The Review of Financial Studies

Banks' Risk Dynamics and Distance to Default

Stefan Nagel
University of Chicago, NBER, and CEPR

Amiyatosh Purnanandam
University of Michigan

We adapt structural models of default risk to take into account the special nature of bank assets. The usual assumption of lognormally distributed asset values is not appropriate for banks. Typical bank assets are risky debt claims with concave payoffs. Because of the payoff nonlinearity, bank asset volatility rises following negative shocks to borrower asset values. As a result, standard structural models with constant asset volatility can severely understate banks' default risk in good times when asset values are high. Additionally, bank equity return volatility is much more sensitive to negative shocks to asset values than in standard structural models. (*JEL* G13, G21, G28)

Received July 10, 2017; editorial decision April 26, 2019 by Editor Itay Goldstein.

The distress that many banks experienced during the recent financial crisis has brought renewed emphasis on the importance of understanding and modeling bank default risk. Assessment of bank default risk is important not only from an investor's viewpoint but also for risk managers analyzing counterparty risks and for regulators gauging the risk of bank failure. Accurate modeling of bank default risk is also required for valuing the benefits that banks derive from implicit and explicit government guarantees.

In many applications of this kind, researchers and analysts rely on structural models of default risk in which equity and debt are viewed as contingent claims on the assets of the firm. Following Merton (1974), the standard approach (which we call the Merton model) is to assume that the value of the assets of the firm follows a lognormal process. The options embedded in the firm's

We are grateful for comments from Andrea Eisfeldt, Itay Goldstein, Zhiguo He, David Lando, Debbie Lucas, Ben Meiselman, George Pennacchi, Uday Rajan, Stephen Schaefer, Pietro Veronesi, and Toni Whited; three anonymous referees; seminar participants at the Bank of Canada, Federal Reserve Board, Georgia State University, MIT, University of Michigan, and University of Texas at Dallas; and conference participants at the American Finance Association Meetings, the Conference on Financial Regulation at the University of Chicago, and the Trinity of Stability Conference at the Deutsche Bundesbank. Send correspondence to Stefan Nagel, Booth School of Business, University of Chicago, Chicago, IL 60637; telephone: 773-834-3709. E-mail: stefan.nagel@chicagobooth.edu.

The Review of Financial Studies 33 (2020) 2421–2467

doi:10.1093/rfs/hhz125 Advance Access publication October 17, 2019

equity and debt can then be valued as in Black and Scholes (1973). Researchers have recently used this approach to value implicit (too-big-to-fail) government guarantees for banks (Acharya, Anginer, and Warburton 2014; Schweikhard, Tsesmelidakis, and Merton 2014) and quasi-governmental institutions (Lucas and McDonald 2006; Lucas and McDonald 2010). An extensive literature has applied this model to price deposit insurance, since Merton (1977), Marcus and Shaked (1984), Ronn and Verma (1986), and Pennacchi (1987).

The Merton model's assumption of lognormally distributed asset values may provide a useful approximation for the asset value process of a typical nonfinancial firm. However, for banks this assumption is clearly problematic. Much of the asset portfolio of a typical bank consist of debt claims such as mortgages. The fact that the upside of the payoffs of these debt claims is limited is not consistent with the unlimited upside implied by a lognormal distribution.

In this paper, we propose a modification of the Merton model that takes into account the capped upside of bank assets. Our approach has three main elements. First, we apply the lognormal distribution assumption not to the assets of the bank, but to the assets of the bank's borrowers that serve as loan collateral. More precisely, we model banks' assets as a pool of no-recourse zero-coupon loans where loan repayments depend on the value of borrowers' collateral assets at loan maturity as in Vasicek (1991). Collateral asset values are subject to common factor shocks as well as idiosyncratic risk. Second, loans have staggered maturities. Every period a fraction of the loan portfolio matures and the bank issues the repayment proceeds as new loans. New loans are issued at a fixed initial loan-to-value ratio. Thus, to the extent that borrowers had excess collateral at loan maturity, this excess collateral is removed completely and is no longer available to back the loan. In the case of a deficiency, collateral is replenished, but only up to the level that the required initial loan-to-value ratio is satisfied. This asymmetry in collateral removal and replenishment reinforces the capped upside of a bank's assets, and it resembles the refinancing ratchet effect for aggregate mortgage portfolios modeled in Khandani, Lo, and Merton (2013). Finally, the bank's assets are contingent claims on borrowers' collateral assets, and the bank's equity and debt are contingent claims on these contingent claims.

This options-on-options feature of bank equity and debt has important consequences for default risk and equity risk dynamics. As as illustration of the main intuition, consider the simplified case in which all borrowers are identical (with perfectly correlated defaults), all loans have identical terms, and the bank has zero-coupon debt outstanding with the same maturity as the loans in its asset portfolio. In this case, Figure 1 illustrates the payoffs at maturity as a function of the borrower's asset value. In this example, the borrowers have loans with a face value of 0.80, and the bank has issued debt with a face value of 0.60. The maximum payoff the bank can receive is the loans' face value, so the bank's asset value is capped at 0.80. Only when the borrower's assets fall below 0.80 is the bank's asset value sensitive to the borrower's asset values.

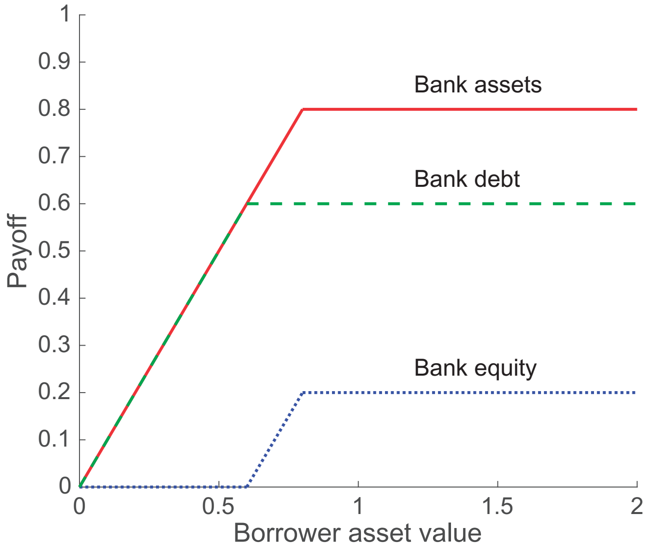


Figure 1
Payoffs at maturity in the simplified model with perfectly correlated borrower defaults

Clearly, the bank asset value cannot have a lognormal distribution (which would imply unlimited upside). Because the bank's borrowers keep the upside of a rise in their asset value above the loan face value, the bank's equity payoff does not resemble a call option on an asset with unlimited upside, but rather a mezzanine claim with two kinks. This mezzanine-like nature of the bank's equity claim has important consequence for the risk dynamics of bank equity and for default risk estimation. Because of the capped upside, bank volatility will be very low in "good times," when asset values are high, and, in Figure 1, asset values at maturity will likely end up toward the right side, where the bank's equity payoff is insensitive to fluctuations in borrower asset values.

A standard Merton model in which equity is a call option on an asset with constant volatility misses these nonlinear risk dynamics. Viewed through the lens of this standard model it might seem that a bank in times of high asset values is many standard deviations away from default. But this conclusion would be misleading because it ignores the fact that bank asset volatility could dramatically rise if asset values fall. Similarly, the standard model would make misleading predictions about the riskiness of banks' assets, equity, and debt.

Going beyond this simple illustrative example, our model incorporates idiosyncratic borrower risks and overlapping cohorts of borrowers with staggered loan maturities where maturing loans are replaced with new loans. This revolving replacement of staggered loans is a quantitatively important and realistic feature of the model. For example, a housing boom raises collateral values and lowers the loan-to-value ratios of mortgage borrowers with existing loans. However, loans issued to new borrowers are typically issued with standard loan-to-value ratios (and new borrowers require bigger loans to purchase houses at appreciated prices). Thus, if a bank's borrowers have high collateral values today, this provides a big safety cushion for a bank's claims only until loans are repaid and rolled over into new loans. The reset of collateral values when new loans are issued thus limits the extent to which an appreciation in borrower collateral values today lowers the default risk of a bank. The periodic reset is, effectively, a cap on the collateral values backing the bank's loans that kicks in when loans are rolled over to new borrowers. This reinforces the mezzanine claim of bank equity and the resultant consequences for risk dynamics and distance to default.

To assess the differences between our modified model and the Merton model, we simulate data from our modified model and ask to what extent an analyst using the Merton model would misjudge the risk-neutral probability of default. We find that this error is particularly stark when asset values are high relative to the face value of the bank's debt. In this case, bank asset payoffs are likely to stay in the flat region in Figure 1, and bank equity payoffs are also likely in the flat region. As a consequence, equity volatility is low. Based on the Merton model, an analyst observing low equity volatility would infer that asset volatility must be low. Furthermore, because the asset's volatility is constant in the Merton model, the analyst would (wrongly) conclude that asset volatility will remain low at this level in the future. What the Merton model misses in this case is that asset volatility could substantially rise following a bad asset value shock, because the region of likely asset payoffs would move closer to or into the downward-sloping region in Figure 1. As a result, the Merton model substantially overestimates the distance to default, and it underestimates the risk-neutral probability of default.

We then calibrate our modified model and the standard Merton model to quarterly bank panel data from 1987 to 2016. In the case of the Merton model, we follow the standard approach and look for asset value and volatility of bank assets that allow the model to match the observed market value of equity and its volatility. In the case of our modified model, we fix the volatility and correlation of borrower asset values and the initial loan-to-value ratio at empirically plausible values. We then look for values for the size of the bank's loan book and a common shock to borrower asset values after loan origination to match the bank's market value of equity and its equity volatility. Even though both models are calibrated to the same equity market data, their implied risk-neutral default probabilities are strikingly different. In line with the simulations we discussed above, the differences are particularly big in the years before the financial crisis when equity values were high and volatility low. Based on the Merton model, the risk-neutral default probability of the average bank in 2006 over a 5-year horizon is roughly 5%. In contrast, the risk-neutral default probability implied by our modified model is three times as high. Translated into credit spreads, this would imply an annualized spread of around 5 basis points (bps) in the Merton model and close to 40 bps according to our modified model.

Once the financial crisis hit in 2007–2008, the models' predictions are not so different anymore. At this point, bad asset value shocks had moved banks into the downward-sloping asset payoff region in Figure 1. In this region, the kink in the asset payoff becomes less relevant and the predictions from our modified model are close to those from the standard Merton model. In periods of the most extreme distress, Merton model default probabilities can even exceed those from our modified model. Default probabilities estimated from the modified model again started to exceed Merton model default probabilities by 2 to 3 times in the post-2012 period as the economy started to recover from the great recession. Returning to the earlier periods, we find the same pattern. The modified model provides a much higher default probability during the 1990s, a period characterized by high equity valuations and low volatility. However, during the savings and loans crisis period of 1987–1993, there are spikes in default probabilities when Merton model predictions approach those from our modified model.

Thus, the key problem with applications of standard structural models to banks is that they understate the risk of default in "good times." This is an important issue , for example, for the estimation of the value of explicit or implicit government guarantees. Based on a standard Merton model calibrated to equity value and volatility data from 2006 (i.e., precrisis times), one may be led to the conclusion that the value of a guarantee is almost nil when, in fact, the value is a lot higher if one takes into account the fact that banks' asset volatility will go up when asset values fall. In fact, the FDIC charged almost zero insurance premium for a number of commercial banks during the precrisis period (see Duffie et al. 2003). Such a policy may seem justified based on Merton model default probabilities, but our modified model will suggest a much higher premium during good times.

We further investigate the plausibility of our modification of the Merton model by comparing the models' predictions about bank equity volatility following a bad shock to the bank's asset value. We calibrate both models to match data on equity values and volatility in 2006Q2. We then add a negative shock to borrower asset values based on the change in house values. We use two measures of shock, one based on Freddie Mac house price index and the other based on unlevered returns on U.S. REITs from 2006Q2 to subsequent quarters. Because the latter measure is market based, we expect this shock to be more informative of the true asset values of the bank's borrowers. The negative shock to borrower asset values then translates, in our modified model, into a shock to the bank's asset value. We then apply an asset value shock of the same magnitude in the Merton model. In the Merton model, the consequences are mild. Using the Freddie Mac housing index as a measure of shock, the average bank's equity volatility rises by about 8–9 percentage points (pp) from 2006Q2 to 2009. This is a modest increase compared to the average bank equity volatility of about 25% in 2006Q2. In contrast, in our model, average bank equity volatility increases by about 14–15 pp because the model takes into account not only the drop in the bank asset values but also the rise in bank asset volatility. This increase is still below the actual realized equity volatility of U.S. banking stocks during the crisis. When we use shocks based on unlevered returns to publicly traded REITs, the match between the volatility implied by our model and the realized volatility dramatically improves. At the peak of the financial crisis, the two measures come within 5% of each other. Merton model-based equity volatility remains considerably lower even with this shock. Taken together, these exercises illustrate that application of a standard structural model with constant asset volatility can severely understate the sensitivity of bank equity risk to negative asset value shocks, and this in turn can lead to a relatively inferior model of default prediction.

Our objective in this paper is to improve structural models of bank default risk in one important aspect—capturing the nonlinearity of bank asset payoffs—but our modified structural model still omits many features—for example, liquidity concerns, interest-rate risk, complex capital structure, and government guarantees—that would be necessary for an entirely realistic modeling of bank default risk. Similarly, our model does not allow banks to rebalance their portfolios to change the volatility of their assets in response to positive or negative shocks. For default prediction that takes into account many of these complications, a reduced-form model rather than a structural one may be the preferred method in practice. But for reduced-form models, too, our results have important implications. Many reduced-form models use a Merton distance-to-default model as one of the state variables driving default intensity (e.g., Duffie, Saita, and Wang 2007; Bharath and Shumway 2008; Campbell, Hilscher, and Szilagyi 2008). Our analysis suggests that for banks the default probability from our modified model may be better suited as a default predictor. For example, it could be included as a predictor within a reduced-form deposit insurance pricing model as in Duffie et al. (2003). The reduced-form approach permits a lot of flexibility to obtain realistic default risk estimates, but the structural approach that we pursue here is useful for understanding the economic drivers of default risk (which in turn may be useful for developing better specifications of reduced-form models).

To understand the relative predictive power of the default risk estimates from our modified model and the Merton model, we estimate a Cox proportional hazard model to predict bank defaults during our sample period using the two default risk measures as predictors. The modified model's default probability does a considerably better job in accurately predicting future default compared to the Merton model's default risk estimates. We also show that the modified model's default probabilities are not simply a monotonic nonlinear transformation of the Merton model estimates. We estimate a nonparametric regression of the Merton model default probabilities on the modified model default probabilities and show that both the predicted values from this regression and the residuals are informative about the actual subsequent defaults of banks. In other words, our modified default risk measure contains additional information that cannot be simply captured by a nonparametric transformation of the Merton model default probabilities.

