Tests for the Number of Common Latent Factors Between Two Short Panels

Alain-Philippe Fortin *

November 22, 2023

Abstract

This paper develops inference tools for the number of common latent factors between two panels having a large cross-sectional dimension n and small time series dimension T. I propose an approach that builds on general tests for the dimension of the intersection of two matrix column spaces, where each matrix is observed with noise. The test statistics are based on canonical correlations, and their asymptotic distributions are derived via perturbation methods. An application to large cross-sectional panels of monthly US stock returns and corporate bond returns finds 4-5 common factors during the 2019-2020 period. These common factors explain around 30% of the cross-sectional variance in stock returns and at least 50% of the cross-sectional variance in bond returns. Moreover, I document a structural break in the factor loadings of stock returns between the pre-COVID and COVID years, while for bond returns I find two factors with constant loadings between the pre-COVID and COVID years.

Keywords: Canonical correlations, latent factor analysis, panel data, large n and fixed T asymptotics, equity returns, bond returns.

JEL codes: C12, C23, C38, C58, G12.

*University of Geneva and Swiss Finance Institute.

1 Introduction

Factor models are widely used as a dimension reduction tool for large panels in economics, see e.g. Bai and Ng (2002), Stock and Watson (2002) and Bai (2003), as well as other fields. In finance, approximate factor models constitute the backbone of the Arbitrage Pricing Theory (APT), see Ross (1976) and Chamberlain and Rothschild (1983). In such models, a small number of latent factors explain a large portion of the panel data variation. The two most popular methods for extracting latent factors in panels are Principal Component Analysis (PCA) and Factor Analysis (FA). A natural question when extracting latent factors on two separate panels is how many latent factors are common to both, and how many are panel specific. For instance, one might be interested in the number of common factors driving returns of stocks and corporate bonds, or between stocks in two different industries. More generally, determining the number of common latent factors between separate panels is crucial to understand common drivers of risk premia among different groups of assets. The question is challenging because (1) latent factors are estimated with noise, and (2) latent factors can only be estimated up to some unknown rotation matrix. In particular, the second point implies that traditional correlations between extracted factors will not provide a meaningful measure of their dependence in general.

The main contribution of this paper is to propose tests for the number of common latent factors between two panels having a large cross-sectional dimension n and small time series dimension T, i.e. two short panels. I propose an approach that builds on general tests for the dimension of the intersection of two matrix column spaces, where each matrix is observed with noise. By working with matrix column spaces I take into account the issue outlined in (2), and by introducing noise I taking into account the issue outlined in (1). The test statistics are based on canonical correlations, defined as the eigenvalues of a matrix constructed from inner products between orthonormal bases for each subspace. The dimension of the subspace intersection, which corresponds to the number of common factors between the two short panels, is shown to be equal to the number of unit canonical correlations, which constitutes the foundation for my tests. I propose two different statistics to test the dimension of the subspace intersection, depending on whether the matrices have (1) a fixed number of rows or (2) an increasing number of rows. For short panels, the first case corresponds to testing for common factors between two cross-sections over the same time period (see section 6.1), while the second case corresponds to testing for common factors in the same cross-section over two different time periods (see Section 6.2). I derive the asymptotic distributions of the two statistics using perturbation methods. I show that the fixed dimensional case yields sample canonical correlations converging to unity at the super-consistent rate $O_p(\frac{1}{n})$, while in the increasing dimensional case convergence occurs at rate $O_p(\frac{1}{\sqrt{n}})$.

I use my methodology to tests for common latent factors between monthly US stock returns and corporate bond returns. I focus on the 2019-2020 period, i.e. the pre-COVID year and the COVID year. I find between 4 and 5 common factors driving the returns of US stocks and bonds during this period. These common factors explain around 30% of the cross-sectional variance in stock returns, and at least 50% of the cross-sectional variance in bond returns. I find that stocks returns had between 1 and 2 specific factors during the pre-COVID and COVID years, explaining between 5% and 8% of their cross-sectional variance. On the other hand, while bond returns had only 1 specific factor during 2019, explaining around 15% for their cross-sectional variance, they had 6 specific factors during the 2020 COVID year, explaining around 40% of their cross-sectional variance. I also document a structural break in the factor loadings of stock returns between the pre-COVID and COVID years, while for bond returns I find two factors with constant loadings between the pre-COVID and COVID years.

My work is most closely related to the literature using canonical correlations to study latent factor models. Bai and Ng (2006) and Pelger (2019) employ canonical correlations to measure similarity between principal components (PCs) and observed factors, whereas Goyal, Perignon, and Villa (2008) use canonical correlations to study similarity of the factor structure driving stock returns on the NYSE and NASDAQ. Andreou et al. (2019) consider inference on the number of common factors in a two group factor setting using canonical correlations computed from PCs, while Choi et al. (2023) propose two selection criteria for the number the common factors when there are possibly more than two groups, robust to the presence of serially correlated and weakly cross-sectionally correlated idiosyncratic errors. Pelger and Xiong (2022) use canonical correlations to construct a test of change of loadings in a state-varying factor model. Other closely related work include Chen (2010, 2012), Wang (2012), Ando and Bai (2015, 2017), Breitung and Eickmeier (2016) and Han (2021). The methodologies put forward in all these papers rely on panels having large cross-section n and large time dimension T.

In recent work, PCA and FA have also been applied to large cross-sectional latent factor models with small time dimension, i.e. short panels with large n but small T. Fortin et al. (2022) show how PCA can be used to conduct inference on the number of factors when errors have a spherical covariance structure (see also Theorem 4 in Bai (2003)), whereas Fortin et al. (2023) rely on FA and a diagonal error covariance structure. The small T perspective is interesting because it mitigates concerns about panel unbalanceness and yields an effective approach to capture general forms of time-variation in factor betas, risk premia and number of factors by performing the analysis in short subperiods (either non-overlapping, or rolling windows) of the sample of interest. On the other hand, the small T setting makes inference on the number of factors more difficult because it induces an error-in-variable problem in the estimation of the factor loadings (see Fortin et al. (2022)). This paper complements the above literature by proposing new tests for the number of common latent factors between two short panels.

The outline of the paper is as follows. In Section 2 I lay out the theoretical framework and introduce test statistics for the dimension of the intersection of two matrix column spaces. In Section 3 I consider testing when the space dimension is fixed. In Section 4 I consider three special cases, i.e. when root-n consistent estimators are available, the number of columns is increasing, or instrumental variables are available. In Section 5 I consider testing when the space dimension is increasing. In Section 6 I provide an empirical application where I test for common latent factors driving monthly US stock returns and corporate bond returns. I provide concluding remarks in Section 7.

2 Framework

Let Π_1 and Π_2 be $m \times k_1$ and $m \times k_2$ matrices respectively. I allow Π_1 and Π_2 to be rank deficient, i.e. $r_1 :=$ rank$(\Pi_1) \leq k_1$ and $r_2 :=$ rank$(\Pi_2) \leq k_2$. I assume that Π_1 and Π_2 are unobserved, but that noisy proxies are available:

$$\hat{\Pi}_j = \Pi_j + \Psi_j, \quad j = 1, 2. \tag{1}$$

For example, $\hat{\Pi}_j$ could represent an estimator. The methods will hold under general assumptions about the noise Ψ_j. Let col(A) denote the column space of matrix A, and let k_0 denote the dimension of col$(\Pi_1) \cap$ col(Π_2). The problem I consider is to test $H_0(k) : k_0 = k$ against $H_1(k) : k_0 < k$, for some $0 \leq k \leq \min(r_1, r_2)$. If Π_1 and Π_2 were observed, then the problem would be degenerate, as k_0 could be determined with certainty[1].

The test strategy relies on canonical correlations, see e.g. Anderson (2003) and Magnus and Neudecker (2007). While canonical correlations are traditionally employed to study linear dependence between two sets of random variables, they can also be used to measure closeness between two linear subspaces, namely the dimension of their intersection.. Let U_j be an $m \times r_j$ matrix with orthonormal columns spanning col(Π_j), $j = 1, 2$.

Definition 1 *The canonical correlations ρ_l between Π_1 and Π_2 are the eigenvalues of $R_{1,2} := U_1' U_2 U_2' U_1$, ordered as $\rho_1 \geq \rho_2 \geq ... \geq \rho_{r_1}$.*

Since the nonzero eigenvalues of $R_{1,2}$ are the same as the nonzero eigenvalues of $R_{2,1} := U_2' U_1 U_1' U_2$, the number of nonzero canonical correlations is at most $\min(r_1, r_2)$. The following proposition provides the theoretical foundation for our tests.

Proposition 1 *Under $H_0(k)$ we have*

1. *The k largest canonical correlations between Π_1 and Π_2 are equal to 1. The other canonical correlations are strictly less than 1.*

2. *If E_1 is a $r_1 \times k$ matrix whose orthonormal columns are eigenvectors of $R_{1,2}$ corresponding to eigenvalues 1, then the orthonormal columns of $U_1 E_1$ span col$(\Pi_1) \cap$ col(Π_2).*

3. *If O_1 is a $r_1 \times (r_1 - k)$ matrix whose orthonormal columns are eigenvectors of $R_{1,2}$ corresponding to eigenvalues strictly less than 1, then the orthonormal columns of $U_1 O_1$ span the orthogonal complement of col$(\Pi_1) \cap$ col(Π_2) in col(Π_1).*

By symmetry Proposition 1 also holds with indices $1, 2$ interchanged, with E_2 and O_2 being defined in terms of $R_{2,1}$. The test statistics I consider for conducting inference on k_0 are function of sample canonical correlation estimators $\hat{\rho}_1 \geq \hat{\rho}_2 \geq ... \geq \hat{\rho}_{r_1}$. The definition of these estimators will depend on the set-up specifics, and are given in Sections 3 and 5. I present the test statistics next.

[1] For example using the formula $k_0 = \text{rank}(\Pi_1) + \text{rank}(\Pi_2) - \text{rank}([\Pi_1 : \Pi_2])$.

Definition 2 *The statistics to test the null hypothesis $H_0(k)$ are (i) $\xi(k) := n(\sum_{l=1}^{k} \hat{\rho}_l - k)$ and (ii) $\zeta(k) := n\sum_{l=1}^{k}(\hat{\rho}_l - 1)^2$.*

Note that statistics $\xi(k)$ and $\zeta(k)$ entail different rates of convergence for the first k sample canonical correlations, namely $O_p(\frac{1}{n})$ and $O_p(\frac{1}{\sqrt{n}})$. Statistic $\xi(k)$ is considered in Sections 3 , while statistic $\zeta(k)$ is considered in Section 5. The test procedures in these sections take the ranks r_1 and r_2 as given. In practice r_1 and r_2 might be unknown. In this case one can build on results of the rank testing literature, see e.g. Cragg and Donald (1996), Robin and Smith (2000), Kleibergen and Paap (2006), Al-Sadoon (2017), to obtain estimators $\hat{r}_1$ and $\hat{r}_2$ satisfying $\hat{r}_1 = r_1$ and $\hat{r}_2 = r_2$ with probability approaching one, and then the asymptotic theory below remains unchanged. Details on how to implement such estimators in the different settings we consider are provided in the Appendix.

Some remarks on notation. We use $\sigma_l(A)$ and $\lambda_l(A)$ (the latter only when A is symmetric) to denote respectively the lth largest singular value and lth largest eigenvalue of a matrix A. We use $vec(A)$ and $vech(A)$ (the latter only when A is symmetric) to denote respectively the vectorization and half-vectorization of a matrix A. We let $K_{p,q}$ denote the commutation matrix of order (p,q), and D_p denote the duplication matrix of order p. We use A^+ to denote the Moore-Penrose inverse (pseudoinverse) of a matrix A, $Tr(A)$ to denote the trace of square matrix A, and $A \otimes B$ to denote the Kronecker product between matrices A and B. The symbol $\Rightarrow$ denotes convergence in distribution, and $N(0,\Sigma)$ denotes the multivariate Gaussian distribution with mean zero and covariance matrix Σ. Except for the symbols introduced in this section, the symbols introduced below are specific to each section, and their definitions do not carry over to other sections.

3 Fixed dimensional case

In this section we take m as fixed and we assume that root-n consistent estimators $\hat{U}_1$ and $\hat{U}_2$ are available for U_1 and U_2.

Assumption 1 *The estimator $\hat{U}_j$ admits the expansion*

$$\hat{U}_j \mathcal{H}_j = U_j + \hat{\Psi}_j + O_p(\frac{1}{n}), \tag{2}$$

with $\hat{U}_j'\hat{U}_j = I_{r_j}$, $\hat{\Psi}_j = O_p(\frac{1}{\sqrt{n}})$, and $\mathcal{H}_j'\mathcal{H}_j = I_{r_j} + O_p(\frac{1}{\sqrt{n}})$, $j = 1, 2$.

Assumption 1 implies in particular that $\hat{U}_j\hat{U}_j' = U_jU_j' + O_p(\frac{1}{\sqrt{n}})$, i.e. root-$n$ consistency of the projectors. Estimators $\hat{U}_j$ satisfying Assumption 1 can be obtained in different settings. We provide three examples in the next section.

