Available online at www.sciencedirect.com
ScienceDirect
Journal of Empirical Finance 15 (2008) 583–612
Journal of EMPIRICAL FINANCE
www.elsevier.com/locate/jempfin

Firm heterogeneity and credit risk diversification ☆

Samuel G. Hanson [a], M. Hashem Pesaran [b,c], Til Schuermann [d,e,*,1]

[a] *Department of Economics, Harvard Business School, Harvard University, United States*
[b] *University of Cambridge, United Kingdom*
[c] *USC, United States*
[d] *Federal Reserve Bank of New York, United States*
[e] *Wharton Financial Institutions Center, United States*

Received 22 June 2006; received in revised form 29 November 2007; accepted 30 November 2007
Available online 8 December 2007

Abstract

This paper examines the impact of *neglected* heterogeneity on credit risk. We show that neglecting heterogeneity in firm returns and/or default thresholds leads to *under*estimation of expected losses (EL), and its effect on portfolio risk is ambiguous. Once EL is controlled for, the impact of neglecting parameter heterogeneity is complex and depends on the source and degree of heterogeneity. We show that ignoring differences in default thresholds results in overestimation of risk, while ignoring differences in return correlations yields ambiguous results. Our empirical application, designed to be typical and representative, combines both and shows that neglected heterogeneity results in *over*estimation of risk. Using a portfolio of U.S. firms we illustrate that heterogeneity in the default threshold or probability of default, measured for instance by a credit rating, is of first order importance in affecting the shape of the loss distribution: including ratings heterogeneity alone results in a 20% drop in loss volatility and a 40% drop in 99.9% VaR, the level to which the risk weights of the New Basel Accord are calibrated.

JEL classification: C33; G13; G21
Keywords: Risk management; Correlated defaults; Factor models; Portfolio choice

☆ We would like to thank Richard Cantor, Paul Embrechts, Michael Gordy, Rafael Repullo, Joshua Rosenberg, Jose Scheinkman, Zhenyu Wang, Alan White, an anonymous referee, the editor (Franz Palm), and especially David Lando, as well as participants at the 13th Annual Conference on Pacific Basin Finance, Economics, and Accounting at Rutgers University, June 2005; the BIS-ECB 4th Joint Central Bank Research Conference in Frankurt, November 2005; the NBER conference on Risks of Financial Institutions, Cambridge, MA, November 2005; the CIRANO-CIREQ Conference of Financial Econometrics, Montreal, Canada, May 2006; the Financial Econometrics Conference at University of York, UK, June 2006; the C.R.E.D.I.T. conference at the University of Venice, Venice, Italy, September 2006; and seminar participants at the Newton Institute, University of Cambridge, UMass Amherst, University of Pennsylvania, University of Montreal, Austrian Central Bank (OeNB), the ECB, and the Federal Reserve Bank of New York for helpful comments and suggestions. We are also grateful to Yue Chen, Ben Iverson, and Chris Metli for research assistance.
* Corresponding author.
E-mail addresses: shanson@fas.harvard.edu (S.G. Hanson), mhp1@cam.ac.uk (M.H. Pesaran), til.schuermann@ny.frb.org (T. Schuermann).
[1] Any views expressed represent those of the authors only and not necessarily those of the Federal Reserve Bank of New York or the Federal Reserve System.

0927-5398/$ - see front matter
doi:10.1016/j.jempfin.2007.11.002

1. Introduction

Practically all credit risk models to date owe an intellectual debt to the options based approach to firm default by Merton (1974).[2] It took more than a decade for the development of solutions to portfolio loss distributions (Vasicek 1987, 1991), and those solutions were obtained under strict homogeneity assumptions regarding the probability distribution of firms asset values and default thresholds. Yet clearly heterogeneity is important and has received considerable attention of late. Two notable examples are Gordy (2003) using a factor-based approach,[3] and Duffie et al. (2007) using a default intensity approach.

To take full account of firm heterogeneity in credit risk places great demands on the data. When firms are public and have traded securities such as stocks, bonds, or even credit default swaps (CDS), as well as third party assessments such as public credit ratings, there is great scope for allowing and accounting for heterogeneity. But this scenario is limited to a small minority of firms; indeed most loans in banks' portfolios are to privately held firms about which we (and the banks) know rather little. In that case one may be forced to settle for the credit portfolio solutions obtained under homogeneity. What then are the consequences of neglecting heterogeneity for the analysis of the loss distribution? What is the impact on expected loss (EL), on risk, whether measured by loss volatility, which we call unexpected loss (UL), or tail quantiles (value at risk, VaR), or the shape of the entire loss distribution? Moreover, which sources of heterogeneity are especially important? This is the focus of our paper, and to our knowledge we are the first to examine the impact of neglected heterogeneity on credit risk.

We consider both observed and unobserved types of heterogeneity. The former is relatively easy to deal with and does not pose any particular technical difficulties.[4] The latter (unobserved heterogeneity) is more difficult and will be the focus of our analysis. Note that parameter heterogeneity refers to differences in population values of the parameters across different firms and prevails even in the absence of estimation uncertainty.[5] We build on the work of Vasicek and Gordy and examine the consequences of incorrectly neglecting the heterogeneity of return correlations and default thresholds across firms for the analysis of loss distributions. The default threshold captures a variety of firm characteristics such as balance sheet structure, including leverage, and intangibles like the quality of management. This heterogeneity can be random – differences in factor loadings are purely random around a fixed mean – and/or the differences could be systematic — mean factor loadings could differ across industries but be randomly distributed around the industry mean, across firms within an industry.

Our theoretical set-up is quite general and imposes few distributional and parametric restrictions. The theoretical results show a complex interaction between the sources of heterogeneity and the resulting loss distribution. We find that neglecting heterogeneity in default thresholds and/or mean returns results in *under*estimation of expected losses, and its effect on portfolio risk is ambiguous. This is a new result and arises due to the nonlinear nature of the relationships that prevail between the return process, the default threshold and the resultant default (and hence loss) process. Differences in asset values and default thresholds across firms do not disappear by cross-section averaging even if the differences across firms are random and the underlying portfolio is sufficiently large.

In comparing heterogeneous loss portfolios it is therefore important that appropriate adjustments are made so that the different portfolios all have the same ELs. In general this can be achieved by allowing for systematic heterogeneity across firms, e.g. by grouping firms into industries, regions, distances to default (e.g. credit rating), or a combination of those. For the same level of EL, the impact of allowing for heterogeneity on the shape of the loss distribution is complex and depends on the source and degree of heterogeneity. We show that random differences in default thresholds result in lower risk, whether measured by UL or typical VaR levels (99%, 99.9%), so long as cross firm return correlations are homogeneous. However, in the more general case where cross firm return correlations are also heterogeneous, the effects of heterogeneity on unexpected loss and VaR levels are ambiguous. The net effect of combining heterogeneity in both sources is hard to predict systematically, but our empirical application, designed to be typical and representative, shows risk to be reduced – both UL and VaR decline – so that neglected heterogeneity results in overestimated risk. Therefore falsely imposing homogeneity could be quite costly. But as the 2007 market

[2] For a summary of models see Saunders and Allen (2002), and for detailed comparisons, see Koyluoglu and Hickman (1998) and Gordy (2000).

[3] Gordy's (2003) result shaped regulatory policy in the specific form of the regulatory capital formula in the New Basel Accord (BCBS 2005, §272).

[4] In our set up an important example of observed heterogeneity is given by credit ratings across firms. These might be either public credit ratings assigned by a rating agency or internal ratings used by a lender.

[5] In this paper we do not allow for parameter estimation uncertainty.

turmoil in structured credit has revealed, mixing heterogeneous assets like subprime mortgage securities with investment grade bonds need not reduce risk, and investors that neglect such heterogeneity do so at their peril.

Along the way we derive analytic solutions to loss distributions under parameter heterogeneity, assuming that the cross-section means and variance/covariances of the firm parameters are known; under homogeneity these variances and covariances are set to zero. Such derivations are important since they allow us to calibrate loss distributions for cases where there is little or no data to estimate the extent of parameter heterogeneity (which is practically the case for most of bank lending), using available estimates based on publicly traded securities. The latter estimates are not perfect and will be subject to errors, but are likely to be more appropriate than setting the variance and covariances to zero. This result marks our second contribution to the literature.

Our work has practical relevance for credit portfolio managers, especially when only limited data is available. The manager will likely have a reasonable grasp of average or expected loss of the portfolio, but might be less confident about the shape of the loss distribution. Yet this shape is critical for risk assessment and is influenced significantly by the heterogeneity of the underlying portfolio. We show how this heterogeneity can be easily parameterized and estimated for portfolios where data is plentiful and then applied to portfolios for which data is limited, as may be the case when the underlying assets are not publicly traded which is the case for the bulk of a bank's loan portfolio; one specific example may be the securitization of small business loans.

The importance of these theoretical insights is illustrated using a portfolio of about 600 publicly traded U.S. firms. Return regressions subject to different degrees of parameter heterogeneity are estimated recursively using six ten-year rolling estimation windows, and for each estimation window the loss distribution is then simulated out-of-sample over a one-year period. The predictions made by theory are confirmed in this application and are found to be robust across the six years. We show that heterogeneity in the default threshold or probability of default (PD), measured for instance by a credit rating, is of first order importance in affecting the shape of the loss distribution: allowing for ratings heterogeneity alone results in a 20% drop in loss volatility (keeping EL's constant) and 40% drop in 99.9% VaR, the level to which the risk weights in the New Basel Accord are calibrated. Allowing for additional heterogeneity results in a further 10% drop in 99.9% VaR. This result has important policy implications as a PD estimate through a credit rating, whether generated by a bank internally or provided by a rating agency externally, is the one parameter (of those considered here) that is allowed to vary in the New Basel Capital Accord.

To analyze the impact of neglected heterogeneity on credit risk, we use a simple multifactor approach which is easily adapted to this task. Multifactor models have been used extensively in finance following Ross (1976) and Chamberlain and Rothschild (1983).[6] Their application to credit risk has been more recent. A notable example is it's use in the CreditMetrics model as set out in Gupton et al. (1997). Gordy (2000) and Schönbucher (2003, ch. 10) provide useful reviews.

A separate line of research has focused on correlated default intensities as in Lando (1998), Schönbucher (1998), Duffie and Singleton (1999), Duffie and Gârleanu (2001), Collin-Dufresne et al. (2004), and Duffie et al. (2007); with a review by Duffie (2005). There are also a host of other approaches, including correlated (but non-systematic) jumps-at-default (Driessen 2005; Jarrow et al., 2005), the contagion model of Davis and Lo (2001) as well as Giesecke and Weber's (2004) indirect dependence approach, where default correlation is introduced through local interaction of firms with their business partners as well as via global dependence on economic risk factors. The idea of generalizing default dependence beyond correlation using copulas is discussed in Li (2000), Embrechts et al. (2001), Schönbucher (2002), Frey and McNeil (2003), and Hull and White (2004). Indeed our approach maps readily to a Gaussian copula approach.[7] The Hull and White paper provides an illustration of the impact of correlation and ratings heterogeneity on an *n*th-to-default CDS.

In short, the literature on modeling default dependence is growing rapidly along different paths, and there is as yet no consensus which approach is best. Our paper does not address that issue, but it does highlight, using a factor approach, the impact of neglected heterogeneity. This issue of neglected heterogeneity clearly also arises in the case of other approaches that focus on correlated default intensities or copulas; we leave that for others to explore. The factor structure considered here does allow us to explore two distinct channels of heterogeneity: one that is shared, namely

[6] Connor and Korajczyk (1995) provide an excellent survey.

[7] Our approach also readily accommodates non-Gaussian distributions for the error processes. See Appendix A. The case of Student-t distributed errors and hence t-copulas, is discussed in an earlier version of this paper, available at http://fic.wharton.upenn.edu/fic/papers/05/p0505.htm.

factor sensitivities, and one which is specific to firms within a given grouping (e.g. credit rating), namely the default threshold or the distance to default.

Our results have implications for risk and capital management as well as for pricing of credit assets. For example, in the case of a commercial bank, ignoring heterogeneity may result in under-provisioning for loan losses since EL is underestimated, and may result in overcapitalization against (bank) default since risk is overestimated. The risk assessment and pricing of complex credit assets such as collateralized debt obligations (CDOs) may be adversely affected since they are driven by the shape of the loss distribution which is then segmented into tranches. In particular, ignoring heterogeneity would result in excessive subordination for senior tranches if risk is overestimated.

The plan for the remainder of the paper is as follows: Section 2 introduces the basic model of firm value and default and considers the problem of correlated defaults. Section 3 derives the portfolio loss distribution under different heterogeneity assumptions, starting with the simple case of a homogeneous portfolio as introduced by Vasicek. These results are illustrated in Section 4 where we explore the impact of heterogeneity using returns for a large sample of publicly traded firms in the U.S. across seven sectors, and we analyze the resulting loss distributions obtained by stochastic simulations. Section 5 provides some concluding remarks. A technical Appendix presents generalizations of some material in Sections 2 and 3.

2. Firm value, default and default dependence

Much of the research on credit risk modeling, including our own, is based on the option theoretic default model of Merton (1974). Merton recognized that a lender is effectively writing a put option on the assets of the borrowing firm; owners and owner-managers (i.e. shareholders) hold the call option. Thus a firm is expected to default when the value of its assets falls below a threshold value determined by its liabilities.[8]

2.1. Firm value and default

Consider a firm i having asset value V_{it} at time t, and an outstanding stock of debt, D_{it}. Under the Merton model default occurs at the maturity date of the debt, $t+h$, if the firm's assets, $V_{i,t+h}$, are less than the face value of the debt at that time, $D_{i,t+h}$. The value of the firm at time t is the sum of debt and equity, namely

$$V_{it} = D_{it} + E_{it}, \quad \text{with} \quad D_{it} > 0. \tag{1}$$

Conditional on time t information, default will take place at time $t+h$ if $V_{i,t+h} \leq D_{i,t+h}$. In the Merton model debt is assumed to be fixed over the horizon h. For simplicity we set $h=1$; extensions to multiple periods can be found in Pesaran et al. (2006), hereafter PSTW. Because default is costly and violations to the absolute priority rule in bankruptcy proceedings are common, in practice debtholders have an incentive to put the firm into receivership even before the equity value of the firm hits the zero value.[9] Importantly, default in a credit relationship is typically a weaker condition than outright bankruptcy. An obligor may meet the technical default condition, e.g. a missed coupon payment, without subsequently going into bankruptcy. As a result we shall assume that default takes place if

$$0 < E_{i,t+1} < C_{i,t+1}, \tag{2}$$

where $C_{i,t+1}$ is a positive default threshold which could vary over time and with the firm's characteristics (such as region or industry sector). Natural candidates that affect the default threshold include observable factors such as leverage, profitability, and firm age (most of which appear in models of firm default), as well as non-observable ones such as management quality.[10]

[8] An alternative to Merton's end of period approach are the first-passage models where default would occur the first time that firm value falls below a default boundary (or threshold) over the period, as in Zhou (2001).

[9] See, for instance, Leland and Toft (1996) who develop a model where default is determined endogenously, rather than by the imposition of a positive net worth condition. More recently, Broadie et al. (2007) show that default can be optimal when $E_{it}>0$ even in the case of a single debt class.

[10] For models of bankruptcy and default at the firm level, see, for instance, Altman (1968), Lennox (1999), Shumway (2001), and Hillegeist et al. (2004).

We are now in a position to consider the evolution of firm equity value which we assume follows a standard geometric random walk model:

$$\ln\left(E_{i,t+1}\right) = \ln(E_{it}) + \mu_i + \xi_{i,t+1}, \quad \xi_{i,t+1} \sim \mathrm{iid}\left(0, \sigma^2_{\xi_i}\right), \tag{3}$$

with a non-zero drift, μ_i, and innovations with a zero mean and firm-specific volatility, $\sigma_{\xi i}$. Typically, it is assumed that the innovations, $\xi_{i,t+1}$, are Gaussian, but as we shall see below our analysis readily carries over to non-Gaussian innovations. Consequently, default occurs if

$$\ln\left(E_{i,t+1}\right) = \ln\left(E_{i,t}\right) + \mu_i + \xi_{i,t+1} < \ln\left(C_{i,t+1}\right), \tag{4}$$

or if the one-period change in equity value or return falls below some threshold defined by

$$\ln\left(\frac{E_{i,t+1}}{E_{it}}\right) < \ln\left(\frac{C_{i,t+1}}{E_{it}}\right) = \lambda_{i,t+1}. \tag{5}$$

We refer to $\lambda_{i,t+1}$ as the "default return threshold," to distinguish it from the default threshold, $C_{i,t+1}$. Eq. (5) tells us that the relative (rather than absolute) decline in firm value must be large enough over the period to result in default. Note that firm-specific information such as leverage and management quality, embedded in the default threshold $C_{i,t+1}$, carry over to $\lambda_{i,t+1}$. Thus for highly levered firms or firms with poor management, the return threshold is higher (less negative) than for well capitalized or well managed firms, so that a smaller negative return will trigger default.

