Testing for Common Trends in Nonstationary Large Datasets

Testing for common trends in non-stationary large datasets *

Matteo Barigozzi
Department of Economics, Università di Bologna
and
Lorenzo Trapani
School of Economics, University of Nottingham

Abstract

We propose a testing based procedure to determine the number of common trends in a large non-stationary dataset. Our procedure is based on a factor representation, where we determine whether there are (and how many) common factors (i) with linear trends, and (ii) with stochastic trends. Cointegration among the factors is also permitted. Our analysis is based on the fact that those 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 largest eigenvalue diverges, based directly on the estimated eigenvalue. The tests only requires minimal assumptions on the data generating process. Monte Carlo evidence shows that our procedure has very good finite sample properties, clearly dominating competing approaches when no common trends are present. We illustrate our methodology through an application to US bond yields with different maturities observed over the last 30 years.

Keywords: non-stationary large panels; common trends; randomised tests.

*We are grateful to the participants to the 1st Italian Workshop on Econometrics and Empirical Economics (Milano, 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); the EEA-ESEM congress at the University of Manchester (Manchester, 26-30 August, 2019).

1 Introduction

Large panels of time series often exhibit co-movements, and can also be non-stationary. In such a case, a convenient way of capturing both aspects (co-movements and non-stationarity) is to employ a factor model where the factors represent common trends. Hence, the first step in the analysis is to determine the presence (or indeed the absence), and the number, of common trends.

The importance of considering common, non-stationary trends when modelling large datasets has been illustrated through numerous applications. For example, focusing only on economics and finance, Bai (2004) studies employment fluctuations across 60 industries in the US; Moon and Perron (2007) consider a panel of interest rates at different maturities in the US and Canada; Engel, Mark, and West (2015) develop a forecasting technique applied to a panel of bilateral US dollar rates against 17 OECD countries; and Zhang, Robinson, and Yao (2019) apply a non-stationary panel factor model to US Industrial Production indices. Whilst, in all the aforementioned applications, the common trends are assumed to be stochastic, panel models with common deterministic, linear trends have also been employed, for example, in the context of analysing macro-economic data (see Maciejowska, 2010), and also in order to model US temperature data (see Chen and Wu, 2019).

In this paper, we consider the following large approximate factor model for an N-dimensional vector of time series X_t:

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

where $\mathcal{F}_t$ is an $r \times 1$ vector of common factors and Λ is an $N \times r$ matrix of factor loadings, with $r < \infty$. We propose a testing procedure to determine the dimension of the following three groups: *(a)* factors with a linear trend and an additional, either $I(1)$ or $I(0)$, zero mean component; *(b)* pure, zero mean $I(1)$ factors, i.e. with no linear trends; and, *(c)* finally, stationary common factors. Each group may well have dimension zero, e.g. factors with linear trends may not be present, etc. If, in model (1), either group *(a)* or *(b)* is non-empty, this entails that X_t is driven by some common trends. We also assume, throughout the paper, that each component of the idiosyncratic vector u_t is $I(0)$, which, in the presence of common stochastic trends, implies that X_t is cointegrated. Although this may be viewed

as a restriction, we point out that our procedure could be employed even in the presence of non-stationary idiosyncratic components, as long as these are only a vanishing fraction of the total, say at most $O(N^{\alpha_0})$, for any $\alpha_0 < 1$ - see also the comments in Section 3.2.

In order to determine the dimension of each of the groups mentioned above, we propose a procedure based on the eigenvalues of the second moment matrix of the data, building on the well-known fact that the number of common factors is related to the number of *spiked* eigenvalues - i.e., the eigenvalues which diverge to infinity as $N \to \infty$. In particular, we develop tests to distinguish whether an eigenvalue is bounded or spiked, and use them in a sequential fashion to estimate the number of common factors belonging in each group. In order to construct the tests, *(i)* we derive the rates of the eigenvalues of the second moment matrix of the data as $\min(N, T) \to \infty$; and *(ii)*, based on those bounds, we propose a randomisation procedure which produces a statistic whose asymptotic behaviour under the null and the alternative hypotheses we are able to study. Whilst details are spelt out hereafter, we point out that our asymptotics does not require virtually any restriction on the relative rate of divergence N and T as they pass to infinity.

Our research question is mainly related to three works. The first contribution in this area is the paper by Bai (2004), who develops the inferential theory for a large panel factor model with common stochastic trends, including a family of information criteria to estimate the number of common (stationary and non-stationary) factors. Whilst linear trends are not considered, Maciejowska (2010) extends the inferential theory to the case of linear trends, although a method to determine their presence is not explicitly derived. Finally, a recent contribution by Zhang et al. (2019) proposes a method based on the ratio of the 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 $\min(N, T) \to \infty$.

Broadly speaking, the literature has considered the presence of common trends in multivariate time series in a plethora of contributions, at least since the seminal articles by Phillips and Ouliaris (1988) and Stock and Watson (1988); more recent extensions to high-dimensional settings include Peña and Poncela (2006), Gengenbach, Palm, and Urbain (2009) and Zhang, Pan, and Gao (2018). Similarly, there is a vast body of literature on unit roots and cointegration in large panels arising from the presence of common stochastic trends -

we refer, *inter alia*, to Moon and Perron (2004), Bai and Ng (2004), Bai, Kao, and Ng (2009), Kapetanios, Pesaran, and Yamagata (2011), and Pesaran, Smith, and Yamagata (2013). Further, the literature has recently extended the theory of cointegration to a high-dimensional setting (see Onatski and Wang, 2018 and Liang and Schienle, 2019); we point out that, while in this strand of the literature, the cointegration rank is implicitly assumed to be finite for any N, in our context the number of cointegration relationships grows with N. Finally, we note that the issue of finding the number of common factors has been studied at length in a stationary setting - examples include Bai and Ng (2002), Onatski (2010), Ahn and Horenstein (2013), and Trapani (2018). However, to the best of our knowledge, in a non-stationary setting like (1), only the information criteria developed by Bai (2004) are available.

The remainder of the paper is organised as follows. In Section 2 we outline the main ideas behind our testing procedure. In Section 3, we spell out the main assumptions and (in Section 3.2) 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 4. Numerical evidence from simulations is in Section 5. In Section 6 we report an empirical illustration on the dimension of the yield curve. Finally, Section 7 concludes. Proofs and further results are in the Supplement available online, where we also report further Monte Carlo evidence.

NOTATION. We define the Euclidean norm of a vector $a = [a_1, ..., a_n]$ as $\|a\| = (\sum_{i=1}^{n} a_i^2)^{1/2}$, and similarly for a matrix A, $\|A\|$ denotes the Frobenius norm; "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 Main ideas

We here present a heuristic preview of how the procedure works, while main arguments are laid out in the remainder of the paper. 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, $\nu_1^{(p)}$, $p = 1, \ldots, N$) 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, $\nu_2^{(p)}$, $p = 1, \ldots, N$), 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 deterministic or stochastic trends. 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 (2018), which would provide an indirect estimate of the number of common stationary factors and therefore of the number of cointegration relations between 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 (2018) 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 virtually no restrictions on the relative rates of divergence of N and T as they pass to infinity (see Assumption 5 below for details), 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, 2018, where a similar restriction is needed); in principle, our procedure can be applied to a wide variety of datasets, being particularly useful when either dimension is much bigger than the other (see also the discussion in Section A.5 of the Supplement).

3 Theory

3.1 Model and assumptions

The scalar version of our model (1) is:

$$X_{i,t} = \lambda_i' \mathcal{F}_t + u_{i,t}, \quad 1 \leq i \leq N, \tag{2}$$

where λ_i and $\mathcal{F}_t$ are $r \times 1$ vectors. We assume that

$$\mathcal{F}_t = a\,(d_1 t) + B\psi_t, \tag{3}$$

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, i.e. with zero mean.

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. The dummy variable d_2 is equal to 1 if the factors with a linear trend have also an $I(1)$ component.

We then have 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.

Lemma 1. *Under Assumption 1, model (2) can be equivalently represented as*

$$X_{i,t} = \lambda_i^{(1)} f_t^{(1)} + \lambda_i^{(2)\prime} f_t^{(2)} + \lambda_i^{(3)\prime} f_t^{(3)} + u_{i,t}, \quad 1 \leq i \leq N, \tag{4}$$

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* $r = r_1 + r_2 + r_3$ *is finite for all* N *and* T, *and* $\lambda_i' = (\lambda_i^{(1)\prime} \lambda_i^{(2)\prime} \lambda_i^{(3)\prime})$ *for all* i.
Moreover, the common non-stationary factors are defined by the following equations

$$f_t^{(1)} = d_1 t + d_2 f_t^{(1)\dagger} + (1 - d_2) g_t, \tag{5}$$

$$f_t^{(1)\dagger} = f_0^{(1)\dagger} + \sum_{j=1}^{t} e_j^{(1)}, \tag{6}$$

$$f_t^{(2)} = f_0^{(2)} + \sum_{j=1}^{t} e_j^{(2)}, \tag{7}$$

where in (5)-(7): $f_t^{(1)\dagger}$, 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, according as $d_1 = 1$ or 0: 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 (5)-(7), 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.

