Systemic Risk: What Defaults Are Telling Us

Kay Giesecke* and Baeho Kim†

Stanford University

September 16, 2009; this draft March 7, 2010‡

Abstract

This paper defines systemic risk as the conditional probability of failure of a large number of financial institutions, and develops maximum likelihood estimators of the term structure of systemic risk in the U.S. financial sector. The estimators are based on a new dynamic hazard model of failure timing that captures the influence of time-varying macro-economic and sector-specific risk factors on the likelihood of failures, and the impact of spillover effects related to missing/unobserved risk factors or the spread of financial distress in a network of firms. In- and out-of-sample tests demonstrate that the fitted risk measures accurately quantify systemic risk for each of several risk horizons and confidence levels, indicating the usefulness of the risk measure estimates for the macro-prudential regulation of the financial system.

*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`.

†Department of Management Science & Engineering, Stanford University, Stanford, CA 94305-4026, USA, Fax (650) 723 1614, email: `baehokim@stanford.edu`, web: `www.stanford.edu/~baehokim`.

‡We are grateful for discussions with Jorge Chan-Lau, Darrell Duffie, Marco Espinosa, Juan Solé, and for comments from Laura Kodres, Brenda González-Hermosillo and participants at the Monetary and Capital Markets Department Seminar at the IMF, and at Bank Negara Malaysia.

1 Introduction

The systemic risk in the financial sector is difficult to measure. This makes it hard for regulators and policy makers to address it effectively. A major challenge is to take account of potential spillover effects when quantifying systemic risk. Information based spillover effects are related to omitted or imperfectly observed risk factors influencing several firms, and the associated Bayesian learning at failure events. Spillover effects may also be caused by contagion in an increasingly opaque network of interbank loans, derivative trading relationships, and other links between firms. The Lehman Brothers and AIG events highlight the importance of these network effects. Spillover concerns were arguably at the center of the government's decision to bail out AIG, whose collapse would presumably have caused a major market disruption.

This paper develops maximum likelihood estimators of the term structure of systemic risk, defined as the conditional probability of failure of a sufficiently large fraction of the total population of financial institutions. Its main contribution over prior work is to incorporate the statistical implications of spillover effects when measuring systemic risk. Unlike the existing estimators, which focus on the influence of observable risk factors on systemic risk, the estimators developed in this paper also account for the role of omitted or imperfectly observed risk factors influencing several firms, and the potential feedback effects from failure events. Applying these estimators to the U.S. financial system during 1987–2008, we find that the part of systemic risk *not* explained by the variation of the trailing return of the S&P 500, the lagged slope of the U.S. yield curve, the default and TED spreads, and other observable variables can be substantial, and tends to be higher in periods of adverse economic conditions. The results indicate that the systemic risk in the U.S. financial sector can be much greater than would be estimated under the standard assumption in the bank failure prediction literature that bank failure clusters arise only from exposure to observable risk factors common to many firms.

Our estimators are based on a new dynamic hazard model of correlated failure timing that extends the traditional proportional hazards specification used by Das, Duffie, Kapadia & Saita (2007), Duffie, Saita & Wang (2007), McDonald & Van de Gucht (1999) and others to predict non-financial corporate default, and by Brown & Dinc (2005), Brown & Dinc (2009), Lane, Looney & Wansley (1986), Whalen (1991), Wheelock & Wilson (2000) and others to forecast bank failures. The distinguishing feature of our formulation is an additional hazard term designed to capture the statistical implications of spillover effects within and between the real and financial sectors. While controlling for the influence of observable risk factors, our specification also incorporates the implications of missing or incompletely observed risk factors, a source of spillover effects emphasized by Collin-Dufresne, Goldstein & Helwege (2009), Duffie, Eckner, Horel & Saita (2009) and others for the real sector, and by Acharya & Yorulmazer (2008), Aharony & Swary (1983), Cooperman, Lee & Wolfe (1992) and others for the financial sector. It also addresses the implications of spillover effects channeled through the complex web of derivatives counterparty relations, interbank loans, trade credit chains, parent-subsidiary relationships, and other links between firms. A traditional proportional hazard formulation ignores spillover effects; it focuses on the dependence of default timing on observable explanatory variables whose dynamics are assumed to be unaffected by failure events.

We estimate our extended hazard model using data on economy-wide default experience in the U.S. for the period January 1987 to December 2008, and a collection of time-varying explanatory covariates that capture the influence of economic conditions on failure timing. To address the implications of industrial defaults for bank failures and vice versa, we develop a two-step maximum likelihood estimation approach. In contrast to a traditional one-step estimation method, our approach does not treat the sequence of financial failure events in isolation, but in the context of the sequence of defaults in the wider economy. It seeks to extract the information contained in industrial defaults relevant for predicting financial failures, and allows us to capture the dynamic interaction between the real and financial sectors when measuring systemic risk.

Statistical tests demonstrate the in-sample fit of our new hazard model of correlated failure timing, the significance of the additional hazard term, and the out-of-sample predictive power of our fitted measures of systemic risk. For example, the fitted measures accurately forecast the quantiles of the fraction of failures in the U.S. financial system during 1998–2009, for each of several confidence levels and forecast horizons. These tests validate our modeling approach and the two-step inference procedure. They indicate that our approach leads to useful measures of systemic risk.

The proven predictive power of our estimators makes them well-suited for monitoring the level of systemic risk in the U.S. financial sector. The estimators provide time-series and term-structure perspectives of systemic risk, information that regulators and policy makers can use to implement a macro-prudential approach to supervising the financial system. The estimators indicate that systemic risk increased dramatically during the second half of 2008, and reached unprecedented levels towards the end of 2008. While the magnitude of economy-wide default risk in 2008 is roughly comparable to the level during the burst of the internet bubble around 2001, the estimators suggest that the systemic risk during that period is dwarfed by the magnitude of systemic risk in 2008. During the entire sample period, the failure of a financial firm is estimated to have a relatively greater impact on systemic risk than the default of an industrial firm.

The remainder of this introduction discusses the related literature. Section 2 introduces our measures of systemic risk and discusses their properties. Section 3 develops the statistical methodology. Section 4 describes our data, the basic estimation results, and their statistical evaluation. Section 5 analyzes the behavior of systemic risk during 1987–2008, provides risk forecasts for future periods, and evaluates these forecasts. Section 6 assesses the impact on systemic risk of the failure of an industrial firm or a financial institution. Section 7 concludes. There are several appendices.

1.1 Related literature

There is a substantial literature on bank failure prediction, which includes Brown & Dinc (2005), Brown & Dinc (2009), Cole & Wu (2009), Lane et al. (1986), Whalen (1991), Wheelock & Wilson (2000) and many others. These papers employ traditional hazard models, in which the timing of bank failures is influenced by observable explanatory covariates, which may be time-varying. They focus on predicting individual bank failures, and do not directly address the correlation between failures, which is the driving force behind systemic risk. To incorporate the different sources of this dependence, we significantly extend the traditional hazard model. Our formulation assumes that firms are exposed to common, time-varying risk factors. Movements of these observable factors affect firms across the board, and induce failure clusters. Our formulation also includes an additional "spillover hazard" term, which seeks to address the clustering of failures not due to the variation of observable risk factors. The spillover hazard depends on the timing of past defaults and the volume of defaulted debt. It models the influence of past defaults on failure timing, with a role for the size of a default. This influence can be caused by information based spillover effects, i.e., Bayesian learning at defaults about risk factors that are unobserved or missing from the list of explanatory covariates as in Aharony & Swary (1996), Collin-Dufresne et al. (2009), Duffie et al. (2009), Giampieri, Davis & Crowder (2005), Giesecke (2004), Koopman, Lucas & Monteiro (2008) and others. It can also be due to the spread of distress from one firm to another, as in Azizpour & Giesecke (2008), Jorion & Zhang (2007), Lando & Nielsen (2009) and others. Distress may be channeled through trade credit or buyer/supplier relationships in the real sector, and derivatives counterparty relations and interbank loans in the financial sector (see Upper & Worms (2004) and others in this regard). The goal of our dynamic hazard model is to capture the statistical implications of spillover effects for failure timing without needing to be precise a priori about the economic mechanisms behind them. Accordingly, we do not provide a decomposition of the spillover effects implied by the data.

Our measures of systemic risk are related to alternative measures discussed in the literature. Adrian & Brunnermeier (2009) propose a family of quantile measures of systemic risk that are based on the distribution of the change of the market value of total financial assets of public financial institutions. They estimate these risk measures from time series of equity returns and balance sheet information using quantile regressions. Acharya, Pedersen, Philippon & Richardson (2009) develop and estimate expected shortfall measures of systemic risk that are based on the distribution of market equity returns of financial institutions. Lehar (2005) takes a related structural perspective, and defines systemic risk in terms of adverse changes in the market values of several institutions.

We propose quantile measures of systemic risk that are predicated on the conditional distribution of the failure rate in the financial system, and are estimated from actual failure experience. This is an important difference to Adrian & Brunnermeier (2009), Acharya et al. (2009) and Lehar (2005), who assess systemic risk in terms of adverse asset price changes across the financial sector, and use equity price data to estimate the corresponding risk measures. By tying systemic risk to clustered failures in the financial sector and focusing on actual failure timing data, our measures tend to be less susceptible to equity market factors unrelated to systemic risk. Nevertheless, they incorporate market information through the explanatory variables.

Since they summarize the relevant properties of a conditional distribution, our risk measures are dynamic, and vary through time with the available information. They also define a term structure of systemic risk over multiple future periods that incorporates the dynamics of the explanatory macro-economic and sector-wide variables. The basic risk measures extend naturally to co-risk measures à la Adrian & Brunnermeier (2009). A co-risk statistic quantifies the impact on systemic risk of an adverse event, such as the collapse of an industrial firm or financial institution.

Avesani, Pascual & Li (2006), Bhansali, Gingrich & Longstaff (2008), Chan-Lau & Gravelle (2005), Huang, Zhou & Zhu (2009) and others estimate alternative risk measures from market rates of credit derivatives. These measures quantify systemic risk relative to a risk-neutral pricing measure, and incorporate the risk premia investors demand for bearing correlated default risk. Our measures are based on actual failure behavior rather than market prices, and do not reflect risk premia.

Elsinger, Lehar & Summer (2006) develop and estimate a static network model of the interbank lending market to incorporate spillover phenomena induced by interbank loans when quantifying systemic risk; see also Eisenberg & Noe (2001).

Staum (2009) considers the total premium required to insure all deposits in the banking system as a measure of systemic risk. A bank's contribution to this risk measure is proposed as the bank's deposit insurance premium.

2 Measures of systemic risk

This section discusses our definition of systemic risk, describes a measure to quantify systemic risk, and examines the basic properties of this measure.

We define systemic risk in the financial sector as the conditional probability of failure of a sufficiently large fraction of the total population of institutions in the financial system. This definition targets the scenario of a failure cluster of financial institutions, potentially as part of a larger cluster of economy-wide defaults. Such a cluster could be due to a severe macro-economic shock, or a contagious spread of distress from one institution to another. Financial distress can be propagated through the informational and contractual relationships within the financial system, or the relationships between financial institutions and other non-financial firms. Lehman Brothers is an example of how the collapse of a single institution can induce distress at multiple other entities.

To provide a quantitative measure of systemic risk, consider the process N counting defaults in the financial system.[1] The value N_t represents the number of defaults in the financial system observed by time t. For a given horizon T, consider the conditional distribution at time $t < T$ of the default rate in the financial system, given by $D_t(T) = (N_T - N_t)/W_t$, where W_t denotes the number of financial institutions existing at t. This distribution gives the likelihood of failure by T of any fraction of the population of financial institutions at t. The right tail of this distribution reflects the magnitude of systemic risk. To measure this magnitude more precisely, we consider statistics that summarize the information in the tail of the distribution. A standard statistic is a quantile of the distribution, or value at risk. The value at risk $V_t(\alpha, T)$ at level $\alpha \in (0, 1)$ is the smallest number $x \geq 0$ such that the conditional probability at t that the default rate $D_t(T)$ during $(t, T]$ exceeds x is no larger than $(1 - \alpha)$.

