

Inference on Risk Premia in the Presence of Omitted Factors*

Stefano Giglio[†]
Chicago Booth

Dacheng Xiu[‡]
Chicago Booth

January 15, 2017

Abstract

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 the observed factors are just a subset of the true factors that drive asset prices. Standard methods to estimate risk premia are biased in the presence of omitted priced factors correlated with the observed factors. We show that the risk premium of a factor can be identified in a linear factor model regardless of the rotation of the other control factors as long as they together span the space of true factors. Motivated by this rotation invariance result, our approach uses principal components to recover the factor space and combines the estimated principal components with each observed factor to obtain a consistent estimate of its risk premium. This 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 linear models.

KEYWORDS: Three-Pass Estimator, Empirical Asset Pricing Models, PCA, Latent Factors, Omitted Factors, Fama-MacBeth Regression

1 Introduction

Asset pricing models often predict that some factors – for example, intermediary capital or aggregate liquidity – should command a risk premium: investors should be compensated for their exposure to those sources of risk, holding constant their exposure to all other risk factors.

In many cases, the model-predicted factors are not tradable (i.e., they are not themselves traded portfolios). The risk premium for each factor can be estimated by constructing a portfolio with unit exposure (beta) to that factor and zero exposure to all other factors, for example via two-pass

*We benefited tremendously from discussions with Svetlana Bryzgalova, John Cochrane, George Constantinides, Gene Fama, Valentin Haddad, Christian Hansen, Lars Hansen, Toby Moskowitz, Stavros Panageas, Ivo Welch, and seminar participants at the Xiamen University and Hong Kong University of Science and Technology and at the AFA.

[†]Booth School of Business, University of Chicago. Address: 5807 S Woodlawn Avenue, Chicago IL 60637, USA. E-mail address: stefano.giglio@chicagobooth.edu.

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

regressions like Fama-MacBeth or by constructing long-short portfolios with multiple sorts (isolating selected factor exposures nonparametrically); the risk premium is then computed as the average excess return of the portfolio exposed only to that factor, and not to the others. The risk premium of a nontradable factor tells us how much investors are willing to pay to hedge that risk, and therefore represents quantitative evidence on the economic importance of that factor.

A fundamental concern when estimating risk premia via cross-sectional regressions is the potential omission of factors (noted, among others, by [Jagannathan and Wang \(1998\)](#)): for the estimation procedure to correctly recover the factor risk premia, all *other* priced factors in the economy need to be controlled for in the two-pass regression. This is an important problem in practice because most asset pricing models are too stylized to explicitly capture *all* sources of risk in the economy.¹ The resulting omitted variable bias affects the estimated magnitude and even sign of the risk premia for the observed factors, and also the test for their statistical significance.

The typical, ad-hoc solution used in the literature to handle this omitted factor problem is to simply add some factors as “controls” (for example, the market return is often included even if when it is outside the theoretical model), or, alternatively, to add firm characteristics to the regressions. This solution involves selecting arbitrary factors or characteristics as controls, with no guidance from the model and no guarantee that the selected controls are the right ones; the results are often strongly dependent on the choice of additional factors to include.

In this paper, we show that by exploiting the large dimensionality of available test assets and a rotation invariance result for linear pricing models, we can correctly recover the risk premium of each observable factor, even when not all true risk factors are observed and included. In other words, we can recover the slope of the ideal two-step cross-sectional regression that includes the observed factor as well as all the omitted factors – even when we cannot directly observe the latter. Our method therefore solves in a systematic way the omitted variable problem in estimating risk premia, avoiding the use of arbitrary choices for control factors or characteristics.

We apply our methodology to a large set of 202 equity portfolios. We estimate and test the significance of the risk premium of tradable and nontradable factors from a number of different models. We show that the conclusions about the magnitude and significance of the risk premia often depend on whether we account for omitted factors (using our estimator) or ignore them (using standard two-pass regressions). 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, an indication of the validity of our procedure. We also decompose the variance of each observed factor into the components due to exposure to the underlying factors pricing the cross-section of returns, 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 exposures to unobservable factors, they command a risk premium of essentially zero. Instead, our results yield strong support for factors related to financial frictions (like the liquidity factor of [Pástor and Stambaugh \(2003\)](#)), whereas the

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

standard methods that ignore omitted factors produce mixed or insignificant results for many of these factors in our sample.

Our solution to the omitted factor problem combines two-pass cross-sectional regressions with principal component analysis (PCA). The premise of our procedure is a simple but useful rotation invariance result that holds in linear factor models. Suppose that returns follow a linear factor model with p factors and we wish to determine the risk premium of one of them (call it g_t). We show that a standard two-pass regression will always correctly recover the risk premium for g_t as long as the two-pass analysis includes any $p - 1$ “control” factors that, together with g_t , span the entire factor space. Because PCA recovers a factor-space rotation (as the number of assets $n \rightarrow \infty$), the factors extracted from PCA represent a natural set of “controls” that allow us to recover the risk premium of g_t . Using PCA also guarantees that the recovered “control” factors are measurement-error free (though subject to some estimation error), an important precondition for the controls to span the relevant space.

This invariance result is unique to estimating risk premia in linear factor models, and it does not hold in standard linear regression settings with omitted variables. For example, in the standard regression setting where the researcher does not observe some variables but uses some linear combinations of the variables as “controls,” the estimated coefficients for the observed variable are *not* invariant to rotations of the controls.² The key difference is due to the fact that any rotation of the factors in two-pass regressions has two offsetting effects on risk exposures and risk premia. The invariance result states that the two effects always offset each other so that risk premia for the observable factors are estimated correctly even when risk exposures are not, as long as the “controls” span the true factor space.

To apply directly, this invariance result requires error-free g_t measurement, since g_t , together with the selected PCs, must span the entire factor space. In practice, it is likely that measurement error affects most empirical factors (and especially non-tradable ones). In this paper, we propose a three-pass estimator that exploits the invariance result while also accounting explicitly for potential measurement error in the observed factor. To do so, we first apply PCA to the set of returns, without using information in g_t ; only then we relate the latent factors and their risk premia to the risk premium of the observable factor g_t . More specifically, we first use principal component analysis (PCA) to extract factors and their loadings from a large panel of testing portfolio returns; we then run a cross-sectional regression (CSR) to find the risk premia of the extracted factors, and finally recover the risk premia of the observable factor(s) from a time-series regression (TSR) that uncovers the relation between the observable and latent factors.

²A simple example may help clarify the intuition. Suppose that a variable Y depends linearly on X and Z , two correlated variables. If Z is not observed, the coefficient of a regression of Y on X alone will contain an omitted variable bias. Suppose Z is not observed, but X and a rotation of the two variables ($aX + bZ$) are observed (where a and b are not known to the econometrician). Together, X and ($aX + bZ$) span the same space spanned by X and Z . However, a regression of Y on X and ($aX + bZ$) will *not* recover the correct coefficient on X , i.e., this regression does not solve the omitted variable bias. Our invariance result states that when X and Z are factors and Y are returns in a linear factor model, the risk premium of X is correctly identified even when the “control” factor is any linear combination ($aX + bZ$).

We show that our estimation procedure yields consistent risk-premium estimates for the observed factors, and we derive their asymptotic distribution when both the number of testing 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, and allows for measurement error on the observed factor. In addition, the increasing dimensionality simplifies the asymptotic variance of the risk-premium estimates, for which we also provide an estimator. Moreover, our inference is valid even when any of the observable factors is spurious or even useless. Finally, we construct a consistent estimator for the number of latent factors, while also showing that even without it, the risk-premium estimates could still be consistent.

While most useful for estimating the economic importance of non-tradable factors, our methodology can also be applied to tradable factors. For tradable factors, the risk premium can be computed in two ways. The first is to average the time-series excess return of the factor; the second is to use cross-sectional regressions, like two-pass estimators or our three-pass methodology, under the assumptions of the linear factor model. Misspecifications of the model (omitted controls; nonlinearities; correlated time-variation in risk exposures and risk premia) affect the latter estimator but not the former. Therefore, if the two estimates are different, it is an indication that the factor model is misspecified. While using standard two-pass regressions the two often differ (even for the market portfolio, see [Lettau and Ludvigson \(2001\)](#)), estimates of risk premia obtained with our three-pass methodology are close to the time series average returns for almost all of the tradable factors we study, including the market portfolio. This gives us confidence in using the same model assumptions to estimate the risk premia of non-tradable factors, for which this direct validation exercise is not feasible.

This paper sits at the confluence of several literatures, combining two-pass cross-sectional regressions 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 \(1996\)](#), [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 when n is large relative to T . [Raponi et al. \(2016\)](#) on the other hand study the ex-post risk premia using large n and fixed T asymptotics. For a review of this literature, see [Shanken \(1996\)](#), [Jagannathan et al. \(2010\)](#), and, more recently, [Kan and Robotti \(2012\)](#). 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 (Kleiberger (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); Kleiberger (2009); Kan and Robotti (2009); Kan et al. (2013); Gospodinov et al. (2013); Kleiberger and Zhan (2014); Gospodinov et al. (2016); Bryzgalova (2015); Burnside (2016)). We study a different model misspecification form – priced factors omitted from the model, which would also bias the estimates for the observed factors. Hou and Kimmel (2006) argue that in this case, the definition of risk premia can be ambiguous. Relying on a large number of testing 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 cross-sectional regression. Moskowitz (2003) explores the relation between the characteristic-based portfolios and the covariance matrix of returns, something that we parallel in our paper when we exploit the correlation of observable factors with the latent factors to estimate their risk premia; this link is also further developed in Pukthuanthong and Roll (2014).

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. (2015) show how few principal components capture a large fraction of the cross-section of expected returns. Clarke (2016) constructs principal components of returns after sorting them by their expected return constructed using characteristics, to better isolate their priced component. 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 (2006b) 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 spanning the latent factors in time series, and their ability to explain the cross-sectional variation of expected returns.

Our paper also relates to the literature on weak factors (see for example Gospodinov et al. (2016) or Bryzgalova (2015)). An essential element of our rotation invariance result is that the space spanned by all priced factors can be recovered and factors spanning that space should be used as controls in the cross-sectional regression. Using principal component analysis to do so as we propose in this paper precludes the possibility of weak latent factors, i.e. latent factors that have positive risk premium but close to zero exposure in the set of test assets. Weak factors are associated with small eigenvalues of returns, and may not be selected by principal components. This is an important

concern, that we do not tackle directly in our work and leave for future analysis. However, we point out three sources of robustness of our analysis with respect to the weak factors problem. First, while our analysis requires that all true priced factors are strong, we allow at least for the factor of interest g_t to be weak or even useless. That is, if g_t is a weak factor, our procedure would correctly classify it as such and make correct inference on its risk premium. Second, while our asymptotic variance results are correct only when the exact number of true factors is recovered, our estimates are still consistent even when “too many” principal components are extracted from the data. Therefore, if one is worried about the presence of weak factors in the data, increasing the number of principal components – adding factors with smaller and smaller eigenvalues – will allow us to capture weaker and weaker factors and verify the robustness of the results. Finally, we can look at tradable factors (for which a model-free estimate of the expected return can be computed directly) to verify that the risk premia estimated using our PCA-based methodology correspond to the time-series average excess return of each factor; in the data, we find relatively small differences between the two approaches, suggesting that we are not omitting weak factors that might substantially bias our estimates.

Our analysis in this paper uses unconditional linear models. A large literature has explored conditional factor models where factor exposures, risk quantities, and risk premia can be time varying (for example, [Jagannathan and Wang \(1996\)](#) and [Lewellen and Nagel \(2006\)](#)). For simplicity, we do not explicitly focus on conditional models in this paper. However, it is useful to note that as long as the correlation between risk premia and factor exposures is small, unconditional models would still yield approximately correct (average) risk premia estimates. In addition, conditional models can often be approximated by unconditional models with extra factors (where the extra factors capture interactions between the true factors and variables in the conditioning information set, as described in [Cochrane \(2009\)](#)); so, to the extent that PCA correctly recovers the relevant factors, they will also recover the factors in the unconditional model that arise due to time variation in the conditioning information.

Finally, our paper shares the same spirit with the vast literature on economic forecasting using factor models, as the first step towards forecasting typically involves a parsimonious representation of a large panel of predictors. PCA is a widely used toolkit for this purpose, see, e.g., [Forni and Reichlin \(1998\)](#), [Stock and Watson \(2002a\)](#), [Stock and Watson \(2002b\)](#), [Bai and Ng \(2006a\)](#), [Bai and Ng \(2008\)](#). Alternatively, [Kelly and Pruitt \(2013\)](#) and [Kelly and Pruitt \(2015\)](#) propose a three-pass regression filter which leads to a substantial improvement in forecasting, e.g., the conditional expectation of market returns.

Section 2 proposes a potentially misspecified beta-pricing model and sets the paper’s objective. Section 3 presents an invariance result, which our identification strategy discussed in Section 4 relies on. Section 5 introduces the estimation procedure. Section 6 provides the asymptotic theory on inference. Section 7 presents Monte Carlo simulations, followed by an empirical study in Section 8. The appendix provides the mathematical proofs.

Throughout the paper, we use $(A : B)$ to denote the concatenation (by columns) of two matrices A and B . 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 $(a_1 : a_2 : \dots : a_T)$, and write $\bar{A} = A - \iota_T^T \bar{a}$ correspondingly. e_i is a vector with 1 in the i th entry and 0 elsewhere, whose dimension depends on the context. ι_k denotes a k -dimensional vector with all entries being 1. We denote $\mathbb{P}_A = A(A^T A)^{-1} A^T$ and $\mathbb{M}_A = \mathbb{I} - \mathbb{P}_A$.

2 Model Setup

We start describing a simple example – a special case of the more general setup considered later in this section – that illustrates the omitted factor bias in two-pass regressions.

Suppose that we want to estimate and test the significance of the risk premium for an observable factor g_t suggested by theory: for example, a liquidity or a financial intermediary capital factor. The true factor model includes g_t but also another, unobserved, factor f_t . g_t and f_t can be arbitrarily correlated in the panel of returns, and the betas with respect to each factor can also be arbitrarily cross-sectionally correlated, as long as they are not perfectly correlated. In addition, we allow for some measurement error in g_t .

Trying to estimate the risk premium of g_t using standard cross-sectional regression methods (but without observing the potentially correlated f_t) causes two problems to arise. First, the time series regression of returns on the observed g_t will yield biased betas (due both to the omission of f_t and measurement error in g_t). The second pass involves a cross-sectional regression of the expected returns onto the estimated betas. Because only the betas corresponding to g_t are included in the regression, another omitted variable bias arises. Eventually, all three biases appear in the estimated risk premium due to the time-series correlation of the factors, the cross-sectional correlation of the betas, and the measurement error in the observable factor g_t . In fact, it is enough that *any* of these issues occurs to bias the risk premium estimate of factor g_t .

Instead, our procedure is able to fully recover the correct risk premium of g_t , correcting all three sources of bias. To do so, it uses PCA on the panel of returns to extract factors that span the *entire factor space* (two factors in this case) and directly account for the variation in g_t that is *not* due to measurement error. Effectively, this methodology allows us to “control” for the unobservable factors in the risk premia estimation, and clean up the observed factors from the measurement error.

We specify that assets are priced by a linear factor model with potentially unobservable factors:

Assumption 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 observable returns of the testing assets. The pricing model satisfies:*

$$r_t = \iota_n \gamma_0 + \alpha + \beta \gamma + \beta v_t + u_t, \quad f_t = \mu + v_t, \quad \mathbb{E}(v_t) = \mathbb{E}(u_t) = 0, \quad \text{and} \quad \text{Cov}(u_t, v_t) = 0, \quad (1)$$

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, α is an $n \times 1$ vector of pricing errors, β is an $n \times p$ factor loading matrix, and γ_0 and γ are the zero-beta rate and the $p \times 1$ risk premia vector, respectively.

We allow for a non-zero pricing error in the cross section of expected returns, so that the linear

factor model is a potentially imperfect approximation of the true model. The focus of this paper is not on testing the null of APT, and allowing for at least some form of potential mispricings yields a more robust inference on the factor risk premia. We discuss in Section 6 which processes and types of pricing errors are allowed in our framework. Most of our results hold for non-stationary processes with heteroscedasticity and dependence in both the time series and the cross-sectional dimensions.

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

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

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.

We allow for measurement error in g_t , because this is often plausible in practice. It captures noise in the construction or measurement of the factors, or exposure to idiosyncratic risks (it can be correlated with u_t). Assumption 2 says that g_t proxies for a set of asset pricing factors in the linear factor model representation: after removing measurement error, g_t captures exactly a linear transformation of the fundamental factors, ηv_t .

This specification implies that we can represent the true asset pricing model – after a rotation – as a model where g_t corresponds to the first d factors (after removing measurement error), together with other $p - d$ factors that are other combinations of the fundamental factors v_t but are potentially not observed. The simple model discussed at the beginning of this section is a special case of the general model described in Assumptions 1 and 2. There was no measurement error, and g_t coincided with one of the factors in v_t , so that η was simply the unit vector e_1 .

3 An Invariance Property

We are interested in the risk premium associated with each observable factor in g_t . Recall that a factor’s risk premium is the expected excess return of a portfolio with no idiosyncratic risk, no alpha, unit beta with respect to that factor, and zero beta with respect to all other risk factors. Because g_t may contain measurement error, we refer to the risk premium of g_t as the risk premium with respect to ηv_t , (i.e., the compensation for the systematic risk to which g_t is exposed).³ To calculate the risk premium of any of the factors in g_t , we rely on a rotation of the fundamental model (1) such that the factor appears directly as one of the p factors, together with $p - 1$ rotated factors. Such a rotation is not unique, but the risk premium is invariant as is shown below. This general result holds regardless of whether v_t is observable or not.

Proposition 1. *Suppose Assumptions 1 and 2 hold. The risk premium of g_t is $\eta\gamma$. Moreover, it is invariant to the choice of factors in Assumption 1, as long as the space spanned by the rotated factors is the same as that of the true factors.*

³Without ambiguity, we do not distinguish the risk premium of ηv_t from the risk premium of g_t .

