Granger Centre
for time series econometrics

Granger Centre Discussion Paper Series

Determining the dimension of factor structures in non-stationary large datasets

by

Matteo Barigozzi and Lorenzo Trapani

Granger Centre Discussion Paper No. 18/01

University of Nottingham
UK | CHINA | MALAYSIA

Determining the dimension of factor structures in non-stationary large datasets

Matteo Barigozzi *
Department of Statistics, LSE

Lorenzo Trapani
School of Economics, University of Nottingham

September 25, 2018

Abstract

We propose a procedure to determine the dimension of the common factor space in a large, possibly non-stationary, dataset. Our procedure is designed to determine whether there are (and how many) common factors (i) with linear trends, (ii) with stochastic trends, (iii) with no trends, i.e. stationary. Our analysis is based on the fact that the largest eigenvalues of a suitably scaled covariance matrix of the data (corresponding to the common factor part) diverge, as the dimension N of the dataset diverges, whilst the others stay bounded. Therefore, we propose a class of randomised test statistics for the null that the p-th eigenvalue diverges, based directly on the estimated eigenvalue. The tests only requires minimal assumptions on the data, and no restrictions on the relative rates of divergence of N and T are imposed. Monte Carlo evidence shows that our procedure has very good finite sample properties, clearly dominating competing approaches when no common factors are present. We illustrate our methodology through an application to US bond yields with different maturities observed over the last 30 years. A common linear trend and two common stochastic trends are found and identified as the classical level, slope and curvature factors.

Keywords: Common factors; Unit roots; Common trends; Randomised tests.

*We wish to thank Giuseppe Cavaliere and Laura Coroneo for helpful comments. We are grateful to the participants to the 1st Italian Workshop on Econometrics and Empirical Economics (Milan, 25-26 January, 2018); the Financial Econometrics Seminar at CREST (Paris, 15 February, 2018); and the Economics Seminar Series at Aarhus University (Aarhus, 1 March, 2018); the Workshop on Big Data in Financial Markets at Cambridge University (Cambridge, 24-25 May, 2018); the Workshop on Macroeconomic and Financial Time Series at Lancaster University (Lancaster, 31 May-1 June, 2018).

1 Introduction and main ideas

In this paper, we propose a methodology to estimate the dimension of the space spanned by the common (possibly non-stationary) factors in the large approximate factor model

$$X_t = \Lambda \mathcal{F}_t + u_t, \tag{1}$$

where $\mathcal{F}_t$ is the $r \times 1$ vector of common factors and Λ is an $N \times r$ matrix of factor loadings, with $r < \infty$. We will also make use of the scalar version of (1)

$$X_{i,t} = \lambda_i' \mathcal{F}_t + u_{i,t}, \tag{2}$$

with $1 \leq i \leq N$ and $1 \leq t \leq T$.

Although the relevant assumptions are detailed in the remainder of the paper, in (1) we are assuming that there are three possible categories of common factors in the vector $\mathcal{F}_t$: factors with a linear trend and an additional, either $I(1)$ or $I(0)$, zero mean component; pure, zero mean $I(1)$ factors with no trends; and, finally, stationary common factors. Each group may well have dimension zero, e.g. factors with linear trends may not be present, etc. We also assume, throughout the paper, that the idiosyncratic terms $u_{i,t}$ are $I(0)$ for each i. Based on the classification above, we develop a technique to estimate the number of common factors which have linear trends, and, separately, the ones which are zero mean $I(1)$. In particular, we use the eigenvalues of the second moment matrix of the data, checking whether they diverge to infinity as $\min(N, T) \to \infty$, due to the presence of common factors, or whether they are bounded. In order to construct tests for the asymptotic behaviour of the eigenvalues, (i) we derive bounds on their divergence rates, and (ii) based on those bounds we propose a randomisation procedure which produces a statistic for which we are able to derive the asymptotic behaviour under the null and the alternative hypotheses.

Determining the presence (or not), and the number, of non-stationary factors in (1) can be useful in a variety of applications. First, we can assess the presence of unit roots in a large panel X_t - see e.g. Moon and Perron (2004), Bai and Ng (2004), Bai and Kao (2006), Bai, Kao, and Ng (2009), Kapetanios, Pesaran, and Yamagata (2011), and Pesaran, Smith, and Yamagata (2013). Second, the presence of common $I(1)$ trends in (1) entails cross-unit cointegration among the components of X_t - see e.g. Stock and Watson (1988), Escribano and Peña (1994), Gengenbach, Palm, and Urbain (2009), and Zhang, Robinson, and Yao (2018). Indeed, if in (1) there are, say, r common $I(1)$ factors (common trends), then there are $(N - r)$ cointegration relations - we also refer to Onatski and Wang (2016) for an alternative approach to cointegration in large VARs, although in that paper r is allowed to diverge. Factor models with common $I(1)$ factors have also been emloyed in several empirical applications: see e.g. Zhang et al. (2018), who studies employment fluctuations across 60 industries in the US, Moon and Perron (2007), who apply (1) to a panel of interest rates at different maturities in the US and Canada, and Engel, Mark, and West (2015), who use (1) as part of a strategy to develop a forecasting technique applied to a panel of bilateral US dollar rates against 17 OECD countries. Panel models with linear trends have also been employed in the context of modelling macro-economic data (see Maciejowska (2010)), and US temperature data (see Chen and Wu (2018)). In all these applications, the first step of the analysis would be the determination of the number of non-stationary and stationary common factors.

Starting at least from Chamberlain and Rothschild (1983), the literature has developed a plethora of contributions to determine the number of common factors in a large panel. Most methodologies focus on the case of stationary datasets, and existing approaches can be broadly grouped into two categories. Several studies rely on setting a threshold for the eigenvalues of the covariance or of other second moment matrices of the $X_{i,t}$s - see Bai and Ng (2002), Hallin and Liška (2007), and Alessi, Barigozzi, and Capasso (2010). In addition to this, the literature has explored the possibility of using ratios of adjacent eigenvalues - see Onatski (2009, 2010), Lam and Yao (2012), and Ahn and Horenstein (2013). Although the two approaches have different merits, the rationale underpinning them is the same: in the presence of, say, r common factors, the largest r eigenvalues of the covariance matrix of the $X_{i,t}$s diverge to infinity as $\min(N,T) \to \infty$, whilst all the remaining eigenvalues stay bounded.

Fewer contributions are available to deal directly with factor models for non-stationary data. In particular, developing an inferential theory for Λ and $\mathcal{F}_t$ in a model similar to (1) has been paid significant attention by the statistical literature: examples include Bai and Ng (2004), Bai (2004), Peña and Poncela (2006), Zhang, Pan, and Gao (2017), and Zhang et al. (2018). More specifically, Bai (2004) develops the inferential theory and a criterion to estimate the number of common stationary and non-stationary factors; linear trends, however, are not considered, and the extension by Maciejowska (2010) develops the inferential theory for this case, but not a criterion to determine the presence of common factors with a linear trend. Zhang et al. (2018) propose a method based on the ratio of eigenvalues of a transformation of the long-run covariance matrix to find the number of $I(d)$ factors for $d \geq 0$. In addition to not considering linear trends, however, the theory developed therein also requires the constraint $\frac{N}{T^\kappa} \to c \in (0,\infty)$, for $\kappa \in \left(0, \frac{1}{2}\right)$, as $\min(N,T) \to \infty$. Finally, in a related contribution, Bai and Ng (2004) propose a method to assess the presence or not of stochastic trends in large panels. However, in their setup it is assumed that at least one stationary factor is always present, and the two-step nature of their approach may entail some efficiency loss.

Our paper fills the gaps mentioned above. Although the main arguments are laid out in the remainder of the paper, here we present a heuristic preview of how the procedure works.

To begin with, in the presence of linear trends, it can be expected that the sample second moment matrix of X_t will diverge as fast as T^3. Also, due to the well known eigenvalue separation property of large factor models, it can be expected that the eigenvalues corresponding to common factors should diverge as fast as N. This suggests considering the eigenvalues of $T^{-3}\sum_t X_t X_t'$ (denoted as, say, $\nu_1^{(p)}$) to decide between

$$\begin{cases} H_{0,1}^{(p)} : \nu_1^{(p)} \to \infty, \\ H_{A,1}^{(p)} : \nu_1^{(p)} < \infty, \end{cases}$$

as $\min(N,T) \to \infty$; the test can be carried out for $p = 1, 2, ...$, stopping as soon as the null is rejected. Similarly, considering the zero mean, $I(1)$ common factors, the Functional Central Limit Theorem (FCLT) suggests that the second moment matrix of X_t will diverge as fast as T^2, again with the eigenvalues corresponding to the common factors diverging as fast as N. Thus, one could study the eigenvalues of $T^{-2}\sum_t X_t X_t'$ (denoted as, say, $\nu_2^{(p)}$), and decide between

$$\begin{cases} H_{0,2}^{(p)} : \nu_2^{(p)} \to \infty, \\ H_{A,2}^{(p)} : \nu_2^{(p)} < \infty, \end{cases}$$

as $\min(N,T) \to \infty$, carrying out the test as above. The output of these two steps is an estimate of the number of common factors which have a linear trend and of those which are genuinely zero mean $I(1)$ processes, respectively. Note that, in both steps, if we reject the null-hypothesis when $p = 1$, we are in fact saying that there are no common factors. This approach could be complemented by using $T^{-1}\sum_t \Delta X_t \Delta X_t'$ and determining the number of total common factors as suggested in Trapani (2017), which would provide an indirect estimate of the number of common stationary factors.

From a technical point of view, the implementation of the algorithm described above presents one difficulty: we are unable to construct test statistics which converge to a distributional limit under the null hypotheses, and the best result we can obtain are rates. Thus, we base our tests on randomising the test statistic. This approach builds on an idea of Pearson (1950), and it has been exploited in numerous contexts - see e.g. Corradi and Swanson (2006), and Trapani (2017) in the context of factor models. A major advantage of this procedure is that only rates are needed, and these can be derived under quite general assumptions. In particular, we derive our rates (and, thus, we are able to apply our test) under no restrictions on the relative rates of divergence of N and T as they pass to infinity, which can be compared with the standard restriction that as $\min(N,T) \to \infty$, $\frac{N}{T} \to c \in (0,\infty)$, often assumed in random matrix theory (see also Onatski and Wang, 2016, where a similar restriction is needed); this entails that our procedure can be applied to virtually any dataset, being particularly useful when either dimension is much bigger than the other. Also, our theory requires milder restrictions on the finiteness of moments than other contributions in the literature (see e.g. Bai (2004)), and it allows for arbitrary levels of (weak) cross-correlation among the idiosyncratic errors $u_{i,t}$.

The remainder of the paper is organised as follows. In Section 2, we spell out the main assumptions and (in Section 2.2) we study the strong rates of convergence of the eigenvalues of various rescalings of the second moment matrix of X_t. The testing algorithm is presented in Section 3. Numerical evidence from simulations is in Section 4, where we also report an empirical illustration. Finally, Section 6 concludes. Additional numerical results are reported in Appendix A. Proofs and technical results are in Appendix B.

NOTATION. We define the Euclidean norm of a vector $a = [a_1, ..., a_n]$ as $\|a\| = \left(\sum_{i=1}^n a_i^2\right)^{1/2}$, and similarly for a matrix A; "a.s." stands for "almost surely"; $I_A(x)$ is the indicator function of a set A; finally, C_0, C_1, etc... denote positive, finite constants whose value may differ from line to line. Other relevant notation is introduced later on in the paper.

2 Theory

This section contains the relevant assumptions (Section 2.1), and the main results on the eigenvalues of various rescaled versions of the sample second moment matrix of X_t (Section 2.2).

2.1 Model and assumptions

Recall the scalar version of our model (2):

$$X_{i,t} = \lambda_i' \mathcal{F}_t + u_{i,t}, \quad 1 \leq i \leq N, \tag{3}$$

where λ_i and $\mathcal{F}_t$ are $r \times 1$ vectors. We begin with a representation result which, essentially, states that the number of common factors with a linear trend (and, possibly, further components which may be $I(0)$ or $I(1)$) can be either zero - no common factors with linear trends - or 1. This result is originally due to Maciejowska (2010), and we report it hereafter, as a lemma, for convenience. We assume that

$$\mathcal{F}_t = A(d_1 t) + B\psi_t, \tag{4}$$

where A is a non-zero $r \times 1$ vector, B an $r \times r$ matrix; finally, d_1 is a dummy variable, which has the purpose to entertain the possibility that there are linear trends or not, according as $d_1 = 1$ or 0, respectively. As far as the r-dimensional vector ψ_t is concerned, its components are allowed to be a mixture of $I(0)$ and $I(1)$ processes, with no linear trends.

We consider the following assumption, which ensures that the $\mathcal{F}_t$s are fully identified.

Assumption 1. *It holds that: (i) A is non-zero; (ii) rank $(B) = r$; (iii) the vector ψ_t can be rearranged and partitioned as $[\psi'_{at}, \psi'_{bt}]'$, where $\psi_{at} \sim I(1)$ has dimension $r_2 + d_2$ and $\psi_{bt} \sim I(0)$ has dimension $r_3 + (1 - d_2)$, where d_2 is a dummy variable.*

By part *(ii)* of Assumption 1, B has full rank, which ensures the identification of the vector $\mathcal{F}_t$ irrespective of whether there is a trend or not. When there are trends, that is when $d_1 = 1$, part *(i)* of the assumption ensures that they do have an impact on $\mathcal{F}_t$. Finally, by part *(iii)* there could be both $I(1)$ and $I(0)$ factors in the vector ψ_t, sorted in no particular order.

Lemma 1. *Under Assumption 1, model (3) can be equivalently represented as*

$$X_{i,t} = \lambda_i^{(1)} f_t^{(1)} + \lambda_i^{(2)\prime} f_t^{(2)} + \lambda_i^{(3)\prime} f_t^{(3)} + u_{i,t}, \quad 1 \leq i \leq N, \tag{5}$$

where $\lambda_i^{(1)}$ and $f_t^{(1)}$ are $r_1 \times 1$ with $0 \leq r_1 \leq 1$, $\lambda_i^{(2)}$ and $f_t^{(2)}$ are $r_2 \times 1$ vectors with $r_2 \geq 0$, $\lambda_i^{(3)}$ and $f_t^{(3)}$ are $r_3 \times 1$ vectors with $r_3 \geq 0$, and such that $r = r_1 + r_2 + r_3$ and $\lambda'_i = (\lambda_i^{(1)'} \lambda_i^{(2)'} \lambda_i^{(3)'})$ for all i.
Moreover, the common non-stationary factors are defined by the following equations

$$f_t^{(1)} = d_1 t + d_2 f_t^{(1)\dagger} + (1 - d_2) g_t, \tag{6}$$

$$f_t^{(1)\dagger} = f_0^{(1)\dagger} + \sum_{j=1}^{t} e_j^{(1)}, \tag{7}$$

$$f_t^{(2)} = f_0^{(2)} + \sum_{j=1}^{t} e_j^{(2)}, \tag{8}$$

where in (6)-(8): $f_t^{(1)\dagger}$, g_t and $e_t^{(1)}$ are $r_1 \times 1$ vectors, $e_t^{(2)}$ is an $r_2 \times 1$ vector, $e_t^{(1)}$, $e_t^{(2)}$, g_t and $f_t^{(3)}$ are $I(0)$, and d_1 and d_2 are dummy variables.

Lemma 1 states that the number of linear trends is either zero or one: if an identified r-dimensional vector of common factors has linear trends, this is tantamount to an identified r-dimensional vector of common factors where only the first factor has a linear trend. When $r_1 = 1$ and $d_1 = 1$, we show in Theorem 1 below, that it does not matter whether the remainder $d_2 f_t^{(1)\dagger} + (1 - d_2) g_t$ is $I(1)$ or $I(0)$: the trend component is the one that dominates. When $r_1 = 0$, there are no linear trends in the factor structure; in this case, $f_t^{(1)}$ can be $I(1)$ or $I(0)$, according as $d_2 = 1$ or 0.

Let us denote as r^* the number of non-stationary factors, and as r the total number of factors. Then, based on (6)-(8), the numbers of common factors in $X_{i,t}$ are summarised in the table below.

Factor type	*Number*
With linear trend	$r_1 d_1$
Zero mean, $I(1)$	$r_2 + r_1(1-d_1) d_2$
Zero mean, $I(0)$	$r_3 + r_1(1-d_1)(1-d_2)$
Total non-stationary	$r^* = r_1 d_1 + r_2 + r_1(1-d_1) d_2$
Total number of common factors	$r = r^* + r_3 + r_1(1-d_1)(1-d_2) = r_1 + r_2 + r_3$

Recall that - in addition to restricting r (and, therefore, r_1, r_2, r_3, r^*) to be finite - we allow for the possibility of having any of the numbers r_1, r_2, r_3, r^*, or even r, to be equal to zero. On the other hand, if there is no linear trend ($d_1 = 0$), we have at most $r_1 + r_2$ zero-mean $I(1)$ factors and $r_1 + r_3$ zero-mean $I(0)$ factors, while if there is a linear trend ($d_1 = 1$), we have at most r_2 zero-mean $I(1)$ factors and r_3 zero-mean $I(0)$ factors.

We now spell out the main assumptions. Consider the vector of zero-mean $I(1)$ factors: f_t^*, where $f_t^* = \left[f_t^{(1)\dagger}, f_t^{(2)\prime}\right]'$, and consider the $I(0)$ vector e_t, where $e_t = \left[e_t^{(1)}, e_t^{(2)\prime}\right]'$. Both f_t^* and e_t are $[r_2 + r_1(1-d_1)d_2] \times 1$ vectors.[1]

We define the long-run covariance matrix associated with f_t^* as

$$\Sigma_{\Delta f^*} = \lim_{T\to\infty} Var\left(T^{-1/2}\sum_{t=1}^{T} e_t\right). \tag{9}$$

Assumption 2. *Let $\kappa > 0$. It holds that (i) $E\|e_t\|^{4+\kappa} < \infty$ for all t; (ii) $E|f_0^*|^{4+\kappa} < \infty$; (iii) $\Sigma_{\Delta f^*}$ is positive definite; (iv) there exists, on a suitably enlarged probability space, an $(r_2 + d_2)$-dimensional standard Wiener process $W(t)$ such that, for some $\epsilon > 0$,*

$$\sup_{1\le t\le T}\left\|f_t^* - \Sigma_{\Delta f^*}^{1/2} W(t)\right\| = O_{a.s.}\left(T^{1/2-\epsilon}\right);$$

(v) $E\left\|\sum_{t=1}^{T} e_t\right\|^{2+\kappa} \le C_0\left(\sum_{t=1}^{T} E\|e_t\|^2\right)^{\frac{2+\kappa}{2}}$; (vi) $E\left\|\sum_{t=1}^{T} f_t^* f_t^{*\prime}\right\|^2 \le C_0 T^4$.

Assumption 2 poses some restrictions on the common $I(1)$ factors. Parts *(i)* and *(ii)* require the existence of at least the 4-th moment of the innovation e_t and of the initial condition f_0^* respectively. Part *(iii)* is a standard requirement, which rules out that the common, zero mean $I(1)$ factors are cointegrated: in essence, this ensures that the number of $I(1)$ common factors is genuinely r_2+d_2. Part *(iv)* states that a strong approximation exists for the partial sums process f_t^*. Although this

[1]Note that if $d_1 = 1$ or $d_2 = 0$ then these vectors have dimension r_2 and are given by $f_t^* = f_t^{(2)}$ and $e_t = e_t^{(2)}$; on the other hand if $d_1 = 0$ and $d_2 = 1$ then the vectors have dimension $r_1 + r_2$, thus become scalars if $r_2 = 0$ and $r_1 = 1$.

is a high-level assumption, we prefer to write it in this form as opposed to spelling out more primitive assumptions, since this makes the set-up more general. Part *(v)* is a Burkholder-type inequality (see e.g. Lin and Bai, 2010, p. 108). Finally, part *(vi)* can be verified e.g. under independence and finite fourth moments.

An important implication is that e_t is allowed to be (weakly) dependent over time. Considering part *(iv)* in particular, starting from the seminal paper by Berkes and Philipp (1979), the literature has developed several refinements of the Strong Invariance Principle (SIP) for random vectors. In particular, Liu and Lin (2009) derive the SIP for stationary causal processes, a wide class which includes e.g. conditional heteroskedasticity models, Volterra series, and data generated by dynamical systems - see Wu (2007) and Berkes, Liu, Wu, et al. (2014). Thus, part *(iv)* of the assumption accommodates for a wide variety of commonly considered DGPs.

Assumption 3. *It holds that: (i) (a)* $\max_{1\leq i\leq N,1\leq t\leq T} E\left|u_{i,t}\right|^4 < \infty$*; (b)* $\max_{1\leq t\leq T} E\left\|f_t^{(3)}\right\|^4 < \infty$*; and (c)* $\max_{1\leq t\leq T} E\left|g_t\right|^4 < \infty$*; (ii) (a)* $\max_{1\leq i\leq N} E\left\|\sum_{t=1}^T f_t^* u_{i,t}\right\|^2 \leq C_0 T^2$*; (b)* $E\left\|\sum_{t=1}^T f_t^* f_t^{(3)\prime}\right\|^2 \leq C_0 T^2$*; and (c)* $E\left\|\sum_{t=1}^T f_t^* g_t\right\|^2 \leq C_0 T^2$*; (iii)* $E\left\|\sum_{t=1}^T t f_t^*\right\|^2 \leq C_0 T^5$*; (iv) (a)* $\max_{1\leq i\leq N} E\left|\sum_{t=1}^T t u_{i,t}\right|^2 \leq C_0 T^3$*; (b)* $E\left\|\sum_{t=1}^T t f_t^{(3)}\right\|^2 \leq C_0 T^3$*; and (c)* $E\left|\sum_{t=1}^T t g_t\right|^2 \leq C_0 T^3$*; (v)* $E\left\|\sum_{t=1}^T f_t^* f_t^{*\prime}\right\|^2 \leq C_0 T^4$*.*

Assumption 3 deals with the idiosyncratic terms $u_{i,t}$ and the stationary factors. Part *(i)* requires the existence of the 4-th moments, which is a milder assumption than the customary 8-th moment existence requirement - see Bai (2004). Part *(ii)* could be shown from more primitive assumptions; indeed, a prototypical assumption would require e_t and $u_{i,t}$ to be independent of each other and *i.i.d.* over time - in such a case, explicit calculations would yield part *(ii)(a)*. Parts *(iii)* and *(iv)* could again be shown from more primitive assumptions; for example, part *(iv)(a)* would automatically follow if $Eu_{i,t}^2 < \infty$ and $u_{i,t}$ is *i.i.d.* across time. Similarly, it could be verified that part *(iii)* holds whenever $E\left\|e_t\right\|^2 < \infty$ and e_t is *i.i.d.* across t.

We now spell out the assumptions for the $N \times r$ loadings matrix $\Lambda = [\lambda_1|...|\lambda_N]'$.

Assumption 4. *The loadings* Λ *are non-stochastic with (i)* $\max_{1\leq i\leq N}\left\|\lambda_i\right\| < \infty$*; (ii)* $\lim_{N\to\infty}\frac{\Lambda'\Lambda}{N} \to \Sigma_\Lambda$*, where the matrix* Σ_Λ *is positive definite.*

Assumption 4 is standard in this literature - see e.g. Bai (2004). One consequence of part *(ii)* and Lemma 1 is that every diagonal block of Σ_Λ, defined by the loadings of $f_t^{(1)}$, $f_t^{(2)}$ or $f_t^{(3)}$, is also positive definite. Note that the assumption requires the loadings to be non-stochastic; however, this could be relaxed to the case of random loadings, with no changes to the main arguments in the paper.

Another, important consequence of Assumption 4 is that the common factors belonging in each category are "strong" or "pervasive". We postpone a discussion of this aspect, and of the possibility of extending this set-up, until Section 3.5.

