Limit Theorems for Factor Models*

Stanislav Anatolyev† Anna Mikusheva‡
CERGE-EI and NES MIT

April 2020

Abstract

The paper establishes the central limit theorems and proposes how to perform valid inference in factor models. We consider a setting where many counties/regions/assets are observed for many time periods, and when estimation of a global parameter includes aggregation of a cross-section of heterogeneous micro-parameters estimated separately for each entity. The central limit theorem applies for quantities involving both cross-sectional and time series aggregation, as well as for quadratic forms in time-aggregated errors. The paper studies the conditions when one can consistently estimate the asymptotic variance, and proposes a bootstrap scheme for cases when one cannot. A small simulation study illustrates performance of the asymptotic and bootstrap procedures. The results are useful for making inferences in two-step estimation procedures related to factor models, as well as in other related contexts. Our treatment avoids structural modeling of cross-sectional dependence but imposes time-series independence.

Keywords: factor models, two-step procedure, dimension asymptotics, central limit theorem.

JEL classification codes: C13, C33, C38, C55.

---

*We thank the Co-Editor Guido Kuersteiner and three anonymous referees for guidance and helpful suggestions that have greatly improved the paper.

†Address: Stanislav Anatolyev, CERGE-EI, Politických vězňů 7, 11121 Prague 1, Czech Republic; e-mail stanislav.anatolyev@cerge-ei.cz. Czech Science Foundation support under grant number 20-28055S is gratefully acknowledged.

‡Address: Department of Economics, M.I.T., 50 Memorial Drive, Building E52, Cambridge, MA, 02142, USA; e-mail amikushe@mit.edu. National Science Foundation support under grant number 1757199 is gratefully acknowledged.
1 INTRODUCTION

Data with an underlying factor structure are increasingly used in empirical macroeconomics and finance. Often these data consist of time series of observations for multiple cross-sectional units (assets, portfolios, regions or industries). Quite a few new estimation strategies have appeared in the empirical literature that use both cross-sectional and time series variation in order to estimate global structural parameters. Often the parameter of interest arises from aggregation or estimation using cross-sectional variation of individual parameters for each entity. One example of such a structure is linear factor pricing model in asset pricing (Fama & MacBeth 1973 and Shanken 1992), where we usually use for estimation time series of excess returns for a number of portfolios or assets priced by a small number of risk factors. Each portfolio or stock may have its own (heterogeneous) exposure to risk, often referred to as betas, which can be estimated separately from time series observations for each portfolio. The parameter of interest, a risk premium, is defined as the coefficient of proportionality in the cross-sectional relation between the average excess return on a portfolio and its individual beta.

A vast majority of macroeconomic shocks are only weakly identified via structural VARs that use only time series observations on leading macro variables. A new approach to the estimation of causal effects of a macro shock on the economy is to use cross-sectional variation in data on regions, countries or industries. For example, Serrato & Wingender (2016) use cross-sectional variation in federal spending programs due to a Census shock to identify the causal impact of government spending on the economy. Cross-sectional variation among counties in government spending and in the accuracy of census-based estimates of population provides a better justified treatment effect framework, allows for the estimation of local fiscal multipliers, and finally gives a better global estimate of the fiscal multiplier via aggregation of local multipliers. Hagedorn, Manovskii & Mitman (2015) estimate the aggregate effect of unemployment-benefit duration on employment and labor force participation using cross-sectional differences across US states. Sarto (2018) discusses how heterogeneous sensitivities of regions to aggregate policy variables, so called micro-global elasticities, can be used to recover macro elasticities of interest such as, for example, a fiscal multiplier.

A shared feature of the above-mentioned examples is the use of time-series observations on multiple entities (stocks, portfolios, counties, states or industries), while data on those entities are not independent and identically distributed. Moreover, variables for different entities often display strong co-movements to the extent that the data have
a factor structure, and estimation of these co-movements is the main goal. Indeed, the realization of a risk factor in the economy moves returns on all portfolios simultaneously, while a federal fiscal shock moves spending in all US counties, though in both cases heterogeneously so. A valid estimation procedure must explicitly model and account for the data’s factor structure to the extent that the error terms (or residuals) can be considered idiosyncratic; see Kleibergen & Zhan (2015) and Anatolyev & Mikusheva (2018) for how a factor structure that is unaccounted for can lead to misleading results. However, idiosyncrasy of the errors usually implies only that the correlation among errors for different entities is relatively small and does not introduce first-order bias to the estimation procedure. Usually, it is not reasonable to assume that errors for different entities are completely independent; indeed, stocks in the same industry are likely to co-move even after global-economy risks are removed, while errors for neighboring counties are more likely to be correlated even after one accounts for federal shocks. At the same time, we typically want to remain agnostic about the correlation structure of shocks and avoid their structural modeling as long as this does not introduce biases.

The second typical feature of the above-mentioned examples is the two-step nature of the estimation procedure, where in the first step we estimate entity-specific coefficients (risk exposures/betas, local fiscal multipliers, micro-global elasticities) by running a time-series regression separately for each entity. In the second step, we estimate the global coefficient of interest by either aggregating entity-specific coefficients (Serrato & Wingender 2016 and Hagedorn, Manovskii & Mitman 2015), or by running an OLS regression on the cross-section of entity-specific coefficients (Fama & MacBeth 1973 and Sarto 2018), or by running an IV regression on the cross-section of entity-specific coefficients (Anatolyev & Mikusheva 2018).

The goal of this paper is to establish a central limit theorem (CLT) and to provide a tool for establishing asymptotic normality of estimates obtained in such two-step estimation procedures and for finding ways to do asymptotically correct inference, while being flexible in modeling the cross-sectional dependence of errors. The main difficulty here is that even though the second step cross-sectional regression has nearly uncorrelated errors (which is usually sufficient to obtain consistency of the two-step estimator), this condition is usually insufficient for a CLT, which typically requires that stronger discipline be imposed on the dependence structure (such as independence, or a martingale-difference structure, or mixing). Our solution to this problem is to restrict the time series behavior while staying agnostic about the cross-sectional dependence. We assume time-series independence of idiosyncratic errors, which is consistent with market efficiency for factor
asset pricing models and the non-predictability of macro shocks in macroeconomic settings. The estimation noise in a two-stage procedure involves aggregation both over time (from the first step) and over entities (from the second step). We show that under certain conditions it is sufficient to have a CLT over just one of these directions, and we use the time-series direction for that.

When the second step uses an OLS or IV estimator, the CLT must adapt to averages of quadratic forms, as both the second-step-dependent variable and the second stage regressor/instrument contain first-stage estimation noise. Our CLT has a linear and a quadratic parts. We also note that a need for a CLT for quadratic forms in factor model sometimes arises also for the first-step estimators (e.g., Pesaran & Yamagata 2018) or in higher order asymptotic derivations (e.g., Bai & Ng 2010).

There is a growing literature that establishes different CLTs while acknowledging the importance of cross-sectional dependence in the data, which stems from spatial relations and/or from the presence of common factors. Kuersteiner & Prucha (2013) establish a CLT for linear sums in a panel data context with growing cross-sectional dimension \( N \) and fixed time-series dimension \( T \) allowing for cross-sectional dependence, and Kuersteiner & Prucha (2018) extend these results to quadratic forms as well. Both papers impose conditional moment restrictions, which allows the authors to construct a martingale-difference sequence in the cross-sectional direction. The main conditional moment restrictions imply a correct specification of an underlying model, which need not be required by our CLT. However, the mentioned papers allow more flexibility in modeling the time dependence, and do not require large \( T \). Another CLT that requires both large \( N \) and large \( T \) is established in Hahn et al. (2020) for linear terms only.

This paper also contributes to the literature on the CLT for quadratic forms. Various types of CLT for quadratic forms have been previously established and used in the many instrument literature (see for example, Chao et al. 2012, Hausman et al. 2012, Sølvsten 2020) and many covariate literature (see Cattaneo et al. 2018), as well as in the literature on semi-parametric estimation (Cattaneo et al, 2014a, 2014b). The CLT used in those papers are established for the cross-sectional dimension only, and rely heavily on the independence assumption. We adapt the ideas used in Chao et al. (2012), specifically the approach of de Jong (1987), to accommodate large cross-sectionally dependent panels (an alternative approach, known as Stein’s method, is used in Sølvsten 2020).

Our second set of results is related to ways of conducting valid statistical inference. Under strengthened conditions on the weakness of the cross-sectional correlation of errors, we show that a conventional variance estimator is consistent, and so the usual asymptotic
inference can be applied. When such strengthened conditions do not hold, we propose instead a variant of a wild bootstrap scheme that replicates the original cross-sectional dependence structure. We also conduct a small simulation experiment that provides evidence on the approximation quality of our CLT and on the empirical size and power of wild bootstrap in a moderately sized panel.

The paper proceeds as follows. Section 2 explains problems with establishing asymptotic Gaussianity for two-step and other estimators and test statistics, and shows how discipline in the time series direction can help. Section 3 introduces assumptions on idiosyncratic errors, states central limit theorems for two cases, and discusses the relevance of those cases to empirical practice. Section 4 discusses estimation of asymptotic variances for asymptotic inference and alternative inference tools based on the bootstrap. Section 5 presents a small simulation experiment that reveals properties of asymptotic and proposed bootstrap inference tools. Section 6 concludes. All proofs appear in the Appendix.

2 GOALS AND EXAMPLES

Let the data contain observations on many units indexed by \( i = 1, \ldots, N \), and observed for multiple time periods \( t = 1, \ldots, T \). We assume that both \( N \) and \( T \) increase to infinity without restrictions on their rates. The goal of this paper is to find the conditions under which the following statement will hold:

\[
Ξ_{N,T} \equiv \frac{1}{\sqrt{N}} \sum_{i=1}^{N} ξ_i \Rightarrow \mathcal{N}(0, Σ_ξ),
\]

(1)

where

\[
ξ_i = \left( \frac{1}{T} \sum_{s=1}^{T} v_s \gamma_i e_{is} \right) \cdot \left( \frac{1}{T} \sum_{s=1}^{T} \sum_{t<s} w_{st} e_{it} e_{is} \right).
\]

Here, \( e_{it} \) are weakly cross-sectionally dependent entity-specific (idiosyncratic)\(^1\) errors with \( \mathbb{E}(e_{it}) = 0 \). Errors \( e_{it} \) are uncorrelated with the variables \( v_t \) and \( w_{st} \) that are common to all units \( i = 1, \ldots, N \) (more exact conditions are to appear in the next Section). We assume \( γ_i, i = 1, \ldots, N \), to be non-random entity-specific weights. Further, we want to study the circumstances when one can also consistently estimate the asymptotic covariance – that is, sufficient conditions for a statement like

\[
\frac{1}{N} \sum_{i=1}^{N} ξ_i ξ_i' \overset{p}{\to} Σ_ξ.
\]

(2)

\(^1\)By idiosyncratic error we mean the factor-removed part of entity-specific variables.
As we argue below (see Examples 1–3), statements (1) and (2) are often needed in order to conduct statistical inferences (testing or confidence set construction) about a structural parameter, $\lambda$, which is estimated in two steps. We consider a case when at the first step a researcher estimates a parameter $\beta_i$ for each entity/unit/state $i = 1, \ldots, N$, typically via running OLS or IV time series regressions. A typical linear estimator can be written as

$$\hat{\beta}_i = \beta_i + \varepsilon_i,$$

where the estimation error has the structure

$$\varepsilon_i = (1 + o_p(1)) \frac{1}{T} \sum_{t=1}^{T} v_t e_{it},$$

with the $o_p(1)$ term uniformly small over the units. In this setting, $v_t$ is either a regressor common to all entities, or a common systematic part of entity-specific regressors that have a factor structure.

Example 1. There is a variety of estimation approaches that can be used at the second step. The simplest of them is weighted averaging of the first step estimates, viz.

$$\hat{\lambda} = \frac{1}{N} \sum_{i=1}^{N} \gamma_i \hat{\beta}_i.$$ Such estimator is used in Sarto (2018). In order to justify asymptotic Gaussianity of $\hat{\lambda}$ and to make statistical inferences about $\lambda$, one needs statements on the asymptotic behavior of

$$\sqrt{\frac{T}{N}} \sum_{i=1}^{N} \gamma_i \varepsilon_i = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} v_t \gamma_i e_{it}.$$

Note that the last expression has a structure of normalized averages stated as the first component of $\Xi_{N,T}$. Such ‘linear’ terms, where only the first component of $\Xi_{N,T}$ is involved, are very common in asymptotic derivations in factor models (e.g., Bai & Ng 2006, 2010).

Example 2. The second estimation step may invoke a more complex estimator involving a sample covariance between multiple first stage estimators or estimators for multiple first stage parameters. For example, the Fama-MacBeth procedure employs the data on excess returns to a set of portfolios $\{r_{it}, i = 1, \ldots, N, t = 1, \ldots, T\}$ and time series of a risk factor $\{F_t, t = 1, \ldots, T\}$. Namely, it uses two collections of first stage parameters – the average return on a portfolio $\beta_i^{(1)} = \mathbb{E} r_{it}$ via the sample average return $\hat{\beta}_i^{(1)} = \frac{1}{T} \sum_{t=1}^{T} r_{it}$, and the risk exposure of a portfolio $\beta_i^{(2)} = \text{var}(F_t)^{-1} \text{cov}(r_{it}, F_t)$ via the time series OLS regression of $r_{it}$ on $F_t$ resulting in an estimate $\hat{\beta}^{(2)}_i$. At the second stage of the Fama-MacBeth

\[\varepsilon_i = \frac{1}{T} \sum_{t=1}^{T} v_{it} e_{it} = \frac{1}{T} \sum_{t=1}^{T} u_{it} e_{it} + \frac{1}{T} \sum_{t=1}^{T} u_{it} e_{it} = \frac{1}{T} \sum_{t=1}^{T} v_{it} e_{it}^*,\]

where $v_{it} = (u_{it}, 1)'$ and $e_{it}^* = (a_i e_{it}, u_{it} e_{it})$. 

\[\frac{1}{N} \sum_{i=1}^{N} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} v_{it} \gamma_i e_{it} = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} v_{it} \gamma_i e_{it},\]

Our setting can accommodate entity-specific regressors, say $v_{it}$, that have factor structure themselves. Assume that $v_{it} = a_i u_{it} + u_{it}$, where $u_{it}$ is a common co-movement in the regressors and $u_{it}$ is idiosyncratic. Then

\[\varepsilon_i = \frac{1}{T} \sum_{t=1}^{T} u_{it} e_{it} = \frac{1}{T} \sum_{t=1}^{T} u_{it} (a_i e_{it}) + \frac{1}{T} \sum_{t=1}^{T} u_{it} e_{it} = \frac{1}{T} \sum_{t=1}^{T} v_{it} e_{it}^*,\]

where $v_{it} = (u_{it}, 1)'$ and $e_{it}^* = (a_i e_{it}, u_{it} e_{it})$. 

\[\sqrt{\frac{T}{N}} \sum_{i=1}^{N} \gamma_i \varepsilon_i = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} v_{it} \gamma_i e_{it}^*.\]
procedure, one runs the OLS regression of the sample average return $\hat{\beta}_i^{(1)}$ on the portfolio risk exposure estimated at the first step $\hat{\beta}_i^{(2)}$. In this case,

$$\hat{\lambda} = \left( \sum_{i=1}^{N} \hat{\beta}_i^{(2)} \hat{\beta}_i^{(2)} \right)^{-1} \sum_{i=1}^{N} \hat{\beta}_i^{(1)} \hat{\beta}_i^{(2)} ,$$

the second step involves two sample covariances. If one wants to derive the asymptotic distribution of $\hat{\lambda}$, one needs to establish the asymptotic distribution for a properly normalized sample covariance of the two first step estimators $\frac{1}{N} \sum_{i=1}^{N} \hat{\beta}_i^{(1)} \hat{\beta}_i^{(2)}$, where $\hat{\beta}_i^{(j)} = \beta_i^{(j)} + \varepsilon_i^{(j)}$, with the estimation error having the structure $\varepsilon_i^{(j)} = (1 + o_p(1))^\frac{1}{T} \sum_{t=1}^{T} \epsilon_{it}^{(j)}$.

The normalized sample covariance of the two first step estimators contains several terms:

$$\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left( \hat{\beta}_i^{(1)} \hat{\beta}_i^{(2)} - E \hat{\beta}_i^{(1)} \hat{\beta}_i^{(2)} \right)$$

$$= \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left( \beta_i^{(1)} \varepsilon_i^{(2)} + \beta_i^{(2)} \varepsilon_i^{(1)} \right) + \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left( \varepsilon_i^{(1)} \varepsilon_i^{(2)} - E \varepsilon_i^{(1)} \varepsilon_i^{(2)} \right). \quad (3)$$

The first term on the right-hand-side of equation (3) is similar to a weighted average of first step estimators and has the form of the first component of $\Xi_{N,T}$ (treating $\beta_i^{(j)}$ as constants similar to constants $\gamma_i$). The second term in equation (3) is more complicated and calls for a Central Limit Theorem for quadratic forms:

$$\frac{T}{\sqrt{N}} \sum_{i=1}^{N} \left( \varepsilon_i^{(1)} \varepsilon_i^{(2)} - E \varepsilon_i^{(1)} \varepsilon_i^{(2)} \right)$$

$$= \frac{1}{T \sqrt{N}} \sum_{i=1}^{N} \sum_{t=1}^{T} v_i^{(1)} \varepsilon_i^{(2)} \epsilon_{it}^2 - E \epsilon_{it}^2 + \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \sum_{s=1}^{T} \sum_{t<s} w_{st} \epsilon_{it} \epsilon_{is},$$

where $w_{st} = v_s^{(1)} v_t^{(2)} + v_t^{(1)} v_s^{(2)}$. Here the first term can be treated as the first component, and the second term as the second component of $\Xi_{N,T}$.

