Spurious cross-sectional dependence in credit spread changes

Marcin Jaskowski\textsuperscript{a}, Michael McAleer\textsuperscript{b,c,d,e,f,∗}

\textsuperscript{a}Faculty of Mathematics and Computer Science, Jagiellonian University, Krakow, Poland
\textsuperscript{b}Department of Finance, Asia University, Taiwan
\textsuperscript{c}Discipline of Business Analytics, University of Sydney Business School, Australia
\textsuperscript{d}Econometric Institute, Erasmus School of Economics Erasmus University Rotterdam, the Netherlands
\textsuperscript{e}Department of Economic Analysis and ICAE, Complutense University of Madrid, Spain
\textsuperscript{f}Institute of Advanced Sciences, Yokohama National University, Japan

\textbf{Article history:}
Received 9 February 2019
Revised 16 September 2019
Accepted 16 September 2019
Available online 14 October 2019

\textbf{JEL classification:}
G12
G13
G17
E43

\textbf{Keywords:}
Credit spread puzzle
Market segmentation
Latent factors
Spurious cross-sectional dependence

\textbf{A B S T R A C T}

In order to understand the lingering credit risk puzzle and the apparent segmentation of the stock market from credit markets, we need to be able to assess the strength of the cross-sectional dependence in credit spreads. This turns out to be a non-trivial task due to the extreme data sparsity that is typical for any panel of credit spreads that is extracted from corporate bond transactions. The problem of data sparsity has led to some erroneous conclusions in the literature, including inferences that have been drawn from spurious cross-sectional dependence in credit spread changes. Understanding the pitfalls leads to improved estimation of the latent factor in credit spread changes and its characteristics.

© 2019 EcoSta Econometrics and Statistics. Published by Elsevier B.V. All rights reserved.

1. Introduction

In a seminal paper, Collin-Dufresne et al. (2001) (hereafter CGM) found a strong common component in the changes in credit spreads of firms that is not accounted for by changes in their stock prices and other controls. Explaining this finding has proved to be challenging, and remains an active focus of research in the credit risk literature. The existence of such a latent factor amounts to evidence for a strong segmentation of credit markets from stock markets. However, it will be shown that a search for a dominant latent factor in the credit spreads has been misguided for two common conceptual reasons, one statistical and the other economic, that has led to the inflated strength of the latent factor. We also find evidence for market segmentation, but it is weaker and more nuanced than in the extant literature.

CGM use a regression model with different regressors that are implied by structural models, such as firm-specific, macro and liquidity measures, on the changes in credit spreads of corporate bonds. They found that more than one-half of the variation is unexplained using their regressors. However, a small \(R^2\) does not seem to be surprising to practitioners or to academics because it is not uncommon in the social sciences, and may reflect a high level of idiosyncratic noise.

\textsuperscript{*} Corresponding author at: Department of Finance, Asia University, Taiwan.
E-mail address: michael.mcaleer@gmail.com (M. McAleer).

https://doi.org/10.1016/j.eosta.2019.09.001
2452-3062/© 2019 EcoSta Econometrics and Statistics. Published by Elsevier B.V. All rights reserved.
What is puzzling in CGM is the existence of a strong and unexplained principal component in the residuals. Additionally, the eigenvector of the main principal component had almost equally-sized elements. Moreover, in a recent study of the European corporate bond market, Castagnetti and Rossi (2013) found evidence of a very strong common unobserved factor in the residuals from time series and panel regressions on credit spread changes.

These results are not supportive of the structural models as they show a misspecified description of credit spreads, and consequently of the capital structure. CGM interpreted their findings as evidence for segmentation of bond and equity markets, so that different investors would trade in stocks rather than in bonds, with prices in both markets driven by independent supply and demand shocks. It is not easy to find convincing justification for such segmentation of the markets, and to explain why equity and bonds should react differently to the same aggregate factors. A practical implication of CGM is that structural models of credit risk are severely misspecified, and hence are untrustworthy and unreliable. Moreover, the unobserved latent factor seems to have a simple structure, as it has been constructed from the eigenvector of almost equally-large entries.

It might be expected that a latent factor given by the eigenvector with equally-large entries should be easy to filter out of the data, especially given the simple structure for the latent factor with an eigenvector of equal weights. The best choice is to use reduced form models which are agnostic, by assumption, and do not attempt to explain the phenomenon, but rather describe it accurately.

The results in CGM are so disturbing that several papers have tried to explain the purported puzzle, the most obvious explanation being the presence of inappropriate regressors. Indeed, Cremers et al. (2008) show that the firm-specific equity volatility is an important determinant of corporate bond spreads, and that the economic effects of volatility are large. They use option-based volatility and the implied volatility skew, while CGM use VIX as the aggregate proxy for the volatility of each company. In such cases, principal component analysis of the residuals from the regressions does not reveal any significant omitted factors (There is an enormous literature on the estimation of factors by means of principal component analysis. See for instance Ando and Bai, 2017; Connor and Korajczyk, 1986; Bai, 2009; Bai and Ng, 2002; Bonhomme and Manresa, 2015; Chamberlain and Rothschild, 1983; Forni et al., 2000; Hallin and Liška, 2011; Moench and Ng, 2011; Pesaran and Tosetti, 2011; Wang, 2010).

A caveat is that Cremers et al. (2008) estimate the regression model using the levels of CDSs and not on their changes, as in CGM. They assume that credit spreads are stationary, which is reasonable, but which is not guaranteed empirically. All the regressions in levels have very high $R^2$, and the regressors have strong explanatory power. Nevertheless, what is most important is that Cremers et al. (2008) find no evidence of a large unidentified factor that is unrelated to credit risk. It is shown below that, even with statistically significant regressors, it is possible to generate a spurious factor in the residuals.

Ericsson et al. (2009) analyze the determinants of 5-year CDS spreads from 1999 to 2002. Using firm leverage, stock volatility and the risk-free rate, they are able to explain 61% of the variation in levels, 22% of the variation in CDS spread changes, and find evidence for a common factor in the residuals. However, the common factor is substantially weaker and explains approximately 32% of the variance in the residuals from regressions on changes in CDSs, and not 76%, as in CGM. Another difference from CGM is that the eigenvector of the main principal component in the residuals has both positive and negative elements.

These two studies also consider panel data estimation. The results from the panel regressions are not directly comparable to the CGM results in all respects. One can compare the explanatory power of the variables, but it is difficult to interpret the differences between systematic latent factors in the residuals from univariate and panel regressions. In general, it is found that $R^2$ from both regressions are roughly the same, but different variables are more successful in explaining the variation in the CDS premia.

Schaefer and Strebulaev (2008) find that the poor performance of structural models may be connected to the influence of non-credit factors that are present in the bond price data. They show that even simple structural models can predict accurate equity hedge ratios, so structural models can estimate the credit exposure of corporate debt reasonably well.

An open question arises here, namely how Avramov et al. (2007), Cremers et al. (2008), Ericsson et al. (2009) and Schaefer and Strebulaev (2008) find such diametrically different conclusions from CGM and Castagnetti and Rossi (2013) regarding the missing factor? Can this difference be attributed only to different datasets? This paper explains why such different conclusions are possible and entirely probable. Consistent with Avramov et al. (2007), we show that forming portfolios from individual regression residuals creates a strong and spurious latent factor. It is shown that using bonds of the same company, but with different maturities in one panel of data, is another reason for the spurious amplification of the strength of a latent factor. Finally, we explain theoretically and empirically why the first principal component uncovered by CGM had almost equally-weighted elements of its eigenvector. It follows that this equally-weighted portfolio does not constitute proof of equal exposure of each company’s corporate bond to the latent factor.

2. Is there a strong first eigenvalue and a flat eigenvector?

Consider a $T \times N$ matrix of credit spreads changes, $\Delta CS_{(T \times N)}$, which are assumed to be driven by some possibly non-linear factors with time-varying loadings. Although the factors may be non-linear and the loadings may be time dependent, we will explain $\Delta CS_{it}$ through the use of linear regressions. It follows that:

$$\Delta CS_{it} = \alpha_i + \Lambda_i X_{it} + \eta_{it},$$

(1)
where it is assumed that $\alpha_i$ and $\Lambda_i$ are time independent, and the matrix $X_j$ contains both company-specific and market-wide observable variables. It is likely that the matrix of regressors $X$ might omit some crucial variables, which will lead to an omitted variable problem in the residuals. Even if the choice of explanatory variables and $X$ contains the appropriate regressors, the non-linear relation of $X$ to $\Delta r$ will still lead to cross-correlations of $\eta_i$ for different $i \in N$.

A simple way to check whether we have omitted some variables is to apply principal component analysis to the matrix of residuals. If principal components uncover a factor that explains the total variance, and has an eigenvector which can be interpreted economically, there will be support for a latent factor. CGM find empirical evidence of such a latent factor.

Apart from the strength of the first eigenvalue, another puzzling feature of the latent factor is the shape of the eigenvector corresponding to the first principal component. CGM find that the first eigenvector has almost equal weights across different maturities and leverage ratios. The existence of the principal component that has an eigenvector with economically interpretable loadings is usually easier to accept. Therefore, an equally-weighted eigenvector seems to provide evidence for the factor with equal risk exposure across different maturities and credit qualities.

