Spurious Relationships in High Dimensional Systems with Strong or Mild Persistence

Jesús Gonzalo
Universidad Carlos III de Madrid
Department of Economics
jesus.gonzalo@uc3m.es

Jean-Yves Pitarakis
University of Southampton
Department of Economics
j.pitarakis@soton.ac.uk

October 18, 2020

Abstract

This paper is concerned with the interactions of persistence and dimensionality in the context of the eigenvalue estimation problem of large covariance matrices arising in cointegration and principal component analysis. Following a review of the early and more recent developments in this area we investigate the behaviour of these eigenvalues in a VAR setting that blends pure unit root, local to unit root and mildly integrated components. Our results highlight the seriousness of spurious relationships that may arise in such Big Data environments even when the degree of persistence of variables involved is mild and is affecting only a small proportion of a large data matrix with important implications for forecasts based on principal component regressions and related methods. We argue that first differencing prior to principal component analysis may be suitable even in stationary or nearly-stationary environments.

Keywords: Spurious cointegration, spurious factors, persistence, high dimensional covariances, principal components.

1. Introduction

The advent of Big Data through the increasing availability of data collection methods and new data types resulting from novel interactions between economic agents has had and continues to have a profound impact on conventional econometric methodologies. In such environments the classical framework of a fixed number of variables p and a large number of observations n is no longer suitable for making best use of massive quantities of information and for accurately disentangling meaningful patterns from noise. Over the past twenty years, this has led to a vast body of research broadly referred to as High Dimensional Statistics and Machine Learning methods that seek to develop novel estimation and forecasting techniques or adapt existing methods to environments where $p \sim n$, $p \gg n$ and more generically $p = p(n) \to \infty$. Important methods that seek to adapt to large p environments include high dimensional covariance estimation techniques via shrinkage and thresholding methods, principal component regressions, regression trees, random forests, penalised estimation to just name a few (see Hastie, Tibshirani and Friedman (2009), Bühlmann and Van de Geer (2011), Koch (2014) amongst others).

A notable feature of these developments has been their emphasis on prediction rather than on *explanation* and the rather limited attention that has been paid to the influence of variables' time series characteristics such as their degree of persistence on the properties and reliability of these methods. It is well known for instance that independent but highly persistent, unit-root or local to unit-root processes tend to display spurious correlations and one may then speculate that these may be further amplified in high dimensional environments. Similarly, the phenomenon of cointegration which has traditionally been viewed as a dimensionality reduction technique and a convenient way of addressing the spurious regression problem may in turn spuriously manifest itself in large p settings.

As recognised in Granger (1998) when Big Data concepts were still in their infancy, the move from classical fixed p/large or moderate n statistics to Big Data environments raises three important questions: (i) which of our commonly used standard procedures should be discarded (e.g. tests of statistical significance, small sample adjustments, simulation based methods such as the jacknife and bootstrap), (ii) which new techniques need to be developed (e.g. for the analysis of mega-panels) and (iii) whether some existing procedures are expected to perform better in such settings (e.g. conventional asymptotic approximations). A casual overview of the recent High Dimensional Statistics literature suggests that these insights very much materialised in the form of buoyant research agendas encompassing all three of the above questions (see for instance Johnstone and Titterington (2009), Varian (2014), Carrasco, Chernozhukov, Goncalves and Renault (2015), Mullainathan and Spiess (2017), Ng (2017) and references therein). At the same time the proliferation of time-series variables being used within many of these *Big Data* techniques raises the additional question of whether the persistent nature of these data may amplify spurious outcomes in high dimensional environments (e.g. findings of spurious relationships and correlations). We view this as an important concern to have when evaluating the properties of these new methods or when adapting

existing methods to high dimensional settings.

In the area of forecasting with Big Data for instance principal components have become an important tool for condensing information from hundreds of time series into one or two dominant linear combinations which are then used as predictors within predictive regressions (e.g. diffusion index based forecasting via principal component regressions). How reliable are these principal components in summarising the underlying data when the latter consist of many persistent time series? How persistent should a group of time series be for persistence to become an important concern when extracting principal components? The issue also bears strong resemblance with the literature on cointegrating regressions since cointegration itself may be viewed as a dimensionality reduction device in inherently persistent systems as illustrated by Gonzalo and Granger (1995)'s common factor representation of a cointegrated system.

Whether one wishes to condense a large data set into few dominant components as in principal component analysis or explain a large system of persistent components with a smaller number of common factors as in cointegration analysis, the essential tools for achieving such objectives rely on the information provided by the eigenvalues of sample covariance matrices. Understanding how these eigenvalues behave in high dimensional environments with persistence is therefore key to understanding the nature of the distortions that may arise is such settings.

The main objective of this paper is to investigate the behaviour of eigenvalues of sample covariance matrices obtained from a simple vector autoregressive system that combines high dimensionality with persistence. We introduce a series of new analytical results that illustrate the potentially serious distortions practitioners may face when techniques relying on principal components and related methods are applied to high dimensional environments characterised by persistence *even* when this persistence is reasonably mild (e.g. findings of spurious factors, spurious cointegration, spurious predictability). An important message also conveyed by our analysis is that growingly popular techniques commonly labelled as machine learning methods need to be analysed through the lens of uncorrelated persistent data prior to recommending their use in a time series modelling context.

The plan of the paper is as follows. Section 2 reviews early and more recent developments on the influence of dimensionality in systems characterised by the presence of persistence and cointegration, starting from our early work in Gonzalo and Pitarakis (1995,1999) and followed by the more recent contributions in Onatski and Wang (2018a, 2018b, 2019). Section 3 develops our main results on the large sample behaviour of sample eigenvalues obtained from high dimensional persistent systems with and without cointegration. These are in turn illustrated through a broad range of simulation experiments documenting empirically the importance of the spuriousness phenomenon. Section 4 concludes. All proofs are relegated to the appendix.

2. Dimensionality induced spuriousness in cointegration analysis

The growth of the unit-root and cointegration literature since the early 80s has placed VAR and VECM based modelling environments at the heart of empirical research involving time series data, spanning areas such as impulse response analysis, forecasting, structural macro modelling, forecast error variance decompositions amongst others. The Granger representation theorem ensuring that systems of cointegrated variables admit a VECM representation also highlighted the dangers of misspecifying the cointegration properties of a system with potentially severe implications for the above types of analyses (e.g. using first differenced cointegrated VARs with omitted error correction components). This has prompted a rich research agenda on testing for cointegration with Johansen's VAR based framework emerging as the workhorse model for conducting such inferences (Johansen (1988, 1991)).

Although in their early days these VAR/VECM models were mainly specified using a small number of variables due to degrees of freedom concerns or the nature of questions being investigated (e.g. Blanchard and Quah (1989)'s bivariate structural VAR, Gali (1992)'s four dimensional cointegrated VAR; King, Plosser, Stock and Watson (1991)'s cointegrated trivariate VAR) the important economic implications of detecting the presence of common trends also led to numerous empirical applications implementing these conventional methods in higher dimensional VARs with limited concern for the potential distortions that may arise in such settings (e.g. Baillie and Bollerslev (1989), Kasa (1992), Diebold, Gardeazabal and Yilmaz (1994) amongst numerous others).

These considerations have motivated our own work on *dimensionality effects* within cointegrated I(1) systems in the early to mid 90s. Our main concerns and agenda at the time were geared towards diagnosing the nature of the distortions that may be affecting traditional inferences implemented on large dimensional VARs/VECMs (Gonzalo and Pitarakis (1995, 1999)) and in particular highlighting the fact that dimensionality induced distortions go well beyond simple complications caused by limited degrees of freedom and the importance of considering asymptotic approximations that explicitly allow both the system dimension and sample size to grow.

Consider the following simple p-dimensional system of I(1) variables formulated in a VECM form

$$\Delta \boldsymbol{X}_t \quad = \quad \boldsymbol{\Pi} \boldsymbol{X}_{t-1} + \boldsymbol{u}_t \tag{1}$$

where the reduced rank nature of the long run impact matrix $\boldsymbol{\Pi}$, say $\boldsymbol{\Pi} = \boldsymbol{\alpha}_{p\times r}\boldsymbol{\beta}'_{r\times p}$ conveys the relevant information about the cointegration properties of the system and the fact that the components of $\boldsymbol{X}_t$ share $(p-r)$ common trends and with $\boldsymbol{\Pi} = \boldsymbol{0}$ under $r = 0$. A likelihood ratio based testing approach of the null hypothesis that $Rank(\boldsymbol{\Pi}) = r$ can then be formulated in terms of the canonical correlations between $\Delta \boldsymbol{X}_t$ and $\boldsymbol{X}_{t-1}$. More specifically, letting $\boldsymbol{S}_{11} = \sum \boldsymbol{X}_{t-1}\boldsymbol{X}'_{t-1}$ $\boldsymbol{S}_{00} = \sum \Delta \boldsymbol{X}_t \Delta \boldsymbol{X}'_t$ and $\boldsymbol{S}_{10} = \sum \boldsymbol{X}_{t-1}\Delta \boldsymbol{X}'_t$, Johansen's LR based test statistic (Johansen (1988, 1991)) is formulated as $\mathrm{LR}_n = -n \sum_{i=r+1}^{p} \ln(1-\hat{\lambda}_i)$

with the $\hat{\lambda}_i's$ denoting the ordered eigenvalues of $\boldsymbol{S}_{11}^{-1}\boldsymbol{S}_{10}\boldsymbol{S}_{00}^{-1}\boldsymbol{S}_{01}$. An important question we were concerned with in Gonzalo and Pitarakis (1995, 1999) was about the adequacy of the commonly used asymptotic distribution of this LR_n statistic in high dimensional settings and the question of how fast could p be allowed to grow with n for the conventional asymptotics to continue to be reliable in the sense of the LR_n statistic not displaying excessive size distortions. We approached the problem by evaluating the behaviour of the asymptotic mean of the LR_n statistics' distribution, establishing that

$$E[\mathrm{LR}_n] \quad \approx \quad p(2p-1) + \frac{p^2(2p-1)}{n-p} \tag{2}$$

with the second term in the right hand side of (2) highlighting the severe degree of mis-centeredness that will characterise $\mathrm{LR}_n's$ distribution in high dimensional settings even under large sample sizes, unless one restricts the rate of increase of p relative to the sample size n (e.g. $p^3/n \to 0$). This dimensionality induced rightward shift of $\mathrm{LR}_n's$ distribution translates into excessive size distortions when inferences are based on conventional asymptotic critical values (fixed p/large n) resulting in a phenomenon we coined as *spurious cointegration*. This has also led us to argue that alternative functionals of the $\hat{\lambda}_i's$ may alleviate these dimensionality induced distortions through a milder tendency to propagate the upward biases in the $\hat{\lambda}_i's$ while maintaining the same limiting distribution as the LR_n statistic (note that in the vicinity of zero we have $-\ln(1-\lambda_i) \approx \lambda_i + \lambda_i^2/2 + \lambda_i^3/3 + O(\lambda_i)^4$). We considered two such statistics. A Pillai-Bartlett type functional given by $\mathrm{PB}_n = n\sum_{i=r+1}^{p}\hat{\lambda}_i$ and a novel linear combination statistic given by $(\mathrm{LR}_n + \mathrm{PB}_n)/n \equiv \mathrm{LC}_n$ and shown to be substantially more robust to dimensionality induced distortions (see Figure 1 in Gonzalo and Pitarakis (1999, p. 217)) maintaining a seemingly perfect size behaviour in large p/small n VAR settings.

Although a lot of these early results were established heuristically through a combination of theoretical and simulation based approaches their relevance has become particularly topical in Big Data time series environments where spuriousness is and needs to be viewed as a major concern. In a series of recent papers Onatski and Wang (2018a, 2018b, 2019) developed a rigorous theoretical analysis of the interactions between cointegration and dimensionality using formal random matrix theory methods, corroborating many of our early conjectures. Operating within an $(n,p) \to \infty$ setting combined with a Marchenko-Pastur type of regime whereby $p/n \to \gamma \in (0,1]$ and assuming $r/p \to 0$ the authors obtained the remarkable result that the empirical distribution function (EDF) $F_{n,p}(\lambda) = \sum_{i=1}^{p} I(\hat{\lambda}_i \leq \lambda)/p$ of the $\hat{\lambda}_i's$ converges to a deterministic limit known as Wachter's distribution with its parameters depending on the magnitude of γ (see Wachter (1978, 1980)).

Wachter's distribution was originally obtained in the context of the roots of determinental equations of the type $|\boldsymbol{S}_1 - \hat{\lambda}\boldsymbol{S}_2| = 0$ for two given covariance matrices $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$. Noting that these roots are the eigenvalues of $\boldsymbol{S}_2^{-1}\boldsymbol{S}_1$ the setting can be seen to be analogous to the canonical correlation based framework for testing for cointegration. Taking $\boldsymbol{S}_1$ as the residual covariance matrix of (1), say $\boldsymbol{S}_1 \equiv \hat{\boldsymbol{\Omega}}_u$,

and $\boldsymbol{S}_2 \equiv \boldsymbol{S}_{00}$ for instance it is straightforward to establish that the eigenvalues of $\boldsymbol{S}_{11}^{-1}\boldsymbol{S}_{10}\boldsymbol{S}_{00}^{-1}\boldsymbol{S}_{01}$ are the same as the eigenvalues of $\boldsymbol{I}_p - \boldsymbol{S}_{00}^{-1}\hat{\boldsymbol{\Omega}}_u$ (see Gonzalo and Pitarakis (1995)). What is then particularly interesting and powerful about the result in Onatski and Wang (2019) is that this early Wachter limit obtained under a stationary Gaussian IID setting continues to hold within nonstationary VARs and in particular in a context where $\boldsymbol{X}_t$ and $\boldsymbol{\Delta X}_t$ are not independent. In hindsight, our observation about the connection between the eigenvalues of $\boldsymbol{S}_{11}^{-1}\boldsymbol{S}_{10}\boldsymbol{S}_{00}^{-1}\boldsymbol{S}_{01}$ and $\boldsymbol{I}_p - \boldsymbol{S}_{00}^{-1}\hat{\boldsymbol{\Omega}}_u$ helps highlight the intuition that under the null hypothesis that $\boldsymbol{\Pi} = \boldsymbol{0}$ the matrices $\boldsymbol{S}_{00}$ and $\hat{\boldsymbol{\Omega}}_u$ will consist solely of components on which standard laws of large numbers apply.