2.2 Asymptotic behavior of eigenvalues

We base inference on the two matrices

$$\Sigma_1 = \frac{1}{T^3}\sum_{t=1}^{T} X_t X_t', \tag{10}$$

$$\Sigma_2 = \frac{1}{T^2}\sum_{t=1}^{T} X_t X_t'. \tag{11}$$

We denote the p-th largest eigenvalues of Σ_1 and Σ_2 as $\nu_1^{(p)}$and $\nu_2^{(p)}$ respectively. Consider the slowly varying sequence

$$l_{N,T} = (\ln N)^{1+\epsilon} (\ln T)^{\frac{3}{2}+\epsilon},$$

where $\epsilon > 0$. The asymptotic behaviour of those eigenvalues is reported in the following Theorem.

Theorem 1. *Under Assumptions 2-4, it holds that there are two random N_0 and T_0 such that, for all $N \geq N_0$ and $T \geq T_0$,*

$$\nu_1^{(p)} \geq C_p N, \qquad \text{for } p \leq r_1 d_1, \tag{12}$$

$$\nu_1^{(p)} = O_{a.s.}\left(\frac{N}{\sqrt{T}} l_{N,T}\right), \qquad \text{for } p > r_1 d_1, \tag{13}$$

and

$$\nu_2^{(p)} \geq C_p \frac{N}{\ln \ln T}, \qquad \text{for } 0 \leq p \leq r^*, \tag{14}$$

$$\nu_2^{(p)} = O_{a.s.}\left(\frac{N}{\sqrt{T}} l_{N,T}\right), \qquad \text{for } p > r^*, \tag{15}$$

with C_p a finite, positive constant.

Theorem 1 is a separation result for the eigenvalues corresponding to common factors in Σ_1 and Σ_2 and is our first contribution.

Equations (12) and (13) refer to the eigenvalues of Σ_1. The results state that the first r_1 eigenvalues diverge to infinity at a rate N; conversely, the remaining eigenvalues have a smaller magnitude. We pose no restrictions on the relative rate of divergence between N and T as they pass to infinity. Thus, the magnitude of $\nu_1^{(p)}$, when $p > r_1$, may be very large; it is however smaller - by a factor $T^{-1/2}$ - compared to that of $\nu_1^{(p)}$ when $p \leq r_1$. In the definition of Σ_1, there is a denominator given by T^3: intuitively, this is due to the fact that the presence of a drift in the common factor $f_t^{(1)}$ creates a linear trend. Norming by T^3 is needed in order to make the trend component converge.

Equations (14) and (15) refer to the eigenvalues of Σ_2. This matrix is normalised by T^2: the main idea is that we wish to separate the eigenvalues corresponding to non-stationary factors from the other ones. The partial sums of $f_t^* f_t^{*\prime}$ should grow at least as fast as T^2 by the CLT in functional spaces; the result in (14) follows from this intuition, although, since we need an a.s. rate, it is based on the Law of the Iterated Logarithm (see Donsker and Varadhan, 1977). Similarly to Σ_1, the remaining eigenvalues may also diverge, but this will happen at a slower rate. Equation (15) illustrates the separation result, through the $T^{-1/2}$ term. Following the proof of the theorem, it could be readily shown that, if the idiosyncratic components $u_{i,t}$ were $I(1)$, the upper bound for $\nu_2^{(p)}$ when $p > r_2 + \max\{r_1, d_2\}$ would be $O_{a.s.}(N l_{N,T})$ - in essence, in this case a separation result could not be shown, whence the need to assume that the $u_{i,t}$s are $I(0)$. On the other hand, one could envisage a situation where only a fraction of the $u_{i,t}$s are $I(1)$ - say $O(N^{\alpha_0})$, with $\alpha_0 < 1$. In such a case, by adapting the proof of Lemma B6 it can be shown that the upper bound in (15) would become $O_{a.s.}(N^{\alpha_0} l_{N,T}) + O_{a.s.}\left(\frac{N}{\sqrt{T}} l_{N,T}\right)$, and thus a separation result would obtain.

Note that Theorem 1 provides only rates: no distributional results are available. When data are stationary, Wang and Fan (2016) derive an asymptotic distribution for the estimates of the diverging eigenvalues of the sample covariance matrix. We do not know, however, if this can also be done for the $\nu_1^{(p)}$s and the $\nu_2^{(p)}$s. Hence, in what follows we will rely only on rates.

Finally, in order to construct the relevant test statistics, we will also make use of the first differenced version of (3):

$$\Delta X_{i,t} = \lambda_i' \Delta \mathcal{F}_t + \Delta u_{i,t}, \quad 1 \leq i \leq N. \tag{16}$$

Assumption 5. *It holds that: (i)* $E(\Delta \mathcal{F}_{j,t} \Delta u_{i,t}) = 0$ *for* $1 \leq j \leq r$ *and* $1 \leq i \leq N$*; (ii)* $\max_{1\leq i\leq N, 1\leq t\leq T} E|\Delta X_{i,t}|^4 \leq C_0$*; (iii)* $E \max_{1\leq \widetilde{t}\leq T}\left|\sum_{t=1}^{\widetilde{t}} \Delta X_{h,t}\Delta X_{j,t} - E(\Delta X_{h,t}\Delta X_{j,t})\right|^2 \leq C_0 T$*; (iv) (a)* $T^{-1}\sum_{t=1}^{T} E(\Delta \mathcal{F}_t \Delta \mathcal{F}_t')$ *is a positive definite matrix; (b) the largest eigenvalue of* $T^{-1}\sum_{t=1}^{T} E(\Delta u_t \Delta u_t')$ *is finite; (c)* $T^{-1}\sum_{t=1}^{T} E(\Delta u_t \Delta u_t')$ *is a positive definite matrix.*

Assumption 5 is the same as Assumptions 1-3 in Trapani (2017), and we refer to that paper for examples in which the assumption is satisfied. Essentially, these are the same examples that hold for e_t in Assumption 3.

3 Estimating the number of common factors

We now present the algorithm to estimate the dimension of the factor space. We begin by determining the presence or not of a common linear trend by estimating r_1

based on $\nu_1^{(p)}$, and then we determine the presence or not of zero-mean $I(1)$ common factors by estimating r^*, based on $\nu_2^{(p)}$.

3.1 Preliminary definitions

Consider the notation $\beta = \frac{\ln N}{\ln T}$, and define

$$\delta \begin{cases} > 0 & \text{when } \beta < \frac{1}{2}, \\ > 1 - \frac{1}{2\beta} & \text{when } \beta \geq \frac{1}{2}. \end{cases} \tag{17}$$

The role played by δ is the following. In view of Theorem 1, the largest eigenvalues are (modulo some slowly varying functions) proportional to N; the others, to $NT^{-1/2}$. When premultiplying eigenvalues by $N^{-\delta}$, the former will be proportional to $N^{1-\delta}$, thereby still diverging; the latter will be proportional to $N^{1-\delta}T^{-1/2}$, which, by construction, will drift to zero. Note that (17) provides a general rule to set δ, and we discuss specific choices in Section 4.

In order to construct our test statistics, we make use the eigenvalues of the matrix

$$\Sigma_3 = \frac{1}{T}\sum_{t=1}^{T} \Delta X_t \Delta X_t', \tag{18}$$

which with the same notation as before, are denoted as $\nu_3^{(p)}$ in decreasing order. In particular, when running our procedure for the p-th largest eigenvalues of Σ_1 or Σ_2, we will extensively use the quantities

$$\overline{\nu}_{3,p}(k) = \frac{1}{4\,(N-k+1)}\sum_{h=k}^{N} \nu_3^{(h)}, \tag{19}$$

for different values of k. Essentially, $\overline{\nu}_{3,p}(k)$ is the average of all (or some) eigenvalues of Σ_3 and will be employed in order to rescale the estimated eigenvalues, so as to render all our test statistics scale invariant. In the numerical analysis of Section 4, we consider rescaling schemes with $k=1$, $k=p$, or $k=(p+1)$ and we discuss the impact of these choices on our results. For simplicity in the rest of this section we do not make explicit the dependence of (19) on k. Finally, note the division by 4 in (19), which is done, heuristically, since it is possible that $\Delta X_{i,t}$ could inflate the variance by over-differencing, and the factor 4 represents the largest inflation factor possible.

3.2 Determining the presence of factors with linear trends

Consider first Σ_1 defined in (10), and its eigenvalues $\nu_1^{(p)}$. Based on (12)-(13), the first $r_1 d_1$ eigenvalues of Σ_1 should diverge to positive infinity, as $\min(N,T) \to \infty$, at a faster rate than the $(N - r_1 d_1)$ remaining ones. Thus, the cornerstone of the algorithm to determine $r_1 d_1$ is based on checking whether $\nu_1^{(p)}$ diverges sufficiently fast. In particular, as suggested by Theorem 1, we want to construct a test for

$$\left\{ \begin{array}{l} H_{0,1}^{(p)} : \nu_1^{(p)} \geq C_p' N, \\ H_{A,1}^{(p)} : \nu_1^{(p)} \leq C_p'' \frac{N}{\sqrt{T}} l_{N,T}, \end{array} \right. \tag{20}$$

for some positive bounded constants C_p' and C_p''. Thus, given r_1 we have that $H_{0,1}^{(p)}$ holds true for $p \leq r_1$, while $H_{A,1}^{(p)}$ holds true for $p > r_1$.

Consider the following transformation of $\nu_1^{(p)}$

$$\phi_1^{(p)} = \exp \left\{ N^{-\delta} \frac{\nu_1^{(p)}}{\overline{\nu}_{3,p}} \right\}. \tag{21}$$

Then, based on (20), equations (12) and (13), and given the definition (17) of δ, we have that

$$\begin{array}{lll} \lim_{\min(N,T) \to \infty} \phi_1^{(p)} = \infty, & \text{under } H_{0,1}^{(p)} \text{ i.e. for } p \leq r_1, \\ \lim_{\min(N,T) \to \infty} \phi_1^{(p)} = C_p < \infty, & \text{under } H_{A,1}^{(p)} \text{ i.e. for } p > r_1. \end{array}$$

In principle, we could then use $\phi_1^{(p)}$ to test $H_{0,1}^{(p)}$. However, since $\phi_1^{(p)}$ either diverges to infinity or not, it does not have any randomness. Therefore, we propose to use the following randomisation algorithm - note that other randomisations schemes would also be possible, in principle; the one we propose, however, has been often considered in this type of literature (see e.g. Corradi and Swanson, 2006, and Trapani, 2017).

Step A1.1. Generate an *i.i.d.* sample $\left\{ \xi_{1,j}^{(p)} \right\}_{j=1}^{R_1}$ from a common distribution G_1, independently across p.

Step A1.2. For any u drawn from a distribution $F_1(u)$, define, for $1 \leq j \leq R_1$,

$$\zeta_{1,j}^{(p)}(u) = I\left[\phi_1^{(p)} \times \xi_{1,j}^{(p)} \leq u \right]$$

Step A1.3. Compute

$$\vartheta_1^{(p)}(u) = \frac{1}{\sqrt{R_1}} \sum_{j=1}^{R_1} \frac{\zeta_{1,j}^{(p)}(u) - G_1(0)}{\sqrt{G_1(0)[1 - G_1(0)]}}.$$

Step A1.4. Compute

$$\Theta_1^{(p)} = \int_{-\infty}^{+\infty} \left|\vartheta_1^{(p)}(u)\right|^2 dF_1(u).$$

The intuition for considering this approach is the following. Under the null, we know that $\phi_1^{(p)}$ diverges; thus, we can expect $\zeta_{1,j}^{(p)}(u)$ to be an *i.i.d.* Bernoulli sequence with expected value exactly equal to $G_1(0)$, and variance $G_1(0)[1 - G_1(0)]$. In such case, a CLT should ensure that $\vartheta_1^{(p)}(u)$ follows a Normal distribution, and consequently $\Theta_1^{(p)}$ should be expected to follow a Chi-squared distribution. By the same token, under the alternative $\phi_1^{(p)}$ is finite, and therefore $\zeta_{1,j}^{(p)}(u)$ should be an *i.i.d.* Bernoulli sequence with expected value different from $G_1(0)$; thus, $\vartheta_1^{(p)}(u)$ should diverge as fast as $\sqrt{R_1}$ by the LLN, and consequently $\Theta_1^{(p)}$ should also diverge at a rate R_1. The random variable $\Theta_1^{(p)}$ is then the statistic that we are going to use.

In order to derive the asymptotic behavior of $\Theta_1^{(p)}$, we need some regularity conditions on the distributions $G_1(\cdot)$ and $F_1(\cdot)$ - see Section 4 for a choice of these functions and of R_1.

Assumption 6. *It holds that: (i) (a) $G_1(\cdot)$ has a bounded density function; (b) $G_1(0) \neq 0$ and $G_1(0) \neq 1$; (ii) $\int_{-\infty}^{\infty} u^2 dF_1(u) < \infty$.*

Let P^* denote the conditional probability with respect to $\{X_{i,t}, 1 \leq t \leq T, 1 \leq i \leq N\}$; we use the notation "$\stackrel{D^*}{\to}$" and "$\stackrel{P^*}{\to}$" to define, respectively, conditional convergence in distribution and in probability according to P^*. It holds that

Theorem 2. *Consider $H_{0,1}^{(p)}$ and $H_{A,1}^{(p)}$ defined in (20). Under Assumptions 2-6, if*

$$\lim_{\min(N,R_1)\to\infty} \sqrt{R_1} \exp\left\{-N^{1-\delta}\right\} = 0, \tag{22}$$

then, for almost all realisations of $\{e_t, u_{i,t}, 1 \leq i \leq N, 1 \leq t \leq T\}$ and for all p, as $\min(N, T, R_1) \to \infty$, under $H_{0,1}^{(p)}$ it holds that

$$\Theta_1^{(p)} \stackrel{D^*}{\to} \chi_1^2, \tag{23}$$

and under $H_{A,1}^{(p)}$ it holds that

$$\frac{1}{R_1} \frac{\int_{-\infty}^{\infty} [G_1(u) - G_1(0)]^2 dF_1(u)}{G_1(0)[1 - G_1(0)]} \Theta_1^{(p)} \stackrel{P^*}{\to} 1. \tag{24}$$

The determination of r_1 follows from an algorithm which is based on a single step.

Step T1.1. Set $p = 1$ and run the test for $H_{0,1}^{(1)} : \nu_1^{(1)} = \infty$ based on $\Theta_1^{(1)}$. If the null is rejected, set $\widehat{r}_1 = 0$ and stop, otherwise set $\widehat{r}_1 = 1$.

The output of this step is $\widehat{r}_1$, which is an estimate of $r_1 d_1$. As discussed above, $r_1 d_1$ can be either 0 or 1, whence the test being stopped at $p = 2$. The procedure based on the single *Step T1.1* can therefore be viewed as a test for the presence of a common factor with a linear trend.

As can be expected, in order to ensure that $\widehat{r}_1$ is consistent, a pivotal role is played by the level of the test, $\alpha_1 := P^*(\Theta_1^{(p)} > c_{\alpha,1})$, through the relevant critical value denoted as $c_{\alpha,1}$.

Lemma 2. *Under the assumptions of Theorem 2, as* $\min(N, T, R_1) \to \infty$*, if* $c_{\alpha,1} \to \infty$ *with* $c_{\alpha,1} = o(R_1)$*, then it holds that* $P^*(\widehat{r}_1 = r_1 d_1) = 1$*, for almost all realisations of* $\{e_t, u_{i,t}, 1 \le i \le N, 1 \le t \le T\}$.

Requiring that $c_{\alpha,1} \to \infty$ is necessary in order to have asymptotically zero Type I error probability, which ensures the consistency result in the lemma; an immediate implication of $c_{\alpha,1} \to \infty$ is that the level of the test is such that

$$\lim_{\min(N,T,R_1)\to\infty} P^*\left(\Theta_1^{(p)} > c_{\alpha,1}\right) = 0. \tag{25}$$

The fact that $c_{\alpha,1}$ diverges has also an interesting consequence on the interpretation of the outcome of our testing procedure. It is well-known that randomised tests will yield different results for different researchers when applied to the same data, since the added randomness does not vanish asymptotically. However, this is not the case with our procedure, since, when $c_{\alpha,1} \to \infty$, (25) holds under $H_{0,1}^{(p)}$. Further, we show in the proof that having $c_{\alpha,1} = o(R_1)$ affords that the probability of a Type II error is asymptotically zero, thus ensuring consistency.

Looking at this from a different angle, the results in Lemma 2 are guaranteed by letting the level of the test $\alpha_1 \to 0$ as $\min(N, T, R_1) \to \infty$ and we refer to Section 4 for the choice of α_1.

3.3 Determining the number of non-stationary common factors

Consider the matrix Σ_2 defined in (11) and its eigenvalues $\nu_2^{(p)}$. Based on Theorem 1, the r^* largest eigenvalues of Σ_2 should diverge to positive infinity, as $\min(N, T) \to \infty$, at a faster rate than the $(N - r^*)$ remaining ones. Therefore, we can construct a the test for

$$\begin{cases} H_{0,2}^{(p)} : \nu_2^{(p)} \ge C_p' \frac{N}{\ln \ln T}, \\ H_{A,2}^{(p)} : \nu_2^{(p)} \le C_p'' \frac{N}{\sqrt{T}} l_{N,T}, \end{cases} \tag{26}$$

for some positive bounded constants C_p' and C_p''. Thus, given r^* we have that $H_{0,2}^{(p)}$ holds true for $p \le r^*$, while $H_{A,2}^{(p)}$ holds true for $p > r^*$.

We exploit this fact, as in the above, by considering the following transformation of $\nu_2^{(p)}$

$$\phi_2^{(p)} = \exp\left\{ N^{-\delta} \left(\ln \ln T \right) \frac{\nu_2^{(p)}}{\overline{\nu}_{3,p}} \right\}, \tag{27}$$

which is very similar to (21) except for the presence of the logarithmic term, which is a consequence of (14). Then, based on (20), equations (14) and (15), and given the definition (17) of δ, we have that

$$\begin{array}{ll} \lim_{\min(N,T)\to\infty} \phi_2^{(p)} = \infty, & \text{under } H_{0,2}^{(p)} \text{ i.e. for } p \leq r^*, \\ \lim_{\min(N,T)\to\infty} \phi_2^{(p)} = C_p < \infty, & \text{under } H_{A,2}^{(p)} \text{ i.e. for } p > r^*. \end{array}$$

We consider the following randomisation procedure.

Step A2.1 Generate an *i.i.d.* sample $\left\{ \xi_{2,j}^{(p)} \right\}_{j=1}^{R_2}$ from a common distribution G_2, independently across p and of $\left\{ \xi_{2,j}^{(p')} \right\}_{j=1}^{R_2}$ for all $p' \neq p$.

Step A2.2 For any u drawn from a distribution $F_2(u)$, define, for $1 \leq j \leq R_2$,

$$\zeta_{2,j}^{(p)}(u) = I\left[\phi_2^{(p)} \times \xi_{2,j}^{(p)} \leq u \right].$$

Step A2.3. Compute

$$\vartheta_2^{(p)}(u) = \frac{1}{\sqrt{R_2}} \sum_{j=1}^{R_2} \frac{\zeta_{2,j}^{(p)}(u) - G_2(0)}{\sqrt{G_2(0)\left[1 - G_2(0)\right]}}.$$

Step A2.4. Compute

$$\Theta_2^{(p)} = \int_{-\infty}^{+\infty} \left| \vartheta_2^{(p)}(u) \right|^2 dF_2(u).$$

The same comments as in the previous algorithm apply: in essence, the procedure exploits the fact that under the null and the alternative, $\phi_2^{(p)}$ diverges or drifts to zero respectively: the former feature ensures (asymptotic) normality of $\vartheta_2^{(p)}(u)$, whereas the latter entails that $\vartheta_2^{(p)}(u)$ diverges under the alternative.

Assumption 7. *It holds that: (i) (a) G_2 has a bounded density function; (b) $G_2(0) \neq 0$ and $G_2(0) \neq 1$; (ii) (a) $\int_{-\infty}^{\infty} u^2 dF_2(u) < \infty$.*

It holds that

Theorem 3. *Consider* $H_{0,2}^{(p)}$ *and* $H_{A,2}^{(p)}$ *defined in (26). Under Assumptions 2-5 and 7, if*

$$\lim_{\min(N,R_2)\to\infty} \sqrt{R_2} \exp\left\{-N^{1-\delta}\right\} = 0, \tag{28}$$

then, for almost all realisations of $\{e_t, u_{i,t}, 1 \leq i \leq N, 1 \leq t \leq T\}$ *and for all* p, *as* $\min(N, T, R_2) \to \infty$, *under* $H_{0,2}^{(p)}$ *it holds that*

$$\Theta_2^{(p)} \overset{D^*}{\to} \chi_1^2, \tag{29}$$

and under $H_{A,2}^{(p)}$ *it holds that*

$$\frac{1}{R_2} \frac{\int_{-\infty}^{\infty} (G_2(u) - G_2(0))^2 \, dF_1(u)}{G_2(0)(1 - G_2(0))} \Theta_2^{(p)} \overset{P^*}{\to} 1 \text{ under } H_1^{(2)}. \tag{30}$$

Note that, conditionally on the sample, the sequence $\left\{\Theta_2^{(p)}\right\}_{p=1}^{N}$ is independent across p. We recommend the following algorithm for the determination of r^*.

<u>*Step T2.1.*</u> Run the test for $H_{0,2}^{(1)} : \nu_2^{(1)} = \infty$ based on $\Theta_2^{(1)}$. If the null is rejected, set $\widehat{r}^* = 0$ and stop, otherwise go to the next step.

<u>*Step T2.2.*</u> Starting from $p = 1$, run the test for $H_{0,2}^{(p+1)} : \nu_2^{(p+1)} = \infty$ based on $\Theta_2^{(p+1)}$, constructed using an artificial sample $\left\{\xi_{2,j}^{(p+1)}\right\}_{j=1}^{R_2}$ generated independently of $\left\{\xi_{2,j}^{(1)}\right\}_{j=1}^{R_2}$, ..., $\left\{\xi_{2,j}^{(p)}\right\}_{j=1}^{R_2}$. If the null is rejected, set $\widehat{r}^* = p$ and stop; otherwise repeat the step until the null is rejected (or until a pre-specified maximum number of factors, say $r_{\max}^*$, is reached).

As can be expected, in this context a pivotal role is played by the level of the individual tests, which should be chosen so that $\widehat{r}^*$ is a good approximation of r^*, at least asymptotically. Similarly to the previous case, let $c_{\alpha,2}$ denote the critical value of the test at each step.

Lemma 3. *Under the assumptions of Theorem 3, as* $\min(N, T, R_2) \to \infty$, *if* $r_{\max}^* \geq r^*$ *and* $c_{\alpha,2} \to \infty$ *with* $c_{\alpha,2} = o(R_2)$, *then it holds that* $P^*(\widehat{r}^* = r^*) = 1$ *for almost all realisations of* $\{e_t, u_{i,t}, 1 \leq i \leq N, 1 \leq t \leq T\}$.

This lemma has the same interpretation - especially when it comes to the condition that $c_{\alpha,2} \to \infty$ - as Lemma 2.

