



Instituto  
Complutense  
de Análisis  
Económico

# Spurious Cross-Sectional Dependence in Credit Spread Changes

**Marcin Jaskowski**

Econometric Institute, Erasmus School of Economics  
Erasmus University Rotterdam, The Netherlands

**Michael McAleer**

Department of Finance Asia University, Taiwan and  
Discipline of Business Analytics University  
of Sydney Business School, Australia and  
Econometric Institute, Erasmus School of Economics  
Erasmus University Rotterdam, The Netherlands and  
Department of Economic Analysis and ICAE  
Complutense University of Madrid, Spain and  
Institute of Advanced Sciences  
Yokohama National University, Japan

## Abstract

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 a new and improved estimator of the latent factor in credit spread changes and its characteristics.

**Keywords** Credit spread puzzle, Market segmentation, Latent factors, Spurious cross-sectional dependence.

**JEL Classification** G12, G13, G17, E43.

**Working Paper nº 1821**  
**September, 2018**



UNIVERSIDAD  
COMPLUTENSE  
MADRID

ISSN: 2341-2356

WEB DE LA COLECCIÓN: <http://www.ucm.es/fundamentos-analisis-economico2/documentos-de-trabajo-del-icae> Working papers are in draft form and are distributed for discussion. It may not be reproduced without permission of the author/s.

# Spurious Cross-Sectional Dependence in Credit Spread Changes\*

Marcin Jaskowski

Econometric Institute, Erasmus School of Economics  
Erasmus University Rotterdam, The Netherlands

Michael McAleer\*\*

Department of Finance  
Asia University, Taiwan  
and

Discipline of Business Analytics  
University of Sydney Business School, Australia  
and

Econometric Institute, Erasmus School of Economics  
Erasmus University Rotterdam, The Netherlands  
and

Department of Economic Analysis and ICAE  
Complutense University of Madrid, Spain  
and

Institute of Advanced Sciences  
Yokohama National University, Japan

Revised: August 2018

---

\***Acknowledgments:** For financial and research support, the second author wishes to thank the Australian Research Council and the Ministry of Science and Technology (MOST), Taiwan.

\*\***Corresponding author:** michael.mcaleer@gmail.com

## Abstract

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 a new and improved estimator of the latent factor in credit spread changes and its characteristics.

**Keywords:** Credit spread puzzle, Market segmentation, Latent factors, Spurious cross-sectional dependence.

**JEL:** G12, G13, G17, E43.

## 1 INTRODUCTION

In a seminal paper, [Collin-Dufresne, Goldstein, and Martin \(2001\)](#) (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 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.

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/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 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 literally so disturbing that several papers have tried to explain the purported puzzle, the most obvious explanation being the presence of inappropriate regressors. Indeed, [Cremers, Driessen, Maenhout, and Weinbaum \(2008\)](#) 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.

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 significant regressors, it is possible to generate a spurious factor in the residuals.

[Ericsson, Jacobs, and Oviedo \(2009\)](#) 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, and 22% of the variation in CDS spread changes, and also 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 corpo-

rate debt reasonably well.

An open question arises here, namely how [Avramov, Jostova, and Philipov \(2007\)](#), [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 entirely possible.

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 loadings may be time dependent, we will explain  $\Delta cs_{i,t}$  through the use of linear regressions. It follows that:

$$\Delta cs_{i,t} = \alpha_i + \Lambda_i X_{i,t} + \eta_{i,t}, \tag{1}$$

where it is assumed that  $\alpha_i$  and  $\Lambda_i$  are time independent, and the matrix  $X_{i,t}$  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.<sup>1</sup>

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

---

<sup>1</sup>Even if the choice of explanatory variables and  $X$  contains the appropriate regressors, the non-linear relation of  $X$  to  $\Delta cs$  will still lead to cross-correlations of  $\eta_i$  across different  $i \in N$ .

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.

The intuition is the following. Suppose that all the variation in the dataset,  $\eta_{(T \times N)}$ , is driven by just one factor 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 be singular as there is only one underlying factor. The eigenvector of the first principal component may take any shape, depending on the loadings of the factor for each time series. However, if we compute either the covariance or the correlation matrix for  $\tilde{\eta}_{(T \times K)}$ , we will obtain a matrix where all the entries are equal to 1, with the first eigenvector corresponding to the only non-zero eigenvalue where all its entries are equal to  $\sqrt{\frac{1}{K}}$ . This is explained formally below.

