Corporate bond defaults are consistent with conditional independence

Florian Kramer*

Allianz Investment Management SE

Reinsburgstraße 19, 70178 Stuttgart, Germany

Phone: +49 711 663 1728 Fax: +49 711 663 81728

Email: florian.kramer@allianz.com

Gunter Löffler

Institute of Finance, Ulm University

Helmholtzstraße 18, 89069 Ulm, Germany

Phone: +49 731 50 23597 Fax: +49 731 50 23950

Email: gunter.loeffler@uni-ulm.de

April 2010

*Corresponding author. We are grateful to Moody's Investors Service and Moody's KMV for providing the data. Furthermore, we thank David Lando and Mads Stenbo Nielsen for helpful comments and discussions, an anonymous referee for many valuable suggestions and Mario Rometsch for technical support. Parts of the paper were written while the second author was enjoying the hospitality of the College of Management at North Carolina State University. The views expressed in this paper do not necessarily reflect the opinions of Allianz SE or its subsidiaries. The usual disclaimer applies.

Abstract

Standard credit risk models rely on a doubly stochastic structure. Conditional on the evolution of common factors, defaults are independent. Recently, tests by Das, Duffie, Kapadia and Saita (2007) have cast doubt on the empirical validity of this assumption. We modify their estimation approach in two ways. First, we model intra-month patterns in observed defaults. Second, we estimate default intensities on an out-of-sample basis, which brings our estimates closer to the ones financial institutions would have used when implementing the approach in the past. Once intensity estimation is modified in these ways, the validity of the doubly stochastic assumption is no longer rejected.

Keywords: Default; Intensity Estimation; Doubly stochastic; Default Correlation; Default Prediction

1 Introduction

Corporate defaults tend to cluster in time and in industries. In 2002, for example, the annual corporate default rate recorded by Moody's Investors Service (2003) was 3%, more than twice the long-run average; telecommunications issuers like GlobalCrossing or Worldcom constituted 31% of all defaulted issuers. This clustering did not come as a surprise, however. In 2002, the US economy was emerging from a recession, debt and capacity levels were still high while equity values had gone down significantly; the telecommunications industry was in particularly bad shape. Many market participants therefore expected high default rates for the year 2002 (e.g. Moody's Investors Service (2002)).

In standard credit risk models, default clusters are explained in much the same way. Common factors like the business cycle or stock market valuations induce correlations in individual firms' default intensities. If average intensities happen to be high, the realized default rate is likely to be high as well. The models do not allow for the possibility that the failure of one company triggers other defaults. Rather, they are characterized by a convenient doubly stochastic structure. Conditional on the stochastic evolution of common factors, defaults are independent. The approach is ubiquitous in financial institutions. Key applications range from industry models of portfolio credit risk (see e.g. Crouhy et al. (2000)) and the new capital requirements (e.g. Gordy (2003)) to the pricing of structured finance instruments like collateralized debt obligations (e.g. Duffie and Gârleanu (2001)).

Das, Duffie, Kapadia and Saita (2007) (hereafter DDKS) conducted the first rigorous empirical test of the doubly stochastic assumption. They test whether defaults of US corporates are consistent with the model-implied default clustering of the intensity model introduced by Duffie, Saita and Wang (2007). DDKS find that observed default clustering exceeds the one implied by the estimated model. Therefore, they conclude that the doubly stochastic assumption is invalid and that credit risk models should be enriched by contagion effects or "frailty", i.e. unobservable variables, in order to account for the extra correlation. Otherwise, one might incur significant errors in the assessment of credit portfolio risk or the pricing of structured finance products.

As pointed out by DDKS, their statistical tests are joint tests of well-specified default intensities and the doubly stochastic assumption. In our paper, we use a data set that is very similar to the one used by DDKS but introduce two modifications to the estimation of default intensities. First, we model a *baseline component* which is able to capture the patterns in intra-month default timing that we find in the data. Specifically, defaults tend to cluster around the 1st and 15th of a month.

Second, we do not estimate intensities in-sample as in DDKS but use only information that was available at the start of the default prediction horizon. For example, the intensity estimated at the end of January 1990 uses only information dated January 1990 or earlier. We consider this to be appropriate because the test then can be interpreted as follows: Would somebody working with the model in the past have made significant errors regarding the clustering of defaults? Tests based on in-sample estimates, by contrast, could bias the results in ways which are difficult to assess. Another motivation for our procedure is that the functional relationship between intensities and explanatory variables might change. For example, the relationship between the term spread and recessions has been found to be unstable over time (cf. Benati and Goodhart (2008)), suggesting that the relationship between the term spread and default rates is unstable, too. Other studies suggest that rating standards may have changed over time (see Blume et al. (1998)). Jorion et al. (2009) have attributed this to a decrease in the quality of accounting data, which is relevant not only for ratings but also for quantitative estimates of default risk. To avoid giving too much weight to possibly outdated historical data, we use a rolling five-year estimation window.

We find that both modifications increase the ability of the intensity model to explain observed default clustering. When both modifications are made, most tests suggested by DDKS no longer reject the validity of the doubly stochastic assumption. The only exception that we observe can be attributed to an overestimation of default intensities rather than to unexpected clustering.

In a recent contribution, Lando and Nielsen (2009) arrive at similar results. Specifically, they modify the intensity model of Duffie, Saita and Wang (2007) by adding explanatory variables. In our data set, adding industrial production and the term spread as suggested by Lando and Nielsen (2009) also renders most model forecasts consistent with observed default clustering. The additional modifications suggested by us lead to a further increase in the models' performance. Taken together, the results presented by Lando and Nielsen (2009) and ourselves provide a strong case for reconsidering the results of DDKS.

The doubly stochastic assumption is not the only line of attack against standard credit risk models. Based on recent research, other frequently made assumptions, too, appear to be less critical than a first look may suggest. Hamerle and Rösch (2005), for example, demonstrate the robustness of the Gaussian copula; Kiefer and Larson (2007) and Frydman and Schuermann (2008) show that the Markov specification for the modeling of rating transitions can work relatively well.

The remainder of the paper is organized as follows. After briefly introducing the doubly stochastic modeling approach in section 2, we describe the data in section 3. Our specification of the default intensities and their estimation follows in section 4; the predictive power of the model is assessed in this section, too. Section 5 analyzes if the estimated models are able to account for the default clustering observed in the data. Section 6 concludes.

2 The doubly stochastic modeling approach

This section provides an introduction to the doubly stochastic modeling approach. Readers familiar with such models or mainly interested in the empirical results may wish to proceed to section 3. Further technical details are provided in the Appendix.

In order to specify a doubly stochastic (or equivalently conditional independence) credit portfolio setup, we follow Lando (1998) and define the firms' default times τ_i as

$$\tau_i = \inf\left\{t : \int_0^t \lambda_i(s, X(s))\, ds \ \geq \ E_i\right\}.$$

Here, λ_i denotes some positive function of time and a stochastic process X, which is independent of the unit-exponentially distributed, mutually independent random variables E_i. Implications of this definition are the following:

- Since X and the E_i are independent, past defaults do not influence the default probabilities of the surviving firms, i.e. contagion is not possible in this setup. Dependence between the default times is only introduced by the dependence of the *intensity* $\lambda_i(t)$ on a common process X. Consequently, conditional on the path of X, defaults are independent, which is the reason why this setup is also often called the conditional independence setup.
- Simultaneous defaults of firms occur with probability 0.

Moreover, $\lambda_i(t)\Delta$ can be interpreted as the instantaneous probability of a firm to default over the next infinitesimally small time step Δ conditional on survival until t.

In the concrete model implementation, which we consider in section 4, the process $X = (Y, Z)$ will be given by the explanatory variables or *covariates* where we assume that these variables are independent of past defaults and where, in this article, Y will represent macroeconomic variables such as the S&P 500 index return and $Z = (Z_1, \ldots, Z_I)$ will summarize all firm-specific variables such as the firms' long-term debt rating. As a consequence, the dimension of Y simply corresponds to the number of macroeconomic variables used, while the dimension of Z is the number of firms in the portfolio times the number of firm-specific variables considered.

