Econometric Theory, **37**, 2021, 1034–1074.
doi:10.1017/S0266466620000468

LIMIT THEOREMS FOR FACTOR MODELS*

STANISLAV ANATOLYEV
CERGE-EI
New Economic School
ANNA MIKUSHEVA
Massachusetts Institute of Technology

This paper establishes central limit theorems (CLTs) 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 microparameters estimated separately for each entity. The CLT applies for quantities involving both cross-sectional and time series aggregation, as well as for quadratic forms in time-aggregated errors. This 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.

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 and MacBeth, 1973 and Shanken, 1992), where for estimation we usually use 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

*We thank the Editor Peter C.B. Phillips, Co-Editor Guido Kuersteiner, and three anonymous referees for guidance and helpful suggestions that have greatly improved this paper. Support for Stanislav Anatolyev by Czech Science Foundation under grants 17-26535S and 20-28055S is gratefully acknowledged. Support for Anna Mikusheva by National Science Foundation under grant 1757199 is gratefully acknowledged. Address correspondence to Stanislav Anatolyev, CERGE-EI, Politickych vězňů 7, 11121 Prague 1, Czech Republic; e-mail: stanislav.anatolyev@cerge-ei.cz.

(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 and 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, and 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 microglobal 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 and Zhan (2015) and Anatolyev and 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, microglobal 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 and Wingender, 2016 and Hagedorn et al., 2015), or by running an OLS regression on the cross-section of entity-specific coefficients (Fama and MacBeth, 1973 and Sarto, 2018), or by running an IV regression on the cross-section of entity-specific coefficients (Anatolyev and Mikusheva, 2018).

The goal of this paper is to establish central limit theorems (CLTs) 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 nonpredictability 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 part. We also note that a need for a CLT for quadratic forms in factor models sometimes arises for the first-step estimators (e.g., Pesaran and Yamagata, 2018) or in higher order asymptotic derivations (e.g., Bai and 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 and 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 and Prucha (2020) 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

https://doi.org/10.1017/S0266466620000468 Published online by Cambridge University Press

and large T is established in Hahn, Kuersteiner, and Mazzocco (2020) for linear terms only.

This paper also contributes to the literature on the CLT for quadratic forms. Various types of CLTs for quadratic forms have been previously established and used in the many instrument literature (see, e.g., Chao et al., 2012 ; Hausman et al., 2012; Sølvsten, 2020) and many covariate literature (see Cattaneo, Jansson, and Newey, 2018), as well as in the literature on semi-parametric estimation (Cattaneo, Crump, and Jansson, 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.

This 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 CLTs 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 $i = 1, ..., N$, and observed for multiple time periods $t = 1, ..., T$. 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:

$$\Xi_{N,T} \equiv \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \xi_i \Rightarrow \mathcal{N}(0, \Sigma_\xi), \tag{1}$$

where

$$\xi_i = \begin{pmatrix} \frac{1}{\sqrt{T}} \sum_{s=1}^{T} v_s \gamma_i e_{is} \\ \frac{1}{T} \sum_{s=1}^{T} \sum_{t<s} w_{st} e_{it} e_{is} \end{pmatrix},$$

https://doi.org/10.1017/S0266466620000468 Published online by Cambridge University Press

and Σ_ξ is an asymptotic variance matrix. 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 $i = 1, ..., N$ (more exact conditions are to appear in the next section). We assume $\gamma_i, i = 1, ..., N$, to be nonrandom 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

$$\frac{1}{N}\sum_{i=1}^{N}\xi_i\xi_i' \xrightarrow{p} \Sigma_\xi. \tag{2}$$

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, λ, which is estimated in two steps. We consider a case when in the first step a researcher estimates a parameter β_i for each entity/unit/state $i = 1, ..., N$, typically via running OLS or IV time series regressions. A typical linear estimator can be written as $\widehat{\beta}_i = \beta_i + \varepsilon_i$, where the estimation error has the structure $\varepsilon_i = \big(1 + o_p(1)\big)\frac{1}{T}\sum_{t=1}^{T} v_t e_{it}$, 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.[2]

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. $\widehat{\lambda} = \frac{1}{N}\sum_{i=1}^{N}\gamma_i\widehat{\beta}_i$. Such an estimator is used in Sarto (2018). In order to justify asymptotic Gaussianity of $\widehat{\lambda}$ and to make statistical inferences about λ, 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 the structure of normalized averages stated as the first component of $\Xi_{N,T}$ from equation (1). 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 and 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

[1] By idiosyncratic error we mean the factor-removed part of entity-specific variables.

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

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

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

https://doi.org/10.1017/S0266466620000468 Published online by Cambridge University Press

employs the data on excess returns to a set of portfolios $\{r_{it},\ i=1,...,N,\ t=1,...,T\}$ and time series of a risk factor $\{F_t,\ t=1,...,T\}$. 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 $\widehat{\beta}_i^{(1)}=\frac{1}{T}\sum_{t=1}^{T}r_{it}$, and the risk exposure of a portfolio $\beta_i^{(2)}=\mathrm{var}(F_t)^{-1}\mathrm{cov}(r_{it},F_t)$ via the time series OLS regression of r_{it} on F_t resulting in an estimate $\widehat{\beta}_i^{(2)}$. At the second stage of the Fama–MacBeth procedure, one runs the OLS regression of the sample average return $\widehat{\beta}_i^{(1)}$ on the portfolio risk exposure estimated at the first step $\widehat{\beta}_i^{(2)}$. In this case,

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

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

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

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

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

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 γ_i). The second term in equation (3) is more complicated and calls for a CLT for quadratic forms:

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

$$=\frac{1}{T\sqrt{N}}\sum_{i=1}^{N}\sum_{t=1}^{T}v_t^{(1)}v_t^{(2)}\left(e_{it}^2-\mathbb{E}[e_{it}^2]\right)+\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{1}{T}\sum_{s=1}^{T}\sum_{t<s}w_{st}e_{it}e_{is},$$

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 and 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

https://doi.org/10.1017/S0266466620000468 Published online by Cambridge University Press

return $\widehat{\beta}_i^{(1)} = \frac{1}{T}\sum_{t=1}^{T} r_{it}$ and two estimates of the portfolio risk exposure computed as OLS estimates in regressions of r_{it} on F_t on different subsamples, say, $\widehat{\beta}_i^{(2)}$ and $\widehat{\beta}_i^{(3)}$. In our notation, the usage of a subsample is accommodated by setting $v_t = 0$ for those t not in the currently used subsample. The second step IV estimator is constructed as an IV estimator in the regression of $\widehat{\beta}_i^{(1)}$ on $\widehat{\beta}_i^{(2)}$ using $\widehat{\beta}_i^{(3)}$ as instrument. That is,

$$\widehat{\lambda} = \left(\sum_{i=1}^{N} \widehat{\beta}_i^{(3)} \widehat{\beta}_i^{(1)}\right)^{-1} \sum_{i=1}^{N} \widehat{\beta}_i^{(3)} \widehat{\beta}_i^{(2)}.$$

In order to make inferences on λ, 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 and Yamagata (2018) suggest a new test for factor pricing models 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 $i = 1, ..., N$, where α_i is a pricing error for the portfolio i. To estimate the pricing errors the authors use OLS estimates $\widehat{\alpha}_i$. A large number of portfolios N does not allow one to establish join Gaussianity of all $\widehat{\alpha}_i$ or to consistently estimate their covariance. Pesaran and Yamagata (2018) propose to test the hypothesis of interest using statistics based on a weighted sum of squares of $\widehat{\alpha}_i$. They create a properly normalized statistic of the form

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

where σ_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 and 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 and 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 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

[3] See Assump. 2 and 3 in Pesaran and Yamagata (2018).

https://doi.org/10.1017/S0266466620000468 Published online by Cambridge University Press

the structural error ε_t, and time-series and cross-sectional dependence in e_{it} is allowed, though restricted. Bai and Ng (2010) consider the bias-corrected GMM estimator that 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 $\{e_{it}, i = 1, \ldots, N, t = 1, \ldots, T\}$ cannot be regarded independent and/or identically distributed. In most realistic applications, one is usually willing to assume that e_{it} do not have a strong (detectable) factor structure, but still allow for some correlation between different units, which would not affect 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 independence assumptions in a time-series direction rather than in a cross-sectional direction. For example, the efficient market hypothesis implies mean nonpredictability 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 σ-algebra that contains at least the σ-algebras 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 γ_i as nonrandom $k_\gamma \times 1$ 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 $\mathrm{tr}(A)$ its trace, by $\max \mathrm{ev}(A)$—its maximal eigenvalue, and by $\mathrm{dg}(A)$ a diagonal matrix of the same

https://doi.org/10.1017/S0266466620000468 Published online by Cambridge University Press

size with the elements from the diagonal of A; $\|\cdot\|$ is the l_2 norm for a vector or the operator norm for a matrix.

Assumption 1. The random k_v-vector v_s and k_w-vector w_{st} are measurable with respect to the σ-algebra $\mathcal{F}$ for all s,t, and

(i) $\frac{1}{T}\sum_{s=1}^{T}\mathbb{E}(v_s v_s') \to \Omega_v$ and $\frac{1}{T^2}\sum_{s=1}^{T}\sum_{t<s}\mathbb{E}(w_{st}w_{st}') \to \Omega_w$, where Ω_v and Ω_w are full rank matrices;

(ii) $\max_{1\le s\le T}\mathbb{E}\big[\|v_s\|^4\big] < C$ and $\max_{1\le t,s\le T}\mathbb{E}[\|w_{st}\|^4] < C$;

(iii) $\mathbb{E}[\|\frac{1}{T^2}\sum_{s=1}^{T}\sum_{t<s}\big(w_{st}w_{st}' - \mathbb{E}[w_{st}w_{st}']\big)\|^2] \to 0$;

(iv) $\mathbb{E}[\|\frac{1}{T}\sum_{s=1}^{T}\big(v_s v_s' - \mathbb{E}[v_s v_s']\big)\|^2] \to 0$.

Assumption 2. $\max_{1\le i\le N}\|\gamma_i\| < C$.

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

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

Assumption 1 imposes very mild restrictions on the time-series behavior of the common (nonentity 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 CLT , 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:

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\xi_i = \begin{pmatrix} \frac{1}{\sqrt{T}}\sum_{s=1}^{T} v_s\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\gamma_i e_{is}\right) \\ \frac{1}{T}\sum_{s=1}^{T}\sum_{t<s} w_{st}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N} e_{it}e_{is}\right) \end{pmatrix},$$

and establish asymptotic convergence in the time-series direction. In order to apply a CLT in the time series direction we need some sort of asymptotic negligibility of summands with different time indexes, in particular, of terms like $\big\{v_s(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\gamma_i e_{is})\big\}_s$ and $\big\{w_{st}(\frac{1}{\sqrt{N}}\sum_{i=1}^{N} e_{it}e_{is})\big\}_{s,t}$. 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 can be 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 heteroscedasticity in e_{it} that can be related to some common

https://doi.org/10.1017/S0266466620000468 Published online by Cambridge University Press

variables from $\mathcal{F}$ producing dependence in higher-order conditional moments. This flexibility comes at the cost of imposing some structure on the cross-sectional behavior of e_{it}.