Proposition 1 states that we can always transform a linear factor model with p factors (v_t) into a representation where g_t appears as one of the p factors, together with $p - 1$ other factors, that are linear combinations of the original factors. In *any* such transformation, as long as it preserves the same span of the factors, the risk premium of g_t is equal to $\eta\gamma$. In contrast, the factor loading with respect to g_t is obviously not invariant, which depends on the correlation between g_t and the other factors.

To see the intuition for the result, derived formally in the appendix, consider one observed factor g_t with no constant or measurement error in it, so that $g_t = \eta v_t$. For any full-rank $p \times p$ matrix H , call $q_t = H v_t$ the factors in the rotation H of the linear factor model (i.e., in the linear-factor-model representation where q_t are the factors). It is easy to observe that if the vector of risk premia of v_t is γ , the vector of risk premia of q_t is $H\gamma$. Now consider any rotations H such that g_t appears as a first factor. There are many such rotations: in fact, any matrix H where the first row is η will produce a rotated model where g_t is the first factor (because $g_t = \eta v_t$). The risk premium of g_t is then $\eta\gamma$, no matter what the *other* $p - 1$ rows of H are, because it is the first element of $H\gamma$. So the risk premium of g_t ($\eta\gamma$) is well-defined in any rotation of the model where g_t is the first factor, as long as any other $p - 1$ linear combinations of v_t are included (the additional rows of H). The risk exposures (betas) to g_t , instead, cannot be determined because they depend on the entire matrix H .

Proposition 1 also implies that in theory we can obtain the risk premium of g_t in two ways, assuming v_t is observed. We can first transform the model so that g_t appears as a factor, and then apply standard two-pass estimator to this transformed model, directly recovering $\eta\gamma$ as the risk premium of g_t . Alternatively, we can first obtain the factor risk premia in the original model expressed in terms of v_t (where g_t may not directly appear), obtaining the risk premia of the factors v_t , γ . Then, we compute the risk premium of g_t by multiplying this γ by η , the exposure of g_t to v_t .

Another implication of this invariance result is that as long as the original model (in terms of v_t) is well identified, then its rotations will also be well identified. For example, if g_t and h_t are two observable factors and are both linear functions of v_t , the risk premia associated with g_t and h_t will be well identified *even if these two factors are highly correlated*. Intuitively, rotating the model from v_t to a model expressed in terms of g_t and h_t (and other factors) will rotate not only the factors and their risk premia but also the risk exposures. The rotations of factor exposures and factor risk premia offset each other, so that if the original model is well identified, the transformed model is also well identified. Our procedure exploits this invariance property to achieve identification of the risk premia of several observable factors even when they are highly correlated.

This invariance result effectively tells us that the risk premium of a factor g_t can be identified as long as we control for the exposures to a set of factors that span the entire factor space. In this paper, we do not assume these factors are directly observable. In the next section, we discuss how to use PCA to identify the space spanned by these latent factors.

4 Identification

There is fundamental indeterminacy in latent factor models. We can multiply β by any invertible matrix H on the right-hand side, and multiply γ and v_t by H^{-1} on the left-hand side, and both βv_t and $\beta\gamma$ will remain the same. Clearly, this implies that it is not possible to directly identify γ when not all factors are observed. The previous section shows that the risk premium associated with g_t is always equal to $\eta\gamma$, no matter how the latent factors are rotated. So to estimate it, we need to recover a rotation of the factor space. Below we show that we can identify $\eta\gamma$ and recover it from observed variables, returns r_t and the observable factor g_t , when $n \rightarrow \infty$.

Despite the potential unobserved heterogeneity due to α , the demeaned time series of each asset follows a standard approximate factor model (cf. [Chamberlain and Rothschild \(1983\)](#)), which, in matrix form, is given by

$$\bar{R} = \beta\bar{V} + \bar{U}. \quad (3)$$

[Bai and Ng \(2002\)](#) discuss identifying the number of latent factors p in a large n and large T setting. [Bai \(2003\)](#) argues that we can recover β and \bar{V} up to some invertible matrix H , only as $n \rightarrow \infty$. We denote them by βH^{-1} and $H\bar{V}$. From the cross-sectional equation:

$$E(r_t) = \iota_n \gamma_0 + \alpha + \beta\gamma = \iota_n \gamma_0 + \alpha + \beta H^{-1} H\gamma,$$

we can recover $H\gamma$ and identify γ_0 , if ι_n and β are not perfectly correlated and the cross-sectional mean of α is zero. On the other hand, [Assumption 2](#) leads to

$$\bar{G} = \eta\bar{V} + \bar{Z} = \eta H^{-1} H\bar{V} + \bar{Z}, \quad (4)$$

so we can recover ηH^{-1} if $\bar{V}\bar{V}^\top$ is non-singular. This implies that we can identify $\eta\gamma = \eta H^{-1} H\gamma$.

The success of the identification strategy is another example of the ‘‘blessings of dimensionality’’ ([Donoho et al. \(2000\)](#)). The large panel of cross-sectional returns certainly presents estimation challenges. However, it also provides a unique opportunity to identify and estimate the span of the latent factors that drive the asset returns. We can also identify and consistently estimate $\eta\gamma$, even without a consistent estimator of p , as long as we use some $\check{p} \geq p$ in estimation in the same spirit of [Moon and Weidner \(2015\)](#).⁴

5 The Three-Pass Estimator

We summarize the parameters of interest in $\Gamma = (\gamma_0 : (\eta\gamma)^\top)^\top$, where γ_0 is the zero-beta rate. We only use the observable data, (i.e., r_t and g_t , $t = 1, 2, \dots, T$). In light of the rotation invariance and

⁴[Bai \(2009\)](#) discusses the identification of finite-dimensional parameters in a linear panel regression model with interactive fixed effects, also in the large n and large T setting with p fixed. Allowing p to increase with n or T is interesting, and we leave it for future work.

identification results, we propose the following three-pass estimation procedure:

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

$$\widehat{V} = T^{1/2}(\xi_1 : \xi_2 : \dots : \xi_{\widehat{p}})^\top, \quad \text{and} \quad \widehat{\beta} = T^{-1}\bar{R}\widehat{V}^\top, \quad (5)$$

where $(\xi_1, \xi_2, \dots, \xi_{\widehat{p}})$ are the eigenvectors corresponding to the largest \widehat{p} eigenvalues of the matrix $n^{-1}T^{-1}\bar{R}^\top\bar{R}$, and \widehat{p} takes the following form:

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

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

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

$$\widetilde{\Gamma} := (\widetilde{\gamma}_0, \widetilde{\gamma}^\top)^\top = \left(\iota_n : \widehat{\beta} \right)^\top \left(\iota_n : \widehat{\beta} \right)^{-1} \left(\iota_n : \widehat{\beta} \right)^\top \bar{r}.$$

- (iii) **TSR.** Run another regression of g_t onto the estimated factors based on (4), so that

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

The estimator of the zero-beta rate and the risk premium for the observable factor g_t is obtained by combining the estimates of the second and third steps, given by

$$\widehat{\Gamma} := \begin{pmatrix} \widehat{\gamma}_0 \\ \widehat{\gamma} \end{pmatrix} := \begin{pmatrix} 1 & 0 \\ 0 & \widehat{\eta} \end{pmatrix} \widetilde{\Gamma} = \begin{pmatrix} \widetilde{\gamma}_0 \\ \widetilde{\eta}\widetilde{\gamma} \end{pmatrix}.$$

This estimator also has a more compact form:

$$\widehat{\gamma}_0 = \left(\iota_n^\top \mathbb{M}_{\widehat{\beta}} \iota_n \right)^{-1} \iota_n^\top \mathbb{M}_{\widehat{\beta}} \bar{r}, \quad \text{and} \quad \widehat{\gamma} = \bar{G}\widehat{V}^\top(\widehat{V}\widehat{V}^\top)^{-1} \left(\widehat{\beta}^\top \mathbb{M}_{\iota_n} \widehat{\beta} \right)^{-1} \widehat{\beta}^\top \mathbb{M}_{\iota_n} \bar{r}.$$

The first step presents an estimator of \widehat{p} , which we will show to estimate p consistently. This estimator is based on a penalty function, similar to the one [Bai and Ng \(2002\)](#) propose. It takes on a simpler form. 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 theory. We prefer this estimator for its simplicity in proofs. Other estimators are equally applicable, including but not limited to those proposed by [Onatski \(2010\)](#) and [Ahn and Horenstein \(2013\)](#). Also, similar to [Bai and Ng \(2002\)](#), when $T > n$, we can consider the PCA of the $n \times n$ matrix $n^{-1}T^{-1}\bar{R}^\top\bar{R}$ instead to accelerate the algorithm. The corresponding estimators in (5) are given by

$$\widehat{\beta} = n^{1/2}(\varsigma_1, \varsigma_2, \dots, \varsigma_{\widehat{p}}), \quad \text{and} \quad \widehat{V} = n^{-1}\widehat{\beta}^\top\bar{R},$$

where $(\varsigma_1, \varsigma_2, \dots, \varsigma_{\hat{p}})$ are the eigenvectors corresponding to the largest \hat{p} eigenvalues. The rest steps are identical.

In the second stage, 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 for the purpose of Γ estimation. This is different from the standard large T and fixed n case because the covariance matrix of u_t only matters at the order of $O_p(n^{-1} + T^{-1})$, whereas the convergence rate of $\hat{\Gamma}$ 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 approach cannot accomplish. Even though g_t can be multi-dimensional, the estimation for each observable factor is separate. Estimating the risk premium for one factor does not affect the estimation for the others at all, something that our estimator achieves without any omitted variable bias.

The three steps of our procedure suggest an alternative interpretation of the invariance result. As discussed in [Cochrane \(2009\)](#), in linear factor models the risk premium of any factor g_t is simply the univariate covariance of g_t with the stochastic discount factor m_t . Formally, in our setting, $m_t = 1 - \gamma^\top \Sigma_v^{-1} v_t$, so that $\eta\gamma = -\text{Cov}(g_t, m_t)$. The first two steps of the procedure (PCA and CSR) recover the stochastic discount factor m_t in the linear factor model (which is invariant to the rotation of the factors). The requirement of spanning the factor space is what allows to estimate the stochastic discount factor consistently. The TSR step effectively computes the univariate covariance between g_t and the stochastic discount factor m_t estimated in the first two stages. The invariance result follows from the fact that the latter stage only involves a univariate covariance with m_t , which itself is invariant to the rotation of the factor space.

6 Asymptotic Theory

In this section, we present the large sample distribution of our estimator as $n, T \rightarrow \infty$. Most 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.

We need more notation. We use $\lambda_j(A)$, $\lambda_{\min}(A)$, and $\lambda_{\max}(A)$ to denote the j th, the minimum, and the maximum eigenvalues of a matrix A . By convention, $\lambda_1(A) = \lambda_{\max}(A)$. In addition, we use $\|A\|_1$, $\|A\|_\infty$, $\|A\|$, and $\|A\|_F$ to denote the \mathbb{L}_1 norm, the \mathbb{L}_∞ norm, the operator norm (or \mathbb{L}_2 norm), and the Frobenius norm of a matrix $A = (a_{ij})$, that is, $\max_j \sum_i |a_{ij}|$, $\max_i \sum_j |a_{ij}|$, $\sqrt{\lambda_{\max}(A^\top A)}$, and $\sqrt{\text{Tr}(A^\top A)}$, respectively. We also use $\|A\|_{\text{MAX}} = \max_{i,j} |a_{ij}|$ to denote the \mathbb{L}_∞ norm of A on the vector space.

Let $(\mathbb{P}, \Omega, \mathcal{F})$ be the probability space. We say a sequence of centered multivariate random

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

6.1 Determining the Number of Factors

We start with assumptions on the idiosyncratic component u_t . Define, for any $t, t' \leq T$, $i, i' \leq n$:

$$\gamma_{n,tt'} = E \left(n^{-1} \sum_{i=1}^n u_{it} u_{it'} \right), \quad E(u_{it} u_{i't}) = \sigma_{ii',t}, \quad \text{and} \quad E(u_{it} u_{i't'}) = \sigma_{ii',tt'}.$$

Assumption 3. *There exists a positive constant K , such that for all n and T ,*

- (i) $T^{-1} \sum_{t=1}^T \sum_{t'=1}^T |\gamma_{n,tt'}| \leq K, \quad \gamma_{n,tt} \leq K.$
- (ii) $|\sigma_{ii',t}| \leq |\sigma_{ii'}|$, for some $\sigma_{ii'}$ and for all t . In addition, $n^{-1} \sum_{i=1}^n \sum_{i'=1}^n |\sigma_{ii'}| \leq K.$
- (iii) $n^{-1} T^{-1} \sum_{i=1}^n \sum_{i'=1}^n \sum_{t=1}^T \sum_{t'=1}^T |\sigma_{ii',tt'}| \leq K.$
- (iv) $E(u_t^\top u_{t'} - E u_t^\top u_{t'})^2 \leq Kn$, for all $t, t'.$

Assumption 3 is similar to Assumption C in Bai (2003), which imposes restrictions on the temporal and cross-sectional dependence and heteroskedasticity of u_t . Stationarity of u_t is not required. Eigenvalues of the residual covariance matrices $E(u_t u_t^\top)$ are not necessarily bounded. In fact, they can grow at the rate $n^{1/2}$. Therefore, this assumption is weaker than those for an approximate factor model in Chamberlain and Rothschild (1983).

Assumption 4. *The factor innovation V satisfies:*

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

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 4 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 5. *The factor loadings matrix β satisfies $\|\beta\|_{\text{MAX}} \leq K$. Moreover,*

$$\left\| n^{-1} \beta^\top \beta - \Sigma^\beta \right\| = o(1), \quad \text{as } n \rightarrow \infty,$$

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

Assumption 5 is the so-called pervasive condition for a factor model. 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 rate n , so that as n increases they can be separated from the idiosyncratic component whose eigenvalues are bounded or grow at a lower rate. This assumption precludes weak but priced latent factors. Onatski (2012) develops the inference methodology in a framework that allows for weak factors using a Pitman-drift-like asymptotic device. We leave the case of weak and latent factors for future work. However, we demonstrate the robustness of our empirical results with respect to the number of factors: the risk premia estimates and their significance remain similar even as more latent factors with lower eigenvalues are added to the estimation.

That said, our setup explicitly allows for weak *observable* factors. 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. Our methodology provides an alternative solution to the weak-identification problem (Kleibergen (2009)), which can be applied when n is large.

Finally, the loadings here are non-random for convenience. In contrast, Gagliardini et al. (2016) consider random loadings because of their sampling scheme from a continuum of assets. Our assumption is more commonly seen in the literature, see, Connor and Korajczyk (1988), Bai (2003), and Fan et al. (2013).

Theorem 1. *Under Assumptions 1, 2, 3, 4, and 5, and suppose that as $n, T \rightarrow \infty$, $\phi(n, T) \rightarrow 0$, and $\phi(n, T)/(n^{-1/2} + T^{-1/2}) \rightarrow \infty$, we have $\hat{p} \xrightarrow{p} p$.*

By a simple conditioning argument, we can assume that $\hat{p} = p$ when developing the limiting distributions of the estimators, see Bai (2003). In the sequel, we assume $\hat{p} = p$. Even though we cannot always find the true number of factors in a finite sample, our derivation in Section 6.4 shows that as long as $\hat{p} \geq p$, we can estimate the parameters Γ consistently.

6.2 Limiting Distribution of $\hat{\Gamma}$

In this section, we derive the asymptotic distribution of the estimator $\hat{\Gamma}$. We need more assumptions that link the factor proxies g_t to the latent factors v_t .

Assumption 6. *The residual innovation Z satisfies:*

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

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

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

Similar to Assumption 4, Assumption 6 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, so that it can be justified for non-tradable factor proxies in the empirical applications.

Assumption 7. For any $t \leq T$, and $i, j \leq p$, $l \leq d$, the following moment conditions hold:

$$\begin{aligned} (i) \quad & \mathbb{E} \left(\sum_{s=1}^T \sum_{k=1}^n v_{js} u_{ks} \right)^2 \leq KnT. \\ (ii) \quad & \mathbb{E} \sum_{k=1}^n \left(\sum_{s=1}^T v_{js} u_{ks} \right)^2 \leq KnT. \\ (iii) \quad & \mathbb{E} \left(\sum_{s=1}^T \sum_{k=1}^n v_{is} u_{ks} \beta_{kj} \right)^2 \leq KnT. \end{aligned}$$

Assumption 7 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 8. For any $t \leq T$, and $k \leq n$, $l \leq d$, define $\sigma_{lk,t}^{zu} = \mathbb{E}(z_{lt} u_{kt})$. The following moment conditions hold:

$$\begin{aligned} (i) \quad & |\sigma_{lk,t}^{zu}| \leq |\sigma_{lk}^{zu}| \leq K, \quad \text{for some } \sigma_{lk}^{zu} \text{ and for all } t. \quad \text{In addition, } \sum_{k=1}^n |\sigma_{lk}^{zu}| \leq K. \\ (ii) \quad & \mathbb{E} \sum_{k=1}^n \left(\sum_{s=1}^T (z_{ls} u_{ks} - \mathbb{E}(z_{ls} u_{ks})) \right)^2 \leq KnT. \\ (iii) \quad & \mathbb{E} \left(\sum_{s=1}^T \sum_{k=1}^n (z_{ls} u_{ks} - \mathbb{E}(z_{ls} u_{ks})) \beta_{kj} \right)^2 \leq KnT. \end{aligned}$$

Similar to Assumption 7, Assumption 8 specifies the restrictions on the covariances between the idiosyncratic components and the measurement error. If z_t and u_t are independent, (i) - (iii) 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 1 and 2. It is thereby reasonable to allow for covariances between z_t and u_t . For nontradable factors, z_{ts} can also be correlated with u_t in general.

Assumption 9. The cross-sectional pricing error α is i.i.d., independent of u and v , with mean 0, standard deviation $\sigma^\alpha > 0$, and a finite fourth moment.

Assumption 9 dictates the behavior of pricing errors in model (1). The APT predicts α to be 0. 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 we 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.

