BANK FOR INTERNATIONAL SETTLEMENTS

BIS Working Papers
No 214

The pricing of portfolio credit risk

by Nikola Tarashev and Haibin Zhu

Monetary and Economic Department

September 2006

Copies of publications are available from:

Bank for International Settlements
Press & Communications
CH-4002 Basel, Switzerland

E-mail: publications@bis.org

Fax: +41 61 280 9100 and +41 61 280 8100

This publication is available on the BIS website (www.bis.org).

ISSN 1020-0959 (print)

ISSN 1682-7678 (online

The Pricing of Portfolio Credit Risk*

Nikola Tarashev†
Haibin Zhu‡

This draft: September 2006

Abstract

Equity and credit-default-swap (CDS) markets are in disagreement as to the extent to which asset returns co-move across firms. This suggests market segmentation and casts ambiguity about the asset-return correlations underpinning observed prices of portfolio credit risk. The ambiguity could be eliminated by – currently unavailable – data that reveal the market valuation of low-probability/large-impact events. At present, judicious assumptions about this valuation can be used to reconcile observed prices with asset-return correlations implied by either equity or CDS markets. These conclusions are based on an analysis of tranche spreads of a popular CDS index, which incorporate a rather small premium for correlation risk.

JEL Classification Numbers: G13, C15
Keywords: CDS index tranche, Joint distribution of asset returns, Correlation risk premium, Copula

*The views presented here are solely those of the authors and do not necessarily represent those of the Bank for International Settlements. We are grateful to Claudio Borio, Andy Filardo, Peter Hoerdahl, Srichander Ramaswamy, Kostas Tsatsaronis, Hao Zhou and seminar participants at the Bank for International Settlements and the 2006 FDIC Derivatives Conference for discussions and comments on earlier versions of the paper. We also thank Frank Zhang for sharing Matlab codes and Marcus Jellinghaus for data support.

†Nikola Tarashev: Research and Policy Analysis, Bank for International Settlements, Basel, Switzerland. Tel.: 41-61-280-9213. Fax: 41-61-280-9100. E-mail: nikola.tarashev@bis.org.

‡Haibin Zhu: Research and Policy Analysis, Bank for International Settlements, Basel, Switzerland. Tel.: 41-61-280-9164. Fax: 41-61-280-9100. E-mail: haibin.zhu@bis.org.

1 Introduction

Portfolio credit risk has three key components: probability of default (PD), loss given default (LGD) and the probability distribution of joint defaults.[1] The last component, which has received least attention owing to the traditional focus of academic researchers and market practitioners on single-name credit events, has recently gained in importance as a result of the rapid development of innovative products in structured finance. Such products, which allow investors to trade portfolio credit risk, include collateralized debt obligations (CDOs), CDOs of CDOs (or CDO^2), n^{th}-to-default credit default swap (CDS) and CDS indices.[2] The prices of these financial instruments rely heavily on estimated probabilities of default clustering (see Hull and White, 2004; Gibson, 2004). There is no consensus, however, on how market participants construct such estimates.

The literature has proposed two main approaches to estimating the likelihood that a particular number of defaults will occur within a portfolio. The first, direct, approach relies exclusively on default data (Daniels et al., 2005; Demey et al., 2004; Jarrow and van Deventer, 2005). Since defaults are rare events, however, this approach leads to large estimation errors, especially for portfolios consisting of investment-grade entities. The second, indirect, approach is based on the Merton (1974) framework and exploits the notion that a default occurs when the assets of a borrower fall below a threshold value. This notion allows one to combine single-name PDs with the corresponding correlations, third and higher moments of asset returns in order to construct a probability distribution of defaults. Two of the building blocks of this distribution, PDs and asset-return correlations, are typically estimated.[3] By contrast, the literature makes assumptions about the third and higher moments of asset returns, which relate to the skewness of density functions and the fatness of their tails. Such, largely arbitrary, assumptions can influence substantially, and potentially unduly, the estimated distribution of defaults for any estimate of PDs and asset-return correlations.

This is the first paper to investigate whether prices of securities associated with *individual* companies – ie credit spreads or equity prices – could shed light on market valuation of credit risk associated with a *portfolio* of companies. To analyze this valuation, we compare observed tranche spreads of a popular CDS index – Dow Jones CDX North America 5-

[1] The mainstream of the credit risk literature focuses on PD: see Duffie and Singleton (2003) for an overview. The growing literature on LGD includes Altman and Kishore (1996), Jarrow (2001) and Covitz and Han (2004).

[2] For a general discussion of products used for trading portfolio credit risk, see BCBS (2004).

[3] In the special case of Gaussian asset returns, the correlation of these returns describes fully their co-movement. As a result, the distribution of joint defaults can be derived solely on the basis of PDs and asset-return correlations. Zhou (2001) goes beyond this special case and studies analytically the link between asset returns and default correlations in a first-passage-time model.

year (CDX.NA.IG.5Y) – to estimates of these spreads, as implied by data from single-name markets. Since a tranche spread is the price of bearing the risk that portfolio default losses will fall within a particular range, this comparison allows us to draw parallels between our estimates of the distribution of default losses and the corresponding distribution perceived by the CDS index market.

To estimate the distribution of default losses that pertains to the CDS index of interest, we use data on market expectations of LGDs and risk-free rates, and adopt the indirect approach to estimating the probability distribution of joint defaults. We construct the latter distribution with three building blocks. The first building block, a set of PDs, we extract from single-name CDS spreads.[4] Our PD estimates, which embed a premium for the risk of individual defaults, change over time and differ across firms. Time variability in these estimates, as well as in LGDs and risk-free rates of return, drives time variability in estimated distributions of (discounted) default losses – and, thus, in estimated tranche spreads.

The second building block of the probability distribution of joint defaults, a matrix of asset-return correlations, comes in two alternatives. We obtain the first alternative from the same CDS spreads that we use for the PD estimates, as the co-movement in credit spreads incorporates information on the co-movement of the underlying asset values. Moody's KMV delivers the second alternative, which is obtained from equity-market data on the basis of the proprietary GCorr model (Das and Ishii, 2001; Crosbie, 2005). Each of the two alternative sets of asset-return correlations remains constant over the sample period and does not incorporate a premium for the risk that correlations may change over time.

The third building block consists of assumptions regarding the third and higher moments of asset returns. Such assumptions relate directly to market valuations of the risk of low probability/high impact events, also known as "tail events". By varying these moments assumptions, we establish what valuation of tail event risk reconciles either of the two estimates of asset-return correlations with the probability distribution of joint defaults underlying observed tranche spreads.

The comparison between observed and implied spreads of the CDS index CDX.NA.IG.5Y validates our PD estimates. For one, there is a close match between the average of these estimates and the average PD implied by spreads of the overall (or single-tranche) CDS index. This match suggests that probable errors in the PD estimates could only lead to a negligible bias in the CDS- and GCorr-implied tranche spreads for the same index. In addition, there is substantial evidence that PDs are by far the main driver of the evolution of observed

[4]The CDS spread is widely considered as a better price of default risk than the bond spread, as it responds more quickly to changes in credit conditions (Blanco et al., 2005; Zhu, 2006) and is less polluted by non-credit factors (Longstaff et al., 2005).

tranche spreads over time. By extension, this evidence supports our maintained assumption that market perceptions of asset-return correlations are constant during the sample period.

In contrast, we are not in a position to pin down the asset return correlations underlying observed tranche spreads of the CDS index CDX.NA.IG.5Y. The reason is that the difference between CDS-implied correlations – which average 13% – and their GCorr counterparts – which average 24% – is not only statistically significant but also results in a material difference between CDS- and GCorr-implied tranche spreads. For given third and higher moments of asset returns, the former set of correlations implies less clustering of defaults than does the latter, which leads to higher spreads for equity tranches (which provide protection against the first several defaults) and lower spreads for senior tranches (which come into play for a large number of defaults). Importantly, this result is robust to probable estimation errors in the correlation coefficients and their interaction with firm-specific PDs.

Judiciously made assumptions about the third and higher moments of asset returns can reconcile either of the two sets of estimated correlations with the data. In particular, a negatively skewed distribution of the common factor in asset returns supports a high clustering of defaults, which reconciles the lower CDS-implied correlations with observed tranche spreads. By contrast, a similar reconciliation is obtained by coupling the larger GCorr correlations with a Gaussian (ie symmetric and thin-tailed) distribution of asset returns. Since the two sets of estimated correlations differ materially, there does not exist a moments assumption that reconciles simultaneously both sets with the data.

These findings have two important implications. First, there seems to be inconsistency in the way different markets incorporate asset return correlations. If asset returns are indeed influenced by a common factor with a negatively skewed distribution, then the inconsistency exists between credit and equity markets (as data from the latter market underlie GCorr correlations). If instead market players regard asset returns as Gaussian when pricing portfolio credit risk, then there is inconsistency between the single-name CDS and CDS index markets. Second, short of identifying an asset class that trades *exclusively* asset-return correlations, the correlations used for pricing credit risk can be pinned down only after analyzing the third and higher moments of asset returns. Unfortunately, we are not aware of existing data that could render such analysis meaningful.

Even though we cannot pin down the asset-return correlations underlying the tranche spreads of the CDS index CDX.NA.IG.5Y, we find evidence that these spreads embed little, if any, premium for correlation risk. This evidence surfaces in our investigation of how such a premium would have changed GCorr- and CDS-implied spreads, which abstract from it. Specifically, we find that extremely small levels of the premium are needed in order to improve the match between observed and GCorr-implied spreads when asset returns are Gaussian. In addition, under any of the considered assumptions regarding the third and higher moments of asset returns, a correlation-risk premium cannot improve the fit between observed and CDS-implied spreads. These results stand in contrast to Driessen et al. (2005), who find a significant correlation-risk premium in option prices.

Longstaff and Rajan (2006) also tackle the issue of how markets price portfolio credit risk. They adopt a flexible empirical model and conclude that three credit risk factors are needed in order to explain *fully* observed tranche spreads of the CDS index CDX.NA.IG.5Y. As two of these factors come into play with a low probability but have a far-reaching impact, this conclusion is in line with an interpretation of the pricing of portfolio credit risk that is proposed here and emphasizes the importance of tail event risk. In contrast to Longstaff and Rajan (2006), however, this paper does not seek to provide an exact match of observed spreads of index tranches but instead evaluates the extent to which these spreads can be explained by information on the pricing of credit risk obtained from single-name markets.

The remainder of the paper is organized as follows. Section 2 outlines the structure of the CDS index market and explains how index tranches are priced. Then, Section 3 outlines how the spreads of index tranches can be constructed on the basis of data from the single-name CDS and equity markets. Section 4 describes the data. Section 5 reports and dissects main empirical findings. The final section concludes.

2 The CDS index market

The market for a CDS index, which allows traders to buy and sell protection against portfolio credit risk, delivers two sets of prices. The first set is a time series of *single-tranche spreads*, which are effectively the prices of protecting the *entire* notional amount of the index against losses caused by defaults of the entities in this index. Thus, single-tranche spreads reveal the market's risk-neutral expectation of default losses but are insensitive to changes in the (risk-neutral) inter-dependence of these losses across entities.

