Multiperiod Corporate Default Prediction with Partially-Conditioned Forward Intensity*

Jin-Chuan Duan† and Andras Fulop‡

(First Draft: September 24, 2012)
(This draft: September 18, 2013)

Abstract

The forward-intensity model of Duan, *et al* (2012) is a parsimonious and practical way for predicting corporate defaults over multiple horizons. However, it has a noticeable shortcoming because default correlations through intensities are conspicuously absent when the prediction horizon is more than one data period. We propose a new forward-intensity approach that builds in correlations among intensities of individual obligors by conditioning all forward intensities on the future values of some common variables, such as the observed interest rate and/or a latent frailty factor. The new model is implemented on a large sample of US industrial and financial firms spanning the period 1991-2011 on the monthly frequency. Our findings suggest that the new model is able to correct the structural biases at longer prediction horizons reported in Duan, *et al* (2012). Not surprisingly, default correlations are also found to be important in describing the joint default behavior.

Keywords: default, forward default intensity, pseudo-bayesian inference, sequential monte carlo, self-normalized asymptotics

---

*We would like to thank M. Rindisbacher, Sanjiv Das, and the seminar and conference participants at the National University of Singapore, CREST, ERCIM 2012 Oviedo, SOFIE-FGV Conference Rio de Janeiro 2012, and NYU Volatility Institute annual conference (2013) for useful comments.

†Duan is with Risk Management Institute and NUS Business School, National University of Singapore. Email: bizdjc@nus.edu.sg.

‡Fulop is with Department of Finance, ESSEC Business School. Email: fulop@essec.fr.

1 Introduction

Credit risk plays a central role in finance. A typical business operation entails some level of credit risk arising from lending/borrowing, counterparty linkages through derivatives, or supply-chain trade credits. Credit risk is in a nutshell the likelihood that a debtor is unable to honor its obligations in full. A natural and common way for analyzing credit risks of individual obligors has been to view credit risk in two steps. First, assess the likelihood of failing to pay in full; that is, probability of default. Second, estimate the extent of damage upon default; that is, loss given default. For the first step, the effort is to differentiate defaulters from non-defaulters whereas the second step focuses on identifying a loss-given-default distribution based on the experience of a pool of defaulters with similar characteristics.

Credit risk analysis often needs to go beyond individual obligors to analyze a credit portfolio's loss distribution. An ideal default prediction model should therefore be aggregatable from individual obligors to a portfolio level. By the credit literature's lingo, constructing a model that addresses both individual obligors and portfolios is a bottom-up approach, in contrast to a top-down approach that treats a portfolio as the basic entity. This paper addresses default prediction by proposing a new bottom-up approach for estimating probabilities of default for corporate obligors over different lengths of prediction horizons, and the proposed model is capable of generating superior performance both at the individual-obligor and portfolio levels.

Default prediction has a long history, and the academic literature goes back to Beaver (1966) and Altman (1968). Earlier models have a number of shortcomings, particularly they are not aggregatable either over time or across different obligors. Of late, default modeling has become more technically sophisticated using tools such as logistic regression and hazard rate models; for example, Shumway (2001), Chava and Jarrow (2004), Campbell, *et al* (2008), and Bharath and Shumway (2008). In these papers though, firms that disappear for reasons other than defaults are left untreated, which will result in censoring biases.

A different line of default prediction models has emerged in recent years which views individual defaults along with their economic predictors as a dynamic panel data; for example, Duffie, *et al* (2007), Duffie, *et al* (2009), and Duan, *et al* (2012). They adopted a Poisson intensity approach (Cox doubly stochastic process) to modeling corporate defaults where defaults/non-defaults are tagged with common risk factors and firm-specific attributes. In this line of research, firm exits for reasons other than default have also been explicitly modeled to avoid censoring biases. Common risk factors can be observable such as the interest rate, or latent such as the frailty factor in Duffie, *et al* (2009). The dynamic data panel (defaults or other exits conditional on stochastic covariates) can also be modeled using forward, as opposed to spot, intensities as in Duan, *et al* (2012). The dynamic data panel underlying this line of Poisson intensity models lends it naturally to aggregation both in time (multiperiod) and over obligors (portfolio).

As argued in Duan, *et al* (2012), the forward-intensity approach is more amenable to handling a dynamic panel with a large number of corporate obligors and/or with many firm-specific attributes used in default prediction. The reason is that by directly considering forward intensities, one in effect bypasses the challenging task of modeling and estimating an auxiliary system of time dynamics for the firm-specific attributes whose dimension (the number of firms times the number of firm attributes) can easily run up to several thousands or larger. Apart from being able to avoid the daunting task of modeling the very high-dimensional auxiliary system, Duan, *et al* (2012) showed that their forward-intensity model has more robust performance, because it directly links, through forward intensity, a default occurrence many periods into future to the current observation of covariates. This contrasts with the spot-intensity model which must deduce the probability of default multiperiods ahead from repeating one-period ahead predictions using the time series model for stochastic covariates.

But the current forward-intensity approach put forward in Duan, *et al* (2012) has a noticeable shortcoming. Due to the forward structure, correlations among spot intensities that could have been induced by common risk factors are conspicuously absent. As reported in Duan, *et al* (2012), their forward-intensity model's prediction on aggregate defaults over longer horizons exhibits systematic biases over time, overestimating (underestimating) defaults when the actual default rate was low (high).

In this paper, we propose a new forward-intensity approach that allows for correlations among forward intensities of individual obligors through some common risk factors (observed and/or latent). This is achieved by first constructing partially-conditioned forward intensities which depend on future values of some common risk factor(s). Partially-conditioned forward intensities are of course not observable at the time of performing default prediction, but they work like the forward intensities in Duan, *et al* (2012) once future values of the conditioning common risk factor(s) are assumed known. In essence, the approach dramatically reduces the dimension of the auxiliary system from all stochastic covariates (common risk factors and firm-specific attributes) down to the conditioning common risk factors, which are presumably the main source of default correlations anyway. As an example, use two common risk factors and four firm-specific attributes to model default behavior of 5,000 firms monthly over a period of twenty years. The spot-intensity approach of Duffie, *et al* (2007) will need to model and estimate a vector time series with a dimension of 20,002 over 240 monthly observations. Needless to say, the chosen model for such a high dimensional time series is at best *ad hoc*, and its estimation inevitably requires of placing parameter restrictions that are hard to justify.

Our partially conditioned forward-intensity approach will rely on a low-dimensional time series model to describe the dynamics of the conditioning common risk factors. For the above example, the partially-conditioned forward-intensity approach will only require a two-dimensional time series model for the two common risk factors on the monthly frequency. With such dramatic dimension reduction, one can (1) more confidently specify the time series model for the conditioning common risk factors, (2) estimate and simulate with relative ease in accordance with the time series dynamics of the conditioning common risk factors, (3) apply the forward-intensity model of Duan, *et al* (2012) on each simulated path, and (4) average the results over different simulations to take care of the randomness in future values of the conditioning common risk factors.

It should be noted that the partially-conditioned forward-intensity approach offers no material benefit for modeling individual defaults, because upon taking expectation of partially-conditioned forward intensities, they are back to full forward intensities. The benefit associated with this new approach lies in capturing joint default behaviors through intensity correlations that are made possible by conditioning on some common risk factors. To see this point, let us consider the partially-conditioned forward-intensity function for the period, say, from three to four months. After conditioning on the interest rate applicable to that particular time period, the partially-conditioned forward intensities are independent across all obligors. The conditional forward intensities of all obligors will move up or down together due to their sharing of the common stochastic interest rate. The conditional joint forward default probability still equals the product of the individual conditional forward default probabilities. It then follows that the joint forward default probability equals the expected value of the conditional joint forward default probability. Since the expected value of a product of random variables is higher than the product of individual expected values when their correlations are all positive, we can expect a higher joint default probability under the new forward-intensity model.

In order to implement this partially-conditioned forward-intensity model, we need to figure out a practical way of conducting parameter estimation and inference. The complication arises from the fact that due to the presence of the conditioning common risk factors, the pseudo-likelihood function is no longer decomposable across different forward starting times as is the case for Duan, *et al* (2012). Hence, we cannot independently estimate the parameters that drive forward intensities for different forward starting times. The large number of parameters in the joint model presents an estimation and inference challenge if one uses the usual gradient-based optimization method. Thus, we turn to a pseudo-Bayesian numerical device. In particular, we note that the sequence of pseudo-likelihoods through time defines a sequence of pseudo-posteriors from which one can sample using a sequential Monte Carlo method. We will then take the full-sample pseudo-posterior means to be our final parameter estimates. Consistency and asymptotic normality of our estimator follow from the results of Chernozhukov and Hong (2003). Further, access to the recursive pseudo-posterior means allows us to conduct inference using the self-normalized approach of Shao (2010), bypassing the delicate task of computing asymptotic variances. To our knowledge, this combination of sequential Monte Carlo and self-normalized asymptotics is new in the literature, and we believe is likely to be useful for the analysis of a wider range of richly parameterized financial and economic models. Adding to the benefit is a simple scheme for conducting periodic updates that is capable of revising the parameter estimates and inference with arrivals of new data along with revisions to the old data set. This is possible because of the sequential nature of our estimation and inference method.

We estimate the partially-conditioned forward-intensity model on a data set identical to the one used in Duan, *et al* (2012). The dataset is a large sample of US public firms over the period from 1991 to 2011 on a monthly frequency. There are 12,268 firms (both financial and non-financial) totaling 1,104,93 firm-month observations. The total number of defaults in this sample equals 994. The firm-specific and macro covariates used in this paper are also identical to that of Duan, *et al* (2012). To keep the number of parameters in the joint estimation manageable, we assume that the term structure of the coefficients for different forward starting times are of the Nelson-Siegel (1987) form, which according to the results reported in Duan, *et al* (2012) is flexible enough to deliver good empirical results. The common variables driving default correlations considered in this paper are the US Treasury rate and a latent common frailty factor.

In the empirical analysis, we compare four different specifications. First, we estimate the specification with independent forward defaults (i.e., switching off partial conditioning), which is that of Duan, *et al* (2012). We in essence reproduce their findings that the model produces a good default prediction performance at short horizons but shows structural biases at the longer end. In light of the frailty-based models (Duffie, *et al* (2009) and Koopman, *et al* (2009)), we investigate three specifications all with current value of the frailty factor, but with or without conditioning on future values of different variables. When the specification involves conditioning, we will use either future path of the interest rate or frailty factor. In all three implementations, current value of the frailty factor affects forward intensities, but its impact depends on whether some variable is used in conditioning. For the in-sample performance, we find that conditioning on future frailty path performs best, having largely corrected the over/under predictions at long horizons, and is good at discriminating good credits from bad ones.

Finally, we gauge the implications of forward default correlations on joint default behaviors. We produce the portfolio default distributions right before and after Lehman Brothers' bankruptcy under two sets of modeling assumptions on the model that is partially conditioned on the frailty factor. The first of these takes forward default correlations properly into account. In the second, the marginal default probabilities are kept fixed, but default correlations through the common risk factor are ignored. When markets are volatile (after Lehman Brothers' bankruptcy) or the horizon of interest is relatively long, the model without forward default correlations tends to yield default distributions that resemble normality. In contrast, the one with forward default correlations produces default distributions with heavy tails that are also right-skewed. We also compute the 99^{th} percentile of the portfolio default distribution implied by the two model specifications over the sample period, and find that forward default correlations have the first-order impact on this risk metric with the effect particularly strong at longer horizons. Thus, ignoring default correlations may lead to an overly optimistic view of the risk from holding credit-sensitive portfolios.

2 Forward intensities conditional on common risk factors

We follow Duffie, *et al* (2007) and Duan, *et al* (2012) to model default and other exits for the i-th firm in a group as being governed by two independent doubly stochastic Poisson processes – M_{it} with stochastic intensity λ_{it} for default and L_{it} with stochastic intensity ϕ_{it} for other exits. λ_{it} and ϕ_{it} are instantaneous intensities and are only known at or after time t. Our model development is along the forward intensity approach of Duan, *et al* (2012), hereafter DSW (2012). They defined the average spot combined exit intensity for default and other exits together for the period $[t, t+\tau]$ as

$$\psi_{it}(\tau) \equiv -\frac{\ln E_t\left[\exp\left(-\int_t^{t+\tau}(\lambda_{is}+\phi_{is})ds\right)\right]}{\tau}. \tag{1}$$

By this definition, $\exp[-\psi_{it}(\tau)\tau]$ becomes the survival probability over $[t, t+\tau]$. They then defined the forward combined exit intensity as

$$g_{it}(\tau) \equiv \psi_{it}(\tau) + \psi'_{it}(\tau)\tau \tag{2}$$

where $\psi_{it}(\tau)$ is assumed to be differentiable. Thus, the survival probability can be alternatively expressed as $\exp\left[-\int_0^\tau g_{it}(s)ds\right]$. Since default cannot be properly modeled without factoring in the censoring effect caused by other exits, the censored forward default intensity was defined by DSW (2012) as

$$f_{it}(\tau) \quad \equiv e^{\psi_{it}(\tau)\tau} \lim_{\Delta t\to 0} \frac{E_t\left[\int_{t+\tau}^{t+\tau+\Delta t} \exp\left(-\int_t^s(\lambda_{iu}+\phi_{iu})du\right)\lambda_{is}ds\right]}{\Delta t} \tag{3}$$

By this definition, the cumulative default probability over $[t, t+\tau]$ naturally becomes $\int_0^\tau e^{-\psi_{it}(s)s} f_{it}(s)ds$.