Definition 3 *The sample canonical correlations $\hat{\rho}_l$ between $\hat{\Pi}_1$ and $\hat{\Pi}_2$ are the eigenvalues of $\hat{R}_{1,2} := \hat{U}_1'\hat{U}_2\hat{U}_2'\hat{U}_1$ ordered as $\hat{\rho}_1 \geq \hat{\rho}_2 \geq ... \geq \hat{\rho}_{r_1}$. Moreover we use $\hat{E}_1$ to denote the $r_1 \times k$ matrix whose orthonormal columns are eigenvectors of $\hat{R}_{1,2}$ corresponding to its k largest eigenvalues.*

Since the nonzero eigenvalues of $\hat{R}_{1,2}$ are the same as the nonzero eigenvalues of $\hat{R}_{2,1} := \hat{U}_2'\hat{U}_1\hat{U}_1'\hat{U}_2$, the number of nonzero sample canonical correlations is at most $\min(r_1, r_2)$. The next proposition provides a second-order expansion for the k largest sample canonical correlations under $H_0(k)$.

Proposition 2 *Let Assumption 1 hold, and let $\hat{\rho}_l$ be as in Definition 3. Under $H_0(k)$ we have*

$$\hat{\rho}_l = 1 + \lambda_l(-W'\hat{\Psi}'QQ'\hat{\Psi}W) + O_p(\frac{1}{n^{3/2}}),$$

for $l = 1, \ldots, k$, where $\hat{\Psi} := \hat{\Psi}_1U_1' - \hat{\Psi}_2U_2'$, W is $m \times k$ with orthonormal columns spanning col$(\Pi_1) \cap$ col(Π_2), and Q is $m \times (m - r_1 - r_2 + k)$ with orthonormal columns spanning the orthogonal complement of col(Π_1) + col(Π_2). Moreover, we have $\hat{U}_1\hat{E}_1 = U_1E_1\mathcal{R}_1 + O_p(\frac{1}{\sqrt{n}})$, where $\mathcal{R}_1$ is $k \times k$ and satisfies $\mathcal{R}_1'\mathcal{R}_1 = I_k + O_p(\frac{1}{\sqrt{n}})$.

The first part Proposition 2 relates the deviations of the k largest sample canonical correlation from unity to the k largest eigenvalue of a negative semi-definite matrix, up to a third order term. The negative sign is related to the fact $\hat{\rho}_l$ is bounded above by 1 by construction. The asymptotic expansion of $\hat{\rho}_l$ does not feature a first order term, which implies the rate $\hat{\rho}_l = 1 + O_p(\frac{1}{n})$ for $l = 1, \ldots, k$ under $H_0(k)$. This is a super-convergence result, and in general we only have $\hat{\rho}_l = \rho_l + O_p(\frac{1}{\sqrt{n}})$ for $l > k$ under $H_0(k)$.

By summing the expansion of $\hat{\rho}_l - 1$ in Proposition 2 for $l = 1, \ldots, k$ and using the properties of trace yields the expansion of $\xi(k)$ under $H_0(k)$:

$$\xi(k) = -nTr(W'\hat{\Psi}'QQ'\hat{\Psi}W) + o_p(1)$$

$$= -nvec(\hat{\Psi})'(WW' \otimes QQ')vec(\hat{\Psi}) + o_p(1).$$

Under regularity conditions, the next assumption is implied by a central limit theorem.

Assumption 2 *We have* $\sqrt{n}vec(\hat{\Psi}) \Rightarrow N(0, \Sigma)$ *as* $n \to \infty$.

We allow Σ to be rank deficient, but it must be nonzero to have a nondegenerate test. In Section 4 we show how Assumption 2 can be satisfied under more primitive assumptions. By the result on the distribution of quadratic forms of Gaussian vectors, we get the next result.

Proposition 3 *Let Assumption 1 and 2 hold. Under* $H_0(k)$ *and as* $n \to \infty$ *we have*

$$\xi(k) \Rightarrow - \sum_{l=1}^{k(m-(r_1+r_2-k))} \lambda_l \chi_l^2,$$

where the χ_l^2 *are independent chi-square variables with one degree of freedom, and the* λ_l *are the eigenvalues of the positive semi-definite matrix* $\Xi := (W \otimes Q)'\Sigma(W \otimes Q)$. *Under* $H_1(k)$ *we have* $\xi(k) \xrightarrow{p} -\infty$ *as* $n \to \infty$.

Since $\text{col}(\Pi_1) + \text{col}(\Pi_2)$ is a subspace of $\mathbb{R}^m$, we have the lower bound $k_0 \geq r_1 + r_2 - m$. Hence we only need to test $H_0(k)$ for k between $\max\{r_1 + r_2 - m, 0\} + 1$ and $\min\{r_1, r_2\}$. For values of k in this range the asymptotic distribution of $\xi(k)$ is nondegenerate, unless Ξ is nil[2].

Next we consider a feasible version for the test. Let $\hat{W} := \hat{U}_1 \hat{E}_1$ and let $\hat{Q}$ denote the $m \times (m - r_1 - r_2 + k)$ matrix whose orthonormal columns are eigenvectors of $\hat{U}_1 \hat{U}_1' + \hat{U}_2 \hat{U}_2'$ corresponding to its $m - r_1 - r_2 + k$ smallest eigenvalues.

Lemma 1 *Let Assumption 1 hold. Under* $H_0(k)$ *we have* $(\hat{W} \otimes \hat{Q})\mathcal{O} = W \otimes Q + o_p(1)$ *where* $\mathcal{O}$ *is* $k \times k$ *and satisfies* $\mathcal{O}'\mathcal{O} = I_k$. *Under* $H_1(k)$ *we have* $(\hat{W} \otimes \hat{Q}) = O_p(1)$.

Let $\hat{\Sigma}$ be a consistent estimator of Σ, and let $\hat{\Xi} := (\hat{W} \otimes \hat{Q})'\hat{\Sigma}(\hat{W} \otimes \hat{Q})$. Then it follows from Lemma 1 and the Lipschitz continuity of eigenvalues for symmetric matrices that $\lambda_l(\hat{\Xi}) = \lambda_l(\Xi) + o_p(1)$ under $H_0(k)$, and $\lambda_l(\hat{\Xi}) = O_p(1)$ under $H_1(k)$. Hence we can consistently estimate critical values of the asymptotic distribution of $\xi(k)$ under $H_0(k)$ by simulating a large number of draws from $\sum_{l=1}^{k(m-(r_1+r_2-k))} \hat{\lambda}_l \chi_l^2$, where $\hat{\lambda}_l$ are the eigenvalues of $\hat{\Xi}$. Under $H_1(k)$ these critical values remain bounded in probability. This yields a feasible version of the test with asymptotic correct size under $H_0(k)$ and asymptotic power 1 under $H_1(k)$.

[2] In fact by Proposition 1 we have $\hat{\rho}_l = 1$ for $l = 1, \ldots, r_1 + r_2 - m$ since $\text{col}(\hat{U}_1) + \text{col}(\hat{U}_2)$ is a subspace of $\mathbb{R}^m$. Hence $\xi(k) = 0$ if $k \leq r_1 + r_2 - m$.

4 Discussion of three special cases

In this section we particularize the general results of Section 3 by showing how to obtain estimators $\hat{U}_j$ satisfying Assumption 1 in three cases: (1) k_1 and k_2 are fixed and the elements of Ψ_j are asymptotically Gaussian after scaling, (2) k_1, k_2 are increasing and the elements of Ψ_j are zero-mean random variables, or (3) instrumental variables are available.

4.1 Fixed number of columns

Let us first consider the case where k_1, k_2 are fixed and $\hat{\Pi}_1$, $\hat{\Pi}_2$ are root-n consistent estimators for Π_1 and Π_2.

Assumption 3 $\begin{bmatrix} \sqrt{n_1} vec(\Psi_1) \\ \sqrt{n_2} vec(\Psi_2) \end{bmatrix} \Rightarrow N(0, \Omega)$ *with* $\frac{n_1}{n_2} \to \mu > 0$ *as* $n_1, n_2 \to \infty$.

Let $n := \min\{n_1, n_2\}$, $\hat{\mu}_j := \sqrt{\frac{n_j}{n}}$, and $\mu_j := \lim_{n\to\infty} \hat{\mu}_j$, $j = 1, 2$. Let $U_j S_j V_j'$ be an SVD decomposition of $\hat{\mu}_j \Pi_j$, where S_j is $r_j \times r_j$ diagonal containing its nonzero singular values, for $j = 1, 2$. Correspondingly let $\hat{U}_j \hat{S}_j \hat{V}_j'$ be a "truncated" SVD decomposition of $\hat{\mu}_j \hat{\Pi}_j$, where $\hat{S}_j$ is $r_j \times r_j$ diagonal containing its r_j largest singular values, for $j = 1, 2$.

Lemma 2 *Let Assumption 3 hold. Then we have*

$$\hat{U}_j \mathcal{H}_j = U_j + \hat{\mu}_j \Psi_j V_j S_j^{-1} + O_p(\frac{1}{n}),$$

where $\mathcal{H}_j$ *satisfies* $\mathcal{H}_j' \mathcal{H}_j = I_{r_j} + O_p(\frac{1}{\sqrt{n}})$*, for* $j = 1, 2$.

It follows from Lemma 2 and Assumption 3 that Assumption 1 is satisfied with $\hat{\Psi}_j := \hat{\mu}_j \Psi_j V_j S_j^{-1}$. Moreover Assumption 3 implies that Assumption 2 is satisfied with $\Sigma := A \Omega A'$ and $A := [(\mu_1 \Pi_1)^{+'} \otimes I_m) : -(\mu_2 \Pi_2)^{+'} \otimes I_m]$. Hence Proposition 3 applies.

From $\hat{\Pi}_j = \Pi_j + o_p(1)$ (Assumption 4) and continuity of the pseudoinverse we have $(\hat{\mu}_j \hat{\Pi}_j)^+ := \hat{V}_j \hat{S}_j^{-1} \hat{U}_j' = (\mu_j \Pi_j)^+ + o_p(1)$. It follows that $\hat{A} := [(\hat{\mu}_1 \hat{\Pi}_1)^{+'} \otimes I_m : -(\hat{\mu}_2 \hat{\Pi}_2)^{+'} \otimes I_m]$ is consistent for A. Hence we can consistently estimate Σ provided we have a consistent estimator for Ω.

4.2 Increasing number of columns

Let us next consider the case where $k_1 = n_1$ and $k_2 = n_2$ with $n_1, n_2 \to \infty$, so that the number of columns in Π_1 and Π_2 is increasing while the number of rows remains fixed. In this case $\text{col}(\Pi_j)$ becomes eventually constant, because the additional columns do not add new dimensions to the space for n_j sufficiently large. In particular $H_0(k)$ and $H_1(k)$ are well-defined and Proposition 1 is still valid, for n_1, n_2 sufficiently large. We consider the following assumptions.

Assumption 4 *We have*

(i)
$$\begin{bmatrix} \sqrt{n_1} vech(\frac{1}{n_1}\Psi_1\Psi_1' - \hat{\Phi}_{1,1}) \\ \sqrt{n_2} vech(\frac{1}{n_2}\Psi_2\Psi_2' - \hat{\Phi}_{2,2}) \\ vec(\frac{1}{\sqrt{n_1}}\Psi_1\Pi_1) \\ vec(\frac{1}{\sqrt{n_2}}\Psi_2\Pi_2) \end{bmatrix} \Rightarrow N(0,\Omega) \text{ with } \frac{n_1}{n_2} \to \mu > 0 \text{ as } n_1, n_2 \to \infty.$$

(ii) $\frac{1}{n_j}\Pi_j\Pi_j' \to \Sigma_{j,j}$ *as* $n_j \to \infty$*, where rank*$(\Sigma_{j,j}) = r_j$, $j = 1, 2$.

We interpret $\hat{\Phi}_{j,j}$ as a root-n consistent estimator for the unobservable centering term $\Phi_{j,j} := E[\frac{1}{n_j}\Psi_j\Psi_j']$. Centering is required because, e.g. diagonal elements of $\frac{1}{n_j}\Psi_j'\Psi_j$ are averages of squared random variables. If $\hat{\Phi}_{j,j} = \Phi_{j,j} + o_p(\frac{1}{\sqrt{n_j}})$, then one can consider Φ_j as known, but in general we only require $\hat{\Phi}_{j,j} = \Phi_{j,j} + O_p(\frac{1}{\sqrt{n_j}})$, so that $\sqrt{n}(\hat{\Phi}_{j,j} - \Phi_{j,j})$ can contribute to Ω.

Let n, $\hat{\mu}_j$ and μ_j be as in Section 4.1. Let $U_jD_jU_j'$ be an eigendecomposition of $\frac{\mu_j}{n_j}\Pi_j\Pi_j'$, where D_j is $r_j \times r_j$ diagonal containing its nonzero eigenvalues, for $j = 1, 2$. Correspondingly let $\hat{U}_j\hat{D}_j\hat{U}_j'$ denote the "truncated" eigendecomposition of $\hat{\mu}_j(\frac{1}{n_j}\hat{\Pi}_j\hat{\Pi}_j' - \hat{\Phi}_{j,j})$, where $\hat{D}_j$ is $r_j \times r_j$ diagonal containing its r_j largest eigenvalues, for $j = 1, 2$.

Lemma 3 *Let Assumption 4 hold. Then we have*

$$\hat{U}_j\mathcal{H}_j = U_j + \hat{\mu}_j[\frac{1}{n_j}\Psi_j\Pi_j' + \frac{1}{n_j}\Pi_j\Psi_j' + \frac{1}{n_j}\Psi_j\Psi_j' - \hat{\Phi}_{j,j}]U_jD_j^{-1} + O_p(\frac{1}{n}),$$

where $\mathcal{H}_j$ *satisfies* $\mathcal{H}_j'\mathcal{H}_j = I_{r_j} + O_p(\frac{1}{\sqrt{n}})$*, for* $j = 1, 2$.

