Panel Data Models with Unobserved Multiple Time- Varying Effects to Estimate Risk Premium of Corporate Bonds

Bada, Oualid and Kneip, Alois

University of Bonn

19 October 2010

Abstract

We use a panel cointegration model with multiple time- varying individual effects to control for the missing factors in the credit spread puzzle. Our model specification enables as to capture the unobserved dynamics of the systematic risk premia in the bond market. In order to estimate the dimensionality of the hidden risk factors jointly with the model parameters, we rely on a modified version of the iterated least squares method proposed by Bai, Kao, and Ng (2009). Our result confirms the presence of four common risk components affecting the U.S. corporate bonds during the period between September 2006 and March 2008. However, one single risk factor is sufficient to describe the data for all time periods prior to mid July 2007 when the subprime crisis was detected in the financial market. The dimensionality of the unobserved risk components therefore seems to reflect the degree of difficulty to diversify the individual bond risks.

1 Introduction

In recent years, the use of panel data has attracted an increasing attention in empirical finance studies. This is motivated by the goal to control the so called *unobserved heterogeneity effect* which is undetectable in pure cross-section or pure time-series data (e.g., Hausman and Taylor (1981)). Recent discussions by Ahn, Lee, and Schmidt (2005), Bai (2009), Pesaran (2006), Bai and Kao and Ng (2009), and Kneip, Sickles and Song (2009) have focused on more advanced panel data models in which the unobservable heterogeneity has a multi-dimensional factor structure. Indeed, for many economic applications, the use of the classical panel data approaches will be inappropriate. When analysing stock or bond prices, for instance, the usual within estimation method assumes the heterogeneity effect to be time-invariant. However, the individual effects of such variables could be influenced by different time-changing and dynamic factors such as stochastic market trends, systematic risks, etc...

Exploiting the aptitude of this new generation of panel models to control for unobserved and complex heterogeneity, we consider in this paper the particular problem of the, so called, *credit spread puzzle*. Defined as the difference between a corporate bond yield and a duration-equivalent government bond yield, the credit spread has been considered for a long time to be a simple compensation for the credit risk default. Empirical evidence shows, however, that default risk can not be the unique determinant to explain such a large gap, see, e.g., Huang and Huang (2003) and Elton, Gruber, Agrawal and Manne (2001). Elton, Gruber, Agrawal, and Mann (2001) generated bond yields by implementing structural credit risk models.[1] They found that default probability could not explain more than 25% of the observed spot spreads[2].

[1]Credit risk models can be classified in two main categories: the structural models and the reduced form models. The basic framework of the first discipline consists to evaluate corporate bond prices by using option theory, see Merton (1974) and Black and Scholes (1973). The second discipline is based however on arbitrage theory, see, e.g., Duffie and Singleton (1997).

[2]Elton et al. (2001) define corporate sport rate and the government sport rate as the yield to maturity on a

Using reduced-form-model approach, Longstaff, Mithal, and Neis (2004) argue that non-default components such as bond-specific illiquidity and overall illiquidity risk do exist. This is because most corporate bonds are traded in thin markets which are related with higher transaction costs in compare to the trade markets of equities and Treasuries securities. Tax effect is also considered to be an important determinant of the credit spread puzzle. In contrary to the Treasury securities, corporate bonds are subject to tax-payments at the state level. Arbitrage theory implies hence that such a cost will be priced in the bond yields. Elton et al. (2001) find that the effect of taxes is depending on the rating level and the maturity. Amato and Remolona (2003) argue, however, that such dependency is weakly significant and that tax-effect is roughly constant across rating classes. Extending the structural models proposed by Longstaff and Schwartz (1995), Collin-Dufresne, Goldstein, and Martin (2001) examined the effect of a large number of risk proxies.[3] They detected high cross-correlations in the residuals of the regressed time series and conjectured that an undefined messing factor is generating this cross-section dependencies. The principal component analysis of the idiosyncratic residuals reveals that 75% of the unexplained variation can be captured by the first component. The authors examined additionally the effects of several macroeconomical and financial determinants and argued that such variables do not explain the detected common effect.

In the finance literature, this enigmatic discrepancy between the spread levels and the expected default risk is called the "'credit spread puzzle"'. According to Elton et al. (2001), Driessen (2004) and Amato and Remolona (2003), a possible explanation of this puzzle lies in the existence of an unavoidable systematic risk factor which is difficult to diversify. Fama and French (1993) investigated the yield spread of bond portfolios by using time series analysis.

---

zero-coupon corporate bond and the yield to maturity on a zero-coupon government bond respictively. The spot spread is accordingly defined as the difference between both rates.

[3]Collin-Dufresne et al. (2001) identify 6 theoretical determinants of the credit spread changes. The proposed components are: Changes in the Sport Rate, Changes in the Slope of the Yield Curve, Changes in Leverage, Changes in the Probability or Magnitude of a Downward Jump in Firm, and Changes in the Business Climate.

They determined two main factors: the first is related to the maturity and the second can be interpreted as the common default risk premium. The high fitting quality of the regression models shows that these two determinants can serve as a good explanation for bond portfolios. The authors did not discuss, however, why these factors can not explain individual bonds when observed separately. Kagraoka (2010) decomposed the credit spread into credit risk, liquidity risk, and an unobservable common component which he defined as systematic risk premium. The later is modelled by a unidimensional factor structure and estimated jointly with the remaining model parameters by using a panel data model with a simple time-varying heterogeneity effect.

Motivated by this enigma, we extend in this paper the empirical development of Kagraoka (2010) and allow for the unobserved systematic risk premium to have a multi-dimensional factor structure. In fact, such model specification enables us to control precisely for the cross-section dependencies detected in Collin-Dufresne et al. (2001). Although the credit spread was the focus of many empirical investigations in the finance literature, none of the previous studies has used a panel data model with multiple time-varying individual effects. Corporate bonds are indeed exposed to divers sources of uncertainty, so that the unobserved risk effect can be generated by multi-dimensional risk components. Moreover, our setting can provide an objective indicator to assess the difficulty of diversification mentioned in Elton et al. (2001) and Amato and Remolona (2003). We will show that by using state-of-the-art panel data models, it is possible to estimate the number of the unobserved common risk factors, say d, simultaneously with the effects of the observed risk components.

Our empirical model will be described in detail in Section 2. But after eliminating the rating class effects, it belongs to a class of panel data regression models which can generally be written in the form:

$$Y_{it} = X'_{it}\beta + \underbrace{F_t\Lambda'_i}_{(1\times d)\times(d\times 1)} + \epsilon_{it} \quad \text{for} \quad i = 1, \cdots, N \quad \text{and} \quad t = 1, \cdots, T. \tag{1}$$

A precise definition of the dependent and the independent variables used in our application are given in Section 2 and 3. The difference between (1) and the classical panel models consists in the unobserved factor structure $F_t\Lambda_i'$. Here, Λ_i is a $1 \times d$ vector of individual scores (or factor loadings) and F_t is $1 \times d$ vector of the unobservable common time-varying factors which we interpret as systematic risk components. ϵ_{it} is the idiosyncratic component.

To estimate panel models of the form (1), Bai (2009) proposes an iterated least squares method. The author considers the stationary case and provides asymptotic theory when N and T are both large . However, assuming the common factors to be $I(0)$ processes can be very restrictive in the practice specially when studying panels of security prices which are mostly affected by unknown stochastic trends. Bai et al. (2009) extend the theoretical development of Bai (2009) and consider the case where the cross sections share common stochastic trends of unit root processes. Using Bai's method, they prove that the asymptotic bias rising from the time series in such case can be consistently estimated and corrected. S. Ahn, Lee, and Schmidt (2006) consider the classical case where T is small and N large and estimate the model by using the Generalized Method of Moments (GMM) based on Instrumental Variables (IV). They show that the GMM estimators are more efficient than the iterated least squares estimator of Bai (2009) under fixed T. In contrast, Bai's method provides an alternative set-up if T is allowed to be large. A second critique on the iterative approach of Bai (2009) and Bai et al. (2009) is that the proposed method considers the number of the unobserved factors to be known which is, of course, not evident in the reality.[4] Pesaran (2006) attempts to control for the hidden structure by introducing additional regressors in the model, which are the cross-section (weighted) averages of the dependent variables and the (weighted) averages of the observed explanatory variables. The advantage of the proposed estimation method is its invariance face to the unknown factor structure dimension. However, the issue of identification requires special

[4]In this context, Bai and Ng (2002) and Bai (2004) propose appropriate panel information criteria in order to assess the number of the significant factors.

rank conditions, which are not always fulfilled in economic and finance data. A new approach based on a semi-parametric method and a (functional) factor analysis is proposed by Kneip, Sickles, and Song (2009).

Our empirical study relies on a large number of U.S. corporate-government bond yield spreads over a period of 397 business days. Our data are extracted from the on-line data base "'Datastream"'. The time series of the collected variables seems to possess unit roots.[5] We therefore choose to implement slightly modified versions of the estimators proposed in Bai et al. (2009), and propose an algorithmic refinement of the existing estimation method. Our algorithm enables us to estimate the number of the unobserved common factors jointly with the remaining model parameters and provides a practical and general iteration scheme which is easy to program. Furthermore, we provide detailed inference procedures. Our result confirms the presence of four uncorrelated systematic risk factors affecting the U.S. corporate bonds during the period between September 2006 and March 2008. However, one simple risk factor is sufficient to describe the data for all time periods prior to mid July 2007 when the subprime crisis was detected in the U.S. market. The dimensionality of the unobserved risk components therefore seems to reflect the degree of difficulty to diversify the individual bond risks in the financial market.

The remainder of this paper is organized as follows: Section 2 describes our basic panel data model and illustrates the explanatory variables to be considered. Section 3 proposes an algorithmic refinement of the estimation method proposed by Bai (2009) and Bai et al. (2009) in order to estimate jointly all the parameter of interest, namely, the common slope estimator, the interactive parameters as well as the optimal factor dimension. Section 4 offers a simple and brief pseudo code to program the estimation method in concrete terms. Section 5 discusses the

[5]The *Panel Analysis of Non-stationarity in Idiosyncratic and Common Components* (PANIC) proposed by Bai and Ng (2004), enables us to examine stationarity not only in the observed variables but also in the hidden time-varying common factors. The preliminary tests performed on the credit spread variable do not reject the unit root hypothesis as we expected.

bias correction procedure to re-center the limiting distribution of the slope estimator. Section 6 discusses the estimation procedure of the pre-eliminated group effects. Section 7 describes the data and presents our empirical results. Finally, conclusions and remarks are provided in Section 8.

2 The Model

We extend the idea of Kagraoka (2010) and decompose the corporate- government yield spread into credit risk, liquidity risk and an unknown number d of time varying systematic risk premia. More precisely, our panel model can be written as follows:

$$\mathrm{CS}_{it} = \mu_t + \mathrm{LR}_{it}\beta + \sum_{k=1}^{K} \delta_{ik}\alpha_{kt} + \sum_{l=1}^{d} \lambda_{il} f_{lt} + \varepsilon_{it} \quad \text{for} \quad i = 1, \cdots, N \quad \text{and} \quad t = 1, \cdots, T. \quad (2)$$

The explained variable CS_{it} is the corporate-government credit spread defined as

$$\mathrm{CS}_{it} = \mathrm{R}_{it} - \mathrm{R}_{G,it}$$

where R_{it} and $\mathrm{R}_{G,it}$ measure the corporate bond yield i at time t and its duration-equivalent government bond, respectively. The explanatory variable LR_{it} measures the liquidity risk of bond i at time t. Several proxies of illiquidity have been considered in the literature. [6] Following Bessembinder et al. (2005), we construct our measure based on the following quoted bid-ask spread:

$$\mathrm{LR}_{it} = \left| \frac{\mathrm{R}_{it}^{A} - \mathrm{R}_{it}^{B}}{\mathrm{R}_{it}^{B}} \right| \times 100$$

where R_{it}^{A} and R_{it}^{B} are the ask-yield and the bid-yield of bond i at time t. It is indeed easy to realize that the larger the spread, the more problematic the immediate trading becomes and vice versa. We expect the credit spread to be larger for less liquid bonds.

