Testing a large number of hypotheses in approximate factor models*

Dante Amengual
CEMFI, Casado del Alisal 5, E-28014 Madrid, Spain
<amengual@cemfi.es>

Luca Repetto
CEMFI, Casado del Alisal 5, E-28014 Madrid, Spain
<repetto@cemfi.es>

December 2014

Abstract

We propose a method to test hypotheses in approximate factor models when the number of restrictions under the null hypothesis grows with the sample size. We use a simple test statistic, based on the sums of squared residuals of the restricted and the unrestricted versions of the model, and derive its asymptotic distribution under different assumptions on the covariance structure of the error term. We show how to standardize the test statistic in the presence of both serial and cross-section correlation to obtain a standard normal limiting distribution. We provide estimators for those quantities that are easy to implement. Finally, we illustrate the small sample performance of these testing procedures through Monte Carlo simulations and apply them to reconsider Reis and Watson (2010)'s hypothesis of existence of a pure inflation factor in the US economy.

Keywords: Approximate factor model, hypothesis testing, principal components, large model analysis, large data sets, inflation.

JEL: C12, C33, C55

*We would like to thank Manuel Arellano, Jushan Bai, Serena Ng, Stéphane Bonhomme, Antonio Díez de los Ríos, Gabriele Fiorentini, Guillermo Pérez-Tellechea, Bernard Salanié, Enrique Sentana and Mark W. Watson as well as audiences at CEMFI Workshop, Columbia Econometrics Colloquium, Barcelona GSE Summer Forum (2014), IAAE Conference (Queen Mary, 2014), EEA-ESEM (Toulouse, 2014) and XXXIX Symposium on Economic Analysis (Palma de Mallorca, 2014) for helpful comments, discussions and suggestions. Of course, the usual caveat applies. Amengual gratefully acknowledges financial support from the Spanish Ministry of Science and Innovation through grant 2011-26342. Repetto acknowledges support from a doctoral grant of the AXA Research Fund.

1 Introduction

Economic and financial time series exhibit many distinctive features such as trends, cycles, serial correlation, time-varying volatility, and breaks. Summarizing those features in a group of variables by means of a few pervasive common components is useful for developing parsimonious models. It can also provide evidence regarding the relevance of economic theories that imply such common features. Indeed, in most applications of large-scale factor models, economic theory has important implications that can be tested with data. An example is given by the restricted factor model for sectoral prices proposed by Reis and Watson (2010). In that paper, the authors model sectoral price indices in the US as a function of few unobserved factors, and ask whether one of them can be understood as a "pure inflation" factor, in the sense that it moves all absolute prices up or down by the same amount, leaving relative prices unchanged. The restriction is then that the loadings of at least one of the factors are constant in all sectors.

A second example is found in Chamberlain and Rothschild (1983) who, in addition to introducing the concept of "approximate factor model", nowadays widely popular in the literature of factor models with large cross-sectional (N) and time-series (T) dimensions, investigate the question of whether investors, by allocating their purchases among many assets, can create a portfolio that is riskless, has unitary cost, and yields a positive return. Specifically, they show that a riskless asset will exist unless the sequence of cross-sectional covariance matrices that characterizes market returns has the same structure as it would have if there were a random event which affected the returns of all assets in precisely the same way. In that context, testing that all the loadings of at least one of the factors are constant amounts to test for the absence of a riskless asset in the economy.[1]

Hypotheses testing in a large N, large T factor model often means, like in the aforementioned examples, restricting an entire vector of factors, factor loadings or both. Because factors and factor loadings are unobserved and asymptotic theory must account for situations in which both N and T go to infinity jointly, standard results need not apply. Moreover, the number of restrictions might grow with N, T or both. In this regard, issues with classical tests when there are many regressors or many restrictions are well documented. For instance, Berndt and Savin (1977) confirm the existence of considerable conflicts among the classical tests when the number of restrictions is comparable to the sample size. Similarly, Evans and Savin (1982) show that conflicts are more likely to appear when the ratio of the number of restrictions to the difference between the number of observations

[1]Intuitively, if all assets are affected by the same random event, the market will not allow investors to diversify risks so effectively that they can create a riskless portfolio with a positive return.

and the number of parameters is large. There are situations, though, in which under some additional conditions, an appropriately standardized version of the classical test statistics is still asymptotically normal (see Anatolyev (2012) for a discussion in the linear regression model).

Another issue is that in an approximate factor model the dependence structure of the errors is generally left unspecified and therefore the likelihood function is unknown. As is well known, the usual asymptotic chi-squared distribution for the likelihood-ratio test statistic is based on the assumptions that the data come from a correctly specified model and that the parameters satisfy the null hypothesis. In general, the likelihood-ratio statistic no longer follows an asymptotic chi-squared distribution when the model is misspecified, even when the null hypothesis is true (Kent (1982)).

The purpose of our paper is to develop a hypothesis testing framework in which the number of restrictions entering the null hypothesis grows with the sample size, possibly at a rate of $N + T$. Specifically, we propose a simple testing procedure based on the –properly recentered and rescaled– difference of the residual sum of squares from the restricted and unrestricted versions of model, and obtain the asymptotic distribution of the test statistic. Interestingly, our framework remains valid even when both times-series and cross-sectional dependence among error terms is allowed, as long as this dependence is weak enough.

We initially assume that idiosyncratic components are *iid*, and then we extend our analysis to the case of weak dependence. Given the similarity of our test statistic with the $F-$test, it is perhaps not surprising that, in the *iid* case, we find that the classical result for testing the linear regression model extends to the contest of large factor models. Specifically, the standard χ^2 result with degrees of freedom equal to the number of restrictions ν is extended to the case $\nu = O(N+T)$ by recentering and rescaling the difference of the restricted and unrestricted residual sum of squares by ν and 2ν, respectively, in order to obtain a standard normal limiting distribution. The asymptotic distribution of the test statistic does not depend on factors nor other unobserved quantities so practical implementation is straightforward, and only requires to estimate the restricted and unrestricted models to compute the corresponding sums of squared residuals.

We then introduce cross-sectional correlation, although maintaining the assumption of time-series independence, in the idiosyncratic components to obtain our second asymptotic result. Interestingly, the presence of cross-sectional weak dependence in the errors introduces a variance correction but leaves the mean ν unchanged. We characterize the variance expression by a simple scalar, that can be interpreted as a measure of the average squared correlations of the error terms, and provide a consistent estimator for this quantity. Although the resulting test statistic could be interpreted as a particular way of pooling individual test statistics performed for each series, it differs from the previous ones in the related literature in that the pooling is done over the cross-section dimension i and time-series dimension t at the same time and, without assuming independence over i.

Finally, we allow for both serial and cross-sectional correlation and obtain that, in contrast to the first two cases, the recentering and rescaling factors depend on both the factors and on the dependence structure of the error components. The rescaling factor can be decomposed into three components: i) the variance of the *iid* case, ii) a term that captures the variance inflation due to cross-sectional dependence in the errors, and iii) another term, which now also involves factors, which corrects for the presence of times-series dependence. Remarkably, although the dependence structure of the unobservables is left partially unmodeled, our test statistic, which resembles a likelihood-ratio type test statistic, still has a limiting standard normal distribution since the scaling factor that affect each of the chi-squared terms is correct on average.

In order to assess size and power properties of the testing procedures in small samples, in Section 4 we perform a series of Monte Carlo experiments for a variety of settings, starting from *iid* error terms and moving on to situations with both heterogeneous cross-sectional and times-series weak dependence in the idiosyncratic components, under different assumptions about factors dynamics. We find that the testing procedures have good size and power properties under correct specification. More importantly, the performance of the most robust version of the test is satisfactory despite the fact that it involves estimation of additional quantities to account for cross-sectional and time-series correlations.

We then provide an application of the test by taking the data in Reis and Watson (2010) and reconsidering their hypothesis of the existence of a pure inflation factor in the US price indices. Not surprisingly, and in line with their finding that individual t-tests rejected the null of unit loading 30% of the times at the 5% level, the null of the presence of one factor with constant loadings is strongly rejected under the three versions of the test. We also estimate a restricted version of the model at a more disaggregated level by allowing for heterogeneity across three and thirteen subcategories and find that again the tests reject the null hypothesis of homogeneous loading within groups. Interestingly, a more detailed inspection of the factor loading estimates suggest that in the model with three groups although there seem to be moderate discrepancy in the loadings of durables, non durables and services, there is substantial heterogeneity within those groups when we allow for a higher level of disaggregation.

Although for clarity of exposition we focus on the null that the loadings of one of the factors are constant in all the series, the testing procedure we develop and the asymptotic results we provide can easily accommodate other situations. For instance, in panel data it is typically assumed that the unobserved individual-specific heterogeneity is fixed across time, but there are several cases in which this assumption may not hold; hence, another potentially interesting case is that of testing for constant individual effects. Analogously, our approach can also be used to test the fixed effect with time dummies panel data model versus the interactive fixed effects model of Bai (2009) in the context of both large N and T.

The structure of the paper is as follows. In Section 2, we introduce the model in more detail, discuss the main assumptions, and define the null hypothesis of interest. In Section 3 we describe the test statistics and derive their asymptotic properties under different assumptions about the errors of the model. Practical implementation issues together with a detailed Monte Carlo evaluation of the testing procedures in terms of size and power are given in Section 4; while in Section 5 we revisit Reis and Watson (2010)'s hypothesis of existence of a pure inflation factor in the US economy. In Section 6 we briefly discuss straightforward extensions of the proposed methodology to test for equality of a subset of loadings, for time-invariant fixed effects as well as for interactive fixed effects. Finally, we present our conclusions and suggestions for future work in Section 7. Proofs and auxiliary results are gathered in the Appendix.

2 Model and assumptions

Consider a factor model for a panel of $i = 1, ..., N$ series and $t = 1, ..., T$ periods of the form

$$Y_t = \mathbf{L}\mathbf{F}_t + e_t \tag{1}$$

or, in matrix form,

$$Y = \mathbf{F}\mathbf{L}' + e,$$

where Y denotes a $T \times N$ matrix of observed variables, $\mathbf{F}$ is a $T \times r$ matrix of r latent factors, $\mathbf{L}$ is an $N \times r$ matrix of factor coefficients (or factor loadings) λ_{ij}'s and e is an $N \times T$ vector of idiosyncratic errors.

Since we focus on a situation in which N and T are both large, our asymptotic results require that $N, T \to \infty$ jointly or, equivalently, that $N = N(T)$ with $\lim_{T\to\infty} N(T) = \infty$. Although no restrictions on the relative rates of N and T are required a priori.

Throughout this paper, the number of factors r is assumed to be fixed and known as N and T grow. We are interested in testing hypotheses involving restrictions on an entire $N \times 1$ vector of loadings, an entire $T \times 1$ vector of factors, or both. One of the first problems to tackle is to find a mathematically efficient way to express the limit of a vector whose size grows to infinity. For instance, if the null hypothesis is

$$H_0 : \lambda_{1i} = \bar{\lambda} \text{ for all } i = 1, ..., N, \tag{2}$$

the dimension of the parameter space diverges to infinity with N. An intuitive way to think about it is to associate a probability measure to the vector $(\lambda_{11}, ..., \lambda_{1N})'$, in this case a degenerate one at $\bar{\lambda}$. Associating a measure to a vector in the way we described is meaningful mostly when one wants to have information about the whole set of values taken by the coordinates of the vector, and not about each coordinate.[2]

We now give a couple of examples in which our proposed testing procedure may be useful.

Example 1. *Reis and Watson (2010) model quarterly changes in sectoral price indices y_{it} in terms of three unobserved factors. They are interested in testing whether a "pure inflation" factor exists or not. A pure inflation factor is a shock that shifts all prices up or down by the same fraction, thus leaving relative prices unchanged. The null hypothesis in this case is (2) against the more general alternative*

$$H_1 : \lambda_{1i} = \bar{\lambda} + \eta_i.$$

The null hypothesis in this case, therefore, imposes N restrictions by requiring all the factor loading associated to the first factor f_{1t} (the pure inflation factor) to be constant.

Example 2. *In a panel data context it is customary to control for common time effects or trends f_t and individual fixed effects λ_i by including time and individual dummies in the estimation. This amounts to modeling the unobserved heterogeneity in the model as $\lambda_i + f_t$. If multiplicative (also known as "interactive") effects of the form $\lambda_i f_t$ are present, however, the within-group estimator is inconsistent because it relies on additivity. An interactive effect arises when, for instance, the impact of a global business cycle factor on the dependent variable is different from one country to another. Testing whether a common trend specification is suitable or whether a more flexible, interactive, specification has to be used is equivalent to a test of additive against interactive fixed effects as in Bai (2009), where*

$$H_0 : [\lambda_i, 1] \begin{bmatrix} 1 \\ f_t \end{bmatrix} = \lambda_i + f_t$$

is tested against the alternative

$$H_1 : [\lambda_{1i}, \lambda_{2i}] \begin{bmatrix} f_{1t} \\ f_{2t} \end{bmatrix} = \lambda_{1i} f_{1t} + \lambda_{2i} f_{2t}.$$

[2] For an insightful discussion, see El Karoui (2008).

In this case, the null hypothesis imposes T restrictions by forcing the first factor to be constant for all $t = 1, ..., T$ plus other N restrictions by forcing the second loading to be constant for each i, for a total of $N + T$ restrictions.

In order to keep the notation simple, from now on we focus on the setup and the null hypothesis of Example 1, in which N restrictions on the matrix of factor loadings are tested.[3] Consider then the following representation for a model with r factors in which f_t denotes the first factor with restricted loadings $\lambda_{1i} = \bar{\lambda}$ for all $i = 1, ..., N$, and g_t denotes the remaining $r-1$ factors with unrestricted factor loadings γ_i,

$$y_{it} = \bar{\lambda} f_t + \gamma_i' g_t + e_{it},$$

or, in matrix form,

$$Y = \mathbf{F}\mathbf{L}' + e = F\Lambda' + \mathbf{G}\Gamma' + e \tag{3}$$

where $\mathbf{F} = (F, \mathbf{G})$, $\mathbf{L} = (\Lambda, \Gamma)$, $\mathbf{f} = (f_1, ..., f_T)'$, $\mathbf{G} = (g_1, ..., g_T)'$, $\Lambda = (\lambda_1, ..., \lambda_N)'$ and $\Gamma = (\gamma_1, ..., \gamma_N)'$.

In this context, under the null the model imposes restrictions on the loadings of the first factor f_t; given that the scale of Λ and F is not identified we fix their relative scales by assuming that Λ is equal to $\ell_N = (1, ..., 1)'$.

We now briefly describe the assumptions of the model which are grouped into three blocks: assumptions about the factors and factor loadings, assumptions about the errors, and assumptions about the dependence among those groups.

Assumption A: *Factors and Factor Loadings*

Let $F_t = (f_t, g_t')'$ denote the vector of latent factors at time t and $\mathbf{L} = [\Lambda, \Gamma]$ be the $N \times r$ matrix of factor loadings. We assume that:

(A1) For each t, F_t is a covariance stationary and ergodic process. Furthermore, $E||F_t||^{8+\epsilon} \leq M \leq \infty$, for some $\epsilon > 0$ and

$T^{-1} \sum_t F_t F_t' \xrightarrow{p} \Sigma_{\mathbf{F}}$, a positive definite, $r \times r$ matrix.

(A2) For each i, $\|L_i\| \leq M \leq \infty$ and $||\mathbf{L}'\mathbf{L}/N - \Sigma_{\mathbf{L}}|| \to 0$ for some positive definite, $r \times r$ diagonal matrix $\Sigma_{\mathbf{L}}$. Alternative, loadings can be random, provided that $E||L_i||^{8+\epsilon} \leq M$.

(A3) The eigenvalues of the $r \times r$ matrix $\Sigma_{\mathbf{L}}\Sigma_{\mathbf{F}}$ are distinct.

These assumptions are in line with Bai (2003). Assumption A serves to identify the factors: The nonsingular limiting values of $\Sigma_{\mathbf{L}}$ and $\Sigma_{\mathbf{F}}$ through A1 and A2 imply that each of the factors provides

[3] Notice that $\bar{\lambda}$ is generally not identified, and therefore, the actual number of hypotheses is $N-1$; however, in a large N, large T context, this discrepancy is negligible.

a nonnegligible contribution to the average variance of y_{it}, where y_{it} is the i'th element of Y_t and the average is taken over both the times-series and the cross-sectional dimension. Notice that assumption A1 allows for factors to be serially correlated, although we maintain the static representation in the sense that lagged values of F_t do not enter into y_{it} directly.[4] Notice also that Assumption A1 rules out stochastic trends, unit roots and other processes with non-constant unconditional second moments. Finally, Assumptions A3 implies that the eigenvalues are assumed to be distinct.

Assumption B: *Time and Cross-section Dependence of the Errors*

There are two sequences of $N \times N$ and $T \times T$ matrices Θ and Φ, respectively, such that disturbances e_{it} are generated as $e = \Phi\varepsilon\Theta$, where ε is a $T \times N$ matrix of *iid* elements such that:

(B1) $E(\varepsilon_{it}) = 0$, $V(\varepsilon_{it}) = 1$, $E(\varepsilon_{it}^4) = (3 + \kappa_\varepsilon)$ and $E(\varepsilon_{it}^{8+\epsilon}) < \infty$.

Let $eig_j(A)$ denote the j'th largest eigenvalue of the square matrix A. Then:

(B2) $eig_1(\Theta\Theta') < M$ and $eig_1(\Phi'\Phi) < M$.

(B3) $eig_N(\Theta\Theta') > 0$ and $eig_T(\Phi'\Phi) > 0$.

Assumption B allows for limited time series and cross-section dependence in the idiosyncratic components as in the approximate factor model of Chamberlain and Rothschild (1983), Connor and Korajczyk (1986) and Connor and Korajczyk (1993). Heteroskedasticity in both dimensions is also allowed. Although normality is not assumed, Assumption B1 limits the size of fourth moments of the idiosyncratic components. Φ is a matrix that captures time series correlation and Θ is a matrix that induces cross-section correlation. As discussed in Harding (2013), this assumption restricts the structure of dependence in that assumes separability of the covariance matrix i.e. $V(vec(e)) = (\Phi'\Phi \otimes \Theta\Theta')$. According to Assumption B, a generic single element e_{it} can be written as

$$e_{it} = \sum_{u=1}^{T} \sum_{h=1}^{N} \theta_{ih} \varepsilon_{hu} \phi_{ut}, \tag{4}$$

so that, if we assume $\Phi = I_T$ errors are serially uncorrelated whereas if $\Theta = I_N$ we have no cross-sectional correlation in the disturbances. Although this structure is somewhat restrictive, it has the nonparametric nature of the corresponding assumptions in Bai and Ng (2002) and Stock and Watson (2002) since the entries of Θ and Φ are left unspecified apart from the requirement implied by Assumptions B2 and B3. Assumption B2, by imposing restrictions on the size of the largest eigenvalues of $\Phi'\Phi$ and $\Theta\Theta'$, assures that the correlation is weak enough for a central limit theorem to hold. In

[4] Allowing for lagged values of the factors to enter the equation for X_{it} can be easily done through a minor change of notation. Consider for instance a standard dynamic factor model in which $X_{it} = \lambda_i^q(L) f_t + e_{it}$ where $\lambda_i^q(L)$ is a lag polynomial of order q in nonnegative power of the lag operator L. Then, this model is equivalent to the static factor model (3) with $F_t = (f_t, f_{t-1}, ..., f_{t-q})'$. Notice, however, that our setup excludes the generalized dynamic factor model considered by Forni, Hallin, Lippi, and Reichlin (2005) where the polynomial distributed lag possibly tends to infinity.

turn, Assumption B3 requires $V(vec(e))$ to be a positive definite matrix.[5]

Structures with dependence in only one dimension are easily accommodated under this parametrization. For instance, Pesaran and Yamagata (2012) and Choi (2012) study factor models with cross-section correlation and no time series dependence i.e. with $\Phi = I_T$, a common assumption in the asset pricing literature. A particular case of our setup is the strict factor model structure introduced in the APT theory of Ross (1976) which assumes e_{it} to be uncorrelated across both dimensions, by, in addition, restricting Θ to be diagonal. In contrast, setting $\Theta = I_N$ and leaving Φ unrestricted allows for serial correlation in the disturbance while imposing cross-sectional independence, a setup considered for instance by Breitung and Tenhofen (2011). More generally, when dealing with macroeconomic series variables are typically serially correlated and often present cross-sectional correlation even after the aggregate factors are controlled for (for instance, because of shocks to sectoral prices or to alternative measures of the money supply).

Assumption C*: Independence*

(C) $\{F_t\}_{t=1}^T$, $\{e_{it}\}_{i,t=1}^{N,T}$ and $\{L_i\}_{i=1}^N$ are three mutually independent groups, although dependence within groups is allowed.

Under Assumption C, F_t, e_{it} and L_i are mutually independent across i and t. This assumption is stronger than the one used in e.g. Stock and Watson (2002) and Bai (2003) which permit e_{it} and F_t to be weakly correlated, but analogous to the ones considered by Bai and Ng (2004) and Harding (2013).

In this context, a natural estimator of the factors and factor loadings $\tilde{F}$ and $\tilde{L}$ in the unrestricted model is given by the first r principal components on the matrix YY'/NT. As shown by Chamberlain and Rothschild (1983), the principal components estimator converges to the maximum likelihood estimator when N increases (though they did not consider sampling variation). Yet the former is usually preferred because it is easier to compute and the asymptotic distribution of the principal-component analysis based (PCA henceforth) estimated factors and factor loadings is well known since Bai (2003).

As for estimation of the restricted model, define the cross-sectional mean at time t as $\overline{y}_t = N^{-1}\sum_{i=1}^N y_{it}$. While the cross-sectional mean of the loadings and the time series mean of the factors are not separately identified, the common component c_{it} is. Hence, we could obtain an estimator of

[5] A milder version of Assumption B3 is $eig_N(\Theta\Theta') \geq 0$ and $eig_T(\Phi'\Phi) \geq 0$ with $N^{-1}\sum_{i=1}^N eig_i(\Theta'\Theta) \to \mu_\Theta > 0$ and $T^{-1}\sum_{t=1}^T eig_t(\Phi'\Phi) \to \mu_\Phi > 0$. In that case, the fact that the means of the empirical spectral density of both matrices converge to a positive constant together with Assumption B2, ensures the asymptotic negligibility of any fixed-N, T array of errors.

the common component for the restricted model by: i) removing cross-sectional means from the data:

$$y_{it} - \overline{y}_t = (\gamma_i - \overline{\gamma})' g_t + (e_{it} - \overline{e}_t),$$

where $\overline{e}_t = \overline{y}_t - f_t - \overline{\gamma}' g_t$ and, ii) performing PCA on the demeaned data and extract the remaining $r-1$ principal components. Notice that the second step yields an estimator of $(\gamma_i - \overline{\gamma})' g_t$ and not of $\gamma_i' g_t$.[6] Hence, the estimator of the common component is then given by $\hat{c}_{it} = \overline{y}_t + (\widehat{\gamma_i - \overline{\gamma}})' \hat{g}_t$, and the estimation error is obtained as

$$\begin{aligned}
\hat{c}_{it} - c_{it} &= \overline{y}_t + (\widehat{\gamma_i - \overline{\gamma}})' \hat{g}_t - (f_t + \gamma_i' g_t) \\
&= \overline{e}_t + (\widehat{\gamma_i - \overline{\gamma}})' \hat{g}_t - (\gamma_i - \overline{\gamma})' g_t.
\end{aligned}$$

Having described the estimation procedure used for both the unrestricted and restricted models, we proceed to describe our test statistic, which is based on the sums of square residuals, and derive its large sample properties in the following Section.

3 Asymptotic results

From the unrestricted and restricted estimates we can compute the average sum of squared residuals

$$\tilde{\sigma}^2_{NT} = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \tilde{e}^2_{it} \quad \text{and} \quad \hat{\sigma}^2_{NT} = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \hat{e}^2_{it}$$

where $\tilde{e}_{it} = y_{it} - \tilde{c}_{it}$ and $\hat{e}_{it} = y_{it} - \hat{c}_{it}$ to form

$$Q_{NT} = NT \left(\frac{\hat{\sigma}^2_{NT} - \tilde{\sigma}^2_{NT}}{\hat{\sigma}^2_{NT}} \right). \tag{5}$$

The choice of the statistic Q_{NT} in (5) is motivated by the analogy with the F-test in the classical linear model (e.g. Evans and Savin (1982)). Recall that the F-test statistic for a sample of N *iid* observations is defined as a properly rescaled difference of restricted and unrestricted residual sum of squares,

$$\frac{RRSS - URSS}{URSS} \frac{N}{\nu},$$

where ν is the number of linear restrictions to test. When the sample size goes to infinity, F is asymptotically distributed as a χ^2 with ν degrees of freedom if the number of hypotheses stays fixed. When the number of restrictions grows with sample size, Anatolyev (2012) shows that classical tests, after

[6] For simplicity, we can assume that the cross-sectional means of loadings associated to all factors but f_t are zero. This normalization is innocuous for our purposes as it leaves the common component unchanged and it is irrelevant for our testing problem.

appropriate recentering and rescaling, still have a limiting distribution that, under some additional conditions, is a standard normal. In the same spirit, our goal is to derive asymptotic results for a recentered and rescaled version of the Q_{NT} statistic.