If $\Phi_{j,j}$ is spherical, say $\Phi_{j,j} = c_jI_m$ with $c_j > 0$, then we can take the "truncated" eigendecomposition of $\frac{\hat{\mu}_j}{n_j}\hat{\Pi}_j\hat{\Pi}_j'$ (without subtracting $\hat{\mu}_j\hat{\Phi}_{j,j}$), because the bias $\hat{\mu}_j\Phi_{j,j}$ can be absorbed into the matrix $\mathcal{H}_j$ in this case, and then the expansion in Lemma 3 holds with $\Phi_{j,j}$ instead of $\hat{\Phi}_{j,j}$. In particular we can take $\hat{\Phi}_{j,j} = \Phi_{j,j}$ in Assumption 4. This trick is used in Theorem 4 of Bai (2003) to show that, under sphericity, PCA is consistent even for fixed T.

It follows from Lemma 3 and Assumption 4 that Assumption 1 is satisfied with $\hat{\Psi}_j := \hat{\mu}_j \big[\frac{1}{n_j}\Psi_j\Pi_j' + \frac{1}{n_j}\Pi_j\Psi_j' + \frac{1}{n_j}\Psi_j\Psi_j' - \hat{\Phi}_j\big]U_j D_j^{-1}$. Moreover Assumption 4 implies that Assumption 2 is satisfied with $\Sigma := A\Omega A'$ and $A := [((\mu_1\Sigma_{1,1})^{+'} \otimes I_m)D_m : -((\mu_2\Sigma_{2,2})^{+'} \otimes I_m)D_m : ((\mu_1\Sigma_{1,1})^{+'} \otimes I_m)(I_{m^2} + K_m) : -((\mu_2\Sigma_{2,2})^{+'} \otimes I_m)(I_{m^2} + K_m)]$. Hence Proposition 3 applies.

From $\frac{1}{n_j}\hat{\Pi}_j\hat{\Pi}_j' - \hat{\Phi}_{j,j} = \Sigma_{j,j} + o_p(1)$ (Assumption 4) and the continuity of the pseudoinverse we have $(\hat{\mu}_j\hat{\Sigma}_{j,j})^+ := \hat{U}_j\hat{D}_j^{-1}\hat{U}_j' = (\mu_j\Sigma_{j,j})^+ + o_p(1)$. It follows that $\hat{A} := [((\hat{\mu}_1\hat{\Sigma}_{1,1})^{+'} \otimes I_m)D_m : -((\hat{\mu}_2\hat{\Sigma}_{2,2})^{+'} \otimes I_m)D_m : ((\hat{\mu}_1\hat{\Sigma}_{1,1})^{+'} \otimes I_m)(I_{m^2} + K_m) : -((\hat{\mu}_2\hat{\Sigma}_{2,2})^{+'} \otimes I_m)(I_{m^2} + K_m)]$ is consistent for A. Hence we can consistently estimate Σ provided we have a consistent estimator for Ω.

4.3 Instrumental variables

Let us consider again the case where where $k_1 = n_1$ and $k_2 = n_2$ with $n_1, n_2 \to \infty$. Instead of Assumption 4, let us assume that there exist $n_j \times K_j$ matrices of instrumental variables Z_j satisfying the next assumption.

Assumption 5 *There exist $n_j \times K_j$ matrices Z_j, $j = 1, 2$, satisfying*

(i) $\begin{bmatrix} vec(\frac{1}{\sqrt{n_1}}\Psi_1 Z_1) \\ vec(\frac{1}{\sqrt{n_2}}\Psi_2 Z_2) \end{bmatrix} \Rightarrow N(0,\Omega)$ *with* $\frac{n_1}{n_2} \to \mu > 0$ *as* $n_1, n_2 \to \infty$.

(ii) $col(\frac{1}{n_j}\Pi_j Z_j) = col(\Pi_j)$ *for n_j sufficiently large,* $j = 1, 2$.

(iii) $\frac{1}{n_j}\Pi_j Z_j \to \Gamma_j$ *as $n_j \to \infty$, where rank*$(\Gamma_j) = r_j$, $j = 1, 2$.

Let n, $\hat{\mu}_j$ and μ_j be as in Section 4.1. Let $U_j S_j V_j'$ be an SVD decomposition of $\frac{\mu_j}{n_j}\Pi_j Z_j$, where S_j is $r_j \times r_j$ diagonal containing its nonzero singular values, for $j = 1, 2$. Correspondingly let $\hat{U}_j\hat{S}_j\hat{V}_j'$ be a "truncated" SVD decomposition of $\hat{\mu}_j\hat{\Gamma}_j := \frac{\hat{\mu}_j}{n_j}\hat{\Pi}_j Z_j$, where $\hat{S}_j$ is $r_j \times r_j$ diagonal containing its r_j largest singular values, for $j = 1, 2$.

Lemma 4 *Let Assumption 5 hold. Then we have*

$$\hat{U}_j\mathcal{H}_j = U_j + \frac{\hat{\mu}_j}{n_j}\Psi_j Z_j V_j S_j^{-1} + O_p(\frac{1}{n}),$$

where $\mathcal{H}_j$ satisfies $\mathcal{H}_j'\mathcal{H}_j = I_{r_j} + O_p(\frac{1}{\sqrt{n}})$, for $j = 1, 2$.

It follows from Lemma 4 and Assumption 5 that Assumption 1 is satisfied with $\hat{\Psi}_j := \frac{\hat{\mu}_j}{n_j}\Psi_j Z_j V_j S_j^{-1}$. Moreover Assumption 4 implies that Assumption 2 is satisfied with $\Sigma := A\Omega A'$ and $A := [(\mu_1\Gamma_1)^{+'} \otimes I_m : -(\mu_2\Gamma_2)^{+'} \otimes I_m)]$. Hence Proposition 3 applies.

From $\hat{\Gamma}_j = \Gamma_j + o_p(1)$ (Assumption 5) and the continuity of the pseudoinverse we have $(\hat{\mu}_j\hat{\Gamma}_j)^+ := \hat{V}_j\hat{S}_j^{-1}\hat{U}_j' = (\mu_j\Gamma_j)^+ + o_p(1)$. It follows that $\hat{A} := [(\hat{\mu}_1\hat{\Gamma}_1)^{+'} \otimes I_m : -(\hat{\mu}_2\hat{\Gamma}_2)^{+'} \otimes I_m]$ is consistent for A. Hence we can consistently estimate Σ provided we have a consistent estimator for Ω.

5 Increasing dimensional case

In this section we consider the case where $m = n$ and k_1, k_2 are fixed, so that the number of rows in Π_1 and Π_2 is increasing while the number of columns in each matrix remains fixed. In this case, although the space dimension is increasing, the dimensions of col(Π_1), col(Π_2), and col$(\Pi_1) \cap$ col(Π_2) become constant for n sufficiently large[3]. In particular $H_0(k)$ and $H_1(k)$ are still well defined in this set-up. Proposition 1 applies: the k largest eigenvalues of $R_{1,2}$ are equal to 1, and the remaining ones are strictly less than 1, for n sufficiently large[4].

Obtaining a consistent basis estimator for col(Π_j) in this set-up is more difficult, because $\hat{\Pi}_j$ is not a consistent estimator for Π_j, and the approaches in Sections 4.2-4.2 are not directly applicable. The strategy is to proceed indirectly by first obtaining a basis estimator for the row space of Π_j, and then a basis estimator for col(Π_j) via the SVD decomposition. Let $U_jS_jV_j'$ be an SVD decomposition of $\frac{1}{\sqrt{n}}\Pi_j$ $j = 1, 2$. The relationship $U_j = \frac{1}{\sqrt{n}}\Pi_j V_j S_j^{-1}$ suggests the estimator $\hat{U}_j := \frac{1}{\sqrt{n}}\hat{\Pi}_j\hat{V}_j\hat{S}_j^{-1}$, where $\hat{V}_j$ and $\hat{S}_j$ are suitably chosen estimators. To this end we consider the following assumptions.

Assumption 6 *We have*

[3]Indeed rank(Π_j) is a nondecreasing bounded sequence, hence must converge. Similarly rank(Π_1) + rank(Π_2) − rank$([\Pi_1 : \Pi_2])$ must converge.

[4]In contrast to Section 3, the $r_1 - k$ smallest eigenvalues of $R_{1,2}$ now depend on n. For consistency of the test against $H_1(k)$ we need to rule out the possibility that these eigenvalues become arbitrarily close to 1 asymptotically. See Assumption 8 below.

(i) The estimator $\hat{V}_j$ satisfies $\hat{V}_j \mathcal{H}_j = V_j + O_p(\frac{1}{\sqrt{n}})$ with $\mathcal{H}_j' \mathcal{H}_j = I_{r_j} + O_p(\frac{1}{\sqrt{n}})$ and $\hat{V}_j' \hat{V}_j = I_{r_j}, j = 1, 2$.

(ii) The estimator $\hat{\Phi}_{i,j}$ satisfies $\frac{1}{n} \Psi_i' \Psi_j - \hat{\Phi}_{i,j} = O_p(\frac{1}{\sqrt{n}})$, $i, j = 1, 2$, with $\hat{\Phi}_{1,2} = \hat{\Phi}_{2,1}'$.

(iii) $\frac{1}{n} \Psi_j' \Pi_j = O_p(\frac{1}{\sqrt{n}})$, $j = 1, 2$.

Estimators $\hat{V}_j$ satisfying Assumption 6 (i) can be obtained using e.g. the approaches in Section 4.2-4.3 applied to Π_j'. As in Section 4.2, we interpret $\hat{\Phi}_{i,j}$ as a root-n consistent estimator for the unobservable centering term $\Phi_{i,j} := E[\frac{1}{n} \Psi_i' \Psi_j]$. For S_j we use the estimator $\hat{S}_j^2 := \hat{V}_j' (\frac{1}{n} \hat{\Pi}' \hat{\Pi} - \hat{\Phi}_{j,j}) \hat{V}_j$, for $j = 1, 2$.

Lemma 5 *Let Assumption 6 hold. Then we have*

$$\hat{U}_j \mathcal{H}_j = U_j + \frac{1}{\sqrt{n}} \Psi_j \mathcal{G}_j, \tag{3}$$

where $\mathcal{G}_j = \hat{V}_j \hat{S}_j^{-1} \mathcal{H}_j$ and $\mathcal{H}_j$ satisfies $\mathcal{H}_j' \mathcal{H}_j = I_{r_j} + O_p(\frac{1}{\sqrt{n}})$, for $j = 1, 2$.

In contrast to Assumption 1, the remainder $\frac{1}{\sqrt{n}} \Psi_j \mathcal{G}_j$ in Lemma 5 is not of order $O_p(\frac{1}{\sqrt{n}})$, and the results in Section 3 do not apply. In fact the term $\frac{1}{n} \mathcal{G}_i' \Psi_i' \Psi_j \mathcal{G}_j$ induces a bias in the inner product $\hat{U}_i' \hat{U}_j$. It explains the separate treatment needed for the problem at hand. The correction of this bias will ultimately lead to a slower convergence rate for the k largest sample canonical correlations. The next definition is a modification of Definition 3.

Definition 4 *The sample canonical correlations $\hat{\rho}_l$ between $\hat{\Pi}_1$ and $\hat{\Pi}_2$ are the eigenvalues of $\hat{R}_{1,2} := (\hat{U}_1' \hat{U}_2 - \hat{B}_{1,2})(\hat{U}_2' \hat{U}_1 - \hat{B}_{2,1})$, ordered as $\hat{\rho}_1 \geq \hat{\rho}_2 \geq ... \geq \hat{\rho}_{r_1}$, where $\hat{B}_{i,j} := \hat{S}_i^{-1} \hat{V}_i' \hat{\Phi}_{i,j} \hat{V}_j \hat{S}_j^{-1}$, $j = 1, 2$. Moreover we use $\hat{E}_1$ to denote the $r_1 \times k$ matrix whose orthonormal columns are eigenvectors of $\hat{R}_{1,2}$ corresponding to its k largest eigenvalues.*

The term $\hat{B}_{i,j}$ corrects for the bias in $\hat{U}_i' \hat{U}_j$. By definition we have $\hat{U}_j' \hat{U}_j - \hat{B}_{j,j} = I_{r_j}$, and so $\hat{R}_{1,2} = (\hat{U}_1' \hat{U}_1 - \hat{B}_{1,1})^{-1} \hat{R}_{1,2} (\hat{U}_2' \hat{U}_2 - \hat{B}_{2,2})^{-1}$, i.e. the weighted an unweighted versions of $\hat{R}_{1,2}$ are numerically equivalent. This property is key in deriving the asymptotic expansion of $\hat{\rho}_l$ below.

Proposition 4 *Let Assumption 6 hold and let $\hat{\rho}_l$ be as in Definition 4. Under $H_0(k)$ we have*

$$\hat{\rho}_l = 1 + \lambda_l(-E' \hat{\Psi} E) + O_p(\frac{1}{n}),$$

for $l = 1, \ldots, k$, *where* $E := [E_1' S_1^{-1} V_1' : -E_2' S_2^{-1} V_2']'$, $E_2 := U_2' U_1 E_1$, $\hat{\Psi} = \begin{bmatrix} \hat{\Psi}_{1,1} & \hat{\Psi}_{1,2} \\ \hat{\Psi}_{2,1} & \hat{\Psi}_{2,2} \end{bmatrix}$, *and* $\hat{\Psi}_{i,j} := \frac{1}{n}\Psi_i' \Psi_j - \hat{\Phi}_{i,j}$, $j = 1, 2$. *Moreover, we have* $\hat{U}_1 \hat{E}_1 = U_1 E_1 \mathcal{R}_1 + \frac{1}{\sqrt{n}} \Psi_1 \hat{V}_1 \hat{S}_1^{-1} \hat{E}_1 + O_p(\frac{1}{\sqrt{n}})$, *where* $\mathcal{R}_1$ *satisfies* $\mathcal{R}_1' \mathcal{R}_1 = I_k + O_p(\frac{1}{\sqrt{n}})$.