Assumption 10. *There exists a $p \times 1$ vector β_0 , such that $\|n^{-1}\beta^\top \iota_n - \beta_0\|_{\text{MAX}} = o(1)$. Moreover, the matrix*

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

The convergence of $n^{-1}\beta^\top \iota_n$ in Assumption 10 resembles the law of large numbers for factor loadings. The rank condition ensures that in the limit the factor loadings are not perfectly correlated in the cross section.

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

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

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

$$\begin{aligned} \Pi_{11} &= \lim_{T \rightarrow \infty} \frac{1}{T} \mathbb{E} (\text{vec}(ZV^\top) \text{vec}(ZV^\top)^\top), \\ \Pi_{12} &= \lim_{T \rightarrow \infty} \frac{1}{T} \mathbb{E} (\text{vec}(ZV^\top) \iota_T^\top V^\top), \\ \Pi_{22} &= \lim_{T \rightarrow \infty} \frac{1}{T} \mathbb{E} (V \iota_T \iota_T^\top V^\top). \end{aligned}$$

Assumption 11 describes the joint asymptotic distribution of ZV^\top and $V \iota_T$. Because the dimensions of these random processes are finite, this assumption is a fairly standard result of some central limit theorem for mixing processes, (e.g., Theorem 5.20 of White (2000)). Not surprisingly, it is stronger than Assumption 4, which is sufficient for identification and consistency.

We now present the main theorem of the paper:

Theorem 2. *Under Assumptions 1 – 11, and suppose $\hat{p} \xrightarrow{p} p$, then as $n, T \rightarrow \infty$, we have*

$$\begin{aligned} n^{1/2} (\hat{\gamma}_0 - \gamma_0) &\xrightarrow{\mathcal{L}} \mathcal{N} \left(0, \left(1 - \beta_0^\top (\Sigma^\beta)^{-1} \beta_0 \right)^{-1} (\sigma^\alpha)^2 \right), \\ (T^{-1}\Phi + n^{-1}\Upsilon)^{-1/2} (\hat{\gamma} - \eta\gamma) &\xrightarrow{\mathcal{L}} \mathcal{N} (0, \mathbb{I}_d), \end{aligned}$$

where the asymptotic covariance matrices Φ and Υ are given by

$$\begin{aligned}\Phi &= \left(\gamma^\top (\Sigma^v)^{-1} \otimes \mathbb{I}_d \right) \Pi_{11} \left((\Sigma^v)^{-1} \gamma \otimes \mathbb{I}_d \right) + \left(\gamma^\top (\Sigma^v)^{-1} \otimes \mathbb{I}_d \right) \Pi_{12} \eta^\top \\ &\quad + \eta \Pi_{21} \left((\Sigma^v)^{-1} \gamma \otimes \mathbb{I}_d \right) + \eta \Pi_{22} \eta^\top, \quad \text{and} \\ \Upsilon &= (\sigma^\alpha)^2 \eta \left(\Sigma^\beta - \beta_0 \beta_0^\top \right)^{-1} \eta^\top.\end{aligned}$$

Remarkably, Theorem 2 does not impose any restrictions on the relative rates of n and T . Moreover, 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. The large cross section of testing assets extracts all the relevant factors from their time-series variations, which help correct the biases due to missing controls and measurement error.

6.3 Goodness-of-Fit Measures

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

$$R_v^2 = \frac{\gamma^\top (\Sigma^\beta - \beta_0 \beta_0^\top) \gamma}{(\sigma^\alpha)^2 + \gamma^\top (\Sigma^\beta - \beta_0 \beta_0^\top) \gamma}.$$

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:

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

To calculate these measures in a sample, we use

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

where $\widehat{G} = g - \bar{g}$ 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 3. *Under Assumptions 1 – 11, and suppose $\widehat{p} \xrightarrow{p} p$, then as $n, T \rightarrow \infty$, we have*

$$\widehat{R}_v^2 \xrightarrow{p} R_v^2 \quad \text{and} \quad \widehat{R}_g^2 \xrightarrow{p} R_g^2.$$

6.4 Robustness of the Choice of p

Although \widehat{p} is a consistent estimator of p , it is possible that in finite sample $\widehat{p} \neq p$. In fact, without a consistent estimator of \widehat{p} , as long as our choice, denoted by \check{p} , is greater than or equal to p , the estimator based on \check{p} , denoted by $\check{\Gamma} = (\check{\gamma}_0 : \check{\gamma}^\top)^\top$, is consistent. This result is similar in spirit to that of [Moon and Weidner \(2015\)](#), who establish that, for inference on the regression coefficients in a linear panel model with interactive fixed effect, it is not necessary to estimate p consistently, as long as the number of factors we use, \check{p} , is greater than or equal to p .

To prove this result, we need to apply the random matrix theory in [Bai \(1999\)](#) to analyze the asymptotic behavior of the extreme eigenvalues of the sample covariance matrix of u_t . For this reason, stronger assumptions on u_t are required.

Theorem 4. *Suppose Assumptions 1 – 11 hold. Also, suppose that u_t are i.i.d. centered random variables with finite fourth moment, and that z_t is independent of v_t and u_t . If $\check{p} \geq p$, we have*

$$\check{\Gamma} - \widehat{\Gamma} = O_p(n^{-1/2} + T^{-1/2}).$$

The above theorem establishes the desired consistency of $\check{\Gamma}$. While we cannot establish its asymptotic distribution, simulation exercises suggest that the differences between the asymptotic variances of $\check{\Gamma}$ and $\widehat{\Gamma}$ are tiny.

6.5 Limiting Distribution of \widehat{g}_t

As discussed above, our framework allows for measurement error in the observable factor proxies g . [Theorem 3](#) above proves that we can clean these errors up with identified latent factors. Moreover, we can conduct inference on g at each t , given additional assumptions. Similar to [Bai \(2003\)](#), these assumptions are essential to derive the central limit result for the rotated factors and their loadings.

Assumption 12. *The following conditions hold:*

$$\begin{aligned} (i) \quad & \sum_{t'=1}^T |\gamma_{n,tt'}| \leq K, \quad \text{for all } t. \\ (ii) \quad & \sum_{i'=1}^n |\tau_{ii'}| \leq K, \quad \text{for all } i. \end{aligned}$$

This assumption is identical to Assumption E in [Bai \(2003\)](#). It restricts the eigenvalues of $E(u_t u_t^\top)$ and $E(u_t^\top u_t)$ to be bounded as the dimension increases, because the \mathbb{L}_∞ -norm is stronger than the operator norm for symmetric matrices.

Assumption 13. *For each t , as $n \rightarrow \infty$,*

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

where, writing $\beta = (\beta_1 : \beta_2 : \dots : \beta_n)^\top$,

$$\Omega_t = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \sum_{i'=1}^n \beta_i \beta_{i'}^\top \mathbb{E}(u_{it} u_{i't}). \quad (6)$$

Assumption 13 is identical to Assumption F3 in Bai (2003), which is used to describe the asymptotic distribution of \hat{v}_t at each point in time.

Theorem 5. *Under Assumptions 1 – 8, 11, 12, and 13, and suppose that $\hat{p} \xrightarrow{p} p$, then as $n, T \rightarrow \infty$, we have*

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

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

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

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

6.6 Asymptotic Variances Estimation

We develop consistent estimators of the asymptotic covariances in Theorems 2 and 5. We can estimate them for risk premia as:

$$\begin{aligned} \hat{\Phi} &= \left(\tilde{\gamma}^\top (\hat{\Sigma}^v)^{-1} \otimes \mathbb{I}_d \right) \hat{\Pi}_{11} \left((\hat{\Sigma}^v)^{-1} \tilde{\gamma} \otimes \mathbb{I}_d \right) + \left(\tilde{\gamma}^\top (\hat{\Sigma}^v)^{-1} \otimes \mathbb{I}_d \right) \hat{\Pi}_{12} \hat{\eta}^\top + \hat{\eta} \hat{\Pi}_{21} \left((\hat{\Sigma}^v)^{-1} \tilde{\gamma} \otimes \mathbb{I}_d \right) + \hat{\eta} \hat{\Pi}_{22} \hat{\eta}^\top, \\ \hat{\Upsilon} &= \hat{\sigma}^{\alpha^2} \hat{\eta} \left(\hat{\Sigma}^\beta - \hat{\beta}_0 \hat{\beta}_0^\top \right)^{-1} \hat{\eta}^\top, \end{aligned}$$

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

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

$$\widehat{\Pi}_{22} = \frac{1}{T} \sum_{t=1}^T \widehat{v}_t \widehat{v}_t^\top + \frac{1}{T} \sum_{m=1}^q \sum_{t=m+1}^T \left(1 - \frac{m}{q+1}\right) (\widehat{v}_{t-m} \widehat{v}_t^\top + \widehat{v}_t \widehat{v}_{t-m}^\top),$$

and

$$\widehat{Z} = \bar{G} - \widehat{\eta} \widehat{V}, \quad \widehat{\Sigma}^\beta = n^{-1} \widehat{\beta}^\top \widehat{\beta}, \quad \widehat{\Sigma}^v = T^{-1} \widehat{V} \widehat{V}^\top, \quad \widehat{\beta}_0 = n^{-1} \widehat{\beta}^\top \iota_n, \quad \widehat{\sigma}^{\alpha^2} = n^{-1} \left\| \bar{r} - (\iota_n : \widehat{\beta}) \widehat{\Gamma} \right\|_F^2,$$

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

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

Assumption 14. *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 $t', t \leq T$,*

$$\mathbb{E}(u_t^\top u_{t'} - \mathbb{E}u_t^\top u_{t'})^4 \leq Kn^2, \quad \mathbb{E}\|\beta^\top u_t\|^4 \leq Kn^2.$$

Assumption 14 ensures that the factors and their loadings are consistent up to some rotations under the max norm. Fan et al. (2011) and Fan et al. (2015) also adopt it.

Theorem 6. *Under Assumptions 1 - 12, 14, and suppose that $\widehat{p} \xrightarrow{p} p$, then as $n, T \rightarrow \infty$, $n^{-3}T \rightarrow 0$, $q(T^{-1/4} + n^{-1/4}) \rightarrow 0$, $\widehat{\Phi} \xrightarrow{p} \Phi$ and $\widehat{\Upsilon} \xrightarrow{p} \Upsilon$.*

To estimate the asymptotic covariance matrices Ψ_{1t} and Ψ_{2t} in Theorem 5, we can simply replace $v_t, \Sigma^v, \Pi_{11}, \Pi_{12}, \Pi_{22}, \eta, \Sigma^\beta$ 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 to impose additional assumptions, because it is rather challenging to estimate, when we allow heteroskedasticity and correlation in both the time series and cross section.

We consider two scenarios that are relevant in practice.

Assumption 15. *Either of the following assumptions holds:*

(i) *The innovation u_{it} is cross-sectionally independent, i.e., $\mathbb{E}(u_{it}u_{jt}) = 0$, for any $t \leq T$, $1 \leq i \neq j \leq n$.*

(ii) *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*

$$s_n = \max_{1 \leq i \leq n} \sum_{i'=1}^n |\Sigma_{ii'}^u|^h, \quad \text{where } s_n = o_p \left(\left(\omega_T^{1-h} + n^{-1} + T^{-1} \right)^{-1} \right).$$

Under Assumption 15(i), (6) and its estimator can be rewritten as

$$\Omega_t = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \beta_i \beta_i^\top \mathbb{E}(u_{it}^2), \quad \text{and} \quad \widehat{\Omega}_t = \frac{1}{n} \sum_{i=1}^n \widehat{\beta}_i \widehat{\beta}_i^\top \widehat{u}_{it}^2, \quad (7)$$

where, writing $\widehat{U} = (\widehat{u}_{it})$, $\widehat{U} := \bar{R} - \widehat{\beta} \widehat{V}$.

With Assumption 15(ii), (6) and its estimator can be rewritten as

$$\Omega = \lim_{n \rightarrow \infty} \frac{1}{n} \beta^\top \Sigma^u \beta, \quad \text{and} \quad \widehat{\Omega}_t = \widehat{\Omega} = \frac{1}{n} \widehat{\beta}^\top \widehat{\Sigma}^u \widehat{\beta}, \quad (8)$$

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

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

and $s_{ii'}(z) : \mathbb{R} \rightarrow \mathbb{R}$ is a general thresholding function with an entry dependent threshold $\tau_{ii'}$ such that (i) $s_{ii'}(z) = 0$ if $|z| < \tau_{ii'}$; (ii) $|s_{ii'}(z) - z| \leq \tau_{ii'}$; and (iii) $|s_{ii'}(z) - z| \leq a\tau_{ii'}^2$, 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:

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

where $\widehat{\Omega}_t$ is given by either (7) or (8).

Theorem 7. *Under Assumptions 1 – 15, we have*

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

7 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 below. 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 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 multivariate normal using the calibrated means and covariances.

We report in Tables 1, 2, and 3 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, $\check{p} = 4, 5,$ 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 estimator with $\check{p} = 5$ dominates the other estimators, in particular when n and T are large. Instead, the two-pass estimates have substantial biases. For example, the bias for the market factor premium is so large that the its two-pass estimates are all negative ($\text{True} + \text{Bias} < 0$) even when n and T are large, which actually matches what we find using real data and has been documented in the literature, as we discuss below. The three-pass estimator with $\check{p} = 4$ has an obvious bias, compared to the cases with $\check{p} = 5$ and $6,$ because an omitted-factor problem still affects it.

We then plot in Figure 1 the histograms of the standardized risk premia estimates using Fama-MacBeth standard errors for the two-pass estimator (left column) and the estimated asymptotic standard errors for the three-pass method with $\check{p} = 5$ (right column).⁵ The histograms on the left deviate substantially from the standard normal distribution, whereas those on the right match the normal distribution very well, which verifies our central limit results. There exist some small higher order biases for γ_0 and the market risk premium, which would disappear with a larger n and T in simulations not included here.

Finally, we report in Table 4 the estimated number of factors. We choose $\phi(n, T) = K(\log n + \log T)(n^{-1/2} + T^{-1/2}),$ where $K = 0.5 \times \hat{\lambda},$ $\hat{\lambda}$ is the median of the first p_{\max} eigenvalues of $n^{-1}T^{-1}\bar{R}^T\bar{R}.$ 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.⁶ 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 4.

8 Empirical Analysis

In this section we apply our three-pass methodology to the cross-section of equities. We estimate the risk premia of several factors, both traded and not traded, and show how our results differ from standard two-pass cross-sectional regressions (or Fama-MacBeth regressions since we use their method for calculating standard errors), which ignore the potential omitted factors in the data.

⁵We 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.

⁶The 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.

8.1 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 momentum, and 25 portfolios sorted by size and beta. 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.⁷

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 do not depend on whether other factors are included in g_t . Here is a list of models and corresponding observable factors g_t included:⁸

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).
5. Four factors from the Q-factor model (*HXZ*) of Hou et al. (2014), which include the market return, ME (size), IA (investment), ROE (profitability).

⁷See the description of all portfolio construction on Kenneth French's website: http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

⁸Factor time series for models 1-4 are obtained from Kenneth French's website; for model 5, from Lu Zhang; for models 6-7, from AQR's website; for model 8, from the Federal Reserve Bank of St. Louis; for model 9, from Sydney Ludvigson's website; for model 10, from Lubos Pastor's website; for model 11, from Bryan Kelly's website.

6. Betting-against-beta factor (*BAB*).
7. Quality-minus-junk factor (*QMJ*).
8. 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.
9. The first three principal components of 279 macro-finance variables constructed by [Jurado et al. \(2015\)](#) (*JLN*), also available at the monthly frequency. We estimate a VAR(1) with those three principal components, and use innovations as factors.
10. The liquidity factor from [Pástor and Stambaugh \(2003\)](#).
11. Two intermediary capital factors, one from [He et al. \(2016\)](#) and one from [Adrian et al. \(2014\)](#).

8.2 Factors from the Large Panel of Returns

The first step for estimating the observable factor risk premia is to determine the latent factor model dimension, p . Figure 2 (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 in the right panel we plot the eigenvalues excluding the first one. We observe a noticeable decrease in the eigenvalues after four and six factors, and our estimator suggests using four factors. As discussed in Section 6, our analysis is consistent as long as the number of factors \hat{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.

After extracting the factors via PCA, the second pass in the three-pass procedure estimates the risk premia of the latent factors via cross-sectional regressions (CSR). We cannot interpret these risk premia in economic terms, as opposed to the risk premia of observable factors. The estimated zero-beta rate from the APT model is 55bp per month, close to the 40bp of the average risk-free rate return over the sample.

The model has a cross-sectional R_p^2 of 65%, indicating that it accounts for much of the cross-sectional variation in expected returns for the 202 test portfolios, but leaving some unexplained variation. We report in Figure 3 the actual and predicted 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 (FF25 and momentum) than others (industry), but overall the factor model with six factors performs relatively well.

8.3 Risk Premia Estimates for Observable Factors

Tables 5 and 6 report the estimates of observable factor risk premia. Each factor (or set of factors) g_t corresponds to a panel of the tables; the tested g_t appears in the first column. In each panel, the rows correspond to the coefficients of the cross-sectional regressions (intercept $\hat{\gamma}_0$ and the risk premia

corresponding to the factors $g_t, \hat{\eta}\hat{\gamma}$). The number of observations T is 630 in all cases except for the *HXZ* model (where $T = 588$), *JLN* ($T = 580$) and the intermediary-capital model ($T = 516$).

Across columns, the tables report information about the average returns of the factors (when traded), standard Fama-MacBeth estimates of the risk premia that ignore potential additional factors, and results of the three-pass procedure using different numbers of latent factors, from four to six.

To illustrate the table content, consider for example the second panel, corresponding to the Fama-French three-factor model. The first column reports the average monthly returns for the three factors (RmRf, SMB, HML) over the sample period: respectively 50bp, 23bp, and 34bp. The number in the “intercept” row reports the average value of the risk-free rate R_f over the sample period (in this case, 40bp).

The second set of columns corresponds to the standard Fama-MacBeth estimation of the intercept and the three risk premia using all of the 202 portfolios. The results of this exercise line up well with the previous literature. The zero-beta rate estimate is approximately 1.5% per month, more than 100bp higher than the average risk-free rate. The risk premium estimate associated with the market return is negative, and significantly so. HML has a high and significant risk premium of 23bp per month, close to the time-series average return of the HML portfolio. Finally, size (SMB) has a smaller and statistically insignificant risk premium.