## 2.1 HEURISTIC ARGUMENT

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. Assume that the entries of this matrix are generated by the following one-factor process:

$$\eta_{i,t} = \beta_i f_t + \varepsilon_{i,t} \quad (2)$$

where  $t \in \{1, \dots, T\}$ ,  $i \in \{1, \dots, N\}$ , with  $f_t \sim N(0, \sigma_f^2)$ , and  $\varepsilon_{i,t} \sim N(0, \sigma_\varepsilon^2)$ . For  $\eta_{(T \times N)}$ , we obtain the following  $N \times N$  covariance matrix:

$$\text{cov}(\eta) = \sigma_f^2 \beta \beta' + \sigma_\varepsilon^2 I_N. \quad (3)$$

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  $\mathcal{N} = \{1, \dots, N\}$ , where  $k \in \{1, \dots, K\}$ . Let  $n_k$  be the cardinality of the  $\mathcal{N}_k$  partition of set  $\mathcal{N}$  to generate a new dataset,  $\tilde{\eta}_{(T \times K)}$ , with each time  $t$  entry of column  $k$  constructed according to:

$$\forall i \in \mathcal{N}_k : \tilde{\eta}_{k,t} = \frac{1}{n_k} \sum_{i=1}^{n_k} \eta_{i,t} \quad (4)$$

$$= \left( \frac{1}{n_k} \sum_{i=1}^{n_k} \beta_i \right) f_t + \frac{1}{n_k} \sum_{i=1}^{n_k} \varepsilon_{i,t} \quad (5)$$

$$= \tilde{\beta}_k f_t + \tilde{\varepsilon}_{k,t}. \quad (6)$$

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

$$\tilde{\varepsilon}_{k,t} \sim N \left( 0, \frac{1}{n_k} \sigma_\varepsilon^2 \right) \quad (7)$$

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

The covariance matrix for the new transformed dataset,  $\tilde{\eta}$ , is given as:

$$\text{cov}(\tilde{\eta}) = \sigma_f^2 \tilde{\beta} \tilde{\beta}' + \sigma_\varepsilon^2 I_K \quad (8)$$

where the off-diagonal elements of the covariance matrix take the form

$$\text{cov}(\tilde{\eta}_i, \tilde{\eta}_j) = \sigma_f^2 \tilde{\beta}_i \tilde{\beta}_j, \quad (9)$$

and the elements on the main diagonal are equal to

$$\text{cov}(\tilde{\eta}_i, \tilde{\eta}_i) = \sigma_f^2 \tilde{\beta}_i^2 + \sigma_\varepsilon^2. \quad (10)$$

Now consider the well-known fact from asset pricing that portfolio betas always tend to 1 when the portfolio is large enough,<sup>2</sup> namely:

$$\tilde{\beta} = \frac{1}{n_k} \sum_{i=1}^{n_k} \beta_i \rightarrow 1 \text{ as } n_k \rightarrow \infty.$$

Consequently, the off-diagonal elements of the covariance matrix, (9), approach the value of the main factor's

---

<sup>2</sup>Beta of each asset is obtained from the regression of this asset on the market factor, where the market factor is defined as an equally weighted portfolio. Hence, if we sample a large enough number of assets, representative of the whole population of assets, then an average over betas will be approximately equal to the market beta, which is 1.



variance,  $\sigma_f^2$ :

$$\text{cov}(\tilde{\eta}_i, \tilde{\eta}_j) = \sigma_f^2 \underbrace{\tilde{\beta}_i}_{\rightarrow 1} \underbrace{\tilde{\beta}_j}_{\rightarrow 1} \rightarrow \sigma_f^2 \quad \forall i, j \text{ as } n_k \rightarrow \infty. \quad (11)$$

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

$$\text{var}(\tilde{\varepsilon}) = \frac{1}{n_k} \sigma_\varepsilon^2 \rightarrow 0 \text{ as } n_k \rightarrow \infty.$$

The main diagonal entries tend to the same value as the off-diagonal entries because

$$\text{cov}(\tilde{\eta}_i, \tilde{\eta}_i) = \sigma_f^2 \underbrace{\tilde{\beta}_i^2}_{\rightarrow 1} + \underbrace{\sigma_\varepsilon^2}_{\rightarrow 0} \rightarrow \sigma_f^2 \quad \forall i \text{ as } n_k \rightarrow \infty. \quad (12)$$

Therefore, as  $n_k$  increases, the covariance matrix,  $\text{cov}(\tilde{\eta})$ , approaches a singular  $K \times K$  matrix with  $\sigma_f^2$  elements everywhere.

#### 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 (13), it follows that:

$$C_{K \times K} = \begin{bmatrix} \psi & \rho & \cdots & \rho \\ \rho & \psi & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \cdots & \psi \end{bmatrix} \quad (13)$$

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

$$v_1(C_{K \times K}) = \sqrt{\frac{1}{K}} [1, 1, \dots, 1]. \quad (14)$$

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 simulated data, but we prefer to use real corporate bond credit spread changes.

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) the assumption that the data generating process contains just one homoscedastic factor,  $f_t$ , is not precisely 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  $\sqrt{1/K}$ .

### 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 to 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 to 2012, from Thomson Reuters Datastream.

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.<sup>3</sup> 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:

$$P_{ijt}^C(T) = e^{-Y_{ijt}(T)T} + C \sum_{j=1}^N e^{-Y_{ijt}(T)t_j} \quad (15)$$

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:

$$cs_{ijt} = Y_{ijt}(T) - r(T). \quad (16)$$

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

---

<sup>3</sup>We are indebted to Yoshio Nozawa for access to part of this dataset.