We construct the partially-conditioned forward intensities first by grouping covariates (common risk factors and firm-specific attributes) into two categories where $\mathbf{Z}_t = (z_{t,1}, z_{t,2}, \cdots, z_{t,m})$ is the set of common risk factors on which we will perform conditioning and $\mathbf{X}_{it} = (x_{it,1}, x_{it,2}, \cdots, x_{it,k})$ represents the remaining covariates attributable to obligor i. The conditioning set of common risk factors can, for example, be the risk-free rate of an economy and/or a latent frailty factor which is commonly used in the survival analysis. For the i-th obligor-specific attributes, $\mathbf{X}_{it}$, it may include common risk factors not used in the conditioning set. In the case of the j-th element being a macroeconomic variable, $x_{it,j}$ will be the same variable for all i's.

Let $E_t\left(\cdot \middle| \mathbf{Z}_u, u \le t+\tau\right)$ denote the expectation conditional on the σ-algebra generated jointly by the information set up to time t and $\{\mathbf{Z}_u, u \le t+\tau\}$. Our partially-conditioned forward-intensity equivalents are

$$\psi_{it}(\tau; \mathbf{Z}_u, u \le t+\tau) \equiv -\frac{\ln E_t\left[\exp\left(-\int_t^{t+\tau}(\lambda_{is}+\phi_{is})ds\right)\Big| \mathbf{Z}_u, u \le t+\tau\right]}{\tau} \tag{4}$$

$$g_{it}(\tau; \mathbf{Z}_u, u \le t+\tau) \equiv \psi_{it}(\tau; \mathbf{Z}_u, u \le t+\tau) + \psi'_{it}(\tau; \mathbf{Z}_u, u \le t+\tau)\tau \tag{5}$$

$$\begin{aligned} & f_{it}(\tau; \mathbf{Z}_u, u \le t+\tau) \\ & \equiv e^{\psi_{it}(\tau;\mathbf{Z}_u,u\le t+\tau)\tau} \lim_{\Delta t \to 0} \frac{E_t\left[\int_{t+\tau}^{t+\tau+\Delta t}\exp\left(-\int_t^s(\lambda_{iu}+\phi_{iu})du\right)\lambda_{is}ds\Big| \mathbf{Z}_u, u \le t+\tau\right]}{\Delta t}. \end{aligned} \tag{6}$$

By the above definitions, the survival probability over $[t, t+\tau]$, denoted by $S_{it}(\tau)$, becomes

$$S_{it}(\tau) = E_t\left[\exp(-\psi_{it}(\tau; \mathbf{Z}_u, u \le t+\tau)\tau)\right] = E_t\left[\exp\left(-\int_0^\tau g_{it}(s; \mathbf{Z}_u, u \le t+s)ds\right)\right]. \tag{7}$$

In addition, the forward default probability over $[t+\tau_1, t+\tau_2]$ evaluated at time t, denoted by $F_{it}(\tau_1, \tau_2)$, becomes

$$F_{it}(\tau_1, \tau_2) = E_t\left[\int_{\tau_1}^{\tau_2} e^{-\psi_{it}(s;\mathbf{Z}_u,u\le t+s)s} f_{it}(s; \mathbf{Z}_u, u \le t+s)ds\right]. \tag{8}$$

Note that the cumulative default probability over $[t, t+\tau]$, denoted by $P_{it}(\tau)$, will be a special case of the forward default probability; that is, $P_{it}(\tau) = F_{it}(0, \tau)$.

To specify the partially-conditioned forward intensities, we need to ensure that $g_{it}(\tau; \mathbf{Z}_u, u \le t+\tau) \ge f_{it}(\tau; \mathbf{Z}_u, u \le t+\tau) \ge 0$. This is of course to reflect the fact that intensities must be non-negative. Moreover, the combined exit intensity must be greater than the default intensity as long as they are conditioned on the same set of state variables.

Potentially, there are many functional forms for modeling intensities, but we keep the spirit of DSW (2012) by only modifying their specifications to accommodate conditioning in a rather simple way.

$$
\begin{aligned}
f_{it}(\tau;\mathbf{Z}_{t+\tau}) &= \exp[\alpha_0(\tau)+\alpha_1(\tau)x_{it,1}+\cdots+\alpha_k(\tau)x_{it,k}+\theta_1(\tau)z_{t,1}+\cdots+\theta_m(\tau)z_{t,m} \\
&\quad +\theta_1^*(\tau)(z_{t+\tau,1}-z_{t,1})+\cdots+\theta_m^*(\tau)(z_{t+\tau,m}-z_{t,m})] \qquad (9)\\
g_{it}(\tau;\mathbf{Z}_{t+\tau}) &= \exp[\beta_0(\tau)+\beta_1(\tau)x_{it,1}+\cdots+\beta_k(\tau)x_{it,k}+\eta_1(\tau)z_{t,1}+\cdots+\eta_m(\tau)z_{t,m} \\
&\quad +\eta_1^*(\tau)(z_{t+\tau,1}-z_{t,1})+\cdots+\eta_m^*(\tau)(z_{t+\tau,m}-z_{t,m})]+f_{it}(\tau;\mathbf{Z}_{t+\tau}) \qquad (10)
\end{aligned}
$$

Note that $f_{it}(\tau;\mathbf{Z}_{t+\tau})$ and $g_{it}(\tau;\mathbf{Z}_{t+\tau})$ do not need to share the same set of covariates, which can be achieved in the above specification by setting some coefficients to zero. By the above specifications, the partially-conditioned intensity functions are not observable at time t, and the unobservability is only due to $\mathbf{Z}_{t+\tau}$.

If we do away with conditioning on $\mathbf{Z}_{t+\tau}$, the model immediately becomes that of DSW (2012). A further restriction of $\tau = 0$ turns the model into the spot-intensity formulation of Duffie, *et al* (2007).[1]

The partially-conditioned forward-intensity model is more complicated to implement, as compared to that of DSW (2012), because we need to specify a dynamic model for $\mathbf{Z}_t$ so that the expectations for the default probabilities (forward and cumulative) can be performed. But the dimension of $\mathbf{Z}_t$ is expected to be small. Naturally, researchers are in a far better position to come up with a suitable time series model to describe some lower-dimensional common risk factors than to attempt a model for extremely high-dimensional firm-specific attributes (easily up to several thousands or more). The partially-conditioned forward-intensity model can accommodate a latent common state variable to build in the so-called frailty factor. Later, we will show how the model can be implemented empirically with and without a frailty factor.

[1]Since the immediate forward intensity is the spot intensity, the family of forward intensities specified in equation (9) or (10) may not be self-consistent with many joint dynamics of the stochastic covariates (common risk factors and firm-specific attributes). The issue of self-consistency applies to Duan, *et al* (2012) as well. For example, it is hard to imagine that when the stochastic covaraites follow a vector autoregressive dynamics, the form for the family of forward intensity functions can be generically specified as in those equations. In theory though, consistency is not an issue, because we do not need to specify the dynamics of the firm-specific attributes or how they interact with the conditioning common risk factors. In other words, it is highly unlikely that our forward intensity functions will be self-inconsistent with all possible joint dynamics of the stochastic covariates. On the practical level, a more relevant question is how well the *ad hoc* specification for the forward intensity functions in this paper or Duan, et al (2012) performs empirically.

3 Estimation and Inference

Firm i may exit the data sample either due to default or a non-default related reason, for example, merger/acquisition. Denote the combined exit time by τ_{Ci} and the default time by τ_{Di}. Naturally, $\tau_{Ci} \leq \tau_{Di}$. We have introduced cross-sectional intensity correlation by way of some common risk factors $\mathbf{Z}_{t+\tau}$. Our standing assumption is that conditional on the future realization of $\mathbf{Z}_{t+\tau}$, the individual event times (default or other exits) are independent.

The sample period $[0, T]$ is assumed to be divisible into $T/\Delta t$ periods. Let N be the total number of companies. For firm i, we let t_{0i} be the first time that it appeared in the sample. If τ is the intended prediction horizon, it amounts to $\tau/\Delta t$ periods with the minimum prediction horizon of Δt. Note that the minimum and maximum forward starting times will be 0 and $\tau - \Delta t$, respectively. We will increase or decrease t and τ by the increment of Δt. In the empirical implementation, we use one month as the basic interval, i.e., $\Delta t = 1/12$.

Let $\mathbf{Z}_{t:t+j} = \{\mathbf{Z}_t, \mathbf{Z}_{t+\Delta t}, \cdots, \mathbf{Z}_{t+j\Delta t}\}$. We define the following partially-conditioned forward probabilities:

$$\begin{aligned} & Prob_t(\tau_{Ci} > t + (j+1)\Delta t | \mathbf{Z}_{t:t+j}, \tau_{Ci} > t + j\Delta t) \\ = \; & e^{-g_{it}(j\Delta t; \mathbf{Z}_{t+j\Delta t})\Delta t} \qquad (11) \\ & Prob_t(t + j\Delta t < \tau_{Ci} = \tau_{Di} \leq t + (j+1)\Delta t | \mathbf{Z}_{t:t+j}, \tau_{Ci} > t + j\Delta t) \\ = \; & 1 - e^{-f_{it}(j\Delta t; \mathbf{Z}_{t+j\Delta t})\Delta t} \qquad (12) \\ & Prob_t(t + j\Delta t < \tau_{Ci} \neq \tau_{Di} \leq t + (j+1)\Delta t | \mathbf{Z}_{t:t+j}, \tau_{Ci} > t + j\Delta t) \\ = \; & e^{-f_{it}(j\Delta t; \mathbf{Z}_{t+j\Delta t})\Delta t} - e^{-g_{it}(j\Delta t; \mathbf{Z}_{t+j\Delta t})\Delta t} \qquad (13) \end{aligned}$$

The distribution of $\mathbf{Z}_t$ and the functions $g_{it}(j\Delta t; \mathbf{Z}_{t+j\Delta t})$ and $f_{it}(j\Delta t; \mathbf{Z}_{t+j\Delta t})$ completely characterize the joint conditional distribution of the exit times. Any relevant individual obligor's default probability can be computed with the conditional probabilities given the common risk factors and then integrating over the common risk factors. For any joint probability, one first comes up with the joint probability conditional on the common risk factors using conditional independence, and then integrates over the common risk factors. For example, individual obligor's forward default probability over the time interval $[t + j\Delta t, t + (j+1)\Delta t]$ can be computed using the following result:

$$\begin{aligned} & Prob_t(t + j\Delta t < \tau_{Ci} = \tau_{Di} \leq t + (j+1)\Delta t) \\ = \; & E_t\left[e^{-\sum_{s=0}^{j-1} g_{it}(s\Delta t; \mathbf{Z}_{t+s\Delta t})\Delta t}\left(1 - e^{-f_{it}(j\Delta t; \mathbf{Z}_{t+j\Delta t})\Delta t}\right)\right] \qquad (14) \end{aligned}$$

We denote the model parameters by θ. The likelihood corresponding to the time period

$j\Delta t$ for the prediction interval $[j\Delta t, j\Delta t + \tau]$ can be written as

$$L_{j,\tau}(\theta) = E_{j\Delta t}\left(\prod_{i=1}^{N} P_{i,j,\tau}(\theta; \mathbf{Z}_{j\Delta t:j\Delta t+\tau})\right) \tag{15}$$

where

$$\begin{aligned}
& P_{i,j,\tau}(\theta; \mathbf{Z}_{j\Delta t:j\Delta t+\tau}) \\
= \; & 1_{\{t_{0i}\leq j\Delta t, \tau_{Ci}>j\Delta t+\tau\}} \exp\left[-\sum_{k=0}^{\tau/\Delta t-1} g_{i,j\Delta t}(k\Delta t; \mathbf{Z}_{(j+k)\Delta t})\Delta t\right] \\
& +1_{\{t_{0i}\leq j\Delta t, j\Delta t<\tau_{Di}=\tau_{Ci}\leq j\Delta t+\tau\}} \exp\left[-\sum_{k=0}^{\tau_{Di}/\Delta t-j-2} g_{i,j\Delta t}(k\Delta t; \mathbf{Z}_{(j+k)\Delta t})\Delta t\right] \\
& \quad \times (1-\exp[-f_{i,j\Delta t}(\tau_{Di}-(j+1)\Delta t; \mathbf{Z}_{\tau_{Di}})\Delta t]) \\
& +1_{\{t_{0i}\leq j\Delta t, \tau_{Di}>\tau_{Ci}, j\Delta t<\tau_{Ci}\leq j\Delta t+\tau\}} \exp\left[-\sum_{k=0}^{\tau_{Ci}/\Delta t-j-2} g_{i,j\Delta t}(k\Delta t; \mathbf{Z}_{(j+k)\Delta t})\Delta t\right] \\
& \quad \times (\exp[-f_{i,j\Delta t}(\tau_{Ci}-(j+1)\Delta t; \mathbf{Z}_{\tau_{Ci}})\Delta t] - \exp[-g_{i,j\Delta t}(\tau_{Ci}-(j+1)\Delta t; \mathbf{Z}_{\tau_{Ci}})\Delta t]) \\
& +1_{\{t_{0i}>j\Delta t\}} + 1_{\{\tau_{Ci}\leq j\Delta t\}}
\end{aligned} \tag{16}$$

The first term on the right-hand side of the above expression is the probability (conditional on $\mathbf{Z}_{j\Delta t:j\Delta t+\tau}$) of surviving both forms of exit beyond $(j+1)\Delta t + \tau$. The second term is the forward probability that firm defaults in a particular period within the prediction horizon. The third term is the probability that firm exits due to other reasons. If the firm does not appear in the sample in month t, $P_{i,j,\tau}(\theta; \mathbf{Z}_{j\Delta t:j\Delta t+\tau})$ is set to 1.

