Self-Exciting Corporate Defaults: Contagion vs. Frailty

Shahriar Azizpour and Kay Giesecke*

Stanford University

February 12, 2008; this draft August 29, 2008†

Abstract

Why do corporate defaults cluster? This paper explores the role of contagion, by which the default of a firm has a direct impact on the conditional default rates of the surviving firms, channeled through the complex web of contractual relationships in the economy. We develop filtered maximum likelihood estimators and goodness-of-fit tests for point processes to measure the additional impact of contagion on default rates, over and beyond that due to firms' exposure to observable or unobservable (frailty) risk factors. For U.S. firms during 1970–2006, we find strong evidence that contagion represents a significant additional source of default clustering.

Keywords: Correlated default, event feedback, contagion, frailty, self-exciting point process, intensity, point process filtering and smoothing, measure change, likelihood, goodness-of-fit, time change.

*Corresponding author. Department of Management Science & Engineering, Stanford University, Stanford, CA 94305-4026, USA, Phone (650) 723 9265, Fax (650) 723 1614, email: `giesecke@stanford.edu`, web: `www.stanford.edu/~giesecke`.

†We are grateful for grants from Moody's Corporation and the Global Association of Risk Professionals, and for data from Moody's. We thank Richard Cantor for providing access to data, Lisa Goldberg for discussions and earlier joint work on fitness tests for point processes, Andreas Eckner, Jean-David Fermanian, Francis Longstaff, George Papanicolaou, Bjorgvin Sigurdsson, Ilya Strebulaev, Stefan Weber, Liuren Wu and seminar participants at the Bachelier World Congress, the Princeton University Conference on the Mathematics of Defaultable Securities and the Stanford University Financial Mathematics Seminar for discussions and comments, and Xiaowei Ding, Baeho Kim and Supakorn Mudchanatongsuk for excellent research assistance.

1 Introduction

Doubly stochastic models of clustered event timing are widely used in theoretical and empirical analyses of correlated corporate default risk. In these models, firms are related through their exposure to observable risk factors. The co-movements of the factors induce correlated changes in firms' default rates. While the doubly stochastic assumption simplifies model computation and estimation, Das, Duffie, Kapadia & Saita (2007) show that it is overly restrictive relative to the default history of U.S. industrial firms between 1979 and 2004. They find evidence of default clustering in excess of that implied by a standard doubly stochastic model. Their results highlight the importance of other channels of default correlation that are ignored in a doubly stochastic model.

There is a growing literature addressing the empirical evidence by assuming that some of the risk factors the firms are exposed to may not be observable. A default reveals information about the unobservable "frailty" factors and therefore has an influence on the conditional default rates of any other firms that depend on the same factors. This impact generates an additional channel of default correlation that is absent in a doubly stochastic setting. For U.S. industrial firms during 1979–2004, Duffie, Eckner, Horel & Saita (2008) find strong evidence for the presence of unobservable risk factors. They estimate that frailty has a large impact on fitted conditional mean default rates, above and beyond those predicted by observable default covariates.

There are other potential sources for the excess default clustering documented by Das et al. (2007). This article explores the empirical role of contagion, by which the default of a firm has a negative impact on the other firms. This impact is channeled through the complex web of business, legal and other contractual relationships in the economy. In such a network, the failure of one firm tends to weaken the others. For instance, given the collapse of a borrower or supplier, a creditor or customer is much more likely to default as well. A recent example is the bankruptcy of auto parts manufacturer Delphi in 2005, which jeopardized production at General Motors, Delphi's main purchaser, and brought General Motors to the brink of bankruptcy.

This paper develops statistical methods to ascertain the degree of contagion that has been present for U.S. corporate defaults, and to measure the relative empirical importance of contagion and frailty for explaining the default clustering not captured by a doubly stochastic model. Both contagion and frailty generate "self-exciting" phenomena, in that a default tends to increase the likelihood of other firms to fail. Yet, they have different economic foundations. While frailty effects are purely informational, contagion is based on the propagation of financial distress through contractual linkages among firms. We wish to gain insights into the prevalence of the economic mechanisms that generate the non-doubly stochastic default correlation.

For U.S. industrial and financial firms during 1970–2006, we find strong evidence for the presence of contagion, even when controlling for firms' exposure to common observable or frailty risk factors. A default is estimated to have a substantial contagious impact on the conditional default rates of the surviving firms, above and beyond the influence due to learning about a frailty factor. Contagion and frailty phenomena are found to be roughly equally important for explaining the default clustering in the data that is not captured by a doubly stochastic model. While frailty has been a major focus of the theoretical and empirical credit risk literature in recent years, these findings suggest that researchers should pay as much attention to modeling contagion as they do to modeling frailty.

The estimation and testing of a default timing model that incorporates firms' exposure to common factors and accounts for both contagion and frailty phenomena poses substantial challenges. A major problem is that due to the contagious impact of an event, the doubly stochastic hypothesis is violated even if the information filtration of the econometrician is artificially enlarged to include the frailty risk factors that are unobserved in practice. Therefore, we cannot reap the significant computational benefits of the doubly stochastic assumption in the complete information filtration. This is in contrast to a model without contagion, for which the doubly stochastic assumption in the complete information filtration allows the econometrician to formulate the estimation problem in a "conditionally doubly stochastic" setting even in the presence of frailty.

We confront this challenge by casting the estimation and testing problems in a stochastic filtering framework. Our event timing model is formulated in terms of an intensity process that represents the aggregate conditional default rate in a universe of firms. The intensity jumps in response to defaults so an event tends to increase the likelihood of further events. Random movements in the intensity between events arise from firms' exposure to a common Feller diffusion risk factor. If the path of this factor cannot be identified from the covariate observations, then the factor is a frailty whose values are filtered from the default arrival data. We use measure change arguments to develop computationally tractable point process filters that facilitate a full maximum likelihood approach to parameter estimation in the presence of frailty and contagion effects. An equivalent change of measure transforms the point process describing event arrivals into a standard Poisson process, allowing us to express the likelihood function in terms of a conditional expectation under the new measure of the corresponding Radon-Nikodym derivative. This argument also facilitates the calculation of the conditional posterior mean of the frailty factor and the intensity given the default arrival observations. The filtering approach enables us to avoid the simulation methods commonly used in the frailty literature.

A stochastic time change argument is used to construct statistical tests for evaluating the fit of different specifications of the complete information intensity and the econometrician's observation filtration, and to identify the model features that are empirically relevant. A change of time defined by the cumulative *filtered* intensity transforms the point process describing event arrivals into a new point process, which must be standard Poisson if the intensity model *and* the econometrician's observation filtration are correctly specified. The Poisson property of the time-scaled arrival times can be tested with a battery of statistical tests. The testing methodology is sufficiently general to allow us to contrast the fit of self-exciting model specifications in different observation filtrations.

Therefore, we can assess models with and without frailty, and examine the relative empirical importance of frailty and contagion effects. Moreover, the fit of any self-exciting model can be measured on the same scale as the fit of a doubly stochastic model, considered in Das et al. (2007) with a time-scaling test based on the doubly stochastic hypothesis with complete information. This enables us to directly compare the ability of these opposing model specifications to capture the default clustering in the arrival data, and the substantial time-series variation of U.S. corporate default rates.

The remainder of this introduction discusses the implications of our work for portfolio credit risk, the related literature and our data set. Section 2 introduces our stochastic model of correlated event timing. Section 3 develops a point process filtering approach to the maximum likelihood estimation of the model parameters. Appendix B provides a recursive algorithm to compute the likelihood function. Section 4 explains the calculation of the filtered intensity, which is an input to the goodness-of-fit tests developed in Section 5. Section 6 discusses the results of the estimation and in-sample goodness-of-fit tests, and the time-series behavior of the fitted filtered intensity. Different model variants and informational hypotheses are contrasted. Section 7 analyzes out-of-sample fitting and prediction performance and examines the empirical implications of the self-exciting property for portfolio credit risk. Section 8 concludes. Proofs are found in Appendix A.

1.1 Portfolio credit risk

We briefly explain the implications of our work for portfolio credit risk. Our estimators indicate that the feedback from defaults induced by contagion or frailty has a large impact on the credit risk of portfolios of U.S. corporate debt. All else equal, default feedback causes a significant increase in the likelihood of large losses due to correlated default, over and above that predicted by a doubly stochastic model. Moreover, we find that the shape and term structure of the fitted conditional distribution of future events strongly depends on the number and timing of past defaults. Our goodness-of-fit tests indicate that the sensitivity of the arrival intensity to defaults is statistically important for replicating the timing and size of observed event clusters. For example, a simple intensity model that incorporates this sensitivity can successfully predict the event bursts during the internet bubble from 1999 to 2002, out-of-sample. This is because fitted default rates respond immediately when an event cluster starts to unfold.

The ongoing credit crisis highlights the need for portfolio credit risk estimates that account for the feedback from defaults. Accurate forecasts of correlated event arrivals over multiple future time periods are particularly important for the risk management of corporate debt portfolios. Standard models used by financial institutions to estimate the amount of risk capital to be held in order to withstand large losses due to correlated default rely on the doubly stochastic hypothesis. Since these models ignore the feedback from defaults, the loss distribution they forecast tends to understate the probability of large losses, especially in times of economic stress. Therefore, these models can lead to insufficient risk capital provisions and an increase in systemic risk. An understanding of default clustering is also important for the risk analysis of structured credit instruments such as collateralized debt obligations (CDOs). Default feedback plays a significant role for investors in senior tranches of CDOs, who experience losses only if total defaults in the underlying portfolio reach extreme levels. If an event timing model used for the risk analysis of these instruments ignores the feedback from defaults, then the risk of these tranches may be severely understated.

1.2 Related literature

The frailty models of Collin-Dufresne, Goldstein & Helwege (2003), Delloye, Fermanian & Sbai (2006), Duffie et al. (2008), Giampieri, Davis & Crowder (2005), Giesecke (2004), Giesecke & Goldberg (2004), Koopman, Lucas & Monteiro (2008), McNeil & Wendin (2007), Schönbucher (2004) and others attribute any influence of a default on the surviving firms to Bayesian updating of the conditional distribution of the unobservable frailty factors. Our model permits an event to have an additional impact on the other firms through contagion in a network of contractually related firms. This influence represents an additional channel for default clustering that is active even if all risk factors are observable. Contagion phenomena are not anecdotal: Lang & Stulz (1992) and Jorion & Zhang (2007*b*) find evidence that intra-industry contagion is often associated with Chapter 11 bankruptcies. Jorion & Zhang (2007*a*) find that a firm's default leads to abnormally negative equity returns for the firm's creditors along with increases in creditors' credit swap spreads, which represent the price for protection against default.

Chava & Jarrow (2004), Das et al. (2007), Das, Freed, Geng & Kapadia (2006), Duffie, Saita & Wang (2006), Duffie et al. (2008), Figlewski, Frydman & Liang (2008) and others fit bottom-up models of correlated default intensities for individual firms.[1] Our approach is conceptually different. We estimate the aggregate intensity, which is given by the sum over firms of the individual firm intensities. This top-down approach leads to parsimony, identifiability and computational tractability of likelihood estimation and event prediction, but requires a further random thinning step (Giesecke & Goldberg (2005)) to obtain the intensities of individual firms. Azizpour & Giesecke (2008) implement this step on roughly the same default data studied here to obtain estimates of firm intensities per firm rating category. These estimates are then used to extract premia for correlated default risk from prices of CDOs.

Our results establishing the empirical significance of contagion provide support for the modeling assumptions in a large literature developing default timing models under risk-neutral probabilities for the purpose of pricing and hedging securities that are referenced on portfolios of corporate debt, for example Arnsdorf & Halperin (2007), Cont & Minca (2008), Davis & Lo (2001), Ding, Giesecke & Tomecek (2006), Errais, Giesecke & Goldberg

[1]Shumway (2001) and Hillegeist, Keating, Cram & Lunstedt (2004) estimate single firm intensity (hazard) models that ignore the effects of default correlation among firms.

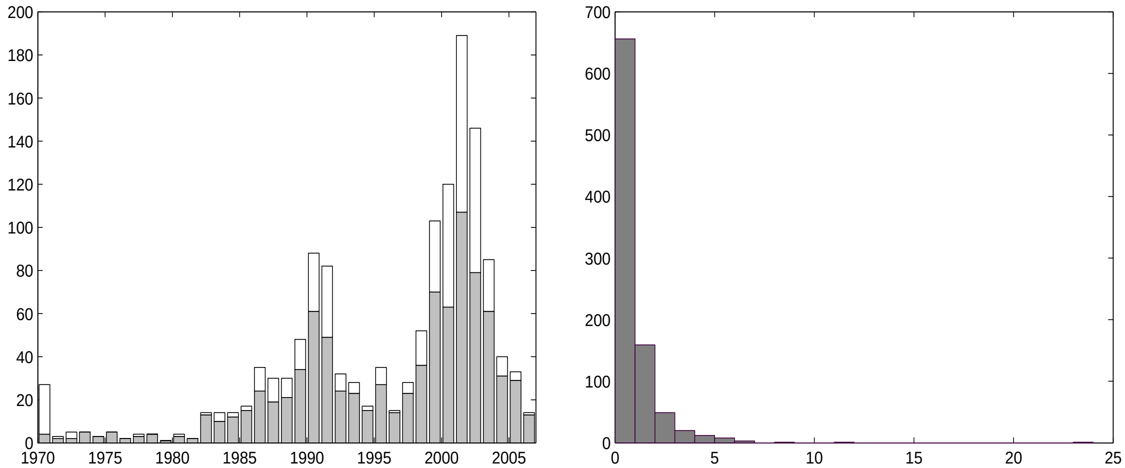

Figure 1: *Left panel*: Annual default data. A bar represents the number of events in a given year. The shaded part of the bar represents the number of days over which these events are distributed. *Right panel*: Number of events per day. A bar represents the number of days in the sample with a given number of events. The largest number of events was observed on June 21, 1970, when 24 railway firms defaulted.

(2006), Giesecke (2007), Giesecke & Tomecek (2005), Frey & Backhaus (2008), Jarrow & Yu (2001), Longstaff & Rajan (2008), Lopatin & Misirpashaev (2007) and Tavella & Krekel (2006). In particular, our empirical results motivate valuation models that account for both contagion and unobserved risk factors, as in Frey & Runggaldier (2007).

1.3 Data

Data on default timing are from Moody's Default Risk Service, which provides detailed issue and issuer information on industry, rating, date and type of default, and other items. Our sample period is January 1970 to October 2006. An issuer is included in our data set if it is not a sovereign and has a senior rating, which is an issuer-level rating generated by Moody's from ratings of particular debt obligations using its Senior Rating Algorithm described in Hamilton (2005). As of October 2006, the data set includes a total of 6048 firms, of which 3215 are investment grade rated issuers.

For our purposes, a "default" is a credit event in any of the following Moody's default categories: (1) A missed or delayed disbursement of interest or principal, including delayed payments made within a grace period; (2) Bankruptcy (Section 77, Chapter 10, Chapter 11, Chapter 7, Prepackaged Chapter 11), administration, legal receivership, or other legal blocks to the timely payment of interest or principal; (3) A distressed exchange occurs where: (i) the issuer offers debt holders a new security or package of securities that amount to a diminished financial obligation; or (ii) the exchange had the apparent purpose of helping the borrower avoid default.

