MANAGEMENT SCIENCE
Vol. 57, No. 8, August 2011, pp. 1387–1405
ISSN 0025-1909 | EISSN 1526-5501 | 11 | 5708 | 1387

DOI 10.1287/mnsc.1110.1375
© 2011 INFORMS

Systemic Risk: What Defaults Are Telling Us

Kay Giesecke
Department of Management Science and Engineering, Stanford University, Stanford, California 94305, giesecke@stanford.edu

Baeho Kim
Korea University Business School, Seoul 136-701, Korea, baehokim@korea.ac.kr

This paper develops dynamic measures of the systemic risk of the financial sector as a whole. It defines systemic risk as the conditional probability of failure of a sufficiently large fraction of the total population of financial institutions. This definition recognizes that the cause of systemic distress is the correlated failure of institutions to meet obligations to creditors, customers, and trading partners. The likelihood estimators of the failure probability are based on a dynamic hazard model of correlated failure timing that captures the influence on failure timing of time-varying macroeconomic and sector-specific risk factors, and of spillover effects. Tests indicate that our measures provide accurate out-of-sample forecasts of the term structure of systemic risk in the United States for the period from 1998 to 2009.

Key words: banks; financial system; correlated failure; systemic risk
History: Received March 7, 2010; accepted April 8, 2011, by Wei Xiong, finance. Published online in *Articles in Advance* June 30, 2011.

1. Introduction

The financial crisis of 2007–2009 highlights the need to measure and manage systemic risk. The measurement of systemic risk involves two problems: the quantification of the systemic risk of the financial sector as a whole, and the allocation of that risk to individual institutions. Much of the recent research has focused on measuring the marginal systemic risk of individual institutions (see Acharya et al. 2010, Adrian and Brunnermeier 2010, Brownlees and Engle 2010, Huang et al. 2009, and others). In this paper, we propose measures of the systemic risk of the financial sector as a whole, and then develop, implement, and validate maximum likelihood estimators of these measures. Tests indicate that our measures provide accurate out-of-sample forecasts of systemic risk in the United States for the period from 1998 to 2009.

We propose to define the systemic risk of the financial sector as the conditional probability of failure of a sufficiently large fraction of the total population of financial institutions. This definition recognizes that the cause of systemic distress is the correlated failure of institutions to meet obligations to creditors, customers, and trading partners. It associates systemic risk with a failure cluster in the financial sector, which is often part of a larger cluster of economy-wide defaults. A failure event of this magnitude may be triggered by an initial economic shock, which is then propagated through a complex web of contractual and informational relationships in the economy. The collapse of Lehman Brothers is an example of how the distress of an institution can spill over to multiple other entities and eventually endanger the stability of the financial system.

To estimate the probability of failure clusters, we formulate a dynamic hazard model of correlated default timing in the economy. Our model is designed to capture the statistical implications of spillover effects. It extends the traditional proportional hazards specification used by Das et al. (2007), Duffie et al. (2007), McDonald and Van de Gucht (1999), and others to predict nonfinancial corporate default; and by Brown and Dinc (2005, 2011), Lane et al. (1986), Whalen (1991), Wheelock and Wilson (2000), and others to forecast bank failures. In our model, the timing of defaults is influenced by time-varying explanatory covariates and past defaults with a role for the size of a defaulter. While controlling for the influence of observable risk factors on default timing, our econometric specification addresses the implications of missing or incompletely observed ("frailty") factors, a source of information-based spillover effects emphasized by Collin-Dufresne et al. (2009), Duffie et al. (2009), Giesecke (2004), Koopman et al. (2008), and others for the real sector; and by Acharya and Yorulmazer (2008), Aharony and Swary (1983), Cooperman et al. (1992), and others for the financial sector. It also addresses the implications of spillover effects channeled through the complex and increasingly opaque network of derivatives trading relationships, interbank loans, trade credit chains, parent-subsidiary relationships, and other contractual links between firms. A traditional proportional hazard formulation ignores these effects.

A likelihood method is used to estimate our hazard model of correlated default timing in the economy. Then, a random thinning method is used to estimate the portion of the fitted economy-wide hazard that is tied to failures in the financial system. The system-wide failure hazard leads to estimators of the conditional distribution of the failure rate in the financial system and the associated measures of systemic risk. By treating the failures in the financial system in the context of the defaults in the broader economy, our inference method recognizes the financial system as an integral part of the economy. The method is designed to address the implications of industrial defaults for financial failures, and vice versa. It ensures that our systemic risk estimators incorporate the dynamic interaction between the real and financial sectors. The importance of this interaction for systemic risk was emphasized by Schwarcz (2008) and many others. Moreover, our inference method quantifies the contribution of systemic risk to economy-wide risk. It generates compatible estimators of both systemic and economy-wide risk.

We apply our model and inference method to data on default experience in the United States for the period from January 1987 to December 2008, and data on explanatory covariates including the trailing return of the S&P 500, the lagged slope of the U.S. yield curve, the default and TED spreads, and several other variables. Comprehensive tests demonstrate the in-sample fit of our estimated hazard models 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 failure rate in the U.S. financial system during 1998–2009 for each of several confidence levels and forecast horizons. These tests validate our hazard model of correlated failure timing and our inference method.

The proven predictive power of our risk measures renders them appropriate for a number of important applications in practice, which we illustrate through several use cases. First, by examining the time-series and term-structure behavior of our measures, regulators and policy makers can monitor the level of systemic risk in the U.S. financial sector. This analysis yields valuable insights into the dynamics of systemic risk. For example, we find that systemic risk in the United States increased dramatically during the second half of 2008 and reached unprecedented levels toward the end of 2008. Although the economy-wide default risk prevailing at the end of 2008 is comparable with that prevailing when the dot-com bubble burst in 2001, the systemic risk at that time is dwarfed by the systemic risk prevailing at the end of 2008. The ability to quantify the contribution of systemic risk to economy-wide default risk provides useful information regarding the relative significance of systemic risk at a given time. This information can guide the timing of interventions. Second, by examining the sensitivity of our measures, regulators can perform various stress analyses quantifying the impact on systemic risk of adverse events, such as the default of an institution or a shock to risk factors. This analysis characterizes the vulnerability of the financial system. It can guide policy decisions, for example whether a failing institution should be bailed out. For instance, we find that the U.S. system was extremely vulnerable during the second half of 2008, foreshadowing the disruptions induced by the collapse of Lehman Brothers in the fall of 2008.

The remainder of this paper is organized as follows. 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 goodness-of-fit tests. Section 5 analyzes the behavior of systemic risk during 1987–2008, provides risk forecasts, and evaluates these forecasts out-of-sample. It also contrasts the predictive performance of our risk measures with that of alternative measures that are not based on the correlated failure of institutions but adverse changes of equity prices of institutions. Section 6 analyzes the impact on systemic risk of a failure event and a shock to risk factors. Section 7 concludes. There are several appendices.

2. Failure-Based Measures of Systemic Risk

We propose to define the systemic risk of the financial sector as the conditional probability of failure of a sufficiently large fraction of the total population of financial institutions, including banks, thrifts, investment management, trading, leasing, mortgage and securities firms, insurance companies, and related firms. This definition recognizes the correlated failure of institutions to meet obligations to creditors, customers, and trading partners as the cause of systemic distress. It associates systemic risk with a cluster of failures in the financial sector. This perception is consistent with the characterization of systemic risk by the U.S. Commodity Futures Trading Commission,[1] the Bank for International Settlements,[2] the International Monetary Fund,[3] the Financial Stability Board,[4] and other major institutions.

[1] See http://www.cftc.gov/ConsumerProtection/EducationCenter/CFTCGlossary/glossary_s.html (last accessed June 24, 2011).

[2] See p. 177 of "64th Annual Report, Bank for International Settlements," 1994.

[3] See Chap. 2 of "Global Financial Stability Report, International Monetary Fund," April 2010.

[4] See p. 2 of "Guidance to Assess the Systemic Importance of Financial Institutions, Markets and Instruments: Initial Considerations," Financial Stability Boards, 2009.

Consider the process N counting failures in the financial system.[5] The value N_t represents the number of failures 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 represents the likelihood of failure by T of any fraction of the population of financial institutions at t. The right tail of this distribution describes 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 $(1 - \alpha)$. Formally, $V_t(\alpha, T) = \inf\{x \geq 0\colon P(D_t(T) > x \mid \mathcal{F}_t) \leq 1 - \alpha\}$.