The default trigger condition, Eq. (5), allows firms with the same default thresholds, $C_{i,t+1}$, to have different default probabilities depending on their initial net worth, E_{i0}. Specifically, firms whose initial net worth is closer to the default threshold, $C_{i,t+1}$, are more likely to default. Under the assumption of Gaussian innovations in Eq. (3), the probability that firm i defaults at the end of the period is given by

$$\pi_{i,t+1} = \Phi\left(\frac{\lambda_{i,t+1} - \mu_i}{\sigma_{\xi_i}}\right), \tag{6}$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution function. Equivalently, solving the equity values in terms of their initial values we have

$$\pi_{i,t+1} = \Phi\left(\frac{\ln\left(C_{i,t+1}/E_{i0}\right) - (t+1)\mu_i}{\sqrt{t+1}\,\sigma_{\xi_i}}\right). \tag{7}$$

It is clear that firms with the same return distribution and default threshold will have different default probabilities if they have different initial equity values. In the theoretical discussions that follows we focus on the first formulation, Eq. (6), and assume that the firm-specific default return thresholds, $\lambda_{i,t+1}$, are given.[11] This formulation also has the advantage that it does not depend on often unobserved initial equity values, and yields identical default probabilities for identical firms with the same default return thresholds, irrespective of their initial equity values.

2.2. Cross firm dependence

Cross firm default dependence can be introduced by assuming that shocks to the value of a firm's equity, $\xi_{i,t+1}$, defined by Eq. (3), have the following multifactor structure:[12]

$$\xi_{i,t+1} = \boldsymbol{\gamma}_i' \mathbf{f}_{t+1} + \sigma_i \varepsilon_{i,t+1}, \quad \varepsilon_{i,t+1} | \mathbf{f}_{t+1} \sim \mathrm{iid} F_\varepsilon(0,1) \tag{8}$$

where $\mathbf{f}_{t+1}$ is an $m\times 1$ vector of common factors, γ_i is the associated vector of factor loadings, and $\varepsilon_{i,t+1}$ is the firm-specific idiosyncratic shock, assumed to be distributed independently across i with mean zero and unit variance; in this

[11] The problem of how to identify and estimate $\lambda_{i,t+1}$ is addressed in Appendix C.

[12] We consider a simple linear model, though nonlinear factor models with the possibility of jumps are also possible.

way the model in Eq. (8) is said to be conditionally independent.[13] The common factors could be treated as either unobserved or observed through macroeconomic variables such as output growth, inflation, interest rates and exchange rates.[14] The cumulative distribution function of $\varepsilon_{i,t+1}$ is denoted by $F_\varepsilon(\cdot)$.

In what follows we suppose the factors are unobserved, distributed independently of $\varepsilon_{i,t+1}$, and have constant variances. Here we assume that $\mathbf{f}_{t+1}\sim \mathrm{iid}F_f(\mathbf{0}, \mathbf{I}_m)$, where $\mathbf{I}_m$ is an identity matrix of order m. For a set of arbitrary fixed factor loadings the variance covariance matrix of $\mathbf{f}_{t+1}$ can always be normalized to $\mathbf{I}_m$ without loss of generality.[15]

Using Eq. (8) in (3) we now have

$$\ln\left(E_{i,t+1}\right) - \ln(E_{it}) = r_{i,t+1} = \mu_i + \gamma_i' \ \mathbf{f}_{t+1} + \sigma_i\varepsilon_{i,t+1}. \tag{9}$$

Under the above assumptions

$$\sigma^2_{\xi_i} = \gamma_i'\gamma_i + \sigma_i^2, \tag{10}$$

which decomposes the return variance into the part due the systematic risk factors, $\gamma_i' \ \gamma_i$, and the residual or idiosyncratic variance, σ_i^2.

The reader is likely familiar with two special cases of Eq. (9). The first is the one-factor Gaussian CAPM

$$r_{i,t+1} = \alpha_i + \beta_i r_{t+1} + \sigma_i\varepsilon_{i,t+1}, \quad \varepsilon_{i,t+1}|r_t \sim \mathrm{iid}N(0,1), \quad r_{t+1} \sim \mathrm{iid}N\left(\bar{r}, \sigma_{\bar{r}}^2\right), \tag{11}$$

where r_{t+1} is the return on the market factor so that the expected return for firm i is $\mu_i=\alpha_i+\beta_i\bar{r}$, and $\gamma_i=\beta_i\sigma_{\bar{r}}$. The second is the one-factor Vasicek model which imposes yet further restrictions:

$$r_{\mathrm{i,t+1}} = \sqrt{\rho} f_{t+1} + \sqrt{1-\rho}\ \varepsilon_{\mathrm{i,t+1}}, \tag{12}$$

where both f_{t+1} and $\varepsilon_{i,t+1}$ are iid standard normal, $\mu_i = 0, \gamma_i = \sqrt{\rho}$, and $\sigma_i = \sqrt{1-\rho}$.

The presence of the common factors also introduces a varying degree of asset return correlations across firms, which in turn leads to variation in cross firm default correlations for a given set of default return thresholds, $\lambda_{i,t+1}$. The extent of default correlation depends on the size of the factor loadings, γ_i, the importance of the idiosyncratic shocks, σ_i, the values of the default return thresholds, $\lambda_{i,t+1}$, and the shape of the distribution assumed for $\varepsilon_{i,t+1}$, particularly its left tail properties. The correlation coefficient of returns of firms i and j is given by

$$\rho_{ij} = \frac{\delta_i'\ \delta_j}{\left(1+\delta_i'\ \delta_i\right)^{1/2}\left(1+\delta_j'\ \delta_j\right)^{1/2}}, \tag{13}$$

where $\delta_i=\gamma_i/\sigma_i$ is the standardized $m\times 1$ vector of factor loadings (systematic risk exposures) of firm i. Note that for the Vasicek model given by Eq. (12), correlation of returns across all firms is the same, namely ρ.

To derive the firm-specific probability of default, we first define an indicator variable $z_{i,t+1}$ to be the default outcome for firm i over a single period such that

$$z_{\mathrm{i,t+1}} = I\left(\lambda_{\mathrm{i,t+1}} - r_{\mathrm{i,t+1}}\right), \tag{14}$$

where $I(A)$ is an indicator function that takes the value of unity if $A\geq 0$, and zero otherwise. Then, if conditional on $\mathbf{f}_{t+1}$, $\varepsilon_{i,t+1}$ and $\varepsilon_{j,t+1}$ are cross-sectionally independent, and $\mathbf{f}_{t+1}$ and $\varepsilon_{i,t+1}$ have a joint Gaussian distribution, we have

$$\pi_{i,t+1} = \Phi\left(\frac{\lambda_{i,t+1}-\mu_i}{\sqrt{\sigma_i^2+\gamma_i'\gamma_i}}\right), \tag{15}$$

which is commonly referred to as *conditionally independent double-Gaussian* model.

[13] Note that conditional independence may not necessarily be attained in an empirical setting (see, for instance, Das et al., 2007), a point we discuss in more detail in Section 4.

[14] For instance, PSTW provide an empirical implementation of this model by linking the (observable) factors, $\mathbf{f}_{t+1}$, to the variables in a global vector autoregressive model comprising around 80% of world output.

[15] The more general case where the factors may exhibit time varying volatility can be readily dealt with by allowing the factor loadings to vary over time, in line with market volatilities. But in this paper we shall not pursue this line of research, primarily because the focus of our empirical analysis is on quarterly and annual default risks, and over such horizons asset return volatility dynamics tend to be rather weak and of second order importance.

3. Losses in a credit portfolio

Consider now a credit portfolio composed of N different credit assets such as loans, each with exposures or weights w_{it}, at time t, for $i=1, 2,\ldots, N$, such that[16]

$$\sum_{i=1}^{N} w_{it} = 1, \quad \sum_{i=1}^{N} w_{it}^2 = O\left(N^{-1}\right), w_{it} \geq 0. \tag{16}$$

A sufficient condition for Eq. (16) to hold is given by $w_{it}=O\,(N^{-1})$, which is the standard granularity condition where no single exposure dominates the portfolio.[17] To simplify the exposition, we impose both here and later in the empirical section that a defaulted asset has no recovery value.[18] Under this set-up the portfolio loss over the period t to $t+1$ is given by

$$\ell_{N,t+1} = \sum_{i=1}^{N} w_{it} z_{i,t+1}. \tag{17}$$

The probability distribution function of $\ell_{N,t+1}$ can now be derived both conditional on an information set available at time t, $\mathcal{I}_t$, or unconditionally. The two types of distributions coincide when the factors, $\mathbf{f}_{t+1}$, are assumed to be serially independent, a case often maintained in the literature.[19]

The focus of our analysis will be on the limit distribution of $\ell_{N,t+1}|\mathcal{I}_t$, as $N\to\infty$. Not surprisingly, this limit distribution depends on the nature of the return process $\{r_{i,t+1}\}$ and the extent to which the returns are cross-sectionally correlated. Our theoretical discussion shall be in terms of the variance of the loss distribution, though occasionally we refer to the standard deviation or loss volatility, known as unexpected loss (UL) in the credit risk literature. In practice, one may also be interested in quantiles of the loss distributions, or VaR, and those can be easily obtained through stochastic simulations, as we show in Section 4.

3.1. Credit risk under firm homogeneity

Vasicek (1987) was one of the first to consider the limit distribution of $\ell_{N,t+1}$ using asset return equations with a factor structure. However, he focused on the perfectly homogeneous case described by Eq. (12) with the same factor loadings, $\gamma_i=\gamma$, the same default return thresholds, $\lambda_{i,t+1}=\lambda$, the same firm-specific volatilities, $\sigma_i=1$, and zero unconditional returns, $\mu_i=0$, for all i and t. Note that a multifactor model with homogeneous factor loadings is equivalent to a single factor model.

In this model the pair-wise asset return correlations, ρ_{ij}, is identical for all obligor pairs in the portfolio. The remaining parameter, λ, is then calibrated to a pre-specified default probability, π, so that the distance to default and default return thresholds are the same for all firms and can be easily estimated from historical default frequency of the portfolio using

$$\lambda = \Phi^{-1}(\pi). \tag{18}$$

Under the Vasicek model portfolio loss variance depends on π and the default correlation ρ^*:[20]

$$\mathrm{Var}\left(\ell_{N,t+1}|\mathcal{I}_t\right) = \pi(1-\pi)\left\{\rho^* + (1-\rho^*)\sum_{j=1}^{N} w_{jt}^2\right\}. \tag{19}$$

[16] The assumption that N is time-invariant is made for simplicity and can be relaxed.

[17] Conditions (16) on the portfolio weights was in fact embodied in the initial proposal of the New Basel Accord in the form of the Granularity Adjustment which was designed to mitigate the effects of significant single-borrower concentrations on the credit loss distribution (BCBS, 2001, Section 8). See also the discussion in Lucas et al. (2001) and Gordy (2004).

[18] The case where default and recovery are correlated through common business cycle effects presents new technical difficulties and is addressed briefly in Appendix A of an earlier version of this paper, available at http://fic.wharton.upenn.edu/fic/papers/05/p0505.html.

[19] However, this assumption precludes the use of any business cycle models in the analysis of credit risk. Our results hold so long as $\mathbf{f}_{t+1}$ follows a covariance stationary process, and $\mathcal{I}_t$ contains at least $\mathbf{f}_t$ and its lagged values, or their determinants when they are unobserved. That structure corresponds to the empirical application in PSTW which makes use of the GVAR, a global macroeconometric model.

[20] $\rho^*(\pi,\rho) = \frac{\Phi_2\left[\Phi^{-1}(\pi),\Phi^{-1}(\pi),\rho\right]-\pi^2}{\pi(1-\pi)}$, where $\Phi_2[\cdot]$ is the standard bivariate normal cdf. See Zhou (2001).

Under the granularity condition, (16), the second term in brackets becomes negligible when N is sufficiently large. Hence, in the limit

$$\lim_{N\to\infty} \quad \mathrm{Var}\,(\ell_{N,t+1}|\mathcal{I}_t) = \pi(1-\pi)\rho^* = \Phi_2[\Phi^{-1}(\pi), \Phi^{-1}(\pi), \rho] - \pi^2, \tag{20}$$

where $\Phi_2\,[\cdot]$ is the standard bivariate normal cdf.

Vasicek's credit loss limit distribution is fully determined by two parameters, namely the average default probability, π, and the pair-wise return correlation coefficient, ρ (see Appendix A.2 for further detail). The former fixes the expected loss of the portfolio, while the latter controls the shape of the loss distribution. In effect one parameter, ρ, controls all aspects relating to the shape of the loss distribution: its volatility, skewness and kurtosis, and so on.

3.2. Credit risk with firm heterogeneity

Building on Vasicek's work we now consider models that allow for firm heterogeneity across a number of relevant parameters. In this section we provide some analytical derivations and show how the theoretical work of Vasicek can be generalized. An empirical evaluation of the importance of allowing for firm heterogeneity in credit risk analysis is discussed in Section 4.

Under the heterogeneous multifactor return process (Eq. (9)), the portfolio loss, $\ell_{N,t+1}$, can be written as

$$\ell_{N,t+1} = \sum_{i=1}^{N} w_{it} I\left(a_{i,t+1} - \boldsymbol{\delta}_i' \mathbf{f}_{t+1} - \varepsilon_{i,t+1}\right), \tag{21}$$

where

$$a_{i,t+1} = \frac{\lambda_{i,t+1} - \mu_i}{\sigma_i}, \quad \boldsymbol{\delta}_i = \frac{\boldsymbol{\gamma}_i}{\sigma_i} \tag{22}$$

are the normalized distance to default and factor loading, respectively. In addition, to allowing for parameter heterogeneity, we also relax the assumption that conditional on $\mathcal{I}_t$ the common factors, $\mathbf{f}_{t+1}$, and the idiosyncratic shocks, $\varepsilon_{i,t+1}$, are normally distributed with zero means. Accordingly we assume that

$$\begin{aligned} &\varepsilon_{i,t+1}|\mathcal{I}_t \sim \mathrm{iid}F_\varepsilon(0,1), \quad \text{for all } i \text{ and } t, \\ &\mathbf{f}_{t+1}|\mathcal{I}_t \sim \mathrm{iid}F_f\left(\boldsymbol{\mu}_{ft}, \mathbf{I}_m\right), \quad \text{for all } t. \end{aligned}$$

Allowing $\boldsymbol{\mu}_{ft}$ to be time-varying enables us to explicitly consider the possible effects of business cycle variations on the loss distribution. In the credit risk literature $\boldsymbol{\mu}_{ft}$ is usually set to zero. For future use we shall denote the $\mathcal{I}_t$-conditional probability density and the cumulative distribution functions of $\varepsilon_{i,t+1}$ and $\mathbf{f}_{t+1}$, by $f_\varepsilon(\cdot)$ and $F_\varepsilon(\cdot)$, and $f_{\mathbf{f}}(\cdot)$ and $F_{\mathbf{f}}(\cdot)$, respectively. See Appendix A for further details.

To deal with parameter heterogeneity across firms we adopt the following random coefficient specification:

$$\boldsymbol{\theta}_i = \boldsymbol{\theta} + \mathbf{v}_i, \quad \mathbf{v}_i \sim \mathrm{iid}(\mathbf{0}, \Omega_{vv}), \quad \text{for } i = 1, 2, \ldots, N, \tag{23}$$

where

$$\boldsymbol{\theta}_i = (a_i, \boldsymbol{\delta}_i')', \quad \boldsymbol{\theta} = (a, \boldsymbol{\delta}')', \quad \mathbf{v}_i = \left(v_{ia,}\mathbf{v}_{i\delta}'\right)', \tag{24}$$

$$\boldsymbol{\Omega}_{vv} = \begin{pmatrix} \omega_{aa} & \boldsymbol{\omega}_{a\delta} \\ \boldsymbol{\omega}_{\delta a} & \boldsymbol{\Omega}_{\delta\delta} \end{pmatrix}, \tag{25}$$

is a positive semi-definite symmetric matrix, and $\mathbf{v}_i$'s are distributed independently of ($\varepsilon_{j,t+1}$, $\mathbf{f}_{t+1}$) for all i, j and t. Variation in parameters over time can be dealt with by allowing the mean parameter vector, $\boldsymbol{\theta}$, and the covariance matrix, $\boldsymbol{\Omega}_{vv}$, to change each period.

Allowing for such parameter heterogeneity may be desirable when firms have different sensitivities to the systematic risk factors $\mathbf{f}_{t+1}$, and those sensitivities or factor loadings are known only up to their distributional properties described in Eq. (23). For instance, a firm's market beta on average should be around one, but any given firm's beta may be larger or smaller than one. An example of an application might be assessing the credit risk for a portfolio of borrowers which are privately held, i.e. not publicly traded. This is typically the case for much of middle market and most of small business lending. For such firms it would be very difficult or even impossible to obtain individual estimates of $\boldsymbol{\theta}_i$, and an average estimate based on $\boldsymbol{\theta}$ and $\boldsymbol{\Omega}_{vv}$ might be the only feasible option. See Section 4.5 for further elaboration and an empirical example.