While the denominator of Q_{NT} in (5), $\hat{\sigma}^2_{NT}$, will converge in probability to some average of the squared residuals, say

$$\sigma^2_\infty = \underset{N,T\to\infty}{p\lim} \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} e^2_{it} = \lim_{N,T\to\infty}\frac{1}{NT} tr(\Phi'\Phi)tr(\Theta\Theta')$$

the numerator can be decomposed in three terms as follows

$$NT\left(\hat{\sigma}^2_{NT} - \tilde{\sigma}^2_{NT}\right) = \sum_{i=1}^{N}\sum_{t=1}^{T}\left(\hat{e}^2_{it} - \tilde{e}^2_{it}\right) = \mathcal{A}_{NT} + \mathcal{B}_{NT} - \mathcal{C}_{NT}$$

where

$$\mathcal{A}_{NT} = 2\sum_{i=1}^{N}\sum_{t=1}^{T} e_{it}\left(\tilde{c}_{it} - \hat{c}_{it}\right), \tag{6}$$

$$\mathcal{B}_{NT} = \sum_{i=1}^{N}\sum_{t=1}^{T}\left(\tilde{c}_{it} - \hat{c}_{it}\right)^2, \tag{7}$$

and

$$\mathcal{C}_{NT} = 2\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\tilde{c}_{it} - \hat{c}_{it}\right)\left(\tilde{c}_{it} - c_{it}\right). \tag{8}$$

Before moving to asymptotic results, we state a useful intermediate result proven in the Appendix:

Lemma 1. *(a) Suppose $H_0 : \lambda_i = \bar{\lambda}$ for all $i = 1, ..., N$. Then, under assumptions A, B and C we have that*

$$\tilde{c}_{it} - \hat{c}_{it} = \frac{1}{T}F'_t\mathbf{P}_T\sum_{s=1}^{T}F_s e_{is} + r_{it} \textit{ where } r_{it} = O_p\left(\frac{1}{\delta^2_{NT}}\right)$$

where $\delta_{NT} = \min(\sqrt{N}, \sqrt{T})$ and

$$\mathbf{P}_T = \left(\frac{\mathbf{F}'\mathbf{F}}{T}\right)^{-1} - \begin{pmatrix} 0 & \mathbf{0}' \\ \mathbf{0} & (\mathbf{G}'\mathbf{G}/T)^{-1}\end{pmatrix}.$$

(b) Moreover, if $\sqrt{T}/N \to 0$ and $\sqrt{N}/T \to 0$, then

$$i)\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\sum_{t=1}^{T} e_{it}r_{it} = o_p(1) \quad \textit{and} \quad ii)\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\tilde{c}_{it} - c_{it}\right)r_{it} = o_p(1).$$

This simple asymptotic representation of the difference of the unrestricted and restricted estimated common components is the building block of the asymptotic derivations detailed in the Appendix. Intuitively, the contribution of the $r-1$ factors with unrestricted loadings to the common component cancels out. As for the part of the common component appearing in H_0, we show that $\tilde{f}_t - \hat{f}_t = O_p(\delta^{-2}_{NT})$ while $\tilde{\lambda}_i - \hat{\lambda}_i = \tilde{\lambda}_i - \lambda_i$ contains the term $T^{-1}F'_t\mathbf{P}_T\sum_{s=1}^{T}F_s e_{is}$, which contributes to the asymptotic distribution of the test statistic plus a $O_p(\delta_{NT}^{-2})$ remainder. Part (b) of Lemma 1 ensures that terms involving the remainder r_{it} can be omitted when deriving the asymptotic distribution of the rescaled and recentered Q_{NT} statistic. In particular, $(b.i)$ ensures that terms that involve r_{it} and arise in the expression for $\mathcal{A}_{NT}$ are negligible, whereas $(b.ii)$ guarantees the same happens in term $\mathcal{C}_{NT}$. For $\mathcal{B}_{NT}$, instead, no additional results are needed.

To facilitate the exposition, we begin in Section 3.1 by assuming that the disturbances e_{it} are *iid*. This assumption is unrealistic though, but it allows us to draw analogies with the classical, fixed ν case that helps us to build intuition. Then, we relax this assumption and allow for some degree of cross-sectional dependence in Section 3.2. Finally, we consider the case with both times-series and cross-sectional dependence in Section 3.3.

3.1 The *iid* case

Finding the asymptotic distribution amounts to verify that the conditions for a central limit theorem for Q_{NT} are met. The proof relies on assuring those conditions hold for parts $\mathcal{A}_{NT}$, $\mathcal{B}_{NT}$ and $\mathcal{C}_{NT}$, and on obtaining expressions for the asymptotic mean and variance.

Theorem 1. *(iid errors) Suppose assumptions A, B and C with $\Theta = I_N$ and $\Phi = I_T$ hold. If, in addition $N/T^2 \to 0$ and $T/N^2 \to 0$, then*

$$Q^1_{NT} = \frac{Q_{NT} - N}{\sqrt{2N}} \xrightarrow{d} N(0,1). \tag{9}$$

The classic linear regression model results suggest that a test statistic like Q_{NT} should be asymptotically χ^2 distributed with degrees of freedoms equal to the number of restrictions. Since our null involves N hypotheses the χ^2_N diverges as $N \to \infty$, convergence to a standard normal can be obtained after the corresponding standardization. The result in Theorem 1, hence, extends the classical result for testing in the linear regression model to the case of a factor model with large N and T. Remarkably, the asymptotic distribution of the Q_{NT} statistic does not depend on the factors or other unobserved quantities, so practical implementation is straightforward, and only requires to estimate the restricted and unrestricted versions of the model and compute the corresponding sums of squared residuals.

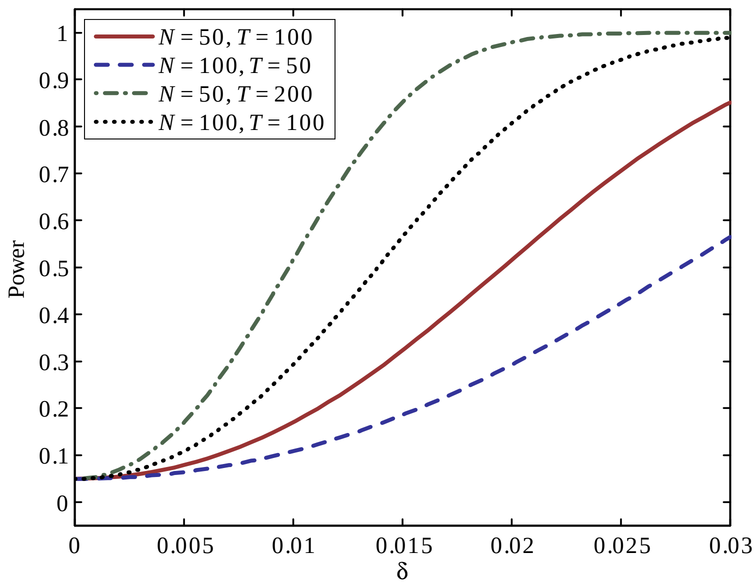
Although we investigate the finite sample properties of the different versions of the tests in detail in Section 4, it is convenient to exploit the simplicity of the expressions in the *iid* case to study its asymptotic power properties. As is well known, the one-sided nature of the test implies that $(Q^1_{NT})^2$ will be asymptotically distributed as a 50:50 mixture of 0 and a χ^2 with one degree of freedom under the null, and as a non-central χ^2 with one degree of freedom and non-centrality parameter

$$\frac{1}{N\sqrt{T}}\sum_{i=1}^{N}\sum_{t=1}^{T}\delta_i^2 f_t^2 \tag{10}$$

under the Pitman sequence of local alternatives H_l : $\boldsymbol{\lambda} = \bar{\lambda} + \boldsymbol{\delta}/\sqrt{NT}$ (see Newey and McFadden (1994)). In this respect, it should be noted that term $\mathcal{B}_{NT}$ –which captures the square of the discrepancy between $\tilde{c}_{it}$ and $\hat{c}_{it}$– is the only one that delivers power to the test. In contrast, Q^1_{NT} will diverge to infinity for fixed alternatives of the form H_f : $\boldsymbol{\lambda} = \boldsymbol{\delta}$, which makes it a consistent test.

We can assess the power implications of different combinations of sample sizes by computing the probability of rejecting the null hypothesis when it is false as a function of δ_i under the assumption that the asymptotic non-central chi-square distribution with one degree of freedom of the square of the test statistic Q^1_{NT} provides reliable rejection probabilities in finite samples. The results at the usual 5% level are plotted in Figure 1 under the fairly innocuous assumptions that $\delta_i = \delta > 0$ for 20% of the loadings and $\delta_i = 0$ for the rest. Not surprisingly, the power of all configurations increases as we depart from the null. More importantly, and in line with (10), the power of the test increases more with the times-series dimension rather than the cross-sectional dimension.

Figure 1: Power of the Q^1_{NT} test



Notes: Results at the 5% level. The non-centrality parameter corresponds to the alternative hypothesis $\lambda_i = \bar{\lambda} + \delta_i$ with $\delta_i = \delta > 0$ for 20% of the loadings and $\delta_i = 0$ for the rest.

3.2 The cross-sectional correlation case

A factor model with independent idiosyncratic errors as the one considered in the previous subsection is a "strict factor model", and use of these models is not new. The new generation of large-dimensional "approximate" factor models differs from the classical ones in at least two important ways: (i) the idiosyncratic errors can be weakly serially and cross-sectionally correlated, and (ii) the number of observations is large in both the cross-section and the time dimensions.

Allowing the errors to be cross-sectionally correlated makes the framework suitable for a wider range of economic applications, but especially in financial applications, where the weak form of the efficient markets hypothesis implies absence of first order times-series correlation in the sense that for financial assets "past performance is not an indicator of future performance". In this respect, the following result states asymptotic normality in the case of errors with cross-sectional dependence:

Theorem 2. *(Weak cross-sectional correlation) Suppose assumptions A, B and C with $\Phi = I_T$ hold. If, in addition $N/T^2 \to 0$ and $T/N^2 \to 0$, then*

$$Q^2_{NT} = \frac{Q_{NT} - N}{\sqrt{2N^2 \times tr[(\Theta\Theta')^2]/[tr(\Theta\Theta')]^2}} \xrightarrow{d} N(0,1). \tag{11}$$

Comparing (11) with the corresponding one in Theorem 1, (9), we note that the presence of cross-sectional correlation in the error terms inflates the variance of Q_{NT} but leaves the mean unchanged. Even though the error structure allows for a largely unrestricted $N \times N$ matrix of cross-sectional covariances $\Theta\Theta'$, in practice what matters are only the traces of $(\Theta\Theta')^2$ and $\Theta\Theta'$. This scalar correction can be seen as a measure of average squared correlation of the error terms (see Pesaran and Yamagata (2012)) and is, by assumption, bounded in N. Similarly to the *iid* case in Theorem 1, factors do not enter the asymptotic distribution of Q_{NT} so that the only quantities that need to be estimated in order to compute the test statistic are $tr(\Theta\Theta')^2$ and $tr(\Theta\Theta')$. Of course, if we were in the *iid* case with $V(e_{it}) = \sigma^2_e$, the numerator of the correction term $tr[(\Theta\Theta')^2]/[tr(\Theta\Theta')]^2$ would be $\sigma^4_e N$, while the square of the trace in the denominator would be $\sigma^4_e N^2$; hence, reducing it to N^{-1}.

There is an interesting analogy between the Q^2_{NT} statistic and the literature on pooling individual test statistics. In fact, the expression (11) could be interpreted as a particular way of pooling individual test statistics performed for each i. Typically, in those situations, one would calculate individual test statistics for each i and then pool (or average) over the cross-section dimension to obtain

$$Q^{pool}_{NT} = \sum_{i=1}^{N} \frac{\hat{\sigma}^2_i - \tilde{\sigma}^2_i}{\tilde{\sigma}^2_i},$$

where, for instance, $\hat{\sigma}_i^2 = T^{-1}\sum_{t=1}^T \hat{e}_{it}^2$. Pooling individual statistics or their p-values is especially frequent in the panel unit root literature (see e.g. Im, Pesaran, and Shin (2003), Bai and Ng (2004) and Maddala and Wu (1999)). For instance, if we let p_i denote the individual p-value associated to a unit-root test for unit i, then the p-value test proposed by Maddala and Wu can be written as

$$P = -2\sum_{i=1}^{N} \ln p_i,$$

which is a χ^2 with N degrees of freedom distributed for fixed N as $T_i \to \infty$. When $N \to \infty$, Choi (2001) proposes to use the following standardized version instead

$$P_m = \frac{1}{2\sqrt{N}}\sum_{i=1}^{N} \left(-2\ln p_i - 2\right),$$

which, by the Lindeberg-Lévy CLT converges in distribution to a standard normal as $T_i \to \infty$ followed by $N \to \infty$, in a sequential manner.

The pooling we implicitly use when constructing Q_{NT}^2 is different from the ones described above because it is done over i and t at the same time (and not over t and then over i) and, in addition, it does not assume independence over i. In the case in which N and T are comparable, Pesaran (2007) proposes to augment the standard Dickey-Fuller regressions with the cross-sectional averages of lagged levels and first-differences of the individual series. Like us and differently from the previous literature, he considers joint limit results for $N, T \to \infty$ for $N/T \to c$, but the proposed pooling statistic is not asymptotically normal and critical values have to be tabulated. In the same spirit, our test statistic in Theorem 2 also accounts for cross-sectional correlation other than that induced by common factors, in the spirit of the approximate factor model theory developed by Bai and Ng (2002), Stock and Watson (2002) and Bai (2003), among others, while retaining, interestingly, asymptotic normality.

3.3 The cross-section and time-series correlation case

Macroeconomic series typically show both serial and cross-sectional correlation, even after controlling for common factors. In our context, allowing for the two types of correlation amounts to leave the matrices Φ and Θ unrestricted apart from fulfilling the requirements in Assumption B. The asymptotic distribution of the corresponding test statistic is given in the following result:

Theorem 3. *(Weak cross-sectional and serial correlation) Suppose assumptions A, B and C hold. If, in addition $N/T^2 \to 0$ and $T/N^2 \to 0$, then*

$$Q_{NT}^3 = \frac{Q_{NT} - \mu_{Q^3}}{\sqrt{\sigma_{Q^3}^2}} \xrightarrow{d} N(0,1),$$

where

$$\mu_{Q^3} = N \times \frac{E\left[tr\left(\mathbf{P}_T\mathbf{F}'\Phi'\Phi\mathbf{F}\right)\right]}{tr(\Phi'\Phi)}$$

and

$$\sigma^2_{Q^3} = 2N \times \frac{N \times tr\left[(\Theta\Theta')^2\right]}{tr\left[(\Theta\Theta')\right]^2} \times \frac{E[tr\left(\mathbf{P}_T\mathbf{F}'\Phi'\Phi\mathbf{F}\right)^2]}{\left[tr(\Phi'\Phi)\right]^2},$$

where $\mathbf{P}_T$ *is defined in Lemma 1.*

Theorem 3 can be applied to testing problems previously encountered in the literature. For instance, Reis and Watson (2010)'s hypothesis on the presence of a pure inflation factor in US sectoral price indices can be rigorously tested jointly by estimating the restricted and unrestricted models and using the residuals to compute Q^3_{NT} as we show in Section 5. This procedure accounts for possible correlation in the error terms induced, for example, by shocks that hit some sectors of the economy at the same time but are not pervasive enough to be captured by a common factor.

In contrast to the results in the previous subsections, now both the asymptotic mean and variance depend on the factors, as well as on the matrices describing the dependence structure of the error components. The orders of magnitude of μ_{Q^3} and $\sigma^2_{Q^3}$ are, however, still of order N because of the assumption of weak correlation in both dimensions. In this context, the correction terms inflates both the mean and variance, essentially capturing the average deviations from the identity of both the cross-sectional and time-series variance matrices.

For what regards the mean μ_{Q^3} it is worth mentioning that, in the presence of times-series dependence in the error terms (i.e. $\Phi \neq I_T$), it remains equal to N when the factors are *iid* because $\mathbf{P}_T\mathbf{F}\mathbf{F}' = \mathbf{e}_1\mathbf{e}_1'$ where $\mathbf{e}_1 = (1, 0, ...0)'$ by the definition of $\mathbf{P}_T$. Secondly, if the first factor f_t is uncorrelated with the remaining ones i.e. those entering in g_t, then the relevant element of $\mathbf{P}_T$ is simply the reciprocal of $\mathbf{f}'\mathbf{f}$ so that $\mathbf{P}_T$ only standardizes an average of the second order cross moments of the first factor by its unconditional second moment. In the more general case of correlated factors, $\mathbf{P}_T$ has the role of performing a regression of the f_t onto g_t so that the orthogonal part is the only one that contributes to μ_{Q^3}.

Analogous considerations apply to the variance term $\sigma^2_{Q^3}$, which can be decomposed into three terms:

i) $2N$, which coincides with the variance in the *iid* case,

ii) $N \times tr\left[(\Theta\Theta')^2\right]/tr\left[(\Theta\Theta')\right]^2$, that captures the variance inflation due to cross-sectional dependence in the errors, and

iii) $E\left[tr(\mathbf{P}_T\mathbf{F}'\Phi'\Phi\mathbf{F})^2\right] / [tr(\Phi'\Phi)]^2$, which performs the same type of correction but with respect to the times-series dependence in the idiosyncratic components.

Q^3_{NT} resembles a likelihood-ratio statistic in which part of the model (dependence in the error terms) is not fully specified. As is well known, the usual asymptotic chi-squared distribution for the likelihood-ratio test statistic is based on the assumptions that the data come from a correctly specified parametric model and that the parameters satisfy the null hypothesis. In general, the likelihood-ratio statistic no longer follows an asymptotic chi-squared distribution even when the null hypothesis is true but the model is misspecified, as discussed in Kent (1982). However, in our context we find that its recentered and rescaled version has an asymptotic standard normal distribution since the scaling factor that affect each of the chi-squared terms can be estimated consistently on average.

4 Implementation and finite sample performance

4.1 Implementation details

Despite it is readily implementable, as no additional quantities are required for computing the test statistic, Theorem 1, relies on the unrealistic assumption of *iid* errors. In the more empirically plausible situation in which there is cross sectional correlation among error terms, as in Theorem 2, second moments that capture the cross sectional dependence of the idiosyncratic components are involved in the additional quantity that has to be estimated. In addition, when both types of dependence are present the additional difficulty of disentangling the cross-sectional dependence from its times-series counterpart arises. For instance, if we consider the element (t,s) of $E(ee'/N)$, straightforward calculations show that $E(ee'/N)_{(t,s)} = E\left(N^{-1}\sum_{i=1}^N e_{it}e_{is}\right) = N^{-1}tr(\Theta\Theta')\sum_{u=1}^T \phi_{ut}\phi_{us}$ so that the times-series covariance matrix $E(ee'/N)$ also depends on the elements of Θ that characterise the cross sectional dimension.

Although the theoretical expressions of the previous Section involve elements of Θ and Φ which translate into second moments of the idiosyncratic terms, it turns out that what matters is the average correlation in each dimension, which is what really characterizes the dependence in e. Considering again the times-series correlation matrix, on can easily note that its diagonal elements are $N^{-1}tr(\Theta\Theta')\sum_{u=1}^T \phi_{ut}^2$, and, since the sum of eigenvalues of $\Phi\Phi'$ is equal to $\sum_{u=1}^T\sum_{t=1}^T \phi_{ut}^2$, a convenient way of normalizing this covariance matrix consists in dividing each column and row by its corresponding average variance, say by defining

$$e^\circ_{it} = \frac{e_{it}}{\sqrt{N^{-1}tr(\Theta\Theta') \times \sum_{u=1}^T \phi_{ut}^2}},$$

in order to obtain a correlation matrix characterizing the times-series dependence by computing

$$E[e^{\circ}e^{\circ\prime}/N]_{(t,s)} = E\left(\frac{1}{N}\sum_{i=1}^{N} e^{\circ}_{it}e^{\circ}_{is}\right) = \frac{\sum_{u=1}^{T}\phi_{ut}\phi_{us}}{\sqrt{\sum_{u=1}^{T}\phi_{ut}^2}\sqrt{\sum_{u=1}^{T}\phi_{us}^2}} = \rho_{t,s},$$

which is the (t,s)'th element of the times-series correlation matrix Υ. Analogously, for the cross-sectional correlation matrix we can again normalize each column and row accordingly, by defining instead

$$e^{\diamond}_{it} = \frac{e_{it}}{\sqrt{T^{-1}tr(\Phi\Phi') \times \sum_{h=1}^{N}\theta_{ih}^2}},$$

so to obtain a correlation matrix Ψ characterizing the cross-sectional dependence with typical element (i,j) given by

$$E[e^{\diamond\prime}e^{\diamond}/T]_{(i,j)} = E\left(\frac{1}{T}\sum_{t=1}^{T} e^{\diamond}_{it}e^{\diamond}_{jt}\right) = \frac{\sum_{h=1}^{N}\theta_{ih}\theta_{jh}}{\sqrt{\sum_{h=1}^{N}\theta_{ih}^2}\sqrt{\sum_{h=1}^{N}\theta_{jh}^2}} = \varrho_{i,j}.$$

It is then clear that

$$N \times \frac{tr\left[(\Theta\Theta')^2\right]}{\left[tr(\Theta\Theta')\right]^2} = \frac{tr(\Psi^2)}{N} \quad \text{and} \quad \frac{E[tr\left(\mathbf{P}_T\mathbf{F}'\Phi'\Phi\mathbf{F}\right)^2]}{\left[tr(\Phi'\Phi)\right]^2} = \frac{E\left[tr(\mathbf{P}_T\mathbf{F}'\Upsilon\mathbf{F})^2\right]}{T^2},$$

where we have used the fact that $tr(\Psi) = N$ and $tr(\Upsilon) = T$ since, by construction, they are correlation matrices. Notice that in the *iid* case, $tr(\Psi^2)/N = E\left[tr(\mathbf{P}_T\mathbf{F}'\Upsilon\mathbf{F})^2\right]/T^2 = 1$. From now on, we can therefore focus on estimating the quantities $tr(\Psi^2)$ and $E\left[tr(\mathbf{P}_T\mathbf{F}'\Upsilon\mathbf{F})^2\right]$.

A natural way to estimate $tr(\Psi^2)$ would be to construct the sample correlation matrix using the estimated residuals and calculate the corresponding trace. Harding (2013) shows that the large N,T limit of those traces, however, differs from the quantities we are interested in.[7] Specifically, using Harding's results and the estimated residuals from the restricted model, $\hat{e}_{it} = y_{it} - \hat{c}_{it}$, we can construct a feasible estimator of $tr(\Psi^2)$ as

$$\frac{1}{N}\widehat{tr(\Psi^2)} = \frac{1}{N}tr[(\hat{\Psi})^2] - \frac{N}{T} \text{ where } \hat{\Psi} = \hat{D}^{-1/2}\widehat{\Theta\Theta'}\hat{D}^{-1/2}$$

with $\widehat{\Theta\Theta'} = \hat{e}'\hat{e}/T$ and $\hat{D} = diag(\widehat{\Theta\Theta'})$. With this estimator at hand, we have all the required elements to construct the Q^2_{NT} statistic.

Unfortunately, for Q^3_{NT} we cannot use Harding (2013)'s results for the estimation of $E\left[tr(\mathbf{P}_T\mathbf{F}'\Upsilon\mathbf{F})^2\right]$ which, in addition to estimation of Υ, requires estimates for $\mathbf{P}_T$ and $\mathbf{F}$. For the last two objects, it

[7] Harding (2013) shows that the moments of the eigenvalue distribution of a covariance matrix of data in which both dimensions N and T are large can be derived analytically from the model assumptions (analogous to ours) using properties of noncommutative random variables. He also provides a recursive expression for the computation of the spectral moments of Θ.

is natural to use their sample counterparts from the restricted model. Estimating that trace is more challenging since: (i) it involves an expectation with respect to the factor distribution; and (ii) estimation errors in the matrix Υ could get amplified when computing the quadratic form $\mathbf{P}_T\mathbf{F}'\Upsilon\mathbf{F}$. As for (i), there is no alternative other than using the estimated factor from the actual sample of size T. As regards to (ii), it is well known that the sample correlation matrix has a number of undesirable properties when the dimensions of the matrix are large. Although there have been recent developments on estimation of large covariance matrices based on shrinkage methods, random matrix theory or adaptive thresholding techniques among others (see e.g. Bai and Shi (2011) for a recent survey), in this section and the next one we follow a parametric approximation in which we capture the times series dependence in the idiosyncratic components through an homogeneous $MA(q)$ process with $q = 5$.

Specifically, we estimate $\Upsilon = \Upsilon(\boldsymbol{\upsilon})$ with $\boldsymbol{\upsilon} = (\upsilon_1, ..., \upsilon_5)'$ by minimizing the distance between the sample times-series correlation matrix and

$$\Upsilon(\boldsymbol{\upsilon}) = \begin{pmatrix} 1 & \upsilon_1 & \cdots & \upsilon_5 & 0 & \cdots & 0 \\ \upsilon_1 & 1 & & & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & & & \ddots & 0 \\ \upsilon_5 & & & & & & \upsilon_5 \\ 0 & \ddots & & & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & & \ddots & 1 & \upsilon_1 \\ 0 & \cdots & 0 & \upsilon_5 & \cdots & \upsilon_1 & 1 \end{pmatrix}.$$

We have also tried the following alternative approaches: 1) using the adaptive thresholding technique for sparse covariance matrices proposed by Fan, Liao, and Mincheva (2011); and 2) the "reduced" correlation matrix discussed in Bai (2010). Through Monte Carlo simulations, our parametric approximation outperforms the others when the object of interest is $\mathbf{P}_T\mathbf{F}'\Upsilon\mathbf{F}$, even when the true DGP presents heterogeneous autocorrelation patterns (e.g. $AR(1)$ processes for half of the sample and $MA(1)$ processes for the remaining part). Naturally, one could also use other estimators for Υ which, depending on the application, may be more suitable.

4.2 Monte Carlo design

In this section we assess the finite sample size and power properties of the Q^1_{NT}, Q^2_{NT} and Q^3_{NT} test statistics under both the null hypothesis $H_0 : \lambda_{1i} = 1$ for all $i = 1, ..., N$, and under the alternative of $\lambda_{1,i} = 1 + u_i,\ u_i \sim iidN(0, .05)$, respectively. The loadings $\gamma_{1j}, ..., \gamma_{r-1j}$ for $j = 1, ..., N$ of the remaining factors are $N(0, 1)$ variates.

In Figure 2, we carry out 10,000 replications with sample sizes $T = 190$ and $N = 187$, as in our empirical application. We collect in tables results for other sample size configurations to assess the general properties of the proposed tests; specifically, we try different combinations of N and T, again based on 10,000 replications. In all cases, the model is generated as a three-factor model under the null of constant loading on the first factor. Results with a different number of factors are similar and hence are not reported.

The framework used in the preceding section was quite rich, allowing for distributed lags of potentially serially correlated factors and error terms that were conditionally heteroscedastic and serially and cross-correlated. The design we use here seeks to incorporate all these features, and the idiosyncratic components are generated accordingly as follows

$$(1 - aL)e_{it} = (1 + b^2)v_{it} + bv_{i+1,t} + bv_{i-1,t},$$

with $v_{it} \sim iidN(0, 1)$, delivering the iid case when $a = b = 0$ as a particular case. In order to introduce some cross-sectional correlation but no times-series dependence in the error terms, we simulate the errors as spatial $MA(1)$ with coefficient $b = 0.5$ and $a = 0$. Finally, we allow for both times-series and cross-sectional correlation by generating disturbances cross-sectionally correlated with $SMA(1)$ coefficient $b = 0.5$ and serially correlated with $AR(1)$ coefficient $a = 0.5$.

