

Credit Spreads, Factors and Noise*

JOB MARKET PAPER

Marcin Jaskowski[†]

Vienna Graduate School of Finance

November 28, 2010

Abstract

Linear regressions and latent factors in the residuals may give untrustworthy evidence against structural models of credit risk. With a large dataset of credit default swap spreads (CDSs), this paper shows that previous studies might have spuriously generated a strong latent factor, which appeared to drive the co-movement in the data. The existence of a strong latent factor is one of the main arguments for the misspecification of structural models of credit risk. However, a careful signal-to-noise analysis of the latent factor and some insights from the random matrix theory reveal that structural models need not be so badly misspecified. Changes in spreads of CDSs are very noisy and difficult to explain in linear regressions even with very good regressors. Moreover, I show that a weak latent factor can also be generated by a structural model of debt.

1 Introduction

The main objective of this paper is to improve the reputation of structural models of debt. In an important calibration study, Huang and Huang (2003) found that several structural models cannot match the true level of credit spreads when their parameters are calibrated to the observed default frequencies. More recent papers have found new ways to solve or at least circumvent the problem of credit spread under-prediction. This paper addresses a related problem of the alleged structural models' misspecification and credit market segmentation. First, I show that the evidence for misspecification is much weaker than it was thought. Second, using simulations from a structural

*I thank Wolfgang Bühler, Alois Geyer, Klaus Ritzberger, Leopold Sögner, Ilya Strbulaev, Josef Zechner and conference participants at EFA, Doctorial Tutorial (Frankfurt 2010) for all comments and discussions.

[†]E-mail: marcin.jaskowski@vgsf.ac.at

model, I show that statistical patterns that might have been regarded as a proof of misspecification are in fact consistent with the structural theory.

The focus of this paper is the empirical performance of structural models of credit risk. One will never be able to show that a particular model is correct, but it is always possible to find evidence that falsifies the model's assumptions and main claims. The seminal paper that had the most devastating effect for the reputation of structural models was Collin-Dufresne, Goldstein and Martin (2001, henceforth CGM). CGM run regressions of various regressors implied by structural models, like firm-specific, macro and liquidity measures, on the changes in credit spreads of corporate bonds. They find that more than half of the variation is unexplained by their regressors. But a small R^2 did not seem to be very surprising, neither to the practitioners nor to the academics: first of all, because it is not uncommon in social sciences, and second, it may just reflect a high level of idiosyncratic noise. However, what was really puzzling in CGM was 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.

That was bad news to the structural models, as it showed that they give a misspecified description of the credit spreads and consequently of capital structure itself. CGM interpreted their findings as evidence for the segmentation of bond and equity markets. Different investors would trade in stocks rather than in bonds. The prices in those markets could be driven by independent supply/demand shocks in both markets. Yet it is not easy to find a convincing justification for such segmentation of the markets, and to answer the question why equity and bonds should react differently to the same aggregate factors. The practical implication of CGM's paper was that structural models of credit risk are severely misspecified and thus untrustworthy. Moreover, the unobserved, latent factor seemed to have a rather simple structure, since it was constructed from the eigenvector of almost equally large entries. One would expect that a latent factor represented by the eigenvector with equally large entries should be easy to filter out of the data, especially, as it is difficult to imagine a simpler structure for the latent factor than the eigenvector with equal weights. On that account, the best choice was to use reduced-form models. Reduced-form models are agnostic by assumption and they do not attempt to explain the phenomenon, they only attempt to describe it accurately. Therefore, we might expect them to fit the data better.

Yet, new empirical studies show that the reduced-form models have a better fit to the data but only in-sample. In a recent paper, Gündüz and Uhrig-Homburg (2008) calibrate both structural and reduced-form models to the information extracted from balance sheet data, equity and corporate bonds. Then in the second step, in order to test the out-of-sample performance of the models, they compare their ability to explain CDS prices. Eventually, it turns out that the difference in the out-of-sample predictive power of the two methods is statistically insignificant.

This paper shows that the evidence for the existence of such a latent factor is dubious. In fact, I find that changes of CDSs are very noisy and that the factor structure of the data is not stable

over time. That means that even with very good regressors, it is still difficult to remove all the correlation from the residuals. However, that does not mean that residuals contain one strong and stable common factor. I present two different ways how one can spuriously generate such a latent factor. This way, the importance of a common factor that explains only a relatively small amount of variance can be unintentionally magnified to a very large proportion of the total variance. The good news is then that structural models of credit risk may after all be better than it was thought.

On the technical side, this paper develops a new numerical bootstrap procedure to determine the number of factors in non-i.i.d. data, under the assumption of no cross-correlation between the residuals. This algorithm allows us to apply insights from random matrix theory to the a priori unspecified non-i.i.d. dataset, where the data generating process is unknown. The essence of the algorithm lies in the construction of the new dataset, from the given data, where any cross-correlation is destroyed, and then using it to construct a null hypothesis. I use this algorithm to remove the correlation from the dataset of CDSs, and then to estimate the possible bias of eigenvalues due to a sample's high-dimension.

The paper is organized as follows. The remainder of this introduction describes related literature and the dataset used in the paper. Section 2 explains the methodology and Section 3 presents results of regressions. Section 4 presents how the strength of the latent factor may be overestimated. Section 5 presents a problem of misspecification and market segmentation from the perspective of the theoretical models. Section 6 concludes. Appendix provides some additional robustness checks, cites theorems from the literature that were used in the main part of the paper and also gives the details of David (2008) model used in Section 5.

1.1 Literature review

Results in CGM were so disturbing that already a couple of papers tried to figure out whether anything might have gone wrong. The most obvious culprits are bad regressors. Indeed, Cremmers, 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 implied-volatility skew, while CGM used just the VIX as the aggregate proxy for the volatility of each company. Then the principal component analysis on the residuals from regressions does not reveal any significant omitted factors. One caveat here is that they run their regressions on the levels of CDSs and not on changes like CGM. They just assume that credit spreads are stationary. Theory says that they should be - although empirically it is not that certain at all. Then all the regressions at levels have a very high R^2 and regressors have strong explanatory power. Nevertheless, what is most important from the point of view of this paper, is that Cremmers et al. (2008) find no evidence of one large unidentified factor that would be unrelated to credit risk. Later on in this paper, I show that even with very good regressors one

might generate a spuriously strong factor in the residuals.

Ericsson, Jacobs and Oviedo (2004) 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 variation at levels and above 22% of variation in CDS spreads changes. They also find some evidence for a common factor in the residuals. However, the common factor is substantially weaker and explains approximately 32% of the variance in residuals from regressions on changes in CDSs and not 76% as in CGM. Another difference to CGM was that the eigenvector of the main principal component in residuals had both positive and negative elements.

The two above studies also perform panel data estimation. Results from panel regressions are not directly comparable to CGM's results in all respects. One can compare the explanatory power of particular theoretical variables. But it may be hard to interpret differences between systematic latent factors left in the residuals from univariate and panel regressions. In general, it was found that R^2 from both techniques are roughly at the same level. But in each technique different theoretical variables appear to be more successful in explaining the variation in CDS premia.

Schaefer and Strebulaev (2008) find that poor performance of structural models may be connected to the influence of non-credit factors present in the bond price data. But they show that even the simplest structural model can predict accurate equity hedge ratios. Thus, structural models can estimate properly at least the credit exposure of corporate debt.

A different but related issue was raised by Drehmann, Patton, Sorensen (2006). They use aggregate data on corporate credit in the UK in order to investigate the non-linear transmission of macroeconomic shocks to aggregate corporate default probabilities. First they generalize the vector autoregression model (VAR) setup and interpret it as a first-order Taylor series approximation of an unknown data generating process. This allows them to estimate more flexible impulse-response functions which capture possible nonlinearities in the data. Then in the next step they find that nonlinearities matter for the level and shape of impulse response functions of credit risk, following small as well as large shocks to systematic risk factors. Barrieu and Giammarino (2008) examine the relation between the iTraxx Europe index and several traded risk factors, like interest rate, stock index returns and volatility index. They find that empirical weights of the systematic factors display sudden jumps during market crises and a less intense time-dependent behavior during normal market conditions. In particular, in normal market conditions the risk factor weights are relatively weak but their signs are consistent with economic intuition and with earlier empirical findings. But in the midst of market turmoil, the magnitudes of the estimated relations change significantly and their directions defy ordinary economic explanations. These two studies show that linear regressions may be insufficient to extract relevant information from the data.

An open question arises here. How is it possible that Cremmers et al. (2008) and Ericsson et al. (2004) found so diametrically different conclusions from CGM about the missing factor? Can this difference be attributed only to the better dataset? This paper explains why such different

conclusions were possible.

1.2 Data

In this paper, I use data on corporate default swap spreads (CDSs) as a proxy for the theoretical credit spread. CDS contracts play the role of insurance contracts that allow investors to buy protection against the event that a firm defaults on its debt. An important advantage of using CDS data is that it is typically more liquid than the corresponding bond market, resulting in more accurate estimates of credit spreads. The claim is that CDSs are a cleaner proxy of the theoretical credit spread than the spreads extracted from corporate bonds. So, if there are any latent factors or liquidity problems present in the CDSs, then they are much less pronounced than in the market of corporate bonds. Furthermore, any biases or estimation artefacts that might affect our conclusions about the properties of credit spreads from CDSs will be at least as strong when we extract spreads from corporate bonds. This claim seems to be generally held and should not be very controversial. The same assumption appeared for instance in Berndt, Douglas, Duffie, Ferguson and Schranz (2004), Hull, Predescu and White (2004) and Longstaff, Mithal and Neis (2004).

All the balance sheet data were collected from COMPUSTAT, while data on stock prices and firms' capitalization and all macroeconomic variables are from Datastream. The data on CDSs come from Markit Group. These are daily data that span approximately 4.5 years from January 1, 2004 until May 30, 2008. However, in this study I restrict attention only to those companies for which it is possible to match the CDS data with COMPUSTAT and Datastream. The final sample retains only US firms that do not take part in M&A activity during the sample period and have a small number of missing values (less than 15%). Altogether the sample comprises 177 US firms. Non of these companies defaulted during the sample period.

