Factor models with many assets: strong factors, weak factors, and the two-pass procedure

Stanislav Anatolyev[*] — CERGE-EI and NES
Anna Mikusheva[†] — MIT

August 2017

PRELIMINARY AND INCOMPLETE. PLEASE, DO NOT DISTRIBUTE!

Abstract

This paper re-examines the problem of estimating risk premia in factor pricing models. A typical feature of data used in the empirical literature to estimate such models is the presence of weak factors that are priced and, at the same time, the presence of unaccounted strong cross-sectional dependence in the errors. Another feature of typically used data is (moderately) high cross-sectional dimensionality. Using an asymptotic framework where the number of asset/portfolios grows proportionately with the time span of the data while the risk exposures of weak factors are local-to-zero, we show that under such circumstances the conventional two-pass estimation procedure delivers inconsistent estimates of the risk premia. We propose a modified two-pass procedure based on sample-splitting instrumental variables estimation at the second pass. The proposed estimator of risk premia is robust to the presence of strong unaccounted cross-sectional error dependence, as well as to the presence of included factors that are priced but weak. We derive the many-asset weak factor asymptotic distribution of the proposed estimator, show how to construct its standard errors, verify its performance in simulations, and apply it to often-used datasets from existing empirical studies.

Keywords: factor models, price of risk, risk premia, two-pass procedure, strong factors, weak factors, dimensionality asymptotics, weak factor asymptotics.

JEL classification codes: C33, C38, C58, G12.

[*]Address: Stanislav Anatolyev, CERGE-EI, Politickych vězňů 7, 11121 Prague 1, Czech Republic; e-mail stanislav.anatolyev@cerge-ei.cz.

[†]Address: Department of Economics, M.I.T., 50 Memorial Drive, Building E52, Cambridge, MA, 02142, USA; e-mail amikushe@mit.edu. We are grateful to David Hughes for research assistance.

1 Introduction

Since the introduction of CAPM by Sharpe (1964) and Linner (1965), linear factor pricing models have grown into a very popular sub-field in asset pricing. Harvey, Liu and Zhu (2016) list hundreds of papers proposing, justifying and estimating various factor pricing models. A typical paper in this area proposes a small set (in the range of 3-5) of observed risk factors. The classical factor is the market portfolio excess return, which was augmented by the size factor 'SMB' (small-minus-big return) and the book-to-market factor 'HML' (high-minus-low return) in Fama and French (1993). Among other well-known pricing factors are the momentum factor 'MOM' (Jegadeesh and Titman, 1993) and the consumption-to-wealth ratio 'cay' (Lettau and Ludvigson, 2001).

Traditionally, the model is estimated using what is commonly known as the two-pass estimation procedure (Fama and MacBeth, 1973; Shanken, 1992).[1] This procedure, however, relies on an idealistic setup with strong identification of risk premia. Empirically, in realistic circumstances such conditions often do not hold. For example, the recent literature shows that the mismeasurement of the true risk factors leads to weakness in the observed factors and strong cross-sectional dependence in the errors (Kleibergen and Zhan (2015)), which may result in all sorts of distortions in estimation and inference in theory and their non-reliability in practice (e.g., Kan and Zhang (1999), Andrews (2005), Kleibergen (2009)). Recent papers by Kan and Zhang (1999), Kleibergen (2009), Bryzgalova (2015), Burnside (2015), Gospodinov, Kan and Robotti (2016) all point out that risk exposures (or betas) to some observed factors tend to be small to such an extent that the their estimation errors are of the same order of magnitude as the betas themselves. The observed phenomenon is very similar to the widely studied weak instruments problem. The remedies for some of these failures proposed in the literature rely either on complicated inference tools robust to weak identification (Kleibergen (2009)), or require the use of dimension-reduction techniques (Bryzgalova (2015)).

Along with a combination of problems of small betas and missing factors, we also consider one very important empirical feature of typically employed datasets – the presence of a large number of assets or portfolios often comparable to the number of periods over which returns are observed. We consider an asymptotic framework where the number of assets/portfolios of which returns data is used for estimation grows with its time-series

[1] Sometimes the two-pass procedure is referred to the Fama-MacBeth procedure (Fama and MacBeth, 1973). See Cochrane (2001, section 12.3) on their numerical equivalence when betas are time invariant. The method of obtaining valid standard errors that account for the two step nature of the procedure was given in Shanken (1992).

dimension. Such dimension asymptotics is likely to provide a more accurate asymptotic approximation to the finite sample properties of estimators and tests. As is known from the conceptually-related literature on many instruments (Bekker, 1994) and many regressors (Anatolyev, 2012), large distortions in conventional estimation and inference may arise when the number of instruments and/or regressors ceases to be a tiny fraction of the sample size, and corrections of estimators and tests require a framework with the dimension asymptotics within which these fractions are asymptotically non-zero. The many-asset asymptotic framework was already utilized previously by Gagliardini, Ossola and Scaillet (2016).

Within the new dimensionality asymptotics we show that the presence of small betas leads to a failure of the classical two-pass procedure, while the additional presence of missing factors exacerbates the problem. We propose econometric procedures that are robust to both of the issues with factors – the weakness of observed factors and the presence of unobserved factors in the errors, – and, in contrast to the remedies proposed elsewhere, are easily implementable using standard regression tools (in particular, instrumental variables regressions and two-stage least squares). The estimators we propose are consistent and asymptotically mixed gaussian; moreover, using the variance estimators whose construction we describe, the standard inference tools such as t- and Wald tests can be applied in a conventional way.

Our usage of dimensionality asymptotics has an important implication: even though we assume that some observed factors have small betas, it does *not* lead to a weak identification problem as currently defined in the literature, and we can obtain a consistent estimator. This distinction is similar to the one between the literature on "weak instruments" and the literature on "many weak instruments" (e.g., Hansen, Hausman and Newey, 2008). In the latter, the weakness of instruments implies identification if these instruments are numerous and, while the classical two stage least squares (TSLS) estimator has large distortions, one can construct a consistent and asymptotically gaussian estimation procedure. Our new estimation approach uses some ideas from the many-weak-instruments literature such as sample-splitting and the use of instrumental variables estimation at the second step. However, we additionally face and solve a distinct problem that one does not encounter in the many weak instrument literature, namely, that of the presence of an unobserved factor structure in the error terms that creates strong cross-sectional dependence in the panel of returns and is similar to the classical omitted-variables problem in the second-pass regression of the two-pass procedure.

The paper is organized as follows. Section 2 introduces notation, discusses the relevance of our asymptotic approach, and argues for the presence of a significant factor structure in the errors. Section 3 introduces and discusses technical assumptions. Section 4 explains the asymptotic failure of the classical two-pass procedure and provides detailed intuition as to why that happens. We propose our estimation method in Section 5, describe what motivates it and explain why it works. We also state the formal theorem on the consistency of the newly-proposed 4-split estimator. Section 6 is devoted to deriving inference procedures that use our 4-split estimator, in particular, we show the asymptotic validity of a properly constructed Wald test. The results of numerous simulations support our theoretical finding and are placed in Section 7. Section 8 revisits some prominent empirical applications of the factor-pricing model.

A word on notation: $0_{l,m}$ stays for a zero matrix of size $l \times m$, I_m is an $m \times m$ identity matrix, for an $m \times l_A$ matrix A and an $m \times l_B$ matrix B symbol $(A; B)$ stays for the $m \times (l_A + l_B)$ matrix obtained from placing the initial matrices side-to-side.

2 Formulation of the problem

As mentioned in the Introduction, a paper in the area of financial factor models typically proposes a small set of observed risk factors described by a $k_F \times 1$ vector F_t, with a (usually) small dimension k_F. An asset or portfolio of assets i with excess return r_{it} has exposure to several risk factors, which is quantified by the asset's betas $\beta_i = var(F_t)^{-1} cov(F_t, r_{it})$. A typical claim in linear factor-pricing theory is that exposure to risk (betas) fully determines the assets' expected returns. Particularly, there exists a $k_F \times 1$-dimensional vector of risk premia λ such that $Er_{it} = \lambda' \beta_i$.

The linear factor-pricing model is equivalent to the following formulation:

$$r_{it} = \lambda' \beta_i + (F_t - EF_t)' \beta_i + \varepsilon_{it}, \tag{1}$$

where the random error terms ε_{it} have mean zero and are uncorrelated with F_t. One special case often mentioned in the literature occurs when the F_t factors are asset returns themselves and are supposed to be priced by the same model; in this case, theoretically $\lambda = EF_t$. We will not make this assumption and will consider the general case when λ may differ from EF_t.

Two-pass procedure The estimation and formation of inferences on risk prices, λ, are usually accomplished by a procedure commonly known as the two-pass estimation procedure (Fama and MacBeth, 1973; Shanken, 1992), which is applied to a data set consisting of a panel of asset excess returns $\{r_{it},\ i = 1, ..., N,\ t = 1, ..., T\}$ and of observations of realized factors $\{F_t,\ t = 1, ..., T\}$. As a first step, one estimates β_i by running a time series OLS regression of r_{it} on a constant and on F_t for each $i = 1, ..., N$. The second step produces an estimate of λ (denote it $\widehat{\lambda}_{TP}$) by regressing the time-average excess return $\frac{1}{T}\sum_{t=1}^{T} r_{it}$ on the first-step estimates, $\widehat{\beta}_i$. Under suitable conditions, $\widehat{\lambda}_{TP}$ is proved to be both consistent and asymptotically gaussian. Discussions of the statistical properties of the two-pass procedure appear in Fama and MacBeth (1973), Shanken (1992), and Chapter 12 of Cochrane (2001).

Recently, several prominent researchers have raised the concern that the two-pass procedure may provide misleading estimates of risk premia; see, for example, Kan and Zhang (1999), Kleibergen (2009), Bryzgalova (2015), Burnside (2015), Gospodinov, Kan and Robotti (2016). They surmise that the reason for these erroneous inferences is attached to the empirical observation that either some column of $\beta = (\beta_1, ..., \beta_N)'$ is close to zero, or, more generally, the $N \times k_F$ matrix β appears close to one of reduced rank (less than k_F) for a majority of well-known linear factor-pricing models. The observed phenomenon is similar to the widely studied weak-instruments problem (Staiger and Stock, 1998): if some of the observed factors F_t are only weakly correlated with all of the returns in the data set, then the noise that arises in the first-stage estimates of the corresponding components of β_i will dominate the signal, and the second-pass estimate of the risk premia λ will be over-sensitive to small variations in the sample.

In order to model the observed phenomenon, Kleibergen (2009), Bryzgalova (2015) and Gospodinov, Kan and Robotti (2016), among others, considered a drifting-parameters framework in which they model some component of β_i to be of order $O(\frac{1}{\sqrt{T}})$ assuming that the number of time periods, T, increases to infinity, while the number of assets, N, stays fixed. In such a setting the first-step estimation error is of order of magnitude $O_p(\frac{1}{\sqrt{T}})$, which is comparable to the size of the coefficients themselves. The two-pass procedure could be reformulated as an IV-type estimation, and then the above-described drifting-sequence asymptotics characterize the weak-instruments case. This framework implies inconsistency of the two-pass estimates for the risk premium on small components, poor coverage of regular confidence sets even for the risk premium of strong factors, and asymptotic invalidity of classical specification tests and tests of hypotheses about risk premia.

Following this tradition and acknowledging empirical evidence provided in Kleibergen and Zhan (2015) and Bryzgalova (2015) we also make use of drifting-parameter modeling. We assume that the $k_F \times 1$ vector of factors F_t can be divided into two subvectors: a

$k_1 \times 1$ dimensional vector $F_{t,1}$ and a $k_2 \times 1$ vector $F_{2,t}$ (here $k_F = k_1 + k_2$) such that the risk exposure $\beta_{i,1}$ to factor $F_{t,1}$ will be strong, while the risk exposure coefficients $\beta_{2,i}$ to factor $F_{2,t}$ will be drifting to zero at speed $\frac{1}{\sqrt{T}}$. We make these order assumptions for risk exposure more accurate in the next section. The most important feature of our modeling will be that the standard error of the first-step estimator of $\beta_{2,i}$ will be of the same order of magnitude as the coefficient itself. A more general treatment of the near-degenerate rank condition considers some k_2-dimensional linear combination of factors (unknown to the researcher) to have a local-to-zero (of order $O(\frac{1}{\sqrt{T}})$) exposure coefficient, while the exposure to risk formed by the orthogonal k_1-dimensional linear combination remains fixed. All our results are easily generalizable to this setting, as we do not assume the researcher knows which factors (or combination of factors) bear small coefficients of exposure. However, to simplify the exposition we will stick to the division of factors into two sub-vectors.

This paper deviates from the previous literature in two directions. First, we consider an asymptotic setting where both N and T grow to infinity. We notice that in many common data sets researchers use in the estimation of factor pricing models, the number of assets, N, is large when compared to the number of time periods. The celebrated Fama-French data set provides returns on $N = 25$ sorted portfolios for about $T = 200$ periods. The often-used Jagannathan-Wang data set (Jagannathan and Wang (1996)) contains observations on $N = 100$ portfolios observed for $T = 330$ periods. Lettau and Ludvigson (2001) use Fama-French $N = 25$ portfolios, the returns for which are observed during $T = 141$ quarters. Gagliardini, Ossola and Scaillet (2016) use $N = 44$ industry portfolios observed during $T = 546$ months. In these cases it is hard to believe that the asymptotic results derived under the assumption that N is fixed would provide an accurate approximation of finite-sample distributions. Indeed, among other things, Kleibergen (2009) discovers that the bias of the two-pass estimate of risk premia is strongly and positively related to the number of assets if the total factor strength is kept constant.

In this paper we consider asymptotics when both N and T increase to infinity without restricting the relative growth between them.

The second and main deviation of this paper from the existing literature is our explicit acknowledgment of high cross-sectional dependence among error terms ε_{it} in model (1). In particular, we assume that errors have a factor structure. Namely, this means that there exists a missing (unknown and unobserved to the researcher) factor v_t and loadings μ_i such that

$$\varepsilon_{it} = v_t' \mu_i + e_{it},$$

where the 'clean' errors e_{it} are only weakly cross-sectionally dependent to such an extent that asymptotically we may ignore their dependence (exact formulation of this assumption appears in the next section). Similar weak-dependence assumptions appear in approximate factor models (e.g., Bai and Ng (2002)). The assumptions on loadings μ_i guarantee that the factor structure will be strong enough to be both detected empirically and asymptotically important for inferences. An insightful discussion of weak vs strong factor structure and cross-sectional dependence appears in Onatski (2012).

Below we provide two theoretical reasons as to why we expect factor structure in many linear factor-pricing models. Then we point to empirical evidence that a missing factor structure is indeed present in some well-known factor-pricing models.

Example 1. If one does not observe the true risk factors that price assets but only proxies for them, this would lead to a factor structure in errors (see Kleibergen and Zhan (2015), who show in particular that this further leads to spuriously large values of second-pass R^2). Assume for a moment that the market is priced by risk premia on risk factors G_t. For expositional simplicity we assume that

$$r_{it} = G_t \beta_i^G + \varepsilon_{it}^G,$$

where the shocks ε_{it}^G are drawn with mean zero and finite variance independently cross-sectionally and time-series and independent from G_t. Also assume that G_t is stationary with variance Σ_G.

Assume that the econometrician does not observe G_t, but rather has a proxy for it, $F_t = \alpha + \delta G_t + \epsilon_t$, where ϵ_t has mean zero and is uncorrelated with G_t and shocks ε_{it}^G. For example, ϵ_t may stand for a measurement error or contamination by other macro variables not connected to asset prices. Denote the variance matrix of ϵ_t by Σ_ϵ. If δ is a full-rank square matrix, then one can show that proxies F_t can price assets as well as the true risk factors G_t. Indeed,

$$\beta_i = Var(F_t)^{-1} Cov(F_t, r_{it}) = (\delta \Sigma_G \delta' + \Sigma_\epsilon)^{-1} \delta \Sigma_G \beta_i^G = A \beta_i^G,$$

where the matrix $A = (\delta \Sigma_G \delta' + \Sigma_\epsilon)^{-1} \delta \Sigma_G$ is a full-rank square matrix. Thus,

$$E r_{it} = E G_t \beta_i^G = \lambda \beta_i,$$

where $\lambda = A^{-1} E G_t$. So we see that if the econometrician is trying to estimate a linear factor-pricing model using factors F_t, she has a correctly-specified model; however, this model (unlike the model with the observed factors G_t) has a factor structure in its error terms. Indeed, using some simple algebra one can show that equation (1) holds with

$$\varepsilon_{it} = (\Sigma_\epsilon \delta^{-1} \Sigma_G^{-1} (G_t - EG_t) - \epsilon_t)' \beta_i + \varepsilon_{it}^G = v_t' \mu_i + \varepsilon_{it}^G.$$

What is interesting here is that while the factors v_t (and the errors ε_{it} themselves) are uncorrelated with the observed factors F_t, the loadings on the error factors, μ_i's, and the original loadings, β_i's, are closely related (in this particular case $\mu_i = \beta_i$). We will make use of this observation in our discussion of the validity of the two-pass procedure. □

Example 2. Consider a situation in which one of the risk factors driving asset returns is fully arbitraged and thus carries a risk premia of zero. If an econometrician does not observe this factor but does have observations on all other relevant risk factors, then her linear factor-pricing model that omits the arbitraged factor may still be correctly specified, while the arbitraged factor is moved to the error term, resulting a missing-factor structure to the errors. □

Kleibergen and Zhan (2015) provide numerous pieces of empirical evidence that residuals from many well-known estimated linear factor-pricing models have non-trivial factor structures. For example, they point out that the first three principle components of the residuals from different pricing-model specifications used in the prominent paper by Lettau and Ludvigson (2001) explain somewhere from 82% to 95% of all residual variation (Kleibergen and Zhan (2015, Table 3)). They also show that the largest eigenvalue of the covariance matrix of residuals in all these examples is very large and strongly separated from other eigenvalues that are bunched together. Combining these results on the largest eigenvalues of the residual covariance matrix with the theoretical results on the limiting distribution of its eigenvalues from Onatski (2012), one would suspect there is at least one strong factor present in the residuals. At least five other prominent factor-pricing studies cited in Kleibergen and Zhan (2015) demonstrate similar evidence of a strong factor structure left in the residuals.

Relation between factor structure and correct specification. One may wonder whether the fact that the errors ε_{it} in the model (1) have a factor structure implies that the pricing model is misspecified. The answer is "No"; the linear factor-pricing model describes the expectations of excess returns, while the factor structure in the errors is related to the covariances or co-movements of the assets' returns. It is easy to see that if the risk exposure and risk premia on the variables F_t price the assets, then the variables

F_t co-move the assets' returns and produce factor-structure dependence in the returns. However, not all co-movements of returns must carry non-zero risk premia; those co-movements can be placed in the error term without causing misspecification of the pricing model.

In this paper we assume that the correct specification of a pricing model requires keeping in the model those pricing factors $F_{t,2}$ that carry small coefficients of exposure $\beta_{2,t}$ and produce only a weak factor structure in returns. We show that dropping such observed factors from the specification (as opposed to what is proposed in Bryzgalova (2015)) leads to asymptotically misleading inferences for both the two-pass procedure and our proposed procedure. Our method is robust to minor misspecifications that allow one to drop those pricing factors that carry loadings of order $o(\frac{1}{\sqrt{NT}})$; the exact formulation appears in Section 5.3.

3 Setup and assumptions

3.1 Model

We consider the problem of estimation and inference on the risk premia λ based on observations of returns $\{r_{it}, i = 1, ..., N, t = 1, ..., T\}$ and factors $\{F_t, t = 1, ..., T\}$ coming from a correctly specified factor-pricing model:

$$r_{it} = \lambda'\beta_i + (F_t - EF_t)'\beta_i + v_t'\mu_i + e_{it}, \tag{2}$$

where the 'correct specification' means that the random unobserved factor v_t has zero mean and is uncorrelated with F_t, the idiosyncratic error terms e_{it} also have zero mean and are uncorrelated with F_t. We also assume that they are uncorrelated with v_t. Denote $\mathcal{F}$ to represent the sigma-algebra generated by the random variables $(F_1, ..., F_T)$ and $(v_1, ..., v_T)$; let $\gamma_i' = \left(\beta_{1i}', \sqrt{T}\beta_{2i}', \mu_i'\right)$, and $\Gamma_N' = (\gamma_1, ..., \gamma_N)$.

3.2 Assumptions

We make the following assumptions.

Assumption FACTORS. The $k_F \times 1$ vector of observed factors F_t is stationary with finite fourth moments, a full-rank covariance matrix Σ_F, and summable auto-covariances. The $k_v \times 1$ vector of unobserved factors v_t is such that the following asymptotic statements hold simultaneously:

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(F_t - EF_t) \Rightarrow N(0, \Omega_F);$$

$$\eta_T = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Sigma_F^{-1}\widetilde{F}_t v_t' \Rightarrow \eta;$$

$$\eta_{v,T} = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} v_t \Rightarrow \eta_v \sim N(0_{k_v,1}, I_{k_v}),$$

where $\mathbf{vec}(\eta) \sim N\left(0_{k_F k_v,1}, \Omega_{vF}\right)$ and $\widetilde{F}_t = F_t - \frac{1}{T}\sum_{s=1}^{T} F_s$.

Assumption LOADINGS. As both N and T increase to infinity, we have $N^{-1}\Gamma_N'\Gamma_N \to \Gamma$, where Γ is a positive-definite $k \times k$ matrix with $k = k_F + k_v$. In addition we assume that $\max_{N,T}\frac{1}{N}\sum_{i=1}^{N}\|\gamma_i\|^4 < \infty$.

We adopt the following notation: $\Gamma_{\beta_2\mu}$ is the $k_2 \times k_\mu$ sub-block of matrix Γ corresponding to the limit of $N^{-1}\sum_{i=1}^{N}\sqrt{T}\beta_{i,2}\mu_i$. Other sub-matrices are denoted similarly.

Assumptions ERRORS.

(i) Conditional on $\mathcal{F}$, the random vectors $e_t = (e_{1t}, ..., e_{Nt})'$ are serially independent, and $E(e_t|\mathcal{F}) = 0$ for all t.

(ii) Let $\rho(t,s) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N} e_{it}e_{is}$. Then, $\sup_t \sup_{s\neq t} E\left[(1 + \|F_t\|^4)(\rho(s,t)^2 + 1)\right] < C$.

(iii) Let $S_t = \frac{1}{N}\sum_{i=1}^{N} e_{it}^2$. Then, $\frac{\sqrt{N}}{T}\sum_{t=1}^{T}\widetilde{F}_t S_t = o_p(1)$ and $\frac{1}{T}\sum_{t=1}^{T}\widetilde{F}_t\widetilde{F}_t' S_t \to^p \Sigma_{SF^2}$.

(iv) Let $W_t = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\gamma_i e_{it}$. Then, $E\left[(1 + \|F_t\|^2)\|W_t\|^2\right] < \infty$.

3.3 Discussion of Assumptions

Assumptions FACTORS. As a part of the error term, v_t is uncorrelated with F_t. One can come up with a variety of assumptions on decaying dependence and moment conditions that would guarantee some Central Limit Theorem stated in (i) and (ii). The restriction that the asymptotic covariance matrix be the identity matrix is just normalization, as neither v_t nor loadings μ_i are observed.

Assumption LOADINGS. In this paper we treat the loadings β_i and μ_i as unknown constant (non-random) vectors the true values of which may change with the sample sizes N and T, which is an example of the so-called "drifting parameters asymptotics." Assumption LOADINGS characterizes the size of the loadings as the sample size increases. Notice that the loadings on the factors $F_{t,1}$ and v_t are treated differently than the loadings on $F_{t,2}$; following Onatski (2012) we will refer to the former as "strong factors" and the latter as "weak factors." The cross-sectional average of squared loadings is closely connected to the explanatory power the factors exhibit in cross-sectional variation. The assumptions we make on the loadings $\beta_{i,1}$ and μ_i guarantee that the explanatory power of the factors $F_{t,1}$ and v_t dominates that of the idiosyncratic error terms. The average squared loading on the factor $F_{2,t}$, however, converges to zero at a rate of $1/T$; if N and T increase proportionally, this will lead to factor $F_{2,t}$ having explanatory power comparable to that of the idiosyncratic errors. One characteristic of a weak factor is the following: if it had not been observed we could neither have consistently estimated it via the method of principle components applied to the estimated cross-sectional covariance matrix nor consistently detected it.

The loadings $\beta_{i,2}$ are asymptotically of the same order of magnitude as $\beta_{i,1}$ divided by $\sqrt{T}$. Assumption LOADINGS makes the standard deviation of the first-step estimate $\widehat{\beta}_{i,2}$ of the same order of magnitude as $\beta_{i,2}$ itself. As we show below, this is enough to make the two-pass estimator of the risk premia λ_2 on the weak factor $F_{t,2}$ inconsistent and to invalidate the classical confidence interval for the risk premia λ_1 on the strong factor $F_{t,1}$. The modeling assumption that makes $\beta_{i,2}$ drift to zero at the $1/\sqrt{T}$ rate is similar to assumptions Kleibergen (2009), Bryzgalova (2015) and Gospodinov, Kan and Robotti (2016) make. In these papers the authors assume that N remains fixed, which makes the asset premia λ_2 a weakly-identified parameter. We, however, assume that N increases to infinity which, together with this assumption, allows one to construct a consistent estimator for λ_2, the two-pass estimator still being inconsistent. Thus, our setting is not a case of weak identification.

It is also important that the assumption on loadings μ_i be such that the unobserved factor v_t in the error terms is strong. This is consistent with the empirical observations Kleibergen and Zhan (2015) present. This also guarantees that the presence of the factor structure plays an important role in the asymptotics of two-pass estimation. The error terms may also have weak factor structure; we do not explicitly specify this because it will not have asymptotic importance for the estimation procedures we consider here.

Assumption ERRORS. Assumptions ERRORS are high-level assumptions the main goal of which is to allow very flexible weak cross-sectional dependence among the idiosyncratic errors, as well as flexible conditional heteroscedasticity and dependence in higher-order moments of errors and factors. The random variables $\rho(s,t)$ stand for a (normalized) empirical analog of the error autocorrelation coefficient, S_t is an empirical variance, and W_t is a (normalized) weighted average error. These variables are normalized so that they are stochastically bounded when the errors are cross-sectionally i.i.d.

Serial independence of errors as stated in Assumption ERRORS(i) is consistent with the efficient market hypothesis and the unpredictability of asset returns; and is generally consistent with empirical evidence and the tradition in the literature. This assumption may be weakened, though we do not pursue this in the current paper.

In order to understand Assumptions ERRORS we provide below a set of more restrictive primitive assumptions that are common in the literature and that guarantee the validity of our high-level Assumptions ERRORS. We also provide an empirically relevant example not covered by the primitive assumptions below but that satisfies the high-level Assumptions ERRORS.

Assumptions ERRORS*

(i) The factors $\{F_t,\ t=1,...,T\}$ are independent from errors $\{e_{it},\ i=1,...,N,\ t=1,...,T\}$; the error terms $e_t=(e_{1t},...,e_{Nt})'$ are serially independent and identically distributed for different t with $Ee_{it}=0$ and $\sup_{i,t} Ee_{it}^4<\infty$.