The value at risk $V_t(\alpha, T)$ of the financial system is intuitive and easily communicated, relying on the popularity of value at risk in the financial industry. There are other advantages. As indicated by the notation, $V_t(\alpha, T)$ depends on the conditioning time t, and thus changes over time as new information is revealed. This leads to a dynamic risk measure. The value $V_t(\alpha, T)$ also depends on the risk horizon T. By varying T for fixed t we obtain a term structure of systemic risk. Further, as shown in Section 6, $V_t(\alpha, T)$ extends naturally to a co-risk measure that quantifies the contribution to systemic risk of a particular event, such as the default of a financial institution.

The quantification of systemic risk need not be predicated on the value at risk. Our statistical methodology focuses on the entire conditional distribution of $D_t(T)$, so our analysis extends to alternative downside risk measures such as the expected shortfall measure estimated by Acharya et al. (2009). This measure is defined as the conditional mean of $D_t(T)$ given $D_t(T) \geq c$, where c is some high level, such as $V_t(\alpha, T)$. While the value at risk is silent about the magnitude of the failure rate in excess of $V_t(\alpha, T)$, expected shortfall provides more detailed information about the severity of large failure clusters. More generally, our analysis extends to any statistic of the conditional distribution at t of the system-wide default rate $D_t(T)$, including the moments and other tail risk measures. Moreover, our analysis extends to risk measures of the conditional distribution of the value-weighted default rate, which takes account of the default volume.

The measures of systemic risk we propose are distinct from the measures discussed in the literature. The fundamental difference is the underlying distribution. While we define systemic risk in terms of the distribution of the failure rate in the financial system, Adrian & Brunnermeier (2009), Acharya et al. (2009) and Lehar (2005) relate systemic risk to the distribution of the change of the market equity value of financial institutions. Avesani et al. (2006), Chan-Lau & Gravelle (2005), Huang et al. (2009) and others define systemic risk in terms of a risk-neutral probability, which reflects the risk premia investors demand for bearing correlated default risk.

[1] We fix a complete probability space $(\Omega, \mathcal{F}, P)$ with an information filtration $(\mathcal{F}_t)_{t \geq 0}$ that satisfies the usual conditions. Here, P denotes the actual (empirical) probability measure.

3 Statistical methodology

This section develops a likelihood approach to estimating the measures of systemic risk proposed in Section 2. In a first step, we formulate and estimate a new hazard, or intensity-based, model of economy-wide default timing. In a second step, we extract the system-wide failure intensity from the economy-wide default intensity. The fitted system-wide intensity then leads to estimators of our systemic risk measures.

3.1 Economy-wide default timing

Consider the process N^* counting defaults in the economy. The value N^*_t is the number of defaults observed by time t. We suppose that N^* has hazard rate or *intensity* λ^*, which represents the conditional mean default rate in the economy and is measured in events per year. We assume that the intensity evolves through time according to the model

$$\lambda^*_t = \exp(\beta^* X^*_t) + \int_0^t e^{-\kappa(t-s)} dJ_s \tag{1}$$

where X^* is a vector of time-varying explanatory covariates specified in Section 4.2, β^* is a vector of constant parameters, κ is a strictly positive parameter, and

$$J_t = \nu_1 + \cdots + \nu_{N^*_t} \tag{2}$$

where $\nu_n = \gamma + \delta \max(0, \log D^*_n)$. Here, γ and δ are non-negative parameters, and D^*_n is the default volume, i.e. the total amount of debt outstanding at default of the n-th defaulter, measured in million dollars.[2]

The intensity (1) is the sum of two terms. The first term, called *baseline hazard* below, takes a standard Cox proportional hazards form. It models the influence on default arrivals of explanatory covariates X^*, and captures the clustering of defaults due to the exposure of different firms to variations in X^*. The proportional hazards formulation is used by Das et al. (2007), Duffie et al. (2007), McDonald & Van de Gucht (1999) and many others to predict industrial defaults, and by Brown & Dinc (2005), Brown & Dinc (2009), Cole & Wu (2009), Lane et al. (1986), Whalen (1991), Wheelock & Wilson (2000) and others to predict bank failures. We follow these references and estimate the coefficient β^* under the assumption that the dynamics of the variables X^* are not affected by defaults.

The second term, called *spillover hazard*, is not present in the traditional proportional hazards formulation. It models the influence of past defaults on current default rates, which is not captured by the baseline hazard term. At an event, the default rate jumps, with magnitude given by γ plus δ times the positive part of the logarithm of the defaulter's total outstanding debt, which is a proxy of the defaulter's firm size.[3] Thus, the bigger a

[2] We assume that each variable $\max(0, \log D^*_n)$ has finite mean, and that each component of X^*_t is finite almost surely. Under these conditions, the process N^* is non-explosive.

[3] For the purposes of our analysis, we found the total amount of debt outstanding at default to be a better measure of firm size than market capitalization, which was used by Shumway (2001) and others to predict non-financial corporate default.

defaulter the greater the impact of the event, with minimum impact governed by γ. After an event, the intensity decays to the baseline hazard, exponentially at rate κ.

The spillover hazard term is motivated by the results of the empirical analyses of Aharony & Swary (1996), Azizpour & Giesecke (2008), Collin-Dufresne et al. (2009), Das et al. (2007), Duffie et al. (2009), Lando & Nielsen (2009) and others. For U.S. corporate defaults, these papers found evidence of the presence of spillover effects related to contagion and unobserved or missing explanatory covariates, called frailties. With contagion, a default increases the likelihood of additional defaults, a process that may be channeled through trade credit or buyer/supplier relationships in the real sector, and derivatives counterparty relations and interbank loans in the financial sector. With frailty, Bayesian updating of the conditional distribution of the relevant but omitted or unobserved explanatory variables leads to a jump of the econometrician's intensity at a default. The spillover hazard term in (1) seeks to capture the statistical implications of these spillover effects for failure timing, by letting the intensity λ^* jump at a default. In particular, it is designed to replicate the excess default clustering not caused by the variation of the observable covariates X^* defining the baseline hazard. An advantage of this reduced-form formulation is that we do not need to be precise a priori about which of the economic mechanisms is behind the spillover effects. On the other hand, when taken to the data, this formulation does not offer information about the relative importance of the sources of the spillover effects. For an analysis of these sources for U.S. corporate defaults, see Azizpour & Giesecke (2008).

The inference problem for the default timing model (1)–(2) is addressed as follows. Letting $\theta = (\beta^*, \kappa, \gamma, \delta)$ be the set of parameters of the intensity $\lambda^* = \lambda^*(\theta)$, Θ be the set of admissible parameters, and $[0, t]$ be the sample period, we solve the log-likelihood problem

$$\sup_{\theta \in \Theta} \int_0^t (\log \lambda^*_{s-}(\theta) dN^*_s - \lambda^*_s(\theta) ds). \tag{3}$$

The calculation of the likelihood function is based on a measure change argument. Given a trajectory of X^*, the log-likelihood function takes a closed form, allowing for computational tractability of estimation. Under technical conditions stated in Ogata (1978), the maximum likelihood estimator of θ is asymptotically normal and efficient.

We have experimented with several alternative model formulations, including a conventional proportional hazards model in which average spillover effects are captured by a covariate given by the trailing 1-year default rate, as in Duffie et al. (2009). We have also tested alternative specifications of the impact variables ν_n in (2). However, based on the in- and out-of-sample tests described in Section 4.3 below, we found these alternatives to be statistically inferior to the model (1)–(2).

3.2 System-wide default timing

Next we extract from the fitted economy-wide model λ^* the dynamics of system-wide defaults, i.e., failures in the financial system. This is based on the following result.

Proposition 3.1. *There is a (predictable) process Z taking values in the unit interval, such that the intensity λ of system-wide failures is given by $\lambda = \lambda^* Z$.*

Proof. The system-wide failure times form a subsequence of the economy-wide default times. The existence and uniqueness of Z follows from the Radon-Nikodym theorem applied to the random measures associated with the time-integrals of the intensities λ^* and λ. The predictability of Z follows from the predictability of the processes generated by these time-integrals. □

The value Z_t is the conditional probability at t that a firm in the financial system defaults next, given a default in the economy in the next instant. For a precise statement, see Proposition 3.1 in Giesecke, Goldberg & Ding (2009). We formulate and estimate a parametric model of Z, which then leads to λ via Proposition 3.1.

We use probit regression to estimate the process Z from the observed economy- and system-wide default counting processes N^* and N, respectively. Letting Y_n be a binary response variable equal to one if the n-th defaulter belongs to the financial system and 0 otherwise, we obtain a value Y_n for each economy-wide default time T_n^* in the sample. Each Y_n is a Bernoulli variable with success probability $Z_{T_n^*}$, where[4]

$$Z_t = Z_t(\beta) = \Phi(\beta X_{t-}) \tag{4}$$

and where Φ is the cumulative distribution function of a standard normal variable, X_t is a vector of time-varying explanatory covariates specified in Section 4.2, and β is a vector of constant parameters. Given observations $(Y_n)_{n=1,\ldots,N_t^*}$ and $(X_s)_{s\leq t}$ during the sample period $[0,t]$, we estimate β by solving the log-likelihood problem

$$\sup_{\beta\in\Sigma} \sum_{n=1}^{N_t^*} \left[Y_{T_n^*} \log(Z_{T_n^*}(\beta)) + (1 - Y_{T_n^*}) \log(1 - Z_{T_n^*}(\beta)) \right] \tag{5}$$

where Σ is the set of admissible parameters. The maximum likelihood estimator of β is consistent, asymptotically normal and efficient if the covariance matrix of the vector of regressors exists and is non-singular. See McCullagh & Nelder (1989) for details. It can also be shown that the log-likelihood function is globally concave in β, and therefore a standard numerical optimization routine converges quickly to the unique maximum.

The two-step approach to estimating λ has a significant advantage over an alternative one-step approach in which λ would be estimated directly based on the historical default experience in the financial system. The two-step approach allows us to extract the

[4]We experimented with several alternative link functions, including a logit model. All these alternatives were found to be statistically inferior to the probit model.

information contained in the observed default times of non-financial firms, which otherwise would not be utilized in the estimation process. Financial firms are intertwined with the real sector, so defaults in that sector clearly have an influence on financial firms, and vice versa. Our estimation approach seeks to capture this influence. It responds to an argument made by Schwarcz (2008) and many others that systemic risk measures should account for the relationship between financial institutions and industrial firms.

The two-step approach has another, statistical advantage. Failures in the financial system are relatively rare. The number of economy-wide defaults is much larger, leading to a greater sample size and more accurate inference.[5]

3.3 Measures of risk

The intensity $\lambda = \lambda^* Z$ governs the dynamics of the system-wide default process N, and hence the measures of systemic risk introduced in Section 2. Given the fitted models of λ^* and Z, we estimate the entire conditional distribution at t of the system-wide default rate $D_t(T)$ by exact Monte Carlo simulation of default times during $(t, T]$.[6] From the conditional distribution we obtain unbiased estimates of the value at risk $V_t(\alpha, T)$ or any other risk measure based on the distribution of $D_t(T)$ or related quantities, including the value-weighted default rate.

