

Market-based Credit Ratings*

Drew D. Creal Robert B. Gramacy Ruey S. Tsay

Booth School of Business, University of Chicago

June 7, 2012

Abstract

We present a methodology for rating the creditworthiness of public companies in the U.S. from the prices of traded assets. Our approach uses asset pricing data to impute a term structure of risk neutral survival functions or default probabilities. Firms are then clustered into ratings categories based on their survival functions using a functional clustering algorithm. This allows all public firms whose assets are traded to be directly rated by market participants. For firms whose assets are not traded, we show how they can be indirectly rated through the use of matching estimators. We also show how the resulting ratings can be used to construct loss distributions for portfolios of bonds. Our approach has the advantages of being transparent, computationally tractable, simple to implement, and easy to interpret economically.

Keywords: credit ratings; clustering; credit default swaps; default risk; survival functions

*We would like to thank Alan Bester, Chris Hansen, André Lucas, Leonard Nakamura, James Nason, Bernd Schwaab as well as seminar participants at the Federal Reserve Bank of Philadelphia and the Society for Financial Econometrics meetings for their helpful comments. Correspondence: Drew D. Creal, University of Chicago Booth School of Business, 5807 South Woodlawn Avenue, Chicago, IL 60637. Email: dcreal@chicagobooth.edu

1 Introduction

Credit ratings are a summary of a firm's expected future creditworthiness. They are commonly used to quantify the credit risk for large institutions and they play an integral role in the operations of the global financial system. For example, the Basel Accord uses credit ratings as a measure of potential losses and makes credit ratings a determinant of capital requirements for financial institutions. Credit ratings also help determine what positions financial institutions such as hedge funds and mutual funds are legally allowed to hold through contractual obligations. Financial institutions may build their own ratings system or they can use the ratings systems of independent agencies (e.g. Moody's, Standard & Poors, and Fitch).

We propose a methodology that allows the prices of traded assets to directly determine the credit ratings for publicly traded firms with a minimal amount of subjective input on a firm-by-firm basis. We call these *market-based credit ratings*. Our methodology uses the information available in credit markets to construct credit ratings in a transparent manner. The essential element is to specify a function that maps the observed value of firms' credit risky assets into a rating category. This function enables any public firm that is currently traded to be rated.

We begin by converting the observed prices of corporate bonds and credit derivatives into survival functions, which equal one minus the cumulative probability of default on or before a given future date. To construct market-based credit ratings, we focus directly on the survival function because it is the object that is common for traded, credit risky assets. A credit default swap (CDS) is a financial agreement between a buyer and seller that is similar to insurance and is intended to protect the buyer of the contract in case a bond or loan defaults. The buyer of the CDS contract makes periodic payments known as premiums to the seller and, in exchange, receives a payoff from the seller if the reference entity defaults on a loan or bond before the CDS contract matures. The reference entity is the organization on which the CDS contract is written and may be a municipal, state, or sovereign government or a public or private corporation. The periodic payment that the buyer makes to the seller is quoted in terms of a spread. Spreads are higher for entities who the market perceives to have either higher probabilities of default or whose default has worse consequences, leading to a correspondingly greater insurance cost.

In recent years, independent ratings agencies Moody's and Standard & Poors also allow financial institutions to purchase either ratings implied by market prices (S&P) or CDS-implied

default probabilities (Moody’s). The ratings agencies therefore do provide a form of “market-based” rating. But the methods used in those market-based ratings are not openly available. In contrast, the emphasis in this paper is developing a method that is transparent and whose function mapping observed asset prices into ratings is publicly known. Furthermore, our methodology will provide a market-based rating for those firms that do not have bonds/CDS traded on the market.

We take market prices and transform them into market implied survival probabilities using standard asset pricing formulas. The survival functions are then clustered into ratings categories using a functional clustering algorithm. An important contribution of this paper is the construction of an appropriate mixture model for clustering firms into ratings categories based on their survival functions. Our mixture model simultaneously respects economic constraints and leads to an inferential procedure that is easy to implement and computationally tractable on large data sets.

The output of the clustering algorithm allows us to define two types of ratings: *absolute* and *relative*. In the absolute ratings system, the definition of each rating category is time-invariant meaning that each firm’s creditworthiness is compared to a pre-specified target. In contrast to many agency-issued credit ratings, the absolute ratings have a publicly known point of reference that does not change through the credit cycle. During a credit crisis when short-term lending to businesses by banks dries up, it may be that few if any firms are able to attain the highest absolute rating. Relative ratings, by contrast, compare the creditworthiness of an individual firm to other firms at that point in time. For example, a firm’s creditworthiness may decline considerably during a recession but they may not decline as much as other firms. In such a case, a firm’s relative rating may increase while its absolute rating declines.

The next contribution of this paper is to develop a method to rate public firms whose assets are not actively traded using firms whose assets are. This amounts to solving a counterfactual. How would the market rate a firm if its assets were traded? Intuitively, two firms who are identical in terms of observable characteristics but only one of which is traded should receive the same or comparable credit ratings. We solve the counterfactual problem using matching estimators. Specifically, we use observable characteristics to match firms for which we do not observe debt or credit derivatives data with other firms for which we do have such data.

We implement our methodology using a large database of daily CDS spreads from 2005

through 2011 along with the London Interbank Offered Rate (LIBOR) deposits and interest rate swaps, which we convert to survival functions. There are many arguments in the literature as to why these securities are informative about default probabilities. Several papers have documented that CDS spreads lead changes in bond yields and credit ratings, e.g. Hull, Predescu, and White (2004) and Blanco, Brennan, and Marsh (2005). Furthermore, CDS contracts for many reference entities are more liquid than corporate bonds as there is little initial cost to enter into a CDS contract. In fact, prior to April 2009, it was a convention in the CDS market that there was no initial cost, making it easily accessible. In addition, CDS spreads are not affected by as many detailed provisions or features associated with corporate bonds such as seniority, callability, etc.. As such, CDS spreads are relatively more homogeneous than bonds across firms. As a practical matter, databases on CDS spreads are more readily available and easier to manipulate than bond trades. Note, however, that because our methodology works with survival functions directly, it is simple to use corporate bond data instead of, or in addition to, CDS data if it is available.

Finally, we describe how our methods can be used to construct loss-distributions, which are important in practice for credit risk management. Over the past decade, a large number of new models and methods have been proposed for constructing loss distributions, e.g. Duffie, Saita, and Wang (2007), Koopman and Lucas (2008), and Duffie, Eckner, Horel, and Saita (2009). Our approach both complements and leverages previous research. We combine the database on CDS spreads with data from Compustat and CRSP which provides information on bankruptcies and defaults. By making default a category to which a firm can transition, we show how one can construct loss distributions directly from the mixture model used to build the ratings. Alternatively, the new ratings can be used inside models for credit ratings transitions, e.g. Gagliardini and Gouriéroux (2005), Koopman, Lucas, and Monteiro (2008), Creal, Schwaab, Koopman, and Lucas (2011) and Creal, Koopman, and Lucas (2012).

In using our method, one should clearly ask whether, and to what extent, securities prices are informative of the underlying firm's credit quality. The answer depends on a number of factors, most notably liquidity. Our ratings are built on a weekly basis from daily data so that day-to-day fluctuations in liquidity of the market should have a smaller impact. Using risk-neutral survival functions to construct ratings also relies on the assumption that observed spreads are not wildly distorted from fundamental values of the firms. Information does not

flow through over-the-counter markets the way that it does through equity markets; see, e.g. Duffie (2012). Importantly, our paper does *not* argue that credit markets are efficient. Rather, a methodology that determines ratings directly from the data is a useful source of information relative to other sources, e.g. ratings agencies. Moreover, it can be transparent since the CDS data are readily available, the survival functions easy to calculate, and the clustering algorithm standard.

The rest of this paper proceeds as follows. In Section 2, we discuss how ratings are constructed from observed asset pricing data both for firms that are traded and for firms that are not. This section takes the risk neutral survival functions as given. In Section 3, we describe the details of the data and how we construct the survival functions from an asset pricing model. In Section 4, we compare our ratings to those of independent agencies and describe how our ratings can be used to construct loss functions for portfolios of bonds. We conclude and discuss some extensions in Section 5.

2 Classification methodologies

Section 2.1 defines a mixture model for survival functions and describes how to estimate the unknown parameters of the mixture model. In Sections 2.2 and 2.3, we show how to use the resulting parameter estimates, together with the original data, to produce two sets of credit ratings with distinct economic interpretations. Section 2.4 extends this methodology to firms whose assets are not traded using a matching estimator.