This result about the Wachter limit of the eigenvalues arising in cointegration testing subsequently allowed the authors to formally establish the mis-centering phenomenon characterising the LR_n statistic under a fixed p setting with the distribution of LR_n/p^2 shown to concentrate in the vicinity of 2 under $n \to \infty$. Exploring the upper boundaries of the Wachter distribution across alternative magnitudes of γ provides a clear rationale for the excessive size distortions characterising the LR_n statistic when operating in a high dimensional setting while relying on fixed p asymptotics (e.g. under $\gamma = 1/5$ the largest canonical correlation under no cointegration ($r = 0$) will exceed 0.7 and under $\gamma = 1/2$ it is expected to lie in the vicinity of unity when in fact it should be zero in the population). These results also offer a clear rationale for the improved performance of the linear combination statistic LC_n introduced in Gonzalo and Pitarakis (1995, 1999) as PB_n and LR_n's distortions in the high-dimensional regime are shown to move in opposite directions in a symmetric manner.

Although the above results have been obtained within simple VAR settings the messages they convey go well beyond the confines of toy models and cointegration/unit-root based analyses. The tendency for spurious relationships to be further amplified in high dimensional environments combined with unit root type of extreme persistence will almost certainly result in false discoveries of high correlation and spurious predictability across most popular regularisation and prediction algorithms commonly used in the machine learning literature.

In what follows we aim to focus on one such regularisation and dimensionality reduction method that has been particularly popular in the area of economic forecasting involving big data sets, namely principal component analysis and principal component regressions. These have given rise to a rich agenda on diffusion index type indicator building for business cycle analysis such as the Chicago Fed's National Activity Index (see also Stock and Watson (1999)), forecasting (Stock and Watson (1989, 2002)), time-series modelling with factor augmented VARs and VECMs (Stock and Watson (2005), Gao and Tsay (2019) amongst others).

Principal component analysis essentially aims to condense the variance structure of a large number of variables into a smaller set of their linear combinations (PCs) that preserve as much information as possible from the original data set, in the sense of capturing a large proportion of its variance. These (potentially small) number of PCs can in turn be used as stand-alone variables proxying the underlying

high-dimensional data (e.g. via predictive principal component regressions that use these PCs as predictors themselves). Naturally the key issue that arises in such contexts is the reliable detection and estimation of this small number of PCs that are able to capture a large proportion of the variance of the original data in a non-spurious manner. Given the close link between PCA and the eigenvalues and eigenvectors of the data covariance matrix our analysis also provides important insights on the interactions between persistence, cointegration and high dimensionality.

3. Dimensionality induced spuriousness under strong and mild persistence in principal component analysis

This section aims to formalise the relationship between persistence and dimensionality through the lens of covariance matrix behaviour and associated principal components. Operating within a series of toy models commonly encountered in the time-series literature and designed to be flexible enough to capture high persistence, pure unit-root behaviour, near stationarity and cointegration we explore the limiting behaviour of key eigenvalues of covariance matrices formed on data generated from these models. These are in turn used to investigate the joint impact of dimensionality and persistence on the principal component analysis of this data. Our analysis is conducted within a high dimensional setting with $p = p(n) \to \infty$

3.1. A High Dimensional Environment with Persistence

We consider an environment where the p dimensional random vector of interest $\boldsymbol{X}_t$ is modelled as

$$\boldsymbol{X}_t \quad = \quad \left(\boldsymbol{I}_p - \frac{\boldsymbol{C}_p}{n^\alpha} \right) \boldsymbol{X}_{t-1} + \boldsymbol{u}_t \quad t = 1, \ldots, n \tag{3}$$

where $\boldsymbol{C}_p = diag(c, \ldots, c)$ for $c \geq 0$, $\alpha \in [0, 1]$ and $\boldsymbol{u}_t$ is the p-dimensional random disturbance vector. The system in (3) specialises to p random walks when $c = 0$ while under $c > 0$ and $\alpha = 1$ its components are said to be nearly integrated. The parameter α controls the speed at which the random walk boundary is approached and helps add an additional layer of fine-tuning to the persistence properties of the system. The case associated with $\alpha = 1$ is commonly referred to as a local to unit root (LUR) model while lower magnitudes of α turn the components of $\boldsymbol{X}_t$ into nearly stationary or stationary processes where it is understood that $|1 - c| < 1$ under $\alpha = 0$ and $c > 0$. Throughout this paper we also operate under the assumption of zero initial conditions, setting $\boldsymbol{X}_0 = \boldsymbol{0}$.

We next let $\boldsymbol{X}_{n \times p}$ denote the $n \times p$ data matrix formed from the components of $\boldsymbol{X}_t$. Each one of its p

columns consists of the observation vectors $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{in})'$ for $i = 1, \ldots, p$. More specifically, we have

$$\boldsymbol{X}_{n\times p} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p) = \begin{bmatrix} x_{11} & x_{21} & x_{31} & \ldots & x_{p1} \\ x_{12} & x_{22} & x_{32} & \ldots & x_{p2} \\ x_{13} & x_{23} & x_{33} & \ldots & x_{p3} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{1n} & x_{2n} & x_{3n} & \ldots & x_{pn} \end{bmatrix} \tag{4}$$

and we also specify the random disturbance matrix $\boldsymbol{U}_{n\times p}$ analogously, with its columns consisting of the disturbance vectors $\boldsymbol{u}_i = (\boldsymbol{u}_{i1}, \ldots, \boldsymbol{u}_{in})'$ with $i = 1, \ldots, p$.

It is now convenient to reformulate (3) in the following matrix form

$$\boldsymbol{R}_n(c, \alpha)\, \boldsymbol{X} = \boldsymbol{U} \tag{5}$$

where the non-random $n \times n$ matrix $\boldsymbol{R}_n(c, \alpha)$ takes the bi-diagonal form

$$\boldsymbol{R}_n(c, \alpha) = \begin{bmatrix} 1 & 0 & 0 & \ldots & 0 & 0 \\ -\rho_n(c, \alpha) & 1 & 0 & \ldots & 0 & 0 \\ 0 & -\rho_n(c, \alpha) & 1 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \\ 0 & 0 & 0 & \vdots & -\rho_n(c, \alpha) & 1 \end{bmatrix} \tag{6}$$

with $\rho_n(c, \alpha) = 1 - c/n^{\alpha}$.

Letting $\boldsymbol{H}_n = \boldsymbol{I}_n - n^{-1}\mathbb{1}\mathbb{1}'$ denote the $n \times n$ centering matrix and $\boldsymbol{X}_H = \boldsymbol{H}_n\boldsymbol{X}$ the corresponding centered data matrix we can now formulate the centered covariance matrix in outer product form as

$$\boldsymbol{S}_n = \frac{1}{p}\boldsymbol{X}_H\boldsymbol{X}_H'. \tag{7}$$

Note that $\boldsymbol{X}_H\boldsymbol{X}_H'$ shares the same non-zero eigenvalues as $\boldsymbol{X}_H'\boldsymbol{X}_H$ which are also closely linked to the cointegration properties of the multivariate system. As $\boldsymbol{S}_n$ is expressed in outer product form, under $p > n$ for instance its eigenvalues will be the same as the nonzero eigenvalues of its inner product based counterpart. Similarly, under $p < n$ its $(n - p)$ non-zero eigenvalues will also be the same as those of its inner product counterpart. Naturally, for $n = p$ the inner and outer product forms will have the same eigenvalues.

Although the econometrics literature traditionally refers to $\boldsymbol{S}_n$ as the covariance matrix of $\boldsymbol{X}$, its

uncentered counterpart, say $\boldsymbol{S}_n^{uc}$

$$\boldsymbol{S}_n^{uc} = \frac{1}{p}\boldsymbol{X}\boldsymbol{X}' \tag{8}$$

is often the main object of interest in the machine learning literature as well as the vast literature on the spectral properties of covariances. As we discuss in greater detail below there are subtle differences between these two constructions, in particular when $\rho_n(c,\alpha)$ falls in the vicinity of unity and therefore when relevant we will be explicitly distinguishing between the properties of $\boldsymbol{S}_n$ and the properties of $\boldsymbol{S}_n^{uc}$.

Using (5) we can now write $\boldsymbol{X}_H = \boldsymbol{H}_n\boldsymbol{R}_n(c,\alpha)^{-1}\boldsymbol{U}$ so that

$$\boldsymbol{S}_n = \frac{1}{p}\boldsymbol{H}_n\boldsymbol{R}_n(c,\alpha)^{-1}\boldsymbol{U}\boldsymbol{U}'(\boldsymbol{R}_n(c,\alpha)^{-1})'\boldsymbol{H}_n \tag{9}$$

and

$$\boldsymbol{S}_n^{uc} = \frac{1}{p}\boldsymbol{R}_n(c,\alpha)^{-1}\boldsymbol{U}\boldsymbol{U}'(\boldsymbol{R}_n(c,\alpha)^{-1})' \tag{10}$$

and we also let $\hat{\lambda}_k$ and $\hat{\lambda}_k^{uc}$ denote the k^{th} largest eigenvalues of $\boldsymbol{S}_n$ and $\boldsymbol{S}_n^{uc}$ respectively.

In what follows we aim to explore the behaviour of the eigenvalues of $\boldsymbol{S}_n$ and $\boldsymbol{S}_n^{uc}$ in a high-dimensional setting allowing the system dimension p to grow with n. These will inform our understanding of how dimensionality and persistence interact, potentially leading to spurious relationships in the form of spurious principal components and spurious cointegrating relationships. More importantly, a specific objective of our analysis is to quantify the degree of persistence of the components included in $\boldsymbol{X}$ beyond which spuriousness becomes a major concern. Throughout this paper we operate under the following set of assumptions.

ASSUMPTION A: (i) $p = p(n) \to \infty$ as $n \to \infty$; (ii) $\boldsymbol{u}_t \sim \text{IID}(\boldsymbol{0}, \boldsymbol{I}_p)$ with $E[u_{it}^4] < \infty$ for $i = 1, \ldots, p$ and $t = 1, \ldots, n$; (iii) For a non-random $n \times n$ symmetric matrix $\boldsymbol{A}_n\boldsymbol{A}_n'$ whose eigenvalues are of order n^δ and a continuous function $\psi(.)$ it holds that $n^{-\delta}|\psi(\boldsymbol{A}_n(\boldsymbol{U}\boldsymbol{U}'/p)\boldsymbol{A}_n) - \psi(\boldsymbol{A}_n\boldsymbol{A}_n')| \xrightarrow{p} 0$.

Part (iii) of Assumption A is a high level assumption that is a law of large number type of statement for particular functionals of the random disturbance matrix $\boldsymbol{U}_n$ (e.g. eigenvalues). It essentially implies that the two matrices share the same limiting spectral distributions so that the eigenvalues of $\boldsymbol{S}_n$ or $\boldsymbol{S}_n^{uc}$ can be analysed via the limiting spectral distribution of $\boldsymbol{A}_n\boldsymbol{A}_n'$. Upon imposing further primitive conditions on the spectrum of $\boldsymbol{A}_n$ it can in fact be shown that part (iii) of Assumption A follows directly from part (ii) and suitable restrictions on the way p and n interact (see Bai and Silverstein (2010, Chapter 3)).

We are now in a position to state our main results about the behaviour of the eigenvalues and trace of the covariance matrices of interest. As the distinction between (7) and (8) becomes relevant for $\rho_n(c,\alpha)$ in the vicinity of unity only, we will be considering these two cases separately solely for such scenarios

while focusing solely on $\boldsymbol{S}_n$ when operating in purely stationary settings.

An important quantity whose behaviour we are particularly interested in exploring is given by $\hat{\lambda}_k/\sum_{k=1}\hat{\lambda}_k \equiv \hat{\lambda}_k/Tr(\boldsymbol{S}_n)$ which estimates the proportion of the total variance captured by the k^{th} principal component (PCk therefater) of $\boldsymbol{X}_H$. This ratio is a fundamental quantity that is omnipresent in the analysis of high dimensional data as it is used to signal how many PCs would be sufficient to proxy for the underlying higher dimensional data. Within an environment of completely independent components we naturally expect this ratio to be negligible while a sufficiently large magnitude for the first few $k's$ would indicate that a small set of PCs will be sufficient to capture most of the information of the underlying high dimensional data. Note that in the context of the uncentered covariance the same ratio denoted as $\hat{\lambda}_k^{uc}/\sum_{k=1}\hat{\lambda}_k^{uc}$ is equally relevant although strictly speaking it would not be referred to as capturing variance proportions but uncentered second moments instead. Nevertheless in what follows we use the same terminology to refer to these two ratios.

Propositions 1 and 2 below treat the case where the components of $\boldsymbol{X}$ are local to unit root processes, an environment we refer to as highly persistent. Proposition 1 focuses on the behaviour of the eigenvalues of the uncentered matrix $\boldsymbol{S}_n^{uc}$ while Proposition 2 deals with the centered scenario associated with $\boldsymbol{S}_n$. As we discuss extensively in the appendix, the limiting behaviour of these sample eigenvalues is determined by the eigenvalues of special structured matrices which although analytically tractable do not always lend themselves to closed form solutions for any magnitude of $\rho_n(c,\alpha)$. These however can be approximated highly accurately through suitable asymptotic expansions and hence our notation below where we use the symbol $\sim$ to emphasise the fact that the quantities in the right hand side of (12) and (14) are approximations for λ_k^{uc} and v_k^{uc} in (11) and (13).