[6] See, e.g.,Bessembinder, Maxwell, and Venkataraman (2005), Chen, Lesmond, and Wei (2002), Houweling, Mentink, and Vorst (2005)and Lesmond (1999).

Following many previous studies, we consider the rating class to be a measure for assessing the credit default risk, see, e.g. Gebhardt, Hvidkjaer, and Swaminathan (2002), Houweling et al. (2005) and Kagraoka (2010). In fact, the credit level constitutes the synthetic evaluation of the rating agencies tacking into the account the default probability as well as the recovery rate. In our model this proxy is presented by the delta function δ_{ik} which we define as follows

$$\delta_{ik} = \begin{cases} 1 & \text{If bond } i \text{ has the rating level } k \text{ and} \\ 0 & \text{else,} \end{cases}$$

where the rating classes are nominally scaled from 1 to K. In our application, 1 stands for the best rating class and K for the worst one. In order to focus our analysis on the unobserved systematic risk premia, we consider only bonds which did not experienced a rating migration during the observation period. Different from the most existing works, we allow for time-varying rating effects α_{kt}. This establishes a general framework which enables us to to assess possible time changes in investors' behavior. In fact, investors may not necessary accord the same importance to the evaluation of the rating agency when acting in upward phases and when acting in crisis time. We expect credit risk to possess higher explanatory power than liquidity risk.

The stochastic process $\{f_{lt}\}$ represents the time pattern of underlying common risk factors which we expect to be non-stationary. Because f_{lt} is not depending on i, we may interpret these components as unavoidable systematic risks affecting the totality of the credit spreads. The scores λ_{il} are the corresponding individual loading parameters describing the effect of f_{lt} on each bond i independently of its rating class. The distinction between the role of the individual effects λ_i and the role of the time invariant rating levels can be expressed mathematically by condition

(R.1): $\sum_k^K \alpha_{kt} = 0$, $\sum_i^N \lambda_{il} = 0$ and $\sum_j^N \lambda_{jl}\delta_{jk} = 0$ for $l = 1, \ldots, d$ and $k = 1, \ldots, K$.

Note that (R.1) does not impose any restriction but only identifies the values α_{kt} and λ_{jl} in model 2. Analogously to Kagraoka (2007), we interpret the interaction between λ_{jl} and f_{lt} as the systematic risk premium imposed by the investors on the bond i at the time t.

The number of the systematic risk components d is considered to be unknown a priori and has to be estimated jointly with the remaining model parameters. In our analysis, we accord to the estimation of d a special attention. The role of d is in fact intended to determine the number of the missing factors explaining the credit spread puzzle. Because d is describing the space dimension of the orthogonal common risk factors, we interpret it as a measure to assess the degree of difficulty to diversify the individual bond risks in the market. The higher the factor dimension d, the more difficult to avoid the systematic risk.

An intrinsic problem of factor models consists in the fact that true factors only identifiable up to rotation. Therefore, in order to ensure the uniqueness of λ_i and f_{lt} (up to a sign change), we impose the following conditions which are commonly used in approximated factor analysis:

(R.2): $\sum_i^N \lambda_{il}\lambda_{ih} = 0$ for $l \neq h$ and $\sum_i^N \lambda_{i1}^2 \geq \sum_i^N \lambda_{i2}^2 \geq \cdots \geq \sum_i^N \lambda_{id}^2 > 0$

(R.3): $\sum_t^T f_{tl} f_{th} = 0$ for $l \neq h$ and $\frac{1}{T}\sum_t^T f_{tl}^2 = 1$ for all $l, h \in \{1, \cdots d\}$.

3 The Estimation Method

In a first step, we concentrate our presentation on estimating $\beta, f_{tl}, \lambda_{li}$ and d. For this purpose, the rating effects are eliminated from our model by using a group mean filtration. This simplifies the estimation procedure since it avoids to revert to constrained optimization techniques explicitly relying on (R.1). The parameters α_{kt} will be estimated in a second step (see Section 6). Let

$$\begin{aligned} y_{it} &= \mathrm{CS}_{it} - \sum_k^K \frac{1}{\sharp\{j|\delta_{ik}=1\}} \sum_j^N \mathrm{CS}_{jt}\delta_{jk} \\ x_{it} &= \mathrm{LR}_{it} - \sum_k^K \frac{1}{\sharp\{j|\delta_{ik}=1\}} \sum_j^N \mathrm{LR}_{jt}\delta_{jk}. \end{aligned} \tag{3}$$

Now, rewriting our model (2) with $Y_i = (y_{i1}, \cdots, y_{iT})'$, $X_i = (x_{i1}, \cdots, x_{iT})'$, $F = \{f_{tl}\}$ and $\Lambda_i = (\lambda_{i1}, \ldots, \lambda_{id})$, we get

$$Y_i = X_i\beta + F\Lambda_i' + \epsilon_i. \tag{4}$$

As outlined in the introduction, we consider here X and F to be $I(1)$ processes. Bai et al. (2009) propose, in this context, two methods to estimate β, F and Λ. The estimators, referred to as CupBC (continuously-updated and bias-corrected) and CupFM (continuously-updated and fully-modified) estimators, are the result of an iterated least squares approach as proposed by Bai (2009) combined with a bias correction technique. However, the proposed methods rely on a known factor dimension d. The authors suggest that, in practice, an appropriate d may be estimated separately by using a suitable information criterion, see, e.g., Bai and Ng (2002) and Bai (2004).

In this section, we propose a refined algorithmic in order to provide a joint estimation of all the parameters of interest, namely, the common slope estimator $\hat{\beta}$, the interactive parameters $\hat{\Lambda}_i$ and $\hat{F}$ as well as an estimate $\hat{d}$ of the factor dimension. The basic idea of our extension is to consider the continuously-updated estimators of Bai et al. (2009) as conditional estimators depending explicitly on the factor dimension. The latter is jointly estimated over all possible $\mathrm{d} = 0, 1, \ldots, d_{max}$ by means of a penalty term integrated directly in the global objective function to be optimized. The final solution of our algorithm is obtained by double iteration: inner iteration to obtain $\hat{\beta}, \hat{F}$ and $\hat{\lambda}$ and an outer loop to select $\hat{d}$. The updating procedure is repeated till convergence of all the parameters. We will show in the appendix that such an extension do not affect the asymptotic result elaborated by Bai et al. (2009). Our optimization criterion can be therefore defined as a penalized least squares objective function of the form:

$$S(\beta, F, \Lambda_i, \mathrm{d}) = \sum_i^N ||Y_i - X_i\beta - F\Lambda_i'||^2 + \mathrm{d}g_{\{N,T\}}, \tag{5}$$

where $g\{N, T\}$ is a penalty function depending exclusively on the sample size N and T. The appropriate choice of $g\{N, T\}$ will be discussed later. Note here, that, for known d, minimizing

$S(\beta, F, \Lambda_i|d)$ with respect to (β, F, Λ_i) corresponds exactly to optimize the objective function proposed by Bai (2009) and Bai et al. (2009). The role of the additional term $\mathrm{d}g_{\{N,T\}}$ in (5) is to pick up the optimal dimension of the unobserved factor structure $F\Lambda_i^{'}$. In order to focus our presentation on the algorithmic aspect of the estimation method, we define a functional hierarchy describing the way in which we will concentrate out our estimators iteratively:

$$\hat{\Lambda}_i \circ \hat{F} \circ \hat{\beta} \circ \hat{d} = \hat{\Lambda}_i \left(\hat{F} \left(\hat{\beta}(\hat{d}) \right) \right). \tag{6}$$

Cao (2010) named this technique as *Parameter Cascading* and used it to estimate complex mixed effects models with multi-level parameter structures. The algorithm is relatively easy to program and can be described in the following steps.

Step 1 (the individual parameters Λ_i): First, we concentrate out the individual parameters by minimizing the objective function $S(\beta, F, \Lambda_i, \mathrm{d})$ with respect to Λ_i for each given F, β and d. Because the penalty term is not depending on Λ_i, the intermediate optimization criterion at this stage can be expressed as:

$$S_1(\lambda_i|\beta, F, \mathrm{d}) = \sum_i^N ||Y_i - X_i\beta - F\Lambda_i^{'}||^2. \tag{7}$$

Now, minimizing for Λ_i and using restriction (R.3), we get

$$\hat{\Lambda}_i^{'}(\beta, F, \mathrm{d}) = \left(F'F\right)^{-1} F' \left(Y_i - X_i\beta\right) = F' \left(Y_i - X_i\beta\right)/T. \tag{8}$$

Step 2 (the time trend effects F): In order to estimate the multi-dimensional time effect F, we make use of result (8) from Step 1 and minimize a concentrated objective function $S_2(F|\beta, \mathrm{d})$ depending only on β and d. In fact, introducing (8) in (5) and neglecting again $\mathrm{d}g_{\{N,T\}}$, the new intermediate criterion $S_2(F|\beta, \mathrm{d})$ can be defined as

$$\begin{aligned} S_2(F|\beta, \mathrm{d}) &= \textstyle\sum_i^N ||Y_i - X_i\beta - F\hat{\Lambda}_i^{'}||^2 \\ &= \textstyle\sum_i^N ||\left[Y_i - X_i\beta\right] - \frac{FF'}{T}\left[Y_i - X_i\beta\right]||^2. \end{aligned} \tag{9}$$

Rearranging (9), we can see that minimizing $S_2(F|\beta, \mathrm{d})$ is equivalent to maximize the term $\sum_i^N ||\frac{FF'}{T}(Y_i - X_i\beta)||^2$. Solving for $F(\beta, \mathrm{d})$ subject to (R.3), we obtain the following result:

$$\hat{F}(\beta, \mathrm{d}) = \sqrt{T}\hat{P}(\beta, \mathrm{d}) \tag{10}$$

where $\hat{P}$ is a $T \times \mathrm{d}$ matrix binding the first d eigenvectors $(\hat{P}_1, \hat{P}_2, \cdots, \hat{P}_{\mathrm{d}})$ which correspond to the first d eigenvalues, $\hat{\rho}_1(\beta, \mathrm{d}) \geq \hat{\rho}_2(\beta, \mathrm{d}), \cdots, \geq \hat{\rho}_{\mathrm{d}}(\beta, \mathrm{d})$, of the empirical covariance matrix

$$\hat{\Sigma}(\beta, \mathrm{d}) = \frac{1}{NT}\sum_{i=1}^{N}\left([Y_i - X_i\beta(\mathrm{d})]\,[Y_i - X_i\beta(\mathrm{d})]'\right) \tag{11}$$

such that

$$\left[\hat{\Sigma}(\beta, \mathrm{d}) - \hat{\rho}_l(\beta, \mathrm{d})\right]\hat{P}_l(\beta, \mathrm{d}) = 0 \quad \text{for all} \quad l = 1, \ldots, \mathrm{d}. \tag{12}$$