3.1. Independence from Common Variables

Assumption 4. (i) The errors $e_t = (e_{1t}, ..., e_{Nt})', t = 1, \ldots, T$ are independent from the σ-algebra $\mathcal{F}$ and identically distributed across t;

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

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

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

THEOREM 3.1. *Under Assumptions 1, 2, 3, and 4, the CLT stated in equation (1) holds with* $\Sigma_\xi = \begin{pmatrix} \Sigma_V & 0 \\ 0 & \Sigma_W \end{pmatrix}$, *where* $\Sigma_V = \Gamma_\sigma \otimes \Omega_v$ *and* $\Sigma_W = a\Omega_w$.

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 and 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 and 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, for example, 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 $g = 1, ..., G$, which may be overlapping, with each having independent shocks $f_{g,t}$ at time t. Stock i has nonzero 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 η_{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 $g = 1, ..., G$. Thus, for this 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.

https://doi.org/10.1017/S0266466620000468 Published online by Cambridge University Press

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 and Brown (1970). There are alternative CLTs 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. Another CLT for quadratic forms for time series data can be obtained using Bhansali, Giraitis, and Kokoszka (2007). The book by Giraitis, Koul, and Surgailis (2012) has a chapter on this subject and allows for long memory time series as well.

3.2. Conditional Heteroscedasticity

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 heteroscedasticity.

Assumption 5. The errors e_{it} have the following weak (unobserved) factor structure:

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

where the following assumptions hold:

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

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

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

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

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

https://doi.org/10.1017/S0266466620000468 Published online by Cambridge University Press

THEOREM 3.2. *Under Assumptions 1, 2, 3, and 5, the statement of the CLT 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 $\mathcal{F}$, 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 heteroscedasticity. Indeed, the conditional cross-sectional covariance is

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

Since we do not restrict $\mathbb{E}(f_t f_t' | \mathcal{F})$ 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 Γ_ω 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 analogue is the difference between the long-run covariance and instantaneous covariance in a classical time series. However, implementing an analogue of long-run covariance estimation here

https://doi.org/10.1017/S0266466620000468 Published online by Cambridge University Press

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 in (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 heteroscedastic case.

THEOREM 4.1. *If in addition to Assumptions 1, 2, 3, 4, we also have that*

$$\|\mathcal{E}_{N,T} - \mathrm{dg}(\mathcal{E}_{N,T})\| \to 0 \quad \text{as } N,T \to \infty, \tag{4}$$

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 ξ_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 $\frac{1}{N} \sum_{i=1}^N \tilde{\eta}_i^2 \xrightarrow{p} \sigma_\eta^2$, as all conditions of Theorems 3.2 and 4.2 are satisfied by cross-sectionally and time independent errors η_{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 $\frac{1}{\sqrt{N}} \sum_{i=1}^N \pi_i \gamma_i' \to \Gamma_{\pi\gamma}$ as $N \to \infty$. Thus,

$$\frac{1}{\sqrt{N}} \sum_{i=1}^N \xi_i^{(1)} \Rightarrow \mathcal{N}\left(0, \Gamma_{\pi\gamma} \Sigma_{fv} \Gamma_{\pi\gamma}' + \sigma_\eta^2\right),$$

while $\frac{1}{N} \sum_{i=1}^N \left(\xi_i^{(1)}\right)^2 \xrightarrow{p} \sigma_\eta^2$ 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),

https://doi.org/10.1017/S0266466620000468 Published online by Cambridge University Press

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 analogues of errors $e_{it}^* = \delta_t e_{it}, i = 1, ..., N, t = 1, ..., T$;
(iii) compute the bootstrap analogues ξ_i^* using the definition of ξ_i, with e_{it}^* in place of e_{it} for all $i = 1, ..., N, t = 1, ..., T$.

Then the distribution of $\Xi_{N,T}^* = \frac{1}{\sqrt{N}} \sum_{i=1}^N \xi_i^*$ has the same asymptotic limit as that of $\Xi_{N,T}$ has, and can be used for inferences. Alternatively, the distribution of $\Xi_{N,T}^*$ normalized by the bootstrap analogue of the variance estimate $\frac{1}{N} \sum_{i=1}^N \xi_i^* \xi_i^{*\prime}$ can be used to approximate the distribution of $\Xi_{N,T}$ normalized by the variance estimate $\frac{1}{N} \sum_{i=1}^N \xi_i \xi_i'$. 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 double/triple summations over i and time index/indices. For example,

$$\Xi_{N,T}^{(1)} = \frac{1}{\sqrt{N}} \sum_{i=1}^N \xi_i^{(1)} = \frac{1}{\sqrt{T}} \sum_{t=1}^T \left(\frac{1}{\sqrt{N}} \sum_{i=1}^N v_t \gamma_i e_{it} \right).$$

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 $\Xi_{N,T}^{(1)}$ is asymptotically Gaussian with mean zero and variance equal to the limit of $\frac{1}{T} \sum_{t=1}^T \mathbb{E}\left[\left(\frac{1}{\sqrt{N}} \sum_i v_t \gamma_i e_{it} \right)^2 \right]$. This limit clearly depends on how much there is cross-sectional correlation between e_{it} and e_{jt}. Now, in the bootstrapped samples,

$$\left(\Xi_{N,T}^{(1)} \right)^* = \frac{1}{\sqrt{N}} \sum_{i=1}^N \left(\xi_i^{(1)} \right)^* = \frac{1}{\sqrt{T}} \sum_{t=1}^T \left(\frac{1}{\sqrt{N}} \sum_{i=1}^N v_t \gamma_i e_{it} \right) \delta_t.$$

Conditional on the original sample, only δ_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 $\frac{1}{T} \sum_{t=1}^T \left(\frac{1}{\sqrt{N}} \sum_{i=1}^N v_t \gamma_i e_{it} \right)^2$ as $N, T \to \infty$. This limit coincides with $\frac{1}{T} \sum_{t=1}^T \mathbb{E}\left[\left(\frac{1}{\sqrt{N}} \sum_{i=1}^N v_t \gamma_i e_{it} \right)^2 \right]$ under relatively weak assumptions, as long as the Law of Large Numbers holds. For example, under Assumptions 5, we have:

$$\lim \frac{1}{T} \sum_{t=1}^T \left(\frac{1}{\sqrt{N}} \sum_{i=1}^N v_t \gamma_i e_{it} \right)^2 = \lim \frac{1}{T} \sum_{t=1}^T \left(\Gamma_{\pi\gamma} v_t f_t + \frac{1}{\sqrt{N}} \sum_{i=1}^N v_t \gamma_i \eta_{it} \right)^2$$
$$\xrightarrow{p} \Gamma_{\pi\gamma} \Sigma_{fv} \Gamma_{\pi\gamma}' + \sigma_\eta^2.$$

https://doi.org/10.1017/S0266466620000468 Published online by Cambridge University Press

A similar argument can be made about the second component as well. Indeed,

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

For a bootstrapped statistic, conditional on the original sample, the weights $\frac{1}{\sqrt{N}}\sum_{i=1}^{N} w_{st}e_{it}e_{is}$ are fixed, while δ_t'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 nontrivial 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 $\widehat{\Sigma}_\xi = \frac{1}{N}\sum_{i=1}^{N}\xi_i\xi_i'$ 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 e_{it} according to the following weak (unobserved) factor structure:

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

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_ω is tuned so that the average cross-sectional variance of η_{it} is unity. The parameter c_π indexes the degree of cross-sectional dependence as measured by the strength of the factor structure. Specifically, as $\sum_{i=1}^{N}\pi_i\pi_i' \to 1 + c_\pi^2$, 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_π. The errors generated this way are cross-sectionally dependent and heteroscedastic, 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 $\epsilon_{v,t} \sim iid\ \mathcal{N}(0,1)$ across $t = 1,...,T$, and $w_{st} = v_s v_t, s,t = 1,...,T$. All the disturbances $f_t, \tau_i, \epsilon_{\eta,it}$ 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)

https://doi.org/10.1017/S0266466620000468 Published online by Cambridge University Press

TABLE 1. Characteristics of simulated finite-sample distribution of $\Xi_{N,T}$ for $N = T = 500$

Element of $\Xi_{N,T} \rightarrow$	Linear				Quadratic			
$c_\pi \downarrow$	Mean	Skew	Kurt	Quant	Mean	Skew	Kurt	Quant
0.5	0.01	−0.05	3.01	1.63	−0.01	0.25	3.20	1.71
1	0.01	−0.04	2.97	1.64	−0.01	0.25	3.20	1.71
2	0.00	−0.00	2.95	1.65	0.02	0.20	3.11	1.71

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 $c_\pi = 0.5, 1$ and 2 correspond to a very weak, weak and moderately strong error factor structure.

requires $c_{fv} = 0$; the nonzero values of c_{fv} index deviations from the null hypothesis $\mathbb{E}(\xi_i) = 0$, 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 $\zeta_t \sim iid\ \mathcal{B}(\frac{1}{2})$ across $t, ..., T$. The bootstrap analogues of γ_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 5,000 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.

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 $\widehat{\Sigma}_\xi$ as a variance estimator and compare it with the symmetric standard Gaussian critical values. We also explore the performance of two wild bootstrap procedures—one that bootstraps $\Xi_{N,T}$ and another that bootstraps the t -statistics (referred to in Table 2 as bootstrap ξ 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

https://doi.org/10.1017/S0266466620000468 Published online by Cambridge University Press

TABLE 2. Simulated rejection rates for asymptotic and wild bootstrap tests

Element →	Linear									Quadratic								
	1%			5%			10%			1%			5%			10%		
$c_\pi \downarrow$	Asy	Bootstrap		Asy	Bootstrap		Asy	Bootstrap		Asy	Bootstrap		Asy	Bootstrap		Asy	Bootstrap	
	t	ξ	t	t	ξ	t	t	ξ	t	t	ξ	t	t	ξ	t	t	ξ	t
Size: $c_{fv} = 0$																		
0.5	2	1	1	8	5	5	14	10	10	0	1	1	3	6	5	6	10	10
1	7	1	1	16	5	5	24	10	10	0	1	1	3	6	5	6	10	10
2	25	1	1	38	5	5	46	10	10	0	1	1	3	6	6	7	11	12
Power: $c_{fv} = 0.1$																		
0.5	9	6	6	22	17	17	32	27	26	0	2	1	3	7	6	7	12	12
1	39	16	16	56	35	35	67	47	46	1	2	2	3	7	6	6	13	11
2	79	29	29	87	51	51	90	63	63	1	6	3	5	14	9	11	20	17
Power: $c_{fv} = 0.2$																		
0.5	35	28	26	58	51	49	68	63	62	1	5	3	5	13	9	10	19	16
1	90	73	71	96	88	87	97	93	93	2	10	5	9	21	15	17	30	25
2	100	93	92	100	98	98	100	99	99	18	45	29	40	61	50	53	68	61

Notes: The table contains actual rates, computed from 5,000 simulations for 10%, 5%, and 1% declared size asymptotic-t ("asy t"), bootstrap ("bootstrap ξ") 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 a very weak, weak, and moderately strong error factor structure.