PROPOSITION 1 (Uncentered Covariance): *Under Assumption A, for given k and* $\rho_n(c,\alpha) = 1 - c/n$ *with* $c > 0$ *we have*

$$\frac{\hat{\lambda}_k^{uc}}{n^2} \xrightarrow{p} \lambda_k^{uc} \tag{11}$$

where

$$\lambda_k^{uc} \sim \frac{-1 + 2c + e^{-2c}}{2c\tanh(c)}\left(\frac{1}{c^2 + (k-\frac{1}{2})^2\pi^2}\right) \equiv \lambda_{*k}^{uc} \tag{12}$$

and

$$\frac{\hat{\lambda}_k^{uc}}{\sum_{k=1}^n \hat{\lambda}_k^{uc}} \xrightarrow{p} v_k^{uc} \tag{13}$$

where

$$v_k^{uc} \sim \frac{2c}{\tanh(c)}\left(\frac{1}{c^2+(k-\frac{1}{2})^2\pi^2}\right) \equiv v_{*k}^{uc}. \tag{14}$$

The key result in (14) provides interesting insights on how the interaction of persistence and dimensionality leads to spurious relationships. Note for instance that under $c = 1$ and $k = 1$ we have $v_{*1}^{uc} = 75.7\%$ highlighting a substantial proportion of variation captured by the first eigenvalue when in fact the data consist of independent processes. More importantly this spuriousness can also be seen to remain significant even for large magnitudes of c. Under $c = 10$ for instance we have $v_{*1}^{uc} = 19.5\%$ and $v_{*2}^{uc} = 16.4\%$ so that the first two eigenvalues spuriously capture 35.9% of the corresponding trace.

It is now also interesting to assess the behaviour of (12) and (14) as $c \to 0$. It is indeed a simple algebraic exercise to establish that

$$\lim_{c\to 0} \lambda_{*k}^{uc} = \frac{4}{(2k-1)^2\pi^2} \tag{15}$$

and

$$\lim_{c\to 0} v_{*k}^{uc} = \frac{8}{(2k-1)^2\pi^2}. \tag{16}$$

The above two quantities are in fact the exact (as opposed to the approximations in (12) and (14)) limiting outcomes for (11) and (13) as the pure unit root scenario lends itself to tractable closed forms as discussed in the appendix. The quantity in (15) is well known in the literature as it corresponds to the eigenvalues that appear in the Karhunen-Loève representation of a Brownian Motion. The expression in (16) is also particularly interesting as it provides an exact formulation for the proportion of "variance" captured by each eigenvalue. It highlights the fact that spuriousness peaks at the exact unit root boundary with 81.1% $(= 8/\pi^2)$ of the trace (spuriously) captured by the first eigenvalue.

We next concentrate on the case of the centered covariance matrix with our results summarised in Proposition 2 below.

PROPOSITION 2 (Centered Covariance): *Under Assumption A, for given k and $\rho_n(c,\alpha) = 1 - c/n$ with $c > 0$ we have*

$$\frac{\hat{\lambda}_k}{n^2} \xrightarrow{p} \lambda_k \tag{17}$$

where

$$\lambda_k \quad \sim \quad \frac{6+2c^2+2e^{-2c}-8e^{-c}+c(-5+e^{-2c})}{2c(-1+c\coth(c))}\left(\frac{1}{c^2+k^2\pi^2}\right) \equiv \lambda_{*k} \tag{18}$$

and

$$\frac{\hat{\lambda}_k}{\sum_{k=1}^{n}\hat{\lambda}_k} \quad \xrightarrow{p} \quad v_k \tag{19}$$

where

$$v_k \quad \sim \quad \frac{2c^2}{(-1+c\coth(c))}\left(\frac{1}{c^2+k^2\pi^2}\right) \equiv v_{*k}. \tag{20}$$

Within this centered setting, the quantity in (20) corresponds to the proportion of variance captured by the k^{th} principal component and in line with the uncentered case we note the substantial spuriousness that is caused by the presence of persistent components in $\boldsymbol{X}$. Under $c = 1$ for instance we can note that the first principal component will capture close to 60% of the total variance when in fact the system is composed of independent local to unit-root processes. Under $c = 10$ the first two principal components will (spuriously) capture $v_{*1} + v_{*2} = 36.1\%$, a substantial proportion despite being away from the unit-root boundary. This also highlights the important point and key motivation of this paper that the spuriousness induced by the presence of persistent variables is not solely confined to highly persistent ones in the sense that (20) can remain very large even for magnitudes of c well away from 0.

It is here useful to also highlight the behaviour of (18) and (20) when $c \to 0$. Indeed we have

$$\lim_{c\to 0}\lambda_{*k} \quad = \quad \frac{1}{k^2\pi^2} \tag{21}$$

and

$$\lim_{c\to 0} v_{*k} \quad = \quad \frac{6}{k^2\pi^2}. \tag{22}$$

The quantity in (21) can now be recognised as the set of eigenvalues that drive the Karhunen-Loève representation of a Brownian Bridge which makes intuitive sense in this centered context if we recall the following equality in law of a Brownian Bridge and a demeaned Brownin Motion

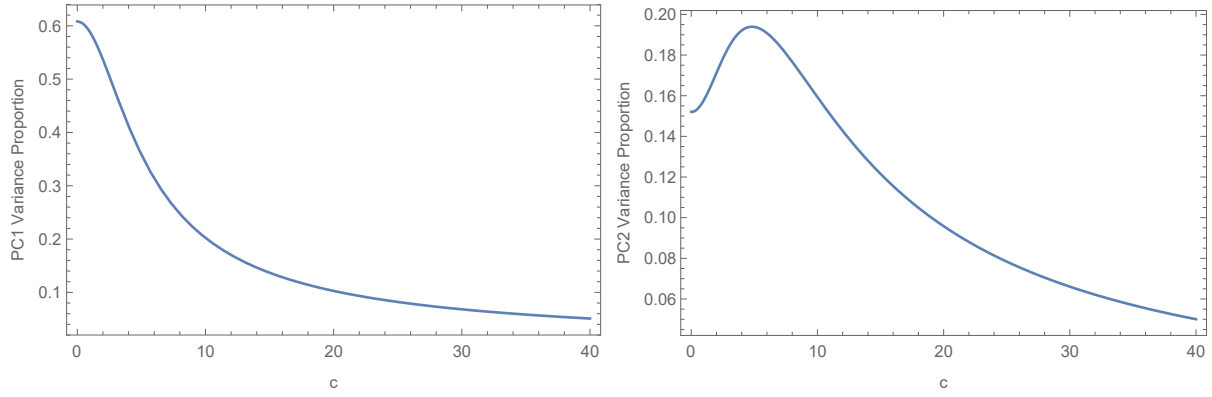
$$\int_0^1 [W(r) - \int_0^1 W(s)ds]^2 dr \quad \overset{law}{\equiv} \quad \int_0^1 [W(r) - rW(1)]^2 dr. \tag{23}$$

The dimensionality/persistence induced spuriousness as illustrated by (20) can clearly be seen to be a decreasing function of the non-centrality parameter c when $k = 1$ and peaks (for $k = 1$) under the pure unit

root scenario $c = 0$. It is also useful to point out that (22) specialises our result to that obtained in Onatski and Wang (2018) who pointed out the striking result that in a pure unit-root setting of independent random walks the first PC spuriously captures more than 60% of the data variance ($6/\pi^2 \approx 60.79\%$).

Here it is also important to mention that researchers from a variety of fields have been interested in the behaviour of PCs in high dimensional random walk settings. Examples include the modelling of the evolution of virus infections in biology (e.g. Moore, Ahmed and Antia (2018)), morphometric analyses of organisms (e.g. Bookstein (2013)) amongst others. In Moore, Ahmed and Antia (2018) for instance the authors also heuristically obtained the result that in a high dimensional pure random walk environment the proportion of variance captured by each PC is given by $6/k^2\pi^2$. Another important contribution in this same area is Antognini and Sohl-Dickstein (2018) who explored similar issues in an environment motivated by neural network training, also establishing the result that in a high dimensional asymptotic framework the proportion of variance captured by the k^{th} PC is given by $6/k^2\pi^2$ when the data consist of random walks. A particularly noteworthy contribution of Onatski and Wang (2018) however is their rigorous linking of the eigenvalues of $\boldsymbol{S}_n$ with their deterministic equivalents via Stieltjes transform methods.

Our result in (20) highlights the fact that this spuriousness could continue to cause important distortions even in non-random walk and more mildly integrated environments. This important point is illustrated in Figure 1 below where we plot the proportion of variance captured by PC1 and PC2 across alternative magnitudes of the near integration parameter c using our expression obtained in (20) for $k = 1$ and $k = 2$. We note that even for reasonably large magnitudes of c the first and second PCs may appear to capture close to 40% of the variance in the data.



(a) PC1 (b) PC2

Figure 1: Proportion of variance captured by PC1 and PC2

Another novel and important observation that follows from our results in Proposition 2 is related to the impact that persistence has on the higher PCks (e.g. $k = 2$ and above). It is indeed interesting to observe that the proportion of variance spuriously captured by higher principal components may be further amplified for magnitudes of c away from $c = 0$ rather than at $c = 0$. This is illustrated by the bump at $c = 4.8$ in the plot of PC2 in Figure 1b and is a phenomenon that affects all PCs beyond $k = 1$.

Under $c = 0$ the pure random walk based result in (21) indicates that the second PC will spuriously capture 15.2% of the total variation whereas if $c = 4.8$ the same PC will spuriously capture 19.4% of the total variation.

It is of course difficult to gauge what these asymptotic settings imply in terms of the specific magnitude of the $\rho_n(c, 1)$'s beyond which we are likely to fall in a spuriousness trap. It is well known that many economic and financial time series that conceptually cannot be modelled as pure unit root processes display first order autocorrelations that often fall within the interval $[0.90, 0.99]$. Familiar examples that have often been included in principal component analyses and diffusion index based forecasting studies include stock market valuation ratios such as price-to-earnings and dividend yields, interest rates, unemployment rates, inflation rates amongst others. Our Figures 1(a)-1(b) suggest that the first two PCs are likely to indicate the presence of spurious factors even for larger magnitudes of c within this local to unit root setting.

To gain more tangible insights on these distortions and to illustrate the validity of our results in Propositions 1 and 2 under finite p and n we implement a series of Monte-Carlo experiments geared towards empirically quantifying the impact of specific magnitudes of $\rho_n(c, \alpha)$ on the spuriousness phenomenon. An important side objective of this exercise is to also assess whether first differencing the data helps remove the spurious outcomes even when $c \neq 0$. In a first instance however our simulations are designed to demonstrate the accuracy of our eigenvalue approximations in (12) and (18) across a broad range of configurations for p, n and c.

Table 1 displays the Monte-Carlo means of the first and second sample eigenvalues for the uncentered and centered covariances and compares them with our theoretical approximations in (12) and (18) respectively. The DGP is given by (3) with $\rho_n(c, \alpha) = 1 - c/n$. As our approximations in (12) and (18) are valid for $c > 0$ we evaluate their accuracy for $c \in \{0.01, 5, 10, 20\}$ across $(n, p) \in \{(150, 150), (500, 100), (100, 500), (50, 50^2)\}$.

The first and third columns of each panel labelled as Uncentered and Centered in Table 1 present the Monte-Carlo means of the first and second eigenvalues while the second and fourth columns present the corresponding theoretical approximations obtained using (12) and (18). We note highly accurate matches between the two pairs of interest across all configurations. Under $(n, p) = (150, 150)$ and $c = 5$ for instance the first suitably normalised sample eigenvalue led to a Monte-Carlo average of 0.032 which can be compared with 0.033 obtained using our approximation in (12). Here it is also interesting to point out the distinct behaviour of these eigenvalues depending on whether the data has been centered or not (e.g. 0.400 versus 0.101 under $c = 0.01$ and $(n, p) = (150, 150)$). More importantly we note that the centering plays an important role when $\rho_n(c, 1)$ gets closer and closer to the unit root boundary while for magnitudes nearer to the stationarity region the spreads become negligible.

We next explore in greater detail the interactions between dimensionality and persistence in order to

Table 1: Monte-Carlo Means of Sample Eigenvalues and their Population Approximations

	Uncentered				Centered			
	$\hat{\lambda}_1^{uc}/n^2$	λ_1^{uc}	$\hat{\lambda}_2^{uc}/n^2$	λ_2^{uc}	$\hat{\lambda}_1/n^2$	λ_1	$\hat{\lambda}_2/n^2$	λ_2
	$(n,p) = (150,150)$				$(n,p) = (150,150)$			
$c = 0.01$	0.400	0.403	0.045	0.045	0.101	0.101	0.025	0.025
$c = 1$	0.195	0.215	0.040	0.032	0.054	0.068	0.024	0.018
$c = 5$	0.032	0.033	0.019	0.019	0.019	0.022	0.012	0.012
$c = 10$	0.010	0.009	0.008	0.008	0.008	0.008	0.006	0.006
$c = 20$	0.003	0.002	0.003	0.002	0.003	0.002	0.002	0.002
	$(n,p) = (500,100)$				$(n,p) = (500,100)$			
$c = 0.01$	0.403	0.403	0.045	0.045	0.101	0.101	0.026	0.025
$c = 1$	0.197	0.215	0.040	0.032	0.055	0.068	0.024	0.018
$c = 5$	0.032	0.033	0.018	0.019	0.019	0.022	0.012	0.012
$c = 10$	0.010	0.009	0.008	0.008	0.008	0.008	0.006	0.006
$c = 20$	0.003	0.002	0.003	0.002	0.003	0.002	0.002	0.002
	$(n,p) = (100,500)$				$(n,p) = (100,500)$			
$c = 0.01$	0.399	0.403	0.045	0.045	0.101	0.101	0.025	0.025
$c = 1$	0.196	0.215	0.040	0.032	0.054	0.068	0.024	0.018
$c = 5$	0.032	0.033	0.019	0.019	0.019	0.022	0.012	0.012
$c = 10$	0.009	0.009	0.008	0.008	0.008	0.008	0.006	0.006
$c = 20$	0.003	0.002	0.002	0.002	0.002	0.002	0.002	0.002
	$(n,p) = (500,100)$				$(n,p) = (500,100)$			
$c = 0.01$	0.394	0.403	0.044	0.045	0.101	0.101	0.025	0.025
$c = 1$	0.196	0.215	0.040	0.032	0.054	0.068	0.024	0.018
$c = 5$	0.032	0.033	0.019	0.019	0.019	0.022	0.013	0.012
$c = 10$	0.009	0.009	0.008	0.008	0.008	0.008	0.006	0.006
$c = 20$	0.003	0.002	0.002	0.002	0.002	0.002	0.002	0.002

empirically gauge how far off the unit root boundary one may need to be for the spuriousness phenomenon not to kick in. For the remainder of this analysis we concentrate solely on the centered covariance based PCA so that $\hat{\lambda}_k/\sum_k \hat{\lambda}_k$ can be interpreted as the proportion of variance captured by the k^{th} principal component. Results for this set of experiments are presented in Table 2 which displays the Monte-Carlo means of $\hat{\lambda}_1/\sum_{k=1}^{n} \hat{\lambda}_k$ across alternative (n,p) configurations and choices of $c \in \{0,5,10,20,40\}$. The columns labelled $\boldsymbol{X}(PC1)$ and $\boldsymbol{X}(PC2)$ correspond to PCA on the raw centered data matrix $\boldsymbol{X}_H$ and the columns labelled $\boldsymbol{\Delta}X(PC1)$ andd $\boldsymbol{\Delta}X(PC2)$ repeat the same exercise on the first differenced data. As the DGP consists of independent components we naturally expect the estimated proportion captured by each PC to remain in the vicinity of zero.