Our paper relates to several strands in the literature. Three papers in the deposit insurance pricing literature anticipate some elements of our approach. Ritchken, Thomson, and Popova (1995) value deposit insurance in a model where banks' assets are risky debt to a representative firm, but they do not derive its implications for default risk or equity volatility dynamics. Dermine and Lajeri (2001) look at the case of a single borrower with loan maturity that matches exactly the bank's debt maturity, as in our illustrative case above in Figure 1. Chen et al. (2006) allow for a portfolio of loans with idiosyncratic default risk, but the maturity of the bank's debt is still tied to be equal to the maturity of the bank's loans. In contrast, our model de-couples the maturities of banks' assets and liabilities. Moreover, the staggering of loan maturities allows us to introduce the collateral ratchet effect. Both features are quantitatively important. Moreover, unlike these earlier papers, we empirically apply and evaluate the model. Gornall and Strebulaev (2018) specify bank assets as loan portfolios in similar ways as we do, albeit without staggering of loan maturities. Their focus is on modeling bank's capital structure choices in equilibrium, while we focus on implications for default risk estimation and valuation of bank's securities.

Our work also relates to recent research that uses data from options markets to understand credit risk and bank risk. Culp, Nozawa, and Veronesi (2014) construct pseudo-firms that have traded securities as assets (e.g., a stock index) and pseudo-bonds—a combination of Treasuries and put options—as liabilities. In one of their applications, they also consider a bank that owns a portfolio of pseudo-bonds. In this way, analogous to our approach, they also capture the options-on-options nature of bank equity and debt, albeit in a nonparametric way rather than in a parametric structural model. Our parametric structure is useful for understanding and counterfactually simulating the economic drivers of bank default risk. Kelly, Lustig, and Van Nieuwerburgh (2016) estimate the value of implicit government guarantees for the banking system by comparing prices of options on a banking index and a portfolio of options on individual bank stocks. Their method involves fitting models with stochastic volatility and jumps to option prices. In these models, the correlation between returns and shocks to volatility (the "leverage" effect) is a reduced-form parameter. Our structural model of bank risk predicts a specific (nonlinear) relation between bank equity returns and bank equity volatility. Finally, our work can be extended in several directions; for example, Peleg-Lazar and Raviv (2017) study the implications of this payoff structure on the risk-shifting incentives of shareholders. In addition, our paper's key insight can be useful in estimating various forward-looking measures of bank risk, a topic that has been extensively examined by a growing literature about bank stress tests (see, e.g., Goldstein and Sapra 2014; Goldstein and Leitner 2018; Leitner 2014; Greenlaw et al. 2012; Acharya, Engle, and Pierret 2014; Boucher et al. 2014; Gofman 2017).

1. Structural Model of Default Risk for Banks

Unlike the simplified case in Figure 1, we now set up a more realistic model in which borrower assets have idiosyncratic risk and banks issue loans with staggered maturity dates. Both of these additional features are important because they lead to some smoothing of the bank asset payoff function in Figure 1.

Consider a setting with continuous time. A bank issues zero-coupon loans with maturity T. Loans are issued in staggered fashion to N cohorts of borrowers. Cohorts are indexed by the time τ that has passed at $t=0$ because their loans were issued, ordered as $\tau = T, T(N-1)/N, ..., T/N$. Each cohort is comprised of a continuum of borrowers indexed by $i \in [0,1]$ with mass $1/N$.

Let $A_t^{\tau,i}$ denote the collateral value of a borrower i in cohort τ at time t. Under the risk-neutral measure, the asset value evolves according to the stochastic differential equation

$$\frac{dA_t^{\tau,i}}{A_t^{\tau,i}} = (r-\delta)dt + \sigma(\sqrt{\rho}dW_t + \sqrt{1-\rho}dZ_t^{\tau,i}), \tag{1}$$

where W and $Z^{\tau,i}$ are independent standard Brownian motions, δ is a depreciation rate, and r is risk-free rate. The $Z^{\tau,i}$ processes are idiosyncratic and independent across borrowers. This is a one-factor model of borrower asset values as in Vasicek (1991). The parameter ρ represents the correlation of asset values that arises from common exposure to W and σ is the instantaneous total volatility.

At the time of the initial loan issue, $t=-\tau$, borrowers in each cohort start out with the same initial collateral asset value. For the purpose of the presentation in this section, we normalize this initial value to $A_{-\tau}^{\tau,i}=1$. One could choose a different normalization (and we do in our empirical analysis), but this does not affect any of the results, it just scales the balance sheet quantities in a different way. We fix an initial loan-to-value ratio of ℓ. The face value of the loan is

$$F_1(\mu) = \ell e^{\mu T}, \tag{2}$$

with μ as the promised yield on the loan (that we will solve for below). In line with standard structural models of credit risk, we assume that borrowers default if the asset value at maturity is lower than the amount owed. The payoff at maturity $t=T-\tau$ received by the bank from an individual borrower in cohort τ then is

$$L_{T-\tau}^{\tau,i}(\mu) = \min\left[A_{T-\tau}^{\tau,i}, F_1(\mu)\right]. \tag{3}$$

We assume that loans are competitively priced, and so the promised yield on the loan is the μ that solves

$$\ell = e^{-rT} E^{\mathbb{Q}}_{-\tau}[L^{\tau,i}_{T-\tau}(\mu)], \tag{4}$$

where $E^{\mathbb{Q}}_{-\tau}[.]$ denotes a conditional expectation under the risk-neutral measure at the time of loan issuance. The expression on the right-hand side is simply the value of risk-free debt with face value $F_1(\mu)$ minus the Black-Scholes value of a put option on the borrower's assets (with the depreciation rate δ as dividend yield).

To analyze the payoff to the bank from the whole loan portfolio, one should first solve for the aggregate value of collateral in cohort τ, which is

$$\begin{aligned} A^{\tau}_{T-\tau} &= \frac{1}{N}\int_0^1 A^{\tau,j}_{T-\tau} dj \\ &= \frac{1}{N}\exp\left\{(r-\delta)T - \frac{1}{2}\rho\sigma^2 T + \sigma\sqrt{\rho}(W_{T-\tau} - W_{-\tau})\right\}, \end{aligned} \tag{5}$$

and the aggregate log asset value, which is

$$\begin{aligned} a^{\tau}_{T-\tau} &= \frac{1}{N}\int_0^1 \log A^{\tau,j}_{T-\tau} dj \\ &= \frac{1}{N}\left[(r-\delta)T - \frac{1}{2}\sigma^2 T + \sigma\sqrt{\rho}(W_{T-\tau} - W_{-\tau})\right]. \end{aligned} \tag{6}$$

Because idiosyncratic risk fully diversifies away with a continuum of borrowers in each cohort, the stochastic component of the aggregate asset value in a cohort depends only on the common factor realization $W_{T-\tau} - W_{-\tau}$.[1]

We now obtain the payoff that the bank receives at maturity from the portfolio of loans given to cohort τ as

$$\begin{aligned} L^{\tau}_{T-\tau}(\mu) &= \frac{1}{N}\int_0^1 L^{\tau,j}_{T-\tau} dj \\ &= \frac{1}{N}\int_0^1 A^{\tau,j}_{T-\tau} dj - \frac{1}{N}\int_0^1 \max\left[A^{\tau,j}_{T-\tau} - F_1(\mu), 0\right] dj \\ &= \frac{1}{N}\left[A^{\tau}_{T-\tau}\Phi\{d_1(\mu)\} + F_1(\mu)\Phi\{d_2(\mu)\}\right], \end{aligned} \tag{7}$$

[1] Formally, for Equations (5) and (6) to hold, we require a law of large numbers such that borrower-specific shocks $dZ^{\tau,i}_t$ cancel out in aggregate within each cohort, conditional on the common factor realization. This can be accomplished, following Uhlig (1996), by defining the integral in Equations (5) and (6) as a Pettis integral (see, e.g., Kogan, Papanikolaou, and Stoffman 2017; Acemoglu and Jensen 2015 for recent applications of the same approach).

where the last equality follows from the properties of the truncated lognormal distribution, $\Phi\{\cdot\}$ denotes the standard normal cumulative distribution function (CDF), and

$$d_1(\mu)=\frac{\log F_1(\mu)-a_{T-\tau}^{\tau}}{\sqrt{1-\rho}\sqrt{T}\sigma}-\sqrt{1-\rho}\sqrt{T}\sigma, \tag{8}$$

$$d_2(\mu)=-\frac{\log F_1(\mu)-a_{T-\tau}^{\tau}}{\sqrt{1-\rho}\sqrt{T}\sigma}. \tag{9}$$

The $\max\left[A_{T-\tau}^{\tau,j}-F_1(\mu),0\right]$ term in Equation (7) reflects the option value for the borrower, that is, the upside of the collateral value that is retained by the borrower. Conditional on $W_{T-\tau}-W_{-\tau}$, there are some borrowers in cohort τ for whom this option is in the money, and others for whom it is not, depending on the realization of their idiosyncratic shocks. This is why d_1 and d_2 are functions of idiosyncratic risk $\sqrt{1-\rho}\sqrt{T}\sigma$. Thus, while idiosyncratic risk is diversified away in the aggregate borrower asset value, it matters for loan payoffs, because borrower default depends on idiosyncratic risk.

At $t=T-\tau$, the bank fully reinvests the proceeds, $L_{T-\tau}^{\tau}$, from the maturing loan portfolio of cohort τ into new loans, with uniform amounts, to members of the same cohort. The new loans carry a face value of

$$F_2(\mu)=L_{T-\tau}^{\tau}e^{\mu T}. \tag{10}$$

We assume that the bank keeps the time-of-issue loan-to-value ratio at the same level, that is, ℓ, as for the initial round of loans. Borrowers reduce or replenish collateral assets accordingly: the asset value of each member of cohort τ is uniformly reset to the same value

$$A_{(T-\tau)^+}^{\tau,i}=\frac{L_{T-\tau}^{\tau}}{\ell} \tag{11}$$

an instant after the reissue of the loans. The cohort-level aggregates A_t^{τ} and a_t^{τ} for $t>\tau$ are based on these reinitialized asset values. With the same loan-to-value ratio, these loans have the same risk as the first-generation loans and hence the same promised yield μ applies.

The aggregate payoff of the portfolio of loans of cohort τ at the subsequent maturity date $2T-\tau$ then follows, along similar lines as above, as

$$\begin{aligned}
L_{2T-\tau}^{\tau}&=\frac{1}{N}\int_0^1 L_{2T-\tau}^{\tau,j}dj\\
&=\frac{1}{N}\int_0^1 A_{2T-\tau}^{\tau,j}dj-\frac{1}{N}\int_0^1 \max(A_{2T-\tau}^{\tau,j}-F_2(\mu),0)dj\\
&=\frac{1}{N}\left[A_{2T-\tau}^{\tau}\Phi(d_3)+F_2(\mu)\Phi(d_4)\right],
\end{aligned} \tag{12}$$

where

$$d_3 = \frac{\log F_2(\mu) - a^{\tau}_{2T-\tau}}{\sqrt{1-\rho}\sqrt{T}\sigma} - \sqrt{1-\rho}\sqrt{T}\sigma, \tag{13}$$

$$d_4 = -\frac{\log F_2(\mu) - a^{\tau}_{2T-\tau}}{\sqrt{1-\rho}\sqrt{T}\sigma}. \tag{14}$$

Thus, after the rollover into new loans, there are two state variables to keep track of that $A^{\tau}_{2T-\tau}$ and F_2 depend on: first, the change of the common factor since rollover, $W_{2T-\tau} - W_{T-\tau}$, and second, $L^{\tau}_{T-\tau}$, which in turn is driven by $W_{\tau} - W_{\tau-T}$.

The payoffs in Equations (12) and (7) together allow us to describe the distribution of the bank's assets. Consider, for example, the aggregate value of the bank's loan portfolio at $t = H$, where $H < T$. Aggregating across all loans outstanding at this time, we obtain

$$V_H = \sum_{\tau < H} e^{-r(\tau+T-H)} E^{\mathbb{Q}}_H[L^{\tau}_{2T-\tau}] + \sum_{\tau \geq H} e^{-r(\tau-H)} E^{\mathbb{Q}}_H[L^{\tau}_{T-\tau}], \tag{15}$$

where the first term aggregates over cohorts whose loans have been rolled over into a second round, whereas the second term aggregates over cohorts that still have the initial first-round loans outstanding. Substituting in from Equations (7) and (12) yields an expression in which the only source of stochastic shocks is the common factor W. Therefore, by simulating W, we can simulate the distribution of V_H under the risk-neutral measure and price contingent claims whose payoffs are functions of V_H.

Now suppose the bank has issued zero-coupon debt maturing at $t = H$ with face value D. Similar to standard structural models, we assume that the bank will pay off its creditors in full if there are sufficient assets available to do so. The bank will default if the asset value at maturity is lower than the face value of debt. To allow for a realistic calibration in our empirical exercise below, we also introduce dividend payouts of the bank to its shareholders. For simplicity, we introduce them as a single payment, just before the bank's debt matures, proportional to the value of the bank's assets at $t = H$,

$$Y_H = V_H(1 - e^{-\gamma H}), \tag{16}$$

The parameter γ determines the payout level. The assets that leave the bank through these payouts are no longer available to pay off the debt holders at $t = H$. In this aspect, our model is similar to standard implementations of the Merton model, where a constant rate of dividends is paid until debt maturity and before the debt holders can access the assets.

Panel (a) of Figure 2 shows the simulated bank asset value, V_H, based on 10,000 draws of the common factor paths plotted against the aggregate borrower asset value at the time of bank debt maturity $t = H$. Parameters are set to $N = 10$, $H = 5$, $T = 10$, $\sigma = 0.2$, $\rho = 0.5$, $r = 0.01$, $\delta = 0.005$, $\ell = 0.66$, $\gamma = 0.002$, and

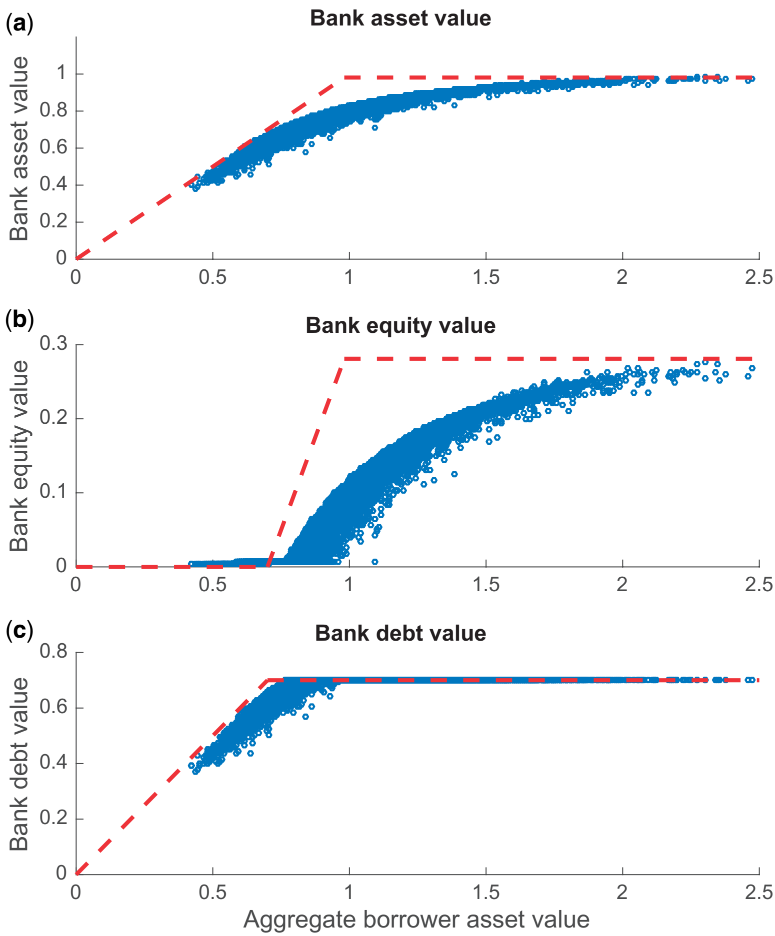


Figure 2
Bank asset, equity, and debt value at bank debt maturity as a function of aggregate borrower asset value at debt maturity
Simulated bank asset values are shown as dots. The dashed lines show the kinked payoffs that would result with perfectly correlated borrower asset values, without staggering of loan maturities and with identical maturity of bank loans and the bank's debt as in Figure 1.