Furthermore, as is common in the literature, we will consider a parametric function for λ_i such that

$$\lambda_i(t, X(t)) := \lambda\left(t, \beta, Y(t), Z_i(t)\right)$$

with $\beta \in \mathbb{R}^d$ some parameter vector. Given β and a path $x = (y, z)$ of X, the likelihood $\mathcal{L}(\beta; y, z)$ of observing some vector $\tau = (\tau_1, \ldots, \tau_I)$ of default times is given by

$$\mathcal{L}(\beta; y, z) = \prod_{i \in I} e^{-\int_0^{\tau_i} \lambda(u, \beta, y(u), z_i(u))du} \lambda\left(\tau_i, \beta, y(\tau_i), z_i(\tau_i)\right), \tag{1}$$

since the probability of firm i to default at $t = \tau_i$ conditional on the path of X is given by $e^{-\int_0^t \lambda(u,\beta,y(u),z_i(u))du} \lambda\left(t, \beta, y(t), z_i(t)\right)$ and conditional on X the default times are independent. In addition, with L denoting the aggregated portfolio loss process

$$L(t) := \sum_{i=1}^{I} 1_{\tau_i \leq t},$$

we have that

$$M(t) := L(f^{-1}(t))$$

is a standard Poisson process where $f(s) := \sum_{i=1}^{I} \int_0^s 1_{u<\tau_i} \lambda_i(u, X(u))du$, i.e. the increments of M

$$M(T) - M(t) \sim Poi(T - t) \tag{2}$$

and the inter-arrival times of M

$$f(\tau_i) - f(\tau_{i-1}) \sim Exp(1) \qquad i.i.d. \tag{3}$$

Inspired by DDKS, we will make use of these relationships when testing in section 5 if the estimated models are able to explain the observed default clustering.

3 Data

We use data on U.S. corporates which have traded equity and a Moody's rating. The data extend from January 1980 to April 2005. The information provided include daily default times and two firm-specific variables that we are going to use as default predictors:

- The firms' 1-year *Expected Default Frequencies* (EDFs) provided by Moody's KMV, which are month-end values ranging from 0.02% to 20%; a firm's 1-year EDF is a non-parametric estimator of its 1-year default probability based on the firm's "distance-to-default", which itself is a leverage measure adjusted for asset volatility that goes back to the firm value model of Merton (1974). For a more detailed description of the EDF as a measure of a firm's default risk see for example Berndt et al. (2005), who use the EDF in order to estimate and analyze risk premia of corporate bonds.

- The Moody's long-term rating of the firms with values {"Aaa", "Aa1", ..., "C"}, which we transform to the log of idealized annual default probabilities used by Moody's (Yoshizawa (2003)). As Moody's does not specify idealized default probabilities for rating grades Ca and C, we estimate them through a quadratic extrapolation[1]. To check robustness, we also considered a simple linear coding ("Aaa"=1, "Aa1"=2,...) and found that it did not affect conclusions.

Our use of both EDFs and ratings as explanatory variables for default prediction is inspired by Löffler (2007), who shows within a static logit regression model that adding ratings to EDFs increases predictive power.

The original data set contains observations for which one of the two variables (EDF, rating) is missing. A possible solution to this problem would be to extrapolate or interpolate missing EDFs or ratings. However, this could make later results sensitive to the chosen interpolation method. Therefore, we remove data points with missing variables. Furthermore, we disregard multiple defaults of firms which can be observed in the data set, i.e. observations after the first default of a firm are not taken into account.

[1]This leads to default probabilities of 59% and 79% for rating grades Ca and C, respectively.

Finally, we end up with a data set consisting of 266462 observation months for 2760 different firms across the time period from 01/1980 to 04/2005. The sample includes 414 defaults which is less than the 495 defaults in DDKS but more than the 370 defaults in Lando and Nielsen (2009). The number of firms in the data set increases from 1980 on, reaches its maximum of approximately 1250 firms around the year 1999 and slowly declines afterwards. Figure 1 shows the monthly default rates. The recessions in 1990-91 and 2001 are visible with monthly default rates peaking at around 0.8%.

For a subset of the data, we use an additional firm-specific variable that has already been used in Duffie, Saita and Wang (2007) and DDKS:

- The firm's month-end 1-year trailing stock return.

Return information is from CRSP. Due to several reasons (delistings before defaults, defaults within 12 month after listing, missing data in CRSP), stock returns are only available for a subset consisting of 248842 observation months for 2657 different firms, comprising 351 defaults.

Apart from the stated firm-specific variables we also use macroeconomic variables for the estimation of the intensities. The macroeconomic variables (all monthly) that we use are[2]:

- The 1-year trailing return on the S&P 500 index.
- The 3-month US-Treasury bill rate.
- The 1-year percentage change in US industrial production, calculated from the gross value of final products and nonindustrial supplies (seasonally adjusted). We use this variable with a 1-month lag since figures on industrial production are published in the middle of the next month.
- The spread between the 10-year and 1-year treasury rate.

The first two variables have been used in Duffie, Saita and Wang (2007) and DDKS. US industrial production has been considered in DDKS, too. In Duffie, Saita and Wang (2007) the 10-year rate was discussed but rejected for lack of significance. By using the spread between the 10-year and 1-year treasury rate we follow Lando and Nielsen (2009).

In the analysis, we will use month-end values of the explanatory variables to estimate the intensities of the following month. Some of these variables (rating information, trailing returns on the S&P 500 and the company's stock, interest rate and interest spread) are observed continuously[3]. In order to investigate the effect of exploiting these variables at a higher frequency, we also conduct an estimation based on month-end and middle-of-the-month values (see section 5).

[2] All macroeconomic variables except for the return on the S&P 500 index have been downloaded from the Federal Reserve Board Website. The S&P 500 index has been obtained from Datastream.

[3] The EDF can also be computed on a daily basis but the database available to us contains only month-end values.

4 Estimation of default intensities

After having decided on the covariates that influence the default intensities λ_i, it remains to specify a parametric function. For our specification, we rely on a *Proportional Hazards Rate Model*, see e.g. Therneau and Grambsch (2001), in which a firm's default intensity $\lambda_i(t)$ is linked to the observable variables $(Y(t), Z_i(t))$ via the relationship

$$\lambda_i(t) = \lambda\left(t, \beta, Y(t), Z_i(t)\right) = \mu_0(t) \exp\left(\sum_{n=1}^{N} \beta_n Y_n(t) + \sum_{m=1}^{M} \beta_{N+m} Z_{mi}(t)\right).$$

As already mentioned, $Y(t) = (Y_1(t), \ldots, Y_N(t))$ denotes the time t values of the macroeconomic variables, $Z_i(t) = (Z_{1i}(t), \ldots, Z_{Mi}(t))$ the values of the firm-specific variables of firm i and $(\beta_1, \ldots, \beta_{N+M})$ the coefficients describing how the variables influence the intensity; $\mu_0(t)$ denotes a possibly time-dependent intercept that may capture recurring intra-month patterns in default times and which we refer to as the *baseline component* in the following. So far, our model is the same as utilized by Duffie, Saita and Wang (2007) and Lando and Nielsen (2009) except for the choice of firm-specific variables and for the baseline component. Differences in firm-specific variables should be minor, though. Both Duffie, Saita and Wang (2007) and Lando and Nielsen (2009) employ a measure of distance-to-default which is also at the heart of the EDF that we use.

The baseline component

Since later model specification tests strongly depend on the clustering of defaults over time – even on the clustering within small time intervals such as months – the estimation of the firms' default intensities described next will take the timing of defaults within the corresponding default months into account, too. Figure 2 shows how defaults are distributed across the single days of a month. Defaults are heavily clustered around the 1st and the 15th of each month, presumably because default is often declared after a missed coupon payment (the 1st and the 15th are common choices for coupon payment dates)[4]. Loosely speaking, if a firm has survived the 15th of a month, it has very good chances to survive the whole month since the dates around the 1st and the 15th account for more than 50% of the defaults.

As baseline hazard rate we therefore use a simple periodical function: A value $\mu_{01}e^{\beta_0}$ for the beginning of the month, $\mu_{02}e^{\beta_0}$ for the first half of the month (excluding its begin), a middle-value $\mu_{03}e^{\beta_0}$ and a value $\mu_{04}e^{\beta_0}$ for the second half of the month (excluding the middle). The μ_{0i}, $i \in \{1, 2, 3, 4\}$ reflect the empirical distribution of the defaults in the considered intra-month time periods and are normalized such that $\int_t^{t+\frac{1}{12}} \mu_0(s)ds = \frac{1}{12}e^{\beta_0}$. The beginning of the month is defined as the 1st and its middle is given by the 15th. Our specification also takes weekends into account in the sense that the period "beginning of the month", for example, extends from the 1st to the 3rd given that the 1st is a Saturday. In the same way each 15th is treated. We thereby model that coupon payments are deferred to the next Monday if the payment date falls

[4]We examined 3782 US corporate bonds issued before May 2005 (the end of our sample period) for which information is available in Datastream. Most bonds have payment dates falling on the 1st (35.3% of all bonds) or 15th (45.3%) of the month.

on a Saturday or Sunday.