Since some of the conditioning variables may be latent, we need to separate the observable variables, denoted by $\mathbf{Y}_t$, from those latent ones $\mathbf{F}_t$ in order to express the observable likelihood. Thus, $\mathbf{Z}_t = (\mathbf{Y}_t, \mathbf{F}_t)$. The expectation in equation (15) therefore only conditions on $\mathbf{Y}_t$ up to time $j\Delta t$ and other observed variables, $\mathbf{X}_t$, also up to time $j\Delta t$, that are not part of the conditioning set. Incorporating latent variables will fundamentally change the pseudo-likelihood function. We will use a fixed-parameter particle filtering scheme to deal with such latency. The fixed-parameter particle filter is then brought into the parameter estimation that is itself particle-based (representing the distribution of parameters by particles) sequential Monte Carlo parameter estimation method. The details are given in Appendix.

We aggregate the likelihoods over different time points into a total sample likelihood by taking a product of the individual likelihoods as follows:

$$\mathscr{L}_{\tau}(\theta;\tau_C,\tau_D,\mathbf{X},\mathbf{Y}) = \prod_{j=1}^{T/\Delta t-1} L_{j,\min(T-j\Delta t,\tau)}(\theta) \tag{17}$$

where $j\Delta t$ and T are the times for the first and last data points. Obviously, the above expression is not a true sample likelihood because the data periods are overlapped. For example, when $\tau = 3\Delta t$ and the data is sampled every Δt, three adjacent likelihoods must be dependent as a result. Due to the expectation operator in equation (15), the decomposability property of the forward-intensity model of DSW (2012) is no longer applicable. Lack of decomposability has important implementation implications. When τ is increased, the effective dimension of parameters will increase linearly with τ. We need to devise a new way to perform parameter estimation. In the Appendix, we explain in detail how the estimation and inference can be carried out.

4 Empirical Analysis

4.1 Data and implementation parameters

The data set used in the paper is the same as that of DSW (2012), which is a large sample of U.S. public firms over the period from 1991 to 2011 assembled from the CRSP monthly and daily files and the Compustat quarterly file. Defaults/bankruptcies and other forms of firm exit are taken by the Credit Research Initiative of Risk Management Institute, National University of Singapore, which collects default cases from Bloomberg, Moody's report, exchange web sites and news search. There are altogether 12,268 companies (including financial firms) giving rise to 1,104,963 firm-month observations in total. The number of active companies ranges from 3,224 in 2011 to 5,703 in 1998. The number of defaults/bankruptcies was as low as 15 (or 0.39%) in 2006 and as high as 160 (or 3.26%) in 2001. Other forms of firm exit are much more often. In 1993, there were 206 cases (4.91%) whereas in 1998 the number stood at 753 (13.2%). Readers are referred to DSW (2012) for more information on this data set.

In our empirical study, parameter estimation relies on a pseudo-Bayesian technique, and for which we set the number of parameter particles to $2,000$. We use a multivariate normal proposal in the move steps and execute three move steps after each resampling.[2] If a model contains a latent frailty factor, we use $1,000$ particles to run the fixed-parameter particle filter. Whenever partially-conditioned forward intensities are used, we simulate the conditioning common risk factors (observed or latent) with $1,000$ sample paths in order to compute the expectation in equation (15).

[2]When a parameter is constrained to be positive as in our later implementation, we will sample it as a truncated normal random variable by simply discarding negative values.

4.2 Smoothing and parameter restrictions

DSW (2012) demonstrated that the parameter estimates exhibit simple term structures. Using the Nelson-Siegel (1987) type of term structure function, they showed that the number of parameters can be significantly reduced while maintaining the model's performance for default prediction.[3] We will take advantage of this finding to impose a term structure on the parameters governing the partially-conditioned forward intensities. This dimension reduction is particularly important because the partially-conditioned forward-intensity model no longer has the kind of decomposability that prevents an accumulation of relevant parameters when the prediction horizon is lengthened. For instance, using 12 covariates for default prediction up to 36 months as in DSW (2012) would necessitate the joint estimation of $13 \times 36 = 468$ parameters where 13 is due to adding an intercept to the intensity function with 12 covariates.[4]

We work with smoothed parameters where smoothing in τ is performed with the Nelson-Siegel (1987) form with four free parameters; that is,

$$h(\tau; \varrho_0, \varrho_1, \varrho_2, d) = \varrho_0 + \varrho_1 \frac{1 - \exp(-\tau/d)}{\tau/d} + \varrho_2 \left[\frac{1 - \exp(-\tau/d)}{\tau/d} - \exp(-\tau/d) \right] \tag{18}$$

With the above smoothing, 36 parameters corresponding to one covariate can be reduced to just four parameters. When there are 12 covariates and one intercept, the total number of default parameters becomes $13 \times 4 = 52$. If the forward-intensity function is partially conditioned on two common risk factors, then there will be eight additional parameters to push the total number to 60. It would be challenging to use the gradient-based method to estimate so many parameters jointly. Nevertheless, the sequential Monte Carlo estimation and inference method described in the Appendix works quite satisfactorily in this case.

In our empirical implementation, we in fact set $\varrho_0 = 0$ and $d > 0$ in the Nelson-Siegel function when it is applied to the coefficients of stochastic covariates (except for those corresponding to the future values of the partial-conditioning variables). We do so because the current value of a stochastic covariate is not supposed to be informative of default or other exits when the forward starting time goes to infinity. In short, the coefficient of such a stochastic covariate should approach zero when τ goes to infinity. Our analysis suggests that setting such parameter restrictions can reduce the number of parameters without adversely

[3]DSW (2012) estimated the Nelson-Siegel function after obtaining estimates for the forward intensity function's parameters corresponding to different forward starting times. Here, we directly estimate the four parameters governing the Nelson-Siegel function without going through the intermediate step.

[4]In DSW (2012), an intervention dummy variable was introduced to account for the transitory effect of the AIG bailout. Our discussion on the number of parameters has excluded this dummy variable.

affecting the model's performance in any material way. Such implementation also facilitates a more sensible extrapolation in τ if a longer-term default prediction is called for.

Given that our main focus is on default behavior, we shut down the dependence of other exit intensities on the frailty factor and future values of the common risk variable. Under this set of restrictions, the only parameters to estimate in equation (10) are the $\beta_l(\tau), l \leq k$. Furthermore, just as in DSW (2012) these can be separately estimated from the default parameters. Here, we also use the Nelson-Siegel form on these parameters, and estimate them by a separate Sequential Monte Carlo run.

4.3 Macro drivers and the frailty factor

The conditioning set of common risk factors can be observable or latent. At the model formulation stage, there is no need to differentiate them. In estimation, however, the latent frailty variable will increase implementation complexity because a nonlinear and non-Gaussian filter is required to tackle latency. Since the frailty factor cannot be directly observed, it must be inferred while estimating the default and other exit intensities. Recall that the conditioning variable set $\mathbf{Z}_t$ consists of observable variables, $\mathbf{Y}_t$, and latent frailty variables, $\mathbf{F}_t$. We assume that there is only one frailty factor, i.e, F_t, but $\mathbf{Y}_t$ may be multi-dimensional. Both $\mathbf{Y}_t$ and F_t are assumed to be of the first-order autoregression, and they are independent of each other. Since F_t is latent, it incurs no loss of generality to make it independent of $\mathbf{Y}_t$.

The time series model is specified on the frequency of Δt to match the empirical setting. In the case of $\mathbf{Y}_t$,

$$\mathbf{Y}_{(j+1)\Delta t} = \mathbf{A} + \mathbf{B}(\mathbf{Y}_{j\Delta t} - \mathbf{A}) + \mathbf{U}_{j+1} \tag{19}$$

where $\mathbf{A}$ and $\mathbf{B}$ are coefficient vector and matrix to match the dimension of $\mathbf{Y}_{j\Delta t}$, and $\mathbf{U}_{j+1}$ is vector of normally distributed random variables with zero means and some covariance structure. Normality is not critical at all, because our method can easily accommodate non-normality. Sensible parameter restrictions may be imposed to make the model more parsimonious. In principle, the transition density of $\mathbf{Y}_{j\Delta t}$ can also enter the overall pseudo-likelihood function to perform a joint parameter estimation. It is not advisable, however, because the number of parameters may significantly increase. We thus estimate the vector autoregressive model for the common risk factors directly using their observations. The set of covariates used in DSW (2012) contains two common risk factors – three-month Treasury bill rate and one-year trailing S&P 500 index return. Here, we only use the three-month Treasury bill rate, because the results, not reported in this paper, suggest that the coefficients on the S&P 500 index return are unstable and this common risk factor does not help in default prediction.

For the frailty factor, we assume

$$F_{(j+1)\Delta t} = cF_{j\Delta t} + \epsilon_{j+1} \tag{20}$$

where ϵ_{j+1} is normally distributed with mean 0 and variance equal to one. Since $F_{j\Delta t}$ is latent, we have set its mean to zero and the variance of ϵ_{j+1} to one without any loss of generality. Again, the normality assumption is not at all essential to our method.

We investigate four different model specifications: (1) a model without correlations among the intensities (DSW), (2) a model where only the current value of frailty affects forward intensities (DSW-F) but still without intensity correlations, (3) a model where both the current and future values of the frailty factor enter the forward-intensity function (PC-F), and (4) one where the current value of the frailty factor and the future value of the interest rate (a macro driver) enters the forward-intensity function (PC-M).

4.4 Prediction accuracy and aggregate rate of default

The main objective of introducing the partially-conditioned forward intensity is to add default correlations to the model of DSW (2012). Improving the prediction accuracy for single-name defaults is not the purpose. Still, it is important to check whether our specifications allowing for default correlations also do a good job in predicting single-name defaults. Hence, in this subsection, we report their prediction accuracies and compare the aggregate default predictions with the realized default rates.

In DSW (2012), a bail-out dummy is used to capture the massive rescue operation launched by the US government when facing the imminent bankruptcy of AIG in September 2008 after Lehman Brothers' collapse. The DSW model implemented here slightly differs from that of DSW (2012) because we do not include the bail-out dummy. By adding a frailty factor to the DSW model, one could argue that the effect of the bail-out dummy may be absorbed by the frailty factor. Our later results seem to suggest that the frailty factor is able to account for the bail-out distortion just as in Koopman, *et al* (2009).

A good credit risk model should also be useful in discriminating good credits from bad ones. A common measure of this is accuracy ratio, the area between the power curve of a given model and the power curve of totally uninformative default predictions over the similar area defined by a perfect ranking model and totally uninformative default predictions. Both the power curve and the accuracy ratio are ordinal measures. They only depend on the rankings implied by default probabilities, not on their absolute magnitudes. A detailed description can be found in Crosbie and Bohn (2002). To examine our models' in-sample performance, we estimate the cumulative default probabilities for each firm-month observation employing the parameter values estimated on the full sample. Panel A of Table 1 reports the in-sample accuracy ratios for various prediction horizons – 1 month, 3 months, 6 months, 12 months, 24 months and 36 months.

As expected, the results for the DSW model are quite similar to the findings reported in DSW (2012). The predictions for short horizons are very accurate with the accuracy ratios for the 1-month and 3-month predictions exceeding 90%. The accuracy ratios for the 6-month and 12-month predictions are also very good with their values staying above 80%. As the horizon is increased to 24 months and 36 months, the accuracy ratios reduce to 73.93% and 66.49%, respectively. The other specifications perform similarly. However, one should note that the accuracy ratio is solely based on ordinal rankings and is also not able to discriminate between joint and separate defaults. Assessing a model's performance needs to be complemented by other means as well.

We further examine the prediction accuracy using two sub-samples – financial and non-financial groups. Panels B and C of Table 1 report the results. The accuracy ratios for the non-financial sub-sample are quite close to those of the full sample. For the financial sub-sample, however, the DSW model has a markedly lower accuracy at longer horizons. Adding a frailty factor to the DSW model, i.e., DSW-F, obviously improves the prediction accuracy for financial firms. The two partially-conditioned forward-intensity models, PC-F and PC-M, perform equally well as the DSW-F model in terms of the in-sample accuracy ratio analysis.

An out-of-sample analysis over time is also conducted, and the results are reported in Panel D of Table 1. Here, we estimate the models each month on an expanding window of data beginning at the end of 2000. We set the evaluation group to be the firm-months after the estimation window. For this out-of-sample accuracy ratio analysis, all models perform similarly.

At each month-end, we compute the predicted rate of default for a prediction horizon among the active firms in the sample. We then compare it with the realized rate of default for the same group of firms in the intended prediction period. For each model, we repeat this for the entire sample period and for different prediction horizons: 1 month, 3 months, 6 months, 12 months, 24 months and 36 months.

The bars in Figure 1 depict the realized rates of default for the intended prediction periods. The DSW model's results (solid curve) are basically same as the findings of DSW (2012), with a good performance at short horizons, but structural overestimation (underestimation) of default rates at the beginning (end) of the sample at longer horizons. In contrast, all other specifications with the frailty factor, go a long way in correcting the biases. The usefulness of a frailty factor in capturing common variation missed by the observable covariates is in line with the findings of Duffie, *et al* (2009) and Koopman, *et al* (2009). In Figure

1, we only present the results for the PC-M model (dashed curve), which is the best among the three specifications with the frailty factor. The model parameter values for both the DSW and PC-M models are in-sample, i.e., estimated using the whole sample. As Figure 1 shows, conditioning the forward intensities on a common risk factor such as frailty is very important in correcting prediction biases for longer horizons.