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. Thus, $V_t(\alpha, T)$ changes over time as new information is revealed. This leads to a dynamic risk measure suitable for monitoring the financial system. The time-series behavior of $V_t(\alpha, T)$ provides a historical perspective on the level of systemic risk in the financial sector. The value $V_t(\alpha, T)$ also depends on the horizon T. By varying T for fixed t, we obtain a term structure of systemic risk. By varying the confidence level α for fixed t and T, we can adjust the threshold defining a systemic event. Finally, the sensitivity of $V_t(\alpha, T)$ to a given event, such as the failure of an institution or a shock to risk factors, describes the vulnerability of the financial system at time t.

It is important to note that the measurement of systemic risk need not be predicated on the value at risk. Our statistical methodology, which is developed in §3 and implemented in §4, leads to estimates of the entire conditional distribution of $D_t(T)$. Thus, our approach covers also alternative risk measures such as expected shortfall. The expected shortfall is defined as the conditional mean of $D_t(T)$ given $D_t(T) \geq l$, where l is some high level, such as $V_t(\alpha, T)$. Although 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. Moreover, our statistical methodology extends to risk measures of the conditional distribution of the value-weighted default rate, which takes account of the default volume.

[5] 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.

The risk measures we propose differ from those in the literature in several respects. First, our measures quantify the systemic risk of the financial sector as a whole. The measures of Acharya et al. (2010), Adrian and Brunnermeier (2010), Brownlees and Engle (2010), Huang et al. (2009), and others quantify the marginal systemic risk of individual institutions. Our system-wide measures complement these marginal measures by providing both time-series and term-structure perspectives on the total systemic risk prevailing in the financial sector. Moreover, they offer a unique perspective on the contribution of systemic risk to economy-wide risk (see §5). This allows one to judge the relative significance of systemic risk. Finally, our measures facilitate stress tests, i.e., analyses of the impact of adverse events on systemic risk (see §6).

Second, our risk measure definition recognizes that it is the correlated failure of institutions to meet obligations to creditors, customers and trading partners that causes systemic distress. Our risk measures associate the systemic risk of the financial sector with a cluster of actual failures in that sector. The measures of Acharya et al. (2010), Adrian and Brunnermeier (2010), Brownlees and Engle (2010), Lehar (2005), and others associate systemic risk with a collapse of asset prices across institutions. The decline of an institution's stock price increases the stress on an institution but it may not cause failure. By focusing on actual failure events, our measures address a broader set of conditions that may cause systemic distress. We seek to understand the factors that influence the timing of failures based on data on actual failure experience in the U.S. economy and data on a number of explanatory state variables. If equity price fluctuations have an impact on failures then our risk measures will capture that.

Third, our risk measures are defined with respect to the actual probability, as are the ones of Acharya et al. (2010), Adrian and Brunnermeier (2010), Brownlees and Engle (2010), Lehar (2005), and others. The alternative measures of Avesani et al. (2006), Chan-Lau and Gravelle (2005), Huang et al. (2009) are defined relative to a risk-neutral pricing probability implied by market rates of credit derivatives. These measures incorporate the risk premia investors demand for bearing correlated default risk. Our measures do not reflect these premia. They are consistent with actual default experience. On the other hand, they tend to be less sensitive to real-time market developments.

3. Statistical Methodology

This section develops a likelihood approach to estimating the measures of systemic risk proposed in

§2. In a first step, we formulate and estimate a hazard, or intensity-based, model of correlated default timing in the economy. 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 §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 nonnegative parameters, and D_n^* is the default volume, i.e., the total amount of debt outstanding at default of the nth defaulter, measured in million dollars. 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 counting process N^* is nonexplosive.

The intensity (1) is the sum of two terms. The first term, which we call *baseline hazard*, takes a standard Cox proportional hazards form. It models the influence on economy-wide default arrivals of explanatory covariates X^*, and it captures the clustering of defaults because of 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 and Van de Gucht (1999), and many others to predict industrial defaults; and by Brown and Dinc (2005, 2011), Cole and Wu (2010), Lane et al. (1986), Whalen (1991), Wheelock and Wilson (2000), and others to predict bank failures.

The second term, which we call *spillover hazard*, is not present in the traditional proportional hazards formulation. It models the influence on economy-wide default arrivals of past defaults. At a default, the hazard rate jumps, with a 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.[6] Thus, the bigger a 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 inclusion of the spillover hazard term is motivated by the results of the empirical analyses of Aharony and Swary (1996), Azizpour et al. (2010), Collin-Dufresne et al. (2009), Das et al. (2007), Duffie et al. (2009), Lando and Nielsen (2010), 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 ("frailties"). With contagion, a default increases the likelihood of additional defaults, an effect 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 missing or unobserved explanatory variables leads to a jump of the (filtered) intensity at a default. The spillover hazard term in (1) seeks to capture the statistical implications of these spillover effects for default timing, by allowing the intensity λ^* to jump at a default. In particular, it is designed to capture 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 the economic mechanisms generating the excess clustering. On the other hand, when taken to the data, this formulation does not offer information about the relative importance of the different sources of the excess clustering. For an analysis of these for U.S. corporate (industrial and financial) defaults, see Azizpour et al. (2010).

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

[6] 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 nonfinancial corporate default.

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 §4.3 below, we found these alternatives to be statistically inferior to the model (1)–(2).

3.2. System-Wide Default Timing

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

Proposition 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 increasing processes $\int_0^\cdot \lambda_s^*\, ds$ and $\int_0^\cdot \lambda_s\, ds$. The predictability of Z follows from the predictability of these processes. □

The value Z_t is the conditional probability at t that a firm in the financial system fails next, given a default in the economy in the next instant. For a precise statement, see Proposition 3.1 in Giesecke et al. (2011). We formulate and estimate a parametric model of Z, which then leads to λ via Proposition 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 1 if the nth 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. The variable Y_n has a Bernoulli distribution with success probability $Z_{T_n^*}$, where[7]

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

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 §4.2, and β is a vector of constant parameters. Given observations $(Y_n)_{n=1,\ldots,N_t^*}$ and $(X_s)_{s\le t}$ during the sample period $[0, t]$, we estimate β by solving the log-likelihood problem

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

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 nonsingular. See McCullagh and 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.

3.3. Failure-Based Risk Measures

The intensity $\lambda = \lambda^* Z$ governs the dynamics of the system-wide default process N, and hence the measures of systemic risk introduced in §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 $(t, T]$.[8] 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.

The estimates of $V_t(\alpha, T)$ take account of the time-variation of the covariates X^* between t and T and the cross-sectional variation of the default volume D_n^*. As explained in Appendices A and B, this is based on a vector autoregressive time-series model for 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).

The risk measure estimates also address the dynamic interaction between the real and financial sectors. Financial firms are intertwined with the real sector, so defaults in that sector clearly have an influence on financial firms, and vice versa. Our two-step inference method seeks to capture this influence by treating the failures in the financial system in the context of the defaults in the broader economy. An alternative approach, in which a model of λ would be estimated directly based on the historical default experience in the financial system, would ignore this influence. Another advantage of our two-step method is that it allows us to exploit a larger data set of economy-wide defaults, leading to a greater sample size and more accurate inference.[9]

4. Empirical Analysis

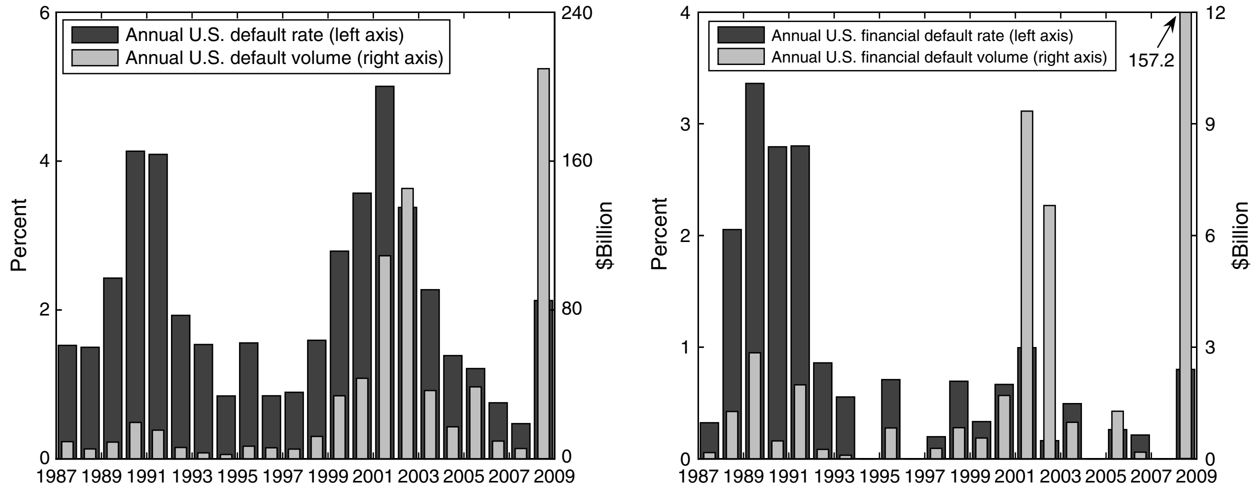
This section describes the default timing data, the data on explanatory covariates, our basic estimation results, and their statistical evaluation.