$c_{fv} = 0$ 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 deviates 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 $\widehat{\Sigma}_\xi$ and becomes more pronounced with stronger cross-sectional dependence. What is surprising is that for the quadratic component of $\xi_{N,T}$, the distortions are relatively minor and not very sensitive to the strength of the factor structure. In contrast, both bootstrap procedures exhibit excellent size control and stability thereof 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. Other interesting areas involve establishing formal properties of the proposed wild bootstrap schemes, exploring the possibility of asymptotic refinements, and examining the superiority of one bootstrap scheme over the other.

APPENDIX: PROOFS

A.1. Preliminary Results

We use the following CLT for a vector valued martingale difference sequence:

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

(1) $(\min \mathrm{ev}(\Sigma_T))^{-2} \sum_{t=1}^T \mathbb{E}\big[\|Z_{t,T}\|^4\big] \to 0$,

(2) $(\min \mathrm{ev}(\Sigma_T))^{-2} \mathbb{E}\big[\big\| \sum_{t=1}^T Z_{t,T} Z_{t,T}' - \Sigma_T \big\|^2\big] \to 0$,

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 nonrandom $r \times 1$ vector λ, 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

https://doi.org/10.1017/S0266466620000468 Published online by Cambridge University Press

martingale difference sequence $z_t = \lambda' Z_{t,T}$ with variance $\sigma_T^2 = \operatorname{var}\big(\sum_{t=1}^T \lambda' Z_{t,T}\big) = \lambda' \Sigma_T \lambda$. Let us check that all conditions of the CLT by Heyde and Brown (1970) are satisfied for $\delta = 1$. Indeed,

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

and

$$\mathbb{E}\left[\left|\frac{\sum_{t=1}^T z_t^2}{\sigma_T^2} - 1\right|^2\right] = \mathbb{E}\left[\left|\frac{\sum_{t=1}^T (\lambda' Z_{t,T})^2}{\lambda' \Sigma_T \lambda} - 1\right|^2\right]$$
$$= \frac{1}{(\lambda' \Sigma_T \lambda)^2} \mathbb{E}\left[\left|\lambda' \left(\sum_{t=1}^T Z_{t,T} Z_{t,T}' - \Sigma_T\right) \lambda\right|^2\right]$$
$$\le \frac{1}{(\|\lambda\|^2 \min \operatorname{ev}(\Sigma_T))^2} \|\lambda\|^4 \mathbb{E}\left[\left\|\sum_{t=1}^T Z_{t,T} Z_{t,T}' - \Sigma_T\right\|^2\right] \to 0.$$

These two conditions imply that $\sigma_T^{-1} \sum_{t=1}^T z_t \Rightarrow \mathcal{N}(0,1)$. This finishes the proof. □

As a preliminary result, we establish a CLT 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 σ-algebra $\mathcal{F}$, and all e_t are independent from each other, conditionally on $\mathcal{F}$. Assume that*

$$\mathbb{E}(W_{st}|\mathcal{F}, e_t) = 0 \text{ and } \mathbb{E}(W_{st}|\mathcal{F}, e_s) = 0. \tag{A.1}$$

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

(i) $\Sigma_{W,T} \to \Sigma_W$, *where Σ_W is a full rank matrix;*
(ii) $T^4 \max_{1 \le t,s \le T} \mathbb{E}[\|W_{st}\|^4] < C$;
(iii) $\mathbb{E}\big[\big\|\sum_{s=1}^T \sum_{t<s} W_{st} W_{st}' - \Sigma_{W,T}\big\|^2\big] \to 0$;
(iv) $T^4 \max_{\substack{s_1 \ne s_2, t_1 \ne t_2 \\ t_1 < s_1, t_2 < s_2}} \big|\mathbb{E}\big(W_{s_1 t_2}' W_{s_2 t_1} W_{s_2 t_2}' W_{s_1 t_1}\big)\big| \to 0.$

Then, as $T \to \infty$,

$$W(T) \Rightarrow \mathcal{N}(0, \Sigma_W).$$

https://doi.org/10.1017/S0266466620000468 Published online by Cambridge University Press

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 σ-algebra $\mathcal{F}$, and $\mathbb{E}(V_s|\mathcal{F}) = 0$. Define $V(T) = \sum_{s=1}^{T} V_s$ and $\Sigma_{V,T} = \mathrm{var}(V(T))$. Assume the following statements hold as $T \to \infty$:*

(a) $\Sigma_{V,T} \to \Sigma_V$, *where Σ_V is a full rank matrix;*
(b) $T \max_{1 \le s \le T} \mathbb{E}[\|V_s\|^4] \to 0$;
(c) $\mathbb{E}[\| \sum_{s=1}^{T} V_s V_s' - \Sigma_{V,T} \|^2] \to 0$;
(d) $T^3 \max_{1 \le t < \min\{s_1, s_2\} \le T} \|\mathbb{E}(W_{s_1 t} V_{s_1}' V_{s_2} W_{s_2 t}')\| \to 0$.

Then, as $T \to \infty$

$$\begin{pmatrix} V(T) \\ W(T) \end{pmatrix} \Rightarrow \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma_V & 0 \\ 0 & \Sigma_W \end{pmatrix} \right).$$

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

$$\mathbb{E}\left(W_{s_1 t_1} \otimes W_{s_2 t_2} \cdots \otimes W_{s_k t_k}\right) = 0$$

when at least one index from the set $\{s_1, t_1, ..., s_k, t_k\}$ 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

$$\begin{aligned} \mathbb{E}\left(W_{s_1 t_1} \otimes W_{s_2 t_2} \otimes \cdots \otimes W_{s_k t_k}\right) &= \mathbb{E}\left[\mathbb{E}(W_{s_1 t_1} \otimes W_{s_2 t_2} \otimes \cdots \otimes W_{s_k t_k} | \mathcal{F}, e_{t_1}, e_{s_2}, e_{t_2}, ..., e_{t_k})\right] \\ &= \mathbb{E}\left[\mathbb{E}(W_{s_1 t_1} | \mathcal{F}, e_{t_1}, e_{s_2}, e_{t_2}, ..., e_{t_k}) \otimes W_{s_2 t_2} \otimes \cdots \otimes W_{s_k t_k}\right] \\ &= \mathbb{E}\left[\mathbb{E}(W_{s_1 t_1} | \mathcal{F}, e_{t_1}) \otimes W_{s_2 t_2} \otimes \cdots \otimes W_{s_k t_k}\right] = 0. \end{aligned}$$

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 $\mathcal{F}_s$ the σ -algebra generated by $\mathcal{F}$ and e_t for all $t < s$. Then, $(Z_{s,T}, \mathcal{F}_s)$ is a martingale difference sequence. Below we check that all conditions of Lemma A.1 are satisfied.

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

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

The last statement follows from the fact that W_{st} is clean, and nonzero 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 $\mathbb{E}\left[\|W_{st_1}\|^2 \|W_{st_2}\|^2\right] \le \frac{1}{2}\left(\mathbb{E}\left[\|W_{st_1}\|^4\right] + \mathbb{E}\left[\|W_{st_2}\|^4\right]\right) \le \max_{1 \le t, s \le T} \mathbb{E}\left[\|W_{st}\|^4\right] < CT^{-4}$ due to condition (ii). Hence, $\mathbb{E}\left[\|Z_{s,T}\|^4\right] \le CT^{-2}$. Thus, $\sum_{s=1}^{T} \mathbb{E}\left[\|Z_{s,T}\|^4\right] \le CT^{-1}$, implying that condition (1) of Lemma A.1 holds.

https://doi.org/10.1017/S0266466620000468 Published online by Cambridge University Press

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

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

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

$$\begin{aligned}
&\mathbb{E}\left[\left\|\sum_{s=1}^{T} Z_{s,T}Z_{s,T}' - \Sigma_{W,T}\right\|_F^2\right] \\
&\quad= \mathbb{E}\left[\left\|\sum_{s=1}^{T}\left(\sum_{t_1<s} W_{st_1}\right)\left(\sum_{t_2<s} W_{st_2}\right)' - \Sigma_{W,T}\right\|_F^2\right] \\
&\quad= \mathbb{E}\left[\left\|\sum_{s=1}^{T}\sum_{t<s}(W_{st}W_{st}' - \mathbb{E}[W_{st}W_{st}']) + \sum_{s=1}^{T}\sum_{t_1\neq t_2} W_{st_1}W_{st_2}'\right\|_F^2\right] \\
&\quad= \mathbb{E}\left[\left\|\sum_{s=1}^{T}\sum_{t<s}(W_{st}W_{st}' - \mathbb{E}[W_{st}W_{st}'])\right\|_F^2\right] + \mathbb{E}\left[\left\|\sum_{s=1}^{T}\sum_{t_1\neq t_2} W_{st_1}W_{st_2}'\right\|_F^2\right]. \qquad \textbf{(A.2)}
\end{aligned}$$

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):

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

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 $C\max_{1\leq t,s\leq T}\mathbb{E}\left[\|W_{st}\|^4\right] < CT^{-4}$. Thus,

$$\mathbb{E}\left[\left\|\sum_{s=1}^{T}\sum_{t_1\neq t_2} W_{st_1}W_{st_2}'\right\|_F^2\right] \to 0. \qquad \textbf{(A.3)}$$

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 $\mathcal{F}_s$ be defined as in the proof of Lemma A.2. We will show that all conditions of Lemma A.1 are satisfied. Notice

https://doi.org/10.1017/S0266466620000468 Published online by Cambridge University Press

that

$$\mathbb{E}\left[V_s W_{st}'\right] = \mathbb{E}\left[\mathbb{E}(V_s W_{st}'|\mathcal{F}, e_s)\right] = \mathbb{E}\left[V_s \mathbb{E}(W_{st}'|\mathcal{F}, e_s)\right] = 0.$$

Thus,

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

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 Σ_T is separated away from zero for large T. Now,