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

$$\Sigma_{\Delta f^*} = \lim_{T \to \infty} Var \left(T^{-1/2} \sum_{t=1}^{T} e_t \right). \tag{8}$$

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

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

Some comments on Assumptions 2 and 3 are in order. Both are high-level assumptions which can be shown to hold under a wide variety of commonly considered dependence as-

sumptions and DGPs. 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 *(v)* is a Burkholder-type inequality (see e.g. Lin and Bai, 2010, p. 108). The most delicate part of the assumption is part *(iv)*, which requires the existence of a strong approximation for the partial sums process f_t^*. In Section A.3 of the Supplement, we show that this part of the assumption can be verified for a wide variety of commonly considered forms of weak dependence (e.g. mixingales, Near Epoch Dependence and causal processes), and also for many commonly employed DGPs (e.g. linear models, threshold autoregressions, random coefficient autoregressive models, and univariate and multivariate conditional heteroskedasticity models).

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). Parts *(ii)*-*(v)* are equivalent to assuming weak dependence; in Section A.3 of the Supplement, we show that these assumptions hold under the same dependence assumptions and DGPs considered for Assumption 2*(iv)*.

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)* $\max_{1 \leq i \leq N} \|\lambda_i\| < \infty$*; (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.

An important consequence of Assumption 4 is that the common factors belonging in each category are "strong" or "pervasive". We point out that our setup can be generalised to consider also the case of weakly pervasive factors (see Section A.5 of the Supplement).

3.2 Asymptotic behaviour of eigenvalues

We introduce the following assumption on the relative rate of divergence of N and T as they pass to infinity.

Assumption 5. *It holds that:* $N = O\left(e^{T^{1/2-\epsilon'}}\right)$, *with* $0 < \epsilon' < \frac{1}{2}$.

In Assumption 5, ϵ' should be intended as arbitrarily close to zero. The assumption imposes a very mild restriction between N and T, which in practice should hold for virtually all datasets. Note that, whilst needing an upper bound for N, we do not need any lower bounds on either N or T as long as they pass to infinity.

We base inference on the two matrices

$$\Sigma_1 = \frac{1}{T^3}\sum_{t=1}^{T} X_t X_t', \tag{9}$$

$$\Sigma_2 = \frac{1}{T^2}\sum_{t=1}^{T} X_t X_t'. \tag{10}$$

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

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

where ϵ is an arbitrarily small number such that $0 < \epsilon < \frac{2\epsilon'}{1-2\epsilon'}$, with ϵ' defined in Assumption 5. 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* $T \geq T_0$,

$$\underline{C}_{1,p} N \leq \nu_1^{(p)} \leq \overline{C}_{1,p} N, \quad \text{for } p \leq r_1 d_1, \tag{11}$$

$$\nu_1^{(p)} = O_{a.s.}\left(\frac{N}{\sqrt{T}} l_{N,T}\right), \quad \text{for } p > r_1 d_1, \tag{12}$$

and

$$\underline{C}_{2,p}\frac{N}{\ln\ln T} \leq \nu_2^{(p)} \leq \overline{C}_{2,p}\frac{N}{\ln\ln T}, \quad \text{for } r_1 d_1 < p \leq r^*, \tag{13}$$

$$\nu_2^{(p)} = O_{a.s.}\left(\frac{N}{\sqrt{T}} l_{N,T}\right), \quad \text{for } p > r^*, \tag{14}$$

with $0 < \underline{C}_{1,p} \leq \overline{C}_{1,p} < \infty$, *and* $0 < \underline{C}_{2,p} \leq \overline{C}_{2,p} < \infty$.

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 (11) and (12) refer to the eigenvalues of Σ_1. The results state that the first r_1 eigenvalues diverge to infinity at a rate N; conversely, by virtue of Assumption 5, the remaining eigenvalues have a smaller magnitude. Indeed, the magnitude of $\nu_1^{(p)}$, when $p > r_1$, may be very large, but Assumption 5 ensures that it is however smaller than 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 (13) and (14) 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 (13) 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 (14) 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 $O_{a.s.}(Nl_{N,T})$ - 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 Theorem 1 it can be shown that the upper bound in (14) 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 (2):

$$\Delta X_{i,t} = \lambda_i' \Delta \mathcal{F}_t + \Delta u_{i,t}, \quad 1 \leq i \leq N, \tag{15}$$

such that the following holds:

Assumption 6. *It holds that: (i)* $E\left(\Delta \mathcal{F}_{j,t} \Delta u_{i,t}\right) = 0$ *for* $1 \leq j \leq r$ *and* $1 \leq i \leq N$*; (ii)* $\max_{1 \leq i \leq N, 1 \leq t \leq T} E\left|\Delta X_{i,t}\right|^4 \leq C_0$*; (iii)* $E \max_{1 \leq \widetilde{t} \leq T} \left|\sum_{t=1}^{\widetilde{t}} \Delta X_{h,t} \Delta X_{j,t} - E\left(\Delta X_{h,t} \Delta X_{j,t}\right)\right|^2 \leq C_0 T$*; (iv) (a)* $T^{-1} \sum_{t=1}^{T} E\left(\Delta \mathcal{F}_t \Delta \mathcal{F}_t'\right)$ *is a positive definite matrix; (b) the largest eigenvalue of* $T^{-1} \sum_{t=1}^{T} E\left(\Delta u_t \Delta u_t'\right)$ *is finite; (c)* $T^{-1} \sum_{t=1}^{T} E\left(\Delta u_t \Delta u_t'\right)$ *is a positive definite matrix.*

Assumption 6 is the same as Assumptions 1-3 in Trapani (2018), and we refer to that paper for examples in which the assumption is satisfied. Note that part *(iv)*(b) can be shown to hold if the largest eigenvalue of $E(u_t u_t')$ is bounded - i.e., in the presence of (weak) cross-sectional dependence in the idiosyncratic term.

4 Estimating the number of common factors

4.1 Preliminary definitions

Consider the notation $\beta = \frac{\ln N}{\ln T}$, and define $\delta < 1$ such that

$$\delta \begin{cases} > 0 & \text{when } \beta < \frac{1}{2}, \\ > 1 - \frac{1}{2\beta} & \text{when } \beta \geq \frac{1}{2}. \end{cases} \tag{16}$$

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.

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

$$\Sigma_3 = \frac{1}{T} \sum_{t=1}^{T} \Delta X_t \Delta X_t', \tag{17}$$

which, with the same notation as before, are denoted as $\nu_3^{(p)}$, $p = 1, \ldots, N$, 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

$$\overline{\nu}_{3,p}(k) = \frac{1}{4\,(N-k+1)} \sum_{h=k}^{N} \nu_3^{(h)}, \tag{18}$$

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 5, we consider rescaling schemes with $k = p$ and $k = p+1$, 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 (18) on k. Finally, note the division by 4 in (18), 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.

4.2 Determining the presence of factors with linear trends

Consider first Σ_1 defined in (9), and its eigenvalues $\nu_1^{(p)}$. Based on (11)-(12), 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

$$\begin{cases} H_{0,1}^{(p)} : \nu_1^{(p)} \geq C_p' N, \\ H_{A,1}^{(p)} : \nu_1^{(p)} \leq C_p'' \frac{N}{\sqrt{T}} l_{N,T}, \end{cases} \tag{19}$$

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 $\nu_1^{(p)}$:

$$\phi_1^{(p)} = \exp\left\{ N^{-\delta} \frac{\nu_1^{(p)}}{\overline{\nu}_{3,p}} \right\}. \tag{20}$$

Then, based on (19), equations (11) and (12), and given the definition (16) of δ, we have that

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

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, 2018).

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

Step A1.2. For a user-defined u, let, for $1 \leq j \leq R_1$, $\zeta_{1,j}^{(p)}(u) = I[\phi_1^{(p)} \times \xi_{1,j}^{(p)} \leq u]$.