We show that the specific structure of the first eigenvector is unrelated to the maturities or leverage ratios, but is related to the use of averages. In fact, any random combination of stationary time series that contains a weak latent factor would generate an equally-weighted first eigenvector from the correlation matrix.

It is relatively easy to explain why the first eigenvalue becomes spuriously so strong in CGM. However, it is less obvious why the first eigenvector turns out to be an equally-weighted flat vector. The intuition is the following. Suppose that all the variation in the dataset, $\eta(T \times N)$, is driven by a linear factor structure plus noise. Then assume that all the noise from the data is eliminated to yield a new noiseless dataset, $\tilde{\eta}(T \times K)$. The covariance matrix of $\tilde{\eta}(T \times K)$ will become singular. The eigenvector of the first principal component for the original data, $\eta(T \times N)$, may take any shape, depending on the loadings of the factor for each time series. However, if we compute the covariance matrix for $\tilde{\eta}(T \times K)$, we will obtain a matrix where all the entries are equal to each other. In this case, covariance matrix becomes singular and of rank 1. Additionally, the first eigenvector, corresponding to the only non-zero eigenvalue, will have all its entries equal to $\sqrt{\frac{T}{K}}$. This is explained formally below.

2.1. Formal explanation

2.1.1. Derivation

In what follows, we show that an equally-weighted eigenvector is a spurious artifact of the procedure used by CGM, that is, it is equally weighted by construction. From the regressions arising from (1), we obtain a matrix of residuals, $\eta(T \times N)$, with $T$ rows and $N$ columns. We make the following two assumptions.

**Assumption 1.** The entries of data matrix, $\eta(T \times N)$, are generated by the following two-factor process:

$$\eta_{it} = \beta_1 f_{1t} + \beta_2 f_{2t} + \epsilon_{it}$$

where $t \in \{1, \ldots, T\}$, $i \in \{1, \ldots, N\}$, with $f_{it} \sim N(0, \sigma_{f_{it}}^2)$, the correlation parameters between the two factors is equal to $\rho_{12}$, and $\epsilon_{it} \sim N(0, \sigma^2_{\epsilon_{it}})$.

**Assumption 2.** For any given factor, $f_{it}$, the factor loadings across different assets are, on average, equal to one. That is, for a portfolio of $n_k$ assets, the loadings on factor $f_{it}$ converge to one as $n_k$ increases:

$$\hat{\beta}_f = \frac{1}{n_k} \sum_{i=1}^{n_k} \beta_{fi} \to 1 \text{ as } n_k \to \infty.$$  

Assumption 1 uses just two factors to make the argument easier to understand, but it can be easily generalized to any number of factors. Assumption 2 about beta has been also made by Harding (2008) and Brown (1989). The empirical beta of each asset is obtained from the regression of this asset on the estimated factor, where the estimated factor is defined as an equally weighted portfolio.

For $\eta(T \times N)$, we obtain the $N \times N$ covariance matrix $\Sigma_{\eta\eta}$. The off-diagonal terms are equal to:

$$cov(\eta_i, \eta_j) = \sum_{t=1}^{T} (\beta_{i1} f_{1t} + \beta_{i2} f_{2t} + \epsilon_{it})(\beta_{j1} f_{1t} + \beta_{j2} f_{2t} + \epsilon_{jt})$$

$$= \beta_{i1}\beta_{j1}\sigma_f^2 + 2\beta_{i1}\beta_{j2}\sigma_f\sigma_{\epsilon} + \beta_{i2}\beta_{j2}\sigma_{\epsilon}^2.$$

while the main-diagonal terms are equal to:

$$cov(\eta_i, \eta_i) = \sum_{t=1}^{T} (\beta_{i1} f_{1t} + \beta_{i2} f_{2t} + \epsilon_{it})^2$$

$$= \beta_{i1}^2\sigma_{f_1}^2 + 2\beta_{i1}\beta_{i2}\sigma_{f_1}\sigma_{f_2} + \beta_{i2}^2\sigma_{f_2}^2 + \sigma_{\epsilon}^2.$$  

Hence, the covariance matrix is equal to $\Sigma_{\eta\eta} = \Sigma_{f\beta}\Sigma_f + \sigma_{\epsilon}^2 I_N$. 


Assume that the information in the dataset, \( \eta \), is transformed by taking averages across specified columns of \( \eta_i \) into \( K \) different portfolios. Each portfolio is formed from the partition, \( \mathcal{N}_k \), of set \( \mathcal{N} = \{1, \ldots, N\} \), where \( k \in \{1, \ldots, K\} \). Let \( n_k \) be the cardinality of the \( \mathcal{N}_k \) partition of set \( \mathcal{N} \). In this way, we generate a new dataset, \( \tilde{\eta}_{(t \times K)} \), where each time \( t \) entry of column \( k \) is constructed according to:

\[
\forall i \in \mathcal{N}_k : \quad \tilde{\eta}_{k,t} = \frac{1}{n_k} \sum_{i=1}^{n_k} \eta_{i,t} = \left( \frac{1}{n_k} \sum_{i=1}^{n_k} \beta_{1i} \right) \eta_{1,t} + \left( \frac{1}{n_k} \sum_{i=1}^{n_k} \beta_{2i} \right) \eta_{2,t} + \left( \frac{1}{n_k} \sum_{i=1}^{n_k} \epsilon_{i,t} \right) = \tilde{\beta}_{1k} \eta_{1,t} + \tilde{\beta}_{2k} \eta_{2,t} + \tilde{\epsilon}_{k,t}.
\]

The new loadings are described by \( \tilde{\beta}_k \), and the new errors by \( \tilde{\epsilon}_{k,t} \). Importantly, observe the distribution of the new error term, which is:

\[
\tilde{\epsilon}_{k,t} \sim N\left(0, \frac{1}{n_k} \sigma^2 \right).
\]

where the variance is \( n_k \) times smaller than for \( \epsilon_{i,t} \). This follows from the independence of \( \epsilon_i \) and \( \text{var}(\tilde{\epsilon}_k) = \left( \frac{1}{n_k} \right)^2 \text{var} \left( \sum_{i=1}^{n_k} \epsilon_{i,t} \right) = \frac{1}{n_k} \sigma^2 \).

Covariance matrix for the transformed data \( \tilde{\eta} \). The covariance matrix for the new transformed dataset, \( \tilde{\eta} \), is given as:

\[
\text{cov}(\tilde{\eta}) = \tilde{\beta} \Sigma_{ff} \tilde{\beta}' + \sigma^2 I_k.
\]

The off-diagonal elements of the covariance matrix take the form:

\[
\text{cov}(\tilde{\eta}_i, \tilde{\eta}_j) = \tilde{\beta}_{1i} \tilde{\beta}_{1j} \sigma^2_{f_1} + 2 \tilde{\beta}_{1i} \tilde{\beta}_{2j} \sigma_{f_1} \sigma_{f_2} \rho_{12} + \tilde{\beta}_{2i} \tilde{\beta}_{2j} \sigma^2_{f_2},
\]

while the elements on the main diagonal are equal to:

\[
\text{var}(\tilde{\eta}_i) = \tilde{\beta}_{1i}^2 \sigma^2_{f_1} + 2 \tilde{\beta}_{1i} \tilde{\beta}_{2i} \sigma_{f_1} \sigma_{f_2} \rho_{12} + \tilde{\beta}_{2i}^2 \sigma^2_{f_2} + \sigma^2_f.
\]

But here we use Assumption 2 that portfolio betas always tend to one when a portfolio is sufficiently large. Consequently, the off-diagonal elements of the covariance matrix, \( (8) \), approach a constant value:

\[
\text{cov}(\tilde{\eta}_i, \tilde{\eta}_j) \to \frac{\tilde{\beta}_{1i} \tilde{\beta}_{1j} \sigma^2_{f_1}}{1} + \frac{2 \tilde{\beta}_{1i} \tilde{\beta}_{2j} \sigma_{f_1} \sigma_{f_2} \rho_{12}}{1} + \frac{\tilde{\beta}_{2i} \tilde{\beta}_{2j} \sigma^2_{f_2}}{1} \to \sigma^2_{f_1} + 2 \sigma_{f_1} \sigma_{f_2} \rho_{12} + \sigma^2_{f_2} = 1' \Sigma_{ff} 1 \quad \forall i, j \text{ as } n_k \to \infty.
\]

Similarly, if the cardinality of each portfolio is increased, the idiosyncratic variance term tends to zero:

\[
\text{var}(\tilde{\epsilon}) = \frac{1}{n_k} \sigma^2 \to 0 \text{ as } n_k \to \infty.
\]

We now observe that the main diagonal entries tend to the same value as the off-diagonal entries because:

\[
\text{var}(\tilde{\eta}_i) \to \frac{\tilde{\beta}_{1i}^2 \sigma^2_{f_1}}{1} + \frac{2 \tilde{\beta}_{1i} \tilde{\beta}_{2i} \sigma_{f_1} \sigma_{f_2} \rho_{12}}{1} + \frac{\tilde{\beta}_{2i}^2 \sigma^2_{f_2}}{1} \to \sigma^2_{f_1} + 2 \sigma_{f_1} \sigma_{f_2} \rho_{12} + \sigma^2_{f_2} = 1' \Sigma_{ff} 1 \quad \forall i \text{ as } n_k \to \infty.
\]

Therefore, as \( n_k \) increases, the covariance matrix, \( \text{cov}(\tilde{\eta}) \), approaches a \( K \times K \) matrix with \( 1' \Sigma_{ff} 1 \) elements everywhere:

\[
\Sigma_{\tilde{\eta}} \to \begin{bmatrix}
1' \Sigma_{ff} 1 & \cdots & 1' \Sigma_{ff} 1 \\
\vdots & \ddots & \vdots \\
1' \Sigma_{ff} 1 & \cdots & 1' \Sigma_{ff} 1
\end{bmatrix}
\]

We observe that matrix \( \Sigma_{\tilde{\eta}} \) from Eq. (14) is singular and has matrix rank equal to one.