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

The first term here is bounded by $T \max_{1\le s\le T} \mathbb{E}\left[\|V_s\|^4\right]$ 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,

$$\begin{aligned}
\mathbb{E}\left[\left\|\sum_{s=1}^{T} Z_s Z_s' - \Sigma_T\right\|^2\right] &\le \mathbb{E}\left[\left\|\sum_{s=1}^{T} Z_s Z_s' - \Sigma_T\right\|_F^2\right] \\
&= \mathbb{E}\left[\left\|\sum_{s=1}^{T} V_s V_s' - \Sigma_{V,T}\right\|_F^2\right] + 2\mathbb{E}\left[\left\|\sum_{s=1}^{T}\left(\sum_{t<s} W_{st}\right) V_s'\right\|_F^2\right] \\
&\quad + \mathbb{E}\left[\left\|\sum_{s=1}^{T}\left(\sum_{t<s} W_{st}\right)\left(\sum_{t<s} W_{st}\right)' - \Sigma_{W,T}\right\|_F^2\right].
\end{aligned}$$

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

$$\mathbb{E}\left[\left\|\sum_{s=1}^{T} V_s V_s' - \Sigma_{V,T}\right\|_F^2\right] \le C\mathbb{E}\left[\left\|\sum_{s=1}^{T} V_s V_s' - \Sigma_{V,T}\right\|^2\right] \to 0.$$

During the proof of Lemma A.2 we show that

$$\mathbb{E}\left[\left\|\sum_{s=1}^{T}\left(\sum_{t<s} W_{st}\right)\left(\sum_{t<s} W_{st}\right)' - \Sigma_{W,T}\right\|_F^2\right] \to 0.$$

Finally,

$$\mathbb{E}\left[\left\|\sum_{s=1}^{T}\left(\sum_{t<s} W_{st}\right) V_s'\right\|_2^2\right] = \sum_{s_1=1}^{T}\sum_{t_1<s_1}\sum_{s_2=1}^{T}\sum_{t_2<s_2} \operatorname{tr}\left(\mathbb{E}\left(W_{s_1 t_1} V_{s_1}' V_{s_2} W_{s_2 t_2}'\right)\right)$$

https://doi.org/10.1017/S0266466620000468 Published online by Cambridge University Press

$$= \sum_{s_1=1}^{T}\sum_{s_2=1}^{T}\sum_{t<\min\{s_1,s_2\}} \operatorname{tr}\left(\mathbb{E}\left(W_{s_1 t}V'_{s_1}V_{s_2}W'_{s_2 t}\right)\right)$$
$$\leq CT^3 \max_{1\leq s_1,s_2,t\leq T}\left\|\mathbb{E}\left(W_{s_1 t}V'_{s_1}V_{s_2}W'_{s_2 t}\right)\right\| \to 0.$$

Here we used that $\mathbb{E}\left(W_{s_1 t_1}V'_{s_1}V_{s_2}W'_{s_2 t_2}\right) = 0$ 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

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

□

A.2. Proofs for Independent Case

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

$$W_{st} = \frac{1}{T\sqrt{N}} w_{st}\sum_{i=1}^{N} e_{it}e_{is} = \frac{1}{T} w_{st}\frac{e'_t e_s}{\sqrt{N}}$$

and

$$V_s = \frac{1}{\sqrt{TN}}\sum_{i=1}^{N}\gamma_i e_{is}\otimes v_s = \frac{1}{\sqrt{T}}\frac{\gamma' e_s}{\sqrt{N}}\otimes v_s.$$

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

$$\mathbb{E}\left[\left(\frac{e'_t e_s}{\sqrt{N}}\right)^2\right] = \frac{\operatorname{tr}\left(\mathcal{E}^2_{N,T}\right)}{N} \to a.$$

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

$$\Sigma_{W,T} = \operatorname{var}\left(\sum_{s=1}^{T}\sum_{t<s} W_{st}\right) = \frac{1}{T^2}\sum_{s=1}^{T}\sum_{t<s}\mathbb{E}(w_{st}w'_{st})\mathbb{E}\left[\left(\frac{e'_t e_s}{\sqrt{N}}\right)^2\right] \to a\Omega_w,$$

and the limit is a positive definite matrix.

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

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

https://doi.org/10.1017/S0266466620000468 Published online by Cambridge University Press

Here we used that $|\mathbb{E}\left(e_{i_1t}e_{i_2t}e_{i_3t}e_{i_4t}\right)| \leq \max_{1\leq i\leq N,\,1\leq t\leq T}\mathbb{E}\left(e_{it}^4\right) < C$ and Assumption 4(iv).

(iii) Next,

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

hence it is enough to prove that $\mathbb{E}\left[\|A_1\|^2\right] \to 0$ and $\mathbb{E}\left[\|A_2\|^2\right] \to 0$. 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,

$$\mathbb{E}\left[\mathrm{tr}(A_1A_1')\right] = \frac{1}{T^4}\sum_{s=1}^{T}\sum_{t<s}\mathbb{E}\left[\|w_{st}\|^4\right]\mathbb{E}\left[\left(\left(\frac{e_s'e_t}{\sqrt{N}}\right)^2 - \frac{\mathrm{tr}(\mathcal{E}_{N,T}^2)}{N}\right)^2\right]$$
$$\leq \frac{1}{T^4}\sum_{s=1}^{T}\sum_{t<s}\mathbb{E}\left[\|w_{st}\|^4\right]\mathbb{E}\left[\left(\frac{e_s'e_t}{\sqrt{N}}\right)^4\right] < \frac{C}{T^2}.$$

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

$$T^4\left|\mathbb{E}\left(W_{s_1t_2}'W_{s_2t_1}W_{s_2t_2}'W_{s_1t_1}\right)\right| \leq \mathbb{E}\left[\|w_{st}\|^4\right]\frac{\mathrm{tr}\left(E\left(e_{s_1}e_{s_1}'e_{t_1}e_{t_1}'e_{s_2}e_{s_2}'e_{t_2}e_{t_2}'\right)\right)}{N^2}$$
$$\leq \frac{C}{N^2}\mathrm{tr}\left(\mathcal{E}_{N,T}^4\right) \leq \frac{C}{N^2}N\max\mathrm{ev}\left(\mathcal{E}_{N,T}^4\right) \leq \frac{C}{N} \to 0.$$

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

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

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

and the limit is a full rank matrix.

(b) Next,

$$T\mathbb{E}\left[\|V_s\|^4\right] = \frac{1}{T}\mathbb{E}\left[\left\|\frac{1}{\sqrt{N}}\gamma'e_s\right\|^4\right]\mathbb{E}\left[\|v_s\|^4\right],$$

where $\mathbb{E}\left[\|v_s\|^4\right] \leq C$ due to Assumption 1(ii). Assumptions 2 and 4(iv) imply that

$$\mathbb{E}\left[\left\|\frac{1}{\sqrt{N}}\gamma'e_s\right\|^4\right] = \frac{1}{N^2}\sum_{i_1=1}^{N}\sum_{i_2=1}^{N}\sum_{i_3=1}^{N}\sum_{i_4=1}^{N}\mathbb{E}\left(e_{i_1t}e_{i_2t}e_{i_3t}e_{i_4t}\right)\gamma_{i_1}'\gamma_{i_2}\gamma_{i_3}'\gamma_{i_4}$$

https://doi.org/10.1017/S0266466620000468 Published online by Cambridge University Press

$$< \max_{1\le i\le N} \|\gamma_i\|^4 \frac{1}{N^2} \sum_{i_1=1}^{N}\sum_{i_2=1}^{N}\sum_{i_3=1}^{N}\sum_{i_4=1}^{N} |\mathbb{E}\left(e_{i_1t}e_{i_2t}e_{i_3t}e_{i_4t}\right)| < C.$$

(c) Next,

$$\begin{aligned}\sum_{s=1}^{T} V_s V_s' - \Sigma_{V,T} &= \left(\frac{1}{N}\gamma'\mathcal{E}_{N,T}\gamma\right) \otimes \left(\frac{1}{T}\sum_{s=1}^{T}\left(v_s v_s' - E\left(v_s v_s'\right)\right)\right) \\ &\quad + \frac{1}{T}\sum_{s=1}^{T}\left(\frac{\gamma' e_s e_s'\gamma}{N} - \frac{\gamma'\mathcal{E}_{N,T}\gamma}{N}\right) \otimes \left(v_s v_s'\right) \\ &= A_1 + A_2.\end{aligned}$$

Notice that A_1 and A_2 are uncorrelated, hence

$$\mathbb{E}\left[\left\|\sum_{s=1}^{T} V_s V_s' - \Sigma_{V,T}\right\|_F^2\right] = \operatorname{tr}\left(\mathbb{E}(A_1'A_1)\right) + \operatorname{tr}\left(\mathbb{E}(A_2'A_2)\right).$$

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

$$\begin{aligned}\operatorname{tr}\left(\mathbb{E}(A_2'A_2)\right) &= \frac{1}{T^2}\sum_{s=1}^{T}\mathbb{E}\left[\left|\frac{\gamma' e_s e_s'\gamma}{N} - \frac{1}{N}\gamma'\mathcal{E}_{N,T}\gamma\right|^2\right]\mathbb{E}\left[\|v_s\|^4\right] \\ &\le \frac{C}{T}\mathbb{E}\left[\left|\frac{\gamma' e_s e_s'\gamma}{N} - \frac{1}{N}\gamma'\mathcal{E}_{N,T}\gamma\right|^2\right].\end{aligned}$$

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

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

Using that scalars can be reshuffled to make two same-index e_t stand back to back and employing time series independence of errors, we obtain that

$$\begin{aligned}\left|\mathbb{E}\left(\frac{e_{s_1}'\gamma}{\sqrt{N}}\frac{\gamma' e_{s_2}}{\sqrt{N}}\frac{e_{s_1}'e_t}{\sqrt{N}}\frac{e_{s_2}'e_t}{\sqrt{N}}\right)\right| &= \frac{1}{N^2}\left|\operatorname{tr}\left(\gamma\gamma'\mathbb{E}(e_{s_2}e_{s_2}')\mathbb{E}(e_te_t')\mathbb{E}(e_{s_1}e_{s_1}')\right)\right| \\ &\le \frac{1}{N^2}\operatorname{tr}(\gamma\gamma')\max\operatorname{ev}\left(\mathcal{E}_{N,T}^3\right) \le \frac{C}{N}.\end{aligned}$$

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

$$T^3\max_{1\le t<\min\{s_1,s_2\}\le T}\left\|\mathbb{E}\left(W_{s_1t}V_{s_1}'V_{s_2}W_{s_2t}'\right)\right\| \le \frac{C}{N} \to 0.$$

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

https://doi.org/10.1017/S0266466620000468 Published online by Cambridge University Press

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 σ_i^2 the diagonal and by σ_{ij} the off-diagonal elements of matrix $\mathcal{E}_{N,T}$. Notice that the additional assumption of Theorem 4.1 implies that