The risk measure estimates take account of the idiosyncratic and clustered default risk of financial institutions. They capture several sources of default clustering, including the exposure of institutions to the common risk factors represented by the covariate vector X^*, and spillover effects within the financial sector and between the industrial and financial sectors. The risk measure estimates reflect the time-variation of X^* and the cross-sectional variation of the default volume D^*_n. As detailed in Appendices A and B, this is based on a vector autoregressive time-series model of the covariates, and a generalized Pareto model of the default volume. The importance for industrial default prediction of incorporating the time-series dynamics of explanatory covariates was emphasized by Duffie et al. (2007).

4 Empirical analysis

This section describes the default timing data, the data on explanatory covariates, our basic estimation results, and their statistical evaluation.

[5]For our sample period 1987-2008, the number of system-wide failures is 83 while the number of economy-wide defaults is 1193.

[6]The simulation is based on an acceptance/rejection scheme. Details are available upon request.

4.1 Default timing data

Our sample period is 1/1/1987 to 12/31/2008.[7] Data on U.S. corporate default timing were obtained from Moody's Default Risk Service. 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. 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. This leaves us with 1193 economy-wide defaults.

For the purpose of analyzing systemic risk, we take the U.S. financial system to be the set of firms classified in Moody's industry category "Banking" or "FIRE" (Finance, Insurance and Real Estate).[8] This set includes commercial and investment banks, bank holding companies, credit unions, thrifts, investment management, trading, leasing, mortgage and securities firms, financial guarantors, insurance and insurance brokerage firms, and REITs and REOCs. Figure 1 shows the 1-year economy- and system-wide default rates during the sample period, along with default volume information obtained from Moody's Default Risk Service.[9]

4.2 Covariates

We examine the influence on systemic risk of two types of macro-economic and sector-wide variables, which are measured monthly. These include:

(1) The trailing 1-year return on the S&P500 index, obtained from Economagic. Duffie et al. (2007) found this variable to be a significant predictor of industrial defaults.

(2) The 1-year lagged slope of the yield curve, computed as the spread between 10-year and 3-month Treasury constant maturity rates, as a forward-looking indicator of

[7] This period was determined by the availability of data for the covariates specified in Section 4.2. Default data for the period 1/1/2009 to 6/30/2009 were used for the out-of-sample analysis.

[8] Moody's uses several industry classifications. Our analysis is based on the "Moody's 11" scheme, which specifies 11 industries: 1. Banking, 2. Capital Industries, 3. Consumer Industries, 4. Energy and Environment, 5. FIRE, 6. Media and Publishing, 7. Retail and Distribution, 8. Sovereign and Public Finance, 9. Technology, 10. Transportation, and 11. Utilities.

[9] As explained by Hamilton (2005), the volume reported by Moodys excludes debt obligations that do not reflect the fundamental default risk of the obligor such as structured finance transactions, short-term debt (e.g., commercial paper), secured lease obligations, and so forth.

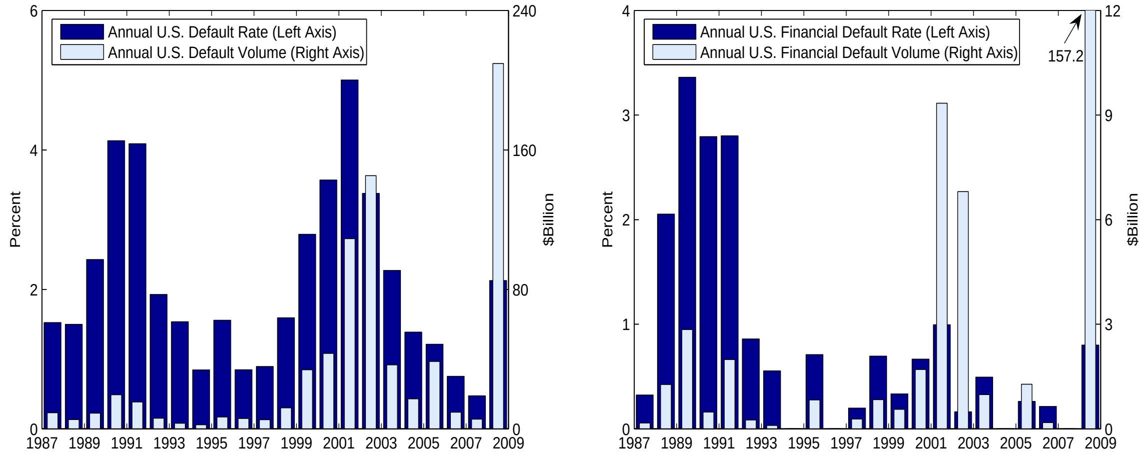

Figure 1: Default timing and volume data. *Left panel*: 1-year economy-wide default rate in the universe of Moody's rated issuers. *Right panel*: 1-year system-wide default rate. The defaults of Lehman Brothers and Washington Mutual contributed to over 80% of the system-wide default volume in 2008. Source: Moody's Default Risk Service.

real economic activity. Estrella & Trubin (2006) found this variable to have strong predictive power for future recessions. We obtained the H.15 release of Treasury rates from the website of the Federal Reserve Bank of New York.

(3) The default spread, defined as the yield differential between Moody's seasoned Aaa-rated and Baa-rated corporate bonds. Chen, Collin-Dufresne & Goldstein (2008) argue that the default spread is a measure of aggregate credit risk that is largely unaffected by bond market frictions such taxes and liquidity. The data were obtained from the website of the Federal Reserve Bank of New York. The left panel of Figure 2 shows the time series of the default spread and the slope of the yield curve.

(4) The TED (Treasury-Eurodollar) spread, defined as the difference between the 3-month LIBOR and 3-month Treasury rates, as an indicator of credit risk in the financial system.[10] We obtained the historical LIBOR rates from Economagic. Figure 3 shows the TED spread during the sample period, with significant events indicated.

(5) The trailing 1-year returns on banking and FIRE portfolios, as a proxy for business cycle activity in the financial system. The data were obtained from the website of Kenneth French.[11] The right panel of Figure 2 shows the return series.

[10] An increase of the TED spread is a sign that lenders believe that the risk of default on interbank loans is increasing. In that case, lenders demand a higher rate of interest, or accept lower returns on risk-free Treasuries. The 3-month LIBOR-OIS (overnight index swap) spread is a similar indicator.

[11] http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

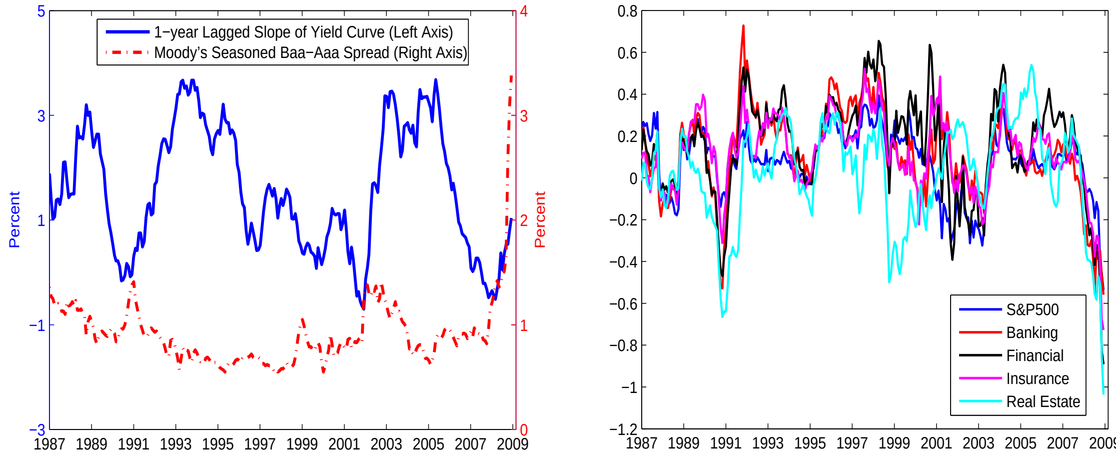

Figure 2: Time-series of explanatory covariates. *Left panel*: The 1-year lagged slope of yield curve and the default spread, given by the difference between Moody's seasoned Baa-rated and Aaa-rated corporate bond yields. *Right panel*: The trailing 1-year returns on the S&P500 index and the banking and FIRE portfolios.

(6) The default ratio $(N_t - N_{t-h})/(N^*_t - N^*_{t-h} + 1)$, which for fixed $h > 0$ relates the number of failures in the financial system during $(t-h, t]$ to one plus the number of economy-wide defaults during that period. It increases at a failure in the financial system, and decreases at a default of a non-financial firm.

We have also considered, and rejected for lack of significance in the presence of the above variables, a number of additional covariates, including the 3-month, 1-year, 10-year, 30-year Treasury rates, the spread between Moody's Baa rate and the 10 year treasury rate, the monthly VIX, and the 3-month LIBOR rate.

4.3 Economy-wide intensity

We start by addressing the likelihood problem (3) for the economy-wide intensity (1), taking the covariate vector X^* to include a constant, the trailing return on the S&P 500, the lagged slope of the yield curve, and the default spread. We have also considered, but rejected for lack of significance in the presence of these variables, the other covariates discussed in Section 4.2. The other covariates are used for the estimation of the process Z in Section 4.5 below.

Table 1 reports the parameter estimates, along with estimates of asymptotic standard errors.[12] The intensity is increasing in the default spread, and decreasing in the trailing

[12]The parameter space $\Theta = (-5, 5)^4 \times (0, 15) \times (0, 5)^2$. The fmincon routine of Matlab was used to search for the optimal parameter set. We performed a search for each of 10 randomly chosen initial parameter sets. Each of these searches converged to the values reported in Table 1.

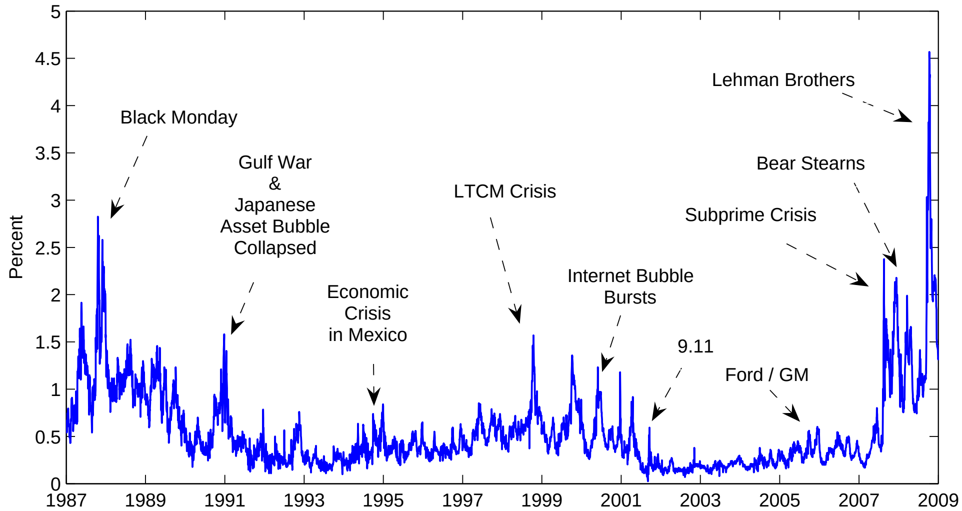

Figure 3: TED spreads during the sample period, along with significant events.

return on the S&P 500 and the lagged slope of the yield curve. The jump of the intensity at a default, measured in events per year, is estimated to be 2.3 plus roughly one half of the logarithm of the default volume, measured in million dollars. The impact of an event fades away exponentially with time: the fitted half life is $\log(2)/6.0592 = 0.1144$ years.