As for the factors, we consider three different DGPs: $DGP1$ corresponds to the standard situation in which factors are normally distributed independent of each other and across time. In $DGP2$ and $DGP3$ we introduce cross-sectional and/or times-series dependence in the factor structure by generating $\mathbf{F}_t$ according to $\mathbf{F}_t = 0.5\mathbf{F}_{t-1} + \mathbf{u}_t$ where $V(\mathbf{u}_t) = I_3(1 - \rho_c) + \rho_c\ell_3\ell_3'$ where I_r denotes the identity matrix of size r while ℓ_r is a vector of r ones. In $DGP2$ we still keep factors independent of each other i.e. $\rho_c = 0$ but we allow them to be serially correlated. Finally, in $DGP3$ factors are not only serially correlated with autocorrelation 0.5 but also correlated with each other through the innovations $\mathbf{u}_t$. The covariance matrix $V(\mathbf{u}_t)$ is calibrated such that $V(\mathbf{F}_t) = 0.75I_3 + 0.25\ell_3\ell_3'$.

As regards estimation we use PCA for the unrestricted model and the procedure described in Section 2 for the restricted model. To adjust for small sample bias, which may be substantial given the number of hypotheses, we rescale the Q_{NT} statistic (see Evans and Savin (1982)) by the factor

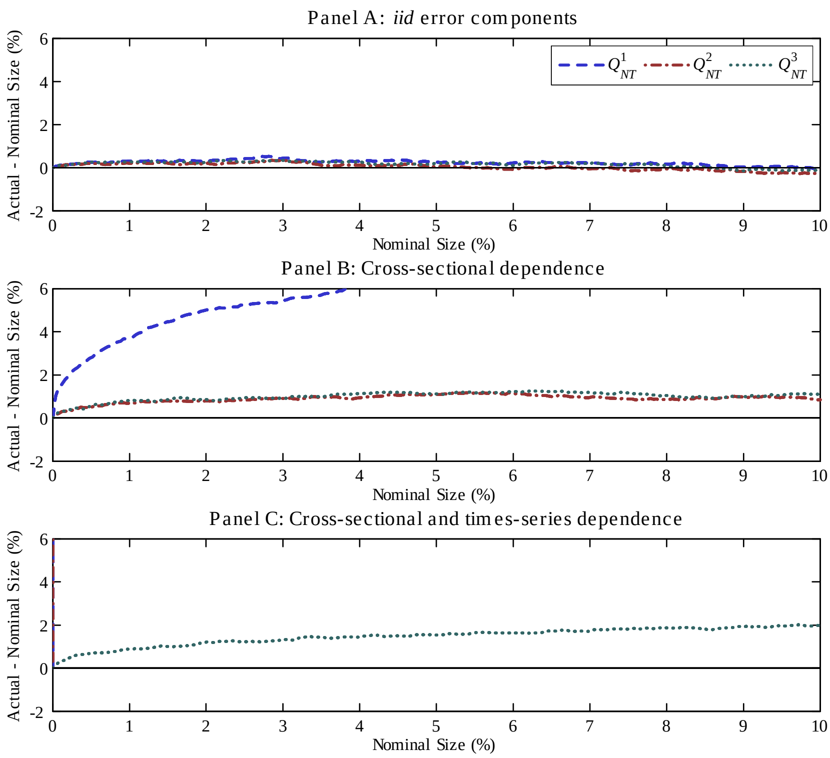
$$\frac{NT - r(N + T) + \nu/2}{NT}$$

where ν is the number of hypotheses ($\nu = N$ in Example 1). Finally, computation details of the test statistics in Theorems 2 and 3 are provided in Section 4.1.

4.3 Size and power properties

The first question that we need to address is whether the asymptotic distribution under the null attributed to the test statistics introduced in Section 3 is reliable in finite samples. To do so, we use the p-value discrepancy plots proposed by Davidson and MacKinnon (1998). Let w_j denote the simulated value of a given test statistic, and let p_j be the asymptotic p-value of w_j, that is the probability of observing a value of the test statistic at least as large as w_j according to its asymptotic distribution under the null. Let also $\hat{F}(x)$ for $x \in (0,1)$ be the empirical distribution function of p_j i.e. the sample proportion of $p_j's$ which are not greater than x. A p-value discrepancy plot is a plot of $\hat{F}(x) - x$ against x, i.e. a plot of the difference between actual and nominal size for a range of nominal sizes. If the candidate distribution for w_j is correct, then the p-value discrepancy should be close to zero.

Figure 2: p-value discrepancy plots of the Q^1_{NT}, Q^2_{NT}, and Q^3_{NT} tests



Notes: Results based on 10,000 replications on a factor model with three factors with $\mathbf{F}_t = 0.5\mathbf{F}_{t-1} + \mathbf{u}_t$, with $V(\mathbf{u}_t)$ s.t. $V(\mathbf{F}_t) = 0.75I_3 + 0.25\ell_3\ell_3'$ where I_r denotes the identity matrix of size r while ℓ_r is a vector of r ones. The model is simulated under the null hypothesis $H_0 : \lambda_{1i} = 1$ for all $i = 1, ..., N$ and the remaining loadings γ_{ij}'s, $j = 1, ..., N$, $i = 2, ..., r$ are $N(0,1)$ variates. In Panel A, error terms are *iid* in both dimensions, whereas in Panel B they are cross-sectionally correlated with $SMA(1)$ coefficient 0.33 for the first $N/2$ series and 0.66 for the rest, and in Panel C errors are also serially correlated with $AR(1)$ coefficient 0.33 and 0.66 for the first and second half of time periods, respectively. A description of estimation procedures as well as computation details of the test statistics are provided in Sections 2 and 4.1, respectively.

The top panel of Figure 2 shows p-value discrepancy plots of the three tests of $H_0 : \lambda_{1i} = \bar{\lambda}$ for all $i = 1, ..., N$, in samples in which the idiosyncratic components are *iid* while factors are generated according to $DGP3$. The most striking fact that we find is that the test statistics Q^2_{NT} and Q^3_{NT}, which involve additional calculations that account for potential cross-sectional and/or times-series dependence, have good size properties that are comparable to Q^1_{NT}. In the middle panel, error terms are cross-sectionally correlated with SMA(1) parameter of 0.5. Not surprisingly, Q^1_{NT} becomes oversized as the variance of Q_{NT} becomes inflated due to the dependence structure in the idiosyncratic terms. As for Q^2_{NT} and Q^3_{NT}, they remain slightly oversized but with overall very satisfactory properties. Finally, in the presense of both cross-sectional and times-series dependence, only Q^3_{NT} provides reliable size properties as can be seen in the bottom panel of Figure 2.

Table 1: Rejection rates under the null at 1%, 5%, and 10% significance levels: *iid* errors

		1%			5%			10%		
N	T	Q^1_{NT}	Q^2_{NT}	Q^3_{NT}	Q^1_{NT}	Q^2_{NT}	Q^3_{NT}	Q^1_{NT}	Q^2_{NT}	Q^3_{NT}
Panel A: $DGP1$, $\mathbf{F}_t \sim iidN(0, I_3)$										
100	100	1.3	1.1	0.7	5.3	4.9	3.3	10.0	9.4	6.5
400	100	1.8	1.4	0.4	7.0	6.0	2.5	13.2	11.6	5.3
100	400	1.1	1.1	0.9	4.5	4.2	3.8	8.3	8.0	7.5
400	400	1.3	1.3	1.1	5.3	5.1	4.2	9.8	9.7	8.1
Panel B: $DGP2$, $\mathbf{F}_t = 0.5\mathbf{F}_{t-1} + \mathbf{u}_t$, $u_{it} \perp u_{js}, i \neq j$ for all t, s s.t. $V(\mathbf{F}_t) = I_3$										
100	100	1.3	1.2	1.3	5.3	4.9	4.9	9.8	9.3	9.3
400	100	1.8	1.3	1.3	6.3	5.3	5.2	12.3	10.8	10.7
100	400	1.4	1.3	1.3	4.6	4.4	4.5	8.6	8.3	8.3
400	400	1.3	1.2	1.2	5.8	5.7	5.8	10.4	10.3	10.6
Panel C: $DGP3$, $\mathbf{F}_t = 0.5\mathbf{F}_{t-1} + \mathbf{u}_t$, with $V(\mathbf{u}_t)$ s.t. $V(\mathbf{F}_t) = 0.75I_3 + 0.25\ell_3\ell_3'$										
100	100	1.5	1.3	1.4	5.2	4.8	4.9	9.7	9.2	9.2
400	100	1.7	1.3	1.5	6.5	5.5	5.6	12.2	10.9	10.6
100	400	1.1	1.0	1.1	4.7	4.4	4.5	8.7	8.5	8.7
400	400	1.3	1.2	1.3	5.2	5.2	5.4	9.7	9.6	9.7
Panel D: $DGP3$ with $V(e_{it}) = r$										
100	100	2.1	1.9	1.1	7.0	6.4	4.3	12.4	11.7	8.3
400	100	1.9	1.4	0.5	6.7	5.6	2.4	12.2	10.7	5.1
100	400	1.7	1.6	1.4	6.3	6.1	5.5	10.9	10.6	9.9
400	400	1.2	1.1	0.8	6.0	5.9	4.7	10.6	10.5	9.1

Notes: Results based on 10,000 replications on factor models with three factors simulated under the null hypothesis $H_0 : \lambda_{1i} = 1$ for all $i = 1, ..., N$ and the remaining loadings γ_{ij}'s, $j = 1, ..., N$, $i = 2, ..., r$ are $N(0, 1)$ variates. A description of estimation procedures as well as computation details of the test statistics are provided in Sections 2 and 4.1, respectively.

We complement the p-value discrepancy plots of Figure 2 with a more detailed Monte Carlo exercise in which we focus on the rejection rates under the null of the test statistics at the 1%, 5% and 10% nominal level. Table 1, 2 and 3 present those figures for the cases in which errors are *iid*, cross-sectionally correlated and with both cross-sectional and times-series dependence, respectively.

For each Table, in Panel A we consider the simple situation of *iid* factors. In Panel B factors follow uncorrelated $AR(1)$ processes with autocorrelation parameter of 0.5, whereas in Panel C a $VAR(1)$ with the same autocorrelation structure and contemporaneous correlation of 0.25.

Not surprisingly, Q^1_{NT} outperforms the others, especially in the case $N = T = 100$ i.e. when the sample is relatively small. This is probably due to the fact that the mean and variance figures needed to standardize Q^1_{NT} are evaluated at their population values N and $2N$, while Q^2_{NT} uses an estimate of the variance and Q^3_{NT} is based on estimates of both. A second noteworthy feature is that Q^3_{NT} is slightly undersized when errors are *iid.* Finally, we should note that the three tests are oversized when N is larger than T irrespective of the DGP assumed for the factors, as can be seen from the second rows in the four panels. Finally, in Panel D we reduce the signal-to-noise ratio of the specification in Panel A to 50% and find the same qualitative results.

Table 2: Rejection rates under the null at 1%, 5%, and 10% significance levels: Cross-sectionally correlated but serially uncorrelated errors

		1%			5%			10%		
N	T	Q^1_{NT}	Q^2_{NT}	Q^3_{NT}	Q^1_{NT}	Q^2_{NT}	Q^3_{NT}	Q^1_{NT}	Q^2_{NT}	Q^3_{NT}
Panel A: $DGP1$, $\mathbf{F}_t \sim iidN(0, I_3)$										
100	100	5.5	1.8	1.2	12.0	6.2	4.7	17.5	11.1	8.5
400	100	5.9	1.8	0.9	13.3	6.5	3.6	19.5	11.8	6.9
100	400	4.7	1.7	1.6	10.6	5.6	5.2	15.3	9.8	9.3
400	400	4.9	1.3	1.2	11.4	5.9	5.1	16.7	10.6	9.3
Panel B: $DGP2$, $\mathbf{F}_t = 0.5\mathbf{F}_{t-1} + \mathbf{u}_t$, $u_{it} \perp u_{js}, i \neq j$ for all t, s s.t. $V(\mathbf{F}_t) = I_3$										
100	100	5.7	1.9	1.9	11.8	6.6	6.5	16.9	11.0	10.8
400	100	5.6	1.4	1.5	13.1	6.3	6.3	19.1	11.5	11.2
100	400	4.9	1.6	1.7	11.2	5.8	6.0	16.2	10.2	10.3
400	400	5.1	1.4	1.5	11.1	6.1	6.2	16.8	10.3	10.6
Panel C: $DGP3$, $\mathbf{F}_t = 0.5\mathbf{F}_{t-1} + \mathbf{u}_t$, with $V(\mathbf{u}_t)$ s.t. $V(\mathbf{F}_t) = 0.75I_3 + 0.25\ell_3\ell_3'$										
100	100	5.4	2.1	2.1	11.9	6.2	6.3	16.9	11.0	11.0
400	100	5.6	1.5	1.6	13.3	6.3	6.4	19.4	11.6	11.6
100	400	5.0	1.6	1.6	11.4	5.7	5.9	16.4	10.6	10.7
400	400	5.1	1.5	1.6	11.6	6.2	6.3	17.2	10.9	11.1
Panel B: $DGP3$, with $e_{it} = (1 + b_i^2)v_{it} + b_i v_{i+1,t} + b_i v_{i-1,t}$										
100	100	5.0	1.8	2.0	11.2	6.2	6.3	16.7	10.6	10.8
400	100	5.4	1.6	1.8	13.0	6.2	6.3	19.5	11.9	11.9
100	400	4.9	1.7	1.8	10.6	6.2	6.2	15.9	10.3	10.3
400	400	4.4	1.3	1.4	10.9	5.6	5.9	16.6	10.5	10.7

Notes: Results based on 10,000 replications. The remaining loadings γ_{ij}'s, $j = 1, ..., N$, $i = 2, ..., r$ are $N(0, 1)$ variates. The $SMA(1)$ coefficient is 0.5 in Panels A, B, and C. In Panel D, error are cross-sectionally correlated with $SMA(1)$ coefficient 0.33 for the first $N/2$ series and 0.66 for the rest. A description of estimation procedures as well as computation details of the test statistics are provided in Sections 2 and 4.1, respectively.

Table 3: Rejection rates under the null at 1%, 5%, and 10% significance levels: Error terms with both times-series and cross-sectional correlation

N	T	1% Q^1_{NT}	1% Q^2_{NT}	1% Q^3_{NT}	5% Q^1_{NT}	5% Q^2_{NT}	5% Q^3_{NT}	10% Q^1_{NT}	10% Q^2_{NT}	10% Q^3_{NT}
Panel A: $DGP1$, $\mathbf{F}_t \sim iidN(0, I_3)$										
100	100	10.1	2.8	1.0	17.4	8.2	3.9	23.1	13.6	8.2
400	100	16.8	2.8	0.1	25.3	7.9	1.1	30.9	13.5	3.7
100	400	6.8	2.3	1.3	13.2	7.0	5.1	18.4	11.4	9.2
400	400	8.6	2.0	0.7	16.0	6.8	3.5	22.0	11.7	7.3
Panel B: $DGP2$, $\mathbf{F}_t = 0.5\mathbf{F}_{t-1} + \mathbf{u}_t$, $u_{it} \perp u_{js}, i \neq j$ for all t, s s.t. $V(\mathbf{F}_t) = I_3$										
100	100	79.9	59.4	2.1	87.9	76.6	7.2	91.1	84.4	12.8
400	100	98.7	90.5	0.6	99.5	96.3	4.1	99.7	98.1	10.2
100	400	83.1	68.4	1.5	90.7	83.6	5.8	93.9	89.2	10.5
400	400	99.9	99.4	0.9	100.0	99.9	4.8	100.0	100.0	9.6
Panel C: $DGP3$, $\mathbf{F}_t = 0.5\mathbf{F}_{t-1} + \mathbf{u}_t$, with $V(\mathbf{u}_t)$ s.t. $V(\mathbf{F}_t) = 0.75I_3 + 0.25\ell_3\ell_3'$										
100	100	80.0	60.7	2.0	88.0	77.0	7.3	91.4	84.1	13.1
400	100	98.8	91.0	0.6	99.5	96.7	4.1	99.7	98.3	10.0
100	400	84.6	71.1	1.7	91.7	85.0	6.0	94.4	90.3	10.8
400	400	99.9	99.4	1.1	100.0	99.9	5.0	100.0	100.0	9.6
Panel D: $DGP3$, with $(1 - a_iL)e_{it} = (1 + b_i^2)v_{it} + b_iv_{i+1,t} + b_iv_{i-1,t}$										
100	100	80.4	61.0	2.5	88.2	76.9	8.2	91.4	84.1	14.4
400	100	98.2	87.2	1.7	99.1	94.9	7.4	99.4	97.3	13.9
100	400	85.4	72.7	2.0	92.5	85.9	6.6	95.0	91.3	11.6
400	400	99.9	99.2	1.5	100.0	99.8	5.6	100.0	99.9	10.9

Notes: Results based on 10,000 replications. The remaining loadings γ_{ij}'s, $j = 1, ..., N$, $i = 2, ..., r$ are $N(0, 1)$ variates. The $SMA(1)$ and $AR(1)$ coefficients are equal to 0.5 in Panels A, B, and C. In Panel D, errors are cross-sectionally correlated with $SMA(1)$ coefficient 0.33 for the first $N/2$ series and 0.66 for the rest, and also serially correlated with $AR(1)$ coefficient 0.33 and 0.66 for the first and second half of time periods, respectively. A description of estimation procedures as well as computation details of the test statistics are provided in Sections 2 and 4.1, respectively.

In Table 2 we report the corresponding figures for the case of cross-sectionally correlated errors. In line with the middle panel of Figure 2 we see that Q^1_{NT} has a large size distortion. Secondly, although Q^2_{NT} is based in the actual mean (N) and its estimator of the variance of Q_{NT} exploits the time-series independence of the error terms, we find a similar pattern in terms of good size properties for both Q^2_{NT} and Q^3_{NT}, which suggest that the effects in terms of efficiency losses of estimating those additional quantities seem to be minor compared to the robustness gains from using the more general version of Theorem 3.3. This result, which holds irrespective of the DGP assumed for the factors, is of practical interest since in applications it is advisable the use of Q^3_{NT} since it accounts for both types of dependence. Finally, in Panel D we introduce some degree of heterogeneity in the cross-sectional dependence of the errors by generating them as cross-sectionally correlated with SMA(1) coefficient $b_1 = 0.33$ for the first $N/2$ series, and $b_2 = 0.66$ for the rest. As expected, since there are no parametric restrictions on Θ in Assumption B, we see that both Q^2_{NT} and Q^3_{NT} have good size properties in that situation.

Table 4: Monte Carlo rejection rates (%) of tests at 1%, 5%, and 10% significance levels under the alternative $H_1 : \lambda_{1i} = 1 + \eta_i$, $\eta_i \sim N(0, 0.05)$

		1%			5%			10%		
N	T	Q^1_{NT}	Q^2_{NT}	Q^3_{NT}	Q^1_{NT}	Q^2_{NT}	Q^3_{NT}	Q^1_{NT}	Q^2_{NT}	Q^3_{NT}
Panel A: $DGP3$, with $e_{it} \sim iidN(0,1)$										
100	100	21.1	20.1	9.8	39.0	38.0	18.3	50.6	49.8	23.6
400	100	68.4	65.8	29.0	84.6	83.3	38.5	90.3	89.6	42.7
100	400	98.0	97.9	73.3	99.4	99.4	79.6	99.7	99.7	82.6
400	400	100.0	100.0	90.1	100.0	100.0	92.0	100.0	100.0	92.8
Panel B: $DGP3$, with $(1 - a_i L)e_{it} = (1 + b_i^2)v_{it} + b_i v_{i+1,t} + b_i v_{i-1,t}$										
100	100	91.9	78.9	4.0	95.8	89.5	10.3	97.3	93.8	15.2
400	100	99.9	97.1	6.2	99.9	99.3	14.4	100.0	99.7	21.0
100	400	100.0	99.8	32.6	100.0	100.0	47.8	100.0	100.0	55.2
400	400	100.0	100.0	60.5	100.0	100.0	67.3	100.0	100.0	70.4

Notes: Results based on 10,000 replications. $DGP3$ corresponds to a factor model with three factors with $\mathbf{F}_t = 0.5\mathbf{F}_{t-1} + \mathbf{u}_t$, with $V(\mathbf{u}_t)$ s.t. $V(\mathbf{F}_t) = 0.75I_3 + 0.25\ell_3\ell_3'$ where I_r denotes the identity matrix of size r while ℓ_r is a vector of r ones. The remaining loadings γ_{ij}'s, $j = 1, ..., N$, $i = 2, ..., r$ are $N(0,1)$ variates. In Panel A, error terms are *iid* in both dimensions, whereas in Panel they are cross-sectionally correlated with $SMA(1)$ coefficient 0.33 for the first $N/2$ series and 0.66 for the rest, and also serially correlated with $AR(1)$ coefficient 0.33 and 0.66 for the first and second half of time periods, respectively. A description of estimation procedures as well as computation details of the test statistics are provided in Sections 2 and 4.1, respectively.

Rejection rates under the null at 1%, 5%, and 10% significance levels in the case of error terms with both times-series and cross-sectional correlation are reported in Table 3. Not surprisingly, only Q^3_{NT} presents good size properties across all the configurations of the different panels. Interestingly, in Panel A, where factors are *iid*, Q^2_{NT} still performs reasonably well, although moderately oversized. This feature can be attributed to the fact that, although the variance of Q_{NT} is underestimated through Q^2_{NT} in the presence of times-series dependence in the error terms, its mean remains equal to N when the factors are *iid* (see Section 3.3).

In order to assess the power properties of the Q^1_{NT}, Q^2_{NT} and Q^3_{NT} test statistics, we simulate and estimate 10,000 samples under under the alternative of $\lambda_{1i} = 1 + u_i$, $u_i \sim iidN(0, .05)$. Table 4 reports the corresponding Monte Carlo rejection rates of the tests at 1%, 5%, and 10% significance levels. In Panel A, error terms are *iid* in both dimensions, whereas in Panel B they are cross-sectionally correlated with $SMA(1)$ coefficient 0.33 for the first $N/2$ series and 0.66 for the rest, and also serially correlated with $AR(1)$ coefficient 0.33 and 0.66 for the first and second half of time periods, respectively. In Panel A, we observe that Q^1_{NT} has more power than Q^2_{NT} which in turns is more powerful than Q^3_{NT}, which suggest that the use of the true recentering and rescaling quantities results in gains of power.

Table 5: Size-adjusted Monte Carlo rejection rates (%) of tests at 1%, 5%, and 10% under the alternative $H_1 : \lambda_{1i} = 1 + \eta_i$, $\eta_i \sim N(0, 0.05)$

		1%			5%			10%		
N	T	Q^1_{NT}	Q^2_{NT}	Q^3_{NT}	Q^1_{NT}	Q^2_{NT}	Q^3_{NT}	Q^1_{NT}	Q^2_{NT}	Q^3_{NT}
		Panel A: $DGP3$, with $e_{it} \sim iidN(0,1)$								
100	100	17.3	17.3	8.3	37.0	37.3	17.9	50.1	50.3	23.7
400	100	60.6	62.1	27.2	81.2	81.9	37.5	87.9	88.4	41.9
100	400	97.7	97.7	72.7	99.4	99.4	79.8	99.8	99.8	83.0
400	400	100.0	100.0	89.8	100.0	100.0	91.8	100.0	100.0	92.8
		Panel B: $DGP3$, with $(1 - a_i L)e_{it} = (1 + b_i^2)v_{it} + b_i v_{i+1,t} + b_i v_{i-1,t}$								
100	100	3.4	3.3	1.9	12.3	12.0	6.6	21.8	21.2	11.7
400	100	4.6	4.2	4.8	16.9	16.2	11.6	27.8	26.8	17.9
100	400	41.1	41.3	24.1	68.6	68.7	42.8	82.0	81.9	52.4
400	400	93.6	93.7	59.1	98.7	98.7	66.8	99.5	99.5	69.5

Notes: Results based on 10,000 replications. $DGP3$ corresponds to a factor model with three factors with $\mathbf{F}_t = 0.5\mathbf{F}_{t-1} + \mathbf{u}_t$, with $V(\mathbf{u}_t)$ s.t. $V(\mathbf{F}_t) = 0.75I_3 + 0.25\ell_3\ell_3'$ where I_r denotes the identity matrix of size r while ℓ_r is a vector of r ones. The remaining loadings γ_{ij}'s, $j = 1, ..., N$, $i = 2, ..., r$ are $N(0,1)$ variates. In Panel A, error terms are *iid* in both dimensions, whereas in Panel they are cross-sectionally correlated with $SMA(1)$ coefficient 0.33 for the first $N/2$ series and 0.66 for the rest, and also serially correlated with $AR(1)$ coefficient 0.33 and 0.66 for the first and second half of time periods, respectively. A description of estimation procedures as well as computation details of the test statistics are provided in Sections 2 and 4.1, respectively.

Interestingly, this is true irrespective of the DGP assumed to generate the error terms.

In view of the substantial size distortions reported in Table 3 of Q^1_{NT} and Q^2_{NT} under the null when there is times-series dependence as in Panel B, we report not only raw rejection rates based on asymptotic critical values as in Table 4 but also size-adjusted ones in Table 5, which exploit the Monte Carlo critical values obtained in simulations under the null. If we focus on this second table, we can conclude that all the tests have significant power and that the aforementioned ranking is preserved.

5 Is there a pure inflation factor in the US economy?

In studying the behavior of inflation, Reis and Watson (2010) decompose the quarterly changes in sectoral goods' prices into a pure inflation component, an aggregate relative price index, and idiosyncratic relative prices. In their factor model, the pure inflation component is modeled as a factor that affects all sectors with the same intensity, whereas the relative price factors are allowed to have different impacts on each sector. The pure inflation factor is then estimated by restricting the loadings to be equal in all sectors.

To test these restrictions, they rely on t-tests obtained from univariate regressions of each price series on the estimated pure inflation factor. The cross-sectional correlation that is likely to be present in sectoral prices may be affecting the idiosyncratic components and, hence, the individual t-statistics.

A more formal approach would be to test all hypotheses jointly, while allowing for correlation in the error terms. To this aim, we begin by estimating a three-factor model under the null of constant loadings on the first factor ($\lambda_{1i} = 1$ for all $i = 1, ..., N$) and under the alternative, as described in Section 2.