Step 3 (the common slope parameter β): To estimate the common slope parameter, we reintegrate (8) and (10) in (5) and optimize the new intermediate objective function

$$S_3(\beta|\mathrm{d}) = \sum_{i}^{N} ||Y_i - X_i\beta - \hat{F}\hat{\Lambda}_i(\beta, \mathrm{d})||^2. \tag{13}$$

Note that $\hat{F}\hat{\Lambda}_i(\beta, \mathrm{d})$ depends nonlinearly on β. Minimizing $S_3(\beta|\mathrm{d})$ with respect to β for each given d leads hence to solve a system of nonlinear equations. Because an analytical solution is not given, optimization needs numerical techniques. Recall from the classical ordinary least squares method that the *infeasible* common slope estimator for known F and Λ_i is given by

$$\hat{\beta}_{infeasible}(\mathrm{d}) = \left[\sum_{i=1}^{N} X_i'X_i\right]^{-1}\left[\sum_{i=1}^{N} X_i'\left(Y_i - F\Lambda_i'\right)\right]. \tag{14}$$

Following Bai (2009), Bai et al. (2009), we propose to estimate β by replacing $F\Lambda_i'$ with $\hat{F}\hat{\Lambda}_i(\beta^{(0)}, \mathrm{d})$ for an appropriate starting value $\beta^{(0)}$ and updating (14) iteratively till convergence. At the optimum, the continuously-updated estimators (Cup) for β, F and Λ satisfy the following equality[7]

$$\hat{\beta}_{Cup}(\mathrm{d}) = \left[\sum_{i=1}^{N} X_i' X_i\right]^{-1} \left[\sum_{i=1}^{N} X_i' \left(Y_i - \hat{F}\hat{\Lambda}_i'(\hat{\beta}_{Cup}, \mathrm{d})\right)\right]. \tag{17}$$

Step 4 (the dimension d): Recall from equation (17) that our setting differs slightly from that proposed by Bai (2009) and Bai et al. (2009). In effect, our estimation algorithm treats $\hat{\beta}_{Cup}(\mathrm{d})$ as an estimator depending explicitly on the unknown factor dimension dimension which has to be jointly estimated from the data. In this regard, functions such as AIC, BIC, Mallows' Cp are well known in the model selection literature. However, such criteria do not consider the case of panel data models with simultaneously diverging N and T. Bai and Ng (2002) propose adjusted criteria in order to ensure consistency of the selection procedure in such cases. The basic idea of their approach consists simply to find a suitable penalty function which re-establishes the undesired variance minimization when d increases[8]. Explicitly, the optimal dimension $\hat{d}$ is obtained by minimizing numerically a panel criterion of the form

$$PC(\mathrm{d}) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(Y_{it} - \hat{Y}_{it}(\mathrm{d}))^2 + \mathrm{d}g_{\{N,T\}},$$

[7]Bai (2009) and Bai et al. (2009) use the analytical form of $\hat{\lambda}_i$ to express $S_3(\beta|\mathrm{d})$ such that

$$S_3(\beta) = \sum_{i}^{N} || (Y_i - X_i\beta) - \frac{1}{T}\hat{F}\hat{F}' (Y_i - X_i\beta) ||^2. \tag{15}$$

Alternatively to (17), the Cup-estimator of β, can be obtained by updating continuously

$$\hat{\beta} = \left(\sum_{i=1}^{N} X_i' M_{\hat{F}} X_i\right)^{-1} \left(\sum_{i=1}^{N} X_i' M_{\hat{F}} Y_i\right) \tag{16}$$

where $M_{\hat{F}} = I_{T\times T} - \hat{F}\hat{F}'/T$. However, programming $\hat{\beta}(\mathrm{d})$ as defined in (16) induces a slower routine which requires updating the inverse matrix $(\sum_{i=1}^{N} X_i' M_F X_i)^{-1}$ during each iteration.

[8]Kapetanios (2009) proposes alternatively a threshold approach based on the empirical distribution properties of the largest eigenvalue. The method requires the idiosyncratic errors to be independent and identically distributed. Onatski (2009) extended the approach of Kapetanios (2009) by allowing the errors to be either serially correlated or cross-sectionally dependent (but not both). Alternatively, S. C. Ahn and Horenstein (2009) propose to estimate d by maximizing the ratio of two adjacent eigenvalues (or the ratio of their growth rate).

where $g_{\{N,T\}}$ is a penalty function depending on the sample size N and T and scaled by a finite parameter. The appropriate choice of $g_{\{N,T\}}$ is however subject to the context in which we are modelling our data. Bai (2004) considers explicitly the case of I(1) common factors and proves under similar assumptions that any penalty function satisfying the following conditions

$$\begin{array}{ll} (i) & \lim_{N,T\to\infty} g_{\{N,T\}} \longrightarrow \infty \qquad \text{and} \\ (ii) & \lim_{N,T\to\infty} \frac{\log\log(T)}{T} g_{\{N,T\}} \longrightarrow 0, \end{array}$$

will be able to pick up the true factor dimension d with probability $P[\hat{d} = d] = 1$, as $N, T \to \infty$. It is however important to notice that the above setting assumes the factors to be extracted directly form observed variables without the presence of additional regressors in the model. In a similar context, Bai (2009) argue that estimating β jointly with F and Λ will not affect the analysis of Bai and Ng (2002) as long as β can be consistently estimated for any bounded $\bar{\mathrm{d}}$, such that $d \leq \bar{\mathrm{d}} \leq d_{max}$. Motivated by this argument, we introduce the penalty term in the global objective function presented in (5). Now, adopting the development of Bai (2004) and making use of result (8), (10) and (17), we define our dimensionality criterion as

$$S_4(\mathrm{d}) = \frac{1}{NT}\sum_{i}^{N} ||Y_i - X_i\hat{\beta}_{Cup}(\mathrm{d}) - \hat{F}\hat{\Lambda}_i(\hat{\beta}_{Cup}, \mathrm{d})||^2 + \mathrm{d}\hat{\sigma}^2 a\frac{\log(b)}{b}, \tag{18}$$

where $a = T/(4\log\log(T))$, $b = NT/(N+T)$ and $\hat{\sigma}^2$ denotes the variance estimator of the idiosyncratic errors ϵ_{it}. In practice, $\hat{\sigma}^2$ has a proper scaling role and can be simply replaced by $\hat{\sigma}^2_{d_{max}} = \frac{1}{NT}\sum_i^N ||Y_i - \hat{Y}_i(d_{max})||^2$. Note that, minimizing $S_4(\mathrm{d})$ numerically by mean of a naive selection procedure for all possible $\mathrm{d} = 0, 1, \ldots, d_{max}$ suffices to minimize entirely our global objective function defined in (5). This requires however extensive computations specially when N and T are large. To leave out such ambiguity, we propose to select $\hat{d}$ simply by replacing $\hat{\beta}_{Cup}(\mathrm{d})$ and $\hat{F}\hat{\Lambda}_i(\hat{\beta}_{Cup}, \mathrm{d})$ in (18) by $\hat{\beta}_{Cup}(\mathrm{d}^{(0)})$ and $\hat{F}\hat{\Lambda}_i(\hat{\beta}_{Cup}, \mathrm{d}^{(0)})$ for an appropriate starting value $\mathrm{d}^{(0)}$ and updating $\hat{d}$ iteratively. In fact, during each iteration stage, say m, the first term on the right hand-side of equation (18) can be simply expressed in term of the eigenvalues $\left\{\hat{\rho}_l(\hat{\beta}_{Cup}, \mathrm{d}^{(m-1)})\,|l = 0, 1, \ldots\right\}$. The latter do not require any extra computations because they are essentially needed to estimate $\hat{F}_{Cup}(\hat{\beta}_{Cup}, \mathrm{d}^{(m-1)})$ during the previous iteration (see Step 2). Optimizing $\mathrm{d}^{(m)}$ returns therefore to select directly the dimension corresponding to the smallest element of the following set

$$\left\{ \sum_{l=\mathrm{d}+1}^{T} \hat{\rho}_l(\hat{\beta}_{Cup}, \mathrm{d}^{(m-1)}) + \mathrm{d}\hat{\sigma}^2 a \frac{\log(b)}{b} \middle| \mathrm{d} = 0, 1, \ldots, d_{max} \right\}. \tag{19}$$

Finally, the global minimizer of the objective function (5) is obtained by double iterations: inner iteration to optimize $\hat{\beta}_{Cup}(\mathrm{d}), \hat{F}_{Cup}(\mathrm{d})$ and $\hat{\lambda}_{Cup}(\mathrm{d})$ for each d and an outer iteration to select the optimal dimension $\hat{d}$. The updating process is repeated entirely till convergence of all the parameters. The starting values and the iteration scheme will be discussed in Section 4. At the optimum, the obtained estimators, referred hereafter to as *entirely updated estimators* and denoted by Eup, satisfy the following equation system:

$$\begin{aligned}
\hat{d} &= \arg\min_{d} \left[\sum_{l=d+1}^{T} \hat{\rho}_l(\hat{\beta}_{Eup}, \hat{d}) \right] + d\hat{\sigma}^2 a \tfrac{\log(b)}{b}, \\
\hat{\beta}_{Eup} &= \left[\sum_{i=1}^{N} X_i' X_i \right]^{-1} \left[\sum_{i=1}^{N} X_i' \left(Y_i - \hat{F}\hat{\Lambda}_i'(\hat{\beta}_{Eup}, \hat{d}) \right) \right], \\
\hat{F}_{Eup} &= \sqrt{T}\hat{P}(\hat{\beta}_{Eup}, \hat{d}), \\
\hat{\Lambda}'_{Eup,i} &= \hat{F}'_{Eup} \left[Y_i - X_i \hat{\beta}_{Eup} \right] / T.
\end{aligned} \tag{20}$$

4 Iteration Scheme

Though the complex structure of our estimators, implementing the algorithm which optimize $S_1(\lambda_i|\beta, F, \mathrm{d}), S_2(F|\beta, \mathrm{d}), S_3(\beta|\mathrm{d})$ and $S_4(\mathrm{d})$ simultaneously is relative easy. In order to converge to the global optimum, the starting values $\beta^{(0)}$ and $\mathrm{d}^{(0)}$ should be however chosen with a minimum of precaution.

In this paper, we distinguish two natural choices of $\beta^{(0)}$: if the observed regressors are supposed to be uncorrelated with the factor structure, then we can start with the classical within estimator; in the contrary, if the unobserved time-varying factors and the observed regressors are expected to be highly correlated, then the within estimator can fail. In this case, we set $\beta^{(0)} = 0$ and start directly with the eigenvectors of the empirical covariance matrix defined in (11). As an initial dimension $\mathrm{d}^{(0)}$, we propose to choose an arbitrary mid-large integer d_{max}.