To develop some insight into the relative statistical importance of the baseline and spillover hazard terms for model fit, we take a Bayesian perspective, following Duffie et al. (2009), Eraker, Johannes & Polson (2003) and others. Specifically, we consider the Bayes factor, given by the ratio of the likelihood of a benchmark model to the likelihood of an alternative model, both evaluated at their respective estimators. The test statistic Ψ is given by twice the natural logarithm of the Bayes factor. According to Kass & Raftery (1995), a value for Ψ between 2 and 6 provides positive evidence, a value between 6 and 10 strong evidence, and a value larger than 10 provides very strong evidence in favor of the benchmark model. Due to the marginal nature of the likelihoods used for computing Ψ, this criterion does not necessarily favor more complex models.

We first test our model against an alternative specification that does not include a spillover hazard term (i.e., a traditional proportional hazards model). When the covariate set of the alternative model includes a constant, the trailing return on the S&P 500, the lagged slope of the yield curve, and the default spread, then the outcome of Ψ is 213.4, providing extremely strong evidence in favor of including the spillover hazard term. When the alternative model is based on an unconstrained covariate set that includes, in addition to the variables just mentioned, the TED spread, the trailing 1-year returns of banking, financial, insurance and real-estate portfolios,[13] then the outcome of Ψ is 131.4,

[13]The parameter estimates are as follows (SE in parentheses): Constant 4.1597 (0.1443), S&P 500

<table>
<tr><th rowspan="2"></th><th colspan="4">Baseline Hazard</th><th colspan="3">Spillover Hazard</th></tr>
<tr><th>Constant</th><th>S&P500</th><th>Yield Slope</th><th>Baa-Aaa</th><th>κ</th><th>γ</th><th>δ</th></tr>
<tr><td>MLE</td><td>2.3026</td><td>-0.4410</td><td>-0.2140</td><td>0.5092</td><td>6.0592</td><td>2.3205</td><td>0.4781</td></tr>
<tr><td>SE</td><td>0.0605</td><td>0.0524</td><td>0.0336</td><td>0.0534</td><td>0.1108</td><td>0.0811</td><td>0.0233</td></tr>
<tr><td>t-stat</td><td>38.04</td><td>-8.42</td><td>-6.37</td><td>9.53</td><td>54.71</td><td>28.60</td><td>20.56</td></tr>
<tr><td rowspan="2">Ψ</td><td></td><td>0.1298</td><td>3.0987</td><td>1.8310</td><td colspan="3" rowspan="2">213.4039</td></tr>
<tr><td colspan="4">26.5308</td></tr>
</table>

Table 1: Maximum likelihood estimates (MLE) of economy-wide intensity parameters, asymptotic standard errors (SE), t-statistics (t-stat), and Bayes factor statistics (Ψ).

still providing strong evidence in favor of including the spillover hazard term. Testing our model against one that does not include the baseline hazard term, the outcome of Ψ is 26.5, providing very strong evidence in favor of including the baseline hazard term. The test results suggest that the default clustering in the data cannot be explained by variations in the observable explanatory variables alone.

The left panel of Figure 4 shows the fitted economy-wide intensity against the number of economy-wide defaults. The fitted intensity tracks the observed arrivals well. The right panel of Figure 4 graphs the decomposition of the fitted intensity into baseline and spillover hazards. The time series behavior of the components is similar. However, during clustering periods, the spillover hazard represents a relatively larger fraction of the total default hazard than the baseline hazard.

4.4 Goodness-of-fit tests

We test the fit of the economy-wide intensity model λ^* to the historical default timing data. The tests are based on a result of Meyer (1971), which implies that the default arrivals follow a standard Poisson process under a change of time given by the cumulative intensity λ^*. Thus, if λ^* is correctly specified, then the time-scaled inter-arrival times are independent standard exponential variables.

The properties of the time-scaled arrival times can be analyzed with a battery of alternative tests. We use a family of tests of the binned arrival time data, following Das et al. (2007) and Lando & Nielsen (2009). For given bin size c, we denote by U_n the number of observed events in the n-th successive time interval lasting for c units of transformed time. With a total of K bins, the null hypothesis is that the $U_1, \ldots, U_K$ are independent Poisson variables with mean c. We consider bin sizes $c = 2, 4, 6, 8$ and 10.

We start with Fisher's dispersion test. Under the null, $W = \sum_{n=1}^{K} (U_n - c)^2/c$ has a chi-squared distribution with $K - 1$ degrees of freedom. Table 2a indicates that there is

-2.1542 (0.3036), Yield Slope -0.2346 (0.0272), Baa-Aaa 0.4612 (0.1370), TED -0.8716 (0.3040), Banking 0.5059 (0.2606), Financial 1.5882 (0.3551), Insurance -0.7845 (0.1548), Real Estate -0.6032 (0.1054). The default ratio was found to be insignificant in the presence of these covariates.

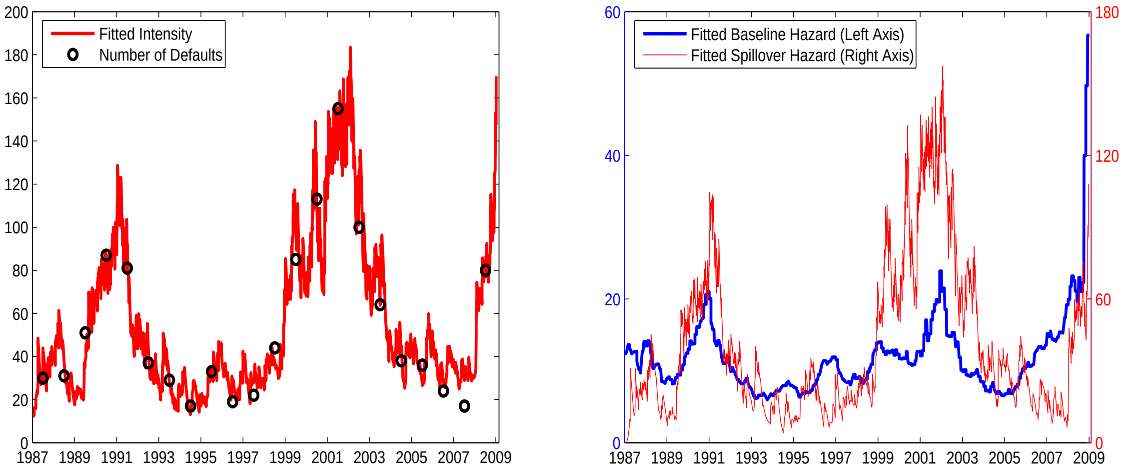

Figure 4: Fitted economy-wide intensity λ^*. *Left panel*: Yearly defaults and fitted intensity. *Right panel*: Intensity decomposition: fitted baseline hazard vs. fitted spillover hazard.

no evidence against the null for bin sizes 4 through 10, at standard confidence levels.

To examine the extent to which our intensity model captures the clustering of defaults, we perform an upper tail test developed by Das et al. (2007). We generate 10,000 data sets by Monte Carlo simulation, each consisting of K iid Poisson random variables with mean c. The p-value of the test is the fraction of the simulated data sets whose sample upper-quantile mean (or median) is above the actual sample mean (or median). The p-values reported in Table 2b suggest that there is no significant deviation of the upper-quartile tails from the theoretical Poisson tails for bin sizes 4 through 10, at standard confidence levels. Furthermore, the null hypothesis cannot be rejected by the joint test across all bin sizes, at conventional confidence levels.

Finally we test for serial dependence of the U_k. To this end, we estimate an autoregressive model, given by $U_k = A + BU_{k-1} + \epsilon_k$ for coefficients A and B. Under the null, $A = c$, $B = 0$, and the ϵ_k are independent, demeaned Poisson random variables. Table 2c shows that the fitted coefficients are not significantly different from their theoretical values for bin sizes 4 through 10, at standard confidence levels.

The results of these tests suggest that the fitted λ^* time-scales most arrival times correctly, indicating a good overall fit of our default timing model (1). Additional experiments suggest that the rejections of the null for bin size 2 are due to events arriving in very short time intervals. On the time scale of the sample period, which stretches over 21 years, these are almost simultaneous arrivals. It appears difficult to match, at the same time, the few extremely short inter-arrival times, and the many longer inter-arrival times that constitute the vast majority of the sample.

Bin Size	Number of Bins	χ^2 Statistic	p-Value
2	596	838.50	0.0000
4	298	332.75	0.0751
6	198	207.17	0.2956
8	149	167.38	0.1316
10	119	125.70	0.2967

(a) Fisher's Dispersion Test

	Mean of Tails			Median of Tails		
Bin Size	Data	Simulation	p-Value	Data	Simulation	p-Value
2	3.9694	3.6740	0.0000	4.0000	3.0524	0.0000
4	6.1739	6.1575	0.3092	6.0000	5.9956	0.0476
6	8.5676	8.8643	0.6254	8.0000	8.5337	0.5419
8	11.6667	11.3794	0.2190	11.0000	10.9284	0.0916
10	14.0313	13.7454	0.2459	13.5000	13.2444	0.2899
All	-	-	0.2799	-	-	0.1942

(b) Mean and Median of Default Upper Quartile Tail Test

Bin Size	Number of Bins	A	(t_A)	B	(t_B)	R^2
2	596	2.3634*	(3.4556)	-0.1847*	(-4.5767)	0.0341
4	298	4.0348	(0.1321)	-0.0121	(-0.2074)	0.0001
6	198	6.1971	(0.4250)	-0.0372	(-0.5203)	0.0014
8	149	8.8132	(1.1613)	-0.1074	(-1.3032)	0.1115
10	119	10.3584	(0.3650)	-0.0378	(-0.4018)	0.0014

(c) Excess Default Autocorrelation Test (t-statistics for A are presented for the test $A = c$ and asterisks indicate significance at the 5% level.)

Table 2: Goodness-of-fit tests of the economy-wide intensity.

4.5 System-wide intensity

Next we address the likelihood problem (5) for the process Z in (4). The value Z_t represents the conditional probability at t that the next defaulter is a financial firm, given that there is a default in the economy in the next instant. We take the covariate vector X to include a constant, the 1-year lagged slope of the yield curve, the TED spread, the trailing 1-year returns of banking and real-estate portfolios, and the default ratio for $h = 1/12$.[14] We have also considered, but rejected for lack of significance in the presence of these variables, the other covariates discussed in Section 4.2.

Table 3 provides the estimates of the coefficient vector β, along with asymptotic standard errors and t-statistics. A likelihood ratio test indicates that the covariates are

[14]We experimented with different window sizes h, but found $h = 1/12$ to work best. This window size is consistent with the frequency of the observations of the other covariates.

Covariate	Coefficient	SE	t-statistic	p-value	Ψ
Constant	-2.0873	0.1484	-14.0659	0.0000	
Yield Slope	0.1256	0.0585	2.1469	0.0318	4.6502
TED Spread	0.3710	0.1506	2.4632	0.0138	5.8223
Banking	0.8952	0.3462	2.5856	0.0097	6.6832
Real Estate	-0.8073	0.2973	-2.7218	0.0065	7.4439
Default Ratio	1.4171	0.4351	3.2572	0.0011	10.1015
Model Fit	LR-ratio (χ^2) = 36.8117			p-value < 0.0001	

Table 3: Maximum likelihood estimates of the coefficients β of the covariate process X governing the thinning process Z in (4), asymptotic standard errors (SE), t-statistics, p-values, and Bayes factor statistics (Ψ).

informative. The coefficient linking the trailing 1-year return of the banking portfolio to the probability Z_t is positive, and of unexpected sign by univariate reasoning. With multiple covariates, however, the sign need not be evidence that a good year in the banking sector foreshadows a higher fraction of bank defaults.

The time-series behavior of the fitted process Z, shown in the left panel of Figure 5, indicates the dramatic increase during the second half of 2008 of the number of defaults in the financial sector relative to the total number of events in the economy.