(ii) Let $\mathcal{E}_{N,T}=E\left[e_te_t'\right]$ be the $N\times N$ covariance matrix when the sample size is N and T (in cross-section and time directions, correspondingly). For some positive constants a, c and C,

$$c<\liminf_{N,T\to\infty}\min eval\left(\mathcal{E}_{N,T}\right)<\limsup_{N,T\to\infty}\max eval\left(\mathcal{E}_{N,T}\right)<C,$$

and $\lim_{N,T}\frac{1}{N}\mathrm{tr}(\mathcal{E}_{N,T})=a$.

(iii) $E\left|\frac{1}{\sqrt{N}}\sum_{i=1}^N(e_{it}^2-Ee_{it}^2)\right|^2<C$.

Lemma 1 *Assumptions LOADINGS and Assumptions ERRORS* imply Assumptions ERRORS.*

The primitive Assumptions ERRORS* are very close to the standard ones in the literature. Numerous papers that establish inferences in factor models, commonly assume that the set of variables $\{F_t,\ t = 1, ..., T\}$ is independent from the set $\{e_{it},\ i = 1, ..., N,$ $t = 1, ..., T\}$, though within-group dependence is allowed; see, for example, Assumption D in Bai and Ng (2006). Many papers allow for both time-series and cross-sectional error dependence. We exclude time-series dependence which is justified by the efficient-market hypothesis in our application. Assumption ERRORS*(ii) is intended to impose only weak dependence cross-sectionally as expressed by the covariance matrix; similar assumptions appear in Onatski (2012), and a stronger form is used in Bai and Ng (2006).

Our high-level Assumptions ERRORS are much more general than the more standard primitive Assumptions ERRORS*. In particular, our assumptions allow for very flexible conditional heteroscedasticity in the error terms and time-varying cross-sectional dependence, which seems relevant when we consider observed factors that characterize market conditions like the momentum factor. Consider the following example.

Example 3. Assume that errors e_{it} have the following weak (unobserved) factor structure:

$$e_{it} = \pi_i' w_t + \eta_{it},$$

where (w_t, F_t) is stationary, w_t is a $k_w \times 1$ serially independent, conditional on $\mathcal{F}$, times series with $E(w_t|\mathcal{F}) = 0$ and $E(w_t w_t') = I_{k_w}$ (which is an innocuous normalization as the factor structure is not observed). Assume $E\left[(\|F_t\|^4 + 1)(\|w_t\|^4 + 1)\right] < \infty$. We assume that the loadings satisfy the condition $\sum_{i=1}^{N} \pi_i \pi_i' \to \Gamma_\pi$ (the factors w_t are weak), and $N^{-1/2} \sum_{i=1}^{N} \pi_i \gamma_i' \to \Gamma_{\pi\gamma}$. Assume that the random variables η_{it} are independent both cross-sectionally and across time, are independent from w_t and F_t, have mean zero and finite fourth moments and variances σ_i^2 that are bounded above and such that $N^{-1} \sum_{i=1}^{N} \sigma_i^2 \to \sigma^2$. As proven in the Appendix, this example satisfies Assumptions ERRORS.

An interesting feature of this example is that it allows the errors to be weakly cross-sectionally dependent to the extent that they may possess a weak factor structure. Moreover, this factor structure may be closely related to the observed factors F_t, which causes the cross-sectional dependence among the errors e_{it} to change with the observed factors F_t and allows a very flexible form of conditional heteroskedasicity. Indeed, the conditional cross-sectional covariance is

$$E(e_{it} e_{jt}|\mathcal{F}) = \pi_i' E(w_t w_t'|\mathcal{F}) \pi_j + \mathbb{I}_{\{i=j\}} \sigma_i^2.$$

Since we do not restrict $E(w_t w_t'|\mathcal{F})$ beyond the proper moment conditions, the strength of any cross-sectional dependence as well as error variances may change stochastically depending on the realizations of the observed factors. This flexibility is extremely relevant for such observed factors as the momentum. For example, one may consider $w_t = \varsigma_t g(F_t, F_{t-1}, ...)$, where $\varsigma_t \sim N(0,1)$ is independent from all other variables; then for a proper choice of the function $g(\cdot)$ one may observe higher volatility and cross-sectional dependence of the idiosyncratic error for higher values of the observed factor F_t.

4 Asymptotic properties of the two-pass procedure

In this section we derive a result concerning the asymptotic properties of the classical two-pass procedure in different models that may or may not include weak observed factors and may or may not have strong missing factors in errors. Let us introduce the following notation:

$$\widetilde{\lambda} = \lambda + \frac{1}{T}\sum_{t=1}^{T} F_t - EF_t, \quad u_i = \frac{1}{T}\sum_{t=1}^{T} \Sigma_F^{-1}\widetilde{F}_t e_{it},$$

where $\widetilde{F}_t = F_t - \frac{1}{T}\sum_{s=1}^{T} F_s$.

Now let us introduce two asymptotically important terms, the meaning and the names of which will be explained in the discussion following Theorem 1. The first term we call "attenuation bias":

$$AB = -\left(\sum_{i=1}^{N} \widehat{\beta}_i\widehat{\beta}_i'\right)^{-1} \sum_{i=1}^{N} u_i u_i'\widetilde{\lambda},$$

while the second is known as "omitted variable bias"

$$OVB = \left(\sum_{i=1}^{N} \widehat{\beta}_i\widehat{\beta}_i'\right)^{-1} \sum_{i=1}^{N} \widehat{\beta}_i \frac{\mu_i'}{\sqrt{T}}(\eta_{v,T} - \eta_T'\widetilde{\lambda}).$$

These terms are not biases in an exact sense as they are random, but rather they are sample analogues of the expressions that are classically called attenuation and omitted variable biases. Notice that both quantities are infeasible – they cannot be calculated from the data alone as they depend on unobserved errors e_{it}, unobserved factors v_t and unknown parameters λ and μ_i. Both terms are $k_F \times 1$ vectors. Let AB_1 and OVB_1 denote $k_1 \times 1$ sub-vector consisting of the first k_1 components, while AB_2 and OVB_2 are $k_2 \times 1$ sub-vectors of the last k_2 components of AB and OVB correspondingly.

Theorem 1 *Assume that the sample $\{r_{it},\, i = 1, ..., N,\, t = 1, ..., T\}$ and $\{F_t,\, t = 1, ..., T\}$ comes from a data-generating process that satisfies the factor-pricing model (2) and assumptions FACTORS, LOADINGS and ERRORS. Let $\widehat{\lambda}_{TP}$ denote the estimate obtained* *via the conventional two-pass procedure. Let both N and T increase to infinity without restrictions on relative rates. Then the following asymptotic statements hold simultaneously:*

$$\begin{pmatrix} \sqrt{T}OVB_1 \\ OVB_2 \end{pmatrix} \Rightarrow \left((I_{k_\beta};\widetilde{\eta})\Gamma(I_{k_\beta};\widetilde{\eta})' + \mathcal{I}_{k_2}\Sigma_u\mathcal{I}_{k_2}\right)^{-1}(\Gamma_{\beta\mu} + \widetilde{\eta}\Gamma_{\mu\mu})(\eta_v - \eta'\lambda),$$
$$\begin{pmatrix} \sqrt{T}AB_1 \\ AB_2 \end{pmatrix} \Rightarrow -\left((I_{k_\beta};\widetilde{\eta})\Gamma(I_{k_\beta};\widetilde{\eta})' + \mathcal{I}_{k_2}\Sigma_u\mathcal{I}_{k_2}\right)^{-1}\mathcal{I}_{k_2}\Sigma_u\lambda,$$
$$\sqrt{T}(\widetilde{\lambda} - \lambda) \Rightarrow N(0, \Omega_F),$$

and

$$\begin{pmatrix} \sqrt{NT}(\widehat{\lambda}_{TP,1} - \widetilde{\lambda}_1 - AB_1 - OVB_1) \\ \sqrt{N}(\widehat{\lambda}_{TP,2} - \widetilde{\lambda}_2 - AB_2 - OVB_2) \end{pmatrix} = O_p(1),$$

where $\Sigma_u = \lim_{N,T\to\infty} \frac{T}{N}\sum_{i=1}^{N} u_i u_i'$, with the last convergence being established in Lemma 4 in the Appendix, $\mathcal{I}_{k_2} = \begin{pmatrix} 0_{k_1,k_1} & 0_{k_1,k_2} \\ 0_{k_2,k_1} & I_{k_2} \end{pmatrix}$ is a $k_F \times k_F$ matrix, $\widetilde{\eta} = \mathcal{I}_{k_2}\eta$ is a $k_F \times k_v$ random matrix (with η described in Assumptions FACTORS), and Ω_F is the long-run variance of F_t.

Theorem 1 states the rate of convergence for different parts of the two-pass estimator. Notice that the theorem does not impose a relative rate of increase between N and T as long as both increase to infinity simultaneously. One observation is that the two-pass procedure cannot estimate λ at a rate faster than $\sqrt{T}$ despite the fact that the dataset has NT observations of portfolio excess returns, and one could expect the $\sqrt{NT}$ rate. This comes from the fact that the correct specification implies that the excess returns satisfy equation (1), which, if averaged across time, gives:

$$\overline{r}_i = \widetilde{\lambda}\beta_i + \overline{\varepsilon}_i. \tag{3}$$

Thus, even if β_i were known, the 'true' coefficient $\widetilde{\lambda}$ in the only ideal regression we have (that is, regression of average return on β_i) differs from the parameter λ we want to estimate, by the term $\frac{1}{T}\sum_{t=1}^{T} F_t - EF_t$, which, if multiplied by $\sqrt{T}$, is asymptotically zero mean gaussian with variance Ω_F. Notice that if all observed factors F_t are excess returns themselves and are assumed to be priced by the same pricing model, then the Asset Pricing Theory provides an alternative way of estimating risk premia. Namely in such a case $\lambda = EF_t$, and the alternative estimate $\widehat{\lambda} = \frac{1}{T}\sum_{t=1}^{T} F_t = \widetilde{\lambda}$. However, this estimate is not valid if factors themselves are not excess returns or are not priced by the same model.

Notice also that if the limits of the normalized OVB and AB are non-zero, then these terms (together with $\widetilde{\lambda}_1$) asymptotically dominate the estimation. Below we consider three cases covered by Theorem 1. The first one is the case with no weak observed factors ($k_2 = 0$). In this case the theorem delivers the validity of the two-pass procedure, namely, the two-pass estimator is consistent and asymptotically mean-zero Gaussian. For the other two (more empirically relevant) cases – one with weak observed factors but no missing factors, the other with weak observed factors and missing strong factors – the two-pass procedure fails. The two-pass estimates of the risk premia on weak factors are inconsistent. The two-pass estimate of the risk premia on the strong observed factor is consistent, but has a bias which is of the same order of magnitude as its standard deviation. This invalidates all standard two-pass inferences in these two cases.

4.1 Case with no weak observed factors

Corollary 1 *Assume that the samples $\{r_{it}, i = 1, ..., N, t = 1, ..., T\}$ and $\{F_t, t = 1, ..., T\}$ come from a data-generating process that satisfies the factor-pricing model (2) and assumptions FACTORS, LOADINGS and ERRORS with $k_2 = 0$ (no weak observed factors). Then*

$$\sqrt{T}(\widehat{\lambda}_{TP} - \lambda) \Rightarrow \Gamma_{\beta\beta}^{-1}\Gamma_{\beta\mu}(\eta_v - \eta'\lambda) + \lim \sqrt{T}(\widetilde{\lambda} - \lambda),$$

where the limit of the right-hand-side is asymptotically gaussian with mean zero. If, in addition to that, there are no strong missing factors in errors (that is, $\mu_i = 0$), then

$$\sqrt{T}(\widehat{\lambda}_{TP} - \lambda) = \sqrt{T}(\widetilde{\lambda} - \lambda) + o_p(1) \Rightarrow N(0, \Omega_F).$$

This is a positive statement about the two-pass procedure, which claims that if all observed factors are strong, then the two-pass procedure is $\sqrt{T}$-consistent and provides asymptotically mean-zero gaussian estimate when both $N, T \to \infty$. If the error terms have a strong factor structure, it does not lead to a bias but may increase the asymptotic variance. If no strong factor structure is present in the error terms, then the two-pass procedure is asymptotically equivalent to the infeasible estimate $\widetilde{\lambda}$ and has asymptotic variance Ω_F.

4.2 Case with weak observed factors but no strong missing factors

Corollary 2 *Assume that the sample consisting of $\{r_{it}, i = 1, ..., N, t = 1, ..., T\}$ and $\{F_t, t = 1, ..., T\}$ comes from a data-generating process that satisfies the factor-pricing* *model (2) and assumptions FACTORS, LOADINGS and ERRORS with $k_2 \geq 1$ (there are weak observed factors) and $k_v = 0$ (no missing factor structure in errors). Then the following asymptotic statements hold simultaneously:*

$$\sqrt{T}(\widehat{\lambda}_{TP,1} - \lambda_1) = \sqrt{T}(\widetilde{\lambda}_1 - \lambda_1) + \sqrt{T}AB_1 + o_p(1),$$
$$\widehat{\lambda}_{TP,2} - \lambda_2 = AB_2 + o_p(1),$$

where

$$\begin{pmatrix} \sqrt{T}AB_1 \\ AB_2 \end{pmatrix} \to^p -\left(\Gamma + \mathcal{I}_{k_2}\Sigma_u\mathcal{I}_{k_2}\right)^{-1}\mathcal{I}_{k_2}\Sigma_u\lambda. \tag{4}$$

In the case when some of the observed factors have relatively small loadings (weak observed factors) the two-pass estimator will deviate from the classical case even if the idiosyncratic errors are not strongly correlated. The limit in (4) is non-random and is a non-zero vector, and thus characterizes the asymptotic bias. The two-pass estimate $\widehat{\lambda}_{TP,2}$ of the risk premia on weak factors $F_{t,2}$ is inconsistent and converges in probability to an incorrect value. The two-pass estimate $\widehat{\lambda}_{TP,1}$ of risk premia on strong factors $F_{t,1}$ is $\sqrt{T}$-consistent but this estimate has a bias of order $\frac{1}{\sqrt{T}}$, the same order of magnitude as the standard deviation of its asymptotically gaussian distribution. This leads to confidence sets being misplaced and standard inferences on the risk premia being invalid.

Intuition for the case with weak observed factors and no factor structure in the error terms. The result of Corollary 2 can be explained in terms of classical error-in-variables bias, or the so-called attenuation bias. Indeed, the first-pass estimate $\widehat{\beta}_i$ of risk exposure coefficients β_i contains estimation errors which are stochastically of order $O_p(1/\sqrt{T})$ each:

$$\widehat{\beta}_i = \left(\sum_{t=1}^{T}\widetilde{F}_t\widetilde{F}_t'\right)^{-1}\sum_{t=1}^{T}\widetilde{F}_t r_{it} = (\beta_i + u_i)(1 + o_p(1)),$$

where the $o_p(1)$ term is related to the difference between $\Sigma_F = E[(F_t - EF_t)(F_t - EF_t)']$ and $T^{-1}\sum_t \widetilde{F}_t\widetilde{F}_t'$. As a result, the second-pass regression encounters an error-in-variables problem. In the case of exposure to a strong observed factor, the estimation error in $\widehat{\beta}_{i,1}$ is asymptotically negligible compared to the size of the coefficient $\beta_{i,1}$ itself, and so this estimation error does not jeopardize consistency. However, the estimation error in $\widehat{\beta}_{i,2}$ is asymptotically of the same order of magnitude as the coefficient itself. The first-pass estimation errors in $\widehat{\beta}_{i,2}$ behave like a classical measurement error in the following sense: the imposed assumptions guarantee that the estimation errors $u_{i,2}$ for different assets are asymptotically uncorrelated and that they are asymptotically uncorrelated[2] with β_i themselves in the sense that the sample correlation between β_i and u_i is asymptotically negligible. The bias we observe in Corollary 2 is classic attenuation bias, with $\mathcal{I}_{k_2}\Sigma_u\mathcal{I}_{k_2}$ corresponding to the variance of the normalized measurement error $\sqrt{T}u_{i,2}$.

Note that if $\Gamma = \Gamma_{\beta\beta}$ is a block diagonal matrix with $\Gamma_{\beta_1\beta_2} = 0_{k_1,k_2}$, the two-pass procedure inferences about λ_1 will not be disturbed; namely, $\widehat{\lambda}_{TP,1}$ will be $\sqrt{T}$-consistent and when multiplied by $\sqrt{T}$ will have an asymptotically mean-zero Gaussian distribution. The block-diagonality assumption, though, is a very strong one: it requires that the values of $\beta_{i,1}$ be unrelated to the values of $\beta_{i,2}$ for the same asset, which is both implausible and not supported in applications. For example, the sample correlation coefficient between portfolios' betas that correspond to the market portfolio and betas that correspond to the SMB (HML) portfolio in the Fama–French dataset is equal to 0.73 (0.47).

4.3 Case with weak observed factors and strong missing factors in errors

Corollary 3 *Assume that the sample* $\{r_{it},\, i = 1, ..., N,\, t = 1, ..., T\}$ *and* $\{F_t,\, t = 1, ..., T\}$ *comes from a data-generating process that satisfies the factor-pricing model (2) and assumptions FACTORS, LOADINGS and ERRORS with* $k_2 \geq 1$ *(there are weak observed factors) and* $k_v \geq 1$ *(there is a missing factor structure in errors). Then the following asymptotic statements hold simultaneously:*

$$\sqrt{T}(\widehat{\lambda}_{TP,1} - \lambda_1) = \sqrt{T}(\widetilde{\lambda}_1 - \lambda_1) + \sqrt{T}AB_1 + \sqrt{T}OVB_1 + o_p(1),$$
$$\widehat{\lambda}_{TP,2} - \lambda_2 = AB_2 + OVB_2 + o_p(1),$$

where

$$\begin{pmatrix} \sqrt{T}OVB_1 \\ OVB_2 \end{pmatrix} \Rightarrow \left((I_{k_\beta};\widetilde{\eta})\Gamma(I_{k_\beta};\widetilde{\eta})' + \mathcal{I}_{k_2}\Sigma_u\mathcal{I}_{k_2}\right)^{-1}\left(\Gamma_{\beta\mu} + \widetilde{\eta}\Gamma_{\mu\mu}\right)(\eta_v - \eta'\lambda),$$
$$\begin{pmatrix} \sqrt{T}AB_1 \\ AB_2 \end{pmatrix} \Rightarrow -\left((I_{k_\beta};\widetilde{\eta})\Gamma(I_{k_\beta};\widetilde{\eta})' + \mathcal{I}_{k_2}\Sigma_u\mathcal{I}_{k_2}\right)^{-1}\mathcal{I}_{k_2}\Sigma_u\lambda.$$

The distributions of the right-hand-side expressions are non-gaussian and are not centered at zero.

This result covers a more general case which, as we argued before, is empirically quite relevant. Here some observed pricing factors may have relatively small loadings

[2]We use the statistical terms 'correlated' and 'uncorrelated' informally here, as the loadings are not formally random variables.

(weak factors), while errors are highly cross-sectionally correlated to the extent that they have strong missing factor structures. The two-pass estimate $\widehat{\lambda}_{TP,2}$ of the risk premia on weak factors $F_{t,2}$ is inconsistent and, asymptotically, has a poorly-centered non-standard distribution. The two-pass estimate $\widehat{\lambda}_{TP,1}$ of risk premia on strong factors $F_{t,1}$ is $\sqrt{T}$-consistent but this estimate has a bias of order $\frac{1}{\sqrt{T}}$ and an asymptotically non-standard distribution. This makes standard inferences (based on usual t-statistics) on the risk premia invalid.

Intuition for the case with factor structure in the error terms. In the presence of a strong factor structure in the errors, first-pass estimates have the following asymptotic representation:

$$\widehat{\beta}_i = \left(\sum_{t=1}^{T} \widetilde{F}_t \widetilde{F}_t'\right)^{-1} \sum_{t=1}^{T} \widetilde{F}_t r_{it} = \left(\beta_i + \frac{\eta_T \mu_i}{\sqrt{T}} + u_i\right)(1 + o_p(1)), \tag{5}$$

where $\eta_T = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \Sigma_F^{-1} \widetilde{F}_t v_t' \Rightarrow \eta$. Again, for the strong observed factors, the estimation error in $\widehat{\beta}_{i,1}$ turns to be asymptotically negligible when compared to the sizes of risk exposure $\beta_{i,1}$ themselves, while the estimation errors in $\widehat{\beta}_{i,2}$ – which are now equal to $\eta_T \mu_i/\sqrt{T} + u_i$ – are of the size $O_p(1/\sqrt{T})$, which is the same order of magnitude as the $\beta_{i,2}$'s themselves.

The estimation errors of $\widehat{\beta}_{i,2}$ distort the asymptotics and invalidate classical inferences. However, unlike the case covered by Corollary 2, the estimation errors in this setting do not behave like classical measurement errors in two respects. First, the estimation errors for different assets are correlated due to the presence of the common component η_T in all of them. Second, unless μ_i is cross-sectionally uncorrelated with β_i (so that $\Gamma_{\beta\mu} = 0_{k_F,k_v}$), the estimation error will be correlated with its own regressor β_i.

There is an additional issue worth noting with the two-pass procedure which is classically known as omitted variable bias. Let us look at the second step (normalized) 'ideal' regression which we can obtain by time-averaging equation (2):

$$\sqrt{T}\overline{r}_i = \sqrt{T}\widetilde{\lambda}'\beta_i + \eta_{v,T}'\mu_i + \sqrt{T}\overline{e}_i, \tag{6}$$

where $\eta_{v,T} = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} v_t \Rightarrow \eta_v \sim N(0_{k_v,1}, I_{k_v})$. Here we introduced normalization $\sqrt{T}$ to make regression (6) more conformable with the classical OLS setup. The regression error terms $\sqrt{T}\overline{e}_i$ all have orders of magnitude of $O_p(1)$, zero means and finite variances. Even though in finite samples $\sqrt{T}\overline{e}_i$ may be weakly cross-sectionally dependent, assumption ERRORS guarantees that they are asymptotically uncorrelated. Imagine for a moment that we know β_i and μ_i for all assets. Then, regression (6) will take the form of a classic OLS regression, with regressors $\sqrt{T}\beta_{i,2}$ and μ_i being of order of magnitude $O(1)$, in the sense expressed in assumption LOADINGS,[3] that in the classical regression setting would lead to a $\sqrt{N}$-consistent and asymptotically gaussian OLS estimator of coefficients on β_i's and μ_i's. The regressor $\sqrt{T}\beta_{i,1}$ is, in contrast, of order $O(\sqrt{T})$ and carries a lot of information, which in the classical regression setting leads to an OLS estimator of the coefficient λ_1 on this regressor that is super-consistent and asymptotically centered gaussian. However, because μ_i is unobserved, it becomes a part of the error term in the second-pass regression, making error terms cross-sectionally correlated; see, for example, Andrews (2005) for a similar phenomenon. A more classical reference for this phenomenon is an omitted variable bias – if $\Gamma_{\beta\mu} \neq 0_{k_F,k_v}$, then even if there were no first-stage estimation error and we knew β_i, running an OLS in a regression of $\sqrt{T}\overline{r}_i$ on $\sqrt{T}\beta_i$ would produce invalid results due to the omission of μ_i.

One question that may arise is whether or not the omitted variable bias is large. The answer to this question is closely related to the size of the cross-sectional correlation between β_i and μ_i as expressed in $\Gamma_{\beta\mu}$. Unfortunately, there is no reliable empirical evidence on this, as μ_i is unobserved and β_i is poorly estimated and biased in the direction of μ_i (see equation (5)). The problem with estimation of μ_i is that the estimator $\widehat{\lambda}_{TP,2}$ is inconsistent, which makes the residuals from the two-pass procedure poor indicators of the true errors, and estimating μ_i via the principle components analysis on the residuals does not produce good estimates. However, even though direct empirical evidence on this matter is absent, we have two indirect arguments which suggest that one should expect a high rather than low correlation between β_i and μ_i. One argument is the empirical observation that for many well-known factor-pricing models the estimated betas for different factors are exceptionally highly correlated. Another argument is related to our theoretical example 1, where the missing factor structure originates as a result of mismeasuring the true risk factor, and the sample correlation between β_i and μ_i equals 1.

5 Newly proposed estimator

5.1 Idea of the proposed solution

The case with no factor structure in the error terms. We begin by solving the easier case when no unobserved factor structure is present in the errors, while some

[3] This is similar to the assumption from classical regression with fixed regressors that $N^{-1}\sum_{i=1}^{N} x_i x_i' \to A_x$, where x_i is i^{th} observation for the regressor, and A_x is a full rank finite matrix.

observed factors are weak. As we discussed before, in such a case the failure of the two-pass procedure can be labeled a classical measurement error-in-variables problem, which is often solved by finding a proper instrument. Apparently, it is relatively easy to find a valid instrument in our setting if one is willing to employ a sample-splitting technique.

Let us divide the set of time indexes $t = 1, ..., T$ into two non-intersecting equal subsets T_1 and T_2. It is more natural to make T_1 the first half of the sample, and T_2 its second half; then the procedure will have greater robustness as we discuss below. Let us run the first step regression twice – separately on each sub-sample:

$$\widehat{\beta}_i^{(j)} = \left(\sum_{t \in T_j} \widetilde{F}_t^{(j)} \widetilde{F}_t^{(j)\prime}\right)^{-1} \sum_{t \in T_j} \widetilde{F}_t^{(j)} r_{it} = (\beta_i + u_i^{(j)})(1 + o_p(1)), \text{ for } j = 1, 2,$$

where $\widetilde{F}_t^{(j)} = F_t - \frac{1}{|T_j|} \sum_{t \in T_j} F_t$, $u_i^{(j)} = \frac{1}{|T_j|} \sum_{t \in T_j} \Sigma_F^{-1} \widetilde{F}_t^{(j)} e_{it}$, and the $o_p(1)$ term is related to the difference between Σ_F and $\frac{1}{|T_j|} \sum_{t \in T_j} \widetilde{F}_t^{(j)} \widetilde{F}_t^{(j)\prime}$.

The assumption ERRORS guarantees that the two sets of estimation uncertainty, $\{u_i^{(1)},\ i = 1, ..., N\}$ and $\{u_i^{(2)},\ i = 1, ..., N\}$, are independent conditionally on $\mathcal{F}$. In fact, the asymptotic independence of the two sets of errors will hold more generally if one makes stationarity assumptions and controls the decay of time-series dependence in errors e_{it}, and the sub-samples are formed to be first and second halves of the sample, correspondingly.

Given the observation about independence of estimation errors obtained from different sub-samples, one may use an estimate of β_i from one sub-sample (for example, $\widehat{\beta}_i^{(1)}$), as a regressor while the other (in this example, $\widehat{\beta}_i^{(2)}$) as an instrument. This would represent a valid IV regression. Indeed, the second step regression we run is:

$$\overline{r}_i = \widetilde{\lambda}' \widehat{\beta}_i^{(1)} + (\overline{e}_i - \widetilde{\lambda}' u_i^{(1)}).$$

In this regression the regressor and the instrument are correlated since they both contain β_i, hence we have a relevance condition. The validity condition holds for two reasons: (i) the part of the second-step regression error $u_i^{(1)}$ is asymptotically uncorrelated with the instrument $\widehat{\beta}_i^{(2)}$; (ii) assumption ERRORS guarantees that $\widehat{\beta}_i^{(2)}$ is asymptotically uncorrelated with $\overline{e}_i$. As we show below, this procedure restores consistency and standard inferences on the estimates of risk premia.