The heterogeneity described in Eqs. (23)–(25) states that firm differences are purely random. However, firms could in addition exhibit systematic parameter differences, say by industry and/or region, so that parameter means and covariances are also industry and/or region specific. For instance, some sectors, say information technology (IT), may have on average high market betas while other sectors, say consumer staples, may have on average lower betas. The dispersion around their sector means may also differ sector by sector. This generalization is taken up in Section 3.4.2.

3.3. Unexpected loss under parameter heterogeneity

Before exploring the entire loss distribution, for reasons of analytical tractability we focus here on loss variance, Var $(\ell_{N,t+1}|\mathcal{I}_t)$, or its square root, unexpected loss. Since the $z_{i,t+1}$ are conditionally independent, under the granularity condition (16),

$$\lim_{N\to\infty} \operatorname{Var}\left(\ell_{N,t+1}|\mathcal{I}_t\right) = \lim_{N\to\infty} \left\{\operatorname{Var}\left[E\left(\ell_{N,t+1}|\mathbf{f}_{t+1},\mathcal{I}_t\right)\right]\right\}, \tag{26}$$

which follows from Proposition 2 in Gordy (2003). So long as the portfolio weights satisfy the granularity condition (16), the limit behavior of the unexpected loss does not depend on the portfolio weights w_{it}. Furthermore, this result holds irrespective of whether a_i and $\boldsymbol{\delta}_i$ are homogeneous or vary randomly across i.

Under the random coefficient model (23), asymptotic loss variance, given by (26), can be obtained by integrating out the heterogeneous effects of a_i and $\boldsymbol{\delta}_i$. First note that $\ell_{N,t+1} = \Sigma_{i=1}^N w_{it} I\left(a_i - \boldsymbol{\delta}_i' \mathbf{f}_{t+1} - \varepsilon_{i,t+1}\right)$, which under Eq. (23) can be written as

$$\ell_{N,t+1} = \sum_{i=1}^{N} w_{it} I\left(a - \boldsymbol{\delta}' \mathbf{f}_{t+1} - \zeta_{i,t+1}\right), \tag{27}$$

where

$$\zeta_{i,t+1} = \varepsilon_{i,t+1} - \mathbf{v}_i' \mathbf{g}_{t+1} \tag{28}$$

captures all innovations, and $\mathbf{g}_{t+1}=(1, -\mathbf{f}_{t+1}')'$. Conditional on $\mathbf{f}_{t+1}$, $\zeta_{i,t+1}$ is distributed independently across i with zero mean and variance

$$\omega_{t+1}^2 = 1 + g_{t+1}' \boldsymbol{\Omega}_{vv} \mathbf{g}_{t+1}, \tag{29}$$

where $\mathbf{g}_{t+1}' \boldsymbol{\Omega}_{vv} \mathbf{g}_{t+1}$ is the variance contribution arising from the random coefficients model (i.e. explicitly due to parameter heterogeneity). The expected loss conditional on the factors $\mathbf{f}_{t+1}$ is given by

$$E\left(\ell_{N,t+1}|\mathbf{f}_{t+1},\mathcal{I}_t\right) = \sum_{i=1}^{N} w_{it} Pr\left(\zeta_{i,t+1} \le a - \boldsymbol{\delta}'\mathbf{f}_{t+1}|\mathbf{f}_{t+1},\mathcal{I}_t\right) = \sum_{i=1}^{N} w_{it} F_{\varkappa}\left(\frac{\boldsymbol{\theta}'\mathbf{g}_{t+1}}{\omega_{t+1}}\right),$$

where $F_{\varkappa}(\cdot)$ is the cumulative distribution function of the standardized composite innovations

$$K_{i,t+1} = \frac{\zeta_{i,t+1}}{\omega_{t+1}} |\mathbf{f}_{t+1},\mathcal{I}_t \sim \text{iid}(0,1), \tag{30}$$

and since $\sum_{i=1}^{N} w_{it}=1$, then

$$E\left(\ell_{N,t+1}|\mathbf{f}_{t+1},\mathcal{I}_t\right) = F_{\varkappa}\left(\frac{\boldsymbol{\theta}'\mathbf{g}_{t+1}}{\omega_{t+1}}\right). \tag{31}$$

Using this loss distribution it is now possible to compute EL and VaR, as well as other risk measures with stochastic simulation by taking independent draws from any assumed distribution of $K_{i,t+1}$. In some cases we are able to analytically study the implications of parameter heterogeneity for the loss distribution, e.g. when heterogeneity is limited to mean returns and/or default return thresholds, or to the factor loadings. Those cases are taken up in Section 3.4.

Finally, it is also worth noting that in the limit (as $N\rightarrow\infty$), using Eq. (26) we have

$$\lim_{N\rightarrow\infty} \mathrm{Var}\left(\ell_{N,t+1}|\mathcal{I}_t\right) = \mathrm{Var}\left[F_{\varkappa}\left(\frac{\boldsymbol{\theta}'\mathbf{g}_{t+1}}{\omega_{t+1}}\right)|\mathcal{I}_t\right], \tag{32}$$

which does not depend on the exposure weights, w_{it}. This result represents a generalization of the limit variance obtained for the homogeneous case, given above by Eq. (20).

3.4. Impact of neglected heterogeneity

Parameter heterogeneity can significantly affect the shape of the loss distribution as well as expected and unexpected losses. This is most easily illustrated with a single factor model. Multifactor generalizations are given in Appendix A. As before, portfolio losses are given by (replacing w_{it} with w_i to simplify the notation)

$$\ell_{N,t+1} = \sum_{i=1}^{N} w_i I\left(a - \delta f_{t+1} - \zeta_{i,t+1}\right),$$

where $a=(\lambda-\mu)/\sigma$, $\delta=\gamma/\sigma$, are the normalized distance to default and factor loading, respectively, and $\zeta_{i,t+1}=\varepsilon_{i,t+1}-v_{ia}+v_{i\delta}f_{t+1}$, is the composite innovation. In the absence of heterogeneity, δ and a can be written in terms of the return correlation, ρ, and the default probability, π:

$$\delta = \sqrt{\frac{\rho}{1-\rho}}, \quad \text{for } \rho > 0, \tag{33}$$

and

$$a = \frac{\Phi^{-1}(\pi)}{\sqrt{1-\rho}} < 0, \quad \text{for } \pi < 1/2, \tag{34}$$

which yields the following useful relationship between a, δ and π:

$$a = \sqrt{1+\delta^2}\,\Phi^{-1}(\pi). \tag{35}$$

Therefore, for a given value of $\pi<1/2$, a and δ are negatively related and can not vary freely of one another.[21]

Under the conditionally independent normal assumption,

$$\begin{pmatrix} \varepsilon_{i,t+1} \\ v_{ia} \\ v_{i\delta} \end{pmatrix} |f_t \sim \mathrm{iid}\, N\left(\mathbf{0}, \begin{bmatrix} 1 & 0 & 0 \\ 0 & \omega_{aa} & \omega_{a\delta} \\ 0 & \omega_{a\delta} & \omega_{\delta\delta} \end{bmatrix}\right),$$

so that $\zeta_{i,t+1}|f_t\sim\mathrm{iid}\, N(0,\, 1+\omega_{aa}+\omega_{\delta\delta}f_t^2-2\omega_{a\delta}f_t)$, and hence as $N\rightarrow\infty$ the loss distribution can be simulated using[22]

$$x(f) = \Phi\left(\frac{a-\delta f}{\sqrt{1+\omega_{aa}+\omega_{\delta\delta}f^2-2\omega_{a\delta}f}}\right), \tag{36}$$

for random draws of $f\sim\mathrm{iid}N\,(0, 1)$. Note that the asymptotic loss distribution is given by the distribution of x (the fraction of the portfolio lost) over (0, 1]. Eq. (36) is a key expression which we use below to analyze the impact of heterogeneity (or its neglect) on the loss distribution, especially its tail.

3.4.1. Heterogeneity in mean returns and/or default return thresholds

Consider first the case where the standardized factor loading is the same for all firms, namely $\delta_i=\delta$, $\forall i$, but we allow for differences in $a_i=(\lambda_i-\mu_i)/\sigma_i$. Fixing δ also imposes $\sigma_i^2=\sigma^2$, $\forall i$, and implies the same pair-wise return correlation,

[21] The condition $\pi<1/2$ covers most real world cases. For instance, even a $\mathcal{CCC}$-rated firm has a one-year expected default rate of about 0.35.

[22] Here to simplify the exposition we have denoted the limit of $\ell_{N,t+1}$ by x, and have abstracted from the subscript t since f_t is serially uncorrelated.

ρ, across all firms. As a result, any variation in a_i is due to cross firm variation in $\lambda_i-\mu_i$, the difference between the default return threshold and the mean return. It is unlikely that one would see differences in firm thresholds, perhaps due to management quality, but not in expected returns, so that variation in λ_i will likely be accompanied by variation in μ_i.[23] In this case portfolio losses are given by

$$\ell_N = \sum_{i=1}^{N} w_i I(a_i - \delta f - \varepsilon_i). \tag{37}$$

Under $\varepsilon_i|f \sim \text{iid}N(0,1)$, and as $N\to\infty$, we have

$$\ell_N \to x = \Phi\left(\tilde{a} - \tilde{\delta} f\right),$$

where

$$\tilde{a} = \frac{a}{\sqrt{1+\omega_{aa}}}, \quad \tilde{\delta} = \frac{\delta}{\sqrt{1+\omega_{aa}}}. \tag{38}$$

In this case the CDF of x would have the same form as Vasicek's loss distribution, namely[24]

$$F_\ell(x) = \Phi\left(\frac{\Phi^{-1}(x) - \tilde{a}}{\tilde{\delta}}\right), = \Phi\left(\frac{\sqrt{1+\omega_{aa}(1-\rho)}\,\Phi^{-1}(x) - \Phi^{-1}(\pi)}{\sqrt{\rho}}\right), \tag{39}$$

where the second expression makes use of Eqs. (34), (33) and (38). This clearly reduces to the CDF of the Vasicek's model for $\omega_{aa}=0$. See Eq. (A.4) in Appendix A.2.

It is also easily seen that in this case EL for the heterogeneous portfolio, denoted $\tilde{\pi}$, is given by

$$\tilde{\pi} = E(x) = \Phi\left(\frac{\tilde{a}}{\sqrt{1+\tilde{\delta}^2}}\right) = \Phi\left(\frac{a}{\sqrt{1+\omega_{aa}+\delta^2}}\right) = \Phi\left(\frac{\Phi^{-1}(\pi)}{\sqrt{1+(1-\rho)\omega_{aa}}}\right), \tag{40}$$

which differs from the EL of the homogeneous portfolio. Since we are interested in losses for values of $\pi<1/2$, for which $\Phi^{-1}(\pi)<0$, an increase in ω_{aa} raises $\tilde{\pi}$

$$\frac{\partial\tilde{\pi}}{\partial\omega_{aa}} = \phi\left(\frac{\Phi^{-1}(\pi)}{\sqrt{1+(1-\rho)\omega_{aa}}}\right)\frac{-\Phi^{-1}(\pi)(1-\rho)}{2(1+(1-\rho)\omega_{aa})^{3/2}} \geq 0, \quad \text{for } \pi < 1/2.$$

It readily follows that as $N\to\infty$, EL is underestimated as $\tilde{\pi} \geq \pi$ when $\omega_{aa} > 0$ and this source of heterogeneity is neglected.

To derive the impact of ω_{aa} on unexpected loss, first without holding EL fixed, note that the pair-wise correlation of asset returns in this case is given by

$$\tilde{\rho} = \frac{\delta^2}{1+\delta^2+\omega_{aa}} = \frac{\rho}{1+\omega_{aa}(1-\rho)},$$

so that $\tilde{\rho}<\rho$, and using the results in Section 3.1 for N sufficiently large we have

$$\text{Var}(x) = \tilde{\pi}(1-\tilde{\pi})\rho^*(\tilde{\pi},\tilde{\rho}),$$

where the default correlation

$$\rho^*(\tilde{\pi},\tilde{\rho}) = \frac{\Phi_2\left[\Phi^{-1}(\tilde{\pi}), \Phi^{-1}(\tilde{\pi}),\tilde{\rho}\right] - \tilde{\pi}^2}{\tilde{\pi}(1-\tilde{\pi})}, \tag{41}$$

[23] Asset pricing theory implies that variation in expected returns across firms are driven entirely by variation in betas or factor loadings. Thus on average we would expect firm alphas to be zero, but any given firm may have a somewhat positive or negative alpha, implying some variation in (firm-specific) expected returns.

[24] See Appendix A.2.

and as before, $\Phi_2[\cdot]$ is the bivariate standard normal cdf. Thus, if we allow EL to vary, the effect of ω_{aa} on loss variance is

$$\frac{\partial \mathrm{Var}(x)}{\partial \omega_{aa}} = \frac{\partial \mathrm{Var}(x)}{\partial \tilde{\pi}} \times \frac{\partial \tilde{\pi}}{\partial \omega_{aa}} + \frac{\partial \mathrm{Var}(x)}{\partial \tilde{\rho}} \times \frac{\partial \tilde{\rho}}{\partial \omega_{aa}} = \left[(1-2\tilde{\pi})\rho^*(\tilde{\pi},\tilde{\rho}) + \tilde{\pi}(1-\tilde{\pi})\left(\frac{\partial \rho^*(\tilde{\pi},\tilde{\rho})}{\partial \tilde{\pi}}\right)\right] \frac{\partial \tilde{\pi}}{\partial \omega_{aa}} + \tilde{\pi}(1-\tilde{\pi})\left(\frac{\partial \rho^*(\tilde{\pi},\tilde{\rho})}{\partial \tilde{\rho}}\right) \frac{\partial \tilde{\rho}}{\partial \omega_{aa}}.$$

The first term of this derivative is positive so long as $0 \le \tilde{\pi} < 1/2$ and $\rho > 0$ (and hence $\rho^*(\tilde{\pi}, \tilde{\rho}) > 0$).[25] However, the second term is negative since $\partial\tilde{\rho}/\partial\omega_{aa} < 0$. Thus, the net effect of heterogeneity in mean returns and/or default return thresholds on portfolio loss variance is ambiguous. There are countervailing location and shape effects: the first term is an EL related increase (the whole distribution is shifted to the right), and the second term is a correlation decrease (the shape of the distribution changes: probability mass is shifted to the left). But once we control for or hold EL fixed, loss variance is unambiguously reduced as we show in the next section.

3.4.2. Within and between type heterogeneity

The parameter heterogeneity considered so far is relatively simple and can be viewed as within type heterogeneity, in the sense that differences across firms are random draws from the same common distribution. Under this set up, asymptotically as $N \to \infty$, the expected loss is invariant to the portfolio weights and it would not be possible to control the EL while experimenting with different degrees of heterogeneity as measured, for example, by different values of ω_{aa}. To control for EL we need to introduce an additional *systematic* source of heterogeneity. One possible approach would be to introduce firm types where for each type a_i's are draws from different distributions or from the same distribution but with different parameters. As an illustration, suppose the loan portfolio contains two types of firms, $\mathcal{H}$ and $\mathcal{L}$, with portfolio weights $w_{i\mathcal{H}}$, $i = 1, 2, \ldots N_{\mathcal{H}}$, and $w_{i\mathcal{L}}$, for $i = 1, 2, \ldots N_{\mathcal{L}}$, (such that $N = N_{\mathcal{H}} + N_{\mathcal{L}}$), and default probabilities, $\pi_{\mathcal{H}}$ and $\pi_{\mathcal{L}}$, respectively, with $0 < \pi_{\mathcal{L}} < \pi_{\mathcal{H}} < 1/2$. The differences in the default probabilities across the two types of firms could be due to differences in leverage or management quality, summarized, for instance, in a credit rating.