**Equally weighted eigenvector.** From the Perron–Frobenius theorem, we know that every real square matrix with positive entries has a unique largest eigenvalue, and the corresponding eigenvector has all strictly positive components. Additionally,
no other eigenvector has strictly positive components. In particular, in the special case of the symmetric matrix, \( C_{K \times K} \), with the form in (15), it follows that:

\[
C_{K \times K} = \begin{bmatrix}
\psi & \rho & \cdots & \rho \\
\rho & \psi & \cdots & \rho \\
\vdots & \vdots & \ddots & \vdots \\
\rho & \rho & \cdots & \psi
\end{bmatrix}
\]

(15)

has the first eigenvector (corresponding to the highest eigenvalue), \( v_1 \), with all equally-weighted strictly positive entries:

\[
v_1(C_{K \times K}) = \sqrt{\frac{1}{K}}[1, 1, \ldots, 1].
\]

(16)

For example, if \( \psi = 1 \) and all \( \rho \approx 1 \) in \( C_{K \times K} \), this yields a matrix with the first eigenvector very close to the equally-weighted \( v_1 \). It is straightforward to demonstrate the result with Monte Carlo simulations, which is presented in Section 2.2.3.

More importantly, we show that the same holds for real data. For real data, we should not expect to obtain an exactly equally-weighted eigenvector for two reasons: (i) the cardinality, \( n_k \), of each portfolio is usually relatively small due to limitations of the available dataset; (ii) Assumptions 1 and 2 that the data generating process contains only homoscedastic factors and factor loadings are, on average, equal to one in each portfolio, need not be true. Nevertheless, despite these two caveats, we can show with real data how an increase in \( n_k \) results in a more equally weighted first eigenvector, and how taking averages leads to an eigenvector with lower variance for each entry that is more tightly centered around the mean of \( \sqrt{1/K} \).

2.2. Relation to random matrix theory

It is possible to consider an alternative explanation for the spurious increase of the first eigenvalue described in this paper. Principal component analysis has been recognized to fail for high-dimensional datasets. More precisely, spectral decomposition of covariance matrices is reliable only in the case when the number of rows or observations, the \( T \)-dimension, is much larger than the number of columns, the \( N \)-dimension. The impact of high-dimensionality on eigenvalues was first described in Marchenko and Pastur (1967). They show that, for the sample covariance matrix computed for a random matrix \( X \) with \( T \times N \) identically and independently distributed random variables with finite variance, as \( T \) and \( N \) tend to infinity and the ratio of matrix dimensions is constant, \( \frac{N}{T} = \gamma \), for any real \( x \):

\[
\frac{1}{N} \text{number of eigenvalues of } \frac{1}{T}X'X \rightarrow F(x).
\]

The function \( F(x) \) is such that \( F'(x) = \frac{1}{2\pi}\sqrt{\frac{1}{\gamma}\frac{1}{\gamma}} \) for \( a < x < b \), where \( a = (1 - \sqrt{\gamma})^2 \) and \( b = (1 + \sqrt{\gamma})^2 \). In cases where \( \gamma > 1 \), the limiting measure has an additional mass \( 1 - \frac{1}{\gamma} \) at zero. This result has been extended by many authors (for reviews see Bai, 1999 and Johnstone, 2006).

We observe that the so-called Marcenko–Pastur distribution describes the upper bound for the first sample eigenvalue, which deviates spuriously from the true population eigenvalue equal to one. In the case considered in this paper, the ratio of rows to columns is equal to \( \gamma_{\text{CM}} = \frac{15}{120} = \frac{1}{8} \), and so the first and highest sample eigenvalue would be equal to \( \left(1 + \sqrt{\frac{1}{8}}\right)^2 = 1.832 \), rather than 1. This implies that the first principal component could explain at most \( \frac{1}{15}\left(1 + \sqrt{1/8}\right)^2 = 0.122 \) of the total variance. In other words, if the residuals in the matrix \( \tilde{\eta} \) were uncorrelated with each other, identically and independently distributed and with a finite variance, then the first principal component might appear to explain approximately 12.2% of the total variance (rather than the true population value of 6.6%), which would be a consequence of the high-dimensionality of data matrix \( \tilde{\eta} \). However, 12.2% is much smaller than 76% found in CM. Therefore, the strong first principal component in CM cannot be attributed to the result of high-dimensional noise alone, as described by the Marcenko–Pastur distribution.

2.2.1. Spiked population model

We have shown that columns of data generated by an uncorrelated noise, even in a high-dimensional setting, cannot explain the strength of the first principal component in CM. Still, it is possible that a high-dimensional dataset generated by a process following a linear factor structure could spuriously produce a strong first principal component with a flat eigenvector. High-dimensional factor models, otherwise called “spiked population models”, have recently attracted a lot of attention in finance and economics. The concept of “spiked population models” was first considered in Johnstone (2001). We will base our analysis on the factor model and parameter calibrations from Harding (2008).

Assume that there are \( N \) columns, \( T \) rows/time periods and \( M \) factors. Asset returns follow a linear factor structure: \( \eta_{T \times N} = F\beta^T + \varepsilon \), where \( \eta_{T \times N} \) is a matrix of \( T \times N \) dimensions, \( F \) is a \( T \times M \) matrix of factor realizations, \( \beta \) is a \( N \times M \) matrix of factor loadings, and \( \varepsilon \) is a matrix of idiosyncratic noise components. Furthermore, factors are assumed to be orthogonal to each other and to the idiosyncratic noise. This implies that the population covariance of returns is
given by $\Sigma_{NN} = \sigma_0^2 \beta' \beta + \sigma_2 I_N$. Additionally, let the columns of the loadings matrix $\beta = [\beta_1 : \beta_2 : \ldots : \beta_N]$ be defined as $\beta_i = \epsilon_i + \sigma_\beta e_i$, where $i = (1, \ldots, 1)$ is the $N \times 1$ vector of ones, and $e_i$ is an i.i.d. vector random variable with mean 0 and variance 1. The population eigenvalues of $\Sigma_{NN}$ are given by:

$$\lambda_1 = \frac{\sigma_0^2}{\sigma_\beta^2} + \frac{\sigma_2^2}{\sigma_\beta^2} = \frac{\sigma_0^2}{\sigma_\beta^2} + \sigma_2^2$$
$$\lambda_i = \frac{\sigma_0^2}{\sigma_\beta^2} + \sigma_2^2$$ for $i = 2, \ldots, M$
$$\lambda_j = \sigma_2^2$$ for $j = M + 1, \ldots, N$

However, as Marchenko and Pastur (1967) observed, for high-dimensional datasets, population eigenvalues differ from sample eigenvalues. In this case, as shown in Harding (2008), sample eigenvalues take the following form:

$$\hat{\lambda}_1 = \frac{N\sigma_0^2}{\sigma_\beta^2} + \frac{\sigma_2^2}{\sigma_\beta^2}$$
$$\hat{\lambda}_i = \frac{N\sigma_0^2}{\sigma_\beta^2} + \sigma_2^2$$ for $i = 2, \ldots, M$ if $N \geq \frac{1}{T} \left( \frac{\sigma_2^2}{\sigma_\beta^2} \right)^2$
$$\hat{\lambda}_i = \sigma_2^2 \left( 1 + \sqrt{\frac{N}{T}} \right)^2$$ for $i = 2, \ldots, M$ if $N < \frac{1}{T} \left( \frac{\sigma_2^2}{\sigma_\beta^2} \right)^2$
$$\hat{\lambda}_j = \sigma_2^2 \left( 1 + \sqrt{\frac{N}{T}} \right)^2$$ for $j = M + 1, \ldots, N$.