In contrast to Proposition 2, the asymptotic expansion of $\hat{\rho}_l$ in Proposition 4 features a first order term, which drives the asymptotic distribution. The quadratic form $-E\hat{\Psi}E$ is not always negative definite, which is related to the fact that $\hat{\rho}_l$ is no longer bounded above by 1 due to the bias correction in Definition 4. Note that the definition of E_2 in Proposition 4 corresponds to a specific choice of eigenvectors of $R_{2,1}$ corresponding to eigenvalue 1. This choice ensures convergence of the off-diagonal blocks in EE' (see Lemma 6 below). Squaring and summing the expansion of $\hat{\rho}_l - 1$ in Proposition 4 for $l = 1, \ldots, k$ and using the properties of trace yields the asymptotic expansion of $\zeta(k)$ under $H_0(k)$:

$$\zeta(k) = nTr((E'\hat{\Psi}E)^2) + o_p(1)$$

$$= nvec(\hat{\Psi})'(EE' \otimes EE')vec(\hat{\Psi}) + o_p(1).$$

It remains to establish the asymptotic behavior of $vec(\hat{\Psi})$ and EE'.

Assumption 7 *We have* $\begin{bmatrix} \sqrt{n}vech(\frac{1}{n}\Psi_1'\Psi_1 - \hat{\Phi}_{1,1}) \\ \sqrt{n}vech(\frac{1}{n}\Psi_2'\Psi_2 - \hat{\Phi}_{2,2}) \\ \sqrt{n}vec(\frac{1}{n}\Psi_1'\Psi_2 - \hat{\Phi}_{1,2}) \end{bmatrix} \Rightarrow N(0, \Omega)$ *as* $n \to \infty$.

From Assumption 7 we have $\sqrt{n}vec(\hat{\Psi}) \Rightarrow N(0, \Sigma)$ as $n \to \infty$, where $\Sigma := A\Omega A'$ and

$$A := \left[\begin{bmatrix} I_{k_1} \\ 0 \end{bmatrix} \otimes \begin{bmatrix} I_{k_1} \\ 0 \end{bmatrix} : \begin{bmatrix} 0 \\ I_{k_2} \end{bmatrix} \otimes \begin{bmatrix} 0 \\ I_{k_2} \end{bmatrix} : [I_{k_1 k_2} + K_{k_1,k_2}] \begin{bmatrix} 0 \\ I_{k_2} \end{bmatrix} \otimes \begin{bmatrix} I_{k_1} \\ 0 \end{bmatrix} \right].$$

The convergence of EE' is considered in the next lemma.

Assumption 8 *We have* $\frac{1}{n}\Pi_i'\Pi_j \to \Sigma_{i,j}$ *as* $n \to \infty$, *where* $rank(\Sigma_{j,j}) = r_j$, $j = 1, 2$. *Moreover, the eigenvalues of* $R_{\Sigma,1,2} := \Sigma_{1,1}^{+}\Sigma_{1,2}\Sigma_{2,2}^{+}\Sigma_{2,1}$ *of order greater than* k *are strictly less than* 1.

Let $\bar{V}_j\bar{S}_j^2\bar{V}_j'$ be an eigendecomposition of $\Sigma_{j,j}$, where $\bar{S}_j^2$ $r_j \times r_j$ diagonal containing its nonzero eigenvalues, for $j = 1, 2$. Moreover let $\bar{E} := [\bar{E}_1'\bar{S}_1^{-1}\bar{V}_1' : -\bar{E}_2'\bar{S}_2^{-1}\bar{V}_2']'$, where $\bar{E}_1$ is an $r_1 \times k$ matrix whose orthonormal columns are eigenvectors of $\bar{R}_{1,2} := \bar{S}_1\bar{V}_1'R_{\Sigma,1,2}\bar{V}_1\bar{S}_1^{-1}$ corresponding to eigenvalue 1, and $\bar{E}_2 := \bar{S}_2^{-1}\bar{V}_2'\Sigma_{2,1}\bar{V}_1\bar{S}_1^{-1}\bar{E}_1$[5].

Lemma 6 *Let Assumption 8 hold. Under $H_0(k)$ we have $EE' = \bar{E}\bar{E}' + o_p(1)$.*

By the result on the distribution of quadratic forms of Gaussian vectors, we get the next result.

Proposition 5 *Let Assumption 6-8 hold and let $\hat{\rho}_l$ be as in Definition 4. Under $H_0(k)$ and as $n \to \infty$ we have*

$$\zeta(k) \Rightarrow \sum_{l=1}^{k^2} \lambda_l \chi_l^2,$$

where the χ_l^2 are independent chi-square variables with one degree of freedom, and the λ_l are the eigenvalues of the positive semi-definite matrix $\Xi := (\bar{E} \otimes \bar{E})'\Sigma(\bar{E} \otimes \bar{E})$. Under $H_1(k)$ we have $\zeta(k) \xrightarrow{p} \infty$ as $n \to \infty$.

Next we consider a feasible version of the test. Let $\hat{E} := [\hat{E}_1'\hat{S}_1^{-1}\hat{V}_1' : -\hat{E}_2'\hat{S}_2^{-1}\hat{V}_2']'$, where $\hat{E}_2 := \hat{U}_2'\hat{U}_1\hat{E}_1$.

Lemma 7 *Let Assumption 6 hold. Under $H_0(k)$ we have $\hat{E}\mathcal{O} = \bar{E} + o_p(1)$, where $\mathcal{O}$ is $k \times k$ and satisfies $\mathcal{O}'\mathcal{O} = I_k$. Under $H_1(k)$ we have $\hat{E} = O_p(1)$.*

Let $\hat{\Omega}$ be a consistent estimator of Ω, $\hat{\Sigma} := A\hat{\Omega}A$, and $\hat{\Xi} := (\hat{E} \otimes \hat{E})'\hat{\Sigma}(\hat{E} \otimes \hat{E})$. It follows from Lemma 7 and the Lipschitz continuity of eigenvalues for symmetric matrices that $\lambda_l(\hat{\Xi}) = \lambda_l(\Xi) + o_p(1)$ under $H_0(k)$, amd $\lambda_l(\hat{\Xi}) = O_p(1)$ under $H_1(k)$. Hence we can consistently estimate critical values of the asymptotic distribution of $\zeta(k)$ under $H_0(k)$ by simulating a large number of draws from $\sum_{l=1}^{k^2} \hat{\lambda}_l \chi_l^2$, where $\hat{\lambda}_l$ are the eigenvalues of $\hat{\Xi}$. Under $H_1(k)$ these critical values remain bounded in probability. This yields a feasible version of the test with asymptotic correct size under $H_0(k)$ and asymptotic power 1 under $H_1(k)$.

[5] $\bar{E}_2$ has orthonormal columns which are eigenvectors of $\bar{R}_{2,1} := \bar{S}_2\bar{V}_2'R_{\Sigma,2,1}\bar{V}_2\bar{S}_2^{-1}$ corresponding to eigenvalue 1, where $R_{\Sigma,2,1} := \Sigma_{2,2}^{+}\Sigma_{2,1}\Sigma_{1,1}^{+}\Sigma_{1,2}$.

6 Empirical application

In this section, I test hypotheses about the number of common latent factors between short panels of monthly US stock returns and corporate bond returns. Monthly stock returns are from the Center for Research in Securities Prices (CRSP), and monthly bond returns are from the Wharton Research Data Services (WRDS) bond database [6]. I focus on 2019-2020 period, i.e. the pre-COVID and COVID years. The stock sample consists of US common stocks trading on the NYSE, AMEX or NASDAQ. I exclude financial firms (Standard Industrial Classification codes between 6000 and 6999) and penny stocks (stock price below 5 USD). The bond sample consists of US corporate bonds linked to non-financial companies having at least one US common stock trading on NYSE, AMEX or NASDAQ. [7] I use the bond return variable RET_L5M, which requires an available bond transaction price within the last 5 trading days of the current and previous months. I remove bonds classified as convertible, with maturity less than one year, with total par-value volume less than 10,000 USD, and with price below 5 USD or above 1,000 USD in a given month.

To stay coherent with notation in the previous sections I use Y_1 and Y_2 to denote panels (matrices) of stock or bond excess returns. The columns of Y_1 and Y_2 represent different stocks or different bonds, and the rows represent consecutive months. I consider the case where each panel follows a static factor model:

$$Y_j = F_j \beta_j' + \varepsilon_j, \qquad j = 1, 2, \tag{4}$$

where F_j is the matrix of unobservable factor values, β_j is matrix of factor loadings, and ε_j is the matrix of error terms. I work conditionally on a given realization of the factor paths, i.e. I treat F_1 and F_2 as unknown matrix parameters. Hence only ε_j is random on the right-hand side in (3).

[6] The WRDS bond database is a cleaned database based on the Enhanced Trade Reporting and Compliance Engine (TRACE) data and the Mergent Fixed Income Securities Database (FISD). It was introduced by WRDS in April 2017.

[7] I use the bond linking table from WRDS to link the WRDS bond data to CRSP.

6.1 Common factors in the returns of stocks and bonds

I first consider the case where panels Y_1 and Y_2 represent respectively the cross-sections of stock and bond returns over the same time period:

$$\begin{bmatrix} Y_1 & Y_2 \end{bmatrix} = \begin{bmatrix} F_{1,s} & F_{2,s} & F_c \end{bmatrix} \begin{bmatrix} \beta_{1,s}' & 0 \\ 0 & \beta_{2,s}' \\ \beta_{1,c}' & \beta_{2,c}' \end{bmatrix} + \begin{bmatrix} \varepsilon_1 & \varepsilon_2 \end{bmatrix},$$

where F_c denotes the $T \times k_c$ matrix of common latent factors between Y_1 and Y_2, and $F_{j,s}$ denotes the $T \times k_{j,s}$ matrix of panel specific factors, $j = 1, 2$. I use n_1 and n_2 to denote cross-section size (the number of columns) in Y_1 and Y_2 respectively. The goal is to conduct inference on k_c when T is fixed and $n_1, n_2 \to \infty$. I use the regularity assumptions listed in Appendix A. In particular, I work with spherical errors (Assumption A.3), which allows to consistently estimate latent factors in each panel with PCA even when T is small, see Bai (2003) Theorem 4 and FGS (2022).

Lemma 8 *Let Assumption A.1 hold. Then k_c equals the dimension of $col(F_1\beta_1') \cap col(F_2\beta_2')$.*

Lemma 8 implies that we are in the set-up of Section 4.2, with $\hat{\Pi}_j := Y_j$, $\Pi_j := F_j\beta_j'$, $\Psi_j := \varepsilon_j$, and $\hat{\Phi}_j := \Phi_j := E[\frac{1}{n}\varepsilon_j\varepsilon_j']$ (see Assumption A.3. and the discussion after Lemma 3 regarding the spherical case). Assumption 4 (ii) is satisfied under Assumption A.1-A.2, and Assumption 4 (i) is equivalent to

$$\begin{bmatrix} \sqrt{n_1} vech(Z_1) \\ \sqrt{n_2} vech(Z_2) \\ \sqrt{n_1} vec(W_1) \\ \sqrt{n_2} vec(W_2) \end{bmatrix} \Rightarrow N(0, \Omega) \quad \text{as} \quad n_1, n_2 \to \infty, \tag{5}$$

where $Z_j := \frac{1}{n_j}\varepsilon_j\varepsilon_j' - E[\frac{1}{n}\varepsilon_j\varepsilon_j']$ and $W_j := \frac{1}{n_j}\varepsilon_j\beta_j$, $j = 1, 2$. In Appendix A.1, I verify that this distributional convergence result holds under cross-sectional independence of the error terms and other regularity conditions, and provide a consistent estimator $\hat{\Omega}$ for Ω.

I start by estimating the rank r_j of Π_j (number of factors) using the estimator $\hat{r}_j$ from Appendix B, which satisfies $\hat{r}_j = r_j$ with probability approaching one. Then, I test for the null hypothesis of k common factors between Y_1 and Y_2 using the statistic $\xi(k)$ computed from the sample canonical correlations in Definition

3, with estimators $\hat{U}_j$ computed as in Lemma 2 for the spherical case (see discussion after Lemma 3). The p-value of the test are obtained by simulating 10,000 draws from the distribution in Proposition 3, using estimators $\hat{\Xi}$ and $\hat{\Sigma}$ at the end of Sections 3 and 4.2. In Appendix C I also provide consistent estimators $\hat{\sigma}^2_{j,c}$ and $\hat{\sigma}^2_{j,s}$ for

$$\sigma^2_{j,c} := \frac{Tr(F_c\beta'_{j,c}\beta'_{j,c}F'_c)}{Tr(Y_jY'_j)}, \quad \sigma^2_{j,s} := \frac{Tr(F_c\beta'_{j,s}\beta'_{j,s}F'_c)}{Tr(Y_jY'_j)}, \quad j = 1, 2. \tag{6}$$

The ratio $\sigma^2_{j,c}$ measures the fraction of average cross-sectional variance explained by the common factors in each panel, while $\sigma^2_{j,s}$ measures the fraction explained by the panel specific factors.