The portfolio loss in this case is given by the sum of losses of the two sub-portfolios,

$$\ell_{N,t+1} = \sum_{i=1}^{N_{\mathcal{H}}} w_{i\mathcal{H}} I\left(a_{i\mathcal{H}} - \delta f_{t+1} - \varepsilon_{i\mathcal{H},t+1}\right) + \sum_{i=1}^{N_{\mathcal{L}}} w_{i\mathcal{L}} I\left(a_{i\mathcal{L}} - \delta f_{t+1} - \varepsilon_{i\mathcal{L},t+1}\right), \tag{42}$$

where $I(\cdot)$ is the indicator function as in Eq. (14),

$$a_{i\mathcal{H}} = a_{\mathcal{H}} + v_{i\mathcal{H}a}, \quad a_{i\mathcal{L}} = a_{\mathcal{L}} + v_{i\mathcal{L}a}, \tag{43}$$

with $f_{t+1} \sim \mathrm{iid}N(0, 1)$, $\varepsilon_{ik,t+1}|f_{t+1} \sim N(0, 1)$ and $v_{ika} \sim \mathrm{iid}N(0, \omega_{aa})$, for $k = \mathcal{H}, \mathcal{L}$. It is also assumed that $\varepsilon_{ik,t+1}$ and v_{ika} are independently distributed across all i and k.[26]

Let $w_{k,N_k} = \sum_{i=1}^{N_k} w_{ik}$ and $w_k = \lim_{N_k \to \infty} w_{k,N_k}$, $k = \mathcal{L}, \mathcal{H}$ where $w_{k,N_k} > 0$ for both finite N_k and as $N_k \to \infty$ so that $w_{\mathcal{H}}, w_{\mathcal{L}} > 0$, $w_{\mathcal{H},N_{\mathcal{H}}} + w_{\mathcal{L},N_{\mathcal{L}}} = 1$. Assuming that the granularity condition (16) holds for each firm type, then as $N_{\mathcal{H}}, N_{\mathcal{L}} \to \infty$ (the within-type portfolio must be large and granular to eliminate within type idiosyncratic risk), we have

$$x|f = \omega_{\mathcal{H}} \Phi\left(\tilde{a}_{\mathcal{H}} - \tilde{\delta} f\right) + w_{\mathcal{L}} \Phi\left(\tilde{a}_{\mathcal{L}} - \tilde{\delta} f\right), \tag{44}$$

where $\tilde{a}_k = a_k (1+\omega_{aa})^{-1/2}$, for $k = \mathcal{H}, \mathcal{L}, w_{\mathcal{H}} + w_{\mathcal{L}} = 1$, and as before $\tilde{\delta} = \delta(1+\omega_{aa})^{-1/2}$. Since $f \sim \mathrm{iid}N(0, 1)$, it is now easily seen that expected loss for the portfolio is just the weighted average of expected losses for the sub-portfolios,

$$E(x) = \tilde{\pi} = w_{\mathcal{H}} \pi_{\mathcal{H}} + w_{\mathcal{L}} \pi_{\mathcal{L}},$$

where

$$\pi_k = \Phi\left(\frac{a_k}{\sqrt{1 + \omega_{aa} + \delta^2}}\right) = \Phi\left(\frac{a_k\sqrt{1-\rho}}{\sqrt{1 + (1-\rho)\omega_{aa}}}\right), \quad \text{for } k = \mathcal{H}, \mathcal{L}.$$

[25] Note that $\partial\tilde{\rho}^*/\partial\tilde{\pi} > 0$.

[26] One could also allow for differences in the variances of v_{ika} across the types, $k = \mathcal{H}, \mathcal{L}$. But to keep the exposition simple here we are assuming that $\mathrm{Var}(v_{ika}) = \omega_{aa}$.

To ensure the same expected losses under the homogeneous and heterogeneous cases we must have

$$\tilde{\pi} = \pi = w_{\mathcal{H}}\pi_{\mathcal{H}} + w_{\mathcal{L}}\pi_{\mathcal{L}}, \tag{45}$$

and this can be achieved, for given values of $\pi_{\mathcal{H}}$ and $\pi_{\mathcal{L}}$, by an appropriate choice of the portfolio weights on the types $\mathcal{L}$ and $\mathcal{H}$, so long as $\pi_{\mathcal{H}} \neq \pi_{\mathcal{L}}$, and $0 < \pi_k < 1$, for $k = \mathcal{H}, \mathcal{L}$.[27] Indeed both $w_{\mathcal{H}}$ and $w_{\mathcal{L}}$ must be positive, so long as $\pi_{\mathcal{H}} \neq \pi_{\mathcal{L}}$, to make the expected loss of the heterogeneous portfolio the same as for the homogeneous portfolio.

Using Eqs. (42) and (23), the variance of the heterogeneous portfolio is

$$V_{\text{het}}(x) = w_{\mathcal{H}}^2 F(\pi_{\mathcal{H}}, \pi_{\mathcal{H}}, \tilde{\rho}) + w_{\mathcal{L}}^2 F(\pi_{\mathcal{L}}, \pi_{\mathcal{L}}, \tilde{\rho}) + 2 w_{\mathcal{H}} w_{\mathcal{L}} F\left(\pi_{\mathcal{H}}, \pi_{\mathcal{L}}, \tilde{\rho}\right) - \pi^2, \tag{46}$$

where

$$F\left(\pi_i, \pi_j, \tilde{\rho}\right) = \Phi_2\left[\Phi^{-1}(\pi_i), \Phi^{-1}\left(\pi_j, \tilde{\rho}\right)\right].$$

Furthermore, the variance of the associated homogeneous portfolio is given by

$$V_{\text{hom}}(x) = F(\pi, \pi, \rho) - \pi^2. \tag{47}$$

It is now easily established that so long as $w_{\mathcal{H}}$ (or $w_{\mathcal{L}}$) is set such that $\tilde{\pi} = \pi$, then for $\rho > 0$, $\omega_{aa} > 0$, and $a_{\mathcal{H}} \neq a_{\mathcal{L}}$, we have

$$V_{\text{hom}}(x) > V_{\text{het}}(x), \tag{48}$$

namely, the risk will be *over*estimated once the EL's of the two portfolios are equalized. A proof of this result is given in Appendix B.

The above result is easily extended to portfolios containing more than two types of firms. Moreover, as the distance between $\pi_{\mathcal{L}}$ and $\pi_{\mathcal{H}}$ widens, the difference between the risks of the two portfolio types increases, suggesting that efficient credit portfolios should follow a "barbell" strategy combining exposures to very high quality credit with very low quality credits, so long as $\pi_k < 1/2$ for $k = \mathcal{H}, \mathcal{L}$. As a result ignoring this type of heterogeneity would result in over estimation of risk when considering loan portfolios with the same EL.

3.4.3. Heterogeneity in cross firm return correlations

Before considering the case of full parameter heterogeneity, it may be instructive to analyze the intermediate case where only pair-wise cross correlation of returns, $\rho_{ij} = Corr(r_{i,t+1}, r_{j,t+1})$, are allowed to vary across firms. With this in mind, consider the following heterogeneous formulation of the Vasicek model:

$$r_{i,t+1} = \sqrt{\rho_i} f_{t+1} + \sqrt{1 - \rho_i}\varepsilon_{i,t+1}, \tag{49}$$

where $\rho_{ij} = \sqrt{\rho_i \rho_j}$. But since firm-specific innovations are all standard normal, then for every i, $r_{i,t+1} \sim \text{iid}N(0, 1)$, and we have

$$\pi_{i,t+1} = Pr\left(r_{i,t+1} < \lambda\right) = \Phi(\lambda) = \pi,$$

namely despite the heterogeneity of ρ_i, the expected loss and default probability is the same for all firms.[28]

However, clearly the loss distribution differs from the case where $\rho_i = \rho > 0$, and in general will depend on the probability distribution of ρ_i across i. Here for illustrative purposes we shall consider a similar example as before where we assume that the portfolio is composed of two groups of firms that only differ in terms of return correlations: the first group is composed of firms with a low return correlation, ρ_L, and a second group with a high return correlation, ρ_H. It turns out that the impact on UL is ambiguous in this case and will depend on ρ_L, ρ_H and π. For instance, if $\pi = 0.01$ and firm types are evenly split between $\rho_L = 0.05$ and $\rho_H = 0.15$, then $\text{UL}_{\text{hom}}(x) = 0.963\% > 0.940\% = \text{UL}_{\text{het}}(x)$. But for higher correlations such as $\rho_L = 0.6$ and $\rho_H = 0.8$, we have $\text{UL}_{\text{hom}}(x) = 5.068\% < 5.112\% = \text{UL}_{\text{het}}(x)$. Similar examples can be constructed for different values of π but keeping ρ_L and ρ_H the same.

[27] Note that the possibility of $\pi_{\mathcal{H}} = \pi_{\mathcal{L}}$ is ruled out only if $a_{\mathcal{H}} \neq a_{\mathcal{L}}$, which requires $a_{i\mathcal{H}}$ and $a_{i\mathcal{L}}$ to be draws from distributions with different means.

[28] A peculiarity of the Vasicek model is that firm returns are always standard normal, no matter what the correlation. This is achieved by exactly off-setting the idiosyncratic risk, which gets weight $\sqrt{1 - \rho_i}$, with systematic risk which gets weight $\sqrt{\rho_i}$ since $f_t + 1$ is also standard normal.

3.4.4. Full parameter heterogeneity

The effect on loss distributions when heterogeneity is introduced at all levels (or existing heterogeneity is ignored) is complex and hard to predict. The empirical application in Section 4 provides a typical case with the net result that risk declines, both in terms of UL and VaR, when one allows for heterogeneity in all the parameters. The effect of allowing for parameter heterogeneity on the loss distribution in the multifactor case is derived in Appendix A.3. For the single factor case, the quantitative implications of having non-zero values of ω_{aa}, $\omega_{\delta\delta}$, and $\omega_{a\delta}$ on the loss distribution can be easily evaluated using Eq. (36) through random draws $f^{(r)}$, $r=1, 2,\ldots, R$, from the assumed distribution of f. Given these simulated values one can readily compute UL, VaR, and other distributional characteristics as desired, and we do so in the empirical application in the next section.

4. Empirical Illustration: The impact of neglected heterogeneity

In this section we consider different types of heterogeneity across firms and illustrate their effects on the resulting loss distribution by simulating losses for credit portfolios comprised of publicly traded U.S. firms. We confirm that the predictions based on the random coefficient model, as set out in Section 3.4, match those obtained from more conventional simulation techniques. Finally, we are also interested in understanding which source of heterogeneity is the most important in affecting the shape of the loss distribution: the firm return process and associated factor loadings or the default return threshold through information either on distance to default or the credit rating.

We start with a brief description of the data and the loan portfolio to be considered. We then set out five models of firm returns that differ in the extent to which they allow for cross firm heterogeneity. After a brief discussion of the econometric results obtained for these models we consider alternative ways that the loss distribution can be simulated. We first simulate the losses by calibrating the asymptotic loss distribution results obtained in Section 3 using the mean, variance and covariances of the parameter estimates from return regressions. These calibrated (asymptotic) loss distributions are then compared to their finite sample counterparts that make direct use of the parameter estimates of the individual firm returns.

4.1. Data and portfolio construction

We form credit portfolios of publicly traded U.S. firms at the end of each year from 1997 to 2002 and then simulate portfolio losses for the following year. Parameters are estimated recursively using 10-year (40-quarter) rolling windows. The simulations are out-of-sample in that the models, fitted over a ten-year sample, are used to simulate losses for the subsequent 11th year. This recursive procedure allows us to explore the robustness of the results to possible time variation in the underlying parameters.

The loss simulations require an estimate of the probability of default for each firm. These are obtained at the level of the credit rating, $\mathcal{R}$, assigned to the firm by the two largest credit rating agencies: Moody's and S&P. In keeping with our overall empirical strategy, we estimate probabilities of default recursively for each grade using 10-year rolling windows of all firm rating histories from S&P. These probabilities are estimated using the time-homogeneous Markov or parametric duration estimator discussed in Lando and Skødeberg (2002) and Jafry and Schuermann (2004). We impose a minimum annual probability of default (PD) of 0.001% or 0.1 basis points which turns out to be binding for both $\mathcal{AAA}$ and $\mathcal{AA}$.

In order to be selected for inclusion in one of our portfolios, a firm needs (a) 10 years of consecutive quarterly equity returns in the CRSP database that match the rolling estimation window, and (b) an active credit rating from either Moody's or S&P at the end of the evaluation window under consideration. In case both ratings are available the S&P rating is chosen.[29] For the first sample or cohort (which ends in 1997) we have 628 firms.

The portfolio composition is adjusted annually, starting with 1998, to reflect defaults, upgrades and downgrades which may have occurred during the year. We also reallocate exposure annually to reflect changes in the distribution of ratings within the universe of rated U.S. firms. For example, $\mathcal{CCC}-$rated firms made up only 2.31% of all rated U.S. firms at the end of 1997, but by year-end 2002 this proportion had risen to 5.89%. In Table 1 we show the average ratings distribution for 1997 and 2002, and we see clearly that there was a systematic deterioration in average credit quality of rated U.S. firms over this period. In addition, estimated probabilities of default for non-investment grade ratings, and for

[29] Our decision rule is driven by the use of S&P ratings histories to compute the default probabilities $\pi_{\mathcal{R}}$.

Table 1
Ratings distributions and probabilities of default

Credit rating	1997		2002	
	w_R(%)	$\hat{\pi}_R$(%)	w_R(%)	$\hat{\pi}_R$(%)
$\mathcal{AAA}$	2.86	0.001	2.37	0.001
$\mathcal{AA}$	10.81	0.001	9.25	0.001
$\mathcal{A}$	25.61	0.005	21.49	0.006
$\mathcal{BBB}$	22.33	0.064	25.67	0.106
$\mathcal{BB}$	16.3	0.481	15.72	0.630
$\mathcal{B}$	19.79	3.343	19.62	5.429
$\mathcal{CCC}$	2.31	36.487	5.89	49.776
Portfolio		1.60		4.12

Note: The table presents the distribution of firms by rating for year-end 1997 and 2002. These distributions are calculated by taking the average of the distribution for Moody's and the distribution for Standard and Poor's. We construct our portfolios so that the exposure weights are consistent with the observed rating distribution. The final column, $\hat{\pi}_R$(%), contains the estimated annual probabilities of default (PD) that are used in the simulation exercises. These PDs are estimated using the time-homogeneous Markov or parametric duration estimator discussed in Lando and Skødeberg (2002) and Jafry and Schuermann (2004). A minimum annual PD of 0.001% or 0.1 basis points is imposed.

$\mathcal{CCC}$ in particular, have risen noticeably over this period. As a result, the weighted average annual probability of default, $\hat{\pi}$, has increased from 1.60% for the year-end 1997 portfolio to 4.12% for the year-end 2002 portfolio. To be sure, since firms choose whether or not to obtain a rating, this sample suffers from self-selection as does any sample which makes use of credit ratings. As a result it is unclear whether these patterns are reflective of the broader population of U.S. firms.

Using two-digit SIC codes we group firms into seven broad industry sectors to ensure a sufficient number of firms per sector. The sectors and percentage of firms by sector at year-end 1997 and 2002 are summarized in Table 2.

4.2. Model specifications

We model firm returns with single or multiple factors, allowing an increasing degree of heterogeneity across the different models. The simplest model (Model I) is the fully homogeneous return specification analogous to the one assumed by Vasicek:

$$r_{ij,t+1} = \alpha + \beta\bar{r}_{t+1} + u_{ij,t+1}, \tag{50}$$

where $r_{ij,t+1}$ is the stock return to firm i in sector j over the quarter t to $t+1$, $\bar{r}_{t+1}$ is the return on the market factor,[30] and $u_{ij,t+1}$~iid(0, σ^2). While Eq. (50) is not exactly the Vasicek model as originally specified by Eq. (12), it is analogous in that all firms have the same return correlation and return default threshold (or distance to default). The coefficients α, β, and σ in Eq. (50) are estimated with pooled regression.

The second model tests the predictions made by theory in Section 3.4 by introducing heterogeneity in default return thresholds by rating (Model II). The third model specification (Model III) allows for full parameter heterogeneity where firm alphas, factor loadings and error variances in (50) are allowed to vary across firms. That is, we estimate a market model separately for each firm. In the fourth specification (Model IV) we add an industry or sector factor so that each firm's return is regressed on, $\bar{r}_{t+1}$ as well as on, $\bar{r}_{j,t+1}$, where $\bar{r}_{j,t+1}$ is a sector return computed as the market-cap weighted average of firm returns in that sector. In this case the return equation for firm i in sector j is given by

$$r_{ij,t+1} = \alpha_i + \beta_{1i}\bar{r}_{t+1} + \beta_{2i}\bar{r}_{j,t+1} + u_{ij,t+1}, \tag{51}$$

[30] The market return is computed as the value-weighted return of all firms in the sample. We compared this return to two other benchmarks: the returns on S&P 500 and on the CRSP value-weighted index, and found their correlations to be very high, namely 0.985 and 0.970, respectively.

Table 2
Industry breakdown

Industry	1997	2002
Agriculture, mining and construction	5.3	5.7
Communication, electric and gas	16.7	12.8
Durable manufacturing	22.1	23.3
Finance, insurance and real estate	23.1	21.7
Non-durable manufacturing	18.2	18.7
Service	4.8	7.2
Wholesale and retail trade	9.9	10.7
Total	100.0	100.0

Note: The table presents the distribution of firms by industry group as of year-end 1997 and 2002.

where error variances are firm-specific, $u_{ij,t+1} \sim \text{iid}N(0, \sigma_{\text{ij}}^2)$ as is the case for Model III, and the market and sector returns are assumed to have a joint normal distribution.

In simulating the loss distributions we must impose conditional independence, namely that conditional on $\overline{r}_t$ and $\overline{r}_{tj}$ the individual firm returns are independently distributed. But if this requirement is not met, the risk is likely to be underestimated. With this in mind, we also consider a fifth purely statistical model (Model V) where the unobserved factors are estimated by the principal components (PC). We selected the number of factors (m), using the IC_1 and IC_2 selection criteria proposed in Bai and Ng (2002), with the maximum number of factors set to 5. Applying the procedure to the 1997 cohort of firms, m was estimated to be 2 for both selection criteria. For tractability the number of factors was kept fixed for the subsequent cohorts, although the actual factors were re-estimated for each cohort.

Table 3 summarizes the five model specifications that we consider: three single and two multiple (two) factor models.