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

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

[9] For our sample period 1987–2008, the number of system-wide failures is 83 and the number of economy-wide defaults is 1,193.

Figure 1 Default Timing and Volume Data



Source. Moody's Default Risk Service.
Notes. Left panel: The 1-year economy-wide default rate in the universe of Moody's rated issuers. Right panel: The 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.

4.1. Default Timing Data

Our sample period is from January 1, 1987, to December 31, 2008.[10] 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 of the U.S. Bankruptcy Act, Chapter 10, Chapter 11, Chapter 7, Prepackaged Chapter 11 of the U.S. Bankruptcy Code) 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 1,193 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).[11] 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.[12]

4.2. Covariates

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

(1) The trailing 1-year return on the S&P 500 index, obtained from Economagic. Duffie et al. (2007) found this variable to be a significant predictor of industrial defaults.

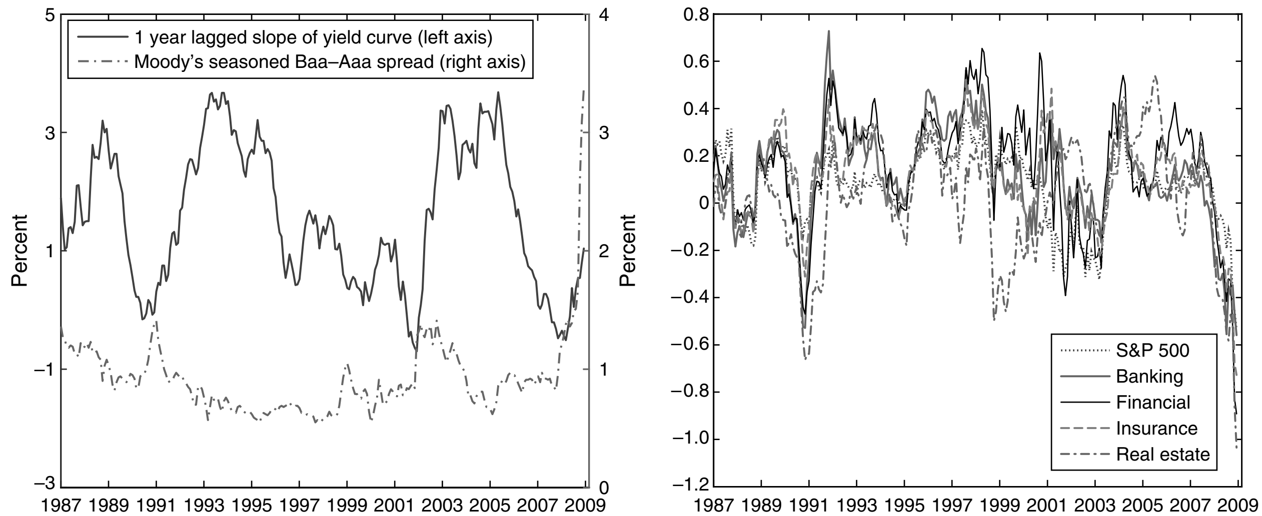
(2) The 1-year lagged slope of the yield curve, computed as the spread between 10-year and 3-month Treasury constant maturity rates, as a forward-looking indicator of real economic activity. Estrella and Trubin (2006) found this variable to have strong predictive power for future recessions. We obtained the H.15 release of Treasury rates from the website of the Federal Reserve Bank of New York.

[10] This period was determined by the availability of data for the covariates specified in §4.2. Default data for the period from January 1, 2009, to June 30, 2009, were used for the out-of-sample analysis.

[11] 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; (11) utilities.

[12] As explained by Hamilton (2005), the volume reported by Moody's 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 2 Time-Series of Explanatory Covariates



Notes. 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&P 500 index and the banking and FIRE portfolios.

(3) The default spread, defined as the yield differential between Moody's seasoned Aaa-rated and Baa-rated corporate bonds. Chen et al. (2009) argue that the default spread is a measure of aggregate credit risk that is largely unaffected by bond market frictions such as 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 Treasury–eurodollar (TED) spread, defined as the difference between the 3-month London Interbank Offered Rate (LIBOR) and 3-month Treasury rates, as an indicator of credit risk in the financial system.[13] 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.[14] The right panel of Figure 2 shows the return series.

(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 nonfinancial firm.

Table 1 provides descriptive statistics of the covariates. 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, the 3-month LIBOR rate, the 3-month general collateral repo rate, the difference between the 3-month repo rate and the 3-month Treasury rate, the ratio of credit to the gross domestic product (GDP) and the monthly change of credit/GDP, and the unemployment rate and its monthly change.

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 §4.2. The other covariates are used for the estimation of the process Z in §4.5.

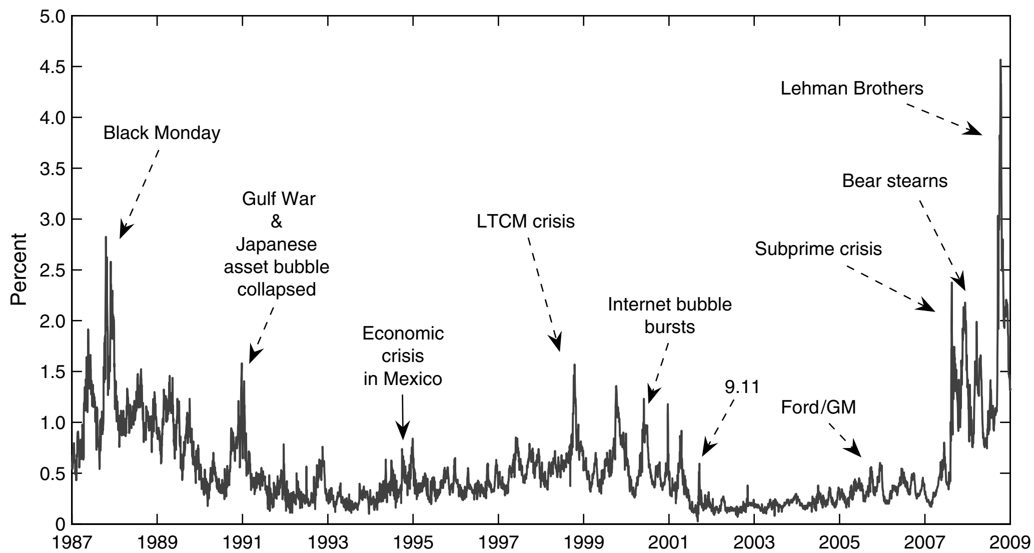
Table 2 reports the parameter estimates, along with estimates of asymptotic standard errors.[15] The intensity is increasing in the default spread, and decreasing in the trailing return on the S&P 500 and the lagged slope of the yield curve. The jump of the intensity at a default, measured in events per year, is estimated to be 2.3 plus roughly one half of the logarithm of the default volume, measured in million

[13] 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.

[14] http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html (last accessed June 24, 2011).

[15] The parameter space $\Theta = (-5, 5)^4 \times (0, 15) \times (0, 5)^2$. The fmincon routine of Matlab was used to optimize the likelihood. We performed an optimization for each of 10 randomly chosen initial parameter sets. Each of these optimizations converged to the values reported in Table 2.

Figure 3 TED Spreads During the Sample Period, Along with Significant Events



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 et al. (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 and 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. Because of 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,[16] then the outcome of Ψ is 131.4, 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 interarrival times are independent standard exponential variables. Das et al. (2007) and Azizpour et al. (2010) use similar approaches to test the fit of their default timing models.

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 and Nielsen (2010).

[16] The parameter estimates are as follows (standard error (SE) in parentheses): constant 4.1597 (0.1443), S&P 500 −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.