2.1 Rating firms whose assets are traded

In this Section, we describe a methodology for rating firms whose credit-risky assets have been traded. We let $S_{it}^Q(\tau)$ denote the survival function under the risk-neutral measure Q for firm i at time t with maturity τ . Further details about the data and the procedure used to construct $S_{it}^Q(\tau)$ are discussed in Section 3. At the moment, we take the vector $S_{it}^Q(\tau)$ as given for $t = 1, \dots, T$. The number of firms for which some assets are traded in period t is N_t and for each firm a maximum of J points on the survival function are observed, i.e. some points on the survival function may be missing for some firms. Elements of the survival functions are ordered

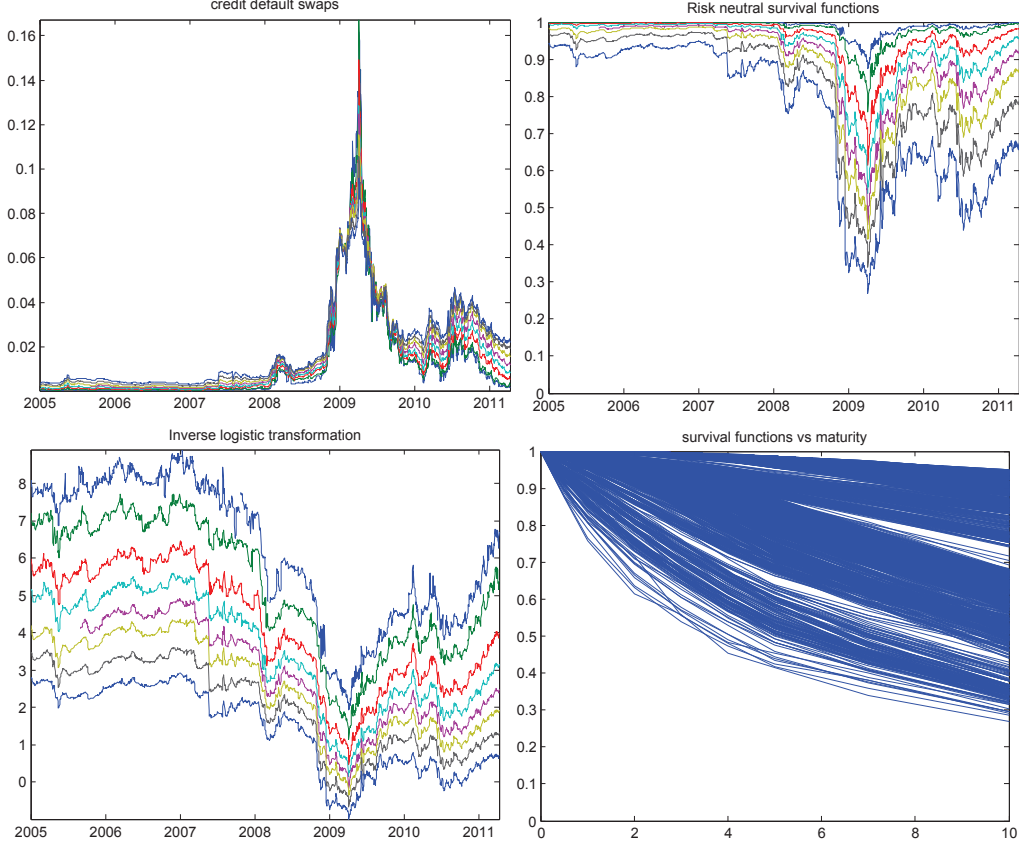


Figure 1: Data for the Alcoa Corporation from January 2005 through February 2011. Upper left: CDS spreads for seven different maturities $\tau = (0.5, 1, 2, 3, 4, 5, 7, 10)$ measured in years; Upper right: market-implied survival functions $S_{it}^Q(\tau)$; Lower left: transformed functions $y_{it}(\tau)$; Lower right: survival functions $S_{it}^Q(\tau)$ as a function of maturity τ for all time periods.

in the vector in terms of a set of increasing maturities $\tau = (\tau_1, \tau_2, \dots, \tau_J)'$. In addition to the survival functions, we also observe a default indicator D_{it} which is equal to 1 if firm i defaults in period t and zero otherwise.

We propose to construct ratings categories for traded firms by clustering a transformation of their survival functions $y_{it}(\tau) = \log \left(S_{it}^Q(\tau) / (1 - S_{it}^Q(\tau)) \right)$. Clustering the entire survival function instead of only a portion of it (e.g. only 5-year default probabilities or the level, slope, and curvature), allows the data to determine which characteristics of the survival function are important. The reason for transforming the survival functions is to make the data closer to being normally distributed and therefore easier to cluster. Figure 1 plots the raw and transformed data for the Alcoa Corporation. The figure illustrates the rapid increase in CDS spreads during

the recent debt crisis as the price for buying protection against the default of Alcoa increased. Increases in CDS spreads imply an increase in the (risk-neutral) default probability of the firm as seen by the survival probabilities during this period.

To cluster the transformed survival functions, we take a model-based clustering approach using mixtures of Gaussian processes. Consider the following finite mixture model

$$y_{it}(\tau) \sim \sum_{k=0}^K \pi_{it,k} \cdot p_N(y|\mu_k(\tau), \Sigma_k(\tau), D_t), \quad \pi_{it,k} = \sum_{\ell=0}^K \pi_{it,\ell k}, \quad (1)$$

$$\pi_{it,\ell k} = P(I_{it} = k | I_{i,t-1} = \ell; Y_{it}), \quad \forall \ell = 1, \dots, K, \quad \forall k = 0, \dots, K \quad (2)$$

$$\pi_{it,00} = P(I_{it} = 0 | I_{i,t-1} = 0; Y_{it}) = 1, \quad (3)$$

where I_{it} is a (latent) indicator function taking values from $k = 0, \dots, K$ while $p_N(y|\mu, \Sigma)$ denotes the multivariate normal distribution with mean μ and covariance matrix Σ . The probabilities π are defined below. The model in Equations (1)-(3) defines a probability distribution over the space of survival functions. We define state $k = 0$ to be default and the largest category K to be the best rating category. While realizations $y_{it}(\tau)$ are vectors of length J , the unknown parameters μ , Σ , and π do not depend on J . Among other things, this allows for unbalanced panels where the vector τ may vary across i and t . It also helps us to cope with missing data, which we encounter in our empirical application.

The probability of transitioning from category ℓ to category k for firm i at time t is $\pi_{it,\ell k}$. The transition probabilities may depend on observable firm, industry, or macroeconomic variables Y_{it} . For state $\ell = 0$ when $D_{it} = 1$, the normal distribution is degenerate with all its mass at the mean (such that $S_{it}^Q = 0$). Equation (3) ensures that default is an absorbing state and once a firm transitions into default it cannot transition out. If a firm transitions out of default, we treat it as a new firm. Combining these results, we let P_{it} denote the matrix of Markov transition probabilities with entries $\pi_{it,\ell k}$ and whose top row is degenerate due to (3).

We let $y_t = (y'_{1t}, y'_{2t}, \dots, y'_{N_t,t})'$, $D_t = (D_{1t}, D_{2t}, \dots, D_{N_t,t})'$, and $I_t = (I_{1t}, I_{2t}, \dots, I_{N_t,t})$ be the stacked vectors of all observations, default indicators, and category indicators. We use $y^t = (y'_1, \dots, y'_t)'$ and $D^t = (D_1, \dots, D_t)'$ to denote the history of all the observable variables up to time t . We also define $\theta^{(k)} = (\mu_k, \text{vech}(\Sigma_k))$ and $\theta = (\theta'_1, \theta'_2, \dots, \theta'_K)'$ to be vectors containing the parameters of the model, where $\text{vech}(\Sigma_k)$ denotes the half-stacking vector of the symmetric

matrix Σ_k . Throughout the paper, the number of categories K is treated as specified by the practitioner. Typical values for K are in the range $8 - 10$, in keeping with conventions set by Moody's, S&P, and Fitch.

To identify the model statistically and to give an economic interpretation to the categories, structure needs to be imposed on the mean vector and covariance matrices. For any particular rating category $k > 0$, the mean vector μ_k is a decreasing function of maturity, i.e. $\mu_k(\tau_1) > \mu_k(\tau_2) > \dots > \mu_k(\tau_J)$ with $\tau_{j-1} < \tau_j$ for $j = 2, \dots, J$. This stems both from the definition of the survival functions and the absence of arbitrage. To assign an ordering to the non-default states for $k = 1, \dots, K$, we also need to impose the fact that better ratings categories have higher survival functions. Roughly, this means that for any maturity τ_j and neighboring categories $k - 1$ and k , we have $\mu_{k-1}(\tau_j) < \mu_k(\tau_j)$. We explain how we impose these restrictions below.

There exists some freedom when determining the parameters θ of the model. One can a priori fix the values of the parameters and treat them as choice variables. For example, the parameters of the mean vector could be chosen to match the values of a representative company's survival functions at key historical times, e.g. they could represent American International Group (AIG), Johnson & Johnson, or Toy's R Us in a good or bad year. Then, the ratings would be endogenously determined by comparing other firms' survival functions to these benchmarks. At the other extreme, the parameters of the mixture model can be freely estimated subject to the restrictions discussed above.

In practice, even for small values of K (the rating categories) and J (the number of maturities), the parameter set can be prohibitively large both computationally and from an identification perspective. It is therefore sensible to impose additional restrictions on the model. We strike a middle ground between the aforementioned two extremes and estimate some parameters of the model while imposing further restrictions. For the simplest model, the covariance matrices are restricted to be the same across ratings categories. The mean vectors μ_1, \dots, μ_K are also restricted such that they are collectively only a function of a small set of parameters. Specifically, we allow the mean vector of the highest rated category μ_K to be freely estimated. Then, each additional mean vector is a deterministic decreasing function of the vector before it, i.e. $\mu_{k-1}(\tau_j) = \mu_k(\tau_j) - \eta$, with η being a choice variable that depends on K . Below, we set $\eta = 1$. This ensures that the components of the mixture are well-separated and cover the full support of the space of survival functions.

With these restrictions in place, the mixture model can be estimated by a number of methods. Techniques for estimating mixture models are well-developed in the literature; see, e.g. McLachlan and Peel (2000) and Frühwirth-Schnatter (2006). We employ Bayesian methods and estimate the model by Gibbs sampling using data augmentation. Our priors for the parameters of the model are conditionally conjugate and draws from the full conditional distributions with the Gibbs sampler are mostly standard. For the mean vector, we use a constrained multivariate normal distribution where the constraints ensure the ordering $\mu_K(\tau_1) > \mu_K(\tau_2) > \dots > \mu_K(\tau_J)$ with $\tau_{j-1} < \tau_j$ for $j = 2, \dots, J$. For the covariance matrix Σ , we use an inverse Wishart distribution. The prior distribution for each row of probabilities in the Markov transition matrix is a Dirichlet distribution. Choices for the prior hyperparameters are discussed below in the empirical section. In practice, some maturities of the observation vector $y_{it}(\tau)$ are missing. We impute the missing values as part of our Gibbs sampling algorithm. We let $\tilde{y}_{it}(\tau_j)$ denote the imputed value for maturity τ_j and we let $y_{c,it}(\tau)$ denote the vector of complete data that combines the observed and imputed values.