The second set consists of time series of *multi-tranche spreads*. Each time series consists of the effective prices of protection against a particular range (or "tranche") of credit losses on the notional amount of the index. For example, the tranche relating to the first losses – and, thus, carrying the highest level of credit risk – is known as the equity tranche. If none of the entities in the index defaults, the investor in this tranche (ie the protection seller) receives quarterly a fixed premium payment (or "spread") on the tranche's principal, which is typically 3% of the total notional amount of the index. If defaults occur, this investor is obliged to pay its counterparty (ie the protection buyer) an amount equal to the losses from default up to a maximum of 3% of the total notional amount of the index. At the same time, the principal value of the tranche is reduced for the remainder of the contract's life to reflect credit losses.[5] Similarly, an investor in the so-called mezzanine tranche is typically responsible (only) for losses between 3% and 7% of the total notional amount, while investors in the two senior and two super-senior tranches are responsible (only) for losses between 7% and 10%, 10% and 15%, 15% and 30%, and 30% and 100% of the total notional amount, respectively. Thus, the higher the seniority of the tranche, the less likely it is that the corresponding investor will need to make payments to the protection buyer.

The second set of prices is of greater use to the analysis in this paper, as the spread of each tranche pertains to a particular segment of the probability distribution of defaults. To see why, observe that in a CDS index consisting of 100 equally-weighted entities with LGDs of 50%, the spread of the equity tranche is effectively the price of protection against the first six defaults in the underlying portfolio. For a given (risk-neutral) expectation of default losses, the less dependent the defaults are across entities, the higher the probability of there being a few (ie up to six) defaults and, as a result, the higher the spread of the equity tranche. Conversely, greater interdependence of defaults increases the probability of default clustering – eg of there being no or a lot of defaults – which lowers the equity tranche spread. At the same time, greater default clustering raises the spread of the senior tranches, as (in the current example) these spreads are the prices of protection against the 14^{th} to the 20^{th} and the 20^{th} to the 30^{th} defaults, respectively.

3 Deriving implied tranche spreads

We process information from single-name asset markets in order to calculate the probability distribution of joint defaults associated with a particular CDS index. Equipped with such a probability distribution and data on LGDs and risk-free rates, we use the numerical methodology developed in Gibson (2004) in order to obtain the *implied* tranche spreads of the CDS index. For each particular tranche, this methodology delivers the expected present value of the principal, EP, and the expected present value of contingent payments, EC, made by the protection seller. Denoting the tranche spread by s, the present value of the expected fee revenue of the protection seller, $s \cdot EP$, has to equal EC. Thus, the tranche spread is calculated as:

[5] For the CDS index contract we consider below, a default triggers an immediate adjustment to the payments by the protection seller and buyer. In our calculations, however, we impose the simplifying assumption that such adjustments are made quarterly.

$$s = \frac{EC}{EP}$$

The heart of the empirical exercise in this paper is the construction of the probability distribution of defaults, which is the key component of implied tranche spreads. This probability distribution has three building blocks: (i) the PDs of individual entities entering a CDS index, (ii) the correlation of these entities' asset returns, and (iii) the third and higher moments of the asset-return distribution. We estimate the PDs from single-name CDS spreads, as described in Section 3.1. Then, we obtain two alternative estimates of asset-return correlations. One of these sets is based on the single-name CDS spreads underlying our PD estimates; its derivation is outlined in Section 3.2. The other set, provided by Moody's KMV, is based on equity market data and its derivation is sketched in Section 3.3. Finally, we make several alternative assumptions regarding the third and higher moments of asset returns. The alternatives differ predominantly in the assumed distribution of the *common* factor of asset returns – Gaussian, Student-t or a mixture of Gaussian distributions. Section 3.4 and Appendices A and B outline how we combine third and higher moments assumptions with a set of PD and correlation estimates to derive a probability distribution of defaults.

3.1 CDS-implied PDs

In order to uncover risk-neutral PDs from single-name CDS spreads, we adopt the simplified framework of Duffie (1999). In a typical CDS contract, the protection buyer agrees to make constant periodic premium payments, determined by the CDS spread s_t, to the protection seller until the contract matures (at time $t+T$) or a pre-specified credit event materializes. In return, if such an event occurs, the protection seller compensates the protection buyer with the realized default loss.

For market clearing, the present value of CDS premium payments (the left-hand side of the next equation) has to equal the present value of protection payments (the right-hand side):

$$s_t \int_t^{t+T} e^{-r_\tau \tau} \Gamma_\tau d\tau = LGD_t \int_t^{t+T} e^{-r_\tau \tau} q_\tau d\tau \tag{1}$$

where r stands for the risk-free rate of return, q denotes the (annualized) default intensity, $\Gamma_\tau \equiv 1 - \int_0^\tau q_v dv$ is the risk-neutral survival probability over the following τ years, and $LGD_t \in [0,1]$ is the date-t expectation of loss given default. Under the standard simplifying assumptions that r and q are *expected to be* constant over time, equation (1) implies that the one-year PD equals:

$$q_t = \frac{as_t}{aLGD_t + bs_t} \tag{2}$$

where $a \equiv \int_t^{t+T} e^{-r\tau} d\tau$ and $b \equiv \int_t^{t+T} \tau e^{-r\tau} d\tau$.

For a particular entity i, we use equation (2) in order to estimate a time series of risk-neutral one-year PDs, $q_{i,t}$, on the basis of time series of CDS spreads and expected LGDs for that entity and a time series of risk-free rates of return.

3.2 CDS-implied asset-return correlations

We map the estimated time series of risk-neutral PDs into time series of asset values. Then we use the latter time series to calculate CDS-implied asset-return correlations. The exact procedure is described in this section.

We start with the assumption that, under the risk-neutral measure, the asset value process of entity i is:

$$\frac{dV_{i,t}}{V_{i,t}} = \mu_i dt + \sigma_i dW_{i,t} \tag{3}$$

where μ_i denotes the drift, σ_i the asset volatility and $W_{i,t}$ a standard Wiener processes. Further, given a default boundary D_i, we define the distance to default as $DD_{i,t} \equiv \frac{lnV_{i,t} - lnD_i}{\sigma_i}$. By Ito's Lemma, $dDD_{i,t}$ has a drift $\mu_i^* = \frac{\mu_i - \sigma_i^2/2}{\sigma_i}$ and a unit variance.

In the spirit of the Merton (1974) framework, we postulate that entity i defaults τ years into the future if, at that time, its distance-to-default is below zero. The asset value process in (3) then implies that the probability of default is:

$$PD_{i,t}(\tau) = 1 - \Phi\left(\frac{DD_{i,t} + \tau\mu_i^*}{\sqrt{\tau}}\right) \tag{4}$$

where $\Phi(\cdot)$ stands for the standard normal CDF. Equipped with time series of PD estimates and setting the horizon $\tau = 1$ year, we calculate asset-return correlations as:

$$\rho_{ij} = corr\left(\Delta \ln V_{i,t}, \Delta \ln V_{j,t}\right) = corr\left(\Delta\Phi^{-1}\left(PD_{i,t}\right), \Delta\Phi^{-1}\left(PD_{j,t}\right)\right) \tag{5}$$

where Δ denotes the first difference in discrete time.

This procedure, which delivers CDS-implied asset return correlations, warrants several remarks. First, the procedure is underpinned by the Merton framework, which is also at the root of the equity-implied correlations delivered by Moody's KMV (see Section 3.3 below). Second, equation (3) assumes that the log of asset values, $ln(V_{i,t})$, follows a unit root process. This assumption is supported by a failure to reject the hypothesis that the time series of $DD_{i,t}$, constructed on the basis of (4), follow a unit root.[6] Third, since $\Delta ln(V_{i,t})$ stands

[6] More precisely, a battery of Phillips-Perron tests fail to reject the unit-root null for 132 of the 136 distance-to-default time series we construct. In addition, a unit root process provides a reasonable approximation to the dynamics in the remaining 4 series.

for an *actual* asset return, (5) delivers *statistical* asset-return correlations. We use these correlations to derive prices of portfolio credit risk, which, as a result, do not incorporate a correlation-risk premium. We will revisit this point in Section 5.3.3. Fourth, consistent with the no-correlation-risk-premium assumption, the estimated correlation coefficients, ρ_{ij}, are constant over time (recall expression (9)). Even though correlations could in principle change over time, we do not find evidence for this in our sample (see Section 5.3.3).[7]

Finally, we would like to think of (4) as a simple but rough mapping from PDs to asset returns. For one, this mapping is inconsistent with our construction of PDs – recall (2) – which assumes that the default density is expected to remain constant in the future. In addition, the mapping is predicated on Gaussian asset returns – recall (3) – while many of our results (below) are based on non-Gaussian distributions of these returns. Despite these inconsistencies, the mapping is likely to fulfill its only objective: ie lead to accurate estimates of asset-return correlations. This is suggested by the results of robustness checks, which establish that these correlations are largely insensitive to changes in (i) the time profile of expected default intensities (see Section 5.2 below) and (ii) the assumed distribution of asset returns.[8]

3.3 Equity-implied asset-return correlations

An alternative set of asset-return correlations can be estimated on the basis of equity market data. Moody's KMV provides such an estimate by using the proprietary GCorr model.

The GCorr correlations are calculated in three steps. The first step uses the Merton (1974) framework to extract *actual* asset values from data on equity prices and balance sheet information (see Crosbie and Bohn, 2002, for details). This step is carried out for each entity in the Moody's KMV universe. The second step estimates the entity-specific loadings of asset returns on 120 common factors, including 2 global economic factors, 5 regional economic factors, 7 sector factors, 61 industry-specific factors and 45 country-specific factors (see Das and Ishii, 2001; Crosbie, 2005). The third step delivers *statistical* asset-return correlations on the basis of the estimated factor loadings.

[7] A recent study by Daniels et al. (2005) also finds evidence that asset return correlations change little over time.

[8] The design and results of robustness checks that experiment with alternative distributions of asset returns are available upon request.

3.4 Common-factor models of asset returns

If asset returns have a joint Gaussian distribution, then PDs and asset-return correlations are sufficient in order to derive the probability distribution of joint defaults. Third and higher moments of asset returns come (independently) into play for more general distributional assumptions, which we introduce via a common-factor model. The model postulates that asset returns are driven by a number of random variables that are common to groups of firms in the reference portfolio as well as by entity-specific variables.[9] In this paper, we estimate common-factor models by fitting the CDS-implied and GCorr asset return correlations.[10,11]

The sole input to this estimation method is a correlation matrix of asset returns with entries ρ_{ij}, where i and $j \in \{1, \cdots, N\}$ and N is the size of the cross section. This matrix is a summary statistic of the joint distribution of asset returns, which is assumed to be underpinned by F common factors $M_t = [M_{1,t}, \cdots, M_{F,t}]'$ and N entity-specific, or idiosyncratic, factors $Z_{i,t}$:

$$\Delta ln(V_{i,t}) = A_i M_t + \sqrt{1 - A_i' A_i} \cdot Z_{i,t} \tag{6}$$

where $A_i \equiv [\alpha_{i,1}, \cdots, \alpha_{i,f}, \cdots, \alpha_{i,F}]$ is the vector of common factor loadings, $\alpha_{i,f} \in [-1, 1]$ and $\sum_{f=1}^{F} \alpha_{i,f}^2 \leq 1$. Without loss of generality, all common and idiosyncratic factors are assumed to be mutually independent and to have zero means and unit standard deviations. Note that, if (3) characterizes the true asset-value process, then (6) is a factorization of the shock $dW_{i,t}$.

We estimate the loading coefficients $\alpha_{i,f}$ $(i = 1, \cdots, N,\ f = 1, \cdots, F)$ by minimizing the mean squared difference between the factor-implied correlation and the target correlation:[12]

$$\min_{A_1 \cdots A_N} \sum_{i=2}^{N} \sum_{j<i}^{N} \left(\rho_{ij} - A_i A_j'\right)^2$$

Besides the "zero mean-unit variance" normalization, this estimation method imposes no restriction on the distribution of the common and idiosyncratic factors. In our empirical exercise, we exploit this fact and alter the assumed distribution of the common factors while preserving the "zero mean-unit variance" normalization as well as the estimated factor

[9] Collin-Dufresne et al. (2003), Das et al. (2006) and Giesecke (2004) discuss possible reasons why the assets of different firms may be driven by common factors.