2 Problems with Eigenvalues in High Dimensional Data

Recently, large quantities of financial data have become more readily available, from high frequency observations and for many firms. However, the data abundance is both an opportunity and a challenge. That is because most of the econometric methods are based on the assumption that the time dimension is many times larger than the cross-section. Yet in practice, the cross-sectional dimension is often very high, sometimes close to the number of time points. An important point is that careless ignorance of the dimensionality problem may backfire. Therefore we need to make use of new methods that accommodate high-dimensional datasets. Moreover, using new and better datasets, it is possible to reconsider some of the older results.

2.1 Marčenko and Pastur distribution for i.i.d. data

Assume that we have a very high dimensional $T \times N$ matrix X of data, where T and N are of the same order of magnitude and all $X_{i,j}$ entries are i.i.d. Then the true covariance matrix is an identity matrix I and obviously the population eigenvalues are all equal to 1. However, the sample eigenvalues distribution in this case will almost certainly be rather bizarre. The distribution function of eigenvalues was provided by Marčenko and Pastur (1967, MP henceforth), for $N, T \rightarrow \infty$ and $\gamma = N/T$ is constant¹, we get

$$\begin{aligned} f^{MP}(\lambda) &= \frac{\sqrt{(\beta_+ - \lambda)(\lambda - \beta_-)}}{2\pi\lambda\gamma} \\ \beta_{\pm} &= (1 \pm \sqrt{\gamma})^2 \\ \lambda &\in [\beta_-, \beta_+] \end{aligned}$$

So for $T = N$, sample eigenvalues will stretch between $(1 - \sqrt{1})^2 = 0$ and $(1 + \sqrt{1})^2 = 4$, although the population eigenvalues are all exactly equal to 1. It is important to note here that the variation of in-sample eigenvalues is around four orders of magnitude. Here one needs some supporting theory to explain the strange behavior of eigenvalues. Otherwise, one might erroneously infer from the sample that the population eigenvalues differ from each other and that there exists at least one strong factor. This has important practical consequences if we intend to use principal component analysis. That is the case because the MP distribution defines the interval where the eigenvalues and corresponding to them eigenvectors should be regarded either as pure noise or as indistinguishable from noise.

The distortion of the sample eigenvalues I will call a bias. For the MP distribution the bias of the highest sample eigenvalue with respect to the population eigenvalue is equal to $\beta_+ - 1$.

2.2 Spiked population model

Another important implication of the MP distribution is that it allows us to determine the number of true factors in the data. If a particular eigenvalue does not fall within the MP distribution then we can regard it as not noisy and thus holding some meaningful information. This result is an implication of MP, but it has been shown only recently. Here, I will briefly describe this result following Paul (2007).

Assume that X_1, X_2, \dots, X_T is an N -variate real Gaussian distribution with mean zero and covariance $\Sigma = \text{diag}(l_1, l_2, \dots, l_M, 1, \dots, 1)$ where $l_1 > l_2 > \dots > l_M > 1$ and $\gamma = N/T$ with $N, T \rightarrow \infty$. This type of dataset is called "spiked population model" (see Johnstone (2006)). It is "spiked", because it has most of the population eigenvalues equal to one and only a few eigenvalues which

¹ N and T are increasing with the same speed, so the ratio $\gamma = N/T$ is constant.

are larger than one. Population eigenvalues equal to one correspond to noisy eigenvectors, while all larger than one contain some signal.

Paul (2007) shows that sample eigenvalues exhibit "phase transition phenomenon". Simply, if the non-unit population eigenvalues l_i are close to one, then their sample versions will behave in roughly the same way as if the true covariance were the identity. However, if the population eigenvalues are larger than $1 + \sqrt{\gamma}$, the sample eigenvalues can be separated and distinguished from the noise. In the appendix, I cite two theorems from Paul (2007), which state that upper boundary of noisy part is exactly equal to β_+ from MP distribution.

Given this information, all we have to do is to compute the covariance matrix of the dataset X and compare its eigenvalues against β_+ . This is possible because the MP distribution has a compact support. In other words, β_+ defines for us the region of no correlation between the columns of the data. So we can design a simple test to determine the number of factors. The main steps of the test are the following,

1. For the sample covariance matrix, compute the set of eigenvalues λ_i and sort them in an ascending order, such that $\lambda_i \leq \lambda_j$ for all $i < j$.
2. Use the upper bound of the MP distribution, $\beta_+ = (1 + \sqrt{\gamma})^2$, as the null hypothesis H_0 that the columns of the dataset are uncorrelated.
3. Any eigenvalue λ_i that is above the value of $\beta_+ < \lambda_i$ indicates that we have found a factor. For $\beta_+ < \lambda_i$ we have exactly $N - i + 1$ factors in the data. For example, if $\beta_+ < \lambda_{N-1}$ then we have 2 factors.

2.3 Non-i.i.d. noise. Onatski's test for the number of factors.

There is, however, one simple problem with the MP distribution - it assumes that the data has i.i.d. distribution. Unfortunately, this does not happen too often. Onatski (2009b) develops a new test for the number of factors in the data, under weaker assumptions than in any other similar paper so far. In Onatski (2009b), the data are assumed to be generated by the following process

$$X = F\Lambda + e$$

where X is a $T \times N$ matrix of data, with N columns and observed over T periods. F is a $T \times r$ matrix of factor values and Λ is an $r \times N$ matrix of loadings. Critical here are the assumptions about the distribution of e . The $T \times N$ matrix of idiosyncratic terms is such that

$$e = A\varepsilon B$$

where A and B are two almost unrestricted deterministic matrices and ε is a $T \times N$ matrix with i.i.d. Gaussian entries. Matrices A and B determine cross-sectional and temporal correlation of the idiosyncratic terms. The gaussianity assumption can be relaxed but then either A or B must be a diagonal matrix, while the other remains relatively unrestricted. The paper by Onatski (2009b) gives all technical descriptions of the properties that matrices A and B must meet in terms of their spectral distributions.

Given these assumptions Onatski (2009b) derives a new result in random matrix theory, that the eigenvalues of the idiosyncratic part will cluster around a single point \hat{u} , while the meaningful eigenvalues will be separated. This helps him to develop the test for the number of factors. Simply put, the test is based on the inspection of the difference between consecutive sample eigenvalues: $\lambda_j - \lambda_{j-1}$. The difference $\lambda_j - \lambda_{j-1}$ should converge to zero with $T, N \rightarrow \infty$ for the group of "idiosyncratic eigenvalues", while diverge for "systematic eigenvalues". For finite samples, an additional parameter δ , is used as a threshold to distinguish between the two groups of eigenvalues. δ can be calibrated with an algorithm also described in Onatski (2009b).

2.4 Random permutation algorithm

The important property of the MP distribution is that the support for eigenvalues of i.i.d. data is compact. Unfortunately, the real data almost never has an i.i.d. structure. Therefore it is difficult to decide whether a particular eigenvalue deviates from the MP distribution due to the correlation of the columns or perhaps due to its non-i.i.d. distribution. The method introduced in this section develops a new numerical algorithm to determine the number of factors in the data under the assumption of no cross-correlation.

What we need here is another dataset \tilde{X} that shares all the specific properties of our original dataset, like jumps and heteroscedasticity, and at the same time is not correlated. One way to achieve this end is by the permutation of the rows in the original data, but separately in each column. This should ensure that all the correlation is wiped out, but keep all the sample moments of the original data intact.

The impact of permutation on the original dataset The intention of the following theorem is to give some intuition on why permutation of each row separately wipes out the correlation structure in the dataset. Here, I use a simple example of the matrix with just two columns.

Theorem 1 Assume that we have a $T \times 2$ matrix X of data. X is generated by a one factor model:

$$X = \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_T \end{bmatrix} \times \begin{bmatrix} a \\ b \end{bmatrix}' + \begin{bmatrix} \varepsilon_1^a & \varepsilon_1^b \\ \varepsilon_2^a & \varepsilon_2^b \\ \vdots & \vdots \\ \varepsilon_T^a & \varepsilon_T^b \end{bmatrix}$$

where f_i is $N(0, \sigma_f)$, $\varepsilon_i^a, \varepsilon_i^b$ is $N(0, \sigma_\varepsilon)$ and f and ε are independent. Let X^1 be the first and X^2 the second column of X . Then, for sufficiently long T , the permutation of any column will kill the correlation structure between X^1 and X^2 .

Proof. In the proof, I will permute the second column, just by moving each entry one step forward. That is, $X_1^2 \rightarrow X_2^2, X_i^2 \rightarrow X_{i+1}^2$ and $X_T^2 \rightarrow X_1^2$, so we get

$$X = \begin{bmatrix} af_1 + \varepsilon_1^a & bf_1 + \varepsilon_1^b \\ af_2 + \varepsilon_2^a & bf_2 + \varepsilon_2^b \\ \vdots & \vdots \\ af_T + \varepsilon_T^a & bf_T + \varepsilon_T^b \end{bmatrix} \rightarrow \tilde{X} = \begin{bmatrix} af_1 + \varepsilon_1^a & bf_T + \varepsilon_T^b \\ af_2 + \varepsilon_2^a & bf_1 + \varepsilon_1^b \\ \vdots & \vdots \\ af_T + \varepsilon_T^a & bf_{T-1} + \varepsilon_{T-1}^b \end{bmatrix}$$

then we have the following off-diagonal elements in the covariance of the original matrix:

$$Cov^{21}(X) = \underbrace{\frac{1}{T} \{ab(f_1^2 + f_2^2 + \dots + f_T^2)\}}_{\rightarrow ab\sigma_f^2} + \underbrace{\frac{1}{T} \{bf_1\varepsilon_1^a + af_1\varepsilon_1^b + \varepsilon_1^2\varepsilon_1^a + \dots\}}_{\rightarrow 0} = ab\sigma_f^2$$

but all those terms disappear when we compute the covariance matrix for the permuted matrix \tilde{X} ,

$$Cov^{21}(\tilde{X}) = \underbrace{\frac{1}{T} \{ab(f_1f_T + f_1f_2 + \dots + f_Tf_{T-1})\}}_{\rightarrow 0} + \underbrace{\frac{1}{T} \{bf_T\varepsilon_1^a + af_1\varepsilon_T^b + \varepsilon_T^2\varepsilon_1^a + \dots\}}_{\rightarrow 0} = 0$$

So, we can see here that simple permutation of rows managed to remove the correlation between the columns. That is, the off-diagonal entries of the covariance matrix will be asymptotically equal to zero for the transformed dataset. ■

The above simple theorem is necessary to convey an intuition that even this simple permutation can destroy cross-sectional covariance. It could be made more general by adding more columns, more factors and applying random permutation of indices. However, I think that it would only make this example more cumbersome without making it more transparent.