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

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

$$\begin{aligned}
&\frac{1}{N} \sum_{i=1}^{N} \xi_{V,i} \xi_{V,i}' - \widetilde{\Sigma}_{V,T} \\
&\quad = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} \left(\left(\gamma_i \gamma_i' e_{is} e_{it}\right) \otimes \left(v_s v_t'\right) - \mathbb{I}\{s=t\} \sigma_i^2 \left(\gamma_i \gamma_i'\right) \otimes \mathbb{E}\left(v_t v_t'\right) \right) \\
&\quad = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(e_{it}^2 - \sigma_i^2\right) \left(\gamma_i \gamma_i'\right) \otimes \left(v_t v_t'\right) + \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{s \neq t} \left(\gamma_i \gamma_i' e_{is} e_{it}\right) \otimes \left(v_s v_t'\right) \\
&\qquad + \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(\gamma_i \gamma_i' \sigma_i^2\right) \otimes \left(v_t v_t' - \mathbb{E}\left(v_t v_t'\right)\right) \\
&\quad = A_1 + A_2 + A_3.
\end{aligned}$$

Notice that the three terms are uncorrelated, so it is enough to prove that $\operatorname{tr}\left(\mathbb{E}(A_j A_j')\right) \to 0$ for $j = 1, 2, 3$. Indeed, if the expectation of the Frobenius norm of a matrix converges to zero, this implies that each entry converges to zero as well. First,

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

https://doi.org/10.1017/S0266466620000468 Published online by Cambridge University Press

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

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_t's and all summands are mean zero. Thus,

$$\begin{aligned}\operatorname{tr}\left(\mathbb{E}(A_2A_2')\right) &= \frac{1}{N^2T^2}\sum_{t=1}^{T}\sum_{s\ne t}\sum_{i,j=1}^{N}\operatorname{tr}\left(\mathbb{E}(\gamma_i\gamma_i'\gamma_j\gamma_j'e_{it}e_{is}e_{jt}e_{js})\otimes\mathbb{E}(v_sv_s'v_tv_t')\right)\\ &= \frac{1}{N^2}\sum_{i,j=1}^{N}\operatorname{tr}(\gamma_i\gamma_i'\gamma_j\gamma_j'\sigma_{ij}^2)\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s\ne t}\operatorname{tr}\left(\mathbb{E}\left(v_sv_t'v_tv_s'\right)\right).\end{aligned}$$

We notice that $T^{-2}\sum_{t=1}^{T}\sum_{s\ne t}\operatorname{tr}\left(\mathbb{E}\left(v_sv_t'v_tv_s'\right)\right) \le \mathbb{E}\left[\|v_t\|^4\right] < C$ due to Assumption 1(ii). Denote r, r^* to be indexes that go over $1,...,k_\gamma$. For any fixed value of r, r^* denote $B^{(r,r^*)} = \left((\gamma_i\gamma_i')_{r,r^*}\right)_{i=1}^N$, an $N\times 1$ vector. Then,

$$\begin{aligned}\frac{1}{N^2}\sum_{i,j=1}^{N}\operatorname{tr}(\gamma_i\gamma_i'\gamma_j\gamma_j'\sigma_{ij}^2) &= \frac{1}{N^2}\sum_{i,j=1}^{N}\sum_{r,r^*}(\gamma_i\gamma_i')_{r,r^*}(\gamma_j\gamma_j')_{r,r^*}\sigma_{ij}^2\\ &= \sum_{r,r^*}\frac{1}{N^2}\sum_{i=1}^{N}(\gamma_i\gamma_i')_{r,r^*}(\gamma_i\gamma_i')_{r,r^*}\sigma_i^4 + \sum_{r,r^*}\frac{1}{N^2}B^{(r,r^*)\prime}\\ &\quad\times\left[(\mathcal{E}_{N,T}-\operatorname{dg}(\mathcal{E}_{N,T}))\odot(\mathcal{E}_{N,T}-\operatorname{dg}(\mathcal{E}_{N,T}))\right]B^{(r,r^*)}\\ &\le k_\gamma^2\max_{1\le i\le N}\|\gamma_i\|^4\left(\frac{1}{N^2}\sum_{i=1}^{N}\sigma_i^4+\frac{\sqrt{N}\|\mathcal{E}_{N,T}-\operatorname{dg}(\mathcal{E}_{N,T})\|}{N^2}\right)\\ &\le \frac{C}{\sqrt{N}}\to 0,\end{aligned}$$

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 $\operatorname{tr}\left(\mathbb{E}(A_2A_2')\right)\to 0$.

Finally, $\operatorname{tr}\left(\mathbb{E}(A_3A_3')\right)\to 0$ due to Assumption 1(iv). This ends the proof of statement (i).

Let us turn to statement (ii):

$$\begin{aligned}\frac{1}{N}\sum_{i=1}^{N}\xi_{W,i}\xi_{W,i}'-\Sigma_{W,T} &= \frac{1}{T^2N}\sum_{i=1}^{N}\sum_{s=1}^{T}\sum_{t<s}\left(e_{it}^2e_{is}^2-\sigma_i^4\right)w_{st}w_{st}'\\ &\quad+\frac{1}{T^2N}\sum_{i=1}^{N}\sum_{s_1=1}^{T}\sum_{t_1<s_1}\sum_{s_2=1}^{T}\sum_{\substack{t_2<s_2,\\ \{s_1,t_1\}\ne\{s_2,t_2\}}}w_{s_1t_1}w_{s_2t_2}'e_{it_1}e_{is_2}e_{it_2}e_{is_1}\end{aligned}$$

https://doi.org/10.1017/S0266466620000468 Published online by Cambridge University Press

$$+\frac{1}{N}\sum_{i=1}^{N}\sigma_i^4\frac{1}{T^2}\sum_{s=1}^{T}\sum_{t<s}\left(w_{st}w'_{st}-\mathbb{E}(w_{st}w'_{st})\right)$$
$$=A_1+A_2+A_3.$$

Again, A_1, A_2, and A_3 are uncorrelated with each other. Thus, we can deal with each one of them separately. We show that the expectation of the Frobenius norm of each matrix converges to zero, this implies that each entry converges to zero as well.

Let us start with

$$\operatorname{tr}\left(\mathbb{E}(A_1A_1')\right)=\frac{1}{T^4N^2}\sum_{i,j=1}^{N}\sum_{s,s^*=1}^{T}\sum_{t<s,t^*<s^*}\operatorname{tr}\left(\mathbb{E}(w_{st}w'_{st}w_{s^*t^*}w'_{s^*t^*})\right)\mathbb{E}(b_{i,t,s}b_{j,t^*,s^*}),$$

where

$$b_{i,t,s}=e_{it}^2e_{is}^2-\sigma_i^4=(e_{it}^2-\sigma_i^2)(e_{is}^2-\sigma_i^2)+\sigma_i^2(e_{is}^2-\sigma_i^2)+\sigma_i^2(e_{it}^2-\sigma_i^2).$$

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 nonzero summands in $\operatorname{tr}\left(\mathbb{E}(A_1A_1')\right)$. 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 nonzero summand is bounded above by $T^{-4}N^{-2}C\max_{1\le t,s\le T}\mathbb{E}[\|w_{st}\|^4]\max_{1\le i\le N,1\le t\le T}\mathbb{E}[(e_{it}^4)^2]$, thus $\operatorname{tr}(\mathbb{E}(A_1A_1'))\le C/T\to 0$.

The term $\operatorname{tr}\left(\mathbb{E}(A_2A_2')\right)$ includes summation over eight time indexes but most of the summands are zeros. The nonzero terms place at least four restrictions on the time indexes. We note that the nontrivial part of the sum in $\operatorname{tr}\left(\mathbb{E}(A_2A_2')\right)$ 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

$$\frac{1}{T^4N^2}\mathbb{E}\left(w_{s_1t_1}w'_{s_1^*t_1^*}w'_{s_2^*t_2^*}w_{s_2t_2}\right)\mathbb{E}\left(e_{it_1}e_{is_1}e_{it_1^*}e_{is_1^*}e_{jt_2}e_{js_2}e_{jt_2^*}e_{js_2^*}\right).$$

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 $T^{-4}N^{-2}\max_{1\le t,s\le T}\mathbb{E}[\|w_{ts}\|^4]\max_{1\le i\le N,1\le t\le T}\mathbb{E}(e_{it}^4)^2$ $\le C/(T^4N^2)$ in absolute value, and as there are at 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

$$C\frac{|\sigma_{ij}|^a(\sigma_i^2)^b(\sigma_j^2)^c}{T^4N^2}\max_{1\le s,t\le T}\mathbb{E}\left[\|w_{st}\|^4\right],$$

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

https://doi.org/10.1017/S0266466620000468 Published online by Cambridge University Press

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

$$\begin{aligned}\frac{C}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}|\sigma_{ij}|^a(\sigma_i^2)^b(\sigma_j^2)^c &\leq \frac{C}{N^2}\max_{1\leq i\leq N}\sigma_i^4\sum_{i=1}^{N}\sum_{j=1}^{N}\sigma_{ij}^2 \\ &= \frac{C}{N^2}\sum_{i=1}^{N}\sigma_i^4 + \frac{C}{N^2}\sum_{i=1}^{N}\sum_{i\neq j}^{N}\sigma_{ij}^2 \\ &\leq \frac{C}{N}\max_{1\leq i\leq N}\sigma_i^4 + \frac{C}{N^2}\|\mathcal{E}_{N,T} - \mathrm{dg}(\mathcal{E}_{N,T})\|_F^2 \leq \frac{C}{N}.\end{aligned} \tag{A.4}$$

In the first inequality, we use $|\sigma_{ij}| \leq \sigma_i\sigma_j$. In the second inequality, we use the definition of the Frobenius norm. In the last inequality, we use that for any symmetric matrix A, we have $\|A\|_F^2 \leq N\|A\|^2$ and assumption stated in Theorem 4.1. Thus, $\mathrm{tr}\left(\mathbb{E}(A_2A_2')\right) \to 0$.

