1

Objectives and Scope

Effective estimation of the likelihoods of default of individual corporate borrowers is crucial to those responsible for granting bank loans or investing in financial products exposed to corporate default. An ability to model the probability distribution of total default losses on portfolios of corporate loans, which depends on the measurement of default correlation across various firms, is an important input to the risk management of corporate loan portfolios, the determination of minimum capital requirements of financial institutions, and investment in structured credit products such as collateralized loan obligations that are exposed to multiple borrowers.

1.1 APPROACH

This book addresses the measurement of corporate default risk based on the empirical estimation of default intensity processes. The default intensity of a borrower is the mean rate of arrival of default, conditional on the available information. For example, a default intensity of 0.1 means an expected arrival rate of one default per 10 years, given all current information. Default intensities change with the arrival of new information about the borrower and its economic environment. I focus on methodologies for estimating default intensities and on some key empirical properties of corporate default risk. I pay special attention to the correlation of default risk across firms. The work summarized here was developed in a series of collaborations over roughly the past decade with Sanjiv Das, Andreas Eckner, Guillaume Horel, Nikunj Kapadia, Leandro Saita, and Ke Wang. Research on the measurement of corporate default risk remains active.

The data reveal, among other findings, that a substantial amount of power for predicting the default of a corporation can be obtained from the firm's "distance to default," a volatility-adjusted measure of leverage that is the basis of the theoretical models of corporate debt pricing of Black and Scholes (1973), Merton

(1974), Fisher, Heinkel, and Zechner (1989), and Leland (1994). Additional explanation is offered by a selection of macroeconomic variables and accounting ratios. Information relevant to the joint defaults of different firms that is not observable, or at least not captured by the chosen estimation approach, can be incorporated into a correlated default model through a statistical device known as "frailty." The last chapter shows that frailty contributes significantly to the likelihood of joint defaults of U.S. corporations.

An alternative approach is the estimation of a structural model of default, by which one directly captures how the managers of a firm opt for bankruptcy protection. For many practical purposes, structural empirical models of default have not yet matured sufficiently, given the complexity of most corporations and of the economic process of default, to be successfully applied to default risk estimation. Instead, the models explained here are based on reduced-form relationships between default risk and default predictors, particularly distance to default, that are suggested by structural models.

I suppose throughout that a firm's default intensity is of the form $\Lambda(X_t, \beta)$, where

- X_t is a list of firm-specific and macroeconomic default covariates, some of which are suggested by structural theories, in addition to unobservable covariates, and
- $\Lambda(\,\cdot\,, \beta)$ is a convenient *ad hoc* function, not necessarily based on a theory of the firm, depending on a parameter vector β to be estimated. Empirical results are also reported for a model estimated with non-parametric dependence of intensity on one of the key covariates, the distance to default.

Typical applications call for estimates of the likelihood of default (or joint default) over various time horizons. For this, it is also necessary to estimate the time-series behavior of the underlying covariate process X, which is treated here as a Markov process whose transition probabilities are governed by additional parameters to be estimated. With "frailty," default correlation is based on the assumption that some elements of the state vector X_t are not observable.

A structural approach could lead instead to an endogenously determined default intensity as a property of the decision problems faced by corporate managers, shareholders, creditors, and regulators. With this structural approach, the parameters to be estimated would specify the primitive technology of the firm, the contracting and capital-markets environment, and the preferences of the firm's managers and shareholders.

Structural model estimation offers the prospect of significant improvements in predictive power and should remain at the top of the research agenda for this subject area, despite not being the focus of the work presented here.

This book does not treat the estimation of the recovery of debt obligations in the event of default, a separate and important topic. Zhang (2009) provides an empirical model of corporate default recovery risk.

Statistical foundations are presented in a stand-alone series of chapters. A separate series of chapters contains the substantive empirical results. Neither of these two series of chapters needs to be read in order to obtain the thrust of the other.

1.2 STATISTICAL FOUNDATION CHAPTERS

Chapters 2, 3, and 6 provide a mathematical foundation for modeling and estimating default events with stochastic intensities. These chapters can be skipped by readers interested mainly in empirical content.

Chapter 2 provides the mathematical foundations for modeling the arrival of events with a stochastic intensity. The intensity of an event such as default is its conditional mean arrival rate λ_t, measured in events per year, given all information currently available to the observer. Under the doubly-stochastic assumption that we sometimes adopt, the probability of survival for t years is $E(e^{-\int_0^t \lambda(s)\, ds})$. Chapter 2 also presents the multi-firm version of the doubly-stochastic hypothesis, under which the sole source of default correlation between two firms is the dependence of their default intensities on common or correlated observable risk factors. The doubly-stochastic property rules out contagion as well as correlation induced by unobservable risk factors. This chapter includes an approach developed by Das, Duffie, Kapadia, and Saita (2007) for testing a model of the default intensity processes of a large number of borrowers.