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 ensures that $\vartheta_1^{(p)}(u)$ follows - asymptotically - a Normal distribution, and consequently $\Theta_1^{(p)}$ is asymptotically Chi-squared. By the same token, under the alternative $\phi_1^{(p)}$ is finite, and therefore $\zeta_{1,j}^{(p)}(u)$ is an *i.i.d.* Bernoulli sequence with expected value different from $G_1(0)$; thus, $\vartheta_1^{(p)}(u)$ diverges as fast as $\sqrt{R_1}$ by the LLN, and consequently $\Theta_1^{(p)}$ also diverges 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 behaviour of $\Theta_1^{(p)}$, we need some regularity conditions on the distributions G_1 and F_1 - see Section 5 for a choice of these functions and of u and R_1.

Assumption 7. *It holds that: (i) (a) G_1 has a bounded density function; (b) $G_1(0) \neq 0$ and $G_1(0) \neq 1$; (ii) $\int_{-\infty}^{\infty} dF_1(u) = 1$, and $\int_{-\infty}^{\infty} u^2 dF_1(u) < \infty$.*

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 "$\xrightarrow{D^*}$" and "$\xrightarrow{P^*}$" 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 (19). Under Assumptions 2-7, if*

$$\lim_{\min(N,R_1)\to\infty} \sqrt{R_1} \exp\left\{-N^{1-\delta}\right\} = 0, \tag{21}$$

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

$$\Theta_1^{(p)} \xrightarrow{D^*} \chi_1^2, \tag{22}$$

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

$$\frac{1}{R_1} \frac{\int_{-\infty}^{\infty} [G_1(u) - G_1(0)]^2 \, dF_1(u)}{G_1(0)[1 - G_1(0)]} \Theta_1^{(p)} \xrightarrow{P^*} 1. \tag{23}$$

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 $\{e_t, u_{i,t}, 1 \leq i \leq N, 1 \leq t \leq T\}$.*

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 under $H_{0,1}^{(p)}$

$$\lim_{\min(N,T,R_1)\to\infty} P^*\left(\Theta_1^{(p)} > c_{\alpha,1}\right) = 0. \tag{24}$$

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 $c_{\alpha,1} \to \infty$, (24) 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 5 for the choice of α_1.

As a final remark, we point out that this part of our procedure can detect whether there exists a common factor with a linear trend. However, our test is not able to directly establish whether such a trend arises from the common factor being an $I(1)$ process with a drift, or a trend-stationary process. Whilst this (important) question goes beyond the main focus of this paper, in Section A.4 of the Supplement we build on the theory developed here in order to propose a further test to discern between the two possible cases.

4.3 Determining the number of non-stationary common factors

Consider the matrix Σ_2 defined in (10) 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

$$\begin{cases} H_{0,2}^{(p)} : \nu_2^{(p)} \geq C'_p \frac{N}{\ln\ln T}, \\ H_{A,2}^{(p)} : \nu_2^{(p)} \leq C''_p \frac{N}{\sqrt{T}} l_{N,T}, \end{cases} \tag{25}$$

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

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

$$\phi_2^{(p)} = \exp\left\{ N^{-\delta} \left(\ln \ln T\right) \frac{\nu_2^{(p)}}{\overline{\nu}_{3,p}} \right\}, \tag{26}$$

which is very similar to (20) except for the presence of the logarithmic term, which is a consequence of (13). Then, based on (19), equations (13) and (14), and given the definition (16) of δ, we have that

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

We consider the following randomisation procedure.

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

Step A2.2 For a user-defined u, let, for $1 \leq j \leq R_2$, $\zeta_{2,j}^{(p)}(u) = I[\phi_2^{(p)} \times \xi_{2,j}^{(p)} \leq u]$.

Step A2.3. Compute $\vartheta_2^{(p)}(u) = \frac{1}{\sqrt{R_2}} \sum_{j=1}^{R_2} \frac{\zeta_{2,j}^{(p)}(u) - G_2(0)}{\sqrt{G_2(0)[1-G_2(0)]}}$.

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 8. *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) $\int_{-\infty}^{\infty} dF_2(u) = 1$ and $\int_{-\infty}^{\infty} u^2 dF_2(u) < \infty$.*

It holds that:

Theorem 3. *Consider $H_{0,2}^{(p)}$ and $H_{A,2}^{(p)}$ defined in (25). Under Assumptions 2-6 and 8, if*

$$\lim_{\min(N,R_2)\to\infty} \sqrt{R_2} \exp\left\{-N^{1-\delta}\right\} = 0, \tag{27}$$

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*

$$\Theta_2^{(p)} \xrightarrow{D^*} \chi_1^2, \tag{28}$$

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

$$\frac{1}{R_2} \frac{\int_{-\infty}^{\infty} \left(G_2(u) - G_2(0)\right)^2 dF_1(u)}{G_2(0)\left(1 - G_2(0)\right)} \Theta_2^{(p)} \xrightarrow{P^*} 1 \text{ under } H_1^{(2)}. \tag{29}$$

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

Step T2.1. 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.

Step T2.2. 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 $\{\xi_{2,j}^{(p+1)}\}_{j=1}^{R_2}$ generated independently of $\{\xi_{2,j}^{(1)}\}_{j=1}^{R_2}, \ldots, \{\xi_{2,j}^{(p)}\}_{j=1}^{R_2}$. 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.

4.4 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

$$\lim_{\min(N,T,R_1,R_2)\to\infty} P^* \left[\widehat{r}^* - \widehat{r}_1 = r_2 + r_1 \left(1 - d_1\right) d_2\right] = 1.$$

As a final remark, on the grounds of Assumption 6 it is possible to use the algorithm proposed in Trapani (2018) 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 (17) in a similar way to the algorithms above and using a simulated sample of size R_3. 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 (2018) and of Lemma 3 above, it follows that

$$\lim_{\min(N,T,R_1,R_2,R_3)\to\infty} P^* \left[\widehat{r} - \widehat{r}^* = r_3 + r_1 \left(1 - d_1\right) \left(1 - d_2\right)\right] = 1.$$

4.5 Estimation

Once we have estimates of the number of common factors, we can estimate the factors via principal component analysis as explained in Bai (2004) and Maciejowska (2010). Define

$$D = \begin{pmatrix} T^{3/2} I_{\widehat{r}_1} & 0 & 0 \\ 0 & T I_{\widehat{r}^* - \widehat{r}_1} & 0 \\ 0 & 0 & \sqrt{T} I_{\widehat{r} - \widehat{r}^*} \end{pmatrix}.$$

Let $\mathcal{F} = (\mathcal{F}_1 \cdots \mathcal{F}_T)'$ be the $T \times \widehat{r}$ matrix of factors and $X = (X_1 \cdots X_T)'$ the $T \times N$ matrix of data. Then, the estimated factors and loadings are solutions to the following constrained minimisation problem

$$(\widehat{\Lambda}, \widehat{\mathcal{F}}) = \arg\min (X - \mathcal{F}\Lambda')'(X - \mathcal{F}\Lambda'), \quad \text{s.t. } D^{-1}\mathcal{F}'\mathcal{F}D^{-1} = I_{\widehat{r}},$$

where, in particular, the solution for $\widehat{\mathcal{F}}$ are the eigenvectors of XX' multiplied by D and $\widehat{\Lambda} = D^{-2}\widehat{\mathcal{F}}'X$. Since we proved that the probability of correctly determining all number of factors tends to 1 as $N, T \to \infty$, the asymptotic results in Bai (2004) and Maciejowska (2010) for the estimated loadings and factors still hold.