Similarly to Harding (2008) and Brown (1989), we assume the following values: $\sigma_0^2 = 0.01$, $\sigma_\beta^2 = 0.000158$ and $\sigma_2^2 = 0.0045$. For two factors only and $T = 120$, $N = 15$, the second eigenvalue falls below the upper bound of the Marcenko–Pastur distribution. The first eigenvalue, scaled by the total variance, explains 14.33% of the total variance. Again, 14.33% is much smaller than the first eigenvalue found in CCM (76%). In this set-up, it would be necessary to increase substantially the signal to noise ratio for the first eigenvalue to be able to be as strong as the one in CCM. More precisely, $\sigma_0^2$ would need to be 41 times larger than the one assumed here, which is unrealistically high. Altogether, we may note that high-dimensionality of “spiked covariance matrices” cannot be solely responsible for the spurious increase in explained variance by the first eigenvector.

### 2.2.2. High-dimensionality of data cannot explain flatness of the first eigenvector

Most of the “random matrix” theory literature concentrated on the distribution of sample eigenvalues. Much less is known about eigenvectors. Ledoit and Péché (2011) address the relationship between the sample and population eigenvectors. But they do not show that the first eigenvector becomes flatter when high-dimensionality of data increases.

We now show that high-dimensionality of a dataset cannot be the only explanation for why the first eigenvector in CCM became spuriously so close to the “flat” eigenvector, that is, the vector with equally-weighted entries. This will be shown in Monte Carlo simulations. We sample 10,000 different datasets separately for three different high-dimensionality, $\gamma = N/T$, parameters. More precisely, we fix the number of columns to be $N = 200$ and let $T$ depend on $\gamma$. For a low $\gamma = 1/100$, we sample a dataset with $N = 200$ columns and $T = 20,000$ rows. For a moderate $\gamma$, we choose the same dimensions as in CCM, that is, for $\gamma = 15/120$ we sample a dataset with $N = 200$ columns and $T = 1,600$ rows. Finally, for truly high-dimensional data with $\gamma = 99/100$ we sample a dataset with $N = 200$ columns and $T = 203$ rows.

Additionally, we introduce the following modifications to the setup of the parameters from Harding (2008). As before, for the simulations we assume that there are two factors, $M = 2$. We also assume that the loadings on the first factor have a substantially higher variance, $\sigma_\beta^2 = 5$, than on the second factor, $\sigma_\beta^2 = 0.01$. We allow a moderately small correlation between the factors, $\rho_{12} = 0.1$. Intuitively, higher $\sigma_\beta$ should cause a larger deviation of the first eigenvector from the equally-weighted flat eigenvector, because greater dispersion in the magnitude of factor loadings across all the assets should pull the first eigenvector away from equal values.

Indeed, for $\sigma_\beta^2 = 5$ and $\sigma_\beta^2 = 0.01$, we find that the angle between the first eigenvector and the flat eigenvector can range from 46.3 to 78.8 degrees (see the histograms in Fig. 1). Crucially, the angle does not depend on the high-dimensionality parameter: $\gamma = N/T$. Fig. 1 shows that, for all three cases of the $\gamma$ parameter, we find a substantial angle in degrees between the first eigenvector from the simulated dataset and the flat equally-weighted vector. In other words, an increase in the $\gamma$ parameter does not translate into a “flatter” first eigenvector. Hence, high-dimensionality in the spiked population models cannot alone explain why the first eigenvector in CCM is so close to the equally-weighted flat vector.

### 2.2.3. Impact of averaging on the first eigenvector and eigenvector in Monte-Carlo simulations

Finally, we show by means of simulations that the process of taking averages, as described above, has a substantial impact on the “flatness” of the first eigenvector and the amount of explained variance by the scaled first eigenvalue. For the simulations, we assume the same parameters as above, and adapt only the size of the data. That is, we start with a
η_T×N dataset, where we fix the number of rows to be T = 120 but let N increase from N = 1 to N = 6,000. Columns of the simulated η_{120×N} dataset are then averaged into a T × N = 120 × 15 of the η_{120×15} dataset. In this way, each column of the η_{120×15} dataset constitutes an averaged portfolio of n_k = N/15 columns sampled from the wider η_{120×N} dataset.

For this setup, Fig. 2 shows that, as N increases, the first eigenvector becomes more “flat”, as measured by its angle with an equally-weighted flat vector, and the amount of explained variance by the scaled first eigenvalue, increases to one.

3. Empirical results - high eigenvalues?

3.1. Data description.

The data on corporate credit spreads is obtained from four different data sources that cover different time periods of bond data. For the period 1992–1997, bond data are obtained from the Lehman Brothers Fixed Income Database. For bond data from 1994 to 1997, and from 2008 to 2011, we access Mergent FISD. We obtained bond data for the period 2008 to 2012 from TRACE. Finally, we downloaded bond data over the entire sample period, 1992–2012, from Thomson Reuters Dastream.

The following order is used to prioritize over the four databases in the case of overlapping data points: the Lehman Brothers Fixed Income Database, TRACE, Mergent FISD, and Thomson Reuters Datastream. Credit spreads are computed in the following two-step procedure using the Federal Reserve constant maturity data. The yield to maturity for corporate bonds with coupons is defined implicitly through the following equation:

\[ p_{ijt}(T) = e^{-\gamma(t)T} + C \sum_{j=1}^{N} e^{-\gamma(t)\eta_{j}} \]

(17)

where C denotes a coupon that is paid out at dates \( t_j, j \in \{1, N\} \), and \( t_N \equiv T \) is the bond’s maturity. Given the yield to maturity and a riskless zero-coupon interest rate with the same maturity, we obtain the credit spread from:

\[ cs_{ijt} = Y_{ijt}(T) - r(T). \]

(18)

Corporate bond issue and issuer level variables are extracted from Mergent FISD. For each corporate bond, we obtained the respective offering date and maturity date. We compute an issue’s time to maturity, and its time since issuance (bond age) for every month in the dataset using this information.

We link the corporate bond with the stock and accounting information, and obtain daily equity data from the Center for Research in Security Prices (CRSP) and quarterly firm fundamentals from Compustat. Firms for which stock data that are not available are excluded (these are mostly privately-held firms). We compute the firm’s market value by the product of the stock price and the number of publicly held shares. Leverage is based on data from Compustat, and is defined in the same way as in Ericsson et al. (2009):

\[ \text{Leverage} = \frac{\text{Book Value of Debt}}{\text{Market Value of Equity+Book Value of Debt}} \]
Impact of averaging on the first Eigenvalue and Eigenvector

Angle (in degrees)  First Eigenvalue

![Graph showing impact of averaging on the first eigenvalue and eigenvector](image)

**Fig. 2.** Simulation results. The figure presents the influence of averaging on the first eigenvalue and eigenvector of the covariance matrix of the \( \hat{\eta} \) dataset. Initial dataset, \( \eta \), is of dimension \( T \times N = 120 \times N \). Columns of this dataset are then averaged into 15 portfolios of the \( \hat{\eta} \) dataset, which is of dimension \( 120 \times 15 \). Each of the columns of the \( \hat{\eta} \) dataset is an average of \( \eta_k = N/15 \) columns. The horizontal x-axis corresponds to the initial number, \( N \), of columns which are then averaged into 15 portfolios. Left panel presents a plot of the angle between the first eigenvector for the averaged dataset, \( \hat{\eta} \), and an equally-weighted flat vector, against the number of columns in the initial dataset \( N \). The right panel presents the amount of explained variance by the first principal component (that is, the first eigenvalue scaled by the total variance of the data) for the averaged dataset, \( \hat{\eta} \), plotted against the number of columns in the initial dataset \( N \). We notice that, as \( N \) increases, the angle drops to zero and the first eigenvalues increase to one.

\[
= \frac{(DD1Q+DLTTQ)}{PRCCM \times CSHOQ + (DD1Q+DLTTQ)}.
\] (20)

In order to obtain the firm’s book value of debt, we follow the literature and assume that it consists of short-term and long-term debt. For short-term debt, we use the Compustat data item, “Long-Term Debt Due in One Year” (DD1Q), which represents the current portion of long-term debt. For long-term debt, we use the Compustat data item, “Long-Term Debt - Total” (DLTTQ).

3.2. Issues with data

The main difficulty with the data is the following. Most of the bonds have very few observations and do not overlap with each other, which is necessary for computing the covariance and correlation matrices. We take bonds that have at least a certain number of observations. Assume that, in order to qualify, the bond must have at least 60 observations, which usually spans more than 5 years because of missing values. Obviously, this method excludes a large number of bonds, especially those with shorter maturities at the origin.

3.3. One bond per company

Table 1 presents the results for the sample of 872 bonds. Importantly, these 872 bonds belong to 872 different companies, so we include in this sample only one bond per company. After collecting all bonds for a company, we include a bond with the highest number of observations. We report the mean and standard deviation of the first \( \lambda_1 \) and second \( \lambda_2 \) eigenvalues for different subsamples. The procedure of computing eigenvalues is the following. First, we construct a new subsample. Second, we apply the probabilistic principal components procedure to the new subsample in order to fill the missing values, and estimate the first and second eigenvalues.