**Example 3.** Anatolyev & Mikusheva (2018) propose a split-sample estimator as an alternative to the Fama-MacBeth procedure for factor asset pricing. There are three sets of parameter estimates produced at the first stage: the sample average return $\hat{\beta}_i^{(1)} = \frac{1}{T} \sum_{t=1}^{T} r_{it}$ and two estimates of the portfolio risk exposure computed as OLS estimates in regressions of $r_{it}$ on $F_t$ on different sub-samples, say, $\hat{\beta}_i^{(2)}$ and $\hat{\beta}_i^{(3)}$. In our notation, the usage of a sub-sample is accommodated by setting $v_t = 0$ for those $t$ not in the currently
used sub-sample. The second step IV estimator is constructed as an IV estimator in the regression of \( \hat{\beta}_1^{(1)} \) on \( \hat{\beta}_1^{(2)} \) using \( \hat{\beta}_1^{(3)} \) as instrument. That is,

\[
\hat{\lambda} = \left( \sum_{i=1}^{N} \hat{\beta}_1^{(3)} \hat{\beta}_1^{(1)} \right)^{-1} \sum_{i=1}^{N} \hat{\beta}_1^{(3)} \hat{\beta}_1^{(2)}.
\]

In order to make inferences on \( \lambda \), one needs to obtain the asymptotic distribution of sample covariances between different first step estimates. Statements like (1) and (2) are instrumental to accomplish this.

The configuration in \( \Xi_{N,T} \) and a need for statements (1) and (2) occur in other situations as well.

Example 4. Pesaran & Yamagata (2018) suggest a new test for factor pricing model that allows many portfolios to be considered simultaneously (with \( N \) and \( T \) both diverging to infinity). The hypothesis of interest \( H_0 : \alpha_i = 0 \) for all \( i = 1, \ldots, N \), where \( \alpha_i \) is a pricing error for the portfolio \( i \). To estimate the pricing errors the authors use OLS estimates \( \hat{\alpha}_i \). A large number of portfolios \( N \) does not allow one to establish joint Gaussianity of all \( \hat{\alpha}_i \) or to consistently estimate their covariance. Pesaran & Yamagata (2018) propose to test the hypothesis of interest using statistics based on a weighted sum of squares of \( \hat{\alpha}_i \). They create a properly normalized statistic of the form

\[
\sum_{i=1}^{N} \left( \frac{\hat{\alpha}_i^2}{\sigma_i^2} - 1 \right),
\]

where \( \sigma_i^2 \) are variances of pricing errors. This statistic is directly related to the sample variance of the first step estimator, and a statement of its asymptotic Gaussianity directly follows from (1) by the same logic as stated above. Pesaran & Yamagata (2018) develop a CLT for quadratic forms that can be applied in this setting. They make an assumption that the idiosyncratic components can be filtered to make them cross-sectionally independent.\(^3\) Here we propose an alternative version of CLT that can be applied under less restrictive assumptions on the cross-sectional dependence of \( e_{it} \)’s.

Example 5. A data-rich IV environment of Bai & Ng (2010) is another example where our linear-quadratic CLTs can be useful. The authors consider an IV setup with many instruments in a panel, where the number of instruments, \( N \), is potentially higher than

\(^3\)See Assumptions 2 and 3 in Pesaran & Yamagata (2018).
the number of observations, $T$. The instruments are generated by a factor model $z_{it} = \lambda_i'F_t + e_{it}$, with $F_t$ and $e_{it}$ independent of the structural error $\varepsilon_t$, and time-series and cross-sectional dependence in $e_{it}$ is allowed, though restricted. Bai & Ng (2010) consider is bias-corrected GMM estimator, which corrects for inconsistency of the baseline GMM. It is consistent when $N/T = O(1)$, however, its asymptotic Gaussianity is established under a more restrictive assumption when $N/T = o(1)$. The challenge is that when $N/T = O(1)$, the asymptotic expansion for this estimator has, in addition to a linear term, a quadratic form in the idiosyncratic components $e_{it}$ similar to the second component of $\Xi_{N,T}$. Thus, using statement (1), an asymptotic theory could be developed for the bias-corrected GMM estimator without having to impose $N/T = o(1)$.

In most of these examples, the set of idiosyncratic components $\{e_{it}, i = 1, ..., N, t = 1, ..., T\}$ cannot be regarded independent and/or identically distributed. In most realistic applications, one is usually willing to assume that $e_{it}$ do not have strong (detectable) factor structure, but still allow for some correlation between different units, which would not affect the consistency. For example, it is reasonable to think that stocks of firms in the same industry or of the same size may react to some local shocks and be correlated, though when averaged over all stocks (and all industries), this co-movement of returns would have no first-order impact on estimation.

Our attempt to be agnostic with regard to possible cross-sectional correlation among errors and to avoid explicit modeling of its structure whenever possible comes at a cost of more restrictive time series assumptions. In many applications of interest, it is more credible to impose strong dependence assumptions in a time-series direction rather than in a cross-sectional direction. For example, the efficient market hypothesis implies mean non-predictability of excess returns given past history, which is equivalent to a martingale difference property for the errors. The definition of shocks in macroeconomics similarly presumes their time-series independence. In this paper, we assume time-series independence, which in some cases may be weakened to the martingale-difference property or stationarity with some proper mixing condition, but we do not pursue this generalization here.

3 CENTRAL LIMIT THEOREM

In this paper we consider asymptotics as both cross-sectional and time-series sample sizes, $N$ and $T$, increase to infinity. We allow the data-generating process for all variables to
vary with \( N \) and \( T \). Define \( \mathcal{F} \) to be a sigma-algebra that contains at least the sigmamodels generated by the full set of variables \( \{ v_s, s = 1, ..., \infty \} \) and \( \{ w_{st}, s, t = 1, ..., \infty \} \) for all \( s \) and \( t \). It may potentially also contain other events related to common shocks and variables, as long as Assumption 3 stated below is satisfied. We treat \( \gamma_i \) as non-random \( k \times 1 \) vectors.

In order to simplify the notation, in what follows we will denote \( C \) to be a positive generic constant, independent of \( N \) and \( T \), which may be different in different equations, but does not depend on or change with \( N \) or \( T \). We will use the following notation: for a square matrix \( A \) we denote by \( \text{tr}(A) \) its trace, by \( \text{max ev}(A) \) – its maximal eigenvalue, and by \( \text{dg}(A) \) a diagonal matrix of the same size with the elements from the diagonal of \( A \), by \( \| \cdot \| \) the \( l_2 \) norm for a vector or the operator norm for a matrix.

**Assumption 1** The random vectors \( v_s \) and \( w_{st} \) are measurable with respect to the sigma-algebra \( \mathcal{F} \) for all \( s, t \), and

\[
\begin{align*}
(i) & \quad \frac{1}{T} \sum_{s=1}^{T} \mathbb{E}(v_s v_s') \to \Omega_v \quad \text{and} \quad \frac{1}{T} \sum_{s=1}^{T} \sum_{t<s} \mathbb{E}(w_{st} w_{st}') \to \Omega_w, \quad \text{where} \quad \Omega_v \quad \text{and} \quad \Omega_w \quad \text{are full rank matrices}; \\
(ii) & \quad \max_{1 \leq s \leq T} \mathbb{E}[\|v_s\|^4] < C \quad \text{and} \quad \max_{1 \leq t, s \leq T} \mathbb{E}[\|w_{st}\|^4] < C; \\
(iii) & \quad \mathbb{E}\left[\left\| \frac{1}{T} \sum_{s=1}^{T} \sum_{t<s} (w_{st} w_{st}' - \mathbb{E}w_{st} w_{st}') \right\|^2 \right] \to 0; \\
(iv) & \quad \mathbb{E}\left[\left\| \frac{1}{T} \sum_{s=1}^{T} (v_s v_s' - \mathbb{E}v_s v_s') \right\|^2 \right] \to 0.
\end{align*}
\]

**Assumption 2** \( \max_{1 \leq i \leq N} \|\gamma_i\| < C \).

**Assumption 3** \( (i) \) Conditional on \( \mathcal{F} \), the random vectors \( e_t = (e_{1t}, ..., e_{Nt})' \) are serially independent, and \( \mathbb{E}(e_t|\mathcal{F}) = 0 \) for all \( t \);

\( (ii) \) \( \max_{1 \leq i \leq N, 1 \leq t \leq T} \mathbb{E}(e_{it}^4) < C \).

Assumption 1 imposes very mild restrictions on the time-series behavior of the common (non-entity specific) variables. For example, the part related to \( v_t \) is trivially satisfied if a time series equal to \( v_t v_t' \) is weakly stationary with summable auto-covariances. Assumption 2 restricts the influence of any one entity in the cross-sectional average and will eventually contribute to asymptotic negligence of the cross-sectional summands needed for the CLT. Assumption 3(i) is a restrictive assumption which imposes discipline on the
time-series structure, and the restriction $\mathbb{E}(e_t|\mathcal{F}) = 0$ is a form of strict exogeneity in the first step regression. Uniform moment boundedness in Assumption 3(ii) is traditional.

Apparently, Assumptions 1, 2 and 3 are insufficient to establish a central limit theorem, and we need to put some restrictions on the cross-sectional dependence and dependence between idiosyncratic errors and common variables. Indeed, we will use a change of summation ordering:

$$
\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \xi_i = \left( \frac{1}{\sqrt{T}} \sum_{s=1}^{T} v_s \left( \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \gamma_i e_{is} \right) \right),
$$

and will establish asymptotic convergence in the time-series direction. In order to apply CLT in time series direction we need some sort of asymptotic negligibility of summands with different time indexes, in particular, of terms like $\left\{ v_s \left( \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \gamma_i e_{is} \right) \right\}$ and $\left\{ w_{st} \left( \frac{1}{\sqrt{N}} \sum_{i=1}^{N} e_{it} e_{is} \right) \right\}$, Our goal is to provide low-level assumptions. There is a trade-off in how much dependence of idiosyncratic errors across entities and how much dependence between idiosyncratic errors and common variables are allowed. Below we consider two particular cases. In the first case, full independence between the $e_{it}$'s and $\mathcal{F}$ is assumed; as a result, we can be agnostic about the structure of cross-sectional dependence, the corresponding assumptions about it are relatively mild. In the second case, we allow for conditional heteroskedasticity in $e_{it}$ that can be related to some common variables from $\mathcal{F}$ producing dependence in higher-order conditional moments. This flexibility comes at a cost of imposing some structure on the cross-sectional behavior of $e_{it}$.

### 3.1 Independence from common variables

**Assumption 4**

(i) The errors $e_t = (e_{1t}, ..., e_{Nt})'$, $t = 1, ..., T$ are independent from the sigma-algebra $\mathcal{F}$ and identically distributed for different $t$; 

(ii) For the $N \times N$ covariance matrix $\mathcal{E}_{N,T} = \mathbb{E} [e_t e_t']$, $\lim \sup_{N,T \to \infty} \max \text{ev} (\mathcal{E}_{N,T}) < \infty$, and $\frac{1}{N} \text{tr}(\mathcal{E}_{N,T}^2) \to a$; 

(iii) $\frac{1}{N} \gamma' \mathcal{E}_{N,T} \gamma \to \Gamma_\sigma$, where $\Gamma_\sigma$ is full rank; 

(iv) $\frac{1}{N^2} \sum_{i_1=1}^{N} \sum_{i_2=1}^{N} \sum_{i_3=1}^{N} \sum_{i_4=1}^{N} |\mathbb{E} \left( e_{i_1t} e_{i_2t} e_{i_3t} e_{i_4t} \right)| < C$.

**Theorem 3.1** Under Assumptions 1, 2, 3 and 4, the central limit theorem stated in equation (1) holds with $\Sigma_\xi = \left( \begin{array}{cc} \Sigma_V & 0 \\ 0 & \Sigma_W \end{array} \right)$, where $\Sigma_V = \Gamma_\sigma \otimes \Omega_v$ and $\Sigma_W = a \Omega_w$. 

10
Numerous papers that establish inferences in factor models commonly assume that the set of factors is independent from the set of idiosyncratic errors, as in Assumption 4(i), though cross-sectional dependence of errors is allowed; see, for example, Assumption D in Bai & Ng (2006). We intended for the first part of Assumption 4(ii) to impose weak cross-sectional dependence as expressed by the covariance matrix; in particular, it means that no strong factor structure is left in the errors; similar assumptions appear in Onatski (2012) and Bai & Ng (2006). The convergence of the trace in Assumptions 4(ii) and 4(iii) is needed for the asymptotic covariance matrix to be properly defined.

Assumption 4(iv) is another way to restrict pervasive dependence in multiple variables, in particular, precluding outliers to realize in too many error terms simultaneously. For example, imagine that the cross-sectional dependence is induced by several groups with a factor structure, e.g., stock returns are correlated because there are industry-specific shocks and geography-specific shocks. Imagine that there are a finite number, say \( G \), groups, indexed by \( g = 1, \ldots, G \), which may be overlapping, with each having independent shocks \( f_{g,t} \) at time \( t \). Stock \( i \) has non-zero loading \( \pi_{i,g} \) only if it belongs to group \( g \). Let the set of groups, to which \( i \) belongs, be denoted by \( G(i) \). That is,

\[
e_{it} = \sum_{g \in G(i)} \pi_{i,g} f_{g,t} + \eta_{it},
\]

where \( \eta_{it} \)'s are independent both cross-sectionally and across time and have finite fourth cumulants. Then, Assumption 4(iv) is essentially equivalent to the following two conditions: \( \mathbb{E} f_{g,t}^4 < C \) and \( \frac{1}{N} \left( \sum_{i=1}^{N} |\pi_{i,g}| \right)^2 < C \) for any \( g = 1, \ldots, G \). Thus, for this toy example, essentially Assumption 4(iv) imposes that the factors \( f_{g,t} \) do not produce outliers too often expressed as the moment condition and a statement about pervasiveness.

One of the important steps in the proof of Theorem 3.1 verifies asymptotic negligibility of time-series summands by checking boundedness of the fourth moments of the cross-sectional sums \( \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \gamma_i e_{is} \) and \( \frac{1}{\sqrt{N}} \sum_{i=1}^{N} e_{it} e_{is} \); that imposes the main way we restrict cross-sectional dependence. The fourth cumulant conditions are reminiscent of those in de Jong (1987), which we follow while proving our CLT using Heyde & Brown (1970). There are alternative CLT for quadratic forms such as Rotar’ (1973) that imposes weaker moment conditions on the summands but stricter assumptions on the negligibility of coefficients and eigenvalues of the quadratic form. In our case, following them would require imposing stronger assumptions on the variables \( w_{st} \), which we would like to avoid.
3.2 Conditional heteroskedasticity

Assumption 4(i) of independence is much stronger than Assumption 3(i) about exogeneity: it does not allow higher conditional moments of $e_{it}$ to co-move with the common variables; in particular, it imposes conditional homoscedasticity. It may be especially problematic in financial applications where time-varying volatility is of strong empirical relevance, and returns on many stocks display patterns of changing volatility driven by some common variables. The assumptions below allow for conditional heteroskedasticity.

**Assumption 5** Assume that the errors $e_{it}$ have the following weak (unobserved) factor structure:

$$e_{it} = \pi_i^t f_t + \eta_{it},$$

where the following assumptions hold.

(i) The $k_f \times 1$ process $f_t$ is serially independent, conditionally on $F$, with $E(f_t|F) = 0$, $\mathbb{E}(f_t f_t') = I_{k_f}$, $\max_{1 \leq t \leq T} \mathbb{E}[(\|v_s\|^4 + 1)\|f_t\|^4] < C$, and $\max_{1 \leq t, t' \leq T} \mathbb{E}[(\|w_{st}\|^4\|f_t\|^8] < C$;

(ii) $\max ev \left( \sum_{i=1}^N \pi_i \pi_i' \right) < C$ and $\frac{1}{\sqrt{N}} \sum_{i=1}^N \pi_i \gamma_i \rightarrow \Gamma_{\pi\gamma}$;

(iii) The random variables $\eta_{it}$ are independent both cross-sectionally and across time, independent from both $f_s$’s and $F$, have mean zero and variances $\text{var}(\eta_{it}) = \omega_i^2$ that are bounded from above and such that $\frac{1}{N} \sum_{i=1}^N \omega_i^4 \rightarrow \omega^4$, $\frac{1}{N} \sum_{i=1}^N \omega_i^2 \gamma_i \gamma_i' \rightarrow \Gamma_{\omega}$, where $\Gamma_{\omega}$ has full rank, and $\max_{1 \leq i \leq N, 1 \leq t \leq T} \mathbb{E}(\eta_{it}^4) < C$;

(iv) Additionally, if $\Gamma_{\pi\gamma} \neq 0$, then there exists a matrix $\Sigma_{fv}$ such that

$$\mathbb{E} \left[ \left\| \frac{1}{T} \sum_{s=1}^T (f_s f_s') \otimes (v_s v_s') - \Sigma_{fv} \right\|^2 \right] \rightarrow 0.$$

**Theorem 3.2** Under Assumptions 1, 2, 3 and 5, the statement of the central limit theorem stated in equation (1) holds with $\Sigma_\xi = \begin{pmatrix} \Sigma_V & 0 \\ 0 & \Sigma_W \end{pmatrix}$, where $\Sigma_W = \omega^4 \Omega_w$ and $\Sigma_V = \left( \Gamma_{\pi\gamma} \otimes I_{k_v} \right) \Sigma_{fv} \left( \Gamma_{\pi\gamma} \otimes I_{k_v} \right) + \Gamma_\omega \otimes \Omega_v$.