5 Monte Carlo and empirical evidence

We generate data as

$$X_{i,t} = \lambda_i^{(1)} f_t^{(1)} + \lambda_i^{(2)'} f_t^{(2)} + \lambda_i^{(3)'} f_t^{(3)} + \sqrt{\theta} u_{i,t}, \quad 1 \leq i \leq N,\ 1 \leq t \leq T, \tag{30}$$

$$f_t^{(1)} = 1 + f_{t-1}^{(1)} + \epsilon_t^{(1)}, \tag{31}$$

$$f_{j,t}^{(2)} = f_{j,t-1}^{(2)} + e_{j,t}^{(2)}, \quad e_{j,t}^{(2)} = \rho_j e_{j,t-1}^{(2)} + \epsilon_{j,t}^{(2)}, \qquad j = 1, \ldots, r_2, \tag{32}$$

$$f_{j,t}^{(3)} = \alpha_j f_{j,t-1}^{(3)} + \epsilon_{j,t}^{(3)}, \qquad j = 1, \ldots, r_3, \tag{33}$$

$$u_{i,t} = a_i u_{i,t-1} + v_{i,t} + b_i \sum_{|k| \leq C_i, k \neq 0} v_{i+k,t}, \tag{34}$$

Note that model (30) 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. The loadings in (30) are simulated such that each nonzero entry is distributed as $\mathcal{N}(0,1)$, with Λ satisfying $\Lambda'\Lambda = NI_r$. In (32) and (33), we use $\rho_j \sim U[0.4, 0.8]$, 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

$$\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} (\lambda_i^{(1)} \Delta f_t^{(1)})^2 = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} (\lambda_i^{(2)'} \Delta f_t^{(2)})^2 = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} (\lambda_i^{(3)'} f_t^{(3)})^2,$$

so that in first differences each non-stationary factor component has, on average, the same weight as the stationary component. In (34) we allow both for serial and cross-sectional dependence in the idiosyncratic errors and for all $1 \leq i \leq N$, in a similar way to Ahn and Horenstein (2013). 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 $\mathcal{N}(0,1)$. Last, we set the noise-to-signal as

$$\theta = 0.5 \, \frac{\sum_{i=1}^{N} \sum_{t=1}^{T} (\lambda_i^{(1)} \Delta f_t^{(1)} + \lambda_i^{(2)'} \Delta f_t^{(2)} + \lambda_i^{(3)'} \Delta f_t^{(3)})^2}{\sum_{i=1}^{N} \sum_{t=1}^{T} (\Delta u_{i,t})^2}.$$

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 4).

We use $N \in \{50, 100, 200\}$ and $T \in \{100, 200, 500\}$, and simulate (30)-(34) 500 times, reporting the average value of $\widehat{r}_1$ or $\widehat{r}_2$ across simulations, as well as the standard deviation and the percentage of time that the estimator is correct. As far as $\widehat{r}_2$ is concerned, we report both the infeasible estimate $\widehat{r}^* - r_1$ and the feasible one $\widehat{r}^* - \widehat{r}_1$, in order to gauge the impact of the estimation error of $\widehat{r}_1$ on the subsequent stage of the procedure.

Our tests are run as follows. When computing $\phi_1^{(p)}$ and $\phi_2^{(p)}$, we rescale the p-th eigenvalue as (see (18))

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

For a given p, we consider two different rescaling schemes:

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

<u>$BT2$</u>: using $k = (p+1)$, i.e. $\bar{\nu}_{3,p}(k) = \frac{1}{4(N-p)} \sum_{h=p+1}^{N} \nu_3^{(h)}$.

Note that, by construction, $\bar{\nu}_{3,p}(p) \geq \bar{\nu}_{3,p}(p+1)$. This entails that, at least for finite samples, tests based on $BT1$ are less conservative than those using $BT2$. In turn, this entails that $BT1$ may have, in finite samples, a tendency to understate the number of common factors more often than $BT2$, and vice versa.

We then divide the eigenvalues by N^δ, where (see (16))

$$\delta = \begin{cases} \delta^*, & \text{when } \frac{\ln N}{\ln T} < \frac{1}{2}, \\ 1 - \frac{1}{2\beta} + \delta^*, & \text{when } \frac{\ln N}{\ln T} \geq \frac{1}{2}, \end{cases} \tag{35}$$

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 with G_1 and G_2 following a standard normal distribution, with $R_1 = 2N$ and $R_2 = 2N$, if $p = 1$ or $R_2 = \lfloor \frac{N}{3} \rfloor$, for $p > 1$. In the second step of the

randomisation algorithm, as weight functions $F_1(\cdot)$ and $F_2(\cdot)$ we choose the distribution of the standard normal; thus, in order to compute $\Theta_1^{(p)}$ and $\Theta_2^{(p)}$, we use

$$\Theta_k^{(p)} = \frac{1}{\sqrt{\pi}} \sum_{s=1}^{n_S} w_s \left(\vartheta_1^{(p)} \left(\sqrt{2} z_s \right) \right)^2, \quad k = 1, 2, \tag{36}$$

where the z_ss are the zeroes of the Hermite polynomial $H_{n_S}(z)$, and the weights w_s are defined as

$$w_s = \frac{\sqrt{\pi} 2^{n_S - 1} (n_S - 1)!}{n_S \left[H_{n_S - 1}(z_s) \right]^2}.$$

In practice, this entails that, when computing $\Theta_1^{(p)}$ (and $\Theta_2^{(p)}$), we construct n_S different versions of $\vartheta_1^{(p)}(u)$ (resp. $\vartheta_2^{(p)}(u)$), using $u = \sqrt{2} z_s$. The values of z_s and the corresponding weights w_s are tabulated e.g. in Salzer, Zucker, and Capuano (1952); in our simulations, we have set $n_S = 4$, which corresponds to $u = \{-2.4, -0.75, 0.75, 2.4\}$, with weights $\pi^{-1/2} w_s$ given by $\{0.05, 0.45, 0.45, 0.05\}$. As well as being a consequence of the Gauss-Hermite quadrature formula, (36) also has an impact on the trade-off between the power and size of each test: as discussed in Trapani (2018), the larger $|u|$, the larger the power and the more oversized the individual tests (thus leading to a potential underestimation of the number of common factors), and vice versa. Heuristically, having "large" values such as ± 2.4, with "small" weights set to 0.05, helps to strike a balance between size and power. Finally, 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.

Results are in Tables 1-5 (see also the further Monte Carlo evidence reported in Section B of the Supplement). As a general point, $BT1$ performs better than $BT2$ at finding no common factors - whether with a linear trend or genuinely $I(1)$ with zero mean - when there are no common factors; this is also consistent with the results in the further simulations in the Supplement. Considering Table 1, there are some cases in which $BT1$ finds a common factor when this is not present, but this error vanishes as T increases. On the other hand, $BT2$ also performs well, but there are instances in which it tends to overstate r_1 when $r_1 = 0$, particularly in the case where there is no common factor with a linear trend but there is one zero-mean $I(1)$ common factor.

As far as determining r_2 is concerned, $BT1$ and $BT2$ are comparable, with $BT1$ dom-

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

		$N=50, T=100$						$N=100, T=100$					
		average		std. dev.		% correct		average		std. dev.		% correct	
r_1	r_2	$BT1$	$BT2$	$BT1$	$BT2$	$BT1$	$BT2$	$BT1$	$BT2$	$BT1$	$BT2$	$BT1$	$BT2$
0	0	0.00	0.00	0.00	0.00	1.00	1.00	0.00	0.00	0.00	0.00	1.00	1.00
0	1	0.19	0.17	0.39	0.37	0.81	0.83	0.11	0.12	0.31	0.32	0.89	0.88
0	2	0.04	0.03	0.19	0.18	0.96	0.97	0.01	0.01	0.10	0.11	0.99	0.99
1	0	1.00	1.00	0.04	0.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00	1.00
1	1	1.00	1.00	0.00	0.00	1.00	1.00	1.00	1.00	0.00	0.04	1.00	1.00
1	2	1.00	1.00	0.00	0.00	1.00	1.00	1.00	1.00	0.06	0.00	1.00	1.00
		$N=200, T=100$						$N=100, T=200$					
		average		std. dev.		% correct		average		std. dev.		% correct	
r_1	r_2	$BT1$	$BT2$	$BT1$	$BT2$	$BT1$	$BT2$	$BT1$	$BT2$	$BT1$	$BT2$	$BT1$	$BT2$
0	0	0.00	0.00	0.00	0.00	1.00	1.00	0.00	0.00	0.00	0.00	1.00	1.00
0	1	0.05	0.33	0.21	0.47	0.95	0.67	0.06	0.07	0.24	0.26	0.94	0.93
0	2	0.01	0.06	0.09	0.23	0.99	0.94	0.01	0.01	0.11	0.10	0.99	0.99
1	0	1.00	1.00	0.00	0.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00	1.00
1	1	1.00	1.00	0.00	0.04	1.00	1.00	1.00	1.00	0.00	0.00	1.00	1.00
1	2	1.00	1.00	0.00	0.00	1.00	1.00	1.00	1.00	0.00	0.04	1.00	1.00
		$N=200, T=200$						$N=200, T=500$					
		average		std. dev.		% correct		average		std. dev.		% correct	
r_1	r_2	$BT1$	$BT2$	$BT1$	$BT2$	$BT1$	$BT2$	$BT1$	$BT2$	$BT1$	$BT2$	$BT1$	$BT2$
0	0	0.00	0.00	0.00	0.00	1.00	1.00	0.00	0.00	0.00	0.00	1.00	1.00
0	1	0.02	0.20	0.15	0.40	0.98	0.80	0.00	0.12	0.00	0.33	1.00	0.88
0	2	0.00	0.02	0.06	0.15	1.00	0.98	0.00	0.01	0.00	0.10	1.00	0.99
1	0	1.00	1.00	0.06	0.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00	1.00
1	1	1.00	1.00	0.04	0.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00	1.00
1	2	1.00	1.00	0.00	0.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00	1.00

