Article:

Spencer, Peter orcid.org/0000-0002-5595-5360 (2016) US Bank Credit Spreads during the Financial Crisis. Journal of Banking & Finance. pp. 168-182. ISSN 1872-6372

https://doi.org/10.1016/j.jbankfin.2016.04.015

US Bank Credit Spreads during the Financial Crisis.

Peter Spencer*

University of York

Abstract

This paper argues that first passage time models are likely to better than affine hazard rate models in modelling stressed credit markets and confirms their superior performance in explaining the behavior of Credit Default Swap rates for the major US banking groups over the period of the financial crisis. Affine models find it hard to deal with periods of exceptionally high or low default risk given their assumption of a constant rate of mean reversion in the hazard rate. In contrast, first passage time models are specified in terms of the distance to default rather than the hazard rate. The persistence of shocks varies with the distance to default, allowing the default curve to invert sharply (compress) when the distance to default is low (high). I use an empirical version of the Collin-Dufresne *et al.* (2003) model, which contains a smoothing parameter that allows it to control the relative effect of these shocks on the short spreads and can be interpreted as an information lag.

*Department of Economics and Related Studies; ps35@york.ac.uk. This paper has benefitted from conversations on this subject with Karim Abadir, John Campbell, Laura Coroneo, Alex Kostakis, Menno Middeldorp, Alistair Milne, Gulcin Ozcan, Tuomas Peltonen, Marco Realdon, Yongcheol Shin, Jacco Thijssen and Mike Wickens. I am also grateful to participants at the 2013 Southampton Conference on the Global Financial Crisis; the 2013 European Financial Management Association conference; the 2014 Infiniti conference and two referees of this Journal for helpful comments and suggestions.

1 Introduction

The global financial crisis provided a stark reminder of the importance of understanding and pricing bank default risk. This paper develops a non-linear econometric model which is designed to capture both the compressed hazard rate structures seen in the run up to the crisis and the strongly inverted structures seen during the crisis itself. The econometric specification is used to model risk-neutral default probabilities implied by the Credit Default Swap (CDS) market for six of the largest US banking institutions since the turn of the millennium[1].

CDS and other credit spreads can be analyzed using structural models, which use accounting information about factors such as profitability and leverage to explain the price of default risk. However, it is difficult to explain the pricing of default risk on a company using accounting data if only because there are many management and economy-wide factors that affect the viability of its business and are not reflected in its accounts. Studies of industrial company spreads suggest that these data can account for only about 60% of the variance (Huang and Huang, 2003). It is necessary to allow for additional factors such as the effect of the business cycle (Collin-Dufresne, Goldstein and Martin, 2001).

Banks are excluded from these studies since they have very high leverage ratios and their capital and other balance sheet ratios are subject to regulatory requirements. Indeed, once the crisis unfolded and liquidity in banking markets evaporated, it became very difficult to value many of the assets in the balance sheet. Asset value uncertainty and the associated counterparty risk caused the interbank deposit market to become extremely stressed over this period (Afonso, Kovner and Schoar, 2011).

In view of these difficulties, I use a reduced form approach to model bank credit

[1] These have been classed as 'globally systemically important' institutions. The CDS data for two other important US banks, State Street and Bank of New York Mellon were too sparse to be used in this study.

risk. The standard reduced form model assumes that under the risk-neutral measure, the instantaneous default or hazard rate h_t follows a diffusion similar to that followed by the spot rate in an Affine Term Structure Model (ATSM).[2] This provides an exponential-affine specification of the cross-section of default probabilities in terms of the hazard rate, which is the analogue of the spot interest rate in an ATSM (Duffie and Singleton, 2003). The hazard rate is modeled as a latent variable that can be estimated using a Kalman filter or simply by assuming that a particular maturity in the cross section of default probabilities is measured without error. A major advantage of the reduced form approach is that it allows me to employ latent variable and other techniques developed in the term structure literature.

Affine hazard rate models usually provide a good empirical explanation of the term structure of credit risk on a particular entity, one that is flexible enough to fit a variety of upward sloping, inverted and hump shaped term structures. However, I find that the extremes exhibited by the default curves seen in the US banking sector since the turn of the millennium cannot be replicated using this approach. These extremes are shown in figures 1 and 2. The first of these figures shows the 1-, 5- and 10-year senior CDS spreads, essentially the cost of insurance against default by these six banks (in percent per annum).[3] The second shows the implied term structure of annual forward default probabilities.[4] It shows that these default curves typically exhibit a gradual upward slope, but became compressed in 2006 before inverting sharply during the crisis.

Figure 3 illustrates the difficulty that affine models have in replicating the ex-

[2] Appendix A in Duffie and Singleton (2003) shows that affine models can also be obtained using jump diffusions and other specifications of the hazard rate process.

[3] Strictly speaking these are the probabilities of a 'credit event', not just default. Besides outright bankruptcy, the 1999 ISDA agreement defined the other credit events that trigger compensation payments under a CDS contract: failure to pay an obligation; obligation acceleration: repudiation, and debt restructuring.

[4] The annual forward default probabilities are calculated as $\ln(p_{t,t+m-12}/p_{t,t+m})$, where $p_{t,t+m}$ is the survivorship probability for maturity $m = 12, 24, ..., 120$ in months.

tremes seen in December 2006 and March 2009, which bound the observations of the sample. It shows the results for Citigroup, which was the most heavily impacted by the Lehman default and JPMorgan which was the least severely affected. While figure 2 shows forward rates (to distinguish the default probabilities in the time series more clearly), figure 3 shows the average default rates over different time horizons. These default rates are the analogue of the discount rates employed in the analysis of the Treasury bond market and are computed as the negative of the log of the probability of survival divided by maturity[5]. The affine models are represented by the well-known model of Cox, Ingersoll and Ross (1985), henceforth CIR and an unrestricted Ordinary Least Squares (OLS) model suggested by recent work on the term structure of interest rates by Hamilton and Wu (2012). The panels on the left show the very poor explanation provided by the affine models in the case of Citigroup. The panels on the right show that the affine models provide a reasonable explanation of the curve seen in December 2006 for JPMorgan, but underestimate the degree of inversion seen in March 2009.

The basic problem with these models lies in their linear structure. This makes them highly tractable but means that shocks to the hazard rate, which have a one for one effect on the short end of the default curves, always have the same proportionate effect on the longer maturities. This reflects the assumption that the degree of mean reversion is constant. To explain the low and flat CDS term structure prior to the crisis the model would need the hazard rate to be low and very persistent. Yet to explain the steeply inverted structure seen during the crisis it would have to be high but much less persistent. Clearly, we need to find a non-linear model that allows the degree of persistence to vary with the initial hazard rate in this nonlinear way.

[5] This calculation gives the maturity-average of the forward default rates shown in figure 2. They are calculated as $-\ln p_{t,t+m}/m$, where $p_{t,t+m}$ are the survivorship probabilities.

This is a characteristic feature of First Passage Time (FPT) models, which specify the risk-neutral dynamics in terms of the distance to default rather than the hazard rate. They assume that the firm's asset value follows a Geometric Brownian Motion and that the firm defaults the first time this reaches a default boundary. Black and Cox (1974) showed that the default probabilities are then given by standard FPT formulae. Consequently, if the initial asset value is close to the boundary, the immediate default probability is very high. However, the forward default probabilities fall back sharply with maturity in this case since the longer the firm survives, the more likely it is that its asset value has diffused away from the boundary. This 'survivorship effect' causes the term structure to invert sharply. On the other hand, if the initial distance to default is high, it is likely to remain so for some time, compressing the default probability structure.

The basic full information version of the FPT model is very sensitive to the initial value of the distance to default, which makes it unsuitable for modelling short credit spreads. However, Duffie and Lando (2001) modified the FPT model by assuming that the investors observe the firm's asset value with a lag. They showed that in this situation, it is important to condition the default probabilities on the observation of no prior default. They allowed the market estimate of the firm's current asset value to be informed by additional signals like credit downgrades. The risk-neutral expectations describing security prices then involve integrals, making it very hard to test this model empirically. Collin-Dufresne, Goldstein and Helwege (2003) showed that if these additional signals are not informative, this simplifies the model considerably. They also abstracted from the effect of the tax system, using the structure of the Black and Cox (1976) model rather than the more complex structure of Leland (1976) that underpins the model of Duffie and Lando (2001). They reported a closed form expression for the conditional probability of default over any future horizon.

I use this basic deferred filtration (DF) specification to model market perceptions of bank default risk during the recent crisis, developing a reduced form model that treats the distance to default as a latent variable. This model also provides a convenient closed form for the instantaneous hazard rate which can be compared with that implied by an affine hazard rate model. It could be fitted to overnight and other inter-bank rates. However, in view of the well documented doubts about the liquidity of the inter-bank markets and the veracity of Libor quotes, I model US bank credit risk using spreads from the CDS market, which were much more liquid over this period. The use of CDS also circumvents the problem of specifying the tax regime (Houweling and Ton Vorst, 2005) and allows me to follow Collin-Dufresne *et al.* (2003) in using the structure of Black and Cox (1976). In their model, the parameter l is added to the maturity in the default rate formulae and has the effect (like maturity) of damping the effect of the distance to default on the default rate structure.

I back out the implied risk-neutral default rates from CDS spreads and data for non-defaultable bond prices and fit them using rival econometric models, based on the principle of risk-neutral pricing. Comparing the fit of the DF model with that of the full information FPT model of Black and Cox (a special case with no information lag), shows that this lag parameter is of crucial importance, allowing the model to capture the relative sensitivity of the short spreads. The performance of the DF model is also superior to that of the affine hazard rate model for five out of the six banks studied in this paper, the exception being JPMorgan.

Figure 4 illustrates this model's non-linear hazard rate reversion effects using some of the empirical results. The central panel shows how the distance to default indicator affects the theoretical value of the default rate across the maturity range using the parameter estimates obtained for Citigroup. With a distance to default of $z = 8$

standard deviations, the (Gaussian) probability mass is well away from the default barrier at all horizons up to 10 years, so the default risk is compressed, as it was prior to the crisis. As the distance to default reduces to one standard deviation, the probability of a near-term default becomes very large and the curve inverts sharply. (The right-hand panel uses the high l value estimated for the RBS, which have the effect of flattening the default curves, making them much less sensitive to maturity.) In contrast, the left-hand panel shows that shocks in the affine hazard rate model are quite persistent, making it very difficult for it to fit the extremes in the data for Citigroup.

The paper is set out along the following lines. The next section describes the basic Black and Cox (1976) model structure and the deferred filtration setting of Collin-Dufresne et al. (2003). It also reviews the two affine models. Section 3 describes the CDS data set and empirical methodology. Section 4 describes the econometric models and reports the empirical results. Section 5 offers a conclusion and suggestions for future research.

2 Theoretical approaches to modeling default risk

This section sets out the various theoretical models that I use to analyze default risk, starting with the FPT model of Black and Cox (1976) and the deferred filtration model of Collin-Dufresne *et al.* (2003). I then give a brief review of the affine model of Cox *et al.* (1985), henceforth CIR.