Similar ideas, such as sample splitting and jackknife-type estimators, have been previously employed in the literature on many weak instruments (e.g., Hansen, Hausman and Newey, 2008). In that literature the term "many instruments" is related to modeling the number of instruments as growing to infinity proportionally (though not always) to the sample size (so-called dimensionality asymptotics), while the term "weak" appears due to a modeling assumption that makes the estimation error of the reduced-form coefficients be of the same order of magnitude as the coefficients themselves (so called local-to-zero asymptotics). This is parallel to the dimensionality asymptotics for a number of portfolios and the local-to-zero asymptotics for risk exposures of weak factors in our setup. In the many-weak-instruments setting, the regular TSLS estimator has a significant bias, and classical inferences are asymptotically invalid. That problem can also be interpreted as a classical measurement error-in-variables problem for the second stage regression, where the regression is run on the fitted values from the first-stage projection of the original regressor on the instruments. Some proposed solutions employ the second-stage instrumental variables regression where, for each observation, the regressor is obtained from a first-stage regression run on a sub-sample that does not include that observation, and the original instrument is still used as an instrument (Angrist, Imbens and Krueger, 1999). This makes the first-stage error in the projection uncorrelated with the instrument for this specific observation. Consistency and classical inferences are restored by sample-splitting or leave-one-out type procedures.

The case with factor structure in the error terms. As we discussed before, the situation in the model with unobserved factor structure has an additional problem that can be described as the presence of omitted (and unobserved) variable μ_i in regression (6). However, after examining formula (5) for the first-pass estimate we may notice that we can obtain a noisy proxy for μ_i if we take the difference between two estimates for the same β_i obtained from different sub-samples. Indeed, consider two non-intersecting subsets of time indexes, T_1 and T_2 and assume they have the same number, say τ, of time indexes. Then

$$\widehat{\beta}_i^{(1)} - \widehat{\beta}_i^{(2)} = \frac{\eta_\tau^{(1)} - \eta_\tau^{(2)}}{\sqrt{\tau}}\mu_i + (u_i^{(1)} - u_i^{(2)}).$$

Notice that both the coefficient on μ_i and the noise term $u_i^{(1)} - u_i^{(2)}$ are of the same order of magnitude $O_p(1/\sqrt{\tau})$. This neither means that the signal dominates the noise, and thus we need a correction to account for the noise, nor does the noise dominate the signal, and thus the proxy is not useless.

Assume that $k_v \leq k_F$, which implies that we have a larger number of proxies than needed and we have a choice among them. Now we assume that we have a fixed and full-rank $k_v \times k_F$ matrix A, and use $A(\widehat{\beta}_i^{(1)} - \widehat{\beta}_i^{(2)})$ as the proxy. It is worth noting that

$$\mu_i = \sqrt{\tau}\left(A(\eta_\tau^{(1)} - \eta_\tau^{(2)})\right)^{-1} A(\widehat{\beta}_i^{(1)} - \widehat{\beta}_i^{(2)}) - \sqrt{\tau}\left(A(\eta_\tau^{(1)} - \eta_\tau^{(2)})\right)^{-1} A(u_i^{(1)} - u_i^{(2)}),$$

where $\eta_\tau^{(1)}$ and $\eta_\tau^{(2)}$ are asymptotically independent non-degenerate $k_F \times k_v$ gaussian (assuming that the size of sub-samples τ increases to infinity with T); the $k_v \times k_v$ matrix $A(\eta_\tau^{(1)} - \eta_\tau^{(2)})$ will be invertible with probability 1.

The idea is to regress the average return $\overline{r}_i$ on $\widehat{\beta}_i^{(1)}$ and $A(\widehat{\beta}_i^{(1)} - \widehat{\beta}_i^{(2)})$ instead of on unobserved β_i and μ_i. This solves the omitted variables part of the problem, but the error-in-variables issue still remains. That problem we solve via instrumental variables upon additional sample splitting. The ultimate idea goes as follows: split the sample into four equal sub-samples along the time dimension; calculate the first-pass estimates of risk exposures for all four sub-samples; run an instrumental variables regression using $\widehat{\beta}_i^{(1)}$ and $A(\widehat{\beta}_i^{(1)} - \widehat{\beta}_i^{(2)})$ as regressors and $\widehat{\beta}_i^{(3)}$ and $(\widehat{\beta}_i^{(3)} - \widehat{\beta}_i^{(4)})$ as instruments.

5.2 Algorithm for constructing a 4-split estimator

Let us divide the set of time indexes into four equal non-intersecting subsets T_j, $j = 1, ..., 4$.

(1) For each asset i and each subset j run a time-series regression to estimate the coefficients of risk exposure:

$$\widehat{\beta}_i^{(j)} = \left(\sum_{t \in T_j} \widetilde{F}_t^{(j)} \widetilde{F}_t^{(j)\prime}\right)^{-1} \sum_{t \in T_j} \widetilde{F}_t^{(j)} r_{it}.$$

(2) Run an IV regression of $\overline{r}_i = \frac{1}{T}\sum_{t=1}^T r_{it}$ on regressors $x_i^{(1)} = \left(\widehat{\beta}_i^{(1)\prime}, (\widehat{\beta}_i^{(1)} - \widehat{\beta}_i^{(2)})' A_1'\right)'$ with instruments $z_i^{(1)} = \left(\widehat{\beta}_i^{(3)\prime}, (\widehat{\beta}_i^{(3)} - \widehat{\beta}_i^{(4)})'\right)'$, where A_1 is a non-random $k_v \times k_F$ matrix of rank k_F. Denote the TSLS estimate of the coefficient on regressor $\widehat{\beta}_i^{(1)}$ by $\widehat{\lambda}^{(1)}$.

(3) Repeat step (2) three more times exchanging indexes 1 to 4 circularly; that is, the first repetition is an IV regression of $\overline{r}_i$ on regressors $x_i^{(2)} = \left(\widehat{\beta}_i^{(2)\prime}, (\widehat{\beta}_i^{(2)} - \widehat{\beta}_i^{(3)})' A_2'\right)'$ with instruments $z_i^{(2)} = \left(\widehat{\beta}_i^{(4)\prime}, (\widehat{\beta}_i^{(4)} - \widehat{\beta}_i^{(1)})'\right)'$; denote the corresponding estimate by $\widehat{\lambda}^{(2)}$, etc.

(4) Obtain the 4-split estimate as $\widehat{\lambda}_{4S} = \frac{1}{4}\sum_{j=1}^4 \widehat{\lambda}^{(j)}$.

(5) In order to compute an estimate of the covariance matrix for $\widehat{\lambda}_{4S}$, denote by $X^{(j)}$ the $N \times k$ matrix of stacked regressors used in the IV regression where $\widehat{\lambda}^{(j)}$ was obtained, and by $Z^{(j)}$ the $N \times k_z$ matrix of instruments from this regression (here $k_z = 2k_F$ is the number of instruments in a single regression, and $k = k_F + k_v$ is the number of regressors). Calculate

$$G = \begin{pmatrix} G_1 & 0_{k,k} & 0_{k,k} & 0_{k,k} \\ 0_{k,k} & G_2 & 0_{k,k} & 0_{k,k} \\ 0_{k,k} & 0_{k,k} & G_3 & 0_{k,k} \\ 0_{k,k} & 0_{k,k} & 0_{k,k} & G_4 \end{pmatrix}, \quad \text{where } G_j = \frac{1}{N} X^{(j)\prime} P_{Z^{(j)}} X^{(j)},$$

and $P_Z = Z\left(Z'Z\right)^{-1} Z'$. Also calculate

$$\widehat{\Sigma}_0 = \frac{1}{N} \sum_{i=1}^{N} \begin{pmatrix} \widetilde{z}_i^{(1)} \widehat{\epsilon}_i^{(1)} \\ \dots \\ \widetilde{z}_i^{(4)} \widehat{\epsilon}_i^{(4)} \end{pmatrix} \begin{pmatrix} \widetilde{z}_i^{(1)} \widehat{\epsilon}_i^{(1)} \\ \dots \\ \widetilde{z}_i^{(4)} \widehat{\epsilon}_i^{(4)} \end{pmatrix}',$$

where $\widehat{\epsilon}_i^{(j)}$ is i^{th} residual from the IV regression where $\widehat{\lambda}^{(j)}$ was obtained, and $\widetilde{z}_i^{(j)} = X^{(j)\prime} Z^{(j)} (Z^{(j)\prime} Z^{(j)})^{-1} z_i^{(j)}$. Also denote $R = (1,1,1,1)' \otimes \begin{pmatrix} \frac{1}{4} I_{k_F} \\ 0_{k_v,k_F} \end{pmatrix}$, which is a $4k \times k_F$ matrix. Finally,

$$\widehat{\Sigma}_{4S} = \frac{1}{N} R' G^{-1} \widehat{\Sigma}_0 G^{-1} R + \frac{1}{T} \widehat{\Omega}_F,$$

where $\widehat{\Omega}_F$ is a consistent estimator of the long-run variance of F_t.

5.3 Consistency of the 4-split estimator

Theorem 2 *Assume that the samples $\{r_{it},\, i = 1, ..., N,\, t = 1, ...T\}$ and $\{F_t,\, t = 1, ..., T\}$ come from a data-generating process that satisfies the factor pricing model (2) and assumptions FACTORS, LOADINGS and ERRORS. Let both N and T increase to infinity then*

$$\sqrt{T}(\widehat{\lambda}_{4S,1} - \lambda_1) = \sqrt{T}(\widetilde{\lambda}_1 - \lambda_1) + O_p(1/\sqrt{N}) \Rightarrow N(0, \Omega_F),$$

and

$$\sqrt{\min\{N,T\}}(\widehat{\lambda}_{4S,2} - \lambda_2) = O_p(1).$$

Discussion. Theorem 2 establishes the speed of consistency for the new 4-split estimator $\widehat{\lambda}_{4S}$ under exactly same assumptions used to show failure of the two-pass estimation procedure. The 4-split estimator for the risk premia on the strong observed factor is

$\sqrt{T}-$ consistent, asymptotically equivalent to $\widetilde{\lambda}_1$ and asymptotically Gaussian, while the 4-split estimate of the risk premia on the weak observed factor is consistent, and the speed of convergence depends on the relative size of N and T. Theorem 2 shows that the 4-split estimator has superior asymptotic properties in comparison to the classical two-pass procedure for the risk premia.

6 Inference procedures using 4-split estimator

Theorem 2 shows that the new 4-split estimator is consistent but does not provide a basis for statistical inference, namely, for confidence set construction or testing. In order to use Theorem 2 the researcher has to know which observed factors are strong, and with that knowledge s/he can construct a confidence set for the risk premia on the strong observed factor only. However, in general there is no a pre-test that successfully discriminates between weak and strong observed factors. The other drawback of Theorem 2 is that it does not provide an asymptotic distribution for the estimator of the risk premia on a weak observed factor. Apparently, the stated assumptions are not enough to obtain the asymptotic distribution of the full 4-split estimator. We additionally need assumptions that will guarantee the validity of some Centra Limit Theorems. Below we formulate the needed high-level assumptions and establish a result about statistical inferences using the 4-split estimator. We also provide primitive assumptions that will guarantee that our high-level assumptions will hold in examples and discuss how one can obtain the needed Central Limit Theorems.

For a set of vectors a_j we denote $(a_j)_{j=1}^4 = (a_1', ..., a_4')'$ as a long vector consisting of the four vectors stacked upon each other, similarly for vectors a_{jj^*} we denote $(a_{jj^*})_{j<j^*} = (a_{12}', a_{13}', a_{14}', a_{23}', a_{24}', a_{34}')'$.

Assumption GAUSSIANITY Assume that the following convergence holds:

$$\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \begin{pmatrix} \sqrt{T}\gamma_i \bar{e}_i \\ (\sqrt{T}\gamma_i u_i^{(j)})_{j=1}^4 \\ (T\bar{e}_i u_i^{(j)})_{j=1}^4 \\ (T u_i^{(j)} u_i^{(j^*)})_{j<j^*} \end{pmatrix} = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \xi_i \Rightarrow \xi = \begin{pmatrix} \xi_{\gamma e} \\ (\xi_{\gamma j})_{j=1}^4 \\ (\xi_{ej})_{j=1}^4 \\ (\xi_{j,j^*})_{j<j^*} \end{pmatrix},$$

where ξ is a Gaussian vector with mean zero and covariance Σ_ξ and assume that

$$\frac{1}{N} \sum_{i=1}^{N} \xi_i \xi_i' \to^p \Sigma_\xi.$$

Theorem 3 *Assume that the samples* $\{r_{it},\, i = 1, ..., N,\, t = 1, ...T\}$ *and* $\{F_t,\, t = 1, ..., T\}$ *come from a data-generating process satisfying factor pricing model (2) and assumptions FACTORS, LOADINGS, ERRORS and GAUSSIANITY as both* N *and* T *increase to infinity. Then*

$$\widehat{\Sigma}_{4S}^{-1/2}(\widehat{\lambda}_{4S} - \lambda) \Rightarrow N(0, I_k).$$

Theorem 3 suggests the use of t and Wald statistics for the construction of confidence sets for the risk premia as well as for testing hypotheses about values of the risk premia. These inference procedures are very standard ones and can be performed using standard econometric software.

From a theoretical perspective, however, the asymptotics of the 4-split estimator are not fully standard. Technically, the asymptotic distribution of the 4-split estimator is not Gaussian but rather mixed Gaussian. This is due to the fact that the limit distribution for this estimator can be written as a Gaussian random vector with random variance. To understand the intuition for why one gets an asymptotically-random covariance matrix one can look at equation (6) and notice that the coefficient $\eta_{v,T}$ on the omitted variable μ_i is random, even asymptotically. This leads to a phenomenon where the amount of information contained in the sample that is used to correct the omitted-variable problem will be random as well, and thus, we have an asymptotically-random covariance matrix. Theorem 3 shows that a properly constructed proxy for the asymptotic variance restores the asymptotic gaussianity of a multidimensional t-statistic even when the estimator itself is not asymptotically gaussian.

Another important aspect of Theorem 3 is that inferences or construction of the proxy for variance do not assume knowledge of the number or identity of strong/weak factors. This is a desirable feature, as we do not have a procedure that can credibly differentiate between weak and strong factors.

As previously discussed, even though the main data set contains NT observations, the risk premia cannot be estimated at a rate better than $\sqrt{T}$. This can be seen from equation (3) as even if we know the true values of β_i the regression of $\overline{r}_i$ on β_i has a true coefficient equal to $\widetilde{\lambda} = \lambda + \overline{F} - EF_t$. This means that the uncertainty associated with the deviations of $\frac{1}{T}\sum_{t=1}^{T} F_t$ from EF_t is unavoidable. This also justifies the presence of the long-run variance of factors, Ω_F, in the formula for variance $\widehat{\Sigma}_{4S}$. Theorem 2 also states that the difference between $\widehat{\lambda}_{4S}$ and $\widetilde{\lambda}$ is of order $\frac{1}{\sqrt{NT}}$. From the proof of Theorem 3 we see that this difference is mixed Gaussian, and the variance can be deduced from $\widehat{\Sigma}_{IV}$.

Typically $\widetilde{\lambda}$ is infeasible. However, if all observed factors are portfolios themselves and are priced by the same model, then we have $\lambda = EF_t$. In such a case the literature suggests the use of alternative feasible estimator $\widehat{\lambda} = \frac{1}{T}\sum_{t=1}^{T} F_t$, which in this case is equal to $\widetilde{\lambda}$. Thus, in this special case we have two competing estimators for λ and can create a test for model specification. In particular, the statistic compares the difference between $\widehat{\lambda}_{4S}$ and $\widetilde{\lambda}$ to zero. The proof of Theorem 3 shows that $\widehat{\lambda}_{4S} - \widetilde{\lambda}$ converges to zero at speed $\frac{1}{\sqrt{NT}}$, is asymptotically mixed Gaussian and $\widehat{\Sigma}_{IV}$ is the proper proxy for the variance that delivers a χ^2- asymptotic distribution to the corresponding Wald statistic. The power properties of such a test are a topic for future research.

6.1 Discussion of Assumption GAUSSIANITY and Central Limit Theorems

Here we provide some sufficient conditions for the validity of Assumption GAUSSIANITY.

Lemma 2 *Assume we have a setting as in Example 3 with additional assumptions that the time series* (F_tF_t', F_t) *is stationary, has summable covariances,* $\max_{i,t} E\eta_{it}^8 < C$, $E(\|F_t\|^8 + 1)(\|w_t\|^8 + 1) < C$, $N^{-1}\sum_{i=1}^{N}\sigma_i^4 \to \sigma_4$ *and* $N^{-1}\sum_{i=1}^{N}\sigma_i^2\gamma_i\gamma_i' \to \Gamma_\sigma$. *Then Assumption GAUSSIANITY holds.*

FIXME: We have to write up the proof of consistency of the variance estimation.

Lemma 3 *Assume that Assumption ERROR* holds as do additional assumptions that the time series* (F_tF_t', F_t) *is stationary, has summable covariances,* $\max_{i,t} Ee_{it}^8 < C$, *and* $N^{-1}\gamma'\mathcal{E}_T\gamma \to \Gamma_\sigma$. *Then Assumption GAUSSIANITY holds.*

FIXME: proof of this result is still missing. Likely we will need an additional assumption that diagonal of $\mathcal{E}_T$ dominates the sum of off-diagonal elements.

So, in our case Assumption GAUSSIANITY results from strengthening moment restrictions on top of already imposed assumptions. However, from a theoretical perspective the derivation of a proper Central Limit Theorem is a major endeavor. The difficulty here is that the assumptions accommodate the quite unrestrictive structure of the cross-sectional dependence of idiosyncratic error terms e_{it} by merely restricting the amount of cross-sectional correlation. This makes ξ_i (introduced in Assumption GAUSSIANITY) cross-sectionally dependent, though the correlation between ξ_i and ξ_{i^*} for $i \neq i^*$ converges to zero for large sample sizes. Without cross-sectional dependence with the proper structure, it is hard to obtain a Central Limit Theorem. We follow a different route and exploit a time-series conditional independence of errors instead.

First, consider the following components of ξ_i:

$$\frac{\sqrt{T}}{\sqrt{N}}\sum_{i=1}^{N}\gamma_i\overline{e}_i = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\gamma_i e_{it} = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{\gamma' e_t}{\sqrt{N}};$$

$$\frac{\sqrt{T}}{\sqrt{N}}\sum_{i=1}^{N}\gamma_i \otimes u_i^{(j)} = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\frac{\gamma' e_t}{\sqrt{N}}\right)\otimes\left(\Sigma_F^{-1}\widetilde{F}_t^{(j)}\right)\mathbb{I}\{t\in T_j\}.$$

Here we changed the order of summation. By collecting all terms of interest into the vector we obtain the following expression for a part of $\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\xi_i$:

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{\gamma' e_t}{\sqrt{N}}\otimes\begin{pmatrix}1\\ \left\{\left(\Sigma_F^{-1}\widetilde{F}_t^{(j)}\right)\mathbb{I}\{t\in T_j\}\right\}_{j=1}^{4}\end{pmatrix},$$

where the summands represent a martingale-difference sequence, and a proper Central Limit Theorem could be used.

Now assume that $j^* > j$ and consider the following sum:

$$vec\left(\frac{T}{\sqrt{N}}\sum_{i=1}^{N}u_i^{(j^*)}u_i^{(j)\prime}\right) = \frac{1}{T\sqrt{N}}\sum_{i=1}^{N}\sum_{t\in T_j}\sum_{s\in T_{j^*}} vec\left(\Sigma_F^{-1}\widetilde{F}_s^{(j^*)}\widetilde{F}_t^{(j)\prime}\Sigma_F^{-1}e_{it}e_{is}\right)$$

$$= \sum_{s\in T_{j^*}}\sum_{t\in T_j}\frac{1}{T}\left(\Sigma_F^{-1}\widetilde{F}_s^{(j^*)}\right)\otimes\left(\Sigma_F^{-1}\widetilde{F}_t^{(j)}\right)\frac{e_t' e_s}{\sqrt{N}}.$$

Similarly consider the following sum:

$$\frac{T}{\sqrt{N}}\sum_{i=1}^{N}\overline{e}_i u_i^{(j)} = \frac{1}{T\sqrt{N}}\sum_{i=1}^{N}\sum_{t_1=1}^{T}\sum_{t_2\in T_j}\Sigma_F^{-1}\widetilde{F}_{t_2}^{(j)}e_{it_1}e_{it_2}$$

$$= \frac{1}{T\sqrt{N}}\sum_{i=1}^{N}\sum_{t\in T_j}\Sigma_F^{-1}\widetilde{F}_t^{(j)}e_{it}^2 + \frac{1}{T}\sum_{t_1=1}^{T}\sum_{t_2\in T_j, t_1\neq t_2}\Sigma_F^{-1}\widetilde{F}_{t_2}^{(j)}\frac{e_{t_1}' e_{t_2}}{\sqrt{N}}.$$

Assumption ERRORS (iii) guarantees that $\left(T\sqrt{N}\right)^{-1}\sum_{i=1}^{N}\sum_{t\in T_j}\Sigma_F^{-1}\widetilde{F}_t^{(j)}e_{it}^2 = o_p(1)$. Thus, we are only interested in gaussianity of the second sum.

In both cases we are interested in making use of a Central Limit Theorem for quadratic forms of $\frac{e_t' e_s}{\sqrt{N}}$ with some random coefficients depending on $\{F_t\}$. That is, the summation goes both over t and s. The idea for deriving such statements comes from the Central Limit Theorem for quadratic forms established by de Jong (1987) who refers to a martingale Central Limit Theorem by Heyde and Brown (1970). We follow a similar path.

6.2 Open question: weak identification

(I am not sure if we need this as a subsection or just a note at the end of the previous as we do not have all the answers yet)

Statement: this is not weak identification.

if betas are even smaller then weak identification.

we'll address this in later research?

7 Simulations

FIXME: this section is to be fully re-done

In this Section, we verify how the asymptotic theory works for finite samples.

Simulation design The simulation design takes as a starting point the 3-factor Fama–French model estimated on $N_0 = 25$ Fama–French portfolios observed during $T_0 = 209$ quarters. We generate the artificial data on factors and excess returns of the size $N = mN_0$, $T = mT_0$, i.e. m times larger than both original dimensions. The 3×1 vectors of N betas are generated independently from a normal distribution with the mean and variance obtained as corresponding estimates from the original sample on the first pass of the two-pass procedure:

$$(m_\beta, v_\beta) \simeq \left(\begin{bmatrix} 0.96 \\ 0.53 \\ 0.19 \end{bmatrix}, \begin{bmatrix} 0.110 & 0.060 & 0.020 \\ & 0.061 & 0.008 \\ & & 0.016 \end{bmatrix} \right).$$

The 3×1 vectors of T factors are generated independently from a normal distribution with the mean and variance obtained as empirical values obtained from the original sample:

$$(m_f, v_f) \simeq \left(\begin{bmatrix} 1.59 \\ 0.89 \\ 0.85 \end{bmatrix}, \begin{bmatrix} 74.7 & 24.9 & -9.2 \\ & 33.5 & 0.3 \\ & & 40.0 \end{bmatrix} \right).$$

The true factor prices λ corresponding to the Market, SMB and HML factors are set to values obtained from the second pass of the two-pass procedure: $\lambda \simeq (2.70, 0.69, 1.96)'$. The idiosyncratic errors are generated from a mean zero normal distribution independently over time and across the assets; the diagonal variance matrix contains the $N \times N$ residual variance matrix replicated m times.

In the baseline design described above, there is no cross-sectional dependence in the errors. To induce cross-sectional dependence with controllable strength, we generate the missing factors that are correlated with the observed factor j ($j = 1, 2, 3$) according to

$$\mu = m_\mu + \rho \frac{v_\mu^{1/2}}{v_{\beta_j}^{1/2}} \left(\beta_j - m_{\beta_j} \right),$$

where ρ is intended correlation coefficient, and m_μ and v_μ are the mean and variance of the loadings corresponding to the first principal component of the residuals from the the first pass of the two-pass procedure applied to the original data. If there are missing factors, the variance of the generated idiosyncratic errors is decreased by the sample variance corresponding to the first principal component of the residuals.

Then the total errors are generated as

$$\varepsilon = p \cdot v \mu' + e,$$

where p is a coefficient of intended inflation of strength of the missing factor, and v is a missing factor vector of IID normals with mean zero and variance corresponding to the variance accountable by the first principal component of the residuals.

Finally, given the he $N \times 1$ vector of returns at time t is generated as

$$r_t = \lambda' \beta + (F_t - EF_t)\beta + \varepsilon_t.$$

Results Tables 1a–1d contain simulation results; Table 1a presents biases of estimates (larger figures) and their standard deviations (smaller figures) from estimation by the conventional two-pass procedure and two variations of the proposed 'average' four-split procedure. One variation uses only the second element of the estimated beta from the first and second subsamples to form the proxy for the missing factor (designated as 'four-split$_2$'), i.e. $A = (0, 1, 0)$; the other variation averages across the three elements, i.e. $A = \frac{1}{3}(1, 1, 1)$. We set up pretty severe correlation between the missing factor and the included (strongest) 'Market' factor (the figures resulting when the missing is correlated with the second strongest 'SMB' factor are very similar), and a quite high inflation coefficient $p = 5$. The four parts of the table correspond to increasing dimension of data $m = 1, 2, 4, 8$, and simultaneous reduction by $\sqrt{m}$ the generated betas corresponding to the weakest 'HML' factor (which means that the average *squared* beta is reduced by m).

When $m = 1$, i.e. the generated samples are of the original size, the two-pass risk premium estimate of the weakest factor, 'HML', is most problematic, and exhibits both high (and negative) bias (-0.37 which amounts for almost 20% of the true value) and high variance. Both the bias and variance are partially offset (by about 30%) from using the four-split$_2$ estimate, and both are further reduced, the bias by a half, when one instead uses averaging the three differences in estimated betas to construct the proxy. When m increases, i.e. the sample size goes up with simultaneous weakening of the 'HML' risk exposures, the bias problem persists and even slightly exacerbates, while the variance slowly goes down. The four-split estimate dominates the conventional two-pass and four-split$_2$ estimates, so from this point on we report evidence only on the averaging four-split variation. Table 1b shows 95% coverage rates of the confidence intervals; the conventional two-pass estimator exploits the Shanken (1992) corrected standard errors. Generally, the four-split estimator has an advantage over the conventional two-pass estimator, though the problem for the weakest factor persists and worsens with the weakness degree deteriorates, and it is not compensated by an increase in the sample size.

In next experiment whose results are reported in Tables 1c and 1d, we further strengthen the cross-sectional dependence in the error term and set $p = 10$. We also further enlarge the sample by setting m to higher powers of two: to 4, 8, 16 and 32, while keeping the weakness of the 'HML' factor at the weakest level of the previous experiment (i.e. corresponding to the factor of 8). As we drive m from 4 to 32 (i.e. 8-fold), both the bias of the HML risk premium estimate from the conventional two-pass procedure diminishes by the factor of three, while the bias of the estimator from the four-split procedure vanishes much faster as the sample size increases, and the ratio between the two increases from 3-fold to 10-fold. The standard deviation of both estimators goes down by about the factor of three as m goes up by 8, while the 95% coverage rates are kept pretty stable along the way. The coverage rates from the four-split procedure are clearly much closer to the nominal one, though still there is a significant gap.

Figure 1 depicts the kernel estimates of the SMB and HML risk premium estimates from the two procedures in one of experiments. One can clearly notice a higher bias and higher variance of the conventional two-pass estimates, as well as more symmetric distributions of the estimates from the four-split procedure for this design. The distributions of four-split estimates are clearly very close to gaussian.