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. \(2009\)](#):

$$\text{Leverage} = \frac{\text{Book Value of Debt}}{\text{Market Value of Equity} + \text{Book Value of Debt}} \quad (17)$$

$$= \frac{(\text{DD1Q} + \text{DLTTQ})}{\text{PRCCM} \times \text{CSHOQ} + (\text{DD1Q} + \text{DLTTQ})} \quad (18)$$

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<sup>4</sup>. 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, Verbeek, and Krose \(2005\)](#)) is that it fills missing values in the data differently for every run. We circumvent this problem by running

---

<sup>4</sup>After collecting all bonds for a company, we include a bond with the highest number of observations.

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<sup>5</sup>. 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 6) use the proper method of measuring the strength of the latent factor in the residuals, and avoid 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-equally distributed  $n_k$  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 portfolio and very large for another, 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.

### 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.

---

<sup>5</sup>On every run, we sample  $r$  bonds randomly from 872 bonds.

Table 1: **Only one bond per company: 872 corporate bonds**

	mean	sd	mean	sd	mean	sd
<b>Panel A. Separate K bonds, without averaging</b>						
	r = 300		r = 200		r = 100	
$\lambda_1$	0,2866	0,0486	0,2866	0,0486	0,2199	0,0593
$\lambda_2$	0,0513	0,0124	0,0584	0,0126	0,0697	0,0130
<b>Panel B. Portfolios</b>						
K portfolios constructed from $n_k$ bonds						
		Random samples				
K = 5, $n_k = 175$	$\lambda_1$	0,7591	0,0614			
	$\lambda_2$	0,0745	0,0205			
K = 15, $n_k = 59$	$\lambda_1$	0,5316	0,0168		15 baskets as in CGM	
	$\lambda_2$	0,0563	0,0074		0,4322	0,0015
K = 30, $n_k = 30$	$\lambda_1$	0,3830	0,0157		0,0902	3,28E-04
	$\lambda_2$	0,0438	0,0037			
K = 50, $n_k = 18$	$\lambda_1$	0,3045	0,0132			
	$\lambda_2$	0,0351	0,0027			
K = 100, $n_k = 9$	$\lambda_1$	0,2683	0,0258			
	$\lambda_2$	0,0115	0,0019			

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 6 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, we might 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,  $\{t_{min}^i, t_{max}^i\}$ , for each bond  $i \in \{1, \dots, N\}$  when the bond starts and when it appears for the last time in the dataset. Based on this pair, we find a bond,  $s \in \{1, \dots, N\}$ , such that its first and last date pair,  $\{t_{min}^s, t_{max}^s\}$ , intersect with the largest number of other bond pairs,  $\exists s \forall i : (t_{min}^s \geq t_{min}^i \text{ and } t_{max}^s \leq t_{max}^i)$ . We select all of those bonds, then omit these bonds that have more than 10 missing observations within the prespecified dates,  $\{t_{min}^s, t_{max}^s\}$ . This procedure necessarily results in a dataset with a short time series. Finally, we obtain a dataset,  $\Xi_{(T \times N)}$ , 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 5 present similar information, but for different samples of data. The dataset used for Table 5 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.

Table 2: All 2983 bonds

	mean	sd	mean	sd	mean	sd
<b>Panel A. Separate K bonds, without averaging</b>						
	r = 300		r = 200		r = 100	
$\lambda_1$	0,3529	0,0681	0,3159	0,0753	0,2689	0,0833
$\lambda_2$	0,0467	0,022	0,0574	0,0242	0,0724	0,0228
<b>Panel B. Portfolios</b>						
K portfolios constructed from $n_k$ bonds						
		Random samples				
K = 5, $n_k = 597$	$\lambda_1$	0,8692	0,0728			
	$\lambda_2$	0,0571	0,0191			
K = 15, $n_k = 199$	$\lambda_1$	0,7486	0,0135		15 baskets as in CGM	
	$\lambda_2$	0,053	0,0116		0,6167	3,95E-04
K = 30, $n_k = 100$	$\lambda_1$	0,6476	0,0107		0,0857	8,21E-05
	$\lambda_2$	0,0372	0,0052			
K = 50, $n_k = 60$	$\lambda_1$	0,5597	0,0102			
	$\lambda_2$	0,0288	0,0032			
K = 100, $n_k = 30$	$\lambda_1$	0,4421	0,0096			
	$\lambda_2$	0,0234	0,0024			