A simple pseudo code which optimize the global objective function $S(\lambda_i, F, \beta, \mathrm{d})$ presented in (5) can be simply described as following:

1. Set $\mathrm{d}^{(m)} = \begin{cases} d_{max} & \text{if} \quad m = 0 \\ \mathrm{d}^{(m-1)} & \text{if} \quad m > 0 \end{cases}$

2. Set $\beta^{(r)} = \begin{cases} \{0, \hat{\beta}_{within}\} & \text{if} \quad r = 0 \\ \beta^{(r-1)} & \text{if} \quad r > 0 \end{cases}$

3. Call (10) to calculate $F^{(r)} = \hat{F}(\beta^{(r)}, \mathrm{d}^{(m)})$

4. Call (8) to calculate $\Lambda_i^{(r)} = \hat{\Lambda}_i(F^{(r)}, \beta^{(r)}, \mathrm{d}^{(m)})$

5. Call (17) to update $\beta^{(r+1)} = \hat{\beta}(\mathrm{d}^{(m)} | F^{(r)} \Lambda_i^{(r)'})$

6. If $(\beta^{(r+1)} = \beta^{(r)})$ go to 7, else repeat 2 - 6 with $(r+1)$ instead of (r)

7. Select $\mathrm{d}^{(m+1)}$ according to (19)

8. If $(\mathrm{d}^{(m+1)} = \mathrm{d}^{(m)})$, exit, else go to 1 with $(m+1)$ instead of (m).

Furthermore, in order to speed up the computation process when $T > N$, we may reconstruct the algorithm with the functional hierarchy $\hat{F} \circ \hat{\Lambda} \circ \hat{\beta} \circ \hat{d}$ instead of $\hat{\Lambda} \circ \hat{F} \circ \hat{\beta} \circ \hat{d}$. The benefit of such modification is to calculate the eigenvectors of a smaller covariance matrix with a dimension $N \times N$ instead of $T \times T$. Both computations lead ultimately to the same result.

5 Inference and Bias Correction

Our theoretical set-up heavily relies on the basic work of Bai et al. (2009). A crucial condition is that X_i and F are generated by $I(1)$ processes. However, theoretical analysis of our model (4) encounters the additional complication that effects of group mean filtration has to be taken in the account.

Asymptotic properties of our estimators along with all underlying assumptions are given in the appendix. Fortunately, it can be shown that filtration effects are asymptotically negligible and that the results of Theorem 1 and 2 in Bai et al. (2009) generalize to our situation. In particular, the slope estimator $\hat{\beta}_{Cup}(d)$ to be obtained for the true factor dimension d is at least T consistent. The limiting distribution of $\sqrt{N}T(\hat{\beta}_{Cup}(d) - \beta)$ is however not centered at zero and given by

$$\sqrt{N}T(\hat{\beta}_{Cup}(d) - \beta) - \sqrt{N}\phi \xrightarrow{d} MN(0, \Sigma_c) \tag{21}$$

for some ϕ and Σ_c where MN denotes the mixed normal distribution .[9] We show in Theorem 2 in the appendix that $\hat{d}$ is a consistent estimator of d and that our final estimator $\hat{\beta}_{Eup} = \hat{\beta}_{Cup}(\hat{d})$ which has the same asymptotic properties as $\hat{\beta}_{Cup}(d)$. Note that performing the usual test statistics such as $t-$, $F-$ and χ^2 tests directly on $\hat{\beta}_{Eup}$, may lead to false interpretations. This is due to the presence of the bias term ϕ in (21). Analogously to Hahn and Kuersteiner (2002), Bai et al. (2009) prove that it is possible to construct a consistent estimator $\hat{\phi}$. Following their suggestion we define our *entirely updated and bias corrected* estimator by:

$$\hat{\beta}_{EupBC} = \hat{\beta}_{Eup} - \frac{1}{T}\hat{\phi}. \tag{22}$$

We want to emphasize that calculating $\hat{\phi}_i$ requires extra exertion because it is depending on unknown matrices, say Ω_i and Ω_i^* which are the long run and one-sided long run covariances of the vector containing the errors of the stochastic processes Y_{it}, X_{it} and F_t. Following Moon and

[9] The exact expression of ϕ and Σ_c is given in the appendix.

Perron (2004) and Bai et al. (2009), we estimate Ω_i and Ω_i^* by using a kernel estimator. The precise formulas for constructing $\hat{\phi}$ are given in the appendix.

The bias corrected estimator $\hat{\beta}_{EupBC}$ now satisfies (see Theorem 2 in the appendix)

$$\sqrt{N}T(\hat{\beta}_{EupBC} - \beta) \xrightarrow{d} MN(0, \Sigma_c). \tag{23}$$

A consistent estimator $\hat{\Sigma}_c$ of Σ_c is also defined in the appendix. This then allow us to calculate t-statistic and the test for the significance of $\hat{\beta}_{EupBC}$. After having calculated $\hat{\beta}_{EupBC}$ a final bias corrected estimators of F and Λ_i can be obtained by $\hat{F}_{EupBC} = \sqrt{T}\hat{P}(\hat{\beta}_{EupBC}, \hat{d})$ and $\hat{\Lambda}'_{EupBC,i} = \hat{F}'_{EupBC}\left(Y_i - X_i\hat{\beta}_{EupBC}\right)/T$, respectively. Theorem 2 in the appendix shows that $\hat{F}_{EupBC}$ provides consistent estimator of the true factors up to rotations.

6 Estimating the Group Effects

Recall from Section 3 that our objective function (5) is used only to estimate Λ, F, β and d. In order to estimate the pre-eliminated rating effects α_{kt} discussed in Section 2, we propose to use a dummy variable regression once $\hat{\beta}_{EupBC}$ and $\hat{F}\hat{\Lambda}'(\hat{\beta}_{EupBC}, d_{opt})$ are obtained. In fact, calculating $\hat{\alpha}_k$ does not require any iteration. This is due to restriction (R.1) which ensures a linear independency between α_{it} and Λ_i. The solution has consequently the same form as the classical fixed effects estimators:

$$\begin{aligned} \hat{\mu}_t &= \overline{\text{CS}}_{.t} - \overline{\text{LR}}_{.t}\hat{\beta}_{EupBC} \\ \hat{\alpha}_{kt} &= \overline{\text{CS}}_{kt} - \overline{\text{LR}}_{kt}\hat{\beta}_{EupBC} - \hat{\mu} \end{aligned} \tag{24}$$

where $\overline{\text{CS}}_{.t} = \frac{1}{N}\sum_i^N \text{CS}_{it}$, $\overline{\text{LR}}_{.t} = \frac{1}{N}\sum_i^N \text{LS}_{it}$, $\overline{\text{CS}}_{kt} = \frac{1}{m_k}\sum_i^N \delta_{ik}\text{CS}_{it}$, and $\overline{\text{LR}}_{kt} = \frac{1}{m_k}\sum_i^N \delta_{ik}\text{LS}_{it}$, with $m_k = \sharp\{i|\delta_{ik} = 1\}$ for $k = 1, \ldots, K$. Under our assumptions given in the appendix, it is easy to show that $\hat{\alpha}_{it}$ is $\sqrt{m_k}$ consistent and has an asymptotic normal distribution, such that

$$\sqrt{m_k}(\hat{\alpha}_{kt} - \alpha_{kt}) \xrightarrow{d} N(0, \sigma^2_{kt}) \tag{25}$$

where $\sigma_{kt}^2 = Var(\epsilon_{it.})$ with $\epsilon_{it.} = \frac{1}{m_k}\sum_i^N \delta_{ik}\varepsilon_{it}$. A consistent estimator of σ_{kt}^2 can be obtained by

$$\hat{\sigma}_{kt}^2 = \frac{1}{m_k}\sum_i^N \delta_{ik}\left(\mathrm{CS}_{it} - \widehat{\mathrm{CS}}_{it}\right)^2. \tag{26}$$

where $\widehat{\mathrm{CS}}_{it} = \hat{\mu}_t + \mathrm{LR}_{it}\hat{\beta}_{EupBC} + \sum_{k=1}^K \delta_{ik}\hat{\alpha}_{kt} + \sum_{l=1}^{\hat{d}} \hat{\lambda}_{il,Eup}\hat{f}_{lt,Eup}$. A 95% confidence interval for α_{it} can be therefore constructed at each time point t as follwoing

$$\left[\hat{\alpha}_{kt} - 1.96\frac{\hat{\sigma}_{kt}}{\sqrt{m_k}}, \hat{\alpha}_{kt} + 1.96\frac{\hat{\sigma}_{kt}}{\sqrt{m_k}}\right]. \tag{27}$$

7 Application: The Unobserved Risk Premia of Corporate Bonds

We write the estimation algorithm in a R-Package (hereafter to be called `phtt`). The routines as well as the R-workspaces and graphic codes are provided on the website of our Institute.

7.1 Data Description

Our data are extracted from Datastream which is an on-line database containing a broad range of financial entities and instruments. Our explained variable is the credit spread defined as the difference between the corporate bond yield and the treasury (or government) bond yield with the same maturity. Because the maturities for most of the bonds for which the spread is calculated will not exactly match the maturity of the available government benchmark bonds, Datastream uses a linear interpolation to estimate the yield of the duration-equivalent government benchmark. The spread is expressed as yield difference in basis points. The explanatory variables are the credit rating levels and the quoted bid-ask yield spread of the corresponding bonds. Our observation period extends from September 18, 2006 to May 25, 2008 including so the (first) alarms of the subprime mortgage problem detected in the U.S. market mid 2007. We choose U.S. corporate bonds rated by S&P. In order to focus our analysis on the impact of the unobserved time varying systematic risk premium, we eliminate bonds which experienced a rating migration during the observation period. Moreover, we choose fixed rate bonds with long remaining time to maturity. This is to marginalize the possible term structure effects. Finally, we neglect securities that have missing prices. We then obtain an equidistant panel data based on 96 U.S. corporate bonds over a period of 397 business days. The retained rating classes are AAA, AA, A, and BBB. The number of bonds by rating is presented in table 1. Descriptive statistics about the credit spread and the liquidity spread variables are summarized in table B.

The non-stationarity of the collected data is tested by using the PANIC-analysis proposed by Bai and Ng (2004). Several procedures for testing unit root hypothesis in panel data models have been proposed in the recent literature, see e.g. Bai and Ng (2004), Moon and Perron (2004), Pesaran (2007) and Chang (2002). However, The PANIC method (Panel Analysis of Non-stationarity in Idiosyncratic and Common Component) enables us to examine stationarity not only in the observed variables but also in the hidden time-varying factor structure. For our data the null hypothesis is not rejected.[10]

Figure 1 displays the 3 dimensional charts of the time series spread curves before and after the within group transformation discussed in Section 2.

7.2 Empirical Results and Interpretations

The results of our estimation method are presented in Table 3. The estimated liquidity risk effect is statistically significant and amounts to 0.1006726 with a standard deviation of 0.0050672. This result confirms the previous findings of Chen et al. (2002), Elton et al. (2001) and Kagraoka (2010). In fact, the more illiquid the bond, the higher the expected credit spread is.

Recall from our Model 2, that the rating effects are considered to be time varying parameters.

[10] The test was applied using X_i and $W_i = Y_i - X_i\hat{\beta}_{Eup}$.