Summarizing, the four-split procedure has definite advantages when compared to the conventional two-pass procedure, in terms of mean bias, dispersion, and coverage rates, when both weak priced factors are included and strong missing factors remain in the errors. However, it still has problems related to the bias and, as a result, to confidence interval coverage. These problems may be attributed to possible weak identification at the second pass of the four-split procedure.

8 Revision of Empirical Applications

FIXME: this section is to be fully re-done

Setup We run the conventional two-pass procedure and proposed 'average four-split' two-pass procedure on a few well-known databases using some classical and extended factors such as the market portfolio excess return ('Market'), size factor (small-minus-big return, 'SMB'), book-to-market factor (high-minus-low return, 'HML') (Fama and French, 1993), consumption-to-wealth ratio (cay) (Lettau and Ludvigson, 2001) and momentum factor (MOM) (Jegadeesh and Titman, 1993). The first three exercises (no. 1–3) use the quarterly data on $N = 25$ Fama–French portfolios from 1963Q3 to 2015Q3 ($T = 209$)[4]; the next two exercises (no. 4–5) use the monthly data on $N = 44$ industry portfolios from 1963:07 to 1990:12 ($T = 546$).[5]

Results Tables 2a–2c contain results of estimation by the conventional two-pass procedure and proposed 'average' four-split procedure, together with some auxiliary output. Table 2 reports average estimated betas and beta squared from the first pass of the conventional two-pass procedure. These figures provide information of how strong or weak the factors being used are. Table 2b reports variance fractions corresponding to five main principal components in the residuals from the first pass of the conventional two-pass procedure. These figures give a feel of the strength of cross-sectional correlation in the errors: according to Onatski (2012), under the dimensionality asymptotics ($N/T \to const$), the eigenvalues corresponding to strong factors are asymptotically isolated and separated from the eigenvalues corresponding to weak factors, the latter ones clustering together. Finally and most importantly, Table 2c reports the risk premia estimates and standard errors by the conventional and proposed procedures.

Experiment 1 is the cleanest case with the three classical factors from Fama and French (1993). All three factors seem to be strong, though the HML factor may raise slight doubt. However, the factor structure of the residuals does not seem to have strong components. Indeed, the penalized estimation methods of Bryzgalova (2015) do not classify the HML factor as useless or weak. The two estimation procedures produce qualitatively similar (though a bit different) estimates of risk premia.

[4]The data on returns and factors are taken from Kenneth French's and Martin Lettau's webpages.

[5]The data on returns are taken from Kenneth French's webpage. The data on factors are taken from Gagliardini, Ossola and Scaillet (2016)'s supplementary materials.

In experiment 2, we exclude the strong factor SMB and expect that this may induce presence of a strong factor in the residuals, and indeed, the largest eigenvalue is well separated from the others. The two estimation procedures produce sharply different estimates of the risk premium for the HML factor. When we add in experiment 3 to the remaining two strong return factors a macroeconomic factor, cay, which is strong but barely priced (see the tiny estimate of its risk premium), the cross-sectional structure of residuals stays the same. The additional factor cay is little priced from the viewpoint of both estimation procedures, and there is less disagreement among the two procedures as far as the other factors are concerned.

Experiment 4 repeats experiment 1 for the industry portfolios. All three factors seem to be strong, and the residual variance seems to be clean of common factors; the estimation procedures tend to agree on values of risk premia though the proposed procedure delivers higher (in absolute value) estimates with a bit higher standard errors. In experiment 5, we add a clearly weak momentum factor, which does not distort the (absence of) factor structure of residuals. As a result, the estimates of risk premia change little, with the proposed procedure being sharper on the SMB factor and the conventional procedure on the momentum factor. While the standard errors are higher with the proposed procedure, the discrepancy in estimation efficiency is small if one takes the width of a confidence interval as an indicator of efficiency.

9 Appendix A with proofs

Proof that Example 3 satisfies Assumptions ERRORS. Assumption ERRORS(i) follows from w_t being serially uncorrelated conditionally on $\mathcal{F}$, and time series independence of η_{it}. For Assumption ERRORS(ii), note that for $t \neq s$,

$$\rho(t,s) = \sum_{i=1}^{N} \pi_i' \frac{w_s w_t'}{\sqrt{N}} \pi_i + \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \eta_{it}\eta_{is} + \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left(\pi_i' w_t \eta_{is} + \pi_i' w_s \eta_{it}\right).$$

By assumptions made we have $E(w_t w_s' | \mathcal{F}) = 0$ and η_{it}'s independent from $\mathcal{F}$ and w_t's with mean zero, so $E(\rho(s,t)|\mathcal{F}) = 0$. This also implies that in $E(\rho(s,t)^2|\mathcal{F})$ all interaction terms are zero, so we have:

$$\begin{aligned} E(\rho(s,t)^2|\mathcal{F}) = & E\left[\left(\sum_{i=1}^{N} \pi_i' \frac{w_s w_t'}{\sqrt{N}} \pi_i\right)^2 |\mathcal{F}\right] + E\left(\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \eta_{it}\eta_{is}\right)^2 \\ & + E\left[\left(\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \pi_i' w_t \eta_{is}\right)^2 |\mathcal{F}\right] + E\left[\left(\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \pi_i' w_s \eta_{it}\right)^2 |\mathcal{F}\right] \end{aligned}$$

We can note that

$$\begin{aligned} \left|\sum_{i=1}^{N} \pi_i' w_s w_t' \pi_i\right| = & \left|\operatorname{tr}(w_s w_t' \sum_{i=1}^{N} \pi_i \pi_i')\right| \leq k_w \max eval\left(w_s w_t' \sum_{i=1}^{N} \pi_i \pi_i'\right) \\ & \leq k_w \|w_s\| \|w_t\| \max eval\left(\sum_{i=1}^{N} \pi_i \pi_i'\right) \leq C \|w_s\| \|w_t\|. \end{aligned}$$

Here we used that the scalar product can be represented as a trace, $\operatorname{tr}(ABC) = \operatorname{tr}(BCA)$, and the trace equals to a sum of eigenvalues and as such is bounded by dimensionality times the maximal eigenvalue. In the last inequality we used that loadings π_i imply only weak factor structure. Due to independence of η_{it}'s, it is easy to see that

$$E\left(\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \eta_{it}\eta_{is}\right)^2 = \frac{1}{N} \sum_{i=1}^{N} \sigma_i^2 < C,$$

and

$$Var\left(\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \pi_i' w_t \eta_{is} | \mathcal{F}\right) = \frac{1}{N} \sum_{i=1}^{N} \pi_i' E(w_t w_t' | \mathcal{F}) \pi_i \sigma_i^2 < \frac{C E(\|w_t\|^2 | \mathcal{F})}{N}.$$

Thus,

$$\begin{aligned} E\left[(\|F_t\|^4 + 1)\rho(s,t)^2\right] = & E\left[(\|F_t\|^4 + 1) E(\rho(s,t)^2 | \mathcal{F})\right] \\ & \leq E\left[(\|F_t\|^4 + 1)\left(C \frac{(\|w_t\|^2 + 1)(\|w_s\|^2 + 1)}{N} + C\right)\right] < \infty, \end{aligned}$$

this proves validity of Assumption ERRORS(ii).

Now consider

$$S_t = \frac{1}{N}\sum_i e_{it}^2 = \sum_{i=1}^N \pi_i' \frac{w_t w_t'}{N}\pi_i + 2w_t' \frac{\sum_i \pi_i \eta_{it}}{N} + \frac{1}{N}\sum_i (\eta_{it}^2 - \sigma_i^2) + \frac{1}{N}\sum_i \sigma_i^2.$$

Denote $\Phi_t = (1, F_t', vec(F_t F_t')')'$. First, let us prove that

$$\frac{\sqrt{N}}{T}\sum_t (S_t - \frac{1}{N}\sum_i \sigma_i^2)\Phi_t = o_p(1). \tag{7}$$

The only non-trivial parts are $\frac{1}{\sqrt{NT}}\sum_{i=1}^N \sum_{t=1}^T \Phi_t w_t' \pi_i \eta_{it} = o_p(1)$ and $\frac{1}{\sqrt{NT}}\sum_{i=1}^N \sum_{t=1}^T \Phi_t(\eta_{it}^2 - \sigma_i^2) = o_p(1)$. For them we use Chebyshev's inequality:

$$\begin{aligned} E\left\|\frac{1}{\sqrt{NT}}\sum_{i=1}^N\sum_{t=1}^T \Phi_t w_t' \pi_i \eta_{it}\right\|^2 &= \frac{1}{NT^2}\sum_{i=1}^N \sigma_i^2 E\left(\sum_{t,s=1}^T \Phi_t' \Phi_s \pi_i' w_t w_s' \pi_i\right) \\ &= \frac{1}{NT^2}\sum_{i=1}^N \sigma_i^2 \pi_i' E\left(\sum_{t=1}^T \|\Phi_t\|^2 E(w_t w_t'|\mathcal{F})\right)\pi_i \\ &\leq \frac{k_w}{NT^2}\max eval\left(\sum_{i=1}^N \pi_i \pi_i' \sigma_i^2\right)\max evalE\left(\sum_{t=1}^T \|\Phi_t\|^2 E(w_t w_t'|\mathcal{F})\right) \end{aligned}$$

For the first equality we used that η_{it} are independent from each other and from all F_t's and w_t's; for the second – that w_t is conditionally serially uncorrelated and have conditional mean zero. By the moment assumptions,

$$\begin{aligned} \max evalE\left(\sum_{t=1}^T \|\Phi_t\|^2 E(w_t w_t'|\mathcal{F})\right) &= \max evalE\left(\sum_{t=1}^T \|\Phi_t\|^2 w_t w_t'\right) \\ &\leq TE(\|F_t\|^4 + \|F_t\|^2 + 1)\|w_t\|^2 \leq CT. \end{aligned}$$

The variances σ_i^2 are all bounded and the factors are weak, which leads to

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^N\sum_{t=1}^T \Phi_t w_t' \pi_i \eta_{it} = o_p(1).$$

Similarly,

$$E\left\|\frac{1}{\sqrt{NT}}\sum_{i=1}^N\sum_{t=1}^T \Phi_t(\eta_{it}^2 - \sigma_i^2)\right\|^2 = \frac{1}{NT^2}\sum_{i=1}^N\sum_{t=1}^T E\left(\eta_{it}^2 - \sigma_i^2\right)^2 E\|\Phi_t\|^2 \to 0.$$

This gives the validity of statement (7).

Let us now prove the first statement in Assumption ERRORS(iii):

$$\frac{\sqrt{N}}{T}\sum_{t=1}^{T}\widetilde{F}_t S_t = \frac{\sqrt{N}}{T}\sum_{t=1}^{T}F_t(S_t - \frac{1}{N}\sum_i \sigma_i^2) + \frac{1}{T}\sum_{t=1}^{T}F_t\sqrt{N}\left(\frac{1}{N}\sum_i \sigma_i^2 - \frac{1}{T}\sum_{t=1}^{T}S_t\right).$$

The first term is $o_p(1)$ according to statement (7) as F_t is part of Φ_t. Statement (7) also implies that $\sqrt{N}(\frac{1}{N}\sum_i \sigma_i^2 - \frac{1}{T}\sum_{t=1}^{T}S_t) = o_p(1)$, that gives negligibility of the second term.

Now consider the second statement in Assumption ERRORS(iii):

$$\frac{1}{T}\sum_{t=1}^{T}\widetilde{F}_t\widetilde{F}_t' S_t = \frac{1}{T}\sum_{t=1}^{T}F_tF_t'S_t - \overline{F}\frac{1}{T}\sum_{t=1}^{T}\widetilde{F}_t'S_t - \frac{1}{T}\sum_{t=1}^{T}F_tS_t\overline{F}'.$$

We have proved above that the second term is $o_p(1)$. By equation (7), the first term equals $\frac{1}{T}\sum_{t=1}^{T}F_tF_t'\frac{1}{N}\sum\sigma_i^2 + o_p(1) \to^p \sigma^2 EF_tF_t'$, while the third term equals to $-\overline{F}\overline{F}'\frac{1}{N}\sum\sigma_i^2 + o_p(1) \to^p -\sigma^2 EF_tEF_t'$. So, the second statement in Assumption(iii) holds with $\Sigma_{SF^2} = \sigma^2 Var(F_t)$.

Finally, for Assumption ERRORS(iv), consider

$$W_t = w_t\frac{1}{\sqrt{N}}\sum_i \pi_i\gamma_i + \frac{1}{\sqrt{N}}\sum_i \gamma_i\eta_{it}.$$

Thus

$$E\left[(1+\|F_t\|^2)\|W_t\|^2\right] = E\left[(1+\|F_t\|^2)\left\|\frac{1}{\sqrt{N}}\sum_i \gamma_i\pi_i'w_t\right\|^2\right] + E\left[(1+\|F_t\|^2)\right]E\left\|\frac{\sum_i\gamma_i\eta_{it}}{\sqrt{N}}\right\|^2.$$

Notice that $E(w_t|\mathcal{F}) = 0$ and that η_{it}'s are independent from all other variables, thus there are no terms containing the first power of η_{it}. Now,

$$\begin{aligned}\left\|\frac{1}{\sqrt{N}}\sum_i \gamma_i\pi_i'w_t\right\|^2 &= \left(\frac{1}{\sqrt{N}}\sum_i\gamma_i\pi_i'\right)w_tw_t'\left(\frac{1}{\sqrt{N}}\sum_i\gamma_i\pi_i'\right)' \\ &= \mathrm{tr}\left[w_tw_t'\left(\frac{1}{\sqrt{N}}\sum_i\gamma_i\pi_i'\right)'\left(\frac{1}{\sqrt{N}}\sum_i\gamma_i\pi_i'\right)\right] \\ &\le k_w\|w_t\|^2 \max eval\left[\left(\frac{1}{\sqrt{N}}\sum_i\gamma_i\pi_i'\right)'\left(\frac{1}{\sqrt{N}}\sum_i\gamma_i\pi_i'\right)\right] < C\|w_t\|^2.\end{aligned}$$

The assumptions on the loadings guarantee that the maximal eigenvalue is bounded above by a constant. Next,

$$E\left\|\frac{\sum_i \gamma_i \eta_{it}}{\sqrt{N}}\right\|^2 = \frac{1}{N}\sum_{i=1}^{N}\|\gamma_i\|^2\sigma_i^2 < C.$$

Thus Assumption ERRORS (iv) is valid as well. □

Proof of Lemma 1. Assumption ERRORS*(i) implies Assumption ERRORS(i). Given the independence between the two groups of variables and the moment condition for F_t stated in Assumption FACTORS, in order to prove Assumption ERRORS(ii) we need to show that $\sup_{s\neq t} E\rho(s,t)^2$ is bounded from above. Indeed,

$$E\rho(s,t)^2 = \frac{1}{N}\sum_{i,j=1}^{N} E[e_{it}e_{is}e_{jt}e_{js}] = \frac{1}{N}\sum_{i,j=1}^{N} E[e_{it}e_{jt}]E[e_{is}e_{js}] = \frac{1}{N}\mathrm{tr}(\mathcal{E}_{N,T}\mathcal{E}_{N,T}).$$

Here we used serial independence in Assumption ERRORS*(i) and the definition of covariance matrices. For any positive definite matrix A we have $\mathrm{tr}(A^2) = \sum_{i=1}^{N}\lambda_i(A)^2 \leq N\left(\max eval(A)\right)^2$, where $\lambda_i(A)$ are eigenvalues of an $N\times N$ matrix A. Thus, due to Assumption ERRORS*(ii), we have $\mathrm{tr}(\mathcal{E}_{N,T}\mathcal{E}_{N,T}) \leq NC^2$. Thus, the right-hand-side of the last displayed equation is bounded from above.

Assumption ERRORS (iii): Notice that since $\sum_{t=1}^{T}\widetilde{F}_t = 0$ we have

$$\frac{\sqrt{N}}{T}\sum_t \widetilde{F}_t S_t = \frac{\sqrt{N}}{T}\sum_t \widetilde{F}_t\left(S_t - \overline{\sigma_N^2}\right),$$

where we denote $\overline{\sigma_N^2} = N^{-1}\sum_{i=1}^{N} Ee_{it}^2$. Let us check the that the second moment of the last expression converges to zero:

$$E\left\|\frac{\sqrt{N}}{T}\sum_t \widetilde{F}_t\left(S_t - \overline{\sigma_N^2}\right)\right\|^2 = \frac{N}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T} E\left[\widetilde{F}_t'\widetilde{F}_s\left(S_t - \overline{\sigma_N^2}\right)\left(S_s - \overline{\sigma_N^2}\right)\right]$$

Given Assumption ERRORS*(i), only those terms survive that have $s = t$:

$$\frac{N}{T^2}\sum_{t=1}^{T} E\left[\widetilde{F}_t\widetilde{F}_t'\right] E\left(S_t - \overline{\sigma_N^2}\right)^2.$$

Notice that

$$NE\left(S_t - \overline{\sigma_N^2}\right)^2 = E\left|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(e_{it}^2 - Ee_{it}^2)\right|^2 < C,$$

using Assumption ERRORS*(iii). Thus, the first statement in ERRORS (iii) holds. For the second statement, note that

$$E\left(\frac{1}{T}\sum_{t=1}^{T}\widetilde{F}_t\widetilde{F}_t'S_t\right)=\Sigma_F\frac{1}{N}\mathrm{tr}(\mathcal{E}_T)\to a\Sigma_F=\Sigma_{SF^2}.$$

In order to prove the second statement in Assumption ERRORS(iii) we will show that $T^{-1}\sum_{t=1}^{T}\widetilde{F}_t\widetilde{F}_t'\left(S_t-\overline{\sigma_N^2}\right)\to^p 0$. Following the same steps as in the proof of the first statement,

$$E\left\|\frac{1}{T}\sum_{t=1}^{T}\widetilde{F}_t\widetilde{F}_t'\left(S_t-\overline{\sigma_N^2}\right)\right\|^2=\frac{1}{T^2}\sum_{t=1}^{T}\|\widetilde{F}_t\|^4E\left(S_t-\overline{\sigma_N^2}\right)^2,$$

and we showed before $E\left(S_t-\overline{\sigma_N^2}\right)^2\to 0$. This finishes a proof of validity of Assumption ERRORS(iii).

Lastly,

$$E\|W_t\|^2=\frac{1}{N}\gamma'\mathcal{E}_{N,T}\gamma\leq\frac{1}{N}\|\gamma\|^2\max eval(\mathcal{E}_{N,T})<C.$$

Thus, Assumption ERRORS (iv) holds as well. □

Lemma 4 *Under Assumptions FACTORS, LOADINGS and ERRORS we have the following convergence as $N,T\to\infty$:*

(1). $\frac{1}{\sqrt{N|T_k||T_j|}}\sum_{i=1}^{N}\sum_{t\in T_j}\sum_{s\in T_k}\widetilde{F}_te_{it}\widetilde{F}_s'e_{is}=O_p(1)$ *for* $T_j\cap T_k=\varnothing$,

(2). $\frac{1}{\sqrt{N|T_k||T_j|}}\sum_{i=1}^{N}\sum_{t\in T_j}\sum_{s\in T_k}\widetilde{F}_te_{it}e_{is}=O_p(1)$ *for* $T_j\cap T_k=\varnothing$,

(3). $\Sigma_F^{-1}\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\widetilde{F}_t\widetilde{F}_t'e_{it}^2\right)\Sigma_F^{-1}\to^p\Sigma_u=\Sigma_F^{-1}\Sigma_{SF^2}\Sigma_F^{-1}$,

(4). $\frac{1}{T\sqrt{N}}\sum_{i=1}^{N}\left(\sum_{t=1}^{T}\sum_{s=1,s\neq t}^{T}\widetilde{F}_t(\widetilde{F}_s',1)e_{it}e_{is}\right)=O_p(1)$,

(5). $\sqrt{\frac{|T_j|}{N}}\sum_{i=1}^{N}\gamma_i\otimes\begin{pmatrix}u_i^{(j)}\\ \frac{1}{|T_j|}\sum_{t\in T_j}e_{it}\end{pmatrix}=O_p(1)$.

Proof of Lemma 4.

Preamble. Notice that due to the absence of serial correlation of idiosyncratic errors stated in Assumption ERRORS(i), for $t\neq s$ and $t_1\neq s_1$ we have

$$E(\rho(s,t)\rho(s_1,t_1)|\mathcal{F})=0$$

unless $t = t_1$ and $s = s_1$ or $t = s_1$ and $s = t_1$.

Part (1). Note that

$$\frac{1}{\sqrt{N|T_k||T_j|}} \sum_{i=1}^{N} \sum_{t \in T_j} \sum_{s \in T_k} \widetilde{F}_t e_{it} \widetilde{F}_s' e_{is} = \frac{1}{\sqrt{|T_k||T_j|}} \sum_{t \in T_j} \sum_{s \in T_k} \widetilde{F}_t \widetilde{F}_s' \rho(t, s).$$

The expectation of the square of the last expression is equal to

$$\frac{1}{|T_k||T_j|} \sum_{t \in T_j} \sum_{t_1 \in T_j} \sum_{s \in T_k} \sum_{s_1 \in T_k} E\left(\widetilde{F}_t \widetilde{F}_s' \widetilde{F}_{t_1} \widetilde{F}_{s_1}' E(\rho(t, s)\rho(t_1, s_1)|\mathcal{F}) \right).$$

Using the preamble statement, we reduce four summation signs to only two, with each summand bounded above by Assumption ERRORS(ii). This implies that the second moment of the last sum is bounded, and hence implies statement (1).

Part (2). Analogously to Part (1).

Part (3). Note that

$$\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \widetilde{F}_t \widetilde{F}_t' e_{it}^2 = \frac{1}{T} \sum_{t=1}^{T} \widetilde{F}_t \widetilde{F}_t' S_t.$$

Part (3) follows from Assumption ERRORS(iii).

Part (4). Note that

$$\frac{1}{T\sqrt{N}} \sum_{i=1}^{N} \left(\sum_{t=1}^{T} \sum_{s=1, s \neq t}^{T} \widetilde{F}_t e_{it} e_{is} \right) = \frac{1}{T} \left(\sum_{t=1}^{T} \sum_{s=1, s \neq t}^{T} \widetilde{F}_t \rho(s, t) \right).$$

The second moment of this expression contains four summations – over t, $s \neq t$, t_1 and $s_1 \neq t_1$. However, by the preamble statement many terms are zero and the expression can be written as a double sum. Assumption ERRORS(ii) guarantees that all summands are bounded by the same constant, which leads to boundedness of the second moment of the expression of interest. Chebyshev's inequality delivers statement (4).

Part (5). Observe that

$$\sqrt{\frac{|T_j|}{N}} \sum_{i=1}^{N} \gamma_i \otimes \begin{pmatrix} u_i^{(j)} \\ \frac{1}{|T_j|} \sum_{t \in T_j} e_{it} \end{pmatrix} = \frac{1}{\sqrt{|T_j|}} \sum_{t \in T_j} \sum_{i=1}^{N} \frac{1}{\sqrt{N}} \left(\gamma_i \otimes \begin{pmatrix} \Sigma_F^{-1} \widetilde{F}_t^{(j)} \\ 1 \end{pmatrix} \right) e_{it}$$
$$= \frac{1}{\sqrt{|T_j|}} \sum_{t \in T_j} W_t \otimes \begin{pmatrix} \Sigma_F^{-1} \widetilde{F}_t^{(j)} \\ 1 \end{pmatrix},$$

where W_t is defined in Assumption ERRORS(iv). Given serial independence of e_t conditional on $\mathcal{F}$, we get that W_t is conditionally serially independent and mean zero, and as such

$$
\begin{aligned}
&E\left\|\frac{1}{\sqrt{|T_j|}}\sum_{t\in T_j}W_t\otimes\left(\begin{array}{c}\Sigma_F^{-1}\widetilde{F}_t^{(j)}\\ 1\end{array}\right)\right\|^2\\
=&\frac{1}{|T_j|}\sum_{t\in T_j}E\left\|W_t\otimes\left(\begin{array}{c}\Sigma_F^{-1}\widetilde{F}_t^{(j)}\\ 1\end{array}\right)\right\|^2\\
\leq&CE\left[(1+\|F_t\|^2)\|W_t\|^2\right]<C.
\end{aligned}
$$

□

Proof of Theorem 1. Assumptions FACTORS guarantee that the Central Limit Theorem holds for sums of F_t and thus $\sqrt{T}(\widetilde{\lambda}-\lambda)\Rightarrow N(0,\Omega_F)$, where Ω_F is the long-run variance of F_t.

The first pass (time series) regression yields

$$
\widehat{\beta}_i=\left(\beta_i+\frac{\eta_T\mu_i}{\sqrt{T}}+u_i\right)(1+o_p(1)), \tag{8}
$$

where we have used assumption FACTORS. The $o_p(1)$ appears from the difference between Σ_F and $T^{-1}\sum_{t=1}^T\widetilde{F}_t\widetilde{F}_t'$.

Denote $Q_T=\left(\begin{array}{cc}I_{k_1}&0_{k_1,k_2}\\ 0_{k_2,k_1}&\sqrt{T}I_{k_2}\end{array}\right)$. Notice that $Q_T/\sqrt{T}\to\mathcal{I}_{k_2}$. Let us prove that as $N,T\to\infty$ we have

$$
N^{-1}\sum_{i=1}^N Q_T\widehat{\beta}_i\widehat{\beta}_i'Q_T\Rightarrow(I_{k_\beta};\widetilde{\eta})\Gamma(I_{k_\beta};\widetilde{\eta})'+\mathcal{I}_{k_2}\Sigma_u\mathcal{I}_{k_2}. \tag{9}
$$

Indeed,

$$
\begin{aligned}
\frac{1}{N}\sum_{i=1}^N Q_T\widehat{\beta}_i\widehat{\beta}_i'Q_T&=\frac{1}{N}\sum_{i=1}^N\left(Q_T\beta_i+Q_T\frac{\eta_T}{\sqrt{T}}\mu_i+Q_Tu_i\right)\left(Q_T\beta_i+Q_T\frac{\eta_T}{\sqrt{T}}\mu_i+Q_Tu_i\right)'\\
&=\frac{1}{N}\sum_{i=1}^N\left((I_{k_\beta};\widetilde{\eta}_T)\gamma_i+Q_Tu_i\right)\left((I_{k_\beta};\widetilde{\eta}_T)\gamma_i+Q_Tu_i\right)', \qquad (10)
\end{aligned}
$$

where $\widetilde{\eta}_T=Q_T\eta_T/\sqrt{T}\Rightarrow\mathcal{I}_{k_2}\eta=\widetilde{\eta}$ is $k_F\times k_v$ gaussian random matrix. Let us show that

$$
\frac{T}{N}\sum_i u_iu_i'\to\Sigma_u. \tag{11}
$$

Indeed, due to statement (4) of Lemma 4 we have that

$$
\frac{1}{TN}\sum_{i=1}^N\sum_{t=1}^T\sum_{s=1,s\neq t}^T\widetilde{F}_t\widetilde{F}_s'e_{it}e_{is}=o_p(1).
$$

Thus,

$$\frac{T}{N}\sum_i u_i u_i' = \Sigma_F^{-1}\left(\frac{1}{TN}\sum_{i=1}^N\sum_{t=1}^T\sum_{s=1}^T \widetilde{F}_t\widetilde{F}_s' e_{it}e_{is}\right)\Sigma_F^{-1}$$
$$= \Sigma_F^{-1}\left(\frac{1}{TN}\sum_{i=1}^N\sum_{t=1}^T \widetilde{F}_t\widetilde{F}_t' e_{it}^2\right)\Sigma_F^{-1} + o_p(1) \to^p \Sigma_u,$$

where the last convergence comes from statement (3) of Lemma 4. Statement (5) of Lemma 4 implies that

$$\frac{\sqrt{T}}{N}\sum_{i=1}^N \gamma_i u_i' \to^p 0_{k,k_F}. \tag{12}$$

Combining equations (10)–(12) with Assumption LOADINGS we arrive to the validity of equation (9).