Table 1 Descriptive Statistics of the Covariates

	Obs	Mean	Std. dev.	Median	Min	Max
S&P 500	264	0.0750	0.1586	0.0974	−0.5364	0.3937
Yield slope	264	1.5663	1.1846	1.4300	−0.7000	3.6800
Default spread	264	0.9169	0.3258	0.8700	0.5500	3.3800
Banking	264	0.1214	0.2120	0.1188	−0.5588	0.7276
Financial	264	0.1512	0.2440	0.1864	−0.8921	0.6542
Insurance	264	0.1159	0.1812	0.1239	−0.7277	0.5217
Real estate	264	0.0197	0.2526	0.0468	−1.0365	0.5388
TED spread	264	0.5509	0.4290	0.3972	0.1249	3.3686

Table 2 Maximum Likelihood Estimates (MLE) of Economy-Wide Intensity Parameters, Asymptotic Standard Errors (SE), t-Statistics (t-Stat), and Bayes Factor Statistics (Ψ)

	Baseline hazard				Spillover hazard		
	Constant	S&P 500	Yield slope	Baa–Aaa	κ	γ	δ
MLE	2.3026	−0.4410	−0.2140	0.5092	6.0592	2.3205	0.4781
SE	0.0605	0.0524	0.0336	0.0534	0.1108	0.0811	0.0233
t-stat.	38.04	−8.42	−6.37	9.53	54.71	28.60	20.56
Ψ		0.1298	3.0987	1.8310		213.4039	
		26.5308					

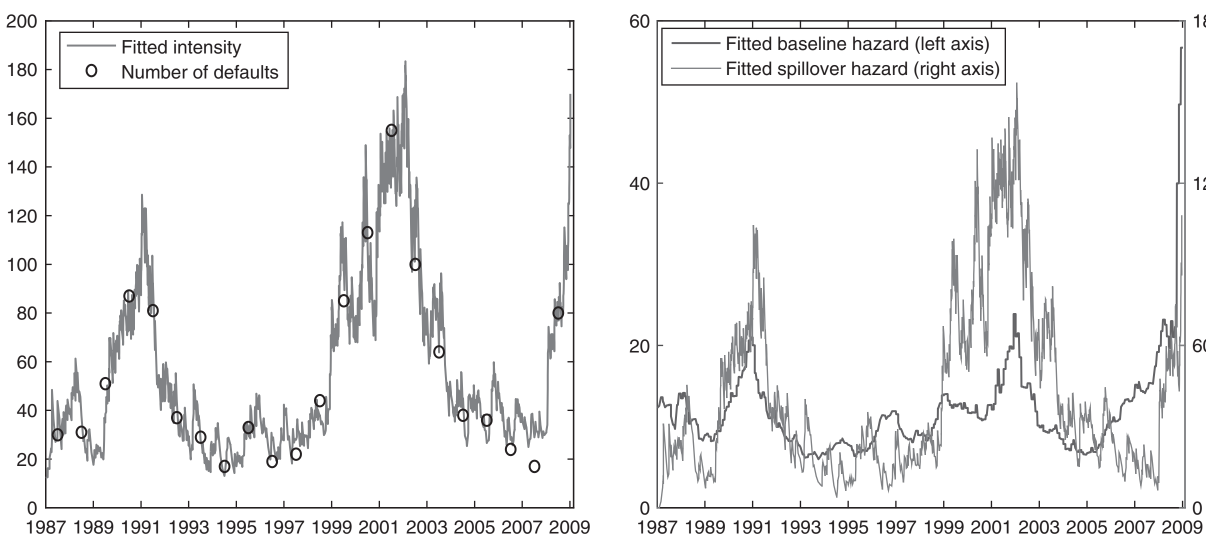
For given bin size c, we denote by U_n the number of observed events in the nth successive time interval lasting for c units of transformed time. With a total of K bins, the null hypothesis is that the $U_1, \ldots, U_K$ are independent Poisson variables with mean c. We BBB consider bin sizes $c = 2$, 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 3(a) indicates that there is no evidence against the null for bin sizes 4–10, 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 independent and identically distributed 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 3(b) suggest that there is no significant deviation of the upper-quartile tails from the theoretical Poisson tails for bin sizes 4–10, 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,

Figure 4 Fitted Economy-Wide Intensity λ^*



Notes. Left panel: Yearly defaults and fitted intensity. Right panel: Intensity decomposition; fitted baseline hazard vs. fitted spillover hazard.

Table 3 Goodness-of-Fit Tests of the Economy-Wide Intensity

(a) Fisher's dispersion test

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

(b) Mean and median of default upper quartile tail 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

(c) Excess default autocorrelation 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

Note. The t-statistics for A are presented for the test $A = c$.
*Significance at the 5% level.

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 3(c) shows that the fitted coefficients are not significantly different from their theoretical values for bin sizes 4–10, 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 because of 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 is challenging to match, at the same time, the few extremely short interarrival times, and the many longer interarrival times that constitute the vast majority of the sample.

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 $h = 1/12$.[17] We have also considered, but rejected for lack of significance in the presence of these variables, the other covariates discussed in §4.2.

Table 4 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 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,

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

Table 4 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)

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

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 a p-value of less than 0.0001.

5. Systemic Risk Analysis

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 begin by examining the fitted system-wide intensity $\lambda_t = \lambda_t^* Z_t$, which measures the level of instantaneous systemic risk prevailing at time t. The time-series behavior of λ_t, shown in the left panel of Figure 6, indicates that the 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, calculated as the fitted ratio of the spillover hazard to the economy-wide default intensity λ_t^*. The estimates indicate the presence of failure clustering not caused by variations in the 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 $D_t(t+\Delta)$ during the period $(t, t+\Delta]$, for given Δ. This is done by exact Monte Carlo simulation[18] (see §3.3), and is based on the models for λ^* and Z fitted with data for the period from January 1, 1987, to t. The fitted distribution 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 fitted conditional distribution of $D_t(t+0.5)$ for conditioning times t varying semiannually between December 31, 1997, and December 31, 2008, for a 6-month horizon. 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 during the forecast period. The time-series behavior of the tail suggests that systemic risk has increased very sharply during the second half of 2008.

We contrast the distribution of the system-wide default rate $D_t(t+0.5)$ with that of the economy-wide default rate for the period $(t, t+0.5]$, shown in Figure 8. Although the increase of economy-wide default risk during the second half of 2008 is clearly visible, the economy-wide risk prevailing at the end of 2008 is comparable with that when the Internet bubble burst in 2001. This is in stark contrast to the behavior of systemic risk shown in Figure 7: The systemic risk prevailing in 2001 is dwarfed by the systemic risk prevailing at the end of 2008. Intuitively, the contribution of systemic risk to economy-wide default risk is much higher during the financial crisis of 2007–2009 than during the Internet bubble period. This relationship is also evident in the time-series behavior of the value at risk $V_t(\alpha, t+0.5)$ of the system-wide default rate $D_t(t+0.5)$, which is shown in the left panel of Figure 9 for conditioning times t varying semiannually between December 31, 1997, and December 31, 2008. Compare with the behavior of the value at risk of the economy-wide default rate, shown in the right panel of Figure 9.

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 December 31, 2008, the end of the sample period, as a function of Δ for each of several confidence levels α.

There are other 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 $(t, T]$. An advantage of this measure is that it does not require the choice of a confidence level. The right

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

Figure 5 Left Panel: Observed Binary Response Variables Y_n and Fitted Process Z; Right Panel: Power Curve for the Fitted Process Z