3.4 Determining the number of zero-mean $I(1)$ and $I(0)$ factors

After estimating r^*, it is possible to estimate the number of common, zero-mean $I(1)$ factors by subtracting the number of those with a linear trend from the total number of non-stationary factors, i.e. as $\widehat{r}^{*}-\widehat{r}_{1}$. Under the conditions of Lemmas 2 and 3, it is immediate to verify that

$$P^{*}\left[\widehat{r}^{*}-\widehat{r}_{1}=r_{2}+r_{1}\left(1-d_{1}\right)d_{2}\right]=1.$$

As a final remark, on the grounds of Assumption 5 it is possible to use the algorithm proposed in Trapani (2017) to estimate the total number of common factors. The algorithm - based on first-differenced data - uses the eigenvalues $\nu_{3}^{(p)}$ of Σ_{3} defined in (18) in a similar way to the algorithms above. Denoting the estimate of the total number of factors as $\widehat{r}$, the number of common $I(0)$ factors can be estimated as $\widehat{r}-\widehat{r}^{*}$. Under the conditions in Trapani (2017) and of Lemma 3 above, it follows that

$$P^{*}\left[\widehat{r}-\widehat{r}^{*}=r_{3}+r_{1}\left(1-d_{1}\right)\left(1-d_{2}\right)\right]=1.$$

3.5 Determining the presence of weak factors

By Assumption 4, all the common factors are assumed to be strongly pervasive. This is a direct consequence of having $\|\Lambda\|^{2}=O\left(N\right)$. It is however possible to imagine a situation in which some of the common factors are "weak", or "less pervasive": this can arise from e.g. having genuinely weak factors, or from having strong factors which impact only on a small number of units - see, for example, Onatski (2012) and the references therein.

In this section, we report some heuristic arguments (similar to Trapani, 2017), on the ability of our procedure to determine weak factors. For the sake of a concise discussion, but with no loss of generality, we consider the case where all r factors are zero-mean $I(1)$, and $\Lambda'\Lambda$ is diagonal, with diagonal elements $C_{p}(N)$ given by

$$C_{p}(N)=\left\{\begin{array}{cl} N & \text{for } 1\leq p\leq p' \\ N^{1-\kappa_{p}} & \text{for } p'<p\leq r \end{array}\right. .$$

Allowing for $\kappa_{p}\in(0,1)$ corresponds to the case of having weak factors, and the larger κ_{p} the weaker the corresponding factor. Suppose that the researcher is using Σ_{2} and its eigenvalues $\nu_{2}^{(p)}$ in order to determine r. Repeating exactly the same arguments in the proof of Theorem 1, it can be shown that

$$\nu_{2}^{(p)}\geq C_{0}\frac{C_{p}(N)}{\ln\ln T}. \tag{31}$$

Equation (31) entails that, whenever $p'<p\leq r$,

$$\nu_{2}^{(p)}\geq C_{0}\frac{N^{1-\kappa_{p}}}{\ln\ln T}. \tag{32}$$

Recall that, our procedure, essentially, is based on testing whether, as $\min(N,T) \to \infty$

$$\begin{cases} H_{0,2}^{(p)} : (\ln \ln T) N^{-\delta} \nu_2^{(p)} \to \infty \\ H_{A,2}^{(p)} : (\ln \ln T) N^{-\delta} \nu_2^{(p)} \to 0 \end{cases},$$

with δ selected as per (17). Thus, based on (32), weak factors can be determined if

$$\lim_{\min(N,T)\to\infty} N^{1-\kappa_p-\delta} \to \infty,$$

which requires

$$\kappa_p < 1 - \delta. \tag{33}$$

On the grounds of (17), the constraint in (33) explains up to which extent weak factors can be detected. When $\beta \leq \frac{1}{2}$, that is $\frac{N}{\sqrt{T}} = O(1)$, then $\delta = 0$, and we need $\kappa_p < 1$. This entails that, when N is much smaller than T, our procedure is able to detect even very weak factors. Conversely, when $\beta > \frac{1}{2}$, that is $\frac{\sqrt{T}}{N} = o(1)$, it is required that $\kappa_p < 1 - \frac{1}{2\beta}$: as β increases, i.e. N increases, the test is less and less able to detect weak factors. Note that when N and T have the same order of magnitude, and thus $\beta = 1$, weak factors can be detected as long as $\kappa_p < \frac{1}{2}$ - that is, when the eigenvalues associated with that factor diverge to infinity a bit faster than $\sqrt{N}$.

4 Monte Carlo and empirical evidence

In our experiments, we use data generated as

$$X_{i,t} = \lambda_i^{(1)} f_t^{(1)} + \lambda_i^{(2)\prime} f_t^{(2)} + \lambda_i^{(3)\prime} f_t^{(3)} + \sqrt{\theta} u_{i,t}, \quad 1 \leq i \leq N,\ 1 \leq t \leq T, \tag{34}$$

$$f_t^{(1)} = 1 + f_{t-1}^{(1)} + \epsilon_t^{(1)}, \tag{35}$$

$$f_{j,t}^{(2)} = f_{j,t-1}^{(2)} + e_{j,t}^{(2)}, \quad e_{j,t}^{(2)} = \rho_j e_{j,t-1}^{(2)} + \epsilon_{j,t}^{(2)}, \qquad j = 1, \ldots, r_2, \tag{36}$$

$$f_{j,t}^{(3)} = \alpha_j f_{j,t-1}^{(3)} + \epsilon_{j,t}^{(3)}, \qquad j = 1, \ldots, r_3, \tag{37}$$

$$u_{i,t} = a_i u_{i,t-1} + v_{i,t} + b_i \sum_{|k| \leq C_i, k \neq 0} v_{i+k,t}, \tag{38}$$

The loadings in (34) are simulated such that each entry - when nonzero - is distributed as $\mathcal{N}(0,1)$ and such that the matrix Λ satisfies the normalization constraint $\Lambda'\Lambda = NI_r$. Note that model (34) represents the most general DGP which we use, where it is understood that, when e.g. no factors with a linear trend are present, we set $\lambda_i^{(1)} = 0$, etc. In (36) and (37), we use $\rho_j \sim U[0, \bar{\rho}]$ with $\bar{\rho} \in \{0, 0.4, 0.8\}$, and $\alpha_j \sim U[-0.5, 0.5]$ respectively. The vector $\epsilon_t = (\epsilon_t^{(1)} \epsilon_{1,t}^{(2)} \ldots \epsilon_{r_2,t}^{(2)} \epsilon_{1,t}^{(3)} \ldots \epsilon_{r_3,t}^{(3)})$ is simulated from $\mathcal{N}(0,\Gamma)$ independently at each t, with Γ diagonal and such that

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(\lambda_i^{(1)}\Delta f_t^{(1)})^2 = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(\lambda_i^{(2)\prime}\Delta f_t^{(2)})^2 = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(\lambda_i^{(3)\prime} f_t^{(3)})^2,$$

so that in first differences each factor component has, on average, the same weight. In (38) we allow both for serial and cross-sectional dependence in the idiosyncratic errors and for all $1 \leq i \leq N$. We fix $a_i = 0.5$, $b_i = 0.5$ and $C_i = \min\left(\left\lfloor \frac{N}{20} \right\rfloor, 10\right)$, and the errors $v_{i,t}$ are simulated from $\mathcal{N}(0,1)$. Note that this model for the idiosyncratic component is the same as in Ahn and Horenstein (2013). Last, we set the noise-to-signal as

$$\theta = 0.5\,\frac{\sum_{i=1}^{N}\sum_{t=1}^{T}(\lambda_i^{(1)}\Delta f_t^{(1)} + \lambda_i^{(2)\prime}\Delta f_t^{(2)} + \lambda_i^{(3)\prime}\Delta f_t^{(3)})^2}{\sum_{i=1}^{N}\sum_{t=1}^{T}(\Delta u_{i,t})^2}.$$

We consider the following cases:

1. we fix $r_3 = 0$ and we let $r_1 \in \{0,1\}$ and $r_2 \in \{0,1,2\}$ and we use the test based on $\phi_1^{(p)}$ to compute $\widehat{r}_1$ (see Table 1);

2. we fix $r_1 = 0$ and we let $r_2 \in \{0,1,2\}$ and $r_3 \in \{0,1,2\}$ and we use the test based on $\phi_2^{(p)}$ to compute $\widehat{r}^* = \widehat{r}_2$ (see Table 2);

3. we fix $r_1 = 1$ and we let $r_2 \in \{0,1,2\}$ and $r_3 \in \{0,1,2\}$ and we use the test based on $\phi_2^{(p)}$ to compute $\widehat{r}^* = \widehat{r}_2 + 1$ (see Table 3).

For each case, we set $N \in \{50, 100, 200\}$ and $T \in \{100, 200, 500\}$, and we simulate model (34)-(38) 500 times, reporting the average value of $\widehat{r}_1$ or $\widehat{r}_2 = (\widehat{r}^* - r_1)$, across simulations. Moreover, when computing $\widehat{r}_2$ we compare our results with the Information Criteria by Bai (2004), denoted as IC - this corresponds to $IC3$ in the original paper; we note that the other criteria, known as IC_1 and IC_2, deliver a similar (or worse) performance and are therefore not reported.

Our tests are run as follows. When computing $\phi_1^{(p)}$ and $\phi_2^{(p)}$, we rescale the p-th eigenvalue as (see (19))

$$\frac{\nu_i^{(p)}}{\bar{\nu}_{3,p}(k)} = \frac{\nu_i^{(p)}}{\frac{1}{4(N-k+1)}\sum_{h=k}^{N}\nu_3^{(h)}}, \qquad i = 1, 2.$$

For a given p, we consider three different rescaling schemes corresponding to three different choices for k:

<u>$BT1$</u>: when $k = 1$, i.e. $\bar{\nu}_{3,p}(k) = \frac{1}{4N}\sum_{h=1}^{N}\nu_3^{(h)}$;

$\underline{BT2}$: when $k = p$, i.e. $\bar{\nu}_{3,p}(k) = \frac{1}{4(N-p+1)} \sum_{h=p}^{N} \nu_3^{(h)}$;

$\underline{BT3}$: when $k = (p+1)$, i.e. $\bar{\nu}_{3,p}(k) = \frac{1}{4(N-p)} \sum_{h=p+1}^{N} \nu_3^{(h)}$.

We then divide the eigenvalues by N^{δ}, where (see (17))

$$\delta = \begin{cases} \delta^*, & \text{when } \frac{\ln N}{\ln T} < \frac{1}{2}, \\ 1 - \frac{1}{2\beta} + \delta^*, & \text{when } \frac{\ln N}{\ln T} \geq \frac{1}{2}, \end{cases} \tag{39}$$

with $\delta^* = 10^{-5}$. Thence, for each p, in the first step of the randomisation algorithm, $\{\xi_{1,j}^{(p)}\}_{j=1}^{R_1}$ and $\{\xi_{2,j}^{(p)}\}_{j=1}^{R_2}$ are generated from a standard normal distribution as typical in this literature (other choices would also be possible), with $R_1 = N$ and $R_2 = N$, if $p = 1$ or $R_2 = \lfloor \frac{N}{3} \rfloor$, for $p > 1$. In the second step of the randomisation algorithm, we set $u = \pm\sqrt{2}$. In Appendix A, we provide an analysis of our results when varying R_1, R_2, δ^* and u, showing that results are robust to these specifications. All tests are carried out at a significance level $\alpha_1 = \alpha_2 = \frac{0.05}{\min(N,T)}$, which corresponds to critical values growing logarithmically with N or T, hence satisfying the conditions in Lemmas 2 and 3.

To save space, here we report only the results when, in (36), we set $\bar{\rho} = 0.4$ - results for $\bar{\rho} = 0$ and $\bar{\rho} = 0.8$ are in Appendix A. As an overall comment, results are in general unaffected but for two cases. The first case is when $r_1 = 0$ and $r_2 = 1$ and we compute $\widehat{r}_1$; in this case, we find that lower values of $\bar{\rho}$ improve the results. Conversely, higher values of $\bar{\rho}$ make the innovations of the zero-mean $I(1)$ factors more persistent, thus making the associated eigenvalues larger: in this case, we are therefore more likely to falsely detect trends. The second case arises when $r_2 = 0$, and we compute $\widehat{r}_2$. In this case, we find the exact opposite. This can be explained upon noting that, for lower values of $\bar{\rho}$, the two $I(1)$ factors become closer to two pure random walks which are highly collinear, thus making the second eigenvalue $\nu_2^{(2)}$ much smaller than the first one $\nu_2^{(1)}$: thus, in this case, we are less likely to detect the second factor. For the same reason, higher values of $\bar{\rho}$ make the two factors less collinear, so that then the second factor is detected more easily.

The tables lend themselves to drawing some general conclusions about the main features of our methodology. First, $BT1$ and $BT2$ are usually very good at finding no common factors - whether with a linear trend or genuinely $I(1)$ with zero mean - when there are no common factors (see Tables 1 and 2), which is also consistent with the results in Trapani (2017). In the case of detecting the presence of genuine zero-mean $I(1)$ common factors, these results can be compared with the ones obtained using IC which invariably finds one common $I(1)$ factor even when such factors are not present (see Table 2). Few exceptions are found in Table 1, in the case where there is no common factor with a linear trend but there is one zero-mean $I(1)$ common factor. Even in this case, both $BT1$ and $BT2$ work extremely well as T increases. Note that $BT3$ also works very well in this case, at least when no common

Table 1: Average estimated number of factors with linear trend, $\widehat{r}_1$.

r_1	r_2	$N=50$, $T=100$ $BT1$	$BT2$	$BT3$	$N=100$, $T=100$ $BT1$	$BT2$	$BT3$	$N=200$, $T=100$ $BT1$	$BT2$	$BT3$
0	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0	1	0.20	0.18	0.19	0.11	0.10	0.41	0.04	0.04	0.31
0	2	0.04	0.05	0.06	0.02	0.02	0.13	0.01	0.00	0.06
1	0	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1	1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1	2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		$N=100$, $T=200$			$N=200$, $T=200$			$N=200$, $T=500$		
r_1	r_2	$BT1$	$BT2$	$BT3$	$BT1$	$BT2$	$BT3$	$BT1$	$BT2$	$BT3$
0	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0	1	0.06	0.06	0.30	0.05	0.04	0.25	0.02	0.01	0.14
0	2	0.00	0.00	0.05	0.00	0.00	0.03	0.00	0.00	0.01
1	0	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1	1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1	2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table 2: Average estimated number of zero-mean $I(1)$ factors, $\widehat{r}_2$, when $r_1 = 0$.

r_2	r_3	$N=50$, $T=100$ $BT1$	$BT2$	$BT3$	IC	$N=100$, $T=100$ $BT1$	$BT2$	$BT3$	IC	$N=200$, $T=100$ $BT1$	$BT2$	$BT3$	IC
0	0	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00
0	1	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00
0	2	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00
1	0	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1	1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1	2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
2	0	1.99	1.99	1.99	2.00	1.98	2.00	1.99	2.00	1.97	1.99	1.99	2.00
2	1	1.98	1.97	1.97	2.00	1.97	2.00	2.00	2.00	1.94	1.98	1.99	2.00
2	2	1.98	1.98	1.98	2.00	1.99	2.00	2.00	2.00	1.98	2.00	2.00	2.00
		$N=100$, $T=200$				$N=200$, $T=200$				$N=200$, $T=500$			
r_2	r_3	$BT1$	$BT2$	$BT3$	IC	$BT1$	$BT2$	$BT3$	IC	$BT1$	$BT2$	$BT3$	IC
0	0	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00
0	1	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00
0	2	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00
1	0	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1	1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1	2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
2	0	2.00	2.00	2.00	2.00	1.99	2.00	2.00	2.00	2.00	2.00	2.00	2.00
2	1	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00
2	2	2.00	2.00	2.00	2.00	2.00	1.99	2.00	2.00	2.00	2.00	2.00	2.00

Table 3: Average estimated number of zero-mean $I(1)$ factors, $\widehat{r}_2$, when $r_1 = 1$.

		$N=50$, $T=100$				$N=100$, $T=100$				$N=200$, $T=100$			
r_2	r_3	$BT1$	$BT2$	$BT3$	IC	$BT1$	$BT2$	$BT3$	IC	$BT1$	$BT2$	$BT3$	IC
0	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0	1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0	2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1	0	1.00	1.00	1.00	1.00	1.00	0.99	1.00	1.00	1.00	1.00	1.00	1.00
1	1	1.00	1.00	1.00	1.00	1.00	0.99	1.00	1.00	1.00	1.00	1.00	1.00
1	2	0.99	0.98	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	0.99	1.00
2	0	1.90	1.97	1.99	1.97	1.91	1.95	1.98	1.99	1.85	1.95	1.98	1.99
2	1	1.64	1.80	1.91	1.85	1.64	1.85	1.94	1.95	1.55	1.76	1.91	1.97
2	2	1.62	1.77	1.87	1.80	1.65	1.76	1.87	1.89	1.54	1.67	1.83	1.94

		$N=100$, $T=200$				$N=200$, $T=200$				$N=200$, $T=500$			
r_2	r_3	$BT1$	$BT2$	$BT3$	IC	$BT1$	$BT2$	$BT3$	IC	$BT1$	$BT2$	$BT3$	IC
0	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0	1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0	2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1	0	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1	1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1	2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
2	0	1.98	2.00	1.99	1.99	1.96	1.99	2.00	2.00	1.99	2.00	2.00	2.00
2	1	1.89	1.98	1.99	1.98	1.87	1.95	1.99	2.00	1.98	1.99	2.00	2.00
2	2	1.89	1.95	1.98	1.96	1.85	1.93	1.96	1.99	1.97	1.99	1.99	2.00

factors with linear trends are present. Conversely, when there is one common factor with a linear trend, $BT3$ tends to overestimated more when one $I(1)$ factor is present, even when T is large. Second, in general our criteria tend to understate, albeit slightly, as opposed to overstate the true number of common factors, this is particularly true when estimating the number of zero-mean $I(1)$ factors in presence of linear trends (see Table 3). In any case, the bias is of the same order as the bias of IC and tends to vanish as T increases. Overall, the performance of all our criteria improves dramatically as T increases: although results are usually good whenever $T = 100$, they markedly improve when $T \geq 200$ for all cases considered. The impact of N is, in general, less clear.

5 An empirical investigation of the dimensions of the yield curve

In this section, we illustrate our methodology through an application to the High Quality Market (HQM) Corporate Bond Yield Curve, available from the Federal

Table 4: Estimated number of factors in the HQM Corporate Bond Yield Curve

		$BT1$	$BT2$	$BT3$	IC
with linear trend	$\widehat{r}_1$	0	0	1	n.a.
non-stationary	$\widehat{r}^*$	1	3	5	5
zero-mean, $I(1)$	$\widehat{r}_2$	1	3	4	n.a.
all factors	$\widehat{r}$	1	5	5	5
zero-mean, $I(0)$	$\widehat{r}_3$	0	2	0	0

Reserve Economic Data (FRED)[2] - details on the construction of the yield curves are available from the US Department of Treasury.[3] We use monthly data on HQM Corporate Bonds with maturities from 6 months up to 100 years ($N = 196$), and spanning the period from January 1985 to September 2017 ($T = 393$). The data are shown in Figure 1, which shows evidence of non-stationarity and co-movements both cross-sectionally and across time.

We use the same settings as in Section 4. In particular, when computing $\widehat{r}_1$, we set $R_1 = N$, while for $\widehat{r}^*$ we set $R_2 = N$ if $p = 1$ and $R_2 = \lfloor N/3 \rfloor$ for $p > 1$. The significance level is $\frac{0.05}{\min(N,T)} = 0.0002551$. Finally, we note that when computing $\widehat{r}$, $BT1$ is equivalent to the test by Trapani (2017).

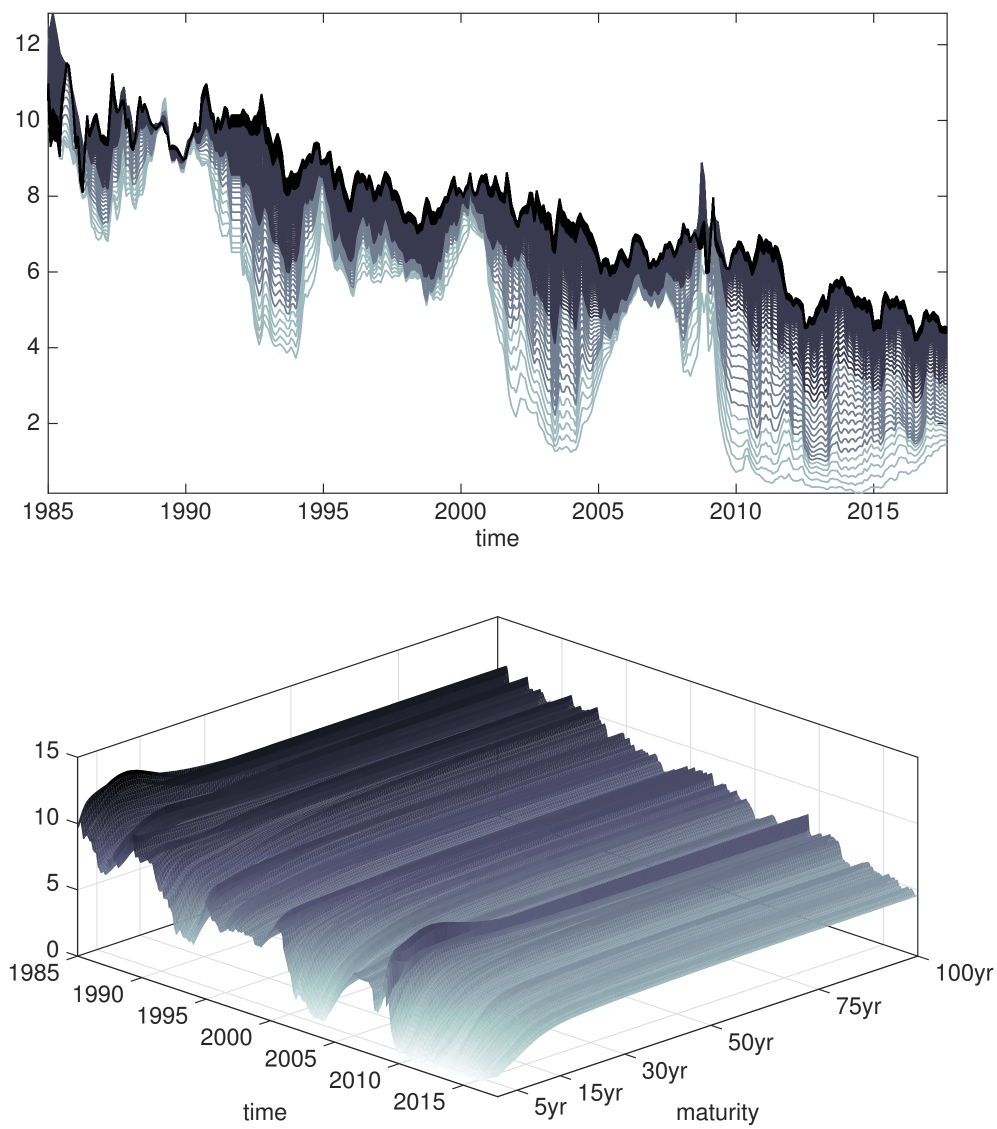
Results are in Table 4, where we have reported our three criteria, and, as a term of comparison, the information criterion IC - when computing $\widehat{r}$, this is equivalent to $IC3$ in Bai and Ng (2002). Based on our findings, there is borderline evidence of a common factor with a linear trend - indeed, this is picked up by $BT3$. Note that $BT3$, in view of our simulations, may have a tendency to overstate the presence of a common factor with a linear trend in small samples when there is one $I(1)$, zero mean common factor. However, in our case the sample sizes are sufficiently large, and there is clear evidence of having several common factors, which suggests that $BT3$ may be correct in indicating the presence of a common factor with a linear trend. As far as the other factors are concerned, both $BT2$ and $BT3$ indicate that there are zero mean $I(1)$ common factors; based on the discrepancy between these two criteria, it may be argued that two of such factors may be only borderline non-stationary. This evidence is in line with the findings from IC; conversely, $BT1$ seems to suggest only one common factor, which is at odds with the stylised factors in this literature where, usually, at least three factors are identified. To sum up, the results in Table 4 indicate the presence of five common factors, which we estimate as the principal components of X_t, using the covariance $T^{-2} \sum_{t=1}^{T} X_t X_t'$ and imposing the identifying constraint $\Lambda' \Lambda = N I_r$, (see Bai, 2004; Maciejowska, 2010).