[10] As noted in Section 3.3, the GCorr correlations are based on a framework with 120 common factors, which is estimated for all the firms in the Moody's KMV universe. In this paper, we work with parsimonious versions of this framework (using only *latent* common factors), and estimate them for a much smaller cross-section (136 firms).

[11] In Appendix A, we outline an alternative method based on the Kalman filter. The results of this method are virtually indistinguishable from those delivered by the factorizing algorithm described in this subsection.

[12] We follow the algorithm proposed in Andersen et al. (2003).

loadings and risk-neutral PDs. This allows us to study the impact of alternative third and higher moments of asset returns on the probability distribution of defaults.[13]

4 Data

The data we use can be divided into three blocks. In addition to these blocks, which are described in this section, we obtain 5-year Treasury rates from Bloomberg in order to proxy for the risk-free rate of return (Figure 1).

The first block of data is provided by JPMorgan Chase and pertains to 5-year contracts written on the CDS index Dow Jones CDX North America investment-grade index (CDX.NA.IG.5Y). These standardized contracts are highly liquid on the secondary market. We use single-tranche spreads for the "on-the-run" CDX.NA.IG.5Y index, as well as spreads for the equity, mezzanine and two senior tranches of the same index.[14] At each point in time the CDS index consists of 125 entities that represent major industrial sectors and are actively traded in the single-name CDS market as well. All entities have equal shares in the total notional principal of the index. The composition of the index is updated semi-annually – in a new release – in order to reflect events such as defaults, rating changes, and mergers and acquisitions. We consider three releases of the CDX.NA.IG.5Y index, launched respectively on 13 November 2003, 23 March 2004 and 21 September 2004. The total number of entities that appear in at least one of these releases is 136.

The second block of data pertains to the single-name CDS market and is provided by Markit, which has constructed a network of leading market participants that contribute pricing information across several thousand credits on a daily basis. Using the contributed quotes, Markit calculates CDS spreads for each credit in its database, as well as market expectations of entity-specific LGDs, at the daily frequency. In line with the contractual terms of the CDX.NA.IG.5Y index, we use time series of 5-year senior unsecured CDS spreads associated with the no-restructuring clause (see ISDA, 2003) and denominated in US dollars. We use CDS spreads from April 24, 2003 to September 27, 2005 (ie 634 business days) for all 136 entities that have belonged to at least one of the CDS index releases we consider.

The LGDs provided by Markit reflect market participants' consensus view on *expected* losses, and therefore need not match *realized* losses. The reported LGDs exhibit little cross-

[13]Appendix B.3 outlines how we incorporate alternative third and higher moments of asset returns in the derivation of the probability distribution of defaults.

[14]We abstract from the two super-senior tranches because the spreads on these tranches are likely to be affected substantially by non-credit factors, such as administrative costs and a liquidity premium. Although the analysis of such factors is important, it is beyond the scope of this paper.

sectional difference and time variation (see Table 1 and Figure 1). For the cross section of 136 time averages of LGDs, the 1st and 99th percentiles equal 60% and 63% respectively. In addition, the time series of cross-sectional averages of LGDs fluctuates within a similarly narrow band. In the light of this and in order to eliminate potential noise in the LGD data, we set LGDs to be the same across entities and smooth the resulting time series via an HP filter with a parameter $\lambda = 64000$.

The third block of data consists of asset-return correlations and is reported by Moody's KMV. These correlations are estimated on the basis of the proprietary GCorr model and are updated monthly. In this study, we use the March 2005 estimate of the GCorr correlations for the 136 firms studied in this paper.

5 Empirical findings

In this section, we report our empirical findings and use them to shed light on the pricing of portfolio credit risk. In Section 5.1, we present two alternative estimates of asset-return correlations (CDS-implied and GCorr) and their implications for tranche spreads under the assumption that asset returns are Gaussian. These preliminary results motivate a battery of robustness checks, which we report in Section 5.2. In Section 5.3, we analyze *observed* tranche spreads of the CDS index CDX.NA.IG.5Y by comparing them to several sets of *implied* spreads. These sets differ owing to the underlying estimate of asset-return correlations and/or the assumption regarding the third and higher moments of asset returns. Finally, in Section 5.3.3, we present evidence that the pricing of the CDS index incorporates a rather small correlation risk premium.

5.1 CDS-implied versus GCorr asset-return correlations

The CDS-implied and GCorr correlations – obtained on the basis of data from the single-name CDS and equity markets, respectively – turn out to be substantially different from each other. In concrete terms, the average CDS-implied correlation coefficient is 13%, whereas the average GCorr correlation coefficient is 24% (Table 1).

We establish that the difference between the two sets of asset-return correlations is *statistically* significant. This conclusion is based on a test that focuses on one entity in the sample at a time and compares (i) the mean of the CDS-implied correlations with the other entities to (ii) the mean of the corresponding GCorr correlations. Specifically, for each of the 136 entities, we start with two cross sections of 135 pairwise correlations and treat each cross section as produced by independent draws from some random variable. By the Central Limit

Theorem, the mean of each of these cross sections is itself a draw from an approximately normal distribution with readily estimable parameters. We calculate the 97.5th percentile of the distribution underlying the CDS-implied mean and the 2.5th percentile of the distribution underlying the GCorr-implied mean. It turns out that, for 126 out of the 136 entities, the former percentile is smaller than the latter. Therefore, we conclude that the average correlation coefficient implied by the single-name CDS spreads is significantly lower than the average GCorr correlation.

The statistically significant difference between the two sets of correlations has important *economic* implications. This is illustrated by the first row in Table 2, which reports the means of CDS- and GCorr-implied tranche spreads for the CDX.NA.IG.5Y index. The two alternative sets of implied spreads differ only owing to the underlying correlation estimates. At the same time, these two sets share the underlying assumptions that asset returns are Gaussian and, in the spirit of the Merton framework, a default can occur only at a particular point in time.[15] The lower CDS-implied asset-return correlations imply a smaller likelihood of default clustering and, as discussed in Section 2 and illustrated in Figure 2, lead to higher (lower) spreads for the equity (senior) tranches. In particular, the average CDS-implied equity spread equals 1,856 basis points, which is 18% higher than the corresponding GCorr-implied spread. In addition, the CDS-implied spreads for the two senior tranches equal 70 and 16 basis points and are, on average, 38% and 62% lower than their GCorr-implied counterparts.[16]

5.2 The difference between implied spreads: A robustness check

The difference between the CDS-implied and GCorr asset-return correlations suggests inconsistency between the single-name CDS and equity markets. The inconsistency implies that choosing from which market to extract asset-return correlation estimates has material consequences for prices of portfolio credit risk. This section demonstrates that these findings change little under alternative estimations of asset-return correlations and alternative mappings of these correlations into tranche spreads. Inter alia, the section reveals that the difference between the CDS- and GCorr-implied spreads is due predominantly to the difference between the *averages* of the underlying correlation coefficients, while the dispersion in these coefficients is of second order. We use this conclusion as a starting point of our subsequent discussion of how the market prices portfolio credit risk.

[15]Recall the discussion leading to equation (4). See Appendix B.1 for further detail.

[16]The impact of correlation on the spread of the intermediate - mezzanine - tranche is, in principle, ambiguous because this tranche shares characteristics of both the equity and senior tranches.

5.2.1 Alternative CDS-implied correlations

Even though we need to take GCorr asset-return correlations as given, there are several ways in which we could estimate CDS-implied correlations. The alternatives may differ owing to the mapping from CDS spreads to PDs or from PDs to asset values. We examine two such alternatives, described below, and find virtually no change in CDS-implied correlations.[17]

For an alternative mapping from CDS spreads to PDs, we follow reported market practice in the context of investment-grade entities. Namely, we replace expression (2) by the following approximation of the default intensity $q_t = \frac{s_t}{LGD_t}$. This changes negligibly our PD estimates and, recalling Section 3.2, our estimates of asset values and CDS-implied correlations.

For an alternative mapping from PDs into asset values and then into CDS-implied asset-return correlations, we allow defaults to occur at any point in time. This modifies (4), which assumed that a default may occur only at a specific point in time. Namely, the probability that entity i defaults over the next τ years becomes:

$$PD_{i,t}\left(\tau; DD_{i,t}, \mu_i^*\right) = 1 - \Phi\left(\frac{DD_{i,t} + \tau\mu_i^*}{\sqrt{\tau}}\right) + \exp\left(-2\tau\mu_i^*\right)\Phi\left(\frac{-DD_{i,t} + \tau\mu_i^*}{\sqrt{\tau}}\right) \qquad (7)$$

We use equation (7) and our CDS-implied estimates of risk-neutral PDs, $q_{i,t}$, in order to derive alternative time series of the distance-to-default variable $DD_{i,t}$. Specifically, we solve the following system of equations for $DD_{i,t}$ and μ_i^*:

$$\begin{aligned} PD_{i,t}\left(1; DD_{i,t}, \mu_i^*\right) d\tau &= q_{i,t} \\ \frac{1}{5} PD_{i,t}\left(5; DD_{i,t}, \mu_i^*\right) d\tau &= q_{i,t} \end{aligned} \qquad (8)$$

Thus, the values of $DD_{i,t}$ and μ_i^* imply a 1-year PD and an average default intensity (over 5 years) that are equal to each other and to the one-year risk-neutral PD, $q_{i,t}$, estimated from single-name CDS spreads.[18]

Mimicking (5) above, we calculate the correlation of asset *returns* between entities i and j as:

$$\rho_{ij} = corr\left(\Delta ln(V_{i,t}), \Delta ln(V_{j,t})\right) = corr\left(\Delta DD_{i,t}, \Delta DD_{j,t}\right) \qquad (9)$$

This alternative mapping method changes little our correlation estimates. It delivers pairwise correlation coefficients that average 13.02% and have a standard deviation of 9.89%, compared with an average of 13% and a standard deviation of 10.28% for the original esti-

[17]The exact results are available upon request.

[18]Recall that the time to maturity of both CDS and CDS-index contracts is 5 years.

mates. The mean absolute difference between alternative and original correlation coefficients is merely 0.0089, which has a negligible effect on derived tranche spreads.

5.2.2 Alternative mappings into tranche spreads

For given correlation estimates, the differences between CDS- and GCorr-implied tranche spreads could be an artefact of the numerical method assumed for the pricing process. This subsection examines whether these differences – reported in the first row of Table 2 – could paint a misleading picture if market participants have in fact adopted alternative numerical methods. The alternatives we examine are generated by: (i) fixing our PD and correlation estimates at their averages in the cross section, or (ii) a simulation of defaults that may occur at any point in time over a given horizon, or (iii) a common-factor approximation to the joint (Gaussian) distribution of asset returns, or (iv) non-Gaussian third and higher moments of asset returns. To anticipate the results, all of these alternatives change little the differences between CDS- and GCorr-implied tranche spreads reported in Section 5.1.

When pricing CDS index tranches, markets reportedly treat all constituent entities as homogeneous, ie as having the same PD and as exhibiting identical correlation coefficients. Table 1 illustrates that entities in the CDS index CDX.NA.IG.5Y are in fact not homogeneous, as the standard deviation in the cross section of PDs equals 50 basis points and correlation coefficients vary between –0.57 and 0.80 (CDS-implied) and between 0.05 and 0.65 (GCorr). Nevertheless, we do examine the pricing implications of assuming away the cross-sectional dispersion in PD and/or correlation estimates. Such an exercise reveals how CDS- and GCorr-implied tranche spreads may be affected by calculation "shortcuts". Indeed, the analysis in Hull and White (2004) implies that the effect could be substantial.