The more important problem is that for real datasets we do not know the true data generating process. Therefore, in order to be able to use the permutation to destroy cross-sectional correlation,

we have to make a new conjecture.

Definition 2 *Let X be a $T \times N$ matrix of data. Then, by random permutation of X , I will define a random permutation applied to each column of X separately. The random permutation is implemented by means of Fisher-Yates shuffle algorithm.*

Conjecture 3 *Let X be a $T \times N$ matrix of data. Assume also that each column in X is generated by a non-integrated and possibly heteroscedastic time series process. Then, by applying random permutation to the columns of matrix X , we are able to destroy cross-sectional correlation among them.*

It is difficult to specify the weakest conditions under which the above conjecture holds. However, there is a very simple way to test it empirically for a given dataset X . If we apply random permutation to $X \rightarrow \tilde{X}$, then we should expect to find that the highest eigenvalue in X is larger than the highest eigenvalue in \tilde{X} . Additionally, we can use intuition from MP distribution and results on spiked population models. The results from Paul (2007) show that in spiked populations, a phase transition will cause different behavior of population eigenvalues that are above and below a certain threshold $(1 + \sqrt{\gamma})$. On the one hand, all the population eigenvalues below the threshold will behave according to the MP distribution. On the other hand, population eigenvalues above the threshold will be separated from the rest of eigenvalues within MP distribution. Therefore, we may expect similar behavior from empirical dataset X and its transform \tilde{X} .

Indeed, Figure 1 shows an example of what happens when we perform random permutation of the entries in each column separately. In this example I use the changes in CDS spreads. The solid line is a distribution of eigenvalues for empirical dataset X and the dotted line shows eigenvalues of the \tilde{X} - permuted matrix. Clearly, the circle, indicating the largest eigenvalue for \tilde{X} , is significantly below the square, which indicates the largest eigenvalue for X . In other words, the first principal component explains just above 18% of the total variance in the original dataset. After permutation, in the new transformed dataset, the first principal explains just below 2% of the total variance. I should emphasize that this significant drop happens every time for each random permutation. Thus the result of random permutation is obvious, it kills the correlation and the eigenvalues are slowly increasing, just as random matrix theory would predict.

The upper boundary for "idiosyncratic eigenvalues" In order to determine the number of factors we need to assume that it is possible to find an equivalent of the upper boundary β_+ from Marčenko and Pastur distribution by a bootstrapping method for some real dataset with an unspecified data-generating process. In order to proceed, the following conjecture is necessary.

Conjecture 4 *Let \bar{X} be a $T \times N$ matrix of data. Assume also that each column in \bar{X} is generated by a non-integrated and possibly heteroscedastic time series. Additionally, assume that there*

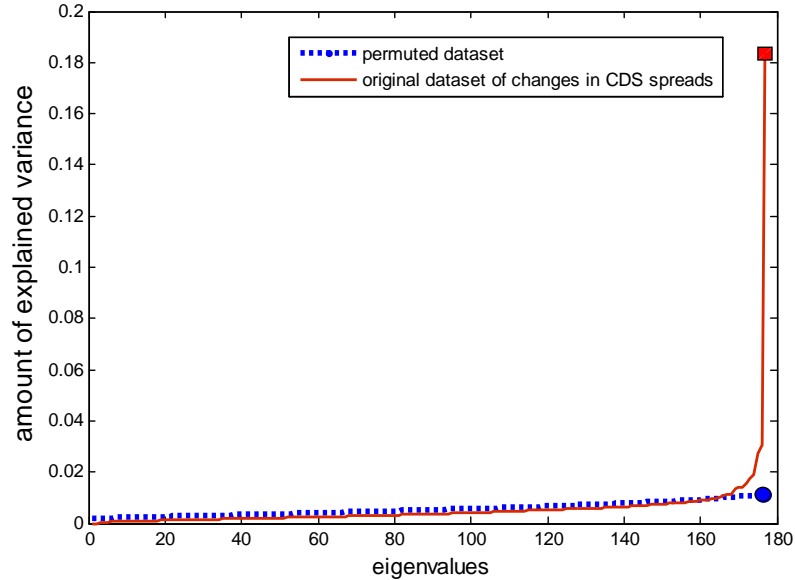


Figure 1: Distribution of eigenvalues for the empirical dataset X and permuted dataset \tilde{X} . On the vertical axis, we can see the amount of variance explained by eigenvalues.

is no cross-sectional correlation between the columns of \bar{X} . Let λ_{\max} be the highest eigenvalue obtained from the correlation matrix of \bar{X} . Then λ_{\max} is an in-sample approximation of β_+ from MP distribution. By applying random permutation of columns to \bar{X} we can investigate how λ_{\max} fluctuates.

The intuition supporting this conjecture comes here from the results on spiked population models in Paul (2007). If the distribution of eigenvalues is slowly increasing and does not exhibit visible spikes (like the dotted line in Figure 1) then intuitively it resembles the compact distribution of MP. Similarly, Onatski (2009b) uses differences between consecutive eigenvalues to tell apart the "idiosyncratic eigenvalues" from "systematic eigenvalues". So, if the group of eigenvalues from the correlation matrix of \bar{X} lay very close to each other then it is a heuristic evidence that these eigenvalues are idiosyncratic and λ_{\max} is an in-sample approximation to β_+ .

Additionally, fluctuations of λ_{\max} obtained from different random permutations of \bar{X} may be informative too. If λ_{\max} does not change significantly in different permutations then we have some ground to belief that it is sufficiently stable and does not rely on any particular order of the entries in the matrix. Hence, it can be used as an approximation of noisy eigenvalues that do not contain cross-sectional information.

Bootstrapping algorithm The algorithm implementing this method is described below in 5 steps.

1. Permute randomly each column of the data matrix X , in order to remove the correlation between the columns. The new artificial data matrix \tilde{X} will still exhibit all the peculiarities of the original dataset, but without the cross-sectional correlation.
2. Compute the correlation matrix $C(\tilde{X})$ and its spectral decomposition. Sort the vector of eigenvalues: v^λ and keep its maximum value λ_{\max}^1 .
3. Iterate steps 1 and 2 at least n times and choose the highest λ_{\max}^i from all iterations:

$$\lambda^{\max} = \max_i (\lambda_{\max}^i)$$

4. Use λ^{\max} as an in-sample equivalent of β_+ from MP distribution.
5. Compare the empirical distribution of the eigenvalues from correlation matrix $C(X)$ to the estimated λ^{\max} . All the empirical eigenvalues that are higher than λ^{\max} can be regarded as meaningful factors², under the assumption that there is no spurious cross-sectional correlation.

On the dataset of changes in CDS spreads, the algorithm usually gives a higher number of factors than Onatski's test. So we can view it as an upper bound for the possible number of factors. In fact it is not surprising that the algorithm finds more factors. Onatski's test also takes into account that residuals might be correlated. If this correlation is not too strong then Onatski's test will still be able to pick the correct number of factors from the shape of the eigenvalues' distribution. However, the numerical permutation algorithm makes no assumptions about the distribution of residuals. It takes account only of non-i.i.d. distribution of data, but does not adjust for the fact that idiosyncratic residuals may be spuriously cross-correlated. However, we might use the algorithm and artificially add some controlled cross-correlation to \tilde{X} in order to check how many meaningful eigenvalues will get separated. Also, we might want to know how many factors drive the co-movement of the data, under the assumption that idiosyncratic noise is not correlated.

This method, may be regarded as a complement to the procedure introduced by Onatski (2009b)³. It is difficult to assess how good it is in finding the true number of factors in the data. Because, first it assumes that Conjecture 4 holds for a particular dataset and second it assumes away any correlation of the error terms. However, later in this paper, I will use the method to assess the maximum possible bias of the highest noisy eigenvalue in the dataset. This application

²By meaningful factors, I understand factors that have population eigenvalues larger than one.

³Onatski's test is sensitive to large idiosyncratic deviations and sometimes it gives rather surprising, and most probably wrong, results. The suggested numerical algorithm is more stable in this sense, but it may lead to overestimation.

should be less controversial, because it does not depend so much on the correctness of Conjecture 4. In this case, Conjecture 3 is more important and this one is easier to test empirically. If we know that in a particular dataset, columns are not correlated with each other then the population eigenvalues are equal to one. If however, we find that sample eigenvalues differ from the value of one then we can use it as an in-sample evidence of the bias.

An obvious disadvantage of the algorithm, shared by all bootstrap methods, is its computational intensity.

3 Regression results

3.1 Theoretically motivated variables used in regressions

For the regressions, the following theoretical variables are used: return on stocks (*ret*), leverage (*lev*), volatility (*vol*), volatility index (*VIX*), 5 year maturity Treasury bond yield (*r*), square of 5 year yield (r^2), the slope of the yield curve (*Slope*), which is a difference between 10-year and 2-year bond yields, and the return on S&P500 (*S&P500*). So there are three firm-specific variables and five macroeconomic ones.