4.3. Return regressions: Recursive estimates

The recursively estimated return regression parameters, using a 10-year rolling window, are summarized in Table 4. We focus our discussion on the average pair-wise correlation of returns and the average pair-wise correlation of residuals as they map naturally into our loss modeling framework. The average pair-wise correlation of residuals is of particular interest since it gives an indication of how close a particular model is to conditional independence.

Starting with the results in Panel A of Table 4, we note that the in-sample average pair-wise correlation of quarterly returns for the first ten years (1988–1997) is 0.1933. The factor models generally do a good job of accounting for the cross-section correlation of returns, at least in-sample. The average pair-wise correlation of residuals for the whole portfolio is around 0.037 for the Vasicek and the single factor CAPM models. Adding an industry factor reduces that residual correlation to 0.022, and the PCA model by construction leaves almost no cross-section residual correlation. To be sure, there is no guarantee that these results will hold out-of-sample.

In-sample goodness of fit across models as measured by $\overline{R}^2$ (not reported in the table) range from 0.112 for the Vasicek to 0.208 for the sector CAPM, and to 0.220 for the PCA model, thus confirming that models that allow for parameter heterogeneity tend to perform better than the homogeneous model.

Table 3
Specifications of return equations

Model		Return specification
I	Vasicek	$r_{ij,t+1} = \alpha_c + \beta\overline{r}_{t+1} + u_{ij,t+1}$
II	Vasicek + Rating	$r_{ij,t+1} = \alpha + \beta\overline{r}_{t+1} + u_{ij,t+1}$
III	CAPM	$r_{ij,t+1} = \alpha_{ij} + \beta_{ij}\overline{\text{r}}_{t+1} + u_{ij,t+1}$
IV	CAPM + Sector	$r_{ij,t+1} = \alpha_{ij} + \beta_{1,ij}\overline{r}_{t+1} + \beta_{2,ij}\overline{r}_{t+1} + u_{ij,t+1}$
V	PCA	$r_{ij,t+1} = \alpha_{ij} + \beta'_{ij}\text{f}_{t+1} + u_{ij,t+1}$

Note: $r_{ij,t+1}$ denotes the return of firm i in sector j over the quarter t to $t+1$. In Models I through IV, $\overline{r}_{t+1}$ denotes the market-cap weighted over the quarter t to $t+1$, and in Model IV $\overline{r}_{j,t+1}$ is the market-cap weighted return for sector j over the same period. For Model I there is a single default threshold for all firms. For Models II to V there is one default threshold per rating.

Table 4
Average pair-wise correlation of returns and in-sample residuals based on ten-year rolling windows

		Avg. pair-wise correlation of returns	Model specifications		Avg. pair-wise correlation of residuals
Panel A	1988–1997	0.1933	I	Vasicek	0.0365
	# of firms		III	CAPM	0.0374
	628		IV	CAPM + Sector	0.0221
			V	PCA	0.0005
Panel B	1989–1998	0.2114	I	Vasicek	0.0440
	# of firms		III	CAPM	0.0456
	633		IV	CAPM + Sector	0.0324
			V	PCA	0.0004
Panel C	1990–1999	0.2237	I	Vasicek	0.0731
	# of firms		III	CAPM	0.0778
	613		IV	CAPM + Sector	0.0621
			V	PCA	−0.0001
Panel D	1991–2000	0.1691	I	Vasicek	0.0783
	# of firms		III	CAPM	0.0821
	588		IV	CAPM + Sector	0.0615
			V	PCA	0.0008
Panel E	1992–2001	0.1633	I	Vasicek	0.0740
	# of firms		III	CAPM	0.0772
	585		IV	CAPM + Sector	0.0624
			V	PCA	−0.0003
Panel F	1993–2002	0.1999	I	Vasicek	0.0772
	# of firms		III	CAPM	0.0811
	600		IV	CAPM + Sector	0.0658
			V	PCA	−0.0009

Note: This table presents the results of recursive estimation of return equations using quarterly return data. All estimation results are calculated using a 40-quarter rolling window. Portfolio determination and sample construction are discussed in Section 4.2. Specification of the return models is discussed in Section 4.3 (see Table 3 for further detail). The data source for returns is CRSP.

Panels B through F in Table 4 show the recursive results using a 10-year rolling window for the next five ten-year periods. We note that average pair-wise cross-sectional correlations of firm returns remain at around 20% through 1999, but starting with the cohort of 1991–2000 (Panel D), the average correlation for the portfolio drops to 0.169. The sudden and substantial market reversals in the U.S. in March 2000 and the subsequent market declines probably play an important role in explaining these results.[31] Similar patterns are also observed across the different models over the subsequent periods.

A further source of systematic heterogeneity across firms is the default probability or distance to default, captured for instance by the differences in their credit rating. It is reasonable to expect that the return processes of firms with a relatively high credit rating should on average exhibit relatively low error variances and vice versa. This is indeed the case when we compare the average estimates of $\sigma_i^2 = \mathrm{Var}(u_{i,t+1})$ across ratings. Table 5 shows the pooled estimate of σ for Models I (Vasicek) and II (Vasicek + Rating) where the estimates do not vary across firms, and an average estimate for the CAPM where they do, based on the final 10-year cohort, 1993–2002. Similar results are obtained for other sample periods. The average estimate of the firm beta, $\bar{\beta}$, for Models I & II (recall this is a pooled estimate) is 0.867, and the average firm error volatility, $\bar{\sigma}$, is 0.194.

The next four rows show the average estimates by credit rating for Model III. Taking the next-to-last row first, the average residual volatility by rating, $\bar{\sigma}_{\mathcal{R}}$, increases monotonically as we descend the credit spectrum, from 0.110 for $\mathcal{AAA}$ and $\mathcal{AA}$ firms, to 0.287 for $\mathcal{B}$−rated and 0.301 for $\mathcal{CCC}$−rated firms. Thus credit ratings seem to sort firms by

[31] Throughout the analysis we have been assuming time invariant volatilities. While it is well known that high frequency (daily, weekly) firm returns exhibit volatility clustering, this effect tends to vanish as the data frequency declines due to temporal aggregation effects. Nonetheless, we conducted standard diagnostic tests for ARCH effects on all firm return regressions in the case of Model III and calculated the percentage of firm-specific return regressions in which the ARCH effects are significant at the 5% level. For most periods the percentage of firms with significant ARCH effects fell between 5 and 10%; the detailed results are available upon request from the authors. Overall the evidence is not sufficiently overwhelming to motivate ARCH modeling across all firms.

Table 5
Parameter estimates by credit ratings based on return regressions 1993–2002

Model	Parameter	$\mathcal{AAA}/\mathcal{AA}$	$\mathcal{A}$	$\mathcal{BBB}$	$\mathcal{BB}$	$\mathcal{B}$	$\mathcal{CCC}$
Vasicek/	$\bar{\beta}$	0.867	0.867	0.867	0.867	0.867	0.867
Vasicek+Rating	$\bar{\sigma}$	0.194	0.194	0.194	0.194	0.194	0.194
CAPM	$\bar{\beta}_{\mathcal{R}}$	0.847	0.821	0.757	0.984	1.161	0.362
	sd $(\bar{\beta}_{\mathcal{R}})$	0.385	0.524	0.507	0.704	0.918	0.987
	$\bar{\sigma}_{\mathcal{R}}$	0.110	0.128	0.149	0.219	0.287	0.301
	Nobs	41	161	199	125	66	8

Note: Averages of estimated parameters from Models I and III; see Table 3. Note that the return specification for Model II (Vasicek+Rating) is the same as for Model I (Vasicek). $\bar{\beta}$ and $\bar{\sigma}$ are the pooled estimates of return factor loading ("beta") and firm error variance, respectively, and are thus invariant across ratings. In the CAPM model these parameters are estimated separately for each firm, so that $\bar{\beta}_{\mathcal{R}}$ and $\bar{\sigma}_{\mathcal{R}}$ are averages by rating with sd $(\bar{\beta}_{\mathcal{R}})$ being the standard deviation of those factor loading estimates by rating based on Nobs number of observations (last row). The mean factor loadings across all investment grades is 0.792 and across all speculative grades is 1.017.

firm-specific risk, so it is reasonable to consider credit rating as being able to distinguish systematic differences in distance to default. Firm betas tend to increase as we descend the credit spectrum, as to be expected when increasing leverage, a main (but far from the only) factor determining credit ratings. Clearly, there is considerable variation of estimated betas within rating, as is shown in the row labeled sd $\left(\bar{\beta}_{\mathcal{R}}\right)$. Therefore, allowing for such heterogeneity, as our approach does, may be important. For example, over the period 1993–2003 the average betas, weighted by number of observations, for the investment grades was 0.792 as compared to 1.102 for the speculative grades. This is as to be expected, and was in fact replicated in the case of the other estimation windows, with the average estimated betas for the investment grades being substantially lower than the estimates obtained for the speculative grades.[32]

Before we move on to analyzing the impact of heterogeneity on credit loss distributions, it is worth asking whether there is any empirical evidence of parameter heterogeneity in firm returns. For instance in the one-factor CAPM (Model III), are the slopes (betas) really different across firms or are we unable to reject the null that they are all the same, as imposed by Model I? In the case of a small number of firms the hypothesis of slope homogeneity can be tested using standard F tests. However, the test breaks down if the number firms is large relative to the sample period, as is the case in our application where N is around 600 and $T=40$. For such large panels Pesaran and Yamagata (in press, PY) have developed dispersion (or delta) statistics that are asymptotically normally distributed even if N is large relative to T. We computed PY delta statistics for 628 and 600 firms over the sample periods 1988–1997 and 1993–2002, respectively, and obtained statistics in excess of 20 (as compared to a 5% critical value of 1.96) in the case of both samples, which suggest massive rejection of the null of slope homogeneity.[33] This degree of slope heterogeneity is unlikely to disappear even if we consider Model IV that allows for sector-specific returns, or even if we were to include firm-specific covariates. This is because such firm-specific covariates would have little correlation with market or sector returns in the time series, and hence their inclusion would be unlikely to have any significant effects on the slope estimates.

4.4. Impact of heterogeneity on credit losses

We are now ready to generate loss distributions using the return and distance to default parameter estimates from Section 4.3. We begin by simulating the limit loss distribution (with $N \to \infty$) presented in Section 3.4 assuming the parameter estimates are random draws from Gaussian processes. We then simulate the corresponding loss distributions using directly the firm-specific estimates from the return regressions discussed above, thus keeping N fixed at around 600. In what follows we shall refer to the former as the "asymptotic portfolios" and the latter as the "finite portfolios." Losses for the asymptotic portfolio are simulated using what we call the "random parameters approach," while losses for the finite portfolio are simulated using the "firm-specific parameters approach." In both approaches we maintain the

[32] We repeated the same exercise using the two alternative market return benchmarks, the S&P 500 and the CRSP value-weighted return, and obtained very similar results.
[33] We are grateful to Takashi Yamagata for computing these test statistics.

double-Gaussian assumption applied to the common factors, $\mathbf{f}_t$, and the idiosyncratic shocks, $\varepsilon_{ij,t+1}$. This allows us to compare these two approaches, and should help shed light on the validity of the Gaussian random coefficient model for the analysis of loss distributions. Moreover, as argued in Section 4.5, the asymptotic approach has important practical merit for portfolios where reliable firm-specific estimates of β and σ can not be obtained because of inadequate return histories, or because some of the firms in the portfolio might not be publicly traded companies.

4.4.1. Simulating losses for asymptotic portfolios

In what follows we focus on unexpected losses given by the square root of Var($\ell_{N,t+1}|\mathcal{I}_t$,) given by (19), and various quantiles or VaRs.

The return process, using the notation introduced in Section 2.2, is defined by

$$r_{i,t+1} = \mu_i + \gamma_i f_{t+1} + \sigma_i \varepsilon_{i,t+1}, \quad \varepsilon_{i,t+1} \sim \mathrm{iid}N(0,1), \quad f_{t+1} \sim N(0,1),$$

or expressed as the more familiar CAPM specification:

$$r_{i,t+1} = \alpha_i + \beta_i \bar{r}_{t+1} + \sigma_i \varepsilon_{i,t+1}, \quad \varepsilon_{i,t+1} \sim \mathrm{iid}N(0,1), \quad \bar{r}_{t+1} \sim N\left(\bar{r}, \sigma^2_{\bar{r}}\right),$$

with the following relationships between the parameters of the two specifications

$$\alpha_i = \mu_i - \beta_i \bar{r}, \text{ and } \beta_i = \gamma_i / \sigma_{\bar{r}}.$$

Thus firm i's default probability is

$$\pi_{i,t+1} = Pr\left(r_{i,t+1} \leq \lambda_{i,t+1}\right) = \Phi\left(\frac{\lambda_{i,t+1} - \mu_i}{\sigma_{\xi_i}}\right) = \Phi\left(\frac{\lambda_{i,t+1} - \alpha_i - \beta_i \bar{r}_{t+1}}{\sigma_{\xi_i}}\right), \tag{52}$$

where $\sigma^2_{\xi_i} = \beta_i^2 \sigma^2_{\bar{r}} + \sigma_i^2$.

Empirical estimates of default return thresholds are obtained using firm credit ratings by assuming that all firms with a given rating, say $\mathcal{R}$, have the same probability of default and hence the same distance to default which is reasonable given the results in Table 5 (discussed in the previous section).[34] With that identifying assumption applied to (52) we obtain the following default return thresholds:

$$\lambda_{i\mathcal{R}} = \sigma_{\xi_i} \Phi^{-1}(\pi_{\mathcal{R}}) + \mu_i. \tag{53}$$

A firm of type $\mathcal{R}$ defaults if $\mu_{i\mathcal{R}} + \gamma_{i\mathcal{R}} f_{t+1} + \sigma_{i\mathcal{R}} \varepsilon_{i,t+1} \leq \lambda_{i\mathcal{R}}$, or if $\delta_{i\mathcal{R}} f_{t+1} + \varepsilon_{i,t+1} \leq a_{i\mathcal{R}}$, where

$$\delta_{i\mathcal{R}} = \frac{\gamma_{i\mathcal{R}}}{\sigma_{i\mathcal{R}}} = \frac{\beta_{i\mathcal{R}} \sigma_{\bar{r}}}{\sigma_{i\mathcal{R}}}, \tag{54}$$

and

$$a_{i\mathcal{R}} = \sqrt{1 + \delta^2_{i\mathcal{R}}} \Phi^{-1}(\pi_{\mathcal{R}}). \tag{55}$$

The reduced form parameters $\delta_{i\mathcal{R}}$ and $a_{i\mathcal{R}}$ can now be estimated using the estimates of β_i and σ_i from the CAPM regressions (categorized by credit rating at the end of the sample), the default probability estimates by rating, $\hat{\pi}_{\mathcal{R}}$, given for 2002 in the last column of Table 1, and the unconditional mean and variance of the market return, $\bar{r}$ and $\hat{\sigma}^2_{\bar{r}}$. The parameters that enter the loss distribution can then be computed as sample moments by rating which we denote by $\hat{a}_{\mathcal{R}}, \hat{\delta}_{\mathcal{R}}, \hat{\omega}_{a\mathcal{R}a\mathcal{R}}, \hat{\omega}_{\delta\mathcal{R}\delta\mathcal{R}}$, and $\hat{\omega}_{a\mathcal{R}\delta\mathcal{R}}$, for $\mathcal{R} = 1, 2, \ldots, K$, where K is the number of credit rating categories (in our application 7), $\hat{\delta}_{i\mathcal{R}} = \hat{\beta}_{i\mathcal{R}} \hat{\sigma} r / \hat{\sigma}_{i\mathcal{R}}$ and $\hat{a}_{i\mathcal{R}} = \sqrt{1 + \hat{\delta}^2_{i\mathcal{R}}} \Phi^{-1}(\hat{\pi}_{\mathcal{R}})$.

[34] For a detailed discussion, please see Appendix C.

Table 6
Parameter means used in asymptotic loss distribution 1993–2002

Parameter	$\mathcal{AAA}/\mathcal{AA}$	$\mathcal{A}$	$\mathcal{BBB}$	$\mathcal{BB}$	$\mathcal{B}$	$\mathcal{CCC}$
$\hat{\alpha}_R$	−5.451	−4.524	−3.392	−2.738	−1.726	−0.006
$\hat{\delta}_R$	0.607	0.489	0.393	0.365	0.317	0.132
$\omega_{a_R a_R}$	0.528	0.253	0.078	0.059	0.021	<0.001
$\omega_{\delta_R \delta_R}$	0.089	0.076	0.052	0.062	0.056	0.064
$\rho_{a_R \delta_R}$	−0.983	−0.968	−0.949	−0.952	−0.941	−0.568

Note: Averages (by rating) for reduced form parameters for CAPM model from Table 5. $\mathcal{AAA}$ and $\mathcal{AA}$ are grouped since their default probabilities are both assigned the minimum of 0.01 bp. For details on how the parameters are computed, please see Section 4.5.