To measure how accurately the fitted model of Z distinguishes between economy- and system-wide events out-of-sample, we construct a power curve, shown in the right panel of Figure 5. The diagonal line represents an uninformative model that sorts events randomly. The larger the area under the curve (AUC), the more accurate the model predictions. For our model, the AUC is 0.7076, with 95% confidence interval given by $[0.6433, 0.7719]$. The standardized AUC is 6.3283, implying that the area is statistically greater than 0.5 with p-value less than 0.0001.

5 Systemic risk

This section analyzes the behavior of systemic risk during the sample period, provides risk forecasts for future periods, and evaluates these forecasts.

5.1 Risk measures

We start by examining the fitted system-wide intensity λ_t, which measures the level of instantaneous systemic risk prevailing at time t. It is calculated as the product of the economy-wide intensity λ_t^* and the thinning variable Z_t, as explained in Section 3. The time-series behavior of λ_t, shown in the left panel of Figure 6, indicates that the level of instantaneous systemic risk reached unprecedented levels during the fall of 2008. The right panel of Figure 6 shows the fitted fraction of λ_t tied to the spillover hazard term,

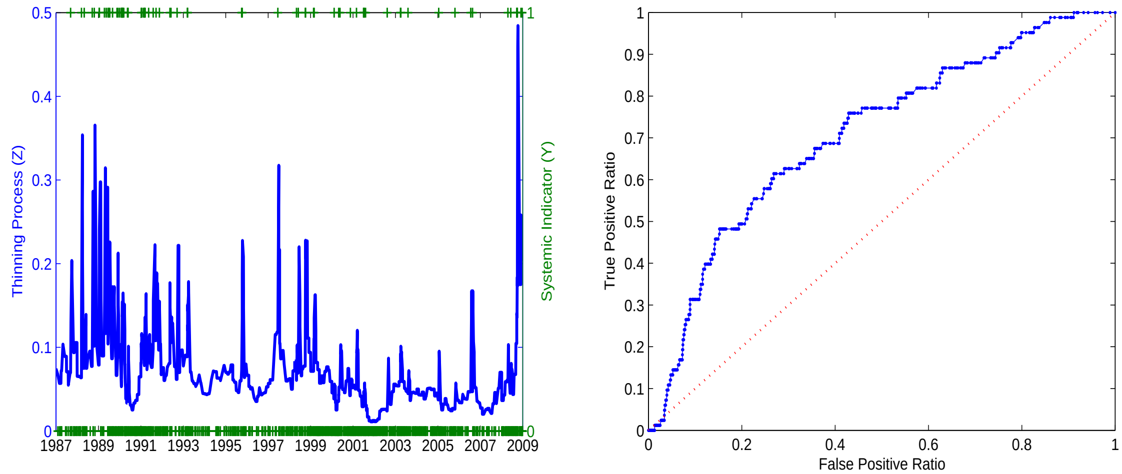

Figure 5: *Left panel*: Observed binary response variables Y_n and fitted process Z. *Right panel*: Power curve for the fitted process Z.

calculated as the fitted ratio of the spillover hazard to the total default intensity λ_t^*. The estimators provide strong evidence for the presence of failure clustering not caused by variations in the observable explanatory covariates. The fraction of systemic risk tied to the spillover term can be substantial, and tends to be higher in periods of adverse economic conditions. Moreover, financial firms tend to fail when the fitted contribution of spillovers to instantaneous systemic risk is relatively large.

Next we estimate the conditional distribution at time t of the default rate in the financial system during the period $(t, t+\Delta]$, for given Δ. As indicated in Section 3.3, this is done by exact Monte Carlo simulation.[15] The estimation is based on the models for λ^* and Z, fitted with data observed from 1/1/1987 to t. As explained above, the estimation takes account of the time-variation of the covariates during the forecast period, and the cross-sectional variation of the default volume.

Figure 7 shows the conditional distribution of the system-wide default rate $D_t(t+0.5)$ for conditioning times t varying semi-annually between 12/31/1997 and 12/31/2008, for a 6-month horizon. As explained in Section 2, the system-wide default rate is obtained by normalizing the number of system-wide defaults during $(t, t+0.5]$ by the number of firms in the financial system at t. The right tail of the distribution indicates the magnitude of systemic risk. The fatter the tail, the greater the likelihood that a large fraction of the financial system fails. The time series behavior suggests that systemic risk has increased very sharply during the second half of 2008.

We contrast the system-wide distribution with the distribution of the economy-wide default rate, shown in Figure 8. While the increase of aggregate default risk in the second half of 2008 is clearly visible, the magnitude of risk is only somewhat greater than that

[15]The estimates are based on 100,000 Monte Carlo replications.

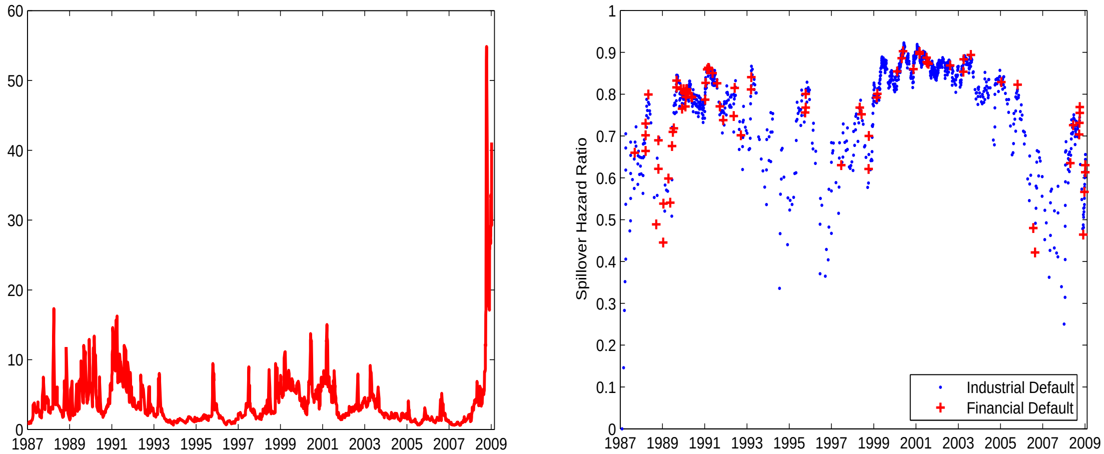

Figure 6: *Left panel*: Fitted system-wide failure intensity λ, based on the parameter estimates reported in Tables 1 and 3. *Right panel*: Fitted fraction of λ tied to the spillover hazard term, with default events indicated.

during the internet bubble. This is in stark contrast to the behavior of systemic risk shown in Figure 7: the systemic risk during the burst of the internet bubble is dwarfed by the systemic risk prevailing at the end of 2008.

Next we consider the value at risk $V_t(\alpha, t + \Delta)$ of the system-wide default rate. For a given level of confidence $\alpha \in (0, 1)$, the conditional probability at time t that the system-wide default rate $D_t(t + \Delta)$ exceeds $V_t(\alpha, t + \Delta)$ is $1 - \alpha$. For conditioning times t varying semi-annually between 12/31/1997 and 12/31/2008, the left panel of Figure 9 shows $V_t(\alpha, t + 0.5)$, along with realized default rates. The right panel of Figure 9 plots the value at risk for economy-wide defaults, defined similarly.

The value at risk defines a term structure of systemic risk. To illustrate this, the left panel of Figure 10 plots $V_t(\alpha, t + \Delta)$ on 12/31/2008, the end of the sample period, as a function of Δ, for each of several α.

There are alternative measures of systemic risk that may be of interest. An example is the conditional probability at t of no failures in the financial system during $(t, T]$. This measure does not require the choice of a confidence level. The right panel of Figure 10 shows this probability during the sample period, for each of several horizons Δ.

5.2 Forecast evaluation

We evaluate the out-of-sample forecast accuracy of the fitted value at risk $V_t(\alpha, t+\Delta)$ by comparing it to the realized default rate. Our selection of tests is informed by the results of the test performance analysis in Berkowitz, Christoffersen & Pelletier (2009).

Let n be the number of forecast periods. Further, let $n_1 \leq n$ be the number of periods for which the corresponding value at risk forecast was violated, i.e. the number

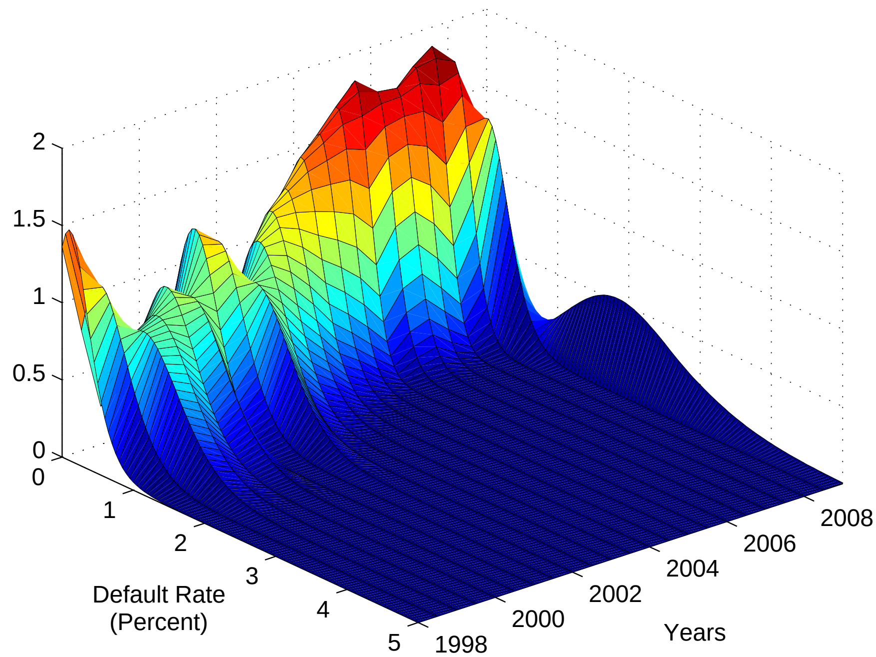

Figure 7: Fitted conditional distribution (kernel-smoothed) of the system-wide 6-month default rate $D_t(t+0.5)$ for conditioning times t varying semi-annually between 12/31/1997 and 12/31/2008.

of periods for which the realized default rate was strictly greater than the fitted value at risk $V_t(\alpha, t+\Delta)$. Then, $n_0 = n - n_1$ denotes the number of periods for which the realized rate was less than or equal to the fitted value at risk. We test whether the actual violation rate n_1/n is significantly different than the theoretical violation rate $(1-\alpha)$, as in Kupiec (1995). Fixing a level $\alpha \in (0,1)$ and assuming violations are independent of one another, the log-likelihood ratio test statistic

$$\mathrm{LR}_{\mathrm{UC}} = -2\log\left(\frac{\alpha^{n_0}(1-\alpha)^{n_1}}{(n_0/n)^{n_0}(n_1/n)^{n_1}}\right) \tag{6}$$

has, asymptotically, a chi-squared distribution with 1 degree of freedom under the null hypothesis of the theoretical $(1-\alpha)$ violation rate.[16]

A test based on the statistic (6) does not address the time-series properties of the sequence of "hit" indicators associated with violations in different periods. The hit indicator I_t for the forecast period $(t, t+\Delta]$ is equal to 1 if the realized default rate for the period is greater than the fitted value at risk $V_t(\alpha, t+\Delta)$, and 0 otherwise. A more stringent conditional coverage test with higher power tests whether the indicators are in-

[16]In case of $n_1 = 0$, we follow the convention $0^0 = 1$ so that the test statistic is well-defined.