Leverage is based on data from Compustat and is defined in the same way as in Ericsson at al (2004), that is

$$\frac{\text{Book Value of Debt} + \text{Book Value of Preferred Equity}}{\text{Market Value of Equity} + \text{Book Value of Debt} + \text{Book Value of Preferred Equity}}$$

Cremmers at al. (2008) show that implied volatility computed from individual option prices contains useful information. Here, I will use an exponentially weighted historical volatility based on 180 days of returns. Ericsson at al. (2004) have shown that historical volatility performs almost as well as the option implied volatility. Volatility h_t for each firm is generated according to

$$h_t = r_t^2(1 - \lambda) + h_{t-1}\lambda$$

where $\lambda = 0.94$ is fixed constant across all companies. All the macro-variables are obtained from the Datastream.

3.2 Regressions

The following regression is estimated for each firm separately, just as in CGM.

$$\begin{aligned} \Delta CDS_{i,t} = & \alpha_i + \beta_i^l ret_{i,t} + \beta_i^l \Delta lev_{i,t} + \beta_i^v \Delta vol_{i,t} + \beta_i^s \Delta skew_{i,t} + \\ & + \beta_i^{vix} \Delta VIX_{i,t} + \beta_i^r \Delta r_{i,t} + \beta_i^{r^2} (\Delta r_{i,t})^2 + \beta_i^{Slope} \Delta Slope_{i,t} + \beta_i^{SP} S\&P500_{i,t} + \varepsilon_{i,t} \end{aligned} \quad (1)$$

but for two different time intervals. First, I estimate the regressions on the whole sample, that is from the very beginning of 2004 until the end of May 2008. Then, I use a shorter time interval from the beginning of 2004 until July 31, 2007. I choose this particular date, in order to separate the impact of the subprime crisis.

| | from 01.01.2004 to 07.31.2007 | | from 01.01.2004 to 05.30.2008 | |
|------------------------|-------------------------------|---------------|-------------------------------|---------------|
| | average beta | % significant | average beta | % significant |
| constant | 0.0000 | 0.00% | 0.0000 | 0.56% |
| ret | 0.0013 | 53.67% | 0.0027 | 82.49% |
| lev | 0.0127 | 33.90% | 0.0227 | 52.54% |
| vol | 0.2787 | 20.90% | 0.2493 | 11.86% |
| VIX | 0.0000 | 2.26% | 0.0000 | 1.13% |
| r | 0.0001 | 5.08% | -0.0014 | 50.28% |
| r ² | 0.0000 | 9.60% | 0.0001 | 38.98% |
| slope | 0.0000 | 3.95% | 0.0004 | 10.17% |
| S&P500 | 0.0000 | 20.34% | 0.0000 | 53.67% |
| average R ² | 0.0743 | | 0.1307 | |

Table 1.

This table presents test statistics (5% significance level) and regression results. The reported coefficients are averages for regression coefficients from time-series regressions. The second and fourth columns report also the percentage of significant coefficients out of 177 companies.

Table 1 presents results of the regressions. The first two columns give the results for the shorter sample, that is from January 1, 2004 until July 31, 2007. The third and fourth columns present the results from regressions on the whole sample period, from January 1, 2004 until May 30, 2008. Also, the first and third columns show the average over the coefficients from all regressions. The second and fourth columns show the fraction out of 177 coefficients, which turned out to be significant at

the level of 0.05. The t -statistics were computed with the Newey-West procedure⁴ to take account of heteroscedasticity and serial correlation. The average Durbin-Watson statistics is 1.54 for the shorter time interval and 1.53 for the whole sample period.

From Table 1 we can clearly see that regressors performed very poorly and also R^2 are far from satisfactory. But this is not unusual for the regressions using daily data. The higher the frequency of the data, the more noisy it is. However, in this paper I do not concentrate on the question of true determinants of the CDS spreads, as this topic was already covered extensively by Ericsson et al (2004). The focus here is on the number and strength of the factors and on the properties of the residuals.

3.3 Distribution of eigenvalues in the residuals

In order to determine the strength of the factors in the residuals, we need to investigate the distribution of their eigenvalues. The following figure depicts these distributions, both for the residuals from the regressions of theoretical variables on the changes in CDSs and also for the changes in CDSs.

I repeat the same exercise for the whole time interval. The plot looks basically the same as the one in Figure 2 and only the highest eigenvalues are different. I will just report how much variance they explain. Table 2 presents the amount of variance explained by the first five principal components for the two subsamples, both in the residuals and in ΔCDS . The last two rows of the table also provide results for the number of factors in the data.

| | from 01.01.2004 to 07.31.2007 | | from 01.01.2004 to 05.30.2008 | |
|----------------------------------|-------------------------------|-----------|-------------------------------|-----------|
| | ΔCDS | residuals | ΔCDS | residuals |
| PC1 | 18.5% | 17.1% | 34.3% | 29.5% |
| PC2 | 3.0% | 2.7% | 3.4% | 3.1% |
| PC3 | 2.7% | 2.6% | 2.9% | 2.8% |
| PC4 | 1.9% | 1.9% | 1.6% | 1.8% |
| PC5 | 1.7% | 1.7% | 1.5% | 1.6% |
| ‡ of factors, Onatski test | 3 | 3 | 5 | 3 |
| ‡ of factors, bootstrapping alg. | 8 | 6 | 6 | 6 |

Table 2.

The amount of variance explained by the first five principal components for the two subsamples.

⁴Following suggestion of Newey and West, the number of lags is set to

$$q = \text{floor} \left(4 \left(\frac{T}{100} \right)^{2/9} \right)$$

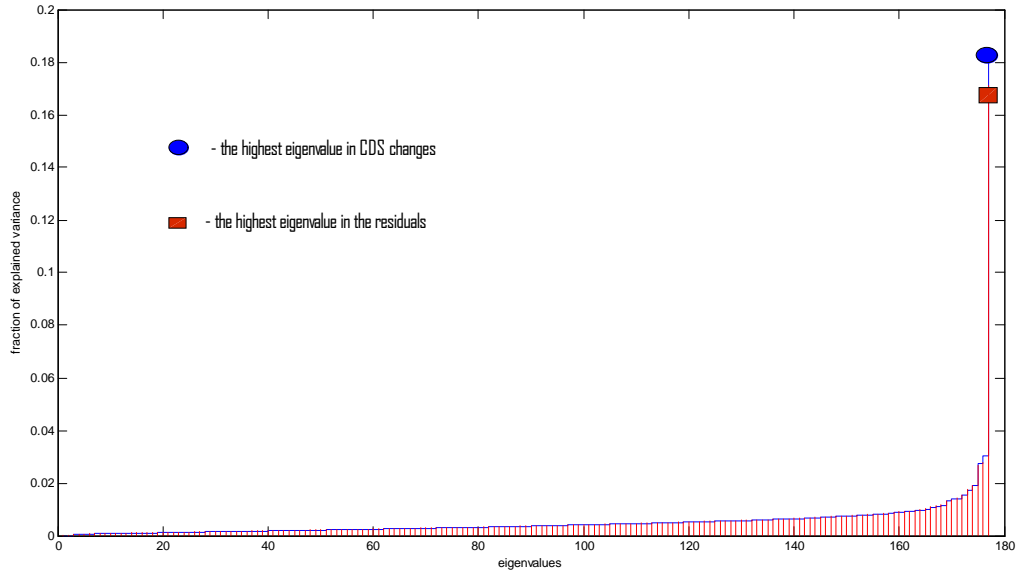


Figure 2: Eigenvalues distribution for changes in CDS spreads and for residuals from the regressions.

Altogether we can see that the changes in CDSs are not strongly correlated, as the first principal component explains only 18% for the shorter time interval - before the subprime crisis and 29% for the whole sample together with the crisis period. It is interesting to note, although not very surprising, that extending the sample period from August 1, 2007 until May 30, 2008 results in a strong increase in the correlation. In other words, the last 10 out of 53 months in the sample contribute to the twice-stronger first principal component.

We know from Harding (2008) that the strength of the first principal components is always biased upward in finite samples. Unfortunately, if we do not know the true data generating process, it may be impossible to assess the magnitude of the bias. However, we can at least try to estimate the number of factors that drive the co-movement. Using Onatski's test, we can find that there are 5 factors in ΔCDS for the time interval between January 1, 2004 and May 30, 2008. In the residuals, according to the test there are only 3 factors. For the shorter time interval, which ends just before the subprime crisis, Onatski's test finds 3 significant factors both in ΔCDS and in the residuals from the regressions, while the numerical algorithm indicates between 6 and 8 factors.

The results of this section are exactly consistent with CGM. I find that regressors motivated by structural models of credit risk can hardly explain daily changes in CDSs. They are also unable to explain the first principal component of the data. This can be seen in Figure 2, where the square

representing highest eigenvalue in the residuals is only slightly below the circle representing highest eigenvalue in CDS changes. But, in the next section, I will show that structural models should not be blamed too much for this dismal state of affairs. That is because even very good regressors cannot fully explain the correlation structure of changes in CDSs.

4 How to generate a spuriously strong latent factor

One of the problems that CGM had to face was a large number of bonds of different maturities, very often issued by the same firms, but a relatively small number of time points. Specifically, they have monthly data spanning 10 years on around 600 bonds, with an average of 56 quotes per bond. The simple solution that they used was to average the results into 15 bins according to two criteria: leverage and maturity in the cross-section⁵. Such a solution may lead to some severe biases. Below I present two independent pitfalls of this procedure. First, I would like to address the problem of high-dimensionality. Second, in subsection 4.2, I will explain in detail how averaging spuriously magnifies the small correlation that exists in the data.

4.1 High dimensional sample in CGM

CGM used residuals from their regressions to estimate by means of principal components the strength of the latent factor. However, their matrix of the residuals was at most of the size 120×15 , or perhaps it had an even smaller number of time points. These are matrices with a non-negligible N/T ratio and we can expect that the highest eigenvalues will have an upward bias, even if the columns are not correlated.