The Gibbs sampler for this model cycles through five individual steps; drawing the missing components of the observation vector $\{\tilde{y}_{it}(\tau_j)\}_{i=1,t=1}^{\tilde{N}_{it,T}}$, the latent state indicators (I_0, I_1, \dots, I_T) , the transition probabilities $\pi_{\ell,k}$, the mean vector μ_K , and the covariance matrix Σ . Details on the Gibbs sampling algorithm for the model are available in the appendix as many of the steps are standard. However, we briefly mention some aspects that are slightly non-standard. Conditional on the parameters θ of the model, the latent indicators are drawn in one-block using a forward-filtering, backward sampling algorithm for Markov-switching models; see, e.g. Frühwirth-Schnatter (2006, p. 342). The only non-standard part of the implementation is handling the degenerate normal distribution when there is a default. In this case, the conditional likelihood is infinite when $D_{it} = 1$ and zero when $D_{it} = 0$. This ensures that the algorithm always draws $I_{it} = 0$ when $D_{it} = 1$ and never draws this state otherwise.

Conditional on the state indicators, the observed data, and the remaining parameters of the model, draws for the missing observations as well as draws for μ_K are from a constrained multivariate normal distribution. Throughout the Gibbs sampling algorithm, the vector of completed observables $\{y_{c,it}(\tau_j)\}_{i=1,t=1}^{N_{it,T}}$ and the components of the mean vector must satisfy the ordering constraints to ensure that they are decreasing functions of maturity. In our work, we use the Gibbs sampler for constrained multivariate normal distributions from Rodriguez-Yam,

Davis, and Scharf (2004).

Since the mixture model is a Markov-switching model, we can use well-known algorithms for computing the prediction $p(I_t|y^{t-1}, D^{t-1}; \theta)$, filtering $p(I_t|y^t, D^t; \theta)$, and smoothing distributions $p(I_t|y^T, D^T; \theta)$ which are the probability that a firm is in rating category k conditional on different information sets, e.g. Hamilton (1989) and Frühwirth-Schnatter (2006). Here we use the notation x^t to denote $\{x_1, \dots, x_t\}$. This enables one to quantify the uncertainty associated with a firm's rating and to obtain forecasts of future ratings transitions. The model-based clustering approach advocated here is also related to discriminant analysis; see, e.g. Hastie, Tibshirani, and Friedman (2009). Consider the marginal filtering distribution,

$$P(I_{it} = k|y^t, D^t; \theta) = \frac{p(y_{it}|\mu_k, \Sigma_k, D^t)P(I_{it} = k|y^{t-1}, D^{t-1}; \theta)}{\sum_{\ell=0}^K P(y_{it}|\mu_\ell, \Sigma_\ell, D^t)P(I_{it} = \ell|y^{t-1}, D^{t-1}; \theta)}.$$

Traditional linear discriminant analysis is the special case when $\Sigma_k = \Sigma$ for all k and the priors over categories are time-invariant.

In the case of a Markov-switching model, the priors over each component depend on the history of a company's past ratings. The conditional decision boundary at time t between any two ratings categories k and ℓ is

$$\begin{aligned} \log \frac{P(I_{it} = k|y^t, D^t; \theta)}{P(I_{it} = \ell|y^t, D^t; \theta)} &= -\frac{1}{2} (\mu_k + \mu_\ell)' \Sigma^{-1} (\mu_k - \mu_\ell) \\ &\quad + y'_{it} \Sigma^{-1} (\mu_k - \mu_\ell) + \log \frac{p(I_{it} = k|y^{t-1}, D^{t-1}; \theta)}{p(I_{it} = \ell|y^{t-1}, D^{t-1}; \theta)}, \end{aligned}$$

which is a linear function of the current observations y_{it} and a weighted function of past observations with declining weights. The Markov-switching characteristic of the model demonstrates how neighboring observations play in an important role in smoothing the ratings intertemporally. If in the last period there was a high probability of being in category k , then that probability persists into this period. A user can eliminate intertemporal smoothing of ratings by simply replacing the Markov dynamics with time-invariant probabilities for each category.

2.2 Loss functions and ratings

The information from the data about a firm’s credit rating is contained in the joint posterior distribution $p(I_0, \dots, I_T | y^T, D^T; \theta)$. To turn this information into credit ratings, we need to specify a decision rule that maps probabilities into indicators that are one for only one category in a given time period. This mapping is intricately related to a researcher’s loss function. In the following, we will propose specific methods that are computationally tractable and imply a specific loss function. However, we emphasize that given the distribution $p(I_0, \dots, I_T | y^T, D^T; \theta)$ other researchers may wish to use their own loss function. We view this as an advantage of this approach.

2.2.1 Absolute ratings

The first type of market-based ratings we construct are *absolute* ratings. In an absolute ratings system, the definition of each category is time invariant, quantitatively explicit, and public knowledge. The absolute ratings provide a point of reference that is interpretable and comparable across time periods. Specifically, we associate each mixture component, k , with its own ratings category, defined by the function μ_k which determines the mean default probabilities for that category as a function of maturity τ . In this paper, we parameterize the functions μ_k to ensure that ratings categories are well separated in terms of their average default probabilities.

To construct our absolute ratings, we use the maximum a posteriori (MAP) estimator, which is the set of indicators I_0, \dots, I_T that maximizes the joint distribution $p(I_0, \dots, I_T | y^T, D^T; \theta)$. This is the optimal decision rule under an absolute error loss function, and can be conveniently computed by the Viterbi (1967) algorithm. The MAP estimator for the joint distribution is different than the MAP estimator for the marginal distribution as it creates intertemporal smoothing of the ratings. Importantly, in order to mimic the information set of an investor who at time t does not observe future asset prices, we run the joint MAP estimator when each new observation becomes available taking the final period’s estimate as the rating for that period. Therefore, ratings at time t will only depend on information up to that time period.

A feature of the absolute ratings system is that a firm’s credit rating does not directly depend on the rating of other firms. The number of firms in any given rating category could be zero at some points in time. This possibility makes sense. If credit ratings are a signal

of real-world default probabilities, then during a credit crisis like 2008-2009 there should be very few firms with the best absolute rating. If the credit market freezes and the willingness for banks to provide short-term lending to businesses deteriorates, this would affect all firms. Conversely, a massive expansion of credit through a combination of lax regulation and perverse incentives in the credit market could lead to an increase in lending to lower quality borrowers. Both of these scenarios would be observable in the time variation of the empirical distribution of the absolute ratings system.

2.2.2 Relative ratings

Another useful ratings system that can be defined from the output of the clustering algorithm are *relative* ratings. In this system, a firm's creditworthiness is only defined relative to other firms in that time period. The relative ratings define an ordinal ranking of firms' credit quality. Given the information in the posterior marginal (filtering) distribution, there are many ways of sorting firms into K^* ordered categories, where K^* may differ from the value of K used to estimate the model. Each sorting scheme corresponds to a loss function. We propose to sort companies into K^* -tiles according to their expected values from the marginal filtering distribution. Specifically, we compute the expected value for each firm

$$\bar{I}_{it} = \sum_{k=0}^K k \cdot p(I_{it} = k | y^t, D^t; \theta),$$

where $p(I_{it} = k | y^t, D^t; \theta)$ is the filtered probability. In practice, we marginalize the parameters θ out of this distribution during the MCMC algorithm. Companies can then be sorted into K^* different categories by sorting their expected values. By sorting based on the posterior means, we are implicitly using a symmetric loss function as we are weighting the posterior categories by their probabilities.

Relative ratings have the potential to behave differently than the absolute ratings. For example, it may be the case that during a recession or credit crunch a firm's overall creditworthiness declines considerably but it may not decline as much or as fast as other firms causing its relative rating to increase rather than decrease. This is despite the fact that its absolute rating is declining. The relative ratings for an individual firm can be compared across time

periods to see how a firm is performing relative to its peers either overall or within an industry. The empirical distribution of the relative ratings of companies in an industry or sub-sector of the economy also provides a simple indicator of that sector's relative performance.

2.3 Rating firms whose assets are not traded

When a firm's debt and credit derivatives are not actively traded, financial markets do not provide a direct signal of a firm's credit quality. A firm's credit rating consequently cannot be directly inferred from the transactions of market participants. Instead, one must infer the market's rating for firms that are not traded using the market prices for firms whose assets are traded. The thought experiment is as follows. Imagine two firms that are identical to one another in every way except for the fact that one firm's assets are traded and the other is not. These firms should have the same or similar credit quality and share the same rating.

The idea of comparing similar economic agents (in our case firms) in two states of the world is a cornerstone of the microeconometrics and statistics literature on estimation of causal effects (treatment effects); see, e.g. Angrist and Pischke (2009) and Wooldridge (2010). Imputing either the rating category or the propensity to be in a rating category is a natural extension of the literature in credit risk management that models the propensity to default. The literature on default estimation and credit risk management does not explicitly use the language of treatment effects and matching estimators. However, all the models and estimators in the literature can be interpreted as matching estimators.

Consider the simplest case where we observe for each firm a sequence of default or bankruptcy indicators and a vector of covariates. Traditional analysis of these data fit a probit or logit model to the default indicators; see, e.g. McNeil, Frey, and Embrechts (2005). Firms have either received the treatment (default) or have not received the treatment (non-default). Firms that have not defaulted but, according to the covariates look similar to firms that have, will have a large default probability. Conditional on the ratings produced in Sections 2.2.1 and 2.2.2, the ratings for non-traded firms can be imputed by matching them to traded firms based on observable characteristics.