The results, summarized in Table 2 (rows 3-5), indicate that assuming homogeneous entities has only a limited impact on implied tranche spreads. Importantly, the difference between CDS- and GCorr-implied spreads tends to change little and actually increases when the dispersion in PDs is removed.

Our second exercise relates to the simulation of defaults for the calculation of their probability distribution. The results we have reported so far are based on a simulation in which defaults can occur only at a specific point in time (Appendix B.1). Alternatively, however, a default can be simulated in a *multi-period* setting under the assumption that it is triggered the first time asset values cross a given threshold.[19] As observed by Duffie and Singleton (2003), this alternative specification may lead to different probabilities of joint defaults and, thus, different prices of portfolio credit risk.

[19] This would be in line with the scenario that leads to the PD in equation (7).

We carry out a multi-period simulation of defaults – described in Appendix B.2 – and report the results in Table 2 (row 6).[20] Even though such a simulation does change noticeably the CDS- and GCorr-implied tranche spreads – especially for the mezzanine and senior tranches – it changes little the quantitative difference between the two sets of spreads.

Our third exercise is motivated by market commentary, which refers regularly to common factor models of asset returns. We investigate whether the use of such a model for pricing purposes could have a material impact on implied tranche spreads. We start by estimating a one-factor structure of the CDS-implied and GCorr correlations (as explained in Section 3.4) and then employ this structure in a Gaussian copula to derive tranche spreads (Appendix B.3). As reported in Table 2 (row 7), the one-factor model leaves implied tranche spreads virtually unchanged. Allowing for more common factors leads to even smaller changes.[21]

Our final robustness exercise examines the impact of alternative third and higher moments of asset returns on CDS- and GCorr-implied tranche spreads. This exercise is based on the one-factor model, which provides a good approximation to our estimates of asset-return correlations and allows us to change the moments assumption while keeping PDs and asset-return correlations fixed (recall Section 3.4). Specifically, we perturb our baseline scenario of Gaussian asset returns by assuming that these returns are driven by (i) a Student-t distributed common factor, or (ii) Student-t idiosyncratic and common factors, or (iii) a common factor driven by a mixture of three Gaussian distributions.[22]

Table 2 (rows 8-10) illustrates the implications of alternative moments assumptions. Even though (some of) these alternatives change materially the CDS- and GCorr-implied tranche spreads, they do not affect the difference between the two sets of spreads. Irrespective of the third and higher moments of asset returns, the lower CDS-implied correlations lead to substantially higher (lower) spreads for the equity (senior) tranches than the GCorr correlations.

[20] For these simulations, we generate 10 intra-day observations for a total of 13200 observations in 5 years. Owing to the computational burden, we calculate the tranche spreads every 20 business days during the period between November 21, 2003 and March 18, 2005.

[21] The one-factor model of the GCorr correlations, which matches exactly the mean and standard deviation of pairwise correlations, has a mean squared error (MSE) of 0.0177. By comparison, a one-factor model of the CDS-implied correlations matches the mean but under-estimates the standard deviation of correlation coefficients, leading to an MSE of 0.08. Adopting a multi-factor model improves the match marginally, eg the latter MSE drops to 0.06 under a three-factor model.

[22] The PDF of the mixture of three Gaussian variables is $0.32\phi(-3,8) + 0.50\phi(1,1) + 0.18\phi(0,1)$, where $\phi(\mu,\sigma)$ represents the PDF of a normally distributed variable with a mean of μ and a standard deviation of σ. As seen below, our choice of a particular mixture of Gaussian variables leads to CDS-implied tranche spreads that closely match the data.

5.3 How does the market price portfolio credit risk?

A natural question to ask is whether the CDS- and GCorr-implied tranche spreads can reveal information as to how markets determine the tranche spreads observed in the data. A partial answer to this question is provided in Section 5.3.1, which demonstrates that the PD estimates derived from the single-name CDS market closely match the PD estimates used for pricing in the CDS index market. The question is explored further in Section 5.3.2, which investigates which of the two sets of asset-return correlations (CDS-implied or GCorr) is used in the index market. Unfortunately, a definitive answer is impossible without further knowledge about the third and higher moments of asset returns. Finally, a by-product of the analysis, discussed in Section 5.3.3, reveals that the market for the CDS index CDX.NA.IG.5Y exhibits a rather small correlation risk premium.

5.3.1 Market estimates of PDs

An important factor in the pricing of portfolio credit risk is the level of default risk of the constituent entities, as captured by individual PDs. A general rise in individual PDs signifies higher credit risk at the portfolio level as well, which raises the spreads for all index tranches (see Figure 3). Given that our PD estimates, which are used for constructing both the CDS- and GCorr-implied tranche spreads, are derived from the single-name CDS market, a first question is whether they are different from the PD estimates used in the CDS index market.

Even though we are not able to pin down the entity-specific PDs used by market participants in the CDS index market, we extract the cross-sectional averages of these PDs. This is done on the basis of a time series of spreads for the overall index (ie, single-tranche spreads), which reveal expected credit losses, and our data on market expectations of LGDs. The average (risk-neutral) PDs implied by the single-tranche spreads are plotted in Figure 4 alongside their analogs implied by the single-name CDS spreads. The two time series differ on average by only 1.22 basis points, which is roughly 1.4% of the average PD implied by single-name CDS spreads. Importantly, this difference has negligible pricing implications: for each of the tranches considered, it shifts implied spread by less than 2% on average over time (Table 2, row 2). Thus, it appears that the PD estimates used in the index market are consistent with the information embedded in the single-name CDS market.

Moreover, the close match between the two series of average PDs drives the similarity of the time paths of implied and observed tranche spreads. This is observed in both Figures 5 and 6 and backed by regression analysis. In particular, we first regress CDS-implied tranche spreads on a constant and average PDs and obtain goodness-of-fit measures (ie adjusted R^2) in the range of 90-99%. Likewise, the levels of the adjusted R^2 are in the range of 95-99%

if GCorr-implied tranche spreads are used as the dependent variable. This indicates that the average PD series is the sole determinant of time variation in implied spreads, with the other time-varying determinants – LGDs and the risk free rate – playing a negligible role. Then, we regress observed tranche spreads on a constant and CDS-implied spreads and obtain goodness-of-fit measures that can be as high as 86% for the equity tranche and attain their lowest level of 65% for the less risky senior tranche (Table 3). Replacing CDS-implied tranche spreads with GCorr-implied spreads yields virtually the same goodness-of-fit results.

5.3.2 Matching observed spreads of the CDS index

Given the substantial difference between asset-return correlations implied, respectively, by the CDS and equity markets (see Section 5.1), we proceed with the question: Which of the two sets of correlations is used by market participants for the pricing of CDS index tranches? Unfortunately, the available data lead to an ambiguous answer, as they reveal insufficient information about the joint (risk-neutral) distribution of asset returns. In particular, the data induce us to make assumptions about the third and higher moments of these returns. By choosing these assumptions judiciously, we can reconcile observed spreads of the CDS index with either CDS- or GCorr-implied spreads. At the same time, it is impossible to identify a single third-and-higher-moments assumption that would reconcile *both* sets of implied spreads with observed spreads.

To be more specific, under the assumption that asset returns are Gaussian, the GCorr correlations, which are based on equity market data, help to explain observed tranche spreads. This is illustrated by Figure 5, which plots time series of these spreads alongside CDS- and GCorr-implied spreads. On average over time, GCorr-implied spreads deviate from the corresponding observed spreads by less than 9% for all tranches. By contrast, the CDS-implied spreads, driven by lower asset-return correlations, undershoot substantially (by 43 and 30 basis points, or 38% and 66%) the observed spreads for the two senior tranches.

The picture is quite different if asset returns are driven by a common factor that is distributed as a mixture of three Gaussian variables (see Figure 6). Such a mixture gives us substantial flexibility and we choose a distribution of the common factor that has a long left tail (skewness of -1.1, see Figure 7, top panel).[23] This alters the third and higher moments of asset returns in such a way that, in comparison to the purely Gaussian case, defaults are more likely to be driven by the common factor. As a result, keeping asset-return correlations fixed, defaults are more likely to be clustered (see Figure 7, middle and bottom

[23] As demonstrated by Geweke and Keane (1999), a mixture of normals is flexible enough to capture a wide range of distributional features.

panels). Alternatively, the new distribution of asset returns captures a high market value of low-probability catastrophic events, in which a large fraction of the reference portfolio defaults.

The alternative distribution of the common factor of asset returns lowers (raises) the spreads for the equity (senior) tranches, leading to smaller differences between CDS-implied and observed spreads. As shown in Figure 6 and Table 2, these differences drop to 6% or less on average over time. By contrast, combining this negatively skewed distribution of the common factor of asset returns with the high GCorr correlations of these returns implies default clustering that is too high to be consistent with observed spreads of the CDS index. Indeed, GCorr-implied spreads undershoot observed ones by 27% for the equity tranche and overshoot by 42% and 90% for the two senior tranches.

The above results reveal that, irrespective of whether CDS-implied or GCorr correlations are used by market participants, there is inconsistency between a single-name market and a market for portfolio credit risk. Under the first scenario, with Gaussian asset returns, there is inconsistency in the way the single-name CDS and CDS-index markets incorporate asset-return correlations. By contrast, such inconsistency exists between the equity and credit markets under the second scenario, which considers asset returns driven by a common factor with a negatively skewed distribution.[24] Either scenario provides evidence for market segmentation in the pricing of portfolio credit risk, which is not arbitraged away.

5.3.3 Is there a correlation risk premium?

For the above analysis, we kept asset-return correlations constant over time and, in a consistent manner, assumed that there is no correlation risk premium – ie a premium for the risk that asset-return correlations will change over time.[25] A priori, such an assumption need not be valid. In fact, Driessen et al. (2005) find strong evidence of a correlation risk premium in the options market. As this section demonstrates, however, there is little evidence for such a premium in the CDS index market.

Assuming that a correlation-risk premium exists, we need to revise all CDS- and GCorr-implied spreads upward. If this revision leads to an improved match between implied and observed spreads, then we have evidence for a correlation risk premium. We carry out such revisions for Gaussian asset returns.[26]

[24] Notice that if asset returns are Student-t-distributed, neither of the implied tranche spreads is able to match observed spreads, suggesting inconsistency across the three markets.

[25] We acknowledge that this claim largely depends on the validity of the assumptions underlying our estimation of asset-return correlations. If these assumptions are violated, the correlations may be polluted by a risk premium.

[26] The distribution of returns is immaterial for the conclusions in this section.

At first glance, the results are mixed. First, recalling Figure 5, CDS-implied tranche spreads overshoot observed ones for the equity tranche but undershoot for senior tranches. Thus, incorporating a correlation risk premium in CDS-implied tranche spreads would improve the match for senior tranches, but would worsen this match for the equity tranche. Second, recalling Figure 5 again, GCorr-implied spreads undershoot observed ones for all tranches when asset returns are assumed to be Gaussian. This is *qualitative* evidence for a correlation risk premium.

In *quantitative* terms, however, a rather small such premium is needed to improve the match between GCorr-implied and observed spreads. To substantiate this statement we calculate the average (risk-neutral) correlation implied by observed spreads of individual tranches, and compare this correlation to its GCorr counterpart.[27] We find that, for the equity and two senior tranches, the implied correlation coefficients are 20%, 24%, and 26%, respectively. The difference between these values and the average GCorr correlation of 24% provides a measure of the correlation risk premium, which is much smaller than the analogous 18-percentage-point difference derived by Driessen et al. (2005) on the basis of option prices.