The calculated baseline component is used in all estimations even if we apply the model out-of-sample since we assume that the intra-month default timing is fixed over time. To support this assumption, as a robustness check, we calculated the empirical distributions of the first 50% and the last 50% of defaults in the data set. These are presented in Figure 3. The distribution seems to remain relatively stable over time. Furthermore, since $\int_t^{t+\frac{1}{12}} \mu_0(s)ds = \frac{1}{12}e^{\beta_0}$, the baseline component leaves a firm's default probability over the next month unchanged and does only very weakly change the estimated values of the coefficients $(\beta_0, \ldots, \beta_{N+M})$, i.e. estimated coefficients $(\hat{\beta}_0, \ldots, \hat{\beta}_{N+M})$ are similar for a model with and without baseline component.

Estimation results with fixed coefficients

Having specified the baseline component, we have finally set the stage for conducting model inference via maximum likelihood. We start by estimating intensities in-sample, i.e. coefficients are fixed for the entire sample period. Since observations for individual firms may start at different dates, the likelihood function (1) that is actually used in the estimation has the form

$$\begin{aligned}\mathcal{L}(\beta; y, z) \quad &= \quad \prod_{i \in I} \exp\left(-\int 1_{u \in \mathcal{T}_i, u < \tau_i} \lambda(u, \beta, y(u), z_i(u)) du\right) \\ &\qquad \cdot \left(1_{\tau_i \in \mathcal{T}_i} \lambda(\tau_i, \beta, y(\tau_i), z_i(\tau_i)) + 1_{\tau_i \notin \mathcal{T}_i}\right),\end{aligned}$$

where $\mathcal{T}_i$ denotes the union of intervals for which all variables of firm i that we want to use (default information, rating, EDF, stock return) are available. Since we disregard multiple defaults, $\max \mathcal{T}_i = \tau_i$ in case of a default of i.

The maximum likelihood estimator $\hat{\beta}$ of β is then given as

$$\hat{\beta} = \arg\max_{\beta} \ \mathcal{L}(\beta; y, z).$$

The estimated default intensity parameters $\hat{\beta}$ as well as their t-statistics are displayed in Table 1 for the basic five model specifications which we will consider throughout the text. Here as in all later tests, we restrict the sample period from 01/1985 to 04/2005 because this is the time period for which we can estimate the intensities with our alternative approach (rolling estimations with a 5-year estimation window). Using the entire available data for the fixed coefficients estimation does not lead to qualitatively different results.

As a first observation, it is interesting to note that if the Moody's KMV EDF model can be taken at face value, the coefficient on the log(EDF) in model I should approximately equal one, while the constant should equal zero, since the 1-year default probability is approximately be given as

$$1 - exp(-exp(0 + 1 \cdot \log(EDF))) = 1 - exp(-EDF) \approx EDF.$$

This only presents an approximation because it uses a Taylor approximation and further assumes that EDFs remain constant over the year. In Table 1, the estimated coefficient on the log(EDF) is about 10 standard deviations away from one. Also, the constant is significantly different from zero. Given the approximations that underlie the hypothesis, this does not necessarily imply that the estimation of EDFs is misspecified, however. When we add more variables, the estimated coefficient on the EDF approaches one. In model V, it is no longer significantly different from one. The results could therefore be consistent with well-specified EDFs which exhibit dynamics that are captured by other variables.

In most cases the coefficients exhibit the expected sign. Default intensities are increasing in the EDF, the rating[5] and they are decreasing in industrial production and the firms' stock returns. Exceptions are the S&P 500 index, the spread between the 10-year and the 1-year rate and the 3-month US-Treasury bill rate. At first sight, one would probably have expected all three coefficients to have a negative sign because boom periods (with defaults below average) are usually accompanied by positive stock returns, high interest rates and an upward sloping yield curve. However, the sign of the S&P 500 index is only positive *after* controlling for the other variables. An increased survival probability of firms during a stock market boom might already be captured by variation in the EDF. Short-term interest rates could go along with higher default rates because they impact borrowing costs or mark the end of a boom. Note that in Duffie, Saita and Wang (2007) the coefficient on the S&P 500 is positive, too; the one on short-term interest rates is negative. The coefficient for the spread between the 10-year and the 1-year rate is positive in Lando and Nielsen (2009), too.

Estimation results with rolling estimation windows

The in-sample estimation from the previous section presumes that the functional relationship between explanatory variables such as the EDF or rating and the intensities is constant over a time period of 25 years. Even though default information providers like Moody's or Moody's KMV aim at the time-consistency of their default risk assessments, it is unclear whether they achieve perfect consistency. Two possible reasons for non-stationary behavior are changes in the information content of accounting measures (cf. Jorion et al. (2009)) or sampling error in the estimation and calibration of risk measures. EDFs, for example, use estimated stock market volatilities and are calibrated on past default behavior which exposes them to sampling error; sampling errors can be correlated across firms and therefore affect the average precision of EDFs. In addition, the information content of macroeconomic variables such as the term spread can change over time (see e.g. Benati and Goodhart (2008)).

Therefore, working with constant coefficients over long time periods is possibly inappropriate. In order to account for possible variations of the covariates' influence on default intensities, we allow the coefficients to change over time. More precisely, for each $\beta(t)$ we determine an estimation period $\mathcal{U}(t)$ which ends in t and estimate $\beta(t)$ based on the observed covariates and defaults during this period. The forecasts of the intensities are therefore obtained on an out-of-sample basis, which provides another motivation for favoring this approach. We only use information that a

[5]Note that a higher rating is a worse rating due to the chosen transformation.

financial institution which implemented the model would have had at the time of implementation. When we later test whether observed default clustering is consistent with the doubly stochastic approach, we therefore can interpret the results in the sense that actual users of that approach would have incurred significant errors in predicting default clusters or not. With in-sample estimates, test results could be biased in some unknown way.

With the estimation sample restricted to $\mathcal{U}(t)$, the likelihood function of $\beta(t)$ has the form

$$\begin{aligned}\mathcal{L}(\beta;\mathcal{U}(t),y,z) &= \prod_{i\in I}\exp\left(-\int 1_{u\in\mathcal{T}_i\cap\mathcal{U}(t)}\lambda(u,\beta,y(u),z_i(u))du\right)\\ &\quad\cdot\left(1_{\tau_i\in\mathcal{T}_i\cap\mathcal{U}(t)}\lambda(\tau_i,\beta,y(\tau_i),z_i(\tau_i))+1_{\tau_i\notin\mathcal{T}_i\cap\mathcal{U}(t)}\right)\end{aligned}$$

and

$$\hat{\beta}(t)=\arg\max_{\beta}\ \mathcal{L}(\beta;\mathcal{U}(t),y,z).$$

Our procedure eventually yields a time series of $\hat{\beta}(t)$. Figure 4 depicts the evolution of two of the most important coefficients – the one on the stock return and the one on the rating – over time. Because of the 5-year estimation period used, the time series $\hat{\beta}$ starts with January 1985 and each of its values $\hat{\beta}(t)$ reflects the past default behavior over the last 5 years. For example, the value of the coefficients at the end of January 1985 are based on the relation between defaults and observable variables during 02/1980 to 01/1985. Though the coefficients fluctuate, they always have the same sign.

We choose the 5-year estimation period because it is a common choice in the financial industry and because it appears to achieve a good balance between opposing effects. On the one hand, increasing the estimation period should lower estimation errors. On the other hand, it could mean that we use more outdated historical data. Also, it would force us to ignore more observations when testing the models' performance because the first estimates are only available after the difference between the current date t and the start of the data is larger than the estimation period.

Predictive power

While the focus of our paper is on the validity of the doubly stochastic assumption, the test we will use is a joint test of the doubly stochastic assumption and well-specified intensities. It is therefore worthwhile to investigate the models' predictive ability. We follow the literature and use the *accuracy ratio AR* to measure the ability to rank firms according to their default probability. The AR is based on the so-called *power curve*. The power curve $pc : [0,1] \mapsto [0,1]$ associated with a model is a function that can be obtained by ranking all firms of the portfolio according to some criterion, which is supposed to carry as much information as possible on the firms' default probability over the period one is interested in. Usually, this criterion would be the firms' default probability over this time period implied by the estimated model. The power curve pc maps the fraction x of worst ranked firms onto the fraction of defaults which these firm account for; it tells us that when we pick the worst ranked firms, a fraction x of all firms accounts for $y = pc(x)$ of the defaults.