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

Finally, we need to prove statement (iii) that

$$\frac{1}{NT^{3/2}}\sum_{i=1}^{N}\sum_{s=1}^{T}\sum_{t<s}\sum_{s^*=1}^{T}(\gamma_i\otimes v_{s^*})w_{st}'e_{is^*}e_{it}e_{is} \to^p 0.$$

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

$$\mathrm{tr}\left(\frac{1}{N^2}\sum_{i,j=1}^{N}\gamma_i\gamma_j'\left(C_1\sigma_{ij}\sigma_i^2\sigma_j^2 + C_2\sigma_{ij}^3\right)\right).$$

The term σ_{ij}^3 appears when $\{s_1, t_1, s_1^*\} = \{s_2, t_2, s_2^*\}$, while $\sigma_{ij}\sigma_i^2\sigma_j^2$ arises when the sets $\{s_1, t_1, s_1^*\}$ and $\{s_2, t_2, s_2^*\}$ have two coinciding indexes each. Therefore,

$$\mathrm{tr}\left(\frac{1}{N^2}\sum_{i,j=1}^{N}\gamma_i\gamma_j'\sigma_{ij}\sigma_i^2\sigma_j^2\right)$$

https://doi.org/10.1017/S0266466620000468 Published online by Cambridge University Press

$$= \operatorname{tr}\left(\frac{1}{N^2}\sum_{i=1}^{N}\gamma_i\gamma_i'\sigma_i^6\right) + \operatorname{tr}\left(\frac{1}{N^2}\sum_{i\neq j}(\gamma_i\sigma_i^2)(\gamma_j'\sigma_j^2)\sigma_{ij}\right)$$

$$= \frac{1}{N^2}\sum_{i=1}^{N}\|\gamma_i\|^2\sigma_i^6 + \frac{1}{N^2}\sum_{i\neq j}\operatorname{tr}(\gamma_i\gamma_j')\sigma_i^2\sigma_j^2\sigma_{ij}$$

$$\leq \max_{1\leq i\leq N}\|\gamma_i\|^2\left(\frac{1}{N}\max_{1\leq i\leq N}\sigma_i^6 + \frac{\|\mathcal{E}_{N,T}-\operatorname{dg}(\mathcal{E}_{N,T})\|}{N}\max_{1\leq i\leq N}\sigma_i^4\right)\to 0.$$

Also,

$$\operatorname{tr}\left(\frac{1}{N^2}\sum_{i,j=1}^{N}\gamma_i\gamma_j'\sigma_{ij}^3\right) = \frac{1}{N^2}\sum_{i=1}^{N}\|\gamma_i\|^2\sigma_i^6 + \frac{1}{N^2}\sum_{i\neq j}\operatorname{tr}(\gamma_i\gamma_j')\sigma_{ij}^3$$

$$\leq \max_{1\leq i\leq N}\|\gamma_i\|^2\left(\frac{1}{N}\max_{1\leq i\leq N}\sigma_i^6 + \frac{1}{N^2}\sum_{i,j=1}^{N}\sigma_{ij}^2\max_{1\leq i\leq N}\sigma_i^2\right)\to 0.$$

Here we used the statement $N^{-2}\sum_{i,j=1}^{N}\sigma_{ij}^2\to 0$, 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_{it} conditionally on $\mathcal{F}$, we have

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

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

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

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

$$\mathbb{E}\left[\left(\frac{e_t'e_s}{\sqrt{N}}\right)^2|\mathcal{F}\right] = \frac{1}{N}\operatorname{tr}\left(\left(\pi\mathbb{E}(f_tf_t'|\mathcal{F})\pi' + \Omega_\eta\right)\left(\pi\mathbb{E}(f_sf_s'|\mathcal{F})\pi' + \Omega_\eta\right)\right)$$

https://doi.org/10.1017/S0266466620000468 Published online by Cambridge University Press

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

where

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

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

$$\begin{aligned}\frac{1}{N}\operatorname{tr}\left(\Omega_\eta \pi \mathbb{E}(f_s f_s'|\mathcal{F})\pi'\right) &\le \frac{1}{N}\max_{1\le i\le N}\omega_i^2 \cdot \operatorname{tr}\left(\mathbb{E}(f_s f_s'|\mathcal{F})\pi'\pi\right)\\ &\le \frac{1}{N}\max_{1\le i\le N}\omega_i^2 \cdot \max \operatorname{ev}(\pi'\pi)\cdot \mathbb{E}\Big[\|f_s\|^2|\mathcal{F}\Big].\end{aligned}$$

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

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

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

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

(ii) Notice that

$$\frac{e_t'e_s}{\sqrt{N}} = \frac{f_t'\pi'\pi f_s}{\sqrt{N}} + \frac{f_t'\pi'\eta_s}{\sqrt{N}} + \frac{f_s'\pi'\eta_t}{\sqrt{N}} + \frac{\eta_t'\eta_s}{\sqrt{N}}.$$

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

$$\mathbb{E}\left[\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\eta_{it}\eta_{is}\right)^4\right] = \frac{1}{N^2}\sum_{i=1}^{N}\mathbb{E}\Big[(\eta_{it}\eta_{is})^4\Big] + C\frac{1}{N^2}\sum_{i_1\ne i_2}\mathbb{E}\big[\eta_{i_1t}^2\eta_{i_1s}^2\eta_{i_2t}^2\eta_{i_2s}^2\big] \le C,$$

$$\begin{aligned}\mathbb{E}\left[\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\pi_i\eta_{is}\right\|^4\right] &\le \frac{1}{N^2}\sum_{i=1}^{N}\mathbb{E}\Big[\|\pi_i\eta_{is}\|^4\Big]\\ &\quad + C\frac{1}{N^2}\sum_{i_1\ne i_2}\|\pi_{i_1}\|^2\|\pi_{i_2}\|^2\mathbb{E}\Big[\eta_{i_1s}^2\eta_{i_2s}^2\Big] \le \frac{C}{N^2},\end{aligned}$$

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

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

https://doi.org/10.1017/S0266466620000468 Published online by Cambridge University Press

Finally, due to Assumption 5(i),

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

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

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

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

We have:

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

so, it is enough to prove convergence of each term separately. Now, $\mathbb{E}\left[\mathrm{tr}(A_1A_1')\right]$ is equal to

$$\frac{1}{T^4}\sum_{s_1,s_2=1}^T\sum_{t_1,t_2}\mathbb{E}\left[\mathrm{tr}(w_{s_1t_1}w_{s_1t_1}'w_{s_2t_2}w_{s_2t_2}')\left(\left(\frac{e_{s_1}'e_{t_1}}{\sqrt{N}}\right)^2 - \vartheta_{s_1t_1}\right)\left(\left(\frac{e_{s_2}'e_{t_2}}{\sqrt{N}}\right)^2 - \vartheta_{s_2t_2}\right)\right].$$

Notice that in order for a summand from the last sum to be nonzero we need that some indexes in the set $\{s_1, s_2, t_1, t_2\}$ coincide, and we obtain at most CT^3 nonzero summands. Each nonzero summand is bounded above by a constant due to the moment assumptions formulated in Assumption 5(i,iii). Thus, $\mathbb{E}\left[\mathrm{tr}(A_1A_1')\right] \to 0$.

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

$$\left|\vartheta_{st} - \frac{1}{N}\sum_{i=1}^N\omega_i^4\right| \le \frac{C}{N}(\|f_s\| + \|f_t\| + 1)^4. \tag{A.5}$$

Thus,

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

https://doi.org/10.1017/S0266466620000468 Published online by Cambridge University Press

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

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

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

Let us check condition (iv):

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

where we used that the scalar products $e'_te_s=e'_se_t$ are scalars and they can be reshuffled to make two same-index e_t stand back to back. Let us bound the $N\times N$ matrix $\mathbb{E}(e_te'_t|\mathcal{F})=\pi\,\mathbb{E}(f_tf'_t|\mathcal{F})\pi'+\Omega_\eta$:

$$\begin{aligned}\max\text{ev}\left(\mathbb{E}(e_te'_t|\mathcal{F})\right)&\le\max\text{ev}\left(\pi'\mathbb{E}(f_tf'_t|\mathcal{F})\pi\right)+\max\text{ev}\left(\Omega_\eta\right)\\&\le\text{tr}\left(\pi'\mathbb{E}(f_tf'_t|\mathcal{F})\pi\right)+\max_{1\le i\le N}\omega_i^2\\&\le\max\text{ev}(\pi\pi')\mathbb{E}\left(\|f_t\|^2|\mathcal{F}\right)+C\\&\le C\mathbb{E}\left(\|f_t\|^2+1|\mathcal{F}\right).\end{aligned}\tag{A.6}$$

As a result,

$$\begin{aligned}\left|\mathbb{E}(e'_{s_1}e_{t_1}e'_{t_1}e_{s_2}e'_{s_2}e_{t_2}e'_{t_2}e_{s_1}|\mathcal{F})\right|&=\left|\text{tr}\left(\mathbb{E}(e_{t_1}e'_{t_1}|\mathcal{F})\mathbb{E}(e_{s_2}e'_{s_2}|\mathcal{F})\mathbb{E}(e_{t_2}e'_{t_2}|\mathcal{F})\mathbb{E}(e_{s_1}e'_{s_1}|\mathcal{F})\right)\right|\\&\le N\max\text{ev}\left(\prod_{t\in\{s_1,s_2,t_1,t_2\}}\mathbb{E}(e_te'_t|\mathcal{F})\right)\\&\le N\prod_{t\in\{s_1,s_2,t_1,t_2\}}\max\text{ev}\left(\mathbb{E}(e_te'_t|\mathcal{F})\right)\\&\le NC\prod_{t\in\{s_1,s_2,t_1,t_2\}}\mathbb{E}\left(\|f_t\|^2+1|\mathcal{F}\right).\end{aligned}$$

Also using Assumption 5(i) we obtain that

$$T^4\left|E\left(W'_{s_1t_2}W_{s_2t_1}W'_{s_2t_2}W_{s_1t_1}\right)\right|\le\frac{C}{N}\max_{1\le s,t,t^*\le T}E\left[\|w_{st}\|^4\|f_{t^*}\|^8\right]\to0.$$

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 $\Sigma_{V,T}$.

$$\begin{aligned}\mathbb{E}\left[\left(\frac{\gamma'e_s}{\sqrt{N}}\right)\left(\frac{\gamma'e_s}{\sqrt{N}}\right)'|\mathcal{F}\right]&=\left(\frac{\gamma'\pi}{\sqrt{N}}\right)\mathbb{E}[f_sf'_s|\mathcal{F}]\left(\frac{\pi'\gamma}{\sqrt{N}}\right)+\frac{1}{N}\sum_{i=1}^N\omega_i^2\gamma_i\gamma_i'\\&\to\Gamma'_{\pi\gamma}\mathbb{E}\left(f_sf'_s|\mathcal{F}\right)\Gamma_{\pi\gamma}+\Gamma_\omega.\end{aligned}$$

https://doi.org/10.1017/S0266466620000468 Published online by Cambridge University Press

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

$$\begin{aligned}\Sigma_{V,T} &= \operatorname{var}\left(\sum_{s=1}^{T} V_s\right) = \mathbb{E}\left[\frac{1}{T}\sum_{s=1}^{T}\mathbb{E}\left(\left(\frac{\gamma' e_s}{\sqrt{N}}\right)\left(\frac{\gamma' e_s}{\sqrt{N}}\right)' |\mathcal{F}\right)\otimes\left(v_s v_s'\right)\right] \\ &= \frac{1}{T}\sum_{s=1}^{T}\mathbb{E}\left[\left(\Gamma_{\pi\gamma}' f_s f_s' \Gamma_{\pi\gamma} + \Gamma_\omega\right)\otimes\left(v_s v_s'\right)\right] \\ &\to (\Gamma_{\pi\gamma}'\otimes I_{k_v})\Sigma_{fv}(\Gamma_{\pi\gamma}\otimes I_{k_v}) + \Gamma_\omega\otimes\Omega_v.\end{aligned}$$

The limit matrix is positive definite since both Γ_ω and Ω_v are positive-definite due to Assumptions 1 (i) and 5(iii).