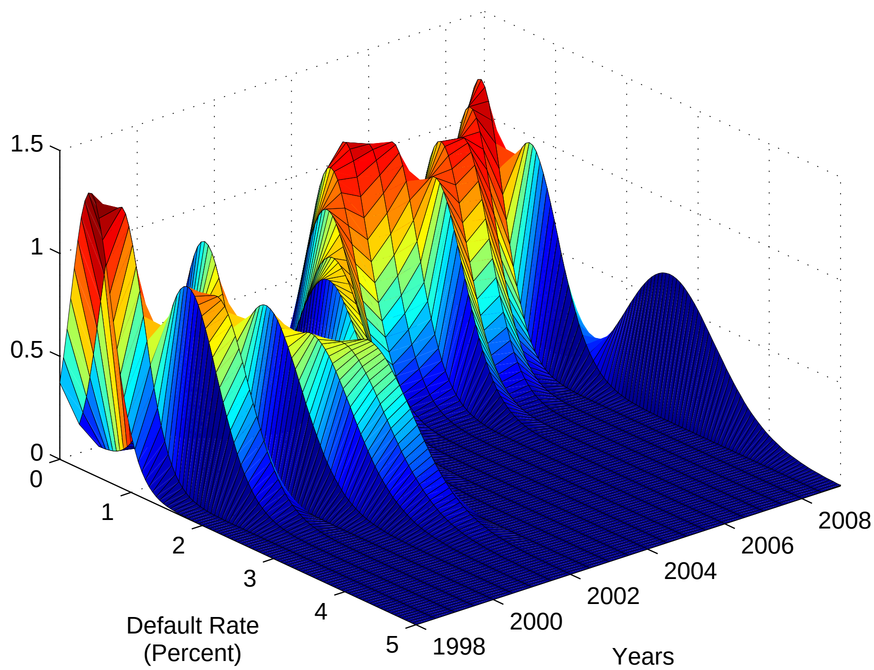

Figure 8: Fitted conditional distribution (kernel-smoothed) of the economy-wide 6-month default rate for conditioning times t varying semi-annually between 12/31/1997 and 12/31/2008. The default rate is obtained by normalizing the number of economy-wide defaults during $(t, t+0.5]$ by the total number of firms in the economy at t.

dependent and identically distributed Bernoulli variables with success probability $(1-\alpha)$. We consider two alternative tests of this property, a Markov test due to Christoffersen (1998) and the CAViaR test of Engle & Manganelli (2004). According to the performance analysis in Berkowitz et al. (2009), the CAViaR test has particularly high power for the relatively small sample sizes we encounter here, for both the 99% and 95% levels.

The Markov test of Christoffersen (1998) tests the Bernoulli distribution of the actual hit indicators and their independence. The test of the Bernoulli property relies on the statistic (6). The independence is tested against an explicit first-order Markov alternative, with log-likelihood ratio test statistic given by

$$\mathrm{LR}_{\mathrm{Ind}} = -2\log\left(\frac{(1-\pi_1)^{n_{00}+n_{10}}\pi_1^{n_{01}+n_{11}}}{(1-\pi_{01})^{n_{00}}\pi_{01}^{n_{01}}(1-\pi_{11})^{n_{10}}\pi_{11}^{n_{11}}}\right). \tag{7}$$

Here, n_{ij} denotes the number of periods with a state of j following a state of i, $\pi_{ij} = n_{ij}/(n_{i0}+n_{i1})$, and $\pi_1 = (n_{01}+n_{11})/(n_{00}+n_{01}+n_{01}+n_{11})$.[17] Under the null of the indicators forming a first-order Markov chain, this statistic has a limiting chi-squared distribution

[17] In case of $n_{10}+n_{11}=0$, we suppose $\pi_{11}=0$ so that the test statistic is well-defined.

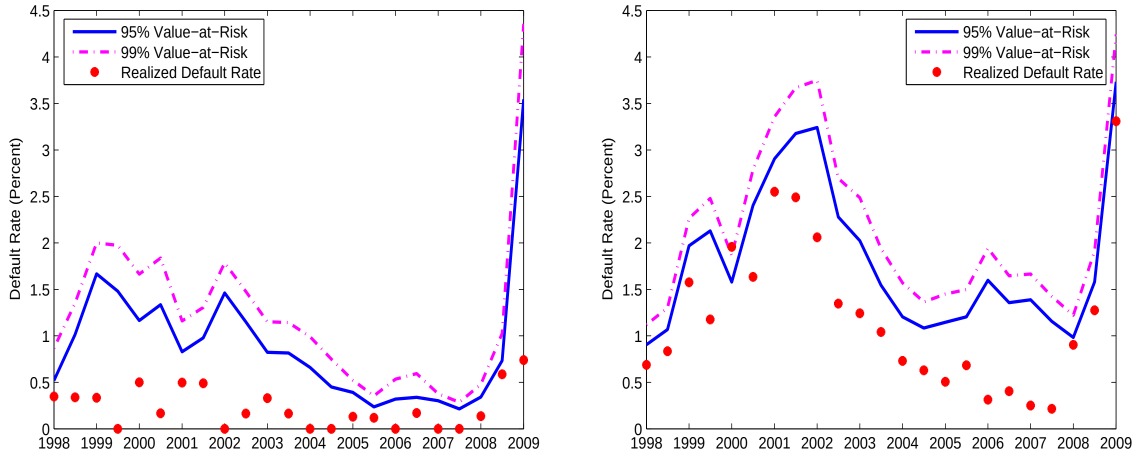

Figure 9: *Left Panel:* Fitted value at risk $V_t(\alpha, t+0.5)$ of the system-wide default rate, for conditioning times t varying semi-annually between 12/31/1997 and 12/31/2008, versus realized default rate. *Right Panel:* Fitted value at risk of the economy-wide default rate versus realized default rate.

with 1 degree of freedom. The combined test of the coverage ratio and independence is based on the statistic

$$\mathrm{LR}_{\mathrm{M}} = \mathrm{LR}_{\mathrm{UC}} + \mathrm{LR}_{\mathrm{Ind}},$$

which has a limiting chi-squared distribution with 2 degrees of freedom.[18]

The CAViaR test described in Berkowitz et al. (2009), which is based on Engle & Manganelli (2004), considers a first-order autoregression for the hit indicator:

$$I_t = \gamma + \beta_1 I_{t-\Delta} + \beta_2 V_t(\alpha, t+\Delta) + \epsilon_t \tag{8}$$

where the error term ϵ_t has a logistic distribution. We test whether the β_i coefficients are statistically significant and whether $P(I_t = 1) = e^{\gamma}/(1+e^{\gamma}) = 1-\alpha$. Denote the ith response variable by Y_i and the corresponding vector of regressors by X_i, for $i = 1, \ldots, n-1$. Also, let $\pi_i = e^{\widehat{\gamma}+\widehat{\beta}X_i}/(1+e^{\widehat{\gamma}+\widehat{\beta}X_i})$, where $(\widehat{\gamma}, \widehat{\beta})$ is the maximum likelihood estimator of $(\gamma, (\beta_1, \beta_2))$ obtained by logistic regression. Then, under the null of $\beta_1 = \beta_2 = 0$ and $\gamma = \log\left(\frac{1-\alpha}{\alpha}\right)$, the log-likelihood ratio test statistic

$$\mathrm{LR}_{\mathrm{CAViaR}} = -2\log\left(\prod_{i=1}^{n-1} \frac{(1-\alpha)^{Y_i}\alpha^{1-Y_i}}{\pi_i^{Y_i}(1-\pi_i)^{1-Y_i}}\right) \tag{9}$$

has a limiting chi-squared distribution with 3 degrees of freedom.

Table 4 reports the test results for the system-wide value at risk $V_t(\alpha, t+\Delta)$, for each of several forecast horizons Δ and confidence levels α.[19] None of the null hypotheses can

[18]This ignores the first observation in the hit sequence.

[19]We also use 2009 default data in the tests: we validate the forecasts obtained on 12/31/2008 on the realized default rates in 2009, which are available for the first 1, 3, and 6 months of 2009.

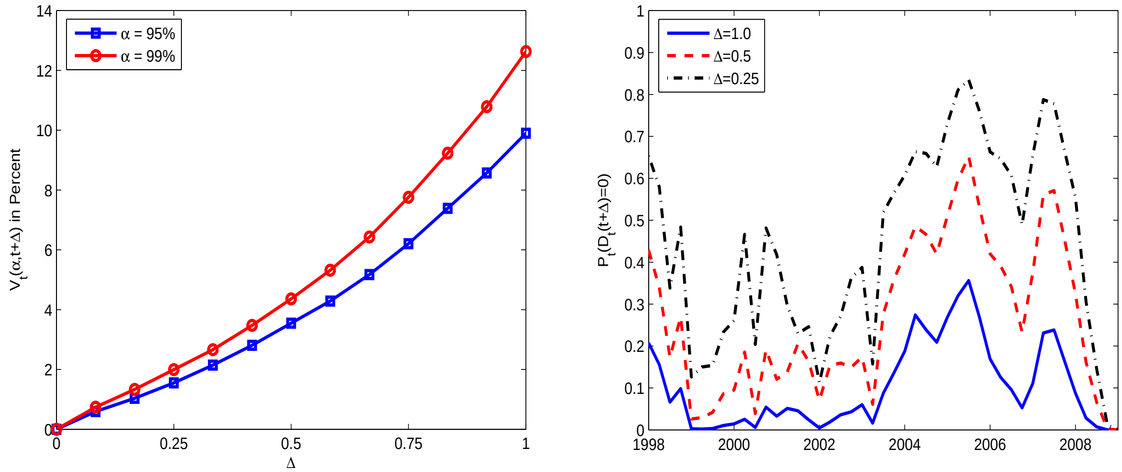

Figure 10: *Left Panel:* Term structure of systemic risk on 12/31/2008: fitted value at risk $V_t(\alpha, t+\Delta)$ on 12/31/2008 as a function of Δ. *Right Panel:* Fitted conditional probability at t of no failures in the financial system during $(t, t+\Delta]$, for conditioning times t varying quarterly between 12/31/1997 and 12/31/2008, for each of several horizons Δ.

be rejected at the 10% level. This suggests that the fitted measures accurately quantify systemic risk, for each of several risk horizons and confidence levels, and this validates our default hazard model (1)–(2) and our two-stage inference procedure. We conclude that the risk measures developed in this paper are useful for monitoring the level of systemic risk in the U.S. financial system by regulators and other supervisory authorities.

6 Sensitivity of systemic risk

We show how to measure the impact of a hypothetical default event on systemic risk. This analysis could be useful to regulatory authorities. For example, regulators could estimate the potential impact on systemic risk of a default of a given financial institution.

Fix a conditioning time t, horizon Δ and confidence level α. We consider the change $\Delta V_t(\alpha, t+\Delta)$ of the value at risk $V_t(\alpha, t+\Delta)$ at t in response to a default at t, which measures the event's impact on systemic risk. To estimate the change, we first estimate the time t value at risk $V_t(\alpha, t+\Delta)$ based on data up to t. Next we enlarge the data set by including a hypothetical default event at t, and then re-estimate $V_t(\alpha, t+\Delta)$ based on the enlarged data set. Finally we calculate $\Delta V_t(\alpha, t+\Delta)$ as the difference between the two risk measure estimates.