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 (latent) factor structure. The condition $\mathbb{E}(f_t f_t') = I_{k_f}$ is a normalization and involves no loss of generality. Assumption 5(ii) forces the factors to be weak to such an extent that the factor structure cannot
be consistently detected; it implies that the covariance matrix of idiosyncratic errors would satisfy the first half of Assumption 4(ii). Moreover, this factor structure may be closely related to the common variables in $F$, which causes the cross-sectional dependence among the errors $e_{it}$ to change with the common variables and allows a very flexible form of conditional heteroskedasticity. Indeed, the conditional cross-sectional covariance is

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

Since we do not restrict $\mathbb{E}(f_t|F)$ beyond proper moment conditions, the strength of any cross-sectional dependence as well as error variances may change stochastically depending on realizations of the common variables.

The moment conditions in Assumption 5(i) help to establish asymptotic negligibility of the time-series summands. Assumption 5(iii) about $\Gamma_\omega$ and Assumption 5(iv) allow us to define properly the asymptotic covariance matrix.

## 4 VALID INFERENCE

In this Section, we first discuss estimation of asymptotic variances for asymptotic inference when this leads to valid inference. Then, we propose alternative tools based on the wild bootstrap to apply in situations when asymptotic inference fails to provide asymptotically correct inference.

### 4.1 Asymptotic inference

Statistical inferences such as confidence set construction and hypotheses testing about the structural parameter typically require consistent estimation of asymptotic variances of all important quantities that are asymptotically Gaussian. The easiest to implement and thus the most appealing from an applied perspective are those that use the same variables and have a structure similar to the original averages, such as the statement in equation (2).

Notice that equation (2) contains the cross-sectional summation outside, and hence it treats the cross-section as nearly uncorrelated observations, or at least it ignores the cross-sectional correlation. A relevant analog is the difference between the long-run covariance and instantaneous covariance in a classical time series. However, implementing an analog of long-run covariance estimation here would be a challenge since we do not have any cross-sectional stationarity or a measure of distance between cross-sectional entities. Rather, we explore under which conditions the convergence (2) holds.
Theorem 4.1 below obtains a statement for the case when the common variables are independent from the idiosyncratic errors, while Theorem 4.2 establishes a similar statement for the conditionally heteroskedastic case.

**Theorem 4.1** If in addition to Assumptions 1, 2, 3, 4 we also have that
\[ \| \mathcal{E}_{N,T} - dg(\mathcal{E}_{N,T}) \| \to 0 \quad \text{as } N,T \to \infty, \] (4)
then consistency statement (2) holds.

**Theorem 4.2** If in addition to Assumptions 1, 2, 3, 5 we also have that \( \Gamma_{\pi\gamma} = 0 \), then consistency statement (2) holds.

The additional assumption (4) in Theorem 4.1 strengthens conditions on the weakness of the cross-sectional correlation; in particular, it requires that the covariance matrix converges to a diagonal one. The additional assumption in Theorem 4.2 requires that the weights used for averaging the cross-sectional entities are orthogonal to the loadings on the latent factor structure, which precludes the latent factor structure (that represents the cross-sectional dependence) from being amplified. This is a necessary assumption for consistency of the variance estimator. Indeed, let assumptions 1, 2, 3, 5 hold, and consider the first component of \( \xi_i \):

\[ \xi_{i}^{(1)} = \frac{1}{\sqrt{T}} \sum_{t} v_{t} \gamma_{i} e_{it} = \pi_{i} \gamma_{i}^{'} \Upsilon_{T} + \tilde{\eta}_{i}, \]

where \( \tilde{\eta}_{i} = \frac{1}{\sqrt{T}} \sum_{t} v_{t} \gamma_{i} \eta_{it} \), and \( \Upsilon_{T} = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} f_{t} v_{t} \). Note that \( \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \tilde{\eta}_{i} \Rightarrow \mathcal{N}(0, \sigma_{\eta}^{2}) \) and \( \frac{1}{N} \sum_{i=1}^{N} \tilde{\eta}_{i}^{2} \overset{p}{\to} \sigma_{\eta}^{2} \), as all conditions of Theorems 3.2 and 4.2 are satisfied by cross-sectionally and time independent errors \( \eta_{is} \). Assumption 5(iv) guarantees that \( \Upsilon_{T} \Rightarrow \mathcal{N}(0, \Sigma_{fv}) \) as \( T \to \infty \), while according to Assumption 5(ii), we have \( \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \pi_{i} \gamma_{i}^{'} \pi_{i} \to \Gamma_{\pi\gamma} \) as \( N \to \infty \). Thus,

\[ \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \xi_{i}^{(1)} \Rightarrow \mathcal{N}(0, \Gamma_{\pi\gamma} \Sigma_{fv} \Gamma_{\pi\gamma}^{'} + \sigma_{\eta}^{2}), \]

while \( \frac{1}{N} \sum_{i=1}^{N} (\xi_{i}^{(1)})^{2} \overset{p}{\to} \sigma_{\eta}^{2} \) because \( \frac{1}{N} \sum_{i=1}^{N} \pi_{i} \gamma_{i}^{'} \gamma_{i} \pi_{i}^{'} \to 0 \) by Assumptions 2 and 5(ii).

**4.2 Bootstrap inference**

As one way to conduct valid inferences in settings when \( \frac{1}{N} \sum_{i=1}^{N} \xi_{i} \xi_{i}^{'} \) is an inconsistent estimator of the variance (in particular, when \( \Gamma_{\pi\gamma} \neq 0 \) under Assumption 5), we propose the following simple wild bootstrap procedure. In each bootstrap repetition,
(i) simulate independent draws of random variables \( \delta_t \sim \{+1, -1\} \), \( t = 1, ..., T \), with success probability 1/2;

(ii) compute the bootstrap analogs of errors \( e^*_{it} = \delta_t e_{it} \), \( i = 1, ..., N \), \( t = 1, ..., T \);

(iii) compute the bootstrap analogs \( \xi^*_{it} \) using the definition of \( \xi_{it} \), with \( e^*_{it} \) in place of \( e_{it} \) for all \( i = 1, ..., N \), \( t = 1, ..., T \).

Then the distribution of \( \Xi^{(1)}_{N,T} = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \xi^{(1)}_{it} \) has the same asymptotic limit as that of \( \Xi_{N,T} \) has, and can be used for inferences. Alternatively, the distribution of \( \Xi^{(1)}_{N,T} \) normalized by the bootstrap analog of the variance estimate \( \frac{1}{N} \sum_{i=1}^{N} \xi_t \xi_t' \) can be used to approximate the distribution of \( \Xi_{N,T} \) normalized by the variance estimate \( \frac{1}{N} \sum_{i=1}^{N} \xi_t \xi_t' \). We call the two described bootstrap procedures bootstrap and bootstrap-t. In the next Section, we implement both variations of the wild bootstrap for the setup of Assumption 5.

The wild bootstrap works because it introduces independence in the time direction while preserving the (unknown) cross-sectional dependence. Specifically, we base our proof of Theorem 3.2 on the change of order of summations in. For example,

\[
\Xi^{(1)}_{N,T} = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \xi^{(1)}_{it} = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left( \frac{1}{\sqrt{N}} \sum_{i=1}^{N} v_t \gamma_i e_{it} \right).
\]

We then argue that our assumptions guarantee that the CLT with respect to summation over \( t \) is applicable. If the assumptions of the CLT hold, then \( \Xi^{(1)}_{N,T} \) is asymptotically Gaussian with mean zero and variance equal to the limit of \( \frac{1}{T} \sum_t \mathbb{E}(\frac{1}{\sqrt{N}} \sum v_t \gamma_i e_{it})^2 \). This limit clearly depends on how much there is cross-sectional correlation between \( e_{it} \) and \( e_{jt} \).

Now, in the bootstrapped samples,

\[
(\Xi^{(1)}_{N,T})^* = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (\xi^{(1)}_{it})^* = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left( \frac{1}{\sqrt{N}} \sum_{i=1}^{N} v_t \gamma_i e_{it} \right) \delta_t.
\]

Conditional on the original sample, only \( \delta_t \)'s are random, and they are independent and have zero mean and unit variance. When \( T \) is large, this bootstrapped normalized sum satisfies the CLT and thus converges to a zero mean Gaussian random variable with the variance equal to the limit of \( \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}(\frac{1}{\sqrt{N}} \sum v_t \gamma_i e_{it})^2 \) as \( N, T \to \infty \). This limit coincides with \( \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}(\frac{1}{\sqrt{N}} \sum v_t \gamma_i e_{it})^2 \) under relatively weak assumptions, as long as the Law of Large Numbers holds. For example, under Assumptions 5, we have:

\[
\lim \frac{1}{T} \sum_{t=1}^{T} \left( \frac{1}{\sqrt{N}} \sum_{i=1}^{N} v_t \gamma_i e_{it} \right)^2 = \lim \frac{1}{T} \sum_{t=1}^{T} \left( \Gamma_{\pi,\gamma} v_t f_t + \frac{1}{\sqrt{N}} \sum_{i=1}^{N} v_t \gamma_i \eta_t \right)^2 \xrightarrow{p} \Gamma_{\pi,\gamma} \Sigma f_t \Gamma_{\pi,\gamma} + \sigma^2_{\eta}.
\]
A similar argument can be made about the second component as well. Indeed,

$$(\Xi_{N,T}^{(2)})^* = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (\xi_i^{(2)})^* = \frac{1}{T} \sum_{s=1}^{T} \sum_{t<s} \left( \frac{1}{\sqrt{N}} \sum_{i=1}^{N} w_{st} \epsilon_i \epsilon_{is} \right) \delta_t \delta_s.$$ 

For a bootstrapped statistic, conditional on the original sample, the weights $\frac{1}{\sqrt{N}} \sum_{i=1}^{N} w_{st} \epsilon_i \epsilon_{is}$ are fixed, while $\delta_i$’s are independent random variables, and all the conditions of Lemma A.1 are satisfied. Thus, a zero mean Gaussian limit obtains as $T \to \infty$. It is straightforward to verify that its variance converges to the asymptotic variance of $\Xi_{N,T}^{(2)}$.

In the simulation experiments in the following Section, we check, among other things, that both wild bootstrap variations deliver correctly sized tests even when the asymptotic-$t$ tests fail to do so, and also that the bootstrap-based tests have a non-trivial power.

5 MONTE-CARLO SIMULATIONS

The goals of this Section are to check finite sample performance of an asymptotic Gaussian approximation for $\Xi_{N,T}$, to explore when the variance estimator $\hat{\Sigma}_\xi = \frac{1}{N} \sum_{i=1}^{N} \xi_i \xi_i'$ allows to construct reliable asymptotic-$t$ inferences, and to evaluate the performance of the wild bootstrap and bootstrap-$t$ procedures in terms of both size and power.

5.1 Setup

Our setup adheres to Assumption 5. We generate the errors $\epsilon_{it}$ according to the following weak (unobserved) factor structure:

$$\epsilon_{it} = \pi_i' f_t + \eta_{it},$$

where $f_t \sim iid \mathcal{N}(0,1)$ across $t = 1, ..., T$, $\gamma_i = 1$ for all $i = 1, ..., N$, and $\eta_{it} = \omega_i \epsilon_{\eta, it}$, where $\epsilon_{\eta, it}$ are $iid \mathcal{N}(0,1)$ across $i$ and $t$. The standard deviations are set to $\omega_i = c_\omega (1 + |\tau_i|)$, and the factor loadings are $\pi_i = (c_\pi + \tau_i) / \sqrt{N}$, where $\tau_i \sim iid \mathcal{N}(0,1)$ across $i = 1, ..., N$. The multiplier $c_\omega$ is tuned so that the average cross-sectional variance of $\eta_{it}$ is unity. The parameter $c_\pi$ indexes the degree of cross-sectional dependence as measured by the strength of factor structure. Specifically, as $\sum_{i=1}^{N} \pi_i \pi_i' \rightarrow 1 + c_\pi^2$, this parameter is assumed to be bounded for the Gaussian approximation to hold. We also have $\Gamma \pi \gamma = c_\pi$, hence we can expect consistency of variance estimation only when $c_\pi = 0$, so we will explore the distortions for different values of $c_\pi$. The errors generated this way are cross-sectionally dependent and heteroskedastic, while still satisfying Assumption 5.
The common variables are generated as follows: $v_t = c_{fv} f_t + \sqrt{1 - c_{fv}^2} \epsilon_{v,t}$, with $\epsilon_{v,t} \sim iid \ N(0,1)$ across $t = 1, ..., T$, and $w_{st} = v_s v_t$, $s, t = 1, ..., T$. All the disturbances $f_t$, $\tau_t$, $\epsilon_{\eta,it}$ and $\epsilon_{v,t}$ are mutually independent. The parameter $c_{fv}$ indexes the dependence between common variables and $e_{it}$. The mean zero Assumption 3(i) requires $c_{fv} = 0$; the non-zero values of $c_{fv}$ index deviations from the null hypothesis $E\xi_i = 0$, and will be used to study the power properties of the proposed wild bootstrap. In wild bootstrap samples, we generate the bootstrap errors by $e_{it}^* = \delta_t e_{it}$, where $\delta_t = 2\zeta_t - 1$, and $\zeta_t \sim iid \ B(\frac{1}{2})$ across $t, ..., T$. The bootstrap analogs of $\gamma_i$, $v_t$ and $w_{st}$ are set equal to their original sample values.

The distribution characteristics are computed from 10,000 simulations, while the rejection rates are based on 2,000 simulation runs. In all simulations, we set $N = T = 500$. The number of bootstrap repetitions is 600.

### 5.2 Results

Table 1 contains distributional characteristics of $\Xi_{N,T}$. We report averages, coefficients of skewness, coefficients of kurtosis, and right 5% quantiles of normalized marginal distributions of both elements of $\Xi_{N,T}$. For the exactly normal distributions, these values are 0, 0, 3 and 1.645, respectively.

<table>
<thead>
<tr>
<th>$c_{x}$</th>
<th>mean</th>
<th>skew</th>
<th>kurt</th>
<th>quant</th>
<th>mean</th>
<th>skew</th>
<th>kurt</th>
<th>quant</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>0.01</td>
<td>-0.05</td>
<td>3.01</td>
<td>1.63</td>
<td>-0.01</td>
<td>0.25</td>
<td>3.20</td>
<td>1.71</td>
</tr>
<tr>
<td>1</td>
<td>0.01</td>
<td>-0.04</td>
<td>2.97</td>
<td>1.64</td>
<td>-0.01</td>
<td>0.25</td>
<td>3.20</td>
<td>1.71</td>
</tr>
<tr>
<td>2</td>
<td>0.00</td>
<td>-0.00</td>
<td>2.95</td>
<td>1.65</td>
<td>0.02</td>
<td>0.20</td>
<td>3.11</td>
<td>1.71</td>
</tr>
</tbody>
</table>

**Notes:** Based on 10,000 simulations. The first component of $\Xi_{N,T}$ is labeled ‘linear’ and second component is labeled ‘quadratic’. ‘Mean’ stands for average, ‘skew’ for skewness coefficient, ‘kurt’ for kurtosis coefficient, and ‘quant’ for right 5% quantile of simulated marginal distribution of components of $\Xi_{N,T}$ normalized by corresponding standard deviations. Rows with $c_x = 0.5, 1$ and 2 correspond to very weak, weak and moderately strong error factor structure. The actual distribution of the linear component of $\Xi_{N,T}$ is very close to Gaussian, in all respects: all the moments and the right tail are almost equal to their theoretical counterparts. The quadratic component of $\Xi_{N,T}$, however, albeit mean unbiased, is somewhat positively skewed and a bit leptokurtic. The shifted right quantile confirms slight
over-dispersion. The distortions, however, do not seem to increase with the strength of the error factor structure.

In Table 2 we document the empirical rejection rates for tests with 10%, 5% and 1% declared size based on the asymptotic-t, wild bootstrap and wild bootstrap-t approaches. In the asymptotic-t approach we create a t-statistic using $\hat{\Sigma}_\xi$ as a variance estimator and compare it with the symmetric standard Gaussian critical values. We also explore the performance of the two wild bootstrap procedures – one that bootstraps $\Xi_{N,T}$ and another that bootstraps the t-statistics (referred to in Table 2 as bootstrap $\xi$ and bootstrap $t$). In both bootstrap procedures, we compare the absolute value of the statistic from the sample to the right quantile of the absolute value of the bootstrapped statistic. We verify the empirical size by setting $c_{fv} = 0$ and empirical power by setting $c_{fv} = 0.1$ for relatively small deviations from the null and $c_{fv} = 0.2$ for relatively large deviations from the null.

As expected, the size of the test based on asymptotic approximations sometimes deviate from nominal rates by a wide margin, especially for the linear component of $\xi_{N,T}$, the gap quickly increasing with the strength of the error factor structure. This happens due to inconsistency of the variance estimator $\hat{\Sigma}_\xi$ becomes more pronounced with stronger cross-sectional dependence. What is surprising is that the asymptotic approximation for the quadratic component of $\xi_{N,T}$, the distortions are relatively minor and not very sensitive to the strength of factor structure. In contrast, both bootstrap procedures exhibit excellent size control and its stability across the strength of the error factor structure for both components of $\xi_{N,T}$. In terms of power, however, the two bootstrap statistics are approximately equally powerful for the linear component of $\xi_{N,T}$, while there is a gap, sometimes sizable, between power figures for its quadratic component. It seems that bootstrapping the statistic itself is preferable.