One reason why spreads of the CDS index may not contain a correlation risk premium is because traders do not expect asset-return correlations to change over time. Evidence in support of this explanation is provided by the regression results reported in Table 3. To understand this evidence, note first that, if market perceptions of asset-return correlations changed over time, then this would affect tranche spreads over and above the effect of changes in PDs. The effect of these perceptions on tranche spreads would then be embedded in the residuals obtained from regressing tranche spreads on PDs. Second, as demonstrated by Figure 2, changes in (the perceptions of) asset-return correlations tend to affect the spreads of the equity tranche and those of senior tranches in opposite directions. Thus, to the extent that the correlation estimates used in the index market did change over time and these changes had a significant pricing impact, we should observe negative correlations between the residuals in the "equity tranche" and "senior tranche" regressions. However, as reported in Table 3, this is not the case: the correlations of regression errors across tranches are

[27] Several clarifications are in order. First, when calculating the implied average correlation, we use the same dispersion of asset-return correlations, as well as the same PDs, LGDs and risk-free rates as those underlying GCorr-implied spreads. Thus, our implied correlation differs from the one calculated by market practitioners under the assumption that the PDs and correlation coefficients do not change in the cross section. Our background analysis reveals that this market practice biases the estimates, causing an artificial "correlation smile" phenomenon. Second, as portrayed in Figure 2, the equity (senior) tranche spreads decrease (increase) in the asset-return correlation. Therefore, a correlation risk premium should lead to a *risk-neutral* correlation for the equity (senior) tranche that is lower (higher) than the statistical correlation. Third, the non-monotone relationship between average correlation and the mezzanine tranche spread leads to a non-informative correlation for this tranche.

invariably positive and highly significant (the lowest correlation coefficient is 65%).

6 Conclusion

This paper demonstrates that asset-return correlations and distributional features pertaining to the third and higher moments of asset returns play equally important roles in the pricing of portfolio credit risk. In addition, the paper finds evidence that there is inconsistency in the implications of equity and single-name CDS markets for asset-return correlations. This creates ambiguity as regards the way the market prices tranches of a popular CDS index. Asset-return correlations based on equity market data are reconciled with observed tranche spreads if the third and higher moments of asset returns are Gaussian. By contrast, in order to reconcile observed tranche spreads with asset-return correlations implied by the single-name CDS market, it is necessary to allow for a negative skew in the distribution of the common factor of asset returns. These results call for further research effort towards understanding market estimates of the correlation as well as the third and higher moments of asset returns.

References

Altman, Edward and Vellore Kishore (1996), "Almost everything you want to know about recoveries on default bonds," *Financial Analysts Journal*, vol. 52, 57–64.

Andersen, Leif and Jakob Sidenius (2005), "Extensions to the Gaussian colupa: random recovery and random factor loadings," *Journal of Credit Risk*, vol. 1, 29–70.

Andersen, Leif, Jakob Sidenius, and Susanta Basu (2003), "All your hedges in one basket," *Risk*, vol. 16, 67–72.

BCBS (2004), "Credit Risk Transfer," *Basel Committee on Banking Supervision Joint Forum*.

Blanco, Roberto, Simon Brennan, and Ian W. March (2005), "An empirical analysis of the dynamic relationship between investment-grade bonds and credit default swaps," *Journal of Finance*, vol. 60, 2255–2281.

Collin-Dufresne, Pierre, Robert Goldstein, and Jean Helwege (2003), "Is credit event risk priced? Modeling contagion via the updating of beliefs," *Working Paper*, Carnegie Mellon University.

Covitz, Dan and Song Han (2004), "An empirical analysis of bond recovery rates: exploring a structural view of default," *Federal Reserve Board Working Paper*.

Crosbie, Peter (2005), "Global correlation factor structure: modeling methodology," *Moody's KMV Documents*.

Crosbie, Peter and Jeffrey Bohn (2002), "Modeling default risk," *KMV White Paper*.

Daniels, Robert, Siem Jan Koopman, and Andrée Lucas (2005), "A non-Gaussian panel time series model for estimating and decomposing default risk," *Working Paper*.

Das, Ashish and Shota Ishii (2001), "Methods for calculating asset correlations: a technical note," *KMV Documents*.

Das, Sanjiv, Darrell Duffie, Nikunj Kapadia, and Leandro Saita (2006), "Commonn failings: how corporate defaults are correlated," *Journal of Finance*.

Demey, Paul, Jean-Frédéric Jouanin, Céline Roget, and Thierry Roncalli (2004), "Maximum likelihood estimate of default correlations," *Risk*, vol. 17, 104–108.

Driessen, Joost, Pascal Maenhout, and Grigory Vilkov (2005), "Option-implied correlation and the price of correlation risk," *Working Paper*.

Duffie, Darrell (1999), "Credit swap valuation," *Financial Analysts Journal*, vol. 55, 73–87.

Duffie, Darrell and Kenneth Singleton (2003), *Credit Risk: Pricing, Measurement, and Management*, Princeton University Press.

Geweke, John and Michael Keane (1999), *Mixture of normals Probit models*, Cambridge University Press, Cambridge.

Gibson, Michael (2004), "Understanding the risk of synthetic CDOs," *Finance and Economics Discussion Series*.

Giesecke, Ikay (2004), "Correlated default with incomplete information," *Journal of Banking and Finance*, vol. 28, 1521–1545.

Hamilton, James (1994), *Time Series Analysis*, Princeton University Press.

Hull, John and Alan White (2004), "Valuation of a CDO and an n-th to default CDS without Monta Carlo simulation," *Journal of Derivatives*, vol. 12, 8–23.

ISDA (2003), "International Swaps and Derivatives Association (ISDA) credit derivatives defintions," *Supplements and Commentaries*.

Jarrow, Robert (2001), "Default parameter estimation using market prices," *Financial Analysts Journal*, vol. 57, 75–92.

Jarrow, Robert and Donald van Deventer (2005), "Estimating default correlations using a reduced-form model," *Risk*, vol. 18, 83–88.

Laurent, Jean-Paul and Jon Gregory (2005), "Basket default swaps, CDOs and factor copulas," *Journal of Risk*, vol. 7, 103–122.

Li, David (2000), "On default correlation: a copula approach," *Journal of Fixed Income*, vol. 9, 43–54.

Longstaff, Francis, Sanjay Mithal, and Eric Neis (2005), "Corporate yield spreads: default risk or liquidity? New evidence from the credit-default-swap market," *Journal of Finance*, vol. 60, 2213–2253.

Longstaff, Francis and Arvind Rajan (2006), "Empirical analysis of the pricing of collateralized debt obligations," *Working Paper*.

Merton, Robert (1974), "On the pricing of corporate debt: the risk structure of interest rates," *Journal of Finance*, vol. 29, 449–470.

Zhou, Chunsheng (2001), "An analysis of default correlation and multiple defaults," *Review of Financial Studies*, vol. 14, 555–576.

Zhu, Haibin (2006), "An empirical comparison of credit spreads between the bond market and the credit default swap market," *Journal of Financial Services Research*, vol. 29, 211–235.

Appendix

A The Kalman filter setup

To implement a Kalman filter, we use time series of asset returns, which are denoted by $\Delta ln(V_{i,t})$. The setup is summarized in the following three equations:

$$\Delta ln(V_{i,t}) = H\xi_t \tag{10}$$

$$\xi_t = F\xi_{t-1} + \upsilon_t \tag{11}$$

$$E(\upsilon_t \upsilon_t') = Q \tag{12}$$

where

$$\Delta v_t \equiv [\Delta ln(V_{1,t}) \cdots \Delta ln(V_{N,t})]', \quad \xi_t \equiv [M_{1,t} \cdots M_{F,t} \;\; Z_{1,t} \cdots Z_{N,t}]'$$

$$\upsilon_t \equiv [\eta_{1,t} \cdots \eta_{F,t} \;\; Z_{1,t} \cdots Z_{N,t}]' \quad \text{is a vector of standard normal variables}$$

$$H \equiv \begin{bmatrix} \alpha_{1,1} & ... & \alpha_{1,F} & \sqrt{1 - A_1' A_1} & 0 & 0 \\ ... & ... & ... & 0 & ... & 0 \\ \alpha_{N,1} & ... & \alpha_{N,F} & 0 & 0 & \sqrt{1 - A_N' A_N} \end{bmatrix}$$

where $A_i \equiv [\alpha_{i,1}, \cdots, \alpha_{i,f}, \cdots, \alpha_{i,F}]$

$$F \equiv \begin{bmatrix} \psi_1 & 0 & 0 & \\ 0 & ... & 0 & \mathbf{0}_{F\times N} \\ 0 & 0 & \psi_F & \\ & \mathbf{0}_{N\times F} & & \mathbf{0}_{N\times N} \end{bmatrix} \text{ and } Q \equiv \begin{bmatrix} 1-\psi_1^2 & 0 & 0 & \\ 0 & ... & 0 & \mathbf{0}_{F\times N} \\ 0 & 0 & 1-\psi_F^2 & \\ & \mathbf{0}_{N\times F} & & \mathbf{I}_N \end{bmatrix}$$

The maximum-likelihood estimation of the common factor loadings $\alpha_{i,f}$ ($i = 1, \cdots, N$, $f = 1, \cdots, F$) directly implies the pairwise correlation coefficients $corr(\Delta ln(V_{i,t}), \Delta ln(V_{j,t})) = A_i A_j'$. These parameters are delivered as part of the joint estimation of the matrices H, F and Q and the unobserved factors ξ_t.

In carrying out this estimation, we start with two preliminary steps:

1. Standardize each time series of asset returns. In other words, for each time series we first de-mean and then divide by the sample standard deviation $\{\Delta v_{i,t}\}_{t=1}^{T}$.

2. Ensure that the estimated loading coefficients belong to the interval $[-1, 1]$, ie para-

meterize H as follow ($i = 1, \cdots, N$, $j = 1, \cdots, F$):

$$
\begin{aligned}
\alpha_{i,1} &= 2\Phi(l_{i,1}) - 1 \\
\alpha_{i,2} &= \sqrt{1 - \alpha_{i,1}^2} \cdot [2\Phi(l_{i,2}) - 1] \\
&\cdots \\
\alpha_{i,F} &= \sqrt{1 - \alpha_{i,1}^2 - \cdots - \alpha_{i,F-1}^2} \cdot [2\Phi(l_{i,F}) - 1] \\
\psi_j &= 2\Phi(b_j) - 1
\end{aligned}
$$

Then we follow Hamilton (1994) to derive the conditional distribution:

$$\Delta v_t | \Delta v_{t-1} \sim N(H\hat{\xi}_{t|t-1}, HP_{t|t-1}H')$$

where $P_{t|t-1} = (F - K_{t-1}H)P_{t-1|t-2}(F' - H'K'_{t-1}) + Q$, in which the gain matrix $K_{t-1} \equiv FP_{t-1|t-2}H'(HP_{t-1|t-2}H')^{-1}$, and $\hat{\xi}_{t|t-1}$ is a linear function of Δv_{t-1}. Thus, the log likelihood function to maximize is

$$
\begin{aligned}
\max_{\{H,F,Q\}} & \quad \sum_{i=1}^{N}\sum_{t=1}^{T} log \;\; f(\Delta v_t|\Delta v_{t-1}) \\
f(\Delta v_t|\Delta v_{t-1}) &\equiv (2\pi)^{-n/2}|HP_{t|t-1}H'|^{-1/2} \\
& \times exp\{-\frac{1}{2}(\Delta v_t - H\hat{\xi}_{i,t|t-1})'(HP_{t|t-1}H')^{-1}(\Delta v_t|\Delta v_{t-1} - H\hat{\xi}_{i,t|t-1})\}
\end{aligned}
$$

B Estimating the probability distribution of defaults

This appendix outlines different methods for estimating the probability distribution of joint defaults in a given portfolio. All methods rely on three inputs: (i) PDs of individual entities; (ii) asset-return correlations across entities; (iii) third and higher moments of the joint distribution of entities' assets. Two of the methods impose the restriction that asset returns are Gaussian. The first of these methods assumes that a default can occur only at a particular point in time, whereas the second one allows for a default to occur at any point in time prior to the maturity of the relevant debt contract. The third method, which allows for more general distributional assumptions, postulates that a default can occur only at a particular point in time, relies on a common-factor model of asset returns and employs a copula framework.