Characteristics used for matching need to be publicly observable although it is possible that regulators have access to proprietary data sources which could be used as additional

information. Our primary resource for characteristics will be data from accounting statements. Details of the characteristics used in this paper are discussed in Section 3. We assume for the moment that we have a vector of characteristics X_{jt} for each firm that is traded and a vector of characteristics \tilde{X}_{it} for each firm i that is not traded, both at time t . These vectors can be separated into sub-vectors of discrete and continuous variables, $X_{jt} = (X_{jt}^{d'}, X_{jt}^{c'})'$ and $\tilde{X}_{it} = (\tilde{X}_{it}^{d'}, \tilde{X}_{it}^{c'})'$.

There are many possible parametric and non-parametric methods that could be employed to produce matches. We use a non-parametric matching estimator from Abadie and Imbens (2006); see also Guo and Fraser (2009). Consider constructing a match for an individual firm i at time t whose assets are not traded. First, we find the set of traded firms that share any common discrete characteristics, e.g. the same industry code. This causes the firm to be matched exactly along this dimension. From this set we choose the firm j that is closest to firm i according to the distance measure d_{ij} defined by

$$d_{ij} = \left(\tilde{X}_{it}^c - X_{jt}^c \right)' V \left(\tilde{X}_{it}^c - X_{jt}^c \right) \quad (4)$$

where V is a positive definite matrix. There are two common choices for the matrix V . The first is the inverse of the sample variance matrix and the second is the inverse of the sample covariance matrix in which case the distance measure (4) is the Mahalanobis distance. In our empirical work below, we use the inverse of the sample variance matrix.

This matching estimator has the advantage of being simple to implement, computationally fast, and it easily handles missing values in the characteristic space \tilde{X}_{it} . With a fully non-parametric estimator, it is also possible to identify both the set of firms that act as a match and exactly which firm is the final match. Looking at which firms are matched together can provide valuable information when thinking about the plausibility of the results.

Using matching estimators requires making an overlap or common support assumption. This assumption states that for each entity that did not receive the treatment, its observable characteristics \tilde{X}_{it} must be similar to the observable characteristics of a firm that did receive the treatment. Otherwise, a sensible match cannot be constructed and we cannot infer from market data what the rating should be. In the context of market-based ratings, we view this as a feature of the approach. It is practically inevitable that there will exist firms that have no

common characteristics with other firms. We would like to identify these firms. Clearly, their ratings will be based entirely on prior beliefs or private information.

3 Data

3.1 Data sources

We use a large database of daily single entity credit default swap (CDS) spreads to construct the firm level survival functions. The data are from Markit Corporation and include foreign and U.S. corporations as well as sovereigns, U.S. states, and U.S. county reference entities. The data covers CDS contracts with four different types of restructuring provisions and with yearly maturities $\tau_j \in \{0.5, 1, 2, 3, 4, 5, 7, 10\}$ some of which may be missing. The data are therefore in the form of an unbalanced panel. For simplicity, we restrict attention to CDS contracts for U.S. corporations in U.S. dollars for senior subordinated debt. And, we consider only CDS spreads under the modified restructuring (MR) contracts because the MR clause is the most commonly traded contract in the U.S. market.

After making these selections, the CDS data includes 1577 U.S. corporations over the period January 3rd, 2005 through February 28th, 2011 making for a total of 1606 days. Of these firms, 1048 are public corporations, 310 are subsidiaries, and 219 are private firms. In some cases, firms have merged or been acquired or have gone public or private. Data on mergers/acquisitions is available from the Center for Research in Security Prices (CRSP), which is an additional data source discussed below. When any of these actions occurs, we treat the firm as a new reference entity. Finally, we build our ratings on a weekly basis (Wednesdays) from the daily data making for a total of 321 weeks over this time period. This limits the influence of daily liquidity in the CDS market on the resulting ratings.

We combine the CDS data with data on defaults, bankruptcies, and credit events, which are collected from three sources. Over the period of our sample, we collect all defaults from the Standard & Poors Ratings Express database for entity ratings, all bankruptcies from the CRSP database stock event files, and all credit events that triggered payments on CDS from the Markit Corporation. The formal definitions of default, bankruptcy, and credit events are not homogenous. Therefore, it is possible for a firm to exhibit any combination of the three events.

Due to the rarity of these events, we will label any one of the three events as a “default.” If more than one of these events occur for a specific firm, the events may not occur on the same date and we take the earliest of the events. The first (and most common) event is typically a default or missed interest payment in the Standard & Poors database followed by a bankruptcy in the CRSP event files and finally a credit event that triggers the payment clause in a CDS contract. For example, a firm may miss an interest payment on a loan or bond causing Standard & Poors to label it a default. However, the firm’s stock may continue trading and the firm may never declare bankruptcy. If the firm does file for Chapter 7 or 11 bankruptcy and has publically traded equity on an exchange, the bankruptcy will appear in the CRSP stock event files when the firm becomes delisted from the exchange. Finally, if the economic event triggered the CDS, the credit event typically follows a few weeks later. The reason that credit events lag the other events is because it takes a review board several weeks to rule on whether a firm did trigger the default clause in the CDS contract; see e.g. Markit (2009).

We invert a CDS pricing formula to calculate the risk neutral survival functions $S_{it}^Q(\tau)$ following the procedure described in O’Kane (2008). This is a procedure in the finance industry called “bootstrapping” and is part of a standard method for marking to market a CDS contract.[†] Since the procedure is standard in finance and available in O’Kane (2008), we leave further details to an online appendix. Importantly, the methodology we propose in this paper treats survival functions S^Q as data and does not otherwise depend on the asset pricing model one uses to extract them from observed market prices. In practice, a researcher may obtain default probabilities using whichever pricing model she prefers and still apply our methodology to obtain credit ratings. Models used for pricing CDS contracts can be found in Duffie and Singleton (2003), McNeil, Frey, and Embrechts (2005) and O’Kane (2008).

Asset pricing formulas for CDS spreads require two additional sources of data. The first are the daily discount factor curves $z_t(\tau)$, or zero-coupon bond rates, which describe the market value today of a dollar delivered on day $t + \tau$. We follow market convention for marking assets to market and construct the discount curves from daily interest rate and credit derivatives data. We use overnight and two-day LIBOR rates, monthly LIBOR rates with maturities from 1 to 12 months, and interest rate swaps with maturities 1 year to 10 years. The data allows us to construct the LIBOR discount curve beginning at day zero and extending to a maturity of ten

[†]This type of bootstrap should not be confused with the bootstrap used in statistics.

years. Both data sets were obtained from Bloomberg.

The second input in CDS pricing formulas is an assumption about the loss-given default (LGD), which is the percentage of principal that an investor loses conditional on default by a firm. This may vary from firm to firm and over the business-cycle. Data on LGD's is relatively scarce as no financial instruments are actively traded that are functions of only the LGD. When a credit event occurs and a CDS gets triggered, an auction is made to sell the residual value of any bonds. The information on LGD from the auctions is publicly available from Markit but there have only been 52 credit events for U.S. corporations over the course of our sample. Therefore, it has been a market convention to assume a value of 60% based on historical estimates collected by Moody's and Standard & Poors. As a standard procedure in the CDS market, the Markit corporation provides LGD estimates which are used by market participants to provide quotes. We use these values in our work and if they are not available we use 60% as a default. Whether or not these are the true LGDs, the resulting survival functions are consistent with how the market quotes spreads. As a practical matter, CDS spreads are not that sensitive to recovery rate assumptions for a large range of values of LGD; see, e.g. O'Kane (2008).

3.2 Data for the matching estimators

To determine which firms have the potential to be rated, we download the entire Compustat and CRSP database from 2005 through 2011. After eliminating the U.S. operations of foreign firms and non-public financial firms that do not disclose accounting statements, there are 6933 firms reporting earnings over the period. Of these, 1048 are public firms for which CDS trades are observed during the sample. This leaves 5885 firms whose ratings need to be imputed.

We obtain characteristics for use in the matching estimators from accounting data available from Compustat and CRSP. The variables we use are taken from Altman, Fargher, and Kalotay (2010) who find evidence that these characteristics help predict the variation in risk-neutral default probabilities from the Merton (1974) model. These variables include two measures of firm specific performance: the ratio of total earnings to total assets and the ratio of retained earnings to total assets. We use two measures to quantify leverage. The first is the ratio of total assets to total liabilities and the second is accounting leverage, which is total assets divided by

total assets minus total liabilities. The debt maturity structure of a firm is measured by the log-ratio of current liabilities to non-current liabilities. We measure relative liquidity of assets by the ratio of working capital to total assets. We also add the four digit Standard Industrial Classification (SIC) code as our industry variable, although the Global Industrial Classification (GIC) codes would be another alternative. Further details on these variables are available in the data appendix.

In theory, characteristics used for matching may be different across industries either due to data availability or to recognize the conventional wisdom that some variables are more relevant in some sectors of the economy than others. For example, banks have different accounting standards than non-banks and they do not report working capital or current liabilities. Credit risk for financial institutions is often associated with the composition of their balance sheets and the liquidity of their assets. Banks do not publicly release detailed information but they are required to release more information than non-financial institutions.

In practice, when implementing the non-parametric matching estimator, we start by using the four-digit SIC code and, if less than five matches exist at the four digit code, we work backwards to the three digit SIC code. If there are less than five matches at the three digit code, we then work back to the two digit code, etc. Throughout our work, we time the arrival of the covariates to coincide with the day earnings are reported by each firm. Our goal is to mimic the information set available to the market. In between reporting dates, we take the last reported data from the earnings statement as the current value of the characteristic, which implies that X_{it} is a step function.