6 CONCLUDING REMARKS

Possible directions for future research may be relaxing the error time-series independence to martingale difference structures and inventing ways to consistently estimate the asymptotic variance matrix when it is not diagonal in the limit. Another glorious research venue is establishing formal properties of the proposed wild bootstrap schemes, such as their asymptotic validity and possibility of asymptotic refinements when the asymptotic testing is valid too, as well as superiority of one bootstrap scheme over the other.
Table 2: Simulated rejection rates for asymptotic and wild bootstrap tests

<table>
<thead>
<tr>
<th>element →</th>
<th>linear</th>
<th>quadratic</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1%</td>
<td>5%</td>
</tr>
<tr>
<td>$c_\pi \downarrow$</td>
<td>asy bootstrap</td>
<td>asy bootstrap</td>
</tr>
<tr>
<td></td>
<td>$t$</td>
<td>$\xi$</td>
</tr>
<tr>
<td>Size: $c_{fv} = 0$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>26</td>
<td>1</td>
</tr>
<tr>
<td>Power: $c_{fv} = 0.1$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>9</td>
<td>6</td>
</tr>
<tr>
<td>1</td>
<td>41</td>
<td>17</td>
</tr>
<tr>
<td>2</td>
<td>80</td>
<td>29</td>
</tr>
<tr>
<td>Power: $c_{fv} = 0.2$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>37</td>
<td>30</td>
</tr>
<tr>
<td>1</td>
<td>90</td>
<td>73</td>
</tr>
<tr>
<td>2</td>
<td>100</td>
<td>92</td>
</tr>
</tbody>
</table>

Notes: The table contains actual rates, computed from 2,000 simulations for 10%, 5% and 1% declared size asymptotic-$t$ (‘asy $t$’), bootstrap (‘bootstrap $\xi$’) and bootstrap-$t$ (‘bootstrap $t$’) two-sided tests for deviations of each component of $\Xi_{N,T}$ from the zero value. The first component of $\Xi_{N,T}$ is labeled ‘linear’ and second component is labeled ‘quadratic’. The size figures are in panel with $c_{fv} = 0$, and power figures are in panels with $c_{fv} \neq 0$. Rows with $c_\pi = 0.5$, 1 and 2 correspond to very weak, weak and moderately strong error factor structure.
A APPENDIX: PROOFS

A.1 Preliminary results

We use the following central limit theorem for vector-valued martingale-difference sequence:

**Lemma A.1** Let the sequence \((Z_{t,T}, F_{t,T}), t = 1, \ldots, T,\) be a martingale difference sequence of \(r \times 1\) random vectors with \(\Sigma_T = \text{var} \left( \sum_{t=1}^T Z_{t,T} \right)\). If the following two conditions hold as \(T \to \infty\),

\[
(1) \quad (\min \text{ev}(\Sigma_T))^{-2} \sum_{t=1}^T \mathbb{E} \left[ \|Z_{t,T}\|^4 \right] \to 0,
\]

\[
(2) \quad (\min \text{ev}(\Sigma_T))^{-2} \mathbb{E} \left[ \left\| \sum_{t=1}^T Z_{t,T} Z_{t,T}' - \Sigma_T \right\|^2 \right] \to 0,
\]

then, as \(T \to \infty\),

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

**Proof of Lemma A.1** Indeed, the statement of Lemma A.1 holds if for any non-random \(r \times 1\) vector \(\lambda\), we have \((\lambda' \Sigma_T \lambda)^{-1/2} \sum_{t=1}^T \lambda' Z_{t,T} \Rightarrow \mathcal{N}(0, 1)\). Let us define a scalar martingale difference sequence \(z_t = \lambda' Z_{t,T}\) with variance \(\sigma^2_T = \text{var} \left( \sum_{t=1}^T \lambda' Z_{t,T} \right) = \lambda' \Sigma_T \lambda\). Let us check that all conditions of the central limit theorem by Heyde & Brown (1970) are satisfied for \(\delta = 1\). Indeed,

\[
\frac{1}{\sigma^2_T} \sum_{t=1}^T \mathbb{E} \left[ |z_t|^4 \right] = \frac{1}{(\lambda' \Sigma_T \lambda)^2} \sum_{t=1}^T \mathbb{E} \left[ |\lambda' Z_{t,T}|^4 \right] \leq \frac{1}{(\|\lambda\|^2 \min \text{ev}(\Sigma_T))^2} \sum_{t=1}^T \|\lambda\|^4 \mathbb{E} \left[ \|Z_{t,T}\|^4 \right] \to 0,
\]

and

\[
\mathbb{E} \left[ \left( \frac{\sum_{t=1}^T z_t^2}{\sigma^2_T} - 1 \right)^2 \right] = \mathbb{E} \left[ \left( \frac{\sum_{t=1}^T (\lambda' Z_{t,T})^2}{\lambda' \Sigma_T \lambda} - 1 \right)^2 \right]
\]

\[
= \frac{1}{(\lambda' \Sigma_T \lambda)^2} \mathbb{E} \left[ \lambda' \left( \sum_{t=1}^T Z_{t,T} Z_{t,T}' - \Sigma_T \right) \lambda \right]^2
\]

\[
\leq \frac{1}{(\|\lambda\|^2 \min \text{ev}(\Sigma_T))^2} \mathbb{E} \left[ \left\| \sum_{t=1}^T Z_{t,T} Z_{t,T}' - \Sigma_T \right\|^2 \right] \to 0.
\]

These two conditions imply that \(\sigma^{-1}_T \sum_{t=1}^T z_t \Rightarrow \mathcal{N}(0, 1)\). This finishes the proof. ■
As a preliminary result, we establish a central limit theorem for quadratic forms. The idea of this result comes from the CLT for quadratic forms by de Jong (1987). All random variables are implicitly indexed by the sample sizes $T$ (or $N,T$ in the further application to factor models), which are omitted to reduce clutter; for example, $W_{st}$ in full notation is indexed as $W_{st,T}$ or $W_{st,N,T}$.

**Lemma A.2** 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 $X_{st}$ is a random vector measurable with respect to the $\sigma$-algebra $\mathcal{F}$, and all $e_t$ are independent from each other, conditionally on $\mathcal{F}$. Assume that $E(W_{st}|\mathcal{F},e_t) = 0$ and $E(W_{st}|\mathcal{F},e_s) = 0$.

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

(i) $\Sigma_{W,T} \rightarrow \Sigma_W$, where $\Sigma_W$ is a full rank matrix;

(ii) $T^4 \max_{1 \leq t,s \leq T} E[||W_{st}||^4] < C$;

(iii) $E\left[\left(\sum_{s=1}^{T} \sum_{t<s} W_{st}W_{st}' - \Sigma_{W,T}\right)^2\right] \rightarrow 0$;

(iv) $T^4 \max_{s_1 \neq s_2,t_1 \neq t_2} E\left|\left|\left(\sum_{s_1}^{s_2} W_{s_1t_1}W_{s_2t_2}'W_{s_1t_1}W_{s_2t_2}'W_{s_1t_1}W_{s_2t_2}'\right)\right|\right| \rightarrow 0$.

Then, as $T \rightarrow \infty$, $W(T) \Rightarrow \mathcal{N}(0,\Sigma_W)$.

**Lemma A.3** Let $W_{st} = W_{st}(X_{st},e_s,e_t)$ satisfy all conditions of Lemma A.2. Let $V_s = V_s(X_s,e_s)$ be a random vector defined for all $s \in \{1,...,T\}$ such that $X_s$ is a random vector measurable with respect to the $\sigma$-algebra $\mathcal{F}$, and $E(V_s|\mathcal{F}) = 0$. Define $V(T) = \sum_{s=1}^{T} V_s$ and $\Sigma_{V,T} = \text{var}(V(T))$. Assume the following statements hold as $T \rightarrow \infty$:

(a) $\Sigma_{V,T} \rightarrow \Sigma_V$, where $\Sigma_V$ is a full rank matrix;

(b) $T \max_{1 \leq s \leq T} E[||V_s||^4] \rightarrow 0$;

(c) $E\left[\left(\sum_{s=1}^{T} V_sV_s' - \Sigma_{V,T}\right)^2\right] \rightarrow 0$;

(d) $T^3 \max_{1 \leq t < \min\{s_1,s_2\} \leq T} \left|\left|E \left( W_{s_1t}V_{s_1}V_{s_2}W_{s_2t} \right) \right|\right| \rightarrow 0$.

Then, as $T \rightarrow \infty$

\[
\begin{pmatrix} V(T) \\ W(T) \end{pmatrix} \Rightarrow \mathcal{N} \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma_V & 0 \\ 0 & \Sigma_W \end{pmatrix} \right).
\]
Proof of Lemma A.2 The proof of this Lemma follows closely the proof and ideas stated in de Jong (1987). Call $W_{st}$ clean if

$$\mathbb{E}(W_{s_1t_1} \otimes W_{s_2t_2} \otimes \ldots \otimes W_{s_kt_k}) = 0$$

when at least one index from the set $\{s_1, t_1, \ldots, s_k, t_k\}$ has a value that occurs only once. The functional form of $W_{st}$ and the condition stated in (A.1) guarantee that in our case $W_{st}$ is clean. Indeed, if, for example, the index $s_1$ occurs only once, then

$$\mathbb{E}(W_{s_1t_1} \otimes W_{s_2t_2} \otimes \ldots \otimes W_{s_kt_k}) = \mathbb{E}[\mathbb{E}(W_{s_1t_1} \otimes W_{s_2t_2} \otimes \ldots \otimes W_{s_kt_k} | F, e_{t_1}, e_{s_2}, e_{t_2}, \ldots, e_{t_k})]
= \mathbb{E}[\mathbb{E}(W_{s_1t_1} | F, e_{t_1}, e_{s_2}, e_{t_2}, \ldots, e_{t_k}) \otimes W_{s_2t_2} \otimes \ldots \otimes W_{s_kt_k}]
= \mathbb{E}[\mathbb{E}(W_{s_1t_1} | F, e_{t_1}) \otimes W_{s_2t_2} \otimes \ldots \otimes W_{s_kt_k}] = 0.$$ 

Now, $W(T) = \sum_{s=1}^{T} \sum_{t<s} W_{st} = \sum_{s=1}^{T} Z_{s,T}$, where $Z_{s,T} = \sum_{t<s} W_{st}$. We denote by $F_s$ the $\sigma$-algebra generated by $F$ and $e_t$ for all $t < s$. Then, $(Z_{s,T}, F_s)$ is a martingale difference sequence. Below we check that all conditions of Lemma A.1 are satisfied.

Condition (i) implies that $\min \text{ev} (\Sigma_{W,T}) \to C > 0$. Now let us check condition (1) of Lemma A.1:

$$\mathbb{E} \left[ \left\| Z_{s,T} \right\|^4 \right] = \mathbb{E} \left[ \left\| \sum_{t<s} W_{st} \right\|^4 \right]
= \mathbb{E} \left[ \left( \sum_{t_1<s} W_{st_1} \right)^4 \left( \sum_{t_2<s} W_{st_2} \right)^4 \left( \sum_{t_3<s} W_{st_3} \right)^4 \left( \sum_{t_4<s} W_{st_4} \right)^4 \right]
\leq \sum_{t<s} \mathbb{E} \left[ \left\| W_{st} \right\|^4 \right] + C \sum_{t_1<s} \sum_{t_2<s} \sum_{t_3<s} \sum_{t_4<s} \mathbb{E} \left[ \left\| W_{st1} \right\|^2 \left\| W_{st2} \right\|^2 \right].$$

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 distinct elements each occurring twice. We also notice that $\mathbb{E}[\left\| W_{st1} \right\|^2 \left\| W_{st2} \right\|^2] \leq \frac{1}{2} (\mathbb{E}[\left\| W_{st1} \right\|^4] + \mathbb{E}[\left\| W_{st2} \right\|^4]) \leq \max_{1 \leq t, s \leq T} \mathbb{E}[\left\| W_{st} \right\|^4] < C T^{-4}$ due to condition (ii). Hence, $\mathbb{E}[\left\| Z_{s,T} \right\|^4] \leq C T^{-2}$. Thus, $\sum_{s=1}^{T} E[\left\| Z_{s,T} \right\|^4] \leq C T^{-1}$, implying that condition (1) of Lemma A.1 holds.

Now let us turn to condition (2). First, notice that

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

22
the last equality holding because $W_{st}$ is clean. Next,

$$
\mathbb{E} \left[ \sum_{s=1}^{T} \left( \sum_{t_1<s} W_{st_1} \right) \left( \sum_{t_2<s} W_{st_2} \right)' - \Sigma_{W,T} \right]_F^2
$$

$$
= \mathbb{E} \left[ \sum_{s=1}^{T} \sum_{t_1<s} (W_{st_1}' - \mathbb{E}[W_{st_1}']) + \sum_{s=1}^{T} \sum_{t_1 \neq t_2} W_{st_1} W_{st_2}' \right]_F^2
$$

$$
= \mathbb{E} \left[ \sum_{s=1}^{T} \sum_{t_1<s} (W_{st_1}' - \mathbb{E}[W_{st_1}']) \right]_F^2 + \mathbb{E} \left[ \sum_{s=1}^{T} \sum_{t_1 \neq t_2} W_{st_1} W_{st_2}' \right]_F^2.
$$

The last equality holds because of the clean form, as the expectation of the Frobenius norm is equal to the trace of the sums of various 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, the first summand in equation (A.2) converges to zero due to condition (iii) of the Lemma. Now consider the second term in (A.2):

$$
\mathbb{E} \left[ \sum_{s=1}^{T} \sum_{t_1 \neq t_2<s} W_{st_1} W_{st_2}' \right]_F^2 = \sum_{s_1=1}^{T} \sum_{s_2=1}^{T} \sum_{t_1 \neq t_2} \sum_{t_3 \neq t_4} \mathbb{E} \left[ \text{tr} \left( W_{s_1 t_1} W_{s_2 t_2}' W_{s_2 t_3} W_{s_3 t_4}' \right) \right]
$$

$$
= C \sum_{s_1=1}^{T} \sum_{s_2=1}^{T} \sum_{t_1 \neq t_2} \mathbb{E} \left[ \text{tr} \left( W_{s_1 t_1} W_{s_2 t_2}' W_{s_2 t_1} W_{s_1 t_2}' \right) \right],
$$

the last equality holding because $W_{st}$ is clean. The last summation can be divided into a category when $s_1 \neq s_2$, the corresponding sum being asymptotically $o(1)$ due to condition (iv), and a category when $s_1 = s_2$, there being at most $CT^3$ of such summands, each smaller than $C \max_{1 \leq t, s \leq T} \mathbb{E} [||W_{st}||^4] < CT^{-4}$. Thus,

$$
\mathbb{E} \left[ \sum_{s=1}^{T} \sum_{t_1 \neq t_2} W_{st_1} W_{st_2}' \right]_F^2 \rightarrow 0.
$$

Putting statements (A.2) and (A.3) together we obtain that condition (2) of Lemma A.1 is satisfied. Thus, the conclusion of Lemma A.2 holds.