Table 3: **Robustness check - Only one bond per company: 263 bonds**

	mean	sd	mean	sd	mean	sd
<b>Panel A. Separate K bonds, without averaging</b>						
	r = 200		r = 100		r = 50	
$\lambda_1$	0,2851	0,0072	0,2878	0,0164	0,2957	0,0261
$\lambda_2$	0,1187	0,0035	0,1207	0,0079	0,1249	0,012 6
<b>Panel B. Portfolios</b>						
K portfolios constructed from $n_k$ bonds						
		Random samples				
K = 5, $n_k = 53$	$\lambda_1$	0,7561	0,0737			
	$\lambda_2$	0,0749	0,0201			
K = 15, $n_k = 18$	$\lambda_1$	0,5318	0,0168		15 baskets as in CGM	
	$\lambda_2$	0,056	0,008		0,5307	2,17E-14
K = 30, $n_k = 9$	$\lambda_1$	0,3515	0,0211		0,1131	5,34E-15
	$\lambda_2$	0,048	0,0056			
K = 50, $n_k = 6$	$\lambda_1$	0,3043	0,0168			
	$\lambda_2$	0,0361	0,0038			
K = 100, $n_k = 3$	$\lambda_1$	0,2714	0,0125			
	$\lambda_2$	0,0261	0,0022			



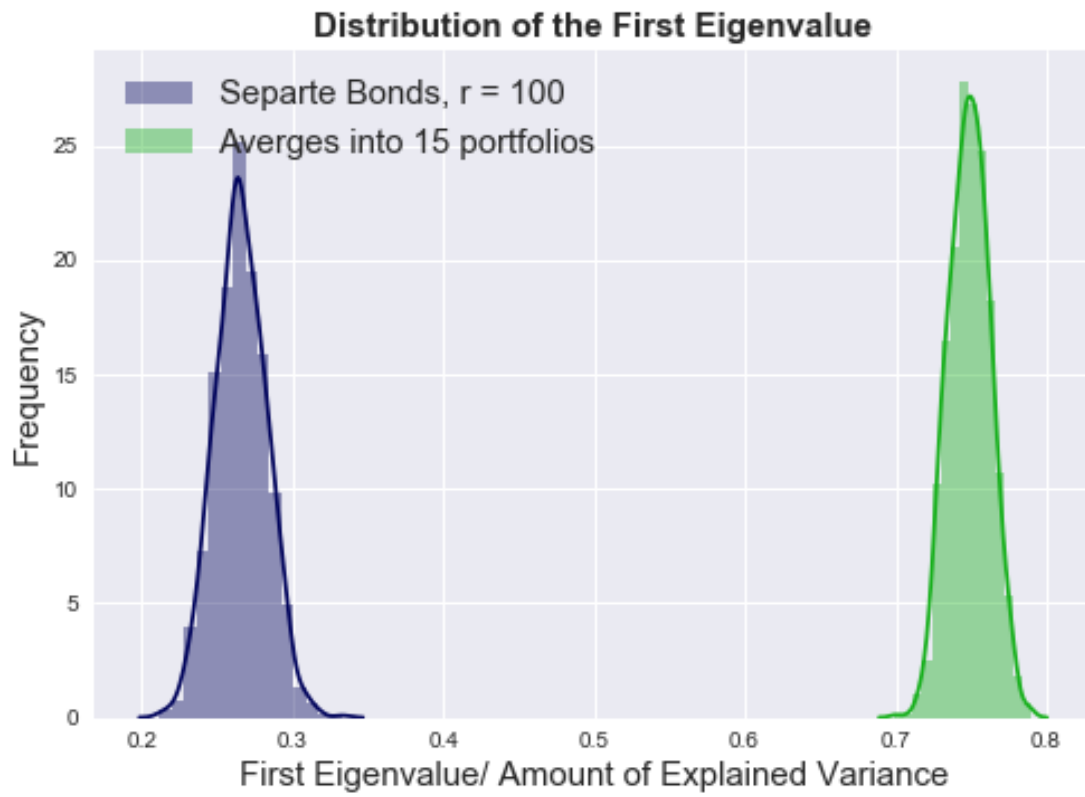


Figure 1: Distribution of first eigenvalue corresponding to Table 4

Figure 1 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 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<sup>6</sup>, with mean eigenvalue of 0.7487.

### 3.6 LONG MATURITY BONDS

Figure 2 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 6 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 6.

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 subsection 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,  $\mathcal{N} = \{1, \dots, N\}$ , into  $K$  partitions with equal cardinality,<sup>7</sup>  $n_k$ , so that we obtain a new transformed,  $\tilde{\Xi}_{(T \times K)}$ , dataset.

The same dataset is used as in Panel B of Table 5, but it is split into two subsamples. In Table 4 for Panel

---

<sup>6</sup>The procedure is as follows. First, we divide the set 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 (using `nanmean()` in Matlab) 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.

<sup>7</sup>If  $\text{mod}(K, n_k) \neq 0$ , then  $n_k = \text{ceil}(\frac{N}{K})$  holds for most of the baskets.