The magnitudes displayed under $\boldsymbol{X}(PC_1)$ and $\boldsymbol{X}(PC_2)$ highlight the severity of the spuriousness phenomenon even for relatively large magnitudes of the non-centrality parameter c. Under $(n,p) = (100,500)$ and $c = 5$ and $c = 10$ for instance we note that the first two PCs spuriously capture 49% and 34% of the data variance respectively. Under $(n,p) = (500,100)$ and $c = 5$ the first two PCs capture more than 50% of the total data variance. The empirical magnitudes of Table 2 can also be compared with their theoretical asymptotic counterpart in (20). Under $c = 0$ and $k = 1$ for instance our estimated variance proportions remained in the close vicinity of 60% across all configurations, closely matching the theoretical value of $6/\pi^2 \approx 60.8$ which can be obtained from (21). Under $c = 10$ and for $k = \{1,2\}$ our estimates of 20.0% and 15.3% can be compared with their theoretical counterparts of 20.2% and 15.9% (as implied from (20)). The outcomes presented in Table 2 also corroborate our earlier point about the higher order PCs whose spuriousness feature may deteriorate when moving from the pure unit root setting to an environment in the vicinity of unity. Note for instance the magnitude of 19.4% associated with $\boldsymbol{X}(PC_2)$ when $c = 5$ which can be compared with the 15.3% outcome in a pure unit root setting.

Table 2: Proportion of variance spuriously captured by PC1 and PC2 under high persistence

$\alpha = 1$	$\boldsymbol{X}(PC_1)$	$\boldsymbol{X}(PC_2)$	$\boldsymbol{\Delta}X(PC_1)$	$\boldsymbol{\Delta}X(PC_2)$	$\boldsymbol{X}(PC_1)$	$\boldsymbol{X}(PC_2)$	$\boldsymbol{\Delta}X(PC_1)$	$\boldsymbol{\Delta}X(PC_2)$
	$(n,p) = (150,150)$				$(n,p) = (500,100)$			
$c = 0$	60.9	15.3	2.6	2.5	61.0	15.4	2.1	2.0
$c = 5$	30.2	19.4	2.8	2.6	30.7	19.6	2.1	2.0
$c = 10$	20.0	15.3	2.6	2.6	20.9	15.6	2.1	2.0
$c = 20$	11.5	9.8	2.7	2.7	12.6	10.5	2.1	2.0
$c = 40$	6.5	5.8	2.7	2.7	7.6	6.7	2.1	2.0
	$(n,p) = (100,500)$				$(n,p) = (50,50^2)$			
$c = 0$	60.8	15.2	2.1	2.0	60.9	15.2	2.7	2.6
$c = 5$	29.6	19.4	2.1	2.0	29.1	19.3	2.7	2.7
$c = 10$	18.9	15.1	2.1	2.0	17.9	14.7	2.8	2.8
$c = 20$	10.0	9.0	2.2	2.1	8.6	8.1	3.1	3.0
$c = 40$	5.0	4.6	2.3	2.1	3.4	3.3	3.9	3.8

Another important message implied by the eigenvalue behaviour of the differenced data is that first differencing eliminates the spuriousness phenomenon even for magnitudes of c that place $\rho_n(c,1)$ far away from the unit-root boundary, essentially corresponding to some mild degree of over-differencing.

From a practical standpoint this latter result suggests that when forecasting with principal component regressions that use a small number of PCA based indices as predictors (e.g. diffusion index forecasting) it may be wise to estimate such indices using first differences of persistent variables regardless of whether these have unit-roots or not. This is expected to alleviate findings of spurious forecastability that may arise from the use of such principal component based predictors whose importance is spuriously inflated. The advantages of over-differencing nearly non-stationary predictors have also been documented in other time-series contexts. Using series modelled as autoregressive processes, Sanchez and Pena (2001) for instance established that overdifferencing nearly-stationary predictors typically leads to forecasts with lower mean squared errors. The VAR based analysis in Marcet (2004) which explores the influence of overdifferencing VARs for impulse response and related analyses also goes in this direction.

Finally, another important practical implication of our analysis is that empirical work involving PCA should aim to eliminate these dimensionality induced biases through preliminary dimensionality reduction stages with techniques such as group based PCA, sparse PCA, thresholding and related approaches (see Zou, Hastie and Tibshirani (2006), Ma (2013) and references therein).

In light of our result in (20) it is also important to gauge more formally the relationship between the *degree* of persistence of the components of $\boldsymbol{X}$ and the spuriousness they induce. Although the above analysis explored alternative magnitudes of the non-centrality parameter c in order to highlight the impact of alternative degrees of persistence, the components remained within the same local to unit root class with $\alpha = 1$. Here we adopt the parameterisation $\rho_n(c, \alpha) = 1 - c/n^\alpha$ but require that $\alpha \in [0, 1)$ so that the unit root boundary is now allowed to be approached more slowly, depending on the magnitude of α. Smaller values of α are associated with larger neighbourhoods away from unity and such series are typically labelled as mildly integrated or near stationary (see Magdalinos and Phillips (2007)). Note that, as we wish to exclude explosive or negative unit root behaviour, this parameterisation implicitly restricts the relationship between c and α so as to avoid such occurrences (e.g. $0 < c/n^\alpha < 2$). The following Proposition establishes the limiting behaviour of the $\hat{\lambda}_k's$ and that of the trace of $\boldsymbol{S}_n$ in this mildly integrated setting.

PROPOSITION 3: *Under Assumption A, for given k and $\alpha \in [0, 1)$ we have*

$$\frac{\hat{\lambda}_k}{n^{1+\alpha}} \xrightarrow{p} 0 \tag{24}$$

and

$$\left| \frac{Tr(\boldsymbol{S}_n)}{n^{1+\alpha}} - \frac{1}{2c\left(1 - \frac{c}{2n^\alpha}\right)} \right| \xrightarrow{p} 0. \tag{25}$$

It is here important to emphasise that the limiting results in (24)-(25) exclude the local to unity scenario associated with $\alpha = 1$. It is now particularly interesting to point out that the outcomes in (24)-(25) smoothly bridge the purely stationary case of $\alpha = 0$ and the mildly integrated cases associated with $\alpha \in (0, 1)$. For $\alpha = 0$ we can note for instance that the limit of $1/(2c - c^2)$ is the familiar expression of the variance of a stationary AR(1) process since $\rho = 1 - c$. Note that this latter observation should not be interpreted as implying that each individual (non-normalised) eigenvalue $\hat{\lambda}_k$ converges to $1/(1 - \rho^2)$ under $\alpha = 0$. It is well known for instance that in high dimensional settings the maximum eigenvalues of covariance matrices are subject to important asymptotic biases that depend on how p and n interact. Our result in (24) refers to suitably normalised versions of these eigenvalues ensuring that they vanish asymptotically while the same normalisation leads to bounded traces.

A useful implication of the results in Proposition 3 is that within this mildly integrated setting we will have

$$\frac{\hat{\lambda}_k}{\sum_{k=1}^{n} \hat{\lambda}_k} \xrightarrow{p} 0 \tag{26}$$

so that the dimensionality/persistence induced spuriousness problem vanishes asymptotically, as expected within this factorless setting. Our analysis of the mildly integrated scenario is particularly useful for highlighting the need to be quite far off the unit root boundary in order to avoid the spuriousness phenomenon. If we take $n = 500$, $c = 5$ and $\alpha = 0.5$ in $\rho_n(c, \alpha) = 1 - c/n^{\alpha}$ for instance we can compare the autocorrelation coefficient of $(1 - 5/\sqrt{500}) = 0.776$ with the magnitude of 0.99 that would arise under a standard local to unity setting having $\alpha = 1$, $(1 - 5/500) = 0.990$.

We next consider a set of simulation experiments to illustrate the finite sample behaviour of (26) across alternative choices of $\alpha \in [0, 1)$. In order to highlight the behaviour of eigenvalues in an increasing sample size/dimension setting we proceed as follows. For a given (n, p) combination we vary $\alpha \in \{0.00, 0.25, 0.50, 0.75\}$ and fine-tune the magnitude of c for each α in a way that guarantees the same $\rho_n(c, \alpha)$ parameter across increasing (n, p) magnitudes. The first five columns of Table 3 detail these alternative parameterisations while the remainder columns present the Monte-Carlo means of $\hat{\lambda}_k / \sum \hat{\lambda}_k$ for $k = 1, 2$ (i.e. the first and second PCs variance capture proportion) using both the level and first differenced data.

The outcomes presented under $\boldsymbol{X}(PC1)$ and $\boldsymbol{X}(PC2)$ corroborate our theoretical results in (24)-(25) as we can observe a clear decline in the upward biases of eigenvalues as α declines towards zero and/or (n, p) increase for given α. Under $\alpha = 0.50$ and $(n, p) = (100, 50)$ for instance we note that the first two PCs spuriously capture a sizeable 37.9% of the variation which subsequently drops to 15.5% under $(n, p) = (500, 50)$ and to just 10.0% under $(n, p) = (500, 250)$. For given (n, p) we also note the substantial decrease in these spurious proportions as α approaches zero. The last two columns of Table 3 illustrate the analogous outcomes when PCA is implemented on first differenced data with outcomes that are in line with our analysis of Table 2. A key message that comes across the outcomes of Table 3 is that the

Table 3: Proportion of variance spuriously captured by PC1 and PC2 under mild persistence

n	p	α	c	$\rho_n(c,\alpha)$	$\boldsymbol{X}(PC1)$	$\boldsymbol{X}(PC2)$	$\boldsymbol{\Delta}X(PC1)$	$\boldsymbol{\Delta}X(PC2)$
100	50	0.00	1.00	0.000	5.60	5.10	6.80	6.10
100	50	0.25	1.00	0.684	10.10	8.60	5.80	5.30
100	50	0.50	1.00	0.900	22.10	15.80	5.60	5.20
100	50	0.75	1.00	0.968	37.00	20.80	5.60	5.20
500	50	0.00	1.00	0.000	3.40	3.20	3.80	3.60
500	50	0.25	1.50	0.684	4.70	4.40	3.50	3.30
500	50	0.50	2.24	0.900	8.30	7.20	3.40	3.30
500	50	0.75	3.34	0.968	16.90	13.10	3.40	3.20
500	250	0.00	1.00	0.000	1.10	1.10	1.40	1.40
500	250	0.25	1.50	0.684	2.20	2.10	1.20	1.10
500	250	0.50	2.24	0.900	5.20	4.80	1.20	1.10
500	250	0.75	3.34	0.968	13.80	11.60	1.20	1.10

dimensionality/persistence induced spuriousness problem is not necessarily confined to highly persistent, pure unit-root or local to unit-root settings. Within factorless settings, having components with first order autocorrelation coefficients in the vicinity of 0.9 or even below may continue to give the impression of meaningful common factors when there are truly none.

3.2. A High dimensional environment with mixed degrees of persistence and cointegration

An important new question that arises from the above analysis is whether the system composition in terms of the relative weights of persistent versus less persistent components in $\boldsymbol{X}$ affects the spuriousness phenomenon and if it does towards which direction. This issue is particularly relevant for empirical applications in economics where PCA is typically implemented on variables with potentially different persistence characteristics and cointegration properties (e.g. the monthly database for Macroeconomic Research known as FRED-MD commonly used in PCA applications and diffusion index construction (see McCracken and Ng (2015)). An important question that arises in this context is whether the spuriousness problem is alleviated if the system dimension is dominated by less persistent variables. Another related issue from a practicioner's point of view is whether first differencing such a mixed data set creates an additional layer of distortions akin to the well known problems that arise when first differencing a cointegrated VAR.

To address these questions we consider a mixed system that consists of variables whose dynamics continue to be captured by (3) but with different parameterisations that blend pure unit-root, local to unit-root (LUR) and stationary components in varying proportions relative to the full system dimension p. It is here important to point out that such a setting can also specialise into a cointegrated system that consists for instance of r cointegrating relationships and $(p-r)$ common trends (e.g. random walk components). Such a system has been investigated in Onatski and Wang (2018) where the authors showed

that the spuriousness phenomenon and limiting result in continues to be present provided that $r/p \to 0$. This can intuitively be explained by the fact that requiring $r/p \to 0$ essentially forces the $I(1)$ components of the system (i.e. $(p-r)/p$) to dominate. As it would typically be impractical to evaluate the joint cointegration properties of hundreds of variables it is also important to assess whether first differencing the entire data matrix $\boldsymbol{X}$ containing series of varying degrees of persistence in addition to pure random walks results in any distortions akin to the well known mispecifications that arise in cointegrated VAR models formulated in first differences.