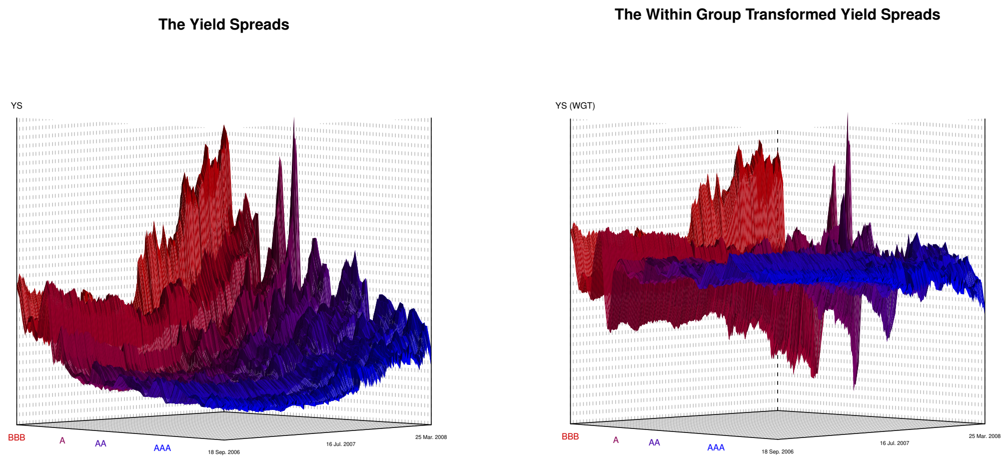


Figure 1: The spread curves before and after the within group mean filtration (in 3D)

Indeed, investors may not accord the same importance to the rating class when planing to invest in an upward phase as when acting in crisis tide. The time series of the estimated parameters $\hat{\alpha}_{kt}$ and their corresponding 95%− confidence intervals are depicted in Figure 2. The confidence intervals of the default risk parameters indicate that the rating effects are statistically significant, except for class A during the short time period from January to February 2008. This finding qualitatively agrees with the existing literature which confirms the significance of the default risk effect, see e.g. Collin-Dufresne et al. (2001), Huang and Hunag (2003) and Kagraoka (2010).

It is however clearly seen, that the time pattern exhibits some structural changes after July 16, 2007. In particular, the volatility of $\hat{\alpha}_{kt}$ over the rating classes $k = 1, \ldots, K$ increases. The stable negative effect of the rating class A during the first period registered some irregularity in the second period of time and became positive in 2008. It is important to note here that these structural changes coincide with the beginning of the subprime problems detected in the U.S. market mid July 2007. This indicates a change of investors' behavior how seem to possess different perception of the external credit rating in stable and crisis time.

The dimensionality criterion presented in (18) is minimized for $\hat{d} = 4$. This result confirms

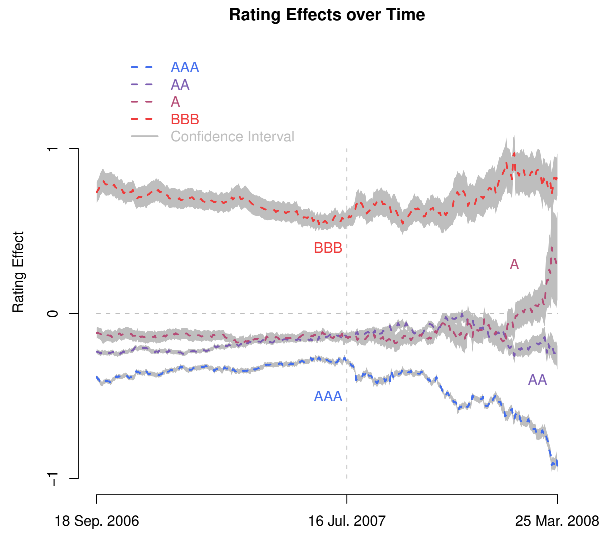
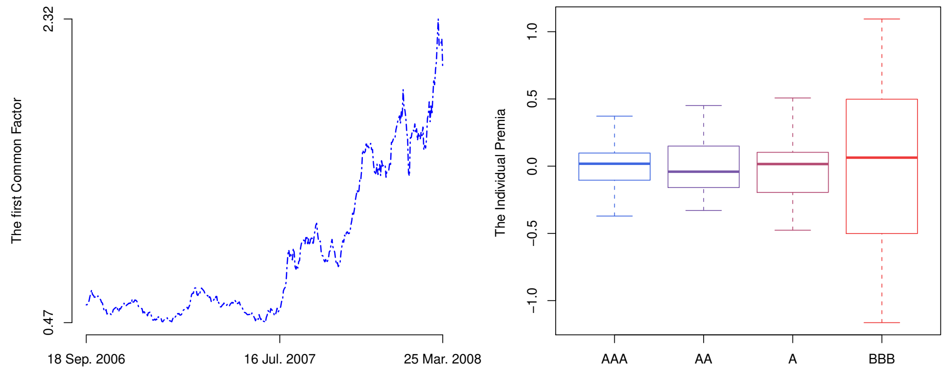


Figure 2: The time series of the estimated rating effects.

the presence of a multi-dimensional systematic risk affecting the yield spread variable. The first common risk factor $\hat{F}_{1t}$ and the boxplots of the corresponding individual loading parameters $\hat{\lambda}_{1i}$ (grouped by rating class) are displayed in Figure 3.[11] The structural change affecting the rating



(a) The first Time Varying Common Factor

(b) Boxplots of the First Loading Parameters by Rating

Figure 3: The first common factor and the loading parameters.

[11]Note that there is an ambiguity to interpret directly $\hat{F}_l$ and $\hat{\lambda}_{il}$ as estimators of F_l and λ_{il} because the sign of F_l is not uniquely identified by imposing the classical restrictions (R.1)- (R.3). To overcome this problem we propose to choose the sign of $\hat{F}_1$ such that $\sum_i^N \sum_t^T \hat{F}_{t1} Y_{it} > 0$. This is just to ensure that the time-varying common factor and the main trend of the credit spread curves maintain essentially the same movement direction.

effects in Figure 2 is explicitly shown in Figure 3-a. The trajectory of $\hat{F}_1$ is quite obviously non-stationary and registers a dramatic increase starting from July 16, 2007. This may be explained by the emergency of the subprime crises at this time. The boxplots presented in Figure 3-b show a high volatility of the loading parameters among the rating class BBB comparing to A, AA and AAA. The standard deviations of the individual parameters corresponding to the rating groups BBB, A, AA and AAA amount to $0.419, 0.0831, 0.0448$ and 0.0264 respectively. The individual time series of the first component $\hat{C}_{i1}$ obtained by

$$\hat{C}_{i1} = \hat{\lambda}_{i1}\hat{F}_1$$

are displayed in Figure 4. Bonds which have positive effects during the period from September

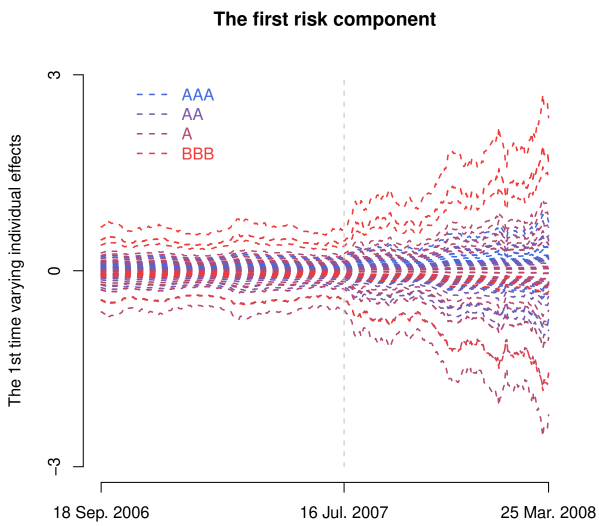


Figure 4: The first principal component.

18, 2006 to July 16, 2007 register an important raise in the next period. In the counterpart, bonds which perform badly in the first period due to the negative amounts of $\hat{C}_{i1}$ show further decreases in the yield spread during the second period. This result confirms the hypothesis of Jegadeesh and Titman (1993) and Chan, Jegadeesh, and Lakonishok (1996) who assert that security returns are affected by a so called *momentum effect*. This term is used to describe a typical investors' behavior which consists in buying stocks which have performed well in the past and selling stocks that have performed poorly. Our analysis thus sheds some light on an on-going discussion in the literature on stock market prices. Harvey and Siddique (2000) and Kang and Li (2009) argue that the momentum effect is an individual aspect which can not be explained by using common components. In the contrary, Chordia and Shivakumar (2002) and Chichernea and Slezak (2010) assert that momentum may effectively stem from some common factors.

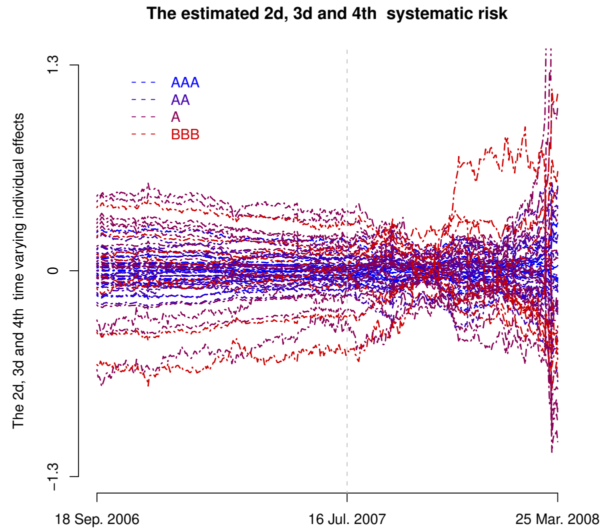


Figure 5: The second, third and fourth principal components.

In order to interpret the second, third and fourth individual risk components, we consider the sum

$$\hat{C}_{i,2,3,4} = \sum_{l=2}^{4} \hat{C}_{il} \quad \text{where} \quad \hat{C}_{il} = \hat{\lambda}_{il}\hat{F}_l.$$

Note that the product $\hat{\lambda}_{il}\hat{F}_l$ is by nature uniquely identifiable and does not require any additional restrictions. The obtained time series are depicted in Figure 5. Different from the first period, the individual time patterns of the vector $\hat{C}_{i,2,3,4}$, show complex and heavily heterogeneous structures during the subprime crisis. More interestingly, when re-estimating our model for the period spanning only the time before July, 16 2007, the dimensionality criterion discussed in

Section 3 detects the presence of just one $I(1)$ factor. The emergence of the additional risk components $\hat{C}_{i2}, \hat{C}_{i3}$ and $\hat{C}_{i4}$ seems hence to reflect the panic behaviour of the market actors during the crises. The number of the detected common risk factors can be interpreted as an index assessing the difficulty of diversification mentioned in Elton et al. (2001) and Amato and Remolona (2003) as a possible explanation of the credit spread puzzle. Indeed, the higher the number of the common risk factors, the more difficult it is to diversify the individual bond risks.

To summarize, we find out that the effects of the default risk account for an amount between 32.87% and 57.52% of the credi spread variation. This result roughly agree with thre previous findings reported by a large number of papers in the existing literature. Our analyis shows, however, that the explanatory power of the credit rating is lower in crisis time, especially for A and BBB bonds. Compared to the results of Elton et al. (2001), Huang and Hunag (2003), we find the contribution of liquidity risk to explain credit spread is relatively low (between 1% and 10.23%). Finally, we are successful to estimate the unobserved time varying systematic risk component which can neither be detected in the pure time series regressions nor in classical panel data analysis. Moreover, by allowing the systematic risk to have a multi-dimensional factor structure, we provide a more general framework than Kagraoka (2010) how assumes that unobserved heterogeneity can be described by exactly one common component. By estimating the factor dimension our approach provides an objective measure to assess the difficulty of diversifying the individual risk of corporate bonds in the financial market.