$D = 0.70$, with initial collateral asset value of $A^{\tau,i}_{-\tau} = 1$ for all cohorts. The dashed lines show the payoffs that would result with perfectly correlated borrower asset values, without staggering of loan maturities and with identical maturity of bank loans and the bank's debt as in Figure 1. As the scatterplot shows, the simulated asset values in our model also exhibit concavity, but without the sharp kink that we had in the simplified case in Figure 1. There are two reasons for the lack of a sharp kink. First, at $t = H$, many loans in the bank's portfolio are not at maturity. For $\tau < T - H$, loans have not matured yet, while for $\tau > T - H$, they have been rolled over into new loans. Hence, the value of these nonmatured loans reflects an expectation, which smooths the kink. Second, the existence of idiosyncratic borrower risk makes the borrower's default option more valuable and the loan less valuable to the bank, particularly when the asset value is close to the face value of the debt.

Moreover, unlike in Figure 1, there is dispersion in the bank asset value conditional on the aggregate borrower asset value. The reason is that for loans that have been rolled over into a second generation of loans, the face value of the loan depends on the path of common factor realizations up to the rollover date $T-\tau$. For example, if $W_{T-\tau}$ is low, there will be more defaults and hence the amount of loans reissued will be lower than if $W_{T-\tau}$ is high. In contrast, if $W_{T-\tau}$ is high and there are virtually no defaults, the face value of the maturing loans is reissued as new loans, but borrowers remove collateral to leave just enough to satisfy the required loan-to-value ratio ℓ. Thus, the bank asset value at $t=H$ depends not only on the level of W_H but also on the path that W followed leading up to $t=H$.

Panel (b) of Figure 2 shows the simulated bank (ex-dividend) equity value,

$$S_H = \max[0, V_H - Y_H - D]. \tag{17}$$

The mezzanine nature of equity is clearly apparent from convex-concave payoff pattern, but the payoff function is smoothed compared to the sharply kinked one in Figure 1. Finally, the bank debt values,

$$B_H = V_H - Y_H - S_H, \tag{18}$$

are shown in panel (c).

The value of the bank's assets, debt, and equity (including the claim to the dividends to be paid just before maturity) $t=0$ then follow as

$$V_0 = e^{-rH} E_0^{\mathbb{Q}}[V_H], \quad B_0 = e^{-rH} E_t^{\mathbb{Q}}[B_H], \quad S_0 = e^{-rH} E_t^{\mathbb{Q}}[S_H] + (1-e^{-\gamma H})V_0. \tag{19}$$

Figure 3, panel (a), shows simulation results for the relationship between S_0 and aggregate borrower asset value. To explore the effect of unanticipated changes in borrower asset value, we set common factor shocks until $t=0$ to zero, we apply a single shock dW_0 at $t=0$ and simulate W from then on forward. More precisely, we set the shock for each cohort equal to dW_0 times the fraction of the loan's life, τ/T, that is completed at $t=0$. This captures the notion that the shock has accumulated over the life of the loan but remained unobserved until its revelation at $t=0$. We vary dW_0 from -0.8 to 0.8 across simulations. As a consequence, we generate variation in aggregate borrower asset value across simulations.

Panel (a) shows the value of bank equity is concave in borrower assets for large values. This is in contrast to the standard Merton model in which the equity value asymptotes toward a slope of one. Thus, we again obtain a mezzanine-like shape, similar to the one in panel (b) of Figure 1.

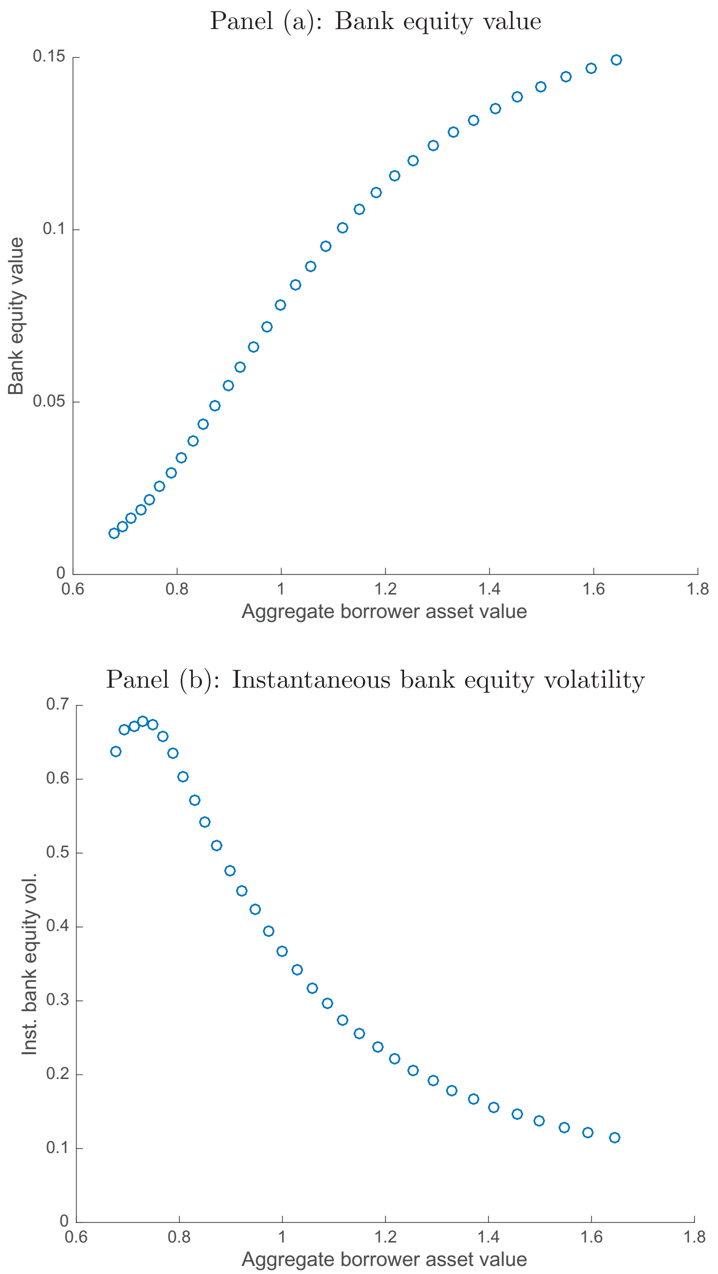


Figure 3
Bank equity value and volatility as a function of aggregate borrower asset value prior to bank debt maturity

Figure 3, panel (b), shows the instantaneous volatility of the bank's equity. Given knowledge of the parameters, one can compute the instantaneous equity volatility as the product of the numerical first derivative of $\log S_0$ with respect to $\log W_0$ and the instantaneous common factor shock volatility $\sqrt{\rho}\sigma$. Because common factor shocks are the only source of stochastic shocks to the bank equity value, the derivative of $\log S_0$ with respect to $\log W_0$ is directly related to the slope of the curve in panel (a). As Panel (b) shows equity volatility converges toward zero for high borrower asset values as the bank's loan portfolio becomes perfectly safe. In the Merton model, in contrast, equity volatility would asymptote toward the (strictly positive) volatility of assets.

This very low equity volatility at high borrower asset values arises from the mezzanine-claim nature of bank equity. Positive shocks raise borrower asset values far above the default thresholds. As a result, bank assets have very low instantaneous risk and bank equity risk resembles the risk of a defaultable bond because the region of concavity dominates. Application of a standard Merton model with lognormal asset value would miss this nonlinearity in bank's equity risk. Our modified model suggests that this nonlinearity is a key property of bank equity risk dynamics.

In particular, our model makes clear that low instantaneous bank equity volatility in good times can quickly turn into high risk in bad times if asset values fall. In the standard model with a lognormal asset value, a fall in asset values would only trigger a moderate rise in bank equity volatility because bank asset volatility is fixed. In our modified model, bank asset volatility goes up as loans fall in value and become riskier. The rise in equity volatility following bad shocks is therefore more dramatic than in the standard Merton model.

The figure also shows that equity volatility is nonmonotonic in asset value. At very low asset value, equity volatility declines as the asset value is lowered. This feature is due to the assumption about dividends (that our model shares with the Merton model): at very low asset values, a substantial portion of the equity value represents the value of the claim to the dividend that is to be paid before the bank's debtholders are paid off. Because of this priority over debtholders, this part of equity payoffs is less risky.

1.1 Default risk assessment: Comparison with the standard Merton model

The highly nonlinear relation between borrower asset value and bank equity risk due to the short put option embedded in bank assets leads to important consequences for distance to default estimation and empirical assessment of default risk. To illustrate these consequences, we now analyze a setting in which our modified model represents the true data generating process. We simulate from our model with parameter values set to the same values as above in Figures 2 and 3. We then study to what extent an analyst applying the (misspecified) standard Merton model would arrive at misleading conclusions about bank default risk.

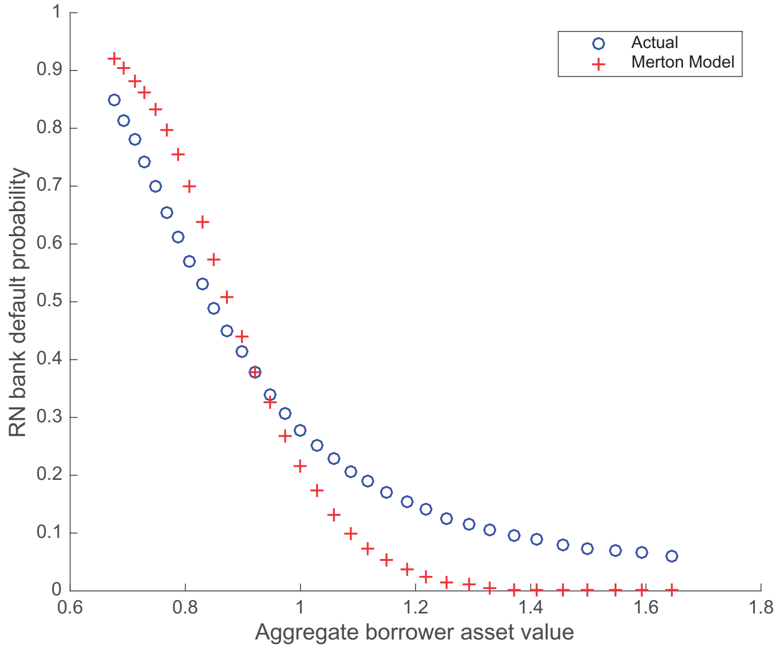


Figure 4
Risk-neutral default probabilities as a function of aggregate borrower asset value: Merton (red) and modified (blue)

Figure 4 shows the simulated true risk-neutral default probabilities (RNPD) in our model (blue) and those estimated based on the Merton model (red) applied to our simulated data. The Merton model default probabilities are obtained by using the simulated equity values and instantaneous volatilities to extract asset values and asset volatilities under the (false) assumption of a lognormal asset value process (see Appendix A.1). This corresponds to the common practice of inverting the Merton model to obtain asset value and asset volatility from empirically observed equity value and volatility. As the figure shows, the Merton model underestimates the probability of default for moderate and low default probabilities. In very good states of the world, the default probabilities are massively understated when the (misspecified) Merton model is applied.

Our modified model produces different predictions for two main reasons. First, the Merton model misses the nonlinearity coming from the mezzanine nature of bank equity. If borrower asset values are relatively high, bank equity volatility is very low because bank asset volatility is very low. However, asset volatility could quickly rise if asset values fall. As a consequence, the bank could reach the default threshold much more quickly than one would think based on the Merton model. Low instantaneous equity volatility hence does not mean that the bank operates at a high distance to default. However, an analyst applying the standard Merton model with constant bank asset volatility would miss these nonlinear risk dynamics. Within the standard Merton model, the analyst would interpret the low instantaneous equity volatility as a high distance to default and hence low default risk. Thus, particularly in good times, application of the Merton model is likely to lead to severe underestimation of bank default risk. Only in a severely distressed situation, when asset values are depressed and default is quite likely, the Merton model overstates risk-neutral default probabilities. Here the above effect works in reverse.

Second, our model features revolving replacement of staggered loans with a collateral reset. When loans roll over in good times after collateral values have risen, some of this collateral is removed between $t=0$ and $t=H$ as new loans are issued at a fixed loan-to-value ratio. Compared with the Merton model, the collateral reset dampens the risk reduction coming from rising asset prices. By the same token, when asset values fall, borrowers replenish collateral when new loans are issued. This dampens the rise of the bank's default risk when asset prices fall.

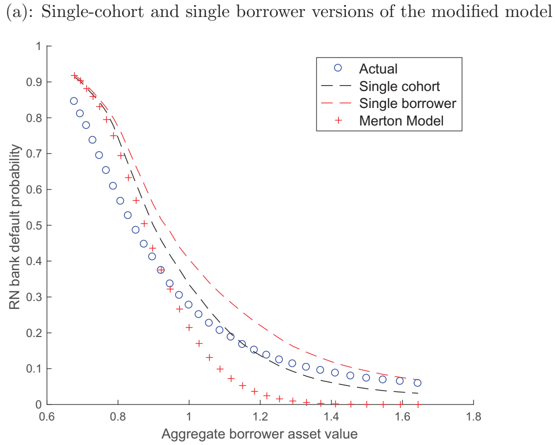
To show in more detail how these different model assumptions play out in generating the wedge in RNPDs between our modified model and the Merton model, the next subsection presents simplified versions of our model that shut off some of the features that differ from the Merton model.

1.2 Decomposing deviations from the Merton model

We start by considering a version of our modified model that retains the asset payoff nonlinearity induced by the borrower default options embedded in the loan portfolio, but it only has a single cohort of borrowers. Without loan rollover, the collateral reset effect disappears. We set $T=5$, equal to the average maturity of loans in our full modified model. The remaining parameters remain unchanged.

The black-dashed line in Figure 5, panel (a), presents the results for this simplified single cohort model. For comparison, we also plot the RNPD from the full modified model and the Merton model. As in Figure 4, the aggregate borrower asset value on the x-axis is the aggregate borrower asset value in the full modified model. For each of these asset values, we compute the RNPD as well as the equity value and volatility from our full modified model. We then choose the borrower asset value at $t=0$ and the loan face value in the single cohort version of the model to exactly match these equity values and equity volatilities.