Attenuation bias:

$$\begin{pmatrix} \sqrt{T}AB_1 \\ AB_2 \end{pmatrix} = Q_T^{-1}\sqrt{T}AB = -\left(\frac{1}{N}Q_T\sum_i \widehat{\beta}_i\widehat{\beta}_i' Q_T\right)^{-1}\frac{Q_T}{\sqrt{T}}\frac{T}{N}\sum_i u_i u_i'\widetilde{\lambda}.$$

Combining equations (9), (11), $\widetilde{\lambda} \to^p \lambda$ and $Q_T/\sqrt{T} \to \mathcal{I}_{k_2}$, we arrive at:

$$\begin{pmatrix} \sqrt{T}AB_1 \\ AB_2 \end{pmatrix} \Rightarrow -\left((I_{k_\beta};\widetilde{\eta})\Gamma(I_{k_\beta};\widetilde{\eta})' + \mathcal{I}_{k_2}\Sigma_u\mathcal{I}_{k_2}\right)^{-1}\mathcal{I}_{k_2}\Sigma_u\lambda.$$

Omitted variable bias:

$$\begin{pmatrix} \sqrt{T}OVB_1 \\ OVB_2 \end{pmatrix} = Q_T^{-1}\sqrt{T}OVB = \left(\frac{1}{N}\sum_i Q_T\widehat{\beta}_i\widehat{\beta}_i' Q_T\right)^{-1}\frac{1}{N}\sum_i Q_T\widehat{\beta}_i\mu_i'(\eta_{v,T} - \eta_T'\widetilde{\lambda}).$$

Let us consider the following expression:

$$\frac{1}{N}\sum_{i=1}^N Q_T\widehat{\beta}_i\mu_i' = \frac{1}{N}\sum_{i=1}^N\left(Q_T\beta_i + Q_T\frac{\eta_T\mu_i}{\sqrt{T}} + Q_T u_i\right)\mu_i'. \tag{13}$$

By assumption LOADINGS, $N^{-1}\sum_i Q_T\beta_i\mu_i' \to \Gamma_{\beta\mu}$ and $N^{-1}\sum_i \mu_i\mu_i' \to \Gamma_{\mu\mu}$, while $Q_T\eta_T/\sqrt{T} \Rightarrow \widetilde{\eta}$. The last term in equation (13) is $o_P(1)$ by statement (5) of Lemma 4. Thus,

$$\frac{1}{N}\sum_{i=1}^N Q_T\widehat{\beta}_i\mu_i' \Rightarrow \Gamma_{\beta\mu} + \widetilde{\eta}\Gamma_{\mu\mu},$$

We also note that

$$\eta_{v,T} - \eta_T'\widetilde{\lambda} \Rightarrow \eta_v - \eta'\lambda.$$

This implies the validity of asymptotic statement about OVB contained in Theorem 1.

The remainder part: By time averaging equation (2) we get:

$$\overline{r}_i = \beta_i'\widetilde{\lambda} + \mu_i'\frac{\eta_{v,T}}{\sqrt{T}} + \overline{e}_i.$$

Combining last equation with equation (8) we obtain

$$\overline{r}_i = \widehat{\beta}_i'\widetilde{\lambda} - u_i'\widetilde{\lambda} + \frac{\mu_i'}{\sqrt{T}}(\eta_{v,T} - \eta_T'\widetilde{\lambda}) + \overline{e}_i.$$

Thus, we arrive at

$$\begin{aligned}\widehat{\lambda}_{TP} - \widetilde{\lambda} - AB - OVB &= \left(\sum_i \widehat{\beta}_i\widehat{\beta}_i'\right)^{-1}\left(-\sum_i(\widehat{\beta}_i - u_i)u_i'\widetilde{\lambda} + \sum_i \widehat{\beta}_i\overline{e}_i\right)\\ &= \left(\sum_i \widehat{\beta}_i\widehat{\beta}_i'\right)^{-1}\left(-\sum_i(\beta_i + \frac{\eta_T\mu_i}{\sqrt{T}} + o_p(1))u_i'\widetilde{\lambda} + \sum_i \widehat{\beta}_i\overline{e}_i\right).\end{aligned}$$

After the proper normalization we get:

$$\begin{aligned}&\sqrt{NT}Q_T^{-1}(\widehat{\lambda}_{TP} - \widetilde{\lambda} - AB - OVB)\\ &= \left(\frac{1}{N}\sum_i Q_T\widehat{\beta}_i\widehat{\beta}_i'Q_T\right)^{-1}\sqrt{\frac{T}{N}}\left(-\sum_i Q_T(\beta_i + \frac{\eta_T\mu_i}{\sqrt{T}})u_i'\widetilde{\lambda} + \sum_i Q_T\widehat{\beta}_i\overline{e}_i\right).\end{aligned}$$

Let us prove that the numerator is asymptotically $O_p(1)$:

$$\begin{aligned}&\sqrt{\frac{T}{N}}\left(-\sum_i Q_T(\beta_i + \frac{\eta_T\mu_i}{\sqrt{T}})u_i'\widetilde{\lambda} + \sum_i Q_T\widehat{\beta}_i\overline{e}_i\right)\\ &= (I_{k_\beta};\widetilde{\eta}_T)\sqrt{\frac{T}{N}}\sum_i \gamma_i(\overline{e}_i - u_i'\widetilde{\lambda}) + \sqrt{\frac{T}{N}}\sum_i Q_T u_i\overline{e}_i + O_p(1). \qquad (14)\end{aligned}$$

According to statement (5) of Lemma 4 we have that $\sqrt{\frac{T}{N}}\sum_i \gamma_i\overline{e}_i = O_p(1)$ and $\sqrt{\frac{T}{N}}\sum_i \gamma_i u_i' = O_p(1)$, which makes the first summand in equation (14) $O_p(1)$. Consider the second term in equation (14) and recall that $Q_T/\sqrt{T} = O(1)$:

$$\begin{aligned}\sqrt{\frac{T}{N}}\sum_i Q_T u_i\overline{e}_i &= \frac{Q_T}{\sqrt{T}}\Sigma_F^{-1}\frac{1}{\sqrt{NT}}\sum_i\sum_t\sum_s \widetilde{F}_s e_{is}e_{it}\\ &= \frac{Q_T}{\sqrt{T}}\Sigma_F^{-1}\frac{1}{\sqrt{NT}}\sum_i\sum_t\sum_{s\neq t} \widetilde{F}_s e_{is}e_{it} + \frac{Q_T}{\sqrt{T}}\Sigma_F^{-1}\frac{\sqrt{N}}{T}\sum_t \widetilde{F}_t S_t.\end{aligned}$$

The first term is $O_p(1)$ by statement (4) of Lemma 4, while the second term is $O_p(1)$ by Assumption ERRORS(iii). This ends the proof of Theorem 1. □

Proof of Corollary 2 When we have weak included factors ($k_2 \geq 1$) but no strong excluded factors ($k_v = 0$), the expression for the omitted variable bias ($\mu_i = 0$) equals to zero exactly: $OVB = 0$. In this case, $v_t = 0$ and hence $\eta_T = 0$ as well as $\eta = 0$ and $\widetilde{\eta} = 0$. That gives the expression (4) for the limit of the attenuation bias. $\square$

Proof of Corollary 1 If all observed factors are strong, then there is no second component to the risk premia, i.e., $k_2 = 0$ and $\lambda = \lambda_1$. We also have $\mathcal{I}_{k_2} = 0_{k_F,k_F}$ and $\widetilde{\eta} = 0$. When applied to the result of Theorem 1 we get that $\sqrt{T}AB \to^p 0$ and

$$\sqrt{T}OVB \Rightarrow (\Gamma_{\beta\beta})^{-1}\Gamma_{\beta\mu}(\eta_v - \eta'\lambda),$$

which is a zero mean gaussian limit. Thus in this case we have

$$\sqrt{T}(\widehat{\lambda} - \lambda) = \sqrt{T}(\widetilde{\lambda} - \lambda) + \sqrt{T}OVB + o_p(1).$$

Finally, if in addition to $k_2 = 0$ we also have $k_v = 0$ (no missing factor structure), then $OVB = 0$ exactly. $\square$

Proof of Theorem 2. We first discuss asymptotics of just one IV regression described on step (2), then this argument will be repeated for the other three IV regressions from step (2) of the algorithm. Denote $\tau = \lfloor \frac{T}{4} \rfloor = |T_j|$.

The time-series regression on a sub-sample j gives us that:

$$\widehat{\beta}_i^{(j)} = \left(\beta_i + u_i^{(j)} + \frac{\eta_{j,T}\mu_i}{\sqrt{\tau}}\right)(1 + o_p(1)),$$

where

$$\eta_{j,T} = \frac{1}{\sqrt{\tau}} \sum_{t \in T_j} \Sigma_F^{-1} \widetilde{F}_t^{(j)} v_t' \Rightarrow \eta_j,$$

η_j is random $k_F \times k_v$ matrix with the following distribution $vec(\eta_j) \sim N(0_{k_F k_v,1}, \Omega_{vF})$, the $o_p(1)$ term is related to the difference between Σ_F and $\frac{1}{\tau}\sum_{t\in T_j}\sum_t \widetilde{F}_t^{(j)}\widetilde{F}_t^{(j)\prime}$.

On step (2) we run an IV regression of $y_i = \frac{1}{T}\sum_{t=1}^T r_{it}$ on regressor

$$x_i^{(1)} = \begin{pmatrix} \widehat{\beta}_i^{(1)} \\ A_1(\widehat{\beta}_i^{(1)} - \widehat{\beta}_i^{(2)}) \end{pmatrix} = \begin{pmatrix} \widehat{\beta}_i^{(1)} \\ A_1\frac{\eta_{1,T}-\eta_{2,T}}{\sqrt{\tau}}\mu_i + A_1(u_i^{(1)} - u_i^{(2)}) \end{pmatrix},$$

with instruments

$$z_i^{(1)} = \begin{pmatrix} \widehat{\beta}_i^{(3)} \\ \widehat{\beta}_i^{(3)} - \widehat{\beta}_i^{(4)} \end{pmatrix} = \begin{pmatrix} \widehat{\beta}_i^{(3)} \\ \frac{\eta_{3,T}-\eta_{4,T}}{\sqrt{\tau}}\mu_i + (u_i^{(3)} - u_i^{(4)}) \end{pmatrix}.$$

The main estimation equation can be written in the following way:

$$\begin{aligned}
y_i =& \frac{1}{T}\sum_{t\in T} r_{it} = \widetilde{\lambda}'\beta_i + \frac{\eta'_{v,T}}{\sqrt{T}}\mu_i + \overline{e}_i \\
=& \widetilde{\lambda}'\widehat{\beta}_i^{(1)} + \left(\frac{\eta'_{v,T}}{\sqrt{T}} - \widetilde{\lambda}'\frac{\eta_{1,T}}{\sqrt{\tau}}\right)\mu_i + \overline{e}_i - \widetilde{\lambda}' u_i^{(1)} \\
=& \widetilde{\lambda}'\widehat{\beta}_i^{(1)} + a_{1,T}A_1(\widehat{\beta}_i^{(1)} - \widehat{\beta}_i^{(2)}) + \overline{e}_i - \widetilde{\lambda}' u_i^{(1)} - a_{1,T}A_1(u_i^{(1)} - u_i^{(2)}).
\end{aligned}$$

Thus, we can write it as follows:

$$y_i = (\widetilde{\lambda}', a_{1,T})x_i^{(1)} + \epsilon_i^{(1)}. \tag{15}$$

Here we use the following notation

$$\begin{aligned}
a_{1,T} =& \left(\frac{\eta'_{v,T}}{\sqrt{T}} - \widetilde{\lambda}'\frac{\eta_{1,T}}{\sqrt{\tau}}\right)\left(A_1\frac{\eta_{1,T} - \eta_{2,T}}{\sqrt{\tau}}\right)^{-1} \\
=& \left(\frac{\eta'_{v,T}}{2} - \eta_{1,T}\right)(A_1(\eta_{1,T} - \eta_{2,T}))^{-1} \Rightarrow \left(\frac{\eta'_v}{2} - \eta_1\right)(A_1(\eta_1 - \eta_2))^{-1},
\end{aligned}$$

and

$$\epsilon_i^{(1)} = \overline{e}_i - \widetilde{\lambda}' u_i^{(1)} - a_{1,T}A_1\left(u_i^{(1)} - u_i^{(2)}\right).$$

Notice that $a_{1,T}$ is random $1\times k_v$ matrix, that is well defined with probability approaching 1 (as $\eta_{1,T}$ and $\eta_{2,T}$ weakly converge to two independent random gaussian matrices) and $a_{1,T}$ is asymptotically of order $O_p(1)$.

The estimator calculated on the step (2) of the 4-split algorithm is:

$$\widehat{\lambda}^{(1)} = (I_{k_F}, 0_{k_F,k_v})\left(X^{(1)\prime}Z^{(1)}(Z^{(1)\prime}Z^{(1)})^{-1}Z^{(1)\prime}X^{(1)}\right)^{-1}X^{(1)\prime}Z^{(1)}(Z^{(1)\prime}Z^{(1)})^{-1}Z^{(1)\prime}Y.$$

Using equation (15) we obtain:

$$\widehat{\lambda}^{(1)} - \widetilde{\lambda} = (I_{k_F}, 0_{k_F,k_v})\left(X^{(1)\prime}P_{Z^{(1)}}X^{(1)}\right)^{-1}X^{(1)\prime}P_{Z^{(1)}}\epsilon^{(1)}, \tag{16}$$

where P_Z is the projection matrix onto Z. Let us introduce two normalizing matrices:

$$Q_x = \begin{pmatrix} Q & 0_{k_F,k_v} \\ 0_{k_v,k_F} & \sqrt{T}I_{k_v} \end{pmatrix}, \qquad Q_z = \begin{pmatrix} Q & 0_{k_F,k_F} \\ 0_{k_F,k_F} & \sqrt{T}I_{k_F} \end{pmatrix}.$$

The dimension of Q_x is $k\times k$, where $k = k_F + k_v$ is the number of regressors in the second stage regression, while Q_z is $2k_F \times 2k_F$, where $2k_F$ is the number of instruments. Matrix $Q = \begin{pmatrix} I_{k_1} & 0_{k_1,k_2} \\ 0_{k_2,k_1} & \sqrt{T}I_{k_2} \end{pmatrix}$ was defined in the proof of Theorem 1.

$$Q_x x_i^{(1)} = \left(\widetilde{A}_{1,T}\gamma_i + \begin{pmatrix} Q & 0_{k_F,k_F} \\ \sqrt{T}A_1 & -\sqrt{T}A_1 \end{pmatrix} \begin{pmatrix} u_i^{(1)} \\ u_i^{(2)} \end{pmatrix} \right),$$

where

$$\widetilde{A}_{1,T} = \begin{pmatrix} I_{k_F} & Q\frac{\eta_{1,T}}{\sqrt{\tau}} \\ 0_{k_v,k_F} & 2A_1(\eta_{1,T} - \eta_{2,T}) \end{pmatrix} \Rightarrow \begin{pmatrix} I_{k_F} & 2\mathcal{I}_{k_2}\eta_1 \\ 0_{k_v,k_F} & 2A_1(\eta_1 - \eta_2) \end{pmatrix} = \widetilde{A}_1,$$
$$\frac{1}{\sqrt{T}} \begin{pmatrix} Q & 0_{k_F,k_F} \\ \sqrt{T}A_1 & -\sqrt{T}A_1 \end{pmatrix} \to \begin{pmatrix} \mathcal{I}_{k_2} & 0_{k_F,k_F} \\ A_1 & -A_1 \end{pmatrix}.$$

Here $\mathcal{I}_{k_2} = \begin{pmatrix} 0_{k_1,k_1} & 0_{k_1,k_2} \\ 0_{k_2,k_1} & I_{k_2} \end{pmatrix}$ is $k_F \times k_F$ matrix which was introduced in Theorem 1. We also have

$$Q_z z_i^{(1)} = \left(A^*_{1,T}\gamma_i + \begin{pmatrix} Q/\sqrt{T} & 0_{k_F,k_F} \\ I_{k_F} & -I_{k_F} \end{pmatrix} \sqrt{T} \begin{pmatrix} u_i^{(3)} \\ u_i^{(4)} \end{pmatrix} \right),$$

where

$$A^*_{1,T} = \begin{pmatrix} I_{k_F} & Q\frac{\eta_{3,T}}{\sqrt{\tau}} \\ 0_{k_F,k_F} & 2(\eta_{3,T} - \eta_{4,T}) \end{pmatrix} \Rightarrow \begin{pmatrix} I_{k_F} & 2\mathcal{I}_{k_2}\eta_3 \\ 0_{k_F,k_F} & 2(\eta_3 - \eta_4) \end{pmatrix} = A^*_1,$$
$$\begin{pmatrix} Q/\sqrt{T} & 0_{k_F,k_F} \\ I_{k_F} & -I_{k_F} \end{pmatrix} \to \begin{pmatrix} \mathcal{I}_{k_2} & 0_{k_F,k_F} \\ I_{k_F} & -I_{k_F} \end{pmatrix}.$$

Statements (1) and (5) of Lemma 4 imply that:

$$\frac{T}{\sqrt{N}} \sum_{i=1}^{N} u_i^{(j)} u_i^{(j^*)\prime} = O_p(1) \quad \text{for} \quad j \neq j^*, \tag{17}$$
$$\sqrt{\frac{T}{N}} \sum_{i=1}^{N} (\gamma_i', 1)' u_i^{(j^*)\prime} = O_p(1). \tag{18}$$

This together with Assumption LOADINGS gives us that

$$\frac{1}{N} \sum_{i=1}^{N} Q_x x_i^{(1)} z_i^{(1)\prime} Q_z \Rightarrow \widetilde{A}_1 \Gamma A_1^{*\prime}. \tag{19}$$

According to Assumption LOADINGS Γ is full rank, matrices $\widetilde{A}_1$ and $A_1^{*\prime}$ are full rank with probability 1. Statements (3) and (4) of Lemma 4 say:

$$\frac{\tau}{N} \sum_{i=1}^{N} u_i^{(j)} u_i^{(j)\prime} \to^p \Sigma_u. \tag{20}$$

Thus we get:

$$\frac{1}{N}\sum_{i=1}^{N} Q_z z_i^{(1)} z_i^{(1)\prime} Q_z \Rightarrow A_1^*\Gamma A_1^{*\prime} + 4\begin{pmatrix} \mathcal{I}_{k_2} & 0_{k_F,k_F} \\ I_{k_F} & -I_{k_F} \end{pmatrix}\begin{pmatrix} \Sigma_u & 0_{k_F,k_F} \\ 0_{k_F,k_F} & \Sigma_u \end{pmatrix}\begin{pmatrix} \mathcal{I}_{k_2} & I_{k_F} \\ 0_{k_F,k_F} & -I_{k_F} \end{pmatrix}$$

$$= A_1^*\Gamma A_1^{*\prime} + 4\begin{pmatrix} \mathcal{I}_{k_2}\Sigma_u\mathcal{I}_{k_2} & \mathcal{I}_{k_2}\Sigma_u \\ \Sigma_u\mathcal{I}_{k_2} & 2\Sigma_u \end{pmatrix}. \quad (21)$$

Let us now show that

$$\sqrt{\frac{T}{N}}\sum_{i=1}^{N} Q_z z_i^{(1)}\epsilon_i^{(1)} = O_p(1). \quad (22)$$

We have

$$\epsilon_i^{(1)} = \overline{e}_i - \widetilde{\lambda}' u_i^{(1)} - a_{1,T}A_1(u_i^{(1)} - u_i^{(2)}).$$

The sum in (22) contains summands of the form $\sqrt{\frac{T}{N}}\sum_i \gamma_i(\overline{e}_i, u_i^{(j)})$, $\frac{T}{\sqrt{N}}\sum_i \overline{e}_i u_i^{(j)}$ and $\frac{T}{\sqrt{N}}\sum_i u_i^{(j^*)\prime} u_i^{(j)}$. All three types of summands are $O_p(1)$ due to Lemma 4 statements (5), (2) and (1) correspondingly. Putting equations (19) and (21) together we get:

$$\begin{aligned} NQ_x^{-1}\Theta_{N,T,1}Q_z^{-1} =& NQ_x^{-1}\left(X^{(1)\prime}Z^{(1)}(Z^{(1)\prime}Z^{(1)})^{-1}Z^{(1)\prime}X^{(1)}\right)^{-1}X^{(1)\prime}Z^{(1)}(Z^{(1)\prime}Z^{(1)})^{-1}Q_z^{-1} \\ =& \left(\frac{Q_xX^{(1)\prime}Z^{(1)}Q_z}{N}\left(\frac{Q_zZ^{(1)\prime}Z^{(1)}Q_z}{N}\right)^{-1}\frac{Q_zZ^{(1)\prime}X^{(1)}Q_x}{N}\right)^{-1} \\ &\cdot\frac{Q_xX^{(1)\prime}Z^{(1)}Q_z}{N}\left(\frac{Q_zZ^{(1)\prime}Z^{(1)}Q_z}{N}\right)^{-1} = O_p(1). \end{aligned}$$

Putting everything together, we have:

$$\sqrt{NT}Q^{-1}(\widehat{\lambda}^{(1)} - \widetilde{\lambda}) = (I_{k_F}, 0_{k_F,k_v})NQ_x^{-1}\Theta_{N,T,1}Q_z^{-1}\sqrt{\frac{T}{N}}\sum_{i=1}^{N} Q_z z_i^{(1)}\epsilon_i^{(1)} = O_p(1).$$

Since $\sqrt{NT}Q^{-1} = \begin{pmatrix} \sqrt{NT}I_{k_1} & 0_{k_1,k_2} \\ 0_{k_2,k_1} & \sqrt{N}I_{k_2} \end{pmatrix}$, we got different rates of estimation of risk premia λ_1 and λ_2 on strong and weak observed factors. We have $\sqrt{NT}(\widehat{\lambda}_1^{(1)} - \widetilde{\lambda}_1) = O_p(1)$, while $\sqrt{N}(\widehat{\lambda}_2^{(1)} - \widetilde{\lambda}_2) = O_p(1)$. Thus

$$\sqrt{T}(\widehat{\lambda}_1^{(1)} - \lambda_1) = \sqrt{T}(\widetilde{\lambda}_1 - \lambda_1) + \sqrt{T}(\widehat{\lambda}_1^{(1)} - \widetilde{\lambda}_1) = \sqrt{T}(\widetilde{\lambda}_1 - \lambda_1) + O_p(1/\sqrt{N}),$$

and

$$\sqrt{T}(\widetilde{\lambda}_1 - \lambda_1) \Rightarrow N(0, \Omega_F).$$

As about the estimator of the risk premia on the weak factor:

$$\widehat{\lambda}_2^{(1)} - \lambda_2 = (\widetilde{\lambda}_2 - \lambda_2) + (\widehat{\lambda}_2^{(1)} - \widetilde{\lambda}_2) = O_p(1/\sqrt{T}) + O_p(1/\sqrt{N}) = O_p(1/\sqrt{\min\{N,T\}}).$$

So far we have proved the statement of Theorem 2 for an estimator obtain using one step of algorithm, but the same line of reasoning applies to $\widehat{\lambda}^{(2)}, \widehat{\lambda}^{(3)}, \widehat{\lambda}^{(4)}$ and their average. This finishes the proof of Theorem 2. □

Proof of Theorem 3. Following the steps of the proof of Theorem 2 we get the following two statements

$$\frac{1}{N}\sum_{i=1}^{N} Q_x x_i^{(j)} z_i^{(j)\prime} Q_z \Rightarrow \widetilde{A}_j \mathbf{\Gamma} A_j^{*\prime}, \tag{23}$$

$$\frac{1}{N}\sum_{i=1}^{N} Q_z z_i^{(j)} z_i^{(j)\prime} Q_z \Rightarrow A_j^{*} \mathbf{\Gamma} A_j^{*\prime} + 4\begin{pmatrix} \mathcal{I}_{k_2}\Sigma_u\mathcal{I}_{k_2} & \mathcal{I}_{k_2}\Sigma_u \\ \Sigma_u\mathcal{I}_{k_2} & 2\Sigma_u \end{pmatrix}, \tag{24}$$

where A_j^* and $\widetilde{A}_j$ are random matrices that are deterministic functions of random vectors $(\eta_1, ..., \eta_4)$. Indeed, let us adopt the following notation. Let $j_1, ..., j_4$ be the circular indexes used for calculating $\widehat{\lambda}^{(j)}$. In particular, estimate $\widehat{\lambda}^{(j)}$ is calculated from the IV regression with regressor $x_i^{(j)} = \left(\widehat{\beta}_i^{(j_1)\prime}, (\widehat{\beta}_i^{(j_1)} - \widehat{\beta}_i^{(j_2)})' A_j'\right)'$ with instruments $z_i^{(j)} = \left(\widehat{\beta}_i^{(j_3)\prime}, (\widehat{\beta}_i^{(j_3)} - \widehat{\beta}_i^{(j_4)})'\right)'$. Then, similarly to the proof of Theorem 2, we obtain:

$$A_j^* = \begin{pmatrix} I_{k_F} & 2\mathcal{I}_{k_2}\eta_{j_3} \\ 0_{k_F,k_F} & 2(\eta_{j_3} - \eta_{j_4}) \end{pmatrix}, \qquad \widetilde{A}_j = \begin{pmatrix} I_{k_F} & 2\mathcal{I}_{k_2}\eta_{j_1} \\ 0_{k_v,k_F} & 2A_j(\eta_{j_1} - \eta_{j_2}) \end{pmatrix}.$$

So,

$$\begin{aligned} &NQ_x^{-1}\left(X^{(j)\prime}Z^{(j)}(Z^{(j)\prime}Z^{(j)})^{-1}Z^{(j)\prime}X^{(j)}\right)^{-1} X^{(j)\prime}Z^{(j)}(Z^{(j)\prime}Z^{(j)})^{-1}Q_z^{-1} \\ =&NQ_x^{-1}\Theta_{N,T,j}Q_z^{-1} \Rightarrow \Theta_j. \end{aligned}$$

The limit Θ_j in the last expression is a known deterministic function of random vectors $(\eta_1, ..., \eta_4)$, which can be explicitly written in terms of A_j^* and $\widetilde{A}_j$.