The change $\Delta V_t(\alpha, t+\Delta)$ reflects the influence of the hypothetical event on the other firms in the financial system and the economy at large, including potential spillover effects. It depends on the characteristics of the hypothetical event, including the sector of the defaulter (industrial vs. financial) and the total debt outstanding at default, which

			Uncond. Coverage		Markov		CAViaR	
	Δ	Obs.	LR	p-value	LR	p-value	LR	p-value
95% VaR	1Y	11	0.3153	0.5744	0.5157	0.7727	2.3203	0.5086
	6M	23	2.3595	0.1245	2.3595	0.3074	2.2569	0.5208
	3M	45	0.9143	0.3390	0.9609	0.6185	5.5926	0.1332
	1M	133	2.6284	0.1050	2.6900	0.2605	4.5851	0.2048
99% VaR	1Y	11	0.2211	0.6382	0.2211	0.8953	0.2211	0.9741
	6M	23	0.4623	0.4965	0.4623	0.7936	0.4422	0.9314
	3M	45	0.9045	0.3416	0.9045	0.6362	0.8844	0.8292
	1M	133	0.0905	0.7636	0.1057	0.9485	0.6689	0.8805

Table 4: Out-of-sample tests of the forecast accuracy of the fitted system-wide value at risk, for each of several horizons. The period considered is January 1998 to June 2009.

proxies the size of the defaulter.

We calculate the change for each of two hypothetical events, a default of a financial institution, and a default of an industrial firm. The left panel of Figure 11 shows the absolute change $\Delta V_t(0.95, t+1)$ for each of the two events, for conditioning times t varying quarterly between 12/31/1997 and 12/31/2008. The right panel of Figure 11 shows the impact of each of these events on term structure of the value at risk $V_t(0.95, t+\Delta)$ on 12/31/2008. The total debt outstanding at default is taken to be the sample mean of the default volumes observed to the conditioning time, for the respective firm class.

The failure of a financial institution has a higher impact on systemic risk than the default of an industrial firm. This means the financial system is more vulnerable to the collapse of a financial firm. The impact of a financial firm default is also more volatile during the sample period. If measured on an absolute scale, the impact of a default has increased dramatically during the second half of 2008, indicating the vulnerability of the financial system during that period.

The sensitivity analysis can be extended to measure the impact on systemic risk of a hypothetical adverse shock to the explanatory covariates (risk factors), along the lines of Avesani et al. (2006), Huang et al. (2009), and others.

7 Conclusion

This paper provides an econometric method for estimating the term structure of systemic risk over multiple future periods. The maximum likelihood estimators incorporate the dependence of failure timing on time-varying macro-economic and sector-specific risk factors. Unlike traditional estimators, they also capture the impact of spillover effects related to missing or unobserved risk factors, and the spread of distress in a network of firms.

Applying our method to data on U.S. firms over 1987 to 2008, we find that the level

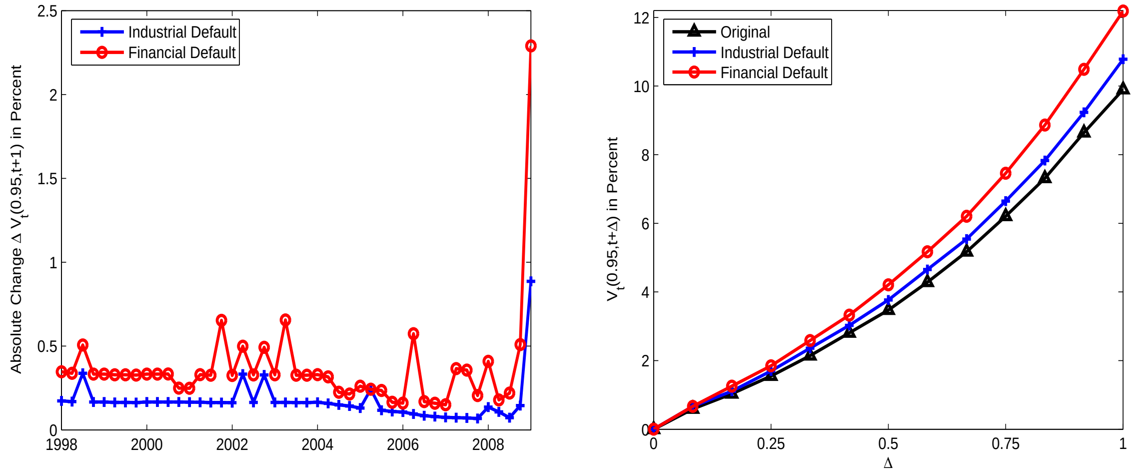

Figure 11: Impact of a default on systemic risk. *Left Panel:* Absolute change $\Delta V_t(0.95, t+1)$ for conditioning times t varying quarterly between 12/31/1997 and 12/31/2008. *Right Panel:* Term structure of value at risk $V_t(0.95, t+\Delta)$ on 12/31/2008, for different scenarios.

and shape of the term structure of systemic risk in the U.S. financial sector depend on the timing and severity of past financial and industrial failures, as well as the current values of observable risk factors including the trailing S&P 500 index return, the lagged slope of the U.S. yield curve, the default and TED spreads, and other sector-wide variables. We find that the variation of these risk factors generally has a less significant impact on systemic risk than spillover effects associated with failures in the past. This highlights the importance of addressing spillovers when managing systemic risk.

Several topics are left for future research, including the estimation of premia for systemic risk. The estimates provided in this paper can be compared to estimates of the risk-neutral probability of failure of a large number of financial institutions, obtained from market rates of credit derivatives contracts. This analysis would shed light on the magnitude of the premia investors demand for bearing systemic risk.

Our econometric method has potential applications in other subject areas requiring estimates of event probabilities in situations where network or information effects may play a role. These applications include the analysis of market transaction data, the analysis of purchase-timing behavior of households, the analysis of unemployment timing, and many others. The extant analyses of these applications in Engle & Russell (1998), Seetharaman & Chintagunta (2003), and Lancaster (1979), respectively, employ standard proportional hazard formulations. Our generalized hazard model could be used to study the implications of network and information effects in these settings.

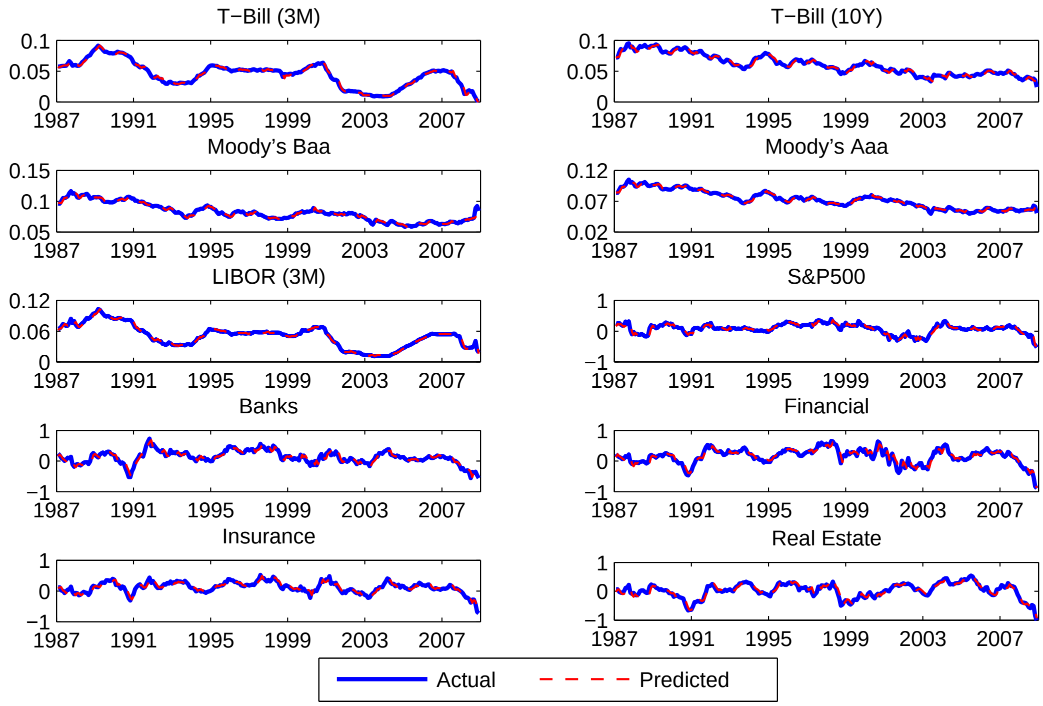

Figure 12: Realized vs. VAR(1) predicted time series (monthly) of covariate components.

A Covariate time-series model

We formulate a vector autoregressive VAR(1) time-series model for the covariates. This model incorporates the dynamic relationships between the different variables. Let Φ_t denote the $(n \times 1)$ vector of covariate values at t. We suppose that

$$\Phi_t = \Pi_0 + \Pi_1 \Phi_{t-1} + \varepsilon_t \tag{10}$$

where Π_0 is an $(n \times 1)$ vector, Π_1 is an $(n \times n)$ coefficient matrix and ε_t is a $(n \times 1)$ zero mean vector of error processes that is serially uncorrelated, and has time-invariant covariance matrix Σ. Table 5 reports the estimators $\widehat{\Pi}_i$ of Π_i for $i = 0, 1$, which are based on monthly observations of Φ_t during the sample period. Given the Φ_t and the $\widehat{\Pi}_i$, we recover the corresponding values of ε_t. From these values, we estimate the covariance matrix Σ, assuming weak stationarity. The fitted Σ is reported in Table 6. An analysis of the error series indicates the appropriateness of the model (10) for our covariates. Figure 12 visualizes the goodness-of-fit by plotting the predicted vs. the realized covariates.

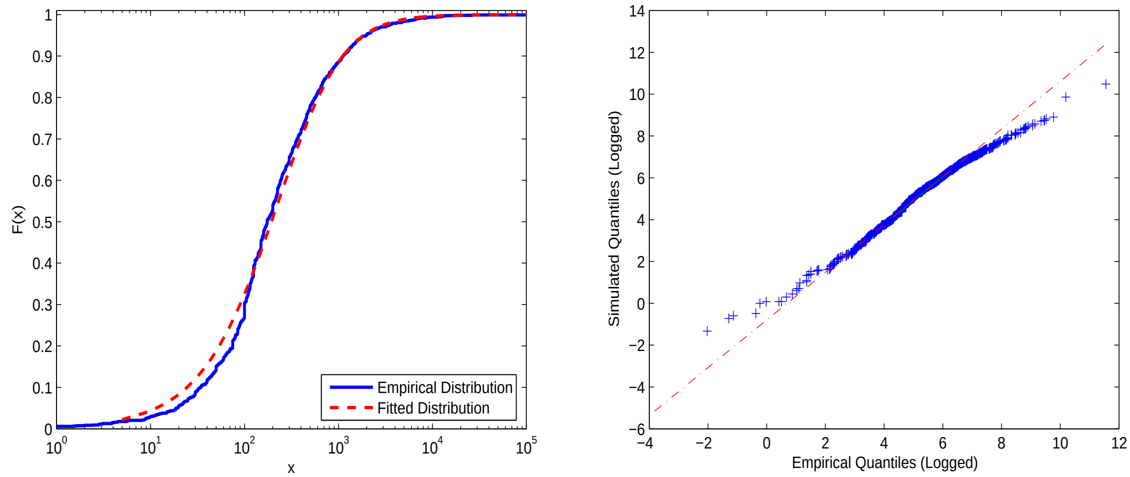

Figure 13: *Left panel*: Empirical default volume distribution vs. fitted generalized Pareto distribution as of 12/31/2008. *Right panel*: Empirical quantiles of the observed default volumes vs. quantiles of realizations of variables from the fitted Pareto distribution.

B Default volume model

We adopt a simple but empirically meaningful model of default volumes. We assume that each D_n^* has a generalized Pareto distribution with shape parameter $\xi > 0$ and scale parameter $\sigma > 0$. We have

$$P(D_n^* > x) = \left(1 + \xi \frac{x}{\sigma}\right)^{-\frac{1}{\xi}} \tag{11}$$

for all $x \geq 0$. The maximum likelihood estimators of (ξ, σ) are given by $(0.5960, 225.8828)$, with standard errors $(0.0427, 10.9864)$ as of 12/31/2008. The left panel of Figure 13 contrasts the fitted Pareto distribution with the empirical distribution of default volumes. The right panel of Figure 13 compares the observed default volumes to the realizations of variables from the fitted Pareto distribution. The plots indicate the statistical appropriateness of our model.