B.1 One-period simulation of defaults

This method estimates the probability distribution of defaults in a portfolio of N entities when a default is driven by a single draw of a Gaussian random variable. The method relies on estimates of asset-return correlations and PDs and is in the spirit of the estimation of asset-return correlations outlined in Section 5.2.1.

1. Generate N random draws x_0 from independent standard normal distributions.
2. Calculate $x = R'x_0$, where R denotes the Cholesky factor of the estimated asset-return correlation matrix for the N entities.
3. Denoting the i-th member of x by x_i ($i = 1, \cdots, N$) and the associated PD by PD_i, entity i is said to default if and only if $x_i < \Phi^{-1}(PD_i)$.
4. Repeat steps 2 to 4 a large number of times to estimate the probability of $n \in \{0, \cdots, N\}$ defaults.

B.2 Multi-period simulation of defaults

We also calculate the probability distribution of defaults when they are allowed to occur at any point in time over a particular period. Specifically, for each entity in a portfolio, we simulate time paths of the distance to default variable – $DD_{i,t}$, defined in Section 3.2 – on the basis of our estimates of the drift μ_i^*. The initial value in each time path is based on equation 7 and thus relies on a PD estimate. For a given time period, each simulation provides a number of defaults: ie the number of i for which $DD_{i,t}$ falls below zero over the time period. Then, Monte Carlo repetitions provide an estimate of the probability distribution of the number of defaults.

B.3 Copula

This appendix outlines the copula method, which relies on a common-factor model of asset returns and has been developed by Li (2000), Laurent and Gregory (2005) and Andersen and Sidenius (2005). We assume that a default is driven by a single draw of asset returns, which are driven by a single common factor. We further impose the following distributional assumptions: (i) all factors have a zero mean and a unit variance, (ii) the idiosyncratic factors are Gaussian or Student-t with 5 degrees of freedom, (iii) the distribution of the common factor is either Gaussian, Student-t with 5 degrees of freedom, or a standardized mixture of three Gaussian variables (see Section 5.2.2).

Denoting the common and idiosyncratic factors, the loading coefficient on the common factor and the unconditional PD by M_t, $Z_{i,t}$, α_i and $PD_{i,t}$, respectively, the joint default probability can be calculated in three steps. The first step is to calculate the conditional default probability for entity i on date t, $PD(i|M_t)$. When the asset value $V_{i,t} = \alpha_i M_t + \sqrt{1 - \alpha_{i,t}^2} \cdot Z_{i,t}$, this probability equals:

$$PD(i|M_t) = G\left(\frac{F^{-1}(PD_{i,t}) - \alpha_i M_t}{\sqrt{1 - \alpha_i^2}}\right) \tag{13}$$

where G and F are the CDFs of $Z_{i,t}$ and $V_{i,t}$, respectively, which need to generated by Monte Carlo simulations if the variables are non-Gaussian.

The second step is to calculate the conditional probability of an arbitrary number of defaults. Suppose we know the probability of $k \in \{0, 1, ..., K\}$ defaults in a portfolio of K entities: $p^K(k|M_t)$. Then, adding one more entity leads to the following update of the distribution of defaults:

$$\begin{aligned}
p^{K+1}(0|M_t) &= p^K(0|M_t)(1 - PD(K+1|M_t)) \\
p^{K+1}(k|M_t) &= p^K(k|M_t)(1 - PD(K+1|M_t)) \\
&\quad + p^K(k-1|M_t)PD(K+1|M_t) \quad \text{for } k = 1, \cdots, K \\
p^{K+1}(K+1|M_t) &= p^K(K|M_t)PD(K+1|M_t)
\end{aligned}$$

This recursion starts with the initial condition $p^0(0|M_t) = 1$.

The final step is to calculate the unconditional probability of k defaults in a portfolio of N entities:

$$p^N(k,t) = \int_{-\infty}^{\infty} p^N(k|M_t)\varphi(M_t)dM_t$$

where φ is the PDF of M_t.

Table 1: **Summary statistics of LGDs, PDs and correlation coefficients**

	mean	*std dev*	*min*	*5%*	*25%*	*50%*	*75%*	*95%*	*max*
LGD (%)									
Daily averages	61.6	0.9	60.3	60.3	60.5	61.7	62.3	62.7	63.6
Averages over time	61.6	0.7	59.0	60.5	61.1	61.5	61.9	62.7	63.7
PDs (bps)									
Daily averages	85.3	12.6	63.7	65.4	74.6	87.9	97.5	102.1	105.4
Averages over time	85.3	50.0	23.5	34.9	58.4	71.2	91.1	216.9	281.1
pairwise correlations									
CDS-implied	0.130	0.099	–0.569	–0.016	0.068	0.124	0.184	0.301	0.796
GCorr	0.238	0.077	0.046	0.131	0.186	0.228	0.278	0.379	0.650

Note: The summary statistics reflect all 136 entities that belonged to any of the first three releases of the CDS index CDX.NA.IG.5Y. For LGDs and PDs, the first row is based on a time series of *daily* cross-sectional averages, whereas the second row is based on a cross section of time averages.

Table 2: **Implied tranche spreads**

	CDS-implied				**GCorr-implied**			
	0-3%	*3-7%*	*7-10%*	*10-15%*	*0-3%*	*3-7%*	*7-10%*	*10-15%*
baseline	1856.3	313.1	69.9	15.9	1572.8	330.4	112.4	42.2
adjust the level of PDs	1886.0	319.0	71.1	16.1	1597.2	335.5	114.4	43.0
remove the dispersion in PDs	1860.2	322.4	75.7	18.3	1499.1	343.2	128.4	52.6
remove the dispersion in correlations	1948.4	299.7	56.6	11.0	1557.9	341.1	113.6	40.8
remove the dispersion in								
PDs and correlations	1915.5	316.8	67.0	14.6	1501.9	353.4	127.2	49.2
multi-period simulation of defaults	1836.4	244.6	42.7	7.9	1568.4	273.7	79.9	26.2
one-factor	1898.1	302.9	65.8	14.9	1577.9	328.8	111.9	42.1
t-copula								
(5,5)	2050.9	230.9	49.4	19.1	1748.2	245.0	77.7	37.5
(5,∞)	1866.6	263.0	63.8	23.6	1541.0	279.8	99.0	47.2
mixture of normals	1656.6	287.0	112.3	44.1	1245.0	298.7	158.0	86.5
memo:								
observed tranche spreads	*1705.4*	*303.9*	*111.1*	*45.5*	*1705.4*	*303.9*	*111.1*	*45.5*

Notes: Unless noted explicitly, the reported averages are based on daily implied tranche spreads, calculated for the period from November 21, 2003 to March 18, 2005 (369 business days). The "baseline" results assume that asset returns are Gaussian and rely on the one-period simulation of defaults (see Appendix B.1). The next five lines of results reflect variations on the baseline scenario and are obtained by: (1) adjusting the level of individual PDs in order to ensure that the cross-sectional average PD equals the PD implied by the single-tranche index spread on each day; (2) removing the dispersion in PDs on each day; (3) removing the dispersion in correlation coefficients on each day; (4) removing the dispersion in both PDs and correlation coefficients on each day; and (5) using a multi-period simulation of defaults to obtain monthly implied spreads (Appendix B.2). The "one-factor" results adopt the one-common-factor correlation structure and rely on a Gaussian copula (Appendix B.3). The "t-copula" and "mixture of normals" results also adopt the one-common-factor structure. The two numbers (in parentheses) qualifying the t-copula results refer to the degrees of freedom of the common and idiosyncratic factors. The "mixture of normals" setup assumes that idiosyncratic factors are normal but the common factor is a mixture of three normally distributed variables with PDF $0.32\phi(-3,8)+0.50\phi(1,1)+0.18\phi(0,1)$, where $\phi(\mu,\sigma)$ is the PDF of a normally distributed variable with mean μ and standard deviation σ.

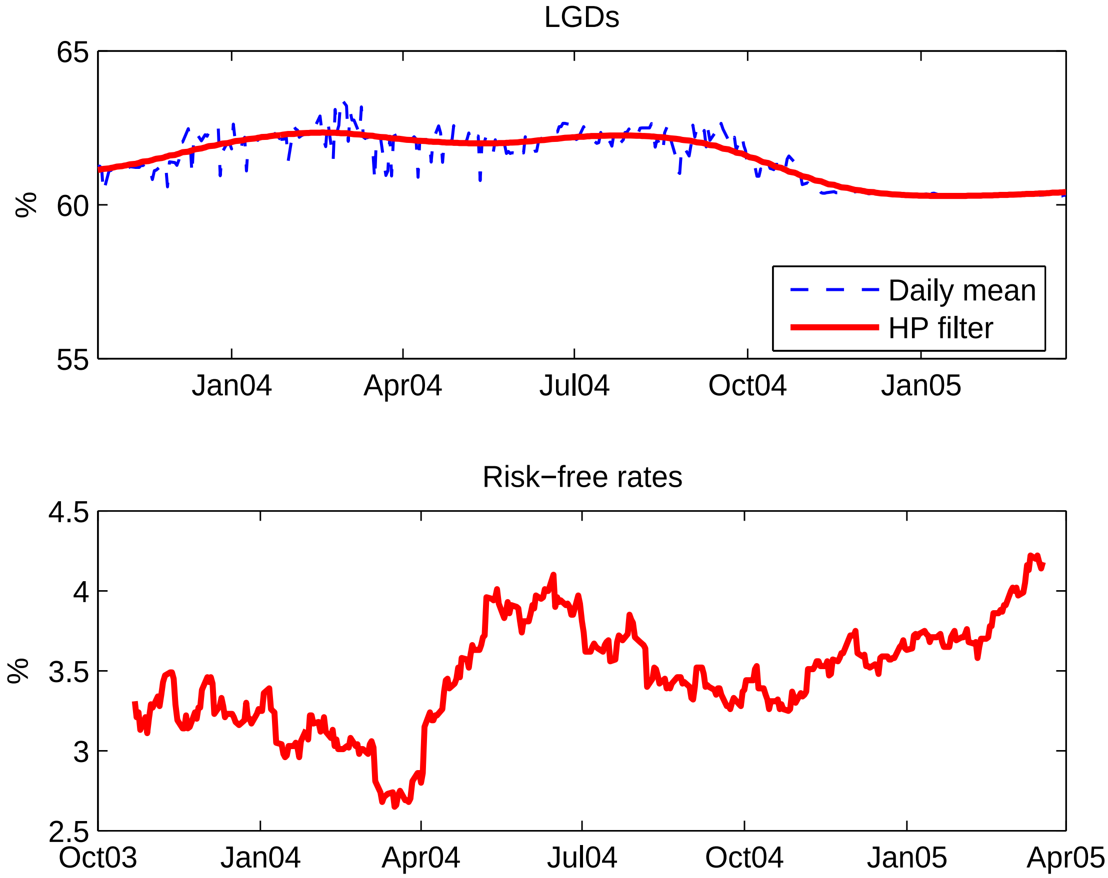
Table 3: **Explaining the time variation in tranche spreads**