We have the following expression for the estimates obtained on the steps (2) and (3) of the 4-split algorithm:

$$\sqrt{NT}Q^{-1}(\widehat{\lambda}^{(j)} - \widetilde{\lambda}) = (I_{k_F}, 0_{k_F,k_v})NQ_x^{-1}\Theta_{N,T,j}Q_z^{-1}\sqrt{\frac{T}{N}}\sum_{i=1}^{N} Q_z z_i^{(j)}\epsilon_i^{(j)},$$

where

$$\begin{aligned} \epsilon_i^{(j)} =& \bar{e}_i - \widetilde{\lambda}' u_i^{(j_1)} - a_{j,T}A_j(u_i^{(j_1)} - u_i^{(j_2)}), \\ Q_z z_i^{(j)} =& \left(A_{j,T}^*\gamma_i + \begin{pmatrix} Q/\sqrt{T} & 0_{k_F,k_F} \\ I_{k_F} & -I_{k_F} \end{pmatrix}\sqrt{T}\begin{pmatrix} u_i^{(j_3)} \\ u_i^{(j_4)} \end{pmatrix}\right). \end{aligned}$$

Consider the following term which could be re-written in terms of ξ_i from Assumption GAUSSIANITY:

$$\sqrt{\frac{T}{N}}\sum_{i=1}^{N} Q_z z_i^{(j)}\epsilon_i^{(j)} = A_{j,T}^{*}\left(\sqrt{\frac{T}{N}}\sum_{i=1}^{N}\gamma_i\begin{pmatrix}\overline{e}_i\\ u_i^{(j_1)}\\ u_i^{(j_2)}\end{pmatrix}\right)'\begin{pmatrix}1\\ -\widetilde{\lambda}-A_j'a_{j,T}'\\ A_j'a_{j,T}'\end{pmatrix}$$

$$+\begin{pmatrix}Q/\sqrt{T} & 0_{k_F,k_F}\\ I_{k_F} & -I_{k_F}\end{pmatrix}\left(\frac{T}{\sqrt{N}}\sum_{i=1}^{N}\begin{pmatrix}u_i^{(j_3)}\\ u_i^{(j_4)}\end{pmatrix}\begin{pmatrix}\overline{e}_i\\ u_i^{(j_1)}\\ u_i^{(j_2)}\end{pmatrix}'\right)\begin{pmatrix}1\\ -\widetilde{\lambda}-A_j'a_{j,T}'\\ A_j'a_{j,T}'\end{pmatrix}$$

$$=\mathcal{A}_{j,T}\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\xi_i,$$

where $\mathcal{A}_{j,T}$ is $k_z\times k_\xi$ matrix which is a deterministic function of $A_{j,T}^{*}, A_j, a_{j,T}, \widetilde{\lambda}$. The exact expression for $\mathcal{A}_{j,T}$ is though obvious but complicated to write down. We have discussed before the convergence of all terms separately, it implies that $\mathcal{A}_{j,T}\Rightarrow\mathcal{A}_j$, where the limit is a deterministic function of $(\eta_1,...,\eta_4)$.

Given assumption GAUSSIANITY we have:

$$\sqrt{\frac{T}{N}}\sum_{i=1}^{N} Q_z z_i^{(j)}\epsilon_i^{(j)}\Rightarrow\mathcal{A}_j\xi.$$

Following step (4) of the 4-split algorithm, we can put all pieces together:

$$\sqrt{NT}Q^{-1}(\widehat{\lambda}_{4S}-\widetilde{\lambda})\Rightarrow(I_{k_F},0_{k_F,k_v})\left(\frac{1}{4}\sum_{j=1}^{4}\Theta_j\mathcal{A}_j\right)\xi. \tag{25}$$

As we can see, the 4-split estimator has asymptotic mixed Gaussian, that is, the limit distribution conditionally on $\eta_1,...,\eta_4$ (which is independent of ξ due to assumption ERRORS) is gaussian with mean zero and variance depending on $\eta_1,...,\eta_4$.

Denote $\widehat{\Sigma}_{IV}=\frac{1}{N}R'G^{-1}\widehat{\Sigma}_0G^{-1}R$. Below we show $\widehat{\Sigma}_{IV}$ has the following asymptotic distribution:

$$NTQ^{-1}\widehat{\Sigma}_{IV}Q^{-1}\Rightarrow(I_{k_F},0_{k_F,k_v})\left(\frac{1}{4}\sum_{j=1}^{4}\Theta_j\mathcal{A}_j\right)\Sigma_\xi\left(\frac{1}{4}\sum_{j=1}^{4}\Theta_j\mathcal{A}_j\right)'(I_{k_F},0_{k_F,k_v})'. \tag{26}$$

Statement (26) implies the statement of the Theorem 3. Indeed, equations (25) and (26) imply that

$$\widehat{\Sigma}_{IV}^{-1/2}(\widehat{\lambda}_{4S}-\widetilde{\lambda})\Rightarrow N(0,I_k)$$

where the limiting gaussian vector is independent from the limiting gaussian vector in the following convergence:

$$\sqrt{T}\Omega_F^{-1/2}(\widetilde{\lambda}-\lambda)\Rightarrow N(0,I_k).$$

The expression $\widehat{\Sigma}_{4S}^{-1/2}(\widehat{\lambda}_{4S}-\lambda)$ is the weighted sum of the expressions staying on the left-hand-side of the last two convergence with weights asymptotically independent from both limiting $N(0,I_k)$. This leads to the validity of the statement of Theorem 3.

In order to prove validity of statement (26) we first notice that

$$\sqrt{T}Q_z z_i^{(j)}\epsilon_i^{(j)}=\mathcal{A}_{j,T}\xi_i,$$

and thus

$$\frac{T}{N}\sum_{i=1}^N\begin{pmatrix}Q_z\widetilde{z}_i^{(1)}\epsilon_i^{(1)}\\ \cdots\\ Q_z\widetilde{z}_i^{(4)}\epsilon_i^{(4)}\end{pmatrix}\begin{pmatrix}Q_z\widetilde{z}_i^{(1)}\epsilon_i^{(1)}\\ \cdots\\ Q_z\widetilde{z}_i^{(4)}\epsilon_i^{(4)}\end{pmatrix}'=\begin{pmatrix}\mathcal{A}_{1,T}\\ \cdots\\ \mathcal{A}_{4,T}\end{pmatrix}\frac{1}{N}\sum_{i=1}^N\xi_i\xi_i'\begin{pmatrix}\mathcal{A}_{1,T}\\ \cdots\\ \mathcal{A}_{4,T}\end{pmatrix}'$$
$$\Rightarrow\begin{pmatrix}\mathcal{A}_1\\ \cdots\\ \mathcal{A}_4\end{pmatrix}\Sigma_\xi\begin{pmatrix}\mathcal{A}_1\\ \cdots\\ \mathcal{A}_4\end{pmatrix}'. \tag{27}$$

Let us consider an infeasible variance estimator $\widetilde{\Sigma}_{IV}$ which is constructed in the same way as $\widehat{\Sigma}_{IV}$ but uses $\epsilon_i^{(j)}$ in place of $\widehat{\epsilon}_i^{(j)}$. That is, denote

$$\widetilde{\Sigma}_0=\frac{1}{N}\sum_{i=1}^N\begin{pmatrix}\widetilde{z}_i^{(1)}\epsilon_i^{(1)}\\ \cdots\\ \widetilde{z}_i^{(4)}\epsilon_i^{(4)}\end{pmatrix}\begin{pmatrix}\widetilde{z}_i^{(1)}\epsilon_i^{(1)}\\ \cdots\\ \widetilde{z}_i^{(4)}\epsilon_i^{(4)}\end{pmatrix}',$$

and consider $\widetilde{\Sigma}_{IV}=\frac{1}{N}R'G^{-1}\widetilde{\Sigma}_0G^{-1}R$. By putting together statements (23), (24) and (27) we obtain:

$$NTQ^{-1}\widetilde{\Sigma}_{IV}Q^{-1}\Rightarrow(I_{k_F},0_{k_F,k_v})\left(\frac{1}{4}\sum_{j=1}^4\Theta_j\mathcal{A}_j\right)\Sigma_\xi\left(\frac{1}{4}\sum_{j=1}^4\Theta_j\mathcal{A}_j\right)'(I_{k_F},0_{k_F,k_v})'.$$

The only thing left to show is that the difference between $\widehat{\Sigma}_{IV}$ and $\widetilde{\Sigma}_{IV}$ is asymptotically negligible. In particular, we will show for any j and j^*

$$\frac{T}{N}\sum_{i=1}^N Q_z z_i^{(j)}z_i^{(j^*)\prime}Q_z\left[\epsilon_i^{(j)}\epsilon_i^{(j^*)}-\widehat{\epsilon}_i^{(j)}\widehat{\epsilon}_i^{(j^*)}\right]\to^p 0, \tag{28}$$

where $\widehat{\epsilon}_i^{(j)}$ are the residuals from the j-th IV regression. Indeed, this last statement implies that

$$\widehat{\Sigma}_{IV}=\widetilde{\Sigma}_{IV}(1+o_p(1)),$$

and usage of residuals in place of true errors does not have asymptotic effect on estimation of variance.

In order to prove (28) we would write down an equation analogous to equation (15):

$$y_i = (\widetilde{\lambda}', a_{j,T})x_i^{(j)} + \epsilon_i^{(j)} = \theta_j' x_i^{(j)} + \epsilon_i^{(j)}.$$

From the proof of Theorem 2 we have:

$$\sqrt{NT} Q_x^{-1}(\widehat{\theta}_j - \theta_j) = O_p(1),$$

where $\widehat{\theta}_j$ is the IV estimator obtained on Steps (2) for $j = 1$ or on Step (3) for $j = 2, 3$ or 4. The residuals for this regression are:

$$\widehat{\epsilon}_i^{(j)} = y_i - \widehat{\theta}_j' x_i^{(j)} = \epsilon_i^{(j)} - (\widehat{\theta}_j - \theta_j)' x_i^{(j)} = \epsilon_i^{(j)} - (Q_x^{-1}(\widehat{\theta}_j - \theta_j))' Q_x x_i^{(j)}.$$

The left hand expression of (28) equal to:

$$\frac{T}{N}\sum_{i=1}^{N} Q_z z_i^{(j)} z_i^{(j^*)\prime} Q_z \left(\epsilon_i^{(j)}(\widehat{\theta}_{j^*} - \theta_{j^*})' x_i^{(j^*)} + \epsilon_i^{(j^*)}(\widehat{\theta}_j - \theta_j)' x_i^{(j)} - (\widehat{\theta}_{j^*} - \theta_{j^*})' x_i^{(j^*)}(\widehat{\theta}_j - \theta_j)' x_i^{(j)}\right). \tag{29}$$

The expression in (29) equals to three sums. We can show that each of these sums is asymptotically negligible. For example, consider the first of the three sums:

$$\frac{1}{N^{3/2}}\sum_{i=1}^{N}\left(\sqrt{T} Q_z z_i^{(j)}\epsilon_i^{(j)}\right)\left(Q_z z_i^{(j^*)}\right)'\left(Q_x x_i^{(j^*)}\right)'\left(\sqrt{NT} Q_x^{-1}(\widehat{\theta}_{j^*} - \theta_{j^*})\right)$$
$$= \frac{1}{N^{3/2}}\sum_{i=1}^{N}\mathcal{A}_{j,T}\xi_i\left(Q_z z_i^{(j^*)}\right)'\left(Q_x x_i^{(j^*)}\right)'\left(\sqrt{NT} Q_x^{-1}(\widehat{\theta}_{j^*} - \theta_{j^*})\right).$$

Notice that $\sqrt{NT} Q_x^{-1}(\widehat{\theta}_{j^*} - \theta_{j^*}) = O_p(1)$. As before $Q_z z_i^{(j)} = O_p(1)\gamma_i + O_p(1)\sqrt{T}(u_i^{(j_3)}, u_i^{(j_4)})'$, while $Q_x x_i^{(j)} = O_p(1)\gamma_i + O_p(1)\sqrt{T}(u_i^{(j_1)}, u_i^{(j_2)})$, where all mentioned $O_p(1)$ terms are not indexed by i. Thus, the sum

$$\frac{1}{N^{3/2}}\sum_{i=1}^{N}\mathcal{A}_{j,T}\xi_i\left(Q_z z_i^{(j^*)}\right)'\left(Q_x x_i^{(j^*)}\right)' = O_p(1)\frac{1}{N^{3/2}}\sum_{i=1}^{N}\xi_i\xi_i' + O_p(1)\frac{1}{N^{3/2}}\sum_{i=1}^{N}\xi_i \otimes (\gamma_i\gamma_i').$$

By Assumption GAUSSIANITY, $\frac{1}{N^{3/2}}\sum_{i=1}^{N}\xi_i\xi_i' \to^p 0$ and thus

$$\frac{1}{N^{3/2}}\|\sum_{i=1}^{N}\xi_i \otimes (\gamma_i\gamma_i')\| \le \frac{1}{N^{3/2}}\sqrt{\sum_{i=1}^{N}\|\xi_i\|^2}\sqrt{\sum_{i=1}^{N}\|\gamma_i\|^4} \to^p 0.$$

This gives the asymptotic negligibility of the first sum in expression (29), the negligibility of the other two sums is proved in a similar manner.

This ends the proof of Theorem 3. □

10 Appendix B: Gaussianity and Central Limit Theorems

This Appendix contains some preliminary statements, including Central Limit Theorem for quadratic forms, and the proof of Lemmas 2 and 3.

10.1 Preliminary Results

The idea coming from the proof of the CLT for quadratic forms by de Jong (1987) who refers to a martingale CLT by Heyde and Brown (1970) and suggests to apply it for $\delta = 1$:

Theorem 4 *(reformulation of the CLT by Heyde and Brown (1970)) Let* $(z_t, \mathcal{F}_t)$, $t = 1, .., T$ *, be a martingale difference sequence with* $\sigma_T^2 = var\left(\sum_{t=1}^T z_t\right)$*. If the following two conditions hold for some* $\delta \in (0, 1]$ *as* $T \to \infty$,

(1) $\frac{1}{\sigma_T^{2+\delta}} \sum_{t=1}^T E|z_t|^{2+2\delta} \to 0$,

(2) $E\left|\frac{\sum_{t=1}^T z_t^2}{\sigma_T^2} - 1\right|^{1+\delta} \to 0$,

then $\frac{1}{\sigma_T} \sum_{t=1}^T z_t \Rightarrow N(0, 1)$.

The following Theorem is a reformulation of this result for the vector case.

Theorem 5 *Let* $(Z_{t,T}, \mathcal{F}_{t,T})$, $t = 1, .., T$, *be a martingale difference sequence of* $r \times 1$ *random vectors for each* T *with,* $\Sigma_T = var\left(\sum_{t=1}^T Z_{t,T}\right)$*. If the following two conditions hold as* $T \to \infty$,

(1) $\frac{1}{(\min eval(\Sigma_T))^2} \sum_{t=1}^T E\|Z_{t,T}\|^4 \to 0$,

(2) $\frac{1}{(\min eval(\Sigma_T))^2} E\left\|\sum_{t=1}^T Z_{t,T} Z_{t,T}' - \Sigma_T\right\|^2 \to 0$,

then $\Sigma_T^{-1/2} \sum_{t=1}^T Z_{t,T} \Rightarrow N(0, I_r)$.

The following useful lemma is a new CLT for quadratic forms. Henceforth, the quantities W_{st} are implicitly indexed by the sample sizes N, T which are omitted to reduce clutter; in full notation they are indexed as $W_{st,N,T}$.

Lemma 5 *Let* $W_{st} = W_{st}(X_{st}, e_s, e_t)$ *be a set of random vectors defined for all* $s > t$*, where* $s, t \in \{1, ..., T\}$*, such that* $W_{st}(\cdot, \cdot, \cdot)$ *is a deterministic function,* X_{st} *is random vector measurable with respect to* σ*-algebra* $\mathcal{F}$*, and all* e_t *are independent from each other conditional on* $\mathcal{F}$*. Assume that*

$$E(W_{st}|\mathcal{F}, e_t) = 0 \text{ and } E(W_{st}|\mathcal{F}, e_s) = 0. \tag{30}$$

Define $W(T, N) = \sum_{s=1}^{T} \sum_{t<s} W_{st}$ *and* $\Sigma_W(T, N) = var(W(T, N))$*. Assume the following statements hold as* $T, N \to \infty$*:*

(i) $\Sigma_W(T, N) \to \Sigma_W$*, where* Σ_W *is a full rank matrix;*

(ii) $T^4 \sup_{s,t} E\|W_{st}\|^4 < C$*;*

(iii)

$$T^4 \max_{t_1<s_1, t_2<s_2, t_1\neq t_2, s_1\neq s_2} \left|cov(\|W_{s_1t_1}\|^2, \|W_{s_2t_2}\|^2)\right| \to 0;$$

(iv)

$$T^4 \max_{s_1\neq s_2, t_1\neq t_2, \max\{t_1,t_2\}<\min\{s_1,s_2\}} \left|EW'_{s_1t_2}W_{s_2t_1}W'_{s_2t_2}W_{s_1t_1}\right| \to 0.$$

Then,

$$W(T, N) \Rightarrow N(0, \Sigma_W)$$

as $T, N \to \infty$*.*

Lemma 6 *Let* $W_{st} = W_{st}(X_{st}, e_s, e_t)$ *be a set of random vectors defined for all* $s > t$*, where* $s, t \in \{1, ..., T\}$*, and satisfying all conditions of Lemma 5. Let* $V_s = V_s(X_s, e_s)$ *be a random vector defined for all* $s \in \{1, ..., T\}$ *such that* $V_s(\cdot, \cdot)$ *is a deterministic function,* X_s *is random vector measurable with respect to* σ*-algebra* $\mathcal{F}$*, and* $E(V_s|\mathcal{F}) = 0$*. Define* $W(T, N) = \sum_{s=1}^{T} \sum_{t<s} W_{st}$*,* $V(T, N) = \sum_{s=1}^{T} V_s$ *,* $\Sigma_W(T, N) = var(W(T, N))$ *and* $\Sigma_V(T, N) = var(V(T, N))$*. Assume the following statements hold as* $T, N \to \infty$*:*

(a) $\Sigma_V(T, N) \to \Sigma_V$*, where* Σ_V *is a full rank matrix;*

(b) $T \sup_s E\|V_s\|^4 \to 0$*;*

(c)

$$E\left\|\sum_{s=1}^{T} V_s V_s' - \Sigma_V(T, N)\right\|^2 \to 0;$$

(d)

$$T^3 \max_{s_1,s_2,t<\min\{s_1,s_2\}} \left\| W_{s_1t} V'_{s_1} V_{s_2} W'_{s_2t} \right\| \to 0.$$

Then,

$$\left(\begin{array}{c} V(T,N) \\ W(T,N) \end{array} \right) \Rightarrow N \left(\binom{\mathbf{0}}{\mathbf{0}}, \left(\begin{array}{cc} \Sigma_V & \mathbf{0} \\ \mathbf{0} & \Sigma_W \end{array} \right) \right)$$

as $T, N \to \infty$.

10.2 Proofs of auxiliary statements

Proof of Theorem 5 Indeed the statement of Theorem 5 holds if for any non-random $r \times 1$ vector λ we have

$$(\lambda' \Sigma_T \lambda)^{-1/2} \sum_{t=1}^{T} \lambda' Z_{t,T} \Rightarrow N(0,1).$$

Let us define $z_t = \lambda' Z_{t,T}$ and $\sigma_T^2 = var\left(\sum_{t=1}^T \lambda' Z_{t,T}\right) = \lambda' \Sigma_T \lambda$. Let us check that all conditions of Theorem 4 are satisfied for $\delta = 1$. Indeed,

$$\frac{1}{\sigma_T^4} \sum_{t=1}^{T} E|z_t|^4 = \frac{1}{(\lambda'\Sigma_T\lambda)^2} \sum_{t=1}^{T} E|\lambda' Z_{t,T}|^4 \le \frac{1}{(\|\lambda\|^2 \min eval(\Sigma_T))^2} \sum_{t=1}^{T} E\|\lambda\|^4 \|Z_{t,T}\|^4 \to 0,$$

and

$$E\left|\frac{\sum_{t=1}^T z_t^2}{\sigma_T^2} - 1\right|^2 = E\left|\frac{\sum_{t=1}^T (\lambda' Z_{t,T})^2}{\lambda'\Sigma_T\lambda} - 1\right|^2 = \frac{1}{(\lambda'\Sigma_T\lambda)^2} E\left|\lambda' \left(\sum_{t=1}^{T} Z_{t,T} Z'_{t,T} - \Sigma_T\right)\lambda\right|^2$$

$$\le \frac{1}{(\|\lambda\|^2 \min eval(\Sigma_T))^2} \|\lambda\|^4 E\left\|\sum_{t=1}^{T} Z_{t,T} Z'_{t,T} - \Sigma_T\right\|^2 \to 0.$$

This finishes the proof of this theorem. □

Proof of Lemma 5 We call W_{st} *clean* if

$$EW_{s_1t_1} \otimes W_{s_2t_2} .... \otimes W_{s_kt_k} = \mathbf{0}$$

when at least one index among $\{s_1, t_1, ..., s_k, t_k\}$ has value occurring only once. Here $\mathbf{0}$ is a zero vector of a proper dimensions. The functional form of W_{st} and ,the condition stated in (30) guarantee that in our case W_{st} is clean. Indeed, if, for example, the index s_1 occurred only once, then

$$\begin{aligned} EW_{s_1t_1} \otimes W_{s_2t_2} \otimes .... \otimes W_{s_kt_k} &= EE(W_{s_1t_1} \otimes W_{s_2t_2} \otimes .... \otimes W_{s_kt_k} | \mathcal{F}, e_{t_1}, e_{s_2}, e_{t_2}, ..., e_{t_k}) \\ &= E\left[E(W_{s_1t_1} | \mathcal{F}, e_{t_1}, e_{s_2}, e_{t_2}, ..., e_{t_k}) \otimes W_{s_2t_2} \otimes .... \otimes W_{s_kt_k}\right] \\ &= E\left[E(W_{s_1t_1} | \mathcal{F}, e_{t_1}) \otimes W_{s_2t_2} \otimes .... \otimes W_{s_kt_k}\right] = 0. \end{aligned}$$

Now,

$$W(T,N) = \sum_{s=1}^{T} \sum_{t<s} W_{st} = \sum_{s=1}^{T} Z_{s,T},$$

where $Z_{s,T} = \sum_{t<s} W_{st}$. We denote by $\mathcal{F}_s$ the σ-algebra generated by $\mathcal{F}$ (all factors) and e_t for all $t < s$. Then, $(Z_{s,T}, \mathcal{F}_s)$ is a martingale difference sequence. Below we check that all conditions of Theorem 5 are satisfied.

Assumption (i) implies that $\min eval(\Sigma(T,N)) \to C > 0$. Now let us check condition 1 of Theorem 5:

$$\begin{aligned}
E\|Z_{s,T}\|^4 &= E\left\|\sum_{t<s} W_{st}\right\|^4 \\
&= E\left[\left(\sum_{t_1<s} W_{st_1}\right)'\left(\sum_{t_2<s} W_{st_2}\right)\left(\sum_{t_3<s} W_{st_3}\right)'\left(\sum_{t_4<s} W_{st_4}\right)\right] \\
&\leq \sum_{t<s} E\|W_{st}\|^4 + C\sum_{t_1<s}\sum_{t_2<s, t_2\neq t_1} E\left(\|W_{st_1}\|^2\|W_{st_2}\|^2\right).
\end{aligned}$$

The last statement follows from the fact that W_{st} is clean, and non-zero summands are only those where either $t_1 = t_2 = t_3 = t_4$ or the set $\{t_1, t_2, t_3, t_4\}$ consists of two index pairs. We also notice that $E\left(\|W_{st_1}\|^2\|W_{st_2}\|^2\right) \leq \frac{1}{2}\left(E\left(\|W_{st_1}\|^4\right) + E\left(\|W_{st_2}\|^4\right)\right) \leq \sup_{s,t} E\|W_{st}\|^4 < CT^{-4}$. Hence, we have $E\|Z_{s,T}\|^4 \leq CT^{-2}$. Thus, $\sum_{s=1}^{T} E\|Z_{s,T}\|^4 \leq CT^{-1}$, implying that condition 1 of Theorem 5 holds.