The remainder of the tables report the estimates for the three-pass procedure. As discussed above, we repeat the exercise for $\check{p} = 4, 5, \text{ and } 6$. The estimates are stable across the number of factors used, consistent with the theoretical result that adding extra factors does not affect the validity of our procedure. We discuss the results for the case $\check{p} = 4$ in detail.

The estimates of the zero-beta rate and the risk premia for the three factors differ substantially from the estimates obtained using the standard Fama-MacBeth regression. First, the zero-beta rate estimate is 55bp, just 15bp per month above the risk-free rate. Second, the market risk premium estimate is positive, significant, and of a magnitude close to the average return in the data (the risk premium estimate is 37bp in the model, whereas the average return of the market portfolio in the data is 50bp over the risk-free rate, and 35bp over the estimated zero-beta rate). The risk premium associated with HML is stable at a significant 21bp; the risk premium associated with size is significant and equal to 23bp, matching exactly the average return in the data. Results for the *FF3* model, therefore, are substantially different when estimated via Fama-MacBeth regression or via three-pass regressions.

The third column of each of the three-pass results reports the R_g^2 of the time-series regression (TSR) of the observed g_t onto the latent factors; we refer to this as R_g^2 . Recall that if the factors driving the returns’ cross section entirely span g_t , we should expect to find R_g^2 close to 100%. However, if the observed g_t is just a noisy proxy for some of the fundamental factors, this R_g^2 will reflect the amount of noise in the observed g_t . In the data, we find interesting heterogeneity among the three factors of *FF3* with respect to their R_g^2 . The market and size portfolios have R_g^2 close to 100%; HML displays greater noise, with an estimated R_g^2 of about 67%.

Figure 4 shows the time series of cumulated (mean-zero) innovations in the original and cleaned factors for the Fama-French three-factor model. The figures present a graphical representation of the variation in the original factors captured by the principal components, corresponding to the R_g^2 reported in the table. The figure shows that all three factors correlate highly with the estimated latent factors.

Tables 5 and 6 report the results for the remaining factors and factor models we study, both traded factors (e.g., MOM) and non-traded factors (e.g., IP). We summarize here the main results, highlighting in particular the differences that emerge when estimating the model using our three-pass procedure rather than the standard Fama-MacBeth regression.

Zero-beta rate. Whereas for most of the models estimated via standard Fama-MacBeth two-pass regression the zero-beta rate is much larger than the observed risk-free rate (typically between 50 and 100bp above it), the zero-beta rate estimated from the three-pass procedure is mostly 15-20bp greater than the risk-free rate on average, and statistically insignificantly so. This is due to the fact that the latent model (with four to six factors) is able to capture a greater fraction of the overall level of equity-portfolio risk premia.

The market risk premium. A classic result in the empirical asset pricing literature is the typically negative estimate of the risk premium for market risk from cross-sectional regressions. This result highlights a potential misspecification for these regressions: under the assumptions of a linear factor model, for tradable factors the cross-sectional estimate of risk premium should correspond to the time-series estimate of the average excess return of the portfolio.

The three-pass approach allows us to control for more factors beyond the observable ones, and at the same time exploit the beta spread across the 202 portfolios to pin down the risk premium of each observable factor better. The result is that the risk premium estimate for exposure to the aggregate stock market is positive and significant at 37bp, close to the average excess return of the market portfolio. It is also useful to note that our procedure guarantees that the estimated risk premium for a factor does *not* depend on whether it is estimated together with other observable factors or by itself; therefore, the market risk premium will be the same when estimating the *CAPM*, the Fama-French three-factor and five-factor models, or the Q-factor model of *HXZ*.

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 aggregate risks.

Other tradable factors. The table shows that using the three-pass method, the cross-sectional risk premia estimates for tradable factors are close to the time-series average excess returns of the portfolios themselves – not only for the market portfolio as described above, but for the vast majority of the tradable factors we examine. For example, the risk premium associated with HML is close to zero in the FF5 model when estimating it using standard two-pass regression, while it is positive, significant, and close to the time-series average return when using the three-pass method.

This result is important because it helps rule out misspecifications of our linear factor model. For tradable factors, risk premia can be computed in two ways: by estimating the time-series average excess return of the factor (a model-free estimator), or by computing the slope of two- and three-pass estimators under the assumptions of the linear factor model. Any misspecification that affects our methodology would bias the latter but not the former. Comparing the two estimates when possible (i.e., for tradable factors) is therefore a simple way to assess whether different types of misspecification – for example, factors with low variance and high risk premia missed by the PCA analysis, nonlinearities, correlated time variation in betas and risk premia – affect our estimator, at least as far as the tradable factors are concerned. The fact that we do not see economically large differences between the two estimators for tradable factors mitigates the misspecification concerns for non-tradable factors (for which this form of validation is not possible).

Macroeconomic factors. We consider two different macroeconomic factors. The first one is real industrial production growth (IP), which captures fluctuations in the real economy and is available at the monthly frequency. In the classic Fama-MacBeth regression, innovations in IP display a significantly *negative* risk premium. The three-pass procedure instead finds it insignificant; in addition, IP is effectively uncorrelated with the factors that seem to price returns: the R_g^2 for IP is about 2%. The three-pass procedure therefore identifies industrial production as essentially a spurious factor.

This can also be seen graphically by looking at the last panel of Figure 4, which reports the cumulated innovations in IP and the version cleaned of measurement error. Most of the variation disappears from the cleaned factor, suggesting that the factor is mostly spurious within our framework.

The same happens for the *JLN* macro factors: standard two-step Fama-MacBeth regression finds a large and statistically significant risk premium for the first factor. However, the three-pass method reveals that that factor is essentially pure noise ($R_g^2 = 1\%$), as are the other three factors. All factors have an insignificant risk premium.

Market frictions. Some of the most interesting results appear with respect to two theoretically motivated nontradable factors related to market frictions: liquidity and intermediation capital.

By simply running Fama-MacBeth regression, the [Pástor and Stambaugh \(2003\)](#) liquidity factor does not appear to be priced in this cross section of 202 portfolios: its risk premium is 2bp per month, with a standard error of 97bp. The three-pass analysis shows instead that the liquidity factor commands a statistically significant risk premium of about 26bp per month.

The prices of the two intermediary factors of [He et al. \(2016\)](#) and [Adrian et al. \(2014\)](#) also vary with the estimation method. Relative to the results obtained using standard Fama-MacBeth regression, the three-pass method finds a slightly smaller (but still very large and significant) risk premium for the [Adrian et al. \(2014\)](#) proxy for intermediary capital. The similar factor built by [He et al. \(2016\)](#), instead, appears to have a risk premium of zero when estimated via Fama-MacBeth

regression, whereas the risk premium appears much larger – 30bp – when estimated using the three-pass method.⁹

Overall, the three-pass procedure shows much stronger support for both types of factors (liquidity-based and intermediary-based) than standard Fama-MacBeth regressions do.

8.4 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 factors, and therefore reveals the exposures of the observable factors to the fundamental priced factors.

In Table 7 we decompose the variance of g_t explained by the set of factors v_t into the components due to each individual factor (which is possible because factors v_t are orthogonal to each other). Each row of the table, therefore, sums up to 100%. This allows us to highlight which fundamental factors are most responsible for the variation of the observable factors. Note that the factors are ordered by their eigenvalues (largest to smallest).

The first row shows that the market return loads mostly onto the first factor, (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 portfolio-based models (such as *FF5* and *HXZ*) show interesting variation in the exposure of observable factors to the latent ones. For example, *SMB* loads on both the first and second factors, *HML* mostly on the third one, and *Momentum* almost exclusively on the fourth factor. *RMW* loads substantially on at least four factors (including the sixth 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 these fundamental factors in nontrivial ways. *IP* is mostly exposed to the sixth factor (to which *RMW* and *CMA* are exposed as well). The first *JLN* factor seems exposed uniformly to all risks sources (but its overall risk premium is insignificant because these exposures are small in absolute level, and the factor is very noisy, as explained above).

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.

8.5 From the Individual Risk Premium to Multifactor Risk Premia

The three-pass method presented in this paper achieves an estimate of risk premia (and their standard error) associated with each factor by relating each factor in g_t *individually* to the priced latent factors v_t . At the same time, as discussed in Section 2, the risk premia we estimate can be interpreted as those of a *multifactor* model in which all observable factors in g_t appear directly (together with some additional latent factors). The rotation-invariance result of Section 2 guarantees that this interpretation always holds. Similarly, the standard errors reported in Tables 5 and 6 are the same

⁹The economic significance is low for this factor in monthly equity data, as was already pointed out in [He et al. \(2016\)](#); our results here match the results in that paper, which only controls for the market in two-pass regressions.

standard errors as those one would obtain in a two-pass cross-sectional regression using as factors \hat{g}_t and any $(p - d)$ latent factor estimates.

Importantly, this is true even when the factors in \hat{g}_t are highly correlated. For example, the market and liquidity factors both load highly on the first principal component and are therefore highly correlated. One might expect that in a two-pass regression where both factors are included, it would be hard to separately identify the two risk premia. Instead, the two individual risk premia are well identified (as can be seen from the standard errors) because these two factors are simply a rotation of a well-identified model, as implied by the invariance result.

9 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 when running cross-sectional regressions. In these cases, the risk premium for observable factors will be consistent 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.

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 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 control for omitted factors in the cross-sectional regression. It also explicitly takes into account the possibility of measurement error in any observed factor.

Application of the methodology to workhorse factor models using equity test assets yields several compelling results. Contrary to most existing estimates, we find out that the risk premium estimate associated with market risk exposure is positive and significant, and close to the time-series average excess return of the market portfolio. This confirms that our methodology correctly recovers the risk premia of the market (and similar results hold for most other tradable factors), thus mitigating misspecification concerns. The most interesting results appear for non-tradable factors. Many standard macroeconomic factors appear insignificant, whereas factors related to various market frictions (like liquidity and intermediary leverage) appear strongly significant when considered as part of richer

linear pricing models that include additional factors.

Although in this paper we apply the three-pass methodology to a standard set of 202 equity test portfolios, this methodology can be directly applied to even larger cross sections, for example those that include other asset classes or international markets. This is because the inference is derived under the assumption that the number of assets n increases to ∞ . We leave for future work a study of larger cross sections that may yield novel insights about the pricing of securities across markets and asset classes.

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.
- Bai, J. (2003). Inferential Theory for Factor Models of Large Dimensions. *Econometrica* 71(1), 135–171.
- Bai, J. (2009). Panel Data Models With Interactive Fixed Effects. *Econometrica* 77(4), 1229–1279.
- 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 (2006a, July). Confidence Intervals for Diffusion Index Forecasts and Inference for Factor-Augmented Regressions. *Econometrica* 74(4), 1133–1150.
- Bai, J. and S. Ng (2006b). Evaluating latent and observed factors in macroeconomics and finance. *Journal of Econometrics* 131(1), 507–537.
- Bai, J. and S. Ng (2008). Forecasting economic time series using targeted predictors. *Journal of Econometrics*.
- Bai, J. and G. Zhou (2015). Fama–MacBeth two-pass regressions: Improving risk premia estimates. *Finance Research Letters* 15, 31–40.
- Bai, Z. (1999). Methodologies in spectral analysis of large dimensional random matrices: A review. *Statistica Sinica* 9, 611–677.
- 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.

- 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.
- Clarke, C. (2016). The level, slope and curve factor model for stocks. *Slope and Curve Factor Model for Stocks (January 23, 2016)*.
- 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?rench 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.
- Donoho, D. L. et al. (2000). High-dimensional data analysis: The curses and blessings of dimensionality. *AMS Math Challenges Lecture*, 1–32.
- 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*.

- Forni, M. and L. Reichlin (1998, July). Let's Get Real: A Factor Analytical Approach to Disaggregated Business Cycle Dynamics. *The Review of Economic Studies* 65(3), 453–473.
- Gagliardini, P., E. Ossola, and O. Scaillet (2016). Time-varying risk premium in large cross-sectional equity datasets. *Econometrica* 84(3), 985–1046.
- 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 (2014). Digesting anomalies: An investment approach. *Review of Financial Studies*, *Forthcoming*.
- Jagannathan, R., G. Skoulakis, and Z. Wang (2010). The analysis of the cross section of security returns. *Handbook of financial econometrics* 2, 73–134.
- Jagannathan, R. and Z. Wang (1996). The Conditional CAPM and the Cross-Section of Expected Returns. *The Journal of Finance* 51(1), 3–53.
- 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.
- Jurado, K., S. C. Ludvigson, and S. Ng (2015). Measuring uncertainty. *The American Economic Review* 105(3), 1177–1216.
- Kan, R. and C. Robotti (2008). Specification tests of asset pricing models using excess returns. *Journal of Empirical Finance* 15(5), 816–838.

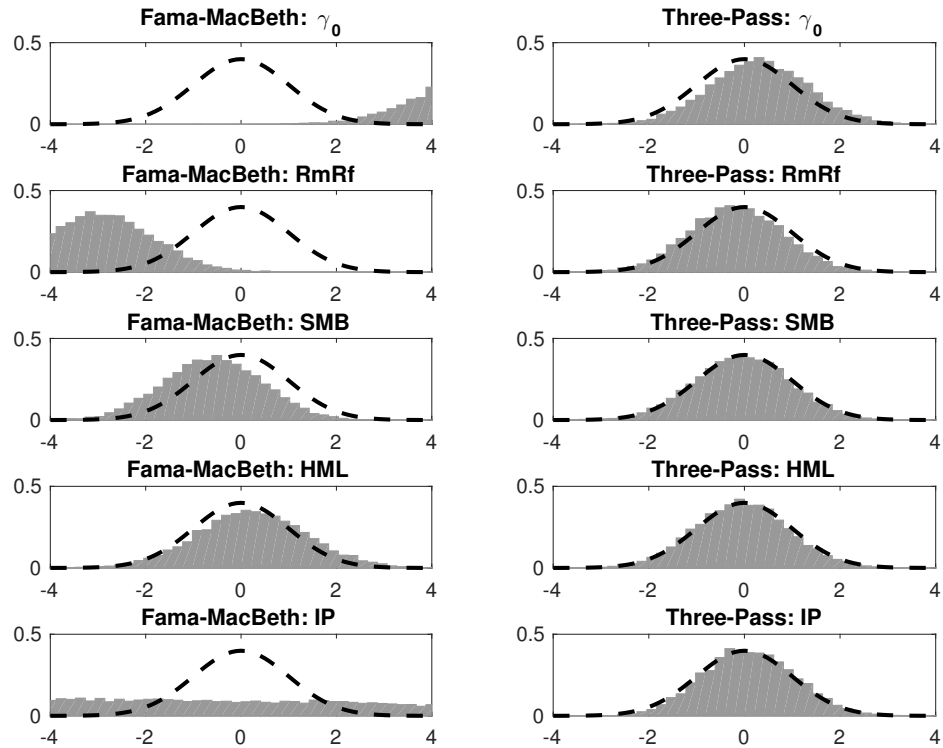
- Kan, R. and C. Robotti (2009). Model comparison using the hansen-jagannathan distance. *Review of Financial Studies* 22(9), 3449–3490.
- Kan, R. and C. Robotti (2012). Evaluation of Asset Pricing Models Using Two-Pass Cross-Sectional Regressions. In *Handbook of Computational Finance*, pp. 223–251. Springer.
- 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. and S. Pruitt (2013). Market expectations in the cross-section of present values. *The Journal of Finance*.
- Kelly, B. and S. Pruitt (2015). The three-pass regression filter: A new approach to forecasting using many predictors. *Journal of Econometrics*.
- 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.
- Kozak, S., S. Nagel, and S. Santosh (2015). Interpreting factor models. Technical report, University of Michigan.
- Lehmann, B. N. and D. M. Modest (1988). The empirical foundations of the arbitrage pricing theory. *Journal of Financial Economics* 21(2), 213–254.
- Lettau, M. and S. Ludvigson (2001). Resurrecting the (c)capm: A cross-sectional test when risk premia are time-varying. *Journal of Political Economy* 109(6), 1238–1287.
- Lewellen, J. and S. Nagel (2006). The conditional CAPM does not explain asset-pricing anomalies. *Journal of Financial Economics* 82(2), 289–314.
- Lewellen, J., S. Nagel, and J. Shanken (2010). A skeptical appraisal of asset pricing tests. *Journal of Financial Economics* 96(2), 175–194.
- 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.
- 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.
- Moskowitz, T. J. (2003). An analysis of covariance risk and pricing anomalies. *Review of Financial Studies* 16(2), 417–457.
- Newey, W. K. and K. D. West (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55, 703–708.
- 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.
- Pukthuanthong, K. and R. Roll (2014). A protocol for factor identification. *Available at SSRN 2342624*.
- Raponi, V., C. Robotti, and P. Zaffaroni (2016). Ex-Post Risk Premia and Tests of Multi-Beta Models in Large Cross-Sections. Technical report, Imperial College London.
- 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. (1996). Statistical methods in tests of portfolio efficiency: A synthesis. *Handbook of statistics* 14, 693–711.
- 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.
- Stock, J. H. and M. W. Watson (2002a). Forecasting Using Principal Components from a Large Number of Predictors. *Journal of the American Statistical Association* 97(460), 1167–1179.
- Stock, J. H. and M. W. Watson (2002b). Macroeconomic Forecasting Using Diffusion Indexes. *Journal of Business & Economic Statistics* 20(2), 147–162.
- 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.