A. Regression of observed spreads on daily average PDs				
Tranche	*0-3%*	*3-7%*	*7-10%*	*10-15%*
slope coefficient	15.40	5.57	2.22	0.91
(t-stat)	(48.7)	(23.5)	(29.6)	(17.6)
R^2	0.88	0.63	0.73	0.49
Correlations of residuals across tranches				
0-3%	1			
3-7%	0.67	1		
7-10%	0.70	0.95	1	
10-15%	0.60	0.96	0.97	1

B. Regression of observed spreads on CDS-implied spreads				
Tranche	*0-3%*	*3-7%*	*7-10%*	*10-15%*
slope coefficient	0.65	0.86	1.10	1.77
(t-stat)	(43.6)	(28.9)	(41.0)	(24.6)
R^2	0.86	0.72	0.84	0.65
Correlations of residuals across tranches				
0-3%	1			
3-7%	0.76	1		
7-10%	0.76	0.90	1	
10-15%	0.65	0.92	0.93	1

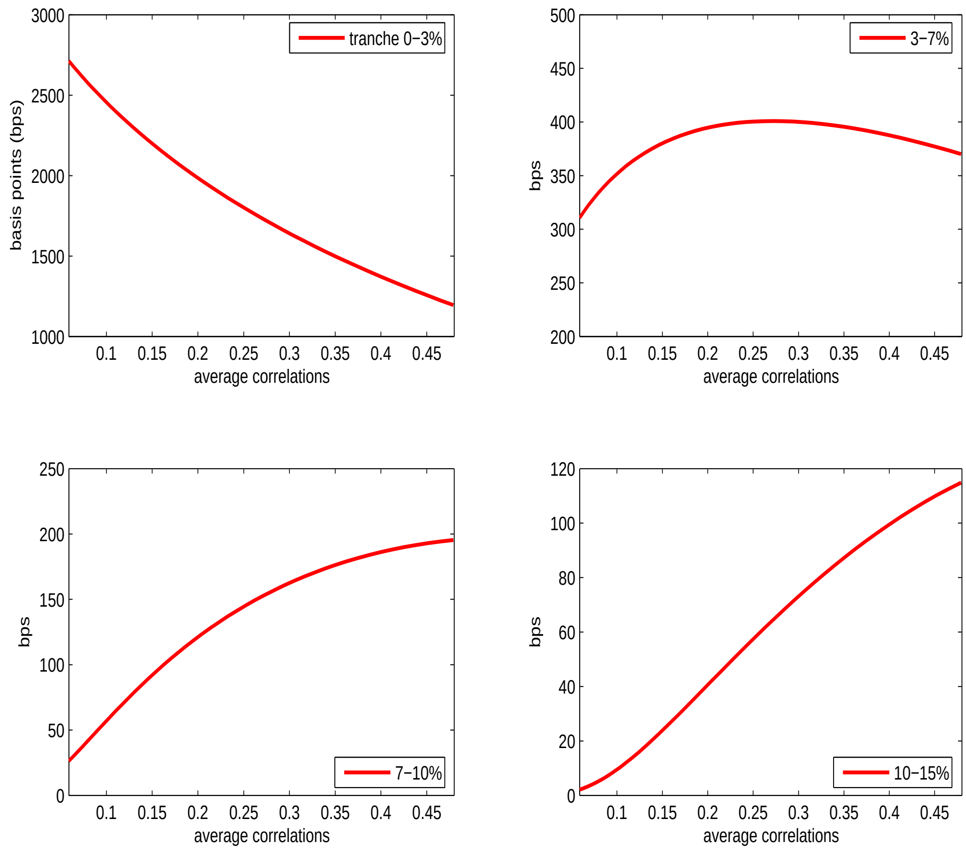
Note: All regressions include a constant term, which is omitted from the table.

Figure 1: LGDs and risk-free rates of return



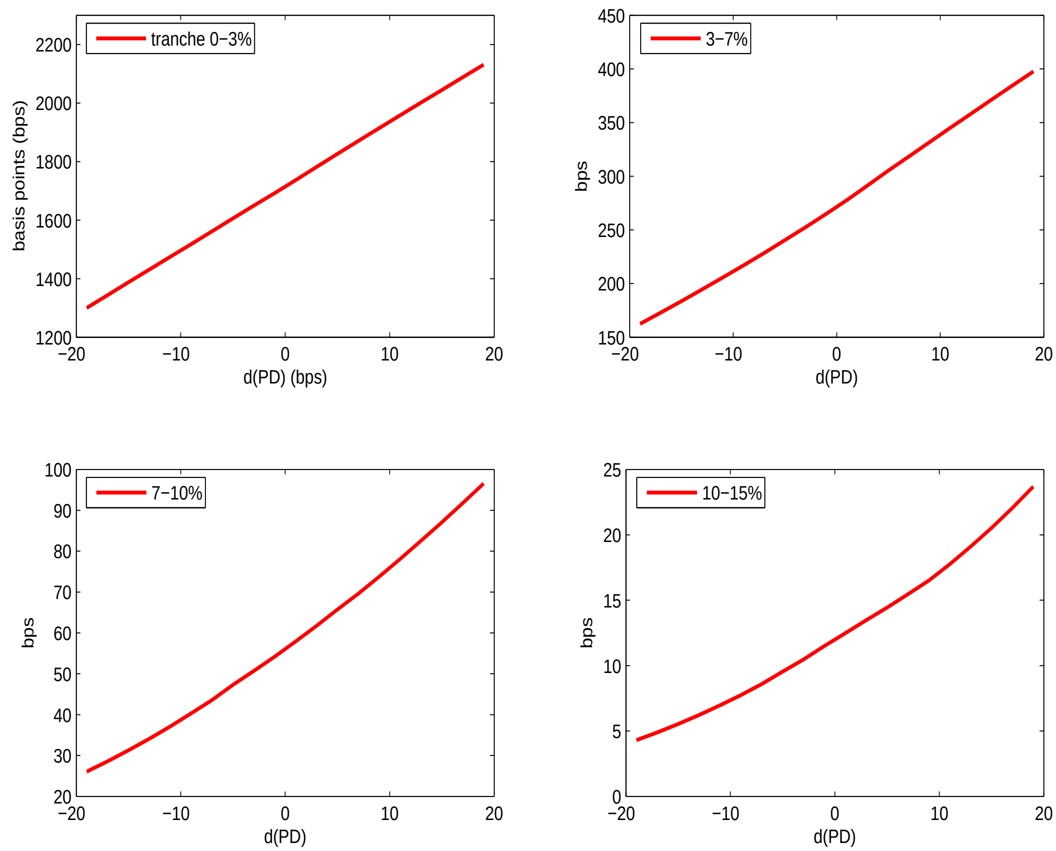
Note: On each day, the loss-given-default (LGD) refers to the cross-sectional average for the 125 entities in the "on-the-run" release of the CDS index CDX.NA.IG.5Y. The HP filter adopts $\lambda = 64000$. The risk-free rate of return is proxied for by 5-year Treasury rates.

Figure 2: The sensitivity of tranche spreads to average correlations



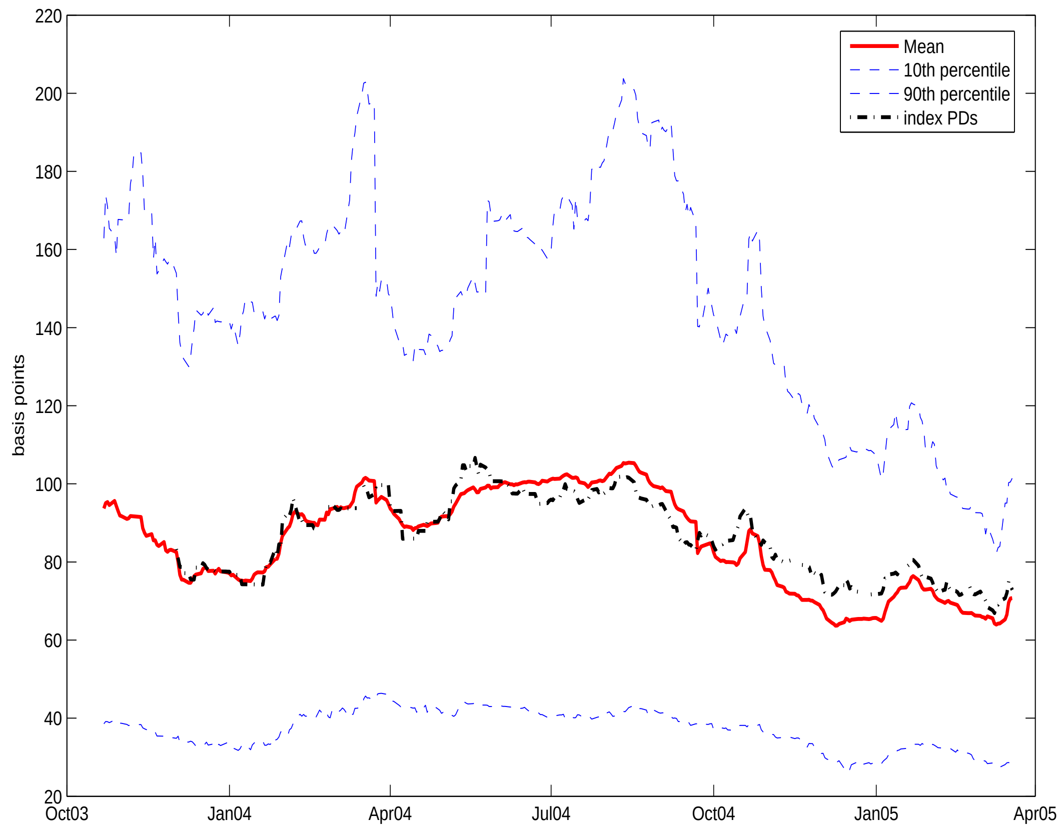
Note: The sample set includes the 125 entities in release 2 of the CDS index CDX.NA.IG.5Y. The pricing of tranche spreads uses the cross section of time averages of PDs, and the average LGD and risk-free rate in the sample. We start with the one-factor approximation of the GCorr correlations, then change all loading coefficients by the same amount (so that the average correlation changes but the correlation structure is maintained) and re-price the tranche spreads.