Over a wide range of relatively high borrower asset values toward the right-hand side of the figure, the simplified model produces a RNPD of only about two thirds to half of the true RNPD, even though at every point the simplified model exactly matches the equity value and equity volatility implied by the full modified model. In line with our discussion of the collateral reset effect above, the simplified model ignores the loss of excess collateral upon loan rollover



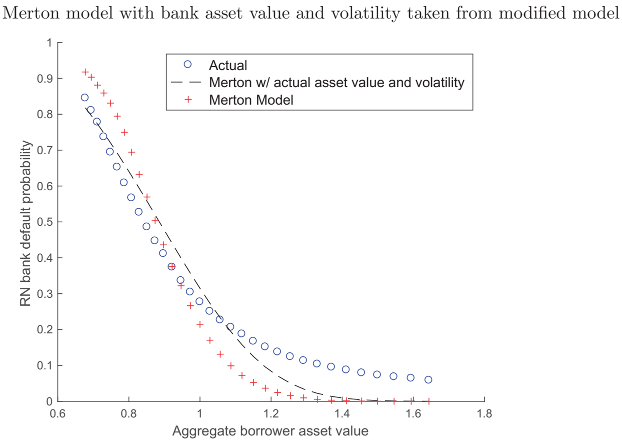


Figure 5
Imperfect approximation with simplified models

and hence underestimates bank credit risk in good times. For very low asset values, the bank effectively becomes an owner of the borrower's assets, so the payoff nonlinearity disappears (as the borrower's assets are far to the left of the kink in the bank's payoff). As a consequence, the RNPD of the single cohort model approaches the RNPD of the Merton model for very low asset values. In our full modified model, this also happens eventually, but for lower asset values than in the single cohort model because of collateral replenishment in the event of loan rollover. The collateral reset is an important ingredient of our model and it is arguably a realistic one.

The red-dashed line in Figure 5, panel (a), presents another simplified version with only a single borrower. The only remaining difference to the Merton model is the bank asset payoff nonlinearity. We again set $T=5$, equal to the average maturity of loans in our full modified model. To let the assets of the single borrower have the same volatility as the aggregate borrower asset portfolio in the full model, we set $\sigma = 0.2 \times \sqrt{0.5}$. With a single borrower, the nonlinearity in the bank's asset payoff is more pronounced than in the single cohort model. At the same level of equity value and equity volatility, we obtain a higher level of tail risk and hence default risk than in the single cohort model. This is why, as the figure shows, the RNPD is substantially higher than in the single cohort model, especially for moderately high asset values where the kink in the bank asset payoff function matters most.

Figure 5, panel (b), considers an alternative simplification that turns off the bank asset payoff nonlinearity by sticking to the standard Merton model. However, rather than inferring asset value and asset volatility from equity value and equity volatility, we plug in the correct bank asset value and (instantaneous) asset volatility that are implied by our modified model.[2] Instantaneously, this simplified model correctly matches the asset risks of our full modified model. However, in terms of the longer-term risks, the model is misspecified because it misses the asset payoff nonlinearity and the resultant risk of changes in future volatility. The plot in panel (b) shows that this leads to drastically different default risk predictions compared with the full modified model. Especially in good times, when asset values are high, the simplified model's predictions are very close to the Merton model and far from our full modified model. Hence, the payoff nonlinearity is very important for default risk prediction, even after inputting the correct current bank asset value and instantaneous volatility.

Finally, another potential simplification approach that might seem promising is to approximate the full modified model's RNPD with a nonlinear transformation of the Merton model RNPD. For example, from Figure 4 a monotone nonlinear transformation of the Merton model RNPD may seem to be all that is needed to achieve the modified models RNPD. However, this is not the case. In Figure 4, we only change the borrower's asset value (dW_0), but we keep the bank's level of debt, D, fixed. But if D varies too—as it would in any typical empirical application with heterogeneously levered banks—the Merton model RNPD does not one-to-one map into the modified model's RNPD.

Figure 6 illustrates this by plotting the RNPD of our modified model against the RNPD obtained from applying the (misspecified) Merton model to data generated from the modified model. In this figure, we vary dW_0, as in Figure 4, which affects the riskiness of the bank's assets and its leverage, and we also vary D, which changes only the bank's leverage, keeping the level and risk of its assets fixed. Each point on the scatterplot represents one (dW_0, D) combination. As the figure shows, there is no one-to-one correspondence between modified

[2] One could imagine that a bank examiner in practice might be able to come up with asset values and volatility estimates by carefully valuing assets bottom-up and looking at short-term volatility of traded proxies for these assets. In this sense, an examination of whether this variant of the Merton model with corrected asset value and volatility could work well as an approximation is also practically relevant.

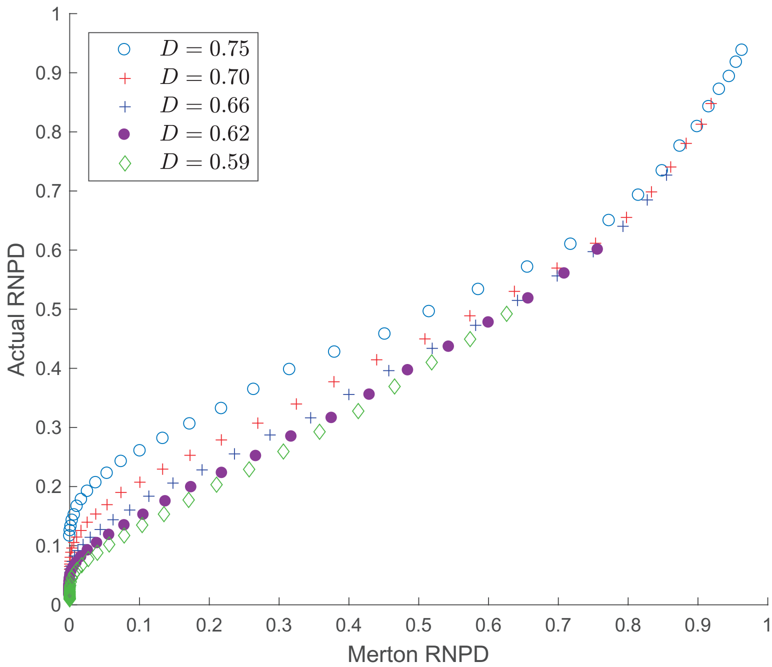


Figure 6
Risk-neutral default probabilities in the modified model (actual) and Merton model

model and Merton model RNPD. A given Merton model RNPD could map into different levels for the modified model RNPD depending on the level of D. The payoff nonlinearities induced by the two layers of leverage within the modified model are too complex to be captured by standard Merton model RNPDs. A change in dW_0 changes the moneyness of the borrower's put option and hence the extent to which assets of the bank deviate from the lognormal assumption of the standard Merton model. In contrast, changing D does not change these asset properties, it only changes the leverage of the bank's balance sheet.

Figure 6 further shows that the wedge between our modified model's RNPD and the Merton model RNPD increases with the bank's leverage. Intuitively, when a bank's leverage is low, borrower asset values need to fall a lot for the bank to advance toward default. In this default-relevant region, where many of the bank's borrowers are distressed the bank's asset payoffs are close to linear. As a consequence, the misspecification error from using the Merton model is small. In contrast, when bank leverage is high, the nonlinearity in the bank's asset payoff is very important.

Overall, these results show that one cannot easily simplify our model without substantial effects on the default risk predictions. Aside from the effect on default risk predictions, the single-borrower and single-cohort versions of our

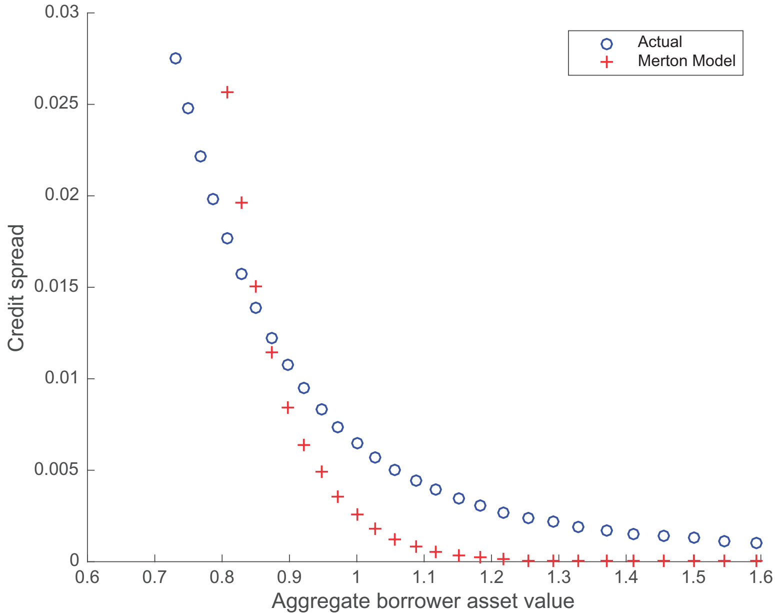


Figure 7
Implied credit spread: Merton (red) and modified (blue)

models are also somewhat awkward in that one cannot change the maturity of the bank's debt, H, without also changing the remaining maturity of the borrower's debt, and hence the risk profile of the bank's assets, at the time the bank's debt matures. The stationary overlapping cohorts setup in our full modified model ensures that the risk profile of the bank's assets at $t=H$ is invariant to the choice of H.

1.3 Pricing of credit instruments

We now return to our full modified model to explore the pricing of credit instruments. Figure 7 takes our baseline case of $D=0.70$ and presents the annualized implied credit spread of the bank's 5-year debt. The credit spread reflects the product of the risk-neutral probability of default (as shown in Figure 4) and the loss given default (which equals one minus the recovery rate). The recovery rate in a Merton-style models can often be quite high because the asset value in default could be just slightly below the face value of the debt. As a consequence, the implied credit spreads are much lower than $1/H$ times the risk-neutral default probabilities (which would be the annualized credit spread with zero recovery). However, by assuming constant asset volatility, the Merton model misses the rise in the bank's asset volatility that is associated with a fall in asset values toward the default boundary. As a consequence, the model underestimates the risk that, conditional on default, asset values could

Table 1
Summary of simulation results

	Borrower asset value shock		
	No shock	Positive shock	Negative shock
A. True properties			
Aggregate borrower asset value	1.06	1.33	0.85
Bank asset value	0.74	0.79	0.66
Bank market equity/market assets	0.12	0.16	0.07
Bank 5Y RN default prob.	0.23	0.11	0.49
Bank credit spread (%)	0.50	0.19	1.39
B. Misspecified estimates based on standard Merton model			
Merton 5Y RN default prob.	0.13	0.01	0.57
Merton credit spread (%)	0.12	0.00	1.50

be far below the face value of the debt. Recovery values in our modified model tend to be lower, which contributes to the higher credit spread.

Table 1 provides the numbers corresponding to some of the points in Figure 7. As the table shows, the differences between the true credit spreads and those extracted via the Merton model are bigger than the differences in risk-neutral default probabilities (RNPD) in Figure 4. For example, in the absence of substantial positive or negative shocks to asset values, with aggregate borrower asset value slightly above one, the true credit spread of 50bp is more than four times the spread of 12bp extracted using the Merton model. In comparison, with 0.23 and 0.13, respectively, the RNPD are only moderately different. The divergence is bigger for credit spreads because the Merton model not only underestimates default probabilities but also overestimates recovery rates. The difference becomes more extreme in good times. With aggregate borrower asset value of 1.33, the true credit spread is 19bp while application of the Merton model yields a spread that is essentially zero. Thus, during economic booms, the application of the standard Merton model could lead an analyst to the conclusion that banks' credit risk is virtually nil, when in fact it is still far from negligible.

Figure 8 shows how application of the standard Merton model would severely underestimate the value of a government guarantee. For illustration, we suppose a 50% risk-neutral probability that the government will fully bail out the debt holders of the bank (and absorb the entire loss given default) in the event of default. The value of the government guarantee then is 0.5 times the value of the bank's default option. To interpret the magnitudes in Figure 8, recall from Table 1 that when the aggregate borrower asset value is around 1.06, the value of the bank's loan portfolio is about 0.74. The value of the guarantee in this case is about 0.01, that is, about 1% of the value of the bank's assets. Estimation based on the (misspecified) Merton model would lead an analyst to conclude that the value is 0.002, that is, roughly one-fifth of the actual value. As in the case of credit spreads, the difference between the actual and Merton-implied values widens in good times when borrower asset values are high.

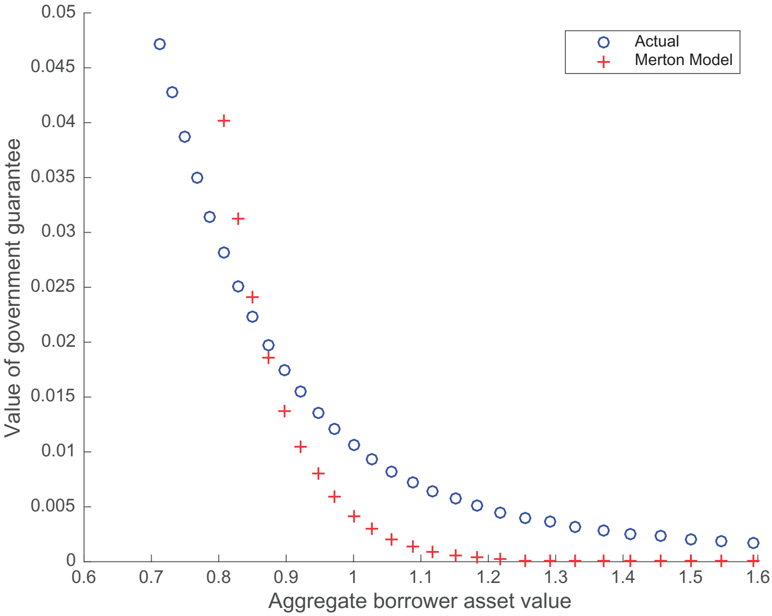


Figure 8
Value of a government guarantee: Merton (red) and modified (blue)

2. Empirical Calibration

To find out how much, quantitatively, the standard Merton model and our modified model differ in their predictions about default probabilities and risk dynamics, we now calibrate these models with empirical data on bank's capital structures and equity volatility.

2.1 Data

Our sample covers all commercial banks listed in the Federal Reserve Bank of New York's CRSP-FRB-linked data set from 1987 to 2016 that are also covered by Compustat Quarterly bank files. We obtain equity returns and market value of equity from CRSP and the accounting data from the Compustat Quarterly files for banks. We take the most recently available data from the quarterly files as of the beginning of the estimation month. We consider the entire outstanding debt of the bank (including demand and time deposits) in our calculation of the debt face value and we do the same for the Merton model.[3] We also add the

[3] Note that this is different from several earlier papers that estimate distance-to-default following the KMV approach of including short-term debt and only *half* of long-term debt in the debt face value calculation. The rationale for the KMV approach is that a substantial portion of long-term debt does not mature and hence won't trigger default during the horizon that is used to calculate default probabilities. Although this may be a reasonable approach for nonfinancial firms, it's less plausible for banks. Banks are funded to a large extent by short-term debt (including

Table 2
Model inputs

	Mean	SD	Min	25th pctl	Median	75 pctl	Max
Market equity	0.15	0.14	0.00	0.10	0.14	0.19	16.09
Equity volatility	0.29	0.10	0.17	0.23	0.27	0.32	0.65
Risk-free rate	0.04	0.02	0.01	0.03	0.04	0.06	0.09
Observations	45,077						

Our sample covers all commercial banks listed in the Federal Reserve Bank of New York's CRSP-FRB-linked data set from 1987 to 2016 that are also covered by Compustat Quarterly Bank Database. Market equity values are normalized by D; that is, the numbers shown in the table are based on the market-equity-to-book-debt ratio. Equity volatility is annualized and estimated from daily stock returns over 1-year moving windows. Our risk-free interest rate proxy is the Federal Reserve Board's 10-year Treasury bond yield series.

book value of preferred equity to the firm's debt if the bank has any outstanding preferred equity. However, relaxation of this assumption makes no qualitative difference to our main results because the amount of preferred equity is very small (less than 0.25% of total assets on average) for the entire sample. To be included in the sample, the bank-year observation must have nonmissing information on book value of debt and a positive value of book equity. To compute the face value of debt from its book value consistent with the zero-coupon debt assumption in the model, we multiply book debt by $\exp(rH)$ where H is the debt maturity we assume in the calibration. Appendix B provides more details on the construction of the bank-level variables.