2.1 The model of Black and Cox (1976)

Consider the structure of the model of Black and Cox (1976) (henceforth BC):

Assumption 1: All agents observe the value of the bank $V(t)$ at time t;

Assumption 2: The logarithm of this value $v = \ln V(t)$ follows a Brownian Motion under the risk-neutral measure:

$$dv = \mu dt + \sigma dw. \tag{1}$$

where: $\mu = (\theta - \eta)$; θ is the expected logarithmic return common to all assets and η the percentage cash flow return to the equity owners;

Assumption 3: The bank has perpetual debt with a face value of L. This value is fixed and cannot be used to finance coupon and dividend payments;

Assumption 4: Interest, dividends and other payments are made continuously;

Assumption 5: There are no taxes;

Assumption 6: Protective covenants or legal restrictions prevent the bank trading with a negative net asset value. In this case, default occurs as the net asset value first reaches zero, or equivalently when the logarithm of the distance to default ratio $x(t) = \ln(V(t)/L)$ first reaches zero;

Assumption 7: In the event of bankruptcy, the banks assets fetch the liquidation value $R(t) = \gamma V(t)$, $\gamma \in [0, 1]$.

This model can be regarded as a simplified version of the model of Leland (1994), which relaxes assumptions 4 and 5 to allow for a lower default trigger value

$V^B < L$ [6] and for a non-zero corporate tax rate. My empirical version of the model uses latent variable techniques to estimate the log distance to default $x(t) = \ln(V(t)/V^B)$ and does not impose a particular value for V^B as a model based on accounting information would. This model is consistent with the geometric Brownian motion used for the firm's net asset value in equity pricing models, allowing comparisons of default risk and asset value across CDS, bond, money and equity

[6] Leland (1994) shows that this simply adjusts the trigger value whilst preserving the mathematical structure of the model.

markets. Implied probabilities of default can be obtained from the equity market in the full information setting using the Black-Scholes formula if we view an equity on a leveraged firm as a call option on the firm's assets with a strike price equal to the debt (Merton, 1974). My use of default swap spreads rather than corporate bonds in the empirical model means that I can abstract from the tax regime, which has a neutral effect on this market (Houweling and Ton Vorst, 2005).

2.2 Default behavior

In this model the probability of default during an investment period of length n and a starting value of $x = z$ is the probability q of a first passage from z to default at zero during the period, which is given by:

$$q(z/\sigma, \mu/\sigma, n) = 1 - \Phi\left[\frac{z + \mu n}{\sigma\sqrt{n}}\right] + \exp\left[-\frac{2\mu z}{\sigma^2}\right]\Phi\left[\frac{-z + \mu n}{\sigma\sqrt{n}}\right] \geq 0; \quad n > 0. \qquad (2)$$

(Duffie and Singleton, 2003), where $\Phi[.]$ is the standard normal distribution function and $\phi[.]$ its density function:

$$\phi(y) = \frac{1}{\sqrt{2\pi}}\exp[-\frac{y^2}{2}].$$

The probability of survival from time s to $t = s + n$ given the observation $x(t) = z$ of the distance to default is thus :

$$p(z/\sigma, \mu/\sigma, n) = \Phi\left[\frac{z + \mu n}{\sigma\sqrt{n}}\right] - \exp\left[-\frac{2\mu z}{\sigma^2}\right]\Phi\left[\frac{-z + \mu n}{\sigma\sqrt{n}}\right] \geq 0; \quad n > 0. \qquad (3)$$

The survivorship function plays a role in the defaultable bond markets that is similar to that played by the discount function in the non-defaultable markets. If the short term interest rate is independent of the default probability structure as in the standard FPT model, then the price $B_{t,u}$ at time t for a defaultable payment at time $u > t$ can be written as the product of non-defaultable discount bond $D_{t,u}$ and the risk-neutral expectation of the payout. If R is the rate of recovery in default and $p_{t,u}$ the probability of the firm surviving from period t to u conditional upon the relevant information set or filtration, then:

$$B_{t,u} = D_{t,u}[p_{t,u} + (1 - p_{t,u})R]. \tag{4}$$

Inverting this relationship allows the risk-neutral survival probability to be determined as $p_{t,u} = (B_{t,u}/D_{t,u} - R)/(1 - R)$. These survival probabilities are in principle tradeable securities.[7]

My empirical model (see section 3) uses latent variable estimation methods that allow the ratios z/σ and μ/σ but not the separate effects of z, μ and σ to be identified. Therefore the rest of this paper uses the normalization $\sigma = 1$. The latent variable z driving the cross section is interpreted as the number of standard deviations to default. The drift parameter μ plays the key role in this model, having a positive effect on the survival probabilities (Spencer, 2013).

As noted in the introduction, this model must be modified to allow it to provide a realistic description of the short spreads. For example, discrete jump processes can be added to the Brownian motion to make the default intensity and short spreads significant when the asset value is within jump-range of the boundary (Baxter, 2007). The Levy distribution can be used to analyze default intensity in this situation and has been used to develop structural default models. Unfortunately, solutions for these distributions are not available in closed form and, in practice, numerical approxima-

[7] A position in the $n-$horizon default probability can be established by buying an n-year defaultable bond and shorting an $n-$year non defaultable bond like a Treasury with the same face value.

tions have to be employed, making them difficult to use in econometric work[8].

Finger, Finkelstein, Lardy, Pan, Ta and Tierney (2002) deal with these problems by assuming that investors observe the balance sheet of the firm but do not know precisely where the default barrier is. This model is extensively used by practitioners like RiskMetrics. The default barrier is related to the observed debt by a lognormally distributed multiplier $\exp[\sigma^2 t + \lambda^2]$, which has the value $\exp[\lambda^2]$ at $t = 0$ at the outset and increases with the time horizon t. Unfortunately this device is problematic because it only conditions the forward default rates on balance sheet variables, not the informative observation that there have been no previous defaults. This means that the algebra is not valid for forward maturities t less than λ^2/σ^2 (see Finger *et al.,* (2002) Reflecting this problem, it is not possible to model the hazard rate or short spreads using this specification.

2.3 The Deferred Filtration model

Duffie and Lando (2001) assume instead that the firm's asset value is uncertain. They deal with these short-maturity problems by conditioning the forward default probabilities on the observation of no prior default as well as balance sheet variables. Their approach is thus valid for modelling the default rate at all horizons. Specifically, the investment decision is conditioned by a 'deferred filtration' or a lagged information set, which could reflect delays in financial reporting, for example. This lag damps the effect of accounting information on market prices and allows room for other risk indicators to affect them.

In a deferred filtration model, the time of default is not affected because the bank manager still observes the net asset value precisely. So assumption 1 is maintained for the bank manager, who declares bankruptcy when x attains zero as in the standard

[8] Moreover, as Duffie and Lando (2001) note, the hazard rate (or instantaneous default intensity) is not well defined in these models.

FPT model. However, investors and other outsiders observe the value of the bank with a lag of length l. The time line is shown in figure 5. Thus we adapt assumption 1 and assume that at time t, investors get a lagged accounting signal $z = v(s)$ where $s = t - l$ (i.e. they have access to the deferred filtration $\mathcal{F}_s$). Suppose that the only other information that they have is that the bank has survived until t. Collin-Dufresne et al. (2003) note that because survival from s until $u > t$ implies survival until t, the joint probability of survival to t and then to $u = s + l + m$ is simply the probability of survival to u: $p(z/\sigma, \mu/\sigma, l + m)$. Using Bayes Law, it follows that the probability of survival to u conditional upon survival to t is obtained by dividing this by the probability $p(z/\sigma, \mu/\sigma, l)$ of surviving to $t = s + l$ given $x(s) = z$:

$$p_{t,t+m} = \frac{p(z/\sigma, \mu/\sigma, l + m)}{p(z/\sigma, \mu/\sigma, l)} \tag{5}$$

Substituting (3) with the normalization $\sigma = 1$ then gives their closed form solution:

$$p_{t,t+m} = p(z, \mu, l + m) = \frac{\Phi\left[\frac{z+\mu(l+m)}{\sqrt{(l+m)}}\right] - \exp\left[-2\mu z\right]\Phi\left[\frac{-z+\mu(l+m)}{\sqrt{(l+m)}}\right]}{\Phi\left[\frac{z+\mu l}{\sqrt{l}}\right] - \exp\left[-2\mu z\right]\Phi\left[\frac{-z+\mu l}{\sqrt{l}}\right]} \tag{6}$$

Taking the logarithm of this function gives the negative of the default probability $q_{t,t+m}$ (since $\ln p_{t,t+m} = \ln(1-q_{t,t+m}) \simeq -q_{t,t+m}$). Changing sign and dividing by the time horizon or maturity m then gives a model of $r_{t,t+m} = -\ln p_{t,t+m}/m \simeq q_{t,t+m}/m$ or the average default rate over the period, which is the analogue of the discount yield in the Treasury bond market. This is the default statistic I use my empirical models to explain. Duffie and Lando (2001) show that although the hazard rate is identically zero in the standard FPT model, this is not the case in the deferred filtration model.

Equation (11) of Appendix 1 describes the forward rate function for the DF model. Importantly, the information lag l adds to the forward maturity m in this formula, damping the effect of the initial distance to default in the same way that forward maturity does in the basic full information version of the model. This parameter is crucially important in controlling the sensitivity of the instantaneous hazard rate and the ultra-short spreads to the distance to default. The instantaneous hazard rate, which I will compare with that of the affine model, follows by taking the limit of (11) as the forward maturity m goes to zero. Appendix 2 describes the likelihood function.

This algebra can be used to represent a situation in which the default barrier (the outstanding liabilities L in the Black-Cox framework) is also uncertain, as in the model of Finger *et al.* (2002). Suppose, for example, that the default barrier $\ln(V^B)$ as well as the asset value $\ln(V)$ follow a Brownian motion resembling (1) and that at time t the investor observes an accurate but lagged value of the distance to default $x(s) = z$. Then it follows that the ratio $x(t) = \ln(V/V^B)$ representing the distance to default also follows a model resembling (1) and that the unconditional survivorship function is given by (3), where μ and σ now represent the combined drift and volatility. The conditional survivorship function is given by (6) but unlike the model of Finger *et al.*, is conditioned by the observation of no prior default and valid over the whole maturity range. Thus the model could capture the effect of changes in the resolution regime, such as the Dodd-Frank Act of 2010, as well as uncertainty about the default barrier implied by the regulatory regime.

Deferred filtration models provide an interesting way of formalizing the effect of accounting and other information lags on asset prices. To estimate my model econometrically, I assume that the information lag l is constant. Formally, this means that in each new period the investor receives a precise observation z of the distance to default that is one period more up to date. However, in view of the doubts about the relevance of accounting information in this area, I adopt an eclectic view of the informational structure, regarding the model as a convenient non–linear reduced form rather than adopting a specific structural interpretation. In addition to accounting information, the information set is likely to include information on the state of the credit and business cycle as well as the regulatory environment. Similarly, I regard l as a smoothing parameter which determines the relative sensitivity of the short spreads to this information, without necessarily representing an information lag. This gives a non-affine reduced form model of the cross section of default rates with one latent variable (z) and two fixed parameters (μ, l) that need to be estimated.