One difficulty with the probabilistic principal components procedure (based on Porta et al., 2005) is that it fills missing values in the data differently for every run. We circumvent this problem by running the procedure 10,000 times, which is why we report both the mean and standard deviation of the first and second eigenvalues.
In Panel A of Table 1, we report the results for three different subsamples of \( r = 300 \), \( r = 200 \), and \( r = 100 \) bonds. On every run, we sample \( r \) bonds randomly from 872 bonds. For all three subsamples, we find that the first eigenvalue explains between 21% and 29% of the total variance in the residuals. The estimates from Panel A (and also Panel A in Tables 3 and 5) use the proper method of measuring the strength of the latent factor in the residuals, thereby avoiding the two main deficiencies of the analysis in CGM.

In Panel B, we construct portfolios, where \( K \) denotes the number of portfolios and \( n_k \) the number of bonds that fall into each portfolio, that is, its cardinality. On the left of Panel B, we present the results for portfolios with randomly-assigned bonds. On the right, we report the results for portfolios that are formed according to maturity and leverage, in the same way as in CGM.

There is a clear pattern that the higher is the cardinality number, \( n_k \), the stronger is the first principal component. For \( K = 100, n_k = 9 \), there seems to be little difference from the “not-averaged” (\( r = 100 \)) residuals from Panel A. However, for \( K = 15, n_k = 59 \), and \( K = 5, n_k = 175 \), the spurious increase in the strength of the eigenvalues is very strong.

It is observed that using the portfolio definitions, as in CGM, does not result in the same increase in the strength of the eigenvalues, as compared with the random portfolios. In fact, it seems to be related to non-equispaced \( n_k \) across different portfolios rather than to the properties of maturity or leverage. For a fixed initial number of bonds, \( N \), different \( n_k \) will filter different amounts of noise across all portfolios. If \( n_k \) is very small for some portfolios and very large for others, there will be noise in the first portfolio while the second will essentially contain only a signal. Consequently, for non-equal \( n_k \), the cross-correlations are also weaker.

<table>
<thead>
<tr>
<th>Table 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Only one bond per company: 872 corporate bonds.</td>
</tr>
<tr>
<td>[ \begin{array}{cccccc}</td>
</tr>
<tr>
<td>[ \text{Panel A. Separate K bonds, without averaging} ]</td>
</tr>
<tr>
<td>[ \text{Panel B. Portfolios} ]</td>
</tr>
<tr>
<td>[ \text{K portfolios constructed from } n_k \text{ bonds} ]</td>
</tr>
<tr>
<td>[ \text{Random samples} ]</td>
</tr>
<tr>
<td>[ \begin{array}{cccccccc}</td>
</tr>
<tr>
<td>[ \text{15 baskets, as in CGM} ]</td>
</tr>
<tr>
<td>[ \text{Panel B. Portfolios} ]</td>
</tr>
<tr>
<td>[ \text{K portfolios constructed from } n_k \text{ bonds} ]</td>
</tr>
<tr>
<td>[ \text{Random samples} ]</td>
</tr>
<tr>
<td>[ \begin{array}{cccccccc}</td>
</tr>
<tr>
<td>[ \text{15 baskets, as in CGM} ]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>All 2983 bonds.</td>
</tr>
<tr>
<td>[ \begin{array}{cccccc}</td>
</tr>
<tr>
<td>[ \text{Panel A. Separate K bonds, without averaging} ]</td>
</tr>
<tr>
<td>[ \text{Panel B. Portfolios} ]</td>
</tr>
<tr>
<td>[ \text{K portfolios constructed from } n_k \text{ bonds} ]</td>
</tr>
<tr>
<td>[ \text{Random samples} ]</td>
</tr>
<tr>
<td>[ \begin{array}{cccccc}</td>
</tr>
<tr>
<td>[ \text{15 baskets, as in CGM} ]</td>
</tr>
</tbody>
</table>
Table 3
Robustness check - Only one bond per company: 263 bonds.

<table>
<thead>
<tr>
<th></th>
<th>mean</th>
<th>sd</th>
<th>mean</th>
<th>sd</th>
<th>mean</th>
<th>sd</th>
</tr>
</thead>
<tbody>
<tr>
<td>Panel A. Separate K bonds, without averaging</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>r = 200</td>
<td></td>
<td></td>
<td>r = 100</td>
<td></td>
<td>r = 50</td>
<td></td>
</tr>
<tr>
<td>(\lambda_1)</td>
<td>0.2851</td>
<td>0.0072</td>
<td>0.2878</td>
<td>0.0164</td>
<td>0.2957</td>
<td>0.0261</td>
</tr>
<tr>
<td>(\lambda_2)</td>
<td>0.1187</td>
<td>0.0035</td>
<td>0.1207</td>
<td>0.0079</td>
<td>0.1249</td>
<td>0.0126</td>
</tr>
</tbody>
</table>

Panel B. Portfolios

<table>
<thead>
<tr>
<th></th>
<th>mean</th>
<th>sd</th>
<th>mean</th>
<th>sd</th>
</tr>
</thead>
<tbody>
<tr>
<td>K portfolios constructed from (n_k) bonds</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>K = 5, (n_k = 53)</td>
<td>(\lambda_1)</td>
<td>0.7561</td>
<td>0.0737</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(\lambda_2)</td>
<td>0.0749</td>
<td>0.0201</td>
<td></td>
</tr>
<tr>
<td>K = 15, (n_k = 18)</td>
<td>(\lambda_1)</td>
<td>0.5318</td>
<td>0.0168</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(\lambda_2)</td>
<td>0.0566</td>
<td>0.0135</td>
<td></td>
</tr>
<tr>
<td>K = 30, (n_k = 9)</td>
<td>(\lambda_1)</td>
<td>0.3515</td>
<td>0.0211</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(\lambda_2)</td>
<td>0.0480</td>
<td>0.0056</td>
<td></td>
</tr>
<tr>
<td>K = 50, (n_k = 6)</td>
<td>(\lambda_1)</td>
<td>0.3043</td>
<td>0.0168</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(\lambda_2)</td>
<td>0.0361</td>
<td>0.0038</td>
<td></td>
</tr>
<tr>
<td>K = 100, (n_k = 3)</td>
<td>(\lambda_1)</td>
<td>0.2714</td>
<td>0.0125</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(\lambda_2)</td>
<td>0.0261</td>
<td>0.0022</td>
<td></td>
</tr>
</tbody>
</table>

Table 4
Robustness check - all 714 bonds.

<table>
<thead>
<tr>
<th></th>
<th>mean</th>
<th>sd</th>
<th>mean</th>
<th>sd</th>
<th>mean</th>
<th>sd</th>
</tr>
</thead>
<tbody>
<tr>
<td>Panel A. Separate K bonds, without averaging</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>r = 300</td>
<td></td>
<td></td>
<td>r = 200</td>
<td></td>
<td>r = 100</td>
<td></td>
</tr>
<tr>
<td>(\lambda_1)</td>
<td>0.2602</td>
<td>0.0080</td>
<td>0.2627</td>
<td>0.0109</td>
<td>0.2654</td>
<td>0.0163</td>
</tr>
<tr>
<td>(\lambda_2)</td>
<td>0.1094</td>
<td>0.0042</td>
<td>0.1112</td>
<td>0.0057</td>
<td>0.1129</td>
<td>0.0086</td>
</tr>
</tbody>
</table>

Panel B. Portfolios

<table>
<thead>
<tr>
<th></th>
<th>mean</th>
<th>sd</th>
<th>mean</th>
<th>sd</th>
</tr>
</thead>
<tbody>
<tr>
<td>K portfolios constructed from (n_k) bonds</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>K = 5, (n_k = 143), r = 100</td>
<td>(\lambda_1)</td>
<td>0.8709</td>
<td>0.0642</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(\lambda_2)</td>
<td>0.0566</td>
<td>0.0189</td>
<td></td>
</tr>
<tr>
<td>K = 15, (n_k = 48), r = 100</td>
<td>(\lambda_1)</td>
<td>0.7487</td>
<td>0.0137</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(\lambda_2)</td>
<td>0.0530</td>
<td>0.0118</td>
<td></td>
</tr>
<tr>
<td>K = 30, (n_k = 24)</td>
<td>(\lambda_1)</td>
<td>0.6482</td>
<td>0.0104</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(\lambda_2)</td>
<td>0.0371</td>
<td>0.0051</td>
<td></td>
</tr>
<tr>
<td>K = 50, (n_k = 15)</td>
<td>(\lambda_1)</td>
<td>0.5061</td>
<td>0.0101</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(\lambda_2)</td>
<td>0.0287</td>
<td>0.0033</td>
<td></td>
</tr>
<tr>
<td>K = 100, (n_k = 8)</td>
<td>(\lambda_1)</td>
<td>0.4422</td>
<td>0.0098</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(\lambda_2)</td>
<td>0.0235</td>
<td>0.0024</td>
<td></td>
</tr>
</tbody>
</table>

3.4. Evidence from the residuals of 2983 bonds from 872 different companies

Mixing bonds of different maturities that belong to the same companies cannot be an accurate measure of residual co-movements. For obvious reasons, credit spreads of different bonds belonging to the same company are highly correlated with each other. If a variable “slope of the Treasuries” is a poor proxy for the term structure of the corporate bond spreads, the residuals will still be strongly correlated.