Table 4: Variance of the first eigenvector as a function of  $n_k$

		K portfolios						K randomly chosen bonds, no averages						
K	$n_k$	$\sqrt{1/K}$	$\min(v_1)$	$\max(v_1)$	$\min(v_1) - \max(v_1)$	$\lambda_1$	$\min(v_1)$	$\max(v_1)$	$\min(v_1) - \max(v_1)$	$\lambda_1$	$\min(v_1)$	$\max(v_1)$	$\min(v_1) - \max(v_1)$	$\lambda_1$
<b>Panel A: Bonds with maturity shorter or equal to 28 years - 585 bonds</b>														
150	4	0,082	0,001	0,117	0,116	0,413	-0,028	0,137	0,165	0,268				
100	6	0,100	0,011	0,138	0,127	0,461	-0,018	0,158	0,175	0,275				
50	12	0,141	0,067	0,185	0,118	0,559	-0,001	0,209	0,210	0,288				
30	20	0,183	0,136	0,217	0,082	0,674	0,002	0,270	0,268	0,289				
15	39	0,258	0,219	0,275	0,056	0,819	-0,082	0,402	0,484	0,282				
10	59	0,316	0,308	0,325	0,018	0,888	-0,017	0,522	0,539	0,286				
5	117	0,447	0,444	0,451	0,007	0,951	-0,227	0,468	0,695	0,293				
<b>Panel B: All bonds</b>														
150	5	0,082	0,002	0,110	0,108	0,438	-0,090	0,128	0,218	0,269				
100	8	0,100	0,015	0,130	0,115	0,488	-0,086	0,148	0,234	0,276				
50	15	0,141	0,058	0,180	0,122	0,583	-0,166	0,227	0,393	0,284				
30	24	0,183	0,141	0,203	0,062	0,732	-0,193	0,275	0,468	0,283				
15	48	0,258	0,227	0,274	0,047	0,830	-0,227	0,366	0,593	0,272				
10	72	0,316	0,306	0,324	0,019	0,900	-0,041	0,496	0,536	0,278				
5	143	0,447	0,445	0,449	0,005	0,957	-0,570	0,465	1,035	0,331				

Table 5: **Robustness check - all 714 bonds**

	mean	sd	mean	sd	mean	sd
<b>Panel A. Separate K bonds, without averaging</b>						
	r = 300		r = 200		r = 100	
$\lambda_1$	0,2602	0,0080	0,2627	0,0109	0,2654	0,0163
$\lambda_2$	0,1094	0,0042	0,1112	0,0057	0,1129	0,0086
<b>Panel B. Portfolios</b>						
K portfolios constructed from $n_k$ bonds						
	Random samples.					
K = 5, $n_k = 143$ , r = 100	$\lambda_1$	0,8709	0,0642			
	$\lambda_2$	0,0566	0,0189			
K = 15, $n_k = 48$ , r = 100	$\lambda_1$	0,	0,0137		15 baskets as in CGM	
	$\lambda_2$	0,053	0,0118		0,8275	5,4428e-15
K = 30, $n_k = 24$	$\lambda_1$	0,6482	0,0104			
	$\lambda_2$	0,037	0,0051			
K = 50, $n_k = 15$	$\lambda_1$	0,5601	0,0101			
	$\lambda_2$	0,0287	0,0033			
K = 100, $n_k = 8$	$\lambda_1$	0,4422	0,0098			
	$\lambda_2$	0,0235	0,0024			

Table 6: **Strength of the latent factor across different maturities**

	mean	sd	mean	sd	mean	sd	mean	sd	mean	sd
	$\tau \in [0, \infty)$		$\tau \in [0, 12)$		$\tau \in [12, 28)$		$\tau \in [12, \infty)$		$\tau \in [28, \infty)$	
<b>Panel A. One bond per company</b>										
$\lambda_1$	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 = 70, N = 147 r = 50		T = 74, N = 26 r = 20		T = 74, N = 45 r = 20		T = 69, N = 53 r = 50	
<b>Panel B. More than one bond per company</b>										
$\lambda_1$	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 = 70, N = 405 r = 50		T = 74, N = 64 r = 50		T = 74, N = 107 r = 50		T = 71, N = 93 r = 50	