Figure 1 shows the results for the year 2019, i.e. the pre-COVID year. The time series dimension is $T = 12$, and the cross-section dimensions are $n_1 = 1,968$ and $n_2 = 2,273$. I find $\hat{r}_1 = 7$ factors driving stock returns and $\hat{r}_2 = 6$ factors driving bond returns during this period. The first panel in Figure 1 shows the $min(\hat{r}_1, \hat{r}_2) = 6$ sample canonical correlations between the two panels. Since $\max(\hat{r}_1 + \hat{r}_2 - T, 0) = 1$ the first sample canonical correlations is equal to 1 by construction (see discussion after Proposition 3). The second panel in Figure 1 shows the statistic $\xi(k)$ scaled by $1/n$. The third panel shows the p-values of the test. Building on the results in Pötscher (1983), we can obtain a consistent estimator of the number of common latent factors between the two panels by allowing the asymptotic size α go to zero as $n_1, n_2 \to \infty$ in the sequential testing procedure. I use the rule $\alpha = \min(10/n_1, 10/n_2)$[8] which gives a size of 0.44%. With this α the third panel of Figure 1 indicate $\hat{k}_c = 5$ common factors during the period. Figure 2 shows the decomposition of the average cross-sectional variance in each panel based on (6). We can observe that around 66% of the cross-sectional variance in stock returns is due to idiosyncratic noise, while for bonds the fraction is around 33% . Moreover, common factors explain around 30% of the cross-sectional variance in stock returns, and around 50% of the cross-sectional variance in bond returns.

Next I repeat the exercise for year 2020, i.e. the COVID year. The results are displayed in Figure 3. The time series dimension is $T = 12$, and the cross-section dimensions are $n_1 = 1,950$ and $n_2 = 2,196$. I find $\hat{r}_1 = 5$ factors driving stock returns and $\hat{r}_2 = 10$ factors driving bond returns during this period. Using the size rule $\alpha = \min(10/n_1, 10/n_2) = 0.46\%$ gives $\hat{k}_c = 4$ common factors during the period. From Figure 4 we can observe that the partition of cross-sectional variance has remained similar to 2019 for stocks. On the other hand, the partition of cross-sectional variance for bonds has changed quite drastically. Bond specific

[8] $\alpha = 10/n$ satisfies the theoretical rule $\log \alpha/n \to 0$ given in Pötscher (1983).

factors now explain more than 33% of the cross-sectional variance, while common factors explain close to 60%. Idiosyncratic noise in bond returns now contributes to only 5% of the cross-sectional variance.

6.2 Stability of the factor structure in the returns of stocks and bonds during COVID

Next I consider the case where Y_1 and Y_2 represent the cross-section of stock or bond returns over two non-overlapping time periods:

$$\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} F_{1,s} & 0 & F_{1,c} \\ 0 & F_{2,s} & F_{2,c} \end{bmatrix} \begin{bmatrix} \beta'_{1,s} \\ \beta'_{2,s} \\ \beta'_c \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix},$$

where $F_{1,c}$ and $F_{2,c}$ are $T_1 \times k_c$ and $T_2 \times k_c$ matrices respectively, and represent paths of the common factors in the two time periods, while $F_{1,s}$ and $F_{2,s}$ are $T_1 \times k_{1,s}$ and $T_2 \times k_{2,s}$ matrices respectively, and represent paths of the period specific factors. Note that common factors have the same loadings in the two periods. Therefore I identify common factors between the two time periods by the stability of their loadings, see also Pelger and Xiong (2022). I use n to denote the cross-section size (number of columns) in Y_1 and Y_2. The goal is to conduct inference on k_c when T_1, T_2 are fixed and $n \to \infty$. I use the regularity assumptions listed in Appendix A. In particular, I work with spherical errors (Assumption A.3), which allows to consistently estimate latent factors in each panel with PCA even when T_1 and T_2 are small, see Bai (2003) Theorem 4 and FGS (2022).

Lemma 9 *Let Assumption A.2 hold. Then k_c equals the dimension of $col(\beta_1 F_1') \cap col(\beta_2 F_2')$.*

Lemma 9 implies that we are in the set-up of Section 5, with $\hat{\Pi}_j := Y_j'$, $\Pi_j := \beta_j F_j'$, $\Psi_j := \varepsilon_j'$, and $\Phi_{i,j} = E[\varepsilon_i \varepsilon_j']$. I consider the estimators $\hat{\Phi}_{i,j} = 0, i \neq j$, and $\hat{\Phi}_{j,j} = \hat{\sigma}_j^2 I_{T_j}$, where $\hat{\sigma}_j^2 = \frac{1}{n(T_j - r_j)} trace(\hat{Q}_j' Y_j Y_j' \hat{Q}_j)$, the orthonormal columns of $\hat{Q}_j$ are eigenvectors of $\frac{1}{n} Y_j Y_j'$ corresponding to its $T - r_j$ smallest eigenvalues, and r_j is the rank of Π_j.

Lemma 10 *Under Assumption A.1-A.5 we have $\hat{\sigma}_j^2 = \frac{1}{n(T_j - r_j)} trace(Q_j' \varepsilon_j \varepsilon_j' Q_j) + O_p(\frac{1}{n})$, where the orthonormal columns of Q_j span the orthogonal complement of $col(F_j)$, $j = 1, 2$.*

It follows from Lemma 7 and the properties of trace that $vech(\frac{1}{n}\Psi_j'\Psi_j - \hat{\Phi}_{j,j}) = X_j vech(\frac{1}{n}\Psi_j'\Psi_j - \Phi_{j,j})$, where $X_j := (I_{T^2} - \frac{1}{T-r_j} vec(I_T)vec(Q_jQ_j')')D_{T_j}$. Hence Assumption 6 (i) is equivalent to

$$\begin{bmatrix} \sqrt{n}X_1 vech(Z_1) \\ \sqrt{n}X_2 vech(Z_2) \\ \sqrt{n}vec(Z_{1,2}) \end{bmatrix} \Rightarrow N(0,\Omega) \quad \text{as} \quad n \to \infty, \tag{7}$$

where $Z_{1,2} := \frac{1}{n}\varepsilon_1\varepsilon_2'$. In Appendix A.2, I verify that this distributional convergence result holds under cross-sectional independence of the error terms and other regularity conditions, and provide a consistent estimator $\hat{\Omega}$ for Ω . Assumption 6 (ii) is met under Assumption A.4, and Assumption 6 (iii) is met under Assumption A.1-A.2. Assumption 6 (iv) is maintained for consistency of the test under the alternative. In Appendix C I also provide consistent estimators $\hat{\sigma}^2_{j,c}$ and $\hat{\sigma}^2_{j,s}$ for

$$\sigma^2_{j,c} := \frac{Tr(F_{j,c}\beta_c'\beta_c F_{j,c}')}{Tr(Y_jY_j')}, \quad \sigma^2_{j,s} = \frac{Tr(F_{j,s}\beta_{j,s}'\beta_{j,s}F_{j,s}')}{Tr(Y_jY_j')}, \quad j = 1, 2, \tag{8}$$

The ratio $\sigma^2_{j,c}$ measures the fraction of average cross-sectional variance explained by the common factors in each period, while $\sigma^2_{j,s}$ measures the fraction explained by the period specific factors.

I start by testing for common factors in stock returns between the years 2019 and 2020, i.e. the pre-COVID year and COVID year. Hence the time series dimensions are $T_1 = 12$ and $T_2 = 12$. The cross-section size is $n = 1,739$ [9]. I use the estimates $\hat{r}_j$ from Section 6.1, namely $\hat{r}_1 = 7$ and $\hat{r}_2 = 5$. The first panel in Figure 5 shows the $min(\hat{r}_1, \hat{r}_2) = 5$ sample canonical correlations between the two panels. The second panel in Figure 5 shows the statistic $\zeta(k)$ scaled by $1/n$. The third panel shows the p-values of the test. Since the p-values are zero for all values of k the test indicates no common factors between the two periods. Figure 6 shows the decomposition of cross-sectional variance in the two periods. We can observe that the partition between systematic and idiosyncratic cross-sectional variance is similar between the two periods.

Next I repeat the exercise for bond returns. The cross-section size is $n = 1,627$ [10]. I use the estimates $\hat{r}_j$ from Section 6.1, namely $\hat{r}_1 = 6$ and $\hat{r}_2 = 10$. Using the size rule $\alpha = 10/n = 0.61\%$ the third

[9] I use only stocks with available returns over the 24 months period from January 2019 to December 2020. Hence the cross-section size is smaller than in Section 6.1

[10] I use only bonds with available returns over the 24 months period from January 2019 to December 2020. Hence the cross-section size is smaller than in Section 6.1.

panel of Figure 7 indicates 2 common factors between the two period. Figure 8 shows the decomposition of cross-sectional variance in the two periods. We can observe that common factors explain around 30% of the cross-sectional varianace in bond returns during the pre-COVID and COVID years.

7 Concluding remarks

In this paper I develop inference tools the number of common latent factors between two panels having a large cross-sectional dimension n and small time series dimension T. I propose an approach that builds on general tests for the dimension of the intersection of two matrix column spaces, where each matrix is observed with noise. An application to large cross-sectional panels of monthly US stock returns and corporate bond returns finds 4-5 common factors during the 2019-2020 period. These common factors explain around 30% of the cross-sectional variance in stock returns and at least 50% of the cross-sectional variance in bond returns. Moreover, I document a structural break in the factor loadings of stock returns between the pre-COVID and COVID years, while for bond returns I find two factors with constant loadings between the pre-COVID and COVID years.

The methodology can be used to address many other relevant questions in asset pricing and other fields. On the theoretical side, on could consider testing for common factors when one set of factors is observable (estimated without noise) and thereby provide a method to compare the factor spaces of statistical and economic factors in short panels, similar to Bai and Ng (2006) and Pelger (2019). Alternatively, one could consider extending the method to test for common latent factors among three or more short panels. I leave these interesting extensions for future research.

References

Al-Sadoon, M., 2017. A unifying theory of tests of rank. Journal of Econometrics 199 (1), 49-62.

Anderson, T. W., 2003. An introduction to multivariate statistical analysis. Wiley.

Ando, T., and Bai, J., 2015. Asset pricing with a general multifactor structure. Journal of Financial Econometrics 13 (3), 556-604.

Ando, T., and Bai, J., 2017. Clustering huge number of financial time series: a panel data approach with high-dimensional predictors and factor structures. Journal of the American Statistical Association, 112 (519), 1182-1198.

Andreou, E., Gagliardini, P., Ghysels, E., and Rubin, M., 2019. Inference in group factor models with an application to mixed-frequency data. Econometrica, 87 (4), 1267-1305.

Bai, J., 2003. Inferential theory for factor models of large dimensions. Econometrica 71 (1), 135-171.

Bai, J., and Ng, S., 2002. Determining the number of factors in approximate factor models. Econometrica 70 (1), 191-221.

Breitung, J., and Eickmeier, S., 2016. Analyzing international business and financial cycles using multi-level factor models: a comparison of alternative approaches. Advances in Econometrics, 35, 177-214.

Bai, J., and Ng, S., 2006. Evaluating latent and observed factors in macroeconomics and finance. Journal of Econometrics 131 (1), 507-537.

Chamberlain, G., and Rothschild, M., 1993. Arbitrage, factor structure, and mean-variance analysis on large asset markets. Econometrica, 51 (5), 1281-1304.

Chen, P., 2010. A Grouped Factor Model. MPRA Paper No. 36082.

Chen, P., 2012. Common Factors and Specific Factors. Working Paper.

Choi, I., Lin, R., and Shin, Y., 2021. Canonical correlation-based model selection for the multilevel factors. Journal of Econometrics 233 (1), 22-44.

Cragg, J., and Donald, S., 1996. On the asymptotic properties of LDU-based tests of the rank of a matrix. Journal of the American Statistical Association 91 (435), 1301-1309.

Fortin, A.-P., Gagliardini, P., and Scaillet, O., 2022. Eigenvalue tests for the number of latent factors in short panels. Journal of Financial Econometrics, forthcoming.

Fortin, A.-P., Gagliardini, P., and Scaillet, O., 2023. Latent Factor Analysis in Short Panels. Working Paper.

Han, X., 2021. Shrinkage estimation of factor models with global and group-specific factors. Journal of Business & Economic Statistics, 39 (1), 1-17.

Kleibergen, F., and Paap, R., 2006. Generalized reduced rank tests using the singular value decomposition. Journal of Econometrics 133 (1), 97-126.

Magnus, J., and Neudecker, H., 2007. Matrix differential calculus, with applications in statistics and econometrics. Wiley.

Pelger, M., 2019. Large-dimensional factor modeling based on high-frequency observations. Journal of Econometrics, 208 (1), 23-42.

Pelger, M., and Xiong, R. 2022. State-varying factor models of large dimensions. Journal of Business & Economic Statistics, 40 (3), 1315-1333.

Pötscher, B., 1983. Order estimation in ARMA-models by Lagrangian multiplier tests, Annals of Statistics 11 (3), 872-885.

Robin, J.-M., and Smith, R., 2000. Tests of rank. Econometric Theory 16 (2), 151-175.

Stock, J., and Watson, M., 2002. Forecasting using principal components from a large number of predictors. Journal of the American Statistical Association 97 (460), 1167-1179.