Now let us turn to condition 2. First, consider

$$\Sigma(T,N) = var(W(T,N)) = var\left(\sum_{s=1}^{T}\sum_{t<s} W_{st}\right) = \sum_{s=1}^{T}\sum_{t<s} var(W_{st}),$$

the last equality holding because W_{st} is clean. Next,

$$\begin{aligned}
&E\left\|\sum_{s=1}^{T} Z_{s,T}Z_{s,T}' - \Sigma(T,N)\right\|_2^2 \\
&= E\left\|\sum_{s=1}^{T}\left(\sum_{t_1<s} W_{st_1}\right)\left(\sum_{t_2<s} W_{st_2}\right)' - \Sigma(T,N)\right\|_2^2 \\
&= E\left\|\sum_{s=1}^{T}\sum_{t<s}(W_{st}W_{st}' - EW_{st}W_{st}') + \sum_{s=1}^{T}\sum_{t_1\neq t_2<s} W_{st_1}W_{st_2}'\right\|_2^2 \\
&= E\left\|\sum_{s=1}^{T}\sum_{t<s}(W_{st}W_{st}' - EW_{st}W_{st}')\right\|_2^2 + E\left\|\sum_{s=1}^{T}\sum_{t_1\neq t_2<s} W_{st_1}W_{st_2}'\right\|_2^2. \qquad (31)
\end{aligned}$$

The last equality again obtains because of the clean form, as the expectation of the Frobenius norm equals the trace of the sums of different products of four terms, and any such product that contains two of the same indexes t and two different indexes $t_1 \neq t_2$ has a zero expectation. Now,

$$
\begin{aligned}
& E\left\|\sum_{s=1}^{T}\sum_{t<s}(W_{st}W_{st}' - EW_{st}W_{st}')\right\|_2^2 \\
&= \sum_{s_1=1}^{T}\sum_{t_1<s_1}\sum_{s_2=1}^{T}\sum_{t_2<s_2} E\mathrm{tr}\left[(W_{s_1t_1}W_{s_1t_1}' - EW_{s_1t_1}W_{s_1t_1}')'(W_{s_2t_2}W_{s_2t_2}' - EW_{s_2t_2}W_{s_2t_2}')\right].
\end{aligned}
$$

Notice that $\mathrm{tr}(W_{st}W_{st}') = \|W_{st}\|^2$, while

$$
\begin{aligned}
\mathrm{tr}(W_{s_1t_1}W_{s_1t_1}'W_{s_2t_2}W_{s_2t_2}') &= \mathrm{tr}(W_{s_1t_1}'W_{s_2t_2}W_{s_2t_2}'W_{s_1t_1}) \\
&= \|W_{s_1t_1}'W_{s_2t_2}\|^2 \leq \|W_{s_1t_1}\|^2\|W_{s_2t_2}\|^2.
\end{aligned}
$$

Thus,

$$
E\left\|\sum_{s=1}^{T}\sum_{t<s}(W_{st}W_{st}' - EW_{st}W_{st}')\right\|_2^2 \leq \sum_{s_1=1}^{T}\sum_{t_1<s_1}\sum_{s_2=1}^{T}\sum_{t_2<s_2}\left|cov(\|W_{s_1t_1}\|^2, \|W_{s_2t_2}\|^2)\right|.
$$

We divide the last summation into a summation where all indexes are different, $s_1 \neq s_2, t_1 \neq t_2$, and a summation where some of indexes $\{s_1, s_2, t_1, t_2\}$ appear twice. In the first summation, there are at most T^4 such summands, each of them being $o(T^{-4})$ as $N, T \to \infty$ according to Assumption (iii) of the Lemma. In the second summation, there are at most CT^3 such summands, each of them being less than $C\sup_{s,t} E\|W_{st}\|^4 < CT^{-4}$. Thus, we obtain

$$
E\left\|\sum_{s=1}^{T}\sum_{t<s}(W_{st}W_{st}' - EW_{st}W_{st}')\right\|_2^2 \to 0 \text{ as } N, T \to \infty. \tag{32}
$$

Now manipulate the second term in (31):

$$
\begin{aligned}
E\left\|\sum_{s=1}^{T}\sum_{t_1\neq t_2<s} W_{st_1}W_{st_2}'\right\|_2^2 &= \sum_{s_1=1}^{T}\sum_{t_1\neq t_2<s_1}\sum_{s_2=1}^{T}\sum_{t_3\neq t_4<s_2} E\mathrm{tr}[W_{s_1t_1}W_{s_1t_2}'W_{s_2t_3}W_{s_2t_4}'] \\
&= C\sum_{s_1=1}^{T}\sum_{s_2=1}^{T}\sum_{t_1\neq t_2, \max\{t_1,t_2\}<\min\{s_1,s_2\}} E\mathrm{tr}[W_{s_1t_1}W_{s_1t_2}'W_{s_2t_1}W_{s_2t_2}'],
\end{aligned}
$$

the last equality holding because W_{st} is clean. The last summation can be divided into the category when $s_1 \neq s_2$, and there are at most CT^4 such summands, each being asymptotically $o(T^{-4})$ according to assumption (iv) of Lemma; and the category when $s_1 = s_2$, and there are at most CT^3 such summands, each being less than $C\sup_{s,t} E\|W_{st}\|^4 < CT^{-4}$. Thus,

$$E\left\|\sum_{s=1}^{T}\sum_{t_1\neq t_2<s} W_{st_1}W'_{st_2}\right\|_2^2 \to 0. \tag{33}$$

Putting statements (31), (32) and (33) together we obtain that condition 2 of Theorem 5 is satisfied. Thus, the CLT holds. □

Proof of Lemma 6. Let us define $Z_s = (V'_s, \sum_{t<s} W'_{st})'$, and let $\mathcal{F}_s$ be defined as in the proof of Lemma 5. We will show that all conditions of Theorem 5 are satisfied. Notice that

$$E[V_sW'_{st}] = EE[V_sW'_{st}|\mathcal{F}, e_s] = E(V_sE[W'_{st}|\mathcal{F}, e_s]) = 0.$$

Thus,

$$\Sigma_T = var\left(\sum_{s=1}^{T} Z_s\right) = \begin{pmatrix} \Sigma_V(T,N) & \mathbf{0} \\ \mathbf{0} & \Sigma_W(T,N) \end{pmatrix} \to \begin{pmatrix} \Sigma_V & \mathbf{0} \\ \mathbf{0} & \Sigma_W \end{pmatrix}.$$

The right-hand-side is a full rank matrix by Assumption (i) of Lemma 5 and Assumption (a) of Lemma 6. Thus, the minimal eigenvalue of Σ_T is separated away from zero for large N and T. Now,

$$\sum_{s=1}^{T} E\|Z_s\|^4 \le C\sum_{s=1}^{T} E\|V_s\|^4 + C\sum_{s=1}^{T} E\left\|\sum_{t<s} W_{st}\right\|^4.$$

The first term here is bounded by $T\sup_s E\|V_s\|^4$ which goes to zero by assumption (b) of Lemma 6, while convergence to zero of the second sum has been already shown in the proof of Lemma 5. Thus, condition 1 of Theorem 5 holds. Next,

$$\begin{aligned} E\left\|\sum_{s=1}^{T} Z_sZ'_s - \Sigma_T\right\|^2 &\le E\left\|\sum_{s=1}^{T} Z_sZ'_s - \Sigma_T\right\|_2^2 \\ &= E\left\|\sum_{s=1}^{T} V_sV'_s - \Sigma_V(T,N)\right\|_2^2 + 2E\left\|\sum_{s=1}^{T}\left(\sum_{t<s} W_{st}\right)V'_s\right\|_2^2 \\ &+ E\left\|\sum_{s=1}^{T}\left(\sum_{t<s} W_{st}\right)\left(\sum_{t<s} W_{st}\right)' - \Sigma_W(T,N)\right\|_2^2. \end{aligned}$$

Here we use that the Frobenius norm of a matrix equals to the sum of squares of all elements and can be decomposed into sums over four blocks of the matrix. Assumption

(c) guarantees that

$$E\left\|\sum_{s=1}^{T}V_sV_s'-\Sigma_V(T,N)\right\|_2^2\leq CE\left\|\sum_{s=1}^{T}V_sV_s'-\Sigma_V(T,N)\right\|^2\to 0.$$

In the proof of Lemma 5 we showed that

$$E\left\|\sum_{s=1}^{T}\left(\sum_{t<s}W_{st}\right)\left(\sum_{t<s}W_{st}\right)'-\Sigma_W(T,N)\right\|_2^2\to 0.$$

Finally,

$$\begin{aligned}E\left\|\sum_{s=1}^{T}\left(\sum_{t<s}W_{st}\right)V_s'\right\|_2^2&=\sum_{s_1=1}^{T}\sum_{t_1<s_1}\sum_{s_2=1}^{T}\sum_{t_2<s_2}\operatorname{tr}\left(E\left(W_{s_1t_1}V_{s_1}'V_{s_2}W_{s_2t_2}'\right)\right)\\&=\sum_{s_1=1}^{T}\sum_{s_2=1}^{T}\sum_{t<\min\{s_1,s_2\}}\operatorname{tr}\left(E\left(W_{s_1t}V_{s_1}'V_{s_2}W_{s_2t}'\right)\right)\\&\leq CT^3\max_{s_1,s_2,t<\min\{s_1,s_2\}}\left\|EW_{s_1t}V_{s_1}'V_{s_2}W_{s_2t}'\right\|\to 0.\end{aligned}$$

Here we used that $E\left(W_{s_1t_1}V_{s_1}'V_{s_2}W_{s_2t_2}'\right)=0$ if $t_1\neq t_2$. To conclude, condition 2 of Theorem 5 also holds. $\square$

10.3 Gaussianity assumption in different examples.

Let us write all ξ_i terms from Assumption GAUSSIANITY in the form used in Lemma 6. First, consider the following desired gaussianity statements:

$$\frac{\sqrt{T}}{\sqrt{N}}\sum_{i=1}^{N}\gamma_i\overline{e}_i=\frac{1}{\sqrt{T}}\sum_{s=1}^{T}\frac{\gamma'e_s}{\sqrt{N}};$$

$$\frac{\sqrt{T}}{\sqrt{N}}\sum_{i=1}^{N}\gamma_i\otimes u_i^{(j)}=\frac{1}{\sqrt{T}}\sum_{s=1}^{T}\left(\frac{\gamma'e_s}{\sqrt{N}}\right)\otimes\left(\Sigma_F^{-1}\widetilde{F}_s^{(j)}\right)\mathbb{I}\{s\in T_j\}.$$

Collect all terms of interest into the vector

$$V_s\equiv\begin{pmatrix}V_s^0\\ \left\{V_s^{(j)}\right\}_{j=1}^4\end{pmatrix}=\frac{1}{\sqrt{T}}\frac{\gamma'e_s}{\sqrt{N}}\otimes v_s,$$

where

$$v_s=\begin{pmatrix}1\\ \left\{v_s^{(j)}\right\}_{j=1}^4\end{pmatrix}$$

and $v_s^{(j)} = \left(\Sigma_F^{-1}\widetilde{F}_s^{(j)}\right)\mathbb{I}\{s \in T_j\}$. Note that $v_s^{(j)}v_s^{(j^*)\prime} = \mathbf{0}$ whenever $j \neq j^*$. Also note that v_s is measurable with respect to the σ-algebra $\mathcal{F}$, and $\|v_s\| \leq C\left(\|\widetilde{F}_s^{(j_s)}\| + 1\right)$, where $j_s \in \{1, ..., 4\}$ is the index of the sub-sample to which s belongs.

Now assume that $j^* > j$ and consider the following sum:

$$vec\left(\frac{T}{\sqrt{N}}\sum_{i=1}^{N} u_i^{(j^*)}u_i^{(j)\prime}\right) = \frac{1}{T\sqrt{N}}\sum_{i=1}^{N}\sum_{t\in T_j}\sum_{s\in T_{j^*}} vec\left(\Sigma_F^{-1}\widetilde{F}_s^{(j^*)}\widetilde{F}_t^{(j)\prime}\Sigma_F^{-1}e_{it}e_{is}\right)$$
$$= \sum_{s\in T_{j^*}}\sum_{t\in T_j}\frac{1}{T}\left(\Sigma_F^{-1}\widetilde{F}_s^{(j^*)}\right)\otimes\left(\Sigma_F^{-1}\widetilde{F}_t^{(j)}\right)\frac{e_t'e_s}{\sqrt{N}}.$$

Let us define for each $1 \leq t < s \leq T$,

$$W_{st}^{(j,j^*)} = \frac{1}{T}\left(\Sigma_F^{-1}\widetilde{F}_s^{(j^*)}\right)\otimes\left(\Sigma_F^{-1}\widetilde{F}_t^{(j)}\right)\mathbb{I}\{t \in T_j, s \in T_{j^*}\}\frac{e_t'e_s}{\sqrt{N}}.$$

Consider the following sum:

$$\frac{T}{\sqrt{N}}\sum_{i=1}^{N}\bar{e}_i u_i^{(j)} = \frac{1}{T\sqrt{N}}\sum_{i=1}^{N}\sum_{t_1=1}^{T}\sum_{t_2\in T_j}\Sigma_F^{-1}\widetilde{F}_{t_2}^{(j)}e_{it_1}e_{it_2}$$
$$= \frac{1}{T\sqrt{N}}\sum_{i=1}^{N}\sum_{t\in T_j}\Sigma_F^{-1}\widetilde{F}_t^{(j)}e_{it}^2 + \frac{1}{T\sqrt{N}}\sum_{i=1}^{N}\sum_{t_1=1}^{T}\sum_{t_2\in T_j, t_1\neq t_2}\Sigma_F^{-1}\widetilde{F}_{t_2}^{(j)}e_{it_1}e_{it_2}.$$

Assumption ERRORS (iii) guarantees that $\left(T\sqrt{N}\right)^{-1}\sum_{i=1}^{N}\sum_{t\in T_j}\Sigma_F^{-1}\widetilde{F}_t^{(j)}e_{it}^2 = o_p(1)$. Thus, we are only interested in gaussianity of the second sum. For those summands when $t_1 < t_2$, we denote $t_1 = t$ and $t_2 = s$, while when the opposite happens, we denote $t_1 = s$ and $t_2 = t$. Then, we obtain the following:

$$\frac{1}{T\sqrt{N}}\sum_{i=1}^{N}\sum_{t_1=1}^{T}\sum_{t_2\in T_j, t_1\neq t_2}\Sigma_F^{-1}\widetilde{F}_{t_2}^{(j)}e_{it_1}e_{it_2} = \sum_{s=1}^{T}\sum_{t<s}\frac{1}{T}\Sigma_F^{-1}(\widetilde{F}_s^{(j)}\mathbb{I}\{s \in T_j\}+\widetilde{F}_t^{(j)}\mathbb{I}\{t \in T_j\})\frac{e_t'e_s}{\sqrt{N}}.$$

Denote

$$W_{st}^{(j)} = \frac{1}{T}\Sigma_F^{-1}\left(\widetilde{F}_s^{(j)}\mathbb{I}\{s \in T_j\} + \widetilde{F}_t^{(j)}\mathbb{I}\{t \in T_j\}\right)\frac{e_t'e_s}{\sqrt{N}}.$$

Now we can stack all vectors $W_{st}^{(j,j^*)}$ on top of each other for different combinations of two indexes $j < j^*$, and below them, all $W_{st}^{(j)}$ for all value of a single index j. Call the resulting vector $W_{st} = (W_{st}^{(1,2)\prime}, W_{st}^{(1,3)\prime}, ..., W_{st}^{(1)}, ..)'$. Notice that

$$W_{st} = \frac{1}{T}w_{st}\frac{e_t'e_s}{\sqrt{N}},$$

where w_{st} is measurable with respect to $\mathcal{F}$. We also have $\|w_{st}\| \leq C(\|\widetilde{F}_s^{(j_s)}\|+1)(\|\widetilde{F}_t^{(j_t)}\|+1)$, where $j_s \in \{1, ..., 4\}$ is a sub-sample number to which index s belongs.

Proof of Lemma 2. In order to apply Lemma 6 we check assumptions (i)-(iv) of Lemma 5 and assumptions (a)-(d) of Lemma 6.

Let us check condition (i) of Lemma 5 for $W(T,N) = \sum_{s=1}^{T} \sum_{t<s} W_{st}$. Indeed,

$$\Sigma_W(T,N) \equiv \sum_{s=1}^{T} \sum_{t<s} var(W_{st}) = \frac{1}{T^2} \sum_{s=1}^{T} \sum_{t<s} E\left[w_{st} w_{st}' E\left(\left(\frac{e_t' e_s}{\sqrt{N}} \right)^2 | \mathcal{F} \right) \right].$$

Notice that $(e_t' e_s)^2 = \text{tr}\left((e_s' e_t)(e_t' e_s)\right) = \text{tr}\left((e_t e_t')(e_s e_s')\right)$, and hence, given the conditional independence assumption,

$$E\left(\left(\frac{e_t' e_s}{\sqrt{N}} \right)^2 | \mathcal{F} \right) = \frac{1}{N} \text{tr}\left[E(e_t e_t' | \mathcal{F}) E(e_s e_s' | \mathcal{F}) \right].$$

Recall that $e_t = \pi w_t + \eta_t$, where π is $N \times k_w$ matrix of π_i's, and $\eta_t = (\eta_{1t}, ..., \eta_{Nt})'$. We also use notation $\Sigma_\eta = E\eta_t \eta_t' = \text{dg}\{\sigma_i^2\}_{i=1}^N$. Then,

$$\begin{aligned} E\left(\left(\frac{e_t' e_s}{\sqrt{N}} \right)^2 | \mathcal{F} \right) &= \frac{1}{N} \text{tr}\left((\pi E(w_t w_t' | \mathcal{F}) \pi' + \Sigma_\eta)(\pi E(w_s w_s' | \mathcal{F}) \pi' + \Sigma_\eta) \right) \\ &= \frac{1}{N} \sum_{i=1}^{N} \sigma_i^4 + \Delta_{N,T}, \end{aligned}$$

where

$$\Delta_{N,T} \leq \frac{C}{N} E((\|w_t\|^2 + 1)(\|w_s\|^2 + 1) | \mathcal{F}).$$

Indeed, $\Delta_{N,T}$ has three terms each of which is easy to bound. For example,

$$\begin{aligned} \frac{1}{N} \text{tr}\left(\Sigma_\eta \pi E(w_s w_s' | \mathcal{F}) \pi' \right) &\leq \frac{1}{N} \max_{1 \leq i \leq N} \sigma_i^2 \cdot \text{tr}\left(E(w_s w_s' | \mathcal{F}) \pi' \pi \right) \\ &\leq \frac{1}{N} \max_{1 \leq i \leq N} \sigma_i^2 \cdot \max eval(\pi' \pi) \cdot E(\|w_s\|^2 | \mathcal{F}), \end{aligned}$$

as we assumed that $\max_{1 \leq i \leq N} \sigma_i^2 < C$ and that $\pi' \pi \to \Gamma_\pi$.

Since $\|w_{st}\| \leq C \left(\|\widetilde{F}_s^{(j_s)}\| + 1 \right) \left(\|\widetilde{F}_t^{(j_t)}\| + 1 \right)$, it follows that

$$\begin{aligned} &\left\| \frac{1}{T^2} \sum_{s=1}^{T} \sum_{t<s} E\left[w_{st} w_{st}' \Delta_{N,T} \right] \right\| \\ &\leq \frac{C}{NT^2} \sum_{s=1}^{T} \sum_{t<s} E(\|\widetilde{F}_t\|^2 + 1)(\|\widetilde{F}_s\|^2 + 1)(\|w_t\|^2 + 1)(\|w_s\|^2 + 1) \leq \frac{C}{N} \to 0, \end{aligned}$$

where the last inequality is due to moment assumptions. So, we obtain that

$$\Sigma_W(T,N) = \frac{1}{T^2}\sum_{s=1}^{T}\sum_{t<s} E\left[w_{st}w_{st}'\right]\frac{1}{N}\sum_{i=1}^{N}\sigma_i^4 + o(1)$$
$$\to \sigma_4^2 \lim_{N,T\to\infty}\frac{1}{T^2}\sum_{s=1}^{T}\sum_{t<s} E\left[w_{st}w_{st}'\right] = \Sigma_W.$$

Consider now various sub-blocks of Σ_W and show that it is a full rank matrix.

First, notice that $W_{st}^{(j,j^*)}W_{st}^{(j_1,j_1^*)\prime}$ is zero matrix when $(j,j^*) \neq (j_1,j_1^*)$. In the same way, $W_{st}^{(j,j^*)}W_{st}^{(j_1)\prime}$ is zero matrix when $j_1 \notin (j,j^*)$. Hence, there are a number of zero blocks in $\lim \Sigma_W(T,N)$. For (j,j^*) block we have and

$$E\left[(\widetilde{F}_t^{(j)}\widetilde{F}_t^{(j)\prime})\otimes(\widetilde{F}_s^{(j^*)}\widetilde{F}_s^{(j^*)\prime})\right] = (E\widetilde{F}_t^{(j)}\widetilde{F}_t^{(j)\prime})\otimes(E\widetilde{F}_s^{(j^*)}\widetilde{F}_s^{(j^*)\prime})$$
$$+ E\left[(\widetilde{F}_t^{(j)}\widetilde{F}_t^{(j)\prime} - E\widetilde{F}_t^{(j)}\widetilde{F}_t^{(j)\prime})\otimes(\widetilde{F}_s^{(j^*)}\widetilde{F}_s^{(j^*)\prime} - E\widetilde{F}_s^{(j^*)}\widetilde{F}_s^{(j^*)\prime})\right].$$

In the $(j,j^*)^{\text{th}}$ block of $\Sigma_W(T,N)$, before the limit is taken, it has $T^{-2}\sum_{s\in T_{j^*}}\sum_{t\in T_j}$ of the last displayed expression. Due to summability of covariances, the double average of covariances coming from the second term becomes negligible in the limit.[6] Thus, the diagonal block corresponding to the variance of the $(j,j^*)^{\text{th}}$ terms comes only from the first term and is, in the limit,

$$\Sigma_{j,j^*}(T,N) \to \lim\left(\frac{|T_j||T_{j^*}|}{T^2}\right)\sigma_4(E\widetilde{F}_t\widetilde{F}_t')\otimes(E\widetilde{F}_t\widetilde{F}_t') = \frac{\sigma_4}{16}\Sigma_F\otimes\Sigma_F.$$

[6] For generic stationary matrix functions ϕ_1 and ϕ_2, denote $\Upsilon_k^\phi = E\left[\phi_1(\widetilde{F}_t)\otimes\phi_2(\widetilde{F}_{t+k})\right]$. If Υ_k^ϕ are summable, then

$$\left\|\frac{1}{T^2}\sum_{t\in T_j}\sum_{s\in T_{j^*}} E\left[\phi_1(\widetilde{F}_t)\otimes\phi_2(\widetilde{F}_s)\right]\right\| \le \frac{1}{T^2}\sum_{t\in T_j}\sum_{s\in T_{j^*}}\left\|\Upsilon_{|t-s|}^\phi\right\| \le \frac{1}{T^2}\sum_{t\in T_j}\sum_{k=-\infty}^{\infty}\|\Upsilon_k^\phi\| \le \frac{C}{T}\to 0.$$

In the text, this is applied for $\phi_1(x) = \phi_2(x) = xx' - Exx'$, $\phi_1(x) = \phi_2(x) = x$, and $\phi_1(x) = xx'$, $\phi_2(x) = x$.

Let us now focus on the block corresponding to the variance of the j^{th} term:

$$\frac{1}{T^2}\sum_{s=1}^{T}\sum_{t<s}E\left[\left(\widetilde{F}_s^{(j)}\mathbb{I}\{s\in T_j\}+\widetilde{F}_t^{(j)}\mathbb{I}\{t\in T_j\}\right)\left(\widetilde{F}_s^{(j)}\mathbb{I}\{s\in T_j\}+\widetilde{F}_t^{(j)}\mathbb{I}\{t\in T_j\}\right)'\right]$$

$$=\frac{1}{T^2}\sum_{s=1}^{T}\sum_{t<s}\Big(E\widetilde{F}_s^{(j)}\widetilde{F}_s^{(j)\prime}\mathbb{I}\{s\in T_j\}+E\widetilde{F}_t^{(j)}\widetilde{F}_t^{(j)\prime}\mathbb{I}\{t\in T_j\}$$
$$+(E\widetilde{F}_s^{(j)}\widetilde{F}_t^{(j)\prime}+E\widetilde{F}_t^{(j)}\widetilde{F}_s^{(j)\prime})\mathbb{I}\{s,t\in T_j\}\Big)$$

$$=\frac{1}{T^2}\sum_{t_1=1}^{T}\sum_{t_2\in T_j}E\widetilde{F}_{t_2}^{(j)}\widetilde{F}_{t_2}^{(j)\prime}+\frac{1}{T^2}\sum_{t_1\in T_j}\sum_{t_2\in T_j}E\widetilde{F}_{t_1}^{(j)}\widetilde{F}_{t_2}^{(j)\prime}.$$

Again, due to stationarity and summability of covariances, the second term is negligible, so we have, in the limit,

$$\Sigma_j(T,N)\to\lim\left(\frac{|T_j|T}{T^2}\right)\sigma_4E\left[\widetilde{F}_t\widetilde{F}_t'\right]=\frac{\sigma_4}{4}\Sigma_F.$$

By the same arguments, the j^{th} and $j^{*\text{th}}$ blocks converge to zero matrices when $j\neq j^*$,

$$\frac{1}{T^2}\sum_{s=1}^{T}\sum_{t<s}E\left[(\widetilde{F}_s^{(j^*)}\mathbb{I}\{s\in T_j^*\})(\widetilde{F}_t^{(j)}\mathbb{I}\{t\in T_j\})'\right]=\frac{1}{T^2}\sum_{s\in T_{j^*}}\sum_{t\in T_j}E\left[\widetilde{F}_t^{(j)}\widetilde{F}_s^{(j^*)\prime}\right]\to 0,$$

as well as the j^{th} and $(j,j^*)^{\text{th}}$ blocks,

$$\frac{1}{T^2}\sum_{s=1}^{T}\sum_{t<s}E\left[\left(\widetilde{F}_t^{(j)}\otimes\widetilde{F}_s^{(j^*)}\right)\mathbb{I}\{s\in T_{j^*},t\in T_j\}\widetilde{F}_t^{(j)\prime}\right]=\frac{1}{T^2}\sum_{s\in T_{j^*}}\sum_{t\in T_j}E\left[(\widetilde{F}_t^{(j)}\widetilde{F}_t^{(j)\prime})\otimes\widetilde{F}_s^{(j^*)}\right]\to 0.$$

To summarize, we have shown that Σ_W is a block diagonal matrix, with each block being a full-rank matrix. Thus, condition (i) of Lemma 5 is satisfied.

Let us check condition (ii). Notice that

$$\frac{e_t'e_s}{\sqrt{N}}=\frac{1}{\sqrt{N}}w_t'\pi'\pi w_s+\frac{1}{\sqrt{N}}w_t'\pi'\eta_s+\frac{1}{\sqrt{N}}w_s'\pi'\eta_t+\frac{1}{\sqrt{N}}\eta_t'\eta_s.$$

Using the Marcinkiewicz–Zygmund inequality with $p=2$ applied twice we notice that in order to bound $E\left(\left(e_t'e_s/\sqrt{N}\right)^4|\mathcal{F}\right)$ from above it is enough to bound 4^{th} moment of each summand. Using serial and cross-sectional conditional independence of η's as well as their conditional independence from w's, we obtain

$$E\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\eta_{it}\eta_{is}\right)^4=\frac{1}{N^2}\sum_{i=1}^{N}E(\eta_{it}\eta_{is})^4+C\frac{1}{N^2}\sum_{i_1\neq i_2}E(\eta_{i_1t}^2\eta_{i_1s}^2\eta_{i_2t}^2\eta_{i_2s}^2)\leq C,$$

$$E\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\pi_i\eta_{is}\right)^4 \leq \frac{1}{N^2}\sum_{i=1}^{N}E(\pi_i\eta_{is})^4 + C\frac{1}{N^2}\sum_{i_1\neq i_2}E(\pi_{i_1}^2\eta_{i_1s}^2\pi_{i_2}^2\eta_{i_2s}^2) \leq \frac{C}{N^2},$$

as $\max_s E|\eta_{is}|^8 < C$ and $\sum_i \pi_i^4 \leq (\sum_i \pi_i^2)^2 \leq C$. Hence,

$$E\left(\left(\frac{e_t'e_s}{\sqrt{N}}\right)^4|\mathcal{F}\right) \leq \frac{C}{N^2}E(\|w_t\|^4\|w_s\|^4|\mathcal{F}) + \frac{C}{N^2}(E(\|w_t\|^4|\mathcal{F}) + E(\|w_s\|^4|\mathcal{F})) + C.$$

Finally,

$$\begin{aligned}T^4E\|W_{st}\|^4 &\leq E\left\|w_{st}\frac{e_t'e_s}{\sqrt{N}}\right\|^4 \leq E\left((\|\widetilde{F}_s^{(j_s)}\|+1)^4(\|\widetilde{F}_t^{(j_t)}\|+1)^4E\left(\left(\frac{e_t'e_s}{\sqrt{N}}\right)^4|\mathcal{F}\right)\right)\\ &\leq CE\left((\|\widetilde{F}_s^{(j_s)}\|+1)^4(\|\widetilde{F}_t^{(j_t)}\|+1)^4(\|w_t\|^4+1)(\|w_s\|^4+1)\right)\\ &\leq C\max_t E\left[(\|F_t\|^8+1)(\|w_t\|^8+1)\right] < C.\end{aligned}$$

Thus, condition (ii) of Lemma 5 holds.