To calibrate the modified model, we need a risk-free interest rate. Both the Merton model and our modified model abstract from interest-rate risk and a term structure of default-free interest rates. As an approximation, we use the Federal Reserve Board's 10-year Treasury yield series.

We also need a good estimate of the conditional equity volatility of the bank on each estimation date. To do so, we first compute the realized value of equity volatility (in annualized form) from daily bank stock returns over backward-looking 1-year moving windows. We then regress these realized volatilities on their 12-month lagged values in a panel regression. We use the fitted values from this regression as proxy for the (forward-looking) conditional equity volatility in our default risk calculations. Appendix C provides further details on this computation. Table 2 provides summary statistics of key inputs used in our estimation.

2.2 Model calibration

For both the standard Merton model and our modified model, we set the maturity of bank debt to $H=5$. It is well known that models with only diffusive shocks do not succeed in delivering realistic default risk and credit spread predictions for short-term debt (Duffie and Lando 2001; Zhou 2001). Our modified model

demand deposits and time deposits) and rollover of short-term debt would likely fail if the outstanding long-term debt renders the bank insolvent. Further, our analysis is in line with capital requirement regulations that are based on total outstanding debt of the bank.

Table 3
Parameters

Parameter	Description	Value
δ	Borrower asset depreciation rate	0.005
γ	Bank payout rate	0.002
T	Bank loan maturity	10 years
H	Bank debt maturity	5 years
ρ	Borrower asset value correlation	0.5
ℓ	Loan-to-value ratio	0.66
σ	Borrower asset volatility	0.20

is no different. Therefore, at short horizons differences between our modified model and the standard model are also relatively small. The nonlinearity in banks' asset payoffs becomes more relevant as the probability distribution of borrower asset values spreads out with longer horizons. Longer horizons also may be relevant even for investors in short-term debt (or guarantors of short-term debt). Solvency problems may not be immediately apparent when bad shocks are realized. Deterioration in asset values may be hidden for a while, perhaps facilitated by regulatory forbearance, and short-term debt may be rolled over even if the bank is actually insolvent. By the time default happens, additional losses may have accumulated.[4]

Based on empirical estimates, we fix the payout rate $\gamma = 0.002$ for all bank-year observations for both the Merton model and our modified model.[5]

We calibrate the standard Merton model by simultaneously solving for asset value and asset volatility that deliver the observed values of a bank's equity and stock return volatility (see Appendix A.1). This approach has been used by prior researchers, such as Jones, Mason, and Rosenfeld (1984), Vassalou and Xing (2004), Campbell, Hilscher, and Szilagyi (2008), and Acharya, Anginer, and Warburton (2014). We solve the model quarterly from 1987Q1 to 2016Q4.

For our modified model, we have several additional parameters that we fix exogenously, as shown in Table 3. We set the depreciation rate $\delta = 0.005$. We assume that the borrower's asset volatility is 20%. This is in line with the implied asset volatility estimates of 17%–21% by Stanton and Wallace (2018), who extract these implied volatilities from newly issued mortgages in the commercial mortgage market. We assume that borrower asset values have a pairwise correlation of $\rho = 0.5$. This implies a factor volatility of $\sqrt{0.5} \times 20\% \approx 14\%$. For comparison, based on unlevered returns of an aggregate index of Real Estate Investment Trusts (REIT), Ling and Naranjo (2015) find an implied asset

[4] Earlier literature in deposit insurance pricing dating back to Merton (1977) often interprets debt maturity as the time until the next regulatory audit, which is typically 1 year or less. See, for example, Marcus and Shaked (1984) and Ronn and Verma (1986). However, in the presence of regulatory forbearance, even if a bank is undercapitalized on an auditing date, it may be allowed to continue for a longer period without additional capital replenishment. This, in turn, justifies a maturity exceeding the time until the next audit.

[5] The payout ratio, computed as the ratio of cash dividend on common equity to the book value of assets, has a mean of 0.0023 and a median of 0.0018 during our sample period.

Table 4
Model-implied risk-neutral probabilities of default

	Mean	SD	Min	25th pctl	Median	75 pctl	Max
Merton model PD	0.14	0.19	0.00	0.03	0.07	0.15	0.98
Modified model PD	0.25	0.16	0.04	0.15	0.21	0.31	0.89
Observations	45,077						

We calibrate the Merton model and our modified model quarterly from 1987 to 2016 based on the data summarized in Table 2. For each bank in each calibration period, we compute the risk-neutral default probability from the two models. The table reports summary statistics for these risk-neutral default probabilities for the whole panel of banks over the full sample period.

volatility of slightly more than 10%. It seems reasonable to assume that bank loan portfolios are not always as well diversified as market-wide REIT portfolio (e.g., in terms of geographic exposure), and so a moderately higher factor shock volatility in our calibration seems appropriate.[6] We fix the loan-to-value ratio at loan origination at 0.66. Finally, we assume that the maturity of loans issued by banks is $T = 10$ years.

Just like the standard Merton model treats asset value and volatility as unobservable, we treat dW_0 (shock to borrower asset values after loan was issued) and F_1 (face value of borrowers' loans) as unobservable. By changing dW_0 and F_1 we can change the value and volatility of the bank's assets. For example, because the loan-to-value ratio at loan origination is fixed, raising F_1 raises the value of the bank's loan portfolio, leaving its volatility constant. Raising dW_0 raises the value of the loan portfolio, while reducing its riskiness. Empirically, we look for values of dW_0 and F_1 that allow us to match the empirically observed equity value and equity volatility of the bank with the model-implied value. Appendix A.2 provides more detail on the invertibility of the mapping from dW_0 and F_1 to equity value and volatility.

2.3 Model-implied risk-neutral default probabilities

We calibrate each model quarterly from 1987 to 2016. We follow standard practice of calibrating the models each quarter, without imposing restrictions across time. We winsorize the estimated RNPDs from both models at 0.5% from both tails to minimize the effect of outliers in our estimation. Table 4 gives the summary statistics of the two calibrated models' implied risk-neutral default probabilities (RNPD) for our panel of banks. The average 5-year RNPD is higher by about 10 pp in our modified model compared with the standard Merton model. Further, the Merton model RNPD is much more positively skewed. The reason is that when a bank is not in distress, the implied RNPD in the Merton model is very low because it is based on the assumption that the bank's assets have constant volatility. In contrast, our model takes into account that bad shocks to asset values in the future would drive up the volatility of

6 By matching borrower's asset volatility to unlevered REITs and assuming a loan maturity of 10 years, our calibration exercise assumes that banks have heavily invested in mortgages or mortgage-related instruments.

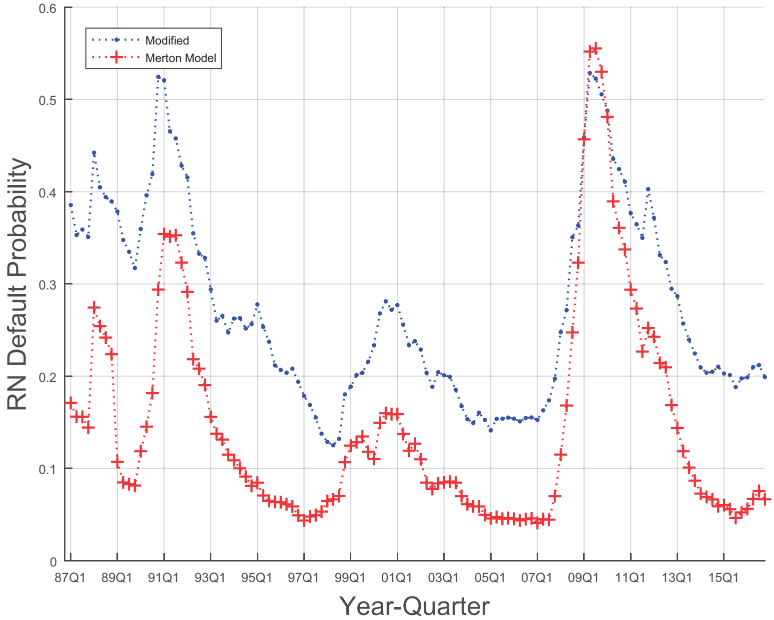


Figure 9
Comparison of calibrated risk-neutral default probabilities (5-year horizon, cumulative)

the bank's assets (as borrowers' default options move into the money), which drastically shrinks the distance to default and hence raises the RNPD in times when asset values are relatively high.

Figure 9 further illustrates the different behavior of the RNPD from the two models over time. The figure shows the average RNPD across all banks each quarter from 1987 to 2016. The modified model's RNPD is two to three times as high as Merton RNPD during the time before the financial crisis of 2007-2008. Expressed in terms of annualized credit spreads, the RNPDs in 2006 would correspond, roughly, to 5 bps for the Merton model and around 40 bps for our modified model.

This behavior of the relative RNPDs is in line with the intuition that bank assets have nonlinear debt-like payoffs. In good times, the Merton model's assumption of constant asset volatility produces very high distance to default and low RNPD. In our model, the analysis recognizes that the volatility of bank assets in good times (in the flat part of the concave asset payoff region in Figure 2) is low, but that it can quickly rise after a bad shock (when the bank enters the downward-sloping asset payoff region). Once the asset value has suffered a sufficiently big bad shock, the asset payoff is dominated by the linear downward-sloping region and the kink is not playing much of a role. In this case, the predictions of the Merton model and our modified model are relatively similar. This is why after the onset of the financial crisis in 2007,

Table 5
Differences in model-implied risk-neutral default probability: Comparison between high- and low-VIX periods

	(1)	(2)	(3)	(4)
Low VIX	0.55		0.52	0.84
	(7.63)		(7.81)	(6.44)
Equity		0.31	0.25	0.20
		(5.55)	(5.26)	(3.77)
Low VIX x Equity				0.16
				(2.34)
Observations	45,059	45,059	45,059	45,059
Within-R^2	0.10	0.02	0.13	0.13
Absorbed FE	Bank	Bank	Bank	Bank
Clustered by	Bank yq	Bank yq	Bank yq	Bank yq

The dependent variable is the log risk-neutral default probability from our modified model minus the log risk-neutral default probability from the standard Merton model for our panel of banks from 1987 to 2016. Explanatory variables include a dummy for quarters with below median VIX index, and the bank's log market equity (normalized by total debt, as in Table 2). t-statistics, estimated with clustering at both bank and year-quarter (yq) level, are presented in the parentheses below the coefficients.

the difference between the RNPD shrinks and eventually inverts for some time period in 2008-2010. But as the economy recovers from the Great Recession, the estimates from the two model begin to diverge in 2014-2016. Similarly, the two models provide roughly similar estimates during the relatively stressful periods of 1999-2000. But the modified model produces a much higher RNPD during the 1993-1998 period when the banking sector performed well.

Going further back in time, the modified model's RNPD is in general higher than the Merton Model RNPD during the stressful years of savings and loans crisis (1987-1992). While this was a stressful time for the banking sector, the extent of distress was not as high as the recent financial crisis period. Second, the S&L crisis was spread out over a number of years in the late 1980s and early 1990s, with significant yearly variations in the extent of stress faced by the sector during this period. Our estimates reflect such variations. During some quarters, the estimates from the two models come close to each other just as in other stressful periods. Yet, in other quarters, our model provides significantly higher estimates of RNPD than the Merton model. Overall a clear pattern emerges from this figure: Merton model RNPDs are significantly lower than modified model RNPD especially during the good times of the economy.

We further illustrate this cyclical behavior of the RNPD differences between the Merton model and our modified model with the following panel regression:

$$\log \frac{RNPD_{it}^{Modified}}{RNPD_{it}^{Merton}} = \alpha_i + \beta_1 \times LowVIX_t + \beta_2 \times \log E_{it} + \beta_3 \times LowVIX_t \times \log E_{it} + \epsilon_{it}. \quad (20)$$

The dependent variable is the log difference in default probabilities for a bank i in quarter t. VIX is the CBOE index of implied volatilities on S&P 100 index options. We compute the average level of this index over the trailing 1 year for each estimation date. *LowVIX* equals one for quarters with below median VIX, zero otherwise. E measures each bank's market equity (normalized by total debt, as in Table 2). The regression results in Table 5 show that the modified model RNPD is significantly higher during "quiet" periods and for high-equity banks. Column 4 further shows the interaction effect: the standard Merton model delivers a lower default probability especially for banks with higher equity during low VIX periods.

The incremental explanatory power of individual banks' market equity in this panel regression also hints at the fact that there is substantial cross-sectional variation in the wedge between the standard model and the modified models' RNPD. Indeed, as we have shown in Figure 6, there is no simple linear or nonlinear mapping from the Merton model RNPD to our modified model's RNPD. This is also true of our empirical estimates, as we show in Appendix D. The empirical size of the wedge between the two RNPDs substantially varies with individual bank characteristics as well as the time-series factors that we have emphasized so far.

In our model we treat dW_0, that is, shocks to borrower asset values, as unobservable, and estimate the RNPD based on the values of dW_0 and F_1 (face value of borrower loans) that best match with the empirically observed equity value and equity volatility of banks. As a reality check, it would be useful to see whether the backed-out dW_0 series has plausible time variation: it should be high when borrower asset values are high. Figure 10 plots the average values of model implied dW_0 shocks with realized 5-year growth rate in national house price index of Freddie Mac.[7] As the figure shows, model-implied asset value shocks line up well with house price growth. In particular, both series show an upward trend during the 2001–2006 period; they both fall during the stressful periods of financial crisis in 2008–2009; and, finally, they both return to their upward trend in 2013–2016. Of course, in reality, a bank's asset portfolio is exposed to other assets as well, not just residential real estate, hence we do not expect the two series to be perfectly correlated. Moreover, house price indices are subject to smoothing and time lags in updating and so one should not expect measured house price growth rates to capture forward-looking expectations in the same instantaneous manner as bank equity values do. Nonetheless, the broad trend presented in Figure 10 shows that our estimation exercise extracts borrower asset values with reasonable time-series properties.

2.4 Model-implied equity risk dynamics

The motivation for our modification of the standard Merton model is based on a priori reasoning that the nature of bank asset payoffs is fundamentally inconsistent with a lognormal process. Improving the model on this dimension seems of first-order importance and should lead to an improvement in the empirical performance of the model. At the same time, it is clear that even

[7] See the FreddieMac house price index at http://www.freddiemac.com/finance/fmhpi/archive.html.