2.4 The affine hazard rate model

The standard reduced form representation models the instantaneous hazard rate h_t directly. This model provides an exponential-affine specification of the default function in terms of the instantaneous hazard rate by assuming that under the risk-neutral measure, this follows a diffusion similar to that followed by the spot interest rate in an exponential-affine specification of the term structure of interest rates (Duffie and Singleton, 2003). To keep default and survivorship probabilities non-negative the reduced form approach typically adopts the CIR square root volatility model of the risk-neutral dynamics:

$$dh = \kappa(\theta - h)dt + \sigma\sqrt{h}dz. \tag{7}$$

This generates the familiar CIR negative exponential solution for the survivorship function:

$$p_{t,t+m} = p(h_t, m; \kappa, \theta, \sigma) = [\frac{2\gamma e^{\delta m/2}}{\delta(e^{\gamma m} - 1) + 2\gamma}]^{\upsilon} \exp[-\beta(m)h_t], \tag{8}$$

where $\beta(m) = 2(e^{\gamma m} - 1)/(\delta(e^{\gamma m} - 1) + 2\gamma); \delta = \gamma + \kappa; \gamma = ((\kappa)^2 + 2\sigma^2)^{1/2}; \upsilon = 2\kappa\theta/\sigma^2$. This model represents the log survivorship function, and hence the default probabilities, as affine functions of the hazard rate. Taking the logarithm of (8), using $\ln p_{t,t+m} = \ln(1 - q_{t,t+m}) \simeq -q_{t,t+m}$ and changing sign gives a linear representation of the default probability. Dividing by the time horizon, or term, then gives a linear model of the average default rate $r_{t,t+m} = -\ln p_{t,t+m}/m \simeq q_{t,t+m}/m$ given any time horizon:

$$\begin{aligned} -\ln p_{t,t+m}/m &= \frac{\upsilon}{m}\ln[\frac{2\gamma e^{\delta m/2}}{\delta(e^{\gamma m}-1)+2\gamma}] - \frac{\beta(m)}{m}h_t \\ &= a(m) + b(m)h_t. \end{aligned} \tag{9}$$

These affine structures have been extensively used in modelling defaultable and non-defaultable bond price structures. However, they are restrictive because they fix the relative effect of the instantaneous hazard rate on the default rates at different maturities independently of the hazard rate. So, for example, the relative effect on the default rates at the respective horizons m and n years given the model (9) is fixed in the ratio $b(m)/b(n)$. We will see this is a serious handicap when using a single factor model to analyze bank default during the crisis.[9]

3 The empirical models

This section describes the CDS data set and the empirical methods employed in this paper. The empirical results are reported in the next section.

3.1 Data

Section 2 sets out several rival econometric models that are designed to explain market data for the cross-section of default rates: $-\ln p_{t,t+m}/m$. These rates could,

[9] This restriction might be relaxed by using a multiple factor model, but empirical models of credit risk typically use a single factor specification. Work on the term structure of interest rates reveals that three factor models also fit inverted curves relatively poorly (see for example Dai and Singleton (2000) Table IV).

in principle, be backed out from the prices of defaultable and non-defaultable bond using (4). But in this paper, they are backed out of the maturity structure of CDS prices. Since the swaps market was more active than the bond market over the period of the crisis (being extensively used by hedge funds as a vehicle for speculation against banks) this is the approach adopted here. Moreover, unlike bond prospectuses, those for CDS contracts are standardized, facilitating liquidity. They are not affected by short-sale restrictions and there is evidence suggesting that they lead the bond market in terms of price discovery (Blanco, Brennan and Marsh, 2005, Forte and Pena, 2009).

The CDS data were provided by Markit Ltd. They have a panel structure, consisting of daily observations on ten annual maturities of US bank debt CDS spreads.[10] The CDS spreads for senior debt are available back to January 2001. Subordinated debt CDS started to trade later, though before the money market crisis of 2007[11]. I use end-month observations for the 6 largest US banking groups. These comprise three large universal banks (Bank of America, Citigroup and JPMorgan), two investment banks (Goldman Sachs and Morgan Stanley) and Wells Fargo, a large regional bank.

I back out the implied default probabilities $q_{t,t+m}$ using standard recursion formulae, given by eq2.7.4) of Hull (2003). These calculations require an assumption about the value of the recovery rate. I use the values of 40% and 20%, respectively, suggested by the calculations in the Markit spreadsheets for senior and subordinated debt (respectively denoted $qsn_{t,t+m}$ and $qsb_{t,t+m}$) These produce estimates for the implicit default probabilities that align reasonably well, with no persistent differ-

[10] The Markit files also contain some observations on 6 month, 15 year and 20 year CDS spreads. However there are a lot of missing observations for these spreads and they were not used for estimation, only for ex post out of sample checks.

[11] The dates of these samples are reported in table 1.

ences, as shown by the observations in figure 2.[12] Table 1 shows that as in many financial data sets, these observations are negatively skewed and have fat tails. They are also persistent.

Figure 1 shows the basic spread data. They show that default risk was high at the turn of the millennium as the dot-com boom ended but then fell to a very low level during the 'search for yield' that preceded the crisis.[13] This compression began to unwind in 2007 after the problems with the Bear Stearns hedge funds emerged. The inter-bank liquidity crisis then followed when problems with the Paribas hedge funds came to light in the first week of August. The implicit one-year default rate moved up to around 2% for many large US banks as the money markets became stressed (Afonso, Kovner and Schoar, 2011), with the five-year rate moving higher in many cases. Then, following the Lehman default in September 2008,[14] the one-year rate spiked up to 9% and 20%, respectively, for the investment banks Goldman Sachs and Morgan Stanley. They were hit most severely by the default, but converted to banks in order to get support from the Fed very soon afterwards, which seems to have has the effect of calming the market's worries.

The effect on the universal banking groups Citigroup and Bank of America was more gradual and did not peak until March 2009, when their one-year default rates reached 11% and 14%, respectively. The impact of the Lehman default on JPMorgan and Wells Fargo was less pronounced, but all six forward default curves inverted sharply during this crisis. Worries about the viability of Goldman Sachs and Morgan Stanley lingered until the end of the year but worries about the other four institutions eased as the Fed's unconventional monetary policies and the Treasury's Troubled

[12] Further work at the estimation stage indicated that the fit of these models could not be significantly improved by adopting different assumptions about recovery rates.

[13] Risk premia also fell back in the Treasury bond market over this period, helping to explain the 'conundrum': an episode in which long rates were stable in the face of rinsing short-term rates.

[14] The Bank of America acquired Merrill Lynch in that month (having acquired the mortgage lender Countrywide in August 2007).

Assets Relief Program took effect (see Krosner and Melick, 2010, Mc Andrews and Wang,2008 and Wu, 2009). The behavior of the spread and default curves was idiosyncratic. At one extreme, Citigroup was badly impacted by the Lehman default while at the other, JPMorgan was less severely affected.

Another high-risk episode began in the summer of 2011 as the European sovereign debt crisis escalated, raising questions about the exposure of US banks to Europe. These worries seem to have been more persistent than in the Lehman crisis since the slope of the forward default curve was broadly flat for most of the major banks. Yet in the case of the Bank of America, worries about the mortgage problems emerging over this period at its subsidiaries Countrywide and Merrill Lynch were an added concern over this period, keeping its forward curve inverted.[15]

3.2 The yield factor method

These default probabilities are then used to compute the respective survivorship probabilities and hence the default rates ($rsn_{t,t+m} = -\ln(1 - qsn_{t,t+m})/m$ and $rsb_{t,t+m} = -\ln(1 - qsb_{t,t+m})/m$) that are the dependent variables in the models tested in this paper. I follow the yield factor literature (Duffie and Kan, 1996), which assumes that the prices and yields of some individual bonds or portfolios of bonds are observed without error. Specifically I assume that the first principal component of the ten rates derived from the senior CDS prices (which can be considered to be a portfolio of senior CDS contracts) is observed without error. I first use this as regressor in an unrestricted OLS regression that explains the 1-, 2-, 3-, 5-, 7- and 10-year default rates implied by the senior spreads and where available the 1-, 2-, 3-, 5-, 7- and 10-year rates implied by the subordinated spreads. This gives a

[15] The CDS spreads for JP Morgan appear to be high over this period, but this effect is exaggerated by the scale of the chart, which reflects the market's relatively low default risk assessment during the earlier episode. Wells Fargo remained relatively immune to default risk during both episodes.

dozen cross-sectional observations for months in the sample when both senior and subordinated CDS data are available. As Hamilton and Wu (2012) note in the the context of the term structure literature, this OLS model is a useful starting point and provides a benchmark for evaluating affine models such as CIR, since these are of the same linear form but with non-linear restrictions across their coefficients.

To estimate the CIR model, I follow the procedure used in a principal component based yield factor model, backing out a time series for the hazard rate h_t from the time series for the first principal component (PC) and substituting this back into the relationships (9) used to fit the cross section. Similarly, to estimate the FPT models I back out the log distance to default indicator z_t from the time series for the first PC and substitute this back into the relationships (6) used to fit the cross section. Appendix 2 derives the likelihood of the various cross-section models using this PC-based yield factor approach and outlines the estimation procedure.

4 The empirical results

This analysis began with a preliminary investigation of the data using principal components analysis and the two linear models of the cross section (OLS and CIR). Next, I estimated the two FPT models of the cross section (DF and BC).

4.1 Affine models

Table 2 shows the likelihood statistics and parameter estimates for the affine models. The first PC for each bank typically explains 90-95% of the variance of the cross section of its senior default rates. Reflecting this, the OLS regression model also provides a reasonable fit. However, it does this by generating significant negative default probabilities during the pre-crisis period. The CIR models are designed to prevent this, but badly fail a Hamilton and Wu (2102) likelihood ratio test.

This test is based on the fact that because the CIR model is nested within the OLS model, (twice) the difference in their loglikelihood values has a χ^2 distribution. However, the large number of observations in large data sets such as this strongly biases this test towards the rejection of the restricted model. The Bayesian Information Criterion (BIC) provides a better performance indicator in this situation since it is asymptotically unbiased. The Schwarz Approximation (SCA, Appendix 1) is based on the difference between the BIC values of two nested models and also has a χ^2 distribution (with $9 = 12 - 3$ degrees of freedom in this case). This test statistic is reported as SCA in the table. The $p-$values of the statistics shown in the table are effectively zero, indicating that the CIR model is mis-specified. Moreover, the parameters of these CIR models are problematic. The positive estimates of the autoregressive coefficient κ indicate that the risk-neutral dynamics are unstable and the negative values of θ reflect the difficulty the CIR model has in explaining the period of yield compression.

4.2 FPT models

Table 3 shows the likelihood statistics and parameters for the FPT models. With the exception of JPMorgan, the loglikelihood and BIC values for model DF are much higher than for the two affine models. Reflecting this, the Vuong (1989) test statistic (Appendix 2), which can be used to compare the performance of non-nested models, such as the CIR and DF model, favours the latter. This statistic has a standard normal distribution in large samples. Table 4 shows how well the DF model fits the senior and subordinated default rates at different maturities. Clearly, this model has problems fitting the very volatile short rates, it but does well at the longer end.

The negative values of the drift parameter μ shown in table 3 indicate investor pessimism about long-term default risk. (Wells Fargo is the exception, in having a positive drift term.) This increases the level of default risk and gives an upward bias to the slope of the term structure of default rates, consistent with the term structure of the mean sample values reported in table 1. The slope is nevertheless negative (i.e. curve inverted) when the value of z is depressed following the Lehman default. The parameter l is highly significant, showing that it is very important to smooth the effect of the distance to default indicator on the short maturity spreads. This is confirmed by the very poor likelihood values shown in the lower panel of table 3 for the Black-Cox model[16].