4.5 Parameter estimates

In the following discussion, we focus on the parameter estimates of the PC-F model. The results for the PC-M model are qualitatively similar. Panel A of Table 2 presents the parameter governing the frailty dynamics, suggesting high persistence with $c = 0.99$. Panel B of Table 2 reports the parameter estimates and confidence intervals of the partially-conditioned default intensity function. The parameter estimates and confidence intervals for the other-exits intensity function are given in Table 3. Note that in this study, the other-exits intensity function remains the same regardless of the model for the default intensity.

The Nelson-Siegel parameters by themselves are somewhat difficult to interpret. Thus, we report in Figures 2 and 3 the implied term structures for the coefficients in the default intensity functions for the firm-specific and common risk factors as well as their 90% confidence bands.

The patterns exhibited by the coefficients for the firm-specific attributes in Figure 2 are largely similar to those of of DSW (2012), and readers are referred to that paper for comparison. An important difference is that our inference is more conservative (wider confidence bands) by allowing unrestricted cross-sectional correlations in our pseudo-scores. Still, the coefficients on the firm-specific attributes remain fairly well identified. The situation changes for the common risk factors reported in Figure 3. The confidence bands are much wider than is the case for the macro variables in DSW (2012). This is likely due to our more conservative but more appropriate way of approaching the statistical inference. Similarly to DSW (2012), we find that coefficient of the 3-month Treasury bill rate switches signs. The current value of the frailty factor has a roughly constant effect on the forward intensities across the maturity spectrum, whereas the additional effect of the future path of the frailty factor increases with maturity.

Figure 4 depicts the path of the filtered values for the frailty factor, using the particle filter computed at the full-sample parameter estimates. We observe positive frailty values in the default waves before 2008, signaling that variation of the observable covariates is insufficient to fully explain the default cycles. Similar to what has been found in Koopman, *et al* (2009), the frailty factor takes up negative values during the recent financial crisis, reflecting fewer than the expected number of defaults given the value of the observable covariates. Hence, the frailty factor is likely to have taken up the role of the bail-out dummy variable of DSW (2012) to capture the massive government intervention post-Lehman Brothers' bankruptcy.

4.6 Portfolio default distributions

Allowing correlations among individual forward default intensities is of first-order importance when one is interested in the joint behavior of defaults. In this subsection, we use the PC-F model to illustrate this point. The PC-M model by partially-conditioning on future value of interest-rates can also produce similar results.

Figure 5 reports the difference in the log-pseudo-likelihoods at different time points between the PC-F and DSW models. In contrast to the measures examined earlier, these log-pseudo-likelihoods reflect the goodness-of-fit of the different specifications on the whole portfolio. One can clearly see that the PC-F model easily beats the DSW model throughout the entire sample period. Its superiority is especially clear at the beginning of the sample period and during the default waves. We note that using the results in Vuong (1989) we could form a formal statistical test based on the full-sample log-pseudo-likelihood difference. However, the Wald test on the parameter restrictions can be more easily conducted with the self-normalized approach described in the Appendix.[5]

Next, we investigate the consequences of forward default correlations on the distribution for the number of defaults in the portfolio. We use a conditional Monte-Carlo simulation technique to compute the default distribution under the PC-F model. Specifically, we first simulate paths of the common factors, and on each simulated path, we compute the exact conditional distribution for the number of defaults using the convolution algorithm given in Duan (2010). The unconditional default distribution then follows by averaging the conditional probabilities. To flush out the effect of default correlations, we also compute the individual default probabilities under the same model, and construct a version of the distribution for the number of defaults by assuming no forward default correlation.

Figure 6 reports these two distributions at two dates for three prediction horizons. The left column shows the predictions in May 2007, before the blow-up of the BNP hedge funds which was one of the first signs of the financial crisis. The right column reports the predictions in October 2008, the month after Lehman Brothers' bankruptcy. One can see in the top panel (one-month prediction horizon) that the distributions for the two dates are very different, with the one after the bankruptcy of Lehman Brothers being much closer to normality.

[5]The quantities that would be needed for the likelihood-based tests are the inverse Hessian and an estimate of the variance of the scores. The former is available from our algorithm (see Appendix) whereas the latter would need to be estimated by a HAC-type approach. Given strong dependence in the scores from the use of overlapping data points, we feel that invoking the usual consistency argument for the resulting variance estimator may be problematic.

The intuition is that much higher individual default probabilities allow the Central Limit Theorem to kick in faster to exhibit an approximate normality, whereas lower individual default probabilities make no default a more likely event and hence the portfolio default distribution exhibits a truncated look. Note also that the portfolio default distributions under two modeling assumptions are very similar. Forward default correlations has a rather minor effect simply because correlations in this case solely come from the latency of the frailty factor. The future value of the partial conditioning variable (i.e., the frailty factor in the PC-F model) cannot play any role for the one-month prediction horizon (i.e., $\tau = 0$).

A sharp contrast can be seen in the middle panel for the 12-month prediction horizon where the two portfolio default distributions show a large difference. The one under the PC-F model without assuming away default correlations shows heavy tails. Forward default correlations arising from partial conditioning obviously make the default distribution much more spread out. Furthermore, the shape of the distribution is also changed. Under the conditional independence assumption, the number of defaults is symmetric and is fairly close to being normally distributed. Once forward default correlations are properly accounted for, a more skewed and heavy-tailed distribution emerges. This result is due to mixing different conditional default distributions over the random future paths of the common factor (i.e., the frailty factor in the PC-F model). The bottom panel suggests that the effect of forward default correlations becomes even stronger when the prediction horizon is increased to 36 months.

To show that these results are not specific to a given date, Figure 7 presents the 99^{th} percentile of the portfolio default distribution, with and without forward default correlations, over the sample period. The upper tail of the portfolio default distribution is of particular interest to risk management (e.g., computing credit VaRs) and the rating and pricing of structured finance instruments like senior CDO tranches. Not surprisingly, Figure 7 suggests that ignoring forward default correlations as in the DSW model can lead to a severe underestimation of the upper tail of the portfolio default distribution. A novel and interesting result conveyed by the plots in Figures 6 and 7 is that the effect of forward default correlations becomes more pronounced when the prediction horizon is increased. Since structured debts may have fairly long maturities, say 5 years, the issue of default correlations should not be taken lightly.

5 Appendix

5.1 Handling the latent frailty factor: particle filtering

As the frailty factor is unobserved, we need a filter to compute conditional expectations. We use the smooth particle filter of Malik and Pitt (2011) to sample from the filtering distribution

$f(F_{j\Delta t}|\mathcal{D}_{j\Delta t})$ where $\mathcal{D}_{j\Delta t}$ is the information set at $j\Delta t$. To understand the algorithm, first consider the the following theoretical recursion:

$$f(F_{j\Delta t}|\mathcal{D}_{j\Delta t}) \propto f(F_{j\Delta t}|F_{(j-1)\Delta t}) \left(\prod_{i=1}^{N} P_{i,j,0}(Y_{(j-1)\Delta t}, F_{(j-1)\Delta t}) \right) f(F_{(j-1)\Delta t}|\mathcal{D}_{(j-1)\Delta t}) \qquad (21)$$

which follows from Bayes' rule and conditional independence of the one-period exit probabilities. Here, the one-period probability, $P_{i,j,0}(Y_{(j-1)\Delta t}, F_{(j-1)\Delta t})$, is defined by equation (16).

Also note that defaults in period $j\Delta t$ only depend on the frailty factor in period $(j-1)\Delta t$, due to our assumption that the variables are observed at the end of month (for instance, defaults in February depend on the value of the macro variables and frailty at the end of January).

This equation can then be implemented by sequential importance sampling with resampling. In particular, start with a set of M particles, denoted by $\{F^{(m)}_{(j-1)\Delta t}; m = 1, 2, \cdots, M\}$, that represents $f(F_{(j-1)\Delta t}|\mathcal{D}_{(j-1)\Delta t})$. Then,

1. Attach importance weights $w^{(m)}_{j\Delta t}$ to the particles

$$w^{(m)}_{j\Delta t} = \prod_{i=1}^{N} P_{i,j,0}(Y_{(j-1)\Delta t}, F^{(m)}_{(j-1)\Delta t}).$$

2. Resample the particles with weights proportional to $w^{(m)}_{j\Delta t}$ using the smooth bootstrap sampling technique of Malik and Pitt (2011).

3. Sample from the transition density

$$F^{(m)}_{j\Delta t} \sim f(F_{j\Delta t}|F^{(m)}_{(j-1)\Delta t}).$$

The resulting particle cloud, $\{F^{(m)}_{j\Delta t}; m = 1, 2, \cdots, M\}$, approximately represents $f(F_{j\Delta t}|\mathcal{F}_{j\Delta t})$. Then, to obtain the expectation over the path of the common variables $\tilde{Z}_{j\Delta t:j\Delta t+\tau}$ as required by equation (15), we simulate M paths to compute the sample average. To decrease Monte Carlo noise, we use the same random numbers for calls to the particle filter at different parameter values.

5.2 Parameter estimation by Sequential Monte Carlo (SMC)

Lack of decomposability means that the number of parameters to be estimated jointly will be very large. The conventional gradient-based optimization methods will not work well and statistical inference becomes challenging. Fortunately, a pseudo-Bayesian device can be used for estimation and inference. Consider the following pseudo-posterior distribution:

$$\gamma_n(\theta) \propto \prod_{j=1}^{n-1} L_{j,\min(T-j\Delta t,\tau)}(\theta)\pi(\theta), \text{ for } n = 2, \cdots, \frac{T}{\Delta t} \tag{22}$$

and apply the sequential batch-resampling routine of Chopin (2002) together with tempering steps as in Del Moral, *et al* (2006). For each n, this procedure yields a weighted sample of N points, $(\theta^{(i,n)}, w^{(i,n)})$ for $i = 1, \cdots, N$, whose empirical distribution function will converge to $\gamma_n(\theta)$ as N increases. In what follows, the superscript i is for $i = 1, \cdots, N$.

***Initialization*:** Draw an initial random sample from the prior: $(\theta^{(i,1)} \sim \pi(\theta), w^{(i,1)} = \frac{1}{N})$ Here, the only role of the prior, $\pi(\theta)$, is to provide the initial particle cloud from which the algorithm can start. Of course, the support of $\pi(\theta)$ must contain the true parameter value θ_0.

***Recursions and defining the tempering sequence*:** Assume we have a particle cloud $(\theta^{(i,n)}, w^{(i,n)})$ whose empirical distribution represents $\gamma_n(\theta)$. Then, we will arrive at a cloud representing $\gamma_{n+1}(\theta)$ by combining importance sampling and MCMC steps. Sometimes moving directly from $\gamma_n(\theta)$ to $\gamma_{n+1}(\theta)$ is too ambitious as the two distributions are too far apart, and will result in highly variable importance weights. We follow Del Moral, *et al* (2006) to build a tempered bridge between the two densities and let our particles evolve through a sequence of densities. In the transition to time $n+1$, we rely on P_{n+1} intermediate densities:

$$\bar{\gamma}_{n+1,p}(\theta) \propto \gamma_n(\theta) L^{\xi_p}_{n+1,\min(T-(n+1)\Delta t,\tau)}(\theta) \tag{23}$$

to define an appropriate bridge. When $\xi_0 = 0$, $\bar{\gamma}_{n+1,0}(\theta) = \gamma_n(\theta)$, and when $\xi_{P_{n+1}} = 1$, $\bar{\gamma}_{n+1,P_{n+1}}(\theta) = \gamma_{n+1}(\theta)$. We can initialize a particle cloud representing $\bar{\gamma}_{n+1,0}(\theta)$ as $(\bar{\theta}^{(i,n+1,0)} = \theta^{(i,n)}, \bar{w}^{(i,n+1,0)} = w^{(i,n)})$. Then, for $p = 1, \ldots, P_{n+1}$ we move through the sequence as follows.

- *Reweighting Step:* In order to arrive at a representation of $\bar{\gamma}_{n+1,p}(\theta)$ we can simply use the particles representing $\bar{\gamma}_{n+1,p-1}(\theta)$ and the importance sampling principle. This leads to

$$\begin{aligned} \bar{\theta}^{(i,n+1,p)} &= \bar{\theta}^{(i,n+1,p-1)} \\ \bar{w}^{(i,n+1,p)} &= \bar{w}^{(i,n+1,p-1)} \frac{\bar{\gamma}_{n+1,p}(\bar{\theta}^{(i,n+1,p)})}{\bar{\gamma}_{n+1,p-1}(\bar{\theta}^{(i,n+1,p)})} = \bar{w}^{(i,n+1,p-1)} L^{\xi_p-\xi_{p-1}}_{n+1,\min(T-(n+1)\Delta t,\tau)}(\bar{\theta}^{(i,n+1,p)}) \end{aligned}$$

If one would keep on repeating the reweighting step, quickly all the weights would concentrate on a few particles, the well known phenomenon of particle impoverishment in sequential importance sampling. In the SMC literature, this is tackled by the resample-move step (see e.g Chopin (2002)), and is triggered whenever a measure of particle diversity, the efficient sample size defined as $ESS = \frac{\left(\sum_{i=1}^{N} \bar{w}^{(i,n+1,p)}\right)^2}{\sum_{i=1}^{N}\left(\bar{w}^{(i,n+1,p)}\right)^2}$, falls below a pre-defined constant, B. Resampling directs the particle cloud towards more likely areas of the sample space, while move enriches particle diversity.