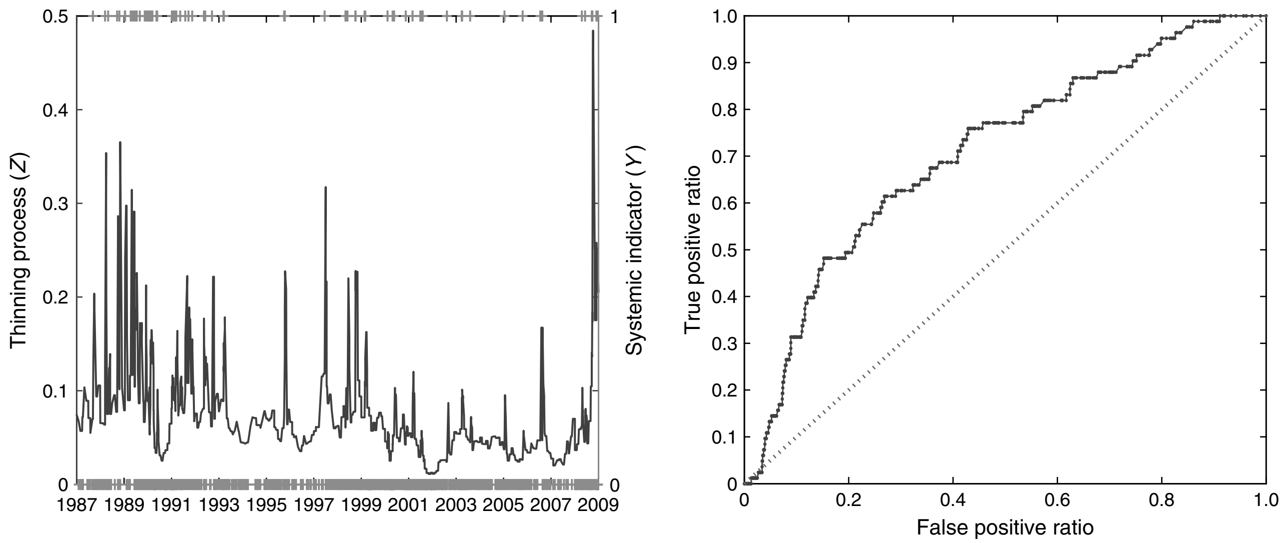
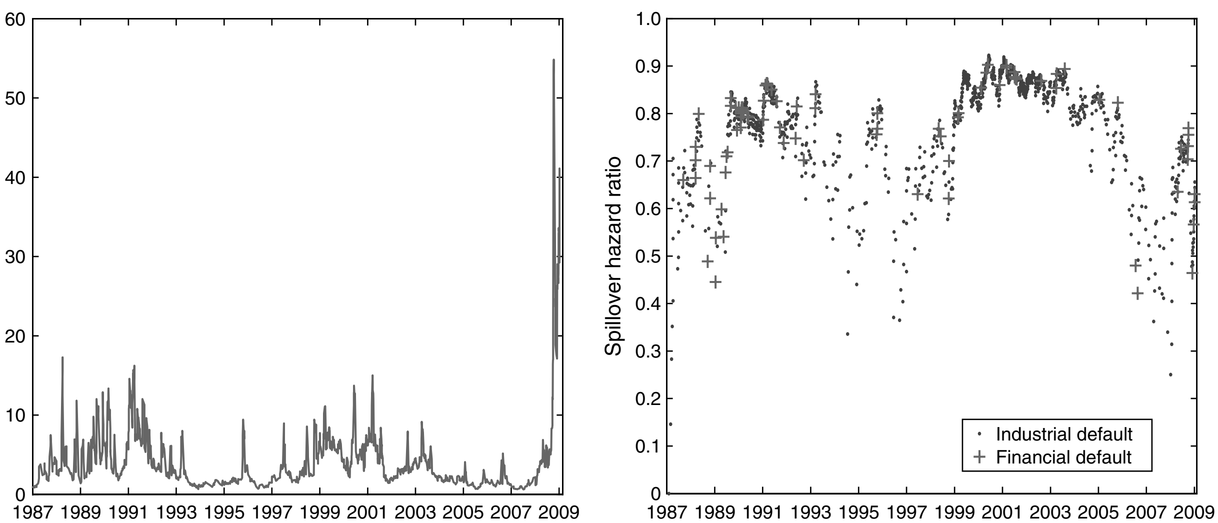


Figure 6 Left Panel: Fitted System-Wide Failure Intensity λ, Based on the Parameter Estimates Reported in Tables 2 and 4; Right Panel: Fitted Fraction of λ Tied to the Spillover Hazard Term, with Default Events Indicated



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 et al. (2009).

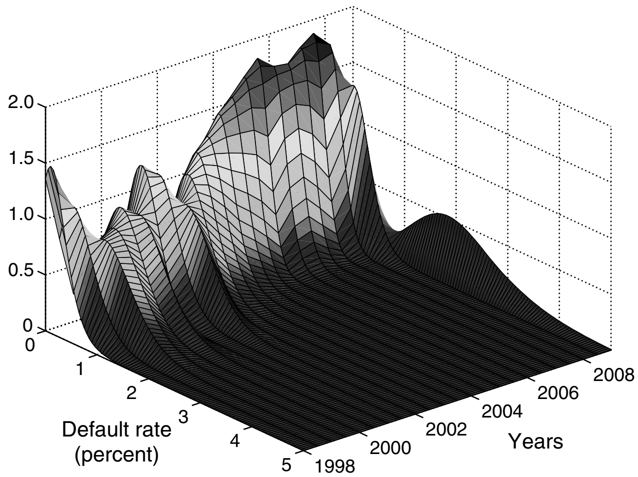
Let n be the number of forecast periods. Further, let $n_1 \leq n$ be the number of periods for which the corresponding value-at-risk forecast was violated, i.e., the number of periods for which the realized default rate was strictly greater than the fitted value at risk $V_t(\alpha, t+\Delta)$. Then, $n_0 = n - n_1$ 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 one degree of freedom under the null hypothesis of the theoretical $(1-\alpha)$ violation rate.[19]

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

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 Semiannually Between December 31, 1997, and December 31, 2008**



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 independent 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 and 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

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

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 $\pi_1 = (n_{01}+n_{11})/(n_{00}+n_{01}+n_{01}+n_{11})$.[20] Under the null of the indicators forming a first-order Markov chain, this statistic has a limiting chi-squared distribution with one 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 two degrees of freedom.[21]

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

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

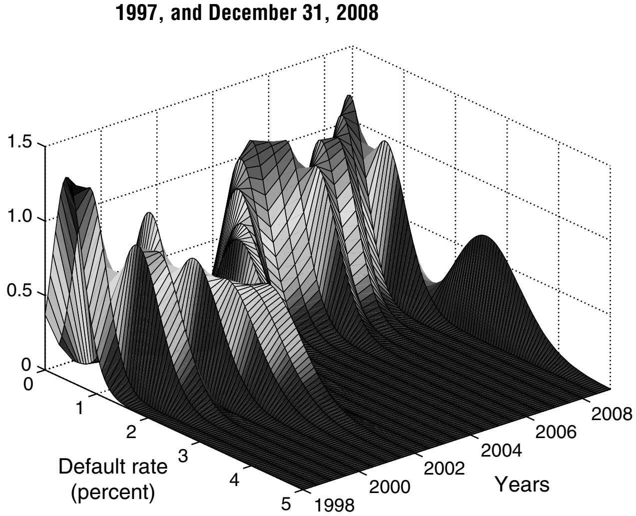
where the error term ϵ_t has a logistic distribution. We test whether the β_i coefficients are statistically significant and whether $P(I_t = 1) = e^{\gamma}/(1+e^{\gamma}) = 1-\alpha$. Denote the ith response variable by Y_i and the corresponding vector of regressors by X_i, for $i = 1, \ldots, n-1$. Also, let $\pi_i = e^{\hat{\gamma}+\hat{\beta}X_i}/(1+e^{\hat{\gamma}+\hat{\beta}X_i})$, where $(\hat{\gamma}, \hat{\beta})$ 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 $\gamma = \log((1-\alpha)/\alpha)$, the log-likelihood ratio test statistic

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

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

Table 5 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.[22] None of

Figure 8 **Fitted Conditional Distribution (Kernel-Smoothed) of the Economy-Wide 6-Month Default Rate for Conditioning Times t Varying Semiannually Between December 31, 1997, and December 31, 2008**



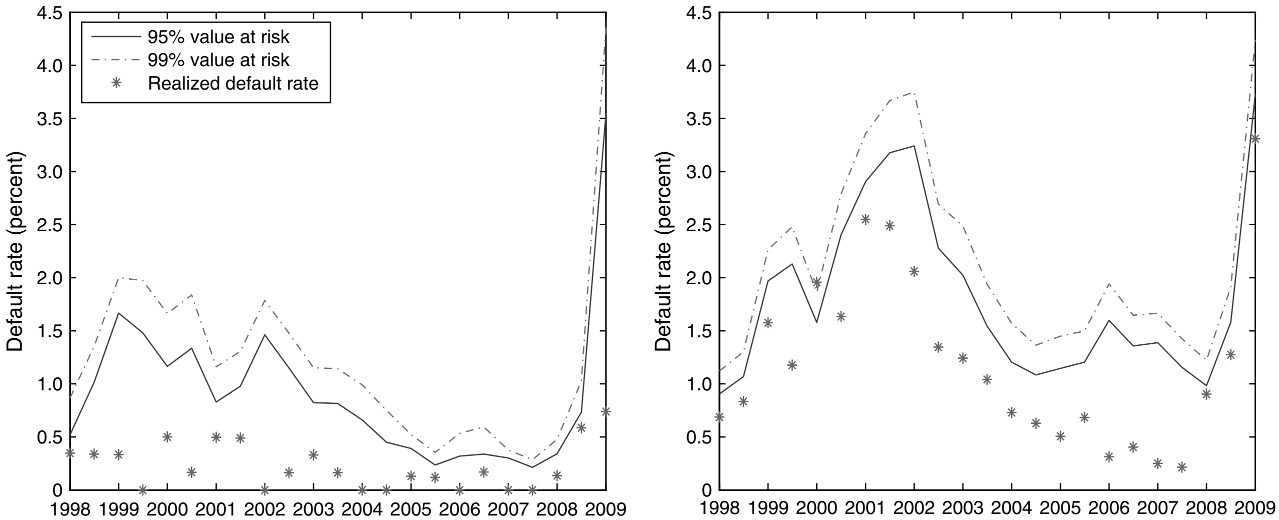
Note. The default rate is obtained by normalizing the number of economy-wide defaults during $(t, t+0.5]$ by the total number of firms in the economy at t.