If we interpret this parameter strictly in terms of the deferred filtration model, this suggests that investors price these default risks as if they observed the distance to default with a lag of between two and four years. However, given the reduced form nature of this literal specification this interpretation may not be appropriate. This parameter acts as a damping factor, reflecting investor uncertainty more broadly. For example, as noted, it may reflect uncertainty about the default barrier.

Figure 6 shows the estimates of the distance to default measures for the six banks. These are highly correlated, suggesting the presence of a common risk factor as proposed by Collin-Dufresne *et al.* (2001). The first PC explains 94% of the variance, although as in the case of the default rates shown in figure 2, the indicators for the two investment banks differ from the rest immediately following the Lehman default. Further linear regression tests were conducted on these estimates to check the lag structure. Stacking the six estimates into a vector and modelling them using first- and second-order Vector Autoregressions (VAR) confirmed that a first order model was appropriate on the basis of the BIC. The off-diagonal elements of this VAR response matrix were insignificant, suggesting that none of these default risk

[16] The SCA statistics are not reported but again effectively have zero $p-$values, providing a decisive rejection of the zero lag restriction.

indicators Granger-caused the risks to other banks. In other words, any contagion effects occurred within the monthly time frame of this analysis and the shocks were in that sense contemporaneous.

4.3 Model comparisons

How does the non-linearity of the DF specification allow it to outperform the linear models in explaining the default risk in stressed institutions like Citigroup? Further analysis of the model residuals reveals that both approaches provide a good explanation of the regular observations found in the centre of the distribution. These are characterized by upward sloping curves. However, the DF model outperforms in the extremes of the distribution. Reflecting this, the affine models perform tolerably well in the case of JPMorgan, which was not as seriously stressed as the other banks. Figure 7 contrasts the behavior of the hazard rates in the DF and CIR models over time. The CIR model models the hazard rate like the spot rate in an affine term structure model and appendix 1 derives the hazard rate function for the DF model. These are similar in the case of JPMorgan (as are the model residuals). For the other institutions, the DF hazard rates tend to be more pronounced than they are in the CIR model. The thin tails of the Gaussian distribution allow the DF model to replicate the very low default rates seen before the crisis by increasing the distance to default until the distribution of future values is sufficiently far from the default boundary.[17]

The central and lower panels of figure 3 show how the two approaches attempt to explain the extreme values of the default curves that are shown for Citigroup and JPMorgan in the top panel and discussed in the introduction. As noted there, these

[17] The default rate becomes negligibly small as the distance to default increases beyond six standard deviations, allowing the DF model to replicate the period of compressed default rates with ease.

extremes occurred in late 2006 as yields were compressed and in March 2009 following the Lehman default. These two curves bound the observations of the sample for these two banks. The panels on the left show that JPMorgan was relatively immune to these developments and the affine models fit tolerably well in this case. However the panels on the right show that they hit Citigroup harder, making it difficult for the affine models to fit these periods using parameters that are largely determined by the need to fit the centre of the distribution. The linear fit tends to pivot around the seven-year time horizon, with large positive residuals at the short end in December 2006 and negative OLS residuals in March 2009.

The DF specification models these extremes relatively well. To see how it is able to do this, the right-hand panel of figure 4 shows how the distance to default indicator (z) affects the theoretical value of the default rate across the maturity range. This panel uses the parameter values for Citigroup shown in tables 2 and 3. With a distance to default of $z = 8$ standard deviations, the risk distribution is well away from the default barrier at all horizons up to 10 years, so the default risk is negligibly low. As the distance reduces to $z = 4$, the risk begins to increase. The effect is felt across the range, not just at the longer horizons, pushing the whole curve bodily upwards, replicating the yield curve shapes seen in the centre of the distribution. As the distance to default reduces to two and then one standard deviation, the probability of a near-term default becomes very large.[18] However, the long-term risk is less severe because if the institution survives the near-term crisis, the odds are that the balance sheet will recover (the survivorship effect). It allows the model to generate the strongly inverted curve shapes seen during the Lehman crisis.

The affine models find it very hard to explain these extremes. The bold line in the left-hand panel of figure 4 shows the intercepts $a(m)$ in (9). These are the default

[18] This effect would obviously be much greater in the absence of the information lag.

rates at $h = 0$, used by the model to generate the CIR estimates shown in the lower left-hand panel of figure 3. Positive hazard rates increase the default term structure but as noted in section 2.4, the relative impact at different horizons is fixed in an affine model. The need to fit the persistent hazard rate shocks that characterize the centre of the distribution means that the compromise effect $b(m)$ is felt fairly evenly across the term structure. But this restricts the ability of the model to explain the inverted curve seen in March 2009. The empirical model selects a value of $h = 0.17$ which allows it to fit the 7- and 10-year rates quite well, but seriously underestimates the shorter rates.

5 Conclusion

This paper reports the first attempt to take the deferred filtration variant of the FPT model to the data. The results confirm the superiority of this model over the standard affine specification, which finds it hard to deal with periods of exceptionally high or low default risk given its assumption of a constant rate of mean reversion in the hazard rate. The dynamics of the FPT model are specified in terms of the distance to default rather than the hazard rate. This means that the persistence of shocks varies with the distance to default, allowing the default curve to invert sharply (compress) when the distance to default is low (high). The DF version of the model uses a smoothing parameter to control the relative effect of these shocks on the short spreads.

I have developed a reduced form variant of the DF model, which treats the distance to default as a latent variable, without specifying the precise information structure. However, the structure of the model is consistent with the view that asymmetric information played a key role in the breakdown of the banking markets during the recent crisis. Investors in the CDS market behaved as if they observed the distance to default with a lag of two to four years.

The success of this approach opens the way to a more reliable and rigorous approach to regulatory issues concerning the banking sector. For example, the barrier uncertainty model of Finger *et al.* (2002) has been employed by Schweikhard and Tsesmelidakis (2011) to estimate the value of government guarantees implicit in equity prices and five year CDS spreads. However, unlike the deferred filtration model, this model is not conditioned on the observation of no prior default and does not handle near-term risk[19]. As we have seen, most of the movement in the recent crisis took place in the short spreads, which the deferred filtration model handles nicely, allowing the whole maturity range to be used to inform estimates of the distance to default and the value of guarantees, rather than just the five year spread. Estimates from the deferred filtration model could also be used to obtain fair value deposit insurance rates (Acharya and Dreyfus, 1989) and throw light on the optimal bank closure decision (Fries and Perraudin, 1997).

Recent US banking sector data may unduly favor the deferred filtration model, which clearly has a comparative advantage in handling stressed credit markets. This feature makes it likely that it will offer a good explanation of other stressed credit markets such as those of the periphery of the Euro area, which experienced default curves similar to the extremes seen in the US banking markets. It might also provide a useful structural interpretation of the data for entities that are less stressed, comparable in terms of fit to that of the standard reduced form approach. However that remains on the agenda for future research and remains to be seen.

Given the simplicity of deferred filtration model used in this study, these results are very encouraging. Moreover, Collin-Dufresne *et al.* (2003) suggest several exten-

[19] Other papers on this topic by Acharya, Anginer and Warburton (2013) for example, have avoided these problems by using linear regression models to explain spreads in terms of equity-based measures of bank default and other relevant regressors.

sions that would be interesting to pursue in future empirical work. Duffie and Lando (2001) show that the deferred filtration approach can allow for other signals of asset value like credit downgrades and defaults of other banks, which become potentially relevant once there are doubts about the accuracy of accounting information. Allowing for these shocks could improve the performance of the model used in this study and provide further insights into the informational structure of the banking markets during the recent financial crisis.

References

Acharya, S. and Dreyfus, J.: 1989, Optimal bank reorganization policies and the pricing of federal deposit insurance, *Journal of Finance* **44**, 1313–1333.

Acharya, V., Anginer, D. and Warburton, A. J.: 2013, The end of market discipline? investor expectations of implicit state guarantees, *Working paper, Syracuse University* .

Afonso, G., Kovner, A. and Schoar, A.: 2011, Stressed, not frozen: The federal funds market in the financial crisis, *Journal of Finance* **66**(4), 1109–1139.

Baxter, M.: 2007, Levy simple structural models, *International Journal of Theoretical and Applied Finance* **10**, 593–606.

Black, F. and Cox, J.: 1976, Valuing corporate securities: Some effects of bond indenture provisions, *Journal of Finance* **31**, 351–367.

Blanco, R., S. B. and Marsh, I.: 2005, An empirical analysis of the dynamic relation between investment-grade bonds and credit default swaps, *Journal of Finance* **60**, 2255Ű2281.

Collin-Dufresne, P., Goldstein, R. and Helwege, J.: 2003, Is credit event risk priced? modeling contagion via the updating of beliefs, *Working paper, Carnegie Mellon University.available at: http://citeseerx.ist.psu.edu/viewdoc/download* .

Collin-Dufresne, P., Goldstein, R. S. and Martin, J. S.: 2001, The determinants of credit spread changes, *The Journal of Finance* **56**, 2177–22074.

Cox, J., Ingersoll, J. and Ross, S.: 1985, A theory of the term structure of interest rates, *Econometrica* **52**, 385–407.

Dai, Q. and Singleton, K.: 2000, Specification analysis of affine term structure models, *Journal of Finance* **55**, 415–41.

Duffie, D. and Kan, R.: 1996, A yield-factor model of interest rates, *Mathematical Finance* **6**, 379–406.

Duffie, D. and Lando, D.: 2001, Term structures and credit spreads with incomplete accounting information, *Econometrica* **69**, 633–664.

Duffie, D. and Singleton, K.: 2003, *Credit Risk*, PUP, Princeton.

Finger, C. C., Finkelstein, V., Lardy, J., Pan, G., Ta, T. and Tierney, J.: 2002, Creditgrades, *Technical Document, Risk Metrics Group.* .

Forte, S. and Pena, J.: 2009, Credit spreads: An empirical analysis on the informational content of stocks, bonds, and cds, *Journal of Banking & Finance,* **33**, 2013Ű2025.

Fries, S. Mella-Barral, P. and Perraudin, W.: 1997, Optimal bank reorganisation and the fair pricing of bank guarantees, *Journal of Banking and Finance* **21**, 441–468.

Hamilton, J.: 1994, *Time Series Analysis*, Princeton University Press, Princeton.

Hamilton, J. and Wu, J.: 2012, Identification and estimation of affine term structure models, *Journal of Econometrics, forthcoming* .

Houweling, P. and Ton Vorst, T.: 2005, Pricing default swaps: Empirical evidence, *Journal of International Money and Finance* **24**, 1200–1225.

Huang, J. and Huang, M.: 2003, How much of the corporate-treasury yield spread is due to credit risk, *Working paper, Penn State University* .

Hull, J.: 2003, *Options, Futures and other Derivatives*, 5th edition edn, Prentice Hall.

Kroszner, R. and Melick, W.: 2010, The response of the fed to the recent banking and financial crisis, *WP University of Chicago* .

Leland, H. E.: 1994, Corporate debt value, bond covenants and optimal capital structure, *Journal of Finance* **49**, 1213–52.

Mc Andrews, J., A. S. and Wang, Z.: 2008, The effect of the term auction facility on libor, *FRBNY Staff Reports 335* .

Merton, R. C.: 1974, Financial innovation and the management and regulation of financial institutions., *Journal of Banking and Finance* **19**, 461–481.