Panel A in Table 2 presents evidence supporting this claim. In Panel A, the first eigenvalue from “non-averaged” matrices is higher than in the corresponding Panel A in Table 1 (the results from Table 5 show this even more strongly for bonds with longer maturities). Thus, the estimates of eigenvalues should not be used as a measure of the strength of the latent factor. Moreover, we notice the same pattern as before, whereby the higher is the cardinality of the portfolio, \(n_k\), the stronger is the increase in the first eigenvalue.

3.5. Robustness check with differently sampled data

There is a risk that, in applying the probabilistic principal component procedure to the dataset with many missing values, it is possible to introduce an upward or downward bias in the estimates of the eigenvalues. Therefore, in this section we construct a new dataset with as few missing values as possible, but with a large cross section so that cardinality, \(n_k\), of each portfolio is sufficiently large.
We filter the data as follows. First, we obtain the pair, $\{t_{\text{min}}^i, t_{\text{max}}^i\}$, for each bond $i \in \{1, \ldots, N\}$ when the bond starts and when it appears for the last time in the dataset. Based on this pair, we find a bond, $s \in \{1, \ldots, N\}$, such that its first and last date pair, $\{t_{\text{min}}^s, t_{\text{max}}^s\}$, intersect with the largest number of other bond pairs, $\exists s \forall i: (t_{\text{min}}^i \geq t_{\text{min}}^s$ and $t_{\text{max}}^i \leq t_{\text{max}}^s)$. We select all of those bonds, then omit those bonds that have more than 10 missing observations within the prespecified dates, $\{t_{\text{min}}^i, t_{\text{max}}^i\}$. This procedure necessarily results in a dataset with a short time series. Finally, we obtain a dataset, $\mathcal{E}_{(T \times N)}$, with $T = 59$ rows (dates) and $N = 714$ corporate bonds. If we filter the bonds that belong to different companies, we obtain a sample of 263 bonds.

3.5.1. Only one bond per company

As before, we check the strength of the eigenvalues for the sample with only one bond per company. In Panel A of Table 3, the first eigenvalue is estimated at the level of roughly 0.28 for all subsamples. In Panel B of Table 3, we present the results for portfolios with different $K$ and $n_k$. In general, the results have the same pattern as before. The strength of the first eigenvalue for the sample of bonds is below 0.3 and, for portfolios, the magnitude of the first eigenvalue increases as the number of portfolios, $K$, decreases and the cardinality $n_k$ of each portfolio increases.

3.5.2. All bonds

Tables 2 and 4 present similar information, but for different samples of data. The dataset used for Table 4 has $N = 714$ columns (bonds) and $T = 59$ rows (time observations). As compared with Table 3, we now have a larger cross section of bonds as all the bonds for each company are used, such that 714 bonds belong to only 263 companies. The results are similar to those in the previous three tables, so we may conclude that the general results seem to be robust with respect to the two sampling methods.

Fig. 3 presents the plot of the distribution of the first eigenvalue for the two subsamples, which are the frequency histograms from 10,000 simulations. On the left (red colored bars) is the frequency histogram of the first eigenvalues for the subsamples of $r = 100$ bonds drawn randomly from the set of 714 bonds. The procedure is as follows. First, we divide the set

---

**Table 5**

Strength of the latent factor across different maturities.

<table>
<thead>
<tr>
<th></th>
<th>mean</th>
<th>sd</th>
<th>mean</th>
<th>sd</th>
<th>mean</th>
<th>sd</th>
<th>mean</th>
<th>sd</th>
<th>mean</th>
<th>sd</th>
</tr>
</thead>
<tbody>
<tr>
<td>$r \in (0, \infty)$</td>
<td></td>
<td></td>
<td>$r \in [0,12)$</td>
<td></td>
<td>$r \in [12,28)$</td>
<td></td>
<td>$r \in [12, \infty)$</td>
<td></td>
<td>$r \in [28, \infty)$</td>
<td></td>
</tr>
</tbody>
</table>

**Panel A. One bond per company**

| $\lambda_1$ | 0.2679 | 0.0248 | 0.2963 | 0.0263 | 0.3012 | 0.0269 | 0.4636 | 0.0576 | 0.2781 | 0.045 |
| $\lambda_2$ | 0.0928 | 0.0098 | 0.0927 | 0.0096 | 0.0981 | 0.0049 | 0.0948 | 0.0112 | 0.1704 | 0.0206 |

| $T = 70, N = 189$ | $r = 50$ |     | $T = 74, N = 26$ | $r = 20$ | $T = 74, N = 45$ | $r = 20$ | $T = 69, N = 53$ | $r = 50$ |

**Panel B. More than one bond per company**

| $\lambda_1$ | 0.2432 | 0.0236 | 0.2582 | 0.0240 | 0.4314 | 0.0217 | 0.5679 | 0.0343 | 0.7421 | 0.0212 |
| $\lambda_2$ | 0.1095 | 0.0112 | 0.1134 | 0.0119 | 0.0605 | 0.0036 | 0.0512 | 0.0066 | 0.0542 | 0.0059 |

| $T = 70, N = 489$ | $r = 50$ |     | $T = 74, N = 405$ | $r = 50$ | $T = 74, N = 64$ | $r = 50$ | $T = 74, N = 107$ | $r = 50$ | $T = 71, N = 93$ | $r = 50$ |
of 714 bonds into 15 partitions/portfolios. The bonds are assigned randomly to their respective partitions, and then averaged to obtain a single time series. This gives 15 time series/portfolios. Finally, the probabilistic principal components procedure is applied to this dataset, and the first two eigenvalues are computed. The process of taking averages eliminates most of the missing values, especially if the cardinality, \( n_k \), is large. The probabilistic principal component procedure has no effect if there are no missing values in the data. The mean eigenvalue for this subsample is 0.2654, with standard deviation (sd) 0.0163. On the right is the distribution of the first eigenvalues for 15 random portfolios, with mean eigenvalue of 0.7487.

3.6. Long maturity bonds

Fig. 4 shows that most of the bonds are issued with a maturity of 10 years. We examine the strength of the first two principal components separately for different sections of the maturity spectrum. When the first two columns in Table 5 are compared, it may appear that there is no difference between the matrix of residuals where each company appears only once, or where all bonds are mixed together in one matrix. However, there is a difference when we evaluate longer maturities. Therefore, mixing bonds of the same companies within a single matrix of residuals will likely lead to an upward bias in the strength of the latent factor.

It seems that the latent factor in long maturity bonds is stronger than for short maturity bonds. However, the pattern of cross-correlations and unexplained co-movements is neither uniform nor monotonic with respect to maturity, which can be seen clearly in Panel A of Table 5.

It seems that the corporate bond credit spread changes are subject to a substantial amount of idiosyncratic noise. Another plausible explanation is that there may be some segmentation within the corporate bond market itself, which cannot be easily discerned through linear regressions.

4. Empirical evidence - flat eigenvector?

The same procedure to sample the dataset is used as in Section 3.5. It is intended to show that taking averages into baskets results in a more equally-weighted first eigenvector than for the original data, as described in Section 2.1. The next two tables show this for randomly chosen samples of bonds. The method of assigning bonds to different baskets/portfolios is random. We permute randomly the set, \( \mathcal{N} = \{1, \ldots, N\} \), into \( K \) partitions with equal cardinality, \( n_k \), so that we obtain a new transformed dataset, \( \tilde{\mathcal{N}}_{(T \times K)} \). If \( \text{mod}(K, n_k) \neq 0 \), then \( n_k = \text{ceil}(\frac{n}{K}) \) holds for most of the baskets.

The same dataset is used as in Panel B of Table 4, but it is split into two subsamples. In Table 6 for Panel A, we use bonds that have maturity of less than 28 years, which reduces the sample to 585 bonds, while in Panel B, we use all 714 bonds. We use two different samples, as Panel B of Table 5 shows that bonds starting with a maturity of 30 years are more strongly correlated with each other than with bonds of other maturities. This means that the signal, \( f_i \), from Eq. (2) is stronger, so it should be easier for the factors to dominate the noise, and the first eigenvector should be “more equally weighted”.

The first and second columns in Table 6 indicate the number of portfolios to form, \( K \), and the number of bonds that fall into each portfolio. The value in the third column follows from Eq. (16), namely the value of an equally-weighted normalized eigenvector of dimension \( K \) for any matrix of the form (15). The \( \max(v_1) \) and \( \min(v_1) \) in the fourth and fifth columns denote the highest and lowest values among the entries of the first eigenvector, \( v_1 \). Column six gives the differences in the two, that is, the width of the interval covered by the entries of \( v_1 \). Column seven gives the first eigenvalue, and columns eight to eleven present the same results as in columns three to seven, respectively, but for randomly chosen \( K \) bonds. For columns eight to eleven, we have \( n_k = 1 \) as there is no averaging into portfolios.
Table 6
Variance of the first eigenvector as a function of $n_k$.