With estimates from both models in hand, we construct the Q_{NT} statistic under the assumptions of *iid* disturbances, of cross-sectionally correlated disturbances and, finally, allowing for both types of correlation. In the first panel of Table 6 we can see that the hypothesis that there exists one factor with constant loadings is strongly rejected under all the three assumptions on the errors, with values of Q^1_{NT}, Q^2_{NT} and Q^3_{NT} well above the 95% critical value of 1.64. This result is in line with the findings in Reis and Watson, in which the individual t-tests rejected the null of unit loading 30% of the times at the 5% level. It should be noted that, since the estimated values for $tr(\Psi^2)/N$ and $E\left[tr(\mathbf{P}_T\mathbf{F}'\Upsilon\mathbf{F})^2\right]/T^2$ are, respectively, 2.82 and 1.30, there seems to be significant correlation in both dimensions, suggesting that Q^3_{NT} is the most appropriate choice.

Table 6: Testing for the existence of a pure inflation factor

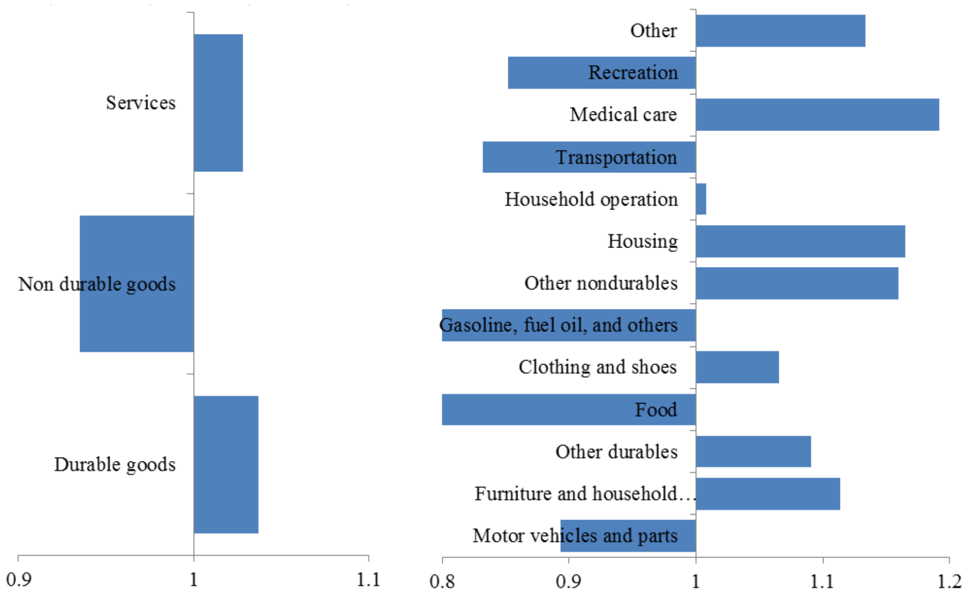
	Q^1_{NT}		Q^2_{NT}		Q^3_{NT}	
Null hypothesis (H_0)	Stat.	p-value	Stat.	p-value	Stat.	p-value
Panel A: Testing for a pure inflation factor						
$\lambda_{1i} = \lambda$	49.7	0.00	31.5	0.00	26.5	0.00
$\lambda_{1i} \in \{\lambda_1, \lambda_2, \lambda_3\}$	43.3	0.00	31.8	0.00	27.0	0.00
$\lambda_{1i} \in \{\lambda_1, ..., \lambda_{13}\}$	45.6	0.00	33.5	0.00	27.7	0.00
Panel B: Sorting loadings						
$\lambda_{1i} \in \{\lambda_1, \lambda_2, \lambda_3\}$	11.6	0.00	8.8	0.00	3.8	0.00
$\lambda_{1i} \in \{\lambda_1, \lambda_2, \lambda_3, \lambda_4\}$	3.8	0.00	2.9	0.00	-0.2	0.58
$\lambda_{1i} \in \{\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5\}$	0.0	0.50	0.0	0.50	-2.2	0.98

Notes: Data are quarterly price changes in each of 187 sectors in the US personal consumption expenditures category of the national income and product accounts from 1959:Q1 to 2006:Q2. Estimation of the restricted model is described in Section 5. Computation details of the test statistics are provided in Section 4.1.

A second exercise we perform is to relax our null of constant loadings in all sectors allowing heterogeneity between the three categories in which US sectoral prices are divided: durable goods, non durable goods, and services. Estimation of the group-restricted model is performed by iteration as follows: first, we initialize the estimate of the first factor by extracting one principal component from the covariance matrix of the data. Then, we estimate the common loading in each of the three groups by pooled OLS by regressing sectoral prices in each group on the factor. With those estimates for the three loadings, we obtain a new estimate for the factor by regressing the sectoral prices on the loadings. We use this new estimate to reinitialize the factor and iterate on the steps described above until the maximum difference between the common components estimated in two consecutive iterations is smaller that 0.001. Once convergence is attained, we estimate the remaining two unrestricted factors using the original data after removing the estimated contribution of the restricted factor.

We also estimate this grouped model at a more disaggregated level by allowing for heterogeneity across thirteen subcategories.[8] The second and third rows of the Panel A of Table 6 show that the test statistic is very large in all cases, suggesting that the hypothesis that price factors affect sectors in the same category or subcategory equally is rejected by the data.

Figure 3: Factor loading estimates. Model with 3 and 13 groups



Notes: Data are quarterly price changes in each of 187 sectors in the US personal consumption expenditures category of the national income and product accounts from 1959:Q1 to 2006:Q2. Estimation of the restricted model is described in Section 5.

Although the tests strongly reject the null hypothesis, it is informative to look at the estimated loadings in each group. Figure 3 plots the estimated loadings for each of the three groups and for each group's subcategories. Loadings in each case are rescaled by dividing them by their cross-sectional mean so to have a mean of one. Interestingly, we see that in the model with three groups there does not seem to be substantial heterogeneity across durables, non durables and services. The picture, however, masks substantial heterogeneity, which becomes apparent once we allow for more disaggregation: within services goods, we see that, for example, the effect of the factor on recreation and transportation sectors is much weaker than that on medical care and housing. For nondurables,

[8] The corresponding subcategories are: motor vehicles, furnitures, other durables, food, clothing, gasoline, other nondurables, housing, household operation, transportation, medical care, recreation and other. For a detailed list of sectors, categories and subcategories, see the Supplemental Appendix.

we see that the loading of the gasoline and other fuels subcategory is much different than the clothing or other nondurables one. The same pattern applies to durable goods subcategories.

As a final exercise, we let the data choose the groups. We begin by estimating a one-factor model by PCA. The loadings are ordered from the smallest to the largest and the $N = 187$ series are divided in three, four and five groups of equal size.[9] Results are reported in the Panel B of Table 6. Although the test rejects the structure with three groups, when we allow for four groups the Q^3_{NT} statistic, which accounts for both time series and cross-sectional correlation, is below the standard normal critical value of 1.64, with a p-value of 0.58. Allowing for five groups is enough for neither of the three test statistics to reject the null of grouped loadings.

These results suggest that there seem to be enough differences in the loadings to reject the existence of a pure inflation factor, but these differences are not as strong as to imply strong heterogeneity: with as little as four groups chosen by the data, the restricted model is not rejected as a valid representation of the data.

6 Extensions

In the previous sections we have focused on the setup and the null hypothesis of Example 1, in which N restrictions on the matrix of factor loadings are tested; here we provide a few additional examples to illustrate how to conduct inference using the proposed methodology in similar but slightly different situations.

A first example could be testing for equality of the factor loading coefficients for a group of the N series, say n, leaving the remaining $N - n$ factor loadings unrestricted, in a context in which n/N converges to a constant. For instance, in our empirical application one could relax the null of constant loadings in all sectors to test for a pure inflation factor within the services category only. In Panel A of Table 7 we provide the corresponding expressions to Lemma 1 as well as mean and variances for Theorems 1, 2 and 3 in the case in which $n = N/2$. Not surprisingly, the test statistic accounts for differences in the restricted and unrestricted common components related to the $N/2$ series whose loading enter the null hypothesis. Although in the current context we suggest the use of an iterative estimation procedure, similar to the group-restricted model estimator described in Section 5 which exploits the N series to estimate the factors, the intution for the quantities involved in the means and variances can be easily understood as being the same as in Section 3, but arising from using the

[9] Clearly, when the number of series is not divisible by the number of groups, some of the groups will have one additional unit.

Table 7: Asymptotic distribution of the test statistic Q_{NT} under different null hypotheses

Unrestricted econometric model: $X_{it} = \lambda_i f_t + \gamma_i g_t + e_{it}$, for $i = 1, ..., N, t = 1, ..., T$

Panel A: $H_0 : \lambda_i = \bar{\lambda}$ for $i = 1, ..., N/2$

$$\tilde{c}_{it} - \hat{c}_{it} = \begin{cases} \frac{1}{T} F_t' \mathbf{P}_T \sum_{s=1}^{T} F_s e_{is} + O_p\left(\frac{1}{\delta_{NT}^2}\right) & \text{for } i = 1, ..., N/2 \\ O_p\left(\frac{1}{\delta_{NT}^2}\right) & \text{for } i = N/2+1, ..., N \end{cases}$$

iid errors ($\Theta = I_N, \Phi = I_T$)

$E(Q_{NT}) = N/2$

$V(Q_{NT}) = N$

Weak cross-sectional correlation ($\Phi = I_T$)

$E(Q_{NT}) = N/2$

$V(Q_{NT}) = N^2/2 \times tr[(\Theta_0\Theta_0')^2]/[tr(\Theta_0\Theta_0')]^2$

Weak cross-sectional and serial correlation

$E(Q_{NT}) = N/2 \times E\left[tr\left(\mathbf{P}_T \mathbf{F}'\Phi'\Phi\mathbf{F}\right)\right]/tr(\Phi'\Phi)$

$V(Q_{NT}) = N^2/2 \times tr\left[(\Theta_0\Theta_0')^2\right]/tr\left[(\Theta_0\Theta_0')\right]^2 \times E[tr\left(\mathbf{P}_T \mathbf{F}'\Phi'\Phi\mathbf{F}\right)^2]/\left[tr(\Phi'\Phi)\right]^2$

Panel B: $H_0 : g_t = \bar{g}$ for all $t = 1, ..., T$

$\tilde{c}_{it} - \hat{c}_{it} = \frac{1}{N}\lambda_i' \mathbf{Q}_N \sum_{j=1}^{N} \lambda_j e_{jt} + O_p\left(\frac{1}{\delta_{NT}^2}\right)$

iid errors ($\Theta = I_N, \Phi = I_T$)

$E(Q_{NT}) = T$

$V(Q_{NT}) = 2T$

Weak cross-sectional correlation ($\Phi = I_T$)

$E(Q_{NT}) = T \times E[tr\left(\mathbf{Q}_N \mathbf{L}'\Theta\Theta'\mathbf{L}\right)]/tr(\Theta\Theta')$

$V(Q_{NT}) = 2T \times E[tr\left(\mathbf{Q}_N \mathbf{L}'\Theta\Theta'\mathbf{L}\right)^2]/[tr(\Theta\Theta')]^2$

Weak cross-sectional and serial correlation

$E(Q_{NT}) = T \times E[tr\left(\mathbf{Q}_N \mathbf{L}'\Theta\Theta'\mathbf{L}\right)]/tr(\Theta\Theta')$

$V(Q_{NT}) = 2T^2 \times tr[(\Phi'\Phi)^2]/tr[(\Phi'\Phi)]^2 \times E[tr\left(\mathbf{Q}_N \mathbf{L}'\Theta\Theta'\mathbf{L}\right)^2]/\left[tr(\Theta\Theta')\right]^2$

Panel C: $H_0 : \lambda_i = \bar{\lambda}$ for all $i = 1, ..., N$ and $g_t = \bar{g}$ for all $t = 1, ..., T$

$\tilde{c}_{it} - \hat{c}_{it} = \frac{1}{T} F_t' \mathbf{P}_T \sum_{s=1}^{T} F_s e_{is} + \frac{1}{N} l_i' \mathbf{Q}_N \sum_{j=1}^{N} l_j e_{jt} + O_p\left(\frac{1}{\delta_{NT}^2}\right)$

iid errors ($\Theta = I_N, \Phi = I_T$)

$E(Q_{NT}) = N + T$

$V(Q_{NT}) = 2(N + T)$

Weak cross-sectional correlation ($\Phi = I_T$)

$E(Q_{NT}) = N + T \times E\left[tr\left(\mathbf{Q}_N \mathbf{L}'\Theta\Theta'\mathbf{L}\right)\right]/tr(\Theta\Theta')$

$V(Q_{NT}) = 2N^2 \times tr[(\Theta\Theta')^2]/tr[(\Theta\Theta')]^2 + 2T \times E[tr\left(\mathbf{Q}_N \mathbf{L}'\Theta\Theta'\mathbf{L}\right)^2]/\left[tr(\Theta\Theta')\right]^2$

Weak cross-sectional and serial correlation

$E(Q_{NT}) = N \times E\left[tr\left(\mathbf{P}_T \mathbf{F}'\Phi'\Phi\mathbf{F}\right)\right]/tr(\Phi'\Phi) + T \times E\left[tr\left(\mathbf{Q}_N \mathbf{L}'\Theta\Theta'\mathbf{L}\right)\right]/tr(\Theta\Theta')$

$$\begin{aligned} V(Q_{NT}) = {} & 2N^2 \times tr[(\Theta\Theta')^2]/tr[(\Theta\Theta')]^2 \times E[tr\left(\mathbf{P}_T \mathbf{F}'\Phi'\Phi\mathbf{F}\right)^2]/\left[tr(\Phi'\Phi)\right]^2 \\ & + 2T^2 \times tr[(\Phi'\Phi)^2]/tr[(\Phi'\Phi)]^2 \times E[tr\left(\mathbf{Q}_N \mathbf{L}'\Theta\Theta'\mathbf{L}\right)^2]/\left[tr(\Theta\Theta')\right]^2 \end{aligned}$$

Notes: $\delta_{NT} = \min(\sqrt{N}, \sqrt{T})$ $\mathbf{L} = [l_1', ..., l_N']'$ with $l_i = (\lambda_i, \gamma_i)'$, $\mathbf{F} = [F_1', ..., F_T']'$ with $F_t = (f_t, g_t)'$, Θ_0 is the $N \times N/2$ matrix containing the first $N/2$ columns of Θ,

$$\mathbf{P}_T = \left(\frac{\mathbf{F}'\mathbf{F}}{T}\right)^{-1} - \begin{pmatrix} 0 & 0 \\ 0 & (\mathbf{G}'\mathbf{G}/T)^{-1} \end{pmatrix} \text{ and } \mathbf{Q}_N = \left(\frac{\mathbf{L}'\mathbf{L}}{N}\right)^{-1} - \begin{pmatrix} (\mathbf{\Lambda}'\mathbf{\Lambda}/N)^{-1} & 0 \\ 0 & 0 \end{pmatrix}$$

where $\mathbf{G} = (g_1, ...g_T)'$ and $\mathbf{\Lambda} = (\lambda_1, ..., \lambda_N)'$. A description of model assumptions is in Section 2.

relevant $N/2$ series only. Indeed, since $\tilde{c}_{it} - \hat{c}_{it} = O_p(\delta_{NT}^{-2})$ for $i = N/2+1, ..., N$ under $H_0 : \lambda_i = \bar{\lambda}$ for $i = 1, ..., N/2$, Q_{NT} is constructed using the restricted and unrestricted squared residuals of the first $N/2$ series only, that is $Q_{NT} = (N/2)T\left(\hat{\sigma}^2_{NT} - \tilde{\sigma}^2_{NT}\right)/\hat{\sigma}^2_{NT}$ where $\tilde{\sigma}^2_{NT} = 2(NT)^{-1}\sum_{i=1}^{N/2}\sum_{t=1}^{T}\tilde{e}^2_{it}$ and $\hat{\sigma}^2_{NT} = 2(NT)^{-1}\sum_{i=1}^{N/2}\sum_{t=1}^{T}\hat{e}^2_{it}$. As a consequence, the correction term for the variance in the presence of cross-sectional correlation is the analogous to the one in Theorem 1 but excluding the series whose residual estimates do not enter in Q_{NT} i.e. Θ_0 only contains the first $N/2$ columns of Θ.

Typically in panel data it is assumed that the unobserved individual-specific heterogeneity is fixed across time, but there are several reasons why this assumption may not hold. Another potentially interesting case is, therefore, that of testing for constant individual effects. Panel B of Table 7 reports the necessary quantities to apply our testing procedure in this context. Not surprisingly, the mean and the variance in Theorems 2 and 3 now depend on the loadings.

Finally, in a panel data context in which both dimensions are large it is customary to control for common time effects or trends f_t and individual fixed effects λ_i by including time and individual dummies in the estimation. This amounts to modeling the unobserved heterogeneity in the model as $\lambda_i + f_t$. If multiplicative (also "interactive") effects of the form $\lambda_i f_t$ are present as in Bai (2009), however, the within-group estimator is inconsistent because it relies on additivity. Our approach provides an alternative to the Hausman type test introduced in that paper. The results we provide in Panel C of Table 7 accommodate this situation by exploiting the orthogonality of eigenvectors, i.e. asymptotic independence of the two set of hypotheses, to combine the results of Section 3 with those in Panel B of Table 7: the number of hypotheses are $N+T$; and, since one eigenvector of each dimension is restricted under the null, the asymptotic distribution of Q^3_{NT} now depends on both factors and factor loadings.

7 Summary and directions for further research

In this paper we develop a hypothesis testing framework in which the number of restrictions entering the null hypothesis grows with the sample size, possibly at a rate of $N+T$. We propose a simple testing procedure based on the sums of squared residuals of, respectively, the restricted and the unrestricted version of the approximate factor model and obtain the asymptotic distribution of the test statistic under different assumptions on the error terms, allowing for both serial and cross-section correlation.

In the *iid* case our procedure extends the classical results for testing in the linear model to the case of a factor model with large N and T. By properly rescaling and recentering the test statistic, we show that our framework remains valid even when the dependence among disturbances is left unspecified. Specifically, we derive the asymptotic properties of our test statistic under relatively flexible forms of weak cross-sectional and times-series dependence that are potentially relevant in many empirical situations. We show that the expressions for the limiting mean and variance can be easily corrected to allow for both types of correlation and provide simple and intuitive estimators for those quantities.

We conduct Monte Carlo exercises to study the finite sample reliability and power of our proposed tests. Our results show that the finite sample size of the different tests is reliable under correct specification. They also imply that the most restrictive versions of the tests, either by assuming that errors are *iid* or by neglecting times-series correlation, dominate in terms of power the more robust test that accounts for both types of dependence.

Finally, we provide an application of the test by taking the data in Reis and Watson (2010) and reconsidering their hypothesis of the existence of a pure inflation factor in the US price indices. The null of the presence of one factor with constant loadings is strongly rejected under the three versions of the test, which is in line with their finding that individual t-tests rejected the null of unit loading 30% of the times at the 5% level. Interestingly, when we let the data choose the groups, we find that, although there seem to be enough differences in the loadings to reject the existence of a pure inflation factor, these differences are not as strong as to imply substantial heterogeneity: with four groups chosen by the data the restricted model is not rejected.

The testing procedures we have developed can be extended in several interesting directions. First, in the context of macroeconomic forecasting, Breitung and Eickmeier (2011), Chen, Dolado, and Gonzalo (2014) and Han and Inoue (2013) have recently proposed testing procedures for structural breaks in approximate factor models, a relevant feature since when macroeconomic forecasts are constructed using a sample that spans a long period, some degree of temporal instability is inevitable and determining whether it is large enough in order to be considered a break is empirically relevant. In this respect, our methodology could be adapted to models with structural breaks in the loadings, in which case one has to endogenize the timing of the break. Moreover, it has the potential for disentangling breaks in factor dynamics from breaks in factor loadings.

Finally, recent papers have proposed factor models that use common and block-specific factors to capture the between- and within-block variations in the data, hence incorporating structure, taking advantage of it in terms of efficiency, and facilitating the interpretation of the results. For instance, Kose, Otrok, and Whiteman (2003) investigate the dynamic properties of business-cycle fluctuations across countries, regions, and the world. Stock and Watson (2009) study housing construction activity in the US through a factor model with a national factor, a regional factor, and a state-specific error; more recently, Moench, Ng, and Potter (2013) propose multilevel factor models to characterize within and between-block variations as well as idiosyncratic noise in large dynamic panels. Our approach could also be extended to develop specification tests for the aforementioned models. All these topics constitute interesting avenues for further research.

References

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

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

——— (2009): "Panel Data Models With Interactive Fixed Effects," *Econometrica*, 77(4), 1229–1279.

——— (2010): "Principal components and factor analysis, evaluating commonly used estimation procedures," Discussion paper, Department of Economics, Columbia University.

Bai, J., and S. Ng (2002): "Determining the Number of Factors in Approximate Factor Models," *Econometrica*, 70(1), 191–221.

——— (2004): "A PANIC Attack on Unit Roots and Cointegration," *Econometrica*, 72(4), 1127–1177.

Bai, J., and S. Shi (2011): "Estimating High Dimensional Covariance Matrices and its Applications," *Annals of Economics and Finance*, 12(2), 199–215.

Berndt, E. R., and N. E. Savin (1977): "Conflict among Criteria for Testing Hypotheses in the Multivariate Linear Regression Model," *Econometrica*, 45(5), 1263–77.

Bradley, R. C. (2005): "Basic properties of strong mixing conditions. A survey and some open questions," *Probability surveys*, 2(2), 107–144.

Breitung, J., and S. Eickmeier (2011): "Testing for structural breaks in dynamic factor models," *Journal of Econometrics*, 163(1)(2009,05), 71–84.

Breitung, J., and J. Tenhofen (2011): "GLS estimation of dynamic factor models," *Journal of the American Statistical Association*, 106(495), 1150–1166.

Chamberlain, G., and M. Rothschild (1983): "Arbitrage, Factor Structure, and Mean-Variance Analysis on Large Asset Markets," *Econometrica*, 51(5), 1281–304.

Chen, L., J. J. Dolado, and J. Gonzalo (2014): "Detecting big structural breaks in large factor models," *Journal of Econometrics*, 180(1)(31344), 30–48.

Choi, I. (2001): "Unit root tests for panel data," *Journal of International Money and Finance*, 20(2), 249–272.

——— (2012): "Efficient estimation of factor models," *Econometric Theory*, 28, 274–308.

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

Connor, G., and R. A. Korajczyk (1993): "A Test for the Number of Factors in an Approximate Factor Model," *Journal of Finance*, 48(4), 1263–91.

Davidson, R., and J. G. MacKinnon (1998): "Graphical Methods for Investigating the Size and Power of Hypothesis Tests," *The Manchester School of Economic & Social Studies*, 66(1), 1–26.

El Karoui, N. (2008): "Spectrum estimation for large dimensional covariance matrices using random matrix theory," *Annals of Statistics*, 36(6), 2757–2790.

Evans, G. B. A., and N. E. Savin (1982): "Conflict among the Criteria Revisited: The W, LR and LM Tests," *Econometrica*, 50(3), 737–48.

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

Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2005): "The Generalized Dynamic Factor Model: One-Sided Estimation and Forecasting," *Journal of the American Statistical Association*, 100, 830–840.

Han, X., and A. Inoue (2013): "Tests for Parameter Instability in Dynamic Factor Models," TERG Discussion Papers 306, Graduate School of Economics and Management, Tohoku University.

Harding, M. (2013): "Estimating the Number of Factors in Large Dimensional Factor Models," Discussion paper, Duke University.

Im, K. S., M. H. Pesaran, and Y. Shin (2003): "Testing for unit roots in heterogeneous panels," *Journal of Econometrics*, 115(1), 53–74.

Kent, J. (1982): "Robust Properties of Likelihood Ratio Test," *Biometrika*, 69, 19–27.

Kose, M. A., C. Otrok, and C. H. Whiteman (2003): "International Business Cycles: World, Region, and Country-Specific Factors," *American Economic Review*, 93(4), 1216–1239.

Maddala, G. S., and S. Wu (1999): "A Comparative Study of Unit Root Tests with Panel Data and a New Simple Test," *Oxford Bulletin of Economics and Statistics*, 61, 631–52.

Moench, E., S. Ng, and S. Potter (2013): "Dynamic hierarchical factor models," *Review of Economics and Statistics*, 95(5)(412), 1811–1817.

Newey, W. K., and D. McFadden (1994): "Large sample estimation and hypothesis testing," in *Handbook of Econometrics*, ed. by R. F. Engle, and D. McFadden, vol. 4 of *Handbook of Econometrics*, chap. 36, pp. 2111–2245. Elsevier.

Pesaran, M. H. (2007): "A simple panel unit root test in the presence of cross-section dependence," *Journal of Applied Econometrics*, 22(2), 265–312.

Pesaran, M. H., and T. Yamagata (2012): "Testing CAPM with a Large Number of Assets," Discussion Papers 12/05, Department of Economics, University of York.

Reis, R., and M. W. Watson (2010): "Relative Goods' Prices, Pure Inflation, and the Phillips Correlation," *American Economic Journal: Macroeconomics*, 2(3), 128–57.

Ross, S. A. (1976): "The arbitrage theory of capital asset pricing," *Journal of Economic Theory*, 13(3), 341–360.

Stock, J., and M. Watson (2002): "Forecasting Using Principal Components From a Large Number of Predictors," *Journal of the American Statistical Association*, 97, 1167–1179.

Stock, J., and M. W. Watson (2009): "The Evolution of National and Regional Factors in U.S. Housing Construction," in *Volatility and Time Series Econometrics: Essays in Honor of Robert F. Engle*, ed. by J. Fagerberg, D. Mowery, and R. Nelson. New York: Oxford University Press, Oxford.

White, H. (2000): *Asymptotic theory for econometricians.* Academic Press; Revised edition (October 11, 2000).