We observe 1374 defaults during the sample period. Each default has a time stamp; the resolution is one day. There are days with multiple events whose exact timing during the day cannot be distinguished. Therefore, we view the default data as a realization of a marked point process $(T_n, D_n)_{n=1,\ldots,909}$, where T_n represents a date with at least one default incidence and D_n is the number of defaults at T_n. Note that the sequence (T_n) is strictly increasing, and that $\sum_{n=1}^{909} D_n = 1374$. The data are visualized in Figure 1. Roughly 51% of the defaults are due to missed interest payments and 25% are due to Chapter 11. Of the defaulted firms, 84% are of Moody's "Industrial" category and 85% are domiciled in the United States.

A repeated default by the same issuer is included in the set of events if it was not within a year of the initial event and the issuer's rating was raised above Caa after the initial default. This treatment of repeated defaults is consistent with that of Moody's. There are a total of 46 repeated defaults in our database.

Data on explanatory covariates that influence the likelihood of default are from various sources. Treasury bill yield data are downloaded from the Fed's website. S&P 500 return data is obtained from Economagic.

2 Default timing model

This section introduces our reduced-form model of the marked point process (T_n, D_n). Defaults arrive at increasing dates T_n that generate a non-explosive counting process N with increments of size 1 and intensity λ, relative to a complete probability space $(\Omega, \mathcal{F}, P)$ and a complete information filtration $\mathbb{F} = (\mathcal{F}_t)_{t\geq 0}$ satisfying the usual conditions of right-continuity and P-completeness, where P is the actual, data-generating probability measure. This means that the compensated process $N - \int_0^{\cdot} \lambda_s ds$ is a local martingale relative to $\mathbb{F}$. Intuitively, conditional on the information set $\mathcal{F}_t$, the probability of events between times t and $t + \Delta$ is approximately $\lambda_t \Delta$ for small Δ.

Motivated by the results of Das et al. (2007), who find that a standard doubly stochastic model of firm default intensities fails to fully capture the clustering of defaults of U.S. industrial firms between 1979 and 2004, we suppose that the intensity evolves through time according to the self-exciting model

$$d\lambda_t = \kappa(c - \lambda_t)dt + \sigma\sqrt{\lambda_t}dW_t + \delta dL_t \tag{1}$$

where $\kappa \geq 0$, $c > 0$, $\sigma \geq 0$, $\delta \geq 0$ and $\lambda_0 > 0$ are parameters such that $2\kappa c \geq \sigma^2$ and $\int_0^\tau (\lambda_s^{-1} - 1)^2 \lambda_s ds$ is finite almost surely for some fixed $\tau > 0$, W is a standard Brownian motion relative to $\mathbb{F}$ and L is a response jump process given by

$$L = \sum_{n=1}^{N} \ell(D_n) \tag{2}$$

where ℓ is a positive and bounded weight function that will be specified later and $D_n \in \mathcal{F}_{T_n}$ is the number of defaults at the event date T_n. We assume that D_n is drawn from a fixed distribution, independently of $\mathcal{F}_{T_n-}$. The response process L jumps on an event date and so does the intensity, representing a temporary increase in the likelihood of further arrivals. Along with the sensitivity parameter δ, the weight function ℓ governs the response of the intensity to the number of defaults at an event date. For example, if ℓ is increasing, then the more events the stronger the feedback from L to λ. After an event the intensity reverts back to the level c, exponentially in mean at rate κ. It exhibits diffusive fluctuations that are driven by the Brownian motion W. The diffusive volatility is governed by σ.[2]

The intensity model (1) incorporates several channels for default correlation. First, through the response term δL, a default has a direct impact on the surviving firms that is propagated through the complex web of contractual relationships in the economy. In this firm network, the failure of an entity tends to weaken the other firms, a phenomenon we refer to as contagion. Second, firms are exposed to a systematic risk factor W, whose movements cause correlated changes in firms' conditional default probabilities, and the cyclical behavior of failure rates. Third, depending on the structure of the econometrician's right-continuous and complete observation filtration $\mathbb{G} = (\mathcal{G}_t)_{t\geq 0}$, there may also be uncertainty about the current value of W. This occurs when $\mathbb{G}$ is coarser than the complete information filtration $\mathbb{F}$ and W cannot be fully identified, i.e., is not adapted to $\mathbb{G}$. Then, W is a frailty whose values must be filtered from the observations.

The econometrician's *filtered intensity* of N is a process h adapted to the observation filtration $\mathbb{G}$ such that $N - \int_0^{\cdot} h_s ds$ is a local martingale relative to $\mathbb{G}$, assuming $\mathbb{G}$ is fine enough to distinguish N. One such process is given by the optional projection of the complete information intensity λ onto $\mathbb{G}$.[3] This is the unique $\mathbb{G}$-optional process[4] such that $E(\lambda_T 1_{\{T<\infty\}}) = E(h_T 1_{\{T<\infty\}})$ for any $\mathbb{G}$-stopping time T.[5] The filtered intensity represents the econometrician's estimated default rate given the observations, in the sense that h_t is equal to the conditional posterior mean $E(\lambda_t \,|\, \mathcal{G}_t)$ almost surely, for each fixed time t. As the complete information intensity λ, the filtered intensity h is revised at events. The response of h to a default is due to the contagious impact of the event, a property that h inherits from the complete information intensity λ, and Bayesian updating of the $\mathbb{G}$-conditional distribution of the frailty factor W. In a situation with complete information, i.e., if the observation filtration $\mathbb{G}$ is fine enough to distinguish W and the D_n in addition to the events, then the filtered intensity $h = \lambda$. In this case, frailty effects vanish and the

[2]The process (N, L) is an affine point process in the sense of Errais et al. (2006), who provide its characteristic function and consider applications to the valuation of portfolio credit derivatives.

[3]See Segall & Kailath (1975) and Segall, Davis & Kailath (1975). These authors take h to be the predictable rather than the optional projection in order to guarantee the predictability of the filtered intensity. Predictability leads to uniqueness of the filtered intensity in the sense of Brémaud (1980, Theorem T12). We do not require uniqueness.

[4]A $\mathbb{G}$-optional process is a process that is measurable with respect to the sigma-field on $\mathbb{R}_+ \times \Omega$ generated by the collection of all right-continuous processes adapted to $\mathbb{G}$. Thus, a $\mathbb{G}$-optional process is adapted to the observation filtration $\mathbb{G}$.

[5]See Dellacherie & Meyer (1982, Numbers 43 and 44) for more details, including a proof of existence and uniqueness of the optional projection.

feedback from events to arrival rates is interpreted as contagion.

A doubly stochastic model of event timing ignores the feedback effects induced by contagion and frailty. Suppose $\delta = 0$ in the complete information intensity model (1). In this case, the contagion channel is inactive. The intensity λ is adapted to the filtration $(\mathcal{H}_t)_{t\geq 0}$ generated by the Brownian motion W and N is doubly stochastic in the complete information filtration $\mathbb{F}$. That is, $P(N_{t+u} - N_t = n \,|\, \mathcal{F}_t \vee \mathcal{H}_{t+u}) = p(n, \int_t^{t+u} \lambda_s ds)$ for $u > 0$, where $p(n, \alpha) = e^{-\alpha}\alpha^n/n!$ is the probability that a Poisson random variable with parameter α takes the value $n = 0, 1, 2, \ldots$. Note, however, that N need not be doubly stochastic in the observation filtration $\mathbb{G} \subseteq \mathbb{F}$. If $\mathbb{G}$ is too coarse to make W adapted, then the doubly stochastic property cannot hold in $\mathbb{G}$ because the $\mathbb{G}$-intensity h is revised at events due to Bayesian updating of the conditional distribution of W. In this case the feedback from events to arrival rates is attributed to the presence of frailty.

3 Filtered likelihood estimators

This section develops maximum likelihood estimators for the intensity model (1) based on the observation filtration $\mathbb{G}$. This is not a standard exercise because the intensity may not be adapted to $\mathbb{G}$. We confront this problem by formulating the estimation problem in a point process filtering setting.

We suppose the observation filtration $\mathbb{G}$ is the right-continuous and complete filtration generated by the market point process (T_n, D_n) and a process X of explanatory covariates. We take X to be of the form $X_t = u(Y_t, t)$, where $u : \mathbb{R} \times \mathbb{R}_+ \to \mathbb{R}$ and Y is a standard $\mathbb{F}$-Brownian motion independent of the D_n. The covariates are chosen as we proceed. They need not bear information about the Brownian motion W driving λ. If they do not, then W must be filtered from the observations.

The parameter vector to be estimated is (θ, γ, ν), where $\theta = (\kappa, c, \sigma, \delta, \lambda_0, w)$ collects the intensity parameters, w represents the parameters of the weight function ℓ in equation (2), γ represents the parameters of the distribution of the D_n, and ν represents the parameters of the distribution of X. Bayes' formula implies that, for the sample period $[0, \tau]$, the likelihood function of the data takes the form

$$\mathcal{L}_\tau(\theta, \gamma, \nu \,|\, N_\tau, T, D, X) = f_\tau(N_\tau, T; \theta \,|\, D, X) \cdot g(D; \gamma) \cdot p_\tau(X; \nu) \tag{3}$$

where $f_\tau(\cdot, \cdot; \theta \,|\, D, X)$ is the conditional "density" of the event date count N_τ and the event date vector $T = (T_1, \ldots, T_{N_\tau})$ given the event count vector $D = (D_1, \ldots, D_{N_\tau})$ and the path $\{X_t : 0 \leq t \leq \tau\}$ of the covariate, $g(\cdot; \gamma)$ is the probability function of D, and $p_\tau(\cdot; \nu)$ is the density of the covariate path over $[0, \tau]$. The three terms in (3) can be maximized separately to give the full likelihood estimates.

Instead of estimating a parametric model for $g(\cdot; \gamma)$, we describe the D_ns by their empirical distribution. The covariate model $p_\tau(\cdot; \nu)$ is made precise as we proceed. To evaluate the density $f_\tau(\cdot, \cdot; \theta \,|\, D, X)$, we develop an equivalent change of probability measure that transforms the counting process $(N, \mathbb{F})$ into a standard $\mathbb{F}$-Poisson process. Proposition 3.1 shows that the density can then be expressed in terms of a conditional expectation under the equivalent measure of the Radon-Nikodym derivative, given the observations. Appendix A contains the proofs of this and the other results stated below.

Proposition 3.1. *Let $\widehat{E}$ denote expectation under the equivalent measure $\widehat{P}$ on $\mathcal{F}_\tau$ defined by the Radon-Nikodym derivative Z_τ, where*

$$Z_t = \exp\left(-\int_0^t \log(\lambda_{s-})dN_s - \int_0^t (1-\lambda_s)ds\right) \quad t \leq \tau. \tag{4}$$

Then the conditional density of the event dates and counts is given by

$$f_\tau(N_\tau, T; \theta \,|\, D, X) = \widehat{E}(Z_\tau^{-1} e^{-\tau} \,|\, N_\tau, T, D, X). \tag{5}$$

Formula (4) and the definition (1) of the complete information intensity λ show that the Radon-Nikodym derivative Z_τ is a function of the observed event times and counts $(T_n, D_n)_{n=1,\ldots,N_\tau}$ and the risk factor $\{W_t : 0 \leq t \leq \tau\}$. If the path of W can be identified from the covariate path $\{X_t : 0 \leq t \leq \tau\}$, then the conditional expectation on the right hand side of equation (5) is trivial, and the conditional density $f_\tau(N_\tau, T; \theta \,|\, D, X) = Z_\tau^{-1} e^{-\tau}$. If the path of W cannot be identified from the covariate path, then W is a frailty and the conditional expectation is a nontrivial filter. To evaluate this filter, we need to understand the relationship between P-martingales and $\widehat{P}$-martingales in the complete information filtration $\mathbb{F}$. Proposition 3.2 below is a corollary of the Girsanov-Meyer theorem. The relevant filtration is $\mathbb{F}$ unless noted otherwise.

Proposition 3.2. *Let V be a P-local martingale such that the quadratic covariation $[V, N]$ is locally of integrable variation. Denote by $\langle V, N\rangle$ the conditional covariation. Then a $\widehat{P}$-local martingale is defined by*

$$V_t - \int_0^t (\lambda_{s-}^{-1} - 1) d\langle V, N\rangle_s, \quad t \leq \tau. \tag{6}$$

Proposition 3.2 has several important corollaries. A P-local martingale V remains a local martingale under $\widehat{P}$ if it does not have jumps in common with the counting process N. This is because $[V, N] = \langle V, N\rangle = 0$ if V and N do not have common jumps. For example, the P-Brownian motion W that drives the diffusive fluctuation in the intensity remains a Brownian motion under $\widehat{P}$, by Lévy's theorem. Similarly, the P-Brownian motion Y that drives the covariate process X is also a Brownian motion under $\widehat{P}$. However, the jump times of the compensated P-local jump martingale $M = N - \int_0^\cdot \lambda_s ds$ coincide with those of N, and therefore $[M, N] = [N, N] = N$. This implies that $\langle M, N\rangle$, the P-compensator to $[M, N]$, is equal to $\int_0^\cdot \lambda_s ds$ and that $M_t - \int_0^t (1-\lambda_s)ds = N_t - t$ defines a $\widehat{P}$-local martingale. Thus, by Watanabe's theorem, the change of measure transforms the counting process N with P-intensity λ into a standard $\widehat{P}$-Poisson process. Further, $L_t - E[\ell(D_n)]t$ defines a

$\widehat{P}$-local martingale, and therefore the response jump process L introduced in equation (2) is a standard compound Poisson process under $\widehat{P}$. Consequently, the *paths* of L and W over $[0,\tau]$ must be independent under $\widehat{P}$. Similarly, the paths of L and X over $[0,\tau]$ must be independent under $\widehat{P}$ because X is driven by a process Y that is a Brownian motion relative to $\widehat{P}$.

We exploit the path independence of L and W under $\widehat{P}$ when evaluating the conditional density $f_\tau(N_\tau, T; \theta \,|\, D, X)$ in case W is a frailty. Intuitively, the conditional $\widehat{P}$-expectation (5) is taken with respect to a Feller-jump diffusion process λ whose jump times T_n and jump sizes $\delta\ell(D_n)$ are given. These times and sizes are $\widehat{P}$-independent of the diffusion component. It is this $\widehat{P}$-independence that facilitates our computational approach. Note that under the original data-generating measure P, the T_n are not independent of the diffusion component, because the latter influences the arrival intensity of the T_n. For constants $0 \le a < b \le \tau$ and positive v and w, let

$$\phi_{a,b}(v,w) = \widehat{E}\left(\exp\left(-\int_a^b \lambda_s ds\right) \,\Big|\, \lambda_a = v, \lambda_b = w\right). \tag{7}$$

This conditional expectation can be calculated explicitly. By Proposition 3.2, if $T_n \le a < b < T_{n-1}$ for some n then the intensity λ follows the same Feller diffusion on $[a,b]$ under both P and the equivalent measure $\widehat{P}$. This property allows us to apply the Feller diffusion transform results in the appendix of Broadie & Kaya (2006) to conclude that

$$\phi_{a,b}(v,w) = \frac{I_q\left(\sqrt{vw}\frac{4\beta e^{-0.5\beta\alpha}}{\sigma^2(1-e^{-\beta\alpha})}\right)}{I_q\left(\sqrt{vw}\frac{4\kappa e^{-0.5\kappa\alpha}}{\sigma^2(1-e^{-\kappa\alpha})}\right)} \cdot \frac{\beta e^{-0.5(\beta-\kappa)\alpha}(1-e^{-\kappa\alpha})}{\kappa(1-e^{-\beta\alpha})} \cdot e^{\frac{v+w}{\sigma^2}\left[\frac{\kappa(1+e^{-\kappa\alpha})}{1-e^{-\kappa\alpha}} - \frac{\beta(1+e^{-\beta\alpha})}{1-e^{-\beta\alpha}}\right]} \tag{8}$$

where $\alpha = b - a$ and $\beta = \sqrt{\kappa^2 + 2\sigma^2}$, and where I_q is the modified Bessel function of the first kind of order $q = 2\kappa c/\sigma^2 - 1$. Next we express the density $f_\tau(N_\tau, T; \theta \,|\, D, X)$ in terms of $\phi_{a,b}(v,w)$, using the $\widehat{P}$-independence of the paths of W and L.

