightarrow{p} \Psi_{2t}.$$

Note that by Fan et al. (2013), we have

$$\left\|\widehat{\Sigma}^u - \Sigma^u\right\| = O_p(s_n \omega_T^{1-h}). \tag{D.42}$$

Then by (D.42) and Lemmas 3(b), 4(b), and using the fact that $\|\beta\|_{\mathrm{F}} = O_p(n^{1/2})$ and $\|\Sigma^u\| \le \|\Sigma^u\|_1 = O_p(s_n)$, writing $\widetilde{\beta} = \widehat{\beta} H$, we have

$$\begin{aligned}
\frac{1}{n}\left\|(\widetilde{\beta}-\beta)^\intercal(\widehat{\Sigma}^u - \Sigma^u)(\widetilde{\beta}-\beta)\right\|_{\mathrm{MAX}} &\le \frac{1}{n}\left\|\widetilde{\beta}-\beta\right\|_{\mathrm{F}}^2 \left\|\widehat{\Sigma}^u - \Sigma^u\right\| = O_p\left(s_n \omega_T^{1-h}(n^{-1}+T^{-1})\right), \\
\frac{1}{n}\left\|(\widetilde{\beta}-\beta)^\intercal \Sigma^u (\widetilde{\beta}-\beta)\right\|_{\mathrm{MAX}} &\le \frac{1}{n}\left\|\widetilde{\beta}-\beta\right\|_{\mathrm{F}}^2 \left\|\Sigma^u\right\| = O_p\left(s_n(n^{-1}+T^{-1})\right), \\
\frac{1}{n}\left\|\beta^\intercal(\widehat{\Sigma}^u - \Sigma^u)\beta\right\|_{\mathrm{MAX}} &\le \frac{1}{n}\|\beta\|_{\mathrm{F}}^2 \left\|\widehat{\Sigma}^u - \Sigma^u\right\| = O_p\left(s_n \omega_T^{1-h}\right), \\
\frac{1}{n}\left\|\beta^\intercal(\widehat{\Sigma}^u - \Sigma^u)(\widetilde{\beta}-\beta)\right\|_{\mathrm{MAX}} &\le \frac{1}{n}\left\|(\widehat{\Sigma}^u - \Sigma^u)(\widetilde{\beta}-\beta)\beta^\intercal\right\| \le \frac{1}{n}\left\|\widehat{\Sigma}^u - \Sigma^u\right\|\left\|\beta^\intercal(\widetilde{\beta}-\beta)\right\| \\
&\le \frac{K}{n}\left\|\beta^\intercal(\widetilde{\beta}-\beta)\right\|_{\mathrm{MAX}}\left\|\widehat{\Sigma}^u - \Sigma^u\right\| = O_p\left(s_n \omega_T^{1-h}(n^{-1}+T^{-1})\right), \\
\frac{1}{n}\left\|\beta^\intercal \Sigma^u(\widetilde{\beta}-\beta)\right\|_{\mathrm{MAX}} &\le \frac{K}{n}\left\|\beta^\intercal(\widetilde{\beta}-\beta)\right\|_{\mathrm{MAX}}\left\|\Sigma^u\right\| = O_p\left(s_n(n^{-1}+T^{-1})\right).
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\left\|H^\intercal \widehat{\Omega} H - \Omega\right\|_{\mathrm{MAX}} = \frac{1}{n}\left\|H^\intercal \widehat{\beta}^\intercal \widehat{\Sigma}^u \widehat{\beta} H - \beta^\intercal \Sigma^u \beta\right\|_{\mathrm{MAX}} \\
\le& \frac{1}{n}\left\|(\widetilde{\beta}-\beta)^\intercal(\widehat{\Sigma}^u - \Sigma^u)(\widetilde{\beta}-\beta) + (\widetilde{\beta}-\beta)^\intercal \Sigma^u(\widetilde{\beta}-\beta)\right\|_{\mathrm{MAX}} + \frac{1}{n}\left\|\beta^\intercal(\widehat{\Sigma}^u - \Sigma^u)\beta\right\|_{\mathrm{MAX}} \\
&+ \frac{1}{n}\left\|\beta^\intercal(\widehat{\Sigma}^u - \Sigma^u)(\widetilde{\beta}-\beta) + (\widetilde{\beta}-\beta)^\intercal(\widehat{\Sigma}^u - \Sigma^u)\beta\right\|_{\mathrm{MAX}} + \frac{1}{n}\left\|\beta^\intercal \Sigma^u(\widetilde{\beta}-\beta) + (\widetilde{\beta}-\beta)^\intercal \Sigma^u \beta\right\|_{\mathrm{MAX}} \\
=& O_p\left(s_n\left(\omega_T^{1-h} + n^{-1} + T^{-1}\right)\right) = o_p(1),
\end{aligned}$$

which concludes the proof. $\square$