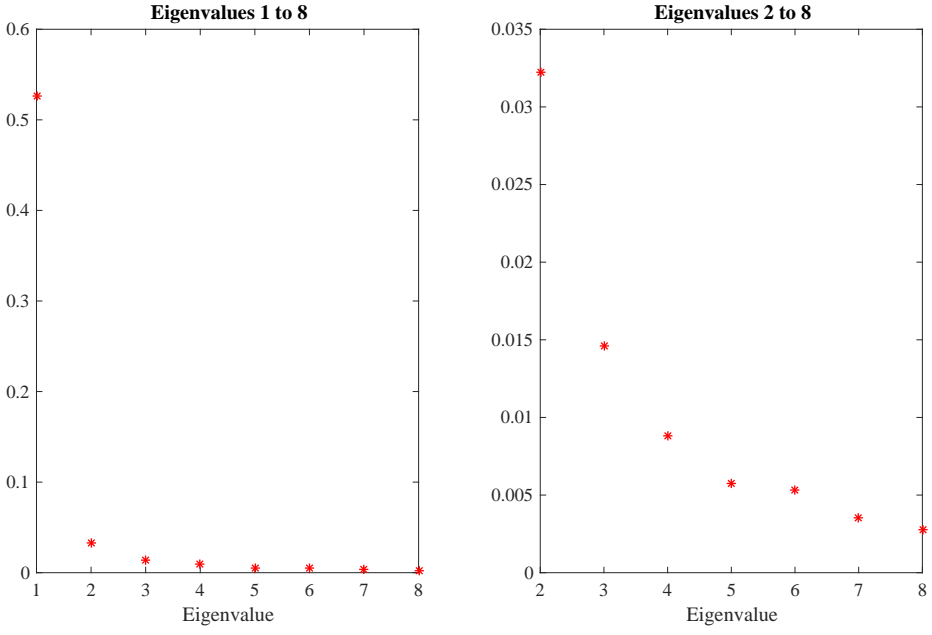
10 Figures and Tables

Figure 1: Histograms of the Standardized Estimates



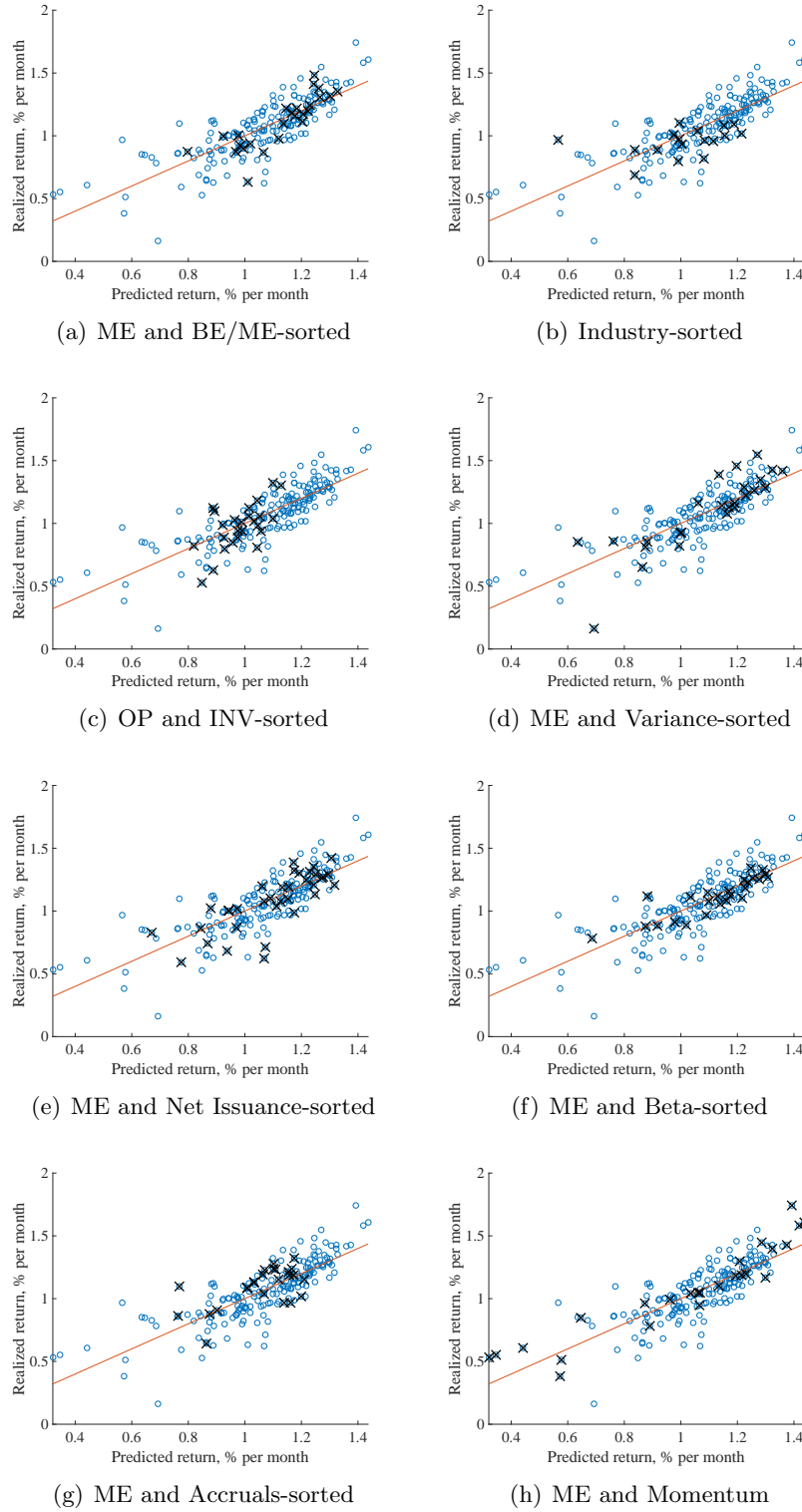
Note: The left panels provide the histograms of the standardized two-pass risk premia estimates using the Fama-MacBeth approach for standard error estimation, whereas the right panels provide the histograms of the standardized three-pass estimates using asymptotic standard errors. We fix $n = 200$ and $T = 600$.

Figure 2: First Eight Eigenvalues of the Covariance Matrix of 202 Equity Portfolios



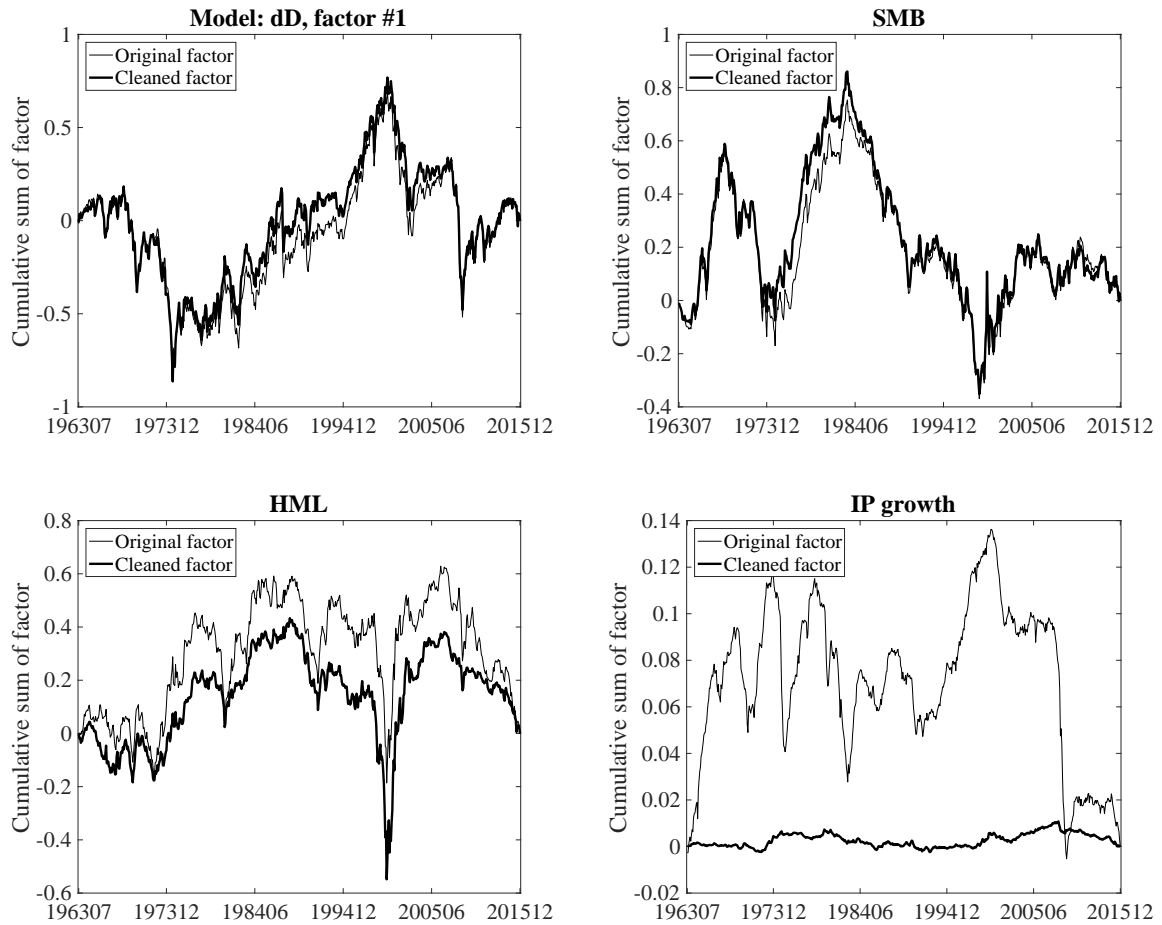
Note: The left panel reports the first eight eigenvalues of the covariance matrix of our 202 test portfolios. The right panel zooms in to the eigenvalues two through eight.

Figure 3: Predicted and Realized Average Excess Returns in a Six-Factor Model



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 4: Cumulative Factor Time Series with and without Measurement Error



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

Table 1: Simulation Results for $n = 50$

T	Param	True	Two-Pass Estimator		Three-Pass Estimators					
			Bias	RMSE	$\check{p} = 4$		$\check{p} = 5$		$\check{p} = 6$	
					Bias	RMSE	Bias	RMSE	Bias	RMSE
50	γ_0	0.546	0.866	0.867	0.476	0.752	0.422	0.707	0.414	0.697
	RmRf	0.372	-0.766	0.853	-0.394	0.790	-0.351	0.759	-0.349	0.752
	SMB	0.229	-0.136	0.262	-0.107	0.418	-0.092	0.416	-0.084	0.416
	HML	0.209	-0.013	0.255	-0.064	0.292	-0.060	0.299	-0.056	0.304
	IP	-0.003	0.001	0.079	0.002	0.015	0.002	0.016	0.002	0.018
200	γ_0	0.546	0.929	0.945	0.165	0.403	0.087	0.368	0.137	0.382
	RmRf	0.372	-0.837	0.842	-0.129	0.449	-0.060	0.430	-0.114	0.442
	SMB	0.229	-0.129	0.130	-0.058	0.221	-0.044	0.218	-0.041	0.218
	HML	0.209	-0.016	0.078	-0.042	0.167	-0.034	0.167	-0.029	0.167
	IP	-0.003	-0.006	0.109	0.001	0.007	0.001	0.007	0.001	0.008
600	γ_0	0.546	0.950	0.990	0.049	0.280	-0.039	0.277	0.049	0.291
	RmRf	0.372	-0.861	0.863	-0.030	0.305	0.048	0.307	-0.040	0.319
	SMB	0.229	-0.129	0.155	-0.043	0.137	-0.030	0.133	-0.030	0.133
	HML	0.209	-0.024	0.026	-0.033	0.108	-0.020	0.106	-0.018	0.105
	IP	-0.003	-0.022	0.165	0.0005	0.004	0.0002	0.004	0.0005	0.004

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 $\check{p} = 4, 5,$ and $6,$ for $n = 50,$ and $T = 50, 200,$ and $600,$ 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 2: Simulation Results for $n = 100$

T	Param	True	Two-Pass Estimator		Three-Pass Estimators					
			Bias	RMSE	$\check{p} = 4$		$\check{p} = 5$		$\check{p} = 6$	
					Bias	RMSE	Bias	RMSE	Bias	RMSE
50	γ_0	0.546	0.802	0.804	0.484	0.666	0.407	0.578	0.387	0.555
	RmRf	0.372	-0.780	0.843	-0.469	0.783	-0.386	0.699	-0.366	0.680
	SMB	0.229	-0.084	0.231	-0.045	0.405	-0.041	0.407	-0.039	0.409
	HML	0.209	0.106	0.258	0.012	0.292	-0.012	0.301	-0.015	0.305
	IP	-0.003	0.001	0.068	0.002	0.015	0.001	0.017	0.001	0.018
200	γ_0	0.546	0.838	0.877	0.418	0.508	0.166	0.279	0.151	0.267
	RmRf	0.372	-0.833	0.834	-0.428	0.581	-0.164	0.387	-0.149	0.380
	SMB	0.229	-0.073	0.073	-0.011	0.214	-0.015	0.215	-0.015	0.215
	HML	0.209	0.147	0.156	0.030	0.163	-0.002	0.164	-0.004	0.165
	IP	-0.003	-0.005	0.092	0.002	0.006	0.001	0.007	0.0005	0.007
600	γ_0	0.546	0.846	0.913	0.412	0.458	0.067	0.194	0.062	0.192
	RmRf	0.372	-0.846	0.853	-0.430	0.498	-0.072	0.253	-0.067	0.252
	SMB	0.229	-0.067	0.112	0.001	0.126	-0.007	0.127	-0.006	0.127
	HML	0.209	0.149	0.153	0.032	0.103	-0.001	0.101	-0.002	0.101
	IP	-0.003	-0.016	0.142	0.002	0.004	0.0004	0.004	0.0004	0.004

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 $\check{p} = 4, 5,$ and $6,$ for $n = 100,$ and $T = 50, 200,$ and $600,$ 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 3: Simulation Results for $n = 200$

T	Param	True	Two-Pass Estimators		Three-Pass Estimators					
			Bias	RMSE	$\check{p} = 4$		$\check{p} = 5$		$\check{p} = 6$	
					Bias	RMSE	Bias	RMSE	Bias	RMSE
50	γ_0	0.546	0.662	0.669	0.330	0.551	0.293	0.429	0.289	0.423
	RmRf	0.372	-0.620	0.681	-0.295	0.683	-0.273	0.591	-0.270	0.589
	SMB	0.229	-0.092	0.229	-0.067	0.413	-0.029	0.411	-0.028	0.412
	HML	0.209	0.028	0.238	-0.028	0.314	-0.030	0.318	-0.030	0.319
	IP	-0.003	0.0004	0.063	0.001	0.016	0.001	0.017	0.001	0.018
200	γ_0	0.546	0.701	0.753	0.039	0.302	0.107	0.186	0.103	0.182
	RmRf	0.372	-0.667	0.667	-0.019	0.411	-0.103	0.334	-0.099	0.332
	SMB	0.229	-0.082	0.082	-0.051	0.221	-0.010	0.214	-0.010	0.214
	HML	0.209	0.036	0.062	-0.010	0.169	-0.014	0.170	-0.014	0.170
	IP	-0.003	-0.010	0.098	0.0001	0.007	0.0005	0.007	0.0005	0.008
600	γ_0	0.546	0.710	0.794	-0.139	0.233	0.039	0.133	0.036	0.132
	RmRf	0.372	-0.679	0.689	0.151	0.294	-0.039	0.217	-0.037	0.217
	SMB	0.229	-0.078	0.121	-0.043	0.134	-0.006	0.126	-0.006	0.126
	HML	0.209	0.034	0.052	0.000	0.100	-0.006	0.100	-0.005	0.100
	IP	-0.003	-0.033	0.161	-0.001	0.005	0.0003	0.004	0.0003	0.004

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 $\check{p} = 4, 5,$ and $6,$ for $n = 200,$ and $T = 50, 200,$ and $600,$ 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 4: Simulation Results for the Number of Factors

T	$n = 50$		$n = 100$		$n = 200$	
	Median	Stderr	Median	Stderr	Median	Stderr
50	3	0.66	3	0.53	5	0.79
200	3	0.64	4	0.83	5	0.14
600	4	0.50	5	0.40	5	0.40

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 5: Three-Pass Regression: Empirical Results (I)

Model	Factors	Avg ret	FM		3-pass, $\check{p} = 4$			3-pass, $\check{p} = 5$			3-pass, $\check{p} = 6$		
			γ	stderr	γ	stderr	R_g^2	γ	stderr	R_g^2	γ	stderr	R_g^2
CAPM	Intercept	0.40	1.28***	(0.21)	0.55	(0.09)		0.55	(0.11)		0.57	(0.13)	
	RmRf	0.50	-0.20	(0.28)	0.37*	(0.20)	98.18	0.37*	(0.21)	98.93	0.35	(0.22)	99.08
FF3	Intercept	0.40	1.53***	(0.17)	0.55	(0.09)		0.55	(0.11)		0.57	(0.13)	
	RmRf	0.50	-0.57**	(0.25)	0.37*	(0.20)	98.18	0.37*	(0.21)	98.93	0.35	(0.22)	99.08
	SMB	0.23	0.17	(0.13)	0.23*	(0.13)	93.90	0.23*	(0.13)	94.88	0.23*	(0.13)	97.19
	HML	0.34	0.23*	(0.13)	0.21*	(0.11)	66.86	0.21*	(0.11)	67.90	0.20*	(0.11)	75.37
FF4	Intercept	0.40	0.90***	(0.15)	0.55	(0.09)		0.55	(0.11)		0.57	(0.13)	
	RmRf	0.50	0.05	(0.23)	0.37*	(0.20)	98.18	0.37*	(0.21)	98.93	0.35	(0.22)	99.08
	SMB	0.23	0.17	(0.13)	0.23*	(0.13)	93.90	0.23*	(0.13)	94.88	0.23*	(0.13)	97.19
	HML	0.34	0.41***	(0.13)	0.21*	(0.11)	66.86	0.21*	(0.11)	67.90	0.20*	(0.11)	75.37
	Mom	0.71	0.81***	(0.17)	0.75***	(0.18)	91.18	0.75***	(0.18)	91.52	0.74***	(0.18)	92.19
	Intercept	0.40	1.01***	(0.15)	0.55	(0.09)		0.55	(0.11)		0.57	(0.13)	
FF5	RmRf	0.50	-0.08	(0.24)	0.37*	(0.20)	98.18	0.37*	(0.21)	98.93	0.35	(0.22)	99.08
	SMB	0.23	0.27**	(0.13)	0.23*	(0.13)	93.90	0.23*	(0.13)	94.88	0.23*	(0.13)	97.19
	HML	0.34	0.02	(0.13)	0.21*	(0.11)	66.86	0.21*	(0.11)	67.90	0.20*	(0.11)	75.37
	RMW	0.25	0.30***	(0.10)	0.13**	(0.06)	33.93	0.13**	(0.07)	37.42	0.13*	(0.07)	45.81
	CMA	0.30	0.37***	(0.09)	0.14*	(0.08)	44.58	0.14*	(0.08)	45.68	0.13	(0.08)	55.38
	Intercept	0.40	0.84***	(0.15)	0.62	(0.10)		0.59	(0.11)		0.62	(0.13)	
HXZ	Mkt	0.49	0.06	(0.25)	0.30	(0.21)	98.37	0.33	(0.22)	98.71	0.30	(0.23)	99.06
	ME	0.31	0.39***	(0.13)	0.31**	(0.14)	90.90	0.30**	(0.13)	92.10	0.31**	(0.13)	94.48
	IA	0.41	0.27***	(0.10)	0.14*	(0.08)	46.14	0.14*	(0.08)	46.68	0.13*	(0.08)	53.68
	ROE	0.56	0.59***	(0.13)	0.28***	(0.09)	50.88	0.28***	(0.09)	51.68	0.28***	(0.09)	55.64

Note: The table reports the results of standard Fama-MacBeth regression and three-pass cross-sectional regression with four, five, and six factors. Each panel corresponds to a different model. The first column shows the average risk-free rate in the data (row “intercept”) and the average excess returns of factors when they are tradable. The “FM” set of results corresponds to standard Fama-MacBeth estimation of the model. The other sets correspond to the three-pass method, using four to six latent factors. For each set of results, the first column reports the zero-beta rate and the risk-premium estimates for the factors. The second column reports the standard error. The column denoted R_g^2 reports the R^2 of the third pass, the regression of g_t onto the estimated latent factors.