Let us check condition (iii). By the analysis of (co)variance formula, for two random variables ξ_1 and ξ_2 and σ-algebra $\mathcal{A}$,

$$cov(\xi_1,\xi_2) = E\left[cov(\xi_1,\xi_2|\mathcal{A})\right] + cov(E(\xi_1|\mathcal{A}),E(\xi_2|\mathcal{A})).$$

Assume that $s_1 \neq s_2$, $t_1 \neq t_2$, and define $\xi_1 = \|W_{s_1t_1}\|^2 = \frac{1}{T^2}\|w_{s_1t_1}\|^2\left(\frac{e_{s_1}'e_{t_1}}{\sqrt{N}}\right)^2$, $\xi_2 = \|W_{s_2t_2}\|^2$ and $\mathcal{A} = \mathcal{F}\cup\{w_t, t=1,...,T\}$. One observation is that our assumptions are sufficient to obtain that $cov(\xi_1,\xi_2|\mathcal{A}) = 0$, because η_{it}'s are independent from each other and independent from F's and w's. In particular, this implies that $\frac{e_{s_1}'e_{t_1}}{\sqrt{N}}$, conditionally on $\mathcal{A}$, is independent from $\frac{e_{s_2}'e_{t_2}}{\sqrt{N}}$, and hence $\left(\frac{e_{s_1}'e_{t_1}}{\sqrt{N}}\right)^2$ is conditionally uncorrelated with $\left(\frac{e_{s_2}'e_{t_2}}{\sqrt{N}}\right)^2$. Let us now write

$$\begin{aligned}T^2E(\|W_{st}\|^2|\mathcal{A}) &= E\left(\|w_{st}\|^2\frac{1}{N}\left[(\pi w_s+\eta_s)'(\pi w_t+\eta_t)\right]^2|\mathcal{A}\right)\\ &= E\left(\|w_{st}\|^2\frac{1}{N}\left[w_s'\pi'\pi w_t\right]^2|\mathcal{A}\right) = \frac{1}{N}\|w_{st}\|^2\left[w_s'\pi'\pi w_t\right]^2.\end{aligned}$$

Then we have

$$\begin{aligned}&T^4\left|cov(\|W_{s_1t_1}\|^2,\|W_{s_2t_2}\|^2)\right|\\ &= \left|cov(T^2E(\|W_{s_1t_1}\|^2|\mathcal{A}),T^2E(\|W_{s_2t_2}\|^2|\mathcal{A}))\right|\\ &= \frac{1}{N^2}\left|cov\left(\|w_{s_1t_1}\|^2\left[w_{s_1}'\pi\pi'w_{t_1}\right]^2,\|w_{s_2t_2}\|^2\left[w_{s_2}'\pi\pi'w_{t_2}\right]^2\right)\right|\\ &\leq \frac{1}{N^2}var\left(\|w_{s_1t_1}\|^2\left[w_{s_1}'\pi\pi'w_{t_1}\right]^2\right)^{\frac{1}{2}}var\left(\|w_{s_2t_2}\|^2\left[w_{s_2}'\pi\pi'w_{t_2}\right]^2\right)^{\frac{1}{2}}.\end{aligned}$$

Notice that

$$
\begin{aligned}
var\left(\|w_{st}\|^2\left[w_s'\pi\pi'w_t\right]^2\right) &\le E\left(\|w_{st}\|^2\left[w_s'\pi\pi'w_t\right]^2\right)^2 \\
&\le CE\left((\|\widetilde{F}_s^{(j_s)}\|+1)^4(\|\widetilde{F}_t^{(j_t)}\|+1)^4\|w_t\|^4\|w_s\|^4\right) \\
&\le C\max_t E\left((\|\widetilde{F}_t\|^8+1)\|w_t\|^8\right) < C.
\end{aligned}
$$

Thus, condition (iii) holds.

Finally, let us check condition (iv):

$$
\begin{aligned}
&T^4E\left[W_{s_1t_2}'W_{s_2t_1}W_{s_2t_2}'W_{s_1t_1}\right] \\
&= \frac{1}{N^2}E\left(w_{s_1t_2}'w_{s_2t_1}w_{s_2t_2}'w_{s_1,t_1}E(e_{s_1}'e_{t_1}e_{t_1}'e_{s_2}e_{s_2}'e_{t_2}e_{t_2}'e_{s_1}|\mathcal{F})\right),
\end{aligned}
$$

where we used that the scalar products $e_t'e_s = e_s'e_t$ are scalars and they can be reshuffled to make two e_t with the same index stand back to back. Let us bound the $N\times N$ matrix $E(e_te_t'|\mathcal{F}) = \pi E(w_tw_t'|\mathcal{F})\pi' + \Sigma_\eta$:

$$
\begin{aligned}
\max eval(E(e_te_t'|\mathcal{F})) &\le \max eval(\pi'E(w_tw_t'|\mathcal{F})\pi) + \max eval\left(\Sigma_\eta\right) \\
&\le \operatorname{tr}(\pi'E(w_tw_t'|\mathcal{F})\pi) + \max_{1\le i\le N}\sigma_i^2 \\
&\le \max eval(\pi\pi')E(\|w_t\|^2|\mathcal{F}) + C \\
&\le CE(\|w_t\|^2+1|\mathcal{F}).
\end{aligned} \tag{34}
$$

As a result,

$$
\begin{aligned}
&\left|E(e_{s_1}'e_{t_1}e_{t_1}'e_{s_2}e_{s_2}'e_{t_2}e_{t_2}'e_{s_1}|\mathcal{F})\right| \\
&= \left|\operatorname{tr}(E(e_{t_1}e_{t_1}'|\mathcal{F})E(e_{s_2}e_{s_2}'|\mathcal{F})E(e_{t_2}e_{t_2}'|\mathcal{F})E(e_{s_1}e_{s_1}'|\mathcal{F}))\right| \\
&\le N\max eval\left\{\prod_{t\in\{s_1,s_2,t_1,t_2\}}E(e_te_t'|\mathcal{F})\right\} \\
&\le N\prod_{t\in\{s_1,s_2,t_1,t_2\}}\max eval\left(E(e_te_t'|\mathcal{F})\right) \\
&\le NC\prod_{t\in\{s_1,s_2,t_1,t_2\}}E\left(\|w_t\|^2+1|\mathcal{F}\right).
\end{aligned}
$$

Also using that $\|w_{st}\| \le C(\|\widetilde{F}_s^{(j_s)}\|+1)(\|\widetilde{F}_t^{(j_t)}\|+1)$, we obtain that

$$
T^4\left|E\left[W_{s_1t_2}'W_{s_2t_1}W_{s_2t_2}'W_{s_1t_1}\right]\right| \le \frac{C}{N}\sup_t E\left[(\|F_t\|^8+1)(\|w_t\|^8+1)\right].
$$

Thus, condition (iv) holds as well.

Now we will check assumptions (a)-(d) of Lemma 6. First, we find the limit of the covariance matrix $\Sigma_V(T,N)$. Note that for any $s=1,...,T$,

$$
\begin{aligned}
E\left(\left(\frac{\gamma' e_s}{\sqrt{N}}\right)^2 |\mathcal{F}\right) &= \frac{1}{N}E\left((\gamma'\pi w_s + \gamma'\eta_s)^2 |\mathcal{F}\right) \\
&= \left(\frac{\gamma'\pi}{\sqrt{N}}\right) E[w_s w_s'|\mathcal{F}] \left(\frac{\pi'\gamma}{\sqrt{N}}\right) + \frac{1}{N}\sum_{i=1}^N \sigma_i^2 \gamma_i\gamma_i' \\
&\to \mathbf{\Gamma}_{\pi\gamma}' E[w_s w_s'|\mathcal{F}]\mathbf{\Gamma}_{\pi\gamma} + \mathbf{\Gamma}_\sigma.
\end{aligned}
$$

Here we used the assumptions that $N^{-1/2}\pi'\gamma \to \Gamma_{\pi\gamma}$ and $N^{-1}\sum_{i=1}^N \sigma_i^2\gamma_i\gamma_i' \to \mathbf{\Gamma}_\sigma$. Therefore,

$$
\begin{aligned}
\Sigma_V(T,N) &= var\left(\sum_{s=1}^T V_s\right) = E\left(\frac{1}{T}\sum_{s=1}^T E\left(\left(\frac{\gamma' e_s}{\sqrt{N}}\right)^2 |\mathcal{F}\right) \otimes (v_s v_s')\right) \\
&= \frac{1}{T}\sum_{s=1}^T E\left[\left(\mathbf{\Gamma}_{\pi\gamma}' w_s w_s' \mathbf{\Gamma}_{\pi\gamma} + \mathbf{\Gamma}_\sigma\right) \otimes (v_s v_s')\right] + o(1).
\end{aligned}
$$

Taking into account the structure of v_s and that $|T_j|/T = \frac{1}{4}$, we compute that

$$
\begin{aligned}
var\left(\sum_{s=1}^T V_s^0\right) &\to \mathbf{\Gamma}_{\pi\gamma}' E[w_s w_s']\mathbf{\Gamma}_{\pi\gamma} + \mathbf{\Gamma}_\sigma, \\
var\left(\sum_{s=1}^T V_s^{(j)}\right) &\to \frac{1}{4}E\left[\left(\mathbf{\Gamma}_{\pi\gamma}' w_t w_t' \mathbf{\Gamma}_{\pi\gamma} + \mathbf{\Gamma}_\sigma\right) \otimes \left(\Sigma_F^{-1}(F_t - EF_t)(F_t - EF_t)'\Sigma_F^{-1}\right)\right], \\
cov\left(\sum_{s=1}^T V_s^0, \sum_{s=1}^T V_s^{(j)}\right) &\to \frac{1}{4}E\left[\left(\mathbf{\Gamma}_{\pi\gamma}' w_t w_t' \mathbf{\Gamma}_{\pi\gamma} + \mathbf{\Gamma}_\sigma\right) \otimes \left(\Sigma_F^{-1}(F_t - EF_t)\right)\right],
\end{aligned}
$$

and

$$
cov\left(\sum_{s=1}^T V_s^{(j)}, \sum_{s=1}^T V_s^{(j^*)}\right) = 0 \text{ for } j \neq j^*.
$$

Our moment assumptions guarantee that the limit matrix Σ_V is finite. One can see that the matrix Σ_V can be written as a covariance matrix of a non-degenerate random vector and hence is of a full rank. Indeed, take a random variable χ which takes values in the set $\{1,2,3,4\}$ with equal probabilities independently of any other variable, and define $\chi_j = \mathbb{I}\{\chi = j\}$, then the covariance of random vector

$$
\left(\mathbf{\Gamma}_{\pi\gamma}' w_t + \mathbf{\Gamma}_\sigma^{1/2}\right) \otimes \begin{pmatrix} 1 \\ \left\{\Sigma_F^{-1}(F_t - EF_t)\chi_j\right\}_{j=1,...,4} \end{pmatrix}
$$

coincides with the matrix under consideration. Thus, condition (a) of Lemma 6 is satisfied.

Now note that

$$E\left(\left\|\frac{\gamma' e_t}{\sqrt{N}}\right\|^4 |\mathcal{F}\right) = \frac{1}{N^2}E\left(\|\gamma'\pi w_t + \gamma'\eta_t\|^4|\mathcal{F}\right) \leq CE\left(\|w_t\|^4|\mathcal{F}\right) + \frac{C}{N^2}E\left(\|\gamma'\eta_t\|^4\right),$$

because, in particular, $N^{-1/2}\pi'\gamma \to \Gamma_{\pi\gamma}$. Now,

$$E\left(\|\gamma'\eta_t\|^4\right) = E\left\|\sum_{i=1}^N \gamma_i'\eta_{it}\right\|^4 \leq \sum_{i=1}^N \|\gamma_i\|^4 E\eta_{it}^4 + C\sum_{i_1,i_2=1}^N \|\gamma_{i_1}\|^2\|\gamma_{i_2}\|^2\sigma_{i_1}^2\sigma_{i_2}^2.$$

Since $N^{-1}\sum_{i=1}^N \|\gamma_i\|^2 < C$ and $N^{-1}\sum_{i=1}^N \|\gamma_i\|^4 < C$ and $\max_{1\leq i\leq N}\sigma_i^2 < C$, we have that $E\left(\|\gamma'\eta_t\|^4\right) \leq CN^2$, and thus

$$E\left(\left\|\frac{\gamma' e_t}{\sqrt{N}}\right\|^4 |\mathcal{F}\right) \leq CE\left(\|w_t\|^4 + 1|\mathcal{F}\right).$$

Collecting the pieces,

$$TE\|V_s\|^4 \leq CTE\left[\frac{1}{T^2}E\left(\left\|\frac{\gamma' e_s}{\sqrt{N}}\right\|^4 |\mathcal{F}\right) \otimes \|v_s\|^4\right] \leq T\frac{C}{T^2}E\left[\left(\|w_s\|^4+1\right)\left(\|F_s\|^4+1\right)\right] \to 0.$$

This gives us the validity of condition (b) of Lemma 6.

Let us establish the validity of condition (c). Denote $\mathbf{\Gamma}_{\sigma,N} = N^{-1}\sum_{i=1}^N \sigma_i^2\gamma_i\gamma_i' \to \mathbf{\Gamma}_\sigma$. Then,

$$\begin{aligned}
&\sum_{t=1}^T V_tV_t' - \Sigma_V(T,N) \\
&= \frac{1}{T}\sum_{t=1}^T\left(\frac{\gamma' e_t}{\sqrt{N}}\frac{e_t'\gamma}{\sqrt{N}}\right)\otimes(v_tv_t') - \frac{1}{T}\sum_{t=1}^T E\left[\left(\frac{\gamma'\pi}{\sqrt{N}}w_tw_t'\frac{\pi'\gamma}{\sqrt{N}} + \mathbf{\Gamma}_{\sigma,N}\right)\otimes(v_tv_t')\right] \\
&= \frac{1}{T}\sum_{t=1}^T\left(\frac{\gamma' e_t}{\sqrt{N}}\frac{e_t'\gamma}{\sqrt{N}} - \frac{\gamma'\pi}{\sqrt{N}}w_tw_t'\frac{\pi'\gamma}{\sqrt{N}} - \mathbf{\Gamma}_{\sigma,N}\right)\otimes(v_tv_t') \\
&+ \frac{1}{T}\sum_{t=1}^T\left[\left(\frac{\gamma'\pi}{\sqrt{N}}w_tw_t'\frac{\pi'\gamma}{\sqrt{N}} + \mathbf{\Gamma}_{\sigma,N}\right) - E\left(\frac{\gamma'\pi}{\sqrt{N}}w_tw_t'\frac{\pi'\gamma}{\sqrt{N}} + \mathbf{\Gamma}_{\sigma,N}\right)\right]\otimes(v_tv_t') \\
&= A_1 + A_2.
\end{aligned}$$

Notice that given the conditional independence of η_{it}'s, the two terms in the last expression, A_1 and A_2 are uncorrelated, so in order to check condition (c) of Lemma 6 we can prove convergence to zero of $E\|A_1\|^2$ and $E\|A_2\|^2$ separately. First,

$$E\|A_1\|^2 = E\left\|\frac{1}{T}\sum_{t=1}^{T}\left(\frac{\gamma'\pi}{\sqrt{N}}w_t\frac{\eta_t'\gamma}{\sqrt{N}} + \frac{\gamma'\eta_t}{\sqrt{N}}w_t'\frac{\pi'\gamma}{\sqrt{N}} + \left(\frac{\gamma'\eta_t}{\sqrt{N}}\frac{\eta_t'\gamma}{\sqrt{N}} - \mathbf{\Gamma}_{\sigma,N}\right)\right)\otimes(v_tv_t')\right\|^2$$
$$\leq \frac{1}{T^2}\sum_{t=1}^{T}E\left\|\left(\frac{\gamma'\pi}{\sqrt{N}}w_t\frac{\eta_t'\gamma}{\sqrt{N}} + \frac{\gamma'\eta_t}{\sqrt{N}}w_t'\frac{\pi'\gamma}{\sqrt{N}} + \left(\frac{\gamma'\eta_t}{\sqrt{N}}\frac{\eta_t'\gamma}{\sqrt{N}} - \mathbf{\Gamma}_{\sigma,N}\right)\right)\right\|^2\cdot\|v_t\|^4$$
$$\leq \frac{1}{T}CE\left[(\|w_t\|^2+1)(\|F_t\|^4+1)\right]\to 0.$$

The former inequality above is due to η_t's being conditionally serially uncorrelated, and thus the summation over t could be taken outside the expectation of the square; the latter inequality uses bounds on the moments of $\eta_t'\gamma/\sqrt{N}$ we derived before. Second,

$$E\|A_2\|^2 = \left\|var\left(\frac{1}{T}\sum_{t=1}^{T}\left(\frac{\gamma'\pi}{\sqrt{N}}w_tw_t'\frac{\pi'\gamma}{\sqrt{N}} + \mathbf{\Gamma}_{\sigma,N}\right)\otimes(v_tv_t')\right)\right\|$$
$$\leq \frac{1}{T^2}\left\|\sum_{t=1}^{T}var\left(\left(\frac{\gamma'\pi}{\sqrt{N}}w_tw_t'\frac{\pi'\gamma}{\sqrt{N}} + \mathbf{\Gamma}_{\sigma,N}\right)\otimes(v_tv_t')\right)\right\| \leq \frac{C}{T}\to 0.$$

Here we use serial conditional independence of w_t's which allows the sum to be taken outside the variance. Putting all terms together we obtain that condition (c) is satisfied.

Finally, we check the condition (d):

$$T^3\left\|EW_{s_1t}V_{s_1}'V_{s_2}W_{s_2t}'\right\| = \left\|E\left[w_{s_1t}v_{s_1}'v_{s_2}w_{s_2t}'E\left(\frac{e_{s_1}'\gamma}{\sqrt{N}}\frac{\gamma'e_{s_2}}{\sqrt{N}}\frac{e_{s_1}'e_t}{\sqrt{N}}\frac{e_{s_2}'e_t}{\sqrt{N}}|\mathcal{F}\right)\right]\right\|.$$

Using that scalars could be reshuffled to make two e_t with the same index stand back to back and employing conditional independence we obtain:

$$\left|E\left(\frac{e_{s_1}'\gamma}{\sqrt{N}}\frac{\gamma'e_{s_2}}{\sqrt{N}}\frac{e_{s_1}'e_t}{\sqrt{N}}\frac{e_{s_2}'e_t}{\sqrt{N}}|\mathcal{F}\right)\right| = \frac{1}{N^2}\left|\mathrm{tr}(\gamma\gamma'E(e_{s_2}e_{s_2}'|\mathcal{F})E(e_te_t'|\mathcal{F})E(e_{s_1}e_{s_1}'|\mathcal{F}))\right|$$
$$\leq \frac{1}{N^2}\mathrm{tr}(\gamma\gamma')\prod_{s\in\{s_1,s_2,t\}}\max eval(E(e_se_s'|\mathcal{F}))$$
$$\leq \frac{C}{N}\prod_{s\in\{s_1,s_2,t\}}E(\|w_s\|^2+1|\mathcal{F})$$
$$= \frac{C}{N}E\left(\prod_{s\in\{s_1,s_2,t\}}(\|w_s\|^2+1)|\mathcal{F}\right).$$

We use the assumption $N^{-1}\mathrm{tr}(\gamma\gamma') < C$ and the bound (34) we derived before. In the last equality we also exploit that w_t's are conditionally independent of each other. Given the bounds on $\|v_s\|$ and $\|w_{st}\|$ and the moment conditions we obtain that

$$T^3 \max_{s_1,s_2,t<\min\{s_1,s_2\}} \left\| EW_{s_1t}V_{s_1}'V_{s_2}W_{s_2t}' \right\| \leq \frac{C}{N} \to 0.$$

Thus, condition (d) of Lemma 6 is satisfied as well. This concludes the proof of Lemma 2.

References

Anatolyev, S. (2012): “Inference in regression models with many regressors,” *Journal of Econometrics*, 170(2), 368-382.

Andrews, D.W.K. (2005): “Cross-Section Regression with Common Shocks,” *Econometrica*, 73(5), 1551-1585.

Angrist, J.D., G.W. Imbens and A. Krueger (1999): “Jackknife instrumental variables estimation,” *Journal of Applied Econometrics*, 14, 57-67.

Bai, J. and Ng, S. (2002): “Determining the Number of Factors in Approximate Factor Models,” *Econometrica*, 70, 191-221.

Bai, J. and S. Ng (2006): “Confidence intervals for diffusion index forecast and inference with factor-augmented regressions”, *Econometrica*, 74, 1133-1155.

Bekker, P. A. (1994): “Alternative approximations to the distributions of instrumental variable estimators,” *Econometrica*, 62, 657-681.

Bryzgalova S. (2015): “Spurious Factors in Linear Asset Pricing Models”, unpublished manuscript.

Burnside, C. (2015): “Identification and Inference in Linear Stochastic Discount Factor Models with Excess Returns,” *Journal of Financial Econometrics*, 2015, 1-36.

Cochrane, J. (2001): *Asset Pricing.* Princeton University Press.

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 J. MacBeth (1973): “Risk, Return and Equilibrium: Empirical Tests,” *Journal of Political Economy*, 81, 607-636.

Gagliardini, P., E. Ossola and O. Scaillet (2016) “Time-Varying Risk Premium in Large Cross-Sectional Equity Data Sets,” *Econometrica* 86, 985–1046.

Gospodinov, N., R. Kan, and C. Robotti (2016): “Spurious Inference in Reduced-Rank Asset-Pricing Models,” unpublished manuscript.

Hansen, C., J. Hausman and W.K. Newey (2008): “Estimation with many instrumental variables,” *Journal of Business & Economics Statistics*, 26, 398-422.

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

Heyde, C. and B. Brown (1970): “On the departure from normality of a certain class of martingales,” *Annals of Mathematical Statistics*, 41(6), 2161–2165.

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

Jegadeesh, N. and S. Titman (1993): "Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency," *Journal of Finance*, 48(1), 65-91.

de Jong, P. (1987): "A central limit theorem for generalized quadratic forms," *Probability Theory and Related Fields*, 75, 261–277.

Kan, R. and Zhang, C. (1999): "Two-Pass Tests of Asset Pricing Models with Useless Factors," *Journal of Finance*, 54, 203-235.

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

Kleibergen, F. and Z. Zhan (2015): "Unexplained factors and their effects on second pass R-squared's," *Journal of Econometrics*, 189, 101-116.

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.

Lintner, J. (1965): "Security Prices, Risk, and Maximal Gains from Diversification," *Journal of Finance*, 20, 587-615.

Onatski, A. (2012): "Asymptotics of the Principal Components Estimator of Large Factor Models with Weakly Influential Factors," *Journal of Econometrics*, 168, 244-258.

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

Sharpe, W. F. (1964): "Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk," *Journal of Finance*, 19, 425-442.

Staiger, D. and J.H. Stock (1997): "Instrumental variables regression with weak instruments," *Econometrica*, 65(3), 557-586.

Table 1a. Mean biases and standard deviations, from simulations. The correlation of missing factor with 'Market' betas is $\rho = 0.9$. The empirical missing factor is multiplied by $p = 5$. The HML betas weakened by $\sqrt{m}$.

m		1			2			4			8	
factor	Mkt	SMB	HML	Mkt	SMB	HML	Mkt	SMB	HML	Mkt	SMB	HML
$\overline{\text{beta}}^2$	0.92	0.51	0.136	0.92	0.51	0.068	0.92	0.51	0.034	0.92	0.51	0.017
two-pass	0.01	0.12	−0.37	0.02	0.06	−0.40	0.03	0.03	−0.43	0.03	0.01	−0.45
	0.89	0.88	1.24	0.60	0.524	1.13	0.41	0.31	1.08	0.29	0.19	1.06
four-split$_2$	0.01	0.06	−0.25	0.02	0.03	−0.27	0.02	0.01	−0.28	0.02	0.00	−0.29
	0.72	0.68	0.92	0.50	0.46	0.87	0.36	0.32	0.85	0.25	0.22	0.83
four-split	−0.05	0.11	−0.11	−0.01	0.05	−0.12	0.01	0.02	−0.15	0.01	0.01	−0.16
	0.67	0.57	0.74	0.47	0.40	0.69	0.33	0.27	0.66	0.24	0.20	0.65

Table 1b. Coverage of confidence intervals, from simulations. The correlation of missing factor μ with 'SMB' betas is $\rho = 0.9$. The empirical missing factor is multiplied by $p = 5$. The HML betas weakened by $\sqrt{m}$.

m		1			2			4			8	
factor	Mkt	SMB	HML	Mkt	SMB	HML	Mkt	SMB	HML	Mkt	SMB	HML
$\overline{\text{beta}}^2$	0.92	0.51	0.136	0.92	0.51	0.068	0.92	0.51	0.034	0.92	0.51	0.017
two-pass	0.87	0.83	0.81	0.87	0.86	0.71	0.87	0.89	0.60	0.86	0.90	0.47
four-split	0.95	0.94	0.94	0.94	0.93	0.88	0.94	0.91	0.77	0.93	0.88	0.62

Table 1c. Mean biases and standard deviations, from simulations. The correlation of missing factor μ with 'Market' betas is $\rho = 0.9$. The empirical missing factor is multiplied by $p = 10$. The HML betas weakened by $\sqrt{8}$.

m	4			8			16			32		
factor	Mkt	SMB	HML	Mkt	SMB	HML	Mkt	SMB	HML	Mkt	SMB	HML
$\overline{\text{beta}^2}$	0.92	0.51	0.017	0.92	0.51	0.017	0.92	0.51	0.017	0.92	0.51	0.017
two-pass	0.05	0.05	−1.05	0.04	0.03	−0.78	0.03	0.01	−0.57	0.02	0.01	−0.38
	0.57	0.55	3.25	0.43	0.37	2.42	0.30	0.23	1.67	0.22	0.14	1.08
four-split	0.03	0.02	−0.38	0.01	0.01	−0.18	0.00	0.01	−0.07	0.00	0.00	−0.04
	0.34	0.28	1.05	0.24	0.20	0.69	0.17	0.14	0.46	0.12	0.10	0.33

Table 1d. Coverage of confidence intervals, from simulations. The correlation of missing factor μ with 'Market' betas is $\rho = 0.9$. The empirical missing factor is multiplied by $p = 10$. The HML betas weakened by $\sqrt{8}$.

m	4			8			16			32		
factor	Mkt	SMB	HML	Mkt	SMB	HML	Mkt	SMB	HML	Mkt	SMB	HML
$\overline{\text{beta}^2}$	0.92	0.51	0.017	0.92	0.51	0.017	0.92	0.51	0.017	0.92	0.51	0.017
two-pass	0.78	0.76	0.28	0.76	0.74	0.25	0.73	0.76	0.25	0.71	0.78	0.28
four-split	0.94	0.93	0.69	0.93	0.90	0.66	0.93	0.89	0.65	0.92	0.87	0.61

Table 2a. Average estimated betas and beta squared from the first pass of the conventional two-pass procedure.

no.	risk factor	Market	SMB	HML	cay	MOM
1	average betas	0.96	0.53	0.19		
	average squared betas	0.92	0.51	0.14		
2	average betas	1.14		0.23		
	average squared betas	1.31		0.15		
3	average betas	1.13		0.23	−0.62	
	average squared betas	1.32		0.15	124.	
4	average betas	1.03	0.18	0.18		
	average squared betas	1.10	0.10	0.12		
5	average betas	1.03	0.18	0.17		−0.036
	average squared betas	1.08	0.10	0.11		0.014

Table 2b. Variance fractions corresponding to five main principal components in the residuals from the first pass of the conventional two-pass procedure.

no.	5 main principal components in residuals				
1	0.29	0.14	0.11	0.07	0.04
2	0.62	0.10	0.05	0.03	0.03
3	0.62	0.10	0.05	0.03	0.03
4	0.14	0.12	0.08	0.07	0.04
5	0.14	0.12	0.08	0.06	0.04

Table 2c. Risk premia estimates by the conventional two pass and proposed average four-split procedures, with standard errors.

no.	risk factor	Market	SMB	HML	cay	MOM
1	conventional two-pass	2.70	0.69	1.96		
		0.61	0.48	0.58		
	average four-split	2.80	0.46	1.29		
		0.62	0.47	0.84		
2	conventional two-pass	2.50		2.01		
		0.61		0.63		
	average four-split	2.32		−4.25		
		0.60		2.60		
3	conventional two-pass	2.55		1.92	0.027	
		0.61		0.62	0.019	
	average four-split	2.06		2.44	−0.009	
		0.63		0.68	0.005	
4	conventional two-pass	1.02	−0.35	−0.07		
		0.20	0.20	0.15		
	average four-split	1.28	−1.00	−0.24		
		0.21	0.28	0.22		
5	conventional two-pass	1.05	−0.27	−0.00		1.05
		0.20	0.19	0.15		0.35
	average four-split	1.15	−1.10	0.03		0.03
		0.21	0.24	0.18		0.40