Proof of Lemma A.3. Let us define $Z_s = (V_s', \sum_{t<s} W_{st}')$, and let $F_s$ be defined as in the proof of Lemma A.2. We will show that all conditions of Lemma A.1 are satisfied.
Notice that
\[\mathbb{E}[V_s W'_s] = \mathbb{E}(\mathbb{E}[V_s W'_s | F, e_s]) = \mathbb{E}(V_s \mathbb{E}[W'_s | F, e_s]) = 0.\]

Thus,
\[\Sigma_T = \text{var} \left( \sum_{s=1}^{T} Z_s \right) = \begin{pmatrix} \Sigma_{V,T} & 0 \\ 0 & \Sigma_{W,T} \end{pmatrix} \rightarrow \begin{pmatrix} \Sigma_V & 0 \\ 0 & \Sigma_W \end{pmatrix}.\]

The right-hand-side is a full rank matrix by condition (i) of Lemma A.2 and condition (a) of Lemma A.3. Thus, the minimal eigenvalue of \(\Sigma_T\) is separated away from zero for large \(T\). Now,
\[\sum_{s=1}^{T} \mathbb{E} \left[ \|Z_s\|^4 \right] \leq C \sum_{s=1}^{T} \mathbb{E} \left[ \|V_s\|^4 \right] + C \sum_{s=1}^{T} \mathbb{E} \left[ \left\| \sum_{t<s} W_{st} \right\|^4 \right].\]

The first term here is bounded by \(T \max_{1 \leq s \leq T} \mathbb{E} \left[ \|V_s\|^4 \right]\) which goes to zero by condition (b) of Lemma A.3, while convergence to zero of the second sum has been already shown during the proof of Lemma A.2. Thus, condition (1) of Lemma A.1 holds. Next,
\[
\mathbb{E} \left[ \left\| \sum_{s=1}^{T} Z_s Z'_s - \Sigma_T \right\|^2 \right] \leq C \mathbb{E} \left[ \left\| \sum_{s=1}^{T} Z_s Z'_s - \Sigma_T \right\|^2 \right]_F
\[
= \mathbb{E} \left[ \left\| \sum_{s=1}^{T} V_s V'_s - \Sigma_{V,T} \right\|^2 \right]_F + 2C \mathbb{E} \left[ \left\| \sum_{s=1}^{T} \left( \sum_{t<s} W_{st} \right) V'_s \right\|^2 \right]_F
\[
+ \mathbb{E} \left[ \left\| \sum_{s=1}^{T} \left( \sum_{t<s} W_{st} \right) \left( \sum_{t<s} W_{st} \right)' - \Sigma_{W,T} \right\|^2 \right]_F.
\]

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. Condition (c) guarantees that
\[\mathbb{E} \left[ \left\| \sum_{s=1}^{T} V_s V'_s - \Sigma_{V,T} \right\|^2 \right]_F \leq C \mathbb{E} \left[ \left\| \sum_{s=1}^{T} V_s V'_s - \Sigma_{V,T} \right\|^2 \right] \rightarrow 0.
\]

During the proof of Lemma A.2 we show that
\[\mathbb{E} \left[ \left\| \sum_{s=1}^{T} \left( \sum_{t<s} W_{st} \right) \left( \sum_{t<s} W_{st} \right)' - \Sigma_{W,T} \right\|^2 \right]_F \rightarrow 0.\]
Finally,

\[
\mathbb{E} \left[ \left\| \sum_{s=1}^{T} \left( \sum_{t<s} W_{st} \right) V_s' \right\|^2 \right] = \sum_{s_1=1}^{T} \sum_{t_1<s_1} \sum_{s_2=1}^{T} \sum_{t_2<s_2} \text{tr} \left( \mathbb{E} \left( W_{s_1 t_1} V_{s_1} V_{s_2} W_{s_2 t_2} V_s' \right) \right)
\]

\[
= T \sum_{s=1}^{T} \sum_{s=1}^{T} \sum_{t=1}^{T} \sum_{t \leq \min \{s_1, s_2\}} \text{tr} \left( \mathbb{E} \left( W_{s_1 t} V_{s_1}^\prime V_{s_2} W_{s_2 t}^\prime \right) \right)
\]

\[
\leq CT^3 \max_{1 \leq s_1, s_2, t \leq T} \left\| \mathbb{E} \left( W_{s_1 t} V_{s_1} V_{s_2} W_{s_2 t} \right) \right\| \to 0.
\]

Here we used that \( \mathbb{E} \left( W_{s_1 t_1} V_{s_1} V_{s_2} W_{s_2 t_2} V_s' \right) = 0 \) if \( t_1 \neq t_2 \) and condition (d) of the Lemma.

To conclude, condition (2) of Lemma A.1 also holds. ■

**Lemma A.4** For an \( N \times N \) symmetric matrix \( A = (a_{ij}) \) denote \( \odot \) to be the Hadamard product. Then \( \|A \odot A\| \leq \sqrt{N} \|A\|^2 \).

**Proof.** Using the equivalence of norms, we have

\[
\|A \odot A\| \leq \|A \odot A\|_F \leq \sqrt{\sum_{1 \leq i, j \leq N} a_{ij}^4} \leq \sqrt{\max_{1 \leq i, j \leq N} a_{ij}^2} \sqrt{\sum_{1 \leq i, j \leq N} a_{ij}^2} \leq \|A\| \|A\|_F \leq \sqrt{N} \|A\|^2.
\]

■

### A.2 Proofs for Independent case

**Proof of Theorem 3.1.** We will check that all conditions of Lemma A.3 are satisfied for

\[
W_{st} = \frac{1}{T \sqrt{N}} w_{st} \sum_{i=1}^{N} e_{it} e_{is} = \frac{1}{T} w_{st} \frac{e_{it} e_{is}}{\sqrt{N}}
\]

and

\[
V_s = \frac{1}{\sqrt{T N}} \sum_{i=1}^{N} \gamma_i e_{is} \otimes v_s = \frac{1}{\sqrt{T \sqrt{N}}} \gamma_{e_s} \otimes v_s.
\]

(i) First notice that due to Assumption 4(ii)

\[
\mathbb{E} \left[ \left( \frac{e_{it} e_{is}}{\sqrt{N}} \right)^2 \right] = \frac{\text{tr}(\mathcal{E}^2_{N,T})}{N} \to a.
\]

Due to the independence between the common variables and \( e_{it} \) and because \( W_{st} \) is clean, we have:

\[
\Sigma_{W,T} = \text{var} \left( \sum_{s=1}^{T} \sum_{t<s} W_{st} \right) = \frac{1}{T^2} \sum_{s=1}^{T} \sum_{t<s} \mathbb{E}(w_{st} w_{st}') \mathbb{E} \left[ \left( \frac{e_{it} e_{is}}{\sqrt{N}} \right)^2 \right] \to a \Omega_w,
\]

25
and the limit is a positive definite matrix.

(ii) By Assumption 4(i) and the i.i.d. nature of $e_t$, we have:

$$T^4 \mathbb{E} \left[ \|W_{st}\|^4 \right] = \mathbb{E} \left[ \|w_{st}\|^4 \right] \mathbb{E} \left[ \left( \frac{e_t e_t'}{\sqrt{N}} \right)^4 \right] \leq \frac{C}{N^2} \sum_{i_1=1}^N \sum_{i_2=1}^N \sum_{i_3=1}^N \sum_{i_4=1}^N \mathbb{E} (e_{i_1 i_2} e_{i_3 i_4})^2 < C.$$ 

Here we used that $|\mathbb{E} (e_{i_1 i_2} e_{i_3 i_4})| \leq \max_{1 \leq i \leq N, 1 \leq t \leq T} \mathbb{E} (e_t^4) < C$ and Assumption 4(iv).

(iii) Next,

$$T \sum_{s=1}^T \sum_{t<s} W_{st} W_{st}' - \Sigma_{W,T} = \frac{1}{T^2} \sum_{s=1}^T \sum_{t<s} w_{st} w_{st}' \left[ \left( \frac{e_s e_t}{\sqrt{N}} \right)^2 - \frac{1}{N} \text{tr}(\mathcal{E}_{N,T}^2) \right] + \frac{1}{N} \text{tr}(\mathcal{E}_{N,T}^2) \left[ \frac{1}{T^2} \sum_{s=1}^T \sum_{t<s} (w_{st} w_{st}' - \mathbb{E} (w_{st} w_{st}')) \right] = A_1 + A_2,$$

hence it is enough to prove that $\mathbb{E} \|A_1\|^2 \to 0$ and $\mathbb{E} \|A_2\|^2 \to 0$. The latter is postulated by Assumption 1(iii). Notice that all summands in $A_1$ are uncorrelated with each other due to Assumptions 3(i) and 4(i). Thus,

$$\mathbb{E} [\text{tr}(A_1 A_1')] = \frac{1}{T^4} \sum_{s=1}^T \sum_{t<s} \mathbb{E} \left[ \|w_{st}\|^4 \right] \mathbb{E} \left[ \left( \frac{e_s e_t}{\sqrt{N}} \right)^2 - \frac{\text{tr}(\mathcal{E}_{N,T}^2)}{N} \right]^2 \leq \frac{1}{T^4} \sum_{s=1}^T \sum_{t<s} \mathbb{E} \left[ \|w_{st}\|^4 \right] \mathbb{E} \left( \frac{e_s e_t}{\sqrt{N}} \right)^4 \leq \frac{C}{T^2}.$$ 

In the last inequality, we use the proof of statement (ii) above. This implies that condition (iii) of Lemma A.2 holds.

(iv) If the set $\{s_1, s_2, t_1, t_2\}$ contains four distinct indexes, then

$$T^4 \left| \mathbb{E} \left( W^r_{s_1 t_2} W_{s_2 t_1} W^r_{s_2 t_2} W_{s_1 t_1} \right) \right| \leq \mathbb{E} \left[ \|w_{st}\|^4 \right] \text{tr} \left( \mathbb{E} \left( e_{s_1 t_1} e_{t_1} e_{s_2 t_2} e_{t_2} e_{s_1 t_1} e_{t_1}' e_{s_2 t_2} e_{t_2}' \right) \right) \leq \frac{C}{N^2} \leq \frac{C}{N^2} \leq \frac{C}{N^2} \leq \frac{C}{N} \to 0.$$ 

We now move to conditions (a)-(d) of Lemma A.3.

(a) By Assumptions 4(iii) and 1(i) we have

$$\Sigma_{V,T} = \left( \frac{1}{N} \gamma' \mathcal{E}_{N,T} \gamma \right) \otimes \left( \frac{1}{T} \sum_{s=1}^T \mathbb{E} (v_{s} v_{s}') \right) \to \Gamma_{\sigma} \otimes \Omega_{v},$$

and the limit is a full rank matrix.
(b) Next,
\[ T \mathbb{E} [\| V_s \|^4] = \frac{1}{T} \mathbb{E} \left[ \left\| \frac{1}{\sqrt{N}} \gamma' e_s \right\|^4 \right] \mathbb{E} [\| v_s \|^4], \]
where \( \mathbb{E} [\| v_s \|^4] \leq C \) due to Assumption 1(ii). Assumptions 2 and 4(iv) imply that
\[ \mathbb{E} \left[ \left\| \frac{1}{\sqrt{N}} \gamma' e_s \right\|^4 \right] \leq C \] due to Assumption 1(ii). Assumptions 2 and 4(iv) imply that
\[ \mathbb{E} \left[ \left\| \frac{1}{\sqrt{N}} \gamma' e_s \right\|^4 \right] = \frac{1}{N^2} \sum_{i_1=1}^{N} \sum_{i_2=1}^{N} \sum_{i_3=1}^{N} \sum_{i_4=1}^{N} \mathbb{E} \left( e_{i_1 t} e_{i_2 t} e_{i_3 t} e_{i_4 t} \right) \gamma'_{i_1} \gamma'_{i_2} \gamma'_{i_3} \gamma'_{i_4} \]
\[ < \max_{1 \leq i \leq N} \| \gamma_i \|^4 \frac{1}{N^2} \sum_{i_1=1}^{N} \sum_{i_2=1}^{N} \sum_{i_3=1}^{N} \sum_{i_4=1}^{N} \mathbb{E} \left( e_{i_1 t} e_{i_2 t} e_{i_3 t} e_{i_4 t} \right) \] \[ < C. \]

(c) Next,
\[ \sum_{s=1}^{T} V_s V'_s - \Sigma_{V,T} = \left( \frac{1}{N} \gamma' \mathcal{E}_{N,T} \gamma \right) \otimes \left( \frac{1}{T} \sum_{s=1}^{T} (v_s v'_s - \mathbb{E}(v_s v'_s)) \right) \]
\[ + \frac{1}{T} \sum_{s=1}^{T} \left( \frac{\gamma' e_s e'_s}{N} - \frac{\gamma' \mathcal{E}_{N,T} \gamma}{N} \right) \otimes (v_s v'_s) \]
\[ = A_1 + A_2. \]

Notice that \( A_1 \) and \( A_2 \) are uncorrelated, hence
\[ \mathbb{E} \left[ \left\| \sum_{s=1}^{T} V_s V'_s - \Sigma_{V,T} \right\|_F^2 \right] = \text{tr} \left( \mathbb{E}(A'_1 A_1) \right) + \text{tr} \left( \mathbb{E}(A'_2 A_2) \right). \]

Assumption 1(iv) guarantees the convergence of the first term. Notice that the summands in \( A_2 \) are uncorrelated due to time independence of errors, hence
\[ \text{tr} \left( \mathbb{E}(A'_2 A_2) \right) = \frac{1}{T^2} \sum_{s=1}^{T} \mathbb{E} \left[ \left\| \gamma' e_s e'_s \gamma - \frac{1}{N} \gamma' \mathcal{E}_{N,T} \gamma \right\|_F^2 \right] \mathbb{E} [\| v_s \|^4] \]
\[ \leq C \frac{1}{T} \mathbb{E} \left[ \left\| \gamma' e_s e'_s \gamma - \frac{1}{N} \gamma' \mathcal{E}_{N,T} \gamma \right\|_F^2 \right]. \]

Given the bounds on the fourth moment of \( N^{-1/2} \gamma' e_s \) derived in the proof of part (b) we get that condition (c) holds.

(d) By Assumption 4(i) we have that
\[ T^3 \mathbb{E} \left( W_{s_1 t} V'_{s_1} V_{s_2} V'_{s_2 t} \right) = \mathbb{E} \left( w_{s_1 t} v'_{s_1} v_{s_2} w'_{s_2 t} \right) \mathbb{E} \left( e'_{s_1 t} \gamma' e_{s_2} e'_{s_1 t} e_{s_2 t} \right). \]
Using that scalars can be reshuffled to make two $e_t$ with the same index stand back to back and employing time series independence of errors, we obtain that

$$\left| \mathbb{E} \left( \frac{e'_{s_1} \gamma' e_{s_2} e'_{s_2} e_t}{\sqrt{N}} \sqrt{N} \frac{e'_t}{\sqrt{N}} \sqrt{N} \frac{e'_s}{\sqrt{N}} \right) \right| = \frac{1}{N^2} \left| \text{tr} \left( \gamma' \mathbb{E} (e_{s_2} e'_{s_2}) \mathbb{E} (e'_t e'_t) \mathbb{E} (e_{s_1} e'_{s_1}) \right) \right| \leq \frac{1}{N^2} \text{tr} (\gamma' \gamma) \max \text{ev} (\mathcal{E}_{N,T}^3) \leq \frac{C}{N}.$$  

Here we use the assumption 2 to get $N^{-1} \text{tr} (\gamma' \gamma) < C$ and Assumption 4(ii). Given Assumption 1(ii) we obtain that

$$T^3 \max_{1 \leq t < \min \{s_1, s_2\} \leq T} \left\| \mathbb{E} \left( W_{s_1} V_{s_1} V_{s_2} W'_{s_2} \right) \right\| \leq \frac{C}{N} \to 0.$$  

Thus, condition (d) of Lemma A.3 is satisfied. This concludes the proof of Theorem 3.1.

\[\blacksquare\]

**Proof of Theorem 4.1.** We will prove the following three statements for

$$\xi_{V,i} = \frac{1}{\sqrt{T}} \sum_{s=1}^{T} \gamma_i e_{is} \otimes v_s$$

and

$$\xi_{W,i} = \frac{1}{T} \sum_{s=1}^{T} \sum_{t<s} w_{st} e_{it} e_{is};$$

(i) $N^{-1} \sum_{i=1}^{N} \xi_{V,i} \xi_{V,i}' \xrightarrow{p} \Sigma_V$;

(ii) $N^{-1} \sum_{i=1}^{N} \xi_{W,i} \xi_{W,i}' \xrightarrow{p} \Sigma_W$;

(iii) $N^{-1} \sum_{i=1}^{N} \xi_{V,i} \xi_{W,i}' \xrightarrow{p} 0$.

Let us start with statement (i). Denote by $\sigma_i^2$ the diagonal and by $\sigma_{ij}$ the off-diagonal elements of matrix $\mathcal{E}_{N,T}$. Notice that the additional assumption of Theorem 4.1 implies that

$$\Gamma_\sigma = \lim \frac{\gamma' \mathcal{E}_{N,T} \gamma}{N} = \lim \frac{1}{N} \sum_{i=1}^{N} \gamma_i \gamma_i' \sigma_i^2.$$  

28
Let us define \( \tilde{\Sigma}_{V,T} = \left( N^{-1} \sum_{i=1}^{N} \gamma_i \gamma_i' \sigma_i^2 \right) \left( T^{-1} \sum_{s=1}^{T} \mathbb{E} \left( v_s v_s' \right) \right) \), and notice that \( \tilde{\Sigma}_{V,T} \rightarrow \Sigma_V \). Thus,

\[
\frac{1}{N} \sum_{i=1}^{N} \xi_{V,i} \xi_{V,i} - \tilde{\Sigma}_{V,T} = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} \left( (\gamma_i \gamma_i' e_{is} e_{it}) \otimes (v_s v_t') - \mathbb{I} \{ s = t \} \sigma_i^2 (\gamma_i \gamma_i') \otimes \mathbb{E} (v_t v_t') \right)
\]

\[
= \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} (e_{it}^2 - \sigma_i^2) (\gamma_i \gamma_i') \otimes (v_t v_t')
\]

\[
+ \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{s \neq t} (\gamma_i \gamma_i' e_{is} e_{it}) \otimes (v_s v_t')
\]

\[
+ \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} (\gamma_i \gamma_i' \sigma_i^2) \otimes (v_t v_t' - \mathbb{E} (v_t v_t'))
\]

\[
= A_1 + A_2 + A_3.
\]

Notice that the three terms are uncorrelated, so it is enough to prove that \( \text{tr} \left( \mathbb{E} (A_j A_j') \right) \rightarrow 0 \) for \( j = 1, 2, 3 \). First,

\[
\text{tr} \left( \mathbb{E} (A_1 A_1') \right) = \text{tr} \left( \left[ \frac{1}{N^2 T^2} \sum_{i,j=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} (\gamma_i \gamma_i' \gamma_j \gamma_j' (e_{it}^2 - \sigma_i^2)(e_{js}^2 - \sigma_i^2)) \otimes (v_t v_t' v_s v_s') \right] \right)
\]

\[
= \frac{1}{N^2 T^2} \sum_{i,j=1}^{N} \sum_{t=1}^{T} \text{tr} \left( \gamma_i \gamma_i' \gamma_j \gamma_j' \text{cov}(e_{it}, e_{jt}) \right) \text{tr} \left( \mathbb{E} (v_t v_t' v_t v_t') \right)
\]

\[
\leq \frac{1}{T^2} \sum_{t=1}^{T} \max_{1 \leq i \leq N} \| \gamma_i \|^4 \max_{1 \leq i \leq N} \mathbb{E} \left[ (e_{it}^2 - \sigma_i^2)^2 \right] \mathbb{E} \left[ \| v_t \|^4 \right] \leq \frac{C}{T}. \]

Here we used that \( e_{it}'s \) are independent from each other for different \( t \) by Assumption 3(i), which forces \( s = t \). The last inequality uses Assumptions 1(ii), 3(ii) and 2.