Table 6: Three-Pass Regression: Empirical Results (II)

Model	Factors	Avg ret	FM		3-pass, $\check{p} = 4$			3-pass, $\check{p} = 5$			3-pass, $\check{p} = 6$		
			γ	stderr	γ	stderr	R_g^2	γ	stderr	R_g^2	γ	stderr	R_g^2
BAB	Intercept	0.40	1.11***	(0.19)	0.55	(0.09)		0.55	(0.11)		0.57	(0.13)	
	Bab	0.84	0.55**	(0.24)	0.55***	(0.11)	45.40	0.55***	(0.11)	45.85	0.54***	(0.12)	49.00
QMJ	Intercept	0.40	1.10***	(0.15)	0.55	(0.09)		0.55	(0.11)		0.57	(0.13)	
	Qmj	0.35	0.03	(0.13)	0.07	(0.08)	59.63	0.07	(0.08)	63.13	0.07	(0.09)	70.43
IP	Intercept	0.40	1.03***	(0.19)	0.55	(0.09)		0.55	(0.11)		0.57	(0.13)	
	IP		-0.13*	(0.07)	-0.00	(0.00)	0.13	-0.00	(0.00)	0.19	-0.00	(0.01)	1.55
JLN	Intercept	0.43	0.94***	(0.19)	0.55	(0.10)		0.56	(0.12)		0.59	(0.13)	
	Factor 1		70.25***	(21.62)	1.74	(1.32)	0.60	1.70	(1.28)	0.82	1.61	(1.28)	0.97
	Factor 2		3.84	(24.02)	-2.25	(1.89)	2.57	-2.22	(1.88)	2.60	-2.01	(1.94)	3.66
	Factor 3		-1.71	(15.04)	1.09	(1.78)	3.58	1.02	(1.91)	4.43	1.09	(2.01)	6.65
Liq.	Intercept	0.40	1.06***	(0.20)	0.55	(0.09)		0.55	(0.11)		0.57	(0.13)	
	Liquidity		0.02	(0.97)	0.26**	(0.12)	11.99	0.26**	(0.12)	12.02	0.25**	(0.12)	12.02
Interm.	Intercept	0.43	0.85***	(0.23)	0.61	(0.10)		0.61	(0.11)		0.63	(0.13)	
	He		0.02	(0.64)	0.30	(0.27)	60.10	0.29	(0.29)	62.08	0.28	(0.30)	62.13
	Adrian		1.25***	(0.32)	0.79***	(0.16)	49.28	0.78***	(0.17)	49.28	0.77***	(0.17)	51.58

Note: The table reports the results of standard Fama-MacBeth regression and three-pass cross-sectional regression with four, five, and six factors. Each panel corresponds to a different model. The first column shows the average risk-free rate in the data (row “intercept”) and the average excess returns of factors when they are tradable. The “FM” set of results corresponds to standard Fama-MacBeth estimation of the model. The other sets correspond to the three-pass method, using four to six latent factors. For each set of results, the first column reports the zero-beta rate and the risk-premium estimates for the factors. The second column reports the standard error. The column denoted R_g^2 reports the R^2 of the third pass, the regression of g_t onto the estimated latent factors.

Table 7: Loading of Observable Factors onto Latent Factors (% of Variation Explained)

Model	Factors	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6
CAPM	RmRf	91.0	6.3	1.7	0.1	0.8	0.2
FF3	RmRf	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
FF4	RmRf	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
FF5	RmRf	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
	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
HXZ	Mkt	91.7	5.7	1.8	0.1	0.3	0.4
	ME	28.4	60.8	5.7	1.3	1.3	2.5
	IA	23.1	1.6	61.2	0.0	1.0	13.0
	ROE	16.2	27.0	0.7	47.5	1.4	7.1
BAB	Bab	0.9	3.6	72.7	15.4	0.9	6.4
QMJ	Qmj	57.3	15.9	2.4	9.0	5.0	10.4
IP	IP Growth	2.2	0.5	4.4	1.0	3.8	88.0
JLN	Factor 1	21.3	10.1	27.3	2.8	22.6	15.9
	Factor 2	59.1	7.1	0.0	4.1	0.7	29.1
	Factor 3	19.9	0.2	21.0	12.8	12.8	33.3
Liq.	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
	Adrian et al.	20.3	6.8	52.0	16.4	0.0	4.5

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%.

Appendix A Mathematical Proofs

Appendix A.1 Proofs of Main Theorems

Proof of Proposition 1. Without loss of generality, we assume that η is a $1 \times p$ vector. Consider an invertible matrix H such that its first row is equal to η . The model (1) can then be written equivalently as:

$$\begin{aligned} r_t &= \iota_n \gamma_0 + \alpha + \beta H^{-1} H \gamma + \beta H^{-1} H v_t + u_t \\ &:= \iota_n \gamma_0 + \alpha + \tilde{\beta} \tilde{\gamma} + \tilde{\beta} \tilde{v}_t + u_t, \end{aligned}$$

where $\tilde{\beta} = \beta H^{-1}$, $\tilde{\gamma} = H \gamma$, and $\tilde{v}_t = H v_t$. The expected excess return of a portfolio with no idiosyncratic risk or alpha, which has unit beta on the i th factor in \tilde{v}_t and zero beta on the rest factors is $E(\gamma_0 + e_i^\top \tilde{\gamma} + e_i^\top \tilde{v}_t) - \gamma_0 = \tilde{\gamma}_i$. By definition, the risk premium of \tilde{v}_t is $\tilde{\gamma}$. Since $\tilde{g}_t = e_1^\top \tilde{v}_t$, its risk premium is $e_1^\top \tilde{\gamma} = e_1^\top H \gamma = \eta \gamma$.

To establish the second statement, suppose $(\tilde{v}_1 : \tilde{v}_2 : \dots : \tilde{v}_T)$ is the new set of factors. Because it shares the same row space as $(v_1 : v_2 : \dots : v_T)$, there exists an invertible matrix H , such that $\tilde{v}_t = H v_t$, for all $t = 1, 2, \dots, T$. By the above result, the risk premium of \tilde{v}_t , $\tilde{\gamma}$, is therefore $H \gamma$. Since $g_t = \eta v_t$, it follows that $g_t = \eta H^{-1} \tilde{v}_t$. Applying the above result again, we have the risk premium of g_t is equal to $\eta H^{-1} H \gamma = \eta \gamma$. \square

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

Step 1: Since

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

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

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

We analyze the terms on the right-hand side one by one.

(i) To begin with, write $\Gamma^u = (\gamma_{n,tt'})$. Note that

$$\|\bar{U}^\top \bar{U} - n \Gamma^u\| \leq \|U^\top U - n \Gamma^u\|_{\text{F}} + 2 \|\iota_T \bar{u}^\top U\|_{\text{F}} + \|\iota_T \bar{u}^\top \bar{u} \iota_T^\top\|_{\text{F}}.$$

By Assumption 3(iv),

$$E \|U^\top U - n \Gamma^u\|_{\text{F}}^2 = \sum_{s=1}^T \sum_{t=1}^T E \left(\sum_{j=1}^n (u_{js} u_{jt} - E(u_{js} u_{jt})) \right)^2 \leq K n T^2, \quad (\text{A.1})$$

and by Assumption 3(i),

$$\mathbb{E} \|\bar{u}\|_{\mathbb{F}}^2 = T^{-2} \mathbb{E} \sum_{i=1}^n \sum_{t=1}^T \sum_{t'=1}^T u_{it} u_{it'} \leq nT^{-2} \sum_{t=1}^T \sum_{t'=1}^T |\gamma_{n,tt'}| \leq KnT^{-1}, \quad (\text{A.2})$$

$$\mathbb{E} \|U\|_{\mathbb{F}}^2 = \sum_{i=1}^n \sum_{t=1}^T \mathbb{E} u_{it}^2 \leq n \sum_{t=1}^T \gamma_{n,tt} \leq KnT, \quad (\text{A.3})$$

it follows that

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

and hence that

$$\|\bar{U}^\top \bar{U} - n\Gamma^u\| = O_p(n^{1/2}T) + O_p(nT^{1/2}). \quad (\text{A.4})$$

Next, writing $\rho_{n,st} = \gamma_{n,st} / \sqrt{\gamma_{n,ss} \gamma_{n,tt}}$, by Assumption 3(i) and the fact that $|\rho_{n,st}| \leq 1$,

$$\begin{aligned} \|\Gamma^u\|_{\mathbb{F}}^2 &= \sum_{s=1}^T \sum_{t=1}^T \gamma_{n,st}^2 = \sum_{s=1}^T \sum_{t=1}^T \gamma_{n,ss} \gamma_{n,tt} \rho_{n,st}^2 \\ &\leq K \sum_{s=1}^T \sum_{t=1}^T |\gamma_{n,ss} \gamma_{n,tt}|^{1/2} |\rho_{n,st}| \leq K \sum_{s=1}^T \sum_{t=1}^T |\gamma_{n,st}| \leq KT, \end{aligned} \quad (\text{A.5})$$

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

Therefore, we obtain

$$\|\bar{U}^\top \bar{U}\| \leq \|\bar{U}^\top \bar{U} - n\Gamma^u\| + n \|\Gamma^u\| = O_p(nT^{1/2}) + O_p(n^{1/2}T). \quad (\text{A.6})$$

(ii) By Assumption 5, we have $\|\beta\|_{\text{MAX}} \leq K$. By Assumption 3(ii)(iii), we have

$$\mathbb{E} \|U^\top \beta\|_{\mathbb{F}}^2 = \mathbb{E} \sum_{j=1}^p \sum_{t=1}^T \left(\sum_{i=1}^n \beta_{ij} u_{it} \right)^2 \leq K \sum_{j=1}^p \sum_{t=1}^T \sum_{i=1}^n \sum_{i'=1}^n |\sigma_{ii',t}| \leq KnT, \quad (\text{A.7})$$

$$\mathbb{E} \|\bar{u}^\top \beta\|_{\mathbb{F}}^2 \leq \mathbb{E} \sum_{k=1}^p \left(\sum_{i=1}^n \bar{u}_i \beta_{ik} \right)^2 \leq KT^{-2} \sum_{k=1}^p \sum_{i=1}^n \sum_{i'=1}^n \sum_{t=1}^T \sum_{t'=1}^T |\sigma_{ii',tt'}| \leq KnT^{-1}, \quad (\text{A.8})$$

it follows that

$$\|\bar{U}^\top \beta\|_{\mathbb{F}} \leq \|U^\top \beta\|_{\mathbb{F}} + \|\iota_T\|_{\mathbb{F}} \|\bar{u}^\top \beta\|_{\mathbb{F}} = O_p(n^{1/2}T^{1/2}). \quad (\text{A.9})$$

Also, by Assumption 4,

$$T^{-1} \|\bar{V} \bar{V}^\top\|_{\text{MAX}} \leq \|T^{-1} V V^\top - \Sigma^v\|_{\text{MAX}} + \|\Sigma^v\|_{\text{MAX}} + \|\bar{v} \bar{v}^\top\|_{\text{MAX}} \leq K, \quad (\text{A.10})$$

we have

$$\|\bar{V}\| \leq \|\bar{V}\bar{V}^\top\|^{1/2} \leq K \|\bar{V}\bar{V}^\top\|_{\text{MAX}}^{1/2} = O_p(T^{1/2}). \quad (\text{A.11})$$

Therefore, we have

$$\|\bar{V}^\top \beta^\top \bar{U}\| = \|\bar{U}^\top \beta \bar{V}\| \leq \|\bar{U}^\top \beta\|_F \|\bar{V}\| = O_p(n^{1/2}T).$$

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

$$n^{-1}T^{-1} |\lambda_j(\bar{R}^\top \bar{R}) - \lambda_j(\bar{V}^\top \beta^\top \beta \bar{V})| = O_p(n^{-1/2} + T^{-1/2}) = o_p(1). \quad (\text{A.12})$$

(iii) Moreover, by Assumption 5, (A.11), and Weyl's inequality again,

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

and combined with Assumption 4, and the fact that $\|\bar{v}\| \leq K \|\bar{v}\|_{\text{MAX}} = O_p(T^{-1/2})$,

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

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

$$\left| n^{-1}T^{-1} \lambda_j(\bar{R}^\top \bar{R}) - \lambda_j \left(\left(\Sigma^\beta \right)^{1/2} \Sigma^v \left(\Sigma^\beta \right)^{1/2} \right) \right| = o_p(1). \quad (\text{A.13})$$

Step 2: By Assumptions 4 and 5, 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 (A.13), for $1 \leq j \leq p$,

$$K_1 < n^{-1}T^{-1} \lambda_j(\bar{R}^\top \bar{R}) < K_2. \quad (\text{A.14})$$

On the other hand, we can write

$$\bar{R} \bar{R}^\top = \tilde{\beta} \bar{V} \bar{V}^\top \tilde{\beta}^\top + \bar{U} (\mathbb{I}_T - \bar{V}^\top (\bar{V} \bar{V}^\top)^{-1} \bar{V}) \bar{U}^\top, \quad (\text{A.15})$$

where $\tilde{\beta} = \beta + U \bar{V}^\top (\bar{V} \bar{V}^\top)^{-1}$. 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

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

Moreover, by (A.6), we have

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

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

$$n^{-1}T^{-1}\lambda_j(\bar{R}^\top\bar{R}) \leq K(n^{-1/2} + T^{-1/2}). \quad (\text{A.16})$$

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

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

(A.14) and (A.16) together imply that for $1 \leq j \leq p$,

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

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

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

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

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

$$H = n^{-1}T^{-1}\hat{\Lambda}^{-1}\hat{V}\bar{V}^\top\beta^\top\beta. \quad (\text{A.17})$$

We use the following decomposition:

$$\begin{aligned} &\tilde{\Gamma} - \begin{pmatrix} \gamma_0 \\ H\gamma \end{pmatrix} \\ &= \left((\iota_n : \hat{\beta})^\top(\iota_n : \hat{\beta})\right)^{-1}(\iota_n : \hat{\beta})^\top \left((\beta - \hat{\beta}H)\gamma + \beta\bar{v} + \alpha + \bar{u} \right) \\ &= \begin{pmatrix} 0 \\ H\bar{v} \end{pmatrix} + \left\{ \frac{1}{n} \begin{pmatrix} \iota_n^\top \iota_n & \iota_n^\top \hat{\beta} \\ \hat{\beta}^\top \iota_n & \hat{\beta}^\top \hat{\beta} \end{pmatrix} \right\}^{-1} \left\{ \frac{1}{n} \begin{pmatrix} \iota_n^\top \alpha \\ H^{-\top} \beta^\top \alpha \end{pmatrix} + \frac{1}{n} \begin{pmatrix} \iota_n^\top \bar{u} + \iota_n^\top (\beta - \hat{\beta}H)\gamma \\ H^{-\top} \beta^\top \bar{u} + H^{-\top} \beta^\top (\beta - \hat{\beta}H)\gamma \end{pmatrix} \right. \\ &\quad \left. + \frac{1}{n} \begin{pmatrix} \iota_n^\top (\beta - \hat{\beta}H)\bar{v} \\ (\hat{\beta} - \beta H^{-1})^\top (\alpha + \bar{u}) + H^{-\top} \beta^\top (\beta - \hat{\beta}H)\bar{v} + (\hat{\beta}^\top - H^{-\top} \beta^\top) (\beta - \hat{\beta}H) (\gamma + \bar{v}) \end{pmatrix} \right\}. \quad (\text{A.18}) \end{aligned}$$

Note that

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

and by Lemma 7(a) and (b), we have

$$\hat{\eta} - \eta H^{-1} = T^{-1} \bar{Z}\bar{V}^\top H^\top + O_p(n^{-1} + T^{-1}). \quad (\text{A.19})$$