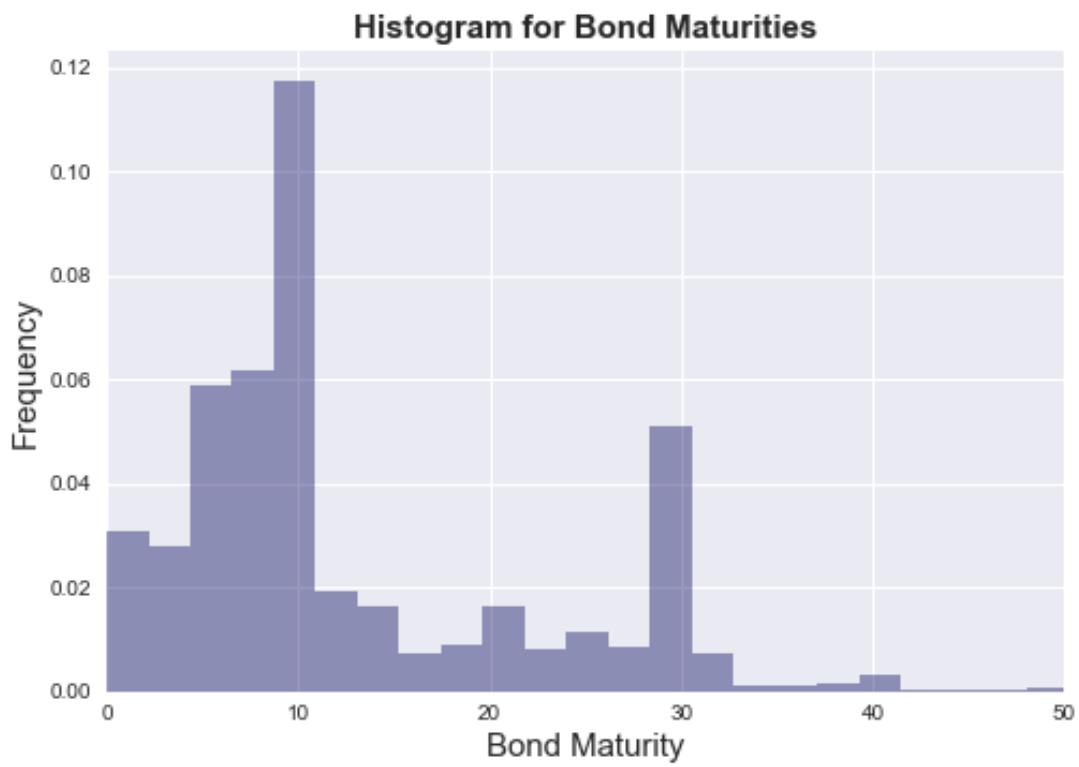


Figure 2: X axis - maturity, Y axis - frequency/number of bonds

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 6 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,  $f_t$ , from equation (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 4 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 equation (14), namely the value of an equally-weighted normalized eigenvector of dimension  $K$  for any matrix of the form (13). 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 difference 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<sup>8</sup>.

From  $\Xi_{(T \times N)}$ , we construct seven different transformed datasets,  $\tilde{\Xi}_{(T \times K)}$ , keeping  $T$  fixed and changing  $K$  and  $n_k$ . For example, the first row in Table 4 represents the results for  $K = 150$  and  $n_k \approx 4$ . 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,  $\max(v_1) - \min(v_1)$ , 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  $\sqrt{1/K}$  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  $e_1$ , 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  $\max(v_1) - \min(v_1)$ . Moreover,  $\max(v_1) - \min(v_1)$  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

---

<sup>8</sup>For columns eight to eleven, we have  $n_k = 1$  as there is no averaging into portfolios.

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

K	r	$\binom{K}{r}$	$n_k$	$\sqrt{1/K}$	$\min(v_1)$	$\max(v_1)$	$\max(v_1) - \min(v_1)$	mean of $s d(v_1)$	$\text{mean}(\lambda_1)$	$\text{sd}(\lambda_1)$
Panel A: r = 15										
150	15	1,62E+20	4	0,258	-0,062	0,395	0,457	0,063	0,446	0,044
100	15	2,53E+17	6	0,258	-0,020	0,380	0,400	0,058	0,490	0,042
50	15	2,25E+12	12	0,258	0,062	0,359	0,297	0,046	0,578	0,040
30	15	1,55E+08	20	0,258	0,171	0,322	0,151	0,025	0,685	0,025
Panel B: r = 5										
150	5	5,92E+08	4	0,447	-0,605	0,686	1,291	0,098	0,521	0,072
100	5	7,53E+07	6	0,447	-0,211	0,687	0,898	0,084	0,558	0,071
50	5	2,12E+06	12	0,447	0,029	0,631	0,602	0,065	0,632	0,071
30	5	1,43E+05	20	0,447	0,285	0,542	0,256	0,032	0,727	0,050
15	5	3 003	39	0,447	0,376	0,486	0,110	0,017	0,844	0,030
10	5	252	59	0,447	0,429	0,461	0,032	0,006	0,900	0,010



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<sup>9</sup>. 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,  $\max(v_1) - \min(v_1)$ , or standard deviation of the entries in  $v_1$ . The results are very clear. In both panels A and B, for  $r = 15$  and  $r = 5$ , respectively, the increase in  $n_k$  leads to a smaller width of  $\max(v_1) - \min(v_1)$ , 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 4 and 7 is that the size of the correlation matrix does not influence the "equal weightedness" of the first eigenvector. The interval,  $\max(v_1) - \min(v_1)$ , 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 equations (11) and (12). 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.

---

<sup>9</sup>More precisely, we repeat it 10,000 times if  $\binom{K}{r} \geq 10,000$ , otherwise  $\binom{K}{r}$  times.

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  $N_{\min,1} = 60$ ,  $N_{\min,2} = 120$ , and  $N_{\min,3} = 180$ .