8 Conclusion

Motivated by the enigma of the credit spread puzzle, we used a panel coeintegration model with multi-time-varying individual effects and esetimate the unobserved systematic risk premium of corporate bonds together with the default risk and liquidity risk effects. Our model specification enables us to control for the cross-section dependencies detected in the most empirical researches applied on the credit spread variable, see e.g. Collin-Dufresne, Goldstein and Martin (2001). In order to estimate the number of factors jointly with the model parameters, we propose an algorithmic refinement of the iterated least squares method proposed Bai (2009). We are successful to estimate the unobserved time varying systematic risk component which can not be detected neither in pure time series regressions nor in the classical panel data analysis. Moreover, the joint estimation of the common factor dimension seems to provid an objective measure to assess the difficulty of diversifying the individual risk of corporate bonds. Our result confirms the presence of four common risk components affecting the U.S. corporate bonds during the period between September 2006 and March 2008. However, one single risk factor is however sufficient to describe the data for all time periods prior to mid July 2007 when the subprime crisis was detected in the U.S. market.

Our analysis neglects, however, the possible effect of taxes. The later can be introduced in the regression function by mean of a reasonable determinant. It is also important to note that the asymptotic properties of the the Eup estimator is elaborated for the case where X and F are generated by I(1) processes. This is, however, a special case of non-stationary processes and imposes a strong restriction of applicability.

A Assumptions and theoretical results

Throughout, we denote by M a finite positive constant, not depending on N and T. We write the Euclidean norm of a vector z as $||z||$. We use $B(.)$ to denote a Brownian motion process defined on $[0, 1]$ and $[\tau]$ to denote the largest integer $\leq \tau$.

We now consider inference of (4) as $(N, T) \to \infty$. Here, $(N, T) \to \infty$ has to be interpreted as a sequential limit: first $T \to \infty$ and then $N \to \infty$. For all N we assume an i.i.d random sample of individuals. In order to establish more generally applicable results we will consider a vector of $p \geq 1$ explanatory variables, i.e., $X_i \in \mathbb{R}^P$.

Our analysis relies on a slight modification of the arguments in Bai et al. (2009). It is important to note that common factors are only identifiable up to rotation. We assume that there are true underlying factors F^0 as well as corresponding loadings Λ^0 such that

$$Y_i = X_i\beta + F^0\Lambda_i^{0'} + \epsilon_i. \tag{28}$$

Our model assumptions will closely follow the setup of Bai (2004) and Bai et al. (2009). Common factors and explanatory variables are assumed to be $I(1)$ variables. However, we have to take into account that (28) has been obtained by subtracting group means.

Let $\delta_{ik} = 1$, if individual i belongs to rating class $k \in \{1, \ldots, K\}$, and $\delta_{ik} = 0$, else. Furthermore, set $m_k = \sum_{i=1}^{N} \delta_{ik}$, $k = 1, \ldots, K$. In this setup, δ_{ik} are i.i.d random variables and $\frac{m_k}{n}$ converges a.s. to $E(\delta_{ik})$ as $n \to \infty$. We will assume that $\inf_{k=1,\ldots,K} E(\delta_{ik}) > 0$.

Let $E_C(\cdot)$ denote conditional expectation given F^0. Our analysis will be based on the following assumptions:

(a) The loadings parameters: $E||\Lambda_i^0||^4 \leq M$; As $N \to \infty$, $E(\Lambda_i^0\delta_{ik}) = 0$ for all $k = 1, \ldots, K$, and $\frac{1}{N}\sum_i \Lambda_i^0 \Lambda_i^{0'} \xrightarrow{d} \Sigma_\Lambda$, a $d \times d$ positive definite matrix.

(b) The common stochastic trends: $F_t = F_{t-1} + \eta_t$, where η_t are zero mean random variables with $E||\eta_t||^{4+\gamma} \leq M$ for some $\gamma > 0$ and for all t; As $T \to \infty$, $\frac{1}{T^2}\sum_t F_t F_t' \xrightarrow{d} \int B_\eta B_\eta'$, a $d \times d$ random matrix, where B_η is a vector of Brownian motions with a positive definite covariance matrix Ω_η. η_t is independent of δ_{ik} for all i, t, k.

(c) Law of iterated logarithm: $\liminf_{T\to\infty} \frac{\log\log(T)}{T^2}\sum_{t=1}^T F_t F_t' = C$ a.s., where C is a non-random positive definite matrix.

(d) Explanatory variables: $X_{it} = X_{i,t-1} + \zeta_{it} - \sum_{k=1}^K \delta_{ik}(\bar{X}_{k,t-1} + \bar{\zeta}_{kt})$, where ζ_{it} are zero mean random variables and $\bar{X}_{k,t-1} = \frac{1}{m_k}\sum_{j=1}^n X_{j,t-1}\delta_{jk}$ as well as $\bar{\zeta}_{kt} = \frac{1}{m_k}\sum_{j=1}^n \zeta_{jt}\delta_{jk}$.

(e) Error term: $\epsilon_{it} = \varepsilon_{it} - \sum_{k=1}^k \delta_{ik}\varepsilon_{kt}$ with $\bar{\varepsilon}_{kt} = \sum_{j=1}^n \frac{1}{m_k}\sum_{j=1}^n \varepsilon_{jt}\delta_{jk}$. Here, ϵ_{it} are zero mean error terms and $E_C(\bar{\varepsilon}_{kt}|\delta_{ik} = 1) = 0$ for all k. Conditional on η_t the error terms ϵ_{it} are cross-sectionally independent, and also independent of X_{it}.

The additional terms $\bar{X}_{k,t-1}$, $\bar{\zeta}_{kt}$ and $\bar{\varepsilon}_{kt}$ in Assumption d) reflect our subtraction of group means. We want to emphasize that such transformation only influences the structure of error terms and explanatory variables, but not the factor F_t^0. The condition of uncorrelatedness of λ_{ij}^0 and λ_{il}^0 for $j \neq l$ just identifies the different common factors and does not impose any restriction. The requirement $E(\Lambda_i^0 \delta_{ik}) = 0$ for all $k = 1, \ldots, K$ is the population version for our condition (R.1) introduced for identifying α_{kt}. If ℓ_i^0 denote the (uncorrelated) loadings of the original model (2), then after subtraction of group means, (28) necessarily holds with loadings $\Lambda_i^0 = \ell_i^0 - \delta_{ik}\bar{\ell}_k$, where $\bar{\ell}_k = \sum_{k=1}^K \frac{1}{m_k}\sum_{j=1}^n \ell_j^0 \delta_{jk}$. If as $N \to \infty$, $\frac{1}{N}\sum_i \ell_i^0 \ell_i^{0'}$ converges to a diagonal matrix, then $\frac{1}{N}\sum_i \Lambda_i^0 \Lambda_i^{0'}$ converges to a diagonal matrix with diagonal entries $E\left((\ell_{ij}^0 - \sum_{j=1}^K \delta_{ik} E(\ell_{ij}^0|\delta_{ik} = 1))^2\right)$, $j = 1, \ldots, d$.

Let $\zeta_{it}^* = \zeta_{it} - \sum_{k=1}^K \delta_{ik}\zeta_{kt}^0$ with $\zeta_{kt}^0 = E_C(\zeta_{kt}|\delta_{ik} = 1)$, $k = 1, \ldots, K$, and define $X_{it}^* = X_{i,t-1}^* + \zeta_{it}^*$, $t = 2, \ldots, T$ with $X_{i1}^* = X_{i1} - \delta_{ik}\bar{X}_{k1}$. Note that if conditional on η_t the original random variables ζ_{it} are cross-sectionally independent, then also the transformed variables ζ^*_{it} are conditionally independent across i.

We need some further assumptions on the joint behavior of the $w_{it} = (\varepsilon_{it}, \zeta^*_{it}, \eta_t)$.

(f) The processes w_{it}: The multivariate processes $w_{it} = (\varepsilon_{it}, \zeta^*_{it}, \eta_t)$ are stationary. For each i, $w_{it} = \sum_{j=0}^{\infty} \Pi_{ij} v_{i,t-j}$, where $v_{it} = (v^{\varepsilon}_{it}, v^{\zeta}_{it}, v^{\eta}_t)$ are mutually independent over i, t as well as identically distributed over t. Furthermore, $E(v_{it}) = 0$, $E(v_{it}v^{'}_{it}) > 0$, and $E(\|v_{it}\|^8) \leq M$, where $M < \infty$ is independent of i, t. v_{it} are independent of λ^0_j for all i, j, t. In addition, all further conditions of Assumptions 2. and 3. of Bai et al. (2009) are satisfied.

(g) No cointegration: $\{X^*_{i,t}, F_t\}$ are not cointegrated.

It will be shown in Theorem 1 below that the additional effects introduced by subtracting group means are negligible, and that the inference results derived in Bai et al. (2009) also apply in our situation. The structure of the asymptotic distribution of $\beta_{Cup}(d)$, already mentioned in (21), involves a bias term ϕ and a covariance matrix Σ_c which are defined in Bai et al. (2009). Before stating the theorem we now define these quantities and show how to construct consistent estimates from given data.

Following again Bai et al. (2009) we first use kernel estimators to approximate the long-run covariance matrices of w_{it}. Estimates of ϵ_{it}, ζ_{it} and η_t are given by the regression residuals $\widehat{\epsilon}_{it}$, $\hat{\zeta}_{it} = \Delta X_{it}$ and $\widehat{\eta}_t = \Delta \widehat{F}_t$. For all $h = -T+1, \ldots, T-1$ and all $i = 1, \ldots, n$ let $\widehat{\Gamma}_{\epsilon,i}(h)$, $\widehat{\Gamma}_{\epsilon,\zeta,i}(h)$, $\widehat{\Gamma}_{\epsilon,\eta,i}(h)$, $\widehat{\Gamma}_{\epsilon,b,i}(h)$ and $\widehat{\Gamma}_{b,i}(h)$ denote the 1×1, $1 \times p$, $1 \times d$, $1 \times (p+d)$ and $(p+d) \times (p+d)$ empirical lag h autocovariance matrices of $(\widehat{\epsilon}_{it}, \widehat{\epsilon}_{i,t+h})$, $(\widehat{\epsilon}_{it}, \hat{\zeta}_{i,t+h})$, $(\widehat{\epsilon}_{it}, \widehat{\eta}_{t+h})$, $(\widehat{\epsilon}_{it}, (\hat{\zeta}'_{i,t+h}, \eta'_{t+h})')$

as well as $((\hat{\zeta}'_{i,t+h}, \hat{\eta}'_{t+h})', (\hat{\zeta}'_{i,t+h}, \hat{\eta}'_{t+h})')$, $t = 1, \ldots, T$. Then define

$$\widehat{\Omega}_{\epsilon,i} = \sum_{j=-T+1}^{T-1} \omega(\frac{j}{\kappa})\widehat{\Gamma}_{\epsilon,i}(j), \quad \widehat{\Omega}_{\epsilon,b,i} = \sum_{j=-T+1}^{T-1} \omega(\frac{j}{\kappa})\widehat{\Gamma}_{\epsilon,b,i}(j)$$

$$\widehat{\Omega}_{b,i} = \sum_{j=-T+1}^{T-1} \omega(\frac{j}{\kappa})\widehat{\Gamma}_{b,i}(j), \quad \widehat{\Omega}_{\epsilon|b,i} = \widehat{\Omega}_{\epsilon,i} - \widehat{\Omega}'_{\epsilon,b,i}\widehat{\Omega}_{b,i}\widehat{\Omega}_{\epsilon,b,i},$$

$$\begin{pmatrix} \widehat{\Delta}^+_{\zeta,\epsilon,i} \\ \\ \widehat{\Delta}^+_{\eta,\zeta,i} \end{pmatrix} = \begin{pmatrix} \sum_{j=0}^{T-1} \omega(\frac{j}{\kappa})\widehat{\Gamma}_{\epsilon,\zeta,i}(h) \\ \\ \sum_{j=0}^{T-1} \omega(\frac{j}{\kappa})\widehat{\Gamma}_{\epsilon,\eta,i}(h) \end{pmatrix} - \sum_{j=0}^{T-1} \omega(\frac{j}{\kappa})\widehat{\Gamma}_{b,i}(h)\widehat{\Omega}^{-1}_{b,i}\widehat{\Omega}_{\epsilon,b,i}.$$

Here, the kernel function ω satisfies the following assumption:

(h) The kernel function $\omega(.) : R \to [-1, 1]$ satisfies (i) $\omega(0) = 1$, $\omega(x) = \omega(-x)$, (ii) $\int_{-1}^{1} \omega(x)^2 dx < \infty$ and with Parzen'exponent $q \in (0, \infty)$ such that $\lim \frac{1-\omega(x)}{|x|^q} < \infty$ and $\liminf_{N,T\leftrightarrow\infty} \left(\frac{log(T)}{log(N)}\right) > 1$; the bandwidth parameter $\kappa \sim N^b$ where $b \in (\frac{1}{2q}, \liminf \left(\frac{log(T)}{log(N)}\right) - 1)$.