Schweikhard, F. and Tsesmelidakis, Z.: 2011, The impact of government interventions on cds and equity markets, *Working paper, Goethe University Frankfurt.* .

Spencer, P. D.: 2013, The mills ratio and the behavior of redeemable bond prices in the black-cox model of corporate default, *University of York Discussion Paper* .

Vuong, Q., H.: 2002, Likelihood ratio tests for model selection and non-nested hypotheses, *Econometrica* **57**, 307Ű333.

Wu, T.: 2009, On the effectiveness of the fed's new liquidity facilities, *WP FRBD* .

Appendix 1: The hazard rate in the Deferred Filtration model

The inter-bank market trades funds at very short maturities, making it important to analyze the behavior of hazard rate (or default arrival intensity). Similarly, the forward default intensity $f_{t,t+m}$ at time t and forward maturity m is the probability of default at any instant $t+m$ conditional on no prior default until then and the information set available to investors at time t. Duffie and Singleton (2003) showed how this is related to the survivorship value $p_{t,t-m}$. Providing that this value is differentiable, then:

$$f_{t,t+m} = -\frac{1}{p_{t,t+m}}\frac{\partial p_{t,t+m}}{\partial m} \quad \Leftrightarrow \quad p_{t,t+m} = \exp[-\int_0^m f_{t,t+u}du]. \tag{10}$$

To specify the forward default intensity structure for the DF model we thus differentiate (6) with respect to m and divide by (6) to get:

$$h_{t,t+m} = f(z/\sigma,\mu/\sigma,l+m) = \frac{\left(\frac{z}{(l+m)^{3/2}}\right)\phi\left[\frac{z+\mu(l+m)}{\sqrt{(l+m)}}\right]}{\Phi\left[\frac{z+\mu(l+m)}{\sqrt{(l+m)}}\right] - \exp\left[-2\mu z\right]\Phi\left[\frac{-z+\mu(l+m)}{\sqrt{(l+m)}}\right]}. \tag{11}$$

The instantaneous hazard rate at time t follows by taking the limit as m tends to zero[20]:

$$h_t = f(z/\sigma,\mu/\sigma,l) = \frac{\left(\frac{z}{l^{3/2}}\right)\phi\left[\frac{z+\mu l}{\sqrt{l}}\right]}{\Phi\left[\frac{z+\mu l}{\sqrt{l}}\right] - \exp\left[-2\mu z\right]\Phi\left[\frac{-z+\mu l}{\sqrt{l}}\right]}. \tag{12}$$

[20] This formula can also be obtained from Duffie and Lando's equation (A1) using the normalization $\tilde{x} = x$ and $\tilde{z} = z$ (which specializes this to the Black-Cox model) and taking the limit in which the variance (a^2) of the additional signal goes to infinity and thus becomes uninformative.

Appendix 2: The likelihood function and the estimation procedure

This appendix sets out the form of the likelihood function and associated test statistics for the models discussed in section 3 and outlines the optimization procedure.

First we use the default function to obtain an estimate of each implied value of the default rate given the relevant model parameters and (depending on the model) the distance to default or the hazard rate. For the FPT model we use (6) to get.

$$r(z_t, \mu, l+m) = -\ln p(z_t, \mu, l+m)/m,$$

(For the affine model we replace this by (9) in what follows.) The senior observations $rsn_{t,t+m} = -psn_{t,t+m}/m$ are stacked in the vectors: $\mathbf{rsn}_t = \{rsn_{t,t+1}, rsn_{t,t+2}, rsn_{t,t+3}, rsn_{t,t+5}, rsn_{t,t+7}, rsn_{t,t+10}\}'$; $t = 1, ..., T$. We next define the conformable vectors of estimates and measurement errors:

$\hat{\mathbf{r}}_t = \hat{\mathbf{r}}(z_t, \mu, l) = \{r(z_t, \mu, l+1), ..., r(z_t, \mu, l+10)\}'$; $\mathbf{esn}_t = \{esn_{t,t+1}, ..., esn_{t,t+10}\}$

to get the econometric relationship:

$$\mathbf{rsn}_t = \hat{\mathbf{r}}(z_t, \mu, l) + \mathbf{esn}_t; \quad t = 1, ..., T. \tag{13}$$

where $\mathbf{esn}_t$ is a vector of $n.i.d.$ measurement and mispecification errors:

$$\mathbf{esn}_t \sim N(\mathbf{0}, \mathbf{D})$$

and where $\mathbf{0}$ is a 6×1 zero vector and $\mathbf{D}$ is a 6×6 diagonal covariance matrix. The principal component c_t is a weighted average of these rates: $c_t = \mathbf{w}'\mathbf{rsn}_t$, where $\mathbf{w}$ is a 6×1 vector of loadings. We follow the yield factor approach and assume that c_t is observed without measurement error:

$$c_t = c(z_t, \mu, l) = \mathbf{w}'\hat{\mathbf{r}}(z_t, \mu, l).$$

This allows z_t to be obtained by numerical inversion:

$$z_t = g(c_t, \mu, l) = c^{-1}(c_t, \mu, l). \tag{14}$$

Substituting this back into (13) gives the model:

$$\begin{aligned}\mathbf{rsn}_t &= \hat{\mathbf{r}}(g(c_t, \mu, l), \mu, l) + \mathbf{esn}_t \\ &= \hat{\mathbf{f}}_t + \mathbf{esn}_t; \quad t = 1, ..., T.\end{aligned}$$

Thus the loglikelihood for the senior rates in period t can be written as:

$$Lsn_t = -\frac{1}{2}\ln(|\mathbf{D}|) - \frac{1}{2}(\mathbf{rsn}_t - \hat{\mathbf{f}}_t)'\mathbf{D}^{-1}(\mathbf{rsn}_t - \hat{\mathbf{f}}_t) \tag{15}$$

(neglecting the intercept $3\ln(2\pi)$ for simplicity). Summing this over T periods gives the loglikelihood for the senior rates over the full estimation period:

$$-\frac{T}{2}\ln(|\mathbf{D}|) - \frac{1}{2}\sum_{t=1}^{T}(\mathbf{rsn}_t - \hat{\mathbf{f}}_t)'\mathbf{D}(\mathbf{rsn}_t - \hat{\mathbf{f}}_t). \tag{16}$$

Similarly, we define a conformable vector of subordinated default rates for the shorter period ($t = N_{sb}, ...T$) in which these are available: $\mathbf{rsb}_t = \{rsb_{t,t+1}, ..., rsb_{t,t+10}\}'$; the loglikelihood for the subordinated rates in period t can be written as:

$$Lsb_t = -\frac{1}{2}\ln(|\mathbf{D}|) - \frac{1}{2}(\mathbf{rsb}_t - \hat{\mathbf{f}}_t)'\mathbf{D}^{-1}(\mathbf{rs}_t - \hat{\mathbf{f}}_t) \tag{17}$$

we obtain the loglikelihood for the entire cross section over the estimation period:

$$L(\mu, l, \mathbf{D}) = -\frac{(2T+1-N_{sb})}{2}\ln(|\mathbf{D}|) - \frac{1}{2}\sum_{t=1}^{T}(\mathbf{rsn}_t - \hat{\mathbf{f}}_t)'\mathbf{D}(\mathbf{rsn}_t - \hat{\mathbf{f}}_t) - \frac{1}{2}\sum_{t=N_{sb}}^{T}(\mathbf{rsb}_t - \hat{\mathbf{f}}_t)'\mathbf{D}(\mathbf{rsb}_t - \hat{\mathbf{f}}_t). \tag{18}$$

Optimizing $\mathbf{D}$ (Hamilton (1994)) gives the concentrated function:

$$L^*(\mu, l) = -\frac{2T+1-N_{sb}}{2}\ln\left(\frac{1}{2T+1-N_{sb}}\left[\sum_{t=1}^{T}(\mathbf{rsb}_t - \hat{\mathbf{f}}_t)'(\mathbf{rsb}_t - \hat{\mathbf{f}}_t) + \sum_{t=N_{sb}}^{T}(\mathbf{rsb}_t - \hat{\mathbf{f}}_t)'(\mathbf{rsb}_t - \hat{\mathbf{f}}_t)\right]\right). \tag{19}$$

This is optimized by minimizing (with respect to μ and l) the double sum in the square brackets. I do this using the *FindMinimum* numerical optimization package on *Matlab.* Standard errors and other diagnostics are obtained using the Hessian generated by *FindMinunc.* I initially used a grid of starting values but soon realized that they always converged to the same optimum which was a unique (for each bank). This is clear from graphs of the likelihood showing how it depends upon its two parameters, revealing that it is essentially quadratic in nature. The likelihood for the CIR model is similar, but uses (9) to define $\hat{\mathbf{f}}_t$ in (19). This is optimized with respect to the parameters κ, θ and σ.

The statistics used for testing the competing models are based on these optimized likelihood ratios. The Vuong (1989) test framework is based on the optimized likelihood for each observation ((15) plus (17)). It can be used for testing two non-nested models such as the CIR and DF models (Vuong, 2002, section 5). This statistic can be computed by calculating the relative loglikelihood of the two models for each observation and regressing these values against a constant. The Vuong non-nested test statistic is equal to the t-statistic in this regression and has a standard normal distribution in large samples. (Strictly speaking, this should be adjusted by a multiplier that depends upon the sample size but this is very close to unity in these large samples and can be neglected here.) The test is naturally two-sided and reported in table 3. In this application a positive value favours the CIR model and a negative one the DF model.

Tests of nested models are more straightforward and Vuong shows that under standard assumptions classical test statistics are valid. These are based on ratios of optimized likelihoods (or differences of loglikelihoods) for the full sample (19). These likelihood values can be adjusted to take account of the sample size and the number of parameters in competing models. For example, the Bayesian Information Criterion shown in tables 2 and 3 is computed as $BIC = 2LogL - N \ln(T)$, where N is the number of model parameters and T the number of observations. The Schwarz test is based on differences in these values. In table 2 for example $SCA = (BIC_{OLS} - BIC_{CIR})/2$ is used to test the CIR against the encompassing OLS model. This test has a χ^2_M distribution, with $M = 10$ being the difference in the parameters of the OLS and CIR models. The probability of observing these test values is effectively zero, indicating that the CIR model restrictions are not accepted by the data.