We follow Duffie, Saita and Wang (2007) and define the accuracy ratio to be twice the area between the power curve and the 45° line, i.e.

$$AR = 2 \int_0^1 (pc(x) - x)\, dx.$$

The 45° line is the power curve one would expect to obtain by ranking firms completely randomly. By definition, the maximum value of the accuracy ratio is 1 minus the fraction of actually defaulted firms.

It remains to choose a criterion for ranking the firms. A natural candidate would be the default probability over the prediction period, e.g. the one-year default probability if the prediction horizon is one year. However, to derive such a probability, we would have to set up a model for the dynamics of the firms' default intensities. Therefore, instead of horizon-matched default probabilities we simply use the firms' current default intensities in order to rank them.

Table 2 summarizes results for default prediction horizons of one and five years, respectively. Accuracy ratios have been calculated separately for each month and then averaged over time. The table shows accuracy ratios for the model with fixed coefficients, which are therefore in-sample, as well as accuracy ratios for the rolling estimation. The accuracy ratios of the model with fixed coefficients are higher, which is unsurprising due to the in-sample nature of these values, but the difference is quite small (for model I there is no visible difference at the displayed precision). Obtained accuracy ratios are similar to the ones reported by Duffie, Saita and Wang (2007) for a comparable data set over the time period from 1993 on. The inclusion of the firm credit rating tends to increase accuracy ratios showing that if one wants to sort firms according to their default probability, the firm rating should be taken into account in the intensity estimation.

5 Comparing model-implied and observed default clustering

We now assess whether the intensity estimates provide default forecasts which are consistent with observed default clustering. For this purpose, we will consider defaults and default intensities on an aggregate, i.e. portfolio, basis. More precisely, we explore the question whether the path of observed portfolio defaults is likely to have been generated by the (aggregated) paths of default intensities that have been estimated in the previous section.

Our analysis of aggregated intensities and defaults is based on four tests which have been introduced by DDKS to the credit risk literature in order to test the doubly stochastic assumption: *Fisher's dispersion test*, an *upper-quartile test*, *Prahl's test* and an *autocorrelation test*. In these tests, one exploits the fact that the time-changed portfolio loss process M is a standard Poisson process (cf. Equation (2)), i.e.

$$M(t+\delta) - M(t) \sim Poi(\delta),$$

with standard-exponentially, i.i.d. distributed inter-arrival times (cf. Equation (3)). In order to test for the exponential distribution of the inter-arrival times $C_1, \ldots, C_R$ we use Prahl's test (cf. Prahl (1999) and DDKS). With C^* denoting the inter-arrival times' sample mean, Prahl (1999) shows that

$$\frac{1}{R} \sum_{\{r:\, C_r < C^*\}} \left(1 - \frac{C_r}{C^*}\right)$$

is asymptotically normally distributed with mean $\mu_R = e^{-1} - \frac{\alpha}{R}$ and variance $\sigma_R^2 = \frac{\xi^2}{R}$, where $\alpha \simeq 0.189$ and $\xi \simeq 0.2427$.

For testing the overall Poisson distribution of the process' increments we apply the following procedure: By dividing the total estimated aggregated intensity $\sum_{i=1}^{I} \int_0^\infty 1_{u<\tau_i} \lambda_i(u) du$ into bins of equal size δ, we obtain a series of G_δ i.i.d. $Poi(\delta)$-distributed random variables. More precisely, we calculate calendar times $t_0, t_1, t_2, \ldots$ such that $\sum_{i=1}^{I} \int_{t_i}^{t_{i+1}} 1_{u<\tau_i} \lambda_i(u) du = \delta$. Then, the

$$U_k := M(k\delta) - M((k-1)\delta) = \sum_{i=1}^{I} 1_{t_{k-1} < \tau_i \le t_k}$$

form a series of i.i.d. $Poi(\delta)$-distributed random variables. A simple test if the U_k, $k \in \mathbb{N}$ are indeed Poisson distributed is Fisher's dispersion test (cf. Cochran (1954)): Given $U_k \sim Poi(\delta)$, $k \in \mathbb{N}$,

$$Z = \sum_{k=1}^{K} \frac{(U_k - \delta)^2}{\delta} \sim \chi^2_{K-1},$$

for arbitrary $\delta > 0$. The upper-quartile test compares for each bin size δ the upper quartile mean of the respective empirical distribution with the one of G_δ simulated i.i.d. $Poi(\delta)$-realizations. Following DDKS, this is done 10,000 times in order to estimate the p-value as the fraction of realizations where the empirical upper quartile mean is larger than the simulated one. Finally, the autocorrelation test checks for independence of defaults in subsequent bins by fitting a linear model of the form

$$U_k = \gamma + \zeta U_{k-1} + \epsilon_k,$$

with coefficients γ and ζ and i.i.d. innovations ϵ_k. If the ζ coefficient is significant, this indicates a violation of the doubly stochastic assumption.

In Table 3, the p-values of Fisher's dispersion test with respect to different bin sizes δ for fixed coefficients as well as for rolling estimations (with 5 year estimation windows) are presented. The tests are performed for 396 defaults (334 defaults in the subset for which the stock return was available) over the time period from January 1985 to April 2005 since for this period both estimates are available.

With fixed coefficients, no baseline component and the variables from DDKS (model III), the levels of the p-values are similar to the ones derived by DDKS for the Duffie, Saita and Wang (2007) model. Based on these p-values one would clearly reject the hypothesis that default times are generated by the estimated default intensities. After adding the macroeconomic variables suggested by Lando and Nielsen (2009), which is done in models IV and V, p-values go up. If the rating is not included (model IV), the test no longer rejects for bin sizes 2 to 10. Thus, the results of Lando and Nielsen (2009) are broadly confirmed in our sample (model IV). With the rating included (model V), p-values increase for some bin sizes, and decrease for others.

Adding the baseline component to the specification of the intensities tends to increase the p-values, in particular for smaller bin sizes. To see why it particularly effects smaller bin sizes, note that the intra-month timing does not affect a model's performance if a test is based on the defaults during an entire month. It only matters if a bin starts or ends with a fractional month. If the bin ends on a 16th, for example, the true expected default count would be almost equal to the default count over the entire month as most defaults occur before the 16th. If intra-month patterns are ignored, however, the expected default count for the time from the 1st to the 16th would be little more than one half of the expected monthly default count. The larger the bin size, the smaller is the weight of any fractional months at the beginning or the end of the bin compared to the full months within the bin, and the lower is their impact on the test.

With rolling estimation, p-values particularly increase for models III (which showed low p-values with fixed coefficients) and V, and p-values are generally less dependent on the variables used in the intensity specification. Even with the choice of macro-variables suggested by DDKS one would not reject for bin sizes one to four. The effect of adding the macro variables suggested by Lando and Nielsen (2009) is weaker than before.

As before, adding the baseline component mostly leads to an increase of p-values. Overall, however, moving from an in-sample estimation to a rolling estimation has a stronger effect on p-values than the inclusion of the baseline component. Our findings are particularly important for practical applications, where one would typically use the most significant variables for the intensity specification. For example, the rating has been found to be strongly significant in the model estimation of section 4. We further found that the inclusion of the rating increases the predictive power of a model. Nevertheless, the rejection of a rating-based model for some bin sizes (models II and V) with fixed coefficients could mislead us to the conclusion that one would have incurred significant errors in predicting default clusters in the past. Our results with rolling estimation show that this is not the case.