The estimated factors are shown in Figure 2 (solid red lines). The first three factors appear to be non-stationary; in particular, as indicated by $BT3$, the first one does seem to be driven by a linear trend. This evidence is consistent with

[2] `https://fred.stlouisfed.org.`

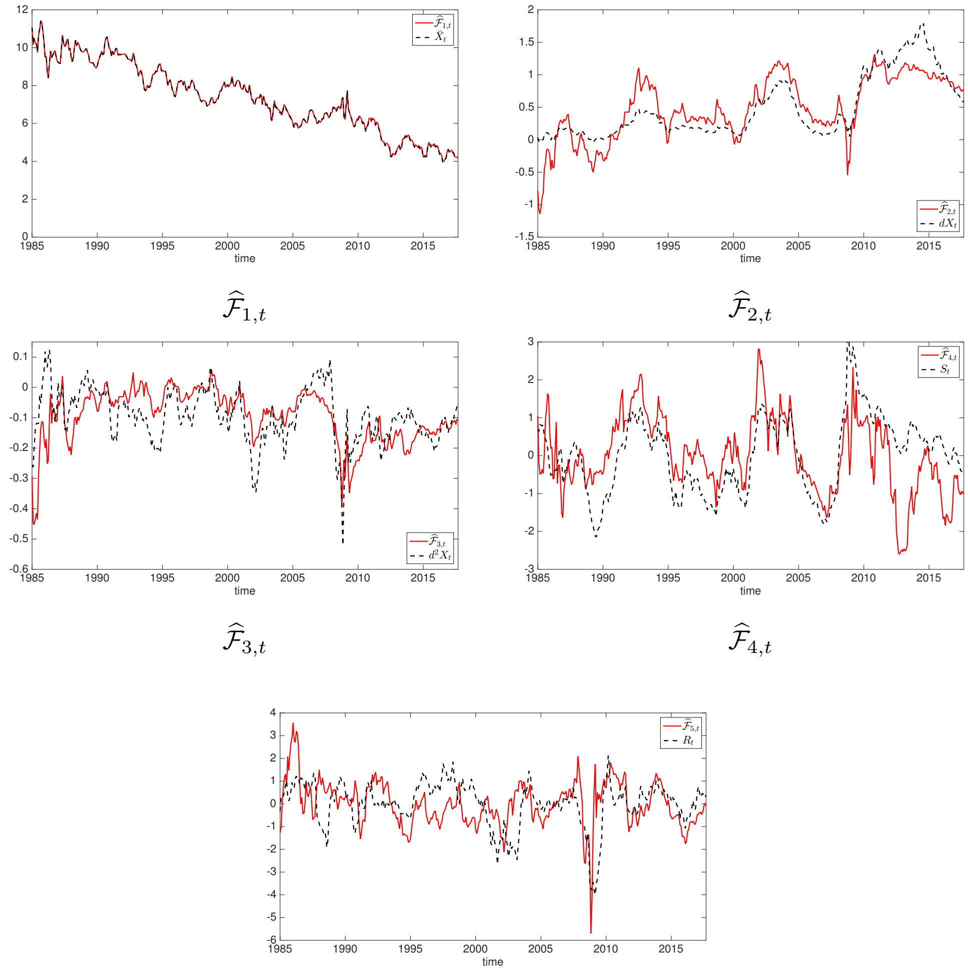
[3] `https://www.treasury.gov/resource-center/economic-policy/corp-bond-yie.`

Figure 1: HQM Corporate Bond Yield Curve



our findings in Table 4 (save for $BT1$), and it implies that the first three factors are highly persistent. In Figure 3 we report the autocorrelation of each estimated factor and the median, 5th and 95th percentiles of the autocorrelations of the idiosyncratic errors together with 95% confidence bands (dashed lines) computed as $\pm\frac{1.96}{\sqrt{T}}$. These results suggest that the fourth and fifth factor are nearly stationary,

Figure 2: Estimated and identified common factors $\widehat{\mathcal{F}}_{j,t}$ with proxies.



$\widehat{\mathcal{F}}_{1,t}$

$\widehat{\mathcal{F}}_{2,t}$

$\widehat{\mathcal{F}}_{3,t}$

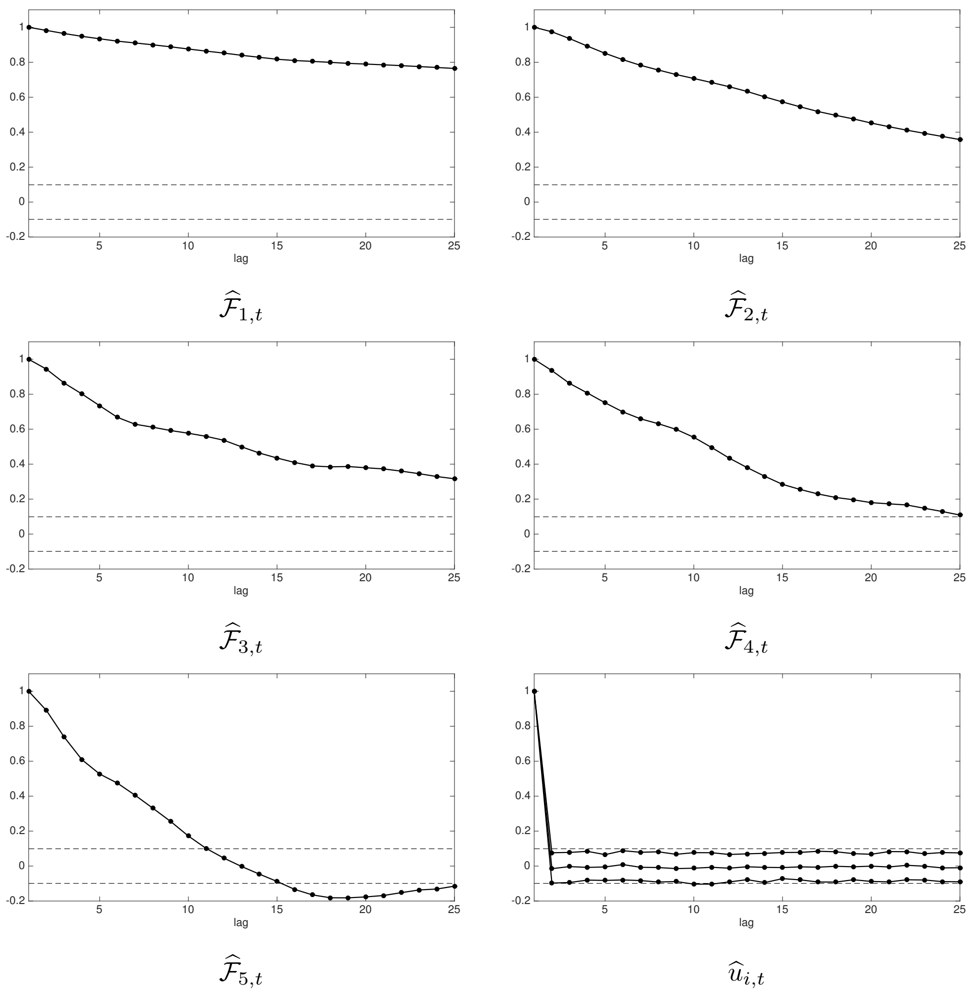
$\widehat{\mathcal{F}}_{4,t}$

$\widehat{\mathcal{F}}_{5,t}$

whilst the idiosyncratic component is clearly stationary since it shows no residual autocorrelation. The presence of common unit roots, and the stationarity of the idiosyncratic error imply cointegration, which in turn implies the factor structure in bond yields - see Dungey, Martin, and Pagan (2000).

Our findings can be contrasted with the stylised facts which are typically found in this literature. In particular, following Nelson and Siegel (1987), it is common to model yield curves by means of three common factors, which are usually interpreted as the level, slope, and curvature of the yield curve in a given time period t – see for example Dai and Singleton (2000) and Diebold and Li (2006). Moreover, when considering corporate bonds it common to find additional factors beyond the

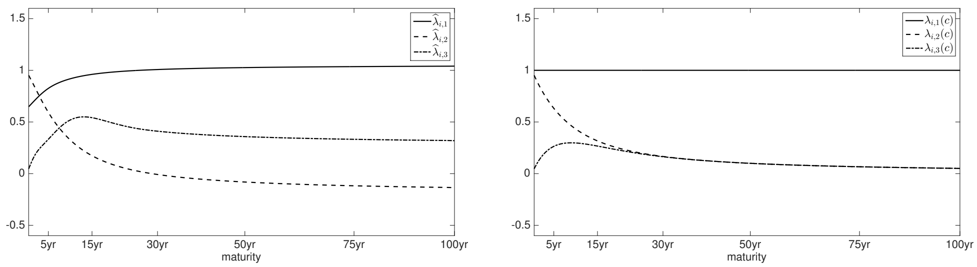
Figure 3: Autocorrelation of common factors $\widehat{\mathcal{F}}_{j,t}$ and idiosyncratic errors $\widehat{u}_{i,t}$.



classical first three – see for example Duffie and Singleton (1999), Duffie, Saita, and Wang (2007), and Christensen and Lopez (2008).

First we analyse the first three estimated common factors. Throughout we assume that at each point in time t, the N elements of X_t are ordered according to their maturity, thus $X_{1,t}$ is the shortest maturity (6 months), while $X_{N,t}$ is the longest maturity (100 years). We compare each estimated factor with a standard proxy as specified by Diebold, Rudebusch, and Aruoba (2006). Results are in the first three panels of Figure 2, where we show both the estimated factors (solid red lines) and the proxies (dashed black lines). In particular, in order to identify $\widehat{\mathcal{F}}_{1,t}$, we consider the proxy $\bar{X}_t = N^{-1}\sum_{i=1}^{N} X_{i,t}$; we found that $\mathrm{Corr}(\bar{X}_t, \widehat{\mathcal{F}}_{1,t}) \simeq 1$, which strongly suggests that $\widehat{\mathcal{F}}_{1,t}$ can be viewed as the *level* of the curve. Turning

Figure 4: Estimated and theoretical factor loadings.



$\widehat{\lambda}_{i,j}$, $j = 1, 2, 3$ $\lambda_{i,j}(c)$, $j = 1, 2, 3$ and $c = 0.2$

to $\widehat{\mathcal{F}}_{2,t}$, we use, as a proxy for the slope, $dX_t = N^{-1} \sum_{i=2}^{N} (\ln X_{i,t} - \ln X_{i-1,t}) = N^{-1}(\ln X_{N,t} - \ln X_{1,t})$. We find that $\text{Corr}(dX_t, \widehat{\mathcal{F}}_{2,t}) = .82$, which suggests that $\widehat{\mathcal{F}}_{2,t}$ can be interpreted as the *slope* of the term structure. Finally, we compare $\widehat{\mathcal{F}}_{3,t}$ to $d^2X_t = (N-2)^{-1} \sum_{i=2}^{N-1} (X_{i+1,t} - 2X_{i,t} + X_{i-1,t})$ as a proxy for the curvature; we find $\text{Corr}(d^2X_t, \widehat{\mathcal{F}}_{3,t}) = .53$, which shows some evidence that $\widehat{\mathcal{F}}_{3,t}$ can be interpreted as the *curvature*. Furthermore, according to Diebold and Li (2006), the first three elements of the i-th row of the loadings matrix Λ should be given by

$$\lambda_{i,1}(c) = 1, \qquad \lambda_{i,2}(c) = \left(\frac{1 - e^{-ci}}{ci} \right), \qquad \lambda_{i,3}(c) = \left(\frac{1 - e^{-ci}}{ci} - e^{-ci} \right), \tag{40}$$

for some $c > 0$. To confirm this finding, in Figure 4, we plot the estimated loadings $(\widehat{\lambda}_{i,1}, \widehat{\lambda}_{i,2}, \widehat{\lambda}_{i,3})$ (left panel) together with the theoretical curves in (40) computed for $c = 0.2$.

As far as the remaining two estimated common factors are concerned, we note that, in addition to level, slope and curvature, macroeconomic and financial factors have also been incorporated in the study of yield curves – see for example Estrella and Mishkin (1998), Ang and Piazzesi (2003), Diebold et al. (2006), Duffie et al. (2007), and Coroneo, Giannone, and Modugno (2016). We evaluate the correlation between S_t - the spread between the 10 years HQM bond rate and the Federal Funds rate - and the fourth factor finding that $\text{Corr}(S_t, \widehat{\mathcal{F}}_{4,t}) = .51$, whence we propose to interpret $\widehat{\mathcal{F}}_{4,t}$ as the *spread factor*. Also, letting R_t be the yearly returns of the Standard & Poor's index, we have $\text{Corr}(R_t, \widehat{\mathcal{F}}_{5,t}) = .30$; this seems to suggest that $\widehat{\mathcal{F}}_{5,t}$ may be viewed as a *financial factor*, or that, at a minimum, $\widehat{\mathcal{F}}_{5,t}$ is intimately related to the financial market.[4] These results are in line with the results by Duffie et al. (2007). In the last two panels of Figure 2 we report the fourth an fifth estimated factors (solid red lines) and the corresponding proxies (dashed black

[4]Data for S_t are available at `https://fred.stlouisfed.org`.
Data for R_t are available at `http://www.econ.yale.edu/~shiller/data.htm`.

lines).

6 Conclusions

In this paper, we propose a methodology to estimate the dimension of the common factor space for a given dataset $X_{i,t}$. We do not assume that the data are stationary or that they have (or not) linear trends: our procedure estimates separately the number of common factors with a linear trend (which can be only 0 or 1), the number of zero mean, $I(1)$ common factors, and the number of zero mean, $I(0)$ common factors.

Since estimation of these dimensions is carried out via testing (as opposed to using an information criterion or some other diagnostic), the results provide several interesting interpretations. For example, having $r_1 = 0$ means that the data have been tested for the presence of common linear trends, and none has been found; finding $r^* = 0$ indicates that the data have been tested for (the null of) non-stationarity, and have been found to be stationary; etc. Our methodology thus complements the results recently derived by Zhang et al. (2017).

Technically, our approach exploits the well-known eigenvalue separation property that characterises the covariance matrix of data with a common factor structure: essentially, the eigenvalues associated to common factors diverge to positive infinity, whereas the other ones are bounded. On top of this, we exploit the also well-known fact that linear trends, unit roots and stationary processes all imply different rates of divergence of the eigenvalues: these two facts allow us not merely to check whether there are common factors (and how many these are) but also to discriminate between those that have a trend, those that have a unit root, and the stationary ones. In this respect, our procedure is akin to the one proposed by Bai (2004) and Zhang et al. (2018), although it is based on tests rather than an information criterion, and it entertains the possibility that linear trends could be present.

Several interesting issues, in the analysis of high dimensional, possibly non-stationary, time series, remain outstanding. In model (1), we made no attempt to allow for the idiosyncratic components, $u_{i,t}$, to be non-stationary, thus relegating all the possible non-stationarity to the common component $\widetilde{\Lambda}\widetilde{F}_t$. Still, it would be worth considering the case where $u_{i,t} \sim I(1)$ for at least some i, so as to be able to disentangle common and idiosyncratic sources of non-stationarity. In such a case, our procedure would not be immediately applicable, since, chiefly, (15) would no longer hold. Also, by proper rescaling of the covariance matrix, our approach can be readily generalized to $I(d)$ factors with $d \geq 1$. These extensions are under current investigation by the authors.

References

Ahn, S. C. and A. R. Horenstein (2013). Eigenvalue ratio test for the number of factors. *Econometrica 81*, 1203–1227.

Alessi, L., M. Barigozzi, and M. Capasso (2010). Improved penalization for determining the number of factors in approximate static factor models. *Statistics and Probability Letters 80*, 1806–1813.

Ang, A. and M. Piazzesi (2003). A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables. *Journal of Monetary Economics 50*, 745–787.

Bai, J. (2004). Estimating cross-section common stochastic trends in nonstationary panel data. *Journal of Econometrics 122*, 137–183.

Bai, J. and C. Kao (2006). On the estimation and inference of a panel cointegration model with cross-sectional dependence. *Contributions to economic analysis 274*, 3–30.

Bai, J., C. Kao, and S. Ng (2009). Panel cointegration with global stochastic trends. *Journal of Econometrics 149*, 82–99.

Bai, J. and S. Ng (2002). Determining the number of factors in approximate factor models. *Econometrica 70*, 191–221.

Bai, J. and S. Ng (2004). A panic attack on unit roots and cointegration. *Econometrica 72*(4), 1127–1177.

Berkes, I., W. Liu, W. B. Wu, et al. (2014). Komlós–major–tusnády approximation under dependence. *The Annals of Probability 42*(2), 794–817.

Berkes, I. and W. Philipp (1979). Approximation thorems for independent and weakly dependent random vectors. *The Annals of Probability*, 29–54.

Chamberlain, G. and M. Rothschild (1983). Arbitrage, factor structure, and mean-variance analysis on large asset markets. *Econometrica*, 1281–1304.

Chen, L. and W. B. Wu (2018). Testing for trends in high-dimensional time series. *Journal of the American Statistical Association*. forthcoming.

Christensen, J. H. E. and J. A. Lopez (2008). Common risk factors in the US treasury and corporate bond markets: An arbitrage-free dynamic Nelson-Siegel modeling approach. Technical report, Federal Reserve Bank of San Francisco.

Coroneo, L., D. Giannone, and M. Modugno (2016). Unspanned macroeconomic factors in the yield curve. *Journal of Business & Economic Statistics 34*, 472–485.

Corradi, V. and N. R. Swanson (2006). The effects of data transformation on common cycle, cointegration, and unit root tests: Monte Carlo and a simple test. *Journal of Econometrics 132*, 195–229.

Dai, Q. and K. J. Singleton (2000). Specification analysis of affine term structure models. *The Journal of Finance 55*, 1943–1978.

Diebold, F. X. and C. Li (2006). Forecasting the term structure of government bond yields. *Journal of Econometrics 130*, 337–364.

Diebold, F. X., G. D. Rudebusch, and S. B. Aruoba (2006). The macroeconomy and the yield curve: a dynamic latent factor approach. *Journal of Econometrics 131*, 309–338.

Donsker, M. and S. Varadhan (1977). On laws of the iterated logarithm for local times. *Communications on Pure and Applied Mathematics 30*(6), 707–753.

Duffie, D., L. Saita, and K. Wang (2007). Multi-period corporate default prediction with stochastic covariates. *Journal of Financial Economics 83*, 635–665.

Duffie, D. and K. J. Singleton (1999). Modeling term structures of defaultable bonds. *The Review of Financial Studies 12*, 687–720.

Dungey, M., V. L. Martin, and A. R. Pagan (2000). A multivariate latent factor decomposition of international bond yield spreads. *Journal of Applied Econometrics 15*(6), 697–715.

Engel, C., N. C. Mark, and K. D. West (2015). Factor model forecasts of exchange rates. *Econometric Reviews 34*(1-2), 32–55.

Escribano, A. and D. Peña (1994). Cointegration and common factors. *Journal of Time Series Analysis 15*(6), 577–586.

Estrella, A. and F. S. Mishkin (1998). Predicting us recessions: Financial variables as leading indicators. *The Review of Economics and Statistics 80*, 45–61.

Gengenbach, C., F. C. Palm, and J.-P. Urbain (2009). Panel unit root tests in the presence of cross-sectional dependencies: comparison and implications for modelling. *Econometric Reviews 29*(2), 111–145.

Hallin, M. and R. Liška (2007). Determining the number of factors in the general dynamic factor model. *Journal of the American Statistical Association 102*, 603–617.

Kapetanios, G., M. H. Pesaran, and T. Yamagata (2011). Panels with non-stationary multifactor error structures. *Journal of Econometrics 160*(2), 326–348.

Lam, C. and Q. Yao (2012). Factor modeling for high-dimensional time series: inference for the number of factors. *The Annals of Statistics 40*(2), 694–726.

Lin, Z. and Z. Bai (2010). Probability inequalities of random variables. In *Probability Inequalities*, pp. 37–50. Springer.

Liu, W. and Z. Lin (2009). Strong approximation for a class of stationary processes. *Stochastic Processes and their Applications 119*(1), 249–280.

Maciejowska, K. (2010). Common factors in nonstationary panel data with a deterministic trend–estimation and distribution theory.

Moon, H. R. and B. Perron (2004). Testing for a unit root in panels with dynamic factors. *Journal of Econometrics 122*(1), 81–126.

Moon, H. R. and B. Perron (2007). An empirical analysis of nonstationarity in a panel of interest rates with factors. *Journal of Applied Econometrics 22*(2), 383–400.

Nelson, C. R. and A. F. Siegel (1987). Parsimonious modeling of yield curves. *Journal of Business 60*, 473–489.

Onatski, A. (2009). Testing hypotheses about the number of factors in large factor models. *Econometrica 77*, 1447–1479.

Onatski, A. (2010). Determining the number of factors from empirical distribution of eigenvalues. *The Review of Economics and Statistics 92*, 1004–1016.

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

Onatski, A. and C. Wang (2016). Alternative asymptotics for cointegration tests in large vars. *arXiv preprint arXiv:1605.08880*.

Pearson, E. (1950). On questions raised by the combination of tests based on discontinuous distributions. *Biometrika 37*, 383–398.

Peña, D. and P. Poncela (2006). Nonstationary dynamic factor analysis. *Journal of Statistical Planning and Inference 136*(4), 1237–1257.

Pesaran, M. H., L. V. Smith, and T. Yamagata (2013). Panel unit root tests in the presence of a multifactor error structure. *Journal of Econometrics 175*(2), 94–115.

Stock, J. H. and M. W. Watson (1988). Testing for common trends. *Journal of the American statistical Association 83*(404), 1097–1107.

Trapani, L. (2017). A randomised testing procedure to determine the number of factors. *Journal of the American Statistical Association*. forthcoming.

Wang, W. and J. Fan (2016). Asymptotics of empirical eigen-structure for high dimensional spiked covariance. *The Annals of Statistics 45*(3), 1342–1374.

Wu, W. B. (2007). Strong invariance principles for dependent random variables. *The Annals of Probability 35*, 2294–2320.

Zhang, B., G. Pan, and J. Gao (2017). CLT for largest eigenvalues and unit root tests for high-dimensional nonstationary time series. *The Annals of Statistics*. forthcoming.

Zhang, R., P. Robinson, and Q. Yao (2018). Identifying cointegration by eigenanalysis. *Journal of the American Statistical Association*. forthcoming.

A Additional numerical results

A.1 The case $\bar{\rho} = 0$

Table 5: Average estimated number of factors with linear trend, $\widehat{r}_1$.

		$N=50,\ T=100$			$N=100,\ T=100$			$N=200,\ T=100$		
r_1	r_2	$BT1$	$BT2$	$BT3$	$BT1$	$BT2$	$BT3$	$BT1$	$BT2$	$BT3$
0	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0	1	0.07	0.07	0.08	0.03	0.03	0.27	0.02	0.02	0.16
0	2	0.00	0.01	0.00	0.00	0.00	0.04	0.00	0.00	0.02
1	0	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1	1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1	2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		$N=100,\ T=200$			$N=200,\ T=200$			$N=200,\ T=500$		
r_1	r_2	$BT1$	$BT2$	$BT3$	$BT1$	$BT2$	$BT3$	$BT1$	$BT2$	$BT3$
0	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0	1	0.01	0.01	0.18	0.00	0.00	0.14	0.00	0.00	0.05
0	2	0.00	0.00	0.01	0.00	0.00	0.01	0.00	0.00	0.00
1	0	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1	1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1	2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table 6: Average estimated number of zero-mean $I(1)$ factors, $\widehat{r}_2$, when $r_1 = 0$.