Appendix

A Proofs

To prove the main results we make use of the following lemmata:

Lemma 2. *Suppose Assumption A holds and let $S_F = (T^{-1}\mathbf{F}'\mathbf{F})$ and*

$$\mathbf{M}_T = \begin{bmatrix} 1 & \mathbf{0}' \\ -(\mathbf{G}'\mathbf{G})^{-1}\mathbf{G}'\mathbf{F} & 0 \end{bmatrix}.$$

Then,

a) $\sum_{t=1}^{T} \mathbf{P}_T F_t F_t' = T\mathbf{M}_T$*, Hence,* $tr\left[\sum_{t=1}^{T} \mathbf{P}_T F_t F_t'\right] = T.$

b) $\sum_{t=1}^{T} S_F^{-1} F_t F_t' = T \times I_r$*. Hence,* $tr\left[\sum_{t=1}^{T} S_F^{-1} F_t F_t'\right] = rT.$

c) $\sum_{t'=1}^{T} \left(F_t'\mathbf{P}_T F_s\right)\left(F_{t'}'\mathbf{P}_T F_{s'}\right)\left(F_{t'}'\mathbf{P}_T F_{s''}\right) = T \times tr\left[\left(\mathbf{P}_T F_{s'} F_{s''}'\right)\left(\mathbf{P}_T F_{s'''} F_t'\right)\right].$

d) $\sum_{t=1}^{T}\sum_{t'=1}^{T} \left(F_t'\mathbf{P}_T F_s\right)\left(F_{t'}'\mathbf{P}_T F_{s'}\right)\left(F_{t'}' S_F^{-1} F_{s''}\right) = T \times tr\left[\left(\mathbf{P}_T F_{s'} F_{s''}'\right)\left(\mathbf{P}_T F_{s'''} F_t'\right)\right].$

e) $\sum_{t=1}^{T}\sum_{t'=1}^{T} \left(F_t'\mathbf{P}_T F_s\right)\left(F_t'\mathbf{P}_T F_{s'}\right)\left(F_{t'}'\mathbf{P}_T F_{s''}\right)\left(F_{t'}'\mathbf{P}_T F_{s'''}\right) = T^2 \times tr\left[\left(\mathbf{P}_T F_{s'} F_{s''}'\right)\left(\mathbf{P}_T F_{s'''} F_s'\right)\right].$

f) $\sum_{t=1}^{T}\sum_{t'=1}^{T} \left(F_t'\mathbf{P}_T F_s\right)\left(F_{t'}'\mathbf{P}_T F_{s'}\right)\left(F_t' S_F^{-1} F_{s''}\right)\left(F_{t'}' S_F^{-1} F_{s'''}\right) = T^2 \times tr\left[\left(\mathbf{P}_T F_{s'} F_{s''}'\right)\left(\mathbf{P}_T F_{s'''} F_s'\right)\right].$

g) $\sum_{t=1}^{T}\sum_{t'=1}^{T} \left(F_t'\mathbf{P}_T F_s\right)\left(F_t'\mathbf{P}_T F_{s'}\right)\left(F_{t'}'\mathbf{P}_T F_{s''}\right)\left(F_{t'}' S_F^{-1} F_{s'''}\right) = T^2 \times tr\left[\left(\mathbf{P}_T F_{s'} F_{s''}'\right)\left(\mathbf{P}_T F_{s'''} F_s'\right)\right].$

Proof of Lemma 2

a) Write

$$\begin{aligned}\sum_{t=1}^{T} \mathbf{P}_T F_t F_t' &= T\mathbf{P}_T S_F = T\left[S_F^{-1} - \begin{pmatrix} 0 & \mathbf{0}' \\ \mathbf{0} & (\mathbf{G}'\mathbf{G}/T)^{-1} \end{pmatrix}\right] S_F = T\left[I_r - \begin{pmatrix} 0 & \mathbf{0}' \\ (\mathbf{G}'\mathbf{G})^{-1}\mathbf{G}'\mathbf{F} & I_{r-1} \end{pmatrix}\right] \\ &= T\begin{bmatrix} 1 & \mathbf{0}' \\ -(\mathbf{G}'\mathbf{G})^{-1}\mathbf{G}'\mathbf{F} & \mathbf{0} \end{bmatrix} = T\mathbf{M}_T.\end{aligned}$$

where the first and second equalities follow from the definition of S_F and $\mathbf{P}_T$, respectively. Since $tr(\mathbf{M}_T) = 1$, the trace result follows immediately.

b) Follows immediately from the definition of S_F.

c) Rearrange terms within the quadratic forms as follows

$$\left(F_t'\mathbf{P}_T F_s\right)\left(F_{t'}'\mathbf{P}_T F_{s'}\right)\left(F_{t'}'\mathbf{P}_T F_{s''}\right) = \left(F_t'\mathbf{P}_T F_s\right)\left(F_{s'}'\mathbf{P}_T F_{t'}\right)\left(F_{t'}'\mathbf{P}_T F_{s''}\right);$$

then using the commutative property of the trace to write

$$tr\left[\left(F_t'\mathbf{P}_T F_s\right)\left(F_{s'}'\mathbf{P}_T F_{t'}\right)\left(F_{t'}'\mathbf{P}_T F_{s''}\right)\right] = tr\left[\left(\mathbf{P}_T F_s F_{s'}'\right)\left(\mathbf{P}_T F_{t'} F_{t'}'\right)\left(\mathbf{P}_T F_{s''} F_t'\right)\right]$$

so that

$$\sum_{t'=1}^{T} tr\left[\left(\mathbf{P}_T F_s F_{s'}'\right)\left(\mathbf{P}_T F_{t'} F_{t'}'\right)\left(\mathbf{P}_T F_{s''} F_t'\right)\right] = T \times tr\left[\left(\mathbf{P}_T F_{t'} F_{t'}'\right)\mathbf{M}_T\left(\mathbf{P}_T F_{s''} F_t'\right)\right],$$

where the last equality follows from 2.a. Now, notice that the matrix $\mathbf{M}_T$ is idempotent and $\mathbf{M}_T\mathbf{P}_T = \mathbf{P}_T$. To see this, write

$$S_F = \frac{1}{T}\left[\begin{pmatrix} \mathbf{F}'\mathbf{F} & \mathbf{F}'\mathbf{G} \\ \mathbf{G}'\mathbf{F} & \mathbf{G}'\mathbf{G} \end{pmatrix}\right] = \frac{1}{T}\left[\begin{pmatrix} \mathbf{A} & \mathbf{C}' \\ \mathbf{C} & \mathbf{D} \end{pmatrix}\right].$$

Using the formula for the partitioned inverse, we have

$$S_F^{-1} = \frac{1}{T}\left[\begin{pmatrix} \left(\mathbf{A}-\mathbf{C}'\mathbf{D}^{-1}\mathbf{C}\right)^{-1} & -\left(\mathbf{A}-\mathbf{C}'\mathbf{D}^{-1}\mathbf{C}\right)^{-1}\mathbf{C}'\mathbf{D}^{-1} \\ -\mathbf{D}^{-1}\mathbf{C}\left(\mathbf{A}-\mathbf{C}'\mathbf{D}^{-1}\mathbf{C}\right)^{-1} & \mathbf{D}^{-1}+\mathbf{D}^{-1}\mathbf{C}\left(\mathbf{A}-\mathbf{C}'\mathbf{D}^{-1}\mathbf{C}\right)^{-1}\mathbf{C}'\mathbf{D}^{-1} \end{pmatrix}\right]$$

so that

$$\mathbf{P}_T = S_F^{-1} - \frac{1}{T}\begin{bmatrix} 0 & \mathbf{0}' \\ \mathbf{0} & (\mathbf{G}'\mathbf{G})^{-1} \end{bmatrix} = \frac{1}{T}\left[\begin{pmatrix} \left(\mathbf{A}-\mathbf{C}'\mathbf{D}^{-1}\mathbf{C}\right)^{-1} & -\left(\mathbf{A}-\mathbf{C}'\mathbf{D}^{-1}\mathbf{C}\right)^{-1}\mathbf{C}'\mathbf{D}^{-1} \\ -\mathbf{D}^{-1}\mathbf{C}\left(\mathbf{A}-\mathbf{C}'\mathbf{D}^{-1}\mathbf{C}\right)^{-1} & \mathbf{D}^{-1}\mathbf{C}\left(\mathbf{A}-\mathbf{C}'\mathbf{D}^{-1}\mathbf{C}\right)^{-1}\mathbf{C}'\mathbf{D}^{-1} \end{pmatrix}\right].$$

From which it readily follows that $\mathbf{M}_T\mathbf{P}_T = \mathbf{P}_T$. Hence,

$$\begin{aligned}\sum_{t'=1}^{T}\left(F_t'\mathbf{P}_T F_s\right)\left(F_{t'}'\mathbf{P}_T F_{s'}\right)\left(F_{t'}'\mathbf{P}_T F_{s''}\right) &= T \times tr\left[\left(\mathbf{P}_T F_{s'} F_{s''}'\right)\left(\mathbf{M}_T\mathbf{P}_T F_{s'''} F_t'\right)\right] \\ &= T \times tr\left[\left(\mathbf{P}_T F_{s'} F_{s''}'\right)\left(\mathbf{P}_T F_{s'''} F_t'\right)\right].\end{aligned}$$

d) The proof is analogous to the one of 2.c but using 2.b instead of 2.a in the last step.

e) Rearrange terms within the quadratic forms as follows:

$$\left(F_t'\mathbf{P}_T F_s\right)\left(F_t'\mathbf{P}_T F_{s'}\right)\left(F_{t'}'\mathbf{P}_T F_{s''}\right)\left(F_{t'}'\mathbf{P}_T F_{s'''}\right) = \left(F_s'\mathbf{P}_T F_t\right)\left(F_t'\mathbf{P}_T F_{s'}\right)\left(F_{s''}'\mathbf{P}_T F_{t'}\right)\left(F_{t'}'\mathbf{P}_T F_{s'''}\right);$$

then, using the commutative property of the trace, write

$$tr\left[\left(F_s'\mathbf{P}_T F_t\right)\left(F_t'\mathbf{P}_T F_{s'}\right)\left(F_{s''}'\mathbf{P}_T F_{t'}\right)\left(F_{t'}'\mathbf{P}_T F_{s'''}\right)\right] = tr\left[\left(\mathbf{P}_T F_t F_t'\right)\left(\mathbf{P}_T F_{s'} F_{s''}'\right)\left(\mathbf{P}_T F_{t'} F_{t'}'\right)\left(\mathbf{P}_T F_{s'''} F_s'\right)\right]$$

so that

$$\sum_{t=1}^{T}\sum_{t'=1}^{T} tr\left[\left(\mathbf{P}_T F_t F_t'\right)\left(\mathbf{P}_T F_{s'} F_{s''}'\right)\left(\mathbf{P}_T F_{t'} F_{t'}'\right)\left(\mathbf{P}_T F_{s'''} F_s'\right)\right] = T^2 \times tr\left[\left(\mathbf{P}_T F_{s'} F_{s''}'\right)\left(\mathbf{P}_T F_{s'''} F_s'\right)\right],$$

where the last equality follows from 2.a and again using the fact that $\mathbf{M}_T\mathbf{P}_T = \mathbf{P}_T$.

f) The proof is analogous to the one of 2.c but using 2.b instead of 2.a in the last step.

g) The proof is analogous to the one of 2.c but using correspondingly 2.b and 2.a in the last step. $\square$

Lemma 3. *Under Assumptions B*

$$\begin{aligned}
E(e_{kv}e_{k'v'}e_{k''v''}e_{k'''v'''}) &= \sum_{u=1}^{T}\sum_{u'=1}^{T}\sum_{h=1}^{N}\sum_{h'=1}^{N}\theta_{kh}\theta_{k'h}\theta_{k''h'}\theta_{k'''h'}\phi_{uv}\phi_{uv'}\phi_{u'v''}\phi_{u'v'''} \\
&+\sum_{u=1}^{T}\sum_{u'=1}^{T}\sum_{h=1}^{N}\sum_{h'=1}^{N}\theta_{kh}\theta_{k'h'}\theta_{k''h}\theta_{k'''h'}\phi_{uv}\phi_{u'v'}\phi_{uv''}\phi_{u'v'''} \\
&+\sum_{u=1}^{T}\sum_{u'=1}^{T}\sum_{h=1}^{N}\sum_{h'=1}^{N}\theta_{kh}\theta_{k'h'}\theta_{k''h'}\theta_{k'''h}\phi_{uv}\phi_{u'v'}\phi_{u'v''}\phi_{uv'''} \\
&+\sum_{u=1}^{T}\sum_{h=1}^{N}\theta_{kh}\theta_{k'h}\theta_{k''h}\theta_{k'''h}\phi_{uv}\phi_{uv'}\phi_{uv''}\phi_{uv'''}\kappa_{\varepsilon}.
\end{aligned}$$

Proof of Lemma 3

From Assumption B we can write $e_{it} = \sum_{h=1}^{N}\sum_{u=1}^{T}\theta_{ih}\varepsilon_{hu}\phi_{ut}$, and hence

$$\begin{aligned}
E(e_{kv}e_{k'v'}e_{k''v''}e_{k'''v'''}) &= \sum_{h=1}^{N}\sum_{h'=1}^{N}\sum_{h''=1}^{N}\sum_{h'''=1}^{N}\theta_{kh}\theta_{k'h'}\theta_{k''h''}\theta_{k'''h'''} \\
&\quad \sum_{u=1}^{T}\sum_{u'=1}^{T}\sum_{u''=1}^{T}\sum_{u'''=1}^{T}\phi_{uv}\phi_{u'v'}\phi_{u''v''}\phi_{u'''v'''}E(\varepsilon_{hu}\varepsilon_{h'u'}\varepsilon_{h''u''}\varepsilon_{h'''u'''}).
\end{aligned}$$

For a given h in each of the four cross-sectional sums, the expectations will be different from zero only if:

1) $u = u' = u'' = u'''$, which are T elements with $\varepsilon_{hu}\varepsilon_{h'u}\varepsilon_{h''u}\varepsilon_{h'''u}$,

2) $u = u' \neq u'' = u'''$, which are $T(T-1)$ elements with $\varepsilon_{hu}\varepsilon_{h'u}\varepsilon_{h''u''}\varepsilon_{h'''u''}$,

3) $u = u'' \neq u' = u'''$, which are $T(T-1)$ elements with $\varepsilon_{hu}\varepsilon_{h'u}\varepsilon_{h''u'}\varepsilon_{h'''u'}$, and

4) $u = u''' \neq u' = u''$, which are $T(T-1)$ elements with $\varepsilon_{hu}\varepsilon_{h'u'}\varepsilon_{h''u'}\varepsilon_{h'''u}$,

while all the remaining ones are zero. Similarly, for a given u in each of the four times-series sums, the expectations will be different from zero only if:

i) $h = h' = h'' = h'''$, which are N elements with $\varepsilon_{hu}\varepsilon_{hu'}\varepsilon_{hu''}\varepsilon_{hu'''}$,

ii) $h = h' \neq h'' = h'''$, which are $N(N-1)$ elements with $\varepsilon_{hu}\varepsilon_{hu'}\varepsilon_{h''u''}\varepsilon_{h''u'''}$,

iii) $h = h'' \neq h' = h'''$, which are $N(N-1)$ elements with $\varepsilon_{hu}\varepsilon_{h'u'}\varepsilon_{hu''}\varepsilon_{h'u'''}$, and

iv) $h = h''' \neq h' = h''$, which are $N(N-1)$ elements with $\varepsilon_{hu}\varepsilon_{h'u'}\varepsilon_{h'u''}\varepsilon_{hu'''}$.

while all the remaining ones are zero. Hence, when considering both types of sums altogether we end up with sixteen cases that we label by (i, j) where $i = 1, ..., 4$ refers to the different cases we classify the u's, and $j = i, ..., iv$ to the different cases we classify the h's.

Then, notice that cases $(2, iii)$, $(2, iv)$, $(3, ii)$, $(3, iv)$, $(4, ii)$ and $(4, iii)$, each of them containing $N(N-1)T(T-1)$ elements, are zero in expectation. Next, Grouping terms $(4.iv)$, $(4, i)$ and $(1, iv)$, we obtain

$$\sum_{u=1}^{T}\sum_{u'=1}^{T}\sum_{h=1}^{N}\sum_{h'=1}^{N}\theta_{kh}\theta_{k'h'}\theta_{k''h'}\theta_{k'''h}\phi_{uv}\phi_{u'v'}\phi_{u'v''}\phi_{uv'''}-\sum_{u=1}^{T}\sum_{h=1}^{N}\theta_{kh}\theta_{k'h}\theta_{k''h}\theta_{k'''h}\phi_{uv}\phi_{uv'}\phi_{uv''}\phi_{uv'''}.$$

Doing the same with $(3, iii)$, $(3, i)$ and $(1, iii)$, and $(2, ii)$, $(2, i)$ and $(1, ii)$, we obtain analogous expressions, which, collected with $(1, i)$ yield the result. $\square$

Lemma 4. *Under Assumption B there exists a constant $M < \infty$ such that*

a) $\sum_{i=1}^{N}\theta_{ih}^2 \leq M$.

b) $\sum_{u=1}^{T}\phi_{ut} \leq M$.

c) $\sum_{u=1}^{T}\phi_{ut}\phi_{us} \leq M$.

d) $\sum_{j=1}^{N}\theta_{jh} \leq M$.

Proof of Lemma 4

a) Assumption B2 implies that all eigenvalues of $\Theta\Theta'$ are bounded and therefore it also implies $tr(\Theta\Theta') = \sum_{i=1}^{N}\sum_{h=1}^{N}\theta_{ih}^2 = O(N)$. This fact, together with Assumption B3, implies that also all row and column sums are bounded, so also $\sum_{i=1}^{N}\theta_{ih}^2$ is.

b) Again by Assumption B2, one has that $tr(\Phi'\Phi) = \sum_{t=1}^{T}\sum_{u=1}^{T}\phi_{ut}^2 = O(T)$ and, hence, $\sum_{u=1}^{T}\phi_{ut}^2 = O(1)$. The result then follows by noticing that, by Hölder's inequality,

$$\sum_{u=1}^{T}\phi_{ut} \leq \left(\sum_{u=1}^{T}\phi_{ut}^2\right)^{1/2} \leq M.$$

c) Similarly,

$$\sum_{u=1}^{T}\phi_{ut}\phi_{us} \leq \left(\sum_{u=1}^{T}\phi_{ut}^2\right)^{1/2}\left(\sum_{u=1}^{T}\phi_{us}^2\right)^{1/2} \leq M$$

where the first inequality follows from Cauchy-Schwarz (CS, henceforth) and the last one from Lemma 4.a applied to Φ.

d) Same as Lemma 4.b but using $tr(\Theta\Theta')$ instead. $\square$

Proof of Lemma 1

a) Start by writing

$$\begin{aligned}\tilde{c}_{it} - \hat{c}_{it} &= \tilde{L}_i'\tilde{F}_t - \hat{L}_i'\hat{F}_t \\ &= (\tilde{F}_t - H'F_t)'H^{-1}L_i - (\hat{F}_t - H_R'F_t)'H_R^{-1}L_i \\ &\quad +F_t'H(\tilde{L}_i - H^{-1}L_i) - F_t'H_R(\hat{L}_i - H_R^{-1}L_i) \\ &\quad +(\tilde{F}_t - H'F_t)'(\tilde{L}_i - H^{-1}L_i) - (\hat{F}_t - H_R'F_t)'(\hat{L}_i - H_R^{-1}L_i),\end{aligned} \tag{12}$$

where the second equality follows from

$$\tilde{c}_{it} - c_{it} = (\tilde{F}_t - H'F_t)'H^{-1}L_i + F_t'H(\tilde{L}_i - H^{-1}L_i) + (\tilde{F}_t - H'F_t)'(\tilde{L}_i - H_i^{-1}L_i)$$

and

$$\hat{c}_{it} - c_{it} = (\hat{F}_t - H_R'F_t)'H_R^{-1}L_i + F_t'H_R(\hat{L}_i - H_R^{-1}L_i) + (\hat{F}_t - H_R'F_t)'(\hat{L}_i - H_R^{-1}L_i).$$

Since the first factor in the restricted model is identified, H_R is block diagonal with zeros in the first line and column except for the element in the $(1,1)$ position. Hence, defining as $\bar{H}_R$ the square submatrix of dimensions $r-1$ obtained by deleting the first row and column from H_R, we can write

$$(\tilde{F}_t - H'F_t)'H^{-1}L_i - (\hat{F}_t - H_R'F_t)'H_R^{-1}L_i = (\tilde{F}_t - H'F_t)'H^{-1}L_i + (\hat{f}_t - f_t)\lambda_i + (\hat{g}_t - \bar{H}_R g_t)'\bar{H}_R\gamma_i.$$

Replacing the expression for the unrestricted estimator, $(\tilde{F}_t - H'F_t)'H^{-1}L_i$, (see e.g. the proof of Theorem 3 Bai (2003) for the corresponding derivation) we obtain

$$L_i H'^{-1}(\tilde{F}_t - H'F_t) = \frac{1}{N} L_i \left(\frac{\mathbf{L}'\mathbf{L}}{N}\right)^{-1} \sum_{j=1}^{N} L_j e_{jt} + O_p\left(\frac{1}{\delta_{NT}^2}\right). \tag{13}$$

Next, defining $S_\Gamma^{-1} = [N^{-1}\sum_i^N(\gamma_i - \overline{\gamma})(\gamma_i - \overline{\gamma})']^{-1}$ and using the partitioned inverse formula we can write

$$\left(\frac{\mathbf{L}'\mathbf{L}}{N}\right)^{-1} = \begin{pmatrix} 1 + \overline{\gamma}'S_\Gamma^{-1}\overline{\gamma} & -\overline{\gamma}'S_\Gamma^{-1} \\ -S_\Gamma^{-1}\overline{\gamma} & S_\Gamma^{-1} \end{pmatrix},$$

and given that under H_0 L_i is equal to $(1, \gamma_i')'$ we can write the first term in the right hand side of (13) as

$$\begin{aligned} \frac{1}{N} L_i \left(\frac{\mathbf{L}'\mathbf{L}}{N}\right)^{-1} \sum_{j=1}^{N} L_j e_{jt} &= \overline{e}_t + \overline{e}_t\overline{\gamma}'S_\Gamma^{-1}\overline{\gamma} - \overline{e}_t\gamma_i'S_\Gamma^{-1}\overline{\gamma} - \frac{1}{N}\overline{\gamma}'S_\Gamma^{-1}\sum_{j=1}^{N}\gamma_j e_{jt} + \frac{1}{N}\gamma_i'S_\Gamma^{-1}\sum_{j=1}^{N}\gamma_j e_{jt} \\ &= \overline{e}_t + \frac{1}{N}(\gamma_i - \overline{\gamma})' S_\Gamma^{-1}\sum_{j=1}^{N}(\gamma_j - \overline{\gamma})e_{jt}, \end{aligned} \tag{14}$$

where $\overline{e}_t = \overline{y}_t - f_t - \overline{\gamma}'g_t$. Using the same asymptotic representation for $\gamma_i\bar{H}_R'^{-1}(\hat{g}_t - \bar{H}_R g_t)$, and accounting for the fact that the $r-1$ factors in the restricted model are also estimated by PCA but using as data the matrix of demeaned data $y_{it} - \bar{y}_t$, we can write

$$\gamma_i\bar{H}_R'^{-1}(\hat{g}_t - \bar{H}_R g_t) = \frac{1}{N}(\gamma_i - \overline{\gamma})' S_\Gamma^{-1}\sum_{j=1}^{N}(\gamma_j - \overline{\gamma})(e_{jt} - \overline{e}_t) + O_p\left(\frac{1}{\delta_{NT}^2}\right). \tag{15}$$

Expand the products in the first term of the right hand side of (15) to obtain

$$\begin{aligned}
\frac{1}{N}\left(\gamma_i-\overline{\gamma}\right)' S_\Gamma^{-1}\sum_{j=1}^{N}(\gamma_j-\overline{\gamma})(e_{jt}-\overline{e}_t) &= \frac{1}{N}\gamma_i' S_\Gamma^{-1}\sum_{j=1}^{N}(\gamma_j e_{jt}-\overline{\gamma}e_{jt}-\gamma_j\overline{e}_t+\overline{\gamma e}_t) \\
&\quad -\frac{1}{N}\overline{\gamma}' S_\Gamma^{-1}\sum_{j=1}^{N}(\gamma_j e_{jt}-\overline{\gamma}e_{jt}-\gamma_j\overline{e}_t+\overline{\gamma e}_t) \\
&= \frac{1}{N}\left(\gamma_i-\overline{\gamma}\right)' S_\Gamma^{-1}\sum_{j=1}^{N}\gamma_j e_{jt}-\frac{1}{N}\left(\gamma_i-\overline{\gamma}\right)' S_\Gamma^{-1}\sum_{j=1}^{N}\overline{\gamma}e_{jt} \\
&= \frac{1}{N}\left(\gamma_i-\overline{\gamma}\right)' S_\Gamma^{-1}\sum_{j=1}^{N}(\gamma_j-\overline{\gamma})e_{jt}. \qquad (16)
\end{aligned}$$

As the estimator for f_t in the restricted model is simply the cross sectional mean, we have that $(\hat{f}_t-f_t)\lambda_i=\overline{e}_t$ which together with (14) and (16) yields

$$(\tilde{F}_t-H'F_t)'H^{-1}L_i-(\hat{F}_t-H_R'F_t)'H_R^{-1}L_i=(\tilde{F}_tH^{-1}-\hat{F}_tH_R^{-1})'L_i=O_p\left(\frac{1}{\delta_{NT}^2}\right). \qquad (17)$$