The upper quartile test (cf. Table 4) leads to qualitatively similar outcomes. Again, results are typically more favorable for the specifications with baseline component and with rolling estimation. For most models, rejections cannot be observed at all. For Prahl's test (cf. Table 5) this is not the case. With fixed coefficients, the null hypothesis cannot be rejected at a level of 5% for models IV and V. When we move to rolling estimation, however, p-values go down, contrary to the pattern that we observed in the other tests. This can be attributed to the fact that our out-of sample estimation routine tends to overestimate defaults, i.e. the total aggregated intensity is larger than the number of defaults actually observed. As a consequence, the mean of the inter-arrival times is larger than one (the hypothetical mean) leading to low p-values[6]. Note that the depicted effect is only observed for the out-of-sample estimation. In-sample, the estimated total aggregated intensity equals the number of defaults in the observation period due to the maximum likelihood estimation routine applied. To conclude, low p-values can originate from an overestimation rather than from an underestimation of risk, which the test is designed to detect, and which is the object of our interest. Therefore, one should not overrate the results of Prahl's test for rolling coefficient estimation. It is worth mentioning that the described overestimation of risk also affects Fisher's dispersion test and the upper quartile test in the sense that the considered bins contain on average fewer defaults than suggested by their bin size. We examine the magnitude of this effect by substituting the average number of defaults for the chosen bin size. In the formula of Fisher's dispersion test, we replace δ by the average number of defaults per bin that we obtain for the chosen bin size; in the upper quartile test, the mean of the simulated Poisson realizations is taken to be the average of the realized defaults per bin rather than the chosen bin size. This re-calibration tends to decrease p-values of the upper quartile test and – for smaller bin sizes only – of Fisher's dispersion test, but it does not affect conclusions. For example, the p-values of model V stay above 0.1409 for all bin sizes in case of Fisher's dispersion test and above 0.1630 in case of the upper quartile test; the averages of the p-values across bin sizes are 0.5472 and 0.4931, respectively.

While the depicted overestimation of defaults is only observed out-of-sample, both in-sample and out-of-sample results of Prahl's test are also affected by the fact that defaults on the same day are treated as simultaneous defaults, which enter the test as zero inter-arrival default times. This is a consequence of the fact that the data base records only the day of a default, not the hour. To investigate the associated effects, we randomly distributed simultaneous defaults across the respective default day. In addition, defaults at the beginning and at the middle of a month (defined as in section 4) were distributed across these periods. This transformation doubled the p-values showing that the precision in which defaults are recorded in our data set, i.e. on a day basis, can have an adverse effect on the test results.

The results of Fisher's dispersion test and the upper quartile test are again broadly confirmed by the outcomes of the autocorrelation test (cf. Table 6). In particular for models IV and V, we cannot find any evidence for dependence between defaults in subsequent bins.

In order to investigate the effect of exploiting the continuously available variables such as the stock return at a higher frequency, we also conducted an estimation based on month-end and middle-of-the-month values. That is, we use month-end values of the explanatory variables to predict defaults in the first half of the next month, and we use middle-of-the-month values to predict defaults in the second half of the same month. We let the first half of the month end at

[6]For example, the mean of the inter-arrival times is 1.15 and 1.14 for models IV and V, respectively.

the 14th as this facilitates the modeling of intra-month default patterns. For the specification of the baseline, we followed our approach from section 4 but only divided each period, i.e. each half of a month, into two parts: Its beginning – marked by the 1st and the 15th, respectively – and the remaining part. Delistings shortly before defaults further decreased our dataset: We finally arrived at 328 defaults for which the tests were performed. For the sake of brevity, we only report the test results for rolling estimation with baseline component, and for the models including the trailing stock return (see Table 7). It turns out that utilizing continuously available information at a higher frequency has a positive effect on p-values, in particular for smaller bin sizes and for specifications without rating, further increasing the p-values of the corresponding specifications. With respect to Prahl's test, the p-value of model V is moved away from rejection; it increases from 0.0013 to a value of 0.0673. This is also true for the specifications not shown in Table 7: For models IV and V with fixed coefficients and baseline component the respective p-values of Prahl's test strongly increase to 0.3922 and 0.3503, respectively (for model III the p-value is still essentially 0). Once again, the results show that a refined estimation of default intensities can strengthen the support for the doubly stochastic assumption.

6 Conclusion

The leading paradigm in credit risk portfolio modeling relies on a doubly stochastic structure. Conditional on the evolution of possibly correlated common factors default intensities and thus defaults are independent. Recent tests by DDKS have questioned the validity of this assumption. In consequence, it seems doubtful whether standard credit risk models provide an adequate framework for measuring and pricing credit risk. In particular, the results suggest that the probability of experiencing large losses might be underestimated.

We modify the estimation approach followed by DDKS in two ways. First, we model intra-month patterns in observed defaults. Second, we estimate default intensities on an out-of-sample basis, which brings our estimates closer to the ones financial institutions implementing the models would actually have used. Once intensity estimation is modified in these ways, the doubly stochastic assumption passes most tests. The exception that is observed can be attributed to an overestimation of default intensities rather than to unexpected clustering.

While the results provide support to the continued use of the standard modeling framework, the results do not imply that contagion or frailty effects are irrelevant in credit risk modeling. Such effects appear small in the data set that we examined, which comprised rated corporate bond issuers with traded equity. Future research should examine whether the results carry over to other portfolios.

As pointed out by Lando and Nielsen (2009), the statistical tests are unfit to uncover a special type of contagion in which the default of one firm affects the factor variables and thus the default intensity of others. This provides a new research avenue for the contagion literature.

References

Benati, L., Goodhart, C., 2008. Investigating time-variation in the marginal predictive power of the yield spread. Journal of Economic Dynamics and Control, 32: 1236–1272.

Berndt, A., Douglas, R., Duffie, D., Ferguson, M., Schranz, D., 2005. Measuring default risk premia from default swap rates and EDFs. Working Paper. Available at: www.defaultrisk.com.

Blume, M. E., Lim, F., Mackinlay, A. C., 1998. The declining credit quality of U.S. corporate debt: Myth or reality?. The Journal of Finance, 53: 1389–1313.

Cochran, W. G., 1954. Some methods of strengthening χ^2 tests. Biometrics, 10: 417–451.

Cont, R., Tankov, P., 2004. Financial modelling with jump processes. Chapman & Hall/CRC , Boca Raton, Fl.

Crouhy, M., Galai, D., Mark, R., 2000. A comparative analysis of current credit risk models. Journal of Banking and Finance, 24: 59–117.

Das, S., Duffie, D., Kapadia, N., Saita, L., 2007. Common failings: How corporate defaults are correlated. The Journal of Finance, 62: 93–117.

Duffie, D., Gârleanu N., 2001. Risk and valuation of collateralized debt obligations. Financial Analysts Journal, 57: 41–59.

Duffie, D., Saita, L., Wang, K., 2007. Multi-period corporate default prediction with stochastic covariates. Journal of Financial Economics, 83: 635–665.

Frydman, H., Schuermann, T., 2008. Credit rating dynamics and Markov mixture models. Journal of Banking and Finance, 32: 1062–1075.

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

Hamerle, A., Rösch, D., 2005. Misspecified copulas in credit risk models: How good is gaussian?. Journal of Risk, 8: 41–58.

Jorion, P., Shi, C., Zhang, S., 2009. Tightening credit standards: The role of accounting quality. Review of Accounting Studies, 14: 123–160 .

Kiefer, N. M., Larson, C. E., 2007. A simulation estimator for testing the time homogeneity of credit rating transitions. Journal of Empirical Finance, 14: 818–835.

Lando, D., 1998. On Cox processes and credit risky securities. Review of Derivatives Research, 2: 99–120.

Lando, D., Nielsen, M. S., 2009. Correlation in corporate defaults: Contagion or conditional independence?. Working Paper, Copenhagen Business School.

Löffler, G., 2007. The complementary nature of ratings and market-based measures of default risk. The Journal of Fixed Income, Summer 2007: 38–47.

Merton, R., 1974. On the pricing of corporate debt: The risk structure of interest rates. The Journal of Finance, 29: 449–470.

Meyer, P.-A., 1971. Démonstration simplifée d'un théorème de Knight. In Séminaire de Probabilités V, Lecture Notes in Mathematics 191, 191–195. Springer, New York, Heidelberg, Berlin.

Moody's Investors Service, 2002. Default & recovery rates of corporate bond issuers. A Statistical Review of Moody's Ratings Performance, 1920-2001.

Moody's Investors Service, 2003. Default & recovery rates of corporate bond issuers. A Statistical Review of Moody's Ratings Performance, 1920-2002.

Prahl, J., 1999. A fast unbinned test on event clustering in Poisson processes. Working Paper, University of Hamburg.

Therneau, T. M., Grambsch, P. M., 2001. Modeling survival data. Extending the Cox model. Statistics for Biology and Heath. Springer, Berlin.

Yoshizawa, Y., 2003. Moody's approach to rating synthetic CDOs. Moody's Investors Service.

Table 1: Maximum likelihood estimates of intensity models in five specifications (in-sample). Each model includes the estimated baseline component. Standard errors in parentheses.