Tables

Table 1(a) Summary statistics for CDS based default probabilities

Type:	Senior						Subordinated					
Term:	1	2	3	5	7	10	1	2	3	5	7	10
Bank of America	January 2001-July 2014						August 2001-July 2014					
Mean	1.022	1.156	1.289	1.555	1.656	1.764	1.196	1.326	1.448	1.694	1.780	1.872
Std.Devn.	1.591	1.517	1.481	1.497	1.454	1.400	1.854	1.763	1.732	1.725	1.685	1.603
Skewness	2.875	2.445	2.065	1.487	1.311	1.160	2.957	2.513	2.211	1.753	1.599	1.455
Kurtosis	8.791	6.450	4.624	2.152	1.583	1.022	9.848	6.856	5.371	3.422	2.822	2.331
Auto.	0.890	0.910	0.922	0.936	0.939	0.941	0.884	0.908	0.917	0.929	0.928	0.924
ADF	-3.070	-2.762	-2.389	-1.869	-1.697	-1.539	-3.170	-2.780	-2.540	-2.170	-2.020	-1.900
Citigroup	January 2001-July 2014						June 2001-July 2014					
Mean	1.210	1.328	1.469	1.732	1.835	1.902	1.364	1.472	1.597	1.831	1.893	1.932
Std.Devn.	2.063	1.864	1.774	1.730	1.643	1.111	2.321	2.122	2.010	1.886	1.778	1.625
Skewness	3.525	2.837	2.302	1.646	1.293	1.025	3.452	2.974	2.636	1.923	1.619	1.354
Kurtosis	15.234	10.307	7.022	3.648	1.904	0.952	14.290	11.163	9.187	5.131	3.501	2.450
Auto.	0.910	0.906	0.913	0.902	0.867	0.780	0.921	0.923	0.929	0.920	0.883	0.801
ADF	-2.744	-2.613	-2.378	-1.977	-1.790	-1.673	-2.580	.2.393	-2.419	-2.110	-2.014	-1.922
JPMorgan	January 2001-July 2014						September 2001-July 2014					
Mean	0.587	0.719	0.855	1.133	1.257	1.403	0.692	0.803	0.920	1.159	1.253	1.363
Std.Devn.	0.594	0.582	0.602	0.669	0.678	0.677	0.638	0.645	0.652	0.705	0.698	0.685
Skewness	3.049	2.024	1.367	0.689	0.437	0.353	2.140	1.685	1.315	0.804	0.546	0.437
Kurtosis	14.509	6.859	2.931	0.030	-0.768	-0.999	7.281	4.489	2.825	0.723	-0.213	-0.580
Auto.	0.840	0.868	0.886	0.914	0.920	0.917	0.889	0.918	0.918	0.932	0.941	0.930
ADF	-2.936	-2.813	-2.541	-2.128	-1.945	-1.682	-2.306	-2.399	-2.305	-2.103	-2.056	-1.872

Table 1(b) Summary statistics continued

Type:	Senior						Subordinated					
Term:	1	2	3	5	7	10	1	2	3	5	7	10
Goldman	January 2001-July 2014						March 2006-July 2014					
Mean	1.264	1.384	1.504	1.737	1.849	1.968	2.172	2.059	2.232	2.483	2.539	2.598
Std.Devn.	1.632	1.517	1.460	1.401	1.335	1.245	1.489	1.516	1.446	1.322	1.246	1.170
Skewness	2.343	1.961	1.698	1.306	1.125	1.040	1.111	1.068	0.966	0.748	0.662	0.476
Kurtosis	5.781	3.754	2.515	1.119	0.673	0.491	0.230	0.272	0.084	-0.026	0.133	0.035
Auto.	0.883	0.892	0.907	0.925	0.923	0.920	0.873	0.869	0.878	0.873	0.845	0.845
ADF	-2.842	-2.769	-2.590	-2.237	-2.102	-2.001	-2.518	-2.792	-2.654	-2.674	-2.776	-2.735
Morgan Stanley	January 2001-July 2014						June 2004-July 2014					
Mean	1.844	1.876	1.925	2.110	2.177	2.265	2.145	2.180	2.253	2.465	2.454	2.498
Std.Devn.	3.188	2.542	2.202	2.010	1.859	1.696	2.860	2.471	2.190	2.081	1.897	1.783
Skewness	4.768	3.165	2.261	1.808	1.630	1.466	2.854	2.169	1.422	1.236	1.091	0.950
Kurtosis	33.872	15.371	7.090	4.364	3.532	2.576	12.588	7.162	1.990	1.772	1.378	0.806
Auto.	0.667	0.777	0.846	0.863	0.868	0.877	0.745	0.782	0.864	0.850	0.858	0.867
ADF	-2.846	-2.686	-2.499	-2.266	-2.174	-2.117	-2.520	-2.475	-2.392	-2.175	-2.168	-2.022
Wells Fargo	January 2001-July 2014						October 2002-July 2014					
Mean	0.573	0.669	0.775	1.018	1.112	1.231	0.725	0.825	0.912	1.129	1.211	1.305
Std.Devn.	0.833	0.778	0.760	0.805	0.768	0.729	0.979	0.985	0.897	0.916	0.873	0.845
Skewness	3.835	2.954	2.276	1.351	0.900	0.630	3.597	3.192	2.180	1.534	1.228	1.277
Kurtosis	19.184	12.451	7.757	2.715	0.767	-0.122	18.198	14.818	7.539	4.244	2.996	3.676
Auto.	0.877	0.889	0.898	0.916	0.924	0.920	0.876	0.892	0.918	0.933	0.933	0.914
ADF	-2.766	-2.578	-2.366	-1.898	-1.670	-1.643	-2.414	-2.368	-2.023	-1.809	-1.757	-1.789

This table shows the basic summary statistics for the default probabilities used in this study (as % p.a.) These are backed out from senior and subordinated debt CDS swap rates provided by Markit. The implied default probabilities $q_{t,t+m}$ for horizon or maturity m are obtained using standard recursion formulae (Hull (2003)). The default rates shown in this table are calculated as $-\ln p_{t,t+m}/m$, where $p_{t,t+m} = (1 - q_{t,t+m})$ are the survivorship probabilities. Mean denotes sample arithmetic mean; Std.Devn standard deviation and Auto. the first order monthly autocorrelation coefficient. Skewness & Kurtosis are standard measures of skewness (the third moment) and excess kurtosis (the fourth moment). ADF is the Adjusted Dickey-Fuller statistic testing the null hypothesis of non-stationarity. The 10% and 5% significance levels are -2.575 and -2.877 respectively.

Table 2: Affine model estimates (t-statistics in parentheses)

Model/Bank	**Loglike -lihood**	**BIC**	**SCA**	**Param -eters**		
OLS						
Bank of America	9795.9	19501.2				
Citigroup	9903.9	19717.0				
Goldman Sachs	8027.4	15966.9				
Morgan Stanley	8492.4	16895.5				
JPMorgan	10995.0	21899.1				
Wells Fargo	10267.0	20443.9				
Cox Ingersoll Ross				κ	θ	σ
Bank of America	9691.0	19291.4	104.9	0.1002	-0.0123	0.1590
				(5.77)	(0.13)	(45.60)
Citigroup	9727.9	19365.0	176.0	0.0290	-0.0633	0.1464
				(3.78)	(3.57)	(47.31)
Goldman Sachs	7977.5	15867.1	49.9	0.0676	-0.0262	0.1602
				(4.39)	(1.03)	(46.71)
Morgan Stanley	8372.1	16654.9	120.3	0.0007	-3.4201	0.1395
				(0.12)	(8.90)	(41.44)
JPMorgan	10955.0	21819.1	40.0	0.2407	-0.0027	0.1770
				(12.39)	(1.63)	(34.99)
Wells Fargo	10142.0	20193.9	125.0	0.1719	-0.0059	0.1933
				(11.49)	(2.17)	(39.13)

The first panel of this table shows the results of estimating an unrestricted OLS benchmark model and the second the results of imposing non-linear parameter restrictions implied by the model of Cox et al (1985) across the parameters of this model. The BIC reports the value of the Bayesian Information Criterion SCA reports the Schwarz test statistic for nested models (appendix 2).

Table 3: First Passage Time model estimates (t-statistics in parentheses)

Model/Bank	Loglike -lihood	BIC	SCA	Vuong test: DF versus:		Parameters	
Deferred Filtration				**OLS**	**CIR**	μ	l
Bank of America	9823.5	19631.9		-1.01	-4.86	-0.0810	3.4839
						(73.63)	(90.02)
Citigroup	10271.0	20526.9		-8.07	-10.32	-0.0300	2.2857
						(25.00)	(103.43)
Goldman Sachs	8028.2	16041.7		-0.04	-1.80	-0.0924	4.3668
						(30.02)	(78.90)
Morgan Stanley	8579.6	17144.3		-2.08	-3.61	-0.0472	2.2831
						(23.51)	(56.71)
JPMorgan	10864.0	21712.8		+3.86	+2.96	-0.0322	3.3660
						(29.27)	(133.04)
Wells Fargo	10370.0	20725.0		-2.65	-4.94	0.0305	2.4358
						(21.11)	(64.33)
Black-Cox							
						μ	
Bank of America	9380.0	18752.4	439.8			0.1254	
						(39.88)	
Citigroup	9742.3	19477.0	524.9			0.1299	
						(36.51)	
Goldman Sachs	7575.9	15144.5	488.6			0.1231	
						(41.00)	
Morgan Stanley	8088.7	16170.0	487.2			0.1024	
						(31.01)	
JPMorgan	10375.0	20742.4	485.2			0.1671	
						(38.98)	
Wells Fargo	9913.8	19820.1	452.5			0.1976	
						(42.12)	

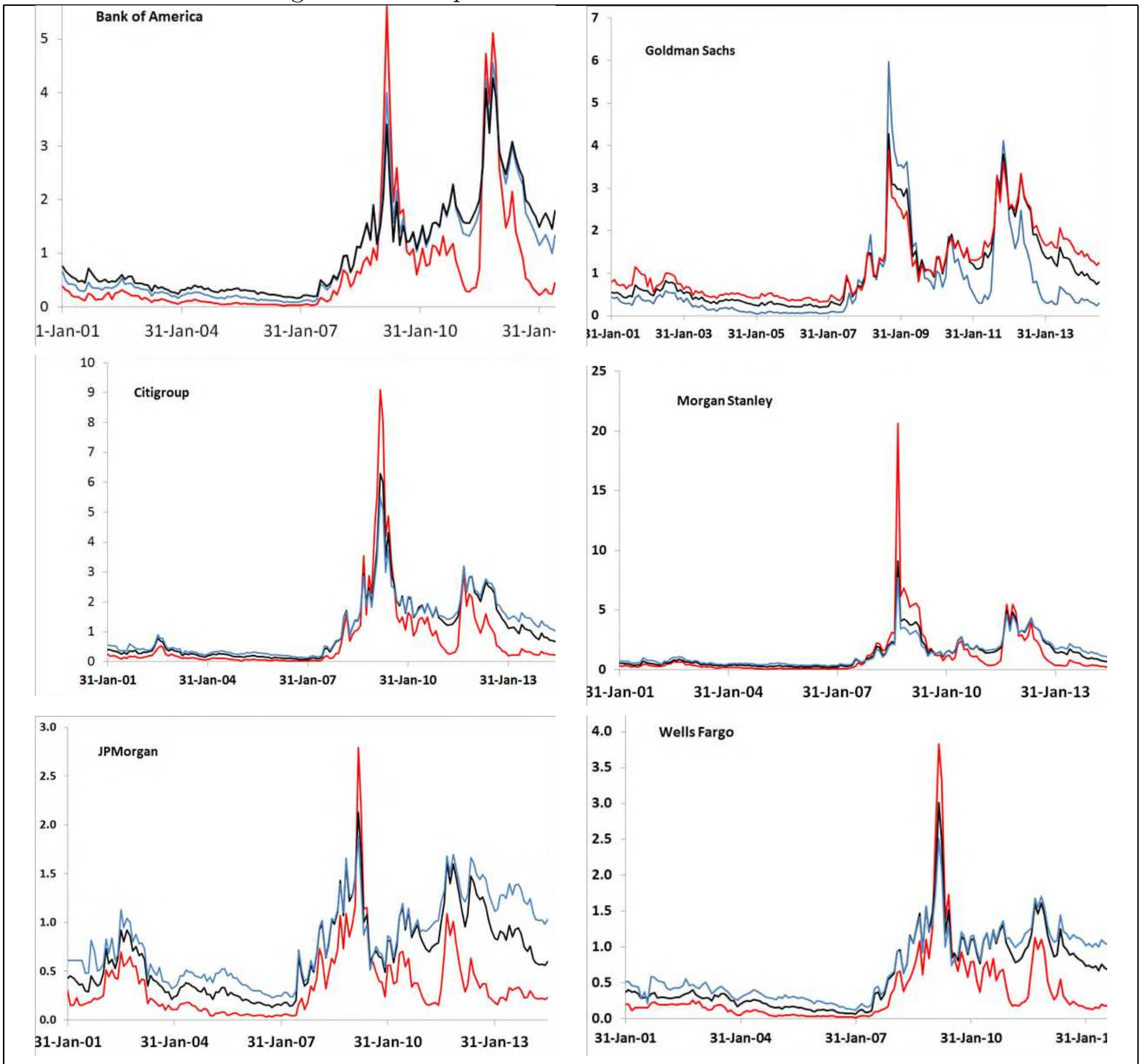
The first panel of this table reports the results for the DF model. This has a much higher likelihood than the standard full information BC model reported in the second panel, reflecting the significance of the accounting lag parameter l in the first panel. Indeed, the BC model is rejected against the DF model on the SCA test. This reveals a very significant degree of investor scepticism about accounting information. The Vuong (1989) non-nested test statistic is used to compare the DF model with the CIR model of table 2. This statistic has a standard normal distribution in large samples (Appendix 2). The positive value shown for JP Morgan favours the CIR model and the negative values for the other banks favour the DF model.