Moreover, by Assumptions 4 and 6, as well as Lemma 2, we have

$$\|T^{-1} \bar{Z}\bar{V}^\top H^\top\|_{\text{MAX}} \leq (\|T^{-1} ZV^\top\|_{\text{MAX}} + \|\bar{z}\bar{v}^\top\|_{\text{MAX}}) \|H\|_{\text{MAX}} = O_p(T^{-1/2}),$$

it follows that

$$\hat{\eta} - \eta H^{-1} = O_p(n^{-1} + T^{-1/2}). \quad (\text{A.20})$$

Using this, and by Lemmas 2, 4, 5, 6, 7, and 8, we have

$$\begin{aligned} \hat{\Gamma} - \begin{pmatrix} \gamma_0 \\ \eta\gamma \end{pmatrix} &= \begin{pmatrix} 0 \\ T^{-1} \bar{Z}\bar{V}^\top (\Sigma^v)^{-1} \gamma + \eta\bar{v} \end{pmatrix} \\ &+ \begin{pmatrix} 1 & 0 \\ 0 & \eta \end{pmatrix} \left\{ \frac{1}{n} \begin{pmatrix} \iota_n^\top \iota_n & \iota_n^\top \beta \\ \beta^\top \iota_n & \beta^\top \beta \end{pmatrix} + o_p(1) \right\}^{-1} \times \left\{ \frac{1}{n} \begin{pmatrix} \iota_n^\top \alpha \\ \beta^\top \alpha \end{pmatrix} + o_p(n^{-1/2} + T^{-1/2}) \right\}. \end{aligned}$$

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

$$n^{-1/2} \begin{pmatrix} \iota_n^\top \alpha \\ \beta^\top \alpha \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N} \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \beta_0^\top \\ \beta_0 & \Sigma^\beta \end{pmatrix} (\sigma^\alpha)^2 \right), \quad (\text{A.21})$$

where we use $\|n^{-1} \beta^\top \iota_n - \beta_0\|_{\text{MAX}} = o(1)$ and $\|n^{-1} \beta^\top \beta - \Sigma^\beta\|_{\text{MAX}} = o(1)$. Therefore, by the Delta method, we have

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

Similarly, we have

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

where

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

On the other hand, since

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

it follows from Assumption 11 that

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

Therefore, by the Delta method, we obtain:

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

where

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

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

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

□

Proof of Theorem 3. By Assumptions 4, 5, and 10, Lemma 4, (A.8), and (A.21), we have

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

it then follows that

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

On the other hand, by Assumption 4, Lemma 3, (A.2), we have

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

Similarly, we have

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

therefore, we obtain

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

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

By Lemma 2, (A.39), (A.20) and the fact that $\|\eta\|_{\text{MAX}} \leq K$, we have

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

Also, by Assumptions 4, 6, and 11, we have

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

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

Proof of Theorem 4. We denote the estimators of \bar{V} and β based on \check{p} as \check{V} and $\check{\beta}$ respectively. We can write $\check{V}^\top = (\widehat{V}^\top : T^{1/2} \check{\xi}_{p+1:\check{p}})$, where $\check{\xi}_{p+1:\check{p}} = (\xi_{p+1} : \xi_{p+2} : \dots : \xi_{\check{p}})$ is a $T \times (\check{p} - p)$ matrix, it follows that

$$\check{\beta} = T^{-1} \left(\bar{R} \widehat{V}^\top : T^{1/2} \bar{R} \check{\xi}_{p+1:\check{p}} \right) = \left(\widehat{\beta} : T^{-1/2} \bar{R} \check{\xi}_{p+1:\check{p}} \right).$$

Consider the singular value decomposition of \bar{R}

$$n^{-1/2} T^{-1/2} \bar{R} = O_1 \text{diag} \left(\lambda_1^{1/2}, \lambda_2^{1/2}, \dots, \lambda_{\min(n, T)}^{1/2} \right) O_2^\top,$$

where O_1 and O_2 are $n \times \min(n, T)$ and $T \times \min(n, T)$ orthogonal matrices, respectively. Given that

$$n^{-1} T^{-1} \bar{R}^\top \bar{R} (\xi_1 : \xi_2 : \dots : \xi_p : \xi_{p+1} \dots : \xi_{\check{p}}) = (\xi_1 : \xi_2 : \dots : \xi_p : \xi_{p+1} \dots : \xi_{\check{p}}) \check{\Lambda},$$

we can select O_2 such that the first \check{p} columns of O_2 coincide with $(\xi_1 : \xi_2 : \dots : \xi_p : \xi_{p+1} \dots : \xi_{\check{p}})$.

Also, we have

$$n^{-1/2} T^{-1/2} \zeta_{p+1:\check{p}}^\top \bar{R} = \check{\lambda}_{p+1:\check{p}}^{1/2} \check{\xi}_{p+1:\check{p}}^\top, \quad \text{and} \quad n^{-1/2} T^{-1/2} \bar{R} \check{\xi}_{p+1:\check{p}} = \check{\zeta}_{p+1:\check{p}} \check{\lambda}_{p+1:\check{p}}^{1/2},$$

where $\check{\lambda}_{p+1:\check{p}} = \text{diag}(\lambda_{p+1}, \lambda_{p+2}, \dots, \lambda_{\check{p}})$, $\check{\zeta}_{p+1:\check{p}}$ is an $n \times (\check{p} - p)$ matrix that corresponds to the $p+1, \dots, \check{p}-1, \check{p}$ columns of O_1 . As a result, we have

$$\check{\zeta}_{p+1:\check{p}}^\top \widehat{\beta} = 0, \quad \widehat{V} \check{\xi}_{p+1:\check{p}} = 0, \quad \text{and} \quad \check{\beta} = \left(\widehat{\beta} : n^{1/2} \check{\zeta}_{p+1:\check{p}} \check{\lambda}_{p+1:\check{p}}^{1/2} \right).$$

Moreover, by inverting block-diagonal matrices, we can decompose

$$\begin{aligned} & \begin{pmatrix} 1 & \\ & \check{\eta} \end{pmatrix} \left\{ \frac{1}{n} \begin{pmatrix} \iota_n^\top \iota_n & \iota_n^\top \check{\beta} \\ \check{\beta}^\top \iota_n & \check{\beta}^\top \check{\beta} \end{pmatrix} \right\}^{-1} \begin{pmatrix} \iota_n^\top \\ \check{\beta}^\top \end{pmatrix} \\ = & \begin{pmatrix} 1 & \\ & \widehat{\eta} \end{pmatrix} \left\{ \frac{1}{n} \begin{pmatrix} \iota_n^\top \iota_n & \iota_n^\top \widehat{\beta} \\ \widehat{\beta}^\top \iota_n & \widehat{\beta}^\top \widehat{\beta} \end{pmatrix} \right\}^{-1} \begin{pmatrix} \iota_n^\top \\ \widehat{\beta}^\top \end{pmatrix} \\ & + \begin{pmatrix} 1 & \\ & \widehat{\eta} \end{pmatrix} \left\{ \frac{1}{n} \begin{pmatrix} \iota_n^\top \iota_n & \iota_n^\top \widehat{\beta} \\ \widehat{\beta}^\top \iota_n & \widehat{\beta}^\top \widehat{\beta} \end{pmatrix} \right\}^{-1} \begin{pmatrix} -\iota_n^\top \check{\zeta}_{p+1:\check{p}} \check{\lambda}_{p+1:\check{p}}^{1/2} \Delta^{-1} \check{\lambda}_{p+1:\check{p}}^{1/2} \check{\zeta}_{p+1:\check{p}}^\top \left(\mathbb{I}_n - \iota_n (\iota_n^\top \mathbb{M}_{\widehat{\beta}} \iota_n)^{-1} \iota_n^\top \mathbb{M}_{\widehat{\beta}} \right) \\ 0 \end{pmatrix} \\ & + \begin{pmatrix} 0 \\ T^{-1/2} n^{1/2} \check{G} \check{\xi}_{p+1:\check{p}} \Delta^{-1} \check{\lambda}_{p+1:\check{p}}^{1/2} \check{\zeta}_{p+1:\check{p}}^\top \left(\mathbb{I}_n - \iota_n (\iota_n^\top \mathbb{M}_{\widehat{\beta}} \iota_n)^{-1} \iota_n^\top \mathbb{M}_{\widehat{\beta}} \right) \end{pmatrix}, \end{aligned}$$

where $\Delta = \check{\lambda}_{p+1:\check{p}}^{1/2} \left(\mathbb{I}_{\check{p}-p} - \check{\zeta}_{p+1:\check{p}}^\top \iota_n (\iota_n^\top \mathbb{M}_{\widehat{\beta}} \iota_n)^{-1} \iota_n^\top \check{\zeta}_{p+1:\check{p}} \right) \check{\lambda}_{p+1:\check{p}}^{1/2}$. This leads to

$$\begin{aligned} \check{\Gamma} - \widehat{\Gamma} &= \begin{pmatrix} 1 & \\ & \widehat{\eta} \end{pmatrix} \left\{ \frac{1}{n} \begin{pmatrix} \iota_n^\top \iota_n & \iota_n^\top \widehat{\beta} \\ \widehat{\beta}^\top \iota_n & \widehat{\beta}^\top \widehat{\beta} \end{pmatrix} \right\}^{-1} \\ & \times \begin{pmatrix} -n^{-1} \iota_n^\top \check{\zeta}_{p+1:\check{p}} \check{\lambda}_{p+1:\check{p}}^{1/2} \Delta^{-1} \check{\lambda}_{p+1:\check{p}}^{1/2} \check{\zeta}_{p+1:\check{p}}^\top \left(\mathbb{I}_n - \iota_n (\iota_n^\top \mathbb{M}_{\widehat{\beta}} \iota_n)^{-1} \iota_n^\top \mathbb{M}_{\widehat{\beta}} \right) \bar{r} \\ 0 \end{pmatrix} \\ & + \begin{pmatrix} 0 \\ T^{-1/2} n^{-1/2} \check{G} \check{\xi}_{p+1:\check{p}} \Delta^{-1} \check{\lambda}_{p+1:\check{p}}^{1/2} \check{\zeta}_{p+1:\check{p}}^\top \left(\mathbb{I}_n - \iota_n (\iota_n^\top \mathbb{M}_{\widehat{\beta}} \iota_n)^{-1} \iota_n^\top \mathbb{M}_{\widehat{\beta}} \right) \bar{r} \end{pmatrix}. \end{aligned}$$

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

$$\begin{aligned} & \check{\zeta}_{p+1:\check{p}}^\top \left(\mathbb{I}_n - \iota_n (\iota_n^\top \mathbb{M}_{\widehat{\beta}} \iota_n)^{-1} \iota_n^\top \mathbb{M}_{\widehat{\beta}} \right) \bar{r} \\ & = \check{\zeta}_{p+1:\check{p}}^\top \left(\mathbb{I}_n - \iota_n (\iota_n^\top \mathbb{M}_{\widehat{\beta}} \iota_n)^{-1} \iota_n^\top \mathbb{M}_{\widehat{\beta}} \right) (\alpha + (\beta - \widehat{\beta} H)(\bar{v} + \gamma) + \bar{u}). \end{aligned}$$

Since α is independent of $\check{\zeta}_{p+1:\check{p}}$, we have

$$\mathbb{E} \left\| \check{\zeta}_{p+1:\check{p}}^\top \alpha \right\|_{\mathbb{F}}^2 = \sum_{j=1}^{\check{p}-p} \mathbb{E} \left(\sum_{i=1}^n \check{\zeta}_{p+1:\check{p},ij} \alpha_i \right)^2 = \|\check{\zeta}_{p+1:\check{p}}\|_{\mathbb{F}}^2 (\sigma^\alpha)^2 \leq K,$$

hence $\left\| \check{\zeta}_{p+1:\check{p}}^\top \alpha \right\|_{\mathbb{F}} = O_p(1)$. By Lemma 3(c) and (A.2), we have

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

Because $\mathbb{I}_n - \iota_n(\iota_n^\top \mathbb{M}_{\widehat{\beta}} \iota_n)^{-1} \iota_n^\top \mathbb{M}_{\widehat{\beta}}$ is idempotent, its operator norm is bounded, so that

$$\left\| \check{\zeta}_{p+1:\check{p}}^\top \left(\mathbb{I}_n - \iota_n(\iota_n^\top \mathbb{M}_{\widehat{\beta}} \iota_n)^{-1} \iota_n^\top \mathbb{M}_{\widehat{\beta}} \right) \bar{r} \right\| = O_p(1 + n^{1/2}T^{-1/2}).$$

Moreover, since $\check{\lambda}_{p+1:\check{p}}^{1/2} \Delta^{-1} \check{\lambda}_{p+1:\check{p}}^{1/2}$ is idempotent, and $\|n^{-1} \iota_n^\top \check{\zeta}_{p+1:\check{p}}\| = O_p(n^{-1/2})$, it follows that

$$\left\| n^{-1} \iota_n^\top \check{\zeta}_{p+1:\check{p}} \check{\lambda}_{p+1:\check{p}}^{1/2} \Delta^{-1} \check{\lambda}_{p+1:\check{p}}^{1/2} \check{\zeta}_{p+1:\check{p}}^\top \left(\mathbb{I}_n - \iota_n(\iota_n^\top \mathbb{M}_{\widehat{\beta}} \iota_n)^{-1} \iota_n^\top \mathbb{M}_{\widehat{\beta}} \right) \bar{r} \right\| = O_p(n^{-1/2} + T^{-1/2}).$$

On the other hand, by Lemma 1,

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

Also, it follows from (A.15) that

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

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

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

Since $\text{rank}(\tilde{\beta}\bar{V}\bar{V}^\top\tilde{\beta}^\top) \leq p$ and $\text{rank}(\bar{U}\bar{V}^\top(\bar{V}\bar{V}^\top)^{-1}\bar{V}\bar{U}^\top) \leq p$, we obtain for $p+1 \leq j \leq \check{p}$,

$$\lambda_{\min(n,T)}(\bar{U}\bar{U}^\top) \leq \lambda_{j+p}(\bar{U}\bar{U}^\top) \leq \lambda_j(\bar{R}\bar{R}^\top) \leq \lambda_{j-p}(\bar{U}\bar{U}^\top) \leq \lambda_1(\bar{U}\bar{U}^\top).$$

Following the random matrix theory in Bai (1999) on the limit of the extreme eigenvalues, we have

$$K'(n^{-1} + T^{-1}) \leq n^{-1}T^{-1}\lambda_j(\bar{R}\bar{R}^\top) \leq K(n^{-1} + T^{-1}),$$

so that $\left\| \check{\lambda}_{p+1:\check{p}}^{-1/2} \right\| = O_p((n^{-1} + T^{-1})^{-1/2})$. By the independence of \bar{Z} and $\check{\xi}_{p+1:\check{p}}$, we have $\left\| \bar{Z} \check{\xi}_{p+1:\check{p}} \right\| = O_p(1)$. It then follows that

$$\begin{aligned} T^{-1/2}n^{-1/2} \left\| \bar{G} \check{\xi}_{p+1:\check{p}} \Delta^{-1} \check{\lambda}_{p+1:\check{p}}^{1/2} \right\| &\leq T^{-1/2}n^{-1/2} \left\| (\eta \bar{V} + \bar{Z}) \check{\xi}_{p+1:\check{p}} \check{\lambda}_{p+1:\check{p}}^{-1/2} \right\| \left\| \check{\lambda}_{p+1:\check{p}}^{1/2} \Delta^{-1} \check{\lambda}_{p+1:\check{p}}^{1/2} \right\| \\ &= (n^{-1/2}T^{-1/2} + n^{-1}) \left\| \check{\lambda}_{p+1:\check{p}}^{-1/2} \right\| = O_p(n^{-1/2}). \end{aligned}$$

We therefore obtain that

$$\left\| T^{-1/2} n^{-1/2} \bar{G}_{\check{\xi}_{p+1:\check{p}}} \check{\Delta}^{-1} \check{\lambda}_{p+1:\check{p}}^{1/2} \check{\xi}_{p+1:\check{p}}^\top \left(\mathbb{I}_n - \iota_n (\iota_n^\top \mathbb{M}_{\hat{\beta}} \iota_n)^{-1} \iota_n^\top \mathbb{M}_{\hat{\beta}} \right) \bar{r} \right\| = O_p(n^{-1/2} + T^{-1/2}),$$

which conclude the proof. \square

Proof of Theorem 5. For any $1 \leq t \leq T$, we have

$$\hat{g}_t - \eta v_t = (\hat{\eta} - \eta H^{-1})(\hat{v}_t - H \bar{v}_t) + (\hat{\eta} - \eta H^{-1}) H \bar{v}_t + \eta H^{-1}(\hat{v}_t - H \bar{v}_t) - \eta \bar{v} \quad (\text{A.22})$$

By (A.34), we have

$$\begin{aligned} \hat{v}_t - H \bar{v}_t &= n^{-1} T^{-1} \hat{\Lambda}^{-1} (\hat{V} - H \bar{V}) (\bar{U}^\top \beta \bar{v}_t + \bar{U}^\top \bar{u}_t) + n^{-1} T^{-1} \hat{\Lambda}^{-1} (H \bar{V} \bar{U}^\top \beta \bar{v}_t + H \bar{V} \bar{U}^\top \bar{u}_t) \\ &\quad + n^{-1} T^{-1} \hat{\Lambda}^{-1} \hat{V} \bar{V}^\top \beta^\top \bar{u}_t. \end{aligned} \quad (\text{A.23})$$

By Assumption 3(ii), we have

$$\mathbb{E} \|\beta^\top u_t\|_{\mathbb{F}}^2 = \mathbb{E} \sum_{i=1}^p \left(\sum_{k=1}^n \beta_{ki} u_{kt} \right)^2 \leq K \sum_{k=1}^n \sum_{k'=1}^n |\sigma_{kk',t}| \leq Kn,$$

so that

$$\|\beta^\top \bar{u}_t\|_{\mathbb{F}} \leq \|\beta^\top u_t\|_{\mathbb{F}} + \|\beta^\top \bar{u}\|_{\mathbb{F}} = O_p(n^{1/2}). \quad (\text{A.24})$$