To address these issues we now assume that the p dimensional vector $\boldsymbol{X}_t$ consists of p_1 and p_2 components, say $\boldsymbol{X}_{1t}$ and $\boldsymbol{X}_{2t}$ modelled as

$$\boldsymbol{X}_{1t} = \boldsymbol{\Phi}_{1n}\boldsymbol{X}_{1t-1} + \boldsymbol{u}_{1t} \tag{27}$$

$$\boldsymbol{X}_{2t} = \boldsymbol{\Phi}_{2n}\boldsymbol{X}_{2t-1} + \boldsymbol{u}_{2t}, \tag{28}$$

with $\boldsymbol{\Phi}_{1n} = \rho_n(c,\alpha)\boldsymbol{I}_{p_1}$ and $\boldsymbol{\Phi}_{2n} = \boldsymbol{I}_{p_2}$. The $n \times p$ data matrix $\boldsymbol{X}$ is now understood to concatenate the $n \times p_1$ and $n \times p_2$ sub-data matrices $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ stacking the components of (27) and (28) in its columns. We also write $\boldsymbol{u}_t = (\boldsymbol{u}_{1t}, \boldsymbol{u}_{2t})'$ so that the $n \times p$ matrix of random disturbance terms $\boldsymbol{U}$ remains as in our earlier analysis.

The system in (27)-(28) can now be conveniently reformulated in first differenced matrix form as

$$\boldsymbol{R}_n(0,0)\boldsymbol{X} = \boldsymbol{D}_n(c,\alpha)\boldsymbol{U}\boldsymbol{L}_p + \boldsymbol{U}\boldsymbol{M}_p \tag{29}$$

where

$$\boldsymbol{D}_n(c,\alpha) = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 & 0 & 0 \\ \delta & 1 & 0 & \dots & 0 & 0 & 0 \\ \delta(1+\delta) & \delta & 1 & \dots & 0 & 0 & 0 \\ \delta(1+\delta)^2 & \delta(1+\delta) & \delta & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & & \\ \delta(1+\delta)^{n-2} & \delta(1+\delta)^{n-3} & \delta(1+\delta)^{n-4} & \vdots & 0 & \delta & 1 \end{bmatrix} \tag{30}$$

$$\boldsymbol{M}_p = \begin{bmatrix} \boldsymbol{0}_{p_1 \times p_1} & \boldsymbol{0}_{p_1 \times p-p_1} \\ \boldsymbol{0}_{p-p_1 \times p_1} & \boldsymbol{I}_{p-p_1 \times p-p_1} \end{bmatrix}, \tag{31}$$

$$\boldsymbol{L}_p = \begin{bmatrix} \boldsymbol{I}_{p_1 \times p_1} & \boldsymbol{0}_{p_1 \times p-p_1} \\ \boldsymbol{0}_{p-p_1 \times p_1} & \boldsymbol{0}_{p-p_1 \times p-p_1} \end{bmatrix} \tag{32}$$

with $\delta = -c/n^{\alpha}$. Note that $\boldsymbol{R}_n(0,0)$ in (29) is a differencing matrix corresponding to (6) with $c = 0$ and $\alpha = 0$. To arrive at the first component in the right hand side of (29) for instance it suffices to combine $\Delta \boldsymbol{X}_{1t} = \delta \boldsymbol{I}_{p_1} \boldsymbol{X}_{1t-1} + \boldsymbol{u}_{1t}$ with the recursion $\boldsymbol{X}_{1t} = \sum_{j=0}^{t} (1+\delta)^{t-j} \boldsymbol{u}_{1j}$ for $t = 1, \ldots, n$. The lower triangular matrix $\boldsymbol{D}_n(c,\alpha)$ in (30) captures the dynamics of the nearly integrated, near stationary or purely stationary components of (27) depending on the chosen magnitudes of α and c. Under $\alpha = 0$ for instance we have $\rho_n(c,0) = 1 - c$ which is understood to satisfy $|1-c| < 1$ and corresponds to a scenario where the first p_1 variables of the data matrix $\boldsymbol{X}$ are purely stationary while the remaining $p - p_1 \equiv p_2$ are random walks. This scenario is analogous to a setting with $r = p_1$ cointegrating relationships and $p - p_1 \equiv p_2$ common trends. Under $\alpha = 1$ the first p_1 components are local to unit-root processes and under $\alpha \in (0,1)$ these p_1 components are mildly integrated.

Using (29) we can now formulate the *centered* data matrix that concatenates $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ as

$$\boldsymbol{X}_H = \boldsymbol{H}_n \boldsymbol{R}_n(0,0)^{-1} \boldsymbol{D}_n(c,\alpha) \boldsymbol{U} \boldsymbol{L}_p + \boldsymbol{H}_n \boldsymbol{R}_n(0,0)^{-1} \boldsymbol{U} \boldsymbol{M}_p \tag{33}$$

which can be viewed as the matrix moving average representation of (27)-(28). Noting the orthogonality of $\boldsymbol{L}_p$ and $\boldsymbol{M}_p$ and their idempotent nature it is now straightforward to express the outer-product of interest as

$$\begin{aligned} \boldsymbol{X}_H \boldsymbol{X}_H' &= \boldsymbol{H}_n \boldsymbol{R}_n(0,0)^{-1} \boldsymbol{D}_n(c,\alpha) \boldsymbol{U} \boldsymbol{L}_p \boldsymbol{U}' \boldsymbol{D}_n(c,\alpha)' (\boldsymbol{R}_n(0,0)^{-1})' \boldsymbol{H}_n \\ &+ \boldsymbol{H}_n \boldsymbol{R}_n(0,0)^{-1} \boldsymbol{U} \boldsymbol{M}_p \boldsymbol{U}' (\boldsymbol{R}_n(0,0)^{-1})' \boldsymbol{H}_n. \end{aligned} \tag{34}$$

and write the (centered) covariance matrix as

$$\begin{aligned} \boldsymbol{S}_n &= \frac{1}{p} \boldsymbol{H}_n \boldsymbol{R}_n(0,0)^{-1} \boldsymbol{D}_n(c,\alpha) \boldsymbol{U} \boldsymbol{L}_p \boldsymbol{U}' \boldsymbol{D}_n(c,\alpha)' (\boldsymbol{R}_n(0,0)^{-1})' \boldsymbol{H}_n \\ &+ \frac{1}{p} \boldsymbol{H}_n \boldsymbol{R}_n(0,0)^{-1} \boldsymbol{U} \boldsymbol{M}_p \boldsymbol{U}' (\boldsymbol{R}_n(0,0)^{-1})' \boldsymbol{H}_n. \end{aligned} \tag{35}$$

It is here straightforward to note that the first component in the right hand side of (35) captures the covariation of the p_1 persistent but also possibly stationary processes associated with (27) depending on how $\rho_n(c,\alpha)$ has been specialised while the second component is associated with the $(p - p_1)$ $I(1)$ components in (28). From our Assumption A and standard algebra it in fact follows that $|\boldsymbol{U}\boldsymbol{L}_p\boldsymbol{U}'/p - p_1/p| \xrightarrow{p} 0$ and $|\boldsymbol{U}\boldsymbol{M}_p\boldsymbol{U}'/p - (p-p_1)/p| \xrightarrow{p} 0$.

We initially evaluate the large sample properties of the trace of $\boldsymbol{S}_n$ under alternative scenarios on the interactions between p_1 and $p_2 \equiv (p - p_1)$ and persistence parameterisations. This will then allow us to establish the large sample behaviour of the proportion of variances captured by the principal components of $\boldsymbol{X}$ across the same scenarios. The types of interactions between p_1 and p_2 that we will operate under are summarised under Assumption B below.

ASSUMPTION B: $p_1/p \to \theta \in [0,1]$.

We note that the boundaries of θ capture pure I(1) environments as in (28) when $\theta = 0$ and purely persistent and/or purely stationary environments as specified in (27) when $\theta = 1$. Intermediate magnitudes of θ accommodate mixed systems blending components that behave as in (27) and (28) in proportions determined by θ. Recalling that the p_1 components are parameterised with $\rho_n(c,\alpha) = 1 - c/n^\alpha$, in what follows we will be solely interested in two alternative scenarios forcing these p_1 components to be either LUR processes (setting $\rho_n(c,1) = 1 - c/n$) in (27) or purely stationary processes (setting $\rho_n(c,0) = 1 - c$ with $|1-c| < 1$) in (27)).

Poposition 4 below treats the boundary cases of $\theta \in \{0,1\}$ while the intermediate scenarios are inferred directly from these further below. Note that under $p_1/p \to 0$ the system essentially evolves towards a pure I(1) setting, capturing an environment with less and less "stationarity" or equivalently an environment where the growth in the number of cointegrating relationships is progressively dominated by pure I(1)'ness. This would for instance be akin to operating under $r/p \to 0$ in a system such as (1) with r cointegrating relationships. Under $p_1/p \to 1$ on the other hand the system evolves towards a setting with no pure I(1) components. The results associated with part (iii) of Proposition 4 are of course known but are maintained for comparison purposes.

PROPOSITION 4: *(i) Under Assumptions A and B and $\theta = 0$ (pure I(1) system) we have*

$$\frac{1}{n^2} Tr(\boldsymbol{S}_n) \quad \xrightarrow{p} \quad \frac{1}{6}. \tag{36}$$

(ii) Under Assumptions A and B, $\theta = 1$ and $\rho_n = 1 - \dfrac{c}{n}$ (pure LUR system) we have

$$\frac{1}{n^2} Tr(\boldsymbol{S}_n) \quad \xrightarrow{p} \quad \frac{6 + 2c^2 + 2e^{-2c} - 8e^{-c} + c(-5 + e^{-2c})}{4c^3}. \tag{37}$$

(iii) Under Assumptions A and B, $\theta = 1$ and $\rho_n = 1 - c$ with $|1-c| < 1$ (pure I(0) system) we have

$$\frac{1}{n} Tr(\boldsymbol{S}_n) \quad \xrightarrow{p} \quad \frac{1}{2c - c^2}. \tag{38}$$

We note that under all of the above scenarios the suitably normalised trace of $\boldsymbol{S}_n$ converges in probability to a constant magnitude. Recalling that under pure stationarity each one of the *independent* components of $\boldsymbol{X}$ follows the same AR(1) process with slope $\rho_n = 1 - c$ the result in (38) is intuitively clear as it implies that the total system variance converges to $1/(1-\rho_n^2) \equiv 1/(2c - c^2)$. The result in (37)

is obtained for a purely LUR system where each component of $\boldsymbol{X}$ consists of nearly integrated AR(1)'s with slope coefficients $\rho_n = (1 - c/n)$.

Before proceeding further it is instructive to empirically assess the adequacy of the above approximations and evaluate the progression of $Tr(\boldsymbol{S}_n)/n^2$ towards its limits in (36)-(38) for θ approaching its $\{0, 1\}$ boundaries. For this purpose we simulate data from the DGP in (27)-(28) using $\theta \in (0.50, 0.25, 0.10, 0.01)$ so as to capture the notion of a θ approaching 0 from above (i.e. progressively decreasing the proportion of non I(1) components in the system). We note from (36) that the normalised trace approaches 1/6 as $p_1/p \to \theta = 0$ (i.e. as the system evolves towards a full I(1) environment) and this result holds regardless of the composition of the p_1 components which we parameterised as pure I(0) processes with autocorrelation coefficients $\rho_n = 0.5$. Under $\theta = 0.1$ for instance the system consists of $p_1 = 150$ I(0) components and $p_2 = 1350$ I(1) components while under $\theta = 0.5$ the system consists of an equal share of I(0) and I(1) components (i.e. $p_1 = p_2 = 150$). From the top panel of Table 4 we can clearly note the Monte-Carlo averages evolving towards 0.167 as $\theta \to 0$. It is here also particularly interesting to note that the entries in the top panel of Table 4 correspond to $(1 - \theta)(1/6)$ i.e. the limiting magnitude of the trace in a pure I(1) setting (see (36)) multiplied by the fraction of I(1) components $(1 - \theta)$.

Our second scenario focuses on the LUR case in (37) and uses $\theta \in (0.50, 0.75, 0.90, 0.99)$ to capture a system whose pure I(1) components are progressively eliminated (see middle panel of Table 4) with the system evolving towards a pure LUR environment (see middle panel of Table 4). The quantities in the rightmost column have been obtained from (37) and we can clearly note that the Monte-Carlo means of $Tr(\boldsymbol{S}_n)/n^2$ evolve towards their expected limit as $\theta \to 1$.

Table 4: Monte-Carlo averages of $Tr(\boldsymbol{S}_n)/n^2$
$(n, p_1, p_2) = (150, 150, \frac{1-\theta}{\theta} p_1)$

	$p_1 \sim$ I(0), $p_2 \sim$ I(1) and $p_1/p \to \theta$				
θ	0.500	0.250	0.100	0.010	Theory
$\rho_n = 0.5$	0.088	0.127	0.151	0.165	0.167
	$p_1 \sim$ LUR, $p_2 \sim$ I(1) and $p_1/p \to \theta$				
θ	0.500	0.750	0.900	0.990	Theory
$\rho_n(c = 1)$	0.142	0.129	0.122	0.118	0.116
$\rho_n(c = 5)$	0.115	0.089	0.087	0.065	0.062
$\rho_n(c = 10)$	0.104	0.072	0.054	0.042	0.039
	$p_1 \sim$ I(0), $p_2 \sim$ I(1) and $p_1/p \to \theta$				
θ	0.500	0.750	0.900	0.990	Theory
$\rho_n = 0.50$	13.152	7.234	3.739	1.607	1.333
$\rho_n = 0.80$	13.788	8.190	4.884	2.865	2.778
$\rho_n = 0.95$	16.215	11.839	9.257	7.670	10.256

The third and last scenario corresponds to a system that evolves into a pure I(0) system with $\rho_n \equiv 1 - c$ (see bottom panel of Table 4 with $\rho_n \in \{0.50, 0.80, 0.95\}$). Here it is particularly interesting to notice that for magnitudes of ρ_n near unity, (37) provides a subtantially better description of the trace's behaviour

than (38).