<table>
<thead>
<tr>
<th>$K$</th>
<th>$n_k$</th>
<th>$\sqrt{TK}$</th>
<th>$\text{min}(v_1)$</th>
<th>$\text{max}(v_1)$</th>
<th>$\text{max}(v_1) - \text{min}(v_1)$</th>
<th>$\lambda_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Panel A: Bonds with maturity shorter or equal to 28 years - 585 bonds</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>150</td>
<td>4</td>
<td>0.082</td>
<td>0.001</td>
<td>0.117</td>
<td>0.116</td>
<td>0.413</td>
</tr>
<tr>
<td>100</td>
<td>6</td>
<td>0.100</td>
<td>0.011</td>
<td>0.138</td>
<td>0.127</td>
<td>0.461</td>
</tr>
<tr>
<td>50</td>
<td>12</td>
<td>0.141</td>
<td>0.067</td>
<td>0.185</td>
<td>0.118</td>
<td>0.559</td>
</tr>
<tr>
<td>30</td>
<td>20</td>
<td>0.183</td>
<td>0.136</td>
<td>0.217</td>
<td>0.082</td>
<td>0.674</td>
</tr>
<tr>
<td>15</td>
<td>39</td>
<td>0.258</td>
<td>0.219</td>
<td>0.275</td>
<td>0.056</td>
<td>0.819</td>
</tr>
<tr>
<td>10</td>
<td>59</td>
<td>0.316</td>
<td>0.308</td>
<td>0.325</td>
<td>0.018</td>
<td>0.888</td>
</tr>
<tr>
<td>5</td>
<td>117</td>
<td>0.447</td>
<td>0.444</td>
<td>0.451</td>
<td>0.007</td>
<td>0.951</td>
</tr>
<tr>
<td>Panel B: All bonds</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>150</td>
<td>5</td>
<td>0.082</td>
<td>0.002</td>
<td>0.110</td>
<td>0.108</td>
<td>0.438</td>
</tr>
<tr>
<td>100</td>
<td>8</td>
<td>0.100</td>
<td>0.015</td>
<td>0.130</td>
<td>0.115</td>
<td>0.488</td>
</tr>
<tr>
<td>50</td>
<td>15</td>
<td>0.141</td>
<td>0.058</td>
<td>0.180</td>
<td>0.122</td>
<td>0.583</td>
</tr>
<tr>
<td>30</td>
<td>24</td>
<td>0.183</td>
<td>0.141</td>
<td>0.203</td>
<td>0.062</td>
<td>0.732</td>
</tr>
<tr>
<td>15</td>
<td>48</td>
<td>0.258</td>
<td>0.227</td>
<td>0.274</td>
<td>0.047</td>
<td>0.830</td>
</tr>
<tr>
<td>10</td>
<td>72</td>
<td>0.316</td>
<td>0.306</td>
<td>0.324</td>
<td>0.019</td>
<td>0.900</td>
</tr>
<tr>
<td>5</td>
<td>143</td>
<td>0.447</td>
<td>0.445</td>
<td>0.449</td>
<td>0.005</td>
<td>0.957</td>
</tr>
</tbody>
</table>

Table 7
Variance of the first eigenvector as a function of $n_k$, fixed $r$.

<table>
<thead>
<tr>
<th>$K$</th>
<th>$r$</th>
<th>$\binom{r}{k}$</th>
<th>$n_k$</th>
<th>$\sqrt{TK}$</th>
<th>$\text{min}(v_1)$</th>
<th>$\text{max}(v_1)$</th>
<th>$\text{max}(v_1) - \text{min}(v_1)$</th>
<th>mean of $s (d(v_1))$</th>
<th>mean($\lambda_1$)</th>
<th>sd($\lambda_1$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Panel A: $r = 15$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>150</td>
<td>15</td>
<td>1.62E+20</td>
<td>4</td>
<td>0.258</td>
<td>-0.062</td>
<td>0.395</td>
<td>0.457</td>
<td>0.063</td>
<td>0.446</td>
<td>0.044</td>
</tr>
<tr>
<td>100</td>
<td>15</td>
<td>2.53E+17</td>
<td>6</td>
<td>0.258</td>
<td>-0.020</td>
<td>0.380</td>
<td>0.400</td>
<td>0.058</td>
<td>0.490</td>
<td>0.042</td>
</tr>
<tr>
<td>50</td>
<td>15</td>
<td>2.25E+12</td>
<td>12</td>
<td>0.258</td>
<td>0.062</td>
<td>0.359</td>
<td>0.297</td>
<td>0.046</td>
<td>0.578</td>
<td>0.040</td>
</tr>
<tr>
<td>30</td>
<td>15</td>
<td>1.55E+08</td>
<td>20</td>
<td>0.258</td>
<td>0.171</td>
<td>0.322</td>
<td>0.151</td>
<td>0.025</td>
<td>0.685</td>
<td>0.025</td>
</tr>
<tr>
<td>Panel B: $r = 5$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>150</td>
<td>5</td>
<td>5.92E+08</td>
<td>4</td>
<td>0.447</td>
<td>-0.605</td>
<td>0.686</td>
<td>1.291</td>
<td>0.098</td>
<td>0.521</td>
<td>0.072</td>
</tr>
<tr>
<td>100</td>
<td>5</td>
<td>7.53E+07</td>
<td>6</td>
<td>0.447</td>
<td>-0.211</td>
<td>0.687</td>
<td>0.898</td>
<td>0.084</td>
<td>0.558</td>
<td>0.071</td>
</tr>
<tr>
<td>50</td>
<td>5</td>
<td>2.12E+06</td>
<td>12</td>
<td>0.447</td>
<td>0.029</td>
<td>0.631</td>
<td>0.602</td>
<td>0.065</td>
<td>0.632</td>
<td>0.071</td>
</tr>
<tr>
<td>30</td>
<td>5</td>
<td>1.43E+05</td>
<td>20</td>
<td>0.447</td>
<td>0.285</td>
<td>0.542</td>
<td>0.256</td>
<td>0.032</td>
<td>0.727</td>
<td>0.050</td>
</tr>
<tr>
<td>15</td>
<td>5</td>
<td>3.003</td>
<td>39</td>
<td>0.447</td>
<td>0.376</td>
<td>0.486</td>
<td>0.110</td>
<td>0.017</td>
<td>0.844</td>
<td>0.030</td>
</tr>
<tr>
<td>10</td>
<td>5</td>
<td>2.522</td>
<td>59</td>
<td>0.447</td>
<td>0.429</td>
<td>0.461</td>
<td>0.032</td>
<td>0.006</td>
<td>0.900</td>
<td>0.010</td>
</tr>
</tbody>
</table>

From $\mathbb{Z}_{(T+n_k)}$, we construct seven different transformed datasets, $\mathbb{Z}_{(T-K)}$, keeping $T$ fixed and changing $K$ and $n_k$. For example, the first row in Table 6 represents the results for $K = 150$ and $n_k = 4$. In columns six and seven, as we decrease $K$ and consequently increase the number of bonds, $n_k$, in each basket, two statistics change. First, the width of the interval, $\text{max}(v_1) - \text{min}(v_1)$, is significantly diminished from 0.116 for $K = 150$ to 0.007 for $K = 5$ in Panel A, and from 0.108 for $K = 150$ to 0.005 for $K = 5$. It is also observed that, as $n_k$ increases, both the highest and smallest entries of the eigenvector tend to approach $\sqrt{TK}$ from above and below. Clearly, the variance of the entries of the first eigenvector diminishes and, at the same time, is more tightly centered around the equally-weighted eigenvector. Second, the magnitude of the total variance explained by the first principal component, as indicated by the eigenvalue $\lambda_1$, increases from 0.413 to 0.951 in Panel A, and from 0.438 to 0.957 in Panel B.

On the other hand, for randomly chosen $K$ bonds (without averaging, $n_k = 1$), we do not observe any such pattern. In column eleven, the first eigenvalues seem to be unrelated to the number of bonds sampled. More importantly, there is an inverse relationship between $K$ and $\text{max}(v_1) - \text{min}(v_1)$. Moreover, $\text{max}(v_1) - \text{min}(v_1)$ is always higher in the tenth than in the sixth column. Therefore, the pattern is clear. Taking averages always changes the shape of the first eigenvector by reducing the variance of its entries, and centering it more tightly around the value of the normalized equally-weighted eigenvector.

Table 7 presents an additional robustness check. First, form $K$ portfolios, each of cardinality $n_k$, and then sample $r$ portfolios for which the first eigenvector and eigenvalue are computed. For each $K$, there are $\binom{K}{r}$ possible samples, not all of them different. We repeat the sampling of $r$ portfolios 10,000 times. More precisely, we repeat it 10,000 times if $\binom{K}{r} \geq 10,000$, otherwise $\binom{K}{r}$ times. For the results, it does not matter that we might sample the same $r$ portfolios, as it just changes the ordering of the entries in the eigenvector without having any effect on the difference, $\text{max}(v_1) - \text{min}(v_1)$, or standard deviation of the entries in $v_1$. The results are very clear. In both panels A and B, for $r = 15$ an $r = 5$, respectively, the increase in $n_k$ leads to a smaller width of $\text{max}(v_1) - \text{min}(v_1)$, and the mean standard deviation of $v_1$ for all 10,000 samples also decreases as $n_k$ increases.