If $ESS < B$, perform the following resampling-move step.

- *Resampling Step:* The particles are resampled proportional to their weights. If $I^{(i,n+1,p)} \in (1, \ldots, N)$ are particle indices sampled proportional to $\bar{w}^{(i,n+1,p)}$, the equally weighted particles are obtained as

$$\begin{aligned}\bar{\theta}^{(i,n+1,p)} &= \bar{\theta}^{(I^{(i,n+1,p)},n+1,p)} \\ \bar{w}^{(i,n+1,p)} &= \frac{1}{N}\end{aligned}$$

- *Move Step:* Each particle is passed through a Markov Kernel $K_{n+1,p}(\bar{\theta}^{(i,n+1,p)}, \cdot)$ that leaves $\bar{\gamma}_{n+1,p}(\theta)$ invariant, typically a Metropolis-Hastings kernel:

  1. Propose $\theta^{*(i)} \sim Q_{n+1,p}(\cdot \mid \bar{\theta}^{(i,n+1,p)})$
  2. Compute the acceptance weight $\alpha = \min\left(1, \frac{\bar{\gamma}_{n+1,p}(\theta^{*(i)})Q_{n+1,p}(\bar{\theta}^{(i,n+1,p)}|\theta^{*(i)})}{\bar{\gamma}_{n+1,p}(\bar{\theta}^{(i,n+1,p)})Q_{n+1,p}(\theta^{*(i)}|\bar{\theta}^{(i,n+1,p)})}\right)$
  3. With probability α, set $\bar{\theta}^{(i,n+1,p)} = \theta^{*(i)}$, otherwise keep the old particle

This step will enrich the support of the particle cloud while conserving its distribution. If the particle set is a poor representation of the target distribution, the move step can help adjust the location of the support. Since we are in an importance sampling setup, the proposal distribution, $Q_{n+1,p}(\cdot \mid \bar{\theta}^{(i,n+1,p)})$, can be adapted using the existing particle cloud. In our implementation, for instance, we use block independent normal distribution proposals that are fitted to the particle cloud before the move. Three (or four) Nelson-Siegel parameters corresponding to each covariate form one block. We need to ensure positive d, and it is accomplished by discarding any non-positive sampled d value along with its block and then resample. Note that the likelihood ratio in the Metropolis-Hastings algorithm will not be affected by this operation, because the ratio of two truncated normal densities is the same as the ratio of two normal densities (i.e., the adjustment term due to truncation is common to both densities).

To improve the support of the particle cloud further, one can execute multiple such Metropolis-Hastings steps each time. In our implementation, additional moves are performed only after ξ_p reaches 1. Each move uses the means implied by the particle set but all standard deviations are increased by a factor of 30%. The number of moves is set to 20 for the first time point and exponentially declines to 3 mid-way to the sample period and stays at 3 for the remainder.

When $p = P_{n+1}$ is reached, we obtain a representation of $\gamma_{n+1}(\theta)$ as $(\theta^{(i,n+1)} = \bar{\theta}^{(i,n+1,P_{n+1})}, w^{(i,n+1)} = \bar{w}^{(i,n+1,P_{n+1})})$. Following Del Moral, *et al* (2011), we set the tempering sequence ξ_p automatically to ensure that the efficient sample size stays close to B. We achieve this by a grid search, where we evaluate the ESS at a grid of candidate ξ_p and then settle with the one that produces the closest ESS to B.

***Pre-Estimated Macro Dynamics*:** It is important to point out an additional complication when the future path of some macro-variables is part of the conditioning set. As suggested in Section 4.3, the parameters of the macro-dynamics are estimated in a preliminary run, separately from the SMC procedure. Further, to properly conduct inference as described in the next section, we need an expanding data structure for all the parameter estimates. Hence, when moving from n to $n+1$, the point estimate for the parameter of the macro-dynamics changes from $\hat{\theta}_n^M$ to $\hat{\theta}_{n+1}^M$. Then, for all $j \leq n+1$, the pseudo-likelihoods change from $L_{j,\min(T-j\Delta t,\tau)}(\theta \mid \hat{\theta}_n^M)$ to $L_{j,\min(T-j\Delta t,\tau)}(\theta \mid \hat{\theta}_{n+1}^M)$. To accommodate this change, the following approach is used:

- First, the parameters of the macro-dynamics are kept fixed at $\hat{\theta}_n^M$ and the algorithm described above allows us to obtain a representation of $\prod_{j=1}^{n} L_{j,\min(T-j\Delta t,\tau)}(\theta \mid \hat{\theta}_n^M)\pi(\theta)$.
- Then, the new macro-parameter, $\hat{\theta}_{n+1}^M$ is computed. We obtain a representation of the new target, $\prod_{j=1}^{n} L_{j,\min(T-j\Delta t,\tau)}(\theta \mid \hat{\theta}_{n+1}^M)\pi(\theta)$, by multiplying the weight with the following factor: $w_{inc} = \frac{\prod_{j=1}^{n} L_{j,\min(T-j\Delta t,\tau)}(\theta|\hat{\theta}_{n+1}^M)}{\prod_{j=1}^{n} L_{j,\min(T-j\Delta t,\tau)}(\theta|\hat{\theta}_{n}^M)}$.

5.3 Inference

The full sample size has $T/\Delta t$ time series data points but one can only make default predictions at $T/\Delta t - 1$ time points; for example, at time point 2, the data is only available for making one-period default prediction at time point 1. Denote the pseudo-posterior mean of the parameter of the whole sample by $\hat{\theta}_{T/\Delta t}$ and for $n = 2, \cdots, T/\Delta t$,

$$\hat{\theta}_n = \frac{1}{\sum_{i=1}^{N} w^{(i,n)}} \sum_{i=1}^{N} w^{(i,n)} \theta^{(i,n)}$$

Note that $\gamma_n(\theta)$ is not a true posterior because the likelihood function in equation (22) is not a true likelihood function. Thus, it cannot directly provide valid Bayesian inference. However, we can turn to the results of Chernozhukov and Hong (2003) to give a classical interpretation to our simulation outcome. In particular, denote the log-pseudo posterior by $l_T(\theta) = \sum_{j=1}^{T/\Delta t-1} \ln L_{j,\tau}(\theta) + \ln \pi(\theta)$. Let the pseudo-score matrix be $s_T(\theta) = \nabla_\theta l_T(\theta)$ and the negative of the scaled Hessian be $J_T(\theta) = -\nabla_{\theta\theta'} l_T(\theta)/(T/\Delta t - 1)$. Then, Theorem 1 of Chernozhukov and Hong (2003) states that for large T, $\gamma_{T/\Delta t}(\theta)$ is approximately a normal density with the random mean parameter: $\theta_0 + J_T(\theta_0)^{-1}s_T(\theta_0)/(T/\Delta t - 1)$. This result implies that the pseudo posterior mean $\hat{\theta}_{T/\Delta t}$ provides a consistent estimate for θ_0, and the scaled estimation error can be characterized by

$$\sqrt{T/\Delta t - 1}(\hat{\theta}_{T/\Delta t} - \theta_0) \approx J_T(\theta_0)^{-1}s_T(\theta_0)/\sqrt{T/\Delta t - 1} \tag{24}$$

In addition to consistency, asymptotic normality follows by applying the Central Limit Theorem (CLT) to the scores. Chernozhukov and Hong (2003) showed that asymptotically, the covariance matrix of the pseudo-scores is equal to the inverse Hessian, $J_T(\theta_0)^{-1}$, hence the latter can be estimated by the former. In turn, the variance of the scores could be estimated by its empirical counterpart. All this would allow us to construct the sandwich matrix necessary for inference.

However, access to the recursive estimates $\hat{\theta}_n$ provides us a more direct way to construct the confidence band using the self-normalized approach of Shao (2010). Thus, we can bypass the delicate issue of estimating asymptotic variance. Knowing that the pseudo-score is asymptotically normal, we assume that the functional CLT applies to the scaled pseudo-score:

$$J_T(\theta_0)^{-1}\frac{1}{\sqrt{T/\Delta t - 1}}s_{[rT]}(\theta_0) \to^d SW_k(r), r \in [0, 1] \tag{25}$$

where S is an unknown lower triangular matrix, k is the number of parameter, and $W_k(r)$ is a k-dimensional Brownian motion.

Following Shao (2010), we define a norming matrix

$$\hat{C}_T = \frac{1}{(T/\Delta t - 1)^2}\sum_{l=2}^{T/\Delta t} l^2(\hat{\theta}_l - \hat{\theta}_{T/\Delta t})(\hat{\theta}_l - \hat{\theta}_{T/\Delta t})' \tag{26}$$

Then, we can form the following asymptotically pivotal statistic

$$(T/\Delta t - 1)(\hat{\theta}_{T/\Delta t} - \theta_0)\hat{C}_T^{-1}(\hat{\theta}_{T/\Delta t} - \theta_0)' \to^d W_k(1)P_k(1)^{-1}W_k(1) \tag{27}$$

where the asymptotic random norming matrix $P_k(1) = \int_0^1 (W_k(r) - rW_k(1))(W_k(r) - rW_k(1))' dr$ is a path functional of the Brownian bridge. The main insight is that the nuisance scale matrix S in equation (25) disappears from this quadratic form, and thus it need not be estimated.

The above result can be used to form tests. For example, we can test the hypothesis of the i-th element of θ_0, denoted by $\theta_0^{(i)}$, equal to a by the following robust analogue to the t-test:

$$t^* = \frac{\sqrt{T/\Delta t - 1}\left(\hat{\theta}^{(i)}_{T/\Delta t} - a\right)}{\sqrt{\hat{\delta}_{i,T}}} \to^d \frac{W(1)}{\left[\int_0^1 (W(r) - rW(1))^2 dr\right]^{1/2}} \tag{28}$$

where $\hat{\delta}_{i,T}$ is the i^{th} diagonal element of $\hat{C}_T$. The right-hand-side random variable does not have a known distribution, but can be easily simulated. Kiefer, *et al* (2000) reported that the 95% quantile is 5.374 and the 97.5% quantile is 6.811. These values can also be used to set up confidence intervals. Note that the same logic applies to any functional transformation of θ; for example, any forward intensity function's coefficient estimate is a functional transformation of some elements of $\hat{\theta}_{T/\Delta t}$ and its corresponding $\hat{C}_T$ and $\hat{\delta}_T$ can be likewisely computed.

We would like to point out a particular strength of our inference procedure. It does not require an explicit computation of the individual scores or the Hessian, and hence can be seamlessly extended to non-smooth objective functions without any extra work. In our case, specifications with multi-dimensional frailty factors (e.g., to allow for industry-level frailties as in Koopman, *et al* (2009)) would constitute such an example, because the smooth particle filter of Malik and Pitt (2011) only works for one-dimensional latency. In general, our combination of recursive parameter estimation and self-normalized inference is likely to be useful for the analysis of any dynamic model with a particle filter and involving pseudo-likelihoods. An alternative approach for such cases will be to use the sandwich matrix and to compute the scores using the marginalized score method advocated in Doucet and Shephard (2012).

5.4 Periodic updates

In reality, portfolio credit risk models need to be updated periodically as new data arrive and/or some old data get revised. One may want to revise the parameter estimates immediately upon the arrival of new data or choose to wait for a few periods before conducting an update. In our notation, this would mean that the final date T of the previous data set is increased to $T + k\Delta t$ ($k \geq 1$). When $k = 1$, the update is performed immediately upon the arrival of new data. A particular strength of our methodology is that the estimation routine does not need to be re-initialized from the beginning, because the pseudo-posterior with data up to T can conveniently serve as a new starting point to run the additional k periods.

Let the pseudo-posterior distribution at T (based on the old data set) be denoted by

$$\gamma_{T/\Delta t}^{(T)}(\theta) \propto \prod_{j=1}^{T/\Delta t-1} L_{j,\min(T-j\Delta t,\tau)}^{(T)}(\theta)\pi(\theta)$$

whereas the pseudo-posterior distribution at $T + k\Delta t$ (based on the new data set) by

$$\gamma_{T/\Delta t+k}^{(T+k\Delta t)}(\theta) \propto \prod_{j=1}^{T/\Delta t+k-1} L_{j,\min(T+k\Delta t-j\Delta t,\tau)}^{(T+k\Delta t)}(\theta)\pi(\theta)$$

The superscript is introduced to differentiate the data sets available at time T and $T+k\Delta t$, respectively. It is important to note that $L_{j,l}^{(T+k\Delta t)}(\theta) \neq L_{j,l}^{(T)}(\theta)$ can be caused by revisions to the old data set. More importantly, there is a generic difference between the pseudo-posterior distribution up to T under the new data set and the corresponding quantity under the old data set specifically due to multiperiod predictions; that is, $\gamma_{T/\Delta t}^{(T+k\Delta t)}(\theta) \neq \gamma_{T/\Delta t}^{(T)}(\theta)$ even without any data revisions to the period covered by the old data set. To put it concretely, using the new data set and at, say, one period before the last (i.e., time $T - \Delta t$), one can make default predictions up to $(k + 1)$ periods, whereas at the same time point, it was only possible to make one-period predictions under the old data set because there were no data beyond time T. Adjustments to the weights are thus necessary to reflect the change in data set before making any sequential updates.