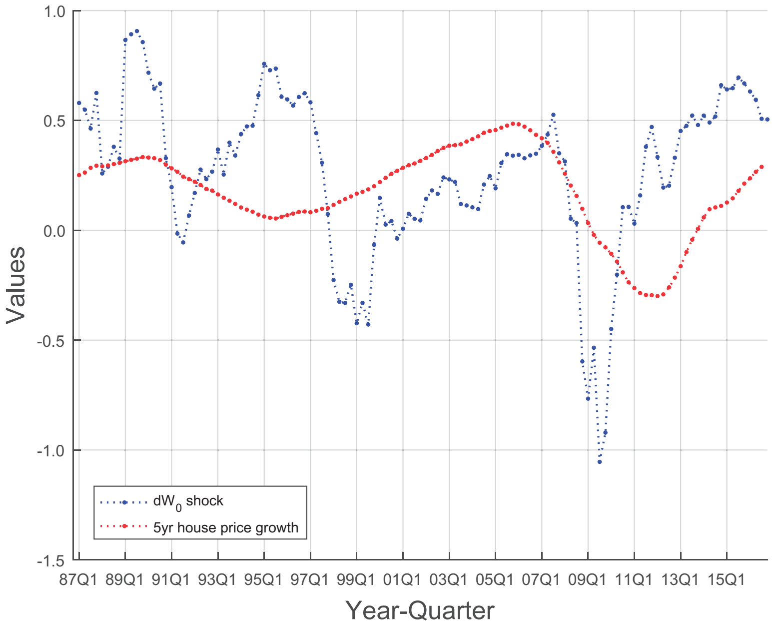


Figure 10
Comparison of asset value shocks in the model with actual house price growth rate

the modified model, in this simple form, is likely to miss important features of a bank's capital structure and of how a distressed bank enters into default. Among other things, our modified model does not take into account the presence of implicit and explicit government guarantees. Furthermore, bank capital structures are a lot more complex than our simple model allows for. A direct comparison of the model implied RNPD with bank CDS rates or credit spreads would therefore be difficult to interpret.

At this point, we do not focus on refining the model to account for these additional complexities. Instead, we evaluate the plausibility of our modification of the Merton model by studying the dynamics of bank equity risk. The risk of the equity claim should be less sensitive to government guarantees and interventions than default risk measures, as their main effect is on the more senior claims in a bank's capital structure.

The modified model and the standard model differ starkly in their predictions of how bank equity risk responds to asset value shocks. To study these differences, we subject bank asset values to a realistic negative asset value shock. We then calculate the models' equity volatility predictions, conditional on this shock, and we compare these predictions with actual data on the trajectory of bank equity volatilities going into the financial crisis around 2008/09. In the standard model, a shock to a bank's asset value leaves the asset volatility unchanged. In contrast, in our modified model a shock to a bank's asset value is associated with a shock to the asset volatility in the opposite direction. As a consequence, the volatility of the bank's equity return rises more in response to a negative asset value shock than in the standard model.

We start by calibrating both models to fit the precrisis data in 2006Q2 for each bank, as we did in our earlier analyses above. In our modified model we then apply a shock to the asset value of the bank's borrowers by modifying dW_0. To obtain a measure of asset value shock, we conduct two tests. In the first test, we take the cumulative log change in the Freddie Mac house price index for the entire country on a quarterly basis from 2006Q2 until a subsequent quarter t as the measure of dW_0 shock. This measure is likely to be a conservative estimate of the actual shock experienced by banks because the house price index may not correctly reflect the market values of assets during the crisis for several reasons such as sellers' reluctance to see the house, delay in foreclosure process, and other financial distress costs incurred by homeowners. Indeed, even during the peak of the financial crisis in 2008–2009, the cumulative percentage drop in the house price index was a modest 20%. In contrast, market-based estimates suggest a much steeper drop in asset values during this period. For example, Giacomini, Ling, and Naranjo (2015) estimate the unlevered return on a sample of U.S. REITs and report a drop of almost 40% from peak to trough during the crisis period. In our second approach, we conduct a stress test that shocks asset values by gradual amount till it drops by 40% by the middle of the financial crisis in 2009. The extent of shock is equivalent to a two standard deviation decline in asset value in our model, and therefore it represents a reasonable left tail event. This approach is akin to the scenario-based stress tests conducted by banking supervisors around the world.

Leaving all other parameters unchanged at the 2006Q2 values, we recalculate the risk-neutral probability distribution of the bank's loan portfolio payoffs, the loan portfolio value, and then the bank's equity value and equity volatility. For the Merton model, we start by calibrating the model to fit the precrisis data in 2006Q2 and we then subject the model to the asset value shock. To compare the Merton model and the modified model on an equal footing, we use the same shock to bank asset values; that is, we use the proportional change in post-shock loan portfolio value from our modified model as the proportional change in the post-shock bank asset value in the Merton model. The crucial difference is that the Merton model features a constant asset volatility. Thus, an analyst making predictions about equity volatility conditional on a shock to the bank's asset value would be led to assume that the asset volatility will remain at its 2006Q2 level.

Panel (a) in Figure 11 shows the trajectories of model-implied equity volatilities (annualized) from the two models, averaged across all banks in the data set when assets are subject to shocks based on Freddie Mac house price index. The differences are quite stark. Even though the drop in the banks' asset values is the same in both models, equity volatility rises only by a modest 8–9 pp in the Merton model, whereas it increases by about 14–15 pp in the modified

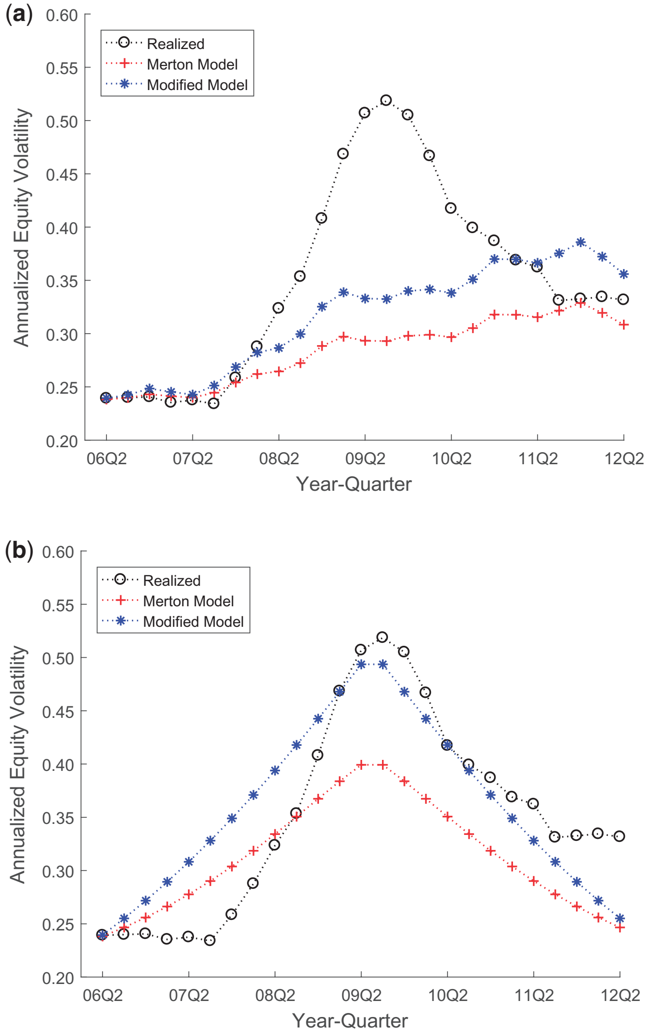


Figure 11
Bank equity volatility after a negative shock to borrowers' asset values

model. Compared to average equity volatility of about 25% in 2006Q2 for all banks in our sample, the modified model produces a substantial increase over time. In the Merton model, equity volatility rises only because of the leverage effect: the drop in banks' asset values leads to higher bank leverage, moving the bank's equity call option on the assets further out of the money. An additional effect occurs in our model: because the fall in bank asset values originates from a fall in borrower asset values, the loan portfolio becomes more risky and hence banks' asset values not only fall but also become riskier. The figure also plots the realized volatility during these quarters.[8] Actual volatilities went up even more than predicted by our calibration of the modified model. This is not surprising because the house price index shocks are likely to underestimate the magnitude of true shocks experienced by bank's assets for reasons mentioned above. In addition, our modified model does not take into account liquidity problems, runs, systemic risks, fire sales, and various other factors that may have led to strongly elevated levels of volatility at the peak of the financial crisis in 2008/09.

Panel (b) in Figure 11 shows the trajectories of model-implied equity volatilities from the two models when assets are subject to a negative 2-standard-deviation shock from their precrisis level. Specifically, starting from 2006Q2, we linearly increase the magnitude of shock such that it reaches a peak value of -40% in 2009Q2 and then gradually reverts to the original value by 2012Q2. The modified model-based equity volatility dynamics is now remarkably closer to the realized volatility. The Merton model-implied volatility, on the other hand, considerably underpredicts the equity volatility even after the full realization of shock in 2009Q2. By subjecting the Merton model to same asset value shock as the one implied by the modified model, we are giving the Merton model a much better shot at explaining the future volatility dynamics. If an analyst simply uses the historical distribution of bank asset shocks without regard to the payoff nonlinearity, he will assign a much smaller probability to such shocks to the asset value of banks. This exercise highlights the usefulness and importance of our modeling approach for stress tests and related counterfactual exercises.

3. Implications for Reduced-Form Models

The insights gained from our analysis are useful beyond the narrow confines of structural modeling of default risk. It is clear from the prior literature that reduced-form models outperform structural models in terms of default prediction performance and in matching market pricing of default risk. As Jarrow and Protter (2004) and Duffie and Lando (2001) argue, part of the reason is that the structural approach assumes, implausibly, that market participants observe a firm's asset value continuously. In many practical applications, a reduced-form model is therefore the preferred approach.

In reduced-form models, a firm's default intensity depends on a vector of state variables. The model is silent about the nature of these state variables. In applications, modelers typically choose covariates relating to the state of the economy and various balance sheet variables and other firm-level predictors of default as elements of the state vector. As Duffie, Saita, and Wang (2007) demonstrate, distance-to-default estimates obtained from structural models can

[8] We plot the average predicted volatility based on the regression model discussed earlier. See Appendix C for details.

be a useful default predictor within a reduced-from model (see also Bharath and Shumway 2008; Campbell, Hilscher, and Szilagyi 2008).

Thus, based on our analysis in this paper, one would expect that default probabilities obtained from our modified model should be a better predictor of bank default than the distance to default obtained from the standard Merton model. The extent to which this makes a difference should also depend on economic conditions. As we showed earlier, the differences in implied RNPD between our modified model and the Merton model are particularly stark in good times when borrower asset values are high.

We assess the relative performance the modified model and the standard Merton model by comparing their ability to predict (pseudo-)bank defaults. For this exercise, properly classifying banks into "default" and "nondefault" groups is crucial. We collect all bank failures during the sample period from the FDIC's failed bank list data set. However, this definition of failure misses several important default events in the sample. It is well known that several banks were in deep distress during the 2008–2009 financial crisis, but they did not default due to government bailouts or government-assisted mergers. These factors are outside of our model. To ensure that we are able to exploit the information contained in these events, we include all banks that experienced very low stock returns during 2008–2009 in the "default" category, in addition to those that actually failed. Specifically, we compute the cumulative stock returns of all banks in our sample in years 2008 and 2009 and classify banks with less than -40% return, that is, returns below the sample average of bank returns during this period, as defaulted. Based on this definition, 285 banks, or a little more than a quarter of all banks, are in the default category. Our results become slightly stronger if we adopt a more conservative cutoff for defaults such as classifying banks below the 75th or 90th percentile of the return distribution as defaults.

We estimate the following Cox-proportional hazard rate model for this test:

$$h(t|rnpd_{i,t}) = h_0(t)\exp(rnpd_{i,t} \times \beta_{rnpd}). \tag{21}$$

$h(t|rnpd_{i,t})$ is the hazard rate at time t conditional on the measure of RNPD estimated at the beginning of the period. We estimate the model at annual frequency with RNPD expressed in percent. As of January 1 of every year, we obtain the measures of RNPD using the standard and modified model, and use these measures to predict default that occurs during the year. Table 6 presents the estimation results. For ease of interpretation, we report one minus the hazard-ratio (i.e., $1-\exp(\beta_{rnpd})$) in the table. Thus, the reported coefficient provides the percentage increase in the odds of default (i.e., the ratio of the probability of default to the probability of no default) for a bank that has 1 pp higher RNPD.

Panel A of Table 6 shows that both measures provide meaningful information about the default likelihood. We find that a 1-pp increase in the estimated Merton RNPD results in an increase of 3.37 pp in the odds ratio for actual default probability, which is highly significant in statistical terms. However, with 4.40

pp, the corresponding effect for the modified model in Column 2 is higher. In Column 3 we include both measures of default and find that the modified RNPD is more important in predicting the eventual failure than the Merton RNPD. In this specification, the effect of the modified model RNPD remains statistically significant and economically large, whereas the Merton model RNPD now has an insignificant coefficient of 0.98 pp, which means that, holding the modified model's RNPD constant, it does not have much incremental ability to predict actual default.

We also provide the area under curve (AUC) for the receiver operating characteristic (ROC) analysis. The ROC curve gives a measure of the accuracy of any predictor by plotting the true positive rate (i.e., defaults in our context) against false positives for all possible cutoff points of a predictor. The larger the AUC, the better the predictor in distinguishing defaulters from nondefaulters: an uninformative predictor has a 50% AUC, whereas a perfect predictor has an AUC of 100%. The modified model has an AUC of 67.62% compared to 64.97% for the Merton model. The difference in AUC of 2.65 pp is statistically significant: a χ^2 test for the equality of the two areas has a *p*-value of .02.

The hazard rate model exploits default information from the entire sample period, but the overwhelming majority of defaults in the sample is clustered during the financial crisis of 2008–2009. To a large extent, the explanatory power of the two models in the hazard rate regression reflects the extent to which they were successful in predicting the huge increase in defaults during the financial crisis. Because the financial crisis quite clearly represented a large unanticipated aggregate shock, a more interesting exercise is to ask which of the models did better, based on precrisis information, in predicting which banks would be most affected by this shock.

In panel B of Table 6, we therefore present results from a single cross-sectional logistic regression focused on explaining (pseudo)-defaults during 2008 using the RNPDs estimated as of the beginning of the year as the explanatory variable. A one pp increase in the Merton RNPD is associated with an increase of 1.13 pp in the odds of default, but the coefficient is insignificant. In contrast, the modified RNPD has a statistically significant and economically strong effect with a coefficient of 3.52 pp. In line with our results based on the hazard rate model, when we include both these RNPDs as explanatory variables for default, it is only the modified RNPD that remains a strong predictor of actual defaults. Based on the cross-sectional test, we find a remarkable improvement in the area under the ROC for the modified model (62.74%) compared with that for the Merton model (54.89%). This indicates that our model is considerably better in extracting cross-sectional differences in default likelihood information. The cross-sectional results do not show out-of-sample predictions, because ex ante one would not have known the coefficient on the explanatory variables. Instead, the results show how the different models fare in discriminating between failed and surviving banks conditional on a crisis hitting.