For the lower bound of observations  $N_{\min,1} = 60$  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  $N_{\min,2} = 120$  and  $N_{\min,3} = 180$ , 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  $\tilde{\eta}_N = \frac{1}{N} \sum_{i=1}^N \eta_i$  is 28.2%, and the distribution of all 978 correlations can be seen in [Figure 3](#). 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,  $N_{\min,2} = 120$  and  $N_{\min,3} = 180$ , this allows the creation of panels of data that are sufficiently balanced for the [Pesaran \(2015\)](#) test to be implemented. The results of the Pesaran test of cross-sectional independence are, respectively, 4.155 and 4.316, both with p-values of 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. [Figure 4](#) 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 [Figure 4](#) 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 \(2012\)](#)).

## 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/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/demand shocks.

Nevertheless, the 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.

Table 8: **Regression Results**

<b>Dep. Variable:</b>	Company_Rating	<b>R-squared:</b>	0.198		
<b>Method:</b>	OLS	<b>Adj. R-squared:</b>	0.197		
<b>No. Observations:</b>	884	<b>F-statistic:</b>	217.2		
<b>df Residuals:</b>	882	<b>Prob (F-statistic):</b>	4.14e-44		
<b>df Model:</b>	1	<b>Log-Likelihood:</b>	-2346.4		
		<b>AIC:</b>	4697.		
		<b>BIC:</b>	4706.		
	coef	std err	t	P> t	[95.0% Conf. Int.]
<b>Intercept</b>	12.9828	0.212	61.348	0.000	12.567 13.398
<b>Corr coeff</b>	-8.0642	0.547	-14.737	0.000	-9.138 -6.990
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		

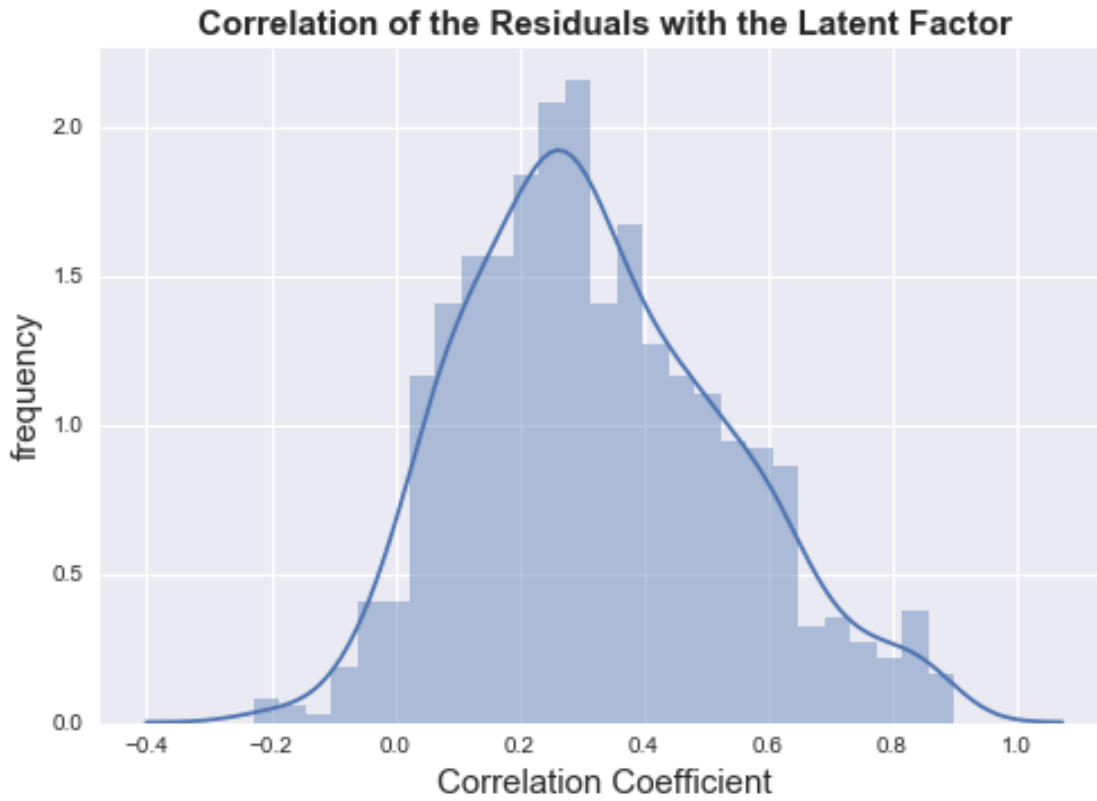


Figure 3: Histogram of correlation coefficients. X axis - Correlation of residuals in vector  $\eta_i$  with the market factor of residuals, that is,  $\tilde{\eta}_N = \frac{1}{N} \sum_{i=1}^N \eta_i$ , Y axis - Number of companies. The figure is for  $N_{\min,1} = 60$ , which results in  $N_1 = 978$  distinct company time series.