Here, I will make use of the numerical algorithm described above. The algorithm allows us to find the upper limit of the bias in eigenvalues.⁶ The logic is the following. First, I remove all the correlation structure from the dataset of residuals obtained from regression (1). Then in the second step, using the bootstrap method, I estimate what could be the possible bias. The results are presented in Table 3. The table presents how much of the total variance is explained by the first two principal components and the sum. Principal components are obtained from correlation matrices in order to standardize variance. The first and second columns are for matrices with $T = 120$ and $T = 56$ rows respectively and $N = 15$ columns. The third column shows us how much variance is explained by the true, population eigenvalues when T is sufficiently long. In other words, when we have a $T \times 15$ matrix and there is no correlation between the columns, then each

⁵It is difficult to say how many time points the matrix of residuals used by CGM had. Almost certainly it was something between 56 (average number of quotes) and 120 (10 years of monthly data).

⁶Instead of the upper limit of eigenvalues generated by the algorithm, one might use any of the quantiles like 95% or 99%. But they are all very close and results would not change, thus I do not report those figures.

principal component should explain exactly $100/15 \approx 6.6\%$ of the total variance.

| | 120×15 | 56×15 | $T \times 15$ |
|--------|-----------------|----------------|---------------|
| PC_1 | 17.7% | 22.1% | 6.6% |
| PC_2 | 9.8% | 10.6% | 6.6% |
| sum | 27.6% | 32.7% | 13.3% |

Table 3.

Possible upward bias of the first two PCs
for purely uncorrelated data.

Here we can see that the first principal component may appear to be three times stronger than it really is. In the case of a matrix with 120×15 entries, the first principal component appears to explain 17.7%, although its true population value is 6.6%. As was explained above, eigenvalues of the correlation matrices are biased upward even for purely uncorrelated time series. One should observe here that these estimates are very conservative, because they were generated for the matrix without any correlation structure. If, however, there is some undesirable cross-correlation between the columns, then this additional correlation will magnify the already biased eigenvalues even further. So, this is one of the reasons why CGM might have overestimated the latent factors. In the next subsection I present even stronger evidence for this overestimation.

4.2 Averaging out the noise in the residuals

4.2.1 Replicate the procedure of CGM

CGM classify firms into 5 different leverage groups and 3 maturities. However, since the 5-year maturity CDSs are the most liquid, I will at first replicate their procedure with 15 leverage groups and just one range of maturities. The bins are chosen to group in one basket firms with similar leverage: under 6.67 percent, from 6.67 to 13.33, from 13.33 to 20 percent and so on. Thus the bins do not contain the same number of firms. Table 4 presents what happens if we apply this procedure

to the residuals that were found in Section 3.

| | from 01.01.2004 to 07.31.2007 | | from 01.01.2004 to 05.30.2008 | |
|-----|-------------------------------|---------------------|-------------------------------|---------------------|
| | 177 columns | averages in 15 bins | 177 columns | averages in 15 bins |
| PC1 | 17.1% | 43.5% | 29.5% | 54.6% |
| PC2 | 2.7% | 7.0% | 3.1% | 6.3% |

Table 4.

This table presents how averaging into smaller baskets spuriously amplifies the first two largest eigenvalues. Here, the empirical regressors were used.

We can see a rather surprising fact here. For instance, in the first time period we know that residuals from the regressions contain some common factors. But the strongest factor in these residuals explains no more than 17.1% of the total variance. However, if we average those 177 columns into 15 baskets, then the strength of the first principal component jumps up to 43.5%. This is actually more than twice as strong. But this result is completely spurious. What has happened? By taking the average into a smaller number of baskets, we magnify the first eigenvalue. That happens because we are averaging out the noisy part and thus we increase the signal-to-noise ratio⁷ of the initially weak factor. Taking averages of different time series into baskets is a good idea when one wants to have a closer look at the main factor driving the changes. But, certainly it is not the best idea to do the same with the residuals and then conclude that there exists a strong unexplained factor.

4.2.2 Principal components as regressors

5 year maturity only One might guess that we would not get such effects if we had used better regressors in the first place. But that is not true, at least not entirely. In order to show that, I will really use better regressors and I show that we still get correlated residuals. Specifically, I will use principal components estimated from the changes in CDSs. Principal components by construction explain the largest part of the common variation in the data and therefore we cannot find better regressors than these. The regressions have the form

$$\Delta CDS_{i,t} = \beta X_t + \varepsilon_{i,t} \quad (2)$$

Here ΔCDS are at first demeaned and so the constant α is not used, and X_t is a set of principal components. Onatski's test indicated that in ΔCDS we can find between 3 to 5 significant factors that drive the co-movement of the data. Therefore, to be on the safe side, I will use two different

⁷This fact is remarkably robust. It changes neither when the baskets are chosen randomly nor with smaller subsamples of the data. This magnifying effect we can also observe on the simulated data.

sets as X regressors - the first 5, and 15 principal components to compare how much information they can extract⁸. Since results are very similar for both time periods, I will report only those for the shorter time interval - without the subprime crisis.

| | first 5 PCs as regressors | | first 15 PCs as regressors | |
|---------------|---------------------------|---------------------|----------------------------|---------------------|
| | 177 columns | averages in 15 bins | 177 columns | averages in 15 bins |
| first factor | 6.8% | 16.2% | 5.4% | 15.9% |
| second factor | 3.6% | 8.6% | 4.3% | 10.8% |

Table 5.

The amplifying impact of averaging on the eigenvalues when in-sample principal components were used as regressors.

The first eigenvector of the average from 15 bins, for the shorter period of time, when 5 PCs are used as regressors (see Figure 3), is the following,

$$e_v^{5 \text{ year maturity, 5 PCs}} = (0.16; 0.27; 0.39; 0.08; 0.46; 0.20; 0.42; 0.33; -0.13; 0.05; 0.19; 0.26; 0.24; 0.09; 0.11)$$

Clearly, using principal components helps to extract more information and leaves less correlation in the residuals. When we used true theoretical regressors then the strongest factor in the residuals still explained 17.1% of the total variance. However, when we use better artificial regressors then the strength of the factor drops from 17.1% to 6.8% for 5 principal components and to 5.4% with 15 PCs. Interestingly, expanding the number of regressors from 5 to 15 principal components only slightly reduces the strength of the main factor in the residuals. It is difficult to explain why the remaining correlation is so insensitive to even a much larger number of principal components. One possible explanation is that there are not more than 5 common factors, as is indicated by Onatski's test, and additional principal components capture only the idiosyncratic part of the variance.

In general, we can observe exactly the same pattern as before. That is, taking the averages causes a spurious increase in the strength of the first factor in the residuals. For instance, we use first 5 principal components as regressors on the changes in CDSs and then average those residuals into 15 bins according to leverage. This involuntarily magnifies the importance of the factor from 6.8% to 16.2%. The noise is averaged out and the signal to noise ratio increases. Accordingly, we

⁸Using principal components as regressors removes so much information that the correlation matrix of the residuals may no longer be positive semidefinite. In fact few eigenvalues are almost zero, but have a negative sign. Therefore I use a numerical algorithm from Sharapov (1997) that finds the positive semidefinite matrix closest to the estimated correlation matrix according to the Froebnius norm.

However, this is not clear why the eigenvalues might be negative. I investigate this problem further in the appendix. The conclusion from the simulated data is that it must be related to some numerical problems, but it is not in any way important for the results.

get a spuriously strong factor that seems to explain a larger part of the variance but is in fact only an artefact of the procedure.

3, 5 and 7 year long maturities Here, I replicate the methodology of CGM even more closely. I use three different maturities: 3, 5 and 7 years long and five leverage groups: under 20%, from 20% to 40% and so on. Altogether I use information from 531 (3×177) columns of data. Just like before, in the first step I extract the first 5 and then 15 principal components from ΔCDS . In the next step, I use residuals to check how much co-movement is still left in the data. The first and the third columns in Table 6 present the strength of the first two factors in all residuals after extracting respectively 5 and 15 principal components. The second and fourth columns show the strength of the first two factors in the residuals, which were grouped into 15 bins according to leverage and maturity, just as in CGM.

| | first 5 PCs as regressors | | first 15 PCs as regressors | |
|---------------|---------------------------|---------------------|----------------------------|---------------------|
| | 531 columns | averages in 15 bins | 531 columns | averages in 15 bins |
| first factor | 12.4% | 44.2% | 3.4% | 16.2% |
| second factor | 3.1% | 11.6% | 2.6% | 14.6% |

Table 6.

The amplifying impact of averaging on the eigenvalues when three different CDS maturities were used.

The first eigenvector, for the second column in Table 6, can be seen in Figure 3. It has the following entries

$$e_v^{\text{different maturities, 5 PCs}} = (0.28; 0.30; 0.18; 0.22; 0.18; 0.26; 0.32; 0.25; 0.25; 0.22; 0.30; 0.32; 0.24; 0.24; 0.22)$$

Not surprisingly, once again we can see that taking averages results in a spurious increase in the strength of the first factor in the residuals. Here, when we use additional information from different maturities, the effect seems to be even stronger. However, there is one interesting difference. When we use additional data from different maturities then the first eigenvector has much more uniform entries⁹. One could quote CGM here: "the first component is approximately an equally weighted portfolio across quality and maturity groups". For comparison I present also the first eigenvector

⁹The uniform distribution of entries in the first eigenvector seems to be a very robust result. I get almost the same equally weighted eigenvector when 5 different maturities are used and 3 different leverage groups. Also, the same result obtains when smaller samples of firms are used divided according to 5 different maturities and 3 random, instead of leverage, groups.