For what concerns the loadings, we can use the asymptotic representation in Bai (2003) directly to write

$$F_t'H(\tilde{L}_i-H^{-1}L_i)=\frac{1}{T}F_t'\left(\frac{\mathbf{F}'\mathbf{F}}{T}\right)^{-1}\sum_{s=1}^{T}F_se_{is}+O_p\left(\frac{1}{\delta_{NT}^2}\right), \qquad (18)$$

for the unrestricted estimator and, using the fact that for the restricted estimator under the null $\hat{\lambda}_i-\lambda_i=0$, we can write compactly

$$\begin{aligned}
\frac{1}{T}g_t'\left(\frac{\mathbf{G}'\mathbf{G}}{T}\right)^{-1}\sum_{s=1}^{T}g_s\left(e_{is}-\overline{e}_t\right) &= \frac{1}{T}g_t'\left(\frac{\mathbf{G}'\mathbf{G}}{T}\right)^{-1}\sum_{s=1}^{T}g_se_{is} \\
&\quad +\frac{1}{NT}g_t'\left(\frac{\mathbf{G}'\mathbf{G}}{T}\right)^{-1}\sum_{s=1}^{T}\sum_{j=1}^{N}g_se_{js}+O_p\left(\frac{1}{\delta_{NT}^2}\right)
\end{aligned}$$

for the restricted estimator. The second term is of order $(NT)^{-1/2}$ and hence negligible since it has mean zero and variance equal to

$$\begin{aligned}
& \frac{1}{N^2T^2}E\left[\sum_{s=1}^{T}\sum_{j=1}^{N}\sum_{s'=1}^{T}\sum_{j'=1}^{N}g_t'\left(\frac{\mathbf{G}'\mathbf{G}}{T}\right)^{-1}g_sg_t'\left(\frac{\mathbf{G}'\mathbf{G}}{T}\right)^{-1}g_{s'}E(e_{js}e_{j's'})\right] \\
=\; & \frac{1}{N^2T^2}E\sum_{s=1}^{T}\sum_{j=1}^{N}\sum_{s'=1}^{T}\sum_{j'=1}^{N}g_t'\left(\frac{\mathbf{G}'\mathbf{G}}{T}\right)^{-1}g_sg_t' \\
& \times\left(\frac{\mathbf{G}'\mathbf{G}}{T}\right)^{-1}g_{s'}\sum_{h=1}^{N}\sum_{h'=1}^{N}\sum_{u=1}^{T}\sum_{u'=1}^{T}\theta_{jh}\theta_{j'h'}\phi_{us}\phi_{u's'}E(\varepsilon_{hu}\varepsilon_{h'u'}) \\
=\; & \frac{1}{N^2T^2}E\sum_{s=1}^{T}\sum_{j=1}^{N}\sum_{s'=1}^{T}\sum_{j'=1}^{N}g_t'\left(\frac{\mathbf{G}'\mathbf{G}}{T}\right)^{-1}g_sg_t'\left(\frac{\mathbf{G}'\mathbf{G}}{T}\right)^{-1}g_{s'}\sum_{h=1}^{N}\sum_{u=1}^{T}\theta_{jh}\theta_{j'h}\phi_{us}\phi_{u's} \\
=\; & O_p\left(\frac{1}{NT}\right)
\end{aligned}$$

by Assumption A1 and C and Lemmas 4.b and 4.d. We can therefore write

$$F_t' H_R(\hat{L}_i - H_R^{-1} L_i) = g_t' \bar{H}_R(\hat{\gamma}_i - \bar{H}_R^{-1}\gamma_i) = \frac{1}{T} g_t' \left(\frac{\mathbf{G}'\mathbf{G}}{T}\right)^{-1} \sum_{s=1}^{T} g_s e_{is} + O_p\left(\frac{1}{\delta_{NT}^2}\right). \tag{19}$$

Using (18) and (19), we can write the difference between the unrestricted and restricted estimated loadings as

$$\begin{aligned} F_t' H(\tilde{L}_i - H^{-1} L_i) - F_t' H_R(\hat{L}_i - H_R^{-1} L_i) &= \frac{1}{T} F_t' \left(\frac{\mathbf{F}'\mathbf{F}}{T}\right)^{-1} \sum_{s=1}^{T} F_s e_{is} \\ & \quad \frac{1}{T} g_t' \left(\frac{\mathbf{G}'\mathbf{G}}{T}\right)^{-1} \sum_{s=1}^{T} g_s e_{is} + O_p\left(\frac{1}{\delta_{NT}^2}\right) \\ &= \frac{1}{T} F_t' \mathbf{P}_T \sum_{s=1}^{T} F_s e_{is} + O_p\left(\frac{1}{\delta_{NT}^2}\right). \end{aligned} \tag{20}$$

Finally, the above calculations also imply that

$$(\tilde{F}_t - H' F_t)'(\tilde{L}_i - H^{-1} L_i) - (\hat{F}_t - H_R' F_t)'(\hat{L}_i - H_R^{-1} L_i) = O_p(\delta_{NT}^{-2}). \tag{21}$$

Then, the result follows by replacing (17), (20) and (21) into (12).

b.i) Start by writing

$$\sum_{i=1}^{N}\sum_{t=1}^{T} e_{it} r_{it} = \sum_{i=1}^{N}\sum_{t=1}^{T} [e_{it} r_{it} - E(e_{it} r_{it})] + \sum_{i=1}^{N}\sum_{t=1}^{T} E(e_{it} r_{it}). \tag{22}$$

Defining $\mathring{r}_{it} = \delta_{NT}^2 r_{it}$, we can write the first term of the right hand side of (22) as

$$\frac{1}{\delta_{NT}^2} \sum_{i=1}^{N}\sum_{t=1}^{T} [e_{it}\mathring{r}_{it} - E(e_{it}\mathring{r}_{it})]$$

where $\mathring{r}_{it} = O_p(1)$ and as a consequence,

$$\begin{aligned} O_p\left(\frac{1}{\delta_{NT}^2} \sum_{i=1}^{N}\sum_{t=1}^{T} [e_{it}\mathring{r}_{it} - E(e_{it}\mathring{r}_{it})]\right) &\leq O_p\left(\frac{1}{\delta_{NT}^2} \sum_{i=1}^{N}\sum_{t=1}^{T} [e_{it}^2 - E(e_{it}^2)]\right) \\ &= O\left(\frac{1}{\delta_{NT}^2}\right) O_p\left(\sqrt{NT}\right) = o(\sqrt{N}) \end{aligned}$$

where the first inequality comes from replacing $\mathring{r}_{it}$ with e_{it} whereas the last equality follows from $\sqrt{T}/N \to 0$ and $\sqrt{N}/T \to 0$. As for the second term of the right hand side of (22), we have to show that $N^{-1/2}\sum_{i=1}^{N}\sum_{t=1}^{T} E(e_{it} r_{it}) = o_p(1)$ for all the terms that compose r_{it}, which in practice involves writing the exact representations for $\tilde{F}_t - H' F_t$, $\hat{F}_t - H_R' F_t$, $(\tilde{L}_i - H^{-1} L_i)$ and $\hat{L}_i - H_R^{-1} L_i$. For the sake of brevity we will only show the argument for the terms involving $\tilde{L}_i - H^{-1} L_i$, noticing that the results for the other terms can be obtained with similar algebra.

Since $\tilde{\mathbf{L}}$ are the eigenvectors of $(NT)^{-1}X'X$, and $\tilde{\mathbf{F}}'\tilde{\mathbf{F}}/T$ is a diagonal matrix with the corresponding eigenvalues on the diagonal, $[(NT)^{-1}X'X]\tilde{\mathbf{L}} = \tilde{\mathbf{L}}(\tilde{\mathbf{F}}'\tilde{\mathbf{F}}/T)$, so that

$$\tilde{\mathbf{L}} = [(NT)^{-1}X'X]\tilde{\mathbf{L}}\left(\tilde{\mathbf{F}}'\tilde{\mathbf{F}}/T\right)^{-1}.$$

Then, by noticing that $X'X = \mathbf{LF}'\mathbf{FL}' + \mathbf{LF}'e + e\mathbf{FL}' + e'e$, we can write the $N \times r$ matrix of estimated loadings of the unrestricted model as

$$\tilde{\mathbf{L}} = \mathbf{L}\Sigma_{\mathbf{F}}\tilde{\Sigma}_{\mathbf{L}\tilde{\mathbf{L}}}\tilde{\Sigma}_{\mathbf{F}}^{-1} + \frac{\mathbf{LF}'e\tilde{\mathbf{L}}}{NT}\tilde{\Sigma}_{\mathbf{F}}^{-1} + \frac{e'\mathbf{F}}{T}\tilde{\Sigma}_{\mathbf{L}\tilde{\mathbf{L}}}\tilde{\Sigma}_{\mathbf{F}}^{-1} + \frac{e'e\tilde{\mathbf{L}}}{NT}\tilde{\Sigma}_{\mathbf{F}}^{-1}.$$

For the i'th row of $\mathbf{L}$ we have

$$\tilde{L}_i - H^{-1}L_i = \tilde{\Sigma}_{\mathbf{F}}^{-1}\tilde{\Sigma}_{\mathbf{L}\tilde{\mathbf{L}}}\frac{1}{T}\sum_{s=1}^{T}F_s e_{is} + \tilde{\Sigma}_{\mathbf{F}}^{-1}\frac{1}{NT}\sum_{j=1}^{N}\sum_{s=1}^{T}\tilde{L}_j e_{js}e_{is} + \tilde{\Sigma}_{\mathbf{F}}^{-1}\frac{1}{NT}\sum_{j=1}^{N}\sum_{s=1}^{T}\tilde{L}_j e_{js}F_s'L_i$$

where the first term delivers the asymptotic distribution of the test while the remaing two contribute to r_{it} through $F_t'H\left(\tilde{L}_i - H^{-1}L_i\right)$. So it remains to show that

$$E\left[\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\sum_{t=1}^{T}F_t'H\tilde{\Sigma}_{\mathbf{F}}^{-1}\frac{1}{NT}\sum_{j=1}^{N}\sum_{s=1}^{T}\tilde{L}_j e_{js}e_{is}e_{it}\right] = o_p(1) \tag{23}$$

and

$$E\left[\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\sum_{t=1}^{T}F_t'H\tilde{\Sigma}_{\mathbf{F}}^{-1}\frac{1}{NT}\sum_{j=1}^{N}\sum_{s=1}^{T}\tilde{L}_j e_{js}F_s'L_i e_{it}\right] = o_p(1). \tag{24}$$

As for (23), add and subtract $(NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{j=1}^{N}\sum_{s=1}^{T}F_t'L_j e_{js}e_{is}e_{it}$ and use the normalization $\tilde{\Sigma}_{\mathbf{F}}^{-1} = I_r$ to write

$$\begin{aligned} E\left[\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\sum_{j=1}^{N}\sum_{s=1}^{T}F_t'H\tilde{L}_j e_{js}e_{is}\right)e_{it}\right] &= E\left[\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{j=1}^{N}\sum_{s=1}^{T}F_t'L_j e_{js}e_{is}e_{it}\right. \\ &\left.+\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{j=1}^{N}\sum_{s=1}^{T}F_t'H(\tilde{L}_j - H^{-1}L_j)e_{js}e_{is}e_{it}\right] \end{aligned} \tag{25}$$

Replace the expressions for e_{js}, e_{is} and e_{it} in the first term of the right hand side of (25) to obtain

$$\begin{aligned} & E\left[\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{j=1}^{N}\sum_{s=1}^{T}F_t'L_j\sum_{u=1}^{T}\sum_{u'=1}^{T}\sum_{u''=1}^{T}\sum_{h=1}^{N}\sum_{h'=1}^{N}\sum_{h''=1}^{N}\theta_{jh}\theta_{ih'}\theta_{ih''}\phi_{us}\phi_{u's}\phi_{u''t}E\left(\varepsilon_{hu}\varepsilon_{h'u'}\varepsilon_{h''u''}|\mathbf{L}\right)\right] \\ \le\ & \frac{1}{N}\max_t E(||F_t||^2)\max_j E(||L_j||^2)\sum_{i=1}^{N}\sum_{h=1}^{N}\theta_{ih}^2\sum_{j=1}^{N}\theta_{jh}\frac{1}{T}\sum_{u=1}^{T}\sum_{s=1}^{T}\phi_{us}^2\sum_{t=1}^{T}\phi_{ut} = O(1) \end{aligned}$$

by Assumption A2, B and C and Lemma 4.a, 4.b and 4.d. The second term of the right hand side of (25) (omitting H since $||H|| = O_p(1)$) is equal to

$$
\begin{aligned}
& E\left[\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{j=1}^{N}\sum_{s=1}^{T} F_t'(\tilde{L}_j - H^{-1}L_j)e_{js}e_{is}e_{it}\right] \\
\leq \;& \max_t E||F_t||^2 \left[E\frac{1}{N}\sum_{j=1}^{N}\left\|\tilde{L}_j - H^{-1}L_j\right\|^2\right]^{1/2}\left[\frac{1}{NT^2}E\sum_{j=1}^{N}\left(\sum_{i=1}^{N}\sum_{s=1}^{T}\sum_{t=1}^{T} e_{js}e_{is}e_{it}\right)^2\right]^{1/2} \\
= \;& O\left(\frac{1}{\delta_{NT}}\right) O\left(\sqrt{T}\right) = o(\sqrt{N}).
\end{aligned}
$$

where the inequality follows from CS and the rates from Theorem 2 in Bai (2003) and the weak dependence of the errors, respectively. Notice that the $O\left(\delta_{NT}^{-1}\sqrt{T}\right)$ term either gives rise to a $O(1)$ term or to a $O(\sqrt{T}/\sqrt{N})$. The first one is clearly $o(\sqrt{N})$ as required, whereas the latter is $o(\sqrt{N})$ as long as $\sqrt{T}/N \to 0$, which is assumed.

As for (24), adding and subtracting $\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{j=1}^{N}\sum_{s=1}^{T} F_t'L_je_{js}F_s'L_ie_{it}$ and using again the normalization $\tilde{\Sigma}_{\mathbf{F}}^{-1} = I_r$ we obtain

$$
\begin{aligned}
E\left[\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{j=1}^{N}\sum_{s=1}^{T} F_t'H\tilde{L}_je_{js}F_s'L_ie_{it}\right] &= E\left[\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{j=1}^{N}\sum_{s=1}^{T} F_t'L_je_{js}F_s'L_ie_{it}\right] \\
&+ E\left[\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{j=1}^{N}\sum_{s=1}^{T} F_t'H(\tilde{L}_j - H^{-1}L_j)e_{js}F_s'L_ie_{it}\right].
\end{aligned}
\tag{26}
$$

The first term of the right hand side of (26) is

$$
\begin{aligned}
E\left[\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{j=1}^{N}\sum_{s=1}^{T} F_t'L_je_{js}F_s'L_ie_{it}\right] &= E\left[\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{j=1}^{N}\sum_{s=1}^{T} F_t'L_jF_s'L_iE\left(e_{it}e_{js}|\mathbf{L},\mathbf{F}\right)\right] \\
&\leq \frac{1}{NT}\left(\max_t E||F_t||^2 \max_j E||L_j||^2\right)^2 \\
&\quad \times \sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{j=1}^{N}\sum_{s=1}^{T} E\left(e_{it}e_{js}|\mathbf{L},\mathbf{F}\right) \\
&= O(1)
\end{aligned}
$$

by Assumptions A, B and C, and by noticing that $E\left(e_{it}e_{js}|\mathbf{L},\mathbf{F}\right) = \sum_{h=1}^{N}\sum_{u=1}^{T}\theta_{ih}\theta_{jh}\phi_{us}\phi_{ut}$ so that

$$
\begin{aligned}
\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{j=1}^{N}\sum_{s=1}^{T}\sum_{h=1}^{N}\sum_{u=1}^{T}\theta_{ih}\theta_{jh}\phi_{us}\phi_{ut} &= \sum_{h=1}^{N}\left(\sum_{i=1}^{N}\theta_{ih}\sum_{j=1}^{N}\theta_{jh}\right)\sum_{u=1}^{T}\left(\sum_{t=1}^{T}\phi_{ut}\sum_{s=1}^{T}\phi_{us}\right) \\
&\leq \frac{1}{NT}M^4 \times N \times T = O(1)
\end{aligned}
$$

by Lemmas 4.b and 4.d. The second term of the right hand side of (26) is also $O(1)$, since

$$\begin{aligned} E\left\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{j=1}^{N}\sum_{s=1}^{T}F_t'(\tilde{L}_j - H^{-1}L_j)e_{js}F_s'L_ie_{it}\right\| &\leq \left[\frac{1}{N}E\sum_{j=1}^{N}\left\|\tilde{L}_j - H^{-1}L_j\right\|^2\right]^{1/2} \\ &\times\left[\frac{1}{NT^2}E\sum_{j=1}^{N}\left\|\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}F_tF_s'L_ie_{js}e_{it}\right\|^2\right]^{1/2} \\ &= O\left(\frac{1}{\delta_{NT}}\right)O\left(\sqrt{N}\right) = o(\sqrt{N}), \end{aligned}$$

where the inequality follows from CS applied to the sum over j, and the rates from Theorem 2 in Bai (2003) and by noticing that

$$\begin{aligned} \frac{1}{NT^2}E\sum_{j=1}^{N}\left\|\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}F_tF_s'L_ie_{js}e_{it}\right\|^2 &\leq \max_{t,s,i}E||F_tF_s'L_i||^4 \\ &\times\frac{1}{NT^2}\sum_{j=1}^{N}E\left[\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{i'=1}^{N}\sum_{t'=1}^{T}\sum_{s'=1}^{T}e_{js}e_{js'}e_{it}e_{i't'}\right] \\ &\leq M\times O(N), \end{aligned}$$

by Assumptions A, B and C and by using Lemma 3 with $k = j, k' = j, k'' = i, k''' = i'$ and $v = s, v' = s', v'' = t, v''' = t'$ and then applying repeatedly Lemma 4. Tedious but otherwise straightforward calculations applied to the other elements of r_{it} show that those terms are all negligible.

b.ii) Use the CS inequality to write

$$\begin{aligned} \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\sum_{t=1}^{T}(\tilde{c}_{it} - c_{it})\,r_{it} &\leq \left[\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\sum_{t=1}^{T}(\tilde{c}_{it} - c_{it})^2\right]^{1/2}\left[\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\sum_{t=1}^{T}r_{it}^2\right]^{1/2} \\ &= \frac{1}{\sqrt{N}}O_p\left(\frac{\sqrt{NT}}{\delta_{NT}}\right)O_p\left(\frac{\sqrt{NT}}{\delta_{NT}^2}\right) = O_p\left(\frac{\sqrt{NT}}{\delta_{NT}^3}\right) = o(\sqrt{N}) \end{aligned}$$

where the rates follows from Theorem 3 in Bai (2003) and Lemma 1.a. Then the result follows from $\sqrt{T}/N \to 0$ and $\sqrt{N}/T \to 0$. □

Proof of Theorem 3

First, we restate Theorem 5.20 in White (2000) for convenience of the reader:

Lemma 5. *Let $\{X_i, i = 1, ...N\}$ with $N \to \infty$ be a sequence of scalars with mean $\mu_i = E(X_i) = 0$ and variance $\sigma_i^2 = var(X_i)$ such that $E|X_i|^{2+\epsilon} < \Delta < \infty$ for some $\epsilon > 0$ and for all i. Also, let $\{X_i\}$ have mixing coefficients $\alpha \to 0$ as $N \to \infty$. If $\overline{\sigma}_N^2 \equiv V\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}X_i\right) > c > 0$ for N sufficiently large, then:*

$$\sqrt{N}\sum_{i=1}^{N}\frac{X_i}{\overline{\sigma}_N} \xrightarrow{d} N(0,1).$$

To apply Lemma 5, we first construct the variables X_i as

$$X_i = \sum_{t=1}^{T} \left[Z_{it} - E(Z_{it}) \right],$$

where $Z_{it} = Z_{it}^{\mathcal{A}} + Z_{it}^{\mathcal{B}} - Z_{it}^{\mathcal{C}}$ with $Z_{it}^{\mathcal{A}} = 2e_{it} \left(\tilde{c}_{it} - \hat{c}_{it} \right)$, $Z_{it}^{\mathcal{B}} = \left(\tilde{c}_{it} - \hat{c}_{it} \right)^2$ and $Z_{it}^{\mathcal{C}} = 2 \left(\tilde{c}_{it} - \hat{c}_{it} \right) \left(\tilde{c}_{it} - c_{it} \right)$. Similarly, we also define $X_i = X_i^{\mathcal{A}} + X_i^{\mathcal{B}} - X_i^{\mathcal{C}}$, where

$$X_i^{\mathcal{A}} = 2 \sum_{t=1}^{T} e_{it} a_{it} - E\left(2 \sum_{t=1}^{T} e_{it} a_{it} \right), \quad X_i^{\mathcal{B}} = \sum_{t=1}^{T} a_{it}^2 - E\left(\sum_{t=1}^{T} a_{it}^2 \right),$$

and

$$X_i^{\mathcal{C}} = 2 \sum_{t=1}^{T} a_{it} \left(\tilde{c}_{it} - c_{it} \right) - E\left[2 \sum_{t=1}^{T} a_{it} \left(\tilde{c}_{it} - c_{it} \right) \right],$$

so that

$$\overline{\sigma}_N^2 = V\left(\frac{1}{\sqrt{N}} \sum_{i=1}^{N} X_i \right) = \frac{1}{N} V\left(\sum_{i=1}^{N} X_i \right) = \frac{1}{N} V\left\{ \sum_{i=1}^{N} \sum_{t=1}^{T} \left[Z_{it} - E(Z_{it}) \right] \right\}.$$

In terms of the mean and the variance of the Q_{NT} test statistic, notice that

$$\mu_{Q_{NT}} = \lim_{N,T \to \infty} \frac{1}{\sigma_\infty^2} \sum_{i=1}^{N} E(X_i) = \lim_{N,T \to \infty} \frac{1}{\sigma_\infty^2} \sum_{i=1}^{N} \sum_{t=1}^{T} E(Z_{it})$$

and

$$\sigma_{Q_{NT}}^2 = \lim_{N,T \to \infty} \frac{1}{\sigma_\infty^4} V\left(\sum_{i=1}^{N} X_i \right) = \lim_{N,T \to \infty} \frac{1}{\sigma_\infty^4} V\left(\sum_{i=1}^{N} \sum_{t=1}^{T} \left[Z_{it} - E(Z_{it}) \right] \right).$$

The array X_i has, therefore, mean $\mu_i = E(X_i) = 0$ by construction and variance $\sigma_i^2 \equiv V(X_i)$. The proof of Theorem 3 consists in the following steps: 1) Calculation of $\mu_{Q_{NT}}$ and $\sigma_{Q_{NT}}^2$, 2) Verifying the conditions for the CLT to hold, and 3) Verifying that the mixing condition holds.

Step 1: *Calculation of* $\mu_{Q_{NT}}$ *and* $\sigma_{Q_{NT}}^2$

Recalling that the error structure is

$$e_{it} = \sum_{u=1}^{T} \sum_{h=1}^{N} \theta_{ih} \varepsilon_{hu} \phi_{ut},$$

and defining $a_{it} = \tilde{c}_{it} - \hat{c}_{it}$, the mean $\mu_{Q_{NT}}$ is

$$\begin{aligned} \frac{1}{\sigma_\infty^2} \sum_{i=1}^{N} \sum_{t=1}^{T} E(Z_{it}) &= \frac{1}{\sigma_\infty^2} E\left[\sum_{i=1}^{N} \sum_{t=1}^{T} 2e_{it} a_{it} + \sum_{i=1}^{N} \sum_{t=1}^{T} a_{it}^2 - 2 \sum_{i=1}^{N} \sum_{t=1}^{T} a_{it} \left(\tilde{c}_{it} - c_{it} \right) \right] \\ &= \frac{1}{\sigma_\infty^2} E\left(\mathcal{A}_{NT} + \mathcal{B}_{NT} - \mathcal{C}_{NT} \right). \end{aligned}$$

Starting with the first term, Lemma 1 allows us to write

$$
\begin{aligned}
E\left(\mathcal{A}_{NT}\right) &= \frac{2}{T}E\left[\sum_{t=1}^{T}\sum_{i=1}^{N}\left(F_t'\mathbf{P}_T\sum_{s=1}^{T}F_s\right)E\left(e_{is}e_{it}\right|\mathbf{F})\right] \qquad (27)\\
&= \frac{2}{T}\sum_{i=1}^{N}\sum_{h=1}^{N}\theta_{ih}^2 E\left[\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{u=1}^{T}(F_t'\mathbf{P}_T F_s)\phi_{ut}\phi_{us}\right]\\
&= \frac{2}{T}tr(\Theta\Theta')E\left[\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{u=1}^{T}(F_t'\mathbf{P}_T F_s)\phi_{ut}\phi_{us}\right],
\end{aligned}
$$

where the first equality use Assumption C and the second equality follows from the definition of the trace. Regarding its variance, write

$$
E(\mathcal{A}_{NT}^2) = \frac{4}{T^2}E\left[\sum_{i=1}^{N}\sum_{i'=1}^{N}\sum_{t=1}^{T}\sum_{t'=1}^{T}\sum_{s=1}^{T}\sum_{s'=1}^{T}\left(F_t'\mathbf{P}_T F_s\right)\left(F_{t'}'\mathbf{P}_T F_{s'}\right)E\left(e_{it}e_{i't'}e_{is}e_{i's'}\right|\mathbf{F})\right]
$$

and invoke Lemma 3 with $v=t$, $v'=t'$, $v''=s$, $v'''=s'$, $k=i$, $k'=i'$, $k''=i$ and $k'''=i'$ to obtain

$$
\begin{aligned}
&\sum_{u=1}^{T}\sum_{u'=1}^{T}\sum_{h=1}^{N}\sum_{h'=1}^{N}\theta_{ih}\theta_{i'h'}\theta_{ih'}\theta_{i'h}\phi_{ut}\phi_{u't'}\phi_{u's}\phi_{us'} + \sum_{u=1}^{T}\sum_{u'\neq u}^{T}\sum_{h=1}^{N}\sum_{h'\neq h}^{N}\theta_{ih}^2\theta_{i'h'}^2\phi_{ut}\phi_{u't'}\phi_{us}\phi_{u's'} \qquad (28)\\
&+\sum_{u=1}^{T}\sum_{u''=1}^{T}\sum_{h=1}^{N}\sum_{h''=1}^{N}\theta_{ih}\theta_{i'h}\theta_{ih''}\theta_{i'h''}\phi_{ut}\phi_{ut'}\phi_{u''s}\phi_{u''s'} + \sum_{u=1}^{T}\sum_{h=1}^{N}\theta_{ih}^2\theta_{i'h}^2\phi_{ut}\phi_{ut'}\phi_{us}\phi_{us'}\kappa_{\varepsilon}.
\end{aligned}
$$

As for the square of the expectation, from (27) we have

$$
\begin{aligned}
E^2\left(\mathcal{A}_{NT}\right) &= \frac{4}{T^2}E_F\left[\sum_{i=1}^{N}\sum_{i'=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{t'=1}^{T}\sum_{s'=1}^{T}\left(F_t'\mathbf{P}_T F_s\right)\left(F_{t'}'\mathbf{P}_T F_{s'}\right)\times \right. \qquad (29)\\
&\left.\left(\sum_{h=1}^{N}\sum_{u=1}^{T}\sum_{h'=1}^{N}\sum_{u'=1}^{T}\theta_{ih}^2\theta_{i'h'}^2\phi_{ut}\phi_{us}\phi_{u't'}\phi_{u's'}\right)\right].
\end{aligned}
$$