Proposition 3.3. *If the Brownian motion W is not adapted to the econometrician's filtration $\mathbb{G}$, then the conditional density of the event count and dates is*

$$f_\tau(N_\tau, T; \theta \,|\, D, X) = \widehat{E}\left(\prod_{n=1}^{N_\tau} \lambda_{T_n^-} \phi_{T_{n-1},T_n}(\lambda_{T_{n-1}}, \lambda_{T_n^-}) \phi_{T_{N_\tau},\tau}(\lambda_{T_{N_\tau}}, \lambda_\tau) \,\Big|\, N_\tau, T, D\right) \tag{9}$$

Note that $\lambda_{T_n^-} = \lambda_{T_n} - \delta\ell(D_n)$. The expectation on the right hand side of equation (9) is taken with respect to the conditional joint $\widehat{P}$-distribution of $(\lambda_{T_1}, \ldots, \lambda_{T_{N_\tau}}, \lambda_\tau)$ given (N_τ, T, D). Since the conditional $\widehat{P}$-distribution of $\lambda_{T_n^-}$ given $\lambda_{T_{n-1}}$ is non-central chi-squared, this expectation could be estimated by exact sampling of intensity values from the non-central chi-squared distribution. While easy to implement, convergence is slow. Appendix B develops an alternative recursive algorithm to calculate this expectation, again exploiting the fact that the intensity λ follows a Feller diffusion between events that is not affected by the measure transformation, and that is independent of the (T_n, D_n)

after the measure change. The idea is to construct a discrete-time, finite state Markov chain approximation to the Feller jump-diffusion λ over $[0,\tau]$, whose jump times and sizes are given.

An alternative approach to the computation of the likelihood function and the filtered intensity in Section 4 below is based on the classical Kushner-Stratonovich equations that describe the time evolution of conditional expectations $E(g(\lambda_t, N_t, L_t) \,|\, \mathcal{G}_t)$ and related filters; see Kliemann, Koch & Marchetti (1990) for general results. As shown in Giesecke, Kakavand & Mousavi (2008), this approach can be used to develop a recursive algorithm to calculate the filter $E(g(\lambda_t, N_t, L_t) \,|\, \mathcal{G}_t)$ for the model (1) at and between event times T_n. The number of required recursions is given by N_t, i.e. the number of event times $T_n \leq t$. With each additional recursion the numerical stability of the algorithm decreases, making this approach computationally intractable in the presence of many default observations.

4 Filtered intensity

This section discusses the calculation of the econometrician's filtered intensity h of the counting process N. The filtered intensity is required in the goodness-of-fit tests we develop in Section 5 below. The process h is given by the optional projection of the complete information intensity λ onto the observation filtration $\mathbb{G} \subseteq \mathbb{F}$, and is unique up to indistinguishability. It satisfies $h_t = E(\lambda_t \,|\, \mathcal{G}_t)$ almost surely, for each fixed time t. We will calculate h_t as this conditional expectation.

Proposition 4.1. *The $\mathbb{G}$-intensity h of N under P satisfies*

$$h_t = \frac{\widehat{E}(Z_t^{-1}\lambda_t \,|\, \mathcal{G}_t)}{\widehat{E}(Z_t^{-1} \,|\, \mathcal{G}_t)} \tag{10}$$

almost surely, for each $t \leq \tau$.

From formula (10) we see that in case W and hence λ are $\mathbb{G}$-adapted, i.e. observable by the econometrician, then so is Z and we have $h = \lambda$, as expected. It is instructive to compare the fitted filtered intensity with the fitted *smoothed* intensity $H_t = E(\lambda_t \,|\, \mathcal{G}_\tau)$, the posterior mean of λ_t conditional on all data. The smoothed intensity can be calculated with an argument similar to the one used in the proof of Proposition 4.1:

$$H_t = \frac{\widehat{E}(Z_\tau^{-1}\lambda_t \,|\, \mathcal{G}_\tau)}{\widehat{E}(Z_\tau^{-1} \,|\, \mathcal{G}_\tau)} = \frac{\widehat{E}(Z_\tau^{-1}\lambda_t \,|\, N_\tau, T, D, X)}{e^\tau f_\tau(N_\tau, T; \theta \,|\, D, X)} \tag{11}$$

almost surely, for each $t \leq \tau$. In formula (11), only the numerator on the right hand side depends on time t. The denominator is given by the scaled likelihood of the event count and times. The conditional $\widehat{P}$-expectations in formulae (10) and (11) are evaluated with a variant of the recursive algorithm developed in Appendix B.

5 Goodness-of-fit tests

This section develops goodness-of-fit tests to evaluate different specifications of the default timing model $(\lambda, \mathbb{G})$, which consists of a complete information, $\mathbb{F}$-adapted intensity process λ and an observation filtration $\mathbb{G} \subseteq \mathbb{F}$ that represents the information flow accessible to the econometrician. Das et al. (2007) have proposed a time-scaling test to assess a doubly stochastic model in a fixed, complete information filtration. We extend this test to general event timing models with arbitrary intensity dynamics and arbitrary observation filtration. This extension allows us to evaluate and directly compare self-exciting models with and without frailty, and to contrast the fitting performance of self-exciting models with that of the doubly stochastic model tested by Das et al. (2007).

The extension is based on a result of Meyer (1971), which implies that any counting process whose Doob-Meyer compensator is continuous and increases to infinity almost surely can be re-scaled into a standard Poisson process by a stochastic *change of time* that is given by the compensator. We apply this result to the counting process N of event dates T_n in the observation filtration $\mathbb{G}$. Relative to $\mathbb{G}$ and the actual, data-generating probability measure P, N has compensator $A = \int_0^{\cdot} h_s ds$, where h is the filtered intensity, calculated in Proposition 4.1 above. This means that the process $N - A$ is a local martingale relative to $(\mathbb{G}, P)$. Let A^{-1} denote the right-continuous inverse to A. Since each A_t^{-1} is a stopping time with respect to $\mathbb{G}$, we can define the stopping time σ-algebra $\mathcal{G}_t^{-1} = \mathcal{G}_{A_t^{-1}}$, which is the smallest σ-algebra containing all right-continuous left limit processes sampled at A_t^{-1}. Meyer's theorem implies that the time-scaled process $N_{A^{-1}}$ is a standard P-Poisson process in the time-scaled filtration $\mathbb{G}^{-1}$, which is the minimal right-continuous completion of $(\mathcal{G}_t^{-1})_{t \geq 0}$. The Poisson property of the time-scaled process can be tested. We summarize the test procedure.

Algorithm 5.1. *To test the goodness-of-fit of a default timing model* $(\lambda, \mathbb{G})$,

(i) Calculate the path $\{h_t(\omega) : 0 \leq t \leq \tau\}$ *of the filtered,* $\mathbb{G}$*-adapted intensity* h *of* N *from the fitted parameters of the complete information intensity* λ *and the data* $\mathcal{G}_\tau$ *based on Proposition 4.1.*

(ii) Transform the observed sequence $(T_n(\omega))_{n=1,\ldots,N_\tau(\omega)}$ *of event dates into a new sequence* $(A_{T_n}(\omega))_{n=1,\ldots,N_\tau(\omega)}$ *of dates given by*

$$A_{T_n}(\omega) = \int_0^{T_n(\omega)} h_s(\omega) ds, \quad n = 1, 2, \ldots, N_\tau(\omega). \tag{12}$$

(iii) Test the hypothesis that the counting process $N_{A^{-1}}(\omega)$ *associated with the transformed sequence* $(A_{T_n}(\omega))_{n=1,\ldots,N_\tau(\omega)}$ *is a realization of a standard* $(\mathbb{G}^{-1}, P)$*-Poisson process on the interval* $[0, A_\tau(\omega))$*.*

The test jointly addresses the specification of the econometrician's observation filtration $\mathbb{G}$ and the specification of the complete information intensity process λ. This is because the time change A that generates the sequence (A_{T_n}) is determined by the filtered intensity $h_t = E(\lambda_t \,|\, \mathcal{G}_t)$, which is governed by the stochastic process followed by λ and the structure of the econometrician's filtration $\mathbb{G}$. If both λ *and* $\mathbb{G}$ are correctly specified, then so is h and the transformed sequence (A_{T_n}) is standard Poisson relative to $\mathbb{G}^{-1}$.

We test the Poisson property relative to the filtration generated by the time-scaled process $N_{A^{-1}}$, which is a subfiltration of $\mathbb{G}^{-1}$. Standard tests include a test of the hypothesis that the number of arrivals in the transformed interval $[0, A_\tau)$ is Poisson distributed with unit intensity. Another test concerns the hypothesis that the inter-arrival times $W_n = A_{T_n} - A_{T_{n-1}}$ of the transformed sequence are independent of one another and standard exponentially distributed. We assess the properties of the inter-arrival times W_n using a Kolmogorov-Smirnov (KS) test and a test due to Prahl (1999), which is particularly sensitive to large deviations of the W_n from their theoretical mean 1. Letting μ be the sample mean of $(W_n)_{n=1,\dots,m}$, Prahl shows that if the W_n are i.i.d. standard exponential, then the random variable

$$M = \frac{1}{m} \sum_{n:W_n<\mu} \left(1 - \frac{W_n}{\mu}\right) \tag{13}$$

is asymptotically normally distributed with mean $\mu_M = e^{-1} - a/m$ and standard deviation $\sigma_M = b/\sqrt{m}$, where $a \approx 0.189$ and $b \approx 0.2427$. With $m = N_\tau = 909$ observations in our sample we obtain for these values

$$\mu_M = 0.3676 \quad \text{and} \quad \sigma_M = 0.0080.$$

We reject the null hypothesis of a correctly specified model $(\lambda, \mathbb{G})$ if the p-value of the KS test is less than 5% and Prahl's test statistic M lies outside a one-standard deviation band around its theoretical mean, i.e. if $M \notin (0.3596, 0.3756)$.

While Algorithm 5.1 describes an in-sample test, an analogous procedure can be used to test the model $(\lambda, \mathbb{G})$ out-of-sample. In this case, we use data observed only up to some horizon $\tau^* \leq \tau$ to estimate the model parameters. Based on the fitted parameters, the filtered intensity h is calculated and used to time-scale the arrival times T_n that fall into the out-of-sample interval $[\tau^*, \tau]$ using formula (12). The Poisson property is then tested on the corresponding transformed sequence.

6 Empirical analysis

This section employs the statistical tools developed above to determine the degree of contagion that has been present for U.S. corporate defaults during 1970–2006, and to measure the relative empirical importance of contagion and frailty for explaining the default clustering not captured by a doubly stochastic model. We begin in Section 6.1 by fitting a simple but well-performing zero-factor intensity model that has no diffusion term ($\sigma = 0$). In this complete information model, any default correlation is attributed to contagion. In Section 6.2 we estimate the unrestricted one-factor model ($\sigma > 0$) for different informational scenarios. With complete information, default correlation is due to contagion and exposure to an observable risk factor. With incomplete information, an additional frailty channel comes into play. Our goodness-of-fit test allows us to evaluate all these alternative specifications with the same metric. A comparison with the test statistics obtained for the benchmark doubly stochastic model by Das et al. (2007) indicates the empirical relevance of the self-exciting model features for explaining the substantial time-series variation of corporate default rates during the sample period.

6.1 Zero-factor model

We estimate the intensity model (1) for $\sigma = 0$, in which case N is a Hawkes (1971) point process. The path of λ is a deterministic function of the marked point process (T_n, D_n) representing our observations; from equation (1) we have by integration by parts

$$\lambda_t = c + (\lambda_0 - c)e^{-\kappa t} + \delta \int_0^t e^{-\kappa(t-s)} dL_s \tag{14}$$

where L is the response jump process (2). Thus, λ is adapted to the observation filtration $\mathbb{G}$ and the filtered intensity $h = \lambda$. By Proposition 3.1, the density $f_\tau(N_\tau, T; \theta \,|\, D, X) = Z_\tau^{-1} e^{-\tau}$. The covariate observations X are irrelevant. We perform a grid search over the discretized parameter space to solve the likelihood problem

$$\sup_\theta \log f_\tau(N_\tau, T; \theta \,|\, D, X) = \sup_\theta \left(\int_0^\tau \log(\lambda_{s-}) dN_s - \int_0^\tau \lambda_s ds \right) \tag{15}$$

where $\lambda = \lambda(\theta)$. Under technical conditions stated in Ogata (1978), the corresponding maximum likelihood estimator of θ is asymptotically normal and efficient.

We consider several specifications of the weight function ℓ, which controls the response of the intensity to the number of defaults at an event date. For the basic model $\ell(n) = 1$, the intensity response is independent of the event count. This choice is tantamount to ignoring the event count observations. Not surprisingly, this model is rejected according to Prahl's and KS tests (p-value 0.004%). The specification $\ell(n) = n$ performs better (p-value 0.14%), but is still rejected. Next we introduce an additional parameter $w \geq 0$ and consider the quadratic model $\ell(n) = n + wn^2$, which assigns a higher weight to the event count at an event date. In this case, the cluster of railway defaults in 1970 strongly affects the likelihood. At T_4, which represents June 21, 1970, the associated event count $D_4 = 24$ takes its maximal value over the whole sample. With this extreme observation, the likelihood function is decreasing in w, all else fixed. We therefore modify our estimation approach as follows: we fix a set of candidate values for w and then estimate the intensity for each of these candidate values. The model is selected according to the quality of the fit rather than the log-likelihood score. The results are summarized in Table 1. Throughout, we assume that the initial intensity $\lambda_0 = c$, after finding that this parameter has vanishing

w	0	0.20	0.25	0.30	0.35	0.40	0.45	0.50	0.60
κ	2.17	1.94	1.91	1.89	1.87	1.85	1.84	1.83	1.81
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
$c = \lambda_0$	2.90	4.35	4.63	4.87	5.10	5.30	5.48	5.65	5.93
	(0.64)	(0.74)	(0.75)	(0.76)	(0.77)	(0.78)	(0.79)	(0.79)	(0.80)
δ	1.28	0.70	0.62	0.56	0.51	0.47	0.43	0.40	0.35
	(0.06)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)
p-value	0.0014	0.0170	0.0244	0.0310	0.0419	0.0481	0.0546	0.0442	0.0439
M	0.3345	0.3540	0.3571	0.3597	0.3621	0.3642	0.3660	0.3676	0.3703
Δ_M	−4.1	−1.7	−1.3	−1.0	−0.7	−0.4	−0.2	0	0.3

Table 1: Estimation results for the zero-factor model with quadratic weight function $\ell(n) = n + wn^2$ for each of several values of w. Estimates of asymptotic standard errors are given parenthetically. The p-value refers to the KS test and Δ_M denotes the distance of Prahl's test statistic M defined in formula (13) from its theoretical mean $\mu_M = 0.3676$, measured in terms of its theoretical standard deviation $\sigma_M = 0.0080$.

influence on the overall likelihood and model fit. The model fit improves with increasing values of w according to Prahl's and KS statistics, with the best fit attained at $w = 0.45$. For values of w larger than 0.45 the fit deteriorates. The optimal parameter estimates are

$$\hat{\kappa} = 1.84\,(0.04), \quad \hat{c} = \hat{\lambda}_0 = 5.48\,(0.79), \quad \hat{\delta} = 0.43\,(0.02), \quad \hat{w} = 0.45; \tag{16}$$

estimates of asymptotic standard errors are given parenthetically. The fitted values of δ and w indicate the significant influence of an event on arrival rates. The zero-factor model (14) attributes this influence to contagion. We propose a more comprehensive interpretation in Section 6.3 below, after examining other model variants in different informational scenarios.