4 Empirical results

4.1 Ratings

We construct the absolute and relative ratings by running the clustering algorithms outlined in Section 2.1 and the matching estimator from Section 2.3. The prior mean $\underline{\mu}_K$ was chosen as the (transformed) survival function for Johnson & Johnson at the beginning of 2006, when it was considered a safe company. The prior covariance matrix $\underline{\Sigma}$ was selected by first decomposing it as $\underline{\Sigma} = SRS$, where S is a diagonal matrix of standard deviations and R is a correlation

Table 1: Estimation results for the mean vectors of the survival function

	μ_1^S	μ_2^S	μ_3^S	μ_4^S	μ_5^S	μ_6^S	μ_7^S	μ_8^S	μ_9^S
$\tau_{\frac{1}{2}}$	0.7121	0.8705	0.9481	0.9803	0.9926	0.9973	0.9990	0.9996	0.9999
τ_1	0.4978	0.7293	0.8799	0.9522	0.9819	0.9932	0.9975	0.9991	0.9997
τ_2	0.2597	0.4881	0.7216	0.8757	0.9504	0.9811	0.9930	0.9974	0.9990
τ_3	0.1528	0.3289	0.5712	0.7836	0.9078	0.9640	0.9864	0.9950	0.9981
τ_4	0.0985	0.2289	0.4466	0.6869	0.8564	0.9419	0.9778	0.9917	0.9969
τ_5	0.0682	0.1659	0.3510	0.5951	0.7998	0.9157	0.9672	0.9877	0.9954
τ_7	0.0424	0.1074	0.2465	0.4707	0.7074	0.8679	0.9470	0.9798	0.9925
τ_{10}	0.0253	0.0659	0.1608	0.3426	0.5862	0.7938	0.9128	0.9660	0.9872

Estimates of the mean components $\mu_k^S(\tau) = \frac{\exp(\mu_k)}{1+\exp(\mu_k)}$ of the mixture model for $k = 1, \dots, K$ as a function of the maturities $\tau = (0.5, 1, 2, 3, 4, 5, 7, 10)$. The maturities are measured in years. State $k = 1$ is the lowest rated state and state $k = 9$ is the highest rated state.

matrix with diagonal elements equal to one. We set the elements in R immediately above and below the diagonal equal to 0.9, to recognize that neighboring maturities in the mean of the survival function are expected to be highly correlated. The remaining elements then decay slowly to 0.2 for the largest set of maturities. We then set the diagonal elements of S equal to 0.4 for all maturities. The rows of the Markov transition matrix other than default have a Dirichlet ($\underline{\alpha}_k$) distribution with off-diagonal elements $\underline{\alpha}_{\ell,j} = 5$ and diagonal elements $\underline{\alpha}_{\ell,\ell} = 800$. This reflects a belief that ratings will be persistent. Finally, the hyperparameter values for the inverse Wishart distribution were $\underline{\kappa} = 2000$ and $\underline{\Omega}$ had diagonal elements equal to 0.05 with off-diagonal elements equal to zero. For the MCMC algorithm, we take $M = 20000$ draws throwing away an initial portion of the draws as a burn-in.

The mixture model as a whole describes a probability distribution over the space of survival functions with each component placing more mass on subsets of this space. Table 1 contains the estimated mean parameters of the components for categories $k = 1, \dots, K$ after they have been transformed back into probabilities via $\mu_k^S = \frac{\exp(\mu_k)}{1+\exp(\mu_k)}$. As required, the mean survival probabilities all start at one for $\tau = 0$ and decline as a function of maturity. The component distributions are reasonably well-separated. The mean vectors of each of these distributions help quantify the ratings categories in a transparent and explicit way. However, some care needs to be taken when interpreting the results as the probabilities in Table 1 are under the risk-neutral probability measure Q . A firm whose absolute rating remains constant through time does not

Table 2: Estimated matrix of transition probabilities.

$\pi_{\ell,k}$	$k = 0$	$k = 1$	$k = 2$	$k = 3$	$k = 4$	$k = 5$	$k = 6$	$k = 7$	$k = 8$	$k = 9$	$\bar{\pi}$
$\ell = 0$	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0002
$\ell = 1$	0.0071	0.8718	0.0763	0.0157	0.0083	0.0049	0.0038	0.0035	0.0027	0.0059	0.0065
$\ell = 2$	0.0049	0.0571	0.8800	0.0407	0.0038	0.0041	0.0028	0.0022	0.0016	0.0027	0.0106
$\ell = 3$	0.0016	0.0051	0.0248	0.9326	0.0306	0.0016	0.0009	0.0010	0.0007	0.0011	0.0255
$\ell = 4$	0.0003	0.0001	0.0008	0.0107	0.9715	0.0152	0.0004	0.0002	0.0002	0.0002	0.0999
$\ell = 5$	0.0001	0.0001	0.0001	0.0003	0.0101	0.9780	0.0106	0.0002	0.0001	0.0001	0.1952
$\ell = 6$	0.0001	0.0001	0.0001	0.0001	0.0002	0.0097	0.9818	0.0077	0.0001	0.0001	0.2732
$\ell = 7$	0.0000	0.0001	0.0001	0.0001	0.0001	0.0002	0.0108	0.9853	0.0031	0.0001	0.3060
$\ell = 8$	0.0002	0.0003	0.0003	0.0003	0.0002	0.0005	0.0005	0.0174	0.9780	0.0024	0.0769
$\ell = 9$	0.0018	0.0055	0.0029	0.0030	0.0029	0.0057	0.0057	0.0090	0.0245	0.9390	0.0060

Posterior means of the transition probabilities $\pi_{\ell,k} = P(I_t = k | I_{t-1} = \ell)$ of the mixture model for $\ell = 0, \dots, K$ to $k = 0, \dots, K$ with $K = 9$. Category $k = 0$ represents default, which is an absorbing state. The far right column is the unconditional average $\bar{\pi} = \ell$ for each state $\ell = 0, \dots, K$ computed from the MCMC draws of the state indicators.

necessarily mean that its actual real-world default probability is constant. Strictly speaking, a constant absolute rating means that the market prices of the firm's traded assets remain in the same range and have not fluctuated dramatically. Movements in asset prices could be due to a combination of changes in real-world default probabilities and investors' appetite to bear risk. However, the absolute ratings still provide important information because a firm whose absolute rating does not decline when the market as a whole moves (as in the recent crisis) is a strong signal.

Table 2 contains the estimated probabilities $\pi_{\ell,k}$ for transitioning from absolute rating category ℓ to category k . While the probabilities in Table 1 are risk-neutral probabilities, the probabilities in this table are real-world probabilities. They represent the probability a firm currently in one ratings category will transition into another. From this matrix, we see that most transitions into default are coming from the lower rated categories $\ell = 1, 2, 3$. This indicates that the market does anticipate the defaults of many firms and that market prices of traded assets (specifically CDS prices) do contain important information. Interestingly, we also find that transitions into default do not monotonically decline as maturity increases. There is a reasonably large increase in the default probability for the best category $\ell = 9$. Some defaults appear to be a surprise to the market.

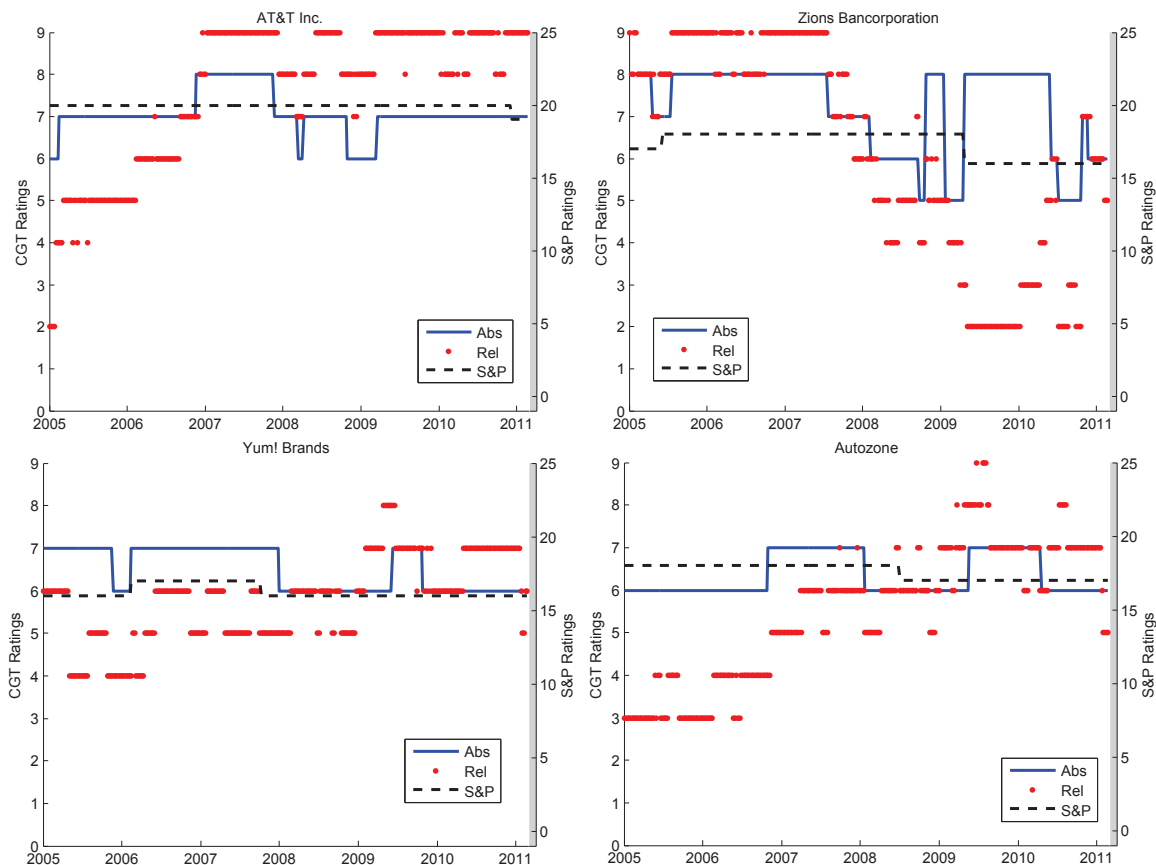


Figure 2: *Absolute (bold line) and relative ratings (dotted line) for four firms from January 2005 through February 2011. Clockwise from the top left are AT&T, Zions Bancorporation, Autozone, and Yum! Brands.*