Wang, P., 2012. Large dimensional factor models with a multi-level factor structure: identification, estimation, and inference. Working Paper.

Figure 1: The first panel displays the sample canonical correlations. The second panel displays the statistic $\xi(k)$ scaled by $\frac{1}{n}$. The third panel displays the p-values of the test. The period is January 2019 to December 2019, i.e. pre-COVID year.

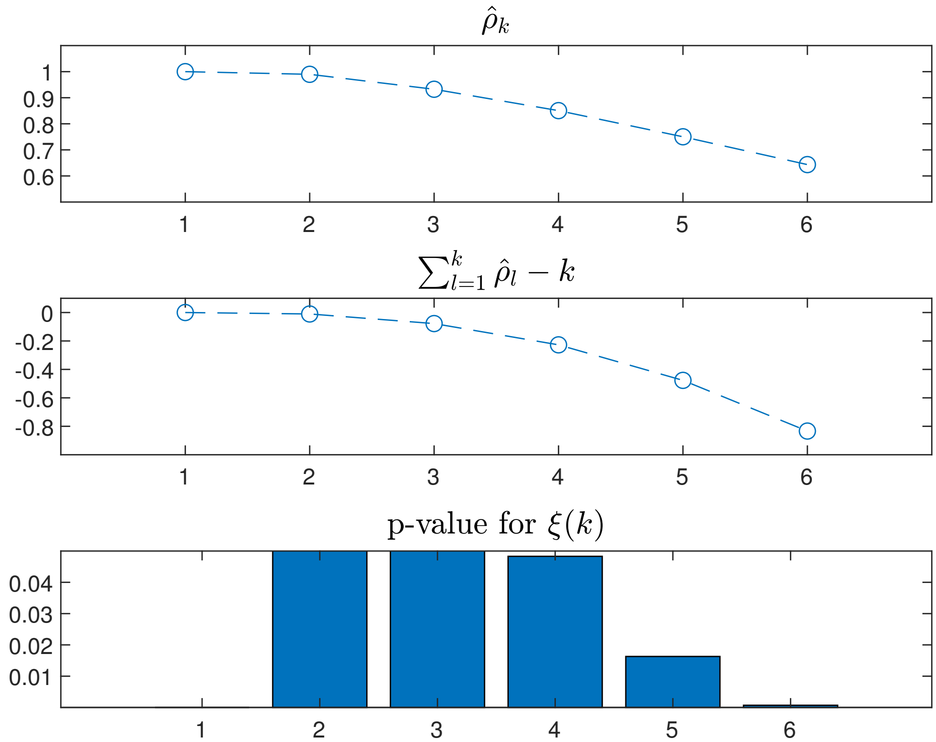


Figure 2: The figure displays the partition of cross-sectional variance for stock and bond returns. The period is January 2019 to December 2019, i.e. pre-COVID year.

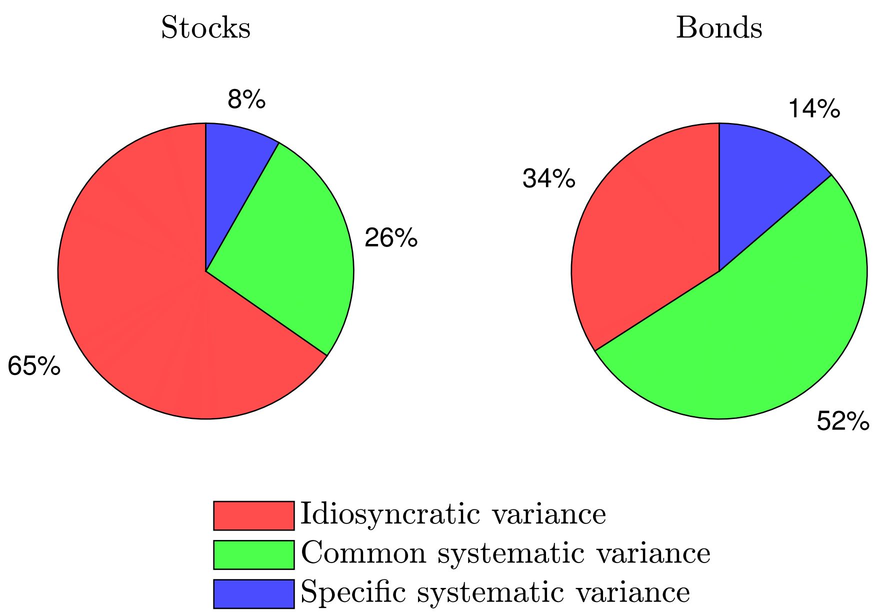


Figure 3: The first panel displays the sample canonical correlations. The second panel displays the statistic $\xi(k)$ scaled by $\frac{1}{n}$. The third panel displays the p-values of the test. The period is January 2020 to December 2020, i.e. COVID year.

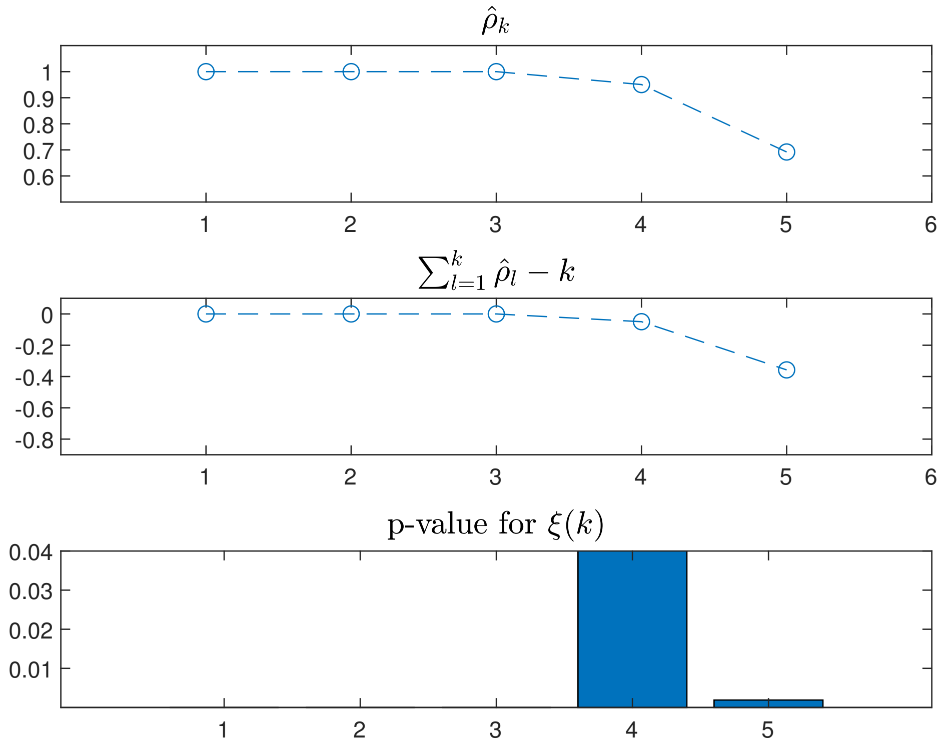


Figure 4: The figure displays the partition of cross-sectional variance for stock and bond returns. The period is January 2020 to December 2020, i.e. COVID year.

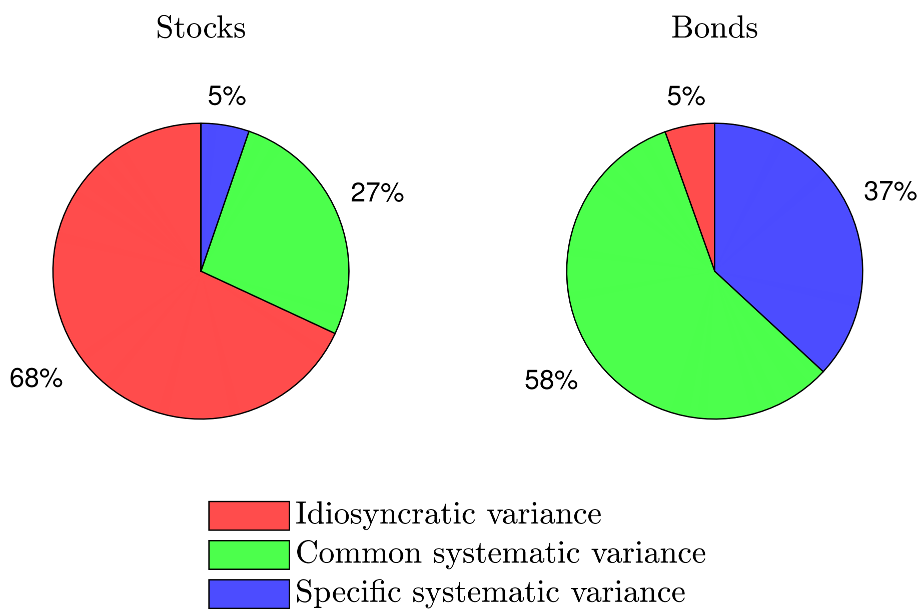


Figure 5: The first panel displays the sample canonical correlations. The second panel panel displays the statistic $\zeta(k)$ scaled by $\frac{1}{n}$. The third panel displays the p-values of the test. The two periods are January 2019 to December 2019 (pre-COVID year) and January 2020 to December 2020 (COVID year). The results are based on the cross-section of stock returns.

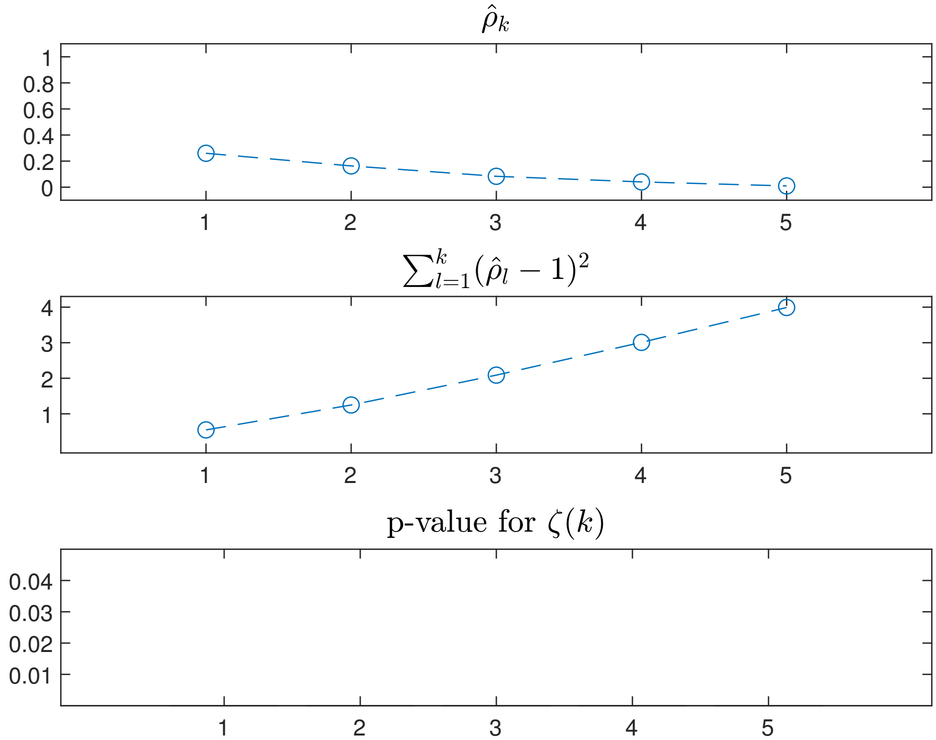


Figure 6: The figure displays the partition of cross-sectional variance for the two periods: January 2019 to December 2019 (pre-COVID year) and January 2020 to December 2020 (COVID year). The results are based on the cross-section of stock returns.

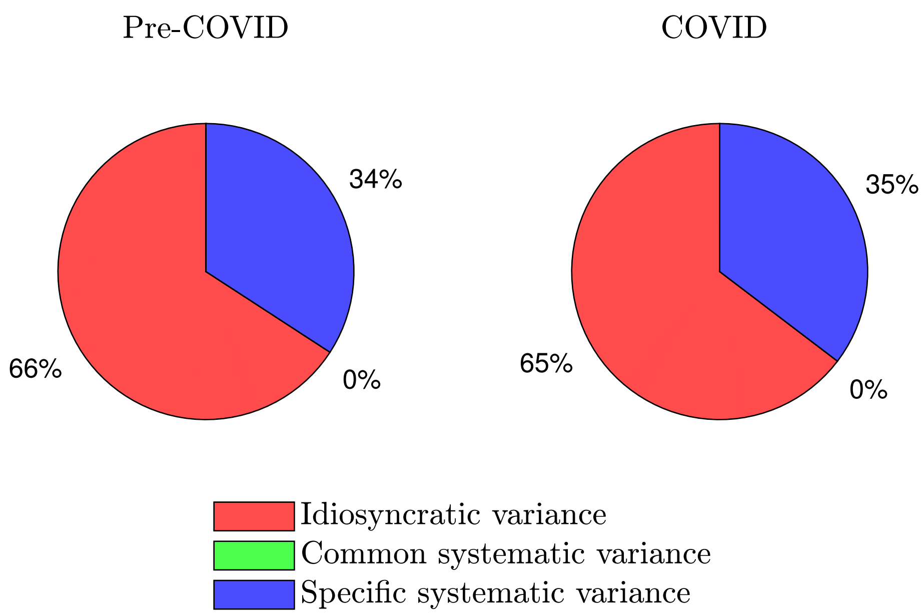


Figure 7: The first panel displays the sample canonical correlations. The second panel panel displays the statistic $\zeta(k)$ scaled by $\frac{1}{n}$. The third panel displays the p-values of the test. The two periods are January 2019 to December 2019 (pre-COVID year) and January 2020 to December 2020 (COVID year). The results are based on the cross-section of bond returns.