The above analysis focused on scenarios where the system composition evolves either towards purely stationary/purely LUR processes ($\theta = 1$) or towards purely I(1) processes ($\theta = 0$). It is now also interesting to explore the behaviour of $Tr(\boldsymbol{S}_n)$ when $\theta \in (0, 1)$ instead. Using Proposition 4 it is straightforward to establish that for $\theta \in (0, 1)$ and $\rho_n = 1 - c/n$ we have

$$\frac{1}{n^2} Tr(\boldsymbol{S}_n) \quad \xrightarrow{p} \quad \theta \left(\frac{6 + 2c^2 + 2e^{-2c} - 8e^{-c} + c(-5 + e^{-2c})}{4c^3} \right) + (1 - \theta)\frac{1}{6} \tag{39}$$

and for $\rho_n = 1 - c$ with $|1 - c| < 1$ we have

$$\frac{1}{n^2} Tr(\boldsymbol{S}_n) \quad \xrightarrow{p} \quad (1 - \theta)\frac{1}{6}. \tag{40}$$

Note that (39) corresponds to an environment where a θ proportion of the system consists of LUR processes and the remainder $(1 - \theta)$ are pure I(1)'s. Similarly (40) corresponds to an environment where a θ proportion of the system consists of purely stationary processes with $\rho_n = 1 - c$ and its remainder are again pure I(1)'s. Under this latter environment it is interesting to point out the dominance of I(1)'ness even for large magnitudes of θ. Indeed, we note that the trace of $\boldsymbol{S}_n$ behaves as in a pure I(1) setting albeit with a scaling factor $(1 - \theta)$ representing the fraction of I(1) components in the system.

These theoretical outcomes are illustrated in Table 5 below whose purpose is to empirically evaluate the approximations in (39) and (40). Note that the first component in the right hand side of (39) converges to 1/6 as $c \to 0$ so that for small magnitudes of c we expect $Tr(\boldsymbol{S}_n)/n^2 \approx 1/6$ as in (36). This clearly makes intuitive sense as for c small the system which consists of p_1 LURs and p_2 pure $I(1)'s$ will essentially behave like a pure $I(1)$ system (as $\theta \to 1$ in particular).

The top panel of Table 5 considers a mixed LUR/I(1) system while its bottom panel considers a mixed I(0)/I(1) system. As expected from the above discussion we note that for small magnitudes of θ (i.e. dominance of I(1)'ness) empirical means almost identically match their theoretical counterparts as formulated in (39) and (40). Focusing on the bottom panel of Table 5 we note a deterioration in the accuracy of the empirical averages for $\theta \approx 0.90$ and above. This corresponds to a scenario where more than 90% of components are purely stationary processes with autocorrelation coefficients of 0.5. Nevertheless these results make it clear that even a small proportion of persistent components in a system will have the ability to generate substantial spuriousness in PCA.

Our next objective is to evaluate the proportion of variance captured by the first few principal components (e.g. $\hat{\lambda}_k/Tr(\boldsymbol{S}_n)$) in such mixed environments. The cases associated with $\theta = 0$ (pure I(1)) and $\theta = 1$ (pure I(0)) can be easily inferred from our results in Propositions 2 and 4 with (22) and (26) characterising these two scenarios's PC behaviour. The main question we wish to explore here is the

Table 5: Monte-Carlo averages of $Tr(\boldsymbol{S}_n)/n^2$ in mixed environments $(n, p_1, p_2) = (150, 150, \frac{1-\theta}{\theta} p_1)$

θ	0.100	0.250	0.750	0.900	0.950
	$p_1 \sim$ LUR and $p_2 \sim$ I(1)				
$c = 1$	0.162	0.154	0.154	0.120	0.119
Theory	0.162	0.154	0.154	0.121	0.118
$c = 5$	0.156	0.141	0.141	0.073	0.068
Theory	0.156	0.140	0.140	0.072	0.067
$c = 10$	0.154	0.136	0.136	0.052	0.047
Theory	0.154	0.135	0.135	0.052	0.045
	$p_1 \sim$ I(0) and $p_2 \sim$ I(1)				
$\rho_n = 0.5$	0.151	0.128	0.048	0.024	0.017
Theory	0.150	0.125	0.042	0.017	0.008

influence of the relative proportion of I(0)/I(1) components on the proportion of variance captured by the first principal component and whether the spuriousness that results from the presence of I(1) components can be alleviated through an increase in I(0) components. For this purpose we consider an environment with p_1 I(0) components having $\rho_n = 1 - c$ with $|1 - c| < 1$ and p_2 I(1) components. Here it continues to be the case that

$$\frac{\hat{\lambda}_k}{\sum_k \hat{\lambda}_k} \quad \xrightarrow{p} \quad \frac{6}{k^2\pi^2} \tag{41}$$

which holds regardless of the proportion of I(0) components in the system and which reflects the fact that I(1)'ness dominates regardless of the proportion of I(0) components. The intuition behind this result can be gained from (35) which when normalised by n^2 sees its first component vanish asymptotically while its second component driven by the presence of I(1) variables remains bounded.

Table 6: Proportion of variance spuriously captured by PC1 and PC2 in Mixed Systems

θ	0.100	0.250	0.500	0.800	0.950	0.990
	$(n, p) = (150, 200)$, $p_1 = \theta p$, $p_2 = (1-\theta)p_2$					
PC1	60.7	60.0	58.0	50.8	33.2	14.6
PC2	15.1	15.1	14.8	13.3	9.1	4.6
	$(n, p) = (500, 500)$, $p_1 = \theta p$, $p_2 = (1-\theta)p_2$					
PC1	60.7	60.7	60.1	57.8	49.0	29.6
PC2	15.3	15.2	15.2	14.5	12.9	6.9

The empirical outcomes in Table 6 are again striking when it comes to the presence of the spuriousness phenomenon. Even when only 1% of the system consists of I(1) variables, the proportion of variance captured spuriously by PC1 and PC2 remains substantial and effectively approaches $6/k^2\pi^2$ asymptotically. Under $(n, p) = (500, 500)$ for instance and $p_1 = 0.99\ p$ and $p_2 = (1 - 0.99)\ p$ (i.e. the system of 500 variables consists of 495 I(0) series with autocorrelation coefficients of 0.5 and only 5 series that are I(1))

we note that the first two PCs spuriously capture close to 40% of the data variation. These outcomes have important implications as they suggest that the presence of only a small fraction of suspected persistent series should not give a false sense of protection against false discoveries.

In the context of this trivially cointegrated system with p_1 stationary components and p_2 I(1) components we have also reconsidered our analysis of Table 6 using fully first differenced data, including (unnecessarily) first differencing the p_1 stationary components. This resulted in PC1 and PC2 proportions lying in the vicinity of 1-2% across all of the above configurations, suggesting that first differencing regardless of the stationarity properties of the underlying series may be an important pre-processing step when implementing such regularisation techniques.

4. Discussion and Conclusions

This paper aimed to highlight some important distortions that are caused by the joint impact of persistence and dimensionality. These distortions take the form of strong spurious relationships between time series that have no connections whatsoever and continue to be present even when these time series are only mildly persistent. Although we have operated within toy models our findings have broad implications for empirical work, in particular when considering Big Data methods relying on regularisation techniques for dimensionality reduction. Such environments are particularly prone to false discoveries when implemented with time series data characterised by even mild degrees of persistence.

Another particularly important finding is that these distortions and tendency to uncover spurious relationships remain very much present even if only a very small fraction of the data contains persistent components. In order to avoid this spuriousness trap our analysis suggests that first differencing the data may be appropriate even for time series that although persistent are away from the unit root boundary. Similarly, even in systems whose majority of components are purely I(0) series and a minor fraction are highly persistent it is important to first difference the entire data matrix to avoid this spuriousness phenomenon. From an empirical strategy point of view if PCA based on first differenced data suggests a factorless outcome then clearly dimensionality reduction is not worth pursuing. If on the other hand the presence of commonalities is supported by PCA on first differenced data and we wish to estimate the common factors numerous options are available, in particular following the PANIC approach of Bai and Ng (2004) (see also Corona, Poncela and Ruiz (2020) and references therein). Our results on the spuriousness of principal components have also a direct analogy with canonical correlations which also rely on the eigenvalues of quantities that include $\boldsymbol{S}_n$ (see for instance Harris (1997) and Gonzalo (1994)).

Granger's early insights on the need to alter existing methods when considering big data environments have become particularly topical. Onatski and Wang (2018)'s recent contributions very much go in that direction within the specific context of cointegration analysis. We expect that methods designed to capture

sparsity via linear combinations and canonical correlations of only a subset of variables will be particularly fruitful for handling a large number of time series data (e.g. sparse principal components, sparse canonical correlations).

It is also important to point out that well known weaknesses of many techniques pointed out in the classical statistics literature may manifest themselves differently in high dimensional settings. A notable example from the model selection literature is the inconsistency of various fixed penalty based information criteria when used to detect the "true" model from a finite selection of competing models (e.g. AIC). In Gonzalo and Pitarakis (2002) for instance we showed that this well known inconsistency vanishes when the system dimension is allowed to increase with the sample size at a suitable rate. Given the usefulness of model selection criteria as tools for dimensionality reduction (e.g. detection of reduced rank structures as explored in Gonzalo and Pitarakis (1998)) we view this agenda as highly promising for tackling the spuriousness problem as recently investigated in Bai, Choi and Fujikoshi (2018), Fujikoshi, Sakurai and Yanagihara (2017) and Fujiskoshi (2017, 2019).

APPENDIX A

This appendix introduces a series of intermediate results which are used for establishing our statements in Propositions 1-4. For notational simplicity we drop the referencing to c and α in $\rho_n(c,\alpha)$ and the associated matrices.

Assumption A ensures that the limiting behaviour of the sample eigenvalues of $\boldsymbol{S}_n^{uc}$ and $\boldsymbol{S}_n$ will be determined by the eigenvalues of the deterministic equivalents $(\boldsymbol{R}_n'\boldsymbol{R}_n)^{-1}$ and $\boldsymbol{H}_n(\boldsymbol{R}_n'\boldsymbol{R}_n)^{-1}\boldsymbol{H}_n$ respectively. Lemma 1 below obtains the exact trace of these two matrices for any n and any magnitude of ρ while Lemma 2 and Lemma 3 obtain their eigenvalues.

From (6) the $n \times n$ dimensional matrix $\boldsymbol{R}_n'\boldsymbol{R}_n$ is given by

$$\boldsymbol{R}_n'\boldsymbol{R}_n = \begin{pmatrix} 1+\rho^2 & -\rho & 0 & \dots & 0 & 0 \\ -\rho & 1+\rho^2 & -\rho & \dots & 0 & 0 \\ 0 & -\rho & 1+\rho^2 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1+\rho^2 & -\rho \\ 0 & 0 & 0 & \dots & -\rho & 1 \end{pmatrix} \tag{42}$$

which is of nearly Toeplitz form with a perturbed bottom right corner. The key quantities of interest are now the eigenvalues of $(\boldsymbol{R}_n'\boldsymbol{R}_n)^{-1}$ and the $(n-1)$ non-zero eigenvalues of $\boldsymbol{H}_n(\boldsymbol{R}_n'\boldsymbol{R}_n)^{-1}\boldsymbol{H}_n$ as $Rank[\boldsymbol{H}_n(\boldsymbol{R}_n'\boldsymbol{R}_n)^{-1}\boldsymbol{H}_n] = n-1$ by construction.

The eigenvalues of $(\boldsymbol{R}_n'\boldsymbol{R}_n)^{-1}$ can be inferred directly from those of (42) for any n while the treatment of $\boldsymbol{H}_n(\boldsymbol{R}_n'\boldsymbol{R}_n)^{-1}\boldsymbol{H}_n$ will require us to operate under large n. More specifically, as $\boldsymbol{H}_n(\boldsymbol{R}_n'\boldsymbol{R}_n)^{-1}\boldsymbol{H}_n$ has one zero eigenvalue we consider the eigenvalues of its pseudo-inverse, say $(\boldsymbol{H}_n(\boldsymbol{R}_n'\boldsymbol{R}_n)^{-1}\boldsymbol{H}_n)^-$, and recover the eigenvalues of interest via their reciprocal. It is here particularly interesting to point out some subtle differences between $(\boldsymbol{R}_n'\boldsymbol{R}_n)^{-1}$ and $\boldsymbol{H}_n(\boldsymbol{R}_n'\boldsymbol{R}_n)^{-1}\boldsymbol{H}_n$. Under $|\rho|< 1$ it is well known that these two matrices will share the same eigenvalues for large n. As ρ approaches unity however this correspondence breaks down. Although an analytical expression for $(\boldsymbol{H}_n(\boldsymbol{R}_n'\boldsymbol{R}_n)^{-1}\boldsymbol{H}_n)^-$ becomes quickly intractable if we wish to operate under any n, using standard properties of pseudo-inverses and the fact that $(\boldsymbol{H}_n(\boldsymbol{R}_n'\boldsymbol{R}_n)^{-1}\boldsymbol{H}_n)(\boldsymbol{H}_n(\boldsymbol{R}_n'\boldsymbol{R}_n)^{-1}\boldsymbol{H}_n)^- = \boldsymbol{H}_n$ in particular, standard but involved algebra allows us

to establish that for sufficiently large n $(\boldsymbol{H}_n(\boldsymbol{R}_n'\boldsymbol{R}_n)^{-1}\boldsymbol{H}_n)^-$ takes the following tridiagonal form

$$(\boldsymbol{H}_n(\boldsymbol{R}_n'\boldsymbol{R}_n)^{-1}\boldsymbol{H}_n)^- \sim \begin{pmatrix} 1 & -\rho & 0 & \dots & 0 & 0 \\ -\rho & 1+\rho^2 & -\rho & \dots & 0 & 0 \\ 0 & -\rho & 1+\rho^2 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1+\rho^2 & -\rho \\ 0 & 0 & 0 & \dots & -\rho & 1 \end{pmatrix}. \tag{43}$$

The right hand side of (43) provides the exact pseudo inverse when $\rho = 1$ while for $\rho \neq 1$ (43) holds for large n and ρ in the vicinity of unity.