Figures 2 and 3 plot the absolute and relative ratings for eight different companies from January 2005 through February 2011. The new ratings (from $k = 0, \dots, 9$) are labeled on the left-hand vertical axis in each plot. In the same graph, we also plot the Standard & Poors (S&P) rating for each of these firms. Direct comparison of our ratings to S&P (or other agencies) is not obvious as they report 25 different letter grades from AAA to C as well as default. We map the letter grades into numeric grades and report them on the right-hand vertical axis of each plot.[‡] By comparing the relative ratings (red dashes) and absolute ratings (blue solid lines) for these companies, we can draw some interesting conclusions about the distribution of

[‡]The grades are mapped into numbers as AAA \rightarrow 25, AA+ \rightarrow 24, AA \rightarrow 23, AA- \rightarrow 22, A+ \rightarrow 21, A \rightarrow 20, A- \rightarrow 19, BBB+ \rightarrow 18, BBB \rightarrow 17, BBB- \rightarrow 16, BB+ \rightarrow 15, BB \rightarrow 14, BB- \rightarrow 13, B+ \rightarrow 12, B \rightarrow 11, B- \rightarrow 10, CCC+ \rightarrow 9, CCC \rightarrow 8, CCC- \rightarrow 7, CC+ \rightarrow 6, CC \rightarrow 5, CC- \rightarrow 4, C+ \rightarrow 3, C \rightarrow 2, C- \rightarrow 1, Default \rightarrow 0, Not Rated \rightarrow -1.

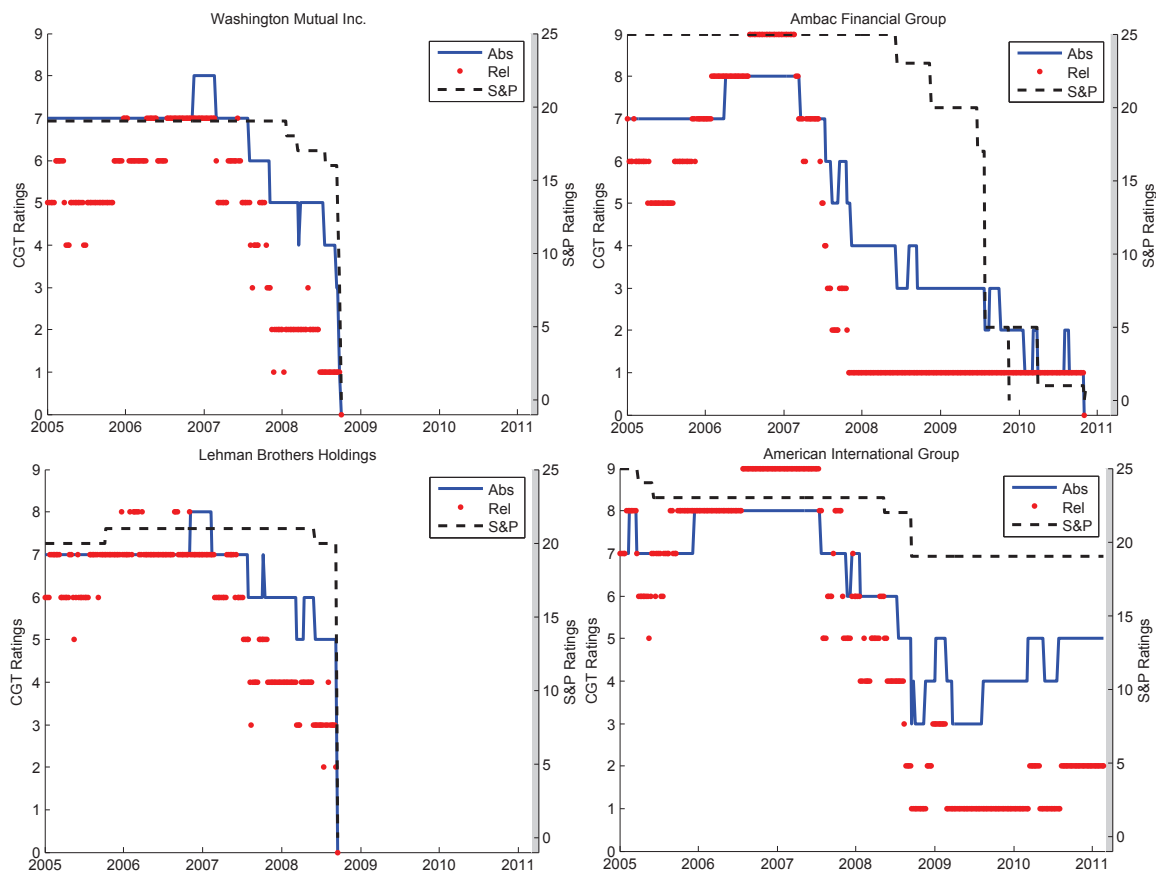


Figure 3: *Absolute (blue line) and relative ratings (red dotted line) for four firms from January 2005 through February 2011. Clockwise from the top left are Washington Mutual Inc., Ambac Financial, American International Group, and Lehman Brothers Inc.*

all companies' ratings. Consider the case of YUM! Brands, AT&T, and Autozone in Figure 2. The absolute ratings for these companies remained roughly stable from 2005 through early 2011 but their relative ratings changed substantially. At the beginning of 2005, they were perceived as companies in the lower grade while by the end of the sample they are some of the better companies. S&P ratings are considered to be relative ratings (i.e. they state that a firm rated AA is regarded as more creditworthy than a firm rated BB). Interestingly, while the market's assessment of these firms relative position changed, S&P ratings for both of these firms did not change as they remained rated BBB or BBB⁺ throughout 2005 to 2011. Finally, notice that the relative ratings change frequently because, by construction, they are not smoothed intertemporally.

Figure 3 includes the relative, absolute, and S&P ratings of three financial firms Washington Mutual, Lehman Brothers, and Ambac Financial Group, which all defaulted over the course of the sample. In addition, we also plot the ratings of American International Group (AIG), which along with the first two firms was at the center of the financial crisis. A striking feature of the ratings for these firms is that the relative ratings declined much faster than the absolute ratings. A comparison of the relative and absolute ratings provide a clear market signal of declining credit quality. Changes in these ratings can be observed in the data as early as the first quarter of 2007, whereas S&P did not substantially change the ratings until immediately prior to their default. Another striking difference in these plots are the ratings of Ambac Financial which had a relative rating equal to 1 for two years prior to finally defaulting. The new market implied ratings in many cases lead the downgrades by S&P by over a year.

Finally, output from the posterior distribution of the latent indicators $p(I_0, \dots, I_t | y^t, D^t; \theta)$ can be used to construct credit indices for different sectors of the economy. Credit indices provide information to market participants, policy makers, and regulators about the state of the credit market. A simple index that is easy to construct is the percentage of firms in each absolute rating category through time. The upper left panel of Figure 4 breaks the empirical distribution of the absolute ratings for all firms into five groups. We can see a dramatic decline in the percentage of firms in the top tier ($k = 7, 8, 9$) beginning in the middle of 2007. There is also a noticeable but more gradual increase in the percentage of firms in the lowest rated categories ($k = 0, 1, 2, 3, 4$) peaking at the end of 2008. The remaining plots of Figure 4 contain similar indices except that SIC industry codes are used to focus on sub-sectors of the economy. The top right graph includes the percentage of financial firms broken down into different categories. The plot illustrates that prior to 2007 most financial firms were well regarded by the market but this changed abruptly at the end of 2007 when financial markets recognized the impending crisis. On the bottom left of Figure 4, we plot similar percentages by absolute ratings category but only for deposit-taking financial institutions (commercial banks). We do the same for non-depository credit/lenders in the bottom right-hand plot. Interestingly, we can see that prior to the end of 2007 almost all depository institutions were regarded as credit-worthy, mainly clustering into the top rated category ($k = 7, 8, 9$), whereas the percentage of firms in these categories in March 2011 is almost zero. Credit indices like these provide a simple summary of debt markets and are a useful tool for regulators and policymakers who need to monitor the