There are several possible ways of advancing the system. Here we propose to decompose the move into two steps. First, we take care of data revisions up to time T and then act as if we were making predictions with data only up to time T. Doing it this way is meant to maintain the same default prediction setting; that is, for example, only makes one-period default prediction at time $T - \Delta t$ even though the new data set permits predictions up to $k + 1$ periods. Thus, we introduce

$$\gamma_{T/\Delta t}^{(T+k\Delta t,T)}(\theta) \propto \prod_{j=1}^{T/\Delta t-1} L_{j,\min(T-j\Delta t,\tau)}^{(T+k\Delta t)}(\theta)\pi(\theta)$$

to denote this pseudo-posterior where the superscript $(T + k\Delta t, T)$ stands for the updated data set available at time $T + k\Delta t$ but making default predictions as if the data were only available up to time T.

From the previous run up to T, one already has a weighted set of particles $(\theta^{(i,T/\Delta t)}, w^{(i,T/\Delta t)})$ representing the pseudo-posterior distribution, $\gamma^{(T)}_{T/\Delta t}(\theta)$. Next, update the weights by

$$w^{*(i,T/\Delta t)} = w^{(i,T/\Delta t)} \times \frac{\gamma^{(T+k\Delta t,T)}_{T/\Delta t}(\theta^{(i,T/\Delta t)})}{\gamma^{(T)}_{T/\Delta t}(\theta^{(i,T/\Delta t)})} \tag{29}$$

Since the denominator is available from the previous run, one only needs to compute the numerator using the new data set up to time T. Then, the weighted set $(\theta^{(i,T/\Delta t)}, w^{*(i,T/\Delta t)})$ represents the revised pseudo-posterior distribution at time T, i.e., $\gamma^{(T+k\Delta t,T)}_{T/\Delta t}(\theta)$, specifically to account for data revisions. From this point onward, one can apply the same recursive procedure, starting from equation (23), to advance from time T to $T + \Delta t$, then to $T + 2\Delta t$ and so on until reaching $T + k\Delta t$ to complete the updating task.

Reweighting may substantially alter the ESS of the particle set due to a large volume of data changes. If the reweighting leads to a satisfactory ESS, i.e., $ESS \geq B$, advancing to $T + \Delta t$ continues as usual. Otherwise, the weighted sample will be discarded to prevent the support from degeneration. One can return to the particle set before reweighting and perform resampling to create an equally-weighted particle set. Then, make the Metropolis-Hastings moves by targeting $\gamma^{(T+k\Delta t,T)}_{T/\Delta t}(\theta)$ using the Gaussian-type sampler described earlier and starting with the mean and variance implied by the resampled particle set. One should make these Metropolis-Hastings moves until the particle set reaches a desirable level of distinctiveness, and perhaps with a preset minimum number of moves to ensure that the resulting particle set is close enough to the target distribution.

Furthermore, one can update all self-normalized statistics in the way as described earlier to reflect the additions of k more pseudo-posterior means to the sequence.

References

[1] Altman, E., 1968, Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy, *Journal of Finance* 23, 589–609.

[2] Beaver, W., 1966, Financial Ratios as Predictors of Failure, *Journal of Accounting Research* 4, 71–111.

[3] Bharath, S. T., and T. Shumway, 2008, Forecasting Default with the Merton Distance to Default Model, *Review of Financial Studies* 21, 1339–1369.

[4] Campbell, J. Y., J. Hilscher, and J. Szilagyi, 2008, In Search of Distress Risk, *Journal of Finance* 63, 2899–2939.

[5] Chava, Sudheer, and Robert A. Jarrow, 2004, Bankruptcy Prediction with Industry Effects, *Review of Finance* 8, 537–569.

[6] Chernozhukov, V. and Hong, H., 2003, An MCMC Approach to Classical Estimation, *Journal of Econometrics* 115, 293–346.

[7] Chopin, N., 2002, A Sequential Particle Filter Method for Static Models, *Biometrika* 89, 539–551.

[8] Crosbie, P., and J. Bohn, 2002, Modeling Default Risk, technical report, KMV LLC.

[9] Del Moral P, Doucet A, Jasra A (2006) Sequential Monte Carlo Samplers, *Journal of Royal Statistical Society B* 68, 411–436.

[10] Del Moral P, Doucet A, Jasra A (2011) An Adaptive Sequential Monte Carlo Method for Approximate Bayesian Computation. Statistics and Computing, DOI: 10.1007/s11222-011-9271-y.

[11] Doucet A, and Shephard N (2012) Robust Inference on Parameters via Particle Filters and Sandwich Covariance Matrices, working paper, Oxford-Man Institute, University of Oxford.

[12] Duan, J. C., 2010, Clustered Defaults, National University of Singapore working paper.

[13] Duan, J.C., Sun, J. and T. Wang, 2012, Multiperiod Corporate Default Prediction - A Forward Intensity Approach, *Journal of Econometrics* 170, 191–209.

[14] Duffie, D., A. Eckner, G. Horel and L. Saita, 2009, Frailty Correlated Default, *Journal of Finance* 64, 2089–2123.

[15] Duffie, D., L. Saita, and K. Wang, 2007, Multi-period Corporate Default Prediction with Stochastic Covariates, *Journal of Financial Economics* 83, 635–665.

[16] Kiefer, N. M., Vogelsang, T. J. and Bunzel, H., 2000, Simple Robust Testing of Regression Hypotheses. *Econometrica* 68, 695-714

[17] Koopman, S.J., A. Lucas, and B. Shwaab, 2009, Macro, Industry and Frailty Effects in Defaults: The 2008 Credit Crisis in Perspective, forthcoming in *Journal of Business and Economic Statistics* (2013).

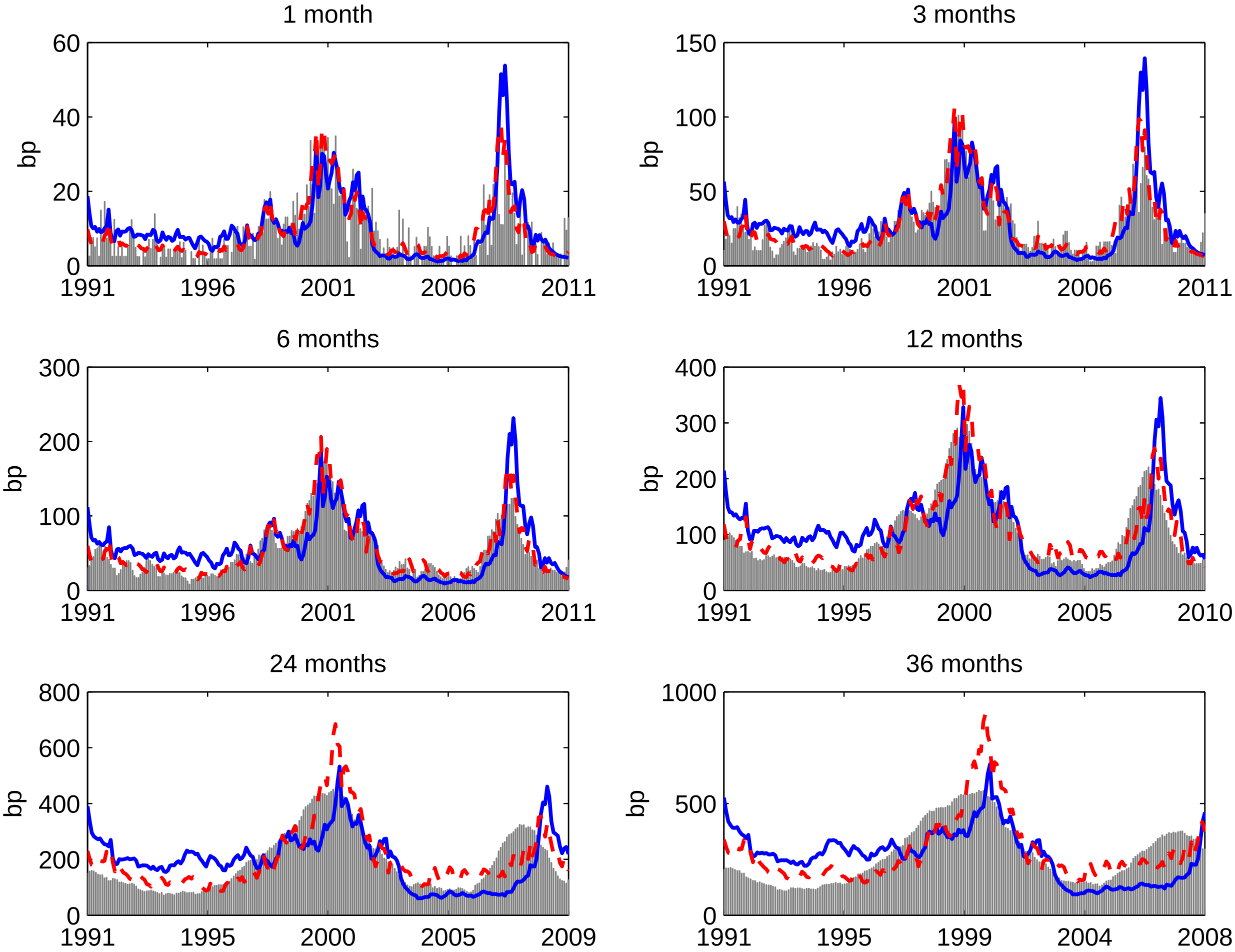
[18] Malik, S., and M. K. Pitt, 2011, Particle Filters for Continuous Likelihood Evaluation and Maximisation, *Journal of Econometrics* 165, 190–209.

[19] Shao, X., 2010, A Self-Normalized Approach to Confidence Interval Construction in Time Series. *Journal of the Royal Statistical Society: Series B* 72, 343–366.

[20] Shumway, T., 2001, Forecasting Bankruptcy More Accurately: A Simple Hazard Model, *Journal of Business* 74, 101–124.

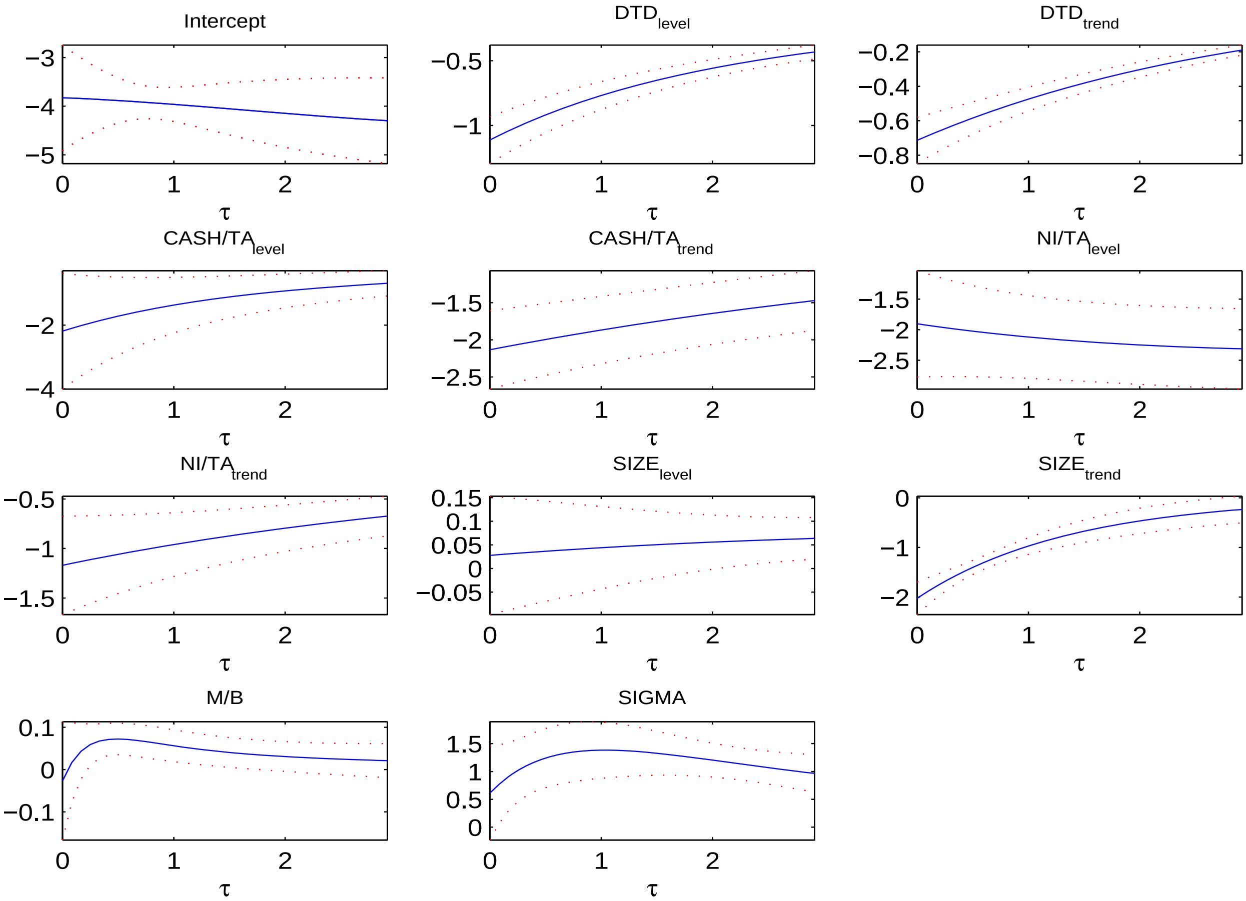
[21] Vuong, Q.H, 1989, Likelihood Ratio Tests for Model Selection and Non-nested Hypotheses, *Econometrica* 57, 307–333.