Figure 2 graphs the fitted intensity λ, obtained from formula (14) evaluated at the observations (T_n, D_n) and the optimal parameters (16). Thanks to the feedback term in λ, the fitted intensity responds quickly to bursts of defaults such as those observed during the internet bubble. This responsiveness benefits the out-of-sample predictive performance of the model, considered in Section 7.2 below. An alternative doubly stochastic model of correlated event timing tends to be less responsive, since defaults are not permitted to directly influence the intensity. Therefore, the fitted doubly stochastic intensity tends to lag event clusters. The sensitivity of a self-exciting model to events can also be a liability, however. Consider the large fitting errors at the beginning of the sample period, caused by the extreme cluster of 24 railway defaults in June 1970. Nevertheless, the goodness-of-fit test, which is based on the fitted intensity, does not reject our specification. This indicates that the simple 4 parameter intensity model replicates the time-variation of U.S. default rates between 1970 and 2006. Figure 3 shows a histogram of the time-scaled

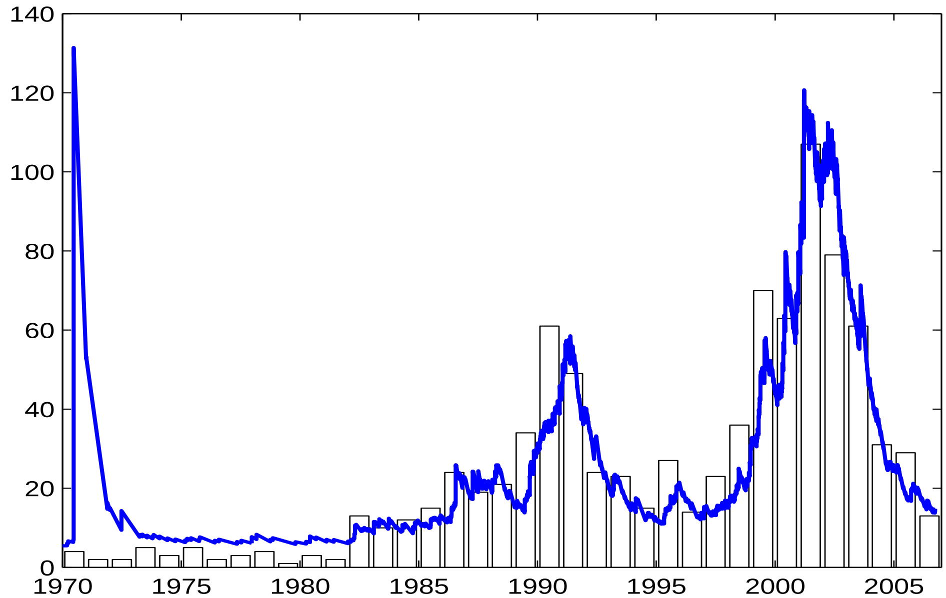

Figure 2: Annual number of days with default events (bars) and fitted intensity λ_t (solid line) for the zero-factor model with quadratic weight function $\ell(n) = n + 0.45n^2$. The peak in 1970 is due to the extreme cluster of 24 railway defaults on June 21, 1970.

inter-arrival times along with their empirical distribution function. Figure 4 plots the empirical quantiles of the time-scaled inter-arrival times against the theoretical standard exponential quantiles. Almost all of the observed 909 inter-arrival times are time-scaled correctly by the fitted model. The three outliers are due to the railway default cluster.

6.2 One-factor model

We turn to the estimation of the unrestricted model (1) to examine the role of the common Brownian motion risk factor W, which generates diffusive movement in the complete information intensity λ between events. The path of λ over $[0, \tau]$ is a deterministic function of the observed event dates and counts $(T_n, D_n)_{n=1,\ldots,N_\tau}$, and the path $\{W_t : 0 \leq t \leq \tau\}$, which may be unobservable. Motivated by the fitting success of the zero-factor model with $\lambda_0 = c$ and quadratic weight function, we rely on these specifications also in this section.

6.2.1 Non-informative covariates

Suppose the covariate X is independent of the Brownian motion W, and therefore non-informative in the presence of the default time and count observations. The relevant covariates might be truly unobservable, or simply be ignored in the set of observable covariates X. In this case, the path of W over $[0, \tau]$ cannot be identified from the observa-

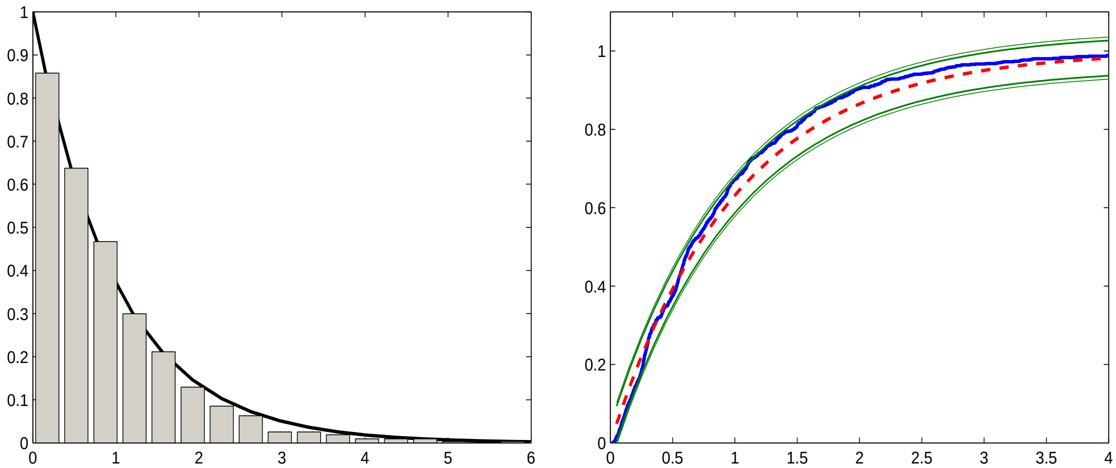

Figure 3: Distribution of time-scaled inter-arrival times for the zero-factor model with quadratic weight function. *Left panel*: Standard exponential density function (solid line) vs. histogram of time-scaled inter-arrival times. *Right panel*: Standard exponential distribution function (dotted line) vs. empirical distribution function of time-scaled inter-arrival times along with 1% and 5% bands.

tions so W is not adapted to the observation filtration $\mathbb{G}$. It is a frailty as in Duffie et al. (2008), Koopman et al. (2008) and others. Default correlation is caused by contagion, exposure of firms to W, and uncertainty about the value of W. Thus, relative to the zero-factor model discussed above, the one-factor model with frailty permits two additional channels of default correlation.

We address the likelihood problem $\sup_\theta \log f_\tau(N_\tau, T; \theta \,|\, D, X)$ using the fitting procedure outlined in Section 6.1. With frailty, the density $f_\tau(N_\tau, T; \theta \,|\, D, X)$ does not take a closed form. We use the algorithm developed in Appendix B to compute the density based on Proposition 3.3. The optimal parameter values are

$$\hat{\kappa} = 1.00\,(0.21), \quad \hat{c} = \hat{\lambda}_0 = 6.20\,(0.97), \quad \hat{\sigma} = 3.50\,(0.51), \quad \hat{\delta} = 0.20\,(0.02), \quad \hat{w} = 0.50;$$

estimates of asymptotic standard errors are given parenthetically. The fitted values of δ and w, which control the sensitivity of the complete information intensity λ to defaults, indicate the significant influence of an event on arrival rates. These values account for firms' exposure to the frailty W, and provide evidence that contagion is statistically and economically highly significant, even when controlling for the presence of frailty. The fitted volatility of the frailty W is substantial.

The filtered intensity $h_t = E(\lambda_t \,|\, \mathcal{G}_t)$ jumps at arrivals due to the direct impact of events, and the Bayesian updating of the $\mathbb{G}$-conditional distribution of the frailty W. Between events, h is a deterministic function of time. Thus, the sample path behavior of the filtered intensity h is similar to that of the zero-factor, complete information intensity examined in Section 6.1 above. The fitted path of h is plotted in Figure 5. Compare with

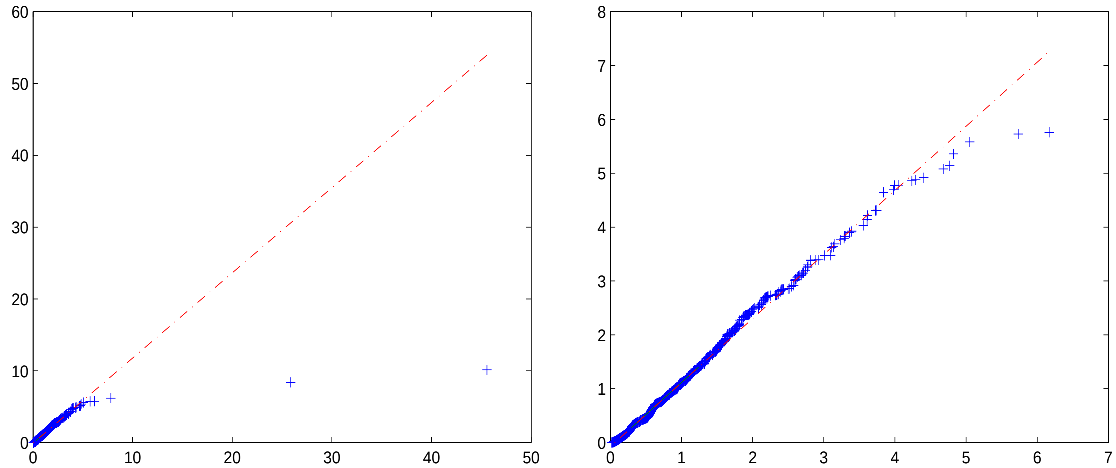

Figure 4: Empirical quantiles of time-scaled inter-arrival times for the zero-factor model with quadratic weight function vs. theoretical standard exponential quantiles. *Left panel*: All 909 observations are shown. The three outliers are due to the cluster of railway defaults on June 21, 1970 (24 events) and June 26, 1972 (4 events). *Right panel*: Only 906 observations are shown. One of the two "outliers" is caused by the 1972 cluster.

the estimated smoothed intensity H, plotted in Figure 6, which is given by the conditional posterior mean $H_t = E(\lambda_t \,|\, \mathcal{G}_\tau)$ conditional on $\mathcal{G}_\tau$, i.e. all the data observed through the final sample time τ. The filtered intensity h_t is based on the smaller information set $\mathcal{G}_t \subseteq \mathcal{G}_\tau$, i.e. all the data observed through time $t \leq \tau$. As expected, H matches the time-series variation of actual default rates better than h.

The goodness-of-fit test requires the fitted filtered intensity h. Figure 7 graphs the empirical distribution of the time-scaled inter-arrival times. Prahl's test statistic $M = 0.3357$ is 4 theoretical standard deviations $\sigma_M = 0.0080$ away from its theoretical mean $\mu_M = 0.3676$. The p-value of the KS test is 0.0012. The test rejects the hypothesis that the re-scaled inter-arrival times are standard exponential.

6.2.2 Informative covariates

Next we estimate the one-factor model in the presence of a covariate X that allows us to identify the path of the Brownian motion risk factor W over the sample period. Then, W is adapted to the observation filtration $\mathbb{G}$ and so is the complete information intensity λ. Default correlation is caused by contagion and exposure of firms to the observable factor W. The frailty channel of default correlation is inactive.

Using a doubly stochastic model of firm intensities, Duffie et al. (2006) found that, besides firm-specific variables, the 3 month Treasury bill yield and the trailing 1 year return on the S&P 500 index have significant explanatory power for defaults of U.S. industrial firms between 1980–2004. Motivated by these findings, we explore as macro-

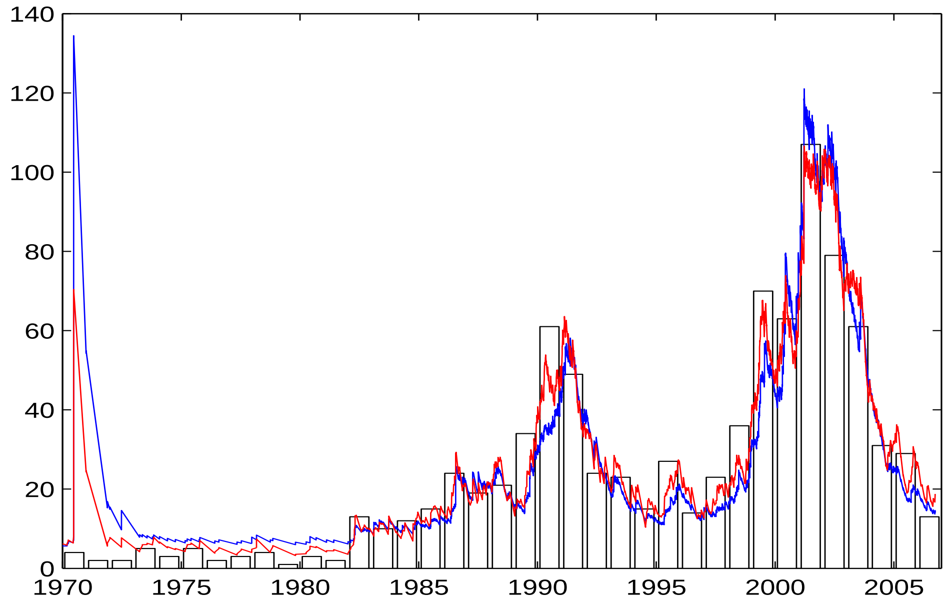

Figure 5: Annual number of days with default events (bars) and fitted filtered intensity $h_t = E(\lambda_t \mid \mathcal{G}_t)$ (solid red), estimated for the one-factor frailty model with quadratic weight function $\ell(n) = n + 0.5n^2$. The fitted complete information intensity λ_t of the zero-factor model (solid blue) is shown for comparison.

economic covariates the 1 year Treasury bill yield and the S&P 500 index value.