	Models				
	Basic dataset		*Subset*		
	I	II	III (DDKS)	IV	V
Constant	2.78	1.95	-1.11	-2.76	-2.36
	(0.192)	(0.280)	(0.407)	(0.470)	(0.448)
S&P 500	–	0.591	1.23	2.48	2.16
	–	(0.339)	(0.339)	(0.380)	(0.379)
3-month rate	–	0.289	0.0490	0.217	0.343
	–	(0.0376)	(0.0331)	(0.0368)	(0.0392)
Industrial production	–	-0.0547	–	-0.0830	-0.0589
	–	(0.0215)	–	(0.0224)	(0.0228)
Spread 10year-1year	–	0.340	–	0.514	0.586
	–	(0.0698)	–	(0.0736)	(0.0740)
Log(EDF)	2.22	1.64	1.47	1.43	0.902
	(0.0995)	(0.104)	(0.114)	(0.113)	(0.118)
Stock return	–	–	-3.64	-4.04	-3.61
	–	–	(0.303)	(0.314)	(0.292)
Rating	–	0.952	–		0.866
	–	(0.0677)	–		(0.0744)
Log-Likelihood	303	430	354	391	465

Table 2: Accuracy ratios of the different regression intensity models averaged from 1985 on and from 1993 on. Rolling estimations have been conducted with five-year estimation windows. Accuracy ratios have been calculated for each month and then averaged over time. Models III, IV and V add the stock return and are based on a data subset for which the stock return was available.

	1985-2005		1993-2005	
	1-year horizon	5-year horizon	1-year horizon	5-year horizon
Fixed coefficients (in-sample)				
Model I	0.864	0.704	0.867	0.675
Model II	0.873	0.732	0.876	0.716
Model III (DDKS)	0.857	0.654	0.843	0.603
Model IV	0.854	0.644	0.838	0.591
Model V	0.873	0.706	0.863	0.673
Rolling estimation (out-of-sample)				
Model I	0.864	0.704	0.867	0.675
Model II	0.872	0.730	0.876	0.712
Model III	0.851	0.655	0.836	0.595
Model IV	0.851	0.655	0.835	0.596
Model V	0.867	0.703	0.857	0.663
Compared to prior studies (out-of-sample)				
Duffie, Saita and Wang (2007)			0.88	0.69

Table 3: p-values of Fisher's dispersion test for bin sizes 1, 2, 4, 6, 8, 10, 16. Intensity models I to V differ in explanatory variables (models II, IV and V add the term spread and industrial production, models II and V add the rating). Models III, IV and V add the stock return and are based on a subset of our basic data set for which the stock return was available. Results are presented for all combinations of two further estimation specifications: in-sample or out-of-sample estimation, estimation with or without baseline component. The baseline component captures the intra-month pattern of defaults. Low p-values would lead to rejection of the doubly stochastic assumption. Asterisks flag rejections at particular significance levels: *** refers to the 0.1%, ** to the 1% and * to the 5% significance level.

i) Fixed coefficients (in-sample):

Without baseline component

	1	2	4	6	8	10	16
Model I	0.0003***	0.0282*	0.0036**	0.0129*	0.0003***	0.0002***	0.0000***
Model II	0.0585	0.3156	0.4668	0.0417*	0.1811	0.0428*	0.0239*
Model III (DDKS)	0.0000***	0.0000***	0.0000***	0.0000***	0.0000***	0.0000***	0.0000***
Model IV	0.3005	0.4207	0.1675	0.0513	0.1970	0.1369	0.0404*
Model V	0.3135	0.4098	0.3118	0.1239	0.0783	0.0323*	0.0909

With baseline component

	1	2	4	6	8	10	16
Model I	0.0008***	0.0420*	0.0071**	0.0101*	0.0003***	0.0002***	0.0000***
Model II	0.2907	0.4866	0.4666	0.0366*	0.2198	0.0428*	0.0239*
Model III	0.0001***	0.0000***	0.0000***	0.0000***	0.0000***	0.0000***	0.0000***
Model IV	0.2872	0.5625	0.1860	0.0781	0.1770	0.1111	0.0333*
Model V	0.2271	0.4854	0.4256	0.2343	0.0601	0.0510	0.0680

ii) Rolling estimation (out-of-sample):

Without baseline component

	1	2	4	6	8	10	16
Model I	0.0011**	0.0413*	0.0217*	0.0044**	0.0001***	0.0003***	0.0000***
Model II	0.2016	0.6718	0.8233	0.4783	0.4014	0.2412	0.1010
Model III	0.3908	0.1350	0.0485*	0.0171*	0.0024**	0.0016**	0.0067**
Model IV	0.7730	0.3308	0.3690	0.0162*	0.0401*	0.1011	0.0022**
Model V	0.7007	0.7587	0.4024	0.7902	0.4983	0.2737	0.5440

With baseline component

	1	2	4	6	8	10	16
Model I	0.0006***	0.0224*	0.0278*	0.0047**	0.0001***	0.0003***	0.0000***
Model II	0.5574	0.9105	0.8795	0.5110	0.5032	0.2120	0.1153
Model III	0.7269	0.3296	0.0517	0.0148*	0.0060**	0.0024**	0.0110*
Model IV	0.6470	0.3677	0.0397*	0.0243*	0.0134*	0.1497	0.0026**
Model V	0.7010	0.8737	0.7095	0.8556	0.6813	0.2362	0.5207

Table 4: p-values of the upper quartile test for bin sizes 1, 2, 4, 6, 8, 10, 16. Intensity models I to V differ in explanatory variables (models II, IV and V add the term spread and industrial production, models II and V add the rating). Models III, IV and V add the stock return and are based on a subset of our basic data set for which the stock return was available. Results are presented for all combinations of two further estimation specifications: in-sample or out-of-sample estimation, estimation with or without baseline component. The baseline component captures the intra-month pattern of defaults. Low p-values would lead to rejection of the doubly stochastic assumption. Asterisks flag rejections at particular significance levels: *** refers to the 0.1%, ** to the 1% and * to the 5% significance level.

i) Fixed coefficients (in-sample):

Without baseline component

	1	2	4	6	8	10	16
Model I	0.0086**	0.1141	0.0799	0.0735	0.0314*	0.0104*	0.0046**
Model II	0.2608	0.4059	0.4063	0.0938	0.2050	0.1628	0.0362*
Model III (DDKS)	0.0062**	0.0016**	0.0000***	0.0000***	0.0000***	0.0000***	0.0000***
Model IV	0.5106	0.5752	0.3019	0.3007	0.2766	0.1995	0.1179
Model V	0.3524	0.4653	0.3002	0.3899	0.1553	0.1344	0.2150

With baseline component

	1	2	4	6	8	10	16
Model I	0.0198*	0.1710	0.1342	0.0608	0.0314*	0.0144*	0.0046**
Model II	0.4557	0.5274	0.4586	0.0735	0.2055	0.1618	0.0361*
Model III	0.0170*	0.0059**	0.0000***	0.0000***	0.0000***	0.0000***	0.0000***
Model IV	0.5101	0.7553	0.3012	0.3432	0.3192	0.2056	0.1175
Model V	0.3157	0.4769	0.3456	0.3946	0.1332	0.1659	0.2165

ii) Rolling estimation (out-of-sample):

Without baseline component

	1	2	4	6	8	10	16
Model I	0.0242*	0.1765	0.0903	0.0373*	0.0100**	0.0102*	0.0027**
Model II	0.6992	0.9581	0.9892	0.9412	0.9748	0.9342	0.9556
Model III	0.7604	0.7830	0.5706	0.4103	0.5334	0.5555	0.8738
Model IV	0.9057	0.8423	0.9134	0.6552	0.7155	0.8510	0.6128
Model V	0.8292	0.9525	0.9424	0.9804	0.9783	0.9050	0.9792

With baseline component

	1	2	4	6	8	10	16
Model I	0.0173*	0.1094	0.1114	0.0476*	0.0100**	0.0102*	0.0027**
Model II	0.8607	0.9903	0.9892	0.9476	0.9840	0.9218	0.9411
Model III	0.8618	0.8538	0.6100	0.3592	0.5819	0.5555	0.8738
Model IV	0.8760	0.9197	0.7530	0.6552	0.6582	0.9173	0.6128
Model V	0.8704	0.9684	0.9575	0.9905	0.9878	0.9073	0.9806

Table 5: p-values of Prahl's test. Intensity models I to V differ in explanatory variables (models II, IV and V add the term spread and industrial production, models II and V add the rating). Models III, IV and V add the stock return and are based on a subset of our basic data set for which the stock return was available. Results are presented for all combinations of two further estimation specifications: in-sample or out-of-sample estimation, estimation with or without baseline component, original and adjusted defaults. The baseline component captures the intra-month pattern of defaults. Low p-values would lead to rejection of the doubly stochastic assumption. Asterisks flag rejections at particular significance levels: *** refers to the 0.1%, ** to the 1% and * to the 5% significance level.