$\widehat{\Omega}_{\epsilon,i}$, $\widehat{\Omega}_{\epsilon,b,i}$, $\widehat{\Omega}_{b,i}$, $\widehat{\Omega}_{\epsilon|b,i}$, $\widehat{\Delta}^+_{\zeta,\epsilon,i}$ and $\widehat{\Delta}^+_{\eta,\zeta,i}$ estimate their theoretical analogues $\Omega_{\epsilon,i}$, $\Omega_{\epsilon,b,i}$, $\Omega_{b,i}$, $\Omega_{\epsilon|b,i}$, $\Delta^+_{\zeta,\epsilon,i}$ and $\Delta^+_{\eta,\zeta,i}$ which are defined by replacing in the above equation the terms $\omega(\frac{\cdot}{\kappa})\widehat{\Gamma}.(\cdot)$ by the corresponding true autocovariance matrices of $w_{it} = (\varepsilon_{it}, \zeta^*_{it}, \eta_t)$. In addition, summation then ranges from $-\infty$ to ∞ (instead of $-T+1$ to $T-1$) and 0 to ∞ (instead of 0 to $T-1$).

Now define the projection matrices of F^0 as $\widehat{M_F} = I_T - F^0(F^{0\prime}F^0)^{-1}F^{0\prime}$ and the scalar a_{ik} as the element i, k of the projection matrix $A_\Lambda = \Lambda^0(\Lambda^{0\prime}\Lambda^0)^{-1}\Lambda^{0\prime}$. Corresponding estimates $\widehat{M_F}$ and $\hat{a}_{ik}$ are obtained by replacing F^0 and Λ^0 by $\widehat{F}$ and $\widehat{\Lambda}$ Then let

$$Z_i = M_F X^*_i - \frac{1}{N}\sum_{j=1}^{N} \widehat{M_F} X^*_j a_{ij}, \quad \widehat{Z}_i = \widehat{M_F} X_i - \frac{1}{N}\sum_{j=1}^{N} \widehat{M_F} X_j \hat{a}_{ij}$$

Conditional on F^0, the bias term ϕ_{NT} is then given by

$$\phi = \left(\frac{1}{NT^2}\sum_{i=1}^{N} Z_i^{'} Z_i\right)^{-1} \frac{1}{N}\sum_{i=1}^{N}\theta_i,$$

$$\theta_i = Z_i^{'}(\Delta b_i)\Omega_{b,i}\Omega_{\epsilon,b,i} + \Delta^{+}_{\zeta,\epsilon,i} - \delta_i^{'}\Delta^{+}_{\eta,\zeta,i},$$

$$\Delta\hat{b}_i = (\Delta X_i^* - \frac{1}{N}\sum_{j=1}^{N}\Delta X_j^* a_{ij}, \Delta F^0), \quad \delta_i = (F^{0'}F^0)^{-1}F^{0'}X_i,$$

and a consistent estimator can be determined by

$$\widehat{\phi}_{NT} = \left(\frac{1}{NT^2}\sum_{i=1}^{N}\hat{Z}_i^{'}\hat{Z}_i\right)^{-1}\frac{1}{N}\sum_{i=1}^{N}\widehat{\theta}_i,$$

$$\widehat{\theta}_i = \hat{Z}_i^{'}(\Delta\hat{b}_i)\widehat{\Omega}_{b,i}\widehat{\Omega}_{\epsilon,b,i} + \widehat{\Delta}^{+}_{\zeta,\epsilon,i} - \hat{\delta}_i^{'}\widehat{\Delta}^{+}_{\eta,\zeta,i},$$

$$\Delta\hat{b}_i = (\Delta X_i - \frac{1}{N}\sum_{j=1}^{N}\Delta X_j\hat{a}_{ij}, \Delta\widehat{F}), \quad \hat{\delta}_i = (\widehat{F}^{'}\widehat{F})^{-1}\widehat{F}^{'}X_i.$$

Conditional on F^0, Bai et al. (2009) show that there exists random matrices R_{Ci}, defined as conditional expectations of integrated Brownian motions with individually different covariance structure, such that as $(N,T)\to\infty$ we have $\frac{1}{NT^2}\sum_{i=1}^{N} Z_i Z_i^{'} \to_d \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N} R_{Ci}$. The covariance matrix Σ_c is then defined by

$$\widehat{\Sigma}_c = \left(\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}R_{Ci}\right)^{-1}\left(\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\Omega_{\epsilon|b,i}R_{Ci}\right)\left(\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}R_{Ci}\right)^{-1}, \tag{29}$$

Bai et al. (2009) do not propose an estimator of Σ_c. However, following their arguments it is straightforward to show that a consistent estimate of the covariance matrix Σ_c is given by

$$\widehat{\Sigma}_c = \left(\frac{1}{NT^2}\sum_{i=1}^{N}\hat{Z}_i^{'}\hat{Z}_i\right)^{-1}\frac{1}{NT^2}\sum_{i=1}^{N}\widehat{\Omega}_{\epsilon|b,i}\hat{Z}_i^{'}\hat{Z}_i\left(\frac{1}{NT^2}\sum_{i=1}^{N}\hat{Z}_i^{'}\hat{Z}_i\right)^{-1},$$

We then obtain the following theorem:

Theorem 1: *Under the above assumptions we have as* $(N,T)\to\infty$

a)

$$\Sigma_c^{-1/2}\left(\sqrt{N}T(\hat{\beta}_{Cup}(d)-\beta)-\sqrt{N}\phi)\right) \xrightarrow{d} N(0,I_p) \tag{30}$$

where ϕ *and* Σ_c *are defined as above.*

b) $\widehat{\Sigma}_c$ *and* $\widehat{\phi}_{NT}$ *constitute consistent estimators of* ϕ_{NT} *and* Σ_c*. Furthermore, the bias corrected estimator* $\hat{\beta}_{CupBC}(d) - \frac{1}{T}\widehat{\phi}_{NT}$ *satisfies*

$$\Sigma_c^{-1/2}\left(\sqrt{N}T(\hat{\beta}_{CupBC}(d) - \beta)\right) \xrightarrow{d} N(0, I_p) \tag{31}$$

Proof: Consider Assertion a) and let $\tilde{Z}_i = M_F X_i - \frac{1}{N}\sum_{j=1}^N \widehat{M}_F X_j a_{ij}$. In view of Proposition 4 and Lemma A.2 of Bai et al. (2009) we only have to show that $\frac{1}{NT^2}\sum_{i=1}^N \tilde{Z}_i'\tilde{Z}_i$ and $\frac{1}{\sqrt{N}T}\sum_{i=1}^N \tilde{Z}_i'\epsilon_i$ have the same limit distributions as $\frac{1}{T^2}Z_i'Z_i$ and $\frac{1}{\sqrt{N}T}\sum_{i=1}^N Z_i'\varepsilon_i$. Note that $X_{it} - X_{it}^* = X_{i,t-1} - X_{i,t-1}^* + (\sum_{k=1}^K \delta_{ik}(\bar{\zeta}_{kt} - \zeta_{kt}^0))$ is also an $I(1)$-process. But the innovations $(\sum_{k=1}^K \delta_{ik}(\bar{\zeta}_{kt} - \zeta_{kt}^0))$ are averages over m_k individuals and hence $var_C((\sum_{k=1}^K \delta_{ik}(\bar{\zeta}_{kt} - \zeta_{kt}^0)) \leq M_1/N$, where $M_1 < \infty$ is some constant independent of i, t. Therefore, as $(N, T) \to \infty$ $\|\frac{1}{NT^2}\sum_{i=1}^N \tilde{Z}_i\tilde{Z}_i' - \frac{1}{NT^2}\sum_{i=1}^N Z_i Z_i'\| = O_P(N^{-1/2})$. Our assumptions imply that conditional on η the random variables ζ_{it}^* and ε_{it} are independent. Consequently, as $(N, T) \to \infty$, we have $\frac{1}{\sqrt{N}T}\sum_{i=1}^N \tilde{Z}_i'(\sum_{k=1}^K \delta_{ik}\bar{\varepsilon}_k) = \sum_{k=1}^K(\frac{1}{\sqrt{N}T}\sum_{i=1}^N \delta_{ik}\tilde{Z}_i')\bar{\varepsilon}_k = o_P(1)$ as well as $\frac{1}{\sqrt{N}T}\sum_{i=1}^N (Z_i' - \tilde{Z}_i')\varepsilon_i = o_P(1)$. One can conclude that $\frac{1}{\sqrt{N}T}\sum_{i=1}^N \tilde{Z}_i'\epsilon_i = \frac{1}{\sqrt{N}T}\sum_{i=1}^N Z_i'\varepsilon_i + o_p(1)$. Assertion a) is an immediate consequence. Assertion b) follows from a straightforward generalization of the arguments used in the proof of Theorem 2 of Bai et al. (2009). All additional terms induced by the differences $X_{it}^* - X_{it}$ and $\varepsilon_{it} - \epsilon_{it}$ are asymptotically negligible. $\Box$

A difference between our approach and the methodology of Bai et al. (2009) consists in the fact that our estimation procedure directly incorporates a dimension estimate. Our final estimator $\hat{\beta}_{EupBC} = \hat{\beta}_{CupBC}(\hat{d})$ thus relies on the estimated dimension $\hat{d}$. The following theorem shows that with high probability $\hat{d}$ will asymptotically coincide with the true dimension d. The asymptotic distributions derived in Theorem 1 thus remain valid when replacing $\hat{\beta}_{CupBC}(d)$ by $\hat{\beta}_{EupBC}$. Furthermore, the final estimator $\widehat{F}_{EupBC}$ yields a consistent estimator of the true factor structure (up to rotations).