The most important difference between the results in Tables 6 and 7 is that the size of the correlation matrix does not influence the "equal weightedness" of the first eigenvector. The interval, $\text{max}(v_1) - \text{min}(v_1)$, is much wider in Panel B for the $r = 5$ columns, as compared with Panel A for the $r = 15$ columns. This can be interpreted as evidence that the main
driving force behind the shape of the first eigenvector is the initial cardinality, \( n_k \), of each portfolio, which is consistent with Eqs. (11) and (13). In short, a higher \( n_k \) filters out more noise and directs the correlation coefficients towards one.

Similarly to CGM, Castagnetti and Rossi (2013) found the latent factor in the residuals averaged according to some observable criteria. CGM divided the residuals into 15 baskets, five baskets based on the leverage ratio and three baskets for different maturities. CGM sample consisted of 688 bonds, or 45 bonds per each basket on average. Castagnetti and Rossi (2013) divided the regression residuals into 9 baskets, with three industrial sector baskets and three maturity baskets. The Castagnetti and Rossi (2013) sample was comprised of 207 bonds with, on average, only 23 bonds per basket. Nevertheless, both papers found the latent factor in the residuals to be very strong. For CGM, the first principal component across 15 residual baskets explained 75.9% of the total variance, while the first principal component across 9 residual baskets in Castagnetti and Rossi (2013) explained 64.9% of the total variance in the residuals. The results reported here show that the procedure of taking averages may be solely responsible for the strength and shape of the first principal component.

5. A more appropriate methodology for panels of credit spreads

We perform two additional tests of cross-sectional dependence in the residuals. The first test is based on Pesaran (2015), namely a test of cross-sectional dependence using the average of all pairwise correlation coefficients of the OLS residuals from the individual regressions in the panel. In our case, these are regressions of the (1) type. The test is generally applicable to a variety of panel data models and, in particular, for datasets where the cross-sectional dimension, \( N \), is large relative to the time series dimension, \( T \). The test can accommodate relatively large numbers of missing values, which is an important feature for panels with credit spreads.

The second test is based on Pesaran (2006). The Common Correlated Effect (CCE) estimator obtains consistent slope coefficients in individual regressions on a pre-defined factor. This estimator is useful in situations when it is difficult, or impossible, to compute latent factors by means of principal component analysis, which holds for panels of credit spreads. The only remaining issue is the form of the pre-defined factor, that is, the vector of weights. The most obvious choice is to use an equally-weighted vector that will mimic the market factor. Such a vector, in the context of the previously described pitfalls, is a logical application of the market portfolio from CGM and Castagnetti and Rossi (2013).

5.1. Another method for creating more balanced panels with credit spread changes

In this section we perform an analysis on a differently constructed panel of credit spreads. A time series of company observations that is as long as possible is constructed. For each company, we collect all the available bonds and, for each date, select a transaction of a bond that has maturity closest to 5 years, if there is more than one available transaction. In this way, we create only one time series of observations for each company. The underlying assumption is typical in the affine term structure literature, namely that the term structure of different credit spread changes obtained from different bonds of the same company can be described with a factor structure. This assumption is not testable as we do not have a sufficient number of overlapping observations, but it nevertheless seems reasonable. Given the methodology of creating panels of data, we implement three different lower bounds of observations per company, namely \( N_{\text{min},1} = 60 \), \( N_{\text{min},2} = 120 \), and \( N_{\text{min},3} = 180 \).

For the lower bound of observations \( N_{\text{min},1} = 60 \) per company, we obtain \( N_1 = 978 \) distinct company time series, so that the \( T \times N = 466 \times 978 \) matrix of residuals from regression (1), where each column represents a unique company, has at least 60 observations. For \( N_{\text{min},2} = 120 \) and \( N_{\text{min},3} = 180 \), we obtain, respectively, \( N_2 = 459 \) and \( N_3 = 201 \) observations.

For the \( T \times N = 466 \times 978 \) matrix of residuals, we cannot test whether they are weakly dependent using the Pesaran (2015) test, as this case is still highly unbalanced. However, using CCE, we can compute the equally-weighted market portfolio, and compute the correlation with each individual column of residuals separately. The median correlation of \( \eta_i \) with \( \bar{\eta}_N = \frac{1}{N} \sum_{j=1}^{N} \eta_j \) is 28.2%, and the distribution of all 978 correlations can be seen in Fig. 5. The median correlation is quantitatively similar to the strength of the average first principal component from the unbalanced portfolio of credit spreads that was obtained previously by means of probabilistic principal component analysis.

We obtain similar results for \( N_2 = 459 \) and \( N_3 = 201 \), with median correlations, respectively, of 25.6% and 21.4%. In the case of datasets with \( N_2 = 459 \) and \( N_3 = 201 \), with lower bounds for the minimum number of observations, \( N_{\text{min},2} = 120 \) and \( N_{\text{min},3} = 180 \), this allows the creation of panels of data that are sufficiently balanced for the Pesaran (2015) test to be implemented. The results of the test of cross-sectional independence are, respectively, 4.155 and 4.316, both with p-values zero, which lead to rejection of the null hypothesis of cross-sectional independence of the \( \eta_i \) residuals.

Finally, we assess the properties of the cross-sectional residual dependence with a simple test, namely the relation between company credit ratings and the correlations of the \( \eta_i \) residuals with the market portfolio of the residuals, where the market portfolio of residuals is simply a mean of all the residuals. Fig. 6 presents the histogram of all such correlation coefficients. This is complemented by regression analysis, where the correlation coefficients are regressed on their corresponding corporate credit ratings. Table 8 and Fig. 6 present details of the regression analysis. The analysis shows that the \( \eta_i \) residuals obtained from the time series regressions on their credit spread changes, tend to co-move much more strongly for higher-rated companies. This is consistent with the fact that structural models have difficulty in explaining credit spreads of highly-rated companies (see, for example, Huang and Huang, 2012).
Fig. 5. Histogram of correlation coefficients. X axis - Correlation of residuals in vector $\eta_i$ with the market factor of residuals, that is, $\hat{\eta}_N = \frac{1}{N} \sum_{i=1}^{N} \eta_i$. Y axis - Number of companies. The figure is for $N_{\text{min}} = 60$, which results in $N_1 = 978$ distinct company time series.

Fig. 6. X axis - Correlation of residuals $\eta_i$ with the market factor of residuals, that is, $\hat{\eta}_N = \frac{1}{N} \sum_{i=1}^{N} \eta_i$. Y axis - Company ratings measures as an average rating across all dates on which corporate bonds of the company are observed. Rating equal to 1 corresponds to AAA and 25 is a bankrupt company. The figure is for $N_{\text{min}} = 60$, which results in $N_1 = 978$ distinct company time series.

Table 8
Regression results.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Company_Rating</td>
<td>OLS</td>
<td>884</td>
<td>882</td>
<td>1</td>
<td>0.198</td>
<td>0.197</td>
<td>217.2</td>
<td>4.14e-44</td>
<td>-2346.4</td>
<td>4697</td>
<td>4706</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>coef</th>
<th>std err</th>
<th>t</th>
<th>P &gt;</th>
<th>t</th>
<th>[95.0% Conf. Int.]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>12.9828</td>
<td>0.212</td>
<td>61.348</td>
<td>0.000</td>
<td>12.567 13.398</td>
</tr>
<tr>
<td>Corr coeff</td>
<td>-8.0642</td>
<td>0.547</td>
<td>-14.737</td>
<td>0.000</td>
<td>-9.138 -6.990</td>
</tr>
</tbody>
</table>

Omnibus: 26.327 Durbin-Watson: 1.973
Prob(Omnibus): 0.000 Jarque-Bera (JB): 13.235
Skew: -0.053 Prob(JB): 0.00134
Kurtosis: 2.410 Cond. No. 5.24
6. Conclusion

A lingering puzzle in the credit risk literature is the existence of a strong latent factor that drives the co-movements in credit spread changes. This single common factor was purportedly driven by local supply and demand shocks, independently of both credit risk factors and proxies for liquidity. The latent factor appears to explain between 25% and 35% of the variance, which is much weaker than the latent factor that CGM (76%) and Castagnetti and Rossi (2013) (65%) claim to have found. We also show that the latent factor is unlikely to represent an equally-weighted market portfolio that is exposed evenly to any supply and demand shocks.

Nevertheless, the empirical results reported above should not be interpreted as evidence against the segmentation of the stock market from credit markets, as conjectured by CGM. The latent factor that has been uncovered here still appears to explain a non-negligible amount of the total variance. Moreover, companies with higher credit ratings were found to have stronger unexplained co-movements. The relationship of this latent factor to other characteristics would seem to be a challenging issue that is left for future research.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors are most grateful to the Associate Editor and two reviewers for very helpful comments and suggestions. For financial and research support, the second author wishes to thank the Australian Research Council and the Ministry of Science and Technology (MOST), Taiwan.

References


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