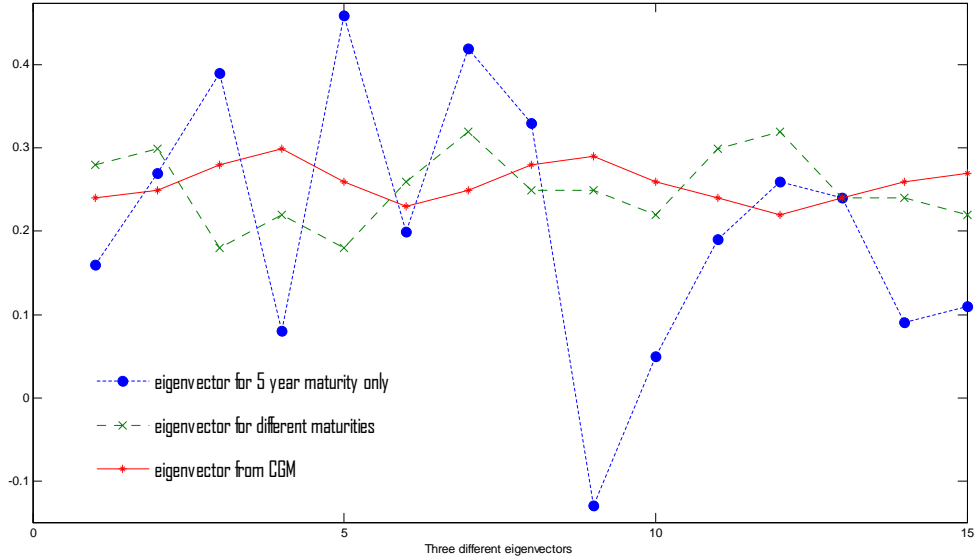


Figure 3: Plot of different eigenvectors.

from the residuals found by CGM (see also Figure 3),

$$e_v^{\text{CGM}} = (0.24; 0.25; 0.28; 0.30; 0.26; 0.23; 0.25; 0.28; 0.29; 0.26; 0.24; 0.22; 0.24; 0.26; 0.27)$$

So we can see here that the two eigenvectors are very similar. This was not the case when we used the data only for 5 year maturity CDSs. This result suggests that the special structure of the eigenvector found by CGM may be just a consequence of insufficient proxies for the term structure.

4.3 Discussion

Results of this section can be summarized in three points. 1) Applying principal component analysis to the correlation matrices of high-dimensional datasets results in the overestimation of the highest eigenvalues. 2) By taking the average into a smaller number of baskets, we unintentionally magnify the signal-to-noise ratio of weaker factors from the data. 3) One needs a lot of very good regressors in order to remove all the correlation from the residuals in changes of CDSs. Therefore we should not be too surprised that the empirical regressors, motivated by structural models, leave in the residuals unexplained variance.

5 Problem of misspecification in a simulated economy

So far, we have seen (Table 4) that the first principal component in residuals of CDS spread changes explains just above 17% of the total variance for the time period before the subprime crisis and around 29% together with the crisis. I have also shown how these estimates can be unintentionally magnified to larger proportions. Therefore we have good reasons to suspect that latent factor found in CGM explains much smaller amount of total variance than 76% that they report in their sample. So, it seems premature to declare credit market segmentation. In order to link the problems of misspecification of models with the concept of market segmentation, econometric tests are not sufficient. We need some additional guidance from theory. In this section, I approach the problem of misspecification from a different angle. Instead of using empirical dataset, I will run the same tests on an artificial dataset simulated by three different structural models.

As was said before, it has been recognized that most of the older structural models have problems with matching the true level of credit spreads when they are calibrated to observed default frequencies. However, recently structural models were generalized to allow for observable and unobservable regime switching, or for more general utility functions like utility with Epstein-Zin-Weil preferences¹⁰. These new theoretical developments allowed structural models to generate both realistic default probabilities and credit spreads. Moreover, some papers like Bhamra et al. (2009), Chen (2008) and David (2008) obtain these realistically high spreads endogenously.

Here I will concentrate on the implications of the paper by David (2008). It was the first paper to investigate the influence of unobservable regime shifts in macroeconomic growth rates on credit spreads. My main observation is that David's model, while improving on the level of credit spreads, simultaneously introduces additional cross-sectional correlation among them. The mechanism responsible for the cross-correlation is the updating of investors' beliefs. The intuition is simple: a representative agent updates her beliefs continuously as new information about the fundamentals arrives. New beliefs about future prospects of the economy are immediately reflected in asset prices. But importantly, all asset prices are updated at the same time as a response to the same information and this results in co-movement.

Simulation experiment I construct a simulation experiment in three steps. First, I generate credit spreads for artificial firms with structural parameters borrowed from David (2008). These artificial firms initially have the same leverage ratios as the firms from my original empirical data sample. In the second step, I run the regressions of everything that is observable in this economy on the first differences of simulated credit spreads. Finally, I examine R^2 , t-statistics and the highest eigenvalues in first changes and in residuals obtained from the regressions.

¹⁰Bhamra, Kuehn and Strebulaev (2009), Chen (2008), Chen, Collin-Dufresne and Goldstein (2009), David (2008), Hackbarth, Miao and Morellec (2006)

A short description of the model and details of the simulation are in the appendix. Here, I will mention only briefly the details that are directly relevant for this paper. David's model is a generalization of Merton's basic structural model. The essential difference is that his model allows for an unobserved regime-switching structure of fundamentals. Fundamentals of the economy are described by the total real earnings and inflation. Drifts of inflation and earnings are hidden and follow a Markov chain process. Investors observe past realizations of earnings and inflation and learn about the drifts over time. Also asset volatilities vary over time as investors update their beliefs about the hidden states and the future growth prospects. At times of higher uncertainty about the current state of fundamentals, investors revise their beliefs more rapidly and consequently they generate higher asset volatility. Inflation plays a signalling role here, because it predicts real growth rates - also called "proxy hypothesis" in Fama (1981).

However, for the purpose of this paper it is important to observe that the belief updating process induces a covariation among all the assets in this economy. That happens for the simple reason that investors update their beliefs simultaneously for all companies in the economy. But this additional covariation of the assets is only indirectly linked to the fundamentals of the economy and directly to investors' beliefs.

For the simulations I make two assumptions. First, all companies initially have the same leverage ratios as the true firms in the empirical sample that I use. Second, I assume that the real earnings of all firms have the same correlation with the total earnings process and the correlation coefficient is $\rho = 0.5$. All other parameters, except for the variance of total earnings are from David (2008). As two benchmark models I use Merton's model and the dynamic capital structure model from Dangl and Zechner (2004). Then the procedure of the simulation experiment follows the following steps:

1. Generate the processes for inflation, total real earnings, real pricing kernel of the investor and the idiosyncratic part for 177 firms' earnings. The data are generated for only one state of the world.
2. Compute the spreads and equity prices separately for the Merton, Dangl-Zechner and David models. Spreads for the Merton and Dangl-Zechner models are also computed for nominal prices.
3. Construct the volatility index specific to David's model. (it is constructed as the assets' volatility - formula (15) from David (2008)).
4. Run the regressions of everything that is observable in these economies on the first changes in credit spreads for all three models.
5. Repeat this procedure for three different values of σ_E , which is a standard deviation of the real earnings process.

| σ_E | 0.25 | 0.2 | 0.15 | 0.1 |
|---|-------|-------|-------|-------|
| A) Results for David's model | | | | |
| mean R^2 | 0.95 | 0.87 | 0.78 | 0.56 |
| PC_1 in changes | 25.4% | 26.8% | 31.2% | 40.1% |
| PC_1 in residuals | 68.9% | 63.5% | 58.6% | 54.6% |
| B) Results for Merton's model | | | | |
| mean R^2 | 0.99 | 0.98 | 0.96 | 0.95 |
| PC_1 in changes | 24.7% | 24.1% | 22.0% | 24.5% |
| PC_1 in residuals | 4.2% | 6.5% | 7.6% | 5.9% |
| C) Results for Dangl and Zechner model | | | | |
| mean R^2 | 0.93 | 0.94 | 0.99 | 0.99 |
| PC_1 in changes | 18.3% | 25.0% | 22.4% | 21.6% |
| PC_1 in residuals | 4.4% | 3.4% | 3.3% | 5.8% |

Table 7. Simulation results.

Essentially, we can see that it is easier to explain changes in credit spreads when σ_E is high. However, for lower values of σ_E linear regressions leave substantially more information in the residuals. For example, when $\sigma_E = 0.1$ then the R^2 for David's model is only 0.37 and the first principal component in the residuals explains 47.7% of variance. For comparison, in the world generated by the Merton model, for $\sigma_E = 0.1$ we have $R^2 = 0.95$, but here PC_1 in the residuals is very small.

Therefore, we may conclude that David's model contains additional dynamics which perturbs first changes in credit spreads in a way that is difficult to capture in linear regressions. However that does not imply that credit spreads react to different shocks than returns on stock prices. On the contrary, it is a heuristic proof of inadequacy of statistical tests. In other words, a factor in the residuals that is not too strong can be reconciled with the theory of structural models and as such should not be used as evidence against them.

6 Conclusion

One of the puzzles in the credit risk literature is the existence of a strong latent factor driving the co-movement in credit spread changes. This single common factor was supposed to be driven by local supply/demand shocks, independent of both credit risk factors and proxies for liquidity. But

the evidence supporting this result was based on the high-dimensional dataset with many missing values. The main conclusion of the present paper is that the single, strong and unexplained factor in the residuals from regressions on the changes in CDSs most probably does not exist.

As shown, one might spuriously magnify the importance of these factors in the residuals either by using datasets with a large ratio of columns to rows or simply by taking the averages over the columns. For instance, in a matrix with 120×15 entries and with all uncorrelated columns, we may have a principal component that appears to explain 17.7% of the total variance, while in fact it explains only 6.6% of the total variance. Taking averages over the columns increases the signal-to-noise ratio of initially weak factors. For example, in the residuals from regressions in Section 3, we find that the strongest principal component explains no more than 17.1% of the total variance. However, if we take the average of the columns into 15 bins, then we find that the first principal component suddenly seems to explain 43.5% of the total variance. Additionally, the paper shows that it is very difficult to remove all the correlation from ΔCDS by means of linear regressions even with very good regressors. When the first 5 or even 15 principal components of changes in CDSs are used as regressors, we still get residuals that are cross-sectionally correlated.