By Assumption 3(i)(iv) and Assumption 12, using the fact that $|\rho_{n,st}| \leq 1$, we have

$$\begin{aligned} \mathbb{E} \|U^\top u_t\|_{\mathbb{F}}^2 &= \mathbb{E} \sum_{s=1}^T \left(n \gamma_{n,st} + \sum_{k=1}^n (u_{ks} u_{kt} - \mathbb{E}(u_{ks} u_{kt})) \right)^2 \\ &\leq Kn^2 \sum_{s=1}^T \gamma_{n,st}^2 + KnT \leq n^2 \sum_{s=1}^T |\gamma_{n,st}| + KnT = Kn^2 + KnT, \\ \mathbb{E} \|u_t\|_{\mathbb{F}}^2 &\leq \sum_{k=1}^n \mathbb{E} u_{kt}^2 \leq \sum_{k=1}^n |\tau_{kk'}| \leq K. \end{aligned} \quad (\text{A.25})$$

Then from (A.2) and (A.51), it follows that

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

The above estimates, along with (A.9), Lemma 1, and $\|\bar{v}_t\| = O_p(1)$, lead to

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

Moreover, it follows from (A.2), (A.41), and (A.46) that

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

We thereby focus on the remaining term, which by Lemma 1, (A.11) and (A.24), satisfies

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

Therefore, we have

$$\|\widehat{v}_t - H\bar{v}_t\|_{\text{MAX}} = O_p(n^{-1/2} + T^{-1}). \quad (\text{A.26})$$

Then by (A.22), (A.23), and (A.19), we have

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

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

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

By (A.17) and Assumptions 5 and 13, we have

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

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

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

$$(\widetilde{\gamma}^\top \otimes \mathbb{I}_d) \widehat{\Pi}_{12} \widehat{\eta}^\top \xrightarrow{p} \left(\gamma^\top (\Sigma^v)^{-1} \otimes \mathbb{I}_d \right) \Pi_{12} \eta^\top. \quad (\text{A.27})$$

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

Note that by (A.50), Lemma 2, Lemma 3(a), and Assumption 4, we have

$$\left\| T^{-1}H^{-1}\widehat{V}\widehat{V}^\top H^{-\top} - \Sigma^v \right\|_{\text{MAX}}$$

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

By (A.20), Lemma 2, and the proof of Theorem 2, we have

$$\|\widehat{\eta}H - \eta\|_{\text{MAX}} = O_p(n^{-1} + T^{-1/2}), \quad \|H^{-1}\widetilde{\gamma} - \gamma\|_{\text{MAX}} = O_p(n^{-1/2} + T^{-1/2}). \quad (\text{A.28})$$

Therefore, to prove (A.27), we only need to show that

$$\widetilde{\Pi}_{12} := (H^{-1} \otimes \mathbb{I}_d)\widehat{\Pi}_{12}H^{-\top} \xrightarrow{p} \Pi_{12}, \quad (\text{A.29})$$

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

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

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

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

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

In fact, to show (A.29), by Lemma 2 we only need to prove for any fixed $1 \leq i, i' \leq p$, and $1 \leq j, j' \leq d$,

$$\widetilde{\Pi}_{12, (i-1)d+j, i'} - T^{-1} \sum_{t=1}^T \sum_{s=1}^T z_{jt} v_{it} Q_{ts} v_{i's} \xrightarrow{p} 0, \quad (\text{A.30})$$

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

$$\begin{aligned}
&\text{the left-hand side of (A.30)} \\
&= T^{-1} \sum_{t=1}^T \sum_{s=1}^T \left\{ (\widehat{z}_{jt} - z_{jt})(\widetilde{v}_{it} - v_{it}) Q_{ts} (\widetilde{v}_{i's} - v_{i's}) + (\widehat{z}_{jt} - z_{jt})(\widetilde{v}_{it} - v_{it}) Q_{ts} v_{i's} \right. \\
&\quad \left. + (\widehat{z}_{jt} - z_{jt}) v_{it} Q_{ts} \widetilde{v}_{i's} + z_{jt} (\widetilde{v}_{it} - v_{it}) Q_{ts} \widetilde{v}_{i's} + z_{jt} \widetilde{v}_{it} Q_{ts} (\widetilde{v}_{i's} - v_{i's}) \right\}.
\end{aligned}$$

We analyze these terms one by one. Since we have

$$\widehat{Z} - \bar{Z} = \eta \bar{V} - \widehat{\eta} \widehat{V} = (\eta H^{-1} - \widehat{\eta}) H \bar{V} - (\widehat{\eta} - \eta H^{-1}) (\widehat{V} - H \bar{V}) - \eta H^{-1} (\widehat{V} - H \bar{V}), \quad (\text{A.31})$$

it follows from (A.20), (A.35), and Lemmas 1 and 2 that

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

Moreover, by Lemma 9(i), Assumption 14, (A.31), and (A.28), we have

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

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

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

Similarly, because of $\|\tilde{V}\|_{\text{F}} \leq O_p(T^{1/2})$ implied by (A.36), $\|Z\|_{\text{MAX}} = O_p((\log T)^{1/a})$ by Assumption 14 and Lemma 2, and by Assumptions 4 and 6, we have

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

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

All the above terms converge to 0, as $T, n \rightarrow \infty$, with $qT^{-1/4} + qn^{-1/4} \rightarrow 0$ and $n^{-3}T \rightarrow 0$, which establishes (A.30).

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

$$\left\| H^{\top} \left(\hat{\Sigma}^{\beta} - \hat{\beta}_0 \hat{\beta}_0^{\top} \right) H - \left(\Sigma^{\beta} - \beta_0 \beta_0^{\top} \right) \right\|_{\text{MAX}} \leq \left\| H^{\top} \hat{\Sigma}^{\beta} H - \Sigma^{\beta} \right\|_{\text{MAX}} + \left\| H^{\top} \hat{\beta}_0 \hat{\beta}_0^{\top} H - \beta_0 \beta_0^{\top} \right\|_{\text{MAX}}.$$

By Lemmas 2, 4(d), 5(d), and Assumption 5,

$$\begin{aligned} & \left\| H^{\top} \hat{\Sigma}^{\beta} H - \Sigma^{\beta} \right\|_{\text{MAX}} \\ & \leq \left\| n^{-1} H^{\top} \hat{\beta}^{\top} \hat{\beta} H - n^{-1} \beta^{\top} \beta \right\|_{\text{MAX}} + \left\| n^{-1} \beta^{\top} \beta - \Sigma^{\beta} \right\|_{\text{MAX}} \\ & \leq \left\| n^{-1} \left(H^{\top} \hat{\beta}^{\top} - \beta^{\top} \right) \left(\hat{\beta} H - \beta \right) + n^{-1} \left(H^{\top} \hat{\beta}^{\top} - \beta^{\top} \right) \beta - n^{-1} \beta^{\top} \left(\beta - \hat{\beta} H \right) \right\|_{\text{MAX}} + o_p(1) \\ & = o_p(1). \tag{A.32} \\ & \left\| H^{\top} \hat{\beta}_0 \hat{\beta}_0^{\top} H - \beta_0 \beta_0^{\top} \right\|_{\text{MAX}} \\ & \leq \left\| \left(H^{\top} \hat{\beta}_0 - \beta_0 \right) \left(\hat{\beta}_0^{\top} H - \beta_0^{\top} \right) + \beta_0 \left(\hat{\beta}_0^{\top} H - \beta_0^{\top} \right) + \left(H^{\top} \hat{\beta}_0 - \beta_0 \right) \beta_0^{\top} \right\|_{\text{MAX}} \\ & = o_p(1), \end{aligned}$$

where we also use Lemma 4(c):

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

Next, by Lemma 3(b) and (A.28), we have

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

Therefore, by (A.20) and the continuous mapping theorem,

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

which concludes the proof. \square

Proof of Theorem 7. For $\widehat{\Psi}_{1t}$, we can follow exactly the same proof as that of Theorem 6, since, similar to (A.28) for $\widetilde{\gamma}$, we have the same estimate for \widehat{v}_t by (A.26).

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

$$\left\| H^\top \widehat{\Omega}_t H - \Omega_t \right\|_{\text{MAX}} = o_p(1).$$

Then by the continuous mapping theorem, along with (A.28) and (A.32), we have

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

We analyze the case with estimator (7) first. Note that

$$\left\| H^\top \widehat{\Omega}_t H - \Omega_t \right\|_{\text{MAX}} \leq \frac{1}{n} \left\| \sum_{i=1}^n \left(H^\top \widehat{\beta}_i \widehat{\beta}_i^\top H \widehat{u}_{it}^2 - \beta_i \beta_i^\top u_{it}^2 \right) \right\|_{\text{MAX}} + \frac{1}{n} \left\| \sum_{i=1}^n \beta_i \beta_i^\top (u_{it}^2 - \text{E}u_{it}^2) \right\|_{\text{MAX}}.$$

On the one hand, by the law of large numbers for i.n.i.d sequences, we have

$$\left\| \frac{1}{n} \sum_{i=1}^n \beta_i \beta_i^\top (u_{it}^2 - \text{E}u_{it}^2) \right\|_{\text{MAX}} = o_p(1).$$

On the other hand, writing $\widetilde{\beta} = \widehat{\beta} H$, for some fixed $t \leq T$, we first show

$$\sum_{i=1}^n (\widehat{u}_{it} - u_{it})^2 = O_p(1 + nT^{-1}).$$

In fact, by (A.26), $\|\beta\|_{\text{F}} = O_p(n^{1/2})$, $\|\bar{v}_t\| = O_p(1)$, and Lemma 3(b),

$$\begin{aligned} \sum_{i=1}^n (\widehat{u}_{it} - \bar{u}_{it})^2 &= \sum_{i=1}^n (\widetilde{\beta}_i^\top \widetilde{v}_t - \beta_i^\top \bar{v}_t)^2 \leq \sum_{i=1}^n \left((\widetilde{\beta}_i - \beta_i)^\top (\widetilde{v}_t - \bar{v}_t) + \beta_i^\top (\widetilde{v}_t - \bar{v}_t) - (\beta_i - \widetilde{\beta}_i)^\top \bar{v}_t \right)^2 \\ &\leq K \|\widetilde{v}_t - \bar{v}_t\|^2 \left(\|\widetilde{\beta} - \beta\|_{\text{F}}^2 + \|\beta\|_{\text{F}}^2 \right) + K \|\widetilde{\beta} - \beta\|_{\text{F}}^2 \|\bar{v}_t\|^2 \\ &= O_p(1 + nT^{-1}). \end{aligned}$$

Hence, by (A.2) we have

$$\sum_{i=1}^n (\widehat{u}_{it} - u_{it})^2 \leq 2 \sum_{i=1}^n (\widehat{u}_{it} - \bar{u}_{it})^2 + 2 \|\bar{u}\|_{\text{F}}^2 = O_p(1 + nT^{-1}).$$

Using this as well as Cauchy-Schwartz inequality, we have

$$\frac{1}{n} \sum_{i=1}^n |\widehat{u}_{it}^2 - u_{it}^2| \leq \left(\frac{1}{n} \sum_{i=1}^n (\widehat{u}_{it} - u_{it})^2 + \frac{2}{n} \left(\sum_{i=1}^n u_{it}^2 \sum_{i=1}^n (u_{it} - \widehat{u}_{it})^2 \right)^{1/2} \right) = O_p(n^{-1/2} + T^{-1/2}),$$

then by Lemmas 3(b) and 9(ii), we have

$$\begin{aligned}
\frac{1}{n} \left\| \sum_{i=1}^n (\tilde{\beta}_i - \beta_i)(\tilde{\beta}_i - \beta_i)^\top (\hat{u}_{it}^2 - u_{it}^2) \right\|_{\text{MAX}} &\leq \|\tilde{\beta} - \beta\|_{\text{MAX}}^2 \frac{1}{n} \sum_{i=1}^n |\hat{u}_{it}^2 - u_{it}^2| = o_p(1), \\
\frac{1}{n} \left\| \sum_{i=1}^n (\tilde{\beta}_i - \beta_i)(\tilde{\beta}_i - \beta_i)^\top u_{it}^2 \right\|_{\text{MAX}} &\leq \frac{1}{n} \|\tilde{\beta} - \beta\|_{\text{MAX}}^2 \sum_{i=1}^n u_{it}^2 = o_p(1). \\
\frac{1}{n} \left\| \sum_{i=1}^n \beta_i (\tilde{\beta}_i - \beta_i)^\top u_{it}^2 \right\|_{\text{MAX}} &\leq \frac{1}{n} \|\beta\|_{\text{MAX}} \|\tilde{\beta} - \beta\|_{\text{MAX}} \sum_{i=1}^n u_{it}^2 = o_p(1). \\
\frac{1}{n} \left\| \sum_{i=1}^n (\tilde{\beta}_i - \beta_i) \beta_i^\top u_{it}^2 \right\|_{\text{MAX}} &\leq \frac{1}{n} \|\beta\|_{\text{MAX}} \|\tilde{\beta} - \beta\|_{\text{MAX}} \sum_{i=1}^n u_{it}^2 = o_p(1). \\
\frac{1}{n} \left\| \sum_{i=1}^n \beta_i (\tilde{\beta}_i - \beta_i)^\top (\hat{u}_{it}^2 - u_{it}^2) \right\|_{\text{MAX}} &\leq \|\beta\|_{\text{MAX}} \|\tilde{\beta} - \beta\|_{\text{MAX}} \frac{1}{n} \sum_{i=1}^n |\hat{u}_{it}^2 - u_{it}^2| = o_p(1). \\
\frac{1}{n} \left\| \sum_{i=1}^n (\tilde{\beta}_i - \beta_i) \beta_i^\top (\hat{u}_{it}^2 - u_{it}^2) \right\|_{\text{MAX}} &\leq \|\beta\|_{\text{MAX}} \|\tilde{\beta} - \beta\|_{\text{MAX}} \frac{1}{n} \sum_{i=1}^n |\hat{u}_{it}^2 - u_{it}^2| = o_p(1). \\
\frac{1}{n} \left\| \sum_{i=1}^n \beta_i \beta_i^\top (\hat{u}_{it}^2 - u_{it}^2) \right\|_{\text{MAX}} &\leq \|\beta\|_{\text{MAX}}^2 \frac{1}{n} \sum_{i=1}^n |\hat{u}_{it}^2 - u_{it}^2| = o_p(1).
\end{aligned}$$

Therefore, we have

$$\begin{aligned}
&\frac{1}{n} \left\| \sum_{i=1}^n \left(H^\top \hat{\beta}_i \hat{\beta}_i^\top H \hat{u}_{it}^2 - \beta_i \beta_i^\top u_{it}^2 \right) \right\|_{\text{MAX}} \\
&\leq \frac{1}{n} \left\| \sum_{i=1}^n (\tilde{\beta}_i - \beta_i)(\tilde{\beta}_i - \beta_i)^\top (\hat{u}_{it}^2 - u_{it}^2) + (\tilde{\beta}_i - \beta_i)(\tilde{\beta}_i - \beta_i)^\top u_{it}^2 + \beta_i (\tilde{\beta}_i - \beta_i)^\top (\hat{u}_{it}^2 - u_{it}^2) \right. \\
&\quad \left. + \beta_i (\tilde{\beta}_i - \beta_i)^\top u_{it}^2 + (\tilde{\beta}_i - \beta_i) \beta_i^\top (\hat{u}_{it}^2 - u_{it}^2) + (\tilde{\beta}_i - \beta_i) \beta_i^\top u_{it}^2 + \beta_i \beta_i^\top (\hat{u}_{it}^2 - u_{it}^2) \right\|_{\text{MAX}} \\
&= o_p(1),
\end{aligned}$$

which leads to the desired result.

Finally, we prove (ii) for the estimator (8). Note that by Fan et al. (2013), we have

$$\left\| \hat{\Sigma}^u - \Sigma^u \right\| = O_p(s_n \omega_T^{1-h}). \tag{A.33}$$

Then by (A.33) and Lemmas 3(b), 4(d), and using the fact that $\|\beta\|_{\text{F}} = O_p(n^{1/2})$ and $\|\Sigma^u\| \leq \|\Sigma^u\|_1 = O_p(s_n)$, we have

$$\begin{aligned}
\frac{1}{n} \left\| (\tilde{\beta} - \beta)^\top (\hat{\Sigma}^u - \Sigma^u) (\tilde{\beta} - \beta) \right\|_{\text{MAX}} &\leq \frac{1}{n} \|\tilde{\beta} - \beta\|_{\text{F}}^2 \|\hat{\Sigma}^u - \Sigma^u\| = O_p \left(s_n \omega_T^{1-h} (n^{-1} + T^{-1}) \right), \\
\frac{1}{n} \left\| (\tilde{\beta} - \beta)^\top \Sigma^u (\tilde{\beta} - \beta) \right\|_{\text{MAX}} &\leq \frac{1}{n} \|\tilde{\beta} - \beta\|_{\text{F}}^2 \|\Sigma^u\| = O_p \left(s_n (n^{-1} + T^{-1}) \right),
\end{aligned}$$

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

Therefore,

$$\begin{aligned}
&\left\| H^\top \widehat{\Omega} H - \Omega \right\|_{\text{MAX}} = \frac{1}{n} \left\| H^\top \widehat{\beta}^\top \widehat{\Sigma}^u \widehat{\beta} H - \beta^\top \Sigma^u \beta \right\|_{\text{MAX}} \\
&\leq \frac{1}{n} \left\| (\tilde{\beta} - \beta)^\top (\widehat{\Sigma}^u - \Sigma^u) (\tilde{\beta} - \beta) + (\tilde{\beta} - \beta)^\top \Sigma^u (\tilde{\beta} - \beta) \right\|_{\text{MAX}} + \frac{1}{n} \left\| \beta^\top (\widehat{\Sigma}^u - \Sigma^u) \beta \right\|_{\text{MAX}} \\
&\quad + \frac{1}{n} \left\| \beta^\top (\widehat{\Sigma}^u - \Sigma^u) (\tilde{\beta} - \beta) + (\tilde{\beta} - \beta)^\top (\widehat{\Sigma}^u - \Sigma^u) \beta \right\|_{\text{MAX}} + \frac{1}{n} \left\| \beta^\top \Sigma^u (\tilde{\beta} - \beta) + (\tilde{\beta} - \beta)^\top \Sigma^u \beta \right\|_{\text{MAX}} \\
&= O_p \left(s_n \left(\omega_T^{1-h} + n^{-1} + T^{-1} \right) \right) = o_p(1),
\end{aligned}$$

which concludes the proof. \square