Table 6
Default prediction: Hazards model

	(1)	(2)	(3)
A. Cox regression model			
Merton EDF	0.0337		0.0098
	(9.20)		(1.73)
Modified EDF		0.0440	0.0323
		(10.39)	(5.17)
Observations	9,809	9,809	9,809
# Banks	1,194	1,194	1,194
# (Pseudo-)defaults	285	285	285
Area under ROC curve	0.6497	0.6762	
B. Logistic regression model			
Merton EDF	0.0113		−0.0235
	(1.26)		(−1.89)
Modified EDF		0.0352	0.0517
		(3.40)	(4.09)
Observations	500	500	500
# (Pseudo-)defaults	226	226	
Area under ROC curve	0.5489	0.6274	

Panel A of the table presents estimates from a Cox-proportional hazard model. The dependent variable is a binary variable that indicates whether or not the bank has defaulted in the following year. The set of defaulted banks includes actual bank failures from 1987 to 2016 as well as pseudo-defaults, which include any bank with a cumulative stock return below -40% in 2008–2009. The table shows one minus hazard ratios (i.e., one minus exponentiated coefficients e^{β}) and associated z-statistics in parentheses. # banks and # (pseudo-)defaults represent the number of unique banks and (pseudo-) defaults in our sample. Panel B presents estimation results from a cross-sectional logistic regression model estimated for year 2008.

Table 7
Default prediction with equity-based classification of distress

	(1)	(2)	(3)
A. Cox regression model			
Merton EDF	0.0372		−0.0136
	(16.24)		(−2.72)
Modified EDF		0.0542	0.0723
		(19.58)	(10.60)
Observations	10,076	10,076	10,076
# Banks	1,194	1,194	1,194
# (Pseudo-)defaults	259	259	259
Area under ROC curve	0.8766	0.9010	
B. Logistic regression model			
Merton EDF	0.0448		−0.0719
	(3.57)		(−3.71)
Modified EDF		0.1374	0.2025
		(5.22)	(8.49)
Observations	500	500	500
# (Pseudo-)defaults	66	66	66
Area under ROC curve	0.6729	0.8639	

Panel A of the table presents estimates from a Cox-proportional hazard model. The dependent variable is a binary variable that indicates whether or not the bank has defaulted in the following year. The set of defaulted banks includes actual bank failures from 1987 to 2016 as well as pseudo-defaults, which include any bank with market-equity-to-book-debt ratio falling below 5% anytime during 2008–2009. The table shows one minus hazard ratios (i.e., one minus exponentiated coefficients e^{β}) and associated z-statistics in parentheses. # Banks and # (Pseudo-)defaults represent the number of unique banks and (pseudo-) defaults in our sample. Panel B presents estimation results from a cross-sectional logistic regression model estimated for year 2008.

In our next test, we modify the definition of failure to include banks that experienced a large fall in their market-equity ratio during the financial crisis of 2008–2009. A bank with large equity capital at the beginning of the crisis is less likely to become distressed for the same level of negative shock to its stock returns compared to a bank with lower levels of equity capital. Our earlier definition, based on stock returns below −40% during the crisis, misses out this feature. Haldane (2011) points out that banks with lower than 5% market equity (as a ratio of the book value of debt) were much more likely to become distressed during the crisis period. Following this approach, we now classify a bank into the "default" category if its market-equity-to-book-debt ratio falls below the 5% threshold anytime during the crisis. We now find a slightly lower number of defaults (259) in our sample. Panel A of Table 7 provides estimation results from the Cox regression model. With the refined definition of default, the point estimate on both Merton model RNPD and the modified model RNPD increases. More important, in a horse race between the two models, the modified model RNPD explains all the variation in default. In fact, conditional on this measure, the Merton model RNPD has just the opposite sign, through the effect is economically small. As we can see from the AUC of ROC curve, the models have better accuracy with this refined definition of default, consistent with the argument in Haldane (2011).

Panel B presents the estimation result of cross-sectional logistic regression model.[9] Again, our results become stronger for the cross-sectional model. The coefficient estimates are 13.74 pp and 4.48 pp for the modified and Merton RNPDs, respectively. Similarly, the area under the ROC is considerably higher for the modified RNPD (86.39%) as compared to the Merton RNPD (67.29%).

In practice, accurate prediction of actual defaults would also need to require taking into account the presence of explicit and implicit government guarantees, including too-big-to-fail (TBTF) subsidies. But even without this extension, the modified model could be useful in reduced-form analyses of the government's role. For example, Acharya, Anginer, and Warburton (2014) use distance to default within a reduced-form model to predict counterfactual no-TBTF credit spreads of large banks by extrapolating, based on the estimated model, from smaller banks. The value of the subsidy then follows from the difference between this counterfactual and the actual credit spread. One of the state variables in their reduced-form model is the Merton distance-to-default model. Our analysis here suggests that using the default probabilities from our modified model should deliver a more accurate assessment of the counterfactual default risk.

9 Based on this definition of default, we have relatively fewer number of defaults in 2008 compared to the market equity returns based definition. Using an equity-to-book-debt ratio-based definition, we find that the majority of defaults occurred in 2009. Our results remain similar if we estimate the cross-sectional regression for 2009. We report results for 2008 to be consistent with the earlier table.

3.1 Comparison with nonlinear transformations of the Merton model RNPD

It is well known that a literal implementation of Merton model-implied RNPD does not match very well with empirical default data. Models such as those of Moody's KMV©typically use a nonparametrically estimated nonlinear transformation of the Merton model RNPD to obtain default probabilities that better match the real-world default frequencies. Is our modified model simply capturing such a nonlinear transformation of the Merton model RNPD? Or, does our model have additional predictive power in explaining future default that is not captured by such transformation of the Merton model? As we show in Appendix D, the modified model RNPD is not a simple monotonic transformation of the Merton model RNPD. Even so, one may still wonder how close a nonparametrically estimated nonlinear transformation of the Merton model RNPD would come to matching the predictive power of our modified model's RNPD.

To examine this, we use a nonparametric regression estimate (see Appendix D for details on the estimation) to break the modified model's RNPD into two parts: the predicted value of the modified RNPD and a residual. The predicted values from this regression model ($\widehat{RNPD}$) gives us a transformation of the Merton model RNPD that best matches the modified model RNPD. The residual ($RNPD_{res}$) is the part of modified RNPD that cannot be attributed to a Merton model transformation, and hence contains additional information about the riskiness of the bank.

Using these two parts separately, we estimate the default prediction regression using the $\widehat{RNPD}$ and $RNPD_{res}$ as the explanatory variables. As shown in Table 8, both $\widehat{RNPD}$ (model 1) and $RNPD_{res}$ (model 2) predict future defaults. Furthermore, conditional on the level of $\widehat{RNPD}$, the marginal effect of the residual is even higher (model 3). This is reasonable because the residual variable by itself does not control for the base level of default risk. These results are stronger for cross-sectional estimations in panel B. Overall, the results show that the information contained in the modified model RNPD cannot simply be captured by a nonlinear transformation of the Merton model RNPD. Key economic features of our model, such as nonlinearity in asset payoffs and replenishment of borrowers' collateral, are simply missing from the Merton model, so statistical transformations are unable to produce default estimates with the same accuracy as those from our modified model.

3.2 Simplified approximations of the modified Model

As we showed in Section 1.2, simplified versions, such as single-cohort or single-borrower models, do not fully reproduce the predictions of our modified model. But are the additional features of the full model that go beyond these simplified versions also empirically relevant for default prediction? To shed

Table 8
Default prediction with nonlinear transformation of Merton model

	(1)	(2)	(3)
A. Cox regression model			
$\widehat{RNPD}$	0.0494		0.0524
	(16.32)		(19.12)
$RNPD_{res}$		0.0549	0.0852
		(9.69)	(12.89)
Observations	10,076	10,076	10,076
# Banks	1,194	1,194	1,194
# (Pseudo-)defaults	259	259	259
Area under ROC curve	0.8765	0.5783	
B. Logistic regression model			
$\widehat{RNPD}$	0.0622		0.0925
	(3.70)		(3.29)
$RNPD_{res}$		0.1699	0.1919
		(8.36)	(8.29)
Observations	500	500	500
# (Pseudo-)defaults	66	66	66
Area under ROC curve	0.6729	0.8060	

Panel A of the table presents estimates from a Cox-proportional hazards model. The dependent variable is a binary variable that indicates whether or not the bank has defaulted in the following year. The set of defaulted banks includes actual bank failures from 1987 to 2016 as well as pseudo-defaults, which include any bank with market-equity-to-book-debt ratio falling below 5% anytime during 2008–2009. The table shows one minus hazard ratios (i.e., one minus exponentiated coefficients e^{β}) and associated z-statistics in parentheses. # Banks and # (Pseudo-)defaults represent the number of unique banks and (pseudo-) defaults in our sample. Panel B presents estimation results from a cross-sectional logistic regression model estimated for year 2008.

light on this question, we estimate RNPDs using these simpler models and compare their default prediction performance with our full model's RNPD.

Appendix Table A.2 provides the results: hazard rate regression estimates in panel A and cross-sectional logistics regression in panel B. In both of these panels, simplified versions have positive and significant coefficient when used alone as the explanatory variable (Columns 2 and 3). These results show that nonlinearity of bank asset payoffs that remains preserved in the simplified models is important. However, when we include the RNPD from the full model, the simplified models' RNPD is no longer significant in predicting future defaults. The results are especially strong for the cross-sectional model, both in terms of marginal effect of the RNPDs and the fit of the Model. The area under the ROC is 86.39% for the full model, compared with 73-74% for the simplified versions of the model. These results confirm the usefulness of all the rich features of our full model, such as collateral resetting, in obtaining the RNPD for banks.

4. Conclusion

The standard assumption that firms have a lognormally distributed asset value is not appropriate when applying structural models of default risk to banks. Banks' assets are risky debt claims with capped upside and hence the asset payoff is nonlinear, with embedded optionality. As a consequence, bad shocks to borrower asset values lead to a rise in bank's asset volatility, unlike in the standard model where asset volatility is constant. A bad shock to asset values therefore reduces the distance to default much more than it would in the standard model. Our modification of the standard model takes this effect into account and leads to substantially different assessment of distance to default and bank risk dynamics. In good times, when asset values are high, the standard model substantially understates risk-neutral default probabilities because it ignores the options-on-options nature of bank equity and debt. For the same reasons, the standard model understates the value of implicit or explicit government guarantees in good times. Furthermore, the standard model also understates the degree to which banks' equity risk rises in response to an adverse shock to asset values. These shortcomings of the standard model can lead to wrong conclusions in stress tests conduct by regulators and an inaccurate assessment of the deposit insurance premium charged by the deposit insurer.

Our results have a number of implications for regulation and policy. The results are directly relevant for pricing of deposit insurance premiums and for valuation of explicit or implicit government subsidies to the banking sector. Models used for these purposes are often based on the Merton model. For example, Marcus and Shaked (1984) use the Merton model to estimate the fair pricing of deposit insurance, and Acharya, Anginer, and Warburton (2014) use the Merton distance-to-default model in an estimation of the credit spread effects of implicit government guarantees. Our approach provides a more accurate assessment of bank credit risk and hence should help obtain improved estimates in these applications. Bank capital adequacy assessment is another relevant area of application. A key goal of bank capital requirements is to limit bad tail outcomes and realistic modeling of nonlinearities in banks' asset payoffs is important to arrive at an accurate assessment of the likelihood of such tail outcomes.

Our focus in this paper is on the fundamental issue that a bank's asset value cannot be lognormally distributed. As we have shown, this issue has first-order consequences for default risk evaluation. Our modified structural model is useful for understanding the economic drivers of bank default risk. Of course, a simple structural model of the kind we use here still omits many additional features that would be necessary to realistically describe banks' default risks. Extensions of the model could explore jumps in asset values, default due to liquidity problems, complex maturity and seniority structures of banks' debt, and various forms of explicit and implicit government support.

Appendix

A. Inverting the Models in Empirical Applications

In empirical applications of Merton-style structural models, we require estimates of the unobservable asset value and asset volatility (of the borrower in our model, of the firm in the standard Merton model). These estimates can be obtained by combining observable equity value and volatility with the model to back out implied asset value and asset volatility. Here, we briefly describe the approach. We also discuss issues regarding the invertibility of the mapping from asset value and volatility to equity value and volatility.

A.1 Inverting the Standard Merton Model

The firm's zero-coupon debt has face value D and matures at T. The asset value V evolves according to

$$\frac{dV_t}{V_t} = (r-\gamma)dt + \sigma_v dB_t \tag{A1}$$

where γ is the cash payout rate. The value of the firm's equity, including the claim to the dividends until T, is

$$S_t = C(V_t, D, r, \gamma, T-t, \sigma_v) + (1-\exp[-\gamma(T-t)])V_t, \tag{A2}$$

where $C(.)$ is the Black-Scholes call option price,

$$C(V_t, D, r, \gamma, T-t, \sigma_v) = V_t \exp[-\gamma(T-t)]N(d_1) - D\exp[-r(T-t)]N(d_2) \tag{A3}$$

and

$$d_1 = \frac{\log V_t - \log D + (r-\gamma+\sigma_v^2/2)(T-t)}{\sigma_v\sqrt{T-t}}$$

$$d_2 = d_1 - \sigma_v\sqrt{T-t}.$$

Equity volatility follows from the leverage ratio of the call option replicating portfolio, modified by including the claim to the dividends until T, as

$$\sigma_{s,t} = \frac{V_t\{\exp[-\gamma(T-t)]N(d_1) + (1-\exp[-\gamma(T-t)]).\}}{S_t}\sigma_v. \tag{A4}$$

In our simulations in Section 1.1, where we apply the Merton model (as a misspecified model) to data generated from our modified model, we use the simulated values of S_t and the instantaneous equity volatility $\sigma_{s,t}$ to solve Equations (A2) and (A4) for V_t and σ_v. Based on V_t and σ_v we can then compute the risk-neutral (RN) distance-to-default as

$$DD_{BSM} = \frac{\log V_t - \log D + (r-\gamma-\sigma_v^2/2)(T-t)}{\sigma_v\sqrt{T-t}}.$$

The corresponding implied RN default probability, also called the expected default frequency (EDF), can be computed as follows:

$$EDF_{BSM} = \Phi\left(\frac{-\log V_t + \log D - (r-\gamma-\sigma_v^2/2)(T-t)}{\sigma_v\sqrt{T-t}},\right)$$

where $\Phi(.)$ is the standard normal CDF.

We take empirically observed equity values and equity volatility (based on predicted values of equity volatility using an AR(1) model; see Appendix C) estimates to solve Equations (A2) and (A4) simultaneously. (The alternative approach of iterating between asset value and asset volatility as in Crosbie and Bohn (2001) and Bharath and Shumway (2008) delivers similar results. An alternative estimation approach that provides similar estimates is proposed by Duan, Gauthier, and Simonato 2004.)

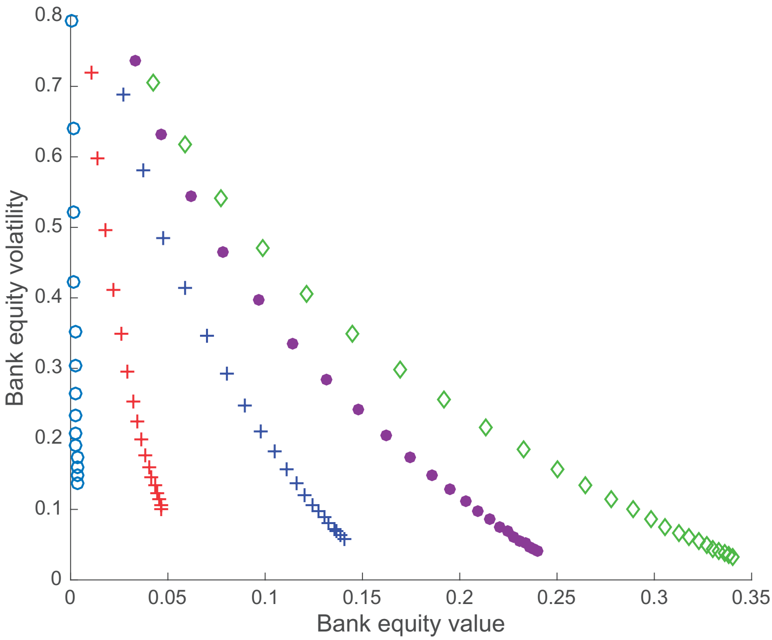


Figure A.1
Equity values and equity volatility for borrower asset value/loan book size combinations

A.2 Inverting our Modified Model

For our modified model, we use essentially the same approach. The only difference is that instead of inverting the model to back out the bank's asset value and volatility, we back out the borrowers' aggregate asset value and the size of the bank's loan book; that is, we look for values for dW_0 shocks and for the loan face value parameter F_1 that allow us to match empirically observed equity value and volatility of the bank.