Using the above parameter estimates losses can be simulated as

$$x^{(r)} = \sum_{\mathcal{R}=1}^{K} w_{\mathcal{R}} \hat{\pi}_{\mathcal{R}} = \sum_{\mathcal{R}=1}^{K} w_{\mathcal{R}} \Phi\left(\frac{\hat{a}_{\mathcal{R}} - \hat{\delta}_{\mathcal{R}} f^{(r)}}{\sqrt{1 + \hat{\omega}_{a_{\mathcal{R}} a_{\mathcal{R}}} + \hat{\omega}_{\delta_{\mathcal{R}} \delta_{\mathcal{R}}} f^{(r)2} - 2\,\hat{\omega}_{a_{\mathcal{R}} \delta_{\mathcal{R}}} f^{(r)}}} \right), \tag{56}$$

where $w_{\mathcal{R}}$ is the weight of $\mathcal{R}$−rated firms in the portfolio $\left(\sum_{\mathcal{R}=1}^{K} w_{\mathcal{R}} = 1\right)$, and $f^{(r)}$, $r=1, 2,\ldots, R$ are random draws from $N(0, 1)$.[35]

Table 6 shows the relevant parameter estimates needed to generate losses in Eq. (56), by rating, for the last 10-year sample window, 1993–2002. The estimate $\hat{a}_{\mathcal{R}}$ is just distance to default scaled by $\sqrt{1+\hat{\delta}_{i\mathcal{R}}^2}$. Since the average standardized factor loading, $\hat{\delta}_{\mathcal{R}}$, varies little across rating, the increase in $\hat{a}_{\mathcal{R}}$ as we descend the rating spectrum is driven mostly by the increase in probability of default, $\hat{\pi}_{\mathcal{R}}$.

A similar effect is driving the declining within-rating parameter variance, $\hat{w}_{a_{\mathcal{R}} a_{\mathcal{R}}}$. When $\hat{\pi}_{\mathcal{R}}$ is high (e.g. for low ratings such as $\mathcal{B}$ and $\mathcal{CCC}$), $\Phi^{-1}(\hat{\pi}_{\mathcal{R}})$ is close to zero and the within-type variance of $\hat{a}_{i\mathcal{R}}, \hat{w}_{a_{\mathcal{R}} a_{\mathcal{R}}}$, shrinks. Note that this within-type variation is assumed to be purely random. We do not expect any systematic pattern with regard to factor loadings, $\hat{\delta}_{\mathcal{R}}$, nor their within-rating dispersion, $\hat{w}_{\delta_{\mathcal{R}} \delta_{\mathcal{R}}}$. Finally, the correlation coefficient between the two sets of parameter estimates, denoted by $\hat{\rho}_{a_{\mathcal{R}} \delta_{\mathcal{R}}}$, is always negative, as expected from Eq. (55).

In Table 7 we present descriptive statistics of the loss distributions. We draw the reader's attention to the first set of columns labeled "Random Parameters/Asymptotic" which summarize the loss distributions using (56) with R=1,000,000 for the three single-factor specifications, namely homogeneous Vasicek (Model I), Vasicek plus rating (Model II), and CAPM (Model III). To allow for easy comparison we hold EL fixed across models. For each model and each year, we show UL and two commonly reported quantiles (VaR), 99.0% and 99.9%.

We see clearly that allowing for parameter heterogeneity reduces risk, whether measured by UL or VaR. Taking for instance the first simulation year, 1998, we see that allowing for only heterogeneity in the a-parameter through rating-specific distance to default, UL drops by nearly 40%, from 1.54% to 0.94%, and 99.9% VaR drops by nearly a half from 12.19% to 6.88%, as expected from the theoretical results in Section 3.4.1. Allowing in addition for factor loading heterogeneity results in a further reduction of about one-third in UL to 0.65%, and of a tenth in 99.9% VaR to 6.28%. So allowing for heterogeneity in all parameters has the net effect of reducing risk, whether measured by UL or VaR. This basic pattern is repeated for subsequent years.

4.4.2. Simulating losses for finite portfolios

In this section we simulate the loss distribution for the firms in our sample portfolio (composed of around 600 firms) using firm-specific parameter estimates. This exercise allows us to assess the performance of the simulated asymptotic loss distribution in predicting tail properties as compared to the asymptotic results.

We simulate firm returns out-of-sample using Eq. (50) for the one-factor models (I–III) and Eq. (51) for the multi-factor models (IV and V), imposing the conditional independence assumption on firm returns. The loss distributions for the different model specifications are then simulated using appropriate default return thresholds, and assuming for simplicity no recovery in the event of default. These simulations are based on 500,000 replications, and the results are reported in the last set of columns labeled "Firm-Specific Parameters/Finite Portfolio" in Table 7. Broadly speaking, in

[35] Clearly, draws from non-Gaussian distributions can also be considered.

Table 7
Out-of-sample simulated annual losses based on 10-year rolling return regressions

	Simulation Year	Using Sample	Model		Random Parameters, Asymptotic Portfolio			Firm-specific Parameters, Finite Portfolio		
						Value-at-Risk			Value-at-Risk	
					UL (%)	99.0% (%)	99.9% (%)	UL	99.0% (%)	99.9% (%)
Panel A	1998	1988–1997	I	Vasicek	1.54	7.46	12.19	1.64	7.79	12.68
		EL=1.60%	II	Vasicek+Rating	0.94	4.79	6.88	1.31	5.55	7.79
		N=628	III	CAPM	0.65	4.06	6.28	1.14	4.92	7.09
			IV	CAPM+Sector	–	–	–	1.16	5.01	7.36
			V	PCA	–	–	–	1.40	6.14	9.61
Panel B	1999	1989–1998	I	Vasicek	2.03	9.83	16.00	2.13	10.07	16.35
		EL=2.04%	II	Vasicek+Rating	1.21	6.13	8.80	1.52	6.84	9.61
		N=633	III	CAPM	0.78	4.96	7.52	1.23	5.73	8.27
			IV	CAPM+Sector	–	–	–	1.29	6.00	8.54
			V	PCA	–	–	–	1.53	7.19	10.99
Panel C	2000	1990–1999	I	Vasicek	2.48	11.96	18.80	2.60	12.33	19.21
		EL=2.70%	II	Vasicek+Rating	1.42	7.32	10.13	1.77	8.09	11.04
		N=613	III	CAPM	0.92	6.05	8.70	1.42	6.90	9.58
			IV	CAPM+Sector	–	–	–	1.51	7.17	9.87
			V	PCA	–	–	–	1.78	8.54	12.84
Panel D	2001	1991–2000	I	Vasicek	2.05	10.04	14.68	2.20	10.48	15.21
		EL=2.94%	II	Vasicek+Rating	1.16	6.46	8.34	1.59	7.35	9.45
		N=588	III	CAPM	1.11	6.57	8.80	1.55	7.38	9.75
			IV	CAPM+Sector	–	–	–	1.64	7.64	10.18
			V	PCA	–	–	–	1.88	8.72	12.29
Panel E	2002	1992–2001	I	Vasicek	2.32	11.40	16.38	2.47	11.85	16.93
		EL=3.48%	II	Vasicek+Rating	1.26	7.21	9.11	1.73	8.09	10.11
		N=585	III	CAPM	1.00	6.75	8.86	1.56	7.66	9.77
			IV	CAPM+Sector	–	–	–	1.61	7.88	10.20
			V	PCA	–	–	–	1.86	8.98	12.28
Panel F	2003	1993–2002	I	Vasicek	3.11	15.04	22.07	3.24	15.43	22.81
		EL=4.12%	II	Vasicek+Rating	1.63	8.99	11.54	1.99	9.67	12.34
		N=600	III	CAPM	1.08	7.97	10.84	1.60	8.77	11.68
			IV	CAPM+Sector	–	–	–	1.64	8.94	11.85
			V	PCA	–	–	–	1.90	10.07	13.95

Note: This table presents results for simulated out-of-sample annual loss distributions. Model specifications, including the return regressions and determination of default thresholds, are discussed in Section 4.3 (see Table 3 for more detail on the model specifications). Simulations are carried out using 1,000,000 replications for the analytical simulations and 500,000 replications for the finite sample replications. For each year all models are calibrated to have the same expect loss given by $\hat{\bar{\pi}}$. For each simulation, the table reports the standard deviation of losses (denoted Unexpected Losses-UL) as well as the 99.0% and 99.9% quantiles of the distribution (denoted Value-at-Risk).

the present exercises where EL's are held fixed, as predicted by theory risk measured either by UL or VaR declines as model heterogeneity increases.[36]

Before embarking on a detailed model-by-model, year-by-year discussion of the loss simulations, it is helpful to consider Fig. 1 to gain an overview of the results. In Fig. 1 we show the loss densities for 2003 across the five different specifications. It is clear that the results from the five model specifications can be grouped into two sets. While the models differ in several ways, the main distinction between the two groups is the use of credit ratings. The more skewed density with the mode closer to the vertical axis is generated by the fully homogeneous Model I, which does not make use of credit rating information while the others do. Indeed Model II differs from Model I only by adding the rating information. Whatever other sources of heterogeneity may be important, an estimate of the probability of, or distance to, default, as provided by a credit rating, clearly has a significant influence on the overall shape of the loss density. Note that the rating serves as a group summary statistic of default risk and need not come from a rating agency. In a banking context it will likely be based on internal models.

[36] In addition to VaR we also calculated expected shortfall; the results are qualitatively very similar and are not reported here.

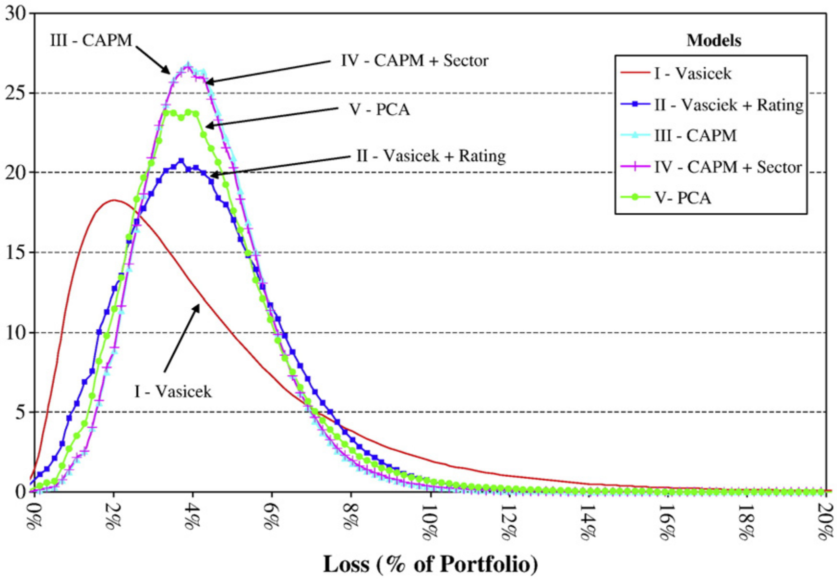

Fig. 1. Simulated finite-sample loss densities using the firm-specific parameters approach (same EL, for 2003).

Consider now the detailed results summarized in Table 7 where we report the loss simulations results for our finite portfolio for each of the six rolling windows. First, comparing the asymptotic and finite-portfolio results for Vasicek (Model I), we see that our finite portfolio is relatively close to an asymptotically diversified portfolio. For example, the portfolio in Panel A has 628 firms, or an effective number of 465 equal-sized exposures.[37] For this portfolio the simulated UL for the Vasicek model is 1.64%, only 10 bps above the asymptotic result, and similarly for the two quantiles 99.0% (within 33 bps) and 99.9% VaR (within 49 bp).[38]

Comparing the first three models within Panel A, we note that the finite portfolio simulations confirm rather precisely the predictions made by the asymptotic theory. The fully homogeneous model of Vasicek (Model I) generates the most extreme losses and exhibit the highest degree of unexpected losses. Adding ratings information (Model II) results in a significant reduction in risk (while controlling for expected losses). If we start in Panel A, UL drops by 20% from 1.64% to 1.31% while 99.9% VaR is reduced by nearly 40% from 12.68% to 7.79%. Credit ratings seem to capture relevant firm-specific information, and this is useful even though the information is grouped together into just a few (seven) rating categories. Models III and IV allow for heterogeneous slopes (factor loadings) and firm-specific error variances, with Model IV also adding an industry return factor. UL falls further by 30% from 1.31% for Model II to 1.14% in the case of Model III, while 99.9% VaR declines another 10% from 7.79% to 7.09%. Thus, the ranking of loss distributions across these three models using simulations from finite portfolios are *exactly* as predicted using the random parameter approach of Section 4.4.1. Adding an industry factor in Model IV results in a very small increase in risk from Model III. However, the loss distributions are very similar: UL is nearly the same, 1.14% vs. 1.16%.

Finally, the principal components Model V generates UL results that are more in line with those obtained for Model II, and generates tail losses which are somewhat higher than Models III and IV. Note that the out-of-sample loss simulations are performed under the maintained assumption of conditional independence. Recall from Table 4 that only Model V has an (in-sample) average pair-wise cross-sectional correlation of residuals which is effectively zero. All other models have some remaining correlation, though less for Model IV than for the one-factor models (I–III). So long as on an out-of-sample basis Model V is still closer to conditional independence than the others, and there is currently no way of verifying this, the other models will generate risk forecasts which are biased downward, meaning that risk would be underestimated since return

[37] If N is the number of obligors in our portfolio, each with exposure weight w_i which is randomly assigned, then $N^* = \left(\sum_{i=1}^{N} w_i^2\right)^{-1}$ is the equivalent number of equally weighted exposures.

[38] This comparison gives a clear indication of the granularity of our finite-sample portfolio since for Vasicek (Model I) $\hat{w}_{a_{\mathcal{R}}a_{\mathcal{R}}} = \hat{w}_{\delta_{\mathcal{R}}\delta_{\mathcal{R}}} = \hat{w}_{a_{\mathcal{R}}\delta_{\mathcal{R}}} = 0$ for all R. Thus, the differences between the two approaches for Model I reflect only the finite size of the portfolio in question.

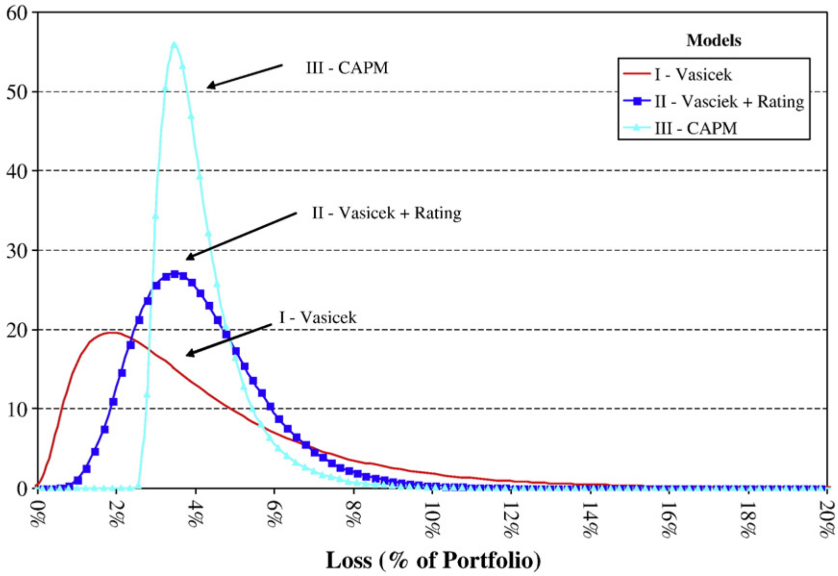

Fig. 2. Calibrated (asymptotic) loss densities under random parameter framework (same EL, for 2003).

correlations are underestimated. Measuring and evaluating out-of-sample conditional dependence is an important topic which requires further investigation.[39]

Returning to Fig. 1, the loss densities are clearly very different. The Vasicek model has only two free parameters (π, ρ), and once credit rating information is included in Model II, the distributional shape changes dramatically. Indeed Model II yields a loss distribution which is remarkably similar to those generated by the fully heterogeneous model specification. Credit ratings seem indeed to be a useful and informative summary statistic for firm-level default risk.

4.5. Heterogeneity and the New Basel Capital Accord

Both sets of results indicate that neglecting firm heterogeneity can have significant impact on risk, and that a dominant source of heterogeneity is the difference in the probability of default across firms as summarized, for instance, in a credit rating. Credit ratings might be obtained from an external rating agency or they could be generated by the bank itself using internal rating models and tools. Indeed the New Basel Capital Accord allows banks, under Pillar 1 (prescribed minimum capital), to assign ratings to their obligors but does not allow them to compute their own factor loadings or return correlations. Those are fixed by the policy makers in the form of the risk weight function which assigns capital to credit exposures. It thus appears that the Basel Committee has allowed banks to model the more important source of heterogeneity.

Nonetheless, full-blown economic capital models, which are limited only by available data, will play an important role in the New Basel Accord under Pillar 2 (supervisory discretion). For obligors that are publicly traded, the finite portfolio simulations which make use of firm-specific parameters as illustrated in Section 4.4.2 would be such an example. However, the bulk of a bank's lending portfolio is to firms which are privately held, and so parameters such as factor loadings cannot be estimated at the firm level. It is still possible to estimate credit ratings using firm financials (these are reported to banks as part of the lending relationship). In this case the theoretical results based on the random parameters approach could be very helpful as the parameter means and their covariances could be estimated using data from publicly traded firms and then applied to the loan portfolio of privately held firms. Moreover, additional (systematic) heterogeneity with respect to country or industry sector could be accommodated in this way. Simply put, our theoretical results provide an easy way of incorporating fairly rich parameter heterogeneity for a portfolio of credit exposures when only limited data is available for parameter estimation.