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

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

[22] We also use 2009 default data in the tests: we validate the forecasts obtained on December 31, 2008, on the realized default rates in 2009, which are available to us for the first one, three, and six months of 2009.

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 Semiannually Between December 31, 1997, and December 31, 2008, vs. Realized Default Rate; Right Panel: Fitted Value at Risk of the Economy-Wide Default Rate vs. Realized Default Rate**



the null hypotheses can be rejected at the 10% level. These results suggest that the fitted measures accurately quantify the systemic risk of the U.S. financial sector for each of several risk horizons and confidence levels. This validates our default hazard model (1)–(2) and our two-stage inference procedure, and indicates that our model successfully captures the sources of failure clustering in the United States. We conclude that our risk measures are useful for monitoring the level of systemic risk in the U.S. financial sector.

5.3. Forecast Accuracy of Equity-Based Measures

To provide some perspective on the predictive performance of our failure-based measures of systemic risk, we test the out-of-sample forecast accuracy of an analogous "equity-based" risk measure associating systemic risk with a collapse of stock prices of financial institutions. Acharya et al. (2010), Adrian and Brunnermeier (2010), Brownlees and Engle (2010), Lehar (2005), and others develop alternative equity-based measures.

We follow the econometric approach of Brownlees and Engle (2010) to construct an equity-based measure that is analogous to our failure-based measure. The value at risk $V_t^e(\alpha, T)$ at time t and at confidence level $\alpha \in (0, 1)$ is defined as the smallest number x such that the conditional probability at t that the return on the S&P 500 Financials Index during $(t, T]$

Figure 10 **Left Panel: Term Structure of Systemic Risk on December 31, 2008: Fitted Value at Risk $V_t(\alpha, t+\Delta)$ on December 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 December 31, 1997, and December 31, 2008, for Each of Several Horizons Δ**

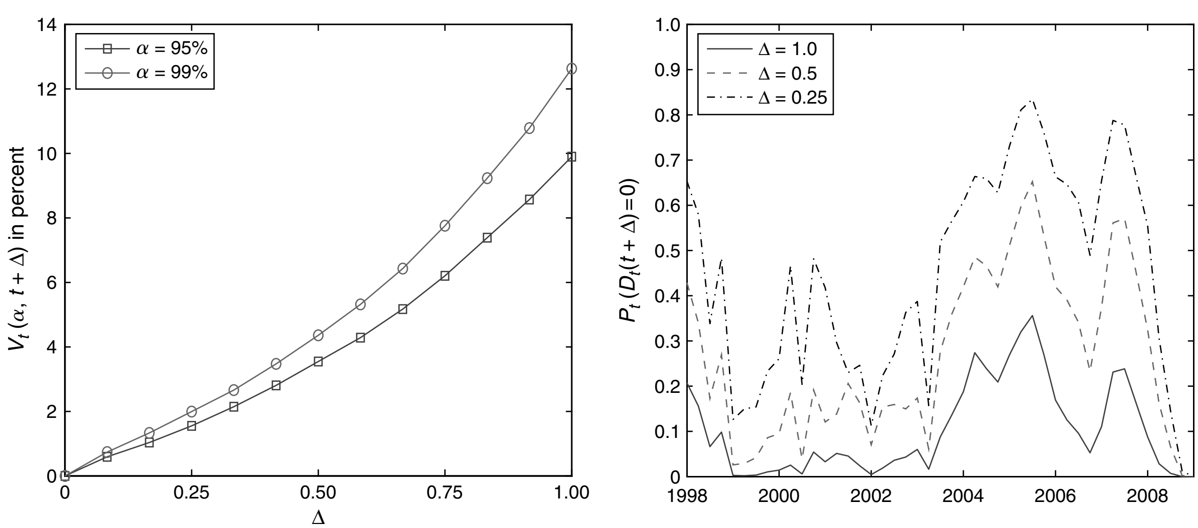


Table 5 Out-of-Sample Tests of the Forecast Accuracy of the Fitted System-Wide Value at Risk for Each of Several Horizons

			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

Note. The period considered is from January 1998 to June 2009.

is less than x is no larger than $(1-\alpha)$. The S&P 500 Financials Index is a capitalization-weighted index that consists of the common stocks of the following industries: banks, diversified financials, brokerage, asset management, insurance, and real estate, including investment trusts. Following Brownlees and Engle (2010), the index return $r_t = r_t(\Delta)$ for the period t to $t+\Delta$ is modeled as

$$r_t = \sigma_t \epsilon_t, \tag{10}$$

where the ϵ_ts are independent and identically distributed random variables with zero mean and unit variance, and the volatility σ_t obeys a TARCH model

$$\sigma_t^2 = \omega_0 + \omega_1 r_{t-\Delta}^2 + \omega_2 r_{t-\Delta}^2 U_{t-\Delta} + \omega_3 \sigma_{t-\Delta}^2, \tag{11}$$

where U_t takes the value 1 if $r_t < 0$ and 0 otherwise, and the ω_is are estimated using a nonparametric kernel estimation approach (see Brownlees and Engle 2010).

We compare the fitted $V_t^e(\alpha, T)$ with the realized index return for the period t to $T = t+\Delta$. The accuracy is formally tested using the tests described in §5.2, which were also used to evaluate the forecasts of our failure-based measure. The results of these tests are reported in Table 6 for each of several forecast horizons Δ and confidence levels α. The p-values of these tests fail to reach the standard 10% level in most tests, and are significantly lower than the p-values of the tests of our failure-based measure $V_t(\alpha, T)$ reported in Table 5. This indicates that the forecasts of the equity-based measure considered here are much less accurate than those of our failure-based measure.

6. Stress Testing Systemic Risk

Our measures can be used to quantify the impact on systemic risk of an adverse event, such as the failure of an institution or a shock to risk factors. The sensitivity of the value at risk $V_t(\alpha, t+\Delta)$, which represents this impact, describes the vulnerability of the financial system. An analysis of this vulnerability can guide policy decisions, for example whether a failing institution should be bailed out.

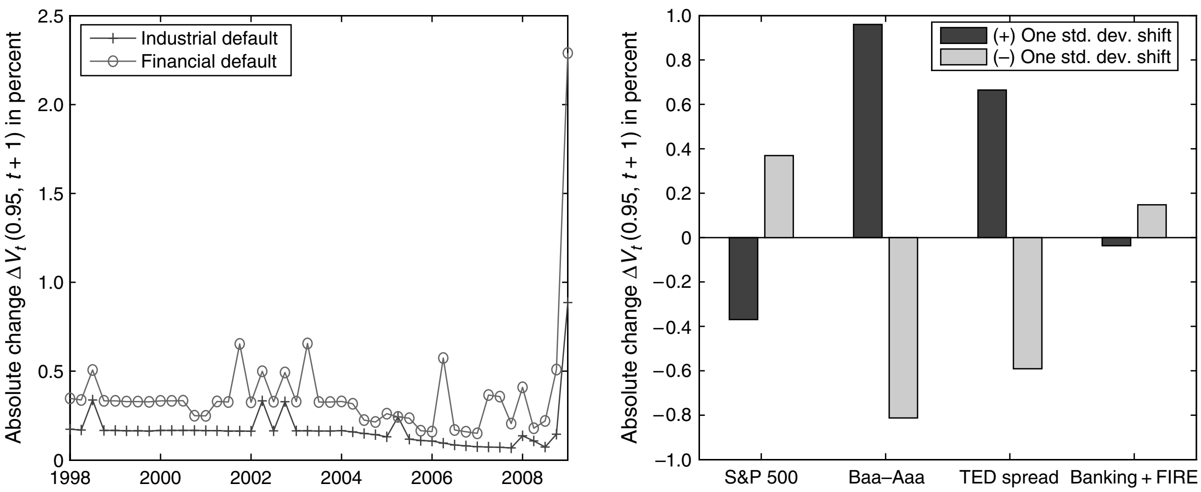
We begin by considering the change $\Delta V_t(\alpha, t+\Delta)$ of the value at risk at t in response to a failure at t. To this end, we first estimate the time t value at risk $V_t(\alpha, t+\Delta)$ based on data up to t. Then we enlarge the data set by including a hypothetical default event at t, and reestimate $V_t(\alpha, t+\Delta)$ based on the enlarged data set. The total debt outstanding at default is taken to be the sample mean of the default volumes observed to t. Finally, we calculate $\Delta V_t(\alpha, t+\Delta)$ as the difference between the post- and predefault risk measure

Table 6 Out-of-Sample Tests of the Forecast Accuracy of the Fitted Equity-Based Value at Risk of the Financial System for Each of Several Horizons

			Uncond. coverage		Markov		CAViaR	
	Δ	Obs.	LR	*p*-value	LR	*p*-value	LR	*p*-value
95% VaR	1Y	11	2.4751	0.1157	6.2042	0.0450	12.8036	0.0051
	6M	23	2.2144	0.1367	8.0855	0.0175	19.9235	0.0001
	3M	45	6.9382	0.0084	11.6048	0.0030	11.8454	0.0079
	1M	133	5.0533	0.0246	7.6562	0.0218	9.5414	0.0229
99% VaR	1Y	11	2.7094	0.0998	2.7094	0.2580	9.3912	0.0245
	6M	23	10.2213	0.0014	16.0924	0.0003	23.8276	0.0000
	3M	45	2.9212	0.0874	3.1117	0.2110	3.3129	0.3459
	1M	133	3.5234	0.0605	3.7735	0.1516	4.9621	0.1746

Note. The period considered is from January 1998 to June 2009.