We immediately notice that (29) simplifies with the second term of (28), so that

$$
\begin{aligned}
V(\mathcal{A}_{NT}) &= E_F\left[\frac{4}{T^2}\sum_{i=1}^{N}\sum_{i'=1}^{N}\sum_{t=1}^{T}\sum_{t'=1}^{T}\sum_{s=1}^{T}\sum_{s'=1}^{T}\left(F_t'\mathbf{P}_T F_s\right)\left(F_{t'}'\mathbf{P}_T F_{s'}\right)\right.\\
&\left. 2\left(\sum_{u=1}^{T}\sum_{u'=1}^{T}\sum_{h=1}^{N}\sum_{h'=1}^{N}\theta_{ih}\theta_{i'h'}\theta_{ih'}\theta_{i'h}\phi_{ut}\phi_{u't'}\phi_{u's}\phi_{us'}\right) + rem_{\mathcal{A}_{NT}}\right],
\end{aligned}
$$

where

$$
\begin{aligned}
rem_{\mathcal{A}_{NT}} &= \frac{4}{T^2}\sum_{t=1}^{T}\sum_{t'=1}^{T}\sum_{s=1}^{T}\sum_{s'=1}^{T}\sum_{u=1}^{T}\left(F_t'\mathbf{P}_T F_s\right)\left(F_{t'}'\mathbf{P}_T F_{s'}\right)\phi_{ut}\phi_{ut'}\phi_{us}\phi_{us'}\sum_{i=1}^{N}\sum_{h=1}^{N}\theta_{ih}^2\sum_{i'=1}^{N}\theta_{i'h}^2\kappa_{\varepsilon},\\
&\leq 4\kappa_{\varepsilon}M^2\times\frac{N}{T^2}\sum_{t=1}^{T}\sum_{t'=1}^{T}\sum_{s=1}^{T}\sum_{s'=1}^{T}\sum_{u=1}^{T}\left(F_t'\mathbf{P}_T F_s\right)\left(F_{t'}'\mathbf{P}_T F_{s'}\right)\phi_{ut}\phi_{ut'}\phi_{us}\phi_{us'}\\
&\leq 4\kappa_{\varepsilon}M^2\times\frac{N}{T^2}||\mathbf{P}_T||^2\max_{t,t',s,s'}E||F_t'F_sF_{t'}'F_{s'}||\sum_{u=1}^{T}\sum_{t=1}^{T}\phi_{ut}\sum_{t'=1}^{T}\phi_{ut'}\sum_{s=1}^{T}\phi_{us}\sum_{s'=1}^{T}\phi_{us'}\\
&= O\left(\frac{N}{T}\right)
\end{aligned}
$$

by Lemma 4.a, Lemma 4.b, Assumption A and the fact that the norm of $\mathbf{P}_T$ is bounded by Assumption B. After rescaling by $\sqrt{N}$ this term $o_p(1)$. This guarantees which are $O(N)$. As for (7), following analogous steps and using Lemma 1 again we can write

$$\begin{aligned} E\left(\mathcal{B}_{NT}\right) &= \frac{1}{T^2}\sum_{i=1}^{N}\sum_{h=1}^{N}\theta_{ih}^2 E\left[\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{s'=1}^{T}\sum_{u=1}^{T}\left(F_t'\mathbf{P}_T F_s\right)\left(F_t'\mathbf{P}_T F_{s'}\right)\phi_{us}\phi_{us'}\right] \\ &= \frac{1}{T^2}tr(\Theta\Theta')E\left[\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{s'=1}^{T}\sum_{u=1}^{T}\left(F_t'\mathbf{P}_T F_s\right)\left(F_t'\mathbf{P}_T F_{s'}\right)\phi_{us}\phi_{us'}\right] \end{aligned}$$

and

$$\begin{aligned} E\left(\mathcal{B}_{NT}^2\right) &= E\left[\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{t'=1}^{T}\sum_{i'=1}^{N}\left(\frac{1}{T}F_t'\mathbf{P}_T\sum_{s=1}^{T}F_s e_{is}\right)^2\left(\frac{1}{T}F_{t'}'\mathbf{P}_T\sum_{s''=1}^{T}F_{s''}e_{i's''}\right)^2\right] \\ &= \frac{1}{T^4}E_F\left[\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{t'=1}^{T}\sum_{i'=1}^{N}\sum_{s=1}^{T}\sum_{s'=1}^{T}\sum_{s''=1}^{T}\sum_{s'''=1}^{T}\left(F_t'\mathbf{P}_T F_s\right)\left(F_t'\mathbf{P}_T F_{s'}\right)\right. \\ &\quad \left.\times\left(F_{t'}'\mathbf{P}_T F_{s''}\right)\left(F_{t'}'\mathbf{P}_T F_{s'''}\right)E\left(e_{is}e_{is'}e_{i's''}e_{i's'''}\right|\mathbf{F})\right]. \end{aligned}$$

By Assumption B we can rewrite $E\left(e_{is}e_{is'}e_{i's''}e_{i's'''}\right|\mathbf{F})$ in terms of the *iid* shocks, and, by invoking Lemma 3 with $v=s$, $v'=s'$, $v''=s''$, $v'''=s'''$, $k=i$, $k'=i$, $k''=i'$ and $k'''=i'$, we obtain

$$\begin{aligned} &\sum_{u=1}^{T}\sum_{u'=1}^{T}\sum_{h=1}^{N}\sum_{h'=1}^{N}\theta_{ih}\theta_{ih'}\theta_{i'h}\theta_{i'h'}\phi_{us}\phi_{u's'}\phi_{u's''}\phi_{us'''}+\sum_{u=1}^{T}\sum_{u''=1}^{T}\sum_{h=1}^{N}\sum_{h''=1}^{N}\theta_{ih}^2\theta_{i'h''}^2\phi_{us}\phi_{us'}\phi_{u''s''}\phi_{u''s'''} \\ &+\sum_{u=1}^{T}\sum_{u'=1}^{T}\sum_{h=1}^{N}\sum_{h'=1}^{N}\theta_{ih}\theta_{ih'}\theta_{i'h}\theta_{i'h'}\phi_{us}\phi_{u's'}\phi_{us''}\phi_{u's'''}+\sum_{u=1}^{T}\sum_{h=1}^{N}\theta_{ih}^2\theta_{i'h}^2\phi_{us}\phi_{us'}\phi_{us''}\phi_{us'''}\kappa_{\varepsilon}. \end{aligned} \tag{30}$$

The square of the expectation is, instead, equal to

$$\begin{aligned} E^2(\mathcal{B}_{NT}) &= \frac{1}{T^4}E_F\left[\sum_{i=1}^{N}\sum_{i'=1}^{N}\sum_{t=1}^{T}\sum_{t'=1}^{T}\sum_{s=1}^{T}\sum_{s'=1}^{T}\sum_{s''=1}^{T}\sum_{s'''=1}^{T}\left(F_t'\mathbf{P}_T F_s\right)\left(F_t'\mathbf{P}_T F_{s'}\right)\right. \\ &\quad \left.\left(F_{t'}'\mathbf{P}_T F_{s''}\right)\left(F_{t'}'\mathbf{P}_T F_{s'''}\right)\sum_{u=1}^{T}\sum_{h=1}^{N}\sum_{u'=1}^{T}\sum_{h'=1}^{N}\theta_{ih}^2\theta_{i'h'}^2\phi_{us}\phi_{us'}\phi_{u's''}\phi_{u's'''}\right]. \end{aligned} \tag{31}$$

We immediately notice that (31) simplifies with the second term of (30), so that the variance is

$$\begin{aligned} V(\mathcal{B}_{NT}) &= E\left[\frac{1}{T^4}\sum_{i=1}^{N}\sum_{i'=1}^{N}\sum_{t=1}^{T}\sum_{t'=1}^{T}\sum_{s=1}^{T}\sum_{s'=1}^{T}\sum_{s''=1}^{T}\sum_{s'''=1}^{T}\left(F_t'\mathbf{P}_T F_s\right)\left(F_t'\mathbf{P}_T F_{s'}\right)\left(F_{t'}'\mathbf{P}_T F_{s''}\right)\left(F_{t'}'\mathbf{P}_T F_{s'''}\right)\right. \\ &\quad \left.2\left(\sum_{u=1}^{T}\sum_{u'=1}^{T}\sum_{h=1}^{N}\sum_{h'=1}^{N}\theta_{ih}\theta_{ih'}\theta_{i'h}\theta_{i'h'}\phi_{us}\phi_{u's'}\phi_{u's''}\phi_{us'''}\right)+rem_{\mathcal{B}_{NT}}\right], \end{aligned}$$

where

$$\begin{aligned} rem_{\mathcal{B}_{NT}} &= \frac{1}{T^4}\sum_{i=1}^{N}\sum_{i'=1}^{N}\sum_{t=1}^{T}\sum_{t'=1}^{T}\sum_{s=1}^{T}\sum_{s'=1}^{T}\sum_{s''=1}^{T}\sum_{s'''=1}^{T}\sum_{u=1}^{T}\sum_{h=1}^{N}\left(F_t'\mathbf{P}_T F_s\right)\left(F_t'\mathbf{P}_T F_{s'}\right) \\ &\quad \times\left(F_{t'}'\mathbf{P}_T F_{s''}\right)\left(F_{t'}'\mathbf{P}_T F_{s'''}\right)\theta_{ih}^2\theta_{i'h}^2\phi_{us}\phi_{us'}\phi_{us''}\phi_{us'''}\kappa_{\varepsilon} \\ &= O\left(\frac{N}{T}\right) \end{aligned}$$

where the last equality follows from Lemma 4.a, Lemma 4.b and Assumption A. Next, to calculate mean and variance of $\mathcal{C}_{NT}$ we use again Lemma 1 to write

$$
\begin{aligned}
\mathcal{C}_{NT} &= 2\sum_{t=1}^{T}\sum_{i=1}^{N}(\tilde{c}_{it}-\hat{c}_{it})(\tilde{c}_{it}-c_{it}) \\
&= \sum_{t=1}^{T}\sum_{i=1}^{N}\left(\frac{1}{T}F_t'\mathbf{P}_T\sum_{s=1}^{T}F_s e_{is}\right)\left(\frac{1}{N}L_i S_L^{-1}\sum_{j=1}^{N}L_j e_{jt}+\frac{1}{T}F_t' S_F^{-1}\sum_{s'=1}^{T}F_{s'}e_{is'}\right)+O_p\left(\frac{1}{\delta_{NT}^2}\right)
\end{aligned}
$$

where we have used the expression for $\tilde{c}_{it}-c_{it}$ in Bai (2003), (see e.g. proof of Theorem 3). Similar calculations to the ones just shown imply that

$$
\begin{aligned}
E(\mathcal{C}_{NT}) &= \frac{2}{NT}\left[\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{s=1}^{T}\sum_{j=1}^{N}\sum_{u=1}^{T}\sum_{h=1}^{N}\left(F_t'\mathbf{P}_T F_s\right)\left(L_i S_L^{-1}L_j\right)\theta_{ih}\theta_{jh}\phi_{us}\phi_{ut}\right] \\
&\quad +\frac{2}{T^2}E\left[\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{s=1}^{T}\sum_{s'=1}^{T}\sum_{u=1}^{T}\sum_{h=1}^{N}\left(F_t'\mathbf{P}_T F_s\right)\left(F_t' S_F^{-1}F_{s'}\right)\theta_{ih}^2\phi_{us}\phi_{us'}\right].
\end{aligned}
$$

Regarding the variance, the expectation of the square involves computing the following four terms

$$
\mathcal{C}_I = E\left[\sum_{i=1}^{N}\sum_{i'=1}^{N}\sum_{t=1}^{T}\sum_{t'=1}^{T}4a_{it}a_{i't'}\left(\frac{1}{N}L_i S_L^{-1}\sum_{j=1}^{N}L_j e_{jt}\right)\left(\frac{1}{N}L_i S_L^{-1}\sum_{j'=1}^{N}L_j e_{j't'}\right)\right],
$$

$$
\mathcal{C}_{II} = E\left[\sum_{i=1}^{N}\sum_{i'=1}^{N}\sum_{t=1}^{T}\sum_{t'=1}^{T}4a_{it}a_{i't'}\left(\frac{1}{N}L_i S_L^{-1}\sum_{j=1}^{N}L_j e_{jt}\right)\left(\frac{1}{T}F_{t'}' S_F^{-1}\sum_{s''=1}^{T}F_{s''}e_{i's''}\right)\right],
$$

$$
\mathcal{C}_{III} = E\left[\sum_{i=1}^{N}\sum_{i'=1}^{N}\sum_{t=1}^{T}\sum_{t'=1}^{T}4a_{it}a_{i't'}\left(\frac{1}{T}F_t' S_F^{-1}\sum_{s''=1}^{T}F_{s''}e_{is''}\right)\left(\frac{1}{N}L_i S_L^{-1}\sum_{j=1}^{N}L_j e_{jt}\right)\right],
$$

and

$$
\mathcal{C}_{IV} = E\left[\sum_{i=1}^{N}\sum_{i'=1}^{N}\sum_{t=1}^{T}\sum_{t'=1}^{T}4a_{it}a_{i't'}\left(\frac{1}{T}F_t' S_F^{-1}\sum_{s''=1}^{T}F_{s''}e_{is''}\right)\left(\frac{1}{T}F_{t'}' S_F^{-1}\sum_{s'''=1}^{T}F_{s'''}e_{i's'''}\right)\right]
$$

where recall $a_{it}=T^{-1}F_t'\mathbf{P}_T\sum_{s=1}^{T}F_s e_{is}$. Then, using Lemma 3 with $v=s$, $v'=s'$, $v''=t$, $v'''=t'$, $k=i$, $k'=i'$, $k''=j$ and $k'''=j'$, and collecting terms we have that

$$
\begin{aligned}
\mathcal{C}_I &= \frac{4}{N^2T^2}E_F\left[\sum_{i=1}^{N}\sum_{i'=1}^{N}\sum_{j=1}^{N}\sum_{j'=1}^{N}\sum_{t=1}^{T}\sum_{t'=1}^{T}\sum_{s=1}^{T}\sum_{s'=1}^{T}\left(F_t'\mathbf{P}_T F_s\right)\left(F_{t'}'\mathbf{P}_T F_{s'}\right)(L_i S_L^{-1}L_j)(L_i S_L^{-1}L_j)\times\right. \\
&\left.\left(3\sum_{u=1}^{T}\sum_{u''=1}^{T}\sum_{h=1}^{N}\sum_{h''=1}^{N}\theta_{ih}\theta_{i'h}\theta_{jh''}\theta_{j'h''}\phi_{us}\phi_{us'}\phi_{u''t}\phi_{u''t'}+\sum_{u=1}^{T}\sum_{h=1}^{N}\theta_{ih}\theta_{i'h}\theta_{jh}\theta_{j'h}\phi_{us}\phi_{us'}\phi_{ut}\phi_{ut'}\kappa_\varepsilon\right)\right].
\end{aligned}
$$

Repeated application of Lemma 3 for the remaining cases delivers

$$\mathcal{C}_{II} = \frac{4}{NT^3}E_F\left[\sum_{i=1}^{N}\sum_{i'=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{t'=1}^{T}\sum_{s=1}^{T}\sum_{s'=1}^{T}\sum_{s''=1}^{T}\left(F_t'\mathbf{P}_TF_s\right)\left(F_{t'}'\mathbf{P}_TF_{s'}\right)(L_iS_L^{-1}L_j)(F_{t'}'S_F^{-1}F_{s''})\times\right.$$
$$\left(2\sum_{u=1}^{T}\sum_{u''=1}^{T}\sum_{h=1}^{N}\sum_{h''=1}^{N}\theta_{ih}\theta_{i'h}\theta_{jh''}\theta_{i'h''}\phi_{us}\phi_{us'}\phi_{u''t}\phi_{u''s''}\right.$$
$$\left.\left.+\sum_{u=1}^{T}\sum_{u'=1}^{T}\sum_{h=1}^{N}\sum_{h'=1}^{N}\theta_{ih}\theta_{i'h'}^2\theta_{jh}\phi_{us}\phi_{u's'}\phi_{ut}\phi_{u's''}+\sum_{u=1}^{T}\sum_{h=1}^{N}\theta_{ih}\theta_{i'h}^2\theta_{jh}\phi_{us}\phi_{us'}\phi_{ut}\phi_{us''}\kappa_\varepsilon\right)\right],$$

and a similar expression for $\mathcal{C}_{III}$, and

$$\mathcal{C}_{IV} = \frac{4}{T^4}\sum_{i=1}^{N}\sum_{i'=1}^{N}\sum_{t=1}^{T}\sum_{t'=1}^{T}\sum_{s=1}^{T}\sum_{s'=1}^{T}\sum_{s''=1}^{T}\sum_{s'''=1}^{T}\left(F_t'\mathbf{P}_TF_s\right)\left(F_{t'}'\mathbf{P}_TF_{s'}\right)\left(F_t'S_F^{-1}F_{s''}\right)\left(F_{t'}'S_F^{-1}F_{s'''}\right)$$
$$\left[2\sum_{u=1}^{T}\sum_{u''=1}^{T}\sum_{h=1}^{N}\sum_{h''=1}^{N}\theta_{ih}\theta_{i'h''}\theta_{ih''}\theta_{i'h}\phi_{us}\phi_{u''s'}\phi_{u''s''}\phi_{us'''}\right.$$
$$\left.+\sum_{u=1}^{T}\sum_{u'=1}^{T}\sum_{h=1}^{N}\sum_{h'=1}^{N}\theta_{ih}^2\theta_{i'h'}^2\phi_{us}\phi_{u's'}\phi_{us''}\phi_{u's'''}+\sum_{u=1}^{T}\sum_{h=1}^{N}\theta_{ih}^2\theta_{i'h}^2\phi_{us}\phi_{us'}\phi_{us''}\phi_{us'''}\kappa_\varepsilon\right].$$

The square of the expectation is

$$\begin{aligned}E^2\left(\mathcal{C}_{NT}\right) = {} & \frac{4}{N^2T^2}\sum_{i=1}^{N}\sum_{i'=1}^{N}\sum_{t=1}^{T}\sum_{t'=1}^{T}\sum_{s=1}^{T}\sum_{s'=1}^{T}\sum_{j=1}^{N}\sum_{j'=1}^{N}\left(F_t'\mathbf{P}_TF_s\right)\left(F_{t'}'\mathbf{P}_TF_{s'}\right) \\ & \times\left(L_iS_L^{-1}L_j\right)\left(L_iS_L^{-1}L_j\right)\sum_{u=1}^{T}\sum_{u'=1}^{T}\sum_{h=1}^{N}\sum_{h'=1}^{N}\theta_{ih}\theta_{i'h'}\theta_{jh}\theta_{j'h'}\phi_{us}\phi_{u's'}\phi_{ut}\phi_{u't'} \\ & +\frac{4}{T^4}\sum_{i=1}^{N}\sum_{i'=1}^{N}\sum_{t=1}^{T}\sum_{t'=1}^{T}\sum_{s=1}^{T}\sum_{s'=1}^{T}\sum_{s''=1}^{T}\sum_{s'''=1}^{T}\left(F_t'\mathbf{P}_TF_s\right)\left(F_{t'}'\mathbf{P}_TF_{s'}\right) \\ & \times\left(F_t'S_F^{-1}F_{s''}\right)\left(F_{t'}'S_F^{-1}F_{s'''}\right)\sum_{u=1}^{T}\sum_{u'=1}^{T}\sum_{h=1}^{N}\sum_{h'=1}^{N}\theta_{ih}^2\theta_{i'h'}^2\phi_{us}\phi_{us'}\phi_{u's''}\phi_{u's'''} \\ & +\frac{8}{NT^3}\sum_{i=1}^{N}\sum_{i'=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{t'=1}^{T}\sum_{s=1}^{T}\sum_{s'=1}^{T}\sum_{s''=1}^{T}\left(F_t'\mathbf{P}_TF_s\right)\left(F_{t'}'\mathbf{P}_TF_{s'}\right) \\ & \times\left(L_iS_L^{-1}L_j\right)\left(F_{t'}'S_F^{-1}F_{s'''}\right)\sum_{u=1}^{T}\sum_{h=1}^{N}\sum_{u'=1}^{T}\sum_{h'=1}^{N}\theta_{ih}\theta_{jh}\phi_{us}\phi_{ut}\theta_{i'h'}^2\phi_{u's''}\phi_{u's'''}.\end{aligned} \tag{32}$$

Now, notice that the first term of (32) cancels out with one third of the first term of $\mathcal{C}_I$, the second term of (32) cancels out with the second term of $\mathcal{C}_{IV}$ while, finally, the third term of (32) cancels with the second term of $\mathcal{C}_{II}$ (and also $\mathcal{C}_{III}$). By repeated application of Lemma 4 and assumption A1, we see that the remaining terms of $\mathcal{C}_I$, $\mathcal{C}_{II}$ and $\mathcal{C}_{III}$ are negligible. The variance is then given by the first term of $\mathcal{C}_{IV}$, that is

$$V(\mathcal{C}_{NT}) = \frac{8}{T^4}\sum_{i=1}^{N}\sum_{i'=1}^{N}\sum_{t=1}^{T}\sum_{t'=1}^{T}\sum_{s=1}^{T}\sum_{s'=1}^{T}\sum_{s''=1}^{T}\sum_{s'''=1}^{T}\left(F_t'\mathbf{P}_T F_s\right)\left(F_{t'}'\mathbf{P}_T F_{s'}\right)\left(F_t' S_F^{-1} F_{s''}\right)\left(F_{t'}' S_F^{-1} F_{s'''}\right)$$
$$\sum_{u=1}^{T}\sum_{u''=1}^{T}\sum_{h=1}^{N}\sum_{h''=1}^{N}\theta_{ih}\theta_{i'h''}\theta_{ih''}\theta_{i'h}\phi_{us}\phi_{u''s'}\phi_{u''s''}\phi_{us'''} + rem_{\mathcal{C}},$$

where, again by Lemma 4.a, Lemma 4.b and Assumption A,

$$rem_{\mathcal{C}_{NT}} = \frac{4}{T^4}\sum_{i=1}^{N}\sum_{i'=1}^{N}\sum_{t=1}^{T}\sum_{t'=1}^{T}\sum_{s=1}^{T}\sum_{s'=1}^{T}\sum_{s''=1}^{T}\sum_{s'''=1}^{T}\left(F_t'\mathbf{P}_T F_s\right)\left(F_{t'}'\mathbf{P}_T F_{s'}\right)$$
$$\times\left(F_t' S_F^{-1} F_{s''}\right)\left(F_{t'}' S_F^{-1} F_{s'''}\right)\sum_{u=1}^{T}\sum_{h=1}^{N}\theta_{ih}^2\theta_{i'h}^2\phi_{us}\phi_{us'}\phi_{us''}\phi_{us'''}\kappa_\varepsilon = O(N/T).$$

As for the covariance terms, tedious but otherwise straightforward computations yield

$$Cov(\mathcal{A}_{NT},\mathcal{B}_{NT}) = \frac{4}{T^3}E_F\left[\sum_{i=1}^{N}\sum_{i'=1}^{N}\sum_{t=1}^{T}\sum_{t'=1}^{T}\sum_{s=1}^{T}\sum_{s'=1}^{T}\sum_{s''=1}^{T}\left(F_t'\mathbf{P}_T F_s\right)\left(F_{t'}'\mathbf{P}_T F_{s'}\right)\left(F_{t'}'\mathbf{P}_T F_{s''}\right)\times\right.$$
$$\left.\sum_{u=1}^{T}\sum_{u'=1}^{T}\sum_{h=1}^{N}\sum_{h'=1}^{N}\theta_{ih}\theta_{i'h'}\theta_{i'h}\theta_{ih'}\phi_{us}\phi_{u's'}\phi_{us''}\phi_{u't}\right] + rem_{\mathcal{AB}},$$

$$Cov(\mathcal{A}_{NT},\mathcal{C}_{NT}) = \frac{8}{NT^2}E_F\left[\sum_{i=1}^{N}\sum_{i'=1}^{N}\sum_{t=1}^{T}\sum_{t'=1}^{T}\sum_{s=1}^{T}\sum_{s'=1}^{T}\sum_{s''=1}^{T}\left(F_t'\mathbf{P}_T F_s\right)\left(F_{t'}'\mathbf{P}_T F_{s'}\right)\left(F_{t'}' S_F^{-1} F_{s''}\right)\times\right.$$
$$\left.\sum_{u=1}^{T}\sum_{u'=1}^{T}\sum_{h=1}^{N}\sum_{h'=1}^{N}\theta_{ih}\theta_{i'h'}\theta_{i'h}\theta_{ih'}\phi_{us}\phi_{u's'}\phi_{us''}\phi_{u't}\right] + rem_{\mathcal{AC}},$$

and

$$Cov(\mathcal{B}_{NT},\mathcal{C}_{NT}) = \frac{4}{T^4}E_F\left[\sum_{i=1}^{N}\sum_{i'=1}^{N}\sum_{t=1}^{T}\sum_{t'=1}^{T}\sum_{s=1}^{T}\sum_{s'=1}^{T}\sum_{s''=1}^{T}\sum_{s'''=1}^{T}\left(F_t'\mathbf{P}_T F_s\right)\left(F_t'\mathbf{P}_T F_{s'}\right)\left(F_{t'}'\mathbf{P}_T F_{s''}\right)\right.$$
$$\left.\times\left(F_{t'}' S_F^{-1} F_{s'''}\right)\sum_{u=1}^{T}\sum_{u''=1}^{T}\sum_{h=1}^{N}\sum_{h''=1}^{N}\theta_{ih}\theta_{i'h''}\theta_{ih''}\theta_{i'h}\phi_{us}\phi_{u''s'}\phi_{u''s''}\phi_{us'''}\right] + rem_{\mathcal{BC}}.$$

where $rem_{\mathcal{AB}}$, $rem_{\mathcal{AC}}$ and $rem_{\mathcal{BC}}$ are $O(N/T)$. Finally, collect terms and apply Lemma 2.f and Lemma 2.g to recognize that $V(\mathcal{A}_{NT}) + V(\mathcal{C}_{NT}) = 2Cov(\mathcal{A}_{NT},\mathcal{C}_{NT})$ which, together with $Cov(\mathcal{A}_{NT},\mathcal{B}_{NT}) = Cov(\mathcal{B}_{NT},\mathcal{C}_{NT})$ imply the only term relevant for $\sigma_{Q_{NT}}^2$ coincides with $V(\mathcal{B}_{NT})$. Then, applying Lemma 2.e to $V(\mathcal{B}_{NT})$, we can write

$$\sigma_{Q_{NT}}^2 = \frac{1}{\sigma_\infty^4}V\left(\sum_{i=1}^{N}X_i\right) = E\left[\frac{2}{T^2}\sum_{i=1}^{N}\sum_{i'=1}^{N}\sum_{s=1}^{T}\sum_{s'=1}^{T}\sum_{s''=1}^{T}\sum_{s'''=1}^{T}tr\left[\left(\mathbf{P}_T F_{s'}F_{s''}'\right)\left(\mathbf{P}_T F_{s'''}F_s'\right)\right]\right.$$
$$\left.\left(\sum_{u=1}^{T}\sum_{u'=1}^{T}\sum_{h=1}^{N}\sum_{h'=1}^{N}\theta_{ih}\theta_{ih'}\theta_{i'h}\theta_{i'h'}\phi_{us}\phi_{u's'}\phi_{u's''}\phi_{us'''}\right)\right]$$
$$= \frac{2}{\sigma_\infty^4 T^2}tr\left[(\Theta\Theta')^2\right]tr\sum_{s=1}^{T}\sum_{s'=1}^{T}\sum_{s''=1}^{T}\sum_{s'''=1}^{T}\sum_{u=1}^{T}\sum_{u'=1}^{T}\left(\mathbf{P}_T F_{s'}F_{s''}'\right)\left(\mathbf{P}_T F_{s'''}F_s'\right)\phi_{us}\phi_{u's'}\phi_{u's''}\phi_{us'''}$$
$$= \frac{2}{\sigma_\infty^4 T^2}tr\left[(\Theta\Theta')^2\right]E\left[tr(\mathbf{P}_T\mathbf{F}'\Phi'\Phi\mathbf{F})^2\right] = 2N\times\frac{N\times tr\left[(\Theta\Theta')^2\right]}{\left[tr(\Theta\Theta')\right]^2}\times E\left[\frac{tr(\mathbf{P}_T\mathbf{F}'\Phi'\Phi\mathbf{F})^2}{\left[tr(\Phi'\Phi)\right]^2}\right].$$