[39] The absence of conditional independence empirically, especially on an out-of-sample basis, we think has been neglected in the literature. Das et al. (2007) also find significant remaining correlation, or default clustering, even after accounting for observable factors. They propose an unobserved factor approach they call "frailty" to absorb the remaining dependence, although that approach makes out-of-sample forecasting challenging.

As an illustration, in Fig. 2 we show the loss densities based on the random parameters approach as in Section 4.4.1 above using the estimates based on the last 10-year window (ending in 2002, thus simulating the loss distribution for 2003). The chart shows the densities for the three models (I: Vasicek; II: Vasicek plus rating; III: CAPM) while keeping EL fixed, and therefore the chart is comparable to the densities in Fig. 1 which make use of the firm-specific parameter estimates. Even with the limited information contained in the means and the variance–covariances of the return parameters and the assumption of multivariate normality of the parameters made by the random coefficient framework, the densities in Fig. 2 are remarkably close to those in Fig. 1. The simple homogeneous Vasicek specification generates the most skewed and fat-tailed distribution. Allowing for ratings (Model II) has a significant effect on the shape of the loss density, as does adding factor loading heterogeneity (Model III). Returning now to Table 7, we note that the difference in VaR between the random parameters and firm-specific parameters approaches is quite small. For the simulation year used in Figs. 1 and 2 (2003), the difference in 99.9% VaR is 3% for Model I, 7% for Model II and 8% for Model III. Similar differences are obtained for the other five years (not reported). However, the results for the homogenous Vasicek model show that the differences between the finite sample and asymptotic outcomes are partly due to portfolio granularity. Thus, the results in Table 7 suggest that for large portfolios the differences between simulations based on the random parameters framework and those that make use of firm-specific parameter estimates are likely to be even smaller.

5. Concluding remarks

In this paper we have considered a simple model of credit risk and derived the limit distribution of losses under different distributional assumptions regarding the structure of systematic and idiosyncratic risks and the nature of firm heterogeneity. The analytical and simulation results point to some interesting conclusions. In either case, *neglecting* parameter heterogeneity can lead to *under*estimation of expected losses. Then once expected losses are controlled for, the effect of neglecting parameter heterogeneity depends on its source and its severity. Neglected heterogeneity in mean returns and/or default thresholds can lead to overestimation of risk, whether measured by unexpected loss (UL) or value-at-risk (VaR). However, the effect of neglected heterogeneity in return correlation is more complex.

In the empirical analysis, designed to be typical and representative, we are able to examine the impact of neglected heterogeneity in both sets of parameters, and in that case risk is overestimated, whether measured by UL or VaR. In either case the loss distribution effectively is more skewed and fat-tailed when heterogeneity is ignored.

In light of these observations a natural question is: which sources of heterogeneity are most important from the perspective of portfolio losses? Here the answer seems clear: allowing for differences in the distance to or probability of default, measured for instance by a credit rating, is of first order importance in affecting the shape of the loss distribution. Including ratings heterogeneity alone results in a drop in loss volatility of 20%, and a drop of nearly 40% in 99.9% VaR, the VaR-level to which the New Basel Accord is calibrated. For policy makers and risk managers alike, this is good news. After all, an obligor default probability, in the form of a rating, whether generated by a bank internally or provided by a rating agency externally, is one of the key parameters in the New Basel Accord which is allowed to vary. Indeed, early U.S. supervisory guidance indicates that banks must group their obligors into at least seven (non-default) grades, each with a unique probability of default (FRB 2003, p.201). Our results suggest that possibly finer differentiation or grouping along these lines may be fruitful, so long as it is properly implemented.[40]

Appendix A. Limit behavior of credit loss distribution

A.1. Loss densities under homogeneous parameters

In order to show how our approach relates to that of Vasicek, here we consider the homogeneous parameter case but do not require f_{t+1} and $\varepsilon_{i,t+1}$ to have Gaussian distributions. Since in the homogeneous case the multifactor model is equivalent to a single factor model, we consider scalar values for δ_i and $\mathbf{f}_{t+1}$ and denote them by δ and f_{t+1}, respectively. In

[40] See, for instance, Hanson and Schuermann (2006) for a discussion on default probability estimation and their grouping into ratings.

this case we note that conditional on f_{t+1}, the random variables $z_{i,t+1}$ are identically and independently distributed as well as being integrable. (Recall that $|w_i z_{i,t+1}| \leq 1$ for all i and t.) Hence, conditional on f_{t+1} and as $N \rightarrow \infty$, we have

$$\ell_{N,t+1}|f_{t+1}, \mathcal{I}_t \overset{a.s}{\rightarrow} F_\varepsilon(a - \delta f_{t+1}).$$

In the limit the probability density function of $\ell_{N,t+1}|\mathcal{I}_t$ can be obtained from the probability density functions of f_{t+1} and $\varepsilon_{i,t+1}$, which we denote here by $f_f(\cdot)$ and $f_\varepsilon(\cdot)$, respectively. It will be helpful to write the loss density $f_\ell(\cdot)$ in terms of the systematic risk factor density $f_f(\cdot)$ and the standardized idiosyncratic shock density $f_\varepsilon(\cdot)$.

Therefore, conditional on $\mathcal{I}_t$ and denoting the limit of $\ell_{N,t+1}$ as $N \rightarrow \infty$, by ℓ_{t+1} we have (with probability 1)

$$\ell_{t+1} = F_\varepsilon(a - \delta f_{t+1}). \tag{A.1}$$

Now making use of standard results on transformation of probability densities, for $\delta \neq 0$ we have

$$f_\ell(\ell_{t+1}|\mathcal{I}_t) = \left| \frac{\partial F_\varepsilon(a - \delta f_{t+1})}{\delta f_{t+1}} \right|^{-1} f_f(f_{t+1}|\mathcal{I}_t),$$

where f_{t+1} is given in terms of ℓ_{t+1}, via Eq. (A.1), namely

$$f_{t+1} = \frac{a - F_\varepsilon^{-1}(\ell_{t+1})}{\delta},$$

and $|\partial F_\varepsilon(a - \delta f_{t+1})/\partial f_{t+1}|$ is the Jacobian of the transformation which is given by

$$\frac{\partial F_\varepsilon(a - \delta f_{t+1})}{\partial f_{t+1}} = -\delta f_\varepsilon(a - \delta f_{t+1}) = -\delta f_\varepsilon\left[F_\varepsilon^{-1}(\ell_{t+1})\right].$$

Hence

$$f_\ell(\ell_{t+1}|\mathcal{I}_t) = \frac{f_f\left(\frac{a - F_\varepsilon^{-1}(\ell_{t+1})}{\delta}|\mathcal{I}_t\right)}{|\delta| f_\varepsilon\left[F_\varepsilon^{-1}(\ell_{t+1})\right]}, \text{ for } 0 < \ell_{t+1} \leq 1. \tag{A.2}$$

A.2. Relation to Vasicek's loss distribution

The above results provide a simple generalization of Vasicek's one-factor loss density distribution, derived in Vasicek (1991, 2002) and Gordy (2000), given by

$$f_\ell(x|\mathcal{I}_t) = \sqrt{\frac{1-\rho}{\rho}} \left\{ \frac{\phi\left[\frac{\sqrt{1-\rho}\Phi^{-1}(x) - \Phi^{-1}(\pi)}{\sqrt{\rho}}\right]}{\phi\left[\Phi^{-1}(x)\right]} \right\}, \text{ for } 0 < x \leq 1, \ \rho \neq 0, \tag{A.3}$$

where x denotes the fraction of the portfolio lost to defaults. The corresponding loss distribution is

$$F_\ell(x|\mathcal{I}_t) = \Phi\left[\frac{\sqrt{1-\rho}\Phi^{-1}(x) - \Phi^{-1}(\pi)}{\sqrt{\rho}}\right]. \tag{A.4}$$

The density Eq. (A.2) reduces to Eq. (A.3) when $\mu_{ft} = 0$, and assuming that the innovations, f_{t+1} and $\varepsilon_{i,t+1}$ are both Gaussian. In this case

$$f_f(f_{t+1}|\mathcal{I}_t) = \phi(f_{t+1})$$
$$f_\varepsilon(\varepsilon_{i,t+1}|\mathcal{I}_t) = \phi(\varepsilon_{i,t+1}), \quad F_\varepsilon(\cdot) = \Phi(\cdot),$$

and

$$f_\ell(x|\mathcal{I}_t) = \frac{1}{|\delta|} \left\{ \frac{\phi\left[\frac{a - \Phi^{-1}(x)}{\delta}\right]}{\phi\left[\Phi^{-1}(x)\right]} \right\}, \text{ for } 0 < x \leq 1, \quad |\delta| \neq 0 \tag{A.5}$$

where we have used x for ℓ_{t+1} where a and δ are as defined in Eqs. (34) and (33). Using Eqs. (34) and (33) in Eq. (A.5) now yields Vasicek's loss density given by Eq. (A.3).

A.3. Loss densities under full parameter heterogeneity

The impact on the loss distribution is considerably more complicated when we consider heterogeneity across the full set of parameters, including for instance the factor loadings. In this case, $\ell_{N,t+1}$, is given by Eq. (27): $\ell_{N,t+1} = \sum_{i=1}^{N} w_i I\left(a - \delta' \mathrm{f}_{t+1} - \zeta_{i,t+1}\right)$. Since conditional on f_{t+1}, the composite errors, $\zeta_{i,t+1}=\varepsilon_{i,t+1}-v_{ia}+\mathbf{v}_{i\delta}'\mathbf{f}_{t+1}$, are independently distributed across i, then

$$\ell_{N,t+1}|\mathbf{f}_{t+1}, \mathcal{I}_t \xrightarrow{a.s.} F_{\varkappa}\left(\frac{\boldsymbol{\theta}' \mathbf{g}_{t+1}}{\omega_{t+1}}\right),$$

where *a.s.* denotes almost sure convergence, and as before $F_{\varkappa}(\cdot)$ denotes the cumulative distribution function of the standardized composite errors, $\varkappa_{i,t+1}$, defined by Eq. (30), $\mathbf{g}_{t+1}=(1,-\mathrm{f}_{t+1}')'$ and ω_{t+1} is given by Eq. (29). Once again the limiting distribution of credit loss depends on the conditional densities of $\zeta_{i,t+1}$ and f_{t+1}. For example, if $(\varepsilon_{i,t+1}, v_{ia}, \mathbf{v}_{i\delta}')$ follows a multivariate Gaussian distribution, then $\varkappa_{i,t+1}|f_{t+1}, \mathcal{I}_t \sim \text{iid}N(0, 1)$.

The probability density of the fraction of the portfolio lost, x, over the range (0,1), can be derived from the (conditional) joint probability density function assumed for the factors, **f**, by application of standard change-of-variable techniques to the non-linear transformation

$$x = F_{\varkappa}\left(\frac{a - \delta' \mathbf{f}}{\sqrt{1 + \omega_{aa} - 2\omega_{a\delta}' \mathrm{f} + \mathrm{f}' \Omega_{\delta\delta} \mathrm{f}}}\right). \tag{A.6}$$

For a general m factor set up analytical derivations are quite complicated and will not be attempted here. Instead, we consider the relatively simple case of a single factor model, where **f** is a scalar, f. Suppose $f=\psi(x)$ satisfies the transformation, Eq. (A.6), and note that

$$f_{\ell}(x|\mathcal{I}_t) = |\psi'(x)| f_f\left[\psi(x) - \mu_{ft}\right], \text{ for } 0 < x \leq 1,$$

where $|\psi'(x)|=|x'(f)|^{-1}$. In other words, $\psi(x)$ is that value of the systematic factor f which generated loss of x. In the double-Gaussian case, for example, we have

$$x'(f) = \left(\frac{f(\delta\omega_{a\delta} - a\omega_{\delta\delta}) + a\omega_{a\delta} - \delta(1 + \omega_{aa})}{(1 + \omega_{aa} - 2\omega_{a\delta} f + \omega_{\delta\delta} f^2)^{3/2}}\right) \times \phi\left(\frac{a - \delta f}{\sqrt{1 + \omega_{aa} - 2\omega_{a\delta} f + \omega_{\delta\delta} f^2}}\right).$$

Hence

$$|\psi'(x)| = \left(\frac{1}{\phi\left[\Phi^{-1}(x)\right]}\right)\left|\frac{\left[1 + \omega_{aa} - 2\omega_{a\delta}\psi(x) + \omega_{\delta\delta}\psi^2(x)\right]^{3/2}}{\psi(x)(\delta\omega_{a\delta} - a\omega_{\delta\delta}) + a\omega_{a\delta} - \delta(1 + \omega_{aa})}\right|.$$

and for $0 < x \leq 1$ we have

$$f_{\ell}(x|\mathcal{I}_t) = \left|\frac{\left[1 + \omega_{aa} - 2\omega_{a\delta}\psi(x) + \omega_{\delta\delta}\psi^2(x)\right]^{3/2}}{\psi(x)(\delta\omega_{a\delta} - a\omega_{\delta\delta}) + a\omega_{a\delta} - \delta(1 + \omega_{aa})}\right|\left\{\frac{\phi\left[\psi(x) - \mu_{ft}\right]}{\phi\left[\Phi^{-1}(x)\right]}\right\}, \tag{A.7}$$

This limiting loss distribution does not depend on the individual values of the portfolio weights, w_i, $i=1, 2,\ldots, N$, so long as the granularity conditions in Eq. (16) are satisfied.

Appendix B. Proposition 1

Suppose that $\rho>0$, $\omega_{aa}>0$, $\pi=w_{\mathcal{H}}\pi_{\mathcal{H}}+w_{\mathcal{L}}\pi_{\mathcal{L}}, w_{\mathcal{H}}+w_{\mathcal{L}}=1$, and $0<\pi_{\mathcal{L}}<\pi_{\mathcal{H}}<1/2$. Assume that the number of obligors of each type grows large and that the relevant granularity condition is met. Let $F(x, y, \rho)=\Phi_2[\Phi^{-1}(x), \Phi^{-1}(y), \rho]$ and $\tilde{\rho}=\rho/[1+\omega_{aa}(1-\rho)]$. Then we have

$$V_{\text{hom}}(x)>V_{\text{het}}(x) \tag{A.8}$$

where

$$V_{\text{hom}(x)}=F(\pi,\pi,\rho)-\pi^2$$

is the estimated loss variance when heterogeneity is neglected and

$$V_{\text{het}}(x)=w_{\mathcal{H}}^2F(\pi_{\mathcal{H}},\pi_{\mathcal{H}},\tilde{\rho})+w_{\mathcal{L}}^2F(\pi_{\mathcal{L}},\pi_{\mathcal{L}},\tilde{\rho})+2w_{\mathcal{H}}w_{\mathcal{L}}F(\pi_{\mathcal{H}},\pi_{\mathcal{L}},\tilde{\rho})-\pi^2$$

is the estimated loss variance when both systematic and idiosyncratic heterogeneity are taken into account.

Proof. First, note that since $\rho>\tilde{\rho}$, and $\partial F(x, y, \rho)/\partial\rho>0$ (Vasicek, 1998), then

$$F(\pi,\pi,\rho)>F(\pi,\pi,\tilde{\rho}).$$

Therefore, to establish $V_{\text{hom}}(x)>V_{\text{het}}(x)$ it suffices to show that

$$F(\pi,\pi,\tilde{\rho})>w_{\mathcal{H}}^2F(\pi_{\mathcal{H}},\pi_{\mathcal{H}},\tilde{\rho})+w_{\mathcal{L}}^2F(\pi_{\mathcal{L}},\pi_{\mathcal{L}},\tilde{\rho})+2w_{\mathcal{H}}w_{\mathcal{L}}F(\pi_{\mathcal{H}},\pi_{\mathcal{L}},\tilde{\rho}). \tag{A.9}$$

The following Lemma shows that for $\rho>0$ we have $\partial^2F(x, y, \rho)/\partial x^2<0$. Thus, for given values of y and ρ, F (x, y, ρ) is a concave function of x. Therefore, by Jensen's inequality we have

$$F(\pi,\pi,\tilde{\rho})=F(w_{\mathcal{H}}\pi_{\mathcal{H}}+w_{\mathcal{L}}\pi_{\mathcal{L}},\pi,\tilde{\rho})>w_{\mathcal{H}}F(\pi_{\mathcal{H}},\pi,\tilde{\rho})+w_{\mathcal{L}}F(\pi_{\mathcal{L}},\pi,\tilde{\rho}). \tag{A.10}$$

Similarly, $\partial^2F(x, y, \rho)/\partial y^2<0$, so that

$$F(\pi_{\mathcal{H}},\pi,\tilde{\rho})>w_{\mathcal{H}}F(\pi_{\mathcal{H}},\pi_{\mathcal{H}},\tilde{\rho})+w_{\mathcal{L}}F(\pi_{\mathcal{H}},\pi_{\mathcal{L}},\tilde{\rho}),$$
$$F(\pi_{\mathcal{L}},\pi,\tilde{\rho})>w_{\mathcal{H}}F(\pi_{\mathcal{L}},\pi_{\mathcal{H}},\tilde{\rho})+w_{\mathcal{L}}F(\pi_{\mathcal{L}},\pi_{\mathcal{L}},\tilde{\rho}).$$

Substituting the two above inequalities in Eq. (A.10), and noting that by symmetry $F(\pi_{\mathcal{H}},\pi_{\mathcal{L}},\tilde{\rho})=F(\pi_{\mathcal{L}},\pi_{\mathcal{H}},\tilde{\rho})$, yields Eq. (A.9) as required.