		$N=50,\ T=100$				$N=100,\ T=100$				$N=200,\ T=100$			
r_2	r_3	$BT1$	$BT2$	$BT3$	IC	$BT1$	$BT2$	$BT3$	IC	$BT1$	$BT2$	$BT3$	IC
0	0	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00
0	1	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00
0	2	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00
1	0	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1	1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	1.00	1.00
1	2	1.00	1.00	1.00	1.00	0.99	0.99	1.00	1.00	0.99	0.99	1.00	1.00
2	0	1.97	1.99	1.98	2.00	1.96	1.99	1.99	2.00	1.94	1.99	1.99	2.00
2	1	1.86	1.87	1.85	2.00	1.91	1.98	1.99	1.99	1.84	1.93	1.98	2.00
2	2	1.91	1.91	1.91	1.99	1.90	1.97	1.99	2.00	1.86	1.94	1.99	2.00
		$N=100,\ T=200$				$N=200,\ T=200$				$N=200,\ T=500$			
r_2	r_3	$BT1$	$BT2$	$BT3$	IC	$BT1$	$BT2$	$BT3$	IC	$BT1$	$BT2$	$BT3$	IC
0	0	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00
0	1	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00
0	2	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00
1	0	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1	1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1	2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
2	0	2.00	2.00	2.00	2.00	1.98	2.00	2.00	2.00	1.99	2.00	2.00	2.00
2	1	1.97	1.99	2.00	2.00	1.97	1.99	2.00	2.00	2.00	2.00	2.00	2.00
2	2	1.98	1.99	2.00	2.00	1.97	1.98	2.00	2.00	1.99	1.99	2.00	2.00

Table 7: Average estimated number of zero-mean $I(1)$ factors, $\widehat{r}_2$, when $r_1 = 1$.

		$N=50$, $T=100$				$N=100$, $T=100$				$N=200$, $T=100$			
r_2	r_3	$BT1$	$BT2$	$BT3$	IC	$BT1$	$BT2$	$BT3$	IC	$BT1$	$BT2$	$BT3$	IC
0	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0	1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0	2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1	0	0.99	1.00	1.00	1.00	0.96	0.95	0.94	1.00	1.00	0.99	1.00	1.00
1	1	0.98	0.98	1.00	1.00	0.88	0.90	0.88	1.00	0.95	0.96	0.99	1.00
1	2	0.97	0.95	0.99	0.99	0.90	0.90	0.88	1.00	0.97	0.96	0.99	1.00
2	0	1.72	1.92	1.96	1.92	1.20	1.20	1.16	1.96	1.66	1.88	1.93	1.97
2	1	1.29	1.52	1.77	1.60	1.01	0.99	0.99	1.82	1.20	1.43	1.72	1.90
2	2	1.30	1.43	1.65	1.56	0.92	0.96	0.93	1.73	1.21	1.35	1.55	1.84
		$N=100$, $T=200$				$N=200$, $T=200$				$N=200$, $T=500$			
r_2	r_3	$BT1$	$BT2$	$BT3$	IC	$BT1$	$BT2$	$BT3$	IC	$BT1$	$BT2$	$BT3$	IC
0	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0	1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0	2	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1	0	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1	1	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1	2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
2	0	1.91	1.99	1.99	1.98	1.89	1.97	1.99	1.99	1.98	1.99	1.99	1.99
2	1	1.70	1.86	1.94	1.89	1.62	1.81	1.92	1.95	1.90	1.96	2.00	2.00
2	2	1.67	1.81	1.91	1.84	1.61	1.75	1.86	1.93	1.88	1.95	1.98	1.98

A.2 The case $\bar{\rho} = 0.8$

Table 8: Average estimated number of factors with linear trend, $\widehat{r}_1$.

		$N=50$, $T=100$			$N=100$, $T=100$			$N=200$, $T=100$		
r_1	r_2	$BT1$	$BT2$	$BT3$	$BT1$	$BT2$	$BT3$	$BT1$	$BT2$	$BT3$
0	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0	1	0.40	0.40	0.40	0.31	0.30	0.67	0.20	0.21	0.57
0	2	0.30	0.28	0.31	0.16	0.15	0.46	0.07	0.07	0.34
1	0	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1	1	1.00	1.00	1.00	1.00	1.00	1.00	0.99	0.99	1.00
1	2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		$N=100$, $T=200$			$N=200$, $T=200$			$N=200$, $T=500$		
r_1	r_2	$BT1$	$BT2$	$BT3$	$BT1$	$BT2$	$BT3$	$BT1$	$BT2$	$BT3$
0	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0	1	0.22	0.21	0.56	0.17	0.18	0.44	0.09	0.09	0.32
0	2	0.07	0.08	0.30	0.05	0.05	0.22	0.00	0.00	0.07
1	0	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1	1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1	2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table 9: Average estimated number of zero-mean $I(1)$ factors, $\widehat{r}_2$, when $r_1 = 0$.

		$N=50$, $T=100$				$N=100$, $T=100$				$N=200$, $T=100$			
r_2	r_3	$BT1$	$BT2$	$BT3$	IC	$BT1$	$BT2$	$BT3$	IC	$BT1$	$BT2$	$BT3$	IC
0	0	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00
0	1	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00
0	2	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00
1	0	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1	1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1	2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
2	0	2.00	2.00	2.00	2.00	1.99	2.00	2.00	2.00	1.99	2.00	1.99	2.00
2	1	2.00	1.99	1.99	2.00	1.99	2.00	2.00	2.00	1.99	2.00	2.00	2.00
2	2	1.99	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00
		$N=100$, $T=200$				$N=200$, $T=200$				$N=200$, $T=500$			
r_2	r_3	$BT1$	$BT2$	$BT3$	IC	$BT1$	$BT2$	$BT3$	IC	$BT1$	$BT2$	$BT3$	IC
0	0	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00
0	1	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00
0	2	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00
1	0	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1	1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1	2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
2	0	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00
2	1	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00
2	2	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00

Table 10: Average estimated number of zero-mean $I(1)$ factors, $\widehat{r}_2$, when $r_1 = 1$.

		$N=50, T=100$				$N=100, T=100$				$N=200, T=100$			
r_2	r_3	$BT1$	$BT2$	$BT3$	IC	$BT1$	$BT2$	$BT3$	IC	$BT1$	$BT2$	$BT3$	IC
0	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0	1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0	2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1	0	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1	1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1	2	1.00	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	1.00
2	0	1.96	1.99	2.00	1.99	1.96	1.98	1.99	2.00	1.93	1.97	1.99	1.99
2	1	1.85	1.91	1.98	1.94	1.84	1.94	1.98	1.97	1.82	1.93	1.97	1.99
2	2	1.84	1.91	1.94	1.93	1.89	1.93	1.96	1.96	1.81	1.89	1.95	1.98
		$N=100, T=200$				$N=200, T=200$				$N=200, T=500$			
r_2	r_3	$BT1$	$BT2$	$BT3$	IC	$BT1$	$BT2$	$BT3$	IC	$BT1$	$BT2$	$BT3$	IC
0	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0	1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0	2	0.00	0.00	-0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1	0	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1	1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1	2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
2	0	2.00	2.00	1.99	2.00	1.98	2.00	2.00	2.00	2.00	2.00	2.00	2.00
2	1	1.98	1.99	2.00	2.00	1.98	1.99	2.00	2.00	2.00	1.99	2.00	2.00
2	2	1.98	1.99	2.00	2.00	1.97	1.98	1.99	2.00	1.99	2.00	1.99	2.00

A.3 Robustness analysis

Table 11: Fraction of wrong detections when estimating r_1

$N = 200$, $T = 200$

			$R_1 = N$, $u = \sqrt{2}$, $\delta^* = 10^{-5}$			$R_1 = N$, $u = \sqrt{2}$, $\delta^* = 10^{-1}$		
r_1	r_2	$\widehat{r}_1$	$BT1$	$BT2$	$BT3$	$BT1$	$BT2$	$BT3$
1	0	0	0.000	0.000	0.000	0.000	0.000	0.000
1	1	0	0.000	0.000	0.000	0.000	0.000	0.000
1	2	0	0.000	0.000	0.002	0.000	0.002	0.002
0	0	1	0.000	0.000	0.000	0.000	0.000	0.000
0	1	1	0.022	0.024	0.252	0.002	0.002	0.104
0	2	1	0.000	0.000	0.026	0.000	0.000	0.002
			$R_1 = N$, $u = 5$, $\delta^* = 10^{-5}$			$R_1 = N$, $u = 5$, $\delta^* = 10^{-1}$		
r_1	r_2	$\widehat{r}_1$	$BT1$	$BT2$	$BT3$	$BT1$	$BT2$	$BT3$
1	0	0	0.000	0.000	0.000	0.000	0.000	0.000
1	1	0	0.000	0.000	0.000	0.398	0.412	0.000
1	2	0	0.000	0.002	0.002	0.412	0.388	0.002
0	0	1	0.000	0.000	0.000	0.000	0.000	0.000
0	1	1	0.000	0.000	0.086	0.000	0.000	0.024
0	2	1	0.000	0.000	0.000	0.000	0.000	0.000
			$R_1 = \lfloor N/2 \rfloor$, $u = \sqrt{2}$, $\delta^* = 10^{-5}$			$R_1 = \lfloor N/2 \rfloor$, $u = \sqrt{2}$, $\delta^* = 10^{-1}$		
r_1	r_2	$\widehat{r}_1$	$BT1$	$BT2$	$BT3$	$BT1$	$BT2$	$BT3$
1	0	0	0.000	0.000	0.000	0.000	0.000	0.000
1	1	0	0.000	0.002	0.000	0.000	0.000	0.000
1	2	0	0.000	0.000	0.000	0.000	0.002	0.000
0	0	1	0.000	0.000	0.000	0.000	0.000	0.000
0	1	1	0.068	0.062	0.326	0.018	0.012	0.192
0	2	1	0.002	0.004	0.050	0.000	0.000	0.010
			$R_1 = \lfloor N/2 \rfloor$, $u = 5$, $\delta^* = 10^{-5}$			$R_1 = \lfloor N/2 \rfloor$, $u = 5$, $\delta^* = 10^{-1}$		
r_1	r_2	$\widehat{r}_1$	$BT1$	$BT2$	$BT3$	$BT1$	$BT2$	$BT3$
1	0	0	0.000	0.000	0.000	0.000	0.000	0.000
1	1	0	0.000	0.000	0.000	0.134	0.148	0.000
1	2	0	0.000	0.002	0.000	0.124	0.162	0.000
0	0	1	0.000	0.000	0.000	0.000	0.000	0.000
0	1	1	0.006	0.004	0.128	0.000	0.000	0.038
0	2	1	0.000	0.000	0.008	0.000	0.000	0.000

Table 12: Fraction of wrong detections when estimating r_1

$N = 200$, $T = 500$

			$R_1 = N$, $u = \sqrt{2}$, $\delta^* = 10^{-5}$			$R_1 = N$, $u = \sqrt{2}$, $\delta^* = 10^{-1}$		
r_1	r_2	$\widehat{r}_1$	$BT1$	$BT2$	$BT3$	$BT1$	$BT2$	$BT3$
1	0	0	0.000	0.000	0.002	0.000	0.000	0.000
1	1	0	0.000	0.000	0.002	0.000	0.002	0.000
1	2	0	0.000	0.002	0.000	0.000	0.000	0.000
0	0	1	0.000	0.000	0.000	0.000	0.000	0.000
0	1	1	0.008	0.008	0.136	0.000	0.000	0.046
0	2	1	0.000	0.000	0.008	0.000	0.000	0.000
			$R_1 = N$, $u = 5$, $\delta^* = 10^{-5}$			$R_1 = N$, $u = 5$, $\delta^* = 10^{-1}$		
r_1	r_2	$\widehat{r}_1$	$BT1$	$BT2$	$BT3$	$BT1$	$BT2$	$BT3$
1	0	0	0.000	0.000	0.002	0.000	0.000	0.002
1	1	0	0.000	0.000	0.002	0.000	0.000	0.002
1	2	0	0.000	0.002	0.000	0.000	0.000	0.000
0	0	1	0.000	0.000	0.000	0.000	0.000	0.000
0	1	1	0.000	0.000	0.032	0.000	0.000	0.006
0	2	1	0.000	0.000	0.000	0.000	0.000	0.000
			$R_1 = \lfloor N/2 \rfloor$, $u = \sqrt{2}$, $\delta^* = 10^{-5}$			$R_1 = \lfloor N/2 \rfloor$, $u = \sqrt{2}$, $\delta^* = 10^{-1}$		
r_1	r_2	$\widehat{r}_1$	$BT1$	$BT2$	$BT3$	$BT1$	$BT2$	$BT3$
1	0	0	0.000	0.000	0.000	0.000	0.000	0.000
1	1	0	0.000	0.000	0.000	0.000	0.002	0.000
1	2	0	0.000	0.000	0.000	0.000	0.000	0.000
0	0	1	0.000	0.000	0.000	0.000	0.000	0.000
0	1	1	0.020	0.028	0.198	0.006	0.002	0.088
0	2	1	0.000	0.000	0.008	0.000	0.000	0.000
			$R_1 = \lfloor N/2 \rfloor$, $u = 5$, $\delta^* = 10^{-5}$			$R_1 = \lfloor N/2 \rfloor$, $u = 5$, $\delta^* = 10^{-1}$		
r_1	r_2	$\widehat{r}_1$	$BT1$	$BT2$	$BT3$	$BT1$	$BT2$	$BT3$
1	0	0	0.000	0.000	0.000	0.000	0.000	0.000
1	1	0	0.000	0.000	0.000	0.000	0.002	0.000
1	2	0	0.000	0.000	0.000	0.000	0.000	0.000
0	0	1	0.000	0.000	0.000	0.000	0.000	0.000
0	1	1	0.002	0.002	0.052	0.000	0.000	0.016
0	2	1	0.000	0.000	0.000	0.000	0.000	0.000

Table 13: Fraction of wrong detections when estimating r_2

$N = 200$, $T = 200$, $r_1 = 0$ - part 1

			$R_2 = N$ $(p=1)$, $R_2 = \lfloor N/3 \rfloor$ $(p>1)$ $u=\sqrt{2}$, $\delta^* = 10^{-5}$			$R_2 = N$ $(p=1)$, $R_2 = \lfloor N/3 \rfloor$ $(p>1)$ $u=\sqrt{2}$, $\delta^* = 10^{-1}$		
r_2	r_3	$\widehat{r}_2$	$BT1$	$BT2$	$BT3$	$BT1$	$BT2$	$BT3$
1	0	$< r_2$	0.000	0.000	0.000	0.000	0.000	0.000
1	1	$< r_2$	0.000	0.002	0.000	0.002	0.000	0.000
1	2	$< r_2$	0.000	0.000	0.000	0.000	0.000	0.000
2	0	$< r_2$	0.002	0.002	0.002	0.024	0.002	0.002
2	1	$< r_2$	0.014	0.004	0.004	0.040	0.012	0.006
2	2	$< r_2$	0.002	0.002	0.000	0.020	0.006	0.002
			$R_2 = N$ $(p=1)$, $R_2 = \lfloor N/3 \rfloor$ $(p>1)$ $u=5$, $\delta^* = 10^{-5}$			$R_2 = N$ $(p=1)$, $R_2 = \lfloor N/3 \rfloor$ $(p>1)$ $u=5$, $\delta^* = 10^{-1}$		
r_2	r_3	$\widehat{r}_2$	$BT1$	$BT2$	$BT3$	$BT1$	$BT2$	$BT3$
1	0	$< r_2$	0.000	0.000	0.000	0.002	0.002	0.000
1	1	$< r_2$	0.002	0.002	0.000	0.018	0.018	0.002
1	2	$< r_2$	0.000	0.000	0.000	0.006	0.008	0.000
2	0	$< r_2$	0.042	0.002	0.002	0.116	0.020	0.014
2	1	$< r_2$	0.132	0.032	0.012	0.406	0.176	0.042
2	2	$< r_2$	0.072	0.024	0.008	0.342	0.164	0.076
			$R_2 = \lfloor N/2 \rfloor$ $(p=1)$, $R_2 = \lfloor N/3 \rfloor$ $(p>1)$ $u=\sqrt{2}$, $\delta^* = 10^{-5}$			$R_2 = \lfloor N/2 \rfloor$ $(p=1)$, $R_2 = \lfloor N/3 \rfloor$ $(p>1)$ $u=\sqrt{2}$, $\delta^* = 10^{-1}$		
r_2	r_3	$\widehat{r}_2$	$BT1$	$BT2$	$BT3$	$BT1$	$BT2$	$BT3$
1	0	$< r_2$	0.000	0.000	0.000	0.000	0.000	0.000
1	1	$< r_2$	0.000	0.000	0.000	0.000	0.000	0.000
1	2	$< r_2$	0.000	0.002	0.000	0.000	0.002	0.000
2	0	$< r_2$	0.006	0.000	0.000	0.020	0.000	0.000
2	1	$< r_2$	0.006	0.002	0.002	0.028	0.008	0.002
2	2	$< r_2$	0.004	0.002	0.004	0.034	0.016	0.006
			$R_2 = \lfloor N/2 \rfloor$ $(p=1)$, $R_2 = \lfloor N/3 \rfloor$ $(p>1)$ $u=5$, $\delta^* = 10^{-5}$			$R_2 = \lfloor N/2 \rfloor$ $(p=1)$, $R_2 = \lfloor N/3 \rfloor$ $(p>1)$ $u=5$, $\delta^* = 10^{-1}$		
r_2	r_3	$\widehat{r}_2$	$BT1$	$BT2$	$BT3$	$BT1$	$BT2$	$BT3$
1	0	$< r_2$	0.000	0.000	0.000	0.000	0.000	0.000
1	1	$< r_2$	0.004	0.002	0.000	0.024	0.022	0.004
1	2	$< r_2$	0.000	0.002	0.000	0.010	0.010	0.000
2	0	$< r_2$	0.034	0.004	0.002	0.096	0.018	0.014
2	1	$< r_2$	0.100	0.022	0.006	0.382	0.134	0.022
2	2	$< r_2$	0.110	0.052	0.018	0.350	0.198	0.098

Table 14: Fraction of wrong detections when estimating r_2

$N = 200$, $T = 200$, $r_1 = 0$ - part 2

			$R_2 = N$ $(p = 1)$, $R_2 = N$ $(p > 1)$ $u = \sqrt{2}$, $\delta^* = 10^{-5}$			$R_2 = N$ $(p = 1)$, $R_2 = N$ $(p > 1)$ $u = \sqrt{2}$, $\delta^* = 10^{-1}$		
r_2	r_3	$\widehat{r}_2$	$BT1$	$BT2$	$BT3$	$BT1$	$BT2$	$BT3$
1	0	$< r_2$	0.000	0.000	0.000	0.000	0.000	0.000
1	1	$< r_2$	0.000	0.000	0.000	0.000	0.000	0.000
1	2	$< r_2$	0.000	0.000	0.000	0.000	0.000	0.000
2	0	$< r_2$	0.028	0.000	0.000	0.052	0.014	0.004
2	1	$< r_2$	0.030	0.014	0.002	0.140	0.044	0.012
2	2	$< r_2$	0.022	0.002	0.000	0.136	0.054	0.016
			$R_2 = N$ $(p = 1)$, $R_2 = N$ $(p > 1)$ $u = 5$, $\delta^* = 10^{-5}$			$R_2 = N$ $(p = 1)$, $R_2 = N$ $(p > 1)$ $u = 5$, $\delta^* = 10^{-1}$		
r_2	r_3	$\widehat{r}_2$	$BT1$	$BT2$	$BT3$	$BT1$	$BT2$	$BT3$
1	0	$< r_2$	0.000	0.000	0.000	0.000	0.000	0.000
1	1	$< r_2$	0.000	0.000	0.000	0.010	0.012	0.004
1	2	$< r_2$	0.000	0.000	0.000	0.026	0.026	0.000
2	0	$< r_2$	0.062	0.016	0.008	0.178	0.034	0.026
2	1	$< r_2$	0.172	0.048	0.014	0.512	0.218	0.060
2	2	$< r_2$	0.172	0.076	0.018	0.522	0.312	0.152
			$R_2 = \lfloor N/2 \rfloor$ $(p = 1)$, $R_2 = \lfloor N/2 \rfloor$ $(p > 1)$ $u = \sqrt{2}$, $\delta^* = 10^{-5}$			$R_2 = \lfloor N/2 \rfloor$ $(p = 1)$, $R_2 = \lfloor N/2 \rfloor$ $(p > 1)$ $u = \sqrt{2}$, $\delta^* = 10^{-1}$		
r_2	r_3	$\widehat{r}_2$	$BT1$	$BT2$	$BT3$	$BT1$	$BT2$	$BT3$
1	0	$< r_2$	0.000	0.000	0.000	0.000	0.000	0.000
1	1	$< r_2$	0.000	0.000	0.000	0.000	0.000	0.000
1	2	$< r_2$	0.000	0.000	0.000	0.000	0.000	0.000
2	0	$< r_2$	0.006	0.000	0.000	0.022	0.002	0.000
2	1	$< r_2$	0.016	0.004	0.002	0.082	0.020	0.004
2	2	$< r_2$	0.010	0.010	0.000	0.072	0.032	0.008
			$R_2 = \lfloor N/2 \rfloor$ $(p = 1)$, $R_2 = \lfloor N/2 \rfloor$ $(p > 1)$ $u = 5$, $\delta^* = 10^{-5}$			$R_2 = \lfloor N/2 \rfloor$ $(p = 1)$, $R_2 = \lfloor N/2 \rfloor$ $(p > 1)$ $u = 5$, $\delta^* = 10^{-1}$		
r_2	r_3	$\widehat{r}_2$	$BT1$	$BT2$	$BT3$	$BT1$	$BT2$	$BT3$
1	0	$< r_2$	0.000	0.000	0.000	0.000	0.000	0.000
1	1	$< r_2$	0.000	0.000	0.000	0.008	0.008	0.000
1	2	$< r_2$	0.000	0.000	0.000	0.004	0.004	0.000
2	0	$< r_2$	0.032	0.006	0.002	0.106	0.016	0.010
2	1	$< r_2$	0.138	0.036	0.008	0.428	0.186	0.048
2	2	$< r_2$	0.122	0.058	0.022	0.418	0.240	0.106