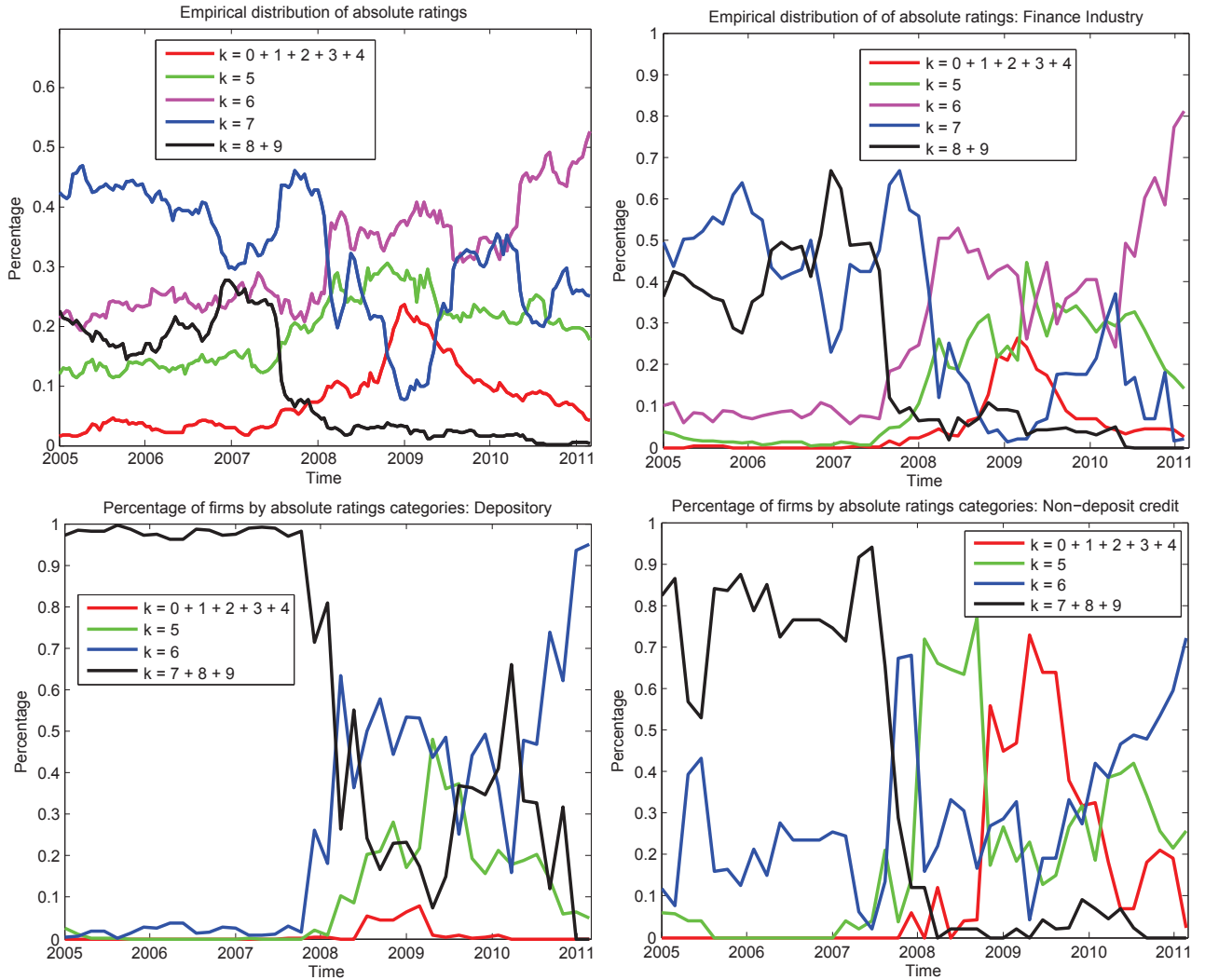


Figure 4: Each graph depicts the percentage of firms in different absolute ratings categories. Top left: all firms. Top right: all financial firms. Bottom left: all deposit-taking financial firms. Bottom right: all non-depository financial firms that are lenders.

health of sub-sectors of the economy.

4.2 Loss distributions

The estimation of loss distributions for portfolios of bonds is an important ingredient in risk management. The past decade has seen a number of advancements in the modeling and estimation of default probabilities. Popular methods may focus only on defaults/bankruptcies, e.g.

Duffie, Saita, and Wang (2007), Koopman and Lucas (2008), and Duffie, Eckner, Horel, and Saita (2009), or default data may be augmented with the addition of credit ratings transitions from Moody's or Standard & Poors; e.g., Gagliardini and Gouriéroux (2005), Koopman, Lucas, and Monteiro (2008), Creal, Schwaab, Koopman, and Lucas (2011), and Creal, Koopman, and Lucas (2012). While it can be challenging to directly estimate models for large databases of CDS's spreads or bonds, it is relatively easier to build models for thousands of firm's credit ratings. As the market-based credit ratings discussed in this paper are direct functions of asset prices, they can be incorporated into models for credit ratings transitions. Loss distributions for portfolios of bonds can then be constructed from credit ratings models. This provides a constructive and computationally tractable way to incorporate data from credit markets. It also leverages previous research on models for ratings transitions.

We propose a simple alternative method for constructing loss distributions directly from the mixture model used for clustering. Consider a portfolio of companies for which we know the absolute ratings. As each rating category corresponds to a distribution over the space of risk-neutral survival functions, one can price a (hypothetical) 10-year zero-coupon, defaultable bond for each firm conditional on the current rating. This provides an estimate of the current value of the portfolio. The loss distribution at any future horizon can then be viewed as a nonlinear function of moments of the predictive distribution from the mixture model (1)-(3). Consistent with Bayesian inference for functions of the predictive distribution, we simulate future paths of the ratings forward through time for each company using the Markov transition matrix. For each simulated path of ratings, we simulate a sequence of survival functions and compute a corresponding path of prices. These prices are then discounted back to the present using a discount factor curve constructed from the current Libor and interest rate swap prices. If a firm transitions into default, a loss-given default is drawn by bootstrapping from the empirical distribution of LGD's. The loss distributions are then built by comparing the current value of the portfolio to the value of the portfolio constructed from the simulated future paths of prices from the predictive distribution.

We illustrate this approach using the portfolio of 4,461 rated companies on 2/23/2011. In practice, we do not condition on the current absolute rating but marginalize over the current ratings using the MCMC draws for the latent category indicators from the final time period $\left\{ I_T^{(j)} \right\}_{j=1}^M$. Figure 5 contains the loss distributions at the six month (left column) and one

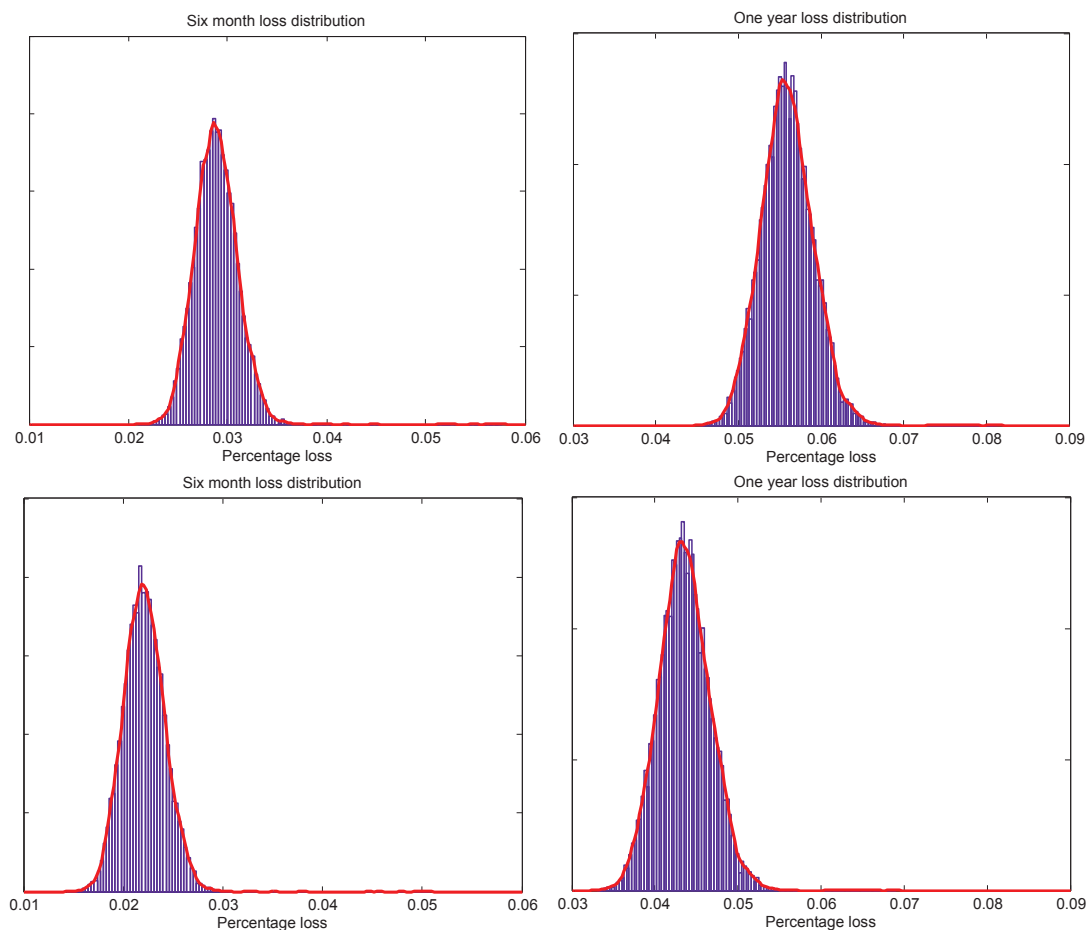


Figure 5: *Six month (left) and one year (right) loss distributions for a portfolio of 4,461 companies rated at the end of the sample on 2/23/2011. The loss distributions were constructed using two different (risk-neutral) discount factor rate curves constructed from Libor and swap rates on 12/17/2008 (top row) and on 2/23/2011 (bottom row).*

year (right column) horizons for two different discount rate curves that existed on 12/17/2008 and 2/23/2011. These are the top and bottom rows of Figure 5, respectively. Consistent with our expectations, the loss distributions during a period like the financial crisis (top row) are worse than a calmer period like early 2011. These graphs illustrate how the loss distributions are sensitive to interest rate variation or interest rate risk. An adverse movement in interest rates causes a rather substantial shift to the right in the loss distribution for bonds. The loss distributions in Figure 5 condition on a known discount rate curve, i.e. the path of future (risk-free) discount rates are assumed to be known. By combining the market based ratings

with a model for the future evolution of the discount rate curve (the term structure of interest rates), one can construct loss distributions that also reflect the uncertainty associated with future discount rates as well.

Another extension that may be beneficial is to allow the Markov transition probabilities of the mixture model to change through time as a function of industry, macroeconomic, and most importantly credit market conditions. It is ultimately the inability of a firm to role over short term debt that results in default. Conditioning on measures of liquidity in the short-term debt market would be useful. Allowing the transition matrices to change through time would enable firms to move more rapidly between ratings categories during some periods rather than others. This is an additional source of uncertainty that could be incorporated into the market-based ratings framework. We leave this extension for future research.