In each cell we report the average and standard deviation of $\widehat{r}_1$ over all Monte Carlo replications, as well as the fraction of times in which $\widehat{r}_1 = r_1$.

inating over $BT2$ when estimating r_2 in the case $r_1 = 0$. In general, our criteria tend to understate r_2, albeit slightly, when $r_1 = 0$, especially in the presence of linear trends; and, conversely, to overstate, albeit again slightly, when $r_1 = 1$. In both cases, the bias tends to vanish as T increases. Overall, the performance of all our criteria improves dramatically as T increases: results markedly improve when $T \geq 200$ for all cases considered. The impact of N is, in general, less clear. Comparing Tables 2-4 with Tables 3-5, the impact of using $\widehat{r}_1$ when estimating $\widehat{r}_2$ is such that when $r_1 = 0$, $\widehat{r}_2$ tends to understate r_2, and vice versa when $r_1 = 1$. However, this error quickly disappears as soon as $T \geq 100$. These results are in line with the observation that in finte samples $BT1$ is less conservative than $BT2$, however this discrepancy quickly becomes negligible as the sample size increases.

As a final remark, we point out that in Section B of the Supplement, we have carried out some sensitivity analysis to assess the impact of different specifications of the DGP and of the tests. In particular, as far as the DGP is concerned, we consider $\rho_j \sim U[\bar{\rho}, 0.8]$ with values of $\bar{\rho}$ equal to 0 and 0.8: results are broadly the same, with few exceptions (see

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

		$N=50, T=100$						$N=100, T=100$					
		average		std. dev.		% correct		average		std. dev.		% correct	
r_2	r_3	$BT1$	$BT2$	$BT1$	$BT2$	$BT1$	$BT2$	$BT1$	$BT2$	$BT1$	$BT2$	$BT1$	$BT2$
0	0	0.00	0.00	0.00	0.00	1.00	1.00	0.00	0.00	0.00	0.00	1.00	1.00
0	1	0.00	0.00	0.00	0.00	1.00	1.00	0.00	0.00	0.00	0.00	1.00	1.00
0	2	0.00	0.00	0.00	0.00	1.00	1.00	0.00	0.00	0.00	0.00	1.00	1.00
1	0	1.00	1.00	0.04	0.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00	1.00
1	1	1.00	1.00	0.00	0.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00	1.00
1	2	1.00	1.00	0.00	0.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00	1.00
2	0	1.99	1.99	0.10	0.09	1.00	0.99	1.98	1.98	0.15	0.15	0.98	0.98
2	1	1.97	1.97	0.19	0.19	0.97	0.97	1.94	1.93	0.24	0.25	0.94	0.93
2	2	1.98	1.98	0.16	0.16	0.98	0.98	1.94	1.93	0.25	0.26	0.94	0.93
		$N=200, T=100$						$N=100, T=200$					
		average		std. dev.		% correct		average		std. dev.		% correct	
r_2	r_3	$BT1$	$BT2$	$BT1$	$BT2$	$BT1$	$BT2$	$BT1$	$BT2$	$BT1$	$BT2$	$BT1$	$BT2$
0	0	0.00	0.00	0.00	0.00	1.00	1.00	0.00	0.00	0.00	0.00	1.00	1.00
0	1	0.00	0.00	0.00	0.00	1.00	1.00	0.00	0.00	0.00	0.00	1.00	1.00
0	2	0.00	0.00	0.00	0.00	1.00	1.00	0.00	0.00	0.00	0.00	1.00	1.00
1	0	1.00	1.00	0.04	0.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00	1.00
1	1	1.00	1.00	0.04	0.04	1.00	1.00	1.00	1.00	0.00	0.00	1.00	1.00
1	2	1.00	1.00	0.04	0.00	1.00	1.00	1.00	1.00	0.04	0.00	1.00	1.00
2	0	1.98	1.99	0.14	0.11	0.99	0.99	1.98	1.99	0.14	0.09	0.99	0.99
2	1	1.93	1.98	0.27	0.16	0.93	0.98	1.99	1.99	0.13	0.09	0.99	0.99
2	2	1.94	1.98	0.24	0.15	0.94	0.98	1.99	1.99	0.10	0.09	0.99	0.99
		$N=200, T=200$						$N=200, T=500$					
		average		std. dev.		% correct		average		std. dev.		% correct	
r_2	r_3	$BT1$	$BT2$	$BT1$	$BT2$	$BT1$	$BT2$	$BT1$	$BT2$	$BT1$	$BT2$	$BT1$	$BT2$
0	0	0.00	0.00	0.00	0.00	1.00	1.00	0.00	0.00	0.00	0.00	1.00	1.00
0	1	0.00	0.00	0.00	0.00	1.00	1.00	0.00	0.00	0.00	0.00	1.00	1.00
0	2	0.00	0.00	0.00	0.00	1.00	1.00	0.00	0.00	0.00	0.00	1.00	1.00
1	0	1.00	1.00	0.00	0.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00	1.00
1	1	1.00	1.00	0.00	0.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00	1.00
1	2	1.00	1.00	0.00	0.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00	1.00
2	0	2.00	2.00	0.04	0.04	1.00	1.00	2.00	2.00	0.04	0.00	1.00	1.00
2	1	1.99	2.00	0.08	0.06	0.99	1.00	1.99	2.00	0.09	0.00	0.99	1.00
2	2	2.00	2.00	0.06	0.00	1.00	1.00	2.00	2.00	0.00	0.00	1.00	1.00

In each cell we report the average and standard deviation of $\widehat{r}_2$ over all Monte Carlo replications, as well as the fraction of times in which $\widehat{r}_2 = r_2$.

discussion in Section B.2 of the Supplement). As far as the test specifications are concerned, we have considered different choices of G_1 and G_2, and of F_1 and F_2 (Section B.3), as well as δ^* (Section B.4) and the R_1 and R_2 (Section B.5). Results show that the tests are virtually unaffected by such specifications, whose impact vanishes as N and T increase. In the Supplement, we have also reported results when estimating r_2 using the information criteria proposed by Bai (2004) (Section B.1); these criteria invariably find one common $I(1)$ factor even when such factors are not present.

Table 3: Estimated number of zero-mean $I(1)$ factors, $\widehat{r}_2$ when $r_1 = 0$, using $\widehat{r}_1$.