Figure 1: Aggregate default rate predictions of the DSW and PC-F models



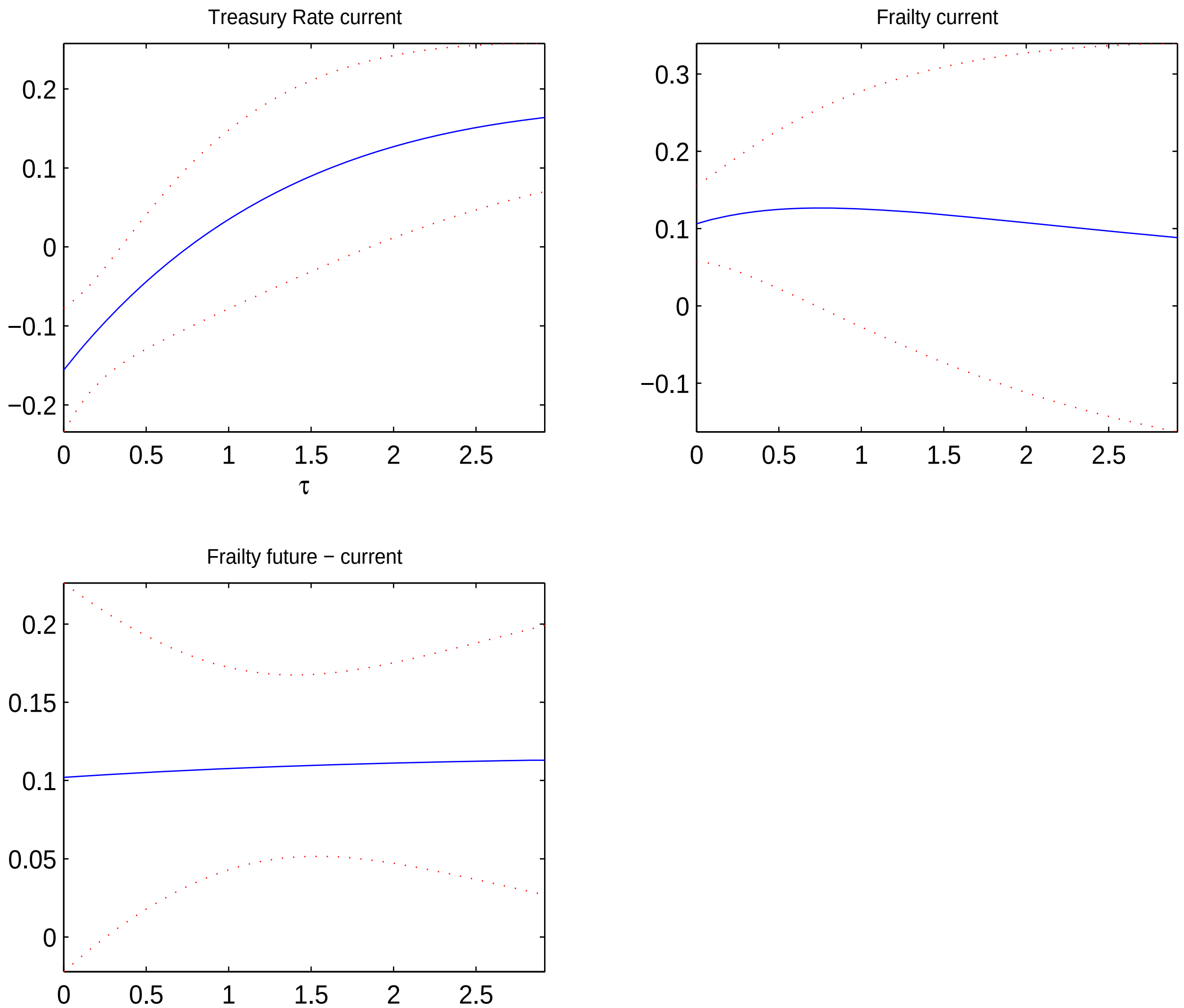
This figure presents the predicted default rates for two models: DSW (solid blue curve) and PC-F (dashed red curve). Realized default rates are the gray bars.

Figure 2: Parameter estimates for the firm-specific attributes in the forward default intensity function of the PC-F model



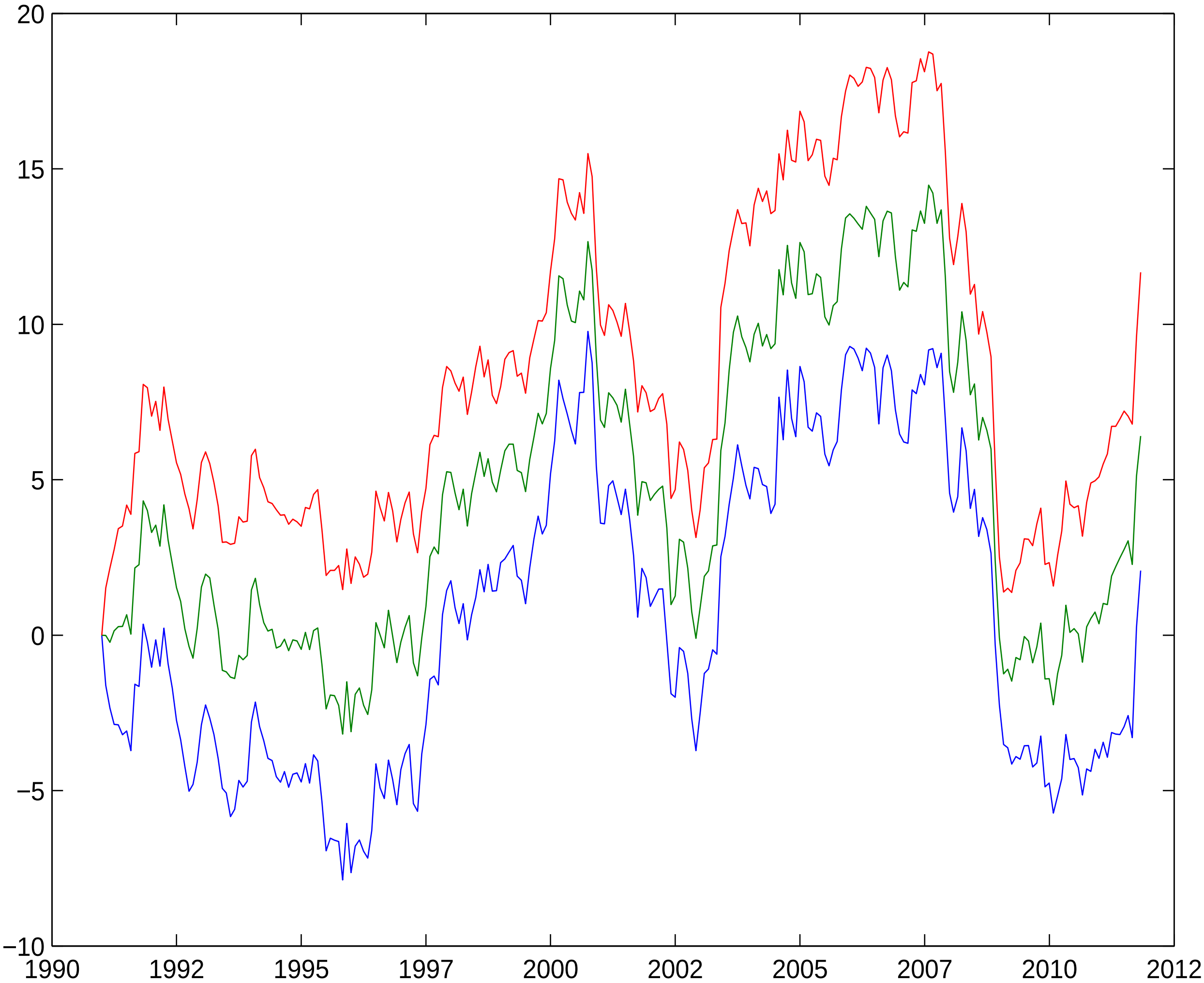
This figure shows the parameter estimates for the forward default intensity function corresponding to different prediction horizons under the PC-F model. DTD is the distance to default, CASH/TA is the sum of cash and short-term investments over the total assets, NI/TA is the net income over the total assets, SIZE is log of firm's market equity over the average market equity value of the S&P500 company, M/B is the market to book equity value ratio, SIGMA is the 1-year idiosyncratic volatility. The subscript "level" denotes the average in the preceding 12 months, "trend" denotes the difference between its current value and the preceding 12-month average. The solid blue curve is for the parameter estimates and the dotted red curves depict the 90% confidence interval.

Figure 3: Parameter estimates for the common risk factors in the forward default intensity function of the PC-F model



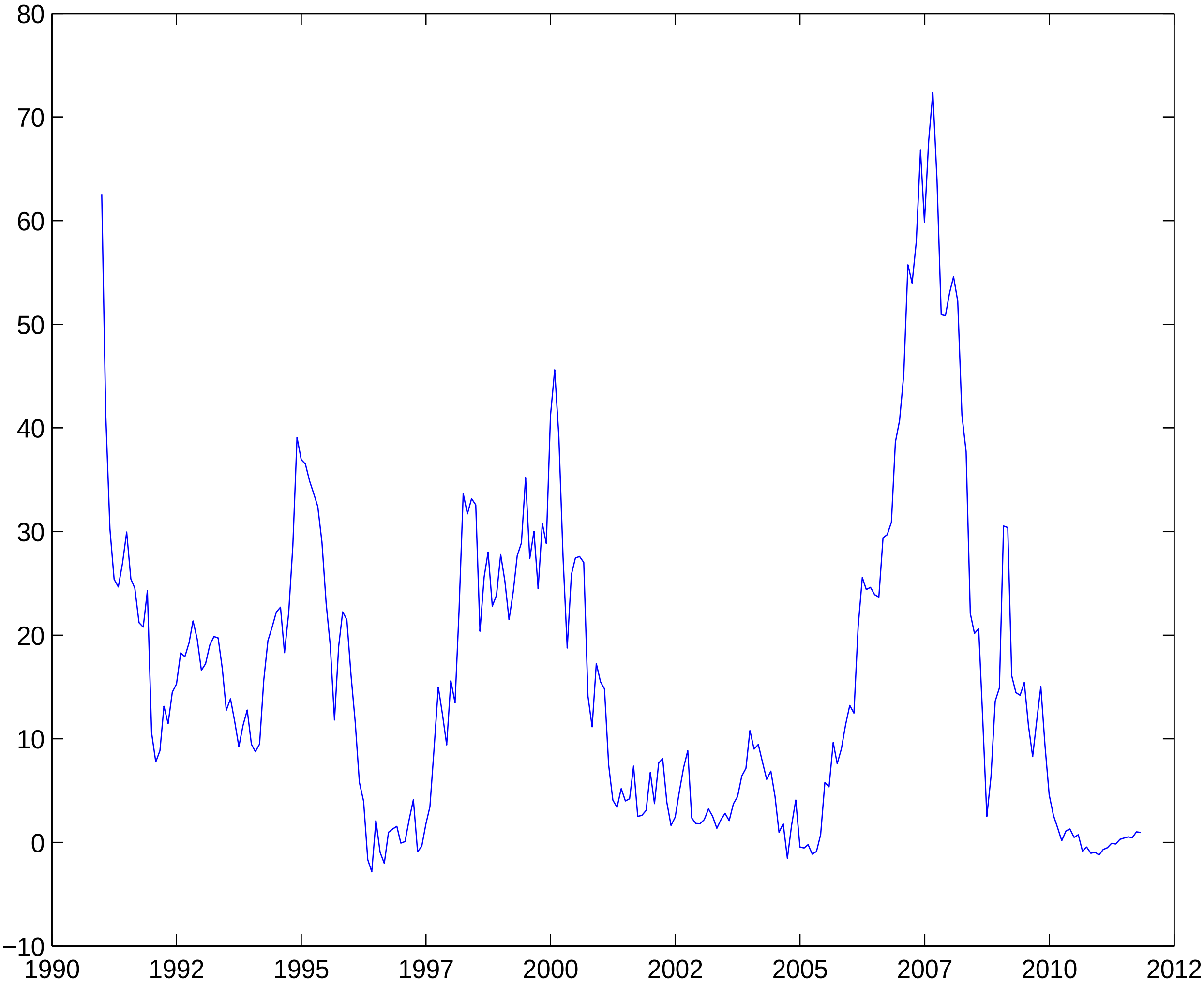
This figure shows the parameter estimates for the forward default intensity function corresponding to different prediction horizons under the PC-F model. "Treasury Rate" is the 3-month US Treasury rate, "Frailty" is the common latent factor. The solid blue curve is for the parameter estimates and the dotted red curves depict the 90% confidence interval.