Now note that due to Assumption 5(ii),

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

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

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

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

Collecting the pieces,

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

This gives us the validity of condition (b) of Lemma A.3.

(c) Denote $\Gamma_{\omega,N} = N^{-1}\sum_{i=1}^{N}\omega_i^2\gamma_i\gamma_i' \to \Gamma_\omega$. Then,

$$\begin{aligned}&\sum_{t=1}^{T} V_t V_t' - \Sigma_{V,T} \\ &= \frac{1}{T}\sum_{t=1}^{T}\left(\frac{\gamma' e_t}{\sqrt{N}}\frac{e_t'\gamma}{\sqrt{N}}\right)\otimes\left(v_t v_t'\right) - \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\left(\frac{\gamma'\pi}{\sqrt{N}} f_t f_t' \frac{\pi'\gamma}{\sqrt{N}} + \Gamma_{\omega,N}\right)\otimes\left(v_t v_t'\right)\right] \\ &= \frac{1}{T}\sum_{t=1}^{T}\left(\frac{\gamma' e_t}{\sqrt{N}}\frac{e_t'\gamma}{\sqrt{N}} - \frac{\gamma'\pi}{\sqrt{N}} f_t f_t' \frac{\pi'\gamma}{\sqrt{N}} - \Gamma_{\omega,N}\right)\otimes\left(v_t v_t'\right) \\ &\quad + \frac{1}{T}\sum_{t=1}^{T}\left[\left(\frac{\gamma'\pi}{\sqrt{N}} f_t f_t' \frac{\pi'\gamma}{\sqrt{N}} + \Gamma_{\omega,N} - \mathbb{E}\left[\frac{\gamma'\pi}{\sqrt{N}} f_t f_t' \frac{\pi'\gamma}{\sqrt{N}} + \Gamma_{\omega,N}\right]\right)\otimes\left(v_t v_t'\right)\right] \\ &= A_1 + A_2.\end{aligned}$$

https://doi.org/10.1017/S0266466620000468 Published online by Cambridge University Press

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

$$\begin{aligned}\mathbb{E}\left[\|A_1\|^2\right] &= \mathbb{E}\left[\left\|\frac{1}{T}\sum_{t=1}^{T}\left(\frac{\gamma'\pi}{\sqrt{N}}f_t\frac{\eta_t'\gamma}{\sqrt{N}}+\frac{\gamma'\eta_t}{\sqrt{N}}f_t'\frac{\pi'\gamma}{\sqrt{N}}+\left(\frac{\gamma'\eta_t}{\sqrt{N}}\frac{\eta_t'\gamma}{\sqrt{N}}-\Gamma_{\omega,N}\right)\right)\otimes\left(v_tv_t'\right)\right\|^2\right]\\ &\le \frac{1}{T^2}\sum_{t=1}^{T}\mathbb{E}\left[\left\|\frac{\gamma'\pi}{\sqrt{N}}f_t\frac{\eta_t'\gamma}{\sqrt{N}}+\frac{\gamma'\eta_t}{\sqrt{N}}f_t'\frac{\pi'\gamma}{\sqrt{N}}+\left(\frac{\gamma'\eta_t}{\sqrt{N}}\frac{\eta_t'\gamma}{\sqrt{N}}-\Gamma_{\omega,N}\right)\right\|^2\|v_t\|^4\right]\\ &\le \frac{1}{T}C\mathbb{E}\left[(\|f_t\|^2+1)\|v_t\|^4\right]\to 0.\end{aligned}$$

The former inequality is due to η_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):

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

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

$$\begin{aligned}\left|\mathbb{E}\left(\frac{e_{s_1}'\gamma}{\sqrt{N}}\frac{\gamma'e_{s_2}}{\sqrt{N}}\frac{e_{s_1}'e_t}{\sqrt{N}}\frac{e_{s_2}'e_t}{\sqrt{N}}|\mathcal{F}\right)\right| &= \frac{1}{N^2}\left|\operatorname{tr}\left(\gamma\gamma'\mathbb{E}(e_{s_2}e_{s_2}'|\mathcal{F})\mathbb{E}(e_te_t'|\mathcal{F})\mathbb{E}(e_{s_1}e_{s_1}'|\mathcal{F})\right)\right|\\ &\le \frac{1}{N^2}\operatorname{tr}(\gamma\gamma')\prod_{s\in\{s_1,s_2,t\}}\max\operatorname{ev}\left(\mathbb{E}(e_se_s'|\mathcal{F})\right)\\ &\le \frac{C}{N}\mathbb{E}\left[\prod_{s\in\{s_1,s_2,t\}}\left(\|f_s\|^2+1\right)|\mathcal{F}\right].\end{aligned}$$

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

$$T^3\max_{1\le t<\min\{s_1,s_2\}\le T}\left\|E\left(W_{s_1t}V_{s_1}'V_{s_2}W_{s_2t}'\right)\right\|\le\frac{C}{N}\to 0.$$

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

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) $N^{-1}\sum_{i=1}^{N}\xi_{V,i}\xi_{W,i}'\to 0$.

https://doi.org/10.1017/S0266466620000468 Published online by Cambridge University Press

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

$$\frac{1}{N}\sum_{i=1}^{N}\xi_{V,i}\xi_{V,i}' = \frac{1}{NT}\sum_{t=1}^{T}\sum_{s=1}^{T}\left(\sum_{i=1}^{N}\gamma_i\gamma_i'(\pi_i'f_s+\eta_{is})(\pi_i'f_t+\eta_{it})\right)\otimes(v_sv_t'). \tag{A.7}$$

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

$$\frac{1}{NT}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{i=1}^{N}\left(\gamma_i\gamma_i'\eta_{is}\eta_{it}\right)\otimes\left(v_sv_t'\right)\xrightarrow{p}\Sigma_V, \tag{A.8}$$

while terms that involve $\pi_i'w_s\pi_i'w_t$ or $\eta_{it}\pi_i'f_s$ converge to zero in probability. Indeed,

$$\begin{aligned}\frac{1}{NT}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{i=1}^{N}\left(\gamma_i\gamma_i'\eta_{is}\eta_{it}\right)\otimes(v_sv_t')-\Sigma_{V,T} &= \frac{1}{NT}\sum_{t=1}^{T}\sum_{s\neq t}^{T}\sum_{i=1}^{N}\left(\gamma_i\gamma_i'\eta_{is}\eta_{it}\right)\otimes(v_sv_t')\\ &\quad+\frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}\left(\gamma_i\gamma_i'\right)\otimes\left(\eta_{it}^2v_tv_t'-\omega_i^2\mathbb{E}\left(v_tv_t'\right)\right).\end{aligned}$$

We check that the first sum in the last expression is negligible:

$$\begin{aligned}&\mathbb{E}\left[\operatorname{tr}\left(\left(\frac{1}{NT}\sum_{t=1}^{T}\sum_{s\neq t}^{T}\sum_{i=1}^{N}\left(\gamma_i\gamma_i'\eta_{is}\eta_{it}\right)\otimes(v_sv_t')\right)^2\right)\right]\\ &\quad\le\frac{1}{N^2T^2}\sum_{t=1}^{T}\sum_{s\neq t}^{T}\sum_{i=1}^{N}\|\gamma_i\|^4\omega_i^4\mathbb{E}\left[\|v_t\|^2\|v_s\|^2\right]\\ &\quad\le\frac{C}{N^2}\sum_{i=1}^{N}\|\gamma_i\|^4\omega_i^4\to0.\end{aligned}$$

Here we use the conditional cross-sectional and temporal independence of η_{it}, that is, for $s\neq t$ we have $\mathbb{E}(\eta_{it}\eta_{is}\eta_{jt^*}\eta_{js^*}|\mathcal{F}) = \omega_i^4$ 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

$$\begin{aligned}&\operatorname{tr}\left(\mathbb{E}\left[\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\gamma_i\gamma_i'\eta_{it}^2\right)\otimes(v_tv_t')-\Sigma_V\right)^2\right]\right)\\ &=\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\operatorname{tr}\left(\mathbb{E}\left[\left(\left(\gamma_i\gamma_i'\eta_{it}^2\right)\otimes(v_tv_t')-\mathbb{E}\left[\left(\gamma_i\gamma_i'\eta_{it}^2\right)\otimes(v_tv_t')\right]\right)^2\right]\right)\\ &\le\frac{C}{N^2T^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\|\gamma_i\|^4\mathbb{E}\left[\|v_t\|^4\right]\to0.\end{aligned}$$

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

https://doi.org/10.1017/S0266466620000468 Published online by Cambridge University Press

Now consider terms in (A.7) that involve $\pi_i' f_s \pi_i' f_t$:

$$\frac{1}{NT}\sum_{t=1}^{T}\sum_{s=1}^{T}\left(\sum_{i=1}^{N}\gamma_i\gamma_i'\pi_i' f_s \pi_i' f_t\right)\otimes(v_s v_t') = \left(\frac{1}{N}\sum_{i=1}^{N}(\gamma_i\gamma_i')\otimes(\pi_i'\otimes\pi_i')\otimes I_{k_v}\right)\left(\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T} I_{k_\gamma}\otimes vec(f_s f_t')\otimes(v_s v_t')\right).$$

Using Assumption 2 and 5(ii) we can show that

$$\left\|\frac{1}{N}\sum_{i=1}^{N}(\gamma_i\gamma_i')\otimes(\pi_i'\otimes\pi_i')\right\| \le \frac{1}{N}\sum_{i=1}^{N}\|\gamma_i\|^2\|\pi_i\|^2 \le C\frac{1}{N}\sum_{i=1}^{N}\|\pi_i\|^2 \to 0.$$

Now observe that

$$\begin{aligned}\mathbb{E}\left[\left\|\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T} vec(f_s f_t')\otimes(v_s v_t')\right\|_F^2\right] &= \operatorname{tr}\left(\frac{1}{T^2}\mathbb{E}\left[\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{t^*=1}^{T}\sum_{s^*=1}^{T}\left(vec(f_s f_t')\,vec(f_{s^*}f_{t^*}')'\right)\otimes(v_s v_t' v_{s^*}v_{t^*}')\right]\right)\\ &\le C\frac{1}{T^2}\mathbb{E}\left[\sum_{t=1}^{T}\sum_{s=1}^{T}\|f_t\|^2\|f_s\|^2\|v_s\|^2\|v_t\|^2\right] < C.\end{aligned}$$