Figure 11 **Sensitivity of Systemic Risk**



Notes. Left panel: Impact of a default—$\Delta V_t(0.95, t+1)$ for conditioning times t varying quarterly between December 31, 1997, and December 31, 2008. Right panel: Impact of a shock to risk factors—$\Delta V_t(0.95, t+1)$ on December 31, 2008. In the case of the "Banking + FIRE" factor, we assume a one standard-deviation shock to banking, financial, insurance, and real-estate portfolio returns.

estimates. The variable $\Delta V_t(\alpha, t+\Delta)$ 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 event, including the sector of the defaulter (industrial versus financial) and the total debt outstanding at default.

We estimate $\Delta V_t(\alpha, t+\Delta)$ for each of two events, a failure of a financial institution and a default of an industrial firm. The left panel of Figure 11 shows $\Delta V_t(0.95, t+1)$ for conditioning times t varying quarterly between December 31, 1997, and December 31, 2008. The failure of a financial institution is estimated to have a larger impact on systemic risk than the default of an industrial firm. The time series of $\Delta V_t(0.95, t+1)$ indicates that the vulnerability of the financial system increased dramatically during the second half of 2008. Lehman Brothers collapsed on September 15, 2008.

Next we consider the impact on $V_t(\alpha, t+\Delta)$ of a hypothetical shock to the explanatory covariates (risk factors). To this end, we assume a one standard deviation shock to a covariate and estimate the corresponding change $\Delta V_t(\alpha, t+\Delta)$ in $V_t(\alpha, t+\Delta)$. The right panel of Figure 11 shows $\Delta V_t(0.95, t+1)$ on December 31, 2008, for each of several risk factors. We see that a shock to the default spread causes the greatest shift in systemic risk at that time. We can extend the analysis to include the effect of joint shocks to several covariates.

7. Conclusion

This paper proposes dynamic measures of the systemic risk of the financial sector as a whole. Systemic risk is defined as the conditional probability of failure of a sufficiently large fraction of the total population of financial institutions. This definition recognizes the correlated failure of institutions to meet obligations to creditors, customers, and trading partners as the cause of systemic distress.

The estimators of the failure probability are based on a hazard model of correlated failure timing. This model incorporates the influence on failure timing of time-varying macroeconomic and sector-specific risk factors, and past defaults. The formulation seeks to capture the spillover effects channeled through a complex network of relationships in the economy. A likelihood-based inference method addresses the statistical implications of industrial defaults for financial failures, and vice versa. It ensures that the systemic risk estimators incorporate the dynamic interaction between the real and financial sectors. Tests indicate that our measures provide accurate out-of-sample forecasts of systemic risk in the United States for the period from 1998 to 2009.

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 a large cluster of failures, 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.

Acknowledgments

The authors are grateful for discussions with Viral Acharya, Jorge Chan-Lau, Darrell Duffie, Marco Espinosa, and Juan Solé; for comments from Monica Banyi, Mathias Drehmann, Brenda González-Hermosillo, Laura Kodres, Mike So, and

Hao Zhou; and for comments from conference and seminar participants at the IMF's Monetary and Capital Markets Department, Bank Negara Malaysia, the 12th Annual Financial Econometrics Conference on Volatility and Systemic Risk in Global Capital Markets, the HKIMR-BIS Conference on Financial Stability: Towards a Macroprudential Approach, the Fourth Annual Risk Management Conference on the Risk Management Paradigm in the Post-Crisis Era, the 13th Conference of the ECB-CFS Research Network on Macroprudential Regulation, and the 23rd Australasian Finance and Banking Conference. The authors thank Myeong Hyeon Kim for excellent research assistance. Baeho Kim is the corresponding author (website: http://biz.korea.ac.kr/~baehokim/).

Appendix 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

$$\Phi_t = \Pi_0 + \Pi_1 \Phi_{t-1} + \epsilon_t, \qquad \text{(A1)}$$

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 A.1 reports the estimators $\hat{\Pi}_i$ 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 $\hat{\Pi}_i$, 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 A.2. An analysis of the error series indicates the appropriateness of the model (A1) for our covariates. Figure A.1 visualizes the goodness-of-fit by plotting the predicted versus the realized covariates.

Table A.1 Fitted Coefficients of VAR(1) Model (A1) as of December 31, 2008

	Constant	TB (3M)	TB (10Y)	Baa	Aaa	LIBOR	S&P 500	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&P 500	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)

Note. The t-statistics are shown in parentheses.

Table A.2 Fitted Covariance Matrix Σ of the VAR(1) Error Term ϵ_t as of December 31, 2008

	TB (3M)	TB (10Y)	Baa	Aaa	LIBOR	S&P 500	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&P 500	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

Figure A.1 Realized vs. VAR(1) Predicted Time Series (Monthly) of Covariate Components

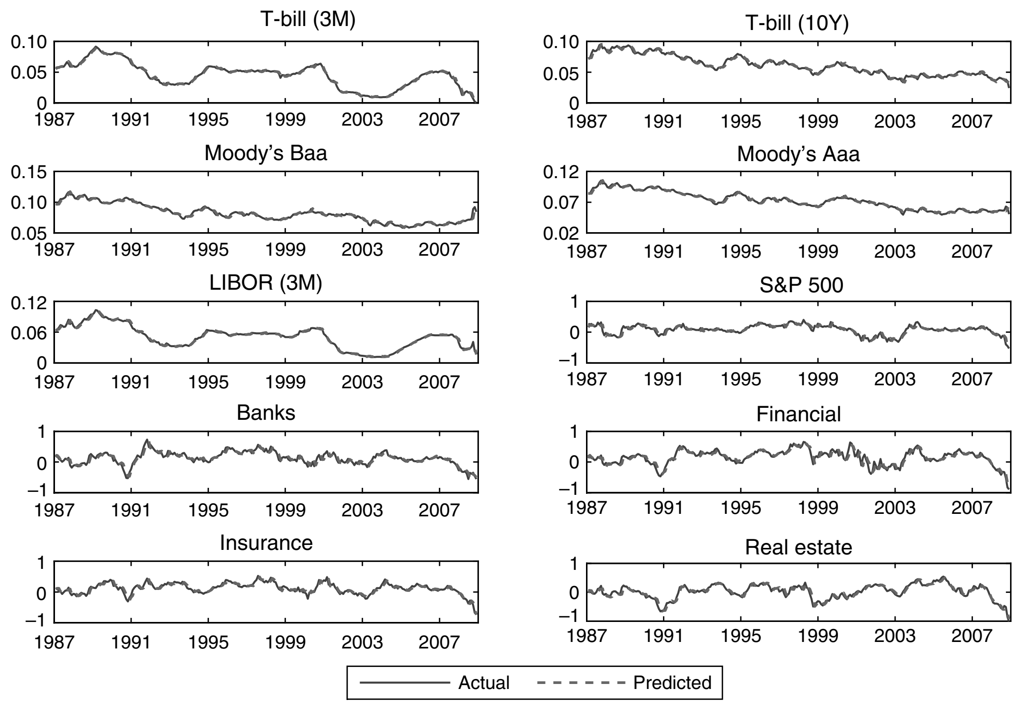
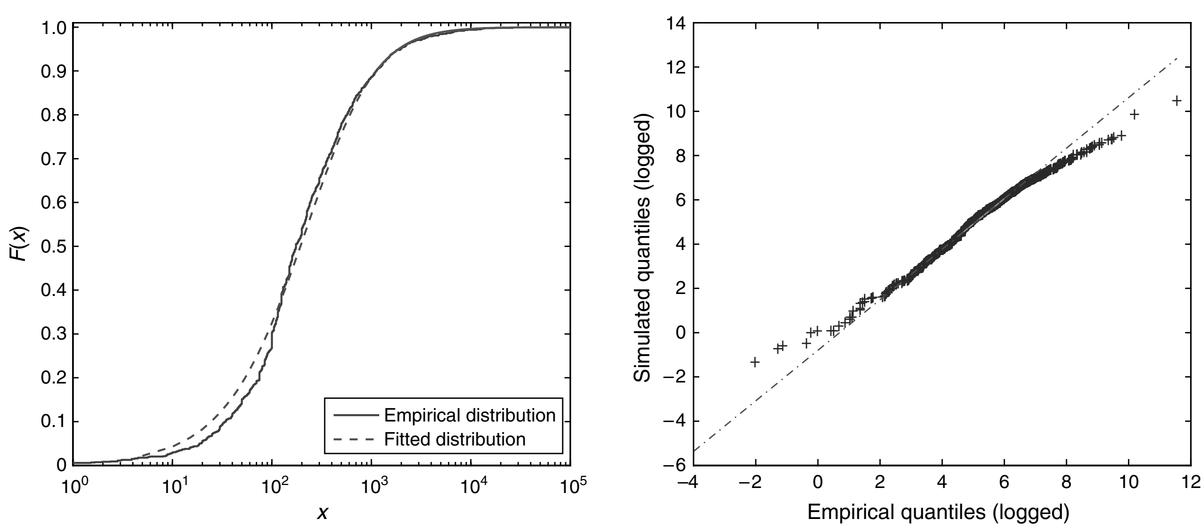