Theorem 2: *Under the above assumptions we have as* $N, T \to \infty$

a)

$$\Sigma_c^{-1/2}\left(\sqrt{N}T(\hat{\beta}_{EupBC}-\beta)\right) \xrightarrow{d} N(0, I_p) \tag{32}$$

b For some $d \times d$ invertible matrix A

$$\frac{1}{T}\sum_{t=1}^{T}\|\frac{1}{T}\widehat{F}_{EupBC,t} - H \cdot F_t^0\|^2 = O_P(\frac{1}{N}) + O_P(\frac{1}{T})$$

Proof: We can infer from the theoretical results of Bai et al. (2009) that $\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(Y_{it} - \hat{Y}_{it}(d))^2 = O_P(1)$ holds for the true dimension d. Since estimates are obtained by least squares, this immediately implies that $\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(Y_{it}-\hat{Y}_{it}(r))^2 = O_P(1)$ for all $r = d, d+1, \ldots, d_{max}$. On the other hand, for $r < d$ a straightforward generalization of the arguments of Bai (2004) shows that, as $(N,T) \to \infty$, $\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(Y_{it} - \hat{Y}_{it}(d))^2$ a.s. tends to infinity with rate at least $T/\log\log(T)$. By definition of the penalty term in (18) we can therefore conclude that $\lim_{(N,T)\to\infty} P(\hat{d} = d) = 1$. Hence,

$$\begin{aligned}\Sigma_c^{-1/2}\left(\sqrt{N}T(\hat{\beta}_{CupBC}(\hat{d})-\beta)\right) &= \Sigma_c^{-1/2}\left(\sqrt{N}T(\hat{\beta}_{CupBC}(d)-\beta)\right) + \Sigma_c^{-1/2}\left(\sqrt{N}T(\hat{\beta}_{CupBC}(\hat{d})-\hat{\beta}_{CupBC}(d)\right) \\ &= \Sigma_c^{-1/2}\left(\sqrt{N}T(\hat{\beta}_{CupBC}(d)-\beta)\right) + o_P(1)\end{aligned}$$

Assertion a) is an immediate consequence. Assertion b) follows from $\lim_{(N,T)\to\infty} P(\hat{d} = d) = 1$ and Proposition 5 of Bai et al. (2009). □

B Data: Descriptive statistics

Rating class	AAA	AA	A	BBB	Total
Number	24	30	27	34	115

Table 1: The number of corporate bonds by rating class

C Results

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	Sd
CS	-5.5890	-1.0990	0.0000	0.07261	1.2530	5.5610	1.5043
LS	-4.2450	-0.0901	0.0000	0.00395	0.0896	3.3560	0.4377

Table 2: Descriptive statistics

References

Ahn, S., Lee, Y. H., & Schmidt, P. (2006). *Panel Data Models with Multiple Time-Varying Individual Effects.*

Ahn, S. C., & Horenstein, A. R. (2009). *Eigenvalue Ratio Test for the Number of Factors.*

Amato, J. D., & Remolona, E. M. (2003, October). The credit-spread puzzle. *BIS Quarterly Review*(December), 51–64.

Bai, J. (2004). Estimating cross-section common stochastic trends in nonstationary panel data. *Journal of Econometrics*, *122*(1), 137–183. Available from http://linkinghub.elsevier.com/retrieve/pii/S0304407603002720

Bai, J. (2009). Panel Data Models With Interactive Fixed Effects. *Econometrica*, *77*(4), 1229–1279. Available from http://doi.wiley.com/10.3982/ECTA6135

Bai, J., Kao, C., & Ng, S. (2009). Panel cointegration with global stochastic trends. *Journal of Econometrics*, *149*(1), 82–99. Available from http://linkinghub.elsevier.com/retrieve/pii/S0304407608002066

Bai, J., & Ng, S. (2002). Determining the Number of Factors in Approximate Factor Models. *Econometrica*, *70*(1), 191–221.

Bai, J., & Ng, S. (2004, July). A PANIC Attack on Unit Roots and Cointegration. *Econometrica*, *72*(4), 1127–1177.

Bessembinder, H., Maxwell, W., & Venkataraman, K. (2005). *Optimal Market Transparency : Evidence from the Initiation of Trade Reporting in Corporate Bonds.*

The Liquidity Effect					
Effekts	Parameters	Estimation	Std	$z-$value	$p-$value
Liq.Spr.	β	0.1081998	0.00556194	16.905	$< 2e - 16$

The Rating Effects					
Effekts	Parameters	1st Qu.	Mean	Median	3rd Qu.
AAA	α_1	-0.4732	-0.4126	-0.3628	-0.3274
AA	α_2	-0.2161	-0.1587	-0.1596	-0.1156
A	α_3	-0.1450	-0.1074	-0.1340	-0.1174
BBB	α_4	0.6137	0.6787	0.6774	0.7261

The Factor Dimension		
Criterion	Parameters	Estimation
IPC	d	4

The Unobserved Multi- Interactive Parameters					
		Min	Mean	Median	Max
1st Fact.	F_1	-3.9658201	-0.4025673	0.2067615	0.2976783
	λ_1	-793.64243	-42.19382	-31.96755	38.34758
2nd Fact.	F_2	-1.3018116	0.8342434	1.0586595	1.5216967
	λ_2	-793.64243	-42.19382	-31.96755	38.34758
3rd Fact.	F_3	-3.9658201	-0.4025673	0.2067615	0.2976783
	λ_3	-793.64243	-42.19382	-31.96755	38.34758
4th Fact.	F_4	-1.3018116	0.8342434	1.0586595	1.5216967
	λ_4	-95.48161	-23.83846	-20.33178	73.02035

Table 3: Parameter Estimation

Black, F., & Scholes, M. (1973). The Pricing of Options and Corporate Liabilities. *The Journal of Political Economy*, *81*(3), 637–654.

Cao, J. (2010, March). Linear Mixed-Effects Modeling by Parameter Cascading. *Journal of the American Statistical Association*, *105*(489), 365–374. Available from `http://pubs.amstat.org/doi/abs/10.1198/jasa.2009.tm09124`

Chan, L. K. C., Jegadeesh, N., & Lakonishok, J. (1996, December). Momentum Strategies. *The Journal of Finance*, *51*(5), 1681. Available from `http://www.jstor.org/stable/2329534?origin=crossref`

Chang, Y. (2002, October). Nonlinear IV unit root tests in panels with cross-sectional dependency. *Journal of Econometrics*, *110*(2), 261–292. Available from `http://linkinghub.elsevier.com/retrieve/pii/S0304407602000957`

Chen, L., Lesmond, D. A., & Wei, J. (2002). Bond Liquidity Estimation and the Liquidity Effect in Yield Spreads. *Wall Street Journal*.

Chichernea, D., & Slezak, S. L. (2010). *Idiosyncratic Risk Premia and Momentum.*

Chordia, T., & Shivakumar, L. (2002, April). Momentum, Business Cycle, and Time-varying Expected Returns. *The Journal of Finance*, *57*(2), 985–1019. Available from `http://doi.wiley.com/10.1111/1540-6261.00449`

Collin-Dufresne, P., Goldstein, R. S., & Martin, J. S. (2001, December). The Determinants of Credit Spread Changes. *The Journal of Finance*, *56*(6), 2177–2207. Available from `http://www.blackwell-synergy.com/doi/abs/10.1111/0022-1082.00402`

Driessen, J. (2004, August). Is Default Event Risk Priced in Corporate Bonds? *Review of Financial Studies*, *18*(1), 165–195. Available from `http://rfs.oupjournals.org/cgi/doi/10.1093/rfs/hhi009`

Duffie, D., & Singleton, K. J. (1997, September). An Econometric Model of the Term Structure of Interest-Rate Swap Yields. *The Journal of Finance*, *52*(4), 1287. Available from `http://www.jstor.org/stable/2329437?origin=crossref`

Elton, E. J., Gruber, M. J., Agrawal, D., & Mann, C. (2001, February). Explaining the Rate Spread on Corporate Bonds. *The Journal of Finance*, *56*(1), 247–277. Available from `http://www.blackwell-synergy.com/doi/abs/10.1111/0022-1082.00324`

Fama, E., & French, R. (1993, February). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, *33*(1), 3–56. Available from `http://linkinghub.elsevier.com/retrieve/pii/0304405X93900235`

Gebhardt, W. R., Hvidkjaer, S., & Swaminathan, B. (2002). Stock and Bond Market Interaction: Does Momentum Spill Over? *SSRN Electronic Journal*. Available from `http://www.ssrn.com/abstract=319009`

Hahn, J., & Kuersteiner, G. (2002). Asymptotically Unbiased Inference for a Dynamic Panel Model with Fixed Effects When Both n and T Are Large. *Econometrica*, *70*(4), 1639–1657.

Harvey, C. R., & Siddique, A. (2000, June). Conditional Skewness in Asset Pricing Tests. *The Journal of Finance*, *LV*(3), 1263–1295. Available from `http://www.blackwell-synergy.com/links/doi/10.1111/0022-1082.00247`

Houweling, P., Mentink, a., & Vorst, T. (2005, June). Comparing possible proxies of corporate bond liquidity. *Journal of Banking & Finance*, *29*(6), 1331–1358. Available from `http://linkinghub.elsevier.com/retrieve/pii/S0378426604000998`

Huang, J.-z., & Hunag, M. (2003). *How Much of the Corporate-Treasury Yield Spread is Due to Credit Risk ? 1.*

Jegadeesh, N., & Titman, S. (1993, March). Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency. *The Journal of Finance*, *48*(1), 65–91. Available from `http://www.jstor.org/stable/2328882?origin=crossref`

Kagraoka, Y. (2010, January). A time-varying common risk factor affecting corporate yield spreads. *The European Journal of Finance*, *16*(6), 527–539.

Kang, Q., & Li, C. (2009). *Predicted Returns and Sources of Momentum Profits* (No. 305).

Kapetanios, G. (2009). A Testing Procedure for Determining the Number of Factors in Approximate Factor Models With Large Datasets. *Journal of Business and Economic Statistics*, 091002081548023–13. Available from http://pubs.amstat.org/doi/abs/10.1198/jbes.2009.07239

Kneip, A., Sickles, R. C., & Song, W. (2009). A New Panel Data Treatment for Heterogeneity in Time Trends. , 1–51.

Lesmond, D. a. (1999, December). A new estimate of transaction costs. *Review of Financial Studies*, *12*(5), 1113–1141. Available from http://rfs.oupjournals.org/cgi/doi/10.1093/rfs/12.5.1113

Longstaff, F. A., Mithal, S., & Neis, E. (2004). *Corporate Yield Spreads: Default Risk or Liquidity? New Evidence from the Credit-Default Swap Market.*

Longstaff, F. A., & Schwartz, E. S. (1995). A Simple Approach to Valuing Risky Fixed and Floating Rate Debt. *The Journal of Finance*, *L*(3).

Merton, R. C. (1974, July). On the Pricing of Corporate Debt: The Risk Structure of Interest Rates. *The Journal of Finance*, *29*(32), 1013.

Moon, H., & Perron, B. (2004, September). Testing for a unit root in panels with dynamic factors. *Journal of Econometrics*, *122*(1), 81–126. Available from http://linkinghub.elsevier.com/retrieve/pii/S0304407603002707

Onatski, A. (2009). *Determining the Number of Factors from Empirical Distribution of Eigenvalues* (No. 212).

Pesaran, H. M. (2006). Estimation and Inference in Large Heterogeneous Panels with a Multifactor Error Structure. *Econometrica*, *74*(4), 967–1012.

Pesaran, H. M. (2007). A SIMPLE PANEL UNIT ROOT TEST IN THE PRESENCE OF CROSS-SECTION DEPENDENCE. *Journal of Applied Econometrics*, *312*, 265– 312.