	Models				
	I	II	III (DDKS)	IV	V
i) Fixed coefficients (in-sample):					
Without baseline component	0.0007***	0.0112*	0.0000***	0.0788	0.0869
With baseline component	0.0010***	0.0177*	0.0000***	0.1058	0.1259
ii) Rolling estimation (out-of-sample):					
Without baseline component	0.0014**	0.0029**	0.0000***	0.0001***	0.0005***
With baseline component	0.0019**	0.0046**	0.0000***	0.0001***	0.0013**

Table 6: p-values of the autocorrelation test for bin sizes 1, 2, 4, 6, 8, 10, 16. Intensity models I to V differ in explanatory variables (models II, IV and V add the term spread and industrial production, models II and V add the rating). Models III, IV and V add the stock return and are based on a subset of our basic data set for which the stock return was available. Results are presented for all combinations of two further estimation specifications: in-sample or out-of-sample estimation, estimation with or without baseline component. The baseline component captures the intra-month pattern of defaults. Low p-values would lead to rejection of the doubly stochastic assumption. Asterisks flag rejections at particular significance levels: *** refers to the 0.1%, ** to the 1% and * to the 5% significance level.

i) Fixed coefficients (in-sample):

Without baseline component

	1	2	4	6	8	10	16
Model I	0.1999	0.0433*	0.0363*	0.0030**	0.0606	0.0227*	0.3682
Model II	0.3670	0.3970	0.0049 **	0.3376	0.0137 *	0.2303	0.9020
Model III (DDKS)	0.0001***	0.0000***	0.0000***	0.0000***	0.0005***	0.0004***	0.0486*
Model IV	0.6335	0.0643	0.4225	0.8357	0.4838	0.5473	0.5269
Model V	0.8740	0.4491	0.1096	0.1463	0.3687	0.7228	0.7839

With baseline component

	1	2	4	6	8	10	16
Model I	0.2718	0.0369*	0.0170*	0.0044**	0.0665	0.0267*	0.3261
Model II	0.6984	0.1073	0.0062 **	0.3173	0.0074 **	0.2303	0.9020
Model III	0.0004***	0.0004***	0.0000***	0.0000***	0.0001***	0.0002***	0.0456*
Model IV	0.4289	0.0470*	0.3249	0.7192	0.4909	0.7394	0.5151
Model V	0.6794	0.3744	0.0338 *	0.0632	0.4625	0.6709	0.8366

ii) Rolling estimation (out-of-sample):

Without baseline component

	1	2	4	6	8	10	16
Model I	0.5918	0.0166*	0.0059**	0.0019**	0.0258*	0.0382*	0.8885
Model II	0.0146 *	0.9313	0.4729	0.8869	0.7428	0.5662	0.5560
Model III	0.0169*	0.0044**	0.0252*	0.0199*	0.1064	0.1745	0.2253
Model IV	0.4337	0.9047	0.0400*	0.3412	0.2804	0.1579	0.8427
Model V	0.8689	0.7509	0.6076	0.6077	0.7640	0.6844	0.2808

With baseline component

	1	2	4	6	8	10	16
Model I	0.6540	0.0387*	0.0046**	0.0025**	0.0197*	0.0495*	0.8116
Model II	0.0142 *	0.4120	0.3394	0.7246	0.4713	0.6971	0.4936
Model III	0.0092**	0.0097**	0.0460*	0.0131*	0.0712	0.2115	0.2414
Model IV	0.4073	0.3001	0.6773	0.2733	0.4077	0.0648	0.8997
Model V	0.5575	0.9597	0.9410	0.3640	0.4844	0.6900	0.4342

Table 7: p-values of Fisher's dispersion test, the upper quartile test and the autocorrelation test for bin sizes 1, 2, 4, 6, 8, 10, 16 and Prahl's test for the estimation based on month-end and middle-of-the-month values. Results are presented for out-of-sample estimation with baseline for the specifications with stock return. Intensity models III to V differ in the other explanatory variables (IV and V add the term spread and industrial production, model V adds the rating). Low p-values would lead to rejection of the doubly stochastic assumption. Asterisks flag rejections at particular significance levels: *** refers to the 0.1%, ** to the 1% and * to the 5% significance level.

	Fisher's dispersion test						
	1	2	4	6	8	10	16
Model III	0.7483	0.5503	0.0962	0.0714	0.1189	0.0841	0.0377*
Model IV	0.8343	0.7399	0.6454	0.2461	0.2489	0.4677	0.2186
Model V	0.7445	0.7906	0.7970	0.6776	0.5848	0.4730	0.2418
	Upper quartile test						
	1	2	4	6	8	10	16
Model III	0.8274	0.9096	0.6834	0.5881	0.7900	0.8538	0.8776
Model IV	0.9240	0.9445	0.9734	0.8727	0.9564	0.9908	0.9697
Model V	0.9304	0.9593	0.9844	0.9943	0.9883	0.9782	0.9400
	Autocorrelation test						
	1	2	4	6	8	10	16
Model III	0.4321	0.0715	0.2667	0.2374	0.5015	0.7571	0.9617
Model IV	0.5889	0.3969	0.3875	0.8499	0.9676	0.8168	0.3172
Model V	0.2586	0.6195	0.6470	0.9461	0.7342	0.5917	0.7510
	Prahl's test						
Model III	0.0081**						
Model IV	0.0150*						
Model V	0.0673						

Figure 1: Monthly default rate in the data set in the time period from 02/1980 to 04/2005.

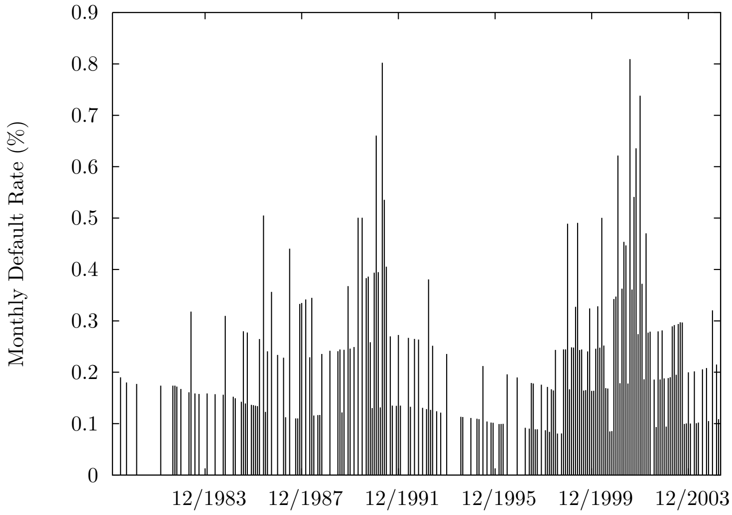


Figure 2: Histogram of intra-month default dates in the data set (left) and corresponding estimated distribution (right).

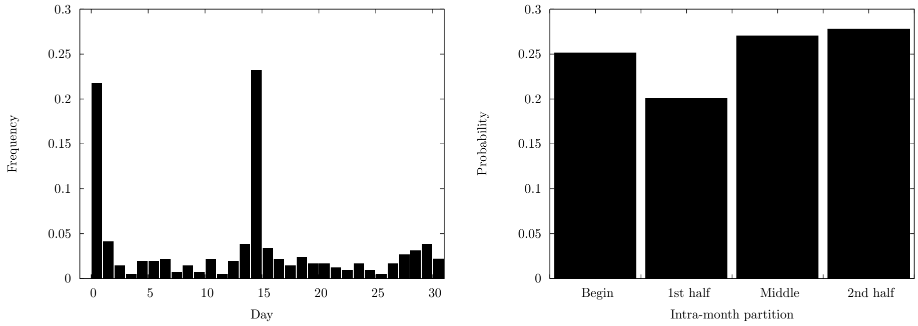


Figure 3: Histogram of intra-month default dates for the first (left) and the second half of defaults (right) in the data set.