Table 4: DF model fit at different horizons

Type:	Senior						Subordinated					
Term:	1	2	3	5	7	10	1	2	3	5	7	10
Bank of America	January 2001-July 2014						August 2001-July 2014					
ESE	0.7100	0.5379	0.3865	0.1875	0.0313	0.1442	0.8992	0.7208	0.5873	0.4334	0.3883	0.4033
R^2	0.8009	0.8743	0.9319	0.9843	0.9995	0.9894	0.7648	0.8328	0.8850	0.9369	0.9469	0.9367
Citigroup	January 2001-July 2014						September 2002-July 2014					
ESE	0.6240	0.4571	0.3421	0.1614	0.0451	0.1346	0.7700	0.5678	0.4599	0.3000	0.2468	0.2847
R^2	0.9085	0.9399	0.9628	0.9913	0.9992	0.9853	0.8899	0.9284	0.9476	0.9747	0.9807	0.9693
JP Morgan Chase	January 2001-July 2014						September 2001-July 2014					
ESE	0.4197	0.3500	0.2814	0.1546	0.0400	0.1057	0.4254	0.3739	0.3036	0.1983	0.1193	0.1294
R^2	0.5007	0.6383	0.7815	0.9466	0.9965	0.9756	0.5555	0.6640	0.7832	0.9209	0.9708	0.9643
Goldman Sachs	January 2001-July 2014						December 2006-July 2014					
ESE	0.7717	0.5449	0.4056	0.1841	0.0495	0.1460	0.6951	0.5444	0.4612	0.2562	0.1926	0.2400
R^2	0.7764	0.8710	0.9228	0.9827	0.9986	0.9862	0.7821	0.8710	0.8983	0.9624	0.9761	0.9579
Morgan Stanley	January 2001-July 2014						June 2004-July 2014					
ESE	1.2916	0.7624	0.5119	0.2127	0.0460	0.1994	0.8760	0.6417	0.5565	0.3730	0.2614	0.2719
R^2	0.8359	0.9101	0.9460	0.9888	0.9994	0.9862	0.9062	0.9326	0.9354	0.9679	0.9810	0.9767
Wells Fargo	January 2001-July 2014						October 2002-July 2014					
ESE	0.3762	0.3038	0.2445	0.1190	0.0460	0.0902	0.4225	0.3957	0.3028	0.2248	0.1753	0.1950
R^2	0.7960	0.8475	0.8965	0.9782	0.9964	0.9847	0.8137	0.8386	0.8860	0.9398	0.9597	0.9467

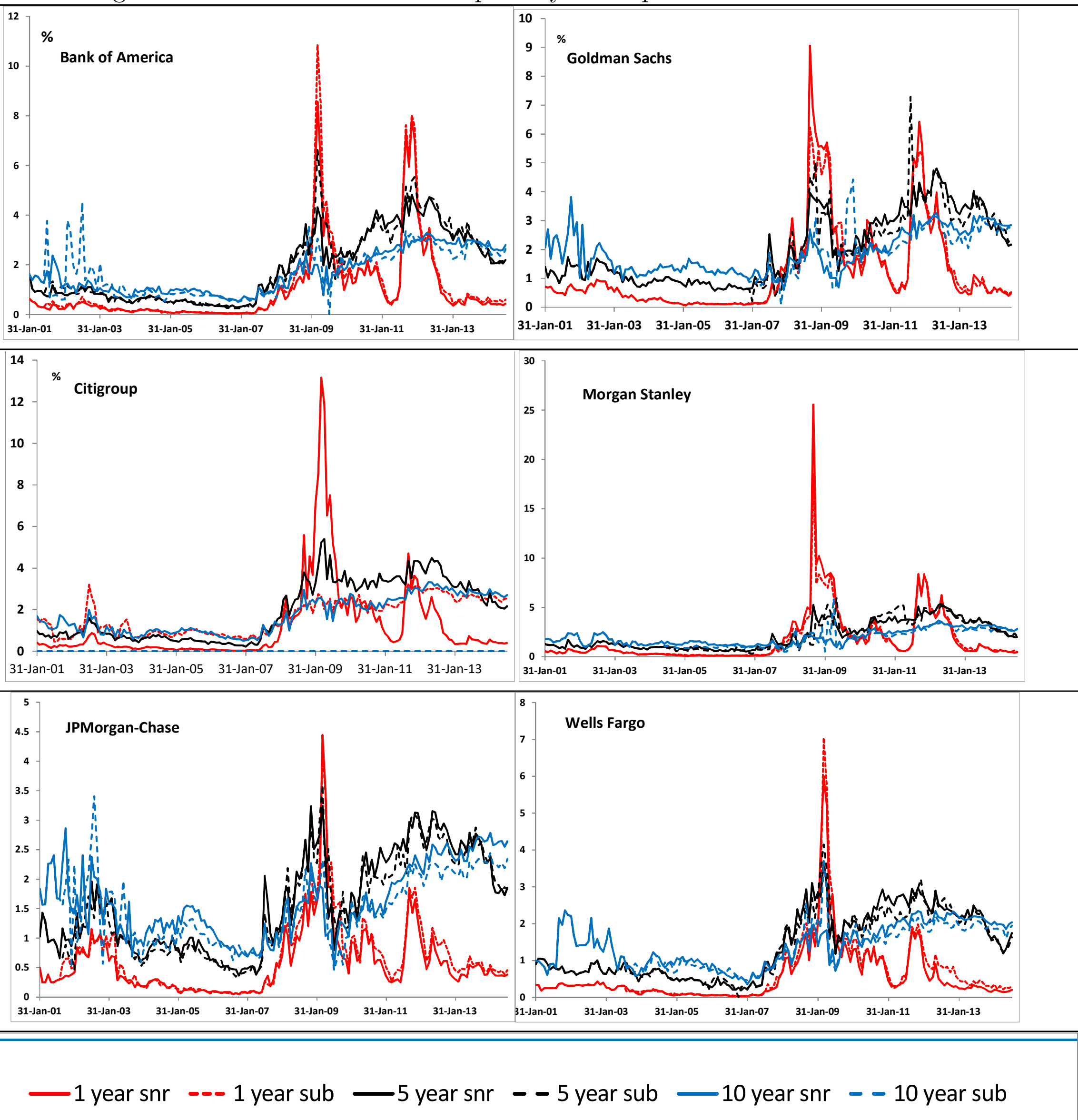
See notes to table 3. ESE denotes the equation standard error and R^2 the coefficient of determination.

Figure 1: CDS spreads on senior debt



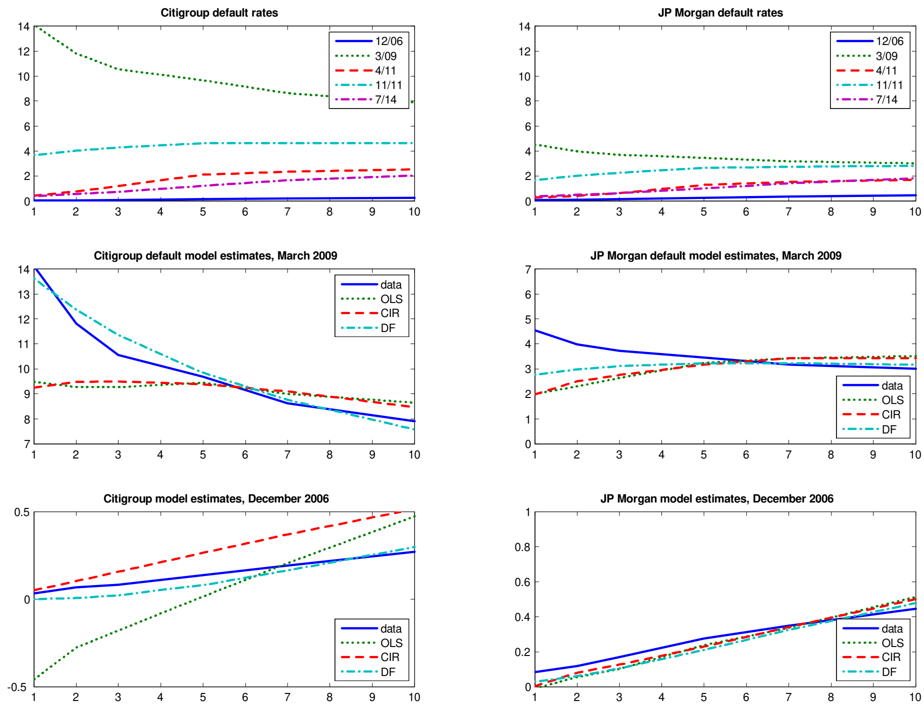
These CDS spread data were provided by Markit Ltd. They have a panel structure, consisting of daily observations on ten annual maturities of US bank debt CDS spreads. This figure shows end-month observations for one-year (red) five-year (black) and ten-year (blue) spreads on the senior debts of three large universal banks (Bank of America, Citigroup and JPMorgan), two investment banks (Goldman Sachs and Morgan Stanley) and Wells Fargo, a large regional bank.