Lemma 1 below obtains the exact traces of $(\boldsymbol{R}_n'\boldsymbol{R}_n)^{-1}$ and $\boldsymbol{H}_n(\boldsymbol{R}_n'\boldsymbol{R}_n)^{-1}\boldsymbol{H}_n$ while Lemma 2 and Lemma 3 obtain their eigenvalues. As the proof of Lemma 1 involves rather tedious but straightforward algebra its proof is omitted from this appendix.

LEMMA 1: *For any n and any magnitude of $\rho_n(c,\alpha)$ the trace of $(\boldsymbol{R}_n'\boldsymbol{R}_n)^{-1}$ is given by*

$$Tr(\boldsymbol{R}_n'\boldsymbol{R}_n)^{-1} = \sum_{j=1}^{n}\left(\sum_{\ell=1}^{j}\rho_n(c,\alpha)^{2(\ell-1)}\right) \tag{44}$$

and the trace of the centered counterpart $\boldsymbol{H}_n(\boldsymbol{R}_n'\boldsymbol{R}_n)^{-1}\boldsymbol{H}_n$ is given by

$$\begin{aligned} Tr(\boldsymbol{H}_n(\boldsymbol{R}_n'\boldsymbol{R}_n)^{-1}\boldsymbol{H}_n) &= \left(1-\frac{1}{n}\right)\sum_{j=1}^{n}\left(\sum_{\ell=1}^{j}\rho_n(c,\alpha)^{2(\ell-1)}\right) \\ &- \frac{2}{n}\sum_{k=1}^{n-1}\left(\sum_{j=1}^{k}\left(\sum_{\ell=1}^{j}\rho_n(c,\alpha)^{2(\ell-1)}\right)\right)\rho_n(c,\alpha)^{n-k}. \end{aligned} \tag{45}$$

REMARKS: The differences between (44) and (45) highlight the impact that centering has on the eigenstructure of covariances. It is straightforward to establish that the second component in the right hand side of (45) vanishes asymptotically provided that $|\rho|< 1$, corroborating our earlier observation that under pure stationarity the eigenvalues of $(\boldsymbol{R}_n'\boldsymbol{R}_n)^{-1}$ and $\boldsymbol{H}_n(\boldsymbol{R}_n'\boldsymbol{R}_n)^{-1}\boldsymbol{H}_n$ will coincide for sufficiently large n. For ρ in the vicinity of unity however this same component becomes equally dominant as the first component causing the (suitably normalised) eigenvalues of the these two matrices to converge to different limits.

LEMMA 2: *(i) The eigenvalues of $(\boldsymbol{R}_n'\boldsymbol{R}_n)^{-1}$ in (42) are given by $\lambda_k^{uc} = (1+\rho_n(c,\alpha)^2+2\rho_n(c,\alpha)\cos\phi_k)^{-1}$*

where the $\phi_k's$ solve $\sin[(n+1)\phi] + \rho_n(c,\alpha)\sin[n\phi] = 0$. *(ii) For* $\rho_n(c,\alpha) = 1$ *the solutions to the trigonometric equation in (i) are given by* $\phi_k = 2k\pi/(2n+1)$ *for* $k = 1, 2, \ldots, n$.

PROOF: We let ν denote the eigenvalues of $\boldsymbol{R}_n'\boldsymbol{R}_n$. Assuming $\rho > 0$ and applying the change of variable

$$2\cos\phi \quad = \quad (1+\rho^2-\nu)/(-\rho) \tag{46}$$

we can write

$$\boldsymbol{R}_n'\boldsymbol{R}_n - \nu\ \boldsymbol{I}_n \quad = \quad (-\rho)\begin{pmatrix} \frac{1+\rho^2-\nu}{-\rho} & 1 & 0 & \ldots & 0 & 0 \\ 1 & \frac{1+\rho^2-\nu}{-\rho} & 1 & \ldots & 0 & 0 \\ 0 & 1 & \frac{1+\rho^2-\nu}{-\rho} & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & \frac{1+\rho^2-\nu}{-\rho} & 1 \\ 0 & 0 & 0 & \ldots & 1 & \frac{1-\nu}{-\rho} \end{pmatrix} \tag{47}$$

$$= \quad (-\rho)\begin{pmatrix} 2\cos\phi & 1 & 0 & \ldots & 0 & 0 \\ 1 & 2\cos\phi & 1 & \ldots & 0 & 0 \\ 0 & 1 & 2\cos\phi & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & 2\cos\phi & 1 \\ 0 & 0 & 0 & \ldots & 1 & (\rho + 2\cos\phi) \end{pmatrix} \tag{48}$$

$$\equiv \quad (-\rho)\boldsymbol{K}_n. \tag{49}$$

It now follows that $\mathrm{Det}[\boldsymbol{R}_n'\boldsymbol{R}_n - \nu\ \boldsymbol{I}_n] = (-\rho)^n\mathrm{Det}[\boldsymbol{K}_n]$. Proceeding with a standard cofactor expansion of the determinant of $\boldsymbol{K}_n$ we have $\mathrm{Det}[\boldsymbol{K}_n] = \mathrm{Det}[\boldsymbol{A}_n(\cos\phi)] + \rho\mathrm{Det}[\boldsymbol{A}_{n-1}(\cos\phi)]$ where

$$\boldsymbol{A}_n(x) = \begin{pmatrix} 2x & 1 & 0 & \ldots & 0 & 0 \\ 1 & 2x & 1 & \ldots & 0 & 0 \\ 0 & 1 & 2x & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & 2x & 1 \\ 0 & 0 & 0 & \ldots & 1 & 2x \end{pmatrix}. \tag{50}$$

We can now recognise that $\mathrm{Det}[\boldsymbol{A}_n(x)] = U_n(x)$ with $U_n(x)$ denoting the n^{th} Chebyshev polynomial of

the second kind. We therefore have

$$\text{Det}[\boldsymbol{R}_n'\boldsymbol{R}_n - \nu\ I_n] \quad = \quad (-\rho)^n[\rho U_{n-1}(\cos\phi) + U_n(\cos\phi)]. \tag{51}$$

From the properties of Chebyshev polynomials, $U_n(\cos\phi) = \sin[(n+1)\phi]/\sin\phi$, so that (49) leads to $\sin[(n+1)\phi] + \rho\sin[n\phi] = 0$ as the eigenvalues must solve $\text{Det}[\boldsymbol{R}_n'\boldsymbol{R}_n - \nu\ I_n] = 0$. The eigenvalues of $(\boldsymbol{R}_n'\boldsymbol{R}_n)$ are now obtained from our earlier change of variable in (46) resulting in $\nu_k = (1+\rho^2+2\rho\cos\phi_k)$ and the eigenvalues of $(\boldsymbol{R}_n'\boldsymbol{R}_n)^{-1}$ follow directly as $\lambda_k^{uc} = (1+\rho^2+2\rho\cos\phi_k)^{-1}$ where the ϕ_k are the n roots of $U_n(\cos\phi) + \rho U_{n-1}(\cos\phi) = 0$.

Using the standard trigonometric toolkit it is straightforward to establish that under $\rho = 1$, $\sin[(n+1)\phi] + \sin[n\phi] = 0$ admits the simple closed form solutions given by $\phi_k = 2k\pi/(2n+1)$ leading to $\lambda_k^{uc} = (1+\rho^2+2\rho\cos[2k\pi/(2n+1)])^{-1}$ while no closed form solutions exist for $\rho \neq 1$.

LEMMA 3: *(i) The eigenvalues of (43) sorted in ascending order are given by* $\omega_1 = 0$ *and* $\omega_k = (1+\rho_n(c,\alpha)^2 + 2\rho_n(c,\alpha)\cos\widetilde{\phi}_k)$ *for* $k = 2,\ldots,n$. *(ii) Under* $\rho_n(c,\alpha) = 1$, $\widetilde{\phi}_k = k\pi/n$.

PROOF: We first note that (43) is symmetric with a zero eigenvalue while its remaining eigenvalues are strictly positive. Let $\mathcal{P}_n(\omega)$ denote the characteristic polynomial of (43) which we wish to obtain an expression for and $\mathcal{S}_n(\omega)$ the characteristic polynomial of (42). It now suffices to observe that $\mathcal{P}_n(\omega)$ can be expressed in terms of $\mathcal{S}_n(\omega)$. Indeed using standard cofactor expansions we have

$$\mathcal{S}_n(\omega) \quad = \quad (1+\rho^2-\omega)\mathcal{S}_{n-1}(\omega) - \rho^2\mathcal{S}_{n-2}(\omega) \tag{52}$$

with initial conditions $\mathcal{S}_0 = 1$ and $\mathcal{S}_1 = 1-\omega$. The characteristic polynomial of (43) is then given by

$$\mathcal{P}_n(\omega) \quad = \quad (1-\omega)\mathcal{S}_{n-1}(\omega) - \rho^2\mathcal{S}_{n-2}(\omega). \tag{53}$$

Solving first (52) using the change of variable $\omega = (1+\rho^2-2\rho\cos\widetilde{\phi})$ and standard difference equation techniques we have

$$\begin{aligned}\mathcal{S}_n(\widetilde{\phi}) \quad = \quad & \frac{-\rho + (\cos\widetilde{\phi} + i\sin\widetilde{\phi})}{2i\sin\widetilde{\phi}}\rho^n(\cos\widetilde{\phi} + i\sin\widetilde{\phi})^n + \\ & \frac{\rho - (\cos\widetilde{\phi} - i\sin\widetilde{\phi})}{2i\sin\widetilde{\phi}}\rho^n(\cos\widetilde{\phi} - i\sin\widetilde{\phi})^n.\end{aligned} \tag{54}$$

Plugging (54) into (53) and solving for $\mathcal{P}_n(\widetilde{\phi}) = 0$ leads to the desired result.

APPENDIX B

PROOF OF PROPOSITION 1: Assumption A ensures that the limiting behaviour of $\hat{\lambda}_k^{uc}/n^2$ is determined by the suitably normalised eigenvalues of $(\boldsymbol{R}_n'\boldsymbol{R}_n)^{-1}$ which as shown below are of order n^2 under $\rho_n(c,1) = 1-(c/n)$. From Lemma 2 we can write the eigenvalues of $(\boldsymbol{R}_n'\boldsymbol{R}_n)^{-1}$ as $\lambda_k^{uc} = (1+\rho_n(c,1)^2 + 2\rho_n(c,1)\cos\phi_k)^{-1}$ with ϕ_k $(k=1,\ldots,n)$ satisfying $\sin[(n+1)\phi] + \rho_n(c,1)\sin[n\phi] = 0$. It is now important to point out that as λ_k^{uc} is formulated in reciprocal form and as these eigenvalues are typically sorted in descending order, ϕ_k above is naturally replaced with ϕ_{n-k+1} so that in descending order the eigenvalues of $(\boldsymbol{R}_n'\boldsymbol{R}_n)^{-1}$ are written as $\lambda_k^{uc} = (1+\rho_n(c,1)^2 + 2\rho_n(c,1)\cos\phi_{n-k+1})^{-1}$.

The trigononometric equation $\sin[(n+1)\phi] + \rho_n(c,1)\sin[n\phi] = 0$ admits the closed form solutions $\phi_{n-k+1} = 2(n-k+1)\pi/(2n+1)$ under $\rho_n(c,1) = 1$ but has no closed form solutions for general $\rho_n(c,1)$. As we operate in the vicinity of unity however we propose to approximate these eigenvalues as $\lambda_k^{uc} \sim (1+(1-c/n)^2 + 2(1-c/n)\cos[2(n-k+1)\pi/(2n+1)])^{-1}$ which we subsequently adjust to ensure that their sum matches the exact trace of $(\boldsymbol{R}_n'\boldsymbol{R}_n)^{-1}$ as formulated in Lemma 1.

A standard Taylor expansion of $(1+\rho_n(c,1)^2 + 2\rho_n(c,1)\cos(2(n-k+1)\pi/(2n+1)))^{-1}$ gives

$$\lambda_k^{uc} \sim \frac{n^2}{c^2 + (1/4)(2k-1)^2\pi^2} + O(n) \tag{55}$$

or equivalently

$$\frac{\lambda_k^{uc}}{n^2} \sim \frac{1}{c^2 + (1/4)(2k-1)^2\pi^2} + O\left(\frac{1}{n}\right). \tag{56}$$

Using (55) it is now useful to note that

$$\lim_{n\to\infty}\sum_{k=1}^{n}\frac{1}{c^2 + (1/4)(2k-1)^2\pi^2} = \frac{\tanh(c)}{2c}. \tag{57}$$

We can now refine the approximation in (55) by adjusting it in a way that ensures that the sum of the adjusted eigenvalues matches the exact normalised trace of $(\boldsymbol{R}_n'\boldsymbol{R}_n)^{-1}$. Using $\rho_n(c,1) = 1-(c/n)$ in (44) gives

$$\frac{1}{n^2}Tr(\boldsymbol{R}_n'\boldsymbol{R}_n)^{-1} = \frac{\left(-c^2n - c^2 + n^2\left(\frac{(c-n)^2}{n^2}\right)^{n+1} + 2cn^2 + 2cn - n^2\right)}{c^2(c-2n)^2} \tag{58}$$

from which it also follows that

$$\lim_{n\to\infty}\frac{1}{n^2}Tr(\boldsymbol{R}_n'\boldsymbol{R}_n)^{-1} = \frac{-1+2c+e^{-2c}}{4c^2}. \tag{59}$$

As a side note it is here useful to contrast (59) with (57) as the spread between these two quantities provides a metric for the distortions induced by our use of $\phi_k = 2k\pi/(2n+1)$ in λ_k^{uc}. For $c = 5$ and $c = 10$ for instance (57) takes the values 0.099 and 0.05 compared with 0.090 and 0.05 for (59).

We can now refine the approximation in (55) as

$$\lambda_k^{uc} \sim \lambda_{*k}^{uc} \equiv \frac{2c}{\tanh c}\left(\frac{-1+2c+e^{-2c}}{4c^2}\right)\left(\frac{1}{c^2+(1/4)(2k-1)^2\pi^2}\right) \tag{60}$$

which establishes our result in (12). Our statement in (14) now follows straightforwardly.