Consequently, we have

$$\overline{\sigma}_N^2 = \frac{\sigma_\infty^4}{N}\sigma_{Q_{NT}}^2 = \frac{2}{NT^2} tr\left[(\Theta\Theta')^2\right] E\left[tr(\mathbf{P}_T\mathbf{F}'\Phi'\Phi\mathbf{F})^2\right] = O(1).$$

As for the mean, applying Lemma 2.a to the sum $\sum_{t=1}^T \mathbf{P}_T F_t' F_t$,

$$\begin{aligned}
\mu_{Q_{NT}} &= \frac{1}{\sigma_\infty^2}\sum_{i=1}^N\sum_{t=1}^T E(Z_{it}) = \frac{1}{\sigma_\infty^2}\left[E(\mathcal{A}_{NT}) + E(\mathcal{B}_{NT}) - E(\mathcal{C}_{NT})\right] \\
&= \left(\frac{NT}{tr(\Phi'\Phi)tr(\Theta\Theta')}\right)\frac{1}{T^2}tr(\Theta\Theta')E\left[tr\sum_{t=1}^T\sum_{s=1}^T\sum_{s'=1}^T\sum_{u=1}^T\left(F_t'\mathbf{P}_T F_t'\right)\left(F_s\mathbf{P}_T F_{s'}\right)\phi_{us}\phi_{us'}\right] \\
&= N \times \frac{E\left[tr(\mathbf{P}_T\mathbf{F}'\Phi'\Phi\mathbf{F})\right]}{tr(\Phi'\Phi)}.
\end{aligned}$$

Step 2: Verifying the conditions for the CLT hold

To check that the condition $E|X_i|^{2+\epsilon} < \Delta < \infty$ for some $\epsilon > 0$ and for all i holds, first notice that, since $|a+b+c|^{2+\epsilon} \leq 9\left(|a|^{2+\epsilon} + |b|^{2+\epsilon} + |c|^{2+\epsilon}\right)$,

$$E\left|X_i\right|^{2+\epsilon} = E\left|X_i^{\mathcal{A}} + X_i^{\mathcal{B}} - X_i^{\mathcal{C}}\right|^{2+\epsilon} \leq 9\left[E\left|X_i^{\mathcal{A}}\right|^{2+\epsilon} + E|X_i^{\mathcal{B}}|^{2+\epsilon} + E|X_i^{\mathcal{C}}|^{2+\epsilon}\right].$$

Then, consider $E\left|X_i^{\mathcal{A}}\right|^{2+\epsilon}$ and write

$$\begin{aligned}
E|X_i^{\mathcal{A}}|^{2+\varepsilon} &= E\left|2\sum_{t=1}^T e_{it}a_{it} - E\left[2\sum_{t=1}^T e_{it}a_{it}\right]\right|^{2+\varepsilon} \qquad (33)\\
&\leq 2^{2+\epsilon}\left[E\left|2\sum_{t=1}^T e_{it}a_{it}\right|^{2+\epsilon} + E\left|-E\left[2\sum_{t=1}^T e_{it}a_{it}\right]\right|^{2+\epsilon}\right].
\end{aligned}$$

The second term of (33) is $O(1)$ since

$$\begin{aligned}
E\left[2\sum_{t=1}^T e_{it}a_{it}\right] &= E(X_i^{\mathcal{A}}) = E\left[\frac{2}{T}\sum_{s=1}^T\sum_{t=1}^T\left(F_t'\mathbf{P}_T F_s\right)E(e_{it}e_{is}|\mathbf{F})\right] \\
&= E\left[\frac{2}{T}\sum_{s=1}^T\sum_{t=1}^T\left(F_t'\mathbf{P}_T F_s\right)\sum_{u=1}^T\sum_{u'=1}^T\sum_{h=1}^N\sum_{h'=1}^N E(\theta_{ih}\theta_{ih'}\phi_{us}\phi_{u't}\varepsilon_{hu}\varepsilon_{h'u'}|\mathbf{F})\right] \\
&= E\left[\frac{2}{T}\sum_{s=1}^T\sum_{t=1}^T\left(F_t'\mathbf{P}_T F_s\right)\sum_{u=1}^T\sum_{h=1}^N\theta_{ih}^2\phi_{us}\phi_{ut}\right] \\
&\leq \frac{2M^2}{T}E\left[\sum_{s=1}^T\sum_{t=1}^T\left(F_t'\mathbf{P}_T F_s\right)\right] = O(1)
\end{aligned}$$

by assumption A and Lemma 4.a and 4.c. Regarding the first term of (33), define $q = 2+\epsilon$ and p its Hölder conjugate (i.e. such that $1/p + 1/q = 1$) and apply Hölder's inequality to obtain

$$E\left[\sum_{t=1}^T e_{it}a_{it}\right]^q \leq E\left[\sum_{t=1}^T|e_{it}a_{it}|^q\right] \leq E\left[\sum_{t=1}^T|e_{it}a_{it}|^4\right]$$

since $q < 4$. Next, focus on the inner expectation ot $E\left(e_{it}a_{it}\right)^4$, that is

$$\frac{1}{T^4}E\left(\sum_{s=1}^{T}\sum_{s'=1}^{T}\sum_{s''=1}^{T}\sum_{s'''=1}^{T}\left(F_t'\mathbf{P}_TF_s\right)\left(F_t'\mathbf{P}_TF_{s'}\right)\left(F_t'\mathbf{P}_TF_{s''}\right)\left(F_t'\mathbf{P}_TF_{s'''}\right)E\left(e_{is}e_{is'}e_{is''}e_{is'''}e_{it}^4\middle|\mathbf{F}\right)\right),$$

and show that this term is $O(T^{-1})$. Specifically, we have that

$$E\left(e_{is}e_{is'}e_{is''}e_{is'''}e_{it}^4\middle|\mathbf{F}\right)\leq\max_{i,t}E(e_{it}^4)E\left(e_{is}e_{is'}e_{is''}e_{is'''}\middle|\mathbf{F}\right),$$

where $\mathbf{F}=\{F_1,...,F_T\}$, and applying Lemma 3 with $k=k'=k''=k'''=i$, $v=s$, $v'=s'$, $v''=s''$ and $v'''=s'''$ we obtain

$$\begin{aligned}E\left(e_{is}e_{is'}e_{is''}e_{is'''}\middle|\mathbf{F}\right) &= \sum_{u=1}^{T}\sum_{u'=1}^{T}\sum_{h=1}^{N}\sum_{h'=1}^{N}\theta_{ih}^2\theta_{ih'}^2\phi_{us}\phi_{us'}\phi_{u's''}\phi_{u's'''}\\ &\quad+\sum_{u=1}^{T}\sum_{u'=1}^{T}\sum_{h=1}^{N}\sum_{h'=1}^{N}\theta_{ih}^2\theta_{ih'}^2\phi_{us}\phi_{u's'}\phi_{us''}\phi_{u's'''}\\ &\quad+\sum_{u=1}^{T}\sum_{u'=1}^{T}\sum_{h=1}^{N}\sum_{h'=1}^{N}\theta_{ih}^2\theta_{ih'}^2\phi_{us}\phi_{u's'}\phi_{u's''}\phi_{us'''}\\ &\quad+\sum_{u=1}^{T}\sum_{h=1}^{N}\theta_{ih}^4\phi_{us}\phi_{us'}\phi_{us''}\phi_{us'''}\kappa_\varepsilon.\end{aligned}$$

Denote the four terms obtained by adding the sums over s and s' and the factors as I, II, III and IV. Then, write term I as follows:

$$\begin{aligned}I &= \frac{1}{T^4}E\left[\sum_{s=1}^{T}\sum_{s'=1}^{T}\sum_{s''=1}^{T}\sum_{s'''=1}^{T}\left(F_t'\mathbf{P}_TF_s\right)\left(F_t'\mathbf{P}_TF_{s'}\right)\left(F_t'\mathbf{P}_TF_{s''}\right)\left(F_t'\mathbf{P}_TF_{s'''}\right)\right.\\ &\quad\left.\sum_{h=1}^{N}\theta_{ih}^2\sum_{h'=1}^{N}\theta_{ih'}^2\sum_{u=1}^{T}\phi_{us}\phi_{us'}\sum_{u'=1}^{T}\phi_{u's''}\phi_{u's'''}\right]\\ &\leq \frac{M^4}{T^4}E\left[\sum_{s=1}^{T}\sum_{s'=1}^{T}\sum_{s''=1}^{T}\sum_{s'''=1}^{T}\left(F_t'\mathbf{P}_TF_s\right)\left(F_t'\mathbf{P}_TF_{s'}\right)\left(F_t'\mathbf{P}_TF_{s''}\right)\left(F_t'\mathbf{P}_TF_{s'''}\right)\right]\end{aligned}$$

since $\sum_{h=1}^N\theta_{ih}^2\leq M=O(1)$ and similarly for $\sum_{h'=1}^N\theta_{ih'}^2$, whereas $\sum_{u=1}^T\phi_{us}\phi_{us'}$ and $\sum_{u'=1}^T\phi_{u's''}\phi_{u's'''}$ by lemmas 4.a and 4.c respectively. The remaining term only involves factors and can be written as

$$\begin{aligned}&\frac{1}{T^4}E\left[\left(F_t'\mathbf{P}_TF_t\right)\sum_{s=1}^{T}\sum_{s'=1}^{T}\sum_{s''=1}^{T}\sum_{s'''=1}^{T}\left(F_s'\mathbf{P}_TF_{s'}\right)\left(F_t'\mathbf{P}_TF_{s''}\right)\left(F_t'\mathbf{P}_TF_{s'''}\right)\right]\\ \leq\ &\frac{1}{T^4}||\mathbf{P}_T||^4E\sum_{s=1}^{T}\sum_{s'=1}^{T}\sum_{s''=1}^{T}\sum_{s'''=1}^{T}F_t'F_tF_s'F_{s'}F_t'F_{s''}F_t'F_{s'''}=O(T^{-1})\end{aligned}$$

by stationarity of F_t and assuming F_t has bounded eight moments, so that finally I is $O(T^{-1})$, implying $E\left[\sum_{t=1}^T|e_{it}a_{it}|^4\right]=O(1)$. The same steps applied to II, III and IV yield the same result. Similar algebra for $X_i^{\mathcal{B}}$ and $X_i^{\mathcal{C}}$ show that also $E|X_i^{\mathcal{C}}|^{2+\varepsilon}$ and $E|X_i^{\mathcal{C}}|^{2+\varepsilon}$ are $O(1)$ so that finally also $E|X_i|^{2+\varepsilon}$ is and the condition is verified.

Step 3: *Verifying that the mixing condition hold*

The second requirement for Lemma 5 is that the α-mixing coefficient go to zero as the sample size increases. Since ρ-mixing implies α-mixing (see e.g. Bradley (2005)), we show the former instead. To do so, for $-\infty \leq J \leq L \leq \infty$, define the σ-algebra $\mathcal{F}_J^L$ as

$$\mathcal{F}_J^L = \sigma(X_k, J \leq k \leq L).$$

For any $\sigma-$field $\mathcal{G}$, let $L_2(\mathcal{G})$ denote the space of square-integrable, $\mathcal{G}$-measurable real valued random variables. For each $m \geq 1$, define the maximal correlation coefficient as follows

$$\rho(m) = \sup |corr(f, g)|,\ f \in L_2(\mathcal{F}_{-\infty}^j),\ g \in L_2(\mathcal{F}_{j+m}^\infty)$$

where the sup is over all f, g and integer j. The sequence $\{X_i\}$ is ρ-mixing process if $\rho(m) \to 0$ as $m \to \infty$. Clearly, the mixing formulation relies on the ordering of the variables. Given that in many applications there is cross-sectional dependence but no natural ordering of the variables, here we assume that there is some permutation of the data for which $\{X_i\}$ is ρ-mixing. Since $E(X_i) = 0$, we check that $corr(X_i, X_j) = E(X_i X_j) / \sqrt{E(X_i^2)E(X_j^2)} \to 0$ as $|i - j| \to \infty$. We first show that $E(X_i^2) = O(1)$ for all i. To do so, recall X_i is equal to $X_i = X_i^{\mathcal{A}} + X_i^{\mathcal{B}} - X_i^{\mathcal{C}}$, so that $E(X_i^2) = V(X_i^{\mathcal{A}}) + V(X_i^{\mathcal{B}}) + V(X_i^{\mathcal{C}}) + 2E(X_i^{\mathcal{A}} X_i^{\mathcal{B}}) - 2E(X_i^{\mathcal{B}} X_i^{\mathcal{C}}) - 2E(X_i^{\mathcal{A}} X_i^{\mathcal{C}})$. Then, consider for instance the first term,

$$V(X_i^{\mathcal{A}}) = E\left[\left(2\sum_{t=1}^{T} e_{it}a_{it}\right)^2\right] - E^2\left[2\sum_{t=1}^{T} e_{it}a_{it}\right]$$

and

$$E\left[\left(2\sum_{t=1}^{T} e_{it}a_{it}\right)^2\right] = E\left[\frac{4}{T^2}\sum_{t=1}^{T}\sum_{t'=1}^{T}\sum_{s=1}^{T}\sum_{s'=1}^{T}\left(F_t'\mathbf{P}_T F_s\right)\left(F_{t'}'\mathbf{P}_T F_s\right) e_{it}e_{it'}e_{is}e_{is'}e_{it}e_{it'}\right] = O(1)$$

$$E^2\left[2\sum_{t=1}^{T} e_{it}a_{it}\right] = E^2\left[\frac{2}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\left(F_t'\mathbf{P}_T F_s\right) e_{it}a_{it}\right] = O(1)$$

by Lemma 3 and Assumption A. The same reasoning yields the same rates for the other terms of $E(X_i^2)$ so that finally $E(X_i^2) = O(1)$ as desired. As for the numerator of the correlation coefficient, the cross-expectations $E(X_i X_j)$ are equal, for a generic i and j, to

$$E(X_i X_j) = E\left[\left(\sum_{t=1}^{T} Z_{it} - E(Z_{it})\right)\left(\sum_{t'=1}^{T} Z_{jt'} - E(Z_{jt'})\right)\right] = \sum_{t=1}^{T}\sum_{t'=1}^{T} E(Z_{it}Z_{jt'}) - E(Z_{it})E(Z_{jt'}).$$

Since $Z_{it} = Z_{it}^{\mathcal{A}} + Z_{it}^{\mathcal{B}} - Z_{it}^{\mathcal{C}}$, computing $E(X_i X_j)$ involves several terms; specifically,

$$E(X_i X_j) = \sum_{a \in \{\mathcal{A},\mathcal{B},\mathcal{C}\}}\ \sum_{b \in \{\mathcal{A},\mathcal{B},\mathcal{C}\}}\ \sum_{t=1}^{T}\sum_{t'=1}^{T}\left[E(Z_{it}^a Z_{jt'}^b) - E(Z_{it}^a)E(Z_{jt'}^b)\right]$$

Consider

$$\begin{aligned}E(Z_{it}^{\mathcal{A}}Z_{jt'}^{\mathcal{A}})-E(Z_{it}^{\mathcal{A}})E(Z_{jt'}^{\mathcal{A}}) &= E\left[\frac{1}{T^2}\sum_{s=1}^{T}\sum_{s'=1}^{T}\left(F_t'\mathbf{P}_TF_s\right)\left(F_{t'}'\mathbf{P}_TF_{s'}\right)E\left(e_{is}e_{it}e_{js'}e_{jt'}\middle|\,\mathbf{F}\right)\right]\\ &\quad -E\left[\frac{1}{T}\sum_{s=1}^{T}\left(F_t'\mathbf{P}_TF_s\right)E\left(e_{is}e_{it}|\mathbf{F}\right)\right]E\left[\frac{1}{T}\sum_{s'=1}^{T}\left(F_{t'}'\mathbf{P}_TF_{s'}\right)E\left(e_{js'}e_{jt'}|\mathbf{F}\right)\right].\end{aligned}$$

Using Lemma 3 with $k=k'=i$, $k''=k'''=j$, $v=s$, $v'=t$, $v''=s'$ and $v'''=t'$, we have that $E(e_{is}e_{js'}e_{it}e_{jt'}\,|\,\mathbf{F})$ is equal to

$$\begin{aligned}&\sum_{u=1}^{T}\sum_{u'=1}^{T}\sum_{h=1}^{N}\sum_{h'=1}^{N}\theta_{ih}^2\theta_{jh'}^2\phi_{us}\phi_{ut}\phi_{u's'}\phi_{u't'}+\sum_{u=1}^{T}\sum_{u'=1}^{T}\sum_{h=1}^{N}\sum_{h'=1}^{N}\theta_{ih}\theta_{ih'}\theta_{jh}\theta_{jh'}\phi_{us}\phi_{u't}\phi_{us'}\phi_{u't'}\\ &+\sum_{u=1}^{T}\sum_{u'=1}^{T}\sum_{h=1}^{N}\sum_{h'=1}^{N}\theta_{ih}\theta_{ih'}\theta_{jh'}\theta_{jh}\phi_{us}\phi_{u't}\phi_{u's'}\phi_{ut'}+\sum_{u=1}^{T}\sum_{h=1}^{N}\theta_{ih}^2\theta_{jh}^2\phi_{us}\phi_{ut}\phi_{us'}\phi_{ut'}\kappa_{\varepsilon}.\end{aligned}$$

As for the cross product of expectations, since

$$E\left(e_{is}e_{it}|\mathbf{F}\right)=\sum_{u=1}^{T}\sum_{h=1}^{N}\sum_{u'=1}^{T}\sum_{h'=1}^{N}\theta_{ih}\theta_{ih'}\phi_{us}\phi_{u't}E(\varepsilon_{hu}\varepsilon_{h'u'})=\sum_{u=1}^{T}\sum_{h=1}^{N}\theta_{ih}^2\phi_{us}\phi_{ut},$$

we have that

$$E\left(e_{is}e_{it}|\mathbf{F}\right)E(e_{js'}e_{jt'}|\mathbf{F})=\sum_{u=1}^{T}\sum_{h=1}^{N}\sum_{u'=1}^{T}\sum_{h'=1}^{N}\theta_{ih}^2\theta_{jh'}^2\phi_{us}\phi_{u's'}\phi_{ut}\phi_{u't'}.$$

Hence, we can write

$$E(Z_{it}^{\mathcal{A}}Z_{jt'}^{\mathcal{A}})=I^{\mathcal{A}}+II^{\mathcal{A}}+III^{\mathcal{A}}+IV^{\mathcal{A}},$$

where

$$I^{\mathcal{A}}=E\left[\frac{1}{T^2}\sum_{s=1}^{T}\left(F_t'\mathbf{P}_TF_s\right)\sum_{s'=1}^{T}\left(F_{t'}'\mathbf{P}_TF_{s'}\right)\sum_{h=1}^{N}\theta_{ih}^2\sum_{h'=1}^{N}\theta_{jh'}^2\sum_{u=1}^{T}\sum_{u'=1}^{T}\phi_{us}\phi_{us'}\phi_{u't}\phi_{u't'}\right]$$

cancels with $E\left(e_{is}e_{it}|\mathbf{F}\right)E(e_{js'}e_{jt'}|\mathbf{F})$, which follows from interchanging indices u and u' in $\phi_{us'}\phi_{u't}$ in the term $\phi_{us}\phi_{us'}\phi_{u't}\phi_{u't'}$. Similarly,

$$\begin{aligned}II^{\mathcal{A}} &= E\left[\frac{1}{T^2}\sum_{s=1}^{T}\left(F_t'\mathbf{P}_TF_s\right)\sum_{s'=1}^{T}\left(F_{t'}'\mathbf{P}_TF_{s'}\right)\sum_{h=1}^{N}\theta_{ih}\theta_{jh}\sum_{h'=1}^{N}\theta_{ih'}\theta_{jh'}\sum_{u=1}^{T}\phi_{us}\phi_{ut}\sum_{u'=1}^{T}\phi_{u's'}\phi_{u't'}\right]\\ &= \frac{1}{T^2}E\left[\left(F_t'\mathbf{P}_TF_{t'}\right)\sum_{s=1}^{T}\sum_{s'=1}^{T}\left(F_s'\mathbf{P}_TF_{s'}\right)\sum_{u=1}^{T}\phi_{us}\phi_{ut}\sum_{u'=1}^{T}\phi_{u's'}\phi_{u't'}\right]\sum_{h=1}^{N}\theta_{ih}\theta_{jh}\sum_{h'=1}^{N}\theta_{ih'}\theta_{jh'}\end{aligned}$$

but

$$\begin{aligned}&\frac{1}{T^2}E\left[(F_t'\mathbf{P}_TF_{t'})\sum_{s=1}^{T}\sum_{s'=1}^{T}\left(F_s'\mathbf{P}_TF_{s'}\right)\sum_{u=1}^{T}\phi_{us}\phi_{ut}\sum_{u'=1}^{T}\phi_{u's'}\phi_{u't'}\right]\\ \le\ &\frac{1}{T^2}\max_{s,s'}E(F_t'\mathbf{P}_TF_{t'}F_s'\mathbf{P}_TF_{s'})\sum_{s=1}^{T}\sum_{u=1}^{T}\phi_{us}\phi_{ut}\sum_{s'=1}^{T}\sum_{u'=1}^{T}\phi_{u's'}\phi_{u't'}\\ =\ &O(1)\end{aligned}$$

by Lemma 4.c and Assumption A. Hence, the order of magnitude of $II^{\mathcal{A}}$ depends on

$$\sum_{h=1}^{N} \theta_{ih}\theta_{jh} \sum_{h'=1}^{N} \theta_{ih'}\theta_{jh'},$$

where $\sum_{h=1}^{N} \theta_{ih}\theta_{jh}$ is the element (i, j) of the matrix $\Theta\Theta'$. By Lemma 4.a, $\sum_{i=1}^{N} \theta_{ih}^2 = O(1)$ and $\sum_{h=1}^{N} \theta_{ih}^2 = O(1)$ which imply that the sum of the squares of each element in each column (and row) of $\Theta\Theta'$ is also bounded. Then, it must be the case that the smallest squared element goes to zero; otherwise, N times the minimum squared element would diverge to infinity and hence $\sum_{i=1}^{N} \theta_{ih}^2 \geq N \min_i \theta_{ih}^2$ contradicts Assumption B. Therefore, there must be a j such that $\theta_{jh} \to 0$; and for this j, the product $\theta_{ih}\theta_{jh} \to 0$ since θ_{ih} is bounded. As a consequence, the mixing condition is satisfied. The same reasoning applies to terms $III^{\mathcal{A}}$ and $IV^{\mathcal{A}}$. To see this, write

$$III^{\mathcal{A}} = E\left[\frac{1}{T^2}\sum_{s=1}^{T}\left(F_t'\mathbf{P}_T F_s\right)\sum_{s'=1}^{T}\left(F_{t'}'\mathbf{P}_T F_{s'}\right)\sum_{h=1}^{N}\theta_{ih}\theta_{jh}\sum_{h'=1}^{N}\theta_{ih'}\theta_{jh'}\sum_{u=1}^{T}\phi_{us}\phi_{ut'}\sum_{u'=1}^{T}\phi_{u's'}\phi_{u't}\right]$$

and

$$IV^{\mathcal{A}} = \frac{1}{T^2}\sum_{h=1}^{N}\theta_{ih}^2\theta_{jh}^2\left(F_t'\mathbf{P}_T F_{t'}\right)\sum_{s=1}^{T}\sum_{s'=1}^{T}\left(F_s'\mathbf{P}_T F_{s'}\right)\sum_{u=1}^{T}\phi_{us}\phi_{us'}\phi_{ut}\phi_{ut'}\kappa_{\varepsilon}.$$

and follow the same steps as for term $II^{\mathcal{A}}$. Similar algebra shows that also all terms involving $\mathcal{B}$ and $\mathcal{C}$ behave in the same way and, under a suitable ordering of the elements, the mixing condition $\rho(m) \to 0$ as $m = |i - j| \to \infty$ is satisfied for j and i sufficiently far away. □

Proof of Theorems 1 and 2

Theorems 1 and 2 are special cases of Theorem 3 imposing, respectively, $\Phi = I_T$ and $\Theta = I_N$ and $\Phi = I_T$. The expressions for mean and variance of the Q_{NT}^1 and Q_{NT}^2 statistics then readily follow by substituting the corresponding identity matrices in the asymptotic mean and variance of Q_{NT}^3. □