It is not immediately obvious, however, that our model implies an invertible relationship. However, based on our numerical computations of the function mapping F_1 and dW_0 to equity value and volatility, we can confirm that the function is invertible in the empirically relevant region. Each (F_1, dW_0) pair is mapped to exactly one combination of equity value and volatility.

Figure A.1 illustrates this. Every scatter point shows a particular equity value-volatility pair. Points with the same symbol have the same F_1 but different dW_0. Different symbols mean different F_1 (always with the same range of borrower asset value shocks), with F_1 ranging from 0.7 to 1.1. The other parameters are set to the same values as in our simulations in the main part of the paper. (In our simulations, we use a much denser grid, the relatively coarse grid in this plot is only for illustration.) Going to lower dW_0 for fixed F_1 results in a move in northwest direction (higher equity volatility, lower equity value), but unless F_1 is very high, it's mostly a volatility effect (i.e., north). In contrast, going to higher F_1, keeping dW_0 fixed, results in a move in southeast direction (lower equity volatility, higher equity value), but unless dW_0 is very low, it's mostly an equity value effect (i.e., east). Overall, each equity value-volatility pair in a roughly triangular region is associated with one (dW_0, F_1) pair.

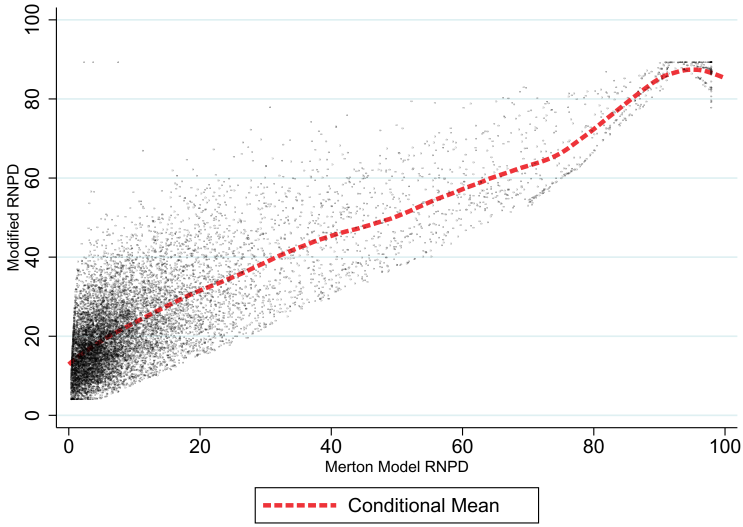


Figure A.2
Modified risk-neutral default probabilities as a nonparametric function of Merton model default probability

B. Data Construction

Table A.1
Variable definitions and construction

Variable	Definition	Source	Construction
E	Market equity value of bank	CRSP	shrcc x shrout
sE	Stock return volatility	CRSP	Predicted Stock return volatility
r	Risk-free rate	FRB	log 10-year risk-free rate
D	Book value of bank debt	Compustat	Short-term debt + long-term debt + deposits + pref. equity (dlcq+dlttq+dptcq+pstkq)

Source: Quarterly Compustat Files for Banks.

C. Computation of Conditional Equity Volatility

We first compute the annualized realized volatility of each bank on every estimation date based on the past 1 year daily stock returns (called $\sigma_{i,t}$) assuming an AR(1) process for daily returns. To minimize the influence of large outliers in the computation of realized equity volatility using daily data, we winsorize these observations at 2.5% from both tails.

We regress realized volatility on 12-month lagged volatility to estimate of conditional equity volatility:

$$\sigma_{i,t+1} = \alpha + \beta\sigma_{i,t} + \epsilon_{i,t}. \tag{C1}$$

We obtain the following best-fit line:

$$\sigma_{i,t+1} = 0.1178 + 0.6438\sigma_{i,t} + \epsilon_{i,t}, \tag{C2}$$

and we use the fitted values from this regression model as conditional volatility in our estimation.

Table A.2
Default predictions with simpler versions of a modified model

	(1)	(2)	(3)	(4)	(5)
A. Cox regression model					
Modified EDF	0.0542			0.0630	0.0514
	(19.58)			(7.47)	(5.52)
Single cohort EDF		0.0456		-0.0077	
		(17.01)		(-1.05)	
Single borrower EDF			0.0512		0.0028
			(17.47)		(0.30)
Observations	10,076	10,076	10,076	10,076	10,076
# Banks	1,194	1,194	1,194	1,194	1,194
# (Pseudo-)defaults	259	259	259	259	259
Area under ROC curve	0.9010	0.8908	0.8961		
B. Logistic regression model					
Modified EDF	0.1374			0.2033	0.1888
	(5.22)			(8.11)	(7.46)
Single cohort EDF		0.0642		−0.0707	
		(4.46)		(-3.24)	
Single borrower EDF			0.0795		-0.0585
			(4.75)		(-2.57)
Observations	500	500	500	500	500
# (Pseudo-)defaults	66	66	66	66	66
Area under ROC curve	0.8639	0.7303	0.7483		

Panel A presents estimates from a Cox-proportional hazards model for some simpler versions of the modified model. The dependent variable is a binary variable that indicates whether or not the bank has defaulted in the following year. The set of defaulted banks includes actual bank failures from 1987 to 2016 as well as pseudo-defaults, which include any bank with market-equity-to-book-debt ratio falling below 5% anytime during 2008–2009. *Modified EDF* is the RNPD from our full model with multiple cohorts of borrowers and multiple borrowers in each cohort. *Single cohort EDF* is the RNPD estimated from a model with only one cohort. *Single borrower EDF* is the corresponding estimate for a single borrower model. The table shows one minus hazard ratios (i.e., one minus exponentiated coefficients e^{β}) and associated z-statistics in parentheses. # banks and # (pseudo-)defaults represent the number of unique banks and (pseudo-) defaults in our sample. Panel B presents estimation results from a cross-sectional logistic regression model estimated for year 2008.

D. Empirical Approximation of Modified Model RNPD by Nonlinear Transformation of Merton Model

To gain further insight into the relationship between the RNPD of the modified model and the Merton model, we estimate a nonparametric regression model linking the two RNPDs. We do so by estimating the conditional mean of the modified model's RNPD for every value of the Merton model's RNPD in our sample: $E(RNPD_i^{Modified}|RNPD_i^{Merton}=x)$ using the following local-linear regression model at each x.

$$min\gamma \sum_{i=1}^{n}\{RNPD_i^{Modified}-\gamma_0-\gamma_1(RNPD_i^{Merton}-x)\}^2 K(RNPD_i^{Merton},x,h) \qquad \text{(D1)}$$

The model minimizes the sum of squared error at each value x of the Merton model's RNPD, where different observations are assigned weights as per the Epanechnikov kernel density function K and an optimally chosen bandwidth parameter h. We obtain the conditional mean γ_0 along with the slope parameter γ_1 for every value of the Merton model RNPD. The slope parameter gives an estimate of the marginal change in the modified model's RNPD for a unit change in the Merton model's RNPD at the specific point. This model allows the parameter values to change at each estimation point, and thus provides us with a flexible nonparametric mapping from the Merton model RNPD to the modified model's RNPD.

Figure A.2 plots the conditional mean of the modified model's RNPD for each value of the Merton model RNPD based on these estimates. As expected, the two measures are positively related. However, as is evident from the scatterplot, the relationship is not a monotonic one. Further, the nonparametric regression model also provides us the marginal effect of the Merton model's RNPD on the modified model's RNPD at each estimation point (i.e., γ_1 above). While these effects vary at each estimation point, the average value of the marginal effects is 0.74, that is, a slope of less than one. At very low values of the Merton model RNPD, the modified model's RNPD is significantly higher. On the other extreme, during bad times and for banks with high leverage the situation reverses. We make use of these estimates later in the paper to further explore the relationship between these risk measures and actual default of banks.

References

Acemoglu, D., and M. K. Jensen. 2015. Robust comparative statics in large dynamic economies. *Journal of Political Economy* 123:587–640.

Acharya, V., R. Engle, and D. Pierret. 2014. Testing macroprudential stress tests: The risk of regulatory risk weights. *Journal of Monetary Economics* 65:36–53.

Acharya, V. V., D. Anginer, and A. J. Warburton. 2014. The end of market discipline? Investor expectations of implicit state guarantees. Working Paper, New York University.

Bharath, S. T., and T. Shumway. 2008. Forecasting default with the Merton distance to default model. *Review of Financial Studies* 21:1339–69.

Black, F., and M. Scholes. 1973. The pricing of options and corporate liabilities. Journal of Political Economy 81:637–54.

Boucher, C. M, J. Daníelsson, P. S. Kouontchou, and B. B. Maillet. 2014. Risk models-at-risk. *Journal of Banking & Finance* 44:72–92.

Campbell, J. Y., J. Hilscher, and J. Szilagyi. 2008. In search of distress risk. *Journal of Finance* 63:2899–939.

Chen, A. H., N. Ju, S. C. Mazumdar, and A. Verma. 2006. Correlated default risks and bank regulations, *Journal of Money, Credit and Banking* 38:375–98.

Crosbie, P. J., and J. R. Bohn. 2001. Modeling default risk. Working Paper.

Culp, C. L., Y. Nozawa, and P. Veronesi. 2014. Option-based credit spreads. Working Paper, University of Chicago.

Dermine, J., and F. Lajeri. 2001. Credit risk and the deposit insurance premium: A note. *Journal of Economics and Business* 53:497–508.

Duan, J.-C., G. Gauthier, and J.-G. Simonato. 2004. On the equivalence of the kmv and maximum likelihood methods for structural credit risk models. Working Paper.

Duffie, D., R. Jarrow, A. Purnanandam, and W. Yang. 2003. Market pricing of deposit insurance. *Journal of Financial Services Research* 24:93–119.

Duffie, D., and D. Lando. 2001. Term structures of credit spreads with incomplete accounting information. *Econometrica* 69:633–64.

Duffie, D., L. Saita, and K. Wang. 2007. Multi-period corporate default prediction with stochastic covariates. *Journal of Financial Economics* 83:635–65.

Giacomini, E., D. C. Ling, and A. Naranjo. 2015. Leverage and returns: A cross-country analysis of public real estate markets. *Journal of Real Estate Finance and Economics* 51:125–59.

Gofman, M. 2017. Efficiency and staility of a financial architecture with too-interconnected-to-fail institutions. *Journal of Financial Economics* 124:113–46.

Goldstein, I., and Y. Leitner. 2018. Stress tests and information disclosure. *Journal of Economic Theory* 177:34–69.

Goldstein, I., and H. Sapra. 2014. Should banks' stress test results be disclosed? An analysis of the costs and benefits. In *Foundations and trends in finance*, eds. I. Goldstein and H. Sapra, 1–54. Boston: Now Publishers.

Gornall, W., and I. A. Strebulaev. 2018. Financing as a supply chain: The capital structure of banks and borrowers. *Journal of Financial Economics* 129:510–30.

Greenlaw, D., A. Kashyap, K. Schoenholtz, and H. Shin. 2012. Stressed out: Macroprudential principles for stress testing. Working Paper.

Haldane, A. G. 2011. Capital discipline. Speech. American Economic Association, Denver. https://www.bankofengland.co.uk/speech/2011/capital-discipline-speech-by-andy-haldane.

Jarrow, R. A., and P. Protter. 2004. Structural vs reduced form models: A new information based perspective. Journal of Investment Management 2:1–10.

Jones, E. P., S. P. Mason, and E. Rosenfeld. 1984. Contingent claims analysis of corporate capital structures: An empirical investigation. *Journal of Finance* 39:611–25.

Kelly, B. T., H. Lustig, and S. Van Nieuwerburgh. 2016. Too-systemic-to-fail: What option markets imply about sector-wide government guarantees. *American Economic Review* 106:1278–319.

Khandani, A. E., A. W. Lo, and R. C. Merton. 2013. Systemic risk and the refinancing ratchet effect. *Journal of Financial Economics* 108:29–45.

Kogan, L., D. Papanikolaou, and N. Stoffman. 2017. Winners and losers: Creative destruction and the stock market. Working Paper.

Leitner, Y. 2014. Should regulators reveal information about banks. Federal Reserve Bank of Philadelphia Business Review, Third Quarter.

Ling, D. C., and A. Naranjo. 2015. Returns and information transmission dynamics in public and private real estate markets. *Real Estate Economics* 43:163–208.

Lucas, D., and R. McDonald. 2006. An options-based approach to evaluating the risk of Fannie Mae and Freddie Mac. *Journal of Monetary Economics* 53:155–76.

Lucas, D., and R. McDonald. 2010. Valuing government guarantees: Fannie and Freddie revisited. In *Measuring and managing federal financial risk*, ed. Deborah Lucas, 131–54. Chicago: Chicago University Press.

Marcus, A. J., and I. Shaked. 1984. The valuation of FDIC deposit insurance using option-pricing estimates. *Journal of Money, Credit and Banking* 16:446–60.

Merton, R. 1974. On the pricing of corporate debt: The risk structure of interest rates. *Journal of Finance* 29:449–70.

Merton, R. C. 1977. An analytic derivation of the cost of deposit insurance and loan guarantees: An application of modern option pricing theory. *Journal of Banking and Finance* 1:3–11.

Peleg-Lazar, S., and A. Raviv. 2017. Bank risk dynamics where assets are risky debt claims. *European Financial Management* 23:3–31.

Pennacchi, G. G. 1987. Alternative forms of deposit insurance: Pricing and bank incentive issues. Journal of Banking and Finance 11:291–312.

Ritchken, P., J. B. Thomson, and I. Popova. 1995. The changing role of banks and the changing value of deposit guarantees. Working Paper.

Ronn, E. I., and A. K. Verma. 1986. Pricing risk-adjusted deposit insurance: An option-based model. *Journal of Finance* 41:871–96.

Schweikhard, F. A., Z. Tsesmelidakis, and R. C. Merton. 2014. The value of implicit guarantees. Working Paper, Oxford University.

Stanton, R., and N. Wallace. 2018. CMBS subordination, ratings inflation, and regulatory-capital arbitrage. *Financial Management* 47:175–201.

Uhlig, H. 1996. A law of large numbers for large economies. *Economic Theory* 8:41–50.

Vasicek, O. 1991. Limiting loan loss probability distribution. Working Paper.

Vassalou, M., and Y. Xing. 2004. Default risk in equity returns. *Journal of Finance* 59:831–68.

Zhou, C. 2001. The term structure of credit spreads with jump risk. *Journal of Banking and Finance* 25:2015–40.