PROOF OF PROPOSITION 2. Our results in (18) and (20) follow from the same steps as in the proof of Proposition 1. From Lemma 3 the (non-zero) eigenvalues of $\boldsymbol{H}_n(\boldsymbol{R}_n'\boldsymbol{R}_n)^{-1}\boldsymbol{H}_n$ sorted in descending order are given by $\lambda_k = (1+\rho_n(c,1)^2 + 2\rho_n(c,1)\cos\widetilde{\phi}_{n-k+1})^{-1}$ and which we now approximate as $\lambda_k \sim (1+(1-c/n)^2 + 2(1-c/n)\cos[(n-k+1)\pi/n])^{-1}$. A standard Taylor expansion now gives

$$\lambda_k \sim \frac{n^2}{c^2+k^2\pi^2} + O(n) \tag{61}$$

from which we also have

$$\frac{\lambda_k}{n^2} \sim \frac{1}{c^2+k^2\pi^2} + O\left(\frac{1}{n}\right). \tag{62}$$

Using (61) it is also useful to note that

$$\lim_{n\to\infty}\sum_{k=1}^{n}\frac{1}{c^2+k^2\pi^2} = \frac{-1+c\coth c}{2c^2}. \tag{63}$$

We now refine the approximation in (62) by adjusting the eigenvalues in a way that ensures that the sum of the adjusted eigenvalues matches the exact trace of $\boldsymbol{H}_n(\boldsymbol{R}_n'\boldsymbol{R}_n)^{-1}\boldsymbol{H}_n$. Using $\rho_n(c,1) = 1-(c/n)$ in (45) gives

$$\begin{aligned} Tr\boldsymbol{H}_n(\boldsymbol{R}_n'\boldsymbol{R}_n)^{-1}\boldsymbol{H}_n) &= \frac{n^3}{c^3(c-2n)^2}\left(c(n-1)\left(\frac{\left(\left(1-\frac{c}{n}\right)^{2n}-1\right)(c-n)^2}{n^2} - \frac{(c-n)^2}{n} + n\right)\right) + \\ &\quad \frac{n^3}{c^3(c-2n)^2}\left(2n\left(\left(1-\frac{c}{n}\right)^{n}-1\right)\left(\left(1-\frac{c}{n}\right)^{n+2}+\frac{2c}{n}-3\right)\left(1-\frac{c}{n}\right)+2c(c-2n)\right) \end{aligned} \tag{64}$$

from which it also follows that

$$\lim_{n\to\infty}\frac{1}{n^2}Tr\boldsymbol{H}_n(\boldsymbol{R}_n'\boldsymbol{R}_n)^{-1}\boldsymbol{H}_n) = \frac{6+2c^2+2e^{-2c}-8e^{-c}+c(-5+e^{-2c})}{4c^3}. \tag{65}$$

We can now refine the approximation in (62) as

$$\lambda_k \sim \lambda_{*k} \equiv \left(\frac{2c^2}{-1+c\coth c}\right)\frac{6+2c^2+2e^{-2c}-8e^{-c}+c(-5+e^{-2c})}{4c^3}\frac{1}{c^2+k^2\pi^2} \tag{66}$$

which establishes our result in (18). Our statement in (20) follows straightforwardly.

PROOF OF PROPOSITION 3. Here we operate under $\rho \equiv \rho_n(c,\alpha) = 1-c/n^\alpha$ for $\alpha \in [0,1)$. It now suffices to replace $\rho_n(c,1) = 1-c/n$ with $\rho_n(c,\alpha) = 1-(c/n^\alpha)$ in $\lambda_k \sim (1+(1-(c/n^\alpha))^2 + 2(1-(c/n^\alpha))\cos[(n-k+1)\pi/n])^{-1}$. A standard Taylor expansion gives

$$\lambda_k \sim \frac{n^{2\alpha}}{c^2+n^\alpha(1+n^\alpha)O(n^{-2})} \tag{67}$$

from which we also have

$$\frac{1}{n^{2\alpha}}\lambda_k = \frac{1}{c^2}+O\left(\frac{1}{n^{2(1-\alpha)}}\right). \tag{68}$$

As $1+\alpha > 2\alpha$ it now follows that $\lim_{n\to\infty}\lambda_k/n^{1+\alpha}=0$ which using Assumption A also implies $\hat{\lambda}_k/n^{1+\alpha} \xrightarrow{p} 0$ as stated in (24).

The boundedness of $Tr(\boldsymbol{S}_n)/n^{1+\alpha}$ as stated in (25) follows directly using $\rho_n(c,\alpha)=1-c/n^\alpha$ in (45) of Lemma 1. We have

$$\begin{aligned} Tr(\boldsymbol{S}_n) = {} & \frac{1}{c^3\left(c-2n^\alpha\right)^2}n^{2\alpha-1}\left(2c^2n^{\alpha+1}+2c^2n^\alpha-9cn^{2a}+6n^{3\alpha}-c^3n^2+c^3+2c^2n^{\alpha+2}-5cn^{2\alpha+1}-\right. \\ & \left. 2\left(1-cn^{-\alpha}\right)^n\left(c-2n^\alpha\right)^2\left(n^\alpha-c\right)+\left(1-cn^{-\alpha}\right)^{2n}\left(c-n^\alpha\right)^2\left(2n^\alpha+c(n-3)\right)\right) \end{aligned} \tag{69}$$

which is of order $n^{1+\alpha}$.

PROOF OF PROPOSITION 4. (i) Using $\rho_n(c,1)=1$ in (46) of Lemma 1 leads to $Tr(\boldsymbol{S}_n)=(n-1)(n+1)/6$ from which (36) follows. (ii) The case for $\rho_n(c,\alpha)=1-c/n$ has been treated in (65). (iii) For $\rho_n(c,\alpha)=|1-c|<1$ we have

$$\begin{aligned} Tr(\boldsymbol{S}_n) = {} & \frac{1}{(c-2)^2c^3n}\left(c(n-1)\left((c-1)^2\left((1-c)^{2n}-1\right)-(c-1)^2n+n\right)\right. \\ & + \left. 2\left((1-c)\left((1-c)^n-1\right)\left((1-c)^{n+2}+2(c-1)-1\right)+(c-2)cn\right)\right) \end{aligned} \tag{70}$$

from which it follows that $Tr(\boldsymbol{S}_n)/n \xrightarrow{p} 1/(2c-c^2)$.

REFERENCES

Antognini, J. M., & Sohl-Dickstein, J. (2018). PCA of high dimensional random walks with comparison to neural network training, 32^{nd} *Conference on Neural Information Processing Systems (NeurIPS 2018), Montréal, Canada.*

Bai, Z., Choi, K. P., & Fujikoshi, Y. (2018). Limiting behavior of eigenvalues in high-dimensional MANOVA via RMT, *Annals of Statistics*, 46, 2985-3013.

Bai, J., & Ng, S. (2004). A PANIC attack on unit roots and cointegration, *Econometrica*, 71, 135-171.

Bai, Z., & Silverstein, J. W. (2010). *Spectral Analysis of Large Dimensional Random Matrices.* Springer Series in Statistics. Springer, New-York, 2010 (second edition).

Baillie, R. T., & Bollerslev, T. (1989). Common Stochastic Trends in a System of Exchange Rates, *Journal of Finance*, 44, 167-181.

Blanchard, O., & Quah, D. 1989. The Dynamic Effects of Aggregate Demand and Supply Disturbances,*American Economic Review*,79, 655-673.

Bookstein, F. L. (2013). Random walk as null model for high-dimensional morphometrics of fossil series: geometrical considerations, *Paleobiology*, 39, 52-74.

Bühlmann, P., & Van de Geer, S. (2011). *Statistics for High-Dimensional Data: Methods, Theory and Applications,* Springer Series in Statistics.

Carrasco, M., Chernozhukov, V., Goncalves, S., & Renault, E. (2015). High dimensional problems in econometrics, *Journal of Econometrics*, 186, 277-279.

Corona, F., Poncela, P., & Ruiz, E. (2020). Estimating Non-stationary common factors: implications for risk sharing,*Computational Economics,* 55, 37-60.

Diebold, F. X., Gardeazabal, J., & Yilmaz, K. (1994), On Cointegration and Exchange Rate Dynamics, *Journal of Finance*, 49, 727-35.

Fujikoshi, Y., Sakurai, T., & Yanagihara, H. (2017). Consistency of high-dimensional AIC-type and

Cp-type criteria in multivariate linear regression, *Journal of Multivariate Analysis*, 123, 184-200.

Fujikoshi, Y. (2017). High-dimensional asymptotic distributions of characteristic roots in multivariate linear models and canonical correlation analysis, *Hiroshima Mathematical Journal*, 47, 249-271.

Fujikoshi, Y. (2019). Strong Consistency of Log-Likelihood-Based Information Criterion in High-Dimensional Canonical Correlation Analysis, *Sankhya A*,

Gali, J., (1992). How well does the IS-LM model fit postwar U.S. data?, *Quarterly Journal of Economics*, 107, 709-738.

Gao, Z. & Tsay, R. S. (2019). A Structural Factor Approach to Modeling High Dimensional Time Series and Space-Time Data, *Journal of Time Series Analysis*, 40, 343-362.

Gonzalo, J. (1994). Five alternative methods of estimating long-run equilibrium relationships, *Journal of Econometrics*,60, 203-233.

Gonzalo, J., & Granger, C. W.J. (1995). Estimation of Common Long-Memory Components in Cointegrated Systems, *Journal of Business and Economic Statistics*,13, 27-35.

Gonzalo, J., & Pitarakis, J. (1995). Comovements in Large Systems, *DES-Working Papers. Statistics and Econometrics. WS 5825. Universidad Carlos III de Madrid.*

Gonzalo, J., & Pitarakis, J. (1998). Specification via model selection in vector error correction models, *Economics Letters*, 60, 321-328.

Gonzalo, J., & Pitarakis, J. (1999). Dimensionality Effect in Cointegration Analysis. Chapter 9 in Engle, R. and H. White (Ed.), *Cointegration, Causality and Forecasting. A Festschrift in Honour of Clive WJ Granger*, Oxford University Press, 1999.

Gonzalo, J. & Pitarakis, J. (2002). Lag length estimation in large dimensional systems, *Journal of Time Series Analysis,* 23, 401-423.

Granger, C. W. J. (1998). Extracting information from mega-panels and high frequency data, *Statistica Neerlandica*, 52, 258-272.

Harris, D. (1997). Principal Components Analysis of cointegrated Time Series, *Econometric Theory*, 13, 529-557.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning (2nd ed.)*, Springer Series in Statistics.

Johansen, S., (1988). Statistical analysis of cointegrating vectors, *Journal of Economic Dynamics and Control*, 12, 231-254.

Johansen, S., (1991). Estimation and hypothesis testing of cointegrating vectors in Gaussian vector autoregressive models, *Econometrica*, 59, 1551-1580.

Johnstone, I., & Titterington, D. M. (2009). Statistical Challenges of high-dimensional data, Philosophical Transactions of the Royal Society A, 367, 4237-4253.

Kasa, K. (1992). Common stochastic trends in international stock markets *Journal of Monetary Economics*, 29, 95-124.

King, R. G., Plosser, C. I., Stock, J. H., & Watson, M. W.(1991). Stochastic Trends and Economic Fluctuations,*American Economic Review*, 81, 819-840.

Koch, I. (2014). *Analysis of Multivariate High-Dimensional Data.* Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press.

Ma, Z. (2013) Sparse Principal Component Analysis and Iterative Thresholding, Annals of Statistics, 41, 772-801.

Magdalinos, A., & Phillips, P. C. B. (2007). Limit theory for moderate deviations from a unit root, *Journal of Econometrics*, 136, 115-130.

Marcet, A. (2004). Overdifferencing VAR's is OK. Unpublished Manuscript.

McCracken, W. W., & Ng, S. (2015). FRED-MD: A monthly database for Macroeconomic Research, *Federal Reserve Bank of St. Louis*, Working Paper 2015-012B.

Moore, J., Ahmed, H., & Antia, R. (2018). High dimensional random walks can appear low dimensional:

Application to influenza H3N2 evolution, *Journal of Theoretical Biology*, 447, 56-64.

Mullainathan, S., & Spiess, J. (2017). Machine Learning: An Applied Econometric Approach, *Journal of Economic Perspectives*, 31, 7-106.

Ng, S. (2017). Opportunities and Challenges: Lessons from Analyzing Terabytes of Scanner Data. NBER Working Paper No. 23673.

Onatski, A., & Wang, C. (2018a). Alternative asymptotics for cointegration tests in Large VARs, *Econometrica*, 86, 1465-1478.

Onatski, A., & Wang, C. (2018b). Spurious Factor Analysis, Manuscript, *University of Cambridge*.

Onatski, A., & Wang, C. (2019). Extreme Canonical correlations and high dimensional cointegration analysis, *Journal of Econometrics*, 212, 307-322.

Sánchez, I., & Peña, D. (2001). Properties of Predictors in overdifferenced nearly nonstationary autoregression, *Journal of Time Series Analysis*, 22, 45-66.

Stock, J. H., & Watson, M. W. (1989), "New Indexes of Coincident and Leading Economic Indicators," NBER Macroeconomics Annual, 351–393.

Stock, J. H., & Watson, M. W. (1999). Forecasting inflation. *Journal of Monetary Economics*, 44, 293–335.

Stock, J. H., & Watson, M. W. (2002). Macroeconomic forecasting using diffusion indexes, *Journal of Business and Economic Statistics,* 20, 147-167.

Stock, J. H., & Watson, M. W. (2005). "Implications of Dynamic Factor Models for VAR Analysis," NBER Working Papers 11467, National Bureau of Economic Research, Inc.

Varian, H. (2014). Big Data: New Tricks for Econometrics, Journal of Economic Perspectives, 28, 2, 3-28.

Wachter, K. W. (1978). The strong limits of random matrix spectra for samplle matrices of independent elements. Annals of Probability, 6, 1-18.

Wachter, K. W. (1980). The limiting empirical measure of multiple discriminant ratios. Annals of

Statistics, 8, 937-957.

Zou, H., Hastie, T., & Tibshirani, R. (2006). Sparse Principal Component Analysis, *Journal of Computational and Graphical Statistics,* 15, 265-286.