Figure 2: Forward default rates implied by CDS spreads



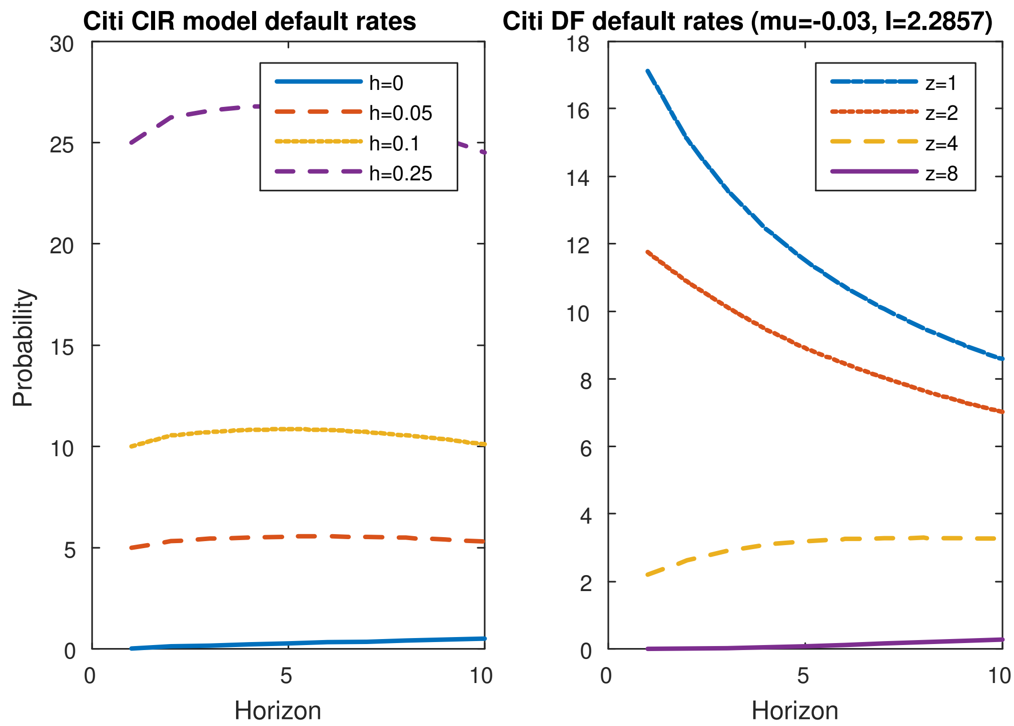
The continuous lines in this figure show one-year (red) five-year (black) and ten-year (blue) forward default rates backed out from the senior spreads shown in the previous figure. The equivalent values from the subordinated spreads (where available) are shown by broken lines. These forward rates are calculated as $\ln(p_{t,t+m-12}/p_{t,t+m})$ where $p_{t,t+m}$ is the probability of the bank surviving for m months. December 2006 sees an exceptionally compressed rate structure, especially in the case of Citigroup. These curves then move much higher and invert by March 2009. Again, this shift was especially pronounced in the case of Citigroup.

Figure 3: Empirical default rates



The two charts shown in the top panel of this figure depict the cross section of the default rates for Citigroup and JPMorgan during representative episodes. This figure shows the average default rates over different time horizons, computed as the negative of the log of the probability of survival divided by maturity. December 2006 shows an exceptionally compressed rate structure, especially in the case of Citigroup. These curves then move much higher and invert by March 2009. Again, this shift was especially pronounced in the case of Citigroup. The middle and bottom panels of the figure show the difficulty that two affine models (CIR and OLS) have in replicating these two extremes. The panels on the right show that JPMorgan was relatively immune to these developments and the affine models fit remarkably well in this case. Those on the right show that they impacted Citigroup harder, making it difficult for the affine models to fit these periods using parameters that are largely determined by the need fit the centre of the distribution. However, the DF model is non-linear and gives a much better representation of these extremes. See notes to the next figure, which shows how it is able to generate a range of different curve shapes.

Figure 4: Theoretical default rates



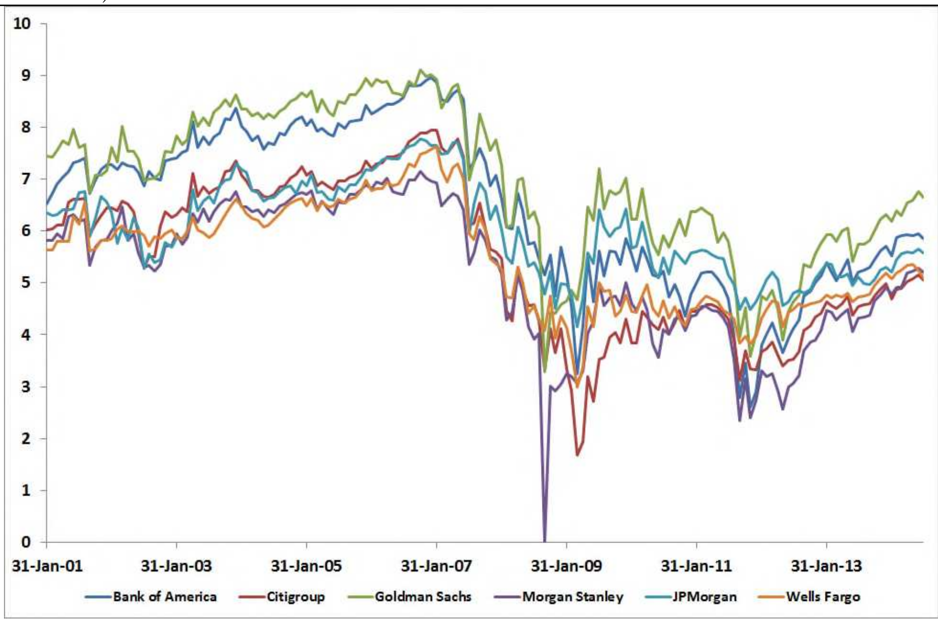
The right-hand panel of this figure shows how the DF specification models the extremes of the data for Citigroup and JPMorgan shown in the previous figure. This figure uses the parameter values shown in table 3. With a distance to default of $z = 8$ standard deviations, the risk distribution is well away from the default boundary area, so the default risk is negligibly small at all horizons up to 10 years. As the distance to default falls, the probability of a near-term default becomes very large but the survivorship effect means the longer term forward risk is much lower and the curve inverts sharply. In contrast, the affine structure of the CIR model shown in the left-hand panel means these curves all have a similar shape. They tend to move up and down in a parallel fashion, with little more variation at the short end than at the long end.

Figure 5: The effect of the deferred filtration in the model of Collin-Dufresne at al. (2003)

	Accounting lag		Maturity	
s	$\xrightarrow{\hspace{2cm}}$ l	t	$\xrightarrow{\hspace{2cm}}$ m	u
past		present		future

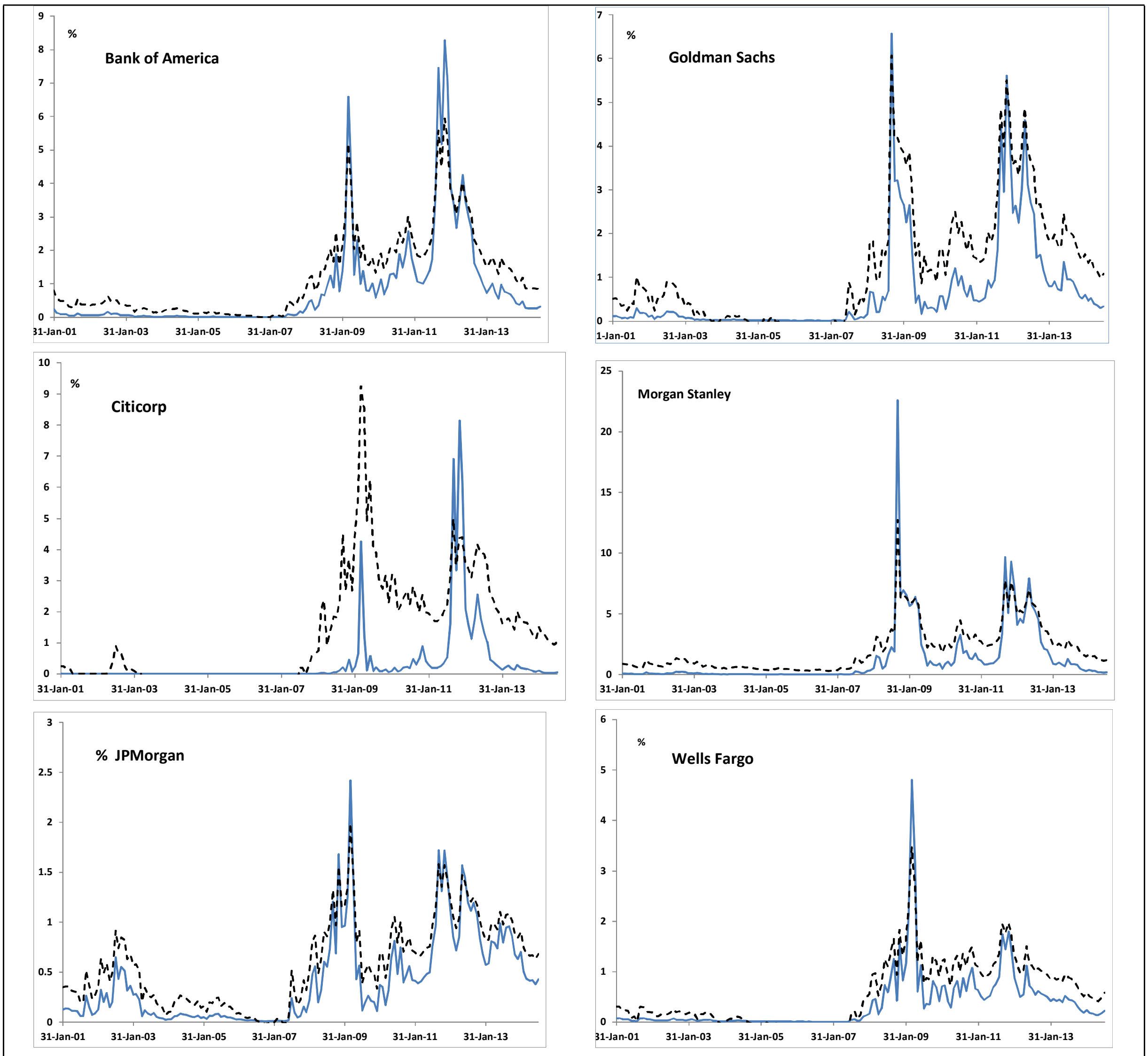
The $m-$maturity survival probability $p_{t,u}$ at t with the accounting lag l conditional upon no prior default is equal to the probability $p_{s,u}$ of survival from s to t and then u divided by the probability $p_{s,t.}$ of survival from s to t in the absence of an accounting lag: $p_{t,u} = p_{s,u}/p_{s,t}$.

Figure 6: Estimates of the distance to default in the DF model (z, in standard deviations)



This figure shows the estimates of the distance to default measures for the six banks from the DF model. These are highly correlated, suggesting that a common risk factor is at work. The first principal component explains 94% of the variance, although as in the case of the default rates shown in the earlier figures, the indicators for the two investment banks differ from the rest following the Lehman default. The previous figure shows the associated estimates of the hazard rates in this model and compares these with the estimates from the CIR model.

Figure 7: The estimated hazard rates



This figure contrasts the behavior of the hazard rates (instantaneous default intensities) in the DF (continuous blue line) and CIR (broken black line) models over time. The CIR model models the hazard rate like the spot rate in an affine term structure model. Appendix 1 derives the hazard rate function for the DF model. These rates are similar in the case of JPMorgan, but movements in the DF hazard rates tend to be more pronounced for the other banks than they are in the CIR model. The thin tails of the Gaussian distribution allow the DF model to replicate the very low default rates seen before the crisis without difficulty, but as noted in section 4.1, the CIR model has difficulty handling this.