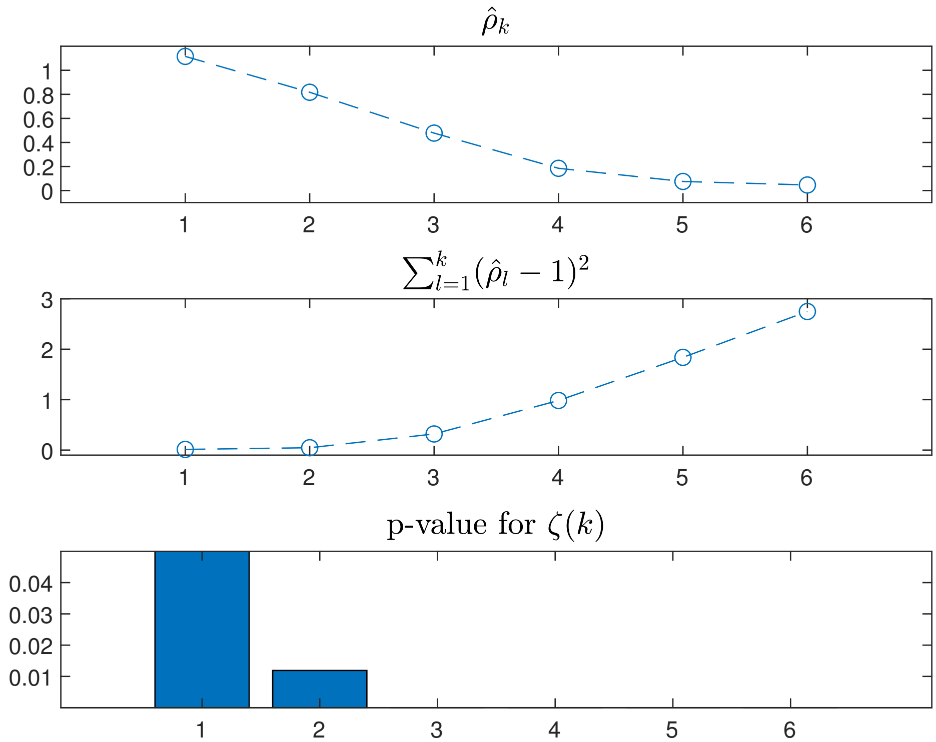
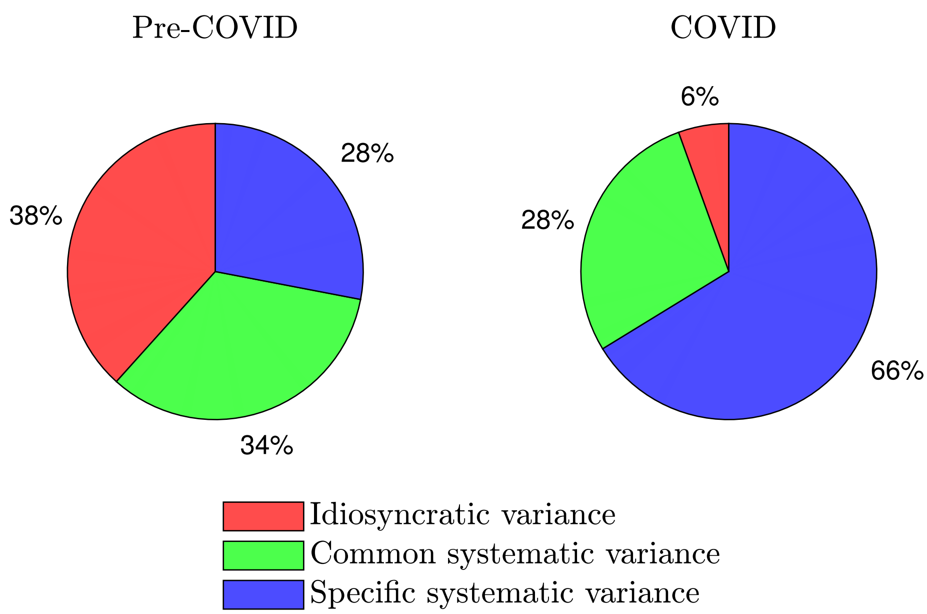


Figure 8: The figure displays the partition of cross-sectional variance for the two periods: January 2019 to December 2019 (pre-COVID year) and January 2020 to December 2020 (COVID year). The results are based on the cross-section of bond returns.



Proof of Proposition 1

Suppose $H_0(k)$ holds. Let U_k be a $T \times k$ matrix with orthonormal columns spanning $\text{col}(\Pi_1) \cap \text{col}(\Pi_2)$. Let Q_1 be a $T \times (r_1 - k)$ matrix with orthonormal columns spanning the orthogonal complement $\text{col}(\Pi_1) \cap \text{col}(\Pi_2)$ in $\text{col}(\Pi_1)$, and similarly for Q_2. Since $\text{col}(\Pi_1) = \text{col}(U_1)$ and $\text{col}(\Pi_2) = \text{col}(U_2)$ there exist orthonormal matrices O_1 and O_2 such that

$$[U_k \, Q_1] = U_1 O_1$$

$$[U_k \, Q_2] = U_2 O_2$$

Therefore,

$$R = O_1 \begin{bmatrix} I_k & 0 \\ 0 & Q_1' Q_2 Q_2' Q_1 \end{bmatrix} O_1'$$

and so the eigenvalues of R are those of I_k and $Q_1' Q_2 Q_2' Q_1$. Now, if x, y are unit vectors then by Cauchy-Schwarz we have $x' Q_1' Q_2 y \leq 1$, with equality holding if and only if $Q_1 x$ and $Q_2 y$ are linearly dependent. The latter case is impossible because otherwise there would exists a nonzero vector lying in $\text{col}(U_k) \cap \text{col}(Q_1) = \{0\}$. It follows that the singular values of $Q_1' Q_2$ are all stricly less than 1. The conclusion follows.

Proof of Proposition 2

Let us define $\tilde{\Psi}_j := \hat{U}_j \mathcal{H}_j - U_j$ and $\tilde{U}_j := U_j + \tilde{\Psi}_j$, so that (2) reads $\hat{U}_j \mathcal{H}_j = \tilde{U}_j$. Substituting into the expression for $\hat{R}_{1,2}$ in Definition 3 and using that $\hat{U}_j' \hat{U}_j = I_{r_j}$ and that $\mathcal{H}_j$ is invertible with probability approaching 1, we get

$$\hat{R}_{1,2} = (\hat{U}_1' \hat{U}_1)^{-1} \hat{U}_1' \hat{U}_2 (\hat{U}_2' \hat{U}_2)^{-1} \hat{U}_2' \hat{U}_1 = \mathcal{H}_1 \tilde{R}_{1,2} \mathcal{H}_1^{-1} \tag{9}$$

where $\tilde{R}_{1,2} := (\tilde{U}_1' \tilde{U}_1)^{-1} \tilde{U}_1' \tilde{U}_2 (\tilde{U}_2' \tilde{U}_2)^{-1} \tilde{U}_2' \tilde{U}_1$. It follows that $\hat{R}_{1,2}$ and $\tilde{R}_{1,2}$ have the same eigenvalues. Note that $\tilde{R}_{1,2}$ is not symmetric, but it can be expanded around the symmetric matrix $R_{1,2}$.

Lemma 11 *Let* $X_{i,j} := U_i'\tilde{\Psi}_j + \tilde{\Psi}_i'U_j + \tilde{\Psi}_i'\tilde{\Psi}_j$ *and suppose that* $\|X_{i,j}\| = O_p(\rho)$*,* $i, j = 1, 2$*, where* $\rho \downarrow 0$*. Then*

$$\tilde{R}_{1,2} = R_{1,2} + \tilde{\Psi}_I + \tilde{\Psi}_{II} + O_p(\rho^3),$$

where

$$\tilde{\Psi}_I = -U_1'M_2\tilde{\Psi}_{1,2} - \tilde{\Psi}_{1,2}'M_2U_1 - \tilde{\Psi}_{1,2}'\tilde{\Psi}_{1,2} + X_{1,1}(I_{r_1} - R_{1,2}) = O_p(\rho),$$
$$\tilde{\Psi}_{II} = (\tilde{\Psi}_{1,2}'(U_2 + \Psi_2) + U_1'M_2\Psi_2)(\tilde{\Psi}_{1,2}'(U_2 + \Psi_2) + U_1'M_2\Psi_2)'$$
$$+X_{1,1}(U_1'M_2\tilde{\Psi}_{1,2} + \tilde{\Psi}_{1,2}'M_2U_1 + \tilde{\Psi}_{1,2}'\tilde{\Psi}_{1,2}) = O_p(\rho^2),$$
$$\tilde{\Psi}_{1,2} = \tilde{\Psi}_1 - \tilde{\Psi}_2U_2'U_1,$$

and $M_2 = I_m - U_2U_2'$*.*

Let $\tilde{P}$ be an $r_1 \times r_1$ matrix whose columns are eigenvectors of $\tilde{R}_{1,2}$. Since $\tilde{R}_{1,2}$ is not symmetric we cannot choose $\tilde{P}$ to have orthonormal columns in general. However from (10) we see that we can choose $\tilde{P} = \mathcal{H}_1^{-1}\hat{P}$, where $\hat{P}$ is the $r_1 \times r_1$ matrix whose orthonormal columns are eigenvectors of $\hat{R}_{1,2}$. By Assumption 1 this choice satisfies $\tilde{P}'\tilde{P} = I_{r_1} + O_p(\frac{1}{\sqrt{n}})$. Note also that Assumption 1 implies that we can take $\rho = \frac{1}{\sqrt{n}}$ in Lemma 8. In particular, using $\mathcal{H}_1'\mathcal{H}_1 = I_{r_1} + O_p(\frac{1}{\sqrt{n}})$ and the Lipschitz continuity of eigenvalues for symmetric matrices, we get

$$\lambda_l(\tilde{R}_{1,2}) = \lambda_l(\hat{R}_{1,2}) = \lambda_l(R_{1,2}) + O_p(\frac{1}{\sqrt{n}}), \quad l = 1, \ldots, r_1.$$

We are now in a position to apply the general eigenvalue expansion result in Proposition 6, with $A := \tilde{R}_{1,2}$ and $\Psi := \tilde{\Psi}_I + \tilde{\Psi}_{II} + O_p(\frac{1}{n^{3/2}}) = O_p(\frac{1}{\sqrt{n}})$. We get

$$\hat{D}_k = I_k + \mathcal{R}^{-1}\big(E_1'\Psi E_1 + E_1'\Psi W_1(I_{m-k} - \Gamma)^{-1}W_1'\Psi E_1\big)\mathcal{R} + O_p(\frac{1}{n^{3/2}}) \tag{10}$$
$$\hat{E}_1 = \mathcal{H}_1 E_1 \mathcal{R} + O_p(\frac{1}{\sqrt{n}}) \tag{11}$$

where $\hat{D}_k$ is $k \times k$ diagonal containing the k largest eigenvalues of $\hat{R}_{1,2}$ and $\hat{E}_1$ is $r_1 \times (r_1 - k)$ with orthonormal columns representing the corresponding eigenvectors, E_1 and W_1 are defined in Proposition 1, Γ is $(r_1 - k) \times (r_1 - k)$ diagonal containing the $r_1 - k$ smallest eigenvalues of $R_{1,2}$, and $\mathcal{R}$ satisfies $\mathcal{R}'\mathcal{R} = I_k + O_p(\frac{1}{\sqrt{n}})$.

Lemma 12 *Under Assumption 1 we have* $E_1'\Psi E_1 + E_1'\Psi W_1(I_{m-k}-\Gamma)^{-1}W_1'\Psi E_1 = -E_1'\hat{\Psi}_{1,2}'\Xi\hat{\Psi}_{1,2}E_1 + O_p(\frac{1}{n^{3/2}})$*, where* $\hat{\Psi}_{1,2}$ *and* Ξ *are defined in Proposition* 2*. Moreover,* Ξ *has the properties stated in Proposition 2.*

The first part of Proposition 2 now follows by substituting the result of Lemma 9 into (11), and using $\hat{\Psi}_{1,2} = O_p(\frac{1}{\sqrt{n}})$ (from Assumption 1), $\mathcal{R}'\mathcal{R} = I_k + O_p(\frac{1}{\sqrt{n}})$, and the Lipschitz continuity of eigenvalues for symmetric matrices. The second part of Proposition 2 follows from $\hat{U}_1\hat{E}_1 = \tilde{U}_1\mathcal{H}_1^{-1}\hat{E}_1 = U_1E_1\mathcal{R} + \frac{1}{\sqrt{n}}\Psi_1\hat{V}_1\hat{S}_1^{-1}\hat{E}_1 + O_p(\frac{1}{\sqrt{n}})$.