Figure B.1 Left Panel: Empirical Default Volume Distribution vs. Fitted Generalized Pareto Distribution as of December 31, 2008; Right Panel: Empirical Quantiles of the Observed Default Volumes vs. Quantiles of Realizations of Variables from the Fitted Pareto Distribution



Appendix 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

$$P(D_n^* > x) = \left(1 + \xi \frac{x}{\sigma}\right)^{-1/\xi} \tag{B1}$$

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 December 31, 2008. The left panel of Figure B.1 contrasts the fitted Pareto distribution with the empirical distribution of default volumes. The right panel of Figure B.1 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, V., T. Yorulmazer. 2008. Information contagion and bank herding. *J. Money, Credit, Banking* **40**(1) 215–231.

Acharya, V., L. Pedersen, T. Philippon, M. Richardson. 2010. Measuring systemic risk. Working paper, New York University, New York.

Adrian, T., M. Brunnermeier. 2010. CoVaR. Working paper, Princeton University, Princeton, NJ.

Aharony, J., I. Swary. 1983. Contagion effects of bank failures: Evidence from capital markets. *J. Bus.* **56**(3) 305–317.

Aharony, J., I. Swary. 1996. Additional evidence on the information-based contagion effects of bank failures. *J. Banking Finance* **20**(1) 57–69.

Avesani, R., A. G. Pascual, J. Li. 2006. A new risk indicator and stress testing tool: A multifactor *n*th-to-default CDS basket. IMF Working Paper 06/105, International Monetary Fund, Washington, DC.

Azizpour, S., K. Giesecke, G. Schwenkler. 2010. Exploring the sources of default clustering. Working paper, Stanford University, Palo Alto, CA.

Berkowitz, J., P. Christoffersen, D. Pelletier. 2009. Evaluating value-at-risk models with desk-level data. *Management Sci.*, ePub ahead of print January 28, http://mansci.journal.informs.org/cgi/content/abstract/mnsc.1080.0964v1.

Brown, C., S. Dinc. 2005. The politics of bank failures: Evidence from emerging markets. *Quart. J. Econom.* **120**(4) 1413–1444.

Brown, C., S. Dinc. 2011. Too many to fail? Evidence of regulatory reluctance in bank failures when the banking sector is weak. *Rev. Financial Stud.* **24**(4) 1378–1405.

Brownlees, C., R. Engle. 2010. Volatility, correlation and tails for systemic risk measurement. Working paper, New York University, New York.

Chan-Lau, J., T. Gravelle. 2005. The END: A new indicator of financial and nonfinancial corporate sector vulnerability. IMF Working Paper 05/231, International Monetary Fund, Washington, DC.

Chen, L., P. Collin-Dufresne, R. S. Goldstein. 2009. On the relation between credit spread puzzles and the equity premium puzzle. *Rev. Financial Stud.* **22**(9) 3367–3409.

Christoffersen, P. 1998. Evaluating interval forecasts. *Internat. Econom. Rev.* **39**(4) 842–862.

Cole, R., Q. Wu. 2010. Is hazard or probit more accurate in predicting bank failures? Evidence from U.S. bank failures. Working paper, DePaul University, Chicago.

Collin-Dufresne, P., R. Goldstein, J. Helwege. 2009. How large can jump-to-default risk premia be? Modeling contagion via the updating of beliefs. Working paper, Columbia University, New York.

Cooperman, E., W. Lee, G. Wolfe. 1992. The 1985 Ohio thrift crisis, the FSLIC's solvency and rate contagion for retail CDs. *J. Finance* **47**(3) 919–941.

Das, S., D. Duffie, N. Kapadia, L. Saita. 2007. Common failings: How corporate defaults are correlated. *J. Finance* **62**(1) 93–117.

Duffie, D., A. Eckner, G. Horel, L. Saita. 2009. Frailty correlated default. *J. Finance* **64**(5) 2089–2123.

Duffie, D., L. Saita, K. Wang. 2007. Multi-period corporate default prediction with stochastic covariates. *J. Financial Econom.* **83**(3) 635–665.

Engle, R., S. Manganelli. 2004. CAViaR: Conditional autoregressive value at risk by regression quantiles. *J. Bus. Econom. Statist.* **22**(4) 367–381.

Eraker, B., M. Johannes, N. Polson. 2003. The impact of jumps in volatility and return. *J. Finance* **58**(3) 1269–1300.

Estrella, A., M. R. Trubin. 2006. The yield curve as a leading indicator: Some practical issues. *Current Issues Econom. Finance* **12**(5) 1–7.

Giesecke, K. 2004. Correlated default with incomplete information. *J. Banking Finance* **28**(7) 1521–1545.

Giesecke, K., L. R. Goldberg, X. Ding. 2011. A top-down approach to multiname credit. *Oper. Res.* **59**(2) 283–300.

Hamilton, D. 2005. Moody's senior ratings algorithm and estimated senior ratings. Report, Moody's Investors Service. Last accessed June 24, 2011, http://www.moodys.com/researchandratings/.

Huang, X., H. Zhou, H. Zhu. 2009. A framework for assessing the systemic risk of major financial institutions. *J. Banking Finance* **33**(11) 2036–2049.

Kass, R., A. Raftery. 1995. Bayes factors. *J. Amer. Statist. Assoc.* **90**(430) 773–795.

Koopman, S. J., A. Lucas, A. Monteiro. 2008. The multistage latent factor intensity model for credit rating transitions. *J. Econometrics* **142**(1) 399–424.

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

Lando, D., M. S. Nielsen. 2010. Correlation in corporate defaults: Contagion or conditional independence? *J. Financial Intermediation* **19**(3) 355–372.

Lane, W., S. Looney, J. Wansley. 1986. An application of the Cox proportional hazards model to bank failure. *J. Banking Finance* **10**(4) 511–531.

Lehar, A. 2005. Measuring systemic risk: A risk management approach. *J. Banking Finance* **29**(10) 2577–2603.

McCullagh, P., J. Nelder. 1989. *Generalized Linear Models*. Chapman and Hall, London.

McDonald, C., L. Van de Gucht. 1999. High-yield bond default and call risks. *Rev. Econom. Statist.* **81**(3) 409–419.

Meyer, P.-A. 1971. Démonstration simplifée d'un théorème de Knight. *Séminaire de Probabilités V, Lecture Note in Mathematics*, Vol. 191. Springer-Verlag, Berlin, 191–195.

Ogata, Y. 1978. The asymptotic behaviour of maximum likelihood estimators of stationary point processes. *Ann. Inst. Statist. Math.* **30**(part A) 243–261.

Schwarcz, S. L. 2008. Systemic risk. *Georgetown Law J.* **97**(1) 193–249.

Shumway, T. 2001. Forecasting bankruptcy more accurately: A simple hazard model. *J. Bus.* **74**(1) 101–124.

Whalen, G. 1991. A proportional hazards model of bank failure: An examination of its usefulness as an early warning model tool. *Federal Reserve Bank of Cleveland Econom. Rev.* **1991**(QI) 21–31.

Wheelock, D., P. Wilson. 2000. Why do banks disappear? The determinants of U.S. bank failures and acquisitions. *Rev. Econom. Statist.* **82**(1) 127–138.