Table 15: Fraction of wrong detections when estimating r_2

$N = 200$, $T = 500$, $r_1 = 0$ - part 1

			$R_2 = N$ $(p = 1)$, $R_2 = \lfloor N/3 \rfloor$ $(p > 1)$ $u = \sqrt{2}$, $\delta^* = 10^{-5}$			$R_2 = N$ $(p = 1)$, $R_2 = \lfloor N/3 \rfloor$ $(p > 1)$ $u = \sqrt{2}$, $\delta^* = 10^{-1}$		
r_2	r_3	$\widehat{r}_2$	$BT1$	$BT2$	$BT3$	$BT1$	$BT2$	$BT3$
1	0	$< r_2$	0.000	0.000	0.002	0.000	0.000	0.002
1	1	$< r_2$	0.002	0.002	0.000	0.002	0.004	0.000
1	2	$< r_2$	0.000	0.000	0.000	0.000	0.000	0.000
2	0	$< r_2$	0.006	0.000	0.000	0.008	0.000	0.002
2	1	$< r_2$	0.000	0.000	0.000	0.008	0.000	0.000
2	2	$< r_2$	0.002	0.004	0.002	0.010	0.006	0.002
			$R_2 = N$ $(p = 1)$, $R_2 = \lfloor N/3 \rfloor$ $(p > 1)$ $u = 5$, $\delta^* = 10^{-5}$			$R_2 = N$ $(p = 1)$, $R_2 = \lfloor N/3 \rfloor$ $(p > 1)$ $u = 5$, $\delta^* = 10^{-1}$		
r_2	r_3	$\widehat{r}_2$	$BT1$	$BT2$	$BT3$	$BT1$	$BT2$	$BT3$
1	0	$< r_2$	0.000	0.000	0.002	0.000	0.000	0.002
1	1	$< r_2$	0.002	0.004	0.000	0.004	0.004	0.002
1	2	$< r_2$	0.000	0.000	0.000	0.000	0.000	0.000
2	0	$< r_2$	0.014	0.004	0.002	0.032	0.006	0.006
2	1	$< r_2$	0.026	0.004	0.000	0.124	0.036	0.004
2	2	$< r_2$	0.022	0.012	0.004	0.094	0.054	0.014
			$R_2 = \lfloor N/2 \rfloor$ $(p = 1)$, $R_2 = \lfloor N/3 \rfloor$ $(p > 1)$ $u = \sqrt{2}$, $\delta^* = 10^{-5}$			$R_2 = \lfloor N/2 \rfloor$ $(p = 1)$, $R_2 = \lfloor N/3 \rfloor$ $(p > 1)$ $u = \sqrt{2}$, $\delta^* = 10^{-1}$		
r_2	r_3	$\widehat{r}_2$	$BT1$	$BT2$	$BT3$	$BT1$	$BT2$	$BT3$
1	0	$< r_2$	0.000	0.000	0.000	0.000	0.000	0.000
1	1	$< r_2$	0.000	0.000	0.000	0.000	0.000	0.000
1	2	$< r_2$	0.000	0.002	0.000	0.000	0.002	0.000
2	0	$< r_2$	0.002	0.000	0.000	0.002	0.000	0.000
2	1	$< r_2$	0.004	0.000	0.000	0.008	0.004	0.000
2	2	$< r_2$	0.000	0.000	0.002	0.002	0.000	0.002
			$R_2 = \lfloor N/2 \rfloor$ $(p = 1)$, $R_2 = \lfloor N/3 \rfloor$ $(p > 1)$ $u = 5$, $\delta^* = 10^{-5}$			$R_2 = \lfloor N/2 \rfloor$ $(p = 1)$, $R_2 = \lfloor N/3 \rfloor$ $(p > 1)$ $u = 5$, $\delta^* = 10^{-1}$		
r_2	r_3	$\widehat{r}_2$	$BT1$	$BT2$	$BT3$	$BT1$	$BT2$	$BT3$
1	0	$< r_2$	0.000	0.000	0.000	0.000	0.000	0.000
1	1	$< r_2$	0.000	0.000	0.000	0.000	0.000	0.000
1	2	$< r_2$	0.000	0.002	0.000	0.000	0.002	0.000
2	0	$< r_2$	0.010	0.000	0.000	0.034	0.002	0.000
2	1	$< r_2$	0.024	0.004	0.000	0.094	0.030	0.004
2	2	$< r_2$	0.012	0.000	0.002	0.088	0.030	0.008

Table 16: Fraction of wrong detections when estimating r_2

$N = 200$, $T = 500$, $r_1 = 0$ - part 2

			$R_2 = N$ $(p = 1)$, $R_2 = N$ $(p > 1)$ $u = \sqrt{2}$, $\delta^* = 10^{-5}$			$R_2 = N$ $(p = 1)$, $R_2 = N$ $(p > 1)$ $u = \sqrt{2}$, $\delta^* = 10^{-1}$		
r_2	r_3	$\widehat{r}_2$	$BT1$	$BT2$	$BT3$	$BT1$	$BT2$	$BT3$
1	0	$< r_2$	0.000	0.000	0.000	0.000	0.000	0.000
1	1	$< r_2$	0.000	0.000	0.000	0.000	0.000	0.000
1	2	$< r_2$	0.000	0.000	0.002	0.000	0.000	0.002
2	0	$< r_2$	0.004	0.000	0.000	0.018	0.000	0.000
2	1	$< r_2$	0.008	0.002	0.000	0.020	0.008	0.000
2	2	$< r_2$	0.004	0.000	0.000	0.034	0.008	0.000
			$R_2 = N$ $(p = 1)$, $R_2 = N$ $(p > 1)$ $u = 5$, $\delta^* = 10^{-5}$			$R_2 = N$ $(p = 1)$, $R_2 = N$ $(p > 1)$ $u = 5$, $\delta^* = 10^{-1}$		
r_2	r_3	$\widehat{r}_2$	$BT1$	$BT2$	$BT3$	$BT1$	$BT2$	$BT3$
1	0	$< r_2$	0.000	0.000	0.000	0.000	0.000	0.000
1	1	$< r_2$	0.000	0.000	0.000	0.002	0.002	0.000
1	2	$< r_2$	0.000	0.000	0.002	0.000	0.002	0.002
2	0	$< r_2$	0.022	0.000	0.000	0.054	0.006	0.000
2	1	$< r_2$	0.030	0.012	0.002	0.152	0.044	0.012
2	2	$< r_2$	0.038	0.010	0.002	0.160	0.074	0.026
			$R_2 = \lfloor N/2 \rfloor$ $(p = 1)$, $R_2 = \lfloor N/2 \rfloor$ $(p > 1)$ $u = \sqrt{2}$, $\delta^* = 10^{-5}$			$R_2 = \lfloor N/2 \rfloor$ $(p = 1)$, $R_2 = \lfloor N/2 \rfloor$ $(p > 1)$ $u = \sqrt{2}$, $\delta^* = 10^{-1}$		
r_2	r_3	$\widehat{r}_2$	$BT1$	$BT2$	$BT3$	$BT1$	$BT2$	$BT3$
1	0	$< r_2$	0.002	0.000	0.000	0.002	0.000	0.000
1	1	$< r_2$	0.000	0.002	0.000	0.000	0.002	0.000
1	2	$< r_2$	0.000	0.000	0.000	0.000	0.000	0.000
2	0	$< r_2$	0.002	0.000	0.000	0.010	0.000	0.002
2	1	$< r_2$	0.004	0.004	0.002	0.012	0.004	0.002
2	2	$< r_2$	0.000	0.000	0.000	0.006	0.002	0.000
			$R_2 = \lfloor N/2 \rfloor$ $(p = 1)$, $R_2 = \lfloor N/2 \rfloor$ $(p > 1)$ $u = 5$, $\delta^* = 10^{-5}$			$R_2 = \lfloor N/2 \rfloor$ $(p = 1)$, $R_2 = \lfloor N/2 \rfloor$ $(p > 1)$ $u = 5$, $\delta^* = 10^{-1}$		
r_2	r_3	$\widehat{r}_2$	$BT1$	$BT2$	$BT3$	$BT1$	$BT2$	$BT3$
1	0	$< r_2$	0.002	0.000	0.000	0.002	0.000	0.000
1	1	$< r_2$	0.000	0.002	0.000	0.000	0.002	0.000
1	2	$< r_2$	0.000	0.000	0.000	0.000	0.000	0.000
2	0	$< r_2$	0.014	0.002	0.002	0.032	0.008	0.006
2	1	$< r_2$	0.018	0.006	0.006	0.122	0.038	0.010
2	2	$< r_2$	0.010	0.004	0.002	0.116	0.040	0.012

B Proofs

B.1 Preliminary lemmas

Henceforth, $\nu^{(p)}(A)$ represent the eigenvalues, sorted in decreasing order, for a matrix A; we occasionally employ the notation $\nu^{(\min)}(A)$ to denote the smallest eigenvalue of A. Also, "$\stackrel{D}{=}$" denotes equality in distribution. We also use the following matrix notation

$$
\begin{aligned}
X_t &= \Lambda^{(1)} f_t^{(1)} + u_t^{(1)} \\
&= \Lambda^{(1)} f_t^{(1)} + \Lambda^{(2)} f_t^{(2)} + u_t^{(2)} \\
&= \Lambda^{(1)} f_t^{(1)} + \Lambda^{(2)} f_t^{(2)} + \Lambda^{(3)} f_t^{(3)} + u_t.
\end{aligned}
$$

As far as the notation is concerned, $\Lambda^{(1)}$ is $N \times r_1$; $\Lambda^{(2)}$ is $N \times r_2$; and, finally, $\Lambda^{(3)}$ is $N \times r_3$.

We begin with the following lemma, which is useful to derive almost sure rates.

Lemma B1. *Consider a multi-index random variable $U_{i_1,...,i_h}$, with $1 \leq i_1 \leq S_1$, $1 \leq i_2 \leq S_2$, etc... Assume that*

$$
\sum_{S_1} \cdot \cdot \sum_{S_h} \frac{1}{S_1 \cdot ... \cdot S_h} P\left(\max_{1 \leq i_1 \leq S_1,...,1 \leq i_h \leq S_h} |U_{i_1,...,i_h}| > \epsilon L_{S_1,...,S_h} \right) < \infty, \tag{B1}
$$

for some $\epsilon > 0$ and a sequence $L_{S_1,...,S_h}$ defined as

$$
L_{S_1,...,S_h} = S_1^{d_1} \cdot ... \cdot S_h^{d_h} l_1(S_1) \cdot ... l_h(S_h),
$$

where d_1, d_2, etc. are non-negative numbers and $l_1(\cdot)$, $l_2(\cdot)$, etc. are slowly varying functions in the sense of Karamata. Then it holds that

$$
\lim \sup_{(S_1,...,S_h) \to \infty} \frac{|U_{S_1,...,S_h}|}{L_{S_1,...,S_h}} = 0 \ a.s. \tag{B2}
$$

Proof. The proof follows similar arguments as the proof of Lemma 2 inTrapani (2017) - see also **?**. We begin by noting that, for every h-tuple $(S_1, ..., S_h)$, there is a h-tuple of integers $(k_1, ..., k_h)$ such that $2^{k_1} \leq S_1 < 2^{k_1+1}$, $2^{k_2} \leq S_2 < 2^{k_2+1}$, etc. Similarly, there is a h-tuple of real numbers defined over $[0, 1)$, say $(\rho_1, ..., \rho_h)$, such that $2^{k_1+\rho_1} = S_1$, $2^{k_2+\rho_2} = S_2$, etc. Consider now the short-hand notation

$$
L_{k_1,...,k_h} = \left(2^{k_1+1}\right)^{d_1} \cdot ... \cdot \left(2^{k_h+1}\right)^{d_h} l_1(S_1) \cdot ... l_h(S_h),
$$

$$
P\left(\max_{1 \leq i_1 \leq S_1,...,1 \leq i_h \leq S_h} |U_{i_1,...,i_h}| > \epsilon L_{k_1,...,k_h} \right) = P_{k_1,...,k_h};
$$

by (B1), we have

$$
\sum_{k_1=0}^{\infty} \cdot ... \cdot \sum_{k_h=0}^{\infty} \frac{2^{k_1+1} \cdot ... \cdot 2^{k_h+1}}{(2^{k_1+1}-1) \cdot ... \cdot (2^{k_h+1}-1)} P_{k_1,...,k_h} < \infty.
$$

This, in turn, entails that

$$\sum_{k_1=0}^{\infty} \cdot ... \cdot \sum_{k_h=0}^{\infty} P_{k_1,...,k_h} \leq \sum_{k_1=0}^{\infty} \cdot ... \cdot \sum_{k_h=0}^{\infty} \frac{2^{k_1+1} \cdot ... \cdot 2^{k_h+1}}{(2^{k_1+1}-1) \cdot ... \cdot (2^{k_h+1}-1)} P_{k_1,...,k_h} < \infty;$$

thus, by the Borel-Cantelli Lemma

$$\frac{\max_{1\leq i_1\leq S_1,...,1\leq i_h\leq S_h} |U_{i_1,...,i_h}|}{L_{S_1,...,S_h}} \to 0 \text{ a.s.}$$

Therefore we have

$$\begin{aligned} \frac{|U_{S_1,...,S_h}|}{L_{S_1,...,S_h}} &\leq \frac{\max_{1\leq i_1\leq S_1,...,1\leq i_h\leq S_h} |U_{i_1,...,i_h}|}{L_{k_1,...,k_h}} \frac{L_{k_1,...,k_h}}{L_{S_1,...,S_h}} \\ &\leq C \frac{\max_{1\leq i_1\leq S_1,...,1\leq i_h\leq S_h} |U_{i_1,...,i_h}|}{L_{k_1,...,k_h}} \to 0 \text{ a.s.,} \end{aligned}$$

which, finally, implies (B2). □

Let now $\gamma^{(p)}$ and $\omega^{(p)}$ denote the p-th largest eigenvalues of $\Lambda T^{-1} \sum_{t=1}^{T} E(\Delta f_t \Delta f_t') \Lambda'$ and $T^{-1} \sum_{t=1}^{T} E(\Delta u_t \Delta u_t')$ respectively. By Assumption 5, it can be easily verified using the arguments in the proof of Lemma 1 in Trapani (2017) that $\gamma^{(p)} = C_p N$ for $1 \leq p \leq r$; $\omega^{(1)} \leq C_1$; and $\liminf_{N\to\infty} \omega^{(N)} > 0$.

We will often need the following lemma, shown in Trapani (2017) (see Lemma A1), which we report here for convenience.

Lemma B2. *Under Assumption 5, it holds that, as* $\min(N,T) \to \infty$

$$\begin{aligned} \limsup_{N\to\infty} \overline{\nu}_{3,p}(k) &= \overline{\nu}_{3,p}^{U}(k) < \infty, \\ \liminf_{N\to\infty} \overline{\nu}_{3,p}(k) &= \overline{\nu}_{3,p}^{L}(k) > 0, \end{aligned}$$

for every p *and* k*, where* $\overline{\nu}_{3,p}(k)$ *is defined in equation (19).*

Proof. We begin by showing that

$$\limsup_{N\to\infty} \frac{1}{N-k+1} \sum_{h=k}^{N} \nu_3^{(h)} = \overline{\nu}_{3,p}^{U}(k) < \infty, \tag{B3}$$

$$\liminf_{N\to\infty} \frac{1}{N-k+1} \sum_{h=k}^{N} \nu_3^{(h)} = \overline{\nu}_{3,p}^{L}(k) > 0. \tag{B4}$$

Letting

$$\overline{\nu}_{3,p}(k) = \frac{1}{N-k+1} \sum_{h=k}^{N} \nu_3^{(h)},$$

note that, by Weyl's inequalities, we have $\gamma^{(h)} + \omega^{(N)} \leq \nu_3^{(h)} \leq \gamma^{(h)} + \omega^{(1)}$. Thus

$$\omega^{(N)} + \frac{1}{N-k+1} \sum_{h=k}^{N} \gamma^{(h)} \leq \overline{\nu}_{3,p}(k) \leq \omega^{(1)} + \frac{1}{N-k+1} \sum_{h=k}^{N} \gamma^{(h)}. \tag{B5}$$

Assumption 5 implies that

$$0 \leq \frac{1}{N-k+1} \sum_{h=k}^{N} \gamma^{(h)} \leq C_{k+1} < \infty,$$

so that (B5) becomes

$$\omega^{(N)} \leq \overline{\nu}_{3,p}(k) \leq C_0 + C_{k+1},$$

whence (B3) and (B4) follow for each k. Hereafter, the proof is exactly the same as that of Lemma A1 in Trapani (2017) and thus omitted. □

Lemma B3. *Under Assumption 2, it holds that*

$$\lim\inf_{T\to\infty} \frac{\ln\ln T}{T^2} \sum_{t=1}^{T} f_t^* f_t^{*\prime} = D \quad a.s.,$$

where D is a positive definite matrix of dimension $[r_2 + r_1(1-d_1)d_2] \times [r_2 + r_1(1-d_1)d_2]$.

Proof. We have

$$\begin{aligned}
f_t^* f_t^{*\prime} &= \left(f_t^* \pm \Sigma_{\Delta f^*}^{1/2} W(t)\right)\left(f_t^* \pm \Sigma_{\Delta f^*}^{1/2} W(t)\right)' \\
&= \Sigma_{\Delta f^*}^{1/2} W(t) W(t)' \Sigma_{\Delta f^*}^{1/2} + \Sigma_{\Delta f^*}^{1/2} W(t)\left(f_t^* - \Sigma_{\Delta f^*}^{1/2} W(t)\right)' + \left(f_t^* - \Sigma_{\Delta f^*}^{1/2} W(t)\right) W(t)' \Sigma_{\Delta f^*}^{1/2} \\
&\quad + \left(f_t^* - \Sigma_{\Delta f^*}^{1/2} W(t)\right)\left(f_t^* - \Sigma_{\Delta f^*}^{1/2} W(t)\right)'. \\
&= I + II + III + IV.
\end{aligned}$$

Let b be a nonzero vector of dimension $r_1 + r_2$, such that $\|b\| < \infty$. We will prove that

$$\lim\inf_{T\to\infty} \frac{\ln\ln T}{T^2} \sum_{t=1}^{T} b' f_t^* f_t^{*\prime} b > 0 \quad a.s.,$$

for every b, thus proving the lemma. Clearly

$$\frac{\ln\ln T}{T^2} \sum_{t=1}^{T} b' \left(f_t^* - \Sigma_{\Delta f^*}^{1/2} W(t)\right)\left(f_t^* - \Sigma_{\Delta f^*}^{1/2} W(t)\right)' b \leq C_0 T^{-2\epsilon} \ln\ln T = o_{a.s.}(1),$$

by Assumption 2*(iv)*. This entails that IV is dominated. Consider now II and III. By the Law of the Iterated Logarithm (henceforth, LIL), we have that there exists a random t_0 such that, for all $t \geq t_0$, there exists a positive finite constant C_0 such that $\|W(t)\|^2 \leq$

$C_0 t^{1/2}(\ln\ln t)^{1/2}$. Thus, using Assumption 2*(iv)*

$$\begin{aligned} & \frac{\ln\ln T}{T^2}\sum_{t=1}^{T} b'\Sigma_{\Delta f^*}^{1/2} W(t)\left(f_t^* - \Sigma_{\Delta f^*}^{1/2} W(t)\right)' b \\ \leq\ & C_0 \frac{\ln\ln T}{T^2}\sum_{t=1}^{T} \|W(t)\| \left\| f_t^* - \Sigma_{\Delta f^*}^{1/2} W(t)\right\| \\ \leq\ & C_0 \frac{\ln\ln T}{T^2}\left(\sup_{1\leq t\leq T}\left\| f_t^* - \Sigma_{\Delta f^*}^{1/2} W(t)\right\|\right)\sum_{t=1}^{T} \|W(t)\| \\ \leq\ & C_0 \frac{\ln\ln T}{T^2} T^{1/2-\epsilon}\sum_{t=1}^{T} t^{1/2}(\ln\ln t)^{1/2} = o_{a.s.}(1). \end{aligned}$$

Finally it holds that

$$\lim\inf_{T\to\infty} \frac{\ln\ln T}{T^2}\sum_{t=1}^{T} b'\Sigma_{\Delta f^*}^{1/2} W(t) W(t)' \Sigma_{\Delta f^*}^{1/2} b = \frac{1}{4}\left(b'\Sigma_{\Delta f^*} b\right) > 0,$$

by noting that $b'\Sigma_{\Delta f^*}^{1/2} W(t) \overset{D}{=} \left(b'\Sigma_{\Delta f^*} b\right)^{1/2} B(t)$ with $B(t)$ a scalar, standard Wiener process and by applying equation (4.6) in Donsker and Varadhan (1977), and by the positive definitness of $\Sigma_{\Delta f^*}$. Since this holds for all b, the Lemma follows. □

We will now make extensive use of the notation $\widetilde{f}_t^{(1)} = d_1 t + d_2 f_t^{(1)\dagger}$.

Lemma B4. *Let* $f_t^{(1,2)} = \left[\widetilde{f}_t^{(1)}, f_t^{(2)\prime}\right]'$. *Under Assumptions 2 and 4-5, it holds that*

$$\nu^{(1)}\left(\frac{1}{T^2}\sum_{t=1}^{T} f_t^{(1,2)} f_t^{(1,2)\prime}\right) \geq C_0 T \text{ if } d_1 = 1, \tag{B6}$$

$$\nu^{(p)}\left(\frac{1}{T^2}\sum_{t=1}^{T} f_t^{(1,2)} f_t^{(1,2)\prime}\right) \geq \frac{C_0}{\ln\ln T}, \text{ for } d_1 + 1 \leq p \leq r_2 + \max\{d_1, d_2\}, \tag{B7}$$

$$\nu^{(r_2+1)}\left(\frac{1}{T^2}\sum_{t=1}^{T} f_t^{(1,2)} f_t^{(1,2)\prime}\right) \leq \frac{C_0}{T}(\ln T)^{3/2+\epsilon} \text{ if } d_1 = d_2 = 0, \tag{B8}$$

for N, T *large enough.*

Proof. Let $\widetilde{d}_1 = [d_1, 0, ..., 0]'$ be an (r_2+1)-dimensional vector. We have

$$\frac{1}{T^2}\sum_{t=1}^{T} f_t^{(1,2)} f_t^{(1,2)\prime} = \frac{1}{T^2}\sum_{t=1}^{T} t^2 \widetilde{d}_1 \widetilde{d}_1{}' + \frac{1}{T^2}\widetilde{d}_1\sum_{t=1}^{T} t f_t^{*\prime} + \frac{1}{T^2}\sum_{t=1}^{T} f_t^* t \widetilde{d}_1{}' + \frac{1}{T^2}\sum_{t=1}^{T} f_t^* f_t^{*\prime}.$$

In the proof, we make repeated use of the lower bound entailed by Weyl's inequality (see **?**, p.181)

$$\nu^{(p)}(A+B) \geq \nu^{(p)}(A) + \nu^{(\min)}(B),$$

for two symmetric matrices A and B. Clearly

$$\nu^{(1)}\left(\frac{1}{T^2}\sum_{t=1}^{T} f_t^{(1,2)} f_t^{(1,2)\prime}\right) \geq \nu^{(1)}\left(\frac{1}{T^2}\sum_{t=1}^{T} t^2 \widetilde{d}_1 \widetilde{d}_1{}'\right) + \nu^{(\min)}(B), \tag{B9}$$

with

$$B = \frac{1}{T^2}\widetilde{d}_1 \sum_{t=1}^{T} t f_t^{*\prime} + \frac{1}{T^2}\sum_{t=1}^{T} f_t^{*} t \widetilde{d}_1{}' + \frac{1}{T^2}\sum_{t=1}^{T} f_t^{*} f_t^{*\prime}.$$

Simple algebra yields

$$\nu^{(1)}\left(\frac{1}{T^2}\sum_{t=1}^{T} t^2 \widetilde{d}_1 \widetilde{d}_1{}'\right) = \frac{d_1^2}{3} T.$$

Also, we have that $\left|\nu^{(\min)}(B)\right| = O_{a.s.}(\ln\ln T)$; indeed

$$\nu^{(\min)}(B) \leq \nu^{(\min)}\left(\frac{1}{T^2}\sum_{t=1}^{T} f_t^{*} f_t^{*\prime}\right) + \nu^{(1)}\left(\frac{1}{T^2}\widetilde{d}_1 \sum_{t=1}^{T} t f_t^{*\prime} + \frac{1}{T^2}\sum_{t=1}^{T} f_t^{*} t \widetilde{d}_1{}'\right).$$

By Donsker and Varadhan (1977, Example 2), it holds that

$$\nu^{(\min)}\left(\frac{1}{T^2}\sum_{t=1}^{T} f_t^{*} f_t^{*\prime}\right) \leq C_0 \ln\ln T.$$

Also, the matrix

$$B_2 = \frac{1}{T^2}\widetilde{d}_1 \sum_{t=1}^{T} t f_t^{*\prime} + \frac{1}{T^2}\sum_{t=1}^{T} f_t^{*} t \widetilde{d}_1{}',$$

is symmetric, and after some algebra it can be shown that

$$\nu^{(1)}(B_2) \leq C_0 \left(\sum_{i=1}^{r_2+1} \left|\frac{1}{T^2}\sum_{t=1}^{T} t f_{i,t}^{*}\right|^2\right)^{1/2}.$$