		$N = 50, T = 100$						$N = 100, T = 100$					
		average		std. dev.		% correct		average		std. dev.		% correct	
r_2	r_3	$BT1$	$BT2$	$BT1$	$BT2$	$BT1$	$BT2$	$BT1$	$BT2$	$BT1$	$BT2$	$BT1$	$BT2$
0	0	0.00	0.00	0.00	0.00	1.00	1.00	0.00	0.00	0.00	0.00	1.00	1.00
0	1	0.00	0.00	0.00	0.00	1.00	1.00	0.00	0.00	0.00	0.00	1.00	1.00
0	2	0.00	0.00	0.00	0.00	1.00	1.00	0.00	0.00	0.00	0.00	1.00	1.00
1	0	0.81	0.82	0.40	0.39	0.81	0.82	0.88	0.88	0.32	0.33	0.88	0.88
1	1	0.93	0.94	0.25	0.24	0.93	0.94	0.98	0.98	0.14	0.15	0.98	0.98
1	2	0.93	0.94	0.25	0.24	0.93	0.94	0.96	0.97	0.19	0.18	0.96	0.97
2	0	1.96	1.96	0.21	0.20	0.96	0.96	1.97	1.97	0.19	0.18	0.97	0.97
2	1	1.96	1.97	0.20	0.19	0.97	0.97	1.94	1.93	0.24	0.25	0.94	0.93
2	2	1.98	1.97	0.17	0.17	0.98	0.98	1.94	1.93	0.25	0.26	0.94	0.93
		$N = 200, T = 100$						$N = 100, T = 200$					
		average		std. dev.		% correct		average		std. dev.		% correct	
r_2	r_3	$BT1$	$BT2$	$BT1$	$BT2$	$BT1$	$BT2$	$BT1$	$BT2$	$BT1$	$BT2$	$BT1$	$BT2$
0	0	0.00	0.00	0.00	0.00	1.00	1.00	0.00	0.00	0.00	0.00	1.00	1.00
0	1	0.00	0.00	0.00	0.00	1.00	1.00	0.00	0.00	0.00	0.00	1.00	1.00
0	2	0.00	0.00	0.00	0.00	1.00	1.00	0.00	0.00	0.00	0.00	1.00	1.00
1	0	0.95	0.67	0.23	0.47	0.95	0.67	0.94	0.94	0.23	0.23	0.94	0.94
1	1	0.99	0.94	0.10	0.24	0.99	0.94	0.98	0.99	0.13	0.12	0.98	0.99
1	2	0.99	0.96	0.11	0.20	0.99	0.96	0.99	0.99	0.12	0.10	0.99	0.99
2	0	1.98	1.93	0.15	0.26	0.98	0.94	1.98	1.99	0.15	0.09	0.98	0.99
2	1	1.93	1.98	0.27	0.17	0.93	0.98	1.99	1.99	0.13	0.09	0.99	0.99
2	2	1.94	1.98	0.24	0.15	0.94	0.98	1.99	1.99	0.10	0.09	0.99	0.99
		$N = 200, T = 200$						$N = 200, T = 500$					
		average		std. dev.		% correct		average		std. dev.		% correct	
r_2	r_3	$BT1$	$BT2$	$BT1$	$BT2$	$BT1$	$BT2$	$BT1$	$BT2$	$BT1$	$BT2$	$BT1$	$BT2$
0	0	0.00	0.00	0.00	0.00	1.00	1.00	0.00	0.00	0.00	0.00	1.00	1.00
0	1	0.00	0.00	0.00	0.00	1.00	1.00	0.00	0.00	0.00	0.00	1.00	1.00
0	2	0.00	0.00	0.00	0.00	1.00	1.00	0.00	0.00	0.00	0.00	1.00	1.00
1	0	0.98	0.79	0.13	0.41	0.98	0.79	0.99	0.88	0.08	0.32	0.99	0.88
1	1	1.00	0.98	0.06	0.13	1.00	0.98	1.00	1.00	0.00	0.04	1.00	1.00
1	2	0.99	0.99	0.08	0.12	0.99	0.99	1.00	0.99	0.00	0.10	1.00	0.99
2	0	2.00	1.98	0.04	0.15	1.00	0.98	2.00	1.99	0.04	0.11	1.00	0.99
2	1	1.99	2.00	0.08	0.06	0.99	1.00	1.99	2.00	0.09	0.00	0.99	1.00
2	2	2.00	2.00	0.06	0.00	1.00	1.00	2.00	2.00	0.00	0.00	1.00	1.00

In each cell we report the average and standard deviation of $\widehat{r}_2$ over all Monte Carlo replications, as well as the fraction of times in which $\widehat{r}_2 = r_2$.

6 On the dimensions of the yield curve

We illustrate our methodology through an application to the High Quality Market (HQM) Corporate Bond Yield Curve, available from the Federal Reserve Economic Data (FRED)[1] - details on the construction of the yield curves are available from the US Department of Treasury.[2] 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.

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

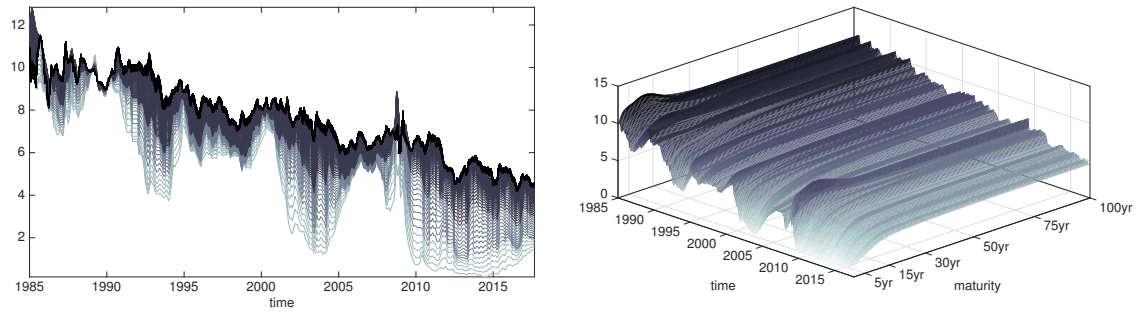
[2] `https://www.treasury.gov/resource-center/economic-policy/corp-bond-yie.`

Table 4: Estimated number of zero-mean $I(1)$ factors, $\widehat{r}_2$ when $r_1 = 1$, using $r_1 = 1$.

		$N = 50, T = 100$						$N = 100, T = 100$					
		average		std. dev.		% correct		average		std. dev.		% correct	
r_2	r_3	$BT1$	$BT2$	$BT1$	$BT2$	$BT1$	$BT2$	$BT1$	$BT2$	$BT1$	$BT2$	$BT1$	$BT2$
0	0	0.00	0.00	0.08	0.12	0.99	0.99	0.02	0.03	0.17	0.17	0.98	0.97
0	1	0.02	0.02	0.14	0.13	0.98	0.98	0.04	0.05	0.22	0.22	0.96	0.95
0	2	0.01	0.01	0.08	0.10	0.99	0.99	0.03	0.04	0.18	0.20	0.97	0.96
1	0	1.00	0.99	0.08	0.10	0.99	1.00	1.01	1.01	0.13	0.12	0.98	0.99
1	1	1.00	1.01	0.08	0.16	0.99	0.98	1.04	1.03	0.20	0.18	0.96	0.97
1	2	1.01	1.01	0.10	0.12	0.99	0.99	1.02	1.03	0.21	0.18	0.97	0.97
2	0	1.86	1.84	0.37	0.40	0.85	0.83	1.92	1.93	0.32	0.33	0.89	0.89
2	1	1.72	1.71	0.47	0.48	0.71	0.69	1.81	1.83	0.44	0.45	0.77	0.79
2	2	1.71	1.66	0.49	0.50	0.72	0.67	1.82	1.82	0.43	0.45	0.79	0.78
		$N = 200, T = 100$						$N = 100, T = 200$					
		average		std. dev.		% correct		average		std. dev.		% correct	
r_2	r_3	$BT1$	$BT2$	$BT1$	$BT2$	$BT1$	$BT2$	$BT1$	$BT2$	$BT1$	$BT2$	$BT1$	$BT2$
0	0	0.00	0.00	0.00	0.04	1.00	1.00	0.02	0.01	0.17	0.11	0.98	0.99
0	1	0.00	0.07	0.06	0.25	1.00	0.93	0.04	0.03	0.21	0.18	0.96	0.97
0	2	0.00	0.00	0.00	0.04	1.00	1.00	0.02	0.04	0.16	0.21	0.97	0.96
1	0	1.00	1.00	0.08	0.04	0.99	1.00	1.04	1.03	0.20	0.20	0.96	0.96
1	1	1.00	1.02	0.06	0.15	1.00	0.98	1.03	1.03	0.19	0.19	0.97	0.97
1	2	1.00	1.00	0.11	0.09	0.99	1.00	1.02	1.03	0.19	0.16	0.97	0.97
2	0	1.99	1.99	0.10	0.09	0.99	0.99	1.99	1.99	0.19	0.23	0.96	0.96
2	1	1.95	1.98	0.27	0.19	0.93	0.96	1.99	1.98	0.28	0.28	0.93	0.92
2	2	1.92	1.97	0.30	0.19	0.92	0.96	1.95	1.96	0.28	0.36	0.92	0.90
		$N = 200, T = 200$						$N = 200, T = 500$					
		average		std. dev.		% correct		average		std. dev.		% correct	
r_2	r_3	$BT1$	$BT2$	$BT1$	$BT2$	$BT1$	$BT2$	$BT1$	$BT2$	$BT1$	$BT2$	$BT1$	$BT2$
0	0	0.00	0.00	0.04	0.00	1.00	1.00	0.00	0.00	0.04	0.00	1.00	1.00
0	1	0.01	0.06	0.08	0.24	0.99	0.94	0.00	0.02	0.06	0.15	1.00	0.98
0	2	0.00	0.01	0.06	0.10	1.00	0.99	0.00	0.00	0.04	0.00	1.00	1.00
1	0	1.00	1.00	0.10	0.09	1.00	1.00	1.00	1.00	0.00	0.06	1.00	1.00
1	1	1.01	1.02	0.13	0.13	0.99	0.98	1.00	1.01	0.00	0.08	1.00	0.99
1	2	1.00	1.00	0.09	0.00	1.00	1.00	1.00	1.00	0.00	0.06	1.00	1.00
2	0	2.00	2.00	0.00	0.06	1.00	1.00	2.00	2.00	0.09	0.06	1.00	1.00
2	1	2.00	2.02	0.13	0.15	0.98	0.98	2.00	2.00	0.04	0.04	1.00	1.00
2	2	1.99	2.00	0.14	0.06	0.99	1.00	1.99	1.99	0.14	0.13	1.00	1.00

In each cell we report the average and standard deviation of $\widehat{r}_2$ over all Monte Carlo replications, as well as the fraction of times in which $\widehat{r}_2 = r_2$.