Proof of Proposition 4

Let us define $\tilde{\Psi}_j := \hat{U}_j\mathcal{H}_j - U_j$ and $\tilde{U}_j := U_j + \tilde{\Psi}_j$, so that (3) reads $\hat{U}_j\mathcal{H}_j = \tilde{U}_j$. Substituting into the expression for $\hat{R}_{1,2}$ in Definition 4, and using that $\hat{U}_j'\hat{U}_j - \hat{B}_{j,j} = I_{r_j}$ and that $\mathcal{H}_j$ is invertible with probability approaching 1, we get

$$\hat{R}_{1,2} = (\hat{U}_1'\hat{U}_1 - \hat{B}_{1,1})^{-1}(\hat{U}_1'\hat{U}_2 - \hat{B}_{1,2})(\hat{U}_2'\hat{U}_1 - \hat{B}_{2,1})(\hat{U}_2'\hat{U}_2 - \hat{B}_{2,2})^{-1} = \mathcal{H}_1\tilde{R}_{1,2}\mathcal{H}_1^{-1}, \qquad (12)$$

where $\tilde{R}_{1,2} := (\tilde{U}_1'\tilde{U}_1 - \tilde{B}_{1,1})^{-1}(\tilde{U}_1'\tilde{U}_2 - \tilde{B}_{1,2})(\tilde{U}_2'\tilde{U}_1 - \tilde{B}_{2,1})(\tilde{U}_2'\tilde{U}_2 - \tilde{B}_{2,2})^{-1}$ and $\tilde{B}_{i,j} := \mathcal{G}_i'\hat{\Phi}_{i,j}\mathcal{G}_j$. Note that $\tilde{R}_{1,2}$ is not symmetric, but it can be expanded around the symmetric matrix $R_{1,2}$.

Lemma 13 *Let* $X_{i,j} := S_i^{-1}V_i'(\frac{1}{n}\Pi_i'\Psi_j + \frac{1}{n}\Psi_i'\Pi_j + \frac{1}{n}\Psi_i'\Psi_j - \hat{\Phi}_{i,j})V_jS_j^{-1}$*. Under Assumption 6 we have*

$$\tilde{R}_{1,2} = R_{1,2} + \tilde{\Psi}_I + O_p(\frac{1}{n})$$

where $\tilde{\Psi}_I = -X_{1,1}R_{1,2} - U_1'U_2X_{2,2}U_2'U_1 + X_{1,2}U_2'U_1 + U_1'U_2X_{2,1}$.

Let $\tilde{P}$ be an $r_1 \times r_1$ matrix whose columns are eigenvectors of $\tilde{R}_{1,2}$. Since $\tilde{R}_{1,2}$ is not symmetric we cannot choose $\tilde{P}$ to have orthonormal columns in general. However from (14) we see that we can choose $\tilde{P} = \mathcal{H}_1^{-1}\hat{P}$, where $\hat{P}$ is the $r_1 \times r_1$ matrix whose orthonormal columns are eigenvectors of $\hat{R}_{1,2}$. Since $\mathcal{H}_1'\mathcal{H}_1 = I_{r_1} + O_p(\frac{1}{\sqrt{n}})$ (Lemma 5) this choice satisfies $\tilde{P}'\tilde{P} = I_{r_1} + O_p(\frac{1}{\sqrt{n}})$. Besides, using Lemma 13, $\mathcal{H}_1'\mathcal{H}_1 = I_{r_1} + O_p(\frac{1}{\sqrt{n}})$, and the Lipschitz continuity of eigenvalues for symmetric matrices, we get

$$\lambda_l(\tilde{R}_{1,2}) = \lambda_l(\hat{R}_{1,2}) = \lambda_l(R_{1,2}) + O_p(\frac{1}{\sqrt{n}}), \quad l = 1, \ldots, r_1.$$

We are now in a position to apply the general eigenvalue expansion result in Proposition 6, with $A := \tilde{R}_{1,2}$ and $\Psi := \tilde{\Psi}_I + O_p(\frac{1}{n}) = O_p(\frac{1}{\sqrt{n}})$. We get

$$\hat{D}_k = I_k + \mathcal{R}^{-1} E_1' \Psi E_1 \mathcal{R} + O_p(\frac{1}{n}) \tag{13}$$

$$\hat{E}_1 = \mathcal{H}_1 E_1 \mathcal{R} + O_p(\frac{1}{\sqrt{n}}) \tag{14}$$

where $\hat{D}_k$ is $k \times k$ diagonal containing the k largest eigenvalues of $\hat{R}_{1,2}$,$\hat{E}_1$ is $r_1 \times (r_1 - k)$ with orthonormal columns representing the corresponding eigenvectors, and $\mathcal{R}$ satisfies $\mathcal{R}'\mathcal{R} = I_k + O_p(\frac{1}{\sqrt{n}})$.

Lemma 14 *We have $E_1' \Psi E_1 = -E' \hat{\Psi} E + O_p(\frac{1}{n})$, where E and $\hat{\Psi}$ are defined in Proposition 4.*

The first part of Proposition 4 now follows by substituting the result of Lemma 14 into (15), and using $\Psi = O_p(\frac{1}{\sqrt{n}})$ (from Assumption 6), $\mathcal{R}'\mathcal{R} = I_k + O_p(\frac{1}{\sqrt{n}})$, and the Lipschitz continuity of eigenvalues for symmetric matrices. The second part of Proposition 4 follows from $\hat{U}_1 \hat{E}_1 = \tilde{U}_1 \mathcal{H}_1^{-1} \hat{E}_1 = U_1 E_1 \mathcal{R}$, using (16) , $\tilde{U}_1 = U_1 + \frac{1}{\sqrt{n}} \Psi_1 \mathcal{G}_1$, and $\mathcal{G}_1 = \hat{V}_1 \hat{S}_1^{-1} \mathcal{H}_1$.

A Feasible central limit theorems

In this section we provide feasible central limit theorems for (6) and (8).

Assumption A.1 $rank(F_j) = r_j$, $j = 1, 2$.

Assumption A.2 $\Sigma_{\beta,j} := \lim_{n_j \to \infty} \frac{1}{n_j} \beta_j' \beta_j$ *is positive definite,* $j = 1, 2$.

Assumption A.3 *The diagonal elements in $E[\frac{1}{n_j} \varepsilon_j \varepsilon_j']$ are equal, and the off-diagonal elements in $E[\frac{1}{n_j} \varepsilon_j \varepsilon_j']$ are $o_p(\frac{1}{\sqrt{n_j}})$, $j = 1, 2$.*

Assumption A.4 $W_j := \frac{1}{n_j} \varepsilon_j \beta_j = O_p(\frac{1}{\sqrt{n_j}})$ *and* $Z_j := \frac{1}{n_j} \varepsilon_j \varepsilon_j' - E[\frac{1}{n} \varepsilon_j \varepsilon_j'] = O_p(\frac{1}{\sqrt{n_j}})$, $j = 1, 2$.

Assumption A.5 $\frac{n_1}{n_2} \to \mu > 0$.

A.1

A.2

C Rank selection

In this Appendix we provide selection criteria for the rank of an $m \times k$ matrix Π when a noisy proxy is available:

$$\hat{\Pi} = \Pi + \Psi$$

The criteria exploit separation of singular values. Let r denote the true rank of Π, and let $g(n)$ denote a generic penality function satisfying the following two conditions as $n \to \infty$:

1. $g(n) \to 0$
2. $ng(n) \to \infty$

Moreover we use $\sigma_l(A)$ to denote the lth largest singular value of an $m \times n$ matrix A, and $\|A\|^2_{>s} = \sum_{l=s+1}^{\min\{m,n\}} \sigma_l^2(A)$ to denote the sum of its singular values of order higher than s.

C.1

Let us first consider the setting where m is fixed and m, k are fixed and $\|\Psi\| = O_p(\frac{1}{\sqrt{n}})$. Then we have the expansion $\hat{\Pi} = \Pi + O_p(\frac{1}{\sqrt{n}})$. From the Lipschitz continuity of singular values and the fact that rank$(\Pi) = r$ we get

$$\sigma_l(\hat{\Pi}) = \begin{cases} \sigma_l(\Pi) + O_p(\frac{1}{\sqrt{n}}), & l = 1, \dots r \\ O_p(\frac{1}{\sqrt{n}}), & l > r \end{cases}$$

In particular $\|\hat{\Pi}\|^2_{>r} = O_p(\frac{1}{n})$. Let $\Delta_l = \|\hat{\Pi}\|^2_{>l} - g(n)$. Then $\Delta_l > 0$ with probability approaching 1 for $l < r$, whereas $\Delta_l < 0$ with probability approaching 1 for $l = r$. It follows that $\hat{r} := \min\{0 \leq l \leq k : \Delta_l < 0\}$ statisfies $\hat{r} = r$ with probability approaching 1.

C.2

Next we consider the setting where m is fixed and $k = n \to \infty$. We assume that we have estimators $\hat{\Phi}(l)$, $0 \le l \le r$, and that the following conditions are satisfied

1. $\frac{1}{n}\Psi\Psi' - \hat{\Phi}(r) = O_p(\frac{1}{\sqrt{n}})$.

2. $\frac{1}{n}\Psi\Pi' = O_p(\frac{1}{\sqrt{n}})$.

3. $\|\frac{1}{n}\hat{\Pi}\hat{\Pi}' - \hat{\Phi}(l)\|^2_{>l} > c$ with probability approaching 1 for some $c > 0$ and all $l < r$.

Using 1 and 2 above we get the expansion $\frac{1}{n}\hat{\Pi}\hat{\Pi}' - \hat{\Phi}(r) = \frac{1}{n}\Pi\Pi' + O_p(\frac{1}{\sqrt{n}})$. Using the the Lipschitz continuity of singular values and the fact that $\text{rank}(\Pi) = r$ for n sufficiently large we get $\sigma_l(\frac{1}{n}\hat{\Pi}\hat{\Pi}' - \hat{\Phi}(r)) = O_p(\frac{1}{\sqrt{n}})$ for $l > r$. Hence $\|\frac{1}{n}\hat{\Pi}\hat{\Pi}' - \hat{\Phi}(r)\|^2_{>r} = O_p(\frac{1}{n})$. Let $\Delta_l = \|\frac{1}{n}\hat{\Pi}\hat{\Pi}' - \hat{\Phi}(l)\|^2_{>l} - g(n)$. Then $\Delta_l > 0$ with probability approaching 1 for $l < r$ (by 3 above), whereas $\Delta_l < 0$ with probability approaching 1 for $l = r$. It follows that $\hat{r} := \min\{0 \le l \le m : \Delta_l < 0\}$ satisfies $\hat{r} = r$ with probability approaching 1.

C.3

We consider again the case where m is fixed and $k = n \to \infty$. We assume that instrumental variables are availabe, i.e. there exists $n \times K$ matrix Z satisfying the following conditions, where $\Gamma := \frac{1}{n}\Pi Z$.

1. $\frac{1}{n}\Psi Z = O_p(\frac{1}{\sqrt{n}})$

2. $\text{col}(\Gamma) = \text{col}(\Pi)$ for n sufficiently large.

3. $\sigma_r(\Gamma) > c$ for some $c > 0$ and n sufficiently large.

From 1 we have the expansion $\hat{\Gamma} = \Gamma + O_p(\frac{1}{\sqrt{n}})$. Using Lipschitz continuity of singular values and the fact that $\text{rank}(\Gamma) = r$ for n sufficiently large (by 2) we get

$$\sigma_l(\hat{\Gamma}) = \begin{cases} \sigma_l(\Gamma) + O_p(\frac{1}{\sqrt{n}}), & l = 1, \ldots, r \\ O_p(\frac{1}{\sqrt{n}}), & l > r \end{cases}$$

Then we can use the same estimator as in C.1, replacing $\hat{\Pi}$ with $\hat{\Gamma}$ and using condition 3 to get $\Delta_l > 0$ with probability approaching 1 for $l < r$.

C.4

Finally we consider the case where $m = n \to \infty$ and k is fixed. Since rank(Π) = rank(Π′) this case reduces to C.2.

D Eigenvalue expansion

Proposition 6 *Let A be an $m \times m$ symmetric matrix whose k largest eigenvalues are exactly 1, while the other eigenvalues are strictly less than 1. Let $\hat{A} = A + \Psi$, where Ψ is a (not necessarily symmetric) perturbation such that $\|\Psi\| = O_p(\frac{1}{\sqrt{n}})$. Moreover suppose that there exist $m \times m$ matrix $\hat{P}$ and $m \times m$ diagonal matrix $\hat{D}$ satisfying*

$$\hat{A}\hat{P} = \hat{P}\hat{D}$$

$$\hat{P}'\hat{P} = I_m + O_p(\frac{1}{\sqrt{n}})$$

$$\hat{D} = \begin{bmatrix} I_k & 0 \\ 0 & \Gamma \end{bmatrix} + O_p(\frac{1}{\sqrt{n}})$$

where Γ is $(m-k) \times (m-k)$ diagonal containing the $m-k$ eigenvalues of A strictly less than 1. Then

$$\hat{D}_k = I_k + \mathcal{R}^{-1}\big(E'\Psi E + E'\Psi W(I_{m-k} - \Gamma)^{-1}W'\Psi E\big)\mathcal{R} + O_p(\frac{1}{n^{3/2}}) \tag{15}$$

$$\hat{E} = E\mathcal{R} + O_p(\frac{1}{\sqrt{n}}) \tag{16}$$

where $\hat{E}$ is $m \times k$ containing the first k columns of $\hat{P}$, $\mathcal{R}$ is a $k \times k$ matrix satisfying $\mathcal{R}'\mathcal{R} = I_k + O_p(\frac{1}{\sqrt{n}})$, E is an $m \times k$ matrix whose orthonormal columns are eigenvectors of A corresponding to eigenvalue 1, and W is an $m \times (m-k)$ matrix whose orthonormal columns are eigenvectors of A corresponding to eigenvalues strictly less than 1.

Proof of Proposition 6