Given that

$$E\left|\frac{1}{T^2}\sum_{t=1}^{T} t f_{i,t}^{*}\right|^2 \leq C_0 \frac{1}{T^3}\sum_{t=1}^{T} t^2 E\left(f_{i,t}^{*}\right)^2 \leq C_1 T,$$

by Lemma B1 it holds that $\nu^{(1)}(B_2) = O_{a.s.}\left(T^{1/2}(\ln T)^{1+\epsilon}\right)$ for every $\epsilon > 0$. Thus, finally, by (B9) it holds that there exists a random T_0 such that for $T \geq T_0$

$$\nu^{(1)}\left(\frac{1}{T^2}\sum_{t=1}^{T} f_t^{(1,2)} f_t^{(1,2)\prime}\right) \geq C_0 T,$$

which proves (B6). Turning to (B7), for each $p>1$

$$\begin{aligned}\nu^{(p)}\left(\frac{1}{T^2}\sum_{t=1}^{T} f_t^{(1,2)} f_t^{(1,2)\prime}\right) &\geq \nu^{(p)}\left(\frac{1}{T^2}\sum_{t=1}^{T} t^2 \widetilde{d}_1 \widetilde{d}_1{}' + \frac{1}{T^2}\sum_{t=1}^{T} f_t^* f_t^{*\prime}\right) \\ &\quad + \nu^{(\min)}\left(\frac{1}{T^2}\widetilde{d}_1 \sum_{t=1}^{T} t f_t^{*\prime} + \frac{1}{T^2}\sum_{t=1}^{T} f_t^* t \widetilde{d}_1{}'\right) \\ &\geq \nu^{(p)}\left(\frac{1}{T^2}\sum_{t=1}^{T} f_t^* f_t^{*\prime}\right) + \nu^{(\min)}\left(\frac{1}{T^2}\sum_{t=1}^{T} t^2 \widetilde{d}_1 \widetilde{d}_1{}'\right) = \nu^{(p)}\left(\frac{1}{T^2}\sum_{t=1}^{T} f_t^* f_t^{*\prime}\right),\end{aligned}$$

so that the desired result follows immediately from Lemma B3. Finally, consider (B8). Let $\widetilde{g}_t = [g_t, 0, ..., 0]'$ and $\widetilde{f}_t = \left[0, f_t^{(2)\prime}\right]'$ be two (r_2+1)-dimensional vectors; in this case we have

$$\frac{1}{T^2}\sum_{t=1}^{T} f_t^{(1,2)} f_t^{(1,2)\prime} = \frac{1}{T^2}\sum_{t=1}^{T} \widetilde{g}_t \widetilde{g}_t' + \frac{1}{T^2}\sum_{t=1}^{T} \widetilde{f}_t \widetilde{f}_t';$$

thus

$$\nu^{(\min)}\left(\frac{1}{T^2}\sum_{t=1}^{T} f_t^{(1,2)} f_t^{(1,2)\prime}\right) \leq \nu^{(1)}\left(\frac{1}{T^2}\sum_{t=1}^{T} \widetilde{g}_t \widetilde{g}_t'\right) + \nu^{(\min)}\left(\frac{1}{T^2}\sum_{t=1}^{T} \widetilde{f}_t \widetilde{f}_t'\right) \leq \frac{1}{T^2}\sum_{t=1}^{T} g_t^2.$$

Assumption 3*(i)* and equation (2.3) in **?** imply that

$$E \max_{1\leq \widetilde{t}\leq T}\left\|\sum_{t=1}^{\widetilde{t}} g_t^2\right\|^2 \leq C_0 (\ln T)^2 T,$$

which, through Lemma B1, yields the desired result. □

Lemma B5. *Under Assumptions 2-4*

$$\max_{1\leq p\leq N}\left|\nu^{(p)}\left(\frac{1}{T^3}\sum_{t=1}^{T} u_t^{(1)} u_t^{(1)\prime} + \frac{1}{T^3}\sum_{t=1}^{T} \Lambda^{(1)} \widetilde{f}_t^{(1)} u_t^{(1)\prime} + \frac{1}{T^3}\sum_{t=1}^{T} u_t^{(1)} \widetilde{f}_t^{(1)} \Lambda^{(1)\prime}\right)\right| = O_{a.s.}\left(\frac{N}{\sqrt{T}} l_{N,T}\right).$$

Proof. We show the lemma for the case $d_1 = d_2 = 1$; when either dummy is zero, calculations become easier and the result can be readily shown. Let

$$\max_{1\leq p\leq N}\left|\nu^{(p)}\left(\frac{1}{T^3}\sum_{t=1}^{T} u_t^{(1)} u_t^{(1)\prime} + \frac{1}{T^3}\sum_{t=1}^{T} \Lambda^{(1)} \widetilde{f}_t^{(1)} u_t^{(1)\prime} + \frac{1}{T^3}\sum_{t=1}^{T} u_t^{(1)} \widetilde{f}_t^{(1)} \Lambda^{(1)\prime}\right)\right| = \nu^{(\max)},$$

for short. It holds that

$$\frac{1}{3}\nu^{(\max)} \leq \frac{1}{3}\left(\sum_{i=1}^{N}\sum_{j=1}^{N}\left|\frac{1}{T^3}\sum_{t=1}^{T}u_{i,t}^{(1)}u_{j,t}^{(1)}+\frac{1}{T^3}\sum_{t=1}^{T}\Lambda_i^{(1)}\widetilde{f}_t^{(1)}u_{j,t}^{(1)}+\frac{1}{T^3}\sum_{t=1}^{T}\Lambda_j^{(1)}\widetilde{f}_t^{(1)}u_{i,t}^{(1)}\right|^2\right)^{1/2} \quad \text{(B10)}$$

$$\leq \left(\sum_{i=1}^{N}\sum_{j=1}^{N}\left|\frac{1}{T^3}\sum_{t=1}^{T}u_{i,t}^{(1)}u_{j,t}^{(1)}\right|^2\right)^{1/2}+\left(\sum_{i=1}^{N}\sum_{j=1}^{N}\left|\frac{1}{T^3}\sum_{t=1}^{T}\sum_{k=1}^{r}\Lambda_i^{(1)}\widetilde{f}_t^{(1)}u_{j,t}^{(1)}\right|^2\right)^{1/2}$$

$$+\left(\sum_{i=1}^{N}\sum_{j=1}^{N}\left|\frac{1}{T^3}\sum_{t=1}^{T}\Lambda_j^{(1)}\widetilde{f}_t^{(1)}u_{i,t}^{(1)}\right|^2\right)^{1/2},$$

where the first passage is the usual spectral norm inequality, and the last passage follows from applying (twice) the C_r-inequality (**?**, p. 140).
Let now

$$u_{i,t}^{(2)}=\lambda^{(1)}g_t+\lambda^{(3)\prime}f_t^{(3)}+u_{i,t} \quad \text{(B11)}$$

and note that

$$\frac{1}{3}\sum_{i=1}^{N}\sum_{j=1}^{N}\left|\frac{1}{T^3}\sum_{t=1}^{T}u_{i,t}^{(1)}u_{j,t}^{(1)}\right|^2$$

$$\leq \sum_{i=1}^{N}\sum_{j=1}^{N}\left|\frac{1}{T^3}\sum_{t=1}^{T}u_{i,t}^{(2)}u_{j,t}^{(2)}\right|^2+\sum_{i=1}^{N}\sum_{j=1}^{N}\left|\frac{1}{T^3}\sum_{t=1}^{T}\sum_{k=1}^{r_2}\Lambda_{i,k}^{(2)}f_{k,t}^{(2)}f_{k,t}^{(2)\prime}\Lambda_{i,k}^{(2)\prime}\right|^2$$

$$+\sum_{i=1}^{N}\sum_{j=1}^{N}\left|\frac{1}{T^3}\sum_{t=1}^{T}\sum_{k=1}^{r_2}\Lambda_{i,k}^{(2)}f_{k,t}^{(2)}u_{j,t}^{(2)}\right|^2+\sum_{i=1}^{N}\sum_{j=1}^{N}\left|\frac{1}{T^3}\sum_{t=1}^{T}\sum_{k=1}^{r_2}\Lambda_{j,k}^{(2)}f_{k,t}^{(2)}u_{i,t}^{(2)}\right|^2.$$

We have

$$E\max_{h_1,h_2,\widetilde{t}}\sum_{i=1}^{h_1}\sum_{j=1}^{h_2}\left|\frac{1}{T^3}\sum_{t=1}^{\widetilde{t}}u_{i,t}^{(2)}u_{j,t}^{(2)}\right|^2 \leq \sum_{i=1}^{N}\sum_{j=1}^{N}E\max_{\widetilde{t}}\left|\frac{1}{T^3}\sum_{t=1}^{\widetilde{t}}u_{i,t}^{(2)}u_{j,t}^{(2)}\right|^2$$

$$\leq C_0T\frac{1}{T^6}\sum_{i=1}^{N}\sum_{j=1}^{N}E\sum_{t=1}^{T}\left|u_{i,t}^{(2)}\right|^2\left|u_{j,t}^{(2)}\right|^2 \leq C_0N^2T^{-4}\max_{1\leq i\leq N}E\left|u_{i,t}^{(2)}\right|^4$$

$$\leq C_0N^2T^{-4}\left(\max_{1\leq i\leq N}E\left|u_{i,t}\right|^4+\max_{1\leq i\leq N}\left\|\lambda_i^{(3)}\right\|^4E\left\|f_t^{(3)}\right\|^4++\max_{1\leq i\leq N}\left\|\lambda_i^{(1)}\right\|^4E\left\|g_t\right\|^4\right)$$

$$\leq C_0N^2T^{-4},$$

so that, by Lemma B1

$$\sum_{i=1}^{N}\sum_{j=1}^{N}\left|\frac{1}{T^3}\sum_{t=1}^{T}u_{i,t}^{(2)}u_{j,t}^{(2)}\right|^2=O_{a.s.}\left(\frac{N^2}{T^4}\ln^{2+\epsilon}N\ln^{1+\epsilon}T\right). \quad \text{(B12)}$$

Also

$$\begin{aligned}
&\sum_{i=1}^{N}\sum_{j=1}^{N}\left|\frac{1}{T^{3}}\sum_{t=1}^{T}\sum_{k=1}^{r_2}\Lambda_{i,k}^{(2)}f_{k,t}^{(2)}f_{k,t}^{(2)\prime}\Lambda_{i,k}^{(2)}\right|^{2} \\
\leq\ & T^{-6}N^{2}\left(\max_{i}\left\|\Lambda_{i,k}^{(2)}\right\|\right)^{2}\left\|\sum_{t=1}^{T}f_{t}^{(2)}f_{t}^{(2)\prime}\right\|^{2};
\end{aligned}$$

on account of Assumption 2*(iv)*, it holds that

$$\left\|\frac{\sum_{t=1}^{T}f_{t}^{(2)}f_{t}^{(2)\prime}}{T^{2}\ln\ln T}\right\|^{2}=\left\|\Sigma_{\Delta f^{*}}^{1/2}\frac{\sum_{t=1}^{T}W\left(t\right)W\left(t\right)^{\prime}}{T^{2}\ln\ln T}\Sigma_{\Delta f^{*}}^{1/2}\right\|^{2}+o_{a.s.}\left(1\right)=O_{a.s.}\left(1\right);$$

the final result follows from Donsker and Varadhan (1977, Example 2)). Thus

$$\sum_{i=1}^{N}\sum_{j=1}^{N}\left|\frac{1}{T^{3}}\sum_{t=1}^{T}\sum_{k=1}^{r_2}\Lambda_{i,k}^{(2)}f_{k,t}^{(2)}f_{k,t}^{(2)\prime}\Lambda_{i,k}^{(2)}\right|^{2}=O_{a.s.}\left(\frac{N^{2}}{T^{2}}\left(\ln\ln T\right)^{2}\right). \tag{B13}$$

Finally, consider

$$\begin{aligned}
&E\max_{h_1,h_2,\widetilde{t}}\sum_{i=1}^{h_1}\sum_{j=1}^{h_2}\left|\sum_{t=1}^{\widetilde{t}}\Lambda_{i}^{(2)\prime}f_{t}^{(2)}u_{j,t}^{(2)}\right|^{2}\leq\sum_{i=1}^{N}\sum_{j=1}^{N}E\max_{\widetilde{t}}\left|\sum_{t=1}^{\widetilde{t}}\Lambda_{i}^{(2)\prime}f_{t}^{(2)}u_{j,t}^{(2)}\right|^{2} \\
\leq\ & C_{0}\left(\ln T\right)^{2}\sum_{i=1}^{N}\sum_{j=1}^{N}E\left|\sum_{t=1}^{T}\Lambda_{i}^{(2)\prime}f_{t}^{(2)}u_{j,t}^{(2)}\right|^{2}\leq C_{0}\left(\ln T\right)^{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\Lambda_{i}^{(2)\prime}\sum_{t=1}^{T}\sum_{s=1}^{T}E\left(f_{t}^{(2)}u_{j,t}^{(2)}f_{s}^{(2)\prime}u_{j,s}^{(2)}\right)\Lambda_{i}^{(2)} \\
\leq\ & C_{1}\left(\max_{i}\left\|\Lambda_{i}^{(2)\prime}\right\|\right)^{2}\left(\ln T\right)^{2}\sum_{i=1}^{N}\sum_{j=1}^{N}E\left\|\sum_{t=1}^{T}f_{t}^{(2)}u_{j,t}^{(2)}\right\|^{2}\leq C_{2}N^{2}T^{2}\left(\ln T\right)^{2},
\end{aligned}$$

having used Assumption 3*(ii)*, so that

$$\sum_{i=1}^{N}\sum_{j=1}^{N}\left|\frac{1}{T^{3}}\sum_{t=1}^{T}\sum_{k=1}^{r_2}\Lambda_{i,k}^{(2)}f_{k,t}^{(2)}u_{j,t}^{(2)}\right|^{2}=O_{a.s.}\left(\frac{N^{2}}{T^{4}}\ln^{2+\epsilon}N\ln^{3+\epsilon}T\right).$$

Putting all together, we have

$$\left(\sum_{i=1}^{N}\sum_{j=1}^{N}\left|\frac{1}{T^{3}}\sum_{t=1}^{T}u_{i,t}^{(1)}u_{j,t}^{(1)}\right|^{2}\right)^{1/2}=O_{a.s.}\left(\frac{N}{T}\ln\ln T\right).$$

Consider now

$$\begin{aligned}
& \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left|\frac{1}{T^{3}}\sum_{t=1}^{T}\Lambda_{i}^{(1)}\widetilde{f}_{t}^{(1)}u_{j,t}^{(1)}\right|^{2} \\
\leq & \sum_{i=1}^{N}\sum_{j=1}^{N}\left|\frac{1}{T^{3}}\sum_{t=1}^{T}\Lambda_{i}^{(1)}tu_{j,t}^{(1)}\right|^{2}+\sum_{i=1}^{N}\sum_{j=1}^{N}\left|\frac{1}{T^{3}}\sum_{t=1}^{T}\Lambda_{i}^{(1)}f_{t}^{(1)\dagger}u_{j,t}^{(1)}\right|^{2}. \qquad \text{(B14)}
\end{aligned}$$

We have

$$\begin{aligned}
& \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left|\frac{1}{T^{3}}\sum_{t=1}^{T}\Lambda_{i}^{(1)}tu_{j,t}^{(1)}\right|^{2} \\
\leq & \sum_{i=1}^{N}\sum_{j=1}^{N}\left|\frac{1}{T^{3}}\sum_{t=1}^{T}\Lambda_{i}^{(1)}tf_{t}^{(2)'}\Lambda_{j}^{(2)\prime}\right|^{2}+\sum_{i=1}^{N}\sum_{j=1}^{N}\left|\frac{1}{T^{3}}\sum_{t=1}^{T}\Lambda_{i}^{(1)}tu_{j,t}^{(2)}\right|^{2}.
\end{aligned}$$

Note that

$$\begin{aligned}
& E\max_{h_1,h_2,\widetilde{t}}\sum_{i=1}^{h_1}\sum_{j=1}^{h_2}\left|\frac{1}{T^{3}}\sum_{t=1}^{\widetilde{t}}\Lambda_{i}^{(1)}tf_{t}^{(2)'}\Lambda_{j}^{(2)\prime}\right|^{2}\leq T^{-6}\sum_{i=1}^{N}\sum_{j=1}^{N}E\max_{\widetilde{t}}\left|\sum_{t=1}^{\widetilde{t}}\Lambda_{i}^{(1)}tf_{t}^{(2)'}\Lambda_{j}^{(2)\prime}\right|^{2} \\
\leq & C_0\left(\ln T\right)^{2}T^{-6}\sum_{i=1}^{N}\sum_{j=1}^{N}E\left|\sum_{t=1}^{T}\Lambda_{i}^{(1)}tf_{t}^{(2)'}\Lambda_{j}^{(2)\prime}\right|^{2}\leq C_0\left(\ln T\right)^{2}T^{-6}\sum_{i=1}^{N}\sum_{j=1}^{N}\Lambda_{i}^{(1)}\sum_{t=1}^{T}\sum_{s=1}^{T}E\left(tf_{s}^{(2)\prime}\right)\Lambda_{j}^{(2)} \\
\leq & C_1N^{2}T^{-6}\left(\max_{i}\left\Vert\Lambda_{i}^{(1)}\right\Vert\right)^{2}\left(\max_{i}\left\Vert\Lambda_{i}^{(2)}\right\Vert\right)^{2}\left(\ln T\right)^{2}E\left\Vert\sum_{t=1}^{T}tf_{t}^{(2)}\right\Vert^{2}\leq C_2N^{2}T^{-1}\left(\ln T\right)^{2},
\end{aligned}$$

having used Assumption 3*(iii)*; Lemma B1 entails that

$$\sum_{i=1}^{N}\sum_{j=1}^{N}\left|\frac{1}{T^{3}}\sum_{t=1}^{T}\Lambda_{i}^{(1)}tf_{t}^{(2)'}\Lambda_{j}^{(2)\prime}\right|^{2}=O_{a.s.}\left(\frac{N^{2}}{T}\ln^{2+\epsilon}N\ln^{3+\epsilon}T\right).$$

Similar passages yield

$$\sum_{i=1}^{N}\sum_{j=1}^{N}\left|\frac{1}{T^{3}}\sum_{t=1}^{T}\Lambda_{i}^{(1)}tu_{j,t}^{(2)}\right|^{2}=O_{a.s.}\left(\frac{N^{2}}{T^{3}}\ln^{2+\epsilon}N\ln^{3+\epsilon}T\right).$$

Thus, finally

$$\sum_{i=1}^{N}\sum_{j=1}^{N}\left|\frac{1}{T^{3}}\sum_{t=1}^{T}\Lambda_{i}^{(1)}tu_{j,t}^{(1)}\right|^{2}=O_{a.s.}\left(\frac{N^{2}}{T}\ln^{2+\epsilon}N\ln^{3+\epsilon}T\right).$$

We now consider the next term in equation (B14). We have

$$\sum_{i=1}^{N}\sum_{j=1}^{N}\left|\frac{1}{T^3}\sum_{t=1}^{T}\Lambda_i^{(1)}f_t^{(1)\dagger}u_{j,t}^{(1)}\right|^2=\sum_{i=1}^{N}\sum_{j=1}^{N}\left|\frac{1}{T^3}\sum_{t=1}^{T}\Lambda_i^{(1)}f_t^{(1)\dagger}f_t^{(2)'}\Lambda_j^{(2)\prime}\right|^2+\sum_{i=1}^{N}\sum_{j=1}^{N}\left|\frac{1}{T^3}\sum_{t=1}^{T}\Lambda_i^{(1)}f_t^{(1)\dagger}u_{j,t}^{(2)}\right|^2.$$

Similar passages as above yield

$$\begin{aligned}&E\max_{h_1,h_2,\widetilde{t}}\sum_{i=1}^{h_1}\sum_{j=1}^{h_2}\left|\frac{1}{T^3}\sum_{t=1}^{\widetilde{t}}\Lambda_i^{(1)}f_t^{(1)\dagger}f_t^{(2)'}\Lambda_j^{(2)\prime}\right|^2\\ \leq\ & C_0N^2T^{-6}\left(\max_i\left\|\Lambda_i^{(1)}\right\|\right)^2\left(\max_i\left\|\Lambda_i^{(2)}\right\|\right)^2(\ln T)^2\,E\left\|\sum_{t=1}^{T}f_t^{(1)\dagger}f_t^{(2)'}\right\|^2\leq C_1N^2T^{-6}(\ln T)^2T^4,\end{aligned}$$

having used Assumption 2*(vi)*. Similarly

$$\begin{aligned}&E\max_{h_1,h_2,\widetilde{t}}\sum_{i=1}^{h_1}\sum_{j=1}^{h_2}\left|\frac{1}{T^3}\sum_{t=1}^{\widetilde{t}}\Lambda_i^{(1)}f_t^{(1)\dagger}u_{j,t}^{(2)}\right|^2\\ \leq\ & C_0N^2T^{-6}\left(\max_i\left\|\Lambda_i^{(1)}\right\|\right)^2(\ln T)^2\,E\left\|\sum_{t=1}^{T}f_t^{(1)\dagger}u_{j,t}^{(2)}\right\|^2\leq C_1N^2T^{-6}(\ln T)^2T^2,\end{aligned}$$

having used Assumption 3*(ii)*. Thus, using Lemma B1

$$\sum_{i=1}^{N}\sum_{j=1}^{N}\left|\frac{1}{T^3}\sum_{t=1}^{T}\Lambda_i^{(1)}f_t^{(1)\dagger}u_{j,t}^{(1)}\right|^2=O_{a.s.}\left(\frac{N}{T}\ln^{1+\epsilon}N\ln^{\frac{3}{2}+\epsilon}T\right).$$

Using (B10) and putting all together, the desired result obtains. □

Lemma B6. *Under Assumptions 2-4*

$$\max_{1\leq p\leq N}\left|\nu^{(p)}\left(\frac{1}{T^2}\sum_{t=1}^{T}u_t^{(2)}u_t^{(2)\prime}+\frac{1}{T^2}\sum_{t=1}^{T}\Lambda^{(1,2)}f_t^{(1,2)}u_t^{(2)\prime}+\frac{1}{T^2}\sum_{t=1}^{T}u_t^{(2)}f_t^{(1,2)\prime}\Lambda^{(1,2)\prime}\right)\right|=O_{a.s.}\left(\frac{N}{\sqrt{T}}l_{N,T}\right),\quad\text{(B15)}$$

where $u_t^{(2)}$ *is defined in (B11).*

Proof. Let

$$\max_{1\leq p\leq N}\left|\nu^{(p)}\left(\frac{1}{T^2}\sum_{t=1}^{T}u_t^{(2)}u_t^{(2)\prime}+\frac{1}{T^2}\sum_{t=1}^{T}\Lambda^{(1,2)}f_t^{(1,2)}u_t^{(2)\prime}+\frac{1}{T^2}\sum_{t=1}^{T}u_t^{(2)}f_t^{(1,2)\prime}\Lambda^{(1,2)\prime}\right)\right|=\nu^{(\max)},$$

for short. As before

$$\frac{1}{3}\nu^{(\max)} \tag{B16}$$

$$\leq \frac{1}{3}\left(\sum_{i=1}^{N}\sum_{j=1}^{N}\left|\frac{1}{T^2}\sum_{t=1}^{T}u_{i,t}^{(2)}u_{j,t}^{(2)}+\frac{1}{T^2}\sum_{t=1}^{T}\sum_{k=1}^{r_2+r_1}\Lambda_{i,k}^{(1,2)}f_{k,t}^{(1,2)}u_{j,t}^{(2)}+\frac{1}{T^2}\sum_{t=1}^{T}\sum_{k=1}^{r_2+r_1}\Lambda_{j,k}^{(1,2)}f_{k,t}^{(1,2)}u_{i,t}^{(2)}\right|^2\right)^{1/2}.$$

$$\leq \left(\sum_{i=1}^{N}\sum_{j=1}^{N}\left|\frac{1}{T^2}\sum_{t=1}^{T}u_{i,t}^{(2)}u_{j,t}^{(2)}\right|^2\right)^{1/2}+\left(\sum_{i=1}^{N}\sum_{j=1}^{N}\left|\frac{1}{T^2}\sum_{t=1}^{T}\sum_{k=1}^{r_2+r_1}\Lambda_{i,k}^{(1,2)}f_{k,t}^{(1,2)}u_{j,t}^{(2)}\right|^2\right)^{1/2}$$

$$+\left(\sum_{i=1}^{N}\sum_{j=1}^{N}\left|\frac{1}{T^2}\sum_{t=1}^{T}\sum_{k=1}^{r_2+r_1}\Lambda_{j,k}^{(1,2)}f_{k,t}^{(1,2)}u_{i,t}^{(2)}\right|^2\right)^{1/2}.$$