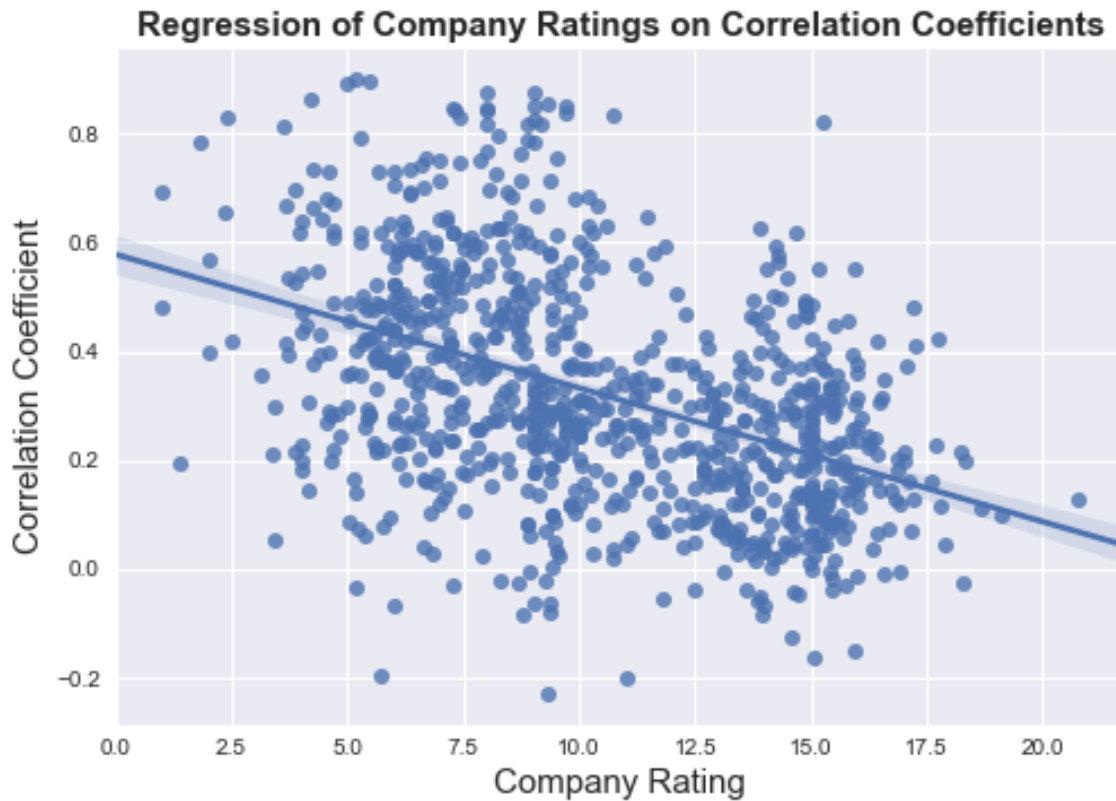


Figure 4: X axis - Correlation of residuals  $\eta_i$  with the market factor of residuals, that is,  $\hat{\eta}_N = \frac{1}{N} \sum_{i=1}^N \eta_i$ , Y axis - Company ratings measures as an average rating across all dates on which corporate bonds of the company are observed. Rating equal to 1 corresponds to AAA and 25 is a bankrupt company. The figure is for  $N_{\min,1} = 60$ , which results in  $N_1 = 978$  distinct company time series.

## REFERENCES

- Avramov, Doron, Gergana Jostova, and Alexander Philipov, 2007, Understanding changes in corporate credit spreads, *Financial Analysts Journal*, 63, 90–105.
- Castagnetti, Carolina, and Eduardo Rossi, 2013, Euro corporate bond risk factors, *Journal of Applied Econometrics*, 28, 372–391.
- Collin-Dufresne, Pierre, Robert S Goldstein, and J Spencer Martin, 2001, The determinants of credit spread changes, *Journal of Finance*, 56, 2177–2207.
- Cremers, Martijn, Joost Driessen, Pascal Maenhout, and David Weinbaum, 2008, Individual stock-option prices and credit spreads, *Journal of Banking and Finance*, 32, 2706–2715.
- Ericsson, Jan, Kris Jacobs, and Rodolfo Oviedo, 2009, The determinants of credit default swap premia, *Journal of Financial and Quantitative Analysis*, 44, 109–132.
- Huang, Jing-Zhi, and Ming Huang, 2012, How much of the corporate-treasury yield spread is due to credit risk?, *Review of Asset Pricing Studies*, 2, 153–202.
- Pesaran, M Hashem, 2006, Estimation and inference in large heterogeneous panels with a multifactor error structure, *Econometrica*, 74, 967–1012.
- Pesaran, M Hashem, 2015, Testing weak cross-sectional dependence in large panels, *Econometric Reviews*, 34, 1089–1117.
- Porta, Josep, Jakob Verbeek, and Ben Krose, 2005, Active appearance-based robot localization using stereo vision, *Autonomous Robots*, 18, 59–80.
- Schaefer, Stephen M., and Ilya A. Strebulaev, 2008, Structural models of credit risk are useful: Evidence from hedge ratios on corporate bonds, *Journal of Financial Economics*, 90, 1–19.