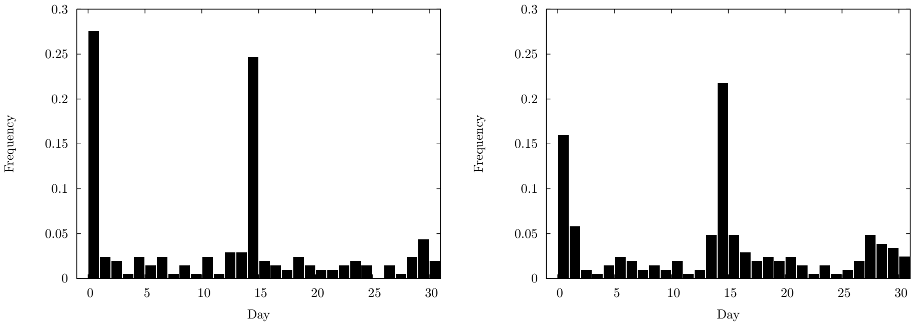
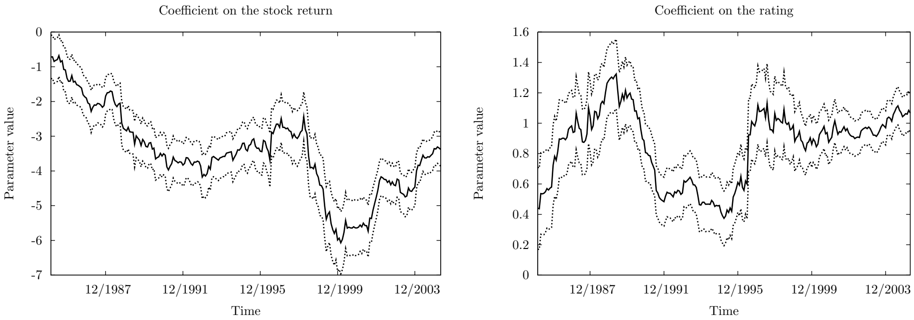


Figure 4: The estimated coefficient on the stock return (solid, left) and the estimated coefficient on the rating (solid, right) with corresponding one-standard-error bands (dotted) for model V.



Appendix

Discussion of the results presented in section 2

In this section we give a mathematically more detailed discussion of the results presented in section 2. To obtain a conditional independence setup, we follow Lando (1998) and proceed as follows: We fix a time horizon T^* and consider a filtered probability space $\left(\Omega, \mathcal{F}, \mathbf{F} = (\mathcal{F}_t)_{0\leq t\leq T^*}, P\right)$ where $\mathbf{F}$ is assumed to satisfy the usual conditions, i.e. P-completeness and right continuity. We further assume that an $\mathbf{F}$-adapted, d-dimensional stochastic process $X = (X_t)_{0\leq t\leq T^*}$, which is assumed to be right continuous with left limits (RCLL), as well as positive, continuous functions $\lambda_i(\cdot,\cdot) : \mathbb{R}\times\mathbb{R}^d \to \mathbb{R}_+$ are given. We define the default time τ_i of firm i as the first jump time of a Cox-process with intensity $\lambda_i(t, X(t))$, i.e.

$$\tau_i = \inf\left\{t : \int_0^t \lambda_i(s, X(s))\, ds \;\geq\; E_i\right\},$$

where E_i is a unit-exponentially distributed random variable independent of X and E_i and E_j are assumed to be independent for two different firms $i \neq j$. Disregarding extreme cases which are practically of no relevance such as exploding intensities, our definition of the default times entails that $\Delta 1_{\tau_i\leq t}\Delta 1_{\tau_j\leq t} = 0$. Furthermore, with $\mathbf{G} = (\mathcal{G}_t)_{0\leq t\leq T^*}$ denoting the subfiltration generated by $(X(t))_{0\leq t\leq T^*}$,

$$\begin{aligned}
P\left(\tau_i > x \,|\, \mathcal{G}_{T^*} \wedge \{\tau_i > t\}\right) &= \exp\left(-\int_t^x \lambda_i(s, X(s))ds\right) \\
P\left(\tau_i > x, \tau_j > x \,|\, \mathcal{G}_{T^*} \wedge \bigcap_{i,j}\{\tau_i > t\}\right) &= \prod_{i,j}\exp\left(-\int_t^x \lambda_i(s, X(s))ds\right) \quad i\neq j, x \leq T^*.
\end{aligned}$$

For the first equality see Lando (1998) and the second equality follows from the fact that conditional on $\mathcal{G}_{T^*}$ the default times are independent since the E_i are mutually independent and independent of X. Thus, contagion is not possible in our setup because $\mathcal{G}_{T^*}$ is independent of $\bigvee_{1\leq i\leq I}\sigma(E_i)$ and since conditional on $\mathcal{G}_{T^*}$ defaults are independent, dependence solely originates from the common dependence of the intensities on the process X. In particular, the likelihood function, given by Equation (1), is a direct consequence of this model property (note that the $\mathcal{G}_{T^*}$-conditional density of τ_i is $\lambda_i(t, X(t))\exp\left(-\int_t^x \lambda_i(s, X(s))ds\right)$ and conditional on $\mathcal{G}_{T^*}$ defaults are independent).

We use these standard results now in order to show that the time-scaled process M of Equation (2) is a Poisson process. It has been pointed out by Lando and Nielsen (2009) that this time transformation result applies to a much broader class of processes (cf. Meyer (1971)) where the intensities of the default times may be affected by defaults in the portfolio and is not restricted to the conditional independence setup. By definition,

$$\int_0^{\tau_i} \lambda_i(s, X(s))ds = E_i,$$

i.e. $\int_0^{\tau_i} \lambda_i(s, X(s))ds \sim Exp(1)$ conditional on $\mathcal{G}_{T^*}$.

With

$$L(t) := \sum_{i=1}^{I} 1_{\tau_i \leq t}$$

and $\tau_{(0)} := 0$ and $\tau_{(i)} := \inf\{t : L(t) \geq i\}$, $1 \leq i \leq I$,

$$\begin{aligned} P\left(\tau_{(i)} - \tau_{(i-1)} > x \middle| \mathcal{G}_{T^*} \wedge \{\tau_{(k)} : k \leq i-1\}\right) &= P\left(\bigcap_{l:\tau_l > \tau_{(i-1)}} \{\tau_l - \tau_{(i-1)} > x\} \middle| \mathcal{G}_{T^*} \wedge \{\tau_{(k)} : k \leq i-1\} \right) \\ &= \prod_{l:\tau_l > \tau_{(i-1)}} \exp\left(- \int_{\tau_{(i-1)}}^{\tau_{(i-1)}+x} \lambda_l(s, X(s))ds \right) \\ &= \exp\left(- \sum_{l:\tau_l > \tau_{(i)}} \int_{\tau_{(i-1)}}^{\tau_{(i-1)}+x} \lambda_l(s, X(s))ds \right). \end{aligned}$$

where we applied the $\mathcal{G}_{T^*}$-conditional independence of the τ_i. This further implies that conditional on $\mathcal{G}_{T^*}$ and $\{\tau_{(k)} : k \leq i-1\}$

$$K_i := \sum_{l:\tau_l > \tau_{(i-1)}} \int_{\tau_{(i-1)}}^{\tau_{(i)}} \lambda_l(s, X(s))ds \ \sim \ Exp(1).$$

Furthermore, independence of the K_i follows by factorizing the characteristic function:

$$\begin{aligned} E\left[\prod_{l=1}^{I} e^{iu_l K_l}\right] &= E\left[E\left[\prod_{l=1}^{I} e^{iu_l K_l} \middle| \mathcal{G}_{T^*} \wedge \{\tau_{(k)} : k \leq I-1\}\right]\right] \\ &= E\left[\prod_{l=1}^{I-1} e^{iu_l K_l} \frac{1}{1 - iu_I}\right] = \frac{1}{1 - iu_I} E\left[\prod_{l=1}^{I-1} e^{iu_l K_l}\right] \\ &= \ldots = \prod_{l=1}^{I} \frac{1}{1 - iu_l}. \end{aligned}$$

Then, by Definition 2.17 of Cont and Tankov (2004), with $T_j := \sum_{i=1}^{j} K_i$, the process

$$\begin{aligned} M(t) &= \sum_{j \geq 1} 1_{T_j \leq t} = \sum_{j \geq 1} 1_{\sum_{i=1}^{I} \int_0^{\tau_{(j)}} 1_{\tau_i > s} \lambda_i(s, X(s))ds \leq t} \\ &= \sum_{j \geq 1} 1_{\tau_{(j)} \leq f^{-1}(t)} = L\left(f^{-1}(t)\right) \qquad \text{with} \\ f(t) &:= \sum_{i=1}^{I} \int_0^t 1_{\tau_i > s} \lambda_i(s, X(s))ds \end{aligned}$$

follows a Poisson process with intensity 1.