5 Conclusion and discussion

We created new classes of credit ratings for publicly traded companies in the U.S. from actively traded assets. The new ratings include both absolute and relative ratings, which have distinct economic interpretations. Our approach uses asset pricing data to impute a term structure of risk neutral survival functions or default probabilities. Firms are then clustered into ratings categories based on their survival functions using functional classification. This allows all public firms whose assets are traded to be directly rated by market participants. Firms whose assets are not traded cannot be directly rated by the market but can be indirectly rated through the use of matching estimators. Our approach has the advantages of being timely (weekly data versus monthly or quarterly), simple to interpret economically, transparent, and computationally tractable. In addition, we have demonstrated that our ratings typically lead those of one of the major ratings agencies.

There are several extensions and further applications of the proposed model-based ratings. The standard matching estimator used here intentionally does not adjust for the presence of selection bias. Selection bias can occur if a firm is not traded for a reason that is unobservable to us and the imputed rating based on observable characteristics is not representative of its actual credit worthiness. Selection biases could be positive or negative. For example, high performing technology companies (Google and Apple) are typically not traded because they

have little debt and are regarded as highly unlikely to fail while other companies are traded because they may be small or the economic consequences of their failing are insignificant. We regard selection bias as an important econometric issue, which we hope to address in future research.

From a statistical perspective, credit ratings are a type of dimension reduction. Ideally, a credit rating would be a sufficient statistic of a firm's creditworthiness. A possible criticism of credit ratings are that they lose information relative to the raw data from which they are built. To address this concern, one can imagine what happens when the number of categories K grows large. In principle, if the number of categories equals the number of firms, then there is no dimension reduction and little pooling of information across firms. The model reduces to a dynamic model for survival functions.

As the number of categories K grows, clustering becomes more difficult in practice due to the number of parameters. However, a practical extension is to recluster firms into further sub-categories after the initial round of clustering. For example, once firms have been clustered into K ratings, it is possible to collect all firms in category j and refine them into \tilde{K}_j sub-categories. As the categories become continually refined, the joint distribution over the space of indicators begins to approximate the joint distribution over the space of survival functions.

Appendix

Details on the Gibbs sampling algorithm

In the Gibbs sampler for the Markov-switching mixture model, there are five conditional distributions from which to draw. The algorithm is initialized by drawing the parameters of the model from their prior distributions. The initial state indicators are drawn from the forward filtering, backward sampling algorithm. All steps of the algorithm also condition on the observed default indicators D^T . The algorithm then proceeds for iterations $m = 1, \dots, M$ as follows:

- Conditional on the observed data, mixture indicators $(I_0, I_1, \dots, I_T)^{(m-1)}$, and parameters $\theta^{(m-1)}$ of the model, the missing observations are drawn from a constrained multivariate normal distribution for $t = 1, \dots, T$ and $i = 1, \dots, \tilde{N}_{it}$. This forms the complete data $y_c = (y_{c,1}, \dots, y_{c,T})$ for this iteration. In this step, we use the Gibbs sampler for constrained multivariate normal distributions of Rodriguez-Yam, Davis, and Scharf (2004) to ensure that the completed observation vector is a decreasing function of maturity.

- Conditional on the complete data $y_c^{(m)}$ and parameters $\theta^{(m-1)}$, draw new mixture indicators $(I_0, I_1, \dots, I_T)^{(m)}$ using the forward filtering backward sampling algorithm for Markov-switching models; see, e.g. Frühwirth-Schnatter (2006, p. 342).
- Conditional on the complete data $y_c^{(m)}$, the mixture indicators $(I_0, I_1, \dots, I_T)^{(m)}$, and mean vector $\mu_K^{(m-1)}$, draw the covariance matrix Σ from an inverse Wishart distribution $\Sigma^{(m)} \sim \text{Inv.-Wishart}(\bar{\kappa}, \bar{\Omega})$ where $\bar{\kappa} = \underline{\kappa} + N_t T$ and $\bar{\Omega} = \underline{\Omega} + \sum_{t=1}^T \sum_{i=1}^{N_t} (y_{it} - \mu_{I_{it}})(y_{it} - \mu_{I_{it}})'$.
- Conditional on the complete data $y_c^{(m)}$, the mixture indicators $(I_0, I_1, \dots, I_T)^{(m)}$, and covariance matrix $\Sigma^{(m)}$, draw the mean vector μ_K from a constrained multivariate normal distribution again following Rodriguez-Yam, Davis, and Scharf (2004).
- Conditional on the mixture indicators $(I_0, I_1, \dots, I_T)^{(m)}$, draw each row of probabilities for the Markov transition matrix from the Dirichlet distribution $\pi_{i,\cdot}^{(m)} \sim \text{Dirichlet}(\bar{\alpha}_j)$ for $j = 1, \dots, K$; see, e.g. Frühwirth-Schnatter (2006, p. 340).

Data

We use the following data items or *symbols* from Compustat/CRSP. The ratio of total earnings to total assets is $(1 - \ln \frac{\text{EBIT}}{\text{TA}})$ where EBIT is item OIBDPQ and total assets (TA) is data item ATQ. The ratio of retained earnings (RE) to total assets is $(1 - \ln \frac{\text{RE}}{\text{TA}})$ where RE is data item REQ. The log-ratio of current liabilities (CL) to non-current liabilities (NCL) is measured by $\ln \frac{\text{CL}}{\text{NCL}}$ where CL is given by data item DLTTQ and NCL is data item LCTQ. Accounting leverage is measured by $\frac{\text{TA}}{\text{TA} - \text{LTA}}$ where long-term assets is data item LTA. Leverage is measured by $1 + \ln \frac{\text{TA}}{\text{TL}}$ where total liabilities (TL) is data item LTQ. Liquidity is measured by working capital divided by total assets where working capital is data item WCAPQ.

References

- Abadie, A. and G. W. Imbens (2006). Large sample properties of matching estimators for average treatment effects. *Econometrica* 74(1), 235–267.
- Altman, E., N. Fargher, and E. Kalotay (2010). A simple empirical model of equity-implied probabilities of default. Unpublished manuscript, Stern School of Business, New York University.
- Angrist, J. D. and J.-S. Pischke (2009). *Mostly Harmless Econometrics*. Princeton, NJ: Princeton University Press.

- Blanco, R., S. Brennan, and I. W. Marsh (2005). An empirical analysis of the dynamic relation between investment-grade bonds and credit-default swaps. *The Journal of Finance* 60, 2255–2281.
- Creal, D. D., S. J. Koopman, and A. Lucas (2012). Generalized autoregressive score models with applications. *Journal of Applied Econometrics*. forthcoming.
- Creal, D. D., B. Schwaab, S. J. Koopman, and A. Lucas (2011). Observation driven mixed measurement dynamic factor models with an application to credit risk. Discussion paper TI 11-042, Tinbergen Institute, Amsterdam.
- Duffie, D. (2012). *Dark Markets: Asset Pricing and Information Transmission in Over-The-Counter Markets*. Princeton, NJ: Princeton University Press.
- Duffie, D., A. Eckner, G. Horel, and L. Saita (2009). Frailty correlated default. *The Journal of Finance* 64(5), 2089–2123.
- Duffie, D., L. Saita, and K. Wang (2007). Multi-period corporate default prediction with stochastic covariates. *Journal of Financial Economics* 83(3), 635–665.
- Duffie, D. and K. J. Singleton (2003). *Credit Risk*. Princeton, NJ: Princeton University Press.
- Frühwirth-Schnatter, S. (2006). *Finite Mixture and Markov Switching Models*. New York, NY: Springer Press.
- Gagliardini, P. and C. Gouriéroux (2005). Stochastic migration models with application to corporate risk. *Journal of Financial Econometrics* 3(2), 188–226.
- Guo, S. Y. and M. W. Fraser (2009). *Propensity Score Analysis: Statistical Methods and Applications*. Thousand Oaks, CA: Sage Publications.
- Hamilton, J. D. (1989). A new approach to the econometric analysis of nonstationary time series and the business cycle. *Econometrica* 57(2), 357–384.
- Hastie, T., R. Tibshirani, and J. Friedman (2009). *The Elements of Statistical Learning* (Second ed.). New York, NY: Springer Press.
- Hull, J., M. Predescu, and A. White (2004). The relationship between credit default swap spreads, bond yields, and credit ratings announcements. *Journal of Banking and Finance* 28, 2789–2811.

- Koopman, S. J. and A. Lucas (2008). A non-Gaussian panel time series model for estimating and decomposing default risk. *Journal of Business and Economic Statistics* 26(4), 510–525.
- Koopman, S. J., A. Lucas, and A. Monteiro (2008). The multi-state latent factor intensity model for credit rating transitions. *Journal of Econometrics* 142(1), 399–424.
- Markit (2009). The CDS Big Bang: understanding the changes to the global CDS contract and North American conventions. www.markit.com/cds/announcements/resource/cds_big_bang.pdf.
- McLachlan, G. and D. Peel (2000). *Finite Mixture Models*. Hoboken, NJ: John Wiley Press.
- McNeil, A. J., R. Frey, and P. Embrechts (2005). *Quantitative Risk Management*. Princeton, NJ: Princeton University Press.
- Merton, R. (1974). On the pricing of corporate debt: the risk structure of interest rates. *The Journal of Finance* 19(2), 449–470.
- O’Kane, D. (2008). *Modelling Single-Name and Multi-Name Credit Derivatives*. West Sussex, England: Wiley Finance.
- Rodriguez-Yam, G., R. A. Davis, and L. L. Scharf (2004). Efficient Gibbs sampling of truncated multivariate normal with application to constrained linear regression. Unpublished manuscript, Department of Statistics, Colorado State University.
- Viterbi, A. J. (1967). Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Transactions in Information Theory* 13, 260–269.
- Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data* (Second ed.). Cambridge: M.I.T. Press.