References

Acharya, Viral, Lasse Pedersen, Thomas Philippon & Matthew Richardson (2009), Measuring systemic risk. Working Paper, New York University.

Acharya, Viral & Tanyu Yorulmazer (2008), 'Information contagion and bank herding', *Journal of Money, Credit and Banking* **40**, 215–231.

Adrian, Tobias & Markus Brunnermeier (2009), CoVaR. Working Paper, Princeton University.

Aharony, Joseph & Itzhak Swary (1983), 'Contagion effects of bank failures: Evidence from capital markets', *Journal of Business* **56**(3), 305–317.

Aharony, Joseph & Itzhak Swary (1996), 'Additional evidence on the information-based contagion effects of bank failures', *Journal of Banking and Finance* **20**, 57–69.

Avesani, Renzo, Antonio Garcia Pascual & Jing Li (2006), A new risk indicator and stress testing tool: A multifactor nth-to-default cds basket. IMF Working Paper No. 105.

Azizpour, Shahriar & Kay Giesecke (2008), Self-exciting corporate defaults: Contagion vs. frailty. Working Paper, Stanford University.

Berkowitz, Jeremy, Peter Christoffersen & Denis Pelletier (2009), Evaluating value-at-risk models with desk-level data. *Management Science*, forthcoming.

Bhansali, Vineer, Robert Gingrich & Francis Longstaff (2008), 'Systemic credit risk: What is the market telling us', *Financial Analysts Journal* **64**(July/August), 16–24.

Brown, Craig & Serdar Dinc (2005), 'The politics of bank failures: Evidence from emerging markets', *Quarterly Journal of Economics* **120**(4), 1413–1444.

Brown, Craig & Serdar Dinc (2009), Too many to fail? Evidence of regulatory reluctance in bank failures when the banking sector is weak. *Review of Financial Studies*, forthcoming.

Chan-Lau, Jorge & Toni Gravelle (2005), The END: A new indicator of financial and nonfinancial corporate sector vulnerability. IMF Working Paper No. 231.

Chen, Long, Pierre Collin-Dufresne & Robert S. Goldstein (2008), On the relation between credit spread puzzles and the equity premium puzzle. *Review of Financial Studies*, forthcoming.

Christoffersen, Peter (1998), 'Evaluating interval forecasts', *International Economic Review* **39**, 842–862.

Cole, Rebel & Qiongbing Wu (2009), Predicting bank failures using a simple dynamic hazard model. Working Paper, DePaul University.

Collin-Dufresne, Pierre, Robert Goldstein & Jean Helwege (2009), How large can jump-to-default risk premia be? Modeling contagion via the updating of beliefs. Working Paper, Columbia University.

Cooperman, Elisabeth, Winson Lee & Glenn Wolfe (1992), 'The 1985 Ohio thrift crisis, the FSLIC's solvency and rate contagion for retail CDs', *Journal of Finance* **47**(3), 919–941.

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

Duffie, Darrell, Andreas Eckner, Guillaume Horel & Leandro Saita (2009), 'Frailty correlated default', *Journal of Finance* **64**, 2089–2123.

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

Eisenberg, Larry & Thomas Noe (2001), 'Systemic risk in financial systems', *Management Science* **47**(2), 236–249.

Elsinger, Helmut, Alfred Lehar & Martin Summer (2006), 'Risk assessment for banking systems', *Management Science* **52**(9), 1301–1314.

Engle, Robert F. & Jeffrey R. Russell (1998), 'Autoregressive conditional duration: A new model for irregularly spaced transaction data', *Econometrica* **66**, 1127–1162.

Engle, Robert & Simone Manganelli (2004), 'Caviar: Conditional autoregressive value at risk by regression quantiles', *Journal of Business and Economic Statictics* **22**(4), 367–381.

Eraker, Bjorn, Michael Johannes & Nicholas Polson (2003), 'The impact of jumps in volatility and return', *Journal of Finance* **58**, 1269–1300.

Estrella, Arturo & Mary R. Trubin (2006), 'The yield curve as a leading indicator: Some practical issues', *Current Issues in Economics and Finance* **12**(5). Federal Reserve Bank of New York.

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, Lisa Goldberg & Xiaowei Ding (2009), A top-down approach to multi-name credit. *Operations Research*, forthcoming.

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

Huang, Xin, Hao Zhou & Haibin Zhu (2009), 'A framework for assessing the systemic risk of major financial institutions', *Journal of Banking and Finance* **33**(11), 2036–2049.

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

Kass, Robert & Adrian Raftery (1995), 'Bayes factors', *Journal of the American Statistical Association* **90**, 773–795.

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.

Kupiec, Paul H. (1995), 'Techniques for verifying the accuracy of risk measurement models', *Journal of Derivatives* **3**(2), 73–84.

Lancaster, Tony (1979), 'Econometric methods for the duration of unemployment', *Econometrica* **47**(4), 939–956.

Lando, David & Mads Stenbo Nielsen (2009), Correlation in corporate defaults: Contagion or conditional independence? Working Paper, Copenhagen Business School.

Lane, William, Stephen Looney & James Wansley (1986), 'An application of the Cox proportional hazards model to bank failure', *Journal of Banking and Finance* **10**, 511–531.

Lehar, Alfred (2005), 'Measuring systemic risk: A risk management approach', *Journal of Banking and Finance* **29**(10), 2577–2603.

McCullagh, Peter & John Nelder (1989), *Generalized Linear Models*, Chapman and Hall, London.

McDonald, Cynthia & Linda Van de Gucht (1999), 'High-yield bond default and call risks', *Review of Economics and Statistics* **81**(3), 409–419.

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.

Schwarcz, Steven L. (2008), 'Systemic risk', *Georgetown Law Journal* **97**(1), 193.

Seetharaman, P. B. & Pradeep K. Chintagunta (2003), 'The proportional hazard model for purchase timing: A comparison of alternative specifications', *Journal of Business and Economic Statistics* **21**(3), 368–382.

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

Staum, Jeremy (2009), Systemic risk components as deposit insurance premia. Working Paper, Northwestern University.

Upper, Christian & Andreas Worms (2004), 'Estimating bilateral exposures in the German interbank market: Is there a danger of contagion?', *European Economic Review* **48**, 827–849.

Whalen, Gary (1991), 'A proportional hazards model of bank failure: an examination of its usefulness as an early warning model tool', *Federal Reserve Bank of Cleveland Economic Review* pp. 21–31.

Wheelock, David & Paul Wilson (2000), 'Why do banks disappear? The determinants of US bank failures and acquisitions', *Review of Economics and Statistics* **82**, 127–138.

	Constant	TB(3M)	TB(10Y)	Baa	Aaa	LIBOR	S&P500	Banks	Fin	Insur	RlEst
TB(3M)	0.172	1.006	0.133	-0.128	0.031	-0.036	0.194	-0.238	0.244	-0.499	0.200
	(1.279)	(21.395)	(2.762)	(-1.776)	(0.351)	(-0.844)	(1.170)	(-1.551)	(1.907)	(-2.872)	(2.689)
TB(10Y)	-0.179	0.133	0.925	0.047	0.037	-0.127	0.622	-0.202	-0.053	-0.128	0.152
	(-1.127)	(2.393)	(16.299)	(0.559)	(0.356)	(-2.564)	(3.177)	(-1.120)	(-0.352)	(-0.623)	(1.733)
Baa	-0.002	-0.018	-0.051	0.999	0.042	0.019	0.435	-0.249	-0.001	-0.043	0.064
	(-0.018)	(-0.379)	(-1.040)	(13.576)	(0.461)	(0.445)	(2.560)	(-1.591)	(-0.006)	(-0.242)	(0.837)
Aaa	-0.088	0.080	-0.025	0.048	0.980	-0.078	0.468	-0.270	0.060	-0.032	0.079
	(-0.704)	(1.837)	(-0.562)	(0.728)	(11.958)	(-1.998)	(3.050)	(-1.904)	(0.502)	(-0.202)	(1.154)
LIBOR	0.148	0.227	0.101	-0.073	0.012	0.752	0.659	-0.564	0.255	-0.614	0.255
	(0.961)	(4.224)	(1.835)	(-0.885)	(0.121)	(15.592)	(3.465)	(-3.212)	(1.737)	(-3.089)	(2.993)
S&P500	0.126	0.020	0.023	-0.047	0.017	-0.016	0.898	0.101	-0.056	-0.080	-0.032
	(3.204)	(1.424)	(1.648)	(-2.256)	(0.662)	(-1.291)	(18.484)	(2.246)	(-1.499)	(-1.573)	(-1.477)
Banks	0.057	0.031	0.010	-0.036	0.026	-0.029	-0.033	0.954	0.016	-0.053	-0.046
	(1.098)	(1.703)	(0.552)	(-1.289)	(0.777)	(-1.812)	(-0.514)	(16.264)	(0.332)	(-0.794)	(-1.607)
Fin	0.128	0.048	0.001	-0.069	0.062	-0.044	0.147	0.111	0.795	-0.121	-0.037
	(2.096)	(2.261)	(0.068)	(-2.119)	(1.548)	(-2.313)	(1.952)	(1.592)	(13.656)	(-1.535)	(-1.106)
Insr	0.017	0.046	-0.008	-0.033	0.042	-0.039	0.040	0.016	0.034	0.801	0.003
	(0.379)	(2.987)	(-0.501)	(-1.422)	(1.446)	(-2.825)	(0.745)	(0.331)	(0.814)	(14.219)	(0.121)
RlEst	0.030	0.059	-0.008	-0.033	0.043	-0.058	0.055	0.092	-0.041	-0.067	0.921
	(0.596)	(3.410)	(-0.459)	(-1.228)	(1.309)	(-3.753)	(0.899)	(1.624)	(-0.870)	(-1.052)	(33.539)

Table 5: Fitted coefficients of VAR(1) model (10) as of 12/31/2008. The t-statistics are shown in parenthesis.

	TB(3M)	TB(10Y)	Baa	Aaa	LIBOR	S&P500	Banks	Fin	Insur	RlEst
TB(3M)	0.0424	0.0220	0.0048	0.0083	0.0257	0.0020	0.0006	0.0036	0.0005	0.0021
TB(10Y)	0.0220	0.0589	0.0396	0.0405	0.0248	0.0010	-0.0009	0.0016	-0.0016	0.0019
Baa	0.0048	0.0396	0.0443	0.0367	0.0214	-0.0021	-0.0023	-0.0030	-0.0031	0.0002
Aaa	0.0083	0.0405	0.0367	0.0362	0.0185	-0.0009	-0.0019	-0.0014	-0.0023	0.0007
LIBOR	0.0257	0.0248	0.0214	0.0185	0.0555	-0.0006	-0.0008	0.0003	-0.0014	0.0023
S&P500	0.0020	0.0010	-0.0021	-0.0009	-0.0006	0.0036	0.0030	0.0038	0.0024	0.0019
Banks	0.0006	-0.0009	-0.0023	-0.0019	-0.0008	0.0030	0.0062	0.0053	0.0043	0.0031
Fin	0.0036	0.0016	-0.0030	-0.0014	0.0003	0.0038	0.0053	0.0087	0.0040	0.0039
Insur	0.0005	-0.0016	-0.0031	-0.0023	-0.0014	0.0024	0.0043	0.0040	0.0045	0.0024
RlEst	0.0021	0.0019	0.0002	0.0007	0.0023	0.0019	0.0031	0.0039	0.0024	0.0058

Table 6: Fitted covariance matrix Σ of the VAR(1) error term ε_t as of 12/31/2008.