Here the equality is due to f_t's being serially independent and mean zero conditionally on $\mathcal{F}$ by Assumption 5(i) and $v_t \in \mathcal{F}$; 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

$$\frac{1}{NT}\sum_{t=1}^{T}\sum_{s=1}^{T}\left(\sum_{i=1}^{N}\gamma_i\gamma_i'\pi_i' f_s\pi_i' f_t\right)\otimes(v_s v_t') \xrightarrow{p} 0.$$

And finally, we show that

$$\frac{1}{NT}\sum_{t=1}^{T}\sum_{s=1}^{T}\left(\sum_{i=1}^{N}\gamma_i\gamma_i'\pi_i' f_s\eta_{it}\right)\otimes(v_s v_t') \xrightarrow{p} 0.$$

This holds because η_{it}'s are mean zero, cross-sectionally independent and independent from f_t conditionally on $\mathcal{F}$. 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

$$\frac{C}{N^2T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{i=1}^{N}\|\gamma_i\|^4\|\pi_i\|^2\omega_i^2\mathbb{E}\left[\|f_t\|^2\|v_t\|^2\|v_s\|^2\right] \to 0.$$

Thus, we proved statement (i).

https://doi.org/10.1017/S0266466620000468 Published online by Cambridge University Press

Let us turn to statement (ii):

$$\begin{aligned}\frac{1}{N}\sum_{i=1}^{N}\xi_{W,i}\xi'_{W,i}-\Sigma_{W,T}&=\frac{1}{T^2N}\sum_{i=1}^{N}\sum_{s=1}^{T}\sum_{t<s}w_{st}w'_{st}\left(e_{it}^2e_{is}^2-\omega_i^4\right)\\&+\frac{1}{T^2N}\sum_{i=1}^{N}\sum_{s=1}^{T}\sum_{t<s}\sum_{s^*=1}^{T}\sum_{t^*<s*,\,\{s,t\}\neq\{s^*,t^*\}}w_{st}w'_{s^*t^*}e_{it}e_{is}e_{it^*}e_{is^*}\\&+\frac{1}{N}\sum_{i=1}^{N}\omega_i^4\frac{1}{T^2}\sum_{s=1}^{T}\sum_{t<s}\left(w_{st}w'_{st}-\mathbb{E}(w_{st}w'_{st})\right)\\&=A_1+A_2+A_3.\end{aligned}$$

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,

$$\begin{aligned}\mathbb{E}\left[\|A_1\|_F^2\right]&=\frac{1}{T^4N^2}\sum_{s=1}^{T}\sum_{t<s}\mathbb{E}\left[\left\|\sum_{i=1}^{N}w_{st}w'_{st}\left(e_{it}^2e_{is}^2-\omega_i^4\right)\right\|_F^2\right]\\&\leq\frac{C}{T^4N^2}\sum_{s=1}^{T}\sum_{t<s}\sum_{i=1}^{N}\begin{pmatrix}\mathbb{E}\left[\|w_{st}\|^4\right]\max_{1\leq i\leq N,\,1\leq t\leq T}\mathbb{E}\left(\eta_{it}^4\right)^2\\+\sum_{j=1}^{N}\|\pi_i\|^4\|\pi_j\|^4\mathbb{E}\left[\|w_{st}\|^4\|f_t\|^4\|f_s\|^4\right]\end{pmatrix}\\&\to 0.\end{aligned}$$

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

$$\frac{C}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\|\pi_i\|^4\|\pi_j\|^4\leq\frac{C}{N^2}\max_{1\leq i\leq N}\|\pi_i\|^4\left(\sum_{i=1}^{N}\|\pi_i\|^2\right)^2\to 0, \tag{A.9}$$

and hence the term A_1 converges to zero.

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

$$\frac{1}{T^4N^2}\sum_{i,j=1}^{N}\sum_{\substack{s_1,s_1^*,\\s_2,s_2^*}}^{T}\sum_{\substack{t_m<s_m,\\t_m^*<s_m^*\\\{s_m,t_m\}\neq\{s_m^*,t_m^*\}}}\mathbb{E}\left[\mathrm{tr}(w_{s_1t_1}w'_{s_1^*t_1^*}w_{s_2t_2}w'_{s_2^*t_2^*})e_{it_1}e_{is_1}e_{it_1^*}e_{is_1^*}e_{jt_2}e_{js_2}e_{jt_2^*}e_{js_2^*}\right]. \tag{A.10}$$

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

$$\mathbb{E}(e_{it_1}e_{is_1}e_{it_1^*}e_{is_1^*}e_{jt_2}e_{js_2}e_{jt_2^*}e_{js_2^*}|\mathcal{F})\neq 0 \tag{A.11}$$

can be nonzero 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 nonzero 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

https://doi.org/10.1017/S0266466620000468 Published online by Cambridge University Press

(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 nonzero terms are bounded above by $\max_{1\le i\le N,\, 1\le s,t,t^*\le T} \mathbb{E}[\|w_{st}\|^4(\mathbb{E}(\eta_{it}^4) + C\|f_{t^*}\|^4)^2] < C$ due to Assumption 5(i,iii). There are at most CT^4N of such terms while the normalization is $N^{-2}T^{-4}$, 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 η_{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 η_{it} are zero, and only a nontrivial part of the term in (A.11) is the one including $\pi_i' f_t$ in place of e_{it}. So, every nonzeros term in the sum (A.10) is bounded by $\|\pi_i\|^4\|\pi_j\|^4\mathbb{E}\big[\|w_{st}\|^4\|f_t\|^8\big]$. 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 $A_2 \xrightarrow{p} 0$. The convergence $A_3 \xrightarrow{p} 0$ comes from Assumption 1(iii). This finishes the proof of (ii).

Finally, let us prove statement (iii):

$$\frac{1}{NT^{3/2}} \sum_{i=1}^{N} \sum_{s=1}^{T} \sum_{t<s} \sum_{s^*=1}^{T} (\gamma_i \otimes v_{s^*}) w_{st}' e_{is^*} e_{it} e_{is} \xrightarrow{p} 0.$$

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 cross-sectional summation (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. If we consider the cases when $i = j$, then all terms are bounded above by a constant and the number of nonzero 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 nontrivial effect is $\pi_i' f_t$; hence this sum is bounded by

$$\frac{1}{N^2} \sum_{i,j=1}^{T} \|\gamma_i\|\|\gamma_j\|\|\pi_i\|^3\|\pi_j\|^3 \max_{1\le s,t,s^*\le T} \mathbb{E}\left[\|v_{s^*}\|^2\|w_{st}\|^2\|f_{s^*}\|^2\|f_s\|^2\|f_t\|^2\right]$$

$$\le C\left(\frac{1}{N}\sum_{i=1}^{N}\|\gamma_i\|\|\pi_i\|^3\right)^2 \le \frac{1}{N^2}\max_{1\le i\le N}\|\gamma_i\|^2 \max_{1\le i\le N}\|\pi_i\|^4 \sum_{i=1}^{N}\|\pi_i\|^2 \to 0.$$

This ends the proof of Theorem 4.2. □

REFERENCES

Anatolyev, S. & A. Mikusheva (2018) Factor Models with Many Assets: Strong Factors, Weak Factors, and the Two-Pass Procedure. Working paper, CERGE-EI and Massachusetts Institute of Technology. arXiv:1807.04094.

https://doi.org/10.1017/S0266466620000468 Published online by Cambridge University Press

Ando, T. & J. Bai (2015) Asset pricing with a general multifactor structure. *Journal of Financial Econometrics* 13, 556–604.

Bai, J. (2009) Panel data models with interactive fixed effects. *Econometrica* 77, 1229–1279.

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

Bai, J. & S. Ng (2010) Instrumental variable estimation in a data rich environment. *Econometric Theory* 26, 1577–1606.

Bhansali, R.J., L. Giraitis, & P.S. Kokoszka (2007) Convergence of quadratic forms with nonvanishing diagonal. *Statistics & Probability Letters* 77, 726–734.

Cattaneo, M.D., R.K. Crump, & M. Jansson (2014a) Bootstrapping density-weighted average derivatives. *Econometric Theory* 30(6), 1135–1164.

Cattaneo, M.D., R.K. Crump, & M. Jansson (2014b) Small bandwidth asymptotics for density-weighted average derivatives. *Econometric Theory* 30(1), 176–200.

Cattaneo, M.D., M. Jansson, & W.K. Newey (2018) Alternative asymptotics and the partially linear model with many regressors. *Econometric Theory* 34(2), 277–301.

Chao, J.C., N.R. Swanson, J.A. Hausman, W.K. Newey, & T. Woutersen (2012) Asymptotic distribution of JIVE in a heteroskedastic IV regression with many instruments. *Econometric Theory* 28, 42–86.

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

Fama, E.F. & J. MacBeth (1973) Risk, return and equilibrium: Empirical tests. *Journal of Political Economy* 81, 607–636.

Giraitis, L., H.L. Koul, & D. Surgailis (2012) *Large Sample Inference for Long Memory Processes*. Imperial College Press.

Hagedorn, M., I. Manovskii, & K. Mitman (2015) The Impact of Unemployment Benefit Extensions on Employment: The 2014 Employment Miracle? Working paper 20884, National Bureau of Economic Research.

Hahn, J., G.M. Kuersteiner, & M. Mazzocco (2020) Central Limit Theory for Combined Cross-Section and Time Series. Working paper, University of California, Los Angeles and University of Maryland. arXiv:1610.01697.

Hausman, J.A., W.K. Newey, T. Woutersen, J.C. Chao, & N.R. Swanson (2012) Instrumental variable estimation with heteroskedasticity and many instruments. *Quantitative Economics* 3, 211–255.

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

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

Kuersteiner, G.M. & I.R. Prucha (2013) Limit theory for panel data models with cross sectional dependence and sequential exogeneity. *Journal of Econometrics* 174, 107–126.

Kuersteiner, G.M. & I.R. Prucha (2020) Dynamic spatial panel models: Networks, common shocks, and sequential exogeneity. *Econometrica* 88(5), 2109–2146.

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

Pesaran, M.H. (2006) Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica* 74, 967–1012.

Pesaran, M.H. & T. Yamagata (2018) Testing for Alpha in Linear Factor Pricing Models with a Large Number of Securities. Working paper, University of Southern California. http://dx.doi.org/10.2139/ssrn.2943640.

Rotar, V.I. (1973) Some limit theorems for polynomials of second degree. *Theory of Probability and Its Applications* 18, 499–507.

Sarto, A.P. (2018) Recovering Macro Elasticities from Regional Data. Working paper, Massachusetts Institute of Technology.

Serrato, J.C.S. & P. Wingender (2016) Estimating Local Fiscal Multipliers. NBER Working paper 22425.

Shanken, J. (1992) On the estimation of beta-pricing models. *Review of Financial Studies* 5, 1–33.

Sølvsten, M. (2020) Robust estimation with many instruments. *Journal of Econometrics* 214, 495–512.