Figure 3: The sensitivity of tranche spreads to PDs



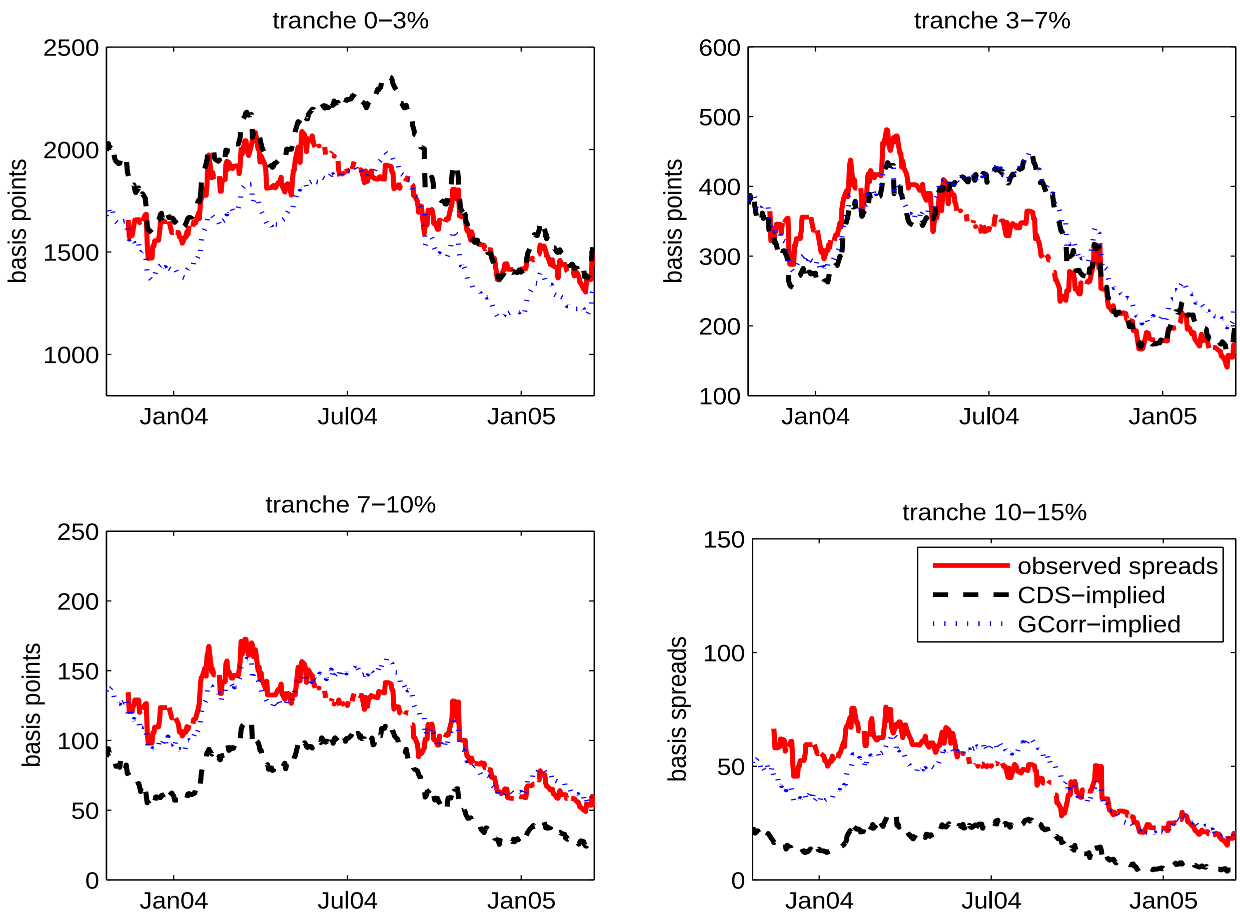
Note: The sample set includes the 125 entities in release 3 of the CDS index CDX.NA.IG.5Y. The pricing of tranche spreads is based on the CDS-implied asset-return correlations and the average LGD and risk-free rate in the sample. We start by fixing the PD of each firm at the corresponding time average. The resulting spreads are plotted for $d\,(PD) = 0$. Then, we change all individual PDs by the same amount and re-price the tranche spreads.

Figure 4: Probabilities of default



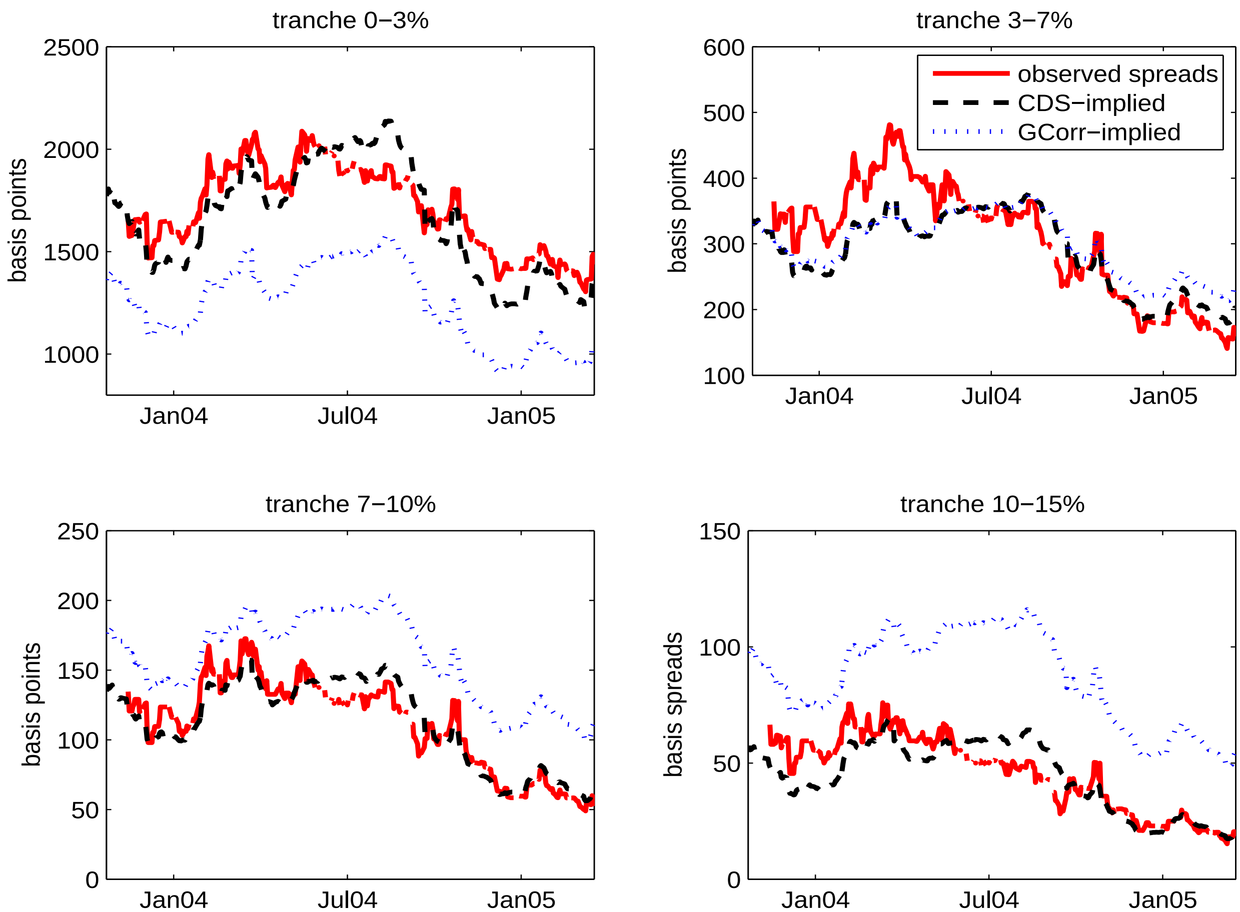
Note: The mean and the percentiles of the daily cross sections of PDs are based on the 125 entities in the "on-the-run" release of the CDS index CDX.NA.IG.5Y. The "index PDs" are estimated from observed single-tranche index spreads.

Figure 5: Observed and implied spreads of CDS index tranches (standard normal)



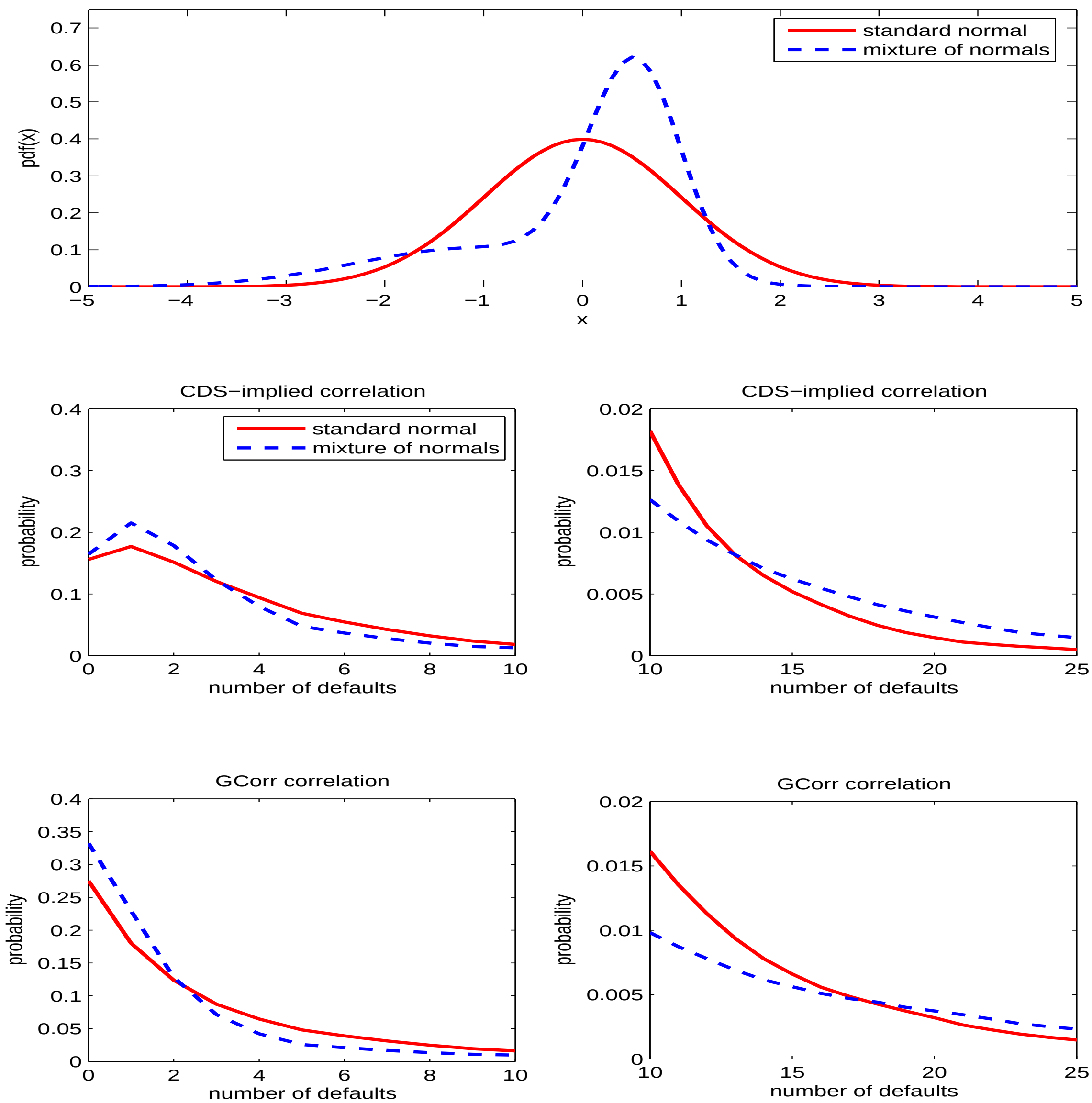
Note: The observed tranche spreads in the CDS index market are provided by JPMorgan Chase. Implied tranche spreads use the CDS-implied or GCorr asset-return correlations. The pricing algorithm adopts the one-period simulation of defaults (Appendix B.1) and assumes normally distributed asset returns.

Figure 6: Observed and implied spreads of CDS index tranches (mixture of normals)



Note: The observed tranche spreads are provided by JPMorgan Chase. Implied tranche spreads are based on the one-common-factor model of CDS-implied or GCorr asset-return correlations. The pricing algorithm adopts a one-period simulation of defaults (Appendix B.1) and assumes that the common factor has a mixture-of-normals distribution, which is specified in Table 2.

Figure 7: Probability distribution of the number of defaults



Note: The top panel plots the PDF of a standard normal variable and a mixture of three normally distributed variables. The PDF of this mixture is $0.32\phi(-3, 8) + 0.50\phi(1, 1) + 0.18\phi(0, 1)$, where $\phi(\mu, \sigma)$ is the PDF of a normal variable with mean μ and standard deviation σ. The other panels plot the probability of joint defaults of a portfolio of 125 entities, using one of the two sets of correlations (CDS-implied or GCorr correlations) and one of the following two distributional assumptions on asset returns: (i) both common and idiosyncratic factors are normally distributed (solid lines); (ii) idiosyncratic factors are normally distributed but the common factor has the mixture-of-normals distribution specified above (dashed lines). The PD is assumed to be the same for all entities and to equal 3%.