Suppose the covariate process X is the S&P 500 index value. Suppose further that X follows a geometric Brownian motion,

$$X_t = X_0 e^{(\mu_X - \sigma_X^2/2)t + \sigma_X W_t} \tag{17}$$

where μ_X is a constant drift parameter, σ_X is a constant volatility parameter and W is the standard Brownian motion risk factor that also influences λ. Based on daily index return observations, the maximum likelihood estimator $\hat{\nu}$ of the covariate parameter vector $\nu = (\mu_X, \sigma_X)$ is given by $\hat{\nu} = (0.089, 0.175)$. Using the relationship (17) with ν substituted by the estimator $\hat{\nu}$, we identify the daily values of W from the observed values of X. We linearly interpolate between these values to approximate the path of W over the sample period $[0, \tau]$.

Treating the estimated path of W as though error-free, we then estimate the intensity λ. The intensity path over $[0, \tau]$ is a deterministic function of the observed path of X and the pairs $(T_n, D_n)_{n=1,\ldots,N_\tau}$. The density $f_\tau(N_\tau, T; \theta \mid D, X) = Z_\tau^{-1} e^{-\tau}$, and the corresponding maximum likelihood problem (15) is solved by performing a grid search over the discretized parameter space. The estimation results are summarized in Table 2. To allow for a negative relationship between the covariate and the intensity while keeping

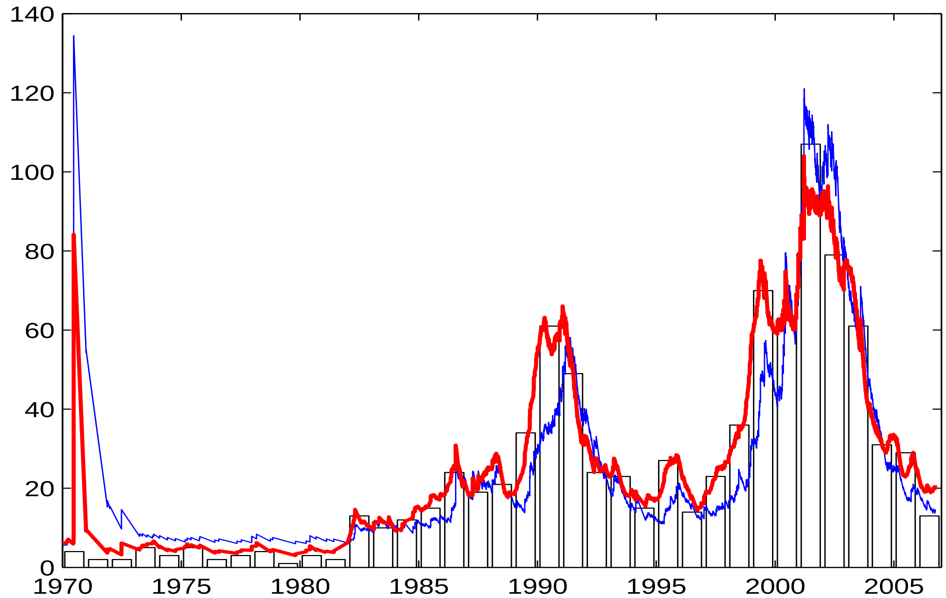

Figure 6: Annual number of days with default events (bars) and fitted smoothed intensity $H_t = E(\lambda_t \,|\, \mathcal{G}_\tau)$ (thick solid red), estimated for the one-factor frailty model with quadratic weight function $\ell(n) = n + 0.5n^2$. The fitted complete information intensity λ_t of the zero-factor model (solid blue) is shown for comparison.

the diffusive volatility $\sigma \geq 0$, we also perform the estimation for the case where the path of W is replaced by the path of the Brownian motion $-W$.

Next suppose the covariate process X is the 1 year Treasury bill yield. Suppose further that the short-term rate of interest r follows a Feller diffusion,

$$dr_t = \kappa_r(c_r - r_t)dt + \sigma_r\sqrt{r_t}dW_t \tag{18}$$

where κ_r is a constant mean-reversion parameter, c_r is a constant mean reversion level and σ_r is a constant volatility parameter such that $2\kappa_r c_r \geq \sigma_r^2$. We assume that under risk neutral probabilities, the short rate follows a Feller diffusion

$$dr_t = \kappa_r^*(c_r^* - r_t)dt + \sigma_r\sqrt{r_t}dW_t^* \tag{19}$$

where $\kappa_r^* = \kappa_r + \eta\sigma_r$ and $c_r^* = c_r\kappa_r/\kappa_r^*$ for some constant parameter η that represents the market price of risk for W and that is chosen such that $\eta\sqrt{r}$ satisfies Novikov's condition for the horizon τ, and where $W_t^* = \eta\int_0^t \sqrt{r_s}ds + W_t$ is a standard Brownian motion under risk neutral probabilities. In this setting, the price at time 0 of a zero coupon bond with unit face value and maturity date t is given by $\exp(\alpha(t) + \beta(t)r_0)$, where the coefficient functions $\alpha(t)$ and $\beta(t)$ depend on the covariate parameter vector $\nu = (\kappa_r, c_r, \sigma_r, \eta)$ and are given in closed form in Cox, Ingersoll & Ross (1985). Based on

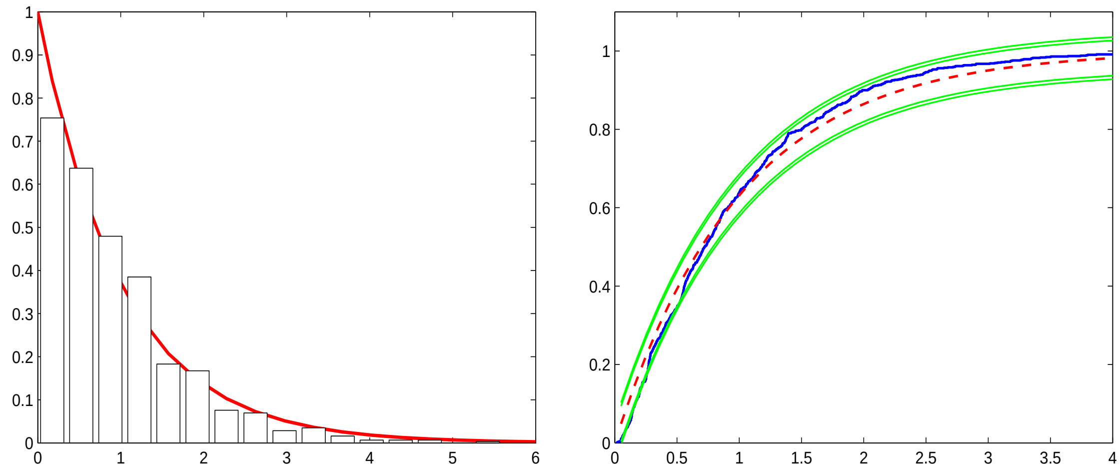

Figure 7: Distribution of time-scaled inter-arrival times for the one-factor frailty model with quadratic weight function. *Left panel*: Standard exponential density function (solid line) vs. histogram of time-scaled inter-arrival times. *Right panel*: Standard exponential distribution function (dotted line) vs. empirical distribution function of time-scaled inter-arrival times along with 1% and 5% bands.

weekly yield observations, a standard maximum likelihood procedure yields the estimate $\hat{\nu} = (0.209, 0.052, 0.059, -1.600)$. We use the bond pricing formula with ν substituted by $\hat{\nu}$ to identify the daily values of r from daily yield observations. The values of W are then inferred from the values of r. We linearly interpolate between the values to approximate the path of W over the sample period $[0, \tau]$.

We then fit λ, again treating the estimated path of W as though error-free. Table 3 summarizes the results. Figure 8 shows the fitted intensity λ for each of the covariates, along with the fitted intensity for the zero-factor model. While difficult to discern in the graph, the sample path behavior of the intensity in the complete information one-factor model differs from the behavior of the intensity in the zero-factor model and the one-factor frailty model. In the complete information one-factor model, the intensity varies randomly between events according to the movements of the covariate. In the other models the intensity evolves deterministically between events: in the zero-factor model random fluctuations are precluded by construction because $\sigma = 0$, and in the one-factor frailty model the random fluctuations of W are "averaged out" by the filter.

The results of the fitness tests stated in Tables 2 and 3 indicate that the complete information one-factor model fits about as well as the zero-factor model, and better than the one-factor frailty model. In terms of fit, the S&P 500 index value covariate performs slightly better than the 1 year Treasury bill yield.

w	0.40	0.40	0.45	0.45	0.50	0.50
Sign W	$+$	$-$	$+$	$-$	$+$	$-$
κ	1.81 (0.07)	1.85 (0.07)	1.80 (0.08)	1.84 (0.08)	1.78 (0.08)	1.83 (0.08)
$c = \lambda_0$	5.25 (0.62)	5.30 (0.63)	5.44 (0.62)	5.48 (0.63)	5.60 (0.63)	5.65 (0.63)
δ	0.47 (0.02)	0.47 (0.02)	0.43 (0.02)	0.43 (0.02)	0.40 (0.02)	0.40 (0.02)
σ	0.60 (0.33)	0.00 (0.35)	0.65 (0.33)	0.00 (0.35)	0.70 (0.33)	0.00 (0.35)
p-value	0.0541	0.0481	0.0507	0.0546	0.0517	0.0442
M	0.3632	0.3642	0.3650	0.3660	0.3665	0.3676
Δ_M	-0.55	-0.43	-0.33	-0.20	-0.14	0.00

Table 2: Estimation results for the complete information one-factor model with quadratic weight function $\ell(n) = n + wn^2$ for various values of w, based on observations of event dates, counts and the S&P 500 index. Estimates of asymptotic standard errors are given parenthetically. The p-value refers to the KS test and Δ_M denotes the distance of Prahl's test statistic M defined in formula (13) from its theoretical mean $\mu_M = 0.3676$ measured in terms of its theoretical standard deviation $\sigma_M = 0.0080$.

6.3 Discussion

Our experiments indicate that the simple zero-factor intensity model is hard to beat in terms of fit to U.S. default rates. The introduction of explanatory covariates does improve the fit relative to the one-factor frailty model, but not relative to the zero-factor model. This finding suggests that the explanatory information content of the default data is substantial. The usual covariates, including the S&P 500 and the Treasury bill yield, do not provide incremental explanatory power for the time-series variation of corporate default rates. This is in strong contrast to a doubly stochastic model of default timing, where the econometrician seeks to explain the time-series variation of default rates solely in terms of movements of exogenous covariates whose evolution must not be influenced by default events. Here, the default observations have no explanatory role at all.

The parameter estimates of all self-exciting models indicate the sensitivity of the intensity to events. The estimates of the frailty model differ from those of the complete information models, which are very similar. In the frailty model, the fitted contagion sensitivity parameter δ is about one half of the value that is obtained for the complete information models. Thus, according to our model, frailty and contagion represent roughly equally important sources for the feedback from events to U.S. default rates, in the sense that on average, they are responsible for an equal share of the increase of fitted default rates at an event. Both contagion and frailty are required to explain the non-doubly stochastic clustering in U.S. defaults. Finally, we observe that the fitted decay rate κ of the intensity is much lower for the frailty model, indicating that the contagious impact of an event is much more persistent in the presence of frailty.

If we are only interested in fitting the time-series variation of corporate default rates,

w	0.40	0.40	0.45	0.45	0.50	0.50
Sign W	$+$	$-$	$+$	$-$	$+$	$-$
κ	1.72 (0.07)	1.85 (0.08)	1.71 (0.07)	1.84 (0.08)	1.72 (0.07)	1.83 (0.08)
$c = \lambda_0$	4.76 (0.62)	5.30 (0.63)	4.96 (0.61)	5.48 (0.63)	5.16 (0.63)	5.65 (0.63)
δ	0.46 (0.02)	0.47 (0.02)	0.42 (0.02)	0.43 (0.02)	0.40 (0.02)	0.40 (0.02)
σ	0.30 (0.21)	0.00 (0.23)	0.30 (0.21)	0.00 (0.22)	0.27 (0.21)	0.00 (0.22)
p-value	0.0469	0.0481	0.0255	0.0546	0.0227	0.0442
M	0.3643	0.3642	0.3660	0.3660	0.3675	0.3676
Δ_M	-0.41	-0.43	-0.20	-0.20	-0.01	0.00

Table 3: Estimation results for the complete information one-factor model with quadratic weight function $\ell(n) = n + wn^2$ for various values of w, based on observations of event dates, counts and 1 year Treasury bill yields. Estimates of asymptotic standard errors are given parenthetically. The p-value refers to the KS test and Δ_M denotes the distance of Prahl's test statistic M defined in formula (13) from its theoretical mean $\mu_M = 0.3676$ measured in terms of its theoretical standard deviation $\sigma_M = 0.0080$.

then the introduction of frailty into the self-exciting model (1) does not appear to be statistically beneficial. Relative to the complete information models, the introduction of frailty does not improve the fit to the data. The reason is that frailty does not introduce additional statistical features into the complete information self-exciting model (1). The empirically important model property is the sensitivity of the intensity to events, and this sensitivity is modeled by the response term in the complete information intensity λ. The filtering of λ preserves the sensitivity, and ascribes part of it to the Bayesian updating of the frailty distribution. The remaining part is attributed to contagion. Our experiments indicate that the filter imposes additional constraints on the model that reduce the quality of the fit relative to the complete information models. The structure of these constraints strongly depends on the parametric model of the frailty.[6]

The complete information intensity in the zero-factor model and the filtered intensity in the frailty model have very similar sample path properties. In fact, we may view the zero-factor intensity as an approximation to the filtered intensity in a more comprehensive frailty model in which random fluctuations of the complete information intensity between events are averaged out by the filter, and the filtered intensity decays exponentially after an event. The zero-factor model is an approximation only to the extent that the filtered intensity does not decay exponentially. In light of this observation, the superior empirical performance of the zero-factor model is perhaps not surprising.

More generally, we may re-interpret the self-exciting $\mathbb{F}$-intensity (1) as the filtered

[6] Testing different frailty specifications is not straightforward, however. The maximum likelihood procedure requires an efficient method for calculating the intensity filter. Such methods seem only available in relatively special cases, including our Feller diffusion formulation.