Consider the first term; by (B12),

$$\left(\sum_{i=1}^{N}\sum_{j=1}^{N}\left|\frac{1}{T^2}\sum_{t=1}^{T}u_{i,t}^{(2)}u_{j,t}^{(2)}\right|^2\right)^{1/2}=O_{a.s.}\left(\frac{N}{T}\left(\ln N\right)^{1+\epsilon}\left(\ln T\right)^{(1+\epsilon)/2}\right).$$

Similarly, considering the second term in (B16) we have

$$E\max_{h_1,h_2,\widetilde{t}}\sum_{i=1}^{h_1}\sum_{j=1}^{h_2}\left|\frac{1}{T^2}\sum_{t=1}^{\widetilde{t}}\sum_{k=1}^{r_2+r_1}\Lambda_{i,k}^{(1,2)}f_{k,t}^{(1,2)}u_{j,t}^{(2)}\right|^2\leq T^{-4}\sum_{i=1}^{N}\sum_{j=1}^{N}E\max_{\widetilde{t}}\left|\sum_{t=1}^{\widetilde{t}}\sum_{k=1}^{r_2+r_1}\Lambda_{i,k}^{(1,2)}f_{k,t}^{(1,2)}u_{j,t}^{(2)}\right|^2$$

$$\leq C_0T^{-4}\left(\ln T\right)^2\sum_{i=1}^{N}\sum_{j=1}^{N}E\left|\sum_{t=1}^{T}\sum_{k=1}^{r_2+r_1}\Lambda_{i,k}^{(1,2)}f_{k,t}^{(1,2)}u_{j,t}^{(2)}\right|^2$$

$$\leq C_0T^{-4}\left(\ln T\right)^2\left(\max_{1\leq i\leq N}\left\|\Lambda_i^{(1,2)}\right\|\right)^2\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}E\left(f_t^{(1,2)}f_s^{(1,2)\prime}u_{j,t}^{(2)}u_{j,s}^{(2)}\right)$$

$$\leq C_0T^{-4}\left(\ln T\right)^2\sum_{i=1}^{N}\sum_{j=1}^{N}E\left\|\sum_{t=1}^{T}f_t^{(1,2)}u_{j,t}^{(2)}\right\|^2\leq C_0N^2T^{-1}\left(\ln T\right)^2,$$

having used equation (2.3) in **?**, Assumption 4*(i)* and Assumption 3*(ii)*. From here henceforth, the proof is the same as for the first tem in (B16); also, the proof for the third term in (B16) is exactly the same, and it is therefore omitted. Putting everything together, the lemma follows. □

B.2 Proofs of main results

Proof of Lemma 1. When $d_1 = 0$, the lemma follows immediately from B having full rank. When $d_1 = 1$, the proof follows the arguments in Maciejowska (2010). Let

$$\mathcal{F}_t = (A|B)\begin{pmatrix} t \\ \psi_t \end{pmatrix} = C\begin{pmatrix} t \\ \psi_t \end{pmatrix};$$

by Assumption 1*(ii)*, C has full rank. It is therefore possible to re-write the expression above as

$$\mathcal{F}_t = P\,(D_1|D_2)\begin{pmatrix} t \\ \psi_t \end{pmatrix},$$

where $D_1 = [1, 0, ..., 0]'$ is $r \times 1$, and P and D_2 are $r \times r$ and have full rank. Among the possible matrices that satisfy this representation one can consider $(D_1|D_2) = (I_r|E)$, where $E = [E_1, ..., E_r]$ is a nonzero vector. The desired result follows immediately after computing

$$P^{-1}\mathcal{F}_t = \begin{pmatrix} t + E_1\psi_{r,t} \\ \psi_{1,t} + E_2\psi_{r,t} \\ \psi_{2,t} + E_3\psi_{r,t} \\ . \\ \psi_{r-1,t} + E_r\psi_{r,t} \end{pmatrix}.$$

□

Proof of Theorem 1. We start with (12)-(13). Weyl's inequality entails that, for $0 \le p \le r_1$

$$\begin{aligned} \nu_1^{(p)} \quad \geq \quad & \nu^{(p)}\left(\frac{1}{T^3}\sum_{t=1}^{T}\Lambda^{(1)}\widetilde{f}_t^{(1)}\widetilde{f}_t^{(1)\prime}\Lambda^{(1)\prime}\right) \\ & +\nu^{(N)}\left(\frac{1}{T^3}\sum_{t=1}^{T}u_t^{(1)}u_t^{(1)\prime} + \frac{1}{T^3}\sum_{t=1}^{T}\Lambda^{(1)}\widetilde{f}_t^{(1)}u_t^{(1)\prime} + \frac{1}{T^3}\sum_{t=1}^{T}u_t^{(1)}\widetilde{f}_t^{(1)\prime}\Lambda^{(1)\prime}\right). \end{aligned}$$

We already know that, by Lemma B5

$$\nu^{(N)}\left(\frac{1}{T^3}\sum_{t=1}^{T}u_t^{(1)}u_t^{(1)\prime} + \frac{1}{T^3}\sum_{t=1}^{T}\Lambda^{(1)}\widetilde{f}_t^{(1)}u_t^{(1)\prime} + \frac{1}{T^3}\sum_{t=1}^{T}u_t^{(1)}\widetilde{f}_t^{(1)\prime}\Lambda^{(1)\prime}\right) = O_{a.s.}\left(\frac{N}{\sqrt{T}}l_{N,T}\right).$$

Consider now

$$\begin{aligned} & \nu^{(p)}\left(\frac{1}{T^3}\sum_{t=1}^{T}\Lambda^{(1)}\widetilde{f}_t^{(1)}\widetilde{f}_t^{(1)\prime}\Lambda^{(1)\prime}\right) \\ \geq \quad & d_1^2\nu^{(p)}\left(\frac{1}{T^3}\sum_{t=1}^{T}t^2\Lambda^{(1)\prime}\Lambda^{(1)}\right) \\ & +\nu^{(N)}\left(2\frac{d_1}{T^3}\sum_{t=1}^{T}tf_t^{(1)\dagger}\Lambda^{(1)}\Lambda^{(1)\prime} + \frac{1}{T^3}\Lambda^{(1)}\sum_{t=1}^{T}f_t^{(1)\dagger}f_t^{(1)\dagger\prime}\Lambda^{(1)\prime}\right) \end{aligned}$$

We have

$$\nu^{(p)}\left(\frac{1}{T^3}\sum_{t=1}^{T}t^2\Lambda^{(1)\prime}\Lambda^{(1)}\right)=\left(\frac{1}{T^3}\sum_{t=1}^{T}t^2\right)\nu^{(p)}\left(\Lambda^{(1)\prime}\Lambda^{(1)}\right)\geq C_0 N,$$

in view of Assumption 4*(ii)*. Consider now

$$\nu^{(N)}\left(2\frac{d_1}{T^3}\sum_{t=1}^{T}tf_t^{(1)\dagger}\Lambda^{(1)}\Lambda^{(1)\prime}+\frac{1}{T^3}\sum_{t=1}^{T}\left(f_t^{(1)\dagger}\right)^2\Lambda^{(1)}\Lambda^{(1)\prime}\right);$$

by Donsker and Varadhan (1977, Example 2) we have

$$\frac{1}{T^3}\sum_{t=1}^{T}\left(f_t^{(1)\dagger}\right)^2=O_{a.s.}\left(\frac{\ln\ln T}{T}\right).$$

Also, by Assumption 3*(iii)* and equation (2.3) in **?** we have

$$E\max_{1\leq t\leq T}\left|\sum_{j=1}^{t}jf_j^{(1)\dagger}\right|^2=C_0T^5\left(\ln T\right)^2,$$

so that by Lemma B1 we have

$$\frac{1}{T^3}\left|\sum_{t=1}^{T}tf_t^{(1)\dagger}\right|=O_{a.s.}\left(T^{-1/2}\left(\ln T\right)^{3/2+\epsilon}\right).$$

The same steps as in the proofs of Lemmas B5 and B6 entail

$$\nu^{(N)}\left(2\frac{d_1}{T^3}\sum_{t=1}^{T}tf_t^{(1)\dagger}\Lambda^{(1)}\Lambda^{(1)\prime}+\frac{1}{T^3}\sum_{t=1}^{T}\left(f_t^{(1)\dagger}\right)^2\Lambda^{(1)}\Lambda^{(1)\prime}\right)=O_{a.s.}\left(\frac{N}{\sqrt{T}}l_{N,T}\right).$$

Putting everything together, the desired result follows. When $p>r_1$

$$\begin{aligned}\nu_1^{(p)} &\leq \nu^{(p)}\left(\frac{1}{T^3}\sum_{t=1}^{T}\Lambda^{(1)}\widetilde{f}_t^{(1)}\widetilde{f}_t^{(1)\prime}\Lambda^{(1)\prime}\right)\\ &\quad+\nu^{(1)}\left(\frac{1}{T^3}\sum_{t=1}^{T}u_t^{(1)}u_t^{(1)\prime}+\frac{1}{T^3}\sum_{t=1}^{T}\Lambda^{(1)}\widetilde{f}_t^{(1)}u_t^{(1)\prime}+\frac{1}{T^3}\sum_{t=1}^{T}u_t^{(1)}\widetilde{f}_t^{(1)\prime}\Lambda^{(1)\prime}\right)\\ &\leq \nu^{(1)}\left(\frac{1}{T^3}\sum_{t=1}^{T}u_t^{(1)}u_t^{(1)\prime}+\frac{1}{T^3}\sum_{t=1}^{T}\Lambda^{(1)}\widetilde{f}_t^{(1)}u_t^{(1)\prime}+\frac{1}{T^3}\sum_{t=1}^{T}u_t^{(1)}\widetilde{f}_t^{(1)\prime}\Lambda^{(1)\prime}\right);\end{aligned}$$

Lemma B5 immediately yields the desired result.

The proof of (14)-(15) is very similar. Whenever $1 \leq p \leq r_1 + r_2 + (1 - r_1)\, d_2$, we have

$$\begin{aligned}\nu_2^{(p)} &\geq \nu^{(p)} \left(\frac{1}{T^2} \sum_{t=1}^{T} \Lambda^{(1,2)} f_t^{(1,2)} f_t^{(1,2)\prime} \Lambda^{(1,2)\prime} \right) \\ &\quad + \nu^{(N)} \left(\frac{1}{T^2} \sum_{t=1}^{T} u_t^{(2)} u_t^{(2)\prime} + \frac{1}{T^2} \sum_{t=1}^{T} \Lambda^{(1,2)} f_t^{(1,2)} u_t^{(2)\prime} + \frac{1}{T^2} \sum_{t=1}^{T} u_t^{(2)} f_t^{(1,2)\prime} \Lambda^{(1,2)\prime} \right).\end{aligned}$$

By Lemma B6 we have

$$\nu^{(N)} \left(\frac{1}{T^2} \sum_{t=1}^{T} u_t^{(2)} u_t^{(2)\prime} + \frac{1}{T^2} \sum_{t=1}^{T} \Lambda^{(1,2)} f_t^{(1,2)} u_t^{(2)\prime} + \frac{1}{T^2} \sum_{t=1}^{T} u_t^{(2)} f_t^{(1,2)\prime} \Lambda^{(1,2)\prime} \right) = O_{a.s.} \left(\frac{N}{\sqrt{T}} l_{N,T} \right).$$

Also, using Theorem 7 in **?** and by Assumption 4*(ii)*

$$\begin{aligned}\nu^{(p)} \left(\frac{1}{T^2} \sum_{t=1}^{T} \Lambda^{(1,2)} f_t^{(1,2)} f_t^{(1,2)\prime} \Lambda^{(1,2)\prime} \right) &\geq \nu^{(p)} \left(\frac{1}{T^2} \sum_{t=1}^{T} f_t^{(1,2)} f_t^{(1,2)\prime} \right) \nu^{(\min)} \left(\Lambda^{(1,2)\prime} \Lambda^{(1,2)} \right) \\ &\geq C_0 \frac{N}{\ln \ln T},\end{aligned}$$

where the last passage follows from equation (B7) in Lemma B4. Equation (14) now follows readily. Turning to (15), whenever $p > r_1 + r_2 + (1 - r_1)\, d_2$,

$$\begin{aligned}\nu_2^{(p)} &\leq \nu^{(p)} \left(\frac{1}{T^2} \sum_{t=1}^{T} \Lambda^{(1,2)} f_t^{(1,2)} f_t^{(1,2)\prime} \Lambda^{(1,2)\prime} \right) \\ &\quad + \nu^{(1)} \left(\frac{1}{T^2} \sum_{t=1}^{T} u_t^{(1)} u_t^{(1)\prime} + \frac{1}{T^2} \sum_{t=1}^{T} \Lambda^{(1)} f_t^{(1)} u_t^{(1)\prime} + \frac{1}{T^2} \sum_{t=1}^{T} u_t^{(1)} f_t^{(1)\prime} \Lambda^{(1)\prime} \right) \\ &= \nu^{(1)} \left(\frac{1}{T^2} \sum_{t=1}^{T} u_t^{(1)} u_t^{(1)\prime} + \frac{1}{T^2} \sum_{t=1}^{T} \Lambda^{(1)} f_t^{(1)} u_t^{(1)\prime} + \frac{1}{T^2} \sum_{t=1}^{T} u_t^{(1)} f_t^{(1)\prime} \Lambda^{(1)\prime} \right),\end{aligned}$$

and Lemma B6 immediately yields the desired result. □

Proof of Theorem 2. The proof is similar to that of related results in other papers - see e.g. Trapani (2017). We begin with (23). Note that, under $H_{0,1}^{(p)}$, (12) and Lemma B2 entail that

$$P \left\{ \omega : \lim_{\min(N,T) \to \infty} \phi_1^{(p)} \exp\left\{ -N^{1-\delta-\varepsilon} \right\} = \infty \right\} = 1,$$

for every $\varepsilon > 0$, and therefore we can henceforth assume that $\lim_{\min(N,T)\to\infty} \phi_1^{(p)} = \infty$ and

$$\left(\phi_1^{(p)} \right)^{-1} = O\left(\exp\left\{ -N^{1-\delta} \right\} \right). \tag{B17}$$

Let E^* and V^* denote, respectively, expectation and variance conditional on P^*; we have, for $1 \leq j \leq R_1$

$$E^*\left(\zeta_{1,j}^{(p)}(u)\right) = G_1(0) \text{ and } V^*\left(\zeta_{1,j}^{(p)}(u)\right) = G_1(0)(1 - G_1(0)).$$

Also

$$\begin{aligned}
& \frac{1}{\sqrt{R_1}} \sum_{j=1}^{R_1} \left(\zeta_{1,j}^{(p)}(u) - G_1(0)\right) \\
= \quad & \frac{1}{\sqrt{R_1}} \sum_{j=1}^{R_1} \left(I\left(\xi_{1,j}^{(p)} \leq 0\right) - G_1(0)\right) + \frac{1}{\sqrt{R_1}} d_u \sum_{j=1}^{R_1} \left(G_1\left(u/\phi_1^{(p)}\right) - G_1(0)\right) \\
& + \frac{1}{\sqrt{R_1}} \sum_{j=1}^{R_1} \left[I\left(0 \leq \left|\xi_{1,j}^{(p)}\right| \leq u/\phi_1^{(p)}\right) - \left(G_1\left(u/\phi_1^{(p)}\right) - G_1(0)\right) d_u\right],
\end{aligned}$$

with $d_u = 1$ for $u \geq 0$ and -1 otherwise. Letting m_{G_1} denote the upper bound for the density of G_1, we have

$$R_1^{-1} \int_{-\infty}^{\infty} \left(\sum_{j=1}^{R_1} \left(G_1\left(u/\phi_1^{(p)}\right) - G_1(0)\right)\right)^2 dF_1(u) \leq m_{G_1}^2 \frac{R_1}{\left(\phi_1^{(p)}\right)^2} \int_{-\infty}^{\infty} u^2 dF_1(u),$$

which drifts to zero under (22) by (B17) and Assumption 6. Also, consider

$$\begin{aligned}
& E^* \int_{-\infty}^{\infty} \left(\frac{1}{\sqrt{R_1}} \sum_{j=1}^{R_1} I\left(0 \leq \left|\xi_{1,j}^{(p)}\right| \leq u/\phi_1^{(p)}\right) - \left(G_1\left(u/\phi_1^{(p)}\right) - G_1(0)\right) d_u\right)^2 dF_1(u) \\
= \quad & E^* \int_{-\infty}^{\infty} \left(I\left(0 \leq \left|\xi_{1,1}^{(p)}\right| \leq u/\phi_1^{(p)}\right) - \left(G_1\left(u/\phi_1^{(p)}\right) - G_1(0)\right) d_u\right)^2 dF_1(u) \\
= \quad & \int_{-\infty}^{\infty} V^*\{I\left(0 \leq \left|\xi_{1,1}^{(p)}\right| \leq u/\phi_1^{(p)}\right)\} dF_1(u)
\end{aligned}$$

by the independence of the $\xi_{1,j}^{(p)}$. Elementary arguments yield

$$\begin{aligned}
V^*\{I\left(0 \leq \xi_{1,1}^{(p)} \leq u/\phi_1^{(p)}\right)\} \quad &= \quad \left|G_1\left(u/\phi_1^{(p)}\right) - G_1(0)\right| \left(1 - \left|G_1\left(u/\phi_1^{(p)}\right) - G_1(0)\right|\right) \\
&\leq \quad \left|G_1\left(u/\phi_1^{(p)}\right) - G_1(0)\right| \leq m_{G_1} \frac{|u|}{\phi_1^{(p)}},
\end{aligned}$$

so that

$$\int_{-\infty}^{\infty} V^*\{I\left(0 \leq \left|\xi_{1,1}^{(p)}\right| \leq u/\phi_1^{(p)}\right)\} dF_1(u) \to 0,$$

as $\phi_1^{(p)} \to \infty$. Thus, by Markov inequality, under (22)

$$\Theta_1^{(p)} = \int_{-\infty}^{\infty} \left(\frac{\sum_{j=1}^{R_1} \left(\zeta_{1,j}^{(p)}(u) - G_1(0) \right)}{\sqrt{R_1}\sqrt{G_1(0)(1-G_1(0))}} \right)^2 dF_1(u)$$

$$= \int_{-\infty}^{\infty} \left(\frac{\sum_{j=1}^{R_1} \left(I\left(\xi_{1,j}^{(p)} \leq 0 \right) - G_1(0) \right)}{\sqrt{R_1}\sqrt{G_1(0)(1-G_1(0))}} \right)^2 dF_1(u) + o_{P^*}(1) \stackrel{D^*}{\to} \chi_1^2,$$

with the last passage following from the CLT for Bernoulli random variables and continuity. This proves (23).

We now turn to (24). By (13) and Lemma B2, we have that, under $H_{A,1}^{(p)}$

$$P\left\{ \omega : \lim_{\min(N,T)\to\infty} \phi_1^{(p)} = 1 \right\} = 1,$$

and therefore we can henceforth assume that

$$\lim_{\min(N,T)\to\infty} \phi_1^{(p)} = 1. \tag{B18}$$

We can write

$$\zeta_{1,j}^{(p)}(u) - G_1(0) = \zeta_{1,j}^{(p)}(u) - G_1(0) \pm G_1\left(u/\phi_1^{(p)} \right),$$

so that

$$E^* \int_{-\infty}^{\infty} \left(\frac{1}{\sqrt{R_1}} \sum_{j=1}^{R_1} \zeta_{1,j}^{(p)}(u) - G_1(0) \right)^2 dF_1(u)$$

$$= E^* \int_{-\infty}^{\infty} \left(\frac{1}{\sqrt{R_1}} \sum_{j=1}^{R_1} \zeta_{1,j}^{(p)}(u) - G_1\left(u/\phi_1^{(p)} \right) \right)^2 dF_1(u) + R_1 \int_{-\infty}^{\infty} \left(G_1\left(u/\phi_1^{(p)} \right) - G_1(0) \right)^2 dF_1(u)$$

$$= \int_{-\infty}^{\infty} V^*\left(\zeta_{1,j}^{(p)}(u) \right) dF_1(u) + R_1 \int_{-\infty}^{\infty} \left(G_1\left(u/\phi_1^{(p)} \right) - G_1(0) \right)^2 dF_1(u),$$

having used again the independence of the $\zeta_{1,j}^{(p)}(u)$. Clearly, $V^*\left(\zeta_{1,j}^{(p)}(u) \right) < \infty$; also, as $\min(N,T) \to \infty$, (B18) yields

$$\int_{-\infty}^{\infty} \left(G_1\left(u/\phi_1^{(p)} \right) - G_1(0) \right)^2 dF_1(u) = \int_{-\infty}^{\infty} \left(G_1(u) - G_1(0) \right)^2 dF_1(u),$$

so that, finally

$$\frac{1}{R_1}\Theta_1^{(p)} = \frac{1}{R_1} \int_{-\infty}^{\infty} \left(\frac{\sum_{j=1}^{R_1} \left(\zeta_{1,j}^{(p)}(u) - G_1(0) \right)}{\sqrt{R_1}\sqrt{G_1(0)(1-G_1(0))}} \right)^2 dF_1(u) = \frac{\int_{-\infty}^{\infty} \left(G_1(u) - G_1(0) \right)^2 dF_1(u)}{G_1(0)(1-G_1(0))} + o(1).$$

□

Proof of Lemma 2. Let Z be a $N(0,1)$ random variable. By (23), using Bernstein concentration inequality we have that

$$P^*\left(\Theta_1^{(p)} > c_{\alpha,1}\right) = P^*\left(Z^2 > c_{\alpha,1}\right) + o_{P^*}(1) \leq 2\exp\left(-\frac{1}{2}c_{\alpha,1}\right) + o_{P^*}(1), \tag{B19}$$

which implies that $P^*\left(\Theta_1^{(p)} > c_{\alpha,1}\right)$ drifts to zero as long as $c_{\alpha,1} \to \infty$. Therefore, under $H_{0,1}^{(1)}$, there is zero probability of a Type I error. Under $H_{A,1}^{(1)}$, by (24) we have

$$\begin{aligned} P^*\left(\Theta_1^{(p)} \leq c_{\alpha,1}\right) &= P^*\left[\left(Z + C_0\sqrt{R_1}\right)^2 \leq c_{\alpha,1}\right] + o_{P^*}(1) \\ &\leq P^*\left(|Z| \leq |c_{\alpha,1}|^{1/2} - C_0\sqrt{R_1}\right) + o_{P^*}(1) \\ &\to P^*\left(|Z| \leq -\infty\right) = 0, \end{aligned}$$

since $c_{\alpha,1} = o(R_1)$. Thus, under the alternative there is zero probability of a Type II error. This proves the desired result. □

Proof of Theorem 3. The proof is exactly the same as the proof of Theorem 2. □

Proof of Lemma 3. The proof is exactly the same as the proof of Theorem 3 in Trapani (2017). □