Two different tests for the number of factors were employed here. Both of these tests are based on insights from random matrix theory. They account for the upward distortion of eigenvalues in high-dimensional datasets. Additionally they also account for non-i.i.d. distribution of the data. The difference is that one of the tests - the numerical bootstrapping algorithm - does not allow for spurious cross-correlation of noise and it indicates between 6 and 8 factors in ΔCDS . The other one, according to Onatski (2009b), which accounts for cross-correlation of noise, finds between 3 and 5 independent factors in the data. The numerical bootstrapping algorithm is a technical contribution of this paper.

In this paper, the strength of the first principal component in the residuals explains very similar amount of total variance (17.7% and 29.5% for different time intervals) to what Cremers et al. (2008) and Ericsson et al. (2004) found (24% and 32% respectively). However, what these two papers could not explain, was the difference between their estimates and the one in CGM. The results of this paper may help to clarify this difference. CGM spuriously magnified importance of their latent factor to 76%, while most probably it explains around 20% to 30%, as it was shown in this paper and in Cremers et al. (2008) and Ericsson et al. (2004).

Finally, I show that weak latent factors are not inconsistent with the theory of structural models. What we call a segmentation of the markets is a matter of definition. But if the segmentation means that different shocks are priced in the credit market and in the stock market then the above simulation experiment demonstrates that we do not have good evidence to support the segmentation hypothesis. The magnitude of unexplained variance and the strength of the first principal component in the residuals clearly do not prove that investors from credit markets respond to different shocks than stock market investors. Obviously, a latent factor as strong as the one

reported in CGM would be much more difficult to explain as consistent with the theory of structural models.

References

- [1] Bai J. and Ng S., (2004), *A panic attack on unit roots and cointegration*, *Econometrica*, Vol. 72, 1127-1177.
- [2] Barrieu P., Giammarino F., (2009), *An adaptive nonparametric model for the systematic factors of portfolio credit risk premia*, *Journal of Empirical Finance*, vol. 6, 665-670
- [3] Berndt A., Douglas R., Duffie D., Ferguson M., and D. Schranz, (2005), *Measuring Default-Risk Premia from Default Swap Rates and EDFs*, Working Paper, Stanford University
- [4] Bhamra H., L. Kuehn and I. Strebulaev, (2009), *The Levered Equity Risk Premium and Credit Spreads: A Unified Framework*, forthcoming in *Review of Financial Studies*
- [5] Black F., *Business cycles and equilibrium*, Wiley 2009
- [6] Chen H., (2009), *Macroeconomic Conditions and the Puzzles of Credit Spreads and Capital Structure*, forthcoming in *Journal of Finance*
- [7] Chen L., Collin-Dufresne P. and R. Goldstein, (2009), *On the relation between the credit spread puzzle and the equity premium puzzle*, *Review of Financial Studies*, 22, 3367-3409
- [8] Collin-Dufresne P., R. Goldstein, and J. Martin, (2001), *The determinants of credit spread changes*, *Journal of Finance* 56, 2177-2208
- [9] Cremers M., Driessen J., Maenhout P., Weinbaum D., (2008), *Individual stock-option and credit spreads*, *Journal of Banking and Finance*, Vol. 32, 2706-2715
- [10] Dangl T. and J. Zechner, (2004), *Credit risk and dynamic capital structure choice*, *Journal of Financial Intermediation*, 2, pp. 183-204
- [11] Das S. R., Hanouna P., (2006), *Credit Default Swap spreads. Survey of the literature*, *Journal of Investment Management*, Vol.4, No.3, 93-105
- [12] David A., (2008), *Inflation uncertainty, asset valuations and the credit spreads puzzle*, *Review of Financial Studies*, 21, 2487-2534
- [13] David A., and P. Veronesi, (2002), *Option prices with uncertain fundamentals*, Working paper, University of Chicago

- [14] Ericsson J., Jacobs K., and Oviedo R.A., (2004), *The determinants of Credit Default Swap premia*, CIRANO, Working paper
- [15] Fama E., (1981), *Stock returns, real activity, inflation and money*, American Economic Review, 71(4):545-65
- [16] Gündüz Y. and Uhrig-Homburg M., (2008), *Does modelling framework matter? A comparative study of structural and reduced-form models*, Working paper
- [17] Hackbarth D., J. Miao and E. Morellec, (2006), *Capital structure, credit risk and macroeconomic conditions*, Journal of Financial Economics, 82, 519-550
- [18] Harding C. M., (2008), *Explaining the single factor bias of arbitrage pricing models in finite samples*, Economic Letters
- [19] Huang J. and M. Huang, (2003), *How much of the corporate-treasury yield spread is due to credit risk?*, Working paper, Penn State and Stanford
- [20] Hull J., Predescu M. and A.White, (2004), *The relationship between credit default swap spreads, bond yields, and credit rating announcements*, Working paper, University of Toronto
- [21] Johnstone I. M., (2006), *High dimensional statistical inference and random matrices*, Working Paper, Stanford University
- [22] Johnstone I. M. and Lu A. Y., (2004), *Sparse principal components analysis*, Working Paper, Stanford University, forthcoming in Journal of American Statistical Association
- [23] Ledoit, O. and Wolf, M., (2002), *Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size*, Annals of Statistics 30, 1081-1102
- [24] Ledoit, O. and Wolf, M., (2009), *Honey, I shrunk the sample covariance matrix*, Journal of Portfolio Management 30, Volume 4, 110-119
- [25] Longstaff F., Mithal S. and E. Neis, (2005), *Corporate yield spreads: Default risk or liquidity? New evidence from the credit default swap market*, Journal of Finance 60, 213-2254
- [26] Marčenko V. A. and Pastur L. A., (1967), *Distributions of eigenvalues of some sets of random matrices*, Math - USSR Sbornik, 1:507-536
- [27] Onatski A., (2009a), *Asymptotics of the Principal Components Estimator of Large Factor Models with Weak Factors*, manuscript, Columbia University
- [28] Onatski A., (2009b), *Determining the number of factors from empirical distribution of eigenvalues*, forthcoming in Review of Economics and Statistics

- [29] Onatski A., (2009c), *Testing hypotheses about the number of factors in large factor models*, forthcoming in *Econometrica*
- [30] Paul D., (2007), *Asymptotics of sample eigenstructure for a large dimensional spiked covariance model*, *Statistica Sinica*, Vol. 17, No. 4, pp. 1617-1642.
- [31] Schaefer S., and I. Strebulaev, (2008), *Structural models of credit risk are useful: evidence from hedge ratios on corporate bonds*, *Journal of Financial Economics*, Vol. 90, 1-20
- [32] Singleton K., *Empirical Dynamic Asset Pricing. Model specification and econometric assessment*, Princeton University Press 2006
- [33] Sharapov I., *Advances in multigrid optimization methods with applications*, Ph.D thesis, UCLA
- [34] Zumbach G., (2009a), *The empirical properties of large covariance matrices*, *RiskMetrics Journal*, Volume 9, Number 1
- [35] Zumbach G., (2009b), *Inference on multivariate ARCH processes with large sizes*, Working Paper, Riskmetrics Group

7 Appendix

7.1 Theorems for spiked population model from Paul (2007)

Theorem 1 from Paul (2007) *Assume that λ_i is a sample eigenvalue and l_i is a population eigenvalue. If $l_i \leq 1 + \sqrt{\gamma}$, where $N, T \rightarrow \infty$, $\gamma = N/T$ and $\gamma \in (0, 1)$ then*

$$\lambda_i \rightarrow (1 + \sqrt{\gamma})^2 \text{ almost surely as } N, T \rightarrow \infty$$

Theorem 2 from Paul (2007) *Suppose that $l_i > 1 + \sqrt{\gamma}$ and $\gamma \in (0, 1)$ then*

$$\lambda_i \rightarrow l_i \left(1 + \frac{\gamma}{l_i - 1} \right) \text{ almost surely as } N, T \rightarrow \infty$$

7.2 Principal components as regressors

In this section I investigate what happens when principal components are used as regressors instead of some true regressors. It is related to the footnote number 8. It was observed that when principal components from changes in CDS spreads are used as regressors in equation (2) then some eigenvalues from the correlation matrix of residuals (ε) are very close to zero but negative. This

is puzzling, because these eigenvalues should be at least zero. If the smallest eigenvalues behave strangely then one might be worried that the highest eigenvalues might also be distorted.

First, I want to check whether this particular approach does not induce some spurious correlation into the residuals from the regressions. In order to show that using principal components as regressors is a safe procedure, I perform two simple tests. First, I generate the data with one single factor. Then I show that once the first principal component is removed, the residuals behave just as they should for an uncorrelated dataset. Second, I compute residuals from changes in CDSs using the method from Zumbach (2009), which is described in details below. This second method also produces strongly correlated residuals. Therefore, I conclude that in order to describe the properties of residuals, one can use principal components as artificial regressors.

Simulation to check whether residuals from PC analysis retain the correlation from original data Here, I generate data with one strong factor,

$$X_t = \Lambda F_t + e_t$$

where F_t is a random number and Λ is an $N \times 1$ vector of loadings. In the simulation I use $T = 1000$ and $N = 200$. Parameters are chosen in such a way that the first and the highest eigenvalue obtained from X 's correlation matrix, explains 21.0% of the total variance, which is similar to empirical values in Table 2. The second highest eigenvalue explains only 1.8% of the total variance. Then using principal component analysis I remove the first, strongest factor and compute the residuals, just like in equation (2). I get a matrix of residuals $R^{(N \times T)}$, which has a $T \times N$ size. The expectation is that residuals $R^{(N \times T)}$ should not be correlated with each other. Indeed, this is the case. The simplest way to show it, is to check the eigenvalues for correlation matrix of $R^{(N \times T)}$. Figure 4 shows the plot of these eigenvalues and we can see that this time the highest eigenvalue explains only around 1% of the total variance. Clearly, there is no evidence for any common factor. Additionally, the lowest eigenvalue is almost zero but negative (-3.5e-16). There is no reason why the eigenvalue should be negative. However, since it is almost zero, we can infer that the problem must be related to some rounding, numerical errors. The main conclusion here is that the principal component analysis was able to remove all the information from the original X simulated data matrix.