Consider the term \( A_2 \) and notice that any two summands in the two-directional sum (over \( t \) and over \( s \)) are uncorrelated due to time series independence of \( e_i's \) and all summands are mean zero. Thus,

\[
\text{tr} \left( \mathbb{E} (A_2 A_2') \right) = \frac{1}{N^2 T^2} \sum_{t=1}^{T} \sum_{s \neq t} \sum_{i,j=1}^{N} \text{tr} \left( \mathbb{E} (\gamma_i \gamma_i' \gamma_j \gamma_j' e_{is} e_{it} e_{js} e_{jt}) \otimes \mathbb{E} (v_t v_t' v_t v_t') \right)
\]

\[
= \frac{1}{N^2} \sum_{i,j=1}^{N} \text{tr}(\gamma_i \gamma_i' \gamma_j \gamma_j' \sigma_i^2) \frac{1}{T^2} \sum_{t=1}^{T} \sum_{s \neq t} \text{tr} \left( \mathbb{E} (v_t v_t' v_t v_t') \right)
\]

We notice that \( T^{-2} \sum_{t=1}^{T} \sum_{s \neq t} \text{tr} \left( \mathbb{E} (v_t v_t' v_t v_t') \right) \leq \mathbb{E} \left[ \| v_t \|^4 \right] < C \) due to Assumption 1(ii).

Denote \( r, r^* \) to be indexes that go over 1, ..., \( \text{dim}(\gamma_i) \). For any fixed value of \( r, r^* \) denote
\[B^{(r,r')} = ((\gamma_i \gamma'_i)_{r,r'})_{i=1}^N, \text{ an } N \times 1 \text{ vector. Then,}\]

\[
\frac{1}{N^2} \sum_{i,j=1}^N \text{tr}(\gamma_i \gamma'_i \gamma_j \gamma'_j \sigma^2_{i,j}) = \frac{1}{N^2} \sum_{i,j=1}^N (\gamma_i \gamma'_i)(\gamma_j \gamma'_j) \sigma^2_{i,j}
\]

\[
= \sum_{r,r'} \frac{1}{N^2} \sum_{i=1}^N (\gamma_i \gamma'_i)(\gamma_j \gamma'_j) \sigma^4_i
\]

\[
+ \sum_{r,r'} \frac{1}{N^2} B^{(r,r')} [(\mathcal{E}_{N,T} - \text{dg} \mathcal{E}_{N,T}) \circ (\mathcal{E}_{N,T} - \text{dg} \mathcal{E}_{N,T})] B^{(r,r')}
\]

\[
\leq (\dim(\gamma_i))^2 \max_{1 \leq i \leq N} \|\gamma_i\|^4 \left( \frac{1}{N^2} \sum_{i=1}^N \sigma^4_i + \frac{\sqrt{N} \|\mathcal{E}_{N,T} - \text{dg} \mathcal{E}_{N,T}\|}{N^2} \right)
\]

\[
\leq \frac{C}{\sqrt{N}} \to 0,
\]

where in the second to last inequality we used Lemma A.4 and the last inequality is due to Assumptions 2, 4(ii) and the additional assumption stated in Theorem 4.1. This shows that \(\text{tr} \left( \mathbb{E}(A_2 A'_2) \right) \to 0.\)

Finally, \(\text{tr} \left( \mathbb{E}(A_3 A'_3) \right) \to 0\) due to Assumption 1(iv). This ends the proof of statement (i).

Let us turn to statement (ii):

\[
\frac{1}{N} \sum_{i=1}^N \xi_{W,i} \xi'_{W,i} - \Sigma_{W,T} = \frac{1}{T^2 N} \sum_{i=1}^N \sum_{s=1}^T \sum_{t<s} \sum_{s=1}^T \sum_{s=1}^T \sum_{t<s} w_{s,t} w'_{s,t} w_{s} w'_{s}
\]

\[
+ \frac{1}{T^2 N} \sum_{i=1}^N \sum_{t=1}^T \sum_{s=1}^T \sum_{t<s} \sum_{s=1}^T \sum_{t<s} w_{s,t} w'_{s,t} w_{s} w'_{s}
\]

\[
= A_1 + A_2 + A_3.
\]

Again, \(A_1, A_2\) and \(A_3\) are uncorrelated with each other. Thus, we can deal with each one of them separately.

Let us start with

\[
\text{tr} \left( \mathbb{E}(A_1 A'_1) \right) = \frac{1}{T^4 N^2} \sum_{i,j=1}^N \sum_{s,s'=1}^T \sum_{t<s,t'<s} \text{tr} \left( \mathbb{E}(w_{s,t} w'_{s,t'} w_{s,t'} w'_{s,t'}) \right) \mathbb{E}(b_{i,t,s} b_{j,t',s'})
\]

where

\[
b_{i,t,s} = e^2_{it} e^2_{is} - \sigma^4_i = (e^2_{it} - \sigma^2_i)(e^2_{is} - \sigma^2_i) + \sigma^2_i (e^2_{is} - \sigma^2_i) + \sigma^2_i (e^2_{it} - \sigma^2_i).
\]
Notice that $\mathbb{E}(b_{i,t,s}b_{j,t^*,s^*}) \neq 0$ only if at least one of the indexes from the set $\{t, t^*, s, s^*\}$ appears twice. Thus, the summation over time index is three-dimensional and there are at most $CT^3N^2$ non-zero summands in $\text{tr} (\mathbb{E}(A_1A_1'))$. Let us bound every summand from above. Notice that since $t < s$ and $t^* < s^*$, all indexes in the set $\{t, t^*, s, s^*\}$ can appear at most twice; also errors with different time indexes are independent from each other, so the largest moment of the error term we will have is the fourth. To sum up, each non-zero summand is bounded above by $T^{-4}N^{-2}C \max_{1 \leq i \leq T} \mathbb{E} [\|w_{is}\|^4] \max_{1 \leq i \leq N, 1 \leq t \leq T} \mathbb{E} (e_{it}^4)^2$, thus $\text{tr} (\mathbb{E}(A_1A_1')) \leq C/T \rightarrow 0$.

The term $\text{tr} (\mathbb{E}(A_2A_2'))$ includes summation over eight time indexes but most of the summands are zeros. The non-zero terms place at least four restrictions on the time indexes. We note that the non-trivial part of the sum in $\text{tr} (\mathbb{E}(A_2A_2'))$ includes summation over $i, j = 1, ..., N$ and over time indexes $\{s_1, s_1^*, t_1, t_1^*, s_2, s_2^*, t_2, t_2^*\}$, where in the last set any distinct index appears at least twice. The summands are

$$\frac{1}{T^4N^2} \mathbb{E} \left( w_{s_1t_1} w_{s_1^* t_1^*} w_{s_2 t_2} w_{s_2 t_2} \right) \mathbb{E} \left( e_{it_1} e_{it_1} e_{it_1} e_{it_1} e_{it_2} e_{it_2} e_{it_2} e_{it_2} \right).$$

Notice also that due to restrictions that $t$’s are strictly smaller than their corresponding $s$’s, each time index can appear at most four times, hence we get at most fourth power of each error term.

First, consider the case when the set $\{s_1, s_1^*, t_1, t_1^*, s_2, s_2^*, t_2, t_2^*\}$ contains at most three distinct indexes (this makes the summation over time three-dimensional). We can show that each summand is bounded by $T^{-4}N^{-2} \max_{1 \leq l, s \leq T} \mathbb{E} [\|w_{is}\|^4] \max_{1 \leq i \leq N, 1 \leq t \leq T} \mathbb{E} (e_{it}^4)^2 \leq C/(T^4N^2)$ in absolute value, and as there are most $N^2T^3$ of them (two-dimensional cross-sectional and three-dimensional over time summations), the sum of such terms will go to zero.

Finally, we consider the case when the set $\{s_1, s_1^*, t_1, t_1^*, s_2, s_2^*, t_2, t_2^*\}$ contains four distinct indexes. Then each summand of this type is bounded in absolute value by

$$C \frac{|\sigma_{ij}^a (\sigma_i^2)^b (\sigma_j^2)^c}{T^4N^2} \max_{1 \leq i, j \leq T} \mathbb{E} [\|w_{is}\|^4],$$

where $a + b + c = 4$, and the values of $a$, $b$ and $c$ depend on which indices coincide with which; however, due to the conditions $\{s_1, t_1\} \neq \{s_2, t_2\}$ and $t_1 < s_1, t_2 < s_2$, we know that the set $\{s_1, s_2, t_1, t_2\}$ contains at least three distinct indexes. Thus, $c$ and $b$ are either 0 or 1 each, and $a \geq 2$. Hence, due to Assumption 1(ii), the corresponding sum is bounded
above by
\[
\frac{C}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} |\sigma_{ij}|^a (\sigma_i^2)^b (\sigma_j^2)^c \leq \frac{C}{N^2} \max_{1 \leq i \leq N} \sigma_i^4 \sum_{i=1}^{N} \sum_{j=1}^{N} \sigma_{ij}^2
\]
\[
= \frac{C}{N^2} \sum_{i=1}^{N} \sigma_i^4 + \frac{C}{N^2} \sum_{i=1}^{N} \sum_{j \neq i} \sigma_{ij}^2
\]
\[
\leq \frac{C}{N} + \frac{C}{N^2} \| E_{N,T} - \delta g(E_{N,T}) \|_F^2 \leq \frac{C}{N}.
\]

(A.4)

In the first inequality we use \(|\sigma_{ij}| \leq \sigma_i \sigma_j\). In the last line we use that for any symmetric matrix \(A\), we have \(\| A \|_F \leq N \| A \|_2\) and assumption stated in Theorem 4.1. Thus, \(\text{tr} (\mathbb{E}(A_2 A'_2)) \rightarrow 0\).

Next, Assumption 1(iii) implies the convergence of \(A_3\). This finishes the proof of (ii).

Finally, we need to prove statement (iii) that
\[
\frac{1}{NT^{3/2}} \sum_{i=1}^{N} \sum_{s=1}^{T} \sum_{t<s} \sum_{s^*=1}^{T} \sum_{t<s} (\gamma_i \otimes v_{s^*}) w_{st} e_{is} e_{it} e_{is} \rightarrow^p 0.
\]

As before, we look at the expectation of the square of the sum above, which involves six-dimensional summation over time indexes and two-dimensional summation over cross-section (over \(i,j\)) and is normalized by \(N^{-2}T^{-3}\). Due to time-series independence of \(e_{it}\), the six-dimensional summation over time indexes has mostly zeros and can be reduced to three-dimensional summation over time indexes as the set \(\{s_1, t_1, s_1^*, s_2, t_2, s_2^*\}\) should have any distinct index to appear at least twice.

First, consider only those terms for which the set \(\{s_1, t_1, s_1^*, s_2, t_2, s_2^*\}\) contains at most two distinct indexes; there are at most \(N^2T^2\) of such terms. Since \(t_1 < s_1\) and \(t_2 < s_2\), each time index can appear at most four times; thus, the highest power of each individual shock can be the fourth. As a result, each summand is bounded above by \(N^{-2}T^{-3} \max_{1 \leq i \leq N} \| \gamma_i \|^2 \max_{1 \leq t \leq T} \mathbb{E}[\| v_t \|^2] \max_{1 \leq t, s \leq T} \mathbb{E}[\| w_{st} \|^2] \max_{1 \leq i \leq N, 1 \leq t \leq T} \mathbb{E}(e_{it}^4)^{3/2}\). Given Assumptions 1(ii) and 3(ii), the sum of these terms is bounded above by \(C/T\).

Finally, consider only those terms for which the set \(\{s_1, t_1, s_1^*, s_2, t_2, s_2^*\}\) contains exactly three distinct indexes. The summation over these indexes is equal to
\[
\text{tr} \left( \frac{1}{N^2} \sum_{i,j=1}^{N} \gamma_i \gamma_j' (C_1 \sigma_{ij}^2 \sigma_i^2 \sigma_j^2 + C_2 \sigma_{ij}^3) \right).
\]
The term $\sigma_{ij}^3$ appears when $\{s_1, t_1, s_1^*\} = \{s_2, t_2, s_2^*\}$, while $\sigma_{ij}^2 \sigma_{ij}^2$ arises when the sets $\{s_1, t_1, s_1^*\}$ and $\{s_2, t_2, s_2^*\}$ have two coinciding indexes each. Therefore,

$$\text{tr} \left( \frac{1}{N^2} \sum_{i,j=1}^{N} \gamma_i \gamma'_j \sigma_{i}^2 \sigma_{j}^2 \right) = \text{tr} \left( \frac{1}{N^2} \sum_{i=1}^{N} \gamma_i \gamma'_6 \sigma_{i}^2 \right) + \text{tr} \left( \frac{1}{N^2} \sum_{i \neq j} \gamma_i \gamma'_j \sigma_{ij} \right) \leq \frac{C}{N} + \frac{1}{N} \sum_{i=1}^{N} \| \gamma_i \|^2 \sigma_{i}^4 + \frac{1}{N} \sum_{i \neq j} \| \gamma_i \| \sigma_{ij} \leq \frac{C}{N} + \max_{1 \leq i \leq N} \| \gamma_i \| \sigma_{i,j} \rightarrow 0. $$

Also,

$$\text{tr} \left( \frac{1}{N^2} \sum_{i,j=1}^{N} \gamma_i \gamma'_j \sigma_{ij}^3 \right) = \frac{1}{N^2} \sum_{i=1}^{N} \| \gamma_i \|^2 \sigma_{i}^6 + \frac{1}{N^2} \sum_{i \neq j} \text{tr}(\gamma_i \gamma'_j) \sigma_{ij}^3 \leq \frac{C}{N} + \max_{1 \leq i \leq N} \| \gamma_i \| \sigma_{i}^2 \max_{1 \leq i \leq N} \| \gamma_i \| \frac{1}{N} \sum_{i,j=1}^{N} \sigma_{ij} \rightarrow 0. $$

Here we used the statement $N^{-2} \sum_{i,j=1}^{N} \sigma_{ij}^2 \rightarrow 0$, which is proved in equation (A.4). This ends the proof of Theorem 4.1. ■

### A.3 Proofs for Conditional Heteroscedasticity case

**Proof of Theorem 3.2.** In order to apply Lemma A.3 we check conditions (i)-(iv) of Lemma A.2 and conditions (a)-(d) of Lemma A.3 for

$$W_{st} = \frac{1}{T} w_{st} \frac{e'_t e_s}{\sqrt{N}}$$

and

$$V_s = \frac{1}{\sqrt{T}} \frac{\gamma' e_s}{\sqrt{N}} \otimes v_s.$$ (i) Due to serial independence of $e_{st}$ conditionally on $F$, we have

$$\Sigma_{W,T} = \frac{1}{T^2} \sum_{s=1}^{T} \sum_{t<s} \mathbb{E} \left[ w_{st} w'_{st} \mathbb{E} \left( \left( \frac{e'_t e_s}{\sqrt{N}} \right)^2 | F \right) \right].$$

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

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

33
Recall that $e_t = \pi f_t + \eta_t$. We will use the notation $\Omega_\eta = \mathbb{E}(\eta_t\eta'_t) = \text{dg}\{\omega_i^2\}_{i=1}^N$. Then,

$$\mathbb{E}\left[\left(\frac{e'_i e_s}{\sqrt{N}}\right)^2 \mid \mathcal{F}\right] = \frac{1}{N} \text{tr}\left((\pi \mathbb{E}(f'_i f'_s)\pi') + \Omega_\eta\right)\left(\pi \mathbb{E}(f'_s f'_s)\pi' + \Omega_\eta\right)$$

$$= \frac{1}{N} \sum_{i=1}^N \omega_i^4 + \Delta_{N,T},$$

where

$$\Delta_{N,T} \leq \frac{C}{N} \mathbb{E}\left[\left(\|f_t\|^2 + 1\right)(\|f_s\|^2 + 1)\mid \mathcal{F}\right].$$

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

$$\frac{1}{N} \text{tr}\left(\Omega_\eta \pi \mathbb{E}(f'_s f'_s)\pi'\right) \leq \frac{1}{N} \max_{1 \leq i \leq N} \omega_i^2 \cdot \text{tr}\left(\mathbb{E}(f'_s f'_s)\pi'\pi\right) \leq \frac{1}{N} \max_{1 \leq i \leq N} \omega_i^2 \cdot \max \pi'\pi \cdot \mathbb{E}\left[\|f_s\|^2 \mid \mathcal{F}\right].$$

Since we assumed that $\max_{1 \leq i \leq N} \omega_i^2 < C$ and from Assumption 5(ii), it follows that

$$\left\|\frac{1}{T^2} \sum_{s=1}^T \sum_{t<s} E\left[w_{st} w'_{st}\Delta_{N,T}\right]\right\| \leq \frac{C}{NT^2} \sum_{s=1}^T \sum_{t<s} E\left[\|w_{st}\|^2 (\|f_t\|^2 + 1)(\|f_s\|^2 + 1)\right] \leq \frac{C}{N} \rightarrow 0,$$

where the last inequality is due to Assumption 5(i). So, we obtain that

$$\Sigma_{W,T}(T, N) = \lim \frac{1}{T^2} \sum_{s=1}^T \sum_{t<s} E\left[w_{st} w'_{st}\right] \frac{1}{N} \sum_{i=1}^N \omega_i^4 = \omega^4 \Omega_w = \Sigma_W.$$