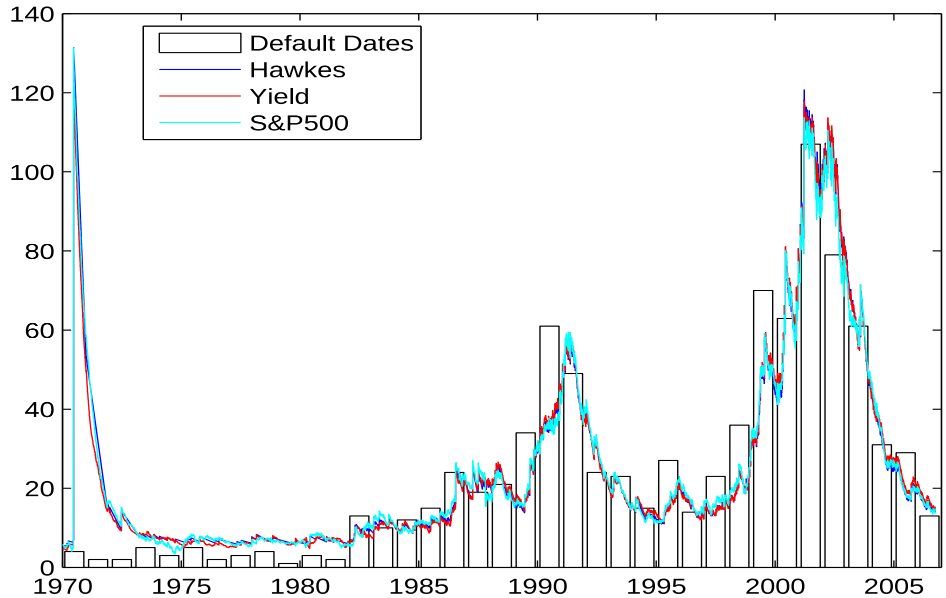

Figure 8: Annual number of days with default events (bars) and fitted intensity λ_t for the complete information one-factor model with quadratic weight function $\ell(n) = n + 0.5n^2$, for each of several covariates. The fitted intensity of the zero-factor model (labeled "Hawkes") is shown for comparison.

intensity of some frailty model formulated in a super-filtration $\mathbb{H} \supseteq \mathbb{F}$ that represents complete information. The filtration $\mathbb{F}$ then takes the role of the observation filtration, and the feedback jumps of the intensity at events can be interpreted in terms of contagion, or Bayesian updating of the frailty distributions. While we lose the ability to empirically distinguish between frailty and contagion effects, the estimation is computationally tractable, and we avoid potential fitting constraints imposed by the filter.

7 Portfolio credit risk

We examine the empirical implications of event feedback for portfolio credit risk. In light of our discussion in Section 6.3, we consider the zero-factor model and interpret the feedback in this model in terms of contagion and Bayesian updating of the distribution of a frailty factor that is not explicitly modeled. We begin by testing the out-of-sample fit of the model and then analyze the forecast conditional distribution of future events during the sample period.

End	01/98	01/99	01/00	01/01	01/02	01/03	01/04	01/05	01/06
Dates	420	456	526	589	696	775	836	867	896
Events	592	644	747	867	1056	1202	1287	1327	1360
κ	1.40	1.59	2.21	1.94	1.89	2.20	2.04	2.01	1.88
	(0.04)	(0.04)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.04)
$c = \lambda_0$	3.03	3.08	3.13	3.10	3.74	4.82	4.74	5.20	5.26
	(0.78)	(0.77)	(0.75)	(0.76)	(0.72)	(0.70)	(0.70)	(0.73)	(0.73)
δ	0.64	0.75	1.07	0.94	0.73	0.65	0.61	0.51	0.48
	(0.03)	(0.03)	(0.03)	(0.04)	(0.04)	(0.06)	(0.07)	(0.06)	(0.05)
w	0.10	0.10	0.10	0.10	0.20	0.30	0.30	0.40	0.40
p-value	0.0971	0.1433	0.1887	0.1581	0.0736	0.0600	0.0586	0.0381	0.0492
M	0.3717	0.3675	0.3602	0.3626	0.3610	0.3659	0.3634	0.3653	0.3651
Δ_M	0.3	0.0	-0.7	-0.5	-0.7	-0.2	-0.5	-0.3	-0.3

Table 4: Estimation results for the zero-factor model with quadratic weight function for observation periods starting at 01/1970 and ending at various dates before 09/2006. Estimates of asymptotic standard errors are given parenthetically. The p-value refers to the in-sample KS test and Δ_M denotes the distance of Prahl's in-sample test statistic M defined in formula (13) from its theoretical mean $\mu_M = 0.3676$ measured in terms of its theoretical standard deviation $\sigma_M = 0.0080$.

7.1 Testing out-of-sample fit

We estimate the zero-factor model from a subset of the observations, following the methodology described in the previous section. We choose observation windows of increasing size, all beginning in January 1970 but ending at different horizons in increments of one year. Table 4 shows the estimation and in sample test results. The test statistics indicate good fits, with p-values exceeding the 5% level, with the exception of the model estimated for the period ending 01/2005, whose p-value is slightly lower at 3.81%.

Next we test the out-of-sample fit of the estimated models. We consider test windows of several sizes. First we test a model on all available out-of-sample observations. For example, the model estimated from observations over the period 01/1970 to 01/2003 is tested on the arrivals observed over the subsequent period 01/2003 to 09/2006. Table 5 summarizes the test results. The models estimated with data observed until 01/2003 and later fit very well, while the models estimated from fewer observations are rejected. The relative size of the out-of-sample test windows used for the models estimated with data until 2003 may be one reason for the rejection. For example, the model estimated with data observed up to 01/1998 is tested on all observations after that date, which make up almost half of the entire sample.

To investigate the out-of-sample fit in more detail, we test the models on one year forecast windows. For example, the model estimated from observations over the period

End	01/98	01/99	01/00	01/01	01/02	01/03	01/04	01/05	01/06
Dates	489	453	383	320	213	134	73	42	13
p-value	0.0013	0.0001	0.0000	0.0000	0.0013	0.2284	0.6513	0.4980	0.8394

Table 5: Out-of-sample goodness-of-fit of the zero-factor models with quadratic weight function estimated for observation periods starting at 01/1970 and ending at various dates before 09/2006. The "Dates" represent the number of event dates T_n ahead of a given end date. The p-value refers to the KS test applied to all observations in the period ahead.

01/1970 to 01/2003 is tested on the arrivals observed over the subsequent one year period 01/2003 to 01/2004. The results, reported in Table 6, indicate the improved out-of-sample fit of the models estimated with data up until 01/2000. The model for 01/2000 fits the peak default events observed during the subsequent year very well, with a p-value of 0.1946. The out-of-sample fit of the two models estimated with data up until 01/2001 and 01/2002 has improved, but both models are still rejected. Additional experiments show that the 01/2001 model achieves a p-value of 0.0284 for a 6 month out-of-sample forecast window (50 event dates), while the 01/2002 model achieves a p-value of 0.0839 for a 3 month window (23 event dates).

7.2 Out-of-sample prediction

We use the fitted zero-factor model to predict event arrivals one year ahead. For example, the model estimated from observations over the period 01/1970 to 01/2003 is used to predict the total number of events in the subsequent period 01/2003 to 01/2004. To this end, we estimate by Monte Carlo simulation of the fitted counting process N the $\mathcal{F}_t$-conditional distribution of future event arrivals during $(t, t+1 \text{ year}]$, given the sequence of $(T_n, D_n)_{n=1,\ldots,N_t}$ observed up to time t. Using the exact simulation algorithm developed by Giesecke & Kim (2007), we generate successive arrival times of N over $(t, t+1 \text{ year}]$. At each arrival time, we draw the number of events at that date from the empirical distribution of the event counts D_n observed up to t.[7]

Figure 9 shows the forecast conditional distribution of future events for the years 1999–2004. The conditional distribution changes its shape in response to the increasing number of arrivals observed in the years up to 2001, and the decreasing trend from 2002 onwards. In particular, the mass of the distribution is shifted from the center into the right tail. This shift indicates a strong increase in the likelihood of a large number of defaults in the year ahead. The distribution with the fattest tail is predicted for 2002, the year following the peak default year 2001.

[7] As an alternative to simulation, we can exploit the semi-analytical transform formulae for N developed by Errais et al. (2006).

End	01/98	01/99	01/00	01/01	01/02	01/03	01/04	01/05	01/06
Dates	36	70	63	107	79	61	31	29	13
p-value	0.0734	0.0222	0.1946	0.0001	0.0020	0.1614	0.2507	0.2743	0.8394

Table 6: Out-of-sample goodness-of-fit of the zero-factor models with quadratic weight function estimated for observation periods starting at 01/1970 and ending at various dates before 09/2006. The "Dates" represent the number of event dates T_n in the year ahead of a given end date. The p-value refers to the KS test applied to the observations in the year ahead.

Our forecasting results are shown in Figure 10, which graphs the mean and the 1% and 99% as well as the 5% and 95% quantiles of the forecast conditional distribution of future events along with the realized number of events from 1998 to 2005. All realizations fall well within the 1% and 99% quantiles of the forecast distribution. For 2000–2005, the realizations fall within the 5% and 95% quantiles, and from 2002–2005 the realized number of events is close to the forecast mean number of events. From 1998–2001, the realized number of events exceeds the forecast mean. After the peak default year 2001 this relationship is reversed and the forecast mean dominates the realized values from above. These results indicate the model's ability to capture the burst of events during 2001–2003. Model forecasts are conservative in that they tend to slightly overestimate the number of arrivals during event bursts.

Next we examine how event feedback influences the conditional event distribution for several future horizons. We compare the forecast distributions by means of their quantiles. A quantile, or *value at risk* is widely used by financial institutions to measure the risk of a position or a portfolio of positions. Figure 11 graphs the 1, 2, 3, 4 and 5 year value at risk at the 95% level for each year between 1998 and 2004. The mass in the tail of the forecast event distribution as measured by the value at risk increased from 1998 to 2002 and decreased afterwards. The change from year to year differs across forecast horizons, and indicates the influence of past events on the term structure of the forecast conditional event distribution.

7.3 Credit portfolio loss distribution

Our self-exciting model of correlated event timing has applications in the risk management of corporate debt portfolios. Suppose that events arrive according to the zero-factor model with fitted parameters given in (16) and current (October 2006) intensity estimated at 14.81. Suppose further that the normalized loss at an event is drawn from a uniform distribution over the values $\{0.4, 0.6, 0.8, 1\}$. Figure 12 shows the conditional distribution of the cumulative loss due to default at all future horizons between one and five years, conditional on all observations in our data set as of October 2006. This "loss surface" represents the conditional distribution of future default loss in a hypothetical portfolio of

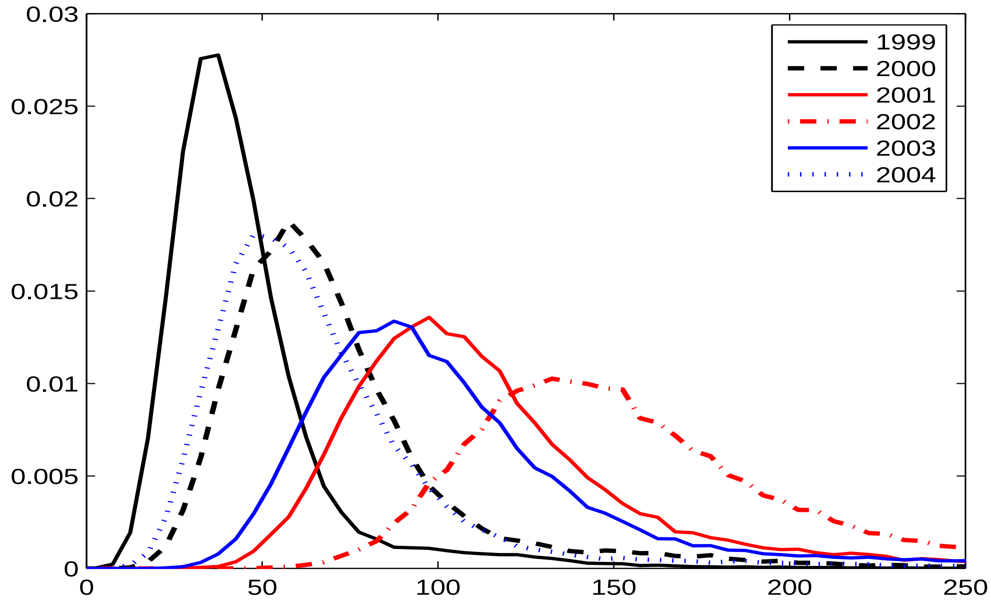

Figure 9: Forecast conditional distribution of future events over one year as implied by the zero-factor model with quadratic weight function. We use observations from 01/1970 to January 1st of each year to estimate the model. Monte Carlo simulation of the counting process N is used to estimate the conditional distribution of events during the year, given the observations. 50,000 paths were simulated.

6048 firms tracked by Moody's as of October 2006, given all 1374 events in the universe of Moody's rated firms that were observed between January 1970 and October 2006. Our model and estimation methodology can be used to generate the loss distribution of any portfolio of firms for which historical event information is available, such as loan or bond portfolios held at banks and other corporate debt investors.

8 Conclusion

We propose statistical tools for analyzing the effects on U.S. corporate defaults of contagion and frailty (unobservable risk factors that are correlated across firms). The inference problem is cast as a stochastic filtering problem. Measure change arguments are used to develop likelihood estimators for point processes describing unpredictable event arrivals. Time change arguments lead to goodness-of-fit tests for alternative specifications of the point process intensity and the observation filtration. Both measure and time changes transform the underlying point process into a simpler Poisson process, and guarantee the computational tractability of likelihood inference and fitness testing.

Applying these tools to U.S. industrial and financial firms during 1970–2006, we find

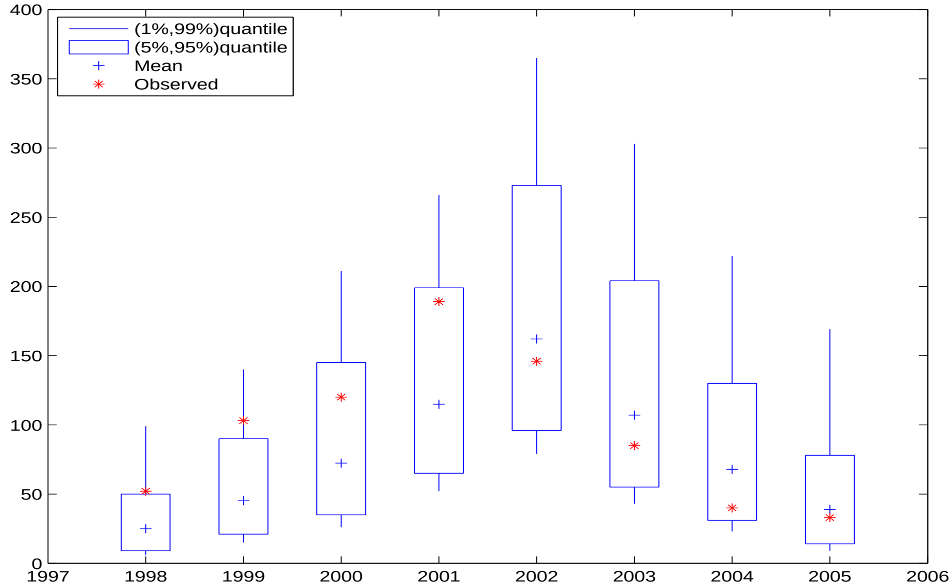

Figure 10: Forecast conditional distribution of event arrivals over one year ahead as implied by the zero-factor model with quadratic weight function vs. actually observed number of events. We use observations from 01/1970 to January 1st of each year to estimate the model. Monte Carlo simulation of the counting process N is used to estimate the conditional distribution of events during the year, given the observations. The graph shows the mean and the upper and lower 1 and 5% quantiles of the forecast distribution along with the realized number of events. 50,000 paths were simulated.

strong evidence for the presence of contagion, even when controlling for firms' exposure to a common unobservable Feller diffusion frailty risk factor. Contagion and frailty phenomena are found to be roughly equally important for explaining the default clustering in the data that is not captured by a traditional doubly stochastic model, in which firms' default times are correlated only as implied by the correlation of observable risk factors determining their default intensities.