Figure 1: HQM Corporate Bond Yield Curve



We use the same settings as in Section 5. 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

Table 5: Estimated number of zero-mean $I(1)$ factors, $\widehat{r}_2$ when $r_1 = 1$, using $\widehat{r}_1$.

		$N=50, T=100$						$N=100, T=100$					
		average		std. dev.		% correct		average		std. dev.		% correct	
r_2	r_3	$BT1$	$BT2$	$BT1$	$BT2$	$BT1$	$BT2$	$BT1$	$BT2$	$BT1$	$BT2$	$BT1$	$BT2$
0	0	0.01	0.00	0.10	0.11	0.99	0.99	0.02	0.03	0.18	0.18	0.97	0.97
0	1	0.02	0.02	0.14	0.14	0.98	0.98	0.04	0.05	0.22	0.22	0.96	0.95
0	2	0.01	0.01	0.08	0.10	0.99	0.99	0.03	0.04	0.18	0.20	0.97	0.96
1	0	1.00	0.99	0.09	0.10	0.99	1.00	1.01	1.01	0.14	0.12	0.98	0.99
1	1	1.00	1.01	0.08	0.16	0.99	0.98	1.04	1.03	0.21	0.18	0.96	0.97
1	2	1.01	1.02	0.11	0.13	0.99	0.98	1.02	1.03	0.23	0.19	0.96	0.96
2	0	1.86	1.84	0.37	0.40	0.85	0.83	1.92	1.93	0.32	0.33	0.89	0.89
2	1	1.73	1.71	0.48	0.48	0.70	0.69	1.82	1.84	0.46	0.46	0.77	0.78
2	2	1.71	1.66	0.49	0.51	0.72	0.66	1.83	1.83	0.44	0.46	0.78	0.77
		$N=200, T=100$						$N=100, T=200$					
		average		std. dev.		% correct		average		std. dev.		% correct	
r_2	r_3	$BT1$	$BT2$	$BT1$	$BT2$	$BT1$	$BT2$	$BT1$	$BT2$	$BT1$	$BT2$	$BT1$	$BT2$
0	0	0.00	0.00	0.04	0.04	1.00	1.00	0.02	0.01	0.17	0.11	0.98	0.99
0	1	0.00	0.07	0.06	0.27	1.00	0.93	0.04	0.03	0.21	0.18	0.96	0.97
0	2	0.00	0.00	0.04	0.04	1.00	1.00	0.02	0.04	0.16	0.21	0.97	0.96
1	0	1.00	1.00	0.09	0.06	0.99	1.00	1.04	1.03	0.20	0.20	0.96	0.96
1	1	1.03	1.02	0.18	0.15	0.97	0.98	1.03	1.03	0.19	0.19	0.97	0.97
1	2	1.04	1.00	0.24	0.10	0.95	1.00	1.02	1.03	0.19	0.16	0.97	0.97
2	0	1.99	1.99	0.10	0.09	0.99	0.99	1.99	1.99	0.19	0.23	0.96	0.96
2	1	1.99	1.99	0.32	0.19	0.90	0.96	1.99	1.98	0.28	0.28	0.93	0.92
2	2	1.96	1.97	0.37	0.21	0.88	0.96	1.95	1.96	0.29	0.36	0.92	0.90
		$N=200, T=200$						$N=200, T=500$					
		average		std. dev.		% correct		average		std. dev.		% correct	
r_2	r_3	$BT1$	$BT2$	$BT1$	$BT2$	$BT1$	$BT2$	$BT1$	$BT2$	$BT1$	$BT2$	$BT1$	$BT2$
0	0	0.00	0.00	0.04	0.04	1.00	1.00	0.00	0.00	0.04	0.00	1.00	1.00
0	1	0.01	0.06	0.09	0.24	0.99	0.94	0.00	0.02	0.06	0.15	1.00	0.98
0	2	0.00	0.01	0.06	0.10	1.00	0.99	0.00	0.00	0.04	0.00	1.00	1.00
1	0	1.00	1.00	0.10	0.09	1.00	1.00	1.00	1.00	0.00	0.06	1.00	1.00
1	1	1.01	1.02	0.13	0.13	0.99	0.98	1.00	1.01	0.00	0.09	1.00	0.99
1	2	1.00	1.00	0.09	0.00	1.00	1.00	1.00	1.00	0.00	0.06	1.00	1.00
2	0	2.00	2.00	0.00	0.06	1.00	1.00	2.00	2.00	0.09	0.06	1.00	1.00
2	1	2.00	2.02	0.13	0.15	0.98	0.98	2.00	2.00	0.04	0.06	1.00	1.00
2	2	1.99	2.00	0.14	0.06	0.99	1.00	1.99	1.99	0.14	0.13	1.00	1.00

In each cell we report the average and standard deviation of $\widehat{r}_2$ over all Monte Carlo replications, as well as the fraction of times in which $\widehat{r}_2 = r_2$.

level is $\frac{0.05}{\min(N,T)} = 0.0002551$.

In our analysis, we also report the information criterion $IC3$ proposed by Bai (2004).[3] Results are in Table 6. Both $BT1$ and $BT2$ find three non-stationary common factors - this is broadly in line with the stylised facts in this literature where a three factor model is customarily employed. Whilst both criteria agree that two of these factors are $I(1)$ with mean zero, there is some evidence that the first common factor may have a linear trend. Indeed, this is picked up by $BT2$ but not by $BT1$, which is compatible with the theory (which stipulates that $BT2$ may have a tendency to find common factors more often than $BT1$) and the evidence from the simulations. Thus, our procedures do not rule out the presence

[3]We note that, when computing $\widehat{r}$, this is also equivalent to $IC3$ in Bai and Ng (2002)

Table 6: Estimated number of factors in the HQM Corporate Bond Yield Curve

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

of a factor with a time trend, a finding in agreement with the debate on the evidence on the secular decline of interest rates (see e.g. the discussion in Rachel and Summers, 2019).

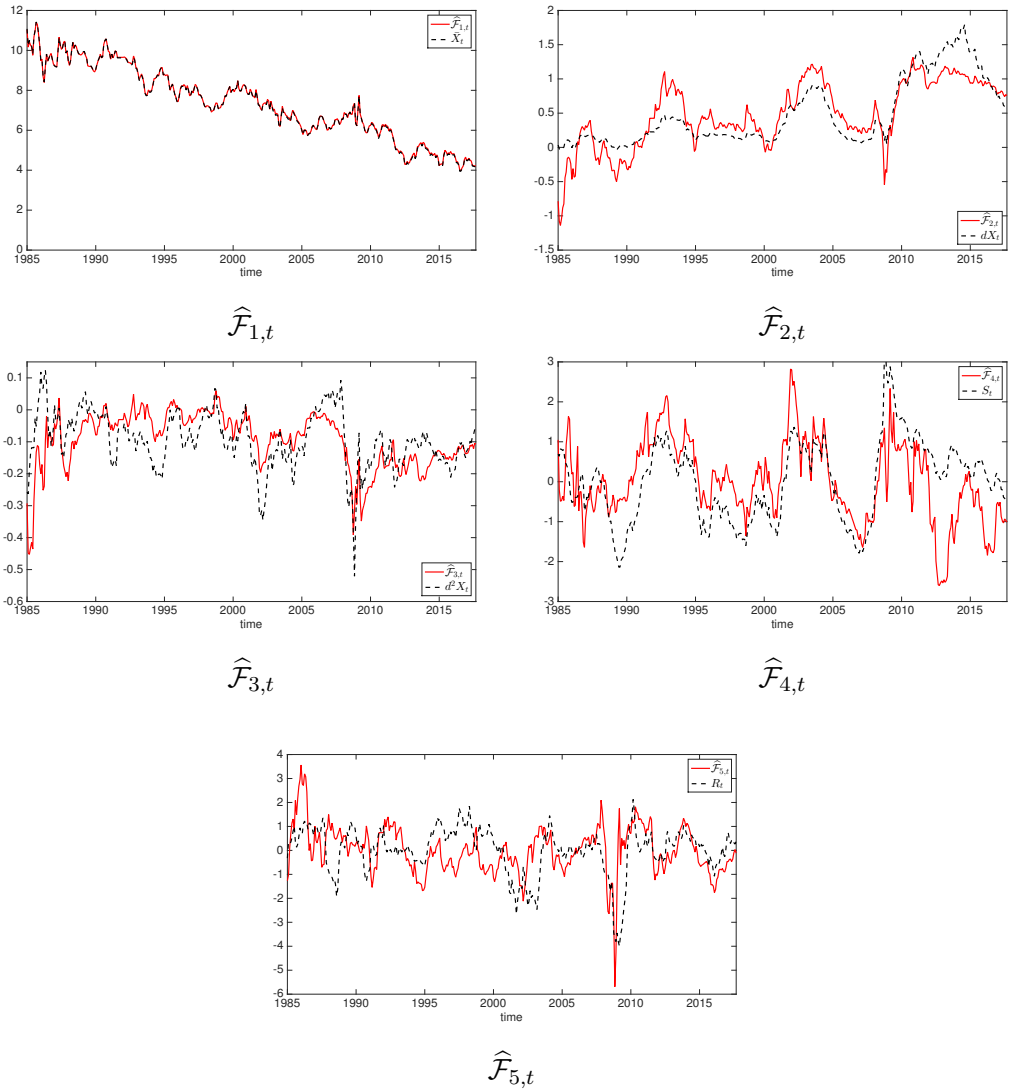
We have also applied the test by Trapani (2018) to the first-differenced data, finding a total number of 5 common factors, which suggests that there are two further $I(0)$ common factors. Note that IC finds 5 non-stationary factors in total, which may suggest that the last two factors are either stationary (as found by $BT1$ and $BT2$), or marginally non-stationary.