(ii) Notice that

$$\frac{e'_i e_s}{\sqrt{N}} = f'_i \pi' f_s + \frac{f'_i \pi' \eta_s}{\sqrt{N}} + \frac{f'_i \eta'_i}{\sqrt{N}} + \frac{\eta'_i \eta_s}{\sqrt{N}}.$$

Using the Marcinkiewicz–Zygmund inequality for a second power applied twice we notice that in order to bound $\mathbb{E}\left[\left(\frac{e'_i e_s}{\sqrt{N}}\right)^4 \mid \mathcal{F}\right]$ from above it is enough to bound the fourth moment of each summand. Using serial and cross-sectional conditional independence of $\eta$’s as well as their conditional independence from $f$’s, we obtain

$$\mathbb{E}\left[\left(\frac{1}{\sqrt{N}} \sum_{i=1}^N \eta_i \eta_i s\right)^4\right] = \frac{1}{N^2} \sum_{i=1}^N \mathbb{E}\left(\eta_i^2 \eta_i s\right) + C \frac{1}{N^2} \sum_{i_1 \neq i_2} \mathbb{E}\left(\eta_i^2 \eta_i s^2 \eta_i s^2 \eta_i^2 s^2\right) \leq C;$$

$$\mathbb{E}\left[\left\|\frac{1}{\sqrt{N}} \sum_{i=1}^N \pi_i \eta_i s\right\|^4\right] \leq \frac{1}{N^2} \sum_{i=1}^N \mathbb{E}\left[\|\pi_i \eta_i s\|^4\right] + C \frac{1}{N^2} \sum_{i_1 \neq i_2} \|\pi_i s\|^2 \|\pi_i s\|^2 \mathbb{E}\left[\eta_i^2 \eta_i s^2 \eta_i^2 s^2\right] \leq \frac{C}{N^2}. $$

34
where we use Assumption 5(ii,iii), and that $\sum_i \|\pi_i\|^4 \leq (\sum_i \|\pi_i\|^2)^2 \leq C$. Hence,

$$
\mathbb{E} \left[ \left( \frac{e_t' e_s}{\sqrt{N}} \right)^4 | \mathcal{F} \right] \leq \frac{C}{N^2} \mathbb{E} \left[ \|f_t\|^4 \|f_s\|^4 | \mathcal{F} \right] + \frac{C}{N^2} \left( \mathbb{E} \left[ \|f_t\|^4 | \mathcal{F} \right] + \mathbb{E} \left[ \|f_s\|^4 | \mathcal{F} \right] \right) + C.
$$

Finally, due to Assumption 5(i),

$$
T^4 \mathbb{E} \left[ \|W_{st}\|^4 \right] \leq \mathbb{E} \left[ \left\| w_{st} \frac{e_t' e_s}{\sqrt{N}} \right\|^4 \right] \leq C \mathbb{E} \left[ \|w_{st}\|^4 (\|f_t\|^4 + 1)(\|f_s\|^4 + 1) \right] < C.
$$

Thus, condition (ii) of Lemma A.2 holds.

(iii) Let us define a sigma-algebra $\mathcal{A} = \mathcal{F} \cup \{ f_t, t = 1, ..., T \}$. Let us now denote

$$
\vartheta_{st} = \mathbb{E} \left[ \left( \frac{e_t' e_s}{\sqrt{N}} \right)^2 | \mathcal{A} \right] = \mathbb{E} \left[ \left( (\pi f_t + \eta_t)(\pi f_s + \eta_s) \right)^2 \right] = \frac{1}{N} \left( f_s' \pi' f_t + f_t' \pi' f_s + f_t' \pi' f_t + \sum_{i=1}^{N} \omega_i^4 \right).
$$

We have:

$$
\sum_{s=1}^{T} \sum_{t<s} W_{st} W_{st}' - \Sigma_{W,T} = \frac{1}{T^2} \sum_{s=1}^{T} \sum_{t<s} w_{st} w_{st}' \left( \left( \frac{e_t' e_s}{\sqrt{N}} \right)^2 - \vartheta_{st} \right)
$$

so, it is enough to prove convergence of each term separately. Now, $\mathbb{E} \left[ \text{tr}(A_1 A_1') \right]$ is equal to

$$
\frac{1}{T^4} \sum_{s_1, s_2=1}^{T} \sum_{t_1, t_2} \mathbb{E} \left[ \text{tr}(w_{s_1 t_1} w_{s_1 t_1}' w_{s_2 t_2} w_{s_2 t_2}') \left( \left( \frac{e_{s_1} e_{t_1}}{\sqrt{N}} \right)^2 - \vartheta_{s_1 t_1} \right) \left( \left( \frac{e_{s_2} e_{t_2}}{\sqrt{N}} \right)^2 - \vartheta_{s_2 t_2} \right) \right].
$$

Notice that in order for a summand from the last sum to be non-zero we need that some indexes in the set $\{ s_1, s_2, t_1, t_2 \}$ coincide, and we obtain at most $C T^3$ non-zero summands. Each non-zero summand is bounded above by a constant due to the moment assumptions formulated in Assumption 5(i,iii). Thus, $\mathbb{E} \left[ \text{tr}(A_1 A_1') \right] \to 0$.

Notice that due to Assumption 5, and similar to the argument above,

$$
\left| \vartheta_{st} - \frac{1}{N} \sum_{i=1}^{N} \omega_i^4 \right| \leq \frac{C}{N} (\|f_s\| + \|f_t\| + 1)^4.
$$

(A.5)
Thus,

\[
A_2 = \left( \frac{1}{N} \sum_{i=1}^{N} \omega_i^2 \right) \frac{1}{T^2} \sum_{s=1}^{T} \sum_{t<s} \left( w_{st}w'_{st} - \mathbb{E} (w_{st}w'_{st}) \right) \\
+ \frac{1}{T^2} \sum_{s=1}^{T} \sum_{t<s} \left( w_{st}w'_{st} (\vartheta_{st} - \frac{1}{N} \sum_{i=1}^{N} \omega_i^4) - \mathbb{E} \left[ w_{st}w'_{st} (\vartheta_{st} \frac{1}{N} - \sum_{i=1}^{N} \omega_i^4) \right] \right),
\]

where the first sum converges to zero due to Assumption 1(iii), while expectation of the second moment of the second term is bounded by

\[
\frac{1}{T^4} \sum_{s_1,s_2,t_1,t_2} \frac{C}{N^2} \mathbb{E} \left[ (\|f_{s_1}\| + \|f_{s_2}\| + 1)^4 (\|f_{s_2}\| + \|f_{t_2}\| + 1)^4 \|w_{s_1t_1}\|^2 \|w_{s_2t_2}\|^2 \right] \leq C N^2,
\]

due to inequality (A.5) and Assumption 5(i). Thus, condition (iii) of Lemma A.2 holds.

Let us check condition (iv):

\[
T^4 \mathbb{E} \left( W'_{s,t_2} W_{s,t_1} W'_{s_2,t_2} W_{s_2,t_1} \right) = \frac{1}{N^2} \mathbb{E} \left[ w'_{s_1t_2} w_{s_2t_1} w'_{s_2t_2} w_{s_1t_1} \mathbb{E} (e'_{s_1} e_{t_1} e_{s_2} e'_{t_2} e_{s_2} e_{t_2} e_{s_1} | \mathcal{F}) \right],
\]

where we used that the scalar products $e'_{i} e_{s} = e'_{i} e_{r}$ 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 $\mathbb{E} (e_{t} e'_{t} | \mathcal{F}) = \pi \mathbb{E} (f_{t} f'_{t} | \mathcal{F}) \pi' + \Omega_{\eta}$:

\[
\max \text{ ev}(\mathbb{E} (e_{t} e'_{t} | \mathcal{F})) \leq \max \text{ ev}(\pi' \mathbb{E} (f_{t} f'_{t} | \mathcal{F}) \pi) + \max \text{ ev}(\Omega_{\eta}) \\
\leq \text{ tr}(\pi' \mathbb{E} (f_{t} f'_{t} | \mathcal{F}) \pi) + \max_{1 \leq i \leq N} \omega_i^2 \\
\leq \max(\pi' \pi') \mathbb{E} (\|f_{t}\|^2 | \mathcal{F}) + C \\
\leq C \mathbb{E} (\|f_{t}\|^2 + 1 | \mathcal{F}). \tag{A.6}
\]

As a result,

\[
\left| \mathbb{E} (e'_{s_1} e_{t_1} e'_{t_1} e_{s_2} e'_{t_2} e_{t_2} e_{s_1} | \mathcal{F}) \right| = \left| \text{ tr}(\mathbb{E} (e_{t_1} e'_{t_1} | \mathcal{F}) \mathbb{E} (e_{s_2} e'_{s_2} | \mathcal{F}) \mathbb{E} (e_{t_2} e'_{t_2} | \mathcal{F}) \mathbb{E} (e_{s_1} e'_{s_1} | \mathcal{F})) \right| \\
\leq N \max \text{ ev} \prod_{t \in \{s_1,s_2,t_1,t_2\}} \mathbb{E} (e_{t} e'_{t} | \mathcal{F}) \\
\leq N \prod_{t \in \{s_1,s_2,t_1,t_2\}} \max \text{ ev} (\mathbb{E} (e_{t} e'_{t} | \mathcal{F})) \\
\leq N C \prod_{t \in \{s_1,s_2,t_1,t_2\}} \mathbb{E} (\|f_{t}\|^2 + 1 | \mathcal{F}).
\]

Also using Assumption 5(i) we obtain that

\[
T^4 \left| E \left( W'_{s,t_2} W_{s_2,t_1} W'_{s_2,t_2} W_{s_1,t_1} \right) \right| \leq \frac{C}{N} \max_{1 \leq s,t \leq T} \mathbb{E} \left[ \|w_{st}\|^4 \|f_{st}\|^8 \right] \rightarrow 0.
\]

36
Thus, condition (iv) holds as well.

Now we will check assumptions (a)-(d) of Lemma A.3. First, we find the limit of the covariance matrix $\Sigma_{V,T}$.

$$\mathbb{E} \left[ \left( \frac{\gamma'e_s}{\sqrt{N}} \right) \left( \frac{\gamma'e_s}{\sqrt{N}} \right)' | \mathcal{F} \right] = \left( \frac{\gamma'\pi}{\sqrt{N}} \right) \mathbb{E} [f_s f_s' | \mathcal{F}] \left( \frac{\pi'\gamma}{\sqrt{N}} \right) + \frac{1}{N} \sum_{i=1}^{N} \omega_i^2 \gamma_i \gamma_i'$$

$$\rightarrow \Gamma_{\pi\gamma} \mathbb{E} (f_s f_s' | \mathcal{F}) \Gamma_{\pi\gamma} + \Gamma_\omega.$$ 

Here we used the Assumptions 5(ii,iii). Therefore,

$$\Sigma_{V,T} = \text{var} \left( \sum_{s=1}^{T} V_s \right) = \mathbb{E} \left[ \frac{1}{T} \sum_{s=1}^{T} \mathbb{E} \left[ \left( \frac{\gamma'e_s}{\sqrt{N}} \right) \left( \frac{\gamma'e_s}{\sqrt{N}} \right)' | \mathcal{F} \right] \odot (v_s v_s') \right]$$

$$= \frac{1}{T} \sum_{s=1}^{T} \mathbb{E} \left[ (\Gamma'_{\pi\gamma} f_s f_s' \Gamma_{\pi\gamma} + \Gamma_\omega) \odot (v_s v_s') \right]$$

$$\rightarrow (\Gamma'_{\pi\gamma} \otimes I_{k_v}) \Sigma_{f_{k_v}} (\Gamma_{\pi\gamma} \otimes I_{k_v}) + \Gamma_\omega \otimes \Omega_v.$$ 

The limit matrix is positive definite since both $\Gamma_\omega$ and $\Omega_v$ are positive-definite due to Assumptions 1(i) and 5(iii).

Now note that due to Assumption 5(ii)

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

$$\mathbb{E} \left( \left\| \gamma' \eta_t \right\|^4 \right) = \mathbb{E} \left[ \left\| \sum_{i=1}^{N} \gamma_i \eta_{i,t} \right\|^4 \right] \leq \sum_{i=1}^{N} \left\| \gamma_i \right\|^4 \mathbb{E} \eta_{i,t}^4 + C \sum_{i_1, i_2=1}^{N} \left\| \gamma_{i_1} \right\|^2 \left\| \gamma_{i_2} \right\|^2 \omega_{i_1}^2 \omega_{i_2}^2.$$

Due to Assumptions 2 and 5(iii) we have that $\mathbb{E} \left( \left\| \gamma' \eta_t \right\|^4 \right) \leq C N^2$, and thus

$$\mathbb{E} \left[ \left\| \frac{\gamma' e_t}{\sqrt{N}} \right\|^4 \right] \leq C \mathbb{E} \left( \left\| f_t \right\|^4 + 1 \right) \mathbb{E} \left( \left\| \gamma' \eta_t \right\|^4 \right).$$

Collecting the pieces,

$$T \mathbb{E} \left( \left\| V_s \right\|^4 \right) \leq C T \mathbb{E} \left[ \frac{1}{T^2} \mathbb{E} \left[ \left\| \frac{\gamma'e_s}{\sqrt{N}} \right\|^4 \mathcal{F} \left( \left\| v_s \right\|^4 \right) \right] \right) \leq T \frac{C}{T^2} \mathbb{E} \left[ \left( \left\| f_s \right\|^4 + 1 \right) \left\| v_s \right\|^4 \right] \rightarrow 0.$$ 

This gives us the validity of condition (b) of Lemma A.3.
Finally, we check the condition (d):

\[
\sum_{t=1}^{T} V_t V'_t - \Sigma_{V;T} = \frac{1}{T} \sum_{t=1}^{T} \left( \frac{\gamma' e_t e'_t \gamma'}{\sqrt{N}} \right) \otimes (v_t v'_t)
\]

\[= \frac{1}{T} \sum_{t=1}^{T} \mathbb{E} \left[ \left( \frac{\gamma' \pi f_t \pi' \gamma}{\sqrt{N}} + \Gamma_{\omega,N} \right) \otimes (v_t v'_t) \right] \]

\[= \frac{1}{T} \sum_{t=1}^{T} \left( \frac{\gamma' e_t e'_t \gamma'}{\sqrt{N}} - \frac{\gamma' \pi f_t \pi' \gamma}{\sqrt{N}} - \Gamma_{\omega,N} \right) \otimes (v_t v'_t) \]

\[+ \frac{1}{T} \sum_{t=1}^{T} \left[ \left( \frac{\gamma' \pi f_t f'_t \pi' \gamma}{\sqrt{N}} + \Gamma_{\omega,N} - \mathbb{E} \left[ \frac{\gamma' \pi f_t f'_t \pi' \gamma}{\sqrt{N}} + \Gamma_{\omega,N} \right] \right) \otimes (v_t v'_t) \right] \]

\[= A_1 + A_2. \]

Notice that given the conditional independence of \( \eta_t \)'s, the two terms in the last expression, \( A_1 \) and \( A_2 \) are uncorrelated, so in order to check condition (c) of Lemma A.3 we can prove convergence to zero of \( \mathbb{E} \|A_1\|^2 \) and \( \mathbb{E} \|A_2\|^2 \) separately. First,

\[\mathbb{E} \|A_1\|^2 = \mathbb{E} \left[ \left\| \frac{1}{T} \sum_{t=1}^{T} \left( \frac{\gamma' \pi f_t \eta' \gamma}{\sqrt{N}} + \frac{\gamma' \eta_t \eta' \gamma}{\sqrt{N}} + \left( \frac{\gamma' \eta_t \eta' \gamma}{\sqrt{N}} - \Gamma_{\omega,N} \right) \right) \otimes (v_t v'_t) \right\|^2 \right] \]

\[\leq \frac{1}{T^2} \sum_{t=1}^{T} \mathbb{E} \left[ \left\| \frac{\gamma' \pi f_t \eta' \gamma}{\sqrt{N}} + \frac{\gamma' \eta_t \eta' \gamma}{\sqrt{N}} + \left( \frac{\gamma' \eta_t \eta' \gamma}{\sqrt{N}} - \Gamma_{\omega,N} \right) \right\|^2 \|v_t\|^4 \right] \]

\[\leq \frac{1}{T} C \mathbb{E} \left[ (\|f_t\|^2 + 1)\|v_t\|^4 \right] \to 0. \]

The former inequality is due to \( \eta_t \)'s being conditionally serially uncorrelated, and thus the summation over \( t \) can 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, the convergence of term \( A_2 \) is due to Assumptions 5(iv) and 1(iv). Putting all terms together, we obtain that condition (c) is satisfied.