These findings have significant implications for the design of theoretical and empirical models of correlated default timing, used for example by banks and other managers of credit portfolios to measure portfolio credit risk and estimate the level of risk capital needed to withstand default losses at high confidence levels.

Our methodology for estimating and testing stochastic point processes for correlated event arrivals could be applied to other situations in which the timing of economic events is influenced by observable and unobservable stochastic risk factors, and in which an event has a direct impact on the arrival intensity. Example applications include the econometric analysis of the timing of corporate events such as mergers, acquisitions or leveraged

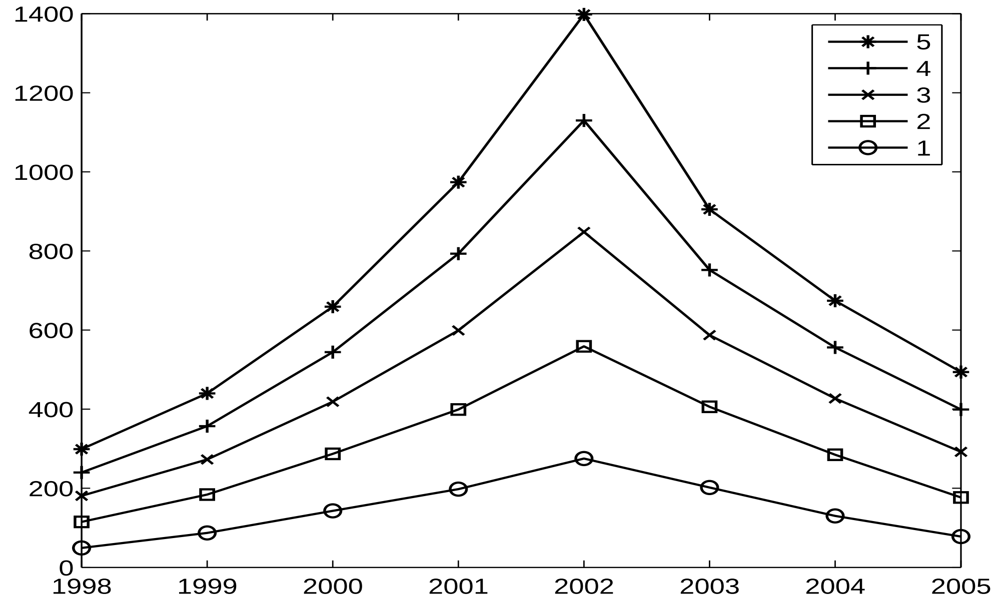

Figure 11: Forecast value at risk at the 95% level for the years 1998–2005 for each of several horizons (1-5 years), as implied by the zero-factor model with quadratic weight function. We use observations from 01/1970 to January 1st of each year to estimate the model. Monte Carlo simulation of the counting process N is used to estimate the conditional distribution of events during the year, given the observations. 50,000 paths were simulated.

buyouts, and the timing of security trades in an order book.

A Proofs

Recall that N is a non-explosive counting process with arrival times T_n and intensity λ. Further (D_n) is a sequence of identically distributed random variables that are independent of one another and independent of the Brownian motion W and the covariate process X, which is of the form $X_t = u(Y_t, t)$, where $u : \mathbb{R} \times \mathbb{R}_+ \to \mathbb{R}$ and Y is a standard Brownian motion. Here and below, the relevant measure is the actual, data-generating measure P and the relevant filtration is the complete information filtration $\mathbb{F}$ unless noted otherwise.

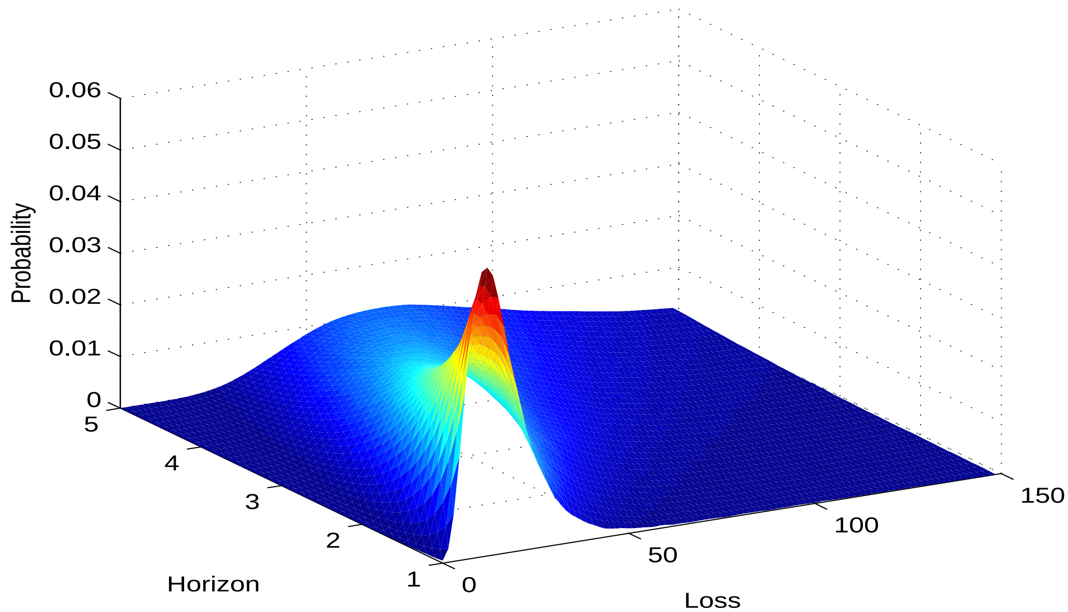

Figure 12: Forecast conditional distribution of the cumulative loss due to default at all future horizons between one and five years, given all observations in our data set as of October 2006, as implied by the zero-factor model with quadratic weight function. 50,000 paths were simulated to estimate this distribution.

Proof of Proposition 3.1. By integration by parts,

$$\begin{aligned} Z_t &= 1 - \int_0^t Z_{s-}(1-\lambda_{s-})ds - \int_0^t Z_s \log(\lambda_{s-})dN_s \\ &= 1 - \int_0^t Z_{s-}(\lambda_{s-}^{-1}-1)\lambda_{s-}ds + \int_0^t Z_{s-}(\lambda_{s-}^{-1}-1)dN_s \\ &= 1 + \int_0^t Z_{s-}(\lambda_{s-}^{-1}-1)dM_s \qquad (20) \end{aligned}$$

where $M = N - \int_0^\cdot \lambda_s ds$ is the compensated P-local jump martingale associated with N. Since the process $|Z_-(\lambda_-^{-1}-1)|$ is predictable and locally bounded, formula (20) implies that Z is a P-local martingale. Further, our hypothesis that $\int_0^\tau (\lambda_s^{-1}-1)^2\lambda_s ds$ is finite almost surely guarantees that Z is a martingale on $[0,\tau]$, see Lemma 19.6 in Liptser & Shiryayev (1989). Therefore, an equivalent probability measure $\widehat{P}$ is well defined by

$$\widehat{P}(B) = E(Z_\tau 1_B), \quad B \in \mathcal{F}_\tau.$$

Proposition 3.2 below implies that the counting process N has unit intensity on $[0,\tau]$ under $\widehat{P}$. Then, from Watanabe's theorem, it follows that N is a $\widehat{P}$-standard Poisson process on $[0, \tau]$. The response point process $L = \sum_{n=1}^{N} \ell(D_n)$ is a $\widehat{P}$-standard compound Poisson process on $[0, \tau]$. Under $\widehat{P}$, the jumps $\{\ell(D_n) : n = 1, \ldots, N_\tau\}$ of L are independent of another and independent of the path $\{N_t : 0 < t \leq \tau\}$, and they have the same distribution as they have under P. Further, Proposition 3.2 implies that Y is a $\widehat{P}$-standard Brownian motion, and therefore the path $\{X_t = u(Y_t, t) : 0 < t \leq \tau\}$ is $\widehat{P}$-independent of the compound Poisson path $\{L_t : 0 < t \leq \tau\}$. Moreover the P- and $\widehat{P}$-distributions of X_t agree. With the standard abuse of notation, these observations imply that

$$\widehat{P}(N_\tau = n, T \in dt, D = d, X \in dx) = \widehat{P}(N_\tau = n, T \in dt) P(D = d) P(X \in dx)$$

where $n \geq 0$ is an integer, $T = (T_1, \ldots, T_n)$ is a vector of event dates, $D = (D_1, \ldots, D_n)$ is a vector of event counts, and $X = (X_1, \ldots, X_m)$ is a vector of covariates on observation dates $(s_1, \ldots, s_m)$. Further, $t = (t_1, \ldots, t_n)$ is a vector of reals such that $0 < t_1 < \ldots < t_n \leq \tau$, $d = (d_1, \ldots, d_n)$ is a vector of non negative integers, and $x = (x_1, \ldots, x_m)$ is a vector of reals. By the Poisson and order statistics properties of N under $\widehat{P}$, we get

$$\widehat{P}(N_\tau = n, T \in dt) = \widehat{P}(T \in dt \,|\, N_\tau = n) \widehat{P}(N_\tau = n) = e^{-\tau} dt.$$

Finally we change the measure to get

$$\begin{aligned} &P(N_\tau = n, T \in dt, D = d, X \in dx) \\ &\qquad = \widehat{E}(Z_\tau^{-1} \,|\, N_\tau = n, T = t, D = d, X = x) \widehat{P}(N_\tau = n, T \in dt, D = d, X \in dx), \end{aligned}$$

and this equality, together with the previous observations, proves our assertion. □

Proof of Proposition 3.2. The Girsanov-Meyer theorem stated as Theorem 36 in Chapter III of Protter (2004) implies that if V is a P-local martingale such that the P-conditional covariation $\langle V, Z\rangle$ exists, then a $\widehat{P}$-local martingale is given by

$$V_t - \int_0^t \frac{1}{Z_{s-}} d\langle V, Z\rangle_s, \quad t \leq \tau. \tag{21}$$

The conditional covariation $\langle V, Z\rangle$ is the unique predictable process such that $[V, Z] - \langle V, Z\rangle$ is a P-local martingale. Using the representation (20),

$$[V, Z]_t = \int_0^t Z_{s-}(\lambda_{s-}^{-1} - 1) d[V, N]_s.$$

Since $[V, N]$ is locally of integrable variation, $\langle V, N\rangle$ exists and we find that

$$\langle V, Z\rangle_t = \int_0^t Z_{s-}(\lambda_{s-}^{-1} - 1) d\langle V, N\rangle_s.$$

Now formula (6) follows from formula (21). □

Proof of Proposition 3.3. It suffices to consider the expectation $\widehat{E}(Z_\tau^{-1}e^{-\tau}\,|\,B)$ for the event $B = \{N_\tau = n, T = t, D = d\}$, using the notation introduced in the proof of Proposition 3.1. For fixed τ, we let

$$\Pi = \exp\left(\int_0^\tau \log(\lambda_{s-})dN_s\right) = \lambda_{T_1-}\cdots\lambda_{T_{N_\tau}-}. \tag{22}$$

By the definition of the density Z, iterated expectations, and the fact that λ follows a $\widehat{P}$-Feller diffusion on $[t_n, \tau]$ that is driven by a $\widehat{P}$-Brownian motion W that is $\widehat{P}$-independent of the event B as argued in Section 3 above, we have

$$\begin{aligned}
\widehat{E}(Z_\tau^{-1}e^{-\tau}|B) &= \widehat{E}\left(\Pi\exp\left(-\int_0^\tau \lambda_s ds\right)\Big|B\right)\\
&= \widehat{E}\left(\widehat{E}\left\{\Pi\exp\left(-\int_0^\tau \lambda_s ds\right)\Big|B, \sigma(\lambda_s : s \le t_n, \lambda_\tau)\right\}\Big|B\right)\\
&= \widehat{E}\left(\Pi\exp\left(-\int_0^{t_n} \lambda_s ds\right)\widehat{E}\left\{\exp\left(-\int_{t_n}^\tau \lambda_s ds\right)\Big|B, \sigma(\lambda_s : s \le t_n, \lambda_\tau)\right\}\Big|B\right)\\
&= \widehat{E}\left(\Pi\exp\left(-\int_0^{t_n} \lambda_s ds\right)\phi_{t_n,\tau}(\lambda_{t_n}, \lambda_\tau)\Big|B\right)
\end{aligned} \tag{23}$$

where $\phi_{a,b}(v, w)$ is given in formula (8). We repeat the conditioning argument leading to formula (23) to obtain the expression

$$\widehat{E}(Z_\tau^{-1}e^{-\tau}\,|\,B) = \widehat{E}\left(\prod_{k=1}^n \lambda_{t_k-}\phi_{t_{k-1},t_k}(\lambda_{t_{k-1}}, \lambda_{t_k-})\phi_{t_n,\tau}(\lambda_{t_n}, \lambda_\tau)\Big|B\right).$$

This concludes the proof. □

Proof of Proposition 4.1. The measure $\widehat{P}$ is equivalent to the actual probability P on $\mathcal{F}_\tau$, with density Z_τ given by formula (4). It follows that P is absolutely continuous with respect to $\widehat{P}$ on $\mathcal{F}_t$ with density Z_t^{-1}, for $t \le \tau$. Now P is also absolutely continuous with respect to $\widehat{P}$ on the econometrician's σ-algebra $\mathcal{G}_t \subseteq \mathcal{F}_t$ for all $t \le \tau$, with density $\widehat{E}(Z_t^{-1}\,|\,\mathcal{G}_t)$. Formula (10) now follows since λ is positive and $\mathbb{F}$-adapted. □

B Evaluating the likelihood

The conditional $\widehat{P}$-expectation in formula (5) is trivial and the conditional density takes an explicit form if the intensity path over $[0, \tau]$ is a deterministic function of $(T_n, D_n)_{n=1,\ldots,N_\tau}$ and X. This is the case for the zero-factor model analyzed in Section 6.1. It is also the case for the one-factor model if X is informative, as in Section 6.2.2. If X is not informative, as in Section 6.2.1, then the $\widehat{P}$-expectation in formula (5) is non-trivial. Proposition 3.3 develops an alternative expression for this expectation. In this appendix, we show

how to evaluate this expression. We propose the following algorithm, which relies on the $\widehat{P}$-Markov property of the inter-arrival intensity and its known transition law under $\widehat{P}$. The idea is to construct a discrete-time, finite state Markov chain approximation to the Feller jump-diffusion λ whose jump times and sizes are given. The convergence of this approximation and the corresponding filters is discussed in Frey & Runggaldier (2001) and Kushner & Dupuis (2001). Below, we let $B = \{N_\tau = n, T = t, D = d\}$.

Initialization

(1) For integers $j = 1, \ldots, J$ and some $\Delta > 0$, let $I_j = [(j - 0.5)\Delta, (j + 0.5)\Delta]$.

(2) For each $t_k \in \{t_1, t_2, \ldots, t_n\}$, compute the non-central chi-squared transition matrix $\widehat{P}^k$ with elements $\widehat{P}_k(i,j) = \widehat{P}(\lambda_{t_k-} \in I_i \,|\, \lambda_{t_{k-1}} \in I_j, B)$ for $i, j = 1, 2, \ldots, J$.