Chapter 3 presents the theory underlying the maximum likelihood estimation of term structures of survival probabilities, including the dependence of default probability on time horizon. The methodology allows the events of concern to be censored by the disappearance of corporations from the data, due for instance to merger or acquisition. The idea is to estimate the parameter vector β determining the default intensity $\lambda_t = \Lambda(X_t, \beta)$ as well as the parameter vector γ determining the transition probabilities of the covariate process X, and then to use the maximum likelihood estimator of (β, γ) to estimate the survival probability $E(e^{-\int_0^t \lambda(s)\, ds})$, for a range of choices of the survival horizon t. The approach, from Duffie et al. (2007), allows the joint estimation of (β, γ) in a relatively tractable manner under the doubly-stochastic property by decomposing the problem into separate estimations of β and for γ.

Chapter 6 presents the foundations for frailty modeling of correlated default in a setting of stochastic intensities. The approach is to assume that default times are jointly doubly stochastic given extra information unavailable to the econometrician. This "hidden" information includes covariates that, although not directly observable, have conditional probability distributions that can be filtered from histories of default times and observable covariates. The dependence of default timing on unobservable covariates allows for sources of default correlation beyond those present in the observed covariates. The methodology relies on Markov Chain Monte Carlo (MCMC) techniques, provided in appendices, for evaluating likelihood functions and for filtering or smoothing hidden (frailty) state information.

1.3 SCOPE OF EMPIRICAL CHAPTERS

Chapters 4, 5, and 7 contain the substantive empirical results for North American non-financial corporations between 1979 and 2005.

Chapter 4 presents an estimated dynamic model of the term structures of conditional default probabilities. The results, based on Duffie, Saita, and Wang (2007) and Duffie, Eckner, Horel, and Saita (2009), show the significant dependence of default probabilities on a firm's distance to default (a volatility-adjusted leverage measure) and, to a lesser extent, on the firm's trailing stock return as well as various macroeconomic variables. In the structural models of Black and Scholes (1973), Merton (1974), Fisher, Heinkel, and Zechner (1989), and Leland (1994), the distance to default is a sufficient statistic for default probabilities. The estimated shape of the term structure of conditional default probabilities reflects the time-series behavior of the covariates, including the mean reversion of macroeconomic performance and leverage targeting by firms. The estimated term structures of default hazard rates are typically upward-sloping at business-cycle peaks and downward-sloping at business-cycle troughs, to a degree that depends on corporate leverage relative to its long-run target. Typical peak-to-trough variation in distances to default have a larger and more persistent impact on default probabilities than does business-cycle variation of the macro-covariates (after controlling for distance to default).

Chapter 5, based on Das, Duffie, Kapadia, and Saita (2007), provides a battery of tests of the ability of the model estimated in Chapter 4 to capture default correlation. Several of these tests are based on a time rescaling by which defaults arrive according to a constant-intensity Poisson process. Additional specification tests from Das, Duffie, Kapadia, and Saita (2007) are found in Appendix C.

The results of Chapter 5 show strong evidence of missing common or correlated default risk factors, some of which may not even have been contemporaneously available. Based on this idea, Chapter 7 provides estimates of a frailty-based model of joint default arrivals, in which default correlation can arise from variables that might have been available to the econometrician but were not included in the model, and also from additional unobservable sources of correlation. The results show substantial dependence of default intensities on common unobservable (or at least un-included) factors whose effects are condensed for modeling purposes into a single dynamic factor, parameterized as an Ornstein–Uhlenbeck frailty process. The estimated parameters governing the mean reversion and volatility of this frailty process, as well as the posterior (filtered) probability distribution of the frailty process, indicate substantial persistence and time variation in unobserved common sources of default risk. Appendices provide extensions of the model that allow for unobserved cross-sectional sources of variation in default intensity and non-linear dependence of a firm's default intensity on its distance to default.

Even after the financial crisis of 2007–2009, traders of structured credit products such as collateralized debt obligations (CDOs) that are directly exposed to default correlation have relied on copula models of default time correlation. As explained in Appendix B, a copula is a simple device for specifying the joint probability distribution of a collection of random variables with given marginal distributions. Despite its advantages for data structuring and for the rapid calculation of multi-firm default probabilities, the copula model is inherently unsuited to portfolio-based risk management and pricing applications, such as CDO pricing and value-at-risk measurement. A key shortcoming of the copula model is that it is unable, even in principle, to capture the risk of changes over time in conditional default probabilities. Models based on correlated intensity processes, although substantially more complicated to use than the industry-standard copula model, are now a sufficiently tractable alternative for many pricing and risk management applications. For example, Eckner (2009) shows how to model the pricing and risk management of CDOs with correlated default intensity processes.