Lemma 2. *Let $F(x_1, y_1, \rho)=\Phi_2(\Phi^{-1}(x_1), \Phi^{-1}(y_1), \rho)$. Then for $\rho>0$, we have $\partial^2F(x_1, y_1, \rho)/\partial x_1^2<0$ and $\partial^2F(x_1, y_1, \rho)/\partial y_1^2<0$.*

Proof. By the symmetry of $F(x_1, y_1, \rho)=\Phi_2(\Phi^{-1}(x_1), \Phi^{-1}(y_1), \rho)$ in x_1 and y_1, it suffices to show that $\partial^2F(x_1, y_1, \rho)/\partial x_1^2<0$. Let $G(x)=\Phi^{-1}(x)$ and note that $G'(x)=\Phi^{-1\prime}(x)=[\phi(\Phi^{-1}(x))]^{-1}$. Note that we may write

$$F(x_1,y_1,\rho)=\Phi_2(G(x_1),G(y_1),\rho)=\int_{-\infty}^{G(x_1)}\int_{-\infty}^{G(y_1)}\frac{1}{\sqrt{1-\rho^2}}\phi\left(\frac{y-\rho x}{\sqrt{1-\rho^2}}\right)\phi(x)dydx.$$

Differentiating with respect to x we obtain

$$\frac{\partial F(x_1,y_1,\rho)}{\partial x}=\int_{-\infty}^{G(y_1)}\frac{G'(x_1)}{\sqrt{1-\rho^2}}\phi\left(\frac{y-\rho G(x_1)}{\sqrt{1-\rho^2}}\right)\phi(G(x_1))dy=\Phi\left(\frac{G(y_1)-\rho G(x_1)}{\sqrt{1-\rho^2}}\right)$$

where the second line follows by first noting that $G'(x_1)\phi(G(x_1))=[\phi(\Phi^{-1}(x))]^{-1}\ \phi(\Phi^{-1}(x_1))=1$ and then integrating the result. Differentiating again, we find that the second partial derivative with respect to x_1 is given by

$$\frac{\partial^2 F(x_1,y_1,\rho)}{\partial x_1^2}=-\frac{\rho}{\sqrt{1-\rho^2}}\phi\left(\frac{\Phi^{-1}(y_1)-\rho\Phi^{-1}(x_1)}{\sqrt{1-\rho^2}}\right)\frac{1}{\phi\left(\Phi^{-1}(x_1)\right)}.$$

The above expression is strictly less than zero for $\rho>0$ as claimed.

Appendix C. Heterogeneity in default thresholds: Specification and identification

The probability of default for the ith firm is given by Eq. (6), which we reproduce here for convenience:

$$\pi_{i,t+1}=\Phi\left(\frac{\lambda_{i,t+1}-\mu_i}{\sigma_{\xi_i}}\right).$$

This provides a functional relationship between a firm's equity returns (as characterized by μ_i and $\sigma_{\xi i}$), its default return threshold, $\lambda_{i,t+1}$, and the default probability, $\pi_{i,t+1}$. In the case of publicly traded companies, μ_i and $\sigma_{\xi i}$ can be consistently estimated from market returns based on historical data using either rolling or expanding observation windows. In general, however, $\lambda_{i,t+1}$ and $\pi_{i,t+1}$ can not be directly observed. One possibility would be to use balance sheet and other accounting data to estimate $\lambda_{i,t+1}$. This approach is taken up by Vassalou and Xing (2004) to cite an academic example, and KMV as an industry example, both of which use just the book value of debt (typically all short plus 1/2 of long term debt). But as argued in PSTW, such accounting information is likely to be noisy and might not be all that reliable due to information asymmetries and agency problems between managers, share-and debtholders.[41] In addition to accounting data, other firm characteristics, such as firm age and perhaps size, as well as management quality could also be important in the determination of default return thresholds, and most if not all of those characteristics typically go into a credit rating. In view of these measurement problems, PSTW propose an alternative estimation approach where firm-specific default return thresholds are obtained using firm-specific credit ratings and historical default frequencies. These credit ratings could be either external, e.g. supplied by a rating agency, or internal from a bank's rating unit.

To be sure, neither our approach nor the results are predicated on the use of credit ratings per se, but rather on some summary measure of firm-specific default risk. PD point estimates, however derived, are very noisy, suggesting an averaging or grouping approach. This is effectively what a credit rating does, whether provided by an external rating agency or a bank-internal model. Moreover, since the bulk of a bank's lending portfolio is to privately held firms, typically only relatively coarse groupings are possible. See Hanson and Schuermann (2006) for a discussion on external ratings, and Trück and Ratchev (2005) on bank-internal ratings.

Broadly two identification schemes are possible, and they imply in turn assumptions about the distance to default, DD. One approach would be to impose the same default return threshold for all firms of a given rating. Alternatively one could impose the same DD for all firms of a given rating, meaning

$$\mathrm{DD}_{i,t+1}=\frac{\lambda_{i,t+1}-\overline{\mu}_i}{\overline{\sigma}_{\xi_i}}=\mathrm{DD}_{\mathcal{R},t+1},\tag{A.11}$$

for all firms i with rating R, where $\overline{\mu}_i$ and $\overline{\sigma}_{\xi i}$ are the unconditional estimates of μ_i and $\sigma_{\xi i}$ obtained using observations on firm-specific returns up to the end of period t. In this case the default return threshold is different for every firm and can be computed using

$$\hat{\lambda}_{i,t+1}=\widehat{\mathrm{DD}}_{\mathcal{R},t+1}\overline{\sigma}_{\xi_i}+\overline{\mu}_i,\quad \text{for } i\in\mathcal{R}_t,\tag{A.12}$$

where

$$\widehat{\mathrm{DD}}_{\mathcal{R},t+1}=\Phi^{-1}\left(\hat{\pi}_{\mathcal{R},t+1}\right),\tag{A.13}$$

[41] With this in mind, Duffie and Lando (2001) allow for the possibility of imperfect information about the firm's assets and default threshold in the context of a first-passage model. Their model is confirmed empirically in Yu (2005).

$\Phi^{-1}(\cdot)$ is the inverse of the cumulative distribution function of the standard normal, and $\hat{\pi}_{\mathcal{R},t+1}$ is the observed default frequency of *R*-rated firms.[42] This approach is analogous to the systematic heterogeneity by types discussed in the theory Section 3.4.2. Once again the estimated default return thresholds, $\hat{\lambda}_{i,t+1}$, will be finite so long as $\hat{\pi}_{\mathcal{R},t+1} \neq \{0,1\}$.

The identification conditions can be summed up as follows: condition (A.11) imposes the same probability of default for each R-rated firm, whereas the alternative strategy simply imposes that this needs to hold on average across R-rated firms in the portfolio. Of the two, the assumption of the same distance-to-default seems more in line with the way credit ratings are established by the main rating agencies and is consistent with the empirical evidence presented in Section 4.3. First, the idea that firms with similar distances-to-default have similar probabilities of default is central to structural models of default. For instance, KMV makes use of a one-to-one mapping from DDs to EDFs (expected default frequencies). Second, rating agencies attempt to group firms according to their probability of default (subject possibly to some adjustments for differences in their expected loss given defaults), and in a structural model this is equivalent to groupingfirms according to distance-to-default.[43]

The default return threshold as specified in Eq. (A.12) incorporates equity market and credit rating information. Empirically, the right-hand-side of Eq. (A.12) is estimated on a rolling-window basis allowing for time variation in $\hat{\lambda}_{i,t+1}$.

References

Altman, Edward I., 1968. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. Journal of Finance 20, 589–609.

Bai, Jushan, Ng, Serena, 2002. Determining the number of factors in approximate factor models. Econometrica 70, 191–221.

Basel Committee on Banking Supervision, 2001. The internal ratings based approach. Basel Committee on Banking Supervision. May; <http://www.bis.org/publ/bcbsca.htm>.

Basel Committee on Banking Supervision, 2005. International Convergence of Capital Measurement and Capital Standards: A Revised Framework. <http://www.bis.org/publ/bcbs118.htm>, November.

Broadie, Mark, Chernov, Mikhail, Sundaresan, Suresh, 2007. Optimal corporate securities values in the presence of chapter 7 and chapter 11. Journal of Finance 62 (3), 1339–1375.

Chamberlain, Gary, Rothschild, Michael, 1983. Arbitrage, factor structure and mean-variance analysis on large asset markets. Econometrica 51, 1281–1304.

Collin-Dufresne, Pierre, Goldstein, Robert S., Hugonnier, Julien, 2004. A general formula for pricing defaultable claims. Econometrica 72, 1377–1407.

Connor, Gregory, Korajczyk, Robert A., 1995. The arbitrage pricing theory and multifactor models of asset returns. ch. 4 In: Jarrow, Robert A., Maksimovic, Vojislav, Ziemba, W.T. (Eds.), Handbooks in Operations Research and Management Science, vol. 9. Elsevier, Amsterdam, NL.

Das, Sajiv. R., Duffie, Darrell, Kapadia, Nikunj, Saita, Leandro, 2007. Common failings: how corporate defaults are correlated. Journal of Finance 62 (1), 93–118.

Davis, Mark, Lo, Violet, 2001. Infectious defaults. Quantitative Finance 1, 382–387.

Driessen, Joost, 2005. Is default event risk priced in corporate bonds? Review of Financial Studies 18, 165–195.

Duffie, Darrell, 2005. Credit risk modeling with affine processes. Journal of Banking & Finance 29, 2751–2802.

Duffie, Darrell, Gârleanu, Nicolae, 2001. Risk and valuation of collateralized debt obligations. Financial Analyst Journal 41–59 January/February.

Duffie, Darrell, Lando, David, 2001. Term structures of credit spreads with incomplete accounting information. Econometrica 69, 633–664.

Duffie, Darrell, Saita, Leandro, Wang, Ke, 2007. Multi-period corporate default prediction with stochastic covariates. Journal of Financial Economics 83 (3), 635–665.

Duffie, Darrell, Singleton, Kenneth J. (1999). "Simulating Correlated Defaults." Working Paper, Stanford Graduate School of Business.

Embrechts, Paul, McNeil, Alexander J., Straumann, D., 2001. Correlation and dependency in risk management: properties and pitfalls. In: Dempster, M.A.H., Moffat, H.K. (Eds.), Risk Management: Value at Risk and Beyond. Cambridge University Press, Cambridge, UK, pp. 176–223.

Federal Reserve Board, 2003. Supervisory Guidance on Internal Ratings-Based Systems for Corporate Credit. Attachment 2 in http://www.federalreserve.gov/boarddocs/meetings/2003/20030711/attachment.pdf.

Frey, Rüdiger, McNeil, Alexander J., 2003. Dependent defaults in models of portfolio credit risk. Journal of Risk 6, 59–92.

Giesecke, Kai, Weber, Stefan, 2004. Cyclical correlations, credit contagion and portfolio losses. Journal of Banking & Finance 28, 3009–3036.

Gordy, Michael B., 2000. A comparative anatomy of credit risk models. Journal of Banking & Finance 24, 119–149.

Gordy, Michael B., 2003. A risk-factor model foundation for ratings-based bank capital rules. Journal of Financial Intermediation 12, 199–232.

Gordy, Michael B., 2004. Granularity adjustment in portfolio credit risk measurement. In: Szegö, Giorgio P. (Ed.), Risk Measures for the 21st Century. John Wiley & Sons, New York, NY.

Gupton, Gregory M., Finger, Christopher C., Bhatia, Mickey, 1997. In: Morgan, J.P. (Ed.), CreditMetrics™ — Technical Document, this version: April 2, 1997. New York.

Hanson, Samuel G., Schuermann, Til, 2006. Confidence intervals for probabilities of default. Journal of Banking & Finance 30 (8), 2281–2301.

Hillegeist, Stephen A., Keating, Elizabeth K., Cram, Donald P., Lundsted, Kyle G., 2004. Assessing the probability of bankruptcy. Review of Accounting Studies 9, 5–34.

[42] Note that $\Phi^{-1}(\pi_{i,t+1})<0$ for $\pi_{i,t+1}<\frac{1}{2}$. In practice $\pi_{i,t+1}$ tends to be quite small.

[43] More detail as well as results using the same-threshold (λ) identifying assumption are given in Pesaran et al. (2006).

Hull, John C., White, Alan, 2004. Valuation of a CDO and *n*th to default CDS without Monte Carlo simulation. Journal of Derivatives 12 (2), 8–23.
Jafry, Yusuf, Schuermann, Til, 2004. Measurement, estimation and comparison of credit migration matrices. Journal of Banking & Finance 28, 2603–2639.
Jarrow, Robert A., Lando, David, Yu, Fan, 2005. Default risk and diversification: theory and empirical implications. Mathematical Finance 15, 1–26.
Koyluoglu, H. Ugur, Hickman, Andrew, 1998. Reconcilable differences. Risk 11, 56–64 October.
Lando, David, 1998. On Cox processes and credit risky securities. Review of Derivatives Research 2, 88–120.
Lando, David, Skødeberg, Torben, 2002. Analyzing ratings transitions and rating drift with continuous observations. Journal of Banking & Finance 26, 423–444.
Leland, Hayne E., Toft, Klaus Bjerre, 1996. Optimal capital structure, endogenous bankruptcy, and the term structure of credit spreads. Journal of Finance 51, 987–1019.
Lennox, Clive, 1999. Identifying failing companies: a re-evaluation of the logit, probit and DA approaches. Journal of Economics and Business 51, 347–364.
Li, David X., 2000. Default correlation: a copula approach. Journal of Fixed Income 43–54 March.
Lucas, André, Klaasen, Pieter, Spreij, Peter, Straetmans, Stefan, 2001. An analytic approach to credit risk of large corporate bond and loan portfolios. Journal of Banking & Finance 25, 1635–1664.
Merton, Robert C., 1974. On the Pricing of corporate debt: the risk structure of interest rates. Journal of Finance 29, 449–470.
Pesaran, M. Hashem, Schuermann, Til, Treutler, Björn-Jakob, 2006. Global business cycles and credit risk. ch. 9 In: Carey, Mark, Stulz, René (Eds.), Risks of Financial Institutions. University of Chicago Press, Chicago, IL, pp. 419–469.
Pesaran, M. Hashem, Schuermann, Til, Treutler, Björn-Jakob, Weiner, Scott M., 2006. Macroeconomic dynamics and credit risk: a global perspective. Journal of Money, Credit and Banking 38 (5), 1211–1261.
Pesaran, M.H., Yamagata, T. (in press). "Testing Slope Homogeneity in Large Panels," forthcoming, Journal of Econometrics.
Ross, Stephen A., 1976. The arbitrage theory of capital asset pricing. Journal of Economic Theory 13, 341–360.
Saunders, Anthony, Allen, Linda, 2002. Credit Risk Measurement-New Approaches to Value at Risk and Other Paradigms, 2nd Ed. John Wiley & Sons, New York.
Schönbucher, Philipp J., 1998. The term structure of defaultable bond prices. Review of Derivatives Research 2, 161–192.
Schönbucher, Philipp J. (2002). "Taken to the Limit: Simple and Not-so-Simple Loan Loss Distributions." Working Paper, Department of Statistics, Bonn Univ.
Schönbucher, Philipp J., 2003. Credit Derivatives Pricing Models: Models, Pricing and Implementation. John Wiley & Sons, New York, NY.
Shumway, Tyler, 2001. Forecasting bankruptcy more accurately: a simple hazard model. Journal of Business 74, 101–124.
Trück, Stefan, Ratchev, Svetzolar T., 2005. Credit portfolio risk and probability of default confidence sets through the business cycle. Journal of Credit Risk 1 (4), 61–88.
Vasicek, Oldrich, 1987. Probability of Loss on Loan Portfolio. KMV Corp. available at www.kmv.com.
Vasicek, Oldrich, 1991. Limiting Loan Loss Distribution. KMV Corp. available at www.kmv.com.
Vasicek, Oldrich, 1998. A series expansion for the bivariate normal integral. Journal of Computational Finance 1, 5–10.
Vasicek, Oldrich, 2002. Loan portfolio value. Risk 15, 160–162.
Vassalou, Maria, Xing, Yuhang, 2004. Default risk in equity returns. Journal of Finance 59, 831–868.
Yu, Fan, 2005. Accounting transparency and the term structure of credit spreads. Journal of Financial Economics 75 (1), 53–84.
Zhou, Chunsheng, 2001. An analysis of default correlations and multiple defaults. Review of Financial Studies 14, 555–576.