The estimated factors are shown in Figure 2 (solid red lines). 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, 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 classical first three – see for example Duffie and Singleton (1999), Duffie, Saita, and Wang (2007), and Christensen and Lopez (2008).

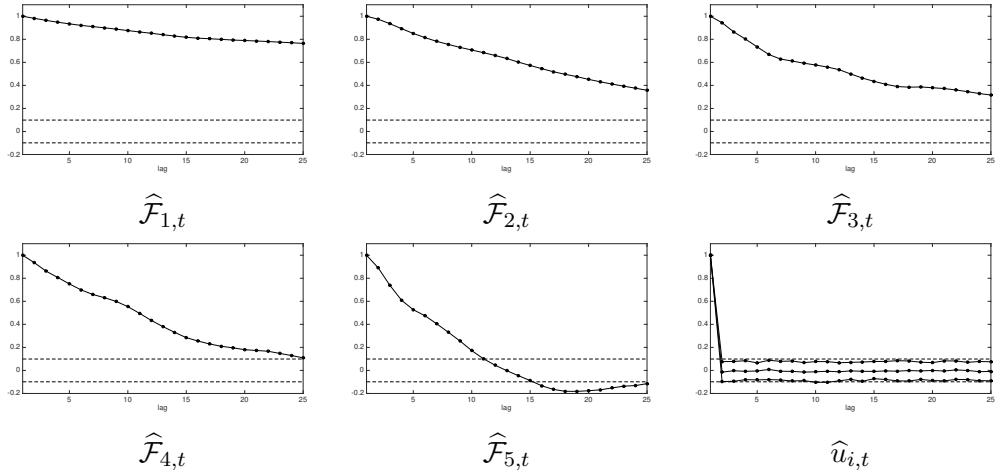
First we analyse the first three estimated common factors. At each point in time t, the N elements of X_t are ordered according to their maturity, thus $X_{1,t}$ is the shortest maturity (6 months), while $X_{N,t}$ is the longest maturity (100 years). We compare each estimated factor

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



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 $\text{Corr}(\bar{X}_t, \widehat{\mathcal{F}}_{1,t}) \simeq 1$, which strongly suggests that $\widehat{\mathcal{F}}_{1,t}$ can be viewed as the *level* of the curve. Turning to $\widehat{\mathcal{F}}_{2,t}$, we use, as a proxy for the slope, $dX_t = (N-1)^{-1}\sum_{i=2}^{N}(\ln X_{i,t} - \ln X_{i-1,t}) = N^{-1}(\ln X_{N,t} - \ln X_{1,t})$. 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

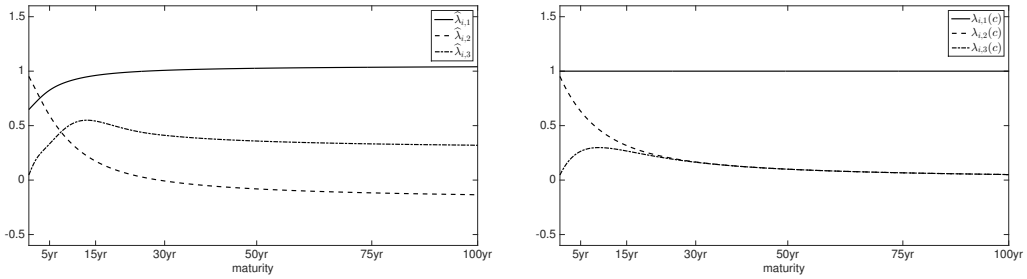
Figure 3: Autocorrelation of common factors $\widehat{\mathcal{F}}_{j,t}$ and idiosyncratic errors $\widehat{u}_{i,t}$.



$\widehat{\mathcal{F}}_{1,t}$ $\widehat{\mathcal{F}}_{2,t}$ $\widehat{\mathcal{F}}_{3,t}$

$\widehat{\mathcal{F}}_{4,t}$ $\widehat{\mathcal{F}}_{5,t}$ $\widehat{u}_{i,t}$

Figure 4: Estimated and theoretical factor loadings.



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

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

$$\lambda_{i,1}(c) = 1, \qquad \lambda_{i,2}(c) = \left(\frac{1 - e^{-ci}}{ci}\right), \qquad \lambda_{i,3}(c) = \left(\frac{1 - e^{-ci}}{ci} - e^{-ci}\right), \tag{37}$$

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 (37) computed for $c = 0.2$.[4]

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,

[4]The chosen value of c is such that it minimizes $\sum_{i=1}^{N}(\widehat{\lambda}_{i,2} - \lambda_{i,2}(c))^2$.

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.[5] 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 lines).

7 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. (2018).

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

[5] 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`.

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. (2019), although it is based on tests rather than an information criterion, and it entertains the possibility that linear trends could be present.

We conclude by mentioning several interesting issues which, albeit not explicitly considered in this paper, can be studied by extending the results derived above. Firstly, all our common factors are assumed to be pervasive; however, based on the discussion in Trapani (2018), the individual tests can be used even in the presence of weak factors. Secondly, as also mentioned in the comments to Theorem 1, it is possible to use our set-up - with no need for changes to it - in the case of $u_{i,t} \sim I(1)$ for at least some i; this would be helpful in order to disentangle common and idiosyncratic sources of non-stationarity. Finally, we note that, by suitably rescaling the second moment matrix of the data, our approach can also be generalised to $I(d)$ factors with $d > 1$.

References

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

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., 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.

Chen, L. and W. B. Wu (2019). Testing for trends in high-dimensional time series. *Journal of the American Statistical Association 114*(526), 869–881.

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.

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.

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

Liang, C. and M. Schienle (2019). Determination of vector error correction models in high dimensions. *Journal of econometrics 208*(2), 418–441.

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

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

Onatski, A. and C. Wang (2018). Alternative asymptotics for cointegration tests in large vars. *Econometrica 86*(4), 1465–1478.

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.

Phillips, P. C. and S. Ouliaris (1988). Testing for cointegration using principal components methods. *Journal of Economic Dynamics and Control 12*(2-3), 205–230.

Rachel, Ł. and L. H. Summers (2019). On secular stagnation in the industrialized world. Technical report, National Bureau of Economic Research.

Salzer, H. E., R. Zucker, and R. Capuano (1952). Table of the zeros and weight factors of the first twenty Hermite polynomials. *Journal of Research of the National Bureau of Standards 48*(2), 111.

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

Trapani, L. (2018). A randomized sequential procedure to determine the number of factors. *Journal of the American Statistical Association 113*(523), 1341–1349.

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

Zhang, B., G. Pan, and J. Gao (2018). CLT for largest eigenvalues and unit root testing for high-dimensional nonstationary time series. *The Annals of Statistics 46*(5), 2186–2215.

Zhang, R., P. Robinson, and Q. Yao (2019). Identifying cointegration by eigenanalysis. *Journal of the American Statistical Association 114*(526), 916–927.