1.4 HISTORICAL RESEARCH DEVELOPMENTS

Altman (1968) and Beaver (1968) were perhaps the first to estimate statistical models of the likelihoods of default of corporations, using financial accounting data. Lane et al. (1986) made an early contribution to the empirical literature on the probability distributions of corporate default times with their work on bank default prediction, using time-independent covariates. Lee and Urrutia (1996) introduced a duration model of default timing based on Weibull distributed default times. Duration models of default with time-varying covariates include those of McDonald and Van de Gucht (1999), who addressed the timing of high-yield bond defaults and call exercises. Duration models were used by

Shumway (2001), Kavvathas (2001), Chava and Jarrow (2004), and Hillegeist, Keating, Cram, and Lundstedt (2004) to predict bankruptcy. Shumway (2001) used a duration model with time-dependent covariates.

Each of these early studies took a "reduced-form" approach, modeling the dependence of default probabilities on explanatory variables through an econometric specification that does not directly model the incentives or ability of the borrower to pay its debt. Some structural models of default timing have the implication that a corporation defaults when its assets drop to a sufficiently low level relative to its liabilities. For example, the models of Black and Scholes (1973), Merton (1974), Fisher, Heinkel, and Zechner (1989), and Leland (1994) model the market value of a firm's assets as a geometric Brownian motion. In these models, a firm's conditional default probability is completely determined by its distance to default, which is the number of standard deviations of annual asset growth by which the asset level (or expected asset level at a given time horizon) exceeds an accounting-based measure of the firm's liabilities. An estimate of this default covariate, using market equity data and accounting data for liabilities, has been adopted in industry practice by Moody's in order to provide estimated probabilities of default for essentially all publicly traded firms. (See Crosbie and Bohn (2002) and Kealhofer (2003).)

In the context of the structural default model of Fisher, Heinkel, and Zechner (1989), Duffie and Lando (2001) modeled the conditional probability distribution of a default time for cases in which a firm's distance to default is imperfectly observed. This model implies the existence of a default intensity process that depends on the currently measured distance to default and on other covariates that may reveal additional information about the firm's condition. More generally, a firm's financial health may have multiple influences over time. For example, firm-specific, sector-wide, and macroeconomic state variables may all influence the evolution of corporate earnings and leverage.

The approach taken here, although not based directly on a structural model of default, is motivated by the structural approach through the inclusion of distance to default as a key covariate, and through the inclusion of additional observable and unobservable default covariates, in an attempt to capture sources of default risk that are not revealed by distance to default.

Duffie, Saita, and Wang (2007) introduced maximum likelihood estimation of term structures of default probabilities based on a joint model of stochastic default intensities and the dynamics of the underlying time-varying covariates. This work was based on the doubly-stochastic assumption, and therefore did not account for unobservable or missing covariates affecting default probabilities. With such incomplete observation, the arrival of a default leads, via Bayes' Rule, to a jump in the conditional distribution of hidden covariates, and therefore a jump in the conditional default probabilities of any other firms whose default intensities depend on the same unobservable covariates. For example, the collapses of Enron and WorldCom could have caused a sudden reduction in the perceived precision of accounting leverage measures of other firms. Collin-Dufresne, Goldstein, and Helwege (2010) and Jorion and Zhang (2007) found that a major credit event at one firm is associated with significant increases in the credit spreads of other firms, consistent with the existence of a frailty effect for actual or risk-neutral default probabilities. Collin-Dufresne, Goldstein, and Huggonier (2004), Giesecke (2004), and Schönbucher (2003) explored learning-from-default interpretations, based on the statistical modeling of frailty, under which default intensities include the expected effect of unobservable covariates. Yu (2005) found empirical evidence that, other things equal, a reduction in the measured precision of accounting variables is associated with a widening of credit spreads.

Delloy, Fermanian, and Sbai (2005) and Koopman, Lucas, and Monteiro (2008) introduced dynamic frailty models of default based on observations of credit ratings for each firm, and assuming that the intensities of changes from one rating to another depend on a common unobservable factor. Because credit ratings are incomplete and lagging indicators of credit quality, as shown for example by Lando and Skødeberg (2002), one would expect to find substantial frailty in ratings-based models such as these. Duffie, Eckner, Horel, and Saita (2009) extended the frailty-based approach to incorporate the variables used by Duffie, Saita, and Wang (2007), and still found substantial sources of frailty-based default correlation. Lando and Nielsen (2009) recently augmented the list of covariates used by Duffie, Eckner, Horel, and Saita (2009) with a selection of accounting ratios, and show an improvement in the ability of the model to capture default correlation with observable covariates. Koopman, Lucas, and Schwaab (2010) explored the role of frailty in the default experience of the recent financial crisis. Azizpour and Giesecke (2010) show the presence of frailty-based correlation in a version of the model that allows for the influence of past default events. Although further improvements in model structure and covariate information are likely, it seems prudent when estimating the likelihood of large portfolio default losses to allow for unobserved sources of default correlation. The financial crisis of 2007–2009 was a severe lesson about potential sources of joint default risk that are not easily observed or captured with simple models.