This is not the case with changes in CDS spreads. Because the residuals exhibit distinctly different behavior, as can be seen from Figure 5.

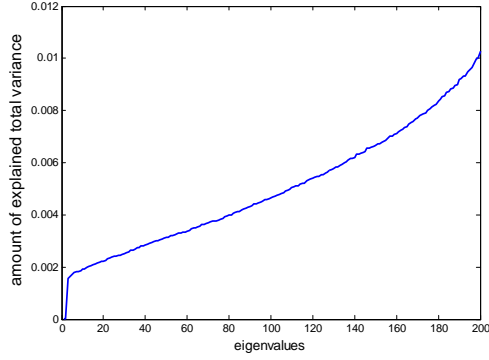


Figure 4: Eigenvalues of the residuals' correlation matrix from the simulated data.

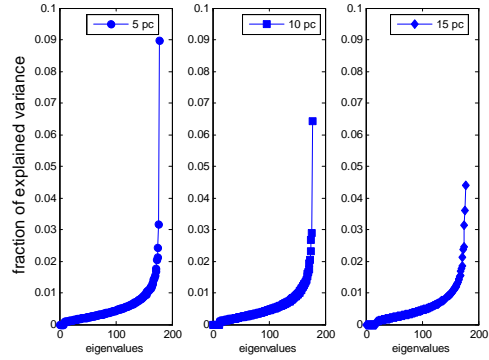


Figure 5: Eigenvalues of the residual matrix from CDS changes after removing 5, 10 and 15 principal components.

Different method to find residuals In order to be on the safe side, I will compute the residuals from the changes in CDS spreads with yet another method. In other words, I want to make sure that spikes in the eigenvalues of residuals that can be seen for example in Figure 5, are not created unintentionally by the method that I use. Hence, I will use the method from Zumbach (2009b) to compute the residuals. First, in order to diminish the impact of outliers, logarithms of the CDS spread levels are computed.

$$\begin{aligned}
 x_t &= \ln(Y_t) \\
 x_{t+\Delta t} &= x_t + r_{t+\Delta t} \\
 r_{t+\Delta t} &= \Sigma_t^{1/2} \varepsilon_{t+\Delta t}
 \end{aligned}$$

where $Y_t = CDS_t$, which is a vector of $1 \times N$ CDS spreads at time t . The vector r_t is the return over the time horizon $\Delta t = 1$ day. The matrix Σ_t is the variance/covariance over the next time period Δt and ε_t is the residual. The covariance should capture both the heteroscedasticity of financial returns and the cross-correlations across time series. The usual hypothesis is that the residuals are independent, but we will see that this is not true for the changes in CDS spreads.

The covariance matrix is given by a cross product of the return vector r_t and λ_i weight for the past returns.

$$\Sigma_t = \sum_{i=0}^{i_{\max}} \lambda_i r_{t-i\Delta t} r'_{t-i\Delta t}$$

The three most common choices for λ_i are equal, exponential and long memory weights¹¹. Here, I will use the long memory process, where the weights decay logarithmically slowly

$$\lambda_i = 1 - \frac{\ln(i\Delta t)}{\ln(\tau_0)}$$

where $i_{\max} = 260$, which is about one year of data. In the second step the covariance matrix is shrunk, using the methodology of Ledoit and Wolf (2009)

$$\Sigma(\gamma) = (1 - \gamma) \Sigma + \gamma \Sigma_{\text{diagonal}}$$

where Σ_{diagonal} is the diagonal part of Σ , which is the volatility of the respective time series. In essence, this equation shrinks only the off-diagonal part by $1 - \gamma$.

In order to compute the matrix of residuals ε , we need to invert Σ , because

$$\varepsilon_{t+\Delta t} = \Sigma_t^{-\frac{1}{2}} r_{t+\Delta t}$$

Provided that all the eigenvalues are positive, the inverse square root covariance matrix is

$$\Sigma_t^{-\frac{1}{2}} = \sum_{j=1}^N \frac{1}{\sqrt{e_j}} v_j v'_j$$

where e_j stands for an eigenvalue v_j is its corresponding eigenvector.

However, in applications most of the eigenvalues are close to zero. Zumbach (2009b) suggests first sorting eigenvalues and then choosing some "cut-off" eigenvalue $e_k > 0$. Then the inverse $\Sigma_t^{-\frac{1}{2}}$

¹¹Simple average is a good example of a long memory weights. Another one, which is used here, is the weights that decay logarithmically slowly.

is computed in the following way

$$\Sigma_t^{-\frac{1}{2}} = \sum_{j=1}^k \frac{1}{\sqrt{e_j}} v_j v_j' + \sum_{j=k+1}^N \frac{1}{\sqrt{e_{k+1}}} v_j v_j'$$

so all eigenvalues smaller than e_k are fixed at the level of e_{k+1} . The intention of the method is to preserve information in the eigenvectors corresponding to very small eigenvalues.

Finally, using the method described above I compute the matrix of residuals ε and then I check the eigenvalues of the $\rho(\varepsilon, \varepsilon)$ correlation matrix. The results are qualitatively just the same as those above for the simple residuals from principal components in Figure 5.

7.2.1 Details of the simulation from David's model

The fundamentals of the economy are described by three log-normal processes: Q_t is the price process, E_t , for the total earnings process and M_t for the real pricing kernel:

$$\begin{aligned} \frac{dQ_t}{Q_t} &= \beta_t dt + \sigma_Q dW_t, \sigma_Q = (\sigma_{Q1}, \sigma_{Q2}, 0) \\ \frac{dE_t}{E_t} &= \theta_t dt + \sigma_E dW_t, \sigma_E = (0, \sigma_{E2}, 0) \\ \frac{dM_t}{M_t} &= -k_t dt + \sigma_M dW_t, \sigma_M = (\sigma_{M1}, \sigma_{M2}, \sigma_{M3}) \end{aligned}$$

with a three dimensional vector of independent Wiener processes: $W_t = (W_{1t}, W_{2t}, W_{3t})$. Drifts β_t and θ_t can take different values and $k_t = \alpha + \alpha_\theta \theta_t + \alpha_\beta \beta_t$ is the real short rate of interest. There are N distinct states of the world and they are described by a vector $v_t = (\beta_t, \theta_t, -k_t)$ and so v_t follows an N state continuous-time finite state Markov chain with generator matrix Λ . The total risk of firm n is assumed to be the same for all firms and equal to the standard deviation of the market: $\sigma_n = \sigma_{E,2}$ and the firm's diffusion is an affine function of two independent Brownian motions:

$$dW^n = \rho dW_{2t} + \sqrt{1 - \rho^2} dW_{n,t}^{id}$$

where $dW_{n,t}^{id}$ is an idiosyncratic part of firm's n earnings and $\rho = 0.5$ for all firms. The belief updating process $\pi_t = (\pi_{1,t}, \dots, \pi_{N,t})$ follows the N -dimensional system of stochastic differential equations,

$$d\pi_{it} = \mu_i(\pi_t) dt + \sigma_i(\pi_t) d\widetilde{W}_t$$

where $\mu_i(\pi_t) = [\pi_t \Lambda]_i$, $\sigma_i(\pi_t) = \pi_{it} [v_i - \bar{v}(\pi_t)]^T (\Sigma^T)^{-1}$, $\Sigma = (\sigma_Q^T, \sigma_E^T, -\sigma_M^T)^T$, $\bar{v}(\pi_t) = \sum_{i=1}^N \pi_{it} v_i$ and $d\widetilde{W}_t = \Sigma^{-1} (v_t - \bar{v}(\pi_t)) dt + dW_t$. David shows that for π_{it} add up to one for every t ,

$$\sum_{i=1}^N \pi_{it} = 1$$

The nominal value of asset n is equal to

$$V_{nt} = \left(\sum_{i=1}^N C_i \pi_{it} \right) E_{nt}$$

where C_i represents investors' expectation of future earnings growth conditional on the state being v_i today and discounted with the pricing kernel process $\{M_t\}$. Hence, a high C_i implies that investors expect high value relative to current earnings. Constants C_i can be obtained from formulas (11) and (12) from David (2008). The nominal asset value volatility for firm n is equal to

$$\sigma_{V_n}^Q = \sigma_E + \sigma_Q + \frac{\sum_{i=1}^N C_i \pi_{it} (v_i - \bar{v}(\pi_t))^T (\Sigma^T)^{-1}}{\sum_{i=1}^N C_i \pi_{it}}$$

Formula for the credit spread is an extension of the Merton model, which accounts for stochastic growth rate and volatility of the asset value,

$$s(Z_t, \pi_t, t, T) = -\frac{1}{T} \log \left[\Pi_2(\cdot, \cdot) + \frac{G(\pi_t, z_t)}{B(\pi_t)} (1 - \Pi_1(\cdot, \cdot)) \right]$$

where

$$\begin{aligned} Z_t &= V_{nt}^Q / D_t^Q \\ z_t &= \log Z_t \\ B(\pi_t) &= f(z_t, 0, \pi_t, T - t) \\ G(\pi_t, z_t) &= f(z_t, 1/i, \pi_t, T - t), \quad i = \sqrt{-1} \\ f(z_t, \omega_1, \pi_t, T - t) &= E \left[e^{-\int_t^T r(\pi_s) ds} e^{i\omega_1 z_T} \mid z_t, \pi_t \right] \end{aligned}$$

Here, $\Pi_1(\cdot, \cdot)$ and $\Pi_2(\cdot, \cdot)$ are Arrow-Debreu securities that pay one dollar if the firm is solvent at time T , under two different measures and they are obtained by Fourier inversion of the $f(z_t, \omega_1, \pi_t, T - t)$ function. The data in the simulation study are generated only for one state of the world: v_1 . Consequently all the variation of beliefs comes only from the diffusion and not from the jumps in drifts.