(3) Define the state probability vectors $\widehat{p}_k$ and $\widehat{p}_k^*$ by $\widehat{p}_k(j) = \widehat{P}(\lambda_{t_k} \in I_j \,|\, B)$ and $\widehat{p}_k^*(j) = \widehat{P}(\lambda_{t_k-} \in I_j \,|\, B)$ for $j = 1, 2, \ldots, J$ and $k = 1, 2, \ldots, n$.

(4) Define the value funtion vectors V_k and V_k^* by $V_k(j) = \widehat{E}(\Pi_k \,|\, \lambda_{t_k} \in I_j, B)$ and $V_k^*(j) = \widehat{E}(\Pi_k \,|\, \lambda_{t_k-} \in I_j, B)$ where $\Pi_k = \prod_{i=1}^k \lambda_{t_i-}\phi_{t_{i-1},t_i}(\lambda_{t_{i-1}}, \lambda_{t_i-})$ for $j = 1, 2, \ldots, J$ and $k = 1, 2, \ldots, n$.

(5) Find the value s such that $s\Delta = \lambda_0$.

(6) Let $\widehat{p}_0(i) = 1, V_0(i) = 1$ if $i = s$ and $\widehat{p}_0(i) = 0, V_0(i) = 0$ otherwise.

For $k = 1, 2, \ldots, n$

(1) Compute $\widehat{p}_k^* = \widehat{P}_k \widehat{p}_{k-1}$.

(2) Compute $\phi_{t_{k-1},t_k}(j\Delta, i\Delta)$ for $i, j = 1, 2, \ldots, J$.

(3) Compute $V_k^*(i) = \frac{1}{\widehat{p}_k^*(i)} \sum_{j=1}^J (i\Delta)\phi_{t_{k-1},t_k}(j\Delta, i\Delta)\widehat{P}_k(i,j)V_{k-1}(j)\widehat{p}_{k-1}(j)$ for $i = 1, 2, \ldots, J$.

(4) Find r such that $r\Delta = \delta\ell(d_k)$.

(5) Let $\widehat{p}_k(j) = \widehat{p}_k^*(j - r), V_k(j) = V_k^*(j - r)$ for $j = r + 1, \ldots, J - 1$ and $\widehat{p}_k(j) = 0, V_k(j) = 0$ for $j = 1, 2, \ldots, r$.

(6) Compute $\widehat{p}_k(J) = \sum_{j=J-r}^J \widehat{p}_k^*(j)$.

(7) Compute $V_k(J) = \sum_{j=J-r}^J V_k^*(j)$.

Termination

(1) Compute $\widehat{P}_{n+1}(i,j) = \widehat{P}(\lambda_\tau \in I_i \,|\, \lambda_{t_n} \in I_j, B)$ for $i, j = 1, 2, \ldots, J$.

(2) Compute $\widehat{p}_{n+1} = \widehat{P}_{n+1}\widehat{p}_n$.

(3) Compute $V_{n+1}(i) = \frac{1}{\widehat{p}_{n+1}(i)} \sum_{j=1}^J \phi_{t_n,\tau}(j\Delta, i\Delta)\widehat{P}_{n+1}(i,j)V_n(j)\widehat{p}_n(j)$ for $i = 1, 2, \ldots, J$.

(4) Compute $\widehat{E}(Z_\tau^{-1}e^{-\tau} \mid B) = \sum_{j=1}^J V_{n+1}(j)\widehat{p}_{n+1}(j)$

The Initialization Step (2) can be implemented by first computing the matrix $\widehat{P}^\delta$ with elements $\widehat{P}^\delta(i,j) = \widehat{P}(\lambda_{t+\delta} \in I_i \mid \lambda_t \in I_j, B)$ for each $\delta \in \{\frac{m}{365} : m = 1, 2, 4, \ldots, 256 \text{ days}\}$, and then calculating the $\widehat{P}_k$ as the products of the appropriate $\widehat{P}^\delta$s.

The formula for $\widehat{V}_k^*(\cdot)$ in the Loop Step (3) is based on the following calculation:

$$
\begin{aligned}
V_k^*(i)\widehat{p}_k^*(i) &= \widehat{E}(\Pi_k \mid \lambda_{t_k-} \in I_i, B)\widehat{p}_k^*(i) \\
&= \widehat{E}(\Pi_k 1_{\{\lambda_{t_k-} \in I_i\}} \mid B) \\
&= \sum_{j=1}^J \widehat{E}(\Pi_k 1_{\{\lambda_{t_k-} \in I_i\}} \mid \lambda_{t_{k-1}} \in I_j, B)\widehat{p}_{k-1}(j) \\
&= \sum_{j=1}^J \widehat{E}(\lambda_{t_k-}\phi_{t_{k-1},t_k}(\lambda_{t_{k-1}}, \lambda_{t_k-})1_{\{\lambda_{t_k-} \in I_i\}} \mid \lambda_{t_{k-1}} \in I_j, B) \\
&\quad \times \widehat{E}(\Pi_{k-1} \mid \lambda_{t_{k-1}} \in I_j, B)\widehat{p}_{k-1}(j) \\
&\approx \sum_{j=1}^J (i\Delta)\phi_{t_{k-1},t_k}(j\Delta, i\Delta)\widehat{P}_k(i,j)V_{k-1}(j)\widehat{p}_{k-1}(j),
\end{aligned}
$$

where, conditional on $\{\lambda \in I_j\}$, we approximate the value of λ with $j\Delta$, the mean value of the interval. This approximation becomes exact for $\Delta \to 0$ (and $J \to \infty$). Thus, conditional on $\{\lambda_{t_{k-1}} \in I_j\}$ and $\{\lambda_{t_k-} \in I_i\}$, the variables $\lambda_{t_{k-1}}$, λ_{t_k-} and $\phi_{t_{k-1},t_k}(\lambda_{t_{k-1}}, \lambda_{t_k-})$ are replaced with $j\Delta$, $i\Delta$ and $\phi_{t_{k-1},t_k}(j\Delta, i\Delta)$ respectively. Both sides are then divided by $\widehat{p}_k^*(i)$ to obtain $V_k^*(i)$.

The Loop Step (5) is based on the relationship $\lambda_{t_k} = \lambda_{t_k-} + \delta\ell(d_k)$.

The Termination Steps (3) and (4) are based on the observation that there is no event during $(t_n, \tau]$, and therefore $\widehat{p}_{n+1} = \widehat{p}_{n+1}^*$ and $V_{n+1} = V_{n+1}^*$.

The optimal values for the discretization parameters J and Δ depend on the value of the parameter w, which is fixed for each likelihood evaluation. These values were chosen such that the discretization grid for the intensity is fine enough for the case at hand. The zero-factor model provided guidance for this choice. The parameters are in the order of 1000 for J and 0.2 for Δ. We have tested the algorithm for candidate discretization parameter settings by estimating $\widehat{E}(Z_\tau^{-1}e^{-\tau} \mid B)$ though simulation of intensity values from the non-central chi-squared distribution.

References

Arnsdorf, Matthias & Igor Halperin (2007), BSLP: markovian bivariate spread-loss model for portfolio credit derivatives. Working Paper, Quantitative Research J.P. Morgan.

Azizpour, Shahriar & Kay Giesecke (2008), Premia for correlated default risk. Working Paper, Stanford University.

Brémaud, Pierre (1980), *Point Processes and Queues – Martingale Dynamics*, Springer-Verlag, New York.

Broadie, Mark & Ozgur Kaya (2006), 'Exact simulation of stochastic volatility and other affine jump diffusion processes', *Operations Research* **54**(2), 217–231.

Chava, Sudheer & Robert Jarrow (2004), 'Bankruptcy prediction with industry effects', *Review of Finance* **8**, 537–569.

Collin-Dufresne, Pierre, Robert Goldstein & Jean Helwege (2003), Are jumps in corporate bond yields priced? Modeling contagion via the updating of beliefs. Working Paper, Carnegie Mellon University.

Cont, Rama & Andreea Minca (2008), Extracting portfolio default rates from CDO spreads. Working Paper, Columbia University.

Cox, John, Jonathan Ingersoll & Stephen Ross (1985), 'A theory of the term structure of interest rates', *Econometrica* **53**, 385–408.

Das, Sanjiv, Darrell Duffie, Nikunj Kapadia & Leandro Saita (2007), 'Common failings: How corporate defaults are correlated', *Journal of Finance* **62**, 93–117.

Das, Sanjiv, Laurence Freed, Gary Geng & Nikunj Kapadia (2006), 'Correlated default risk', *Journal of Fixed Income* **16**, 7–32.

Davis, Mark & Violet Lo (2001), Modeling default correlation in bond portfolios, *in* C.Alexander, ed., 'Mastering Risk Volume 2: Applications', Prentice Hall, pp. 141–151.

Dellacherie, Claude & Paul-André Meyer (1982), *Probabilities and Potential*, North Holland, Amsterdam.

Delloye, Martin, Jean-David Fermanian & Mohammed Sbai (2006), Estimation of a reduced-form credit portfolio model and extensions to dynamic frailties. Working Paper, BNP Paribas.

Ding, Xiaowei, Kay Giesecke & Pascal Tomecek (2006), Time-changed birth processes and multi-name credit derivatives. *Operations Research*, forthcoming.

Duffie, Darrell, Andreas Eckner, Guillaume Horel & Leandro Saita (2008), Frailty correlated default. *Journal of Finance*, forthcoming.

Duffie, Darrell, Leandro Saita & Ke Wang (2006), 'Multi-period corporate default prediction with stochastic covariates', *Journal of Financial Economics* **83**(3), 635–665.

Errais, Eymen, Kay Giesecke & Lisa Goldberg (2006), Pricing credit from the top down with affine point processes. Working Paper, Stanford University.

Figlewski, Stephen, Halina Frydman & Weijian Liang (2008), Modeling the effect of macroeconomic factors on corporate default and credit rating transitions. Working Paper, New York University.

Frey, Rüdiger & Jochen Backhaus (2008), Pricing and hedging of portfolio credit derivatives with interacting default intensities. *International Journal of Theoretical and Applied Finance*, forthcoming.

Frey, Rüdiger & Wolfgang Runggaldier (2001), 'Nonlinear filtering techniques for volatility estimation with a view towards high frequency data', *International Journal of Theoretical and Applied Finance* **4**, 271–300.

Frey, Rüdiger & Wolfgang Runggaldier (2007), Credit risk and incomplete information: a nonlinear-filtering approach. Working Paper, Department of Mathematics, Universität Leipzig.

Giampieri, Giacomo, Mark Davis & Martin Crowder (2005), 'A hidden markov model of default interaction', *Quantitative Finance* **5**, 27–34.

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

Giesecke, Kay (2007), The correlation-neutral measure for portfolio credit. Working Paper, Stanford University.

Giesecke, Kay & Baeho Kim (2007), Estimating tranche spreads by loss process simulation, *in* S. G.Henderson, B.Biller, M.-H.Hsieh, J.Shortle, J. D.Tew & R. R.Barton, eds, 'Proceedings of the 2007 Winter Simulation Conference', IEEE Press, pp. 967–975.

Giesecke, Kay, Hossein Kakavand & Mohammad Mousavi (2008), Simulating point processes by intensity projection, *in* S. J.Mason, R. R.Hill, L.Moench & O.Rose, eds, 'Proceedings of the 2008 Winter Simulation Conference', IEEE Press.

Giesecke, Kay & Lisa Goldberg (2004), 'Sequential defaults and incomplete information', *Journal of Risk* **7**(1), 1–26.

Giesecke, Kay & Lisa Goldberg (2005), A top down approach to multi-name credit. Working Paper, Stanford University.

Giesecke, Kay & Pascal Tomecek (2005), Dependent events and changes of time. Working Paper, Stanford University.

Hamilton, David (2005), Moodys senior ratings algorithm and estimated senior ratings. Moodys Investors Service.

Hawkes, Alan G. (1971), 'Spectra of some self-exciting and mutually exciting point processes', *Biometrika* **58**(1), 83–90.

Hillegeist, Stephen, Elizabeth Keating, Donald Cram & Kyle Lunstedt (2004), 'Assessing the probability of bankruptcy', *Review of Accounting Studies* **9**, 5–34.

Jarrow, Robert A. & Fan Yu (2001), 'Counterparty risk and the pricing of defaultable securities', *Journal of Finance* **56**(5), 555–576.

Jorion, Philippe & Gaiyan Zhang (2007*a*), Credit correlations from counterparty risk. Working Paper, University of California at Irvine.

Jorion, Philippe & Gaiyan Zhang (2007*b*), 'Good and bad credit contagion: Evidence from credit default swaps', *Journal of Financial Economics* **84**(3), 860–883.

Kliemann, Wolfgang, Giorgio Koch & Federico Marchetti (1990), 'On the unnormalized solution of the filtering problem with counting process observations', *IEEE Transactions on Information Theory* **36**(6), 1415–1425.

Koopman, Siem Jan, Andre Lucas & A. Monteiro (2008), 'The multi-stage latent factor intensity model for credit rating transitions', *Journal of Econometrics* **142**(1), 399–424.

Kushner, Harold & Paul Dupuis (2001), *Numerical Methods for Stochastic Control Problems in Continuous Time*, Springer-Verlag, New York.

Lang, Larry & Rene Stulz (1992), 'Contagion and competitive intra-industry effects of bankruptcy announcements', *Journal of Financial Economics* **32**, 45–60.

Liptser, Robert & Albert Shiryayev (1989), *Statistics of Stochastic Processes*, Springer-Verlag, Berlin.

Longstaff, Francis & Arvind Rajan (2008), 'An empirical analysis of collateralized debt obligations', *Journal of Finance* **63**(2), 529–563.

Lopatin, Andrei & Timur Misirpashaev (2007), Two-dimensional Markovian model for dynamics of aggregate credit loss. Working Paper, Numerix.

McNeil, Alexander & Jonathan Wendin (2007), 'Bayesian inference for generalized linear mixed models of portfolio credit risk', *Journal of Empirical Finance* **14**(2), 131–149.

Meyer, Paul-André (1971), Démonstration simplifée d'un théorème de Knight, *in* 'Séminaire de Probabilités V, Lecture Note in Mathematics 191', Springer-Verlag Berlin, pp. 191–195.

Ogata, Yosihiko (1978), 'The asymptotic behavior of maximum likelihood estimators of stationary point processes', *Annals of the Institute of Statistical Mathematics* **30**(A), 243–261.

Prahl, Jürgen (1999), A fast unbinned test of event clustering in poisson processes. Working Paper, Universität Hamburg.

Protter, Philip (2004), *Stochastic Integration and Differential Equations*, Springer-Verlag, New York.

Schönbucher, Philipp (2004), Information-driven default contagion. Working Paper, ETH Zürich.

Segall, Adrian, Mark Davis & Thomas Kailath (1975), 'Nonlinear filtering with counting observations', *IEEE Transactions on Information Theory* **21**(2), 143–149.

Segall, Adrian & Thomas Kailath (1975), 'The modeling of randomly modulated jump processes', *IEEE Transactions on Information Theory* **21**(2), 135–143.

Shumway, Tyler (2001), 'Forecasting bankruptcy more accurately: A simple hazard model', *Journal of Business* **74**, 101–124.

Tavella, Domingo & Martin Krekel (2006), Pricing nth-to-default credit derivatives in the PDE framework. Working Paper, Octanti Associates.