Finally, we check the condition (d):

\[T^3 \left\| \mathbb{E} \left( W_{s_1} V'_{s_1} V_{s_2} W'_{s_2 t} \right) \right\| = \left\| \mathbb{E} \left[ w_{s_1} v'_{s_1} v_{s_2} w'_{s_2 t} \mathbb{E} \left( \frac{e'_{s_1} \gamma e'_{s_2} e'_{s_1} e_t e'_{s_2} e_t}{\sqrt{N} \sqrt{N} \sqrt{N} \sqrt{N}} \right) \right) \right\|. \]

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:

\[
\left| \mathbb{E} \left( \frac{e'_s \gamma e'_t e'_s e'_t}{\sqrt{N} \sqrt{N} \sqrt{N} \sqrt{N}} \mathcal{F} \right) \right| = \frac{1}{N^2} \left| \text{tr}(\gamma' \mathbb{E}(e'_s e'_t | \mathcal{F}) \mathbb{E}(e'_t e'_s | \mathcal{F})) \right| \\
\leq \frac{1}{N^2} \text{tr}(\gamma') \prod_{s \in \{s_1, s_2\}} \max \text{ev}(\mathbb{E}(e_s e'_s | \mathcal{F})) \\
\leq \frac{C}{N} \mathbb{E} \left[ \prod_{s \in \{s_1, s_2\}} (\|f_s\|^2 + 1) | \mathcal{F} \right].
\]

We use Assumption 2 that \( N^{-1} \text{tr}(\gamma' \gamma) \leq C \) and the bound (A.6) we derived before. In the last equality, we also exploit that \( f_i \)'s are conditionally independent of each other. Thus, Assumption 5 (i) implies that

\[
T^3 \max_{1 \leq t < \min\{s_1, s_2\} \leq T} \| \mathbb{E} \left( W_{s_i,t} V_{s_1} V_{s_2} W_{s_2,t} \right) \| \leq \frac{C}{N} \to 0.
\]

Thus, condition (d) of Lemma A.3 is satisfied. This concludes the proof of Theorem 3.2.

\[\square\]

Proof of Theorem 4.2. We will prove three statements:

(i) \( N^{-1} \sum_{i=1}^N \xi_{V,i} \xi'_{V,i} \to \Sigma_V \);

(ii) \( N^{-1} \sum_{i=1}^N \xi_{W,i} \xi'_{W,i} \to \Sigma_W \);

(iii) \( N^{-1} \sum_{i=1}^N \xi_{V,i} \xi'_{W,i} \to 0 \).

(i) Given assumption \( \Gamma_{\pi \gamma} = 0 \), we have \( \Sigma_V = \Gamma_\omega \otimes \Omega_\pi \). Then,

\[
\frac{1}{N} \sum_{i=1}^N \xi_{V,i} \xi'_{V,i} = \frac{1}{NT} \sum_{t=1}^T \sum_{s=1}^T \left( \sum_{i=1}^N \gamma_i \gamma'_i (\pi'_i f_s + \eta_i s)(\pi'_i f_t + \eta_i t) \right) \otimes (v_s v'_t). \tag{A.7}
\]

After we open up the brackets there will be three different types of terms. We will show that

\[
\frac{1}{NT} \sum_{t=1}^T \sum_{s=1}^T \sum_{i=1}^N (\gamma_i \gamma'_i \eta_i s \eta_i t) \otimes (v_s v'_t) \overset{p}{\to} \Sigma_V, \tag{A.8}
\]

while terms that involve \( \pi'_i w_s \pi'_i w_t \) or \( \eta_i \pi'_i f_s \) converge to zero in probability. Indeed,

\[
\frac{1}{NT} \sum_{t=1}^T \sum_{s=1}^T \sum_{i=1}^N (\gamma_i \gamma'_i \eta_i s \eta_i t) \otimes (v_s v'_t) - \Sigma_{V,T} = \frac{1}{NT} \sum_{t=1}^T \sum_{s \neq t}^T \sum_{i=1}^N (\gamma_i \gamma'_i \eta_i s \eta_i t) \otimes (v_s v'_t) \\
+ \frac{1}{NT} \sum_{t=1}^T \sum_{i=1}^N (\gamma_i \gamma'_i) \otimes (\eta_i^2 v_t v'_t - \omega_i^2 \mathbb{E}(v_t v'_t)).
\]
We check that the first sum in the last expression is negligible:
\[
\mathbb{E} \left[ \frac{1}{NT} \sum_{t=1}^{T} \sum_{s \neq t} \sum_{i=1}^{N} \left( \gamma_i \gamma_i' \eta_{it} \eta_{it} \otimes (v_s v_t') \right)^2 \right] \leq \frac{1}{N^2 T^2} \sum_{t=1}^{T} \sum_{s \neq t} \sum_{i=1}^{N} \|\gamma_i\| \omega_i^4 \mathbb{E} \|v_t\|^2 \|v_s\|^2 \leq \frac{C}{N^2} \sum_{i=1}^{N} \|\gamma_i\| \omega_i^4 \to 0.
\]

Here we use the conditional cross-sectional and temporal independence of \( \eta_{it} \), that is, for \( s \neq t \) we have \( \mathbb{E}(\eta_{it} \eta_{is} \eta_{is^*} | \mathcal{F}) = \omega_i^4 \) if \( i = j \) and \( \{t, s\} = \{t^*, s^*\} \), and zero otherwise. We also use Assumptions 2 and 5(iii). As for the second sum, we notice that all summands in the expression below are uncorrelated with each other, hence
\[
\begin{align*}
\text{tr} \left( \mathbb{E} \left[ \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left( \gamma_i \gamma_i' \eta_{it}^2 \otimes (v_t v_t') - \Sigma_V \right)^2 \right] \right) \\
= \frac{1}{N^2 T^2} \sum_{i=1}^{N} \sum_{t=1}^{T} \text{tr} \left( \mathbb{E} \left[ \left( \gamma_i \gamma_i' \eta_{it}^2 \otimes (v_t v_t') - \mathbb{E} \left[ \left( \gamma_i \gamma_i' \eta_{it}^2 \otimes (v_t v_t') \right)^2 \right] \right) \right] \\
\leq \frac{C}{N^2 T^2} \sum_{i=1}^{N} \sum_{t=1}^{T} \|\gamma_i\| \omega_i^4 \mathbb{E} \left[ \|v_t\|^4 \right] \to 0.
\end{align*}
\]

Thus, we showed the convergence (A.8).

Now consider terms in (A.7) that involve \( \pi_i' f_s \pi_i' f_t \):
\[
\begin{align*}
\frac{1}{NT} \sum_{t=1}^{T} \sum_{s=1}^{T} \left( \sum_{i=1}^{N} \gamma_i \gamma_i' \pi_i' f_s \pi_i' f_t \right) \otimes (v_s v_t') \\
= \left( \frac{1}{N} \sum_{i=1}^{N} \left( \gamma_i \gamma_i' \otimes (\pi_i' \otimes \pi_i') \otimes I_{K_v} \right) \right) \left( \frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} I_{K_v} \otimes \text{vec} (f_s f_t') \otimes (v_s v_t') \right).
\end{align*}
\]

Using Assumption 2 and 5(ii) we can show that
\[
\left\| \frac{1}{N} \sum_{i=1}^{N} (\gamma_i \gamma_i' \otimes (\pi_i' \otimes \pi_i')) \right\| \leq \frac{1}{N} \sum_{i=1}^{N} \|\gamma_i\|^2 \|\pi_i\|^2 \leq C \frac{1}{N} \sum_{i=1}^{N} \|\pi_i\|^2 \to 0.
\]
Now observe that
\[
\mathbb{E} \left[ \left\| \frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} \text{vec}(f_{st}) \otimes (v_{s} v'_t) \right\|_F^2 \right] \\
= \text{tr} \left( \frac{1}{T^2} \mathbb{E} \left[ \sum_{t=1}^{T} \sum_{s=1}^{T} \sum_{t^*=1}^{T} \sum_{s^*=1}^{T} (\text{vec}(f_{st}) \text{vec}(f_{s^* t^*})') \otimes (v_{s} v'_t v_{s^*} v'_t) \right] \right) \\
\leq C \frac{1}{T^2} \mathbb{E} \left[ \sum_{t=1}^{T} \sum_{s=1}^{T} \|f_t\|^2 \|f_s\|^2 \|v_s\|^2 \|v_t\|^2 \right] < C.
\]

Here the equality is due to \(f_t\)'s being serially independent and mean zero conditionally on \(F\) by Assumption 5(i) and \(v_t \in F\); hence, among the four summation indexes at most two may be distinct. The last inequality is due to Assumption 5(i). Thus, we showed that
\[
\frac{1}{NT} \sum_{t=1}^{T} \sum_{s=1}^{T} \left( \sum_{i=1}^{N} \gamma_i \gamma'_i f_{si} f'_{st} \right) \otimes (v_{s} v'_t) \overset{p}{\rightarrow} 0.
\]

And finally, we show that
\[
\frac{1}{NT} \sum_{t=1}^{T} \sum_{s=1}^{T} \left( \sum_{i=1}^{N} \gamma_i \gamma'_i f_{si} \eta_{it} \right) \otimes (v_{s} v'_t) \overset{p}{\rightarrow} 0.
\]

This holds because \(\eta_{it}\)'s are mean zero, cross-sectionally independent and independent from \(f_t\) conditionally on \(F\). This implies that the mean of the sum above is zero, and all summands are uncorrelated with each other. The second moment of the sum is bounded above by
\[
\frac{C}{N^2 T^2} \sum_{t=1}^{T} \sum_{s=1}^{T} \sum_{i=1}^{N} \|\gamma_i\|^4 \|\pi_i\|^2 \omega_i^2 \mathbb{E} \left[ \|f_t\|^2 \|v_t\|^2 \|v_s\|^2 \right] \to 0.
\]

Thus, we proved statement (i).

Let us turn to statement (ii):
\[
\frac{1}{N} \sum_{i=1}^{N} \xi_{W,i} \xi'_{W,i} - \Sigma_{W,T} = \frac{1}{T^2 N} \sum_{i=1}^{N} \sum_{s=1}^{T} \sum_{t<s} \sum_{i<s} \sum_{s^*=1}^{T} \sum_{t^*=1}^{T} w_{st} w'_{st} (e_i^2 e_{i s} - \omega_i^4) \\
+ \frac{1}{T^2 N} \sum_{i=1}^{N} \sum_{s=1}^{T} \sum_{t<s} \sum_{s^*=1}^{T} \sum_{t^*=1}^{T} \sum_{i\neq s^*, t^*} w_{st} w'_{st} e_i e_{i s} e_i e_{i t} e_i e_{i t} \\
+ \frac{1}{N} \sum_{i=1}^{N} \omega_i^4 \frac{1}{T^2} \sum_{s=1}^{T} \sum_{t<s} (w_{st} w'_{st} - \mathbb{E}(w_{st} w'_{st})) \\
= A_1 + A_2 + A_3.
\]
As for $A_1$, we can notice that all summands with indexes $\{s, t\} \neq \{s^*, t^*\}$ are uncorrelated with each other, so the correlation for summands with different indexes $i$ can come only from the $\pi_i f_t$ part. Thus,

$$
\mathbb{E} \left[ \|A_1\|^2 \right] = \frac{1}{T^4 N^2} \sum_{s=1}^{T} \sum_{t<s} \sum_{i=1}^{N} \mathbb{E} \left[ \left\| \sum_{i=1}^{N} w_{st} w_{st}' \left( \epsilon_{it}^2 \epsilon_{is}^2 - \omega_i^4 \right) \right\|_F^2 \right] 
\leq \frac{C}{T^4 N^2} \sum_{s=1}^{T} \sum_{t<s} \sum_{i=1}^{N} \left( \mathbb{E} \left[ \|w_{st}\|^4 \right] \max_{1 \leq i \leq N, 1 \leq l \leq T} \mathbb{E} \left( \eta_{lt}^4 \right) + \sum_{j=1}^{N} \|\pi_i\|^4 \|\pi_j\|^4 \mathbb{E} \left[ \|w_{st}\|^4 \|f_{l}\|^4 \|f_{s}\|^4 \right] \right) 
\to 0.
$$

In the last convergence we used that due to Assumption 5,

$$
\frac{C}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \|\pi_i\|^4 \|\pi_j\|^4 \leq \frac{C}{N^2} \max_{1 \leq i \leq N} \|\pi_i\|^4 \left( \sum_{i=1}^{N} \|\pi_i\|^2 \right)^2 \to 0,
$$

and hence the term $A_1$ converges to zero.

Term $\mathbb{E} \left[ \|A_2\|^2 \right]$ equals the following expression:

$$
\frac{1}{T^4 N^2} \sum_{i,j=1}^{N} \sum_{s_1, s_2} \sum_{t_1, t_2} \mathbb{E} \left[ \sum_{m, m' \neq \{s_1, t_1\}, \{s_2, t_2\}} \text{tr}(w_{s_1 t_1} w_{s_2 t_2}' w_{s_1 t_1} w_{s_2 t_2}') \epsilon_{it_1 e_i s_1 e_i s_1^* e_i t_2 e_j s_2 e_j s_2^*} \right].

(A.10)
$$

Notice that if $s_m < t_m$, $s^*_m < t^*_m$ and $\{s_m, t_m\} \neq \{s^*_m, t^*_m\}$ for $m = 1, 2$, the only ways when the expectation

$$
\mathbb{E}(\epsilon_{it_1 e_i s_1 e_i t_1^* e_j s_2 e_j t_2^*} | \mathcal{F}) \neq 0
$$

(A.11)

can be non-zero is when we place at least four restrictions on the time indexes. Indeed, if $\{s_1, s^*_1, t_1, t^*_1\}$ are all distinct, then to get a non-zero expectation we need indexes to coincide as sets: $\{s_1, s^*_1, t_1, t^*_1\} = \{s_2, s^*_2, t_2, t^*_2\}$. If the set $\{s_1, s^*_1, t_1, t^*_1\}$ contains three distinct indexes, for example, $s_1 = s^*_1$ (this is one restriction), then the set $\{s_2, s^*_2, t_2, t^*_2\}$ should contain $(t_1, t^*_1)$ (these are two restrictions), and the remaining indexes should be either equal to each other (one restriction) or equal to the ones previously mentioned (two restrictions). Thus, instead of eight-dimensional summation over time indexes in equation (A.10) we have a four-dimensional summation.

Let us consider those terms in (A.10) when the summation index $j$ is equal to $i$. Notice that since each $t$ index is strictly smaller than the corresponding $s$ index, then any
distinct time index can appear in the set \( \{s_1, s_1^*, t_1, t_1^*, s_2, s_2^*, t_2, t_2^*\} \) at most four times, thus any individual error term \( e_{it} \) may appear in at most power four. Thus, all non-zero terms are bounded above by \( \max_{1 \leq i \leq N, 1 \leq s,t,t^* \leq T} \mathbb{E} \left[ \|w_{st}\|^4 (\mathbb{E} (\eta_{it}^4) + C\|f_{t*}\|^4)^2 \right] < C \) due to Assumption 5(i,iii). There are at most \( CT^4N \) of such terms while the normalization is \( N^{-2T-4} \), hence that sum converges to zero.

Now consider those terms in (A.10) when \( i \neq j \). Since \( e_{it} = \pi_i^*f_t + \eta_{it} \), with \( \eta_{it} \)'s independent of each other both cross-sectionally and temporally, \( i \neq j \) and \( \{s_m, t_m\} \neq \{s_m^*, t_m^*\} \), we have that all terms including \( \eta_{it} \) are zero, and only a non-trivial part of the term in (A.11) is the one including \( \pi_i^*f_t \) in place of \( e_{it} \). So, every non-zero term in the sum (A.10) is bounded by \( \|\pi_i^*\|^4\|\pi_j^*\|^4\mathbb{E}[\|w_{st}\|^4\|f_{t*}\|^8] \). So, the sum in (A.10) over \( j \neq i \) is bounded above in the same manner as stated in equation (A.9). Thus, we showed that \( A_2 \overset{p}{\to} 0 \). The convergence \( A_3 \overset{p}{\to} 0 \) comes from Assumption 1(iii). This finishes the proof of (ii).

Finally, let us prove statement (iii):

\[
\frac{1}{NT^{3/2}} \sum_{i=1}^{N} \sum_{s=1}^{T} \sum_{t<s}^{T} (\gamma_{is} \otimes v_{s*}) w_{st} e_{it} e_{is} e_{it} e_{is} \overset{p}{\to} 0.
\]

As before, we look at the expectation of the square of the sum above, which involves six-dimensional summation over time indexes and two-dimensional cross-sectional summation (over \( i, j \)) and is normalized by \( N^{-2T-3} \). Due to time-series independence of \( e_{it} \), the six-dimensional summation over time indexes has mostly zeros and can be reduced to three-dimensional summation over time indexes as the set \( \{s_1, t_1, s_1^*, t_1^*, s_2, t_2, s_2^*, t_2^*\} \) should have any distinct index to appear at least twice. If we consider the cases when \( i = j \), then all terms are bounded above by a constant and the number of non-zero terms is \( NT^3 \); given the normalization, this sum converges to zero. When we sum over \( i \neq j \), the only part of \( e_{it} \) that yields a non-trivial effect is \( \pi_i^*f_t \); hence this sum is bounded by

\[
\frac{1}{N^2} \sum_{i,j=1}^{T} \|\gamma_i\| \|\gamma_j\| \|\pi_i\|^3 \|\pi_j\|^3 \max_{1 \leq s,t,s^* \leq T} \mathbb{E} \left[ \|v_{s*}\|^2 \|w_{st}\|^2 \|f_{s*}\|^2 \|f_t\|^2 \right] \leq C \left( \frac{1}{N} \sum_{i=1}^{N} \|\gamma_i\| \|\pi_i\|^3 \right)^2 \leq \frac{1}{N^2} \max_{1 \leq i \leq N} \|\gamma_i\|^2 \max_{1 \leq i \leq N} \|\pi_i\|^4 \sum_{i=1}^{N} \|\pi_i\|^2 \to 0.
\]

This ends the proof of Theorem 4.2. ■
REFERENCES


44