Figure 4: Estimates of the frailty factor time series under the PC-F model



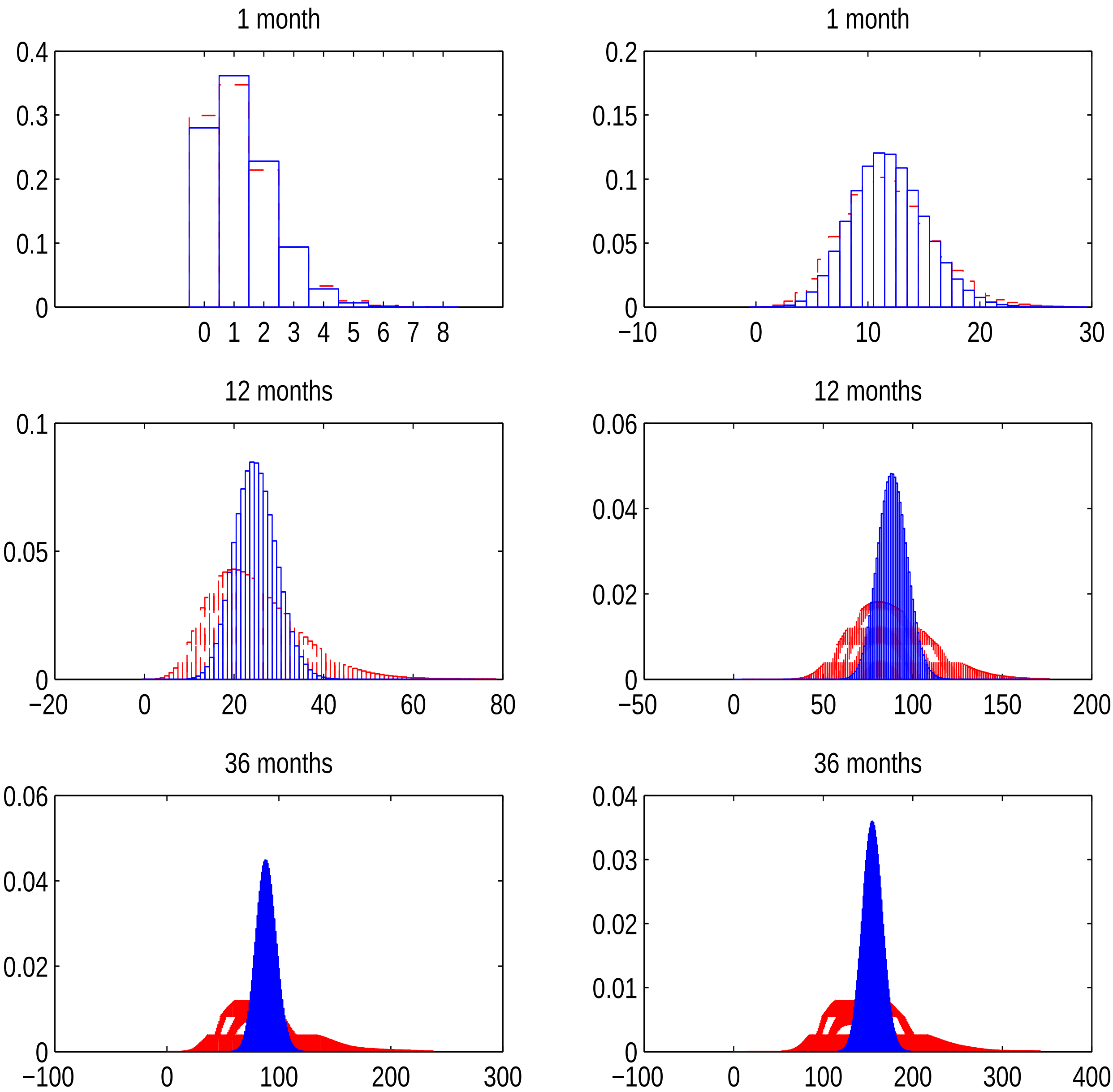
This figure shows the 5, 50 and 95 % quantiles of the filtered frailty factor under the PC-F model. The frailty factor is obtained by applying the full-sample parameter estimates to the particle filter.

Figure 5: Log-pseudo-likelihood differences between the PC-F and DSW models



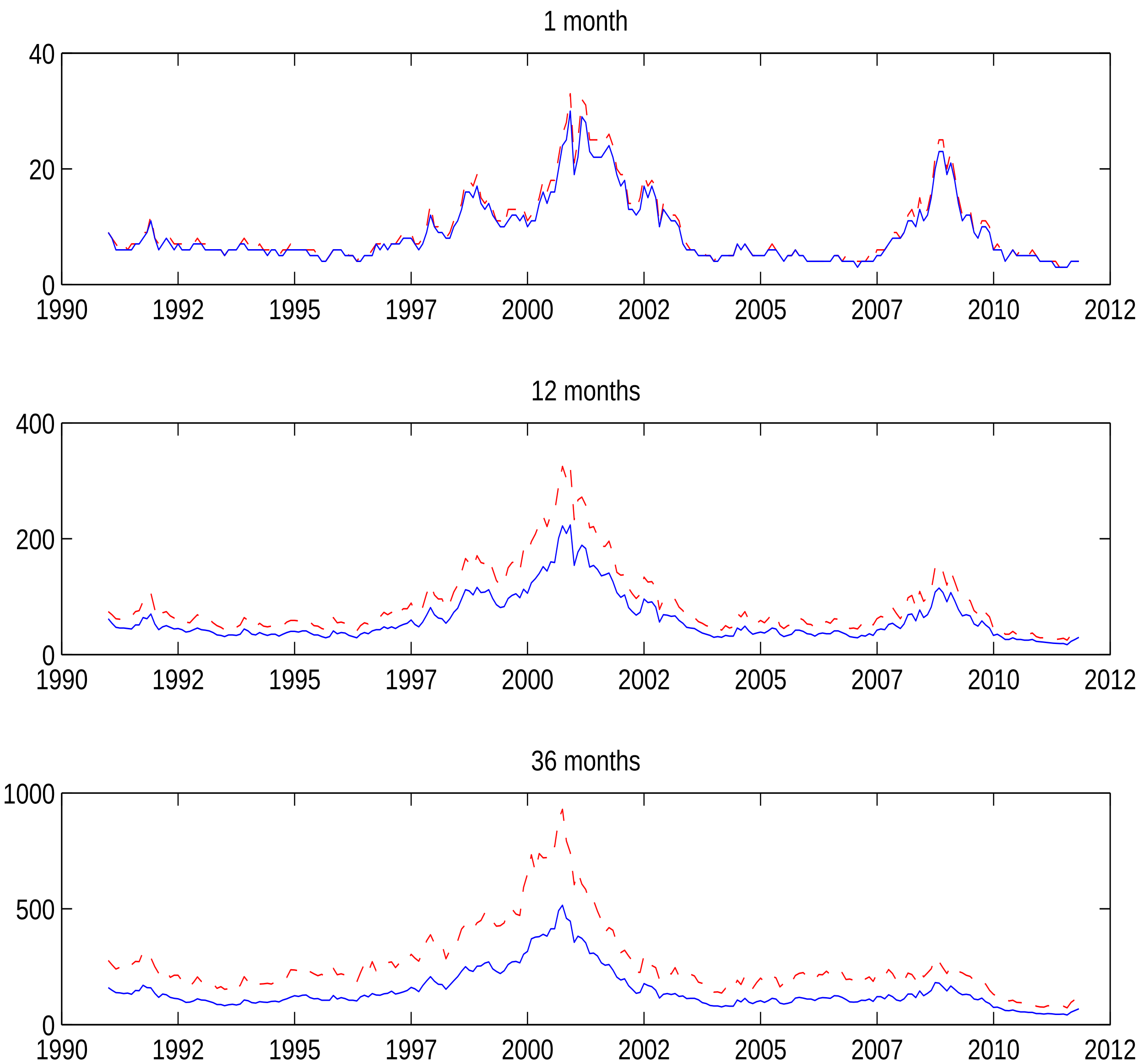
This figure shows the difference in the log-pseudo-likelihoods between the PC-F and DSW models.

Figure 6: Portfolio default distributions implied by the PC-F model with and without default correlations under two market conditions



This figure shows the distribution for the number of defaults in the full population implied by the two models at two different dates. The dotted histogram depicts the distribution implied by the PC-F model, whereas the solid histogram is the distribution when default correlations are ignored. The left column shows the predicted distributions in May 2007, before the onset of financial crisis, whereas the right column is for the distributions in October 2008, after Lehman Brothers' bankruptcy. The top panel shows the results for the prediction horizon of 1 month, the middle 12 months and the bottom 36 months.

Figure 7: Portfolio default distribution's 99^{th} percentile implied by the PC-F model with and without default correlations



This figure shows the 99^{th} percentile of the distribution for the number of defaults in the full population implied by two models. The dotted curve depicts the distribution implied by the PC-F model, whereas the solid curve depicts the distribution when default correlations are ignored. The upper panel shows the results at the prediction horizon of 1 month, the middle 12 months and the lower 36 months.

Table 1: Accuracy ratios

	1 month	3 months	6 months	12 months	24 months	36 months
Panel A: In-sample results for the whole sample						
DSW	93.02%	91.13%	88.49%	83.4%	73.92%	66.49%
DSW-F	93.66%	91.54%	88.84%	84.13%	75.31%	67.6%
PC-F	93.5%	91.49%	88.91%	84.29%	75.51%	68.05%
PC-M	93.48%	91.47%	88.89%	84.27%	75.45%	67.82%
Panel B: In-sample results for the non-financial subsample						
DSW	93.08%	91.1%	88.26%	82.95%	73.87%	66.76%
DSW-F	93.7%	91.53%	88.72%	83.91%	75.42%	67.78%
PC-F	93.57%	91.51%	88.8%	84.03%	75.6%	68.29%
PC-M	93.58%	91.52%	88.81%	84.04%	75.59%	68.1%
Panel C: In-sample results for the financial subsample						
DSW	92.49%	91.18%	90.29%	86.87%	73.51%	60.13%
DSW-F	93.53%	91.85%	90.34%	87.17%	76.96%	67.05%
PC-F	93.09%	91.49%	90.26%	87.36%	77.09%	66.87%
PC-M	93.08%	91.47%	90.19%	87.26%	77.05%	67.14%
Panel D: Out-of-sample (over time) results for the whole sample						
DSW	92.85%	91.31%	88.95%	85.02%	77.15%	72.26%
DSW-F	93.54%	91.93%	89.64%	85.92%	77.46%	70.27%
PC-F	93.37%	91.87%	89.7%	86.2%	78.63%	72.04%
PC-M	93.46%	91.95%	89.75%	86.19%	78.52%	71.61%

This table reports the accuracy ratios for four model specifications. Panel A reports the in-sample results for all firms and the entire sample period (1991-2011). Panels B and C report the in-sample performance on the non-financial and financial subsamples, respectively, with the same parameter values as the full sample. Panel D reports the out-of-sample (over time) results for the sample period (2001-2011). Here the models are estimated with an expanding window.

Table 2: Maximum pseudo-likelihood estimates for the PC-F model

Panel A: Estimates for the frailty dynamics

	c
	0.98
	[0.73 1.22]

Panel B: Estimates for the forward default intensity function

	ϱ_0	ϱ_1	ϱ_2	d
Intercept	-5.044	1.214	1.09	1.137
	[-5.998 -4.089]	[-0.6158 3.045]	[-4.26 6.441]	[-1.176 3.45]
DTD_{level}	0	-1.112	0.1376	1.427
		[-1.293 -0.9311]	[-1.039 1.314]	[-0.3189 3.172]
DTD_{trend}	0	-0.7143	1.109	3.234
		[-0.8482 -0.5803]	[-0.01907 2.236]	[-0.5347 7.002]
CASH/TA_{level}	0	-2.192	0.4337	1.182
		[-4.005 -0.3788]	[-0.3968 1.264]	[0.4244 1.94]
CASH/TA_{trend}	0	-2.134	0.2753	4.142
		[-2.661 -1.608]	[-0.4802 1.031]	[2.521 5.764]
NI/TA_{level}	0	-1.902	-4.27	4.441
		[-2.772 -1.033]	[-7.537 -1.004]	[1.344 7.538]
NI/TA_{trend}	0	-1.169	0.6252	3.832
		[-1.665 -0.6725]	[0.07699 1.173]	[2.286 5.379]
SIZE_{level}	0	0.02789	0.2143	5.028
		[-0.09755 0.1533]	[-0.2081 0.6366]	[1.71 8.345]
SIZE_{trend}	0	-2.021	2.044	1.375
		[-2.349 -1.693]	[1.091 2.997]	[0.5222 2.227]
M/B	0	-0.02654	0.2829	0.2416
		[-0.1666 0.1136]	[-0.005938 0.5717]	[-1.672 2.155]
SIGMA	0	0.612	3.615	0.7228
		[-0.231 1.455]	[0.495 6.736]	[-3.41 4.856]
Treasury Rate (current)	0	-0.1559	0.8175	1.853
		[-0.2342 -0.0776]	[0.5088 1.126]	[0.4004 3.305]
Frailty (current)	0	0.1062	0.2233	0.8891
		[0.05696 0.1554]	[-0.5828 1.029]	[-1.252 3.03]
Frailty (future - current)	0.01204	0.09005	0.1664	5.674
	[-0.775 0.7991]	[-0.6697 0.8498]	[-1.047 1.38]	[2.824 8.525]

Treasury Rate is the 3-month US Treasury rate, DTD is the distance-to-default, CASH/TA is cash and short-term investments over the total assets, NI/TA is the net income over the total assets, SIZE is log of firm's market equity value over the average market equity value of an S&P500 company, M/B is the market to book ratio, SIGMA is the 1-year idiosyncratic volatility. The subscript "level" denotes the average in the previous 12 months, and "trend" is the difference between current value and its previous 12-month average. The 90% confidence bands are reported in brackets.

Table 3: Estimates for the forward other-exits intensity function

Intercept	-2.169	-2.49	-4.002	0.2703
	[-2.425 -1.913]	[-2.946 -2.033]	[-5.694 -2.311]	[-0.04147 0.5821]
DTD_{level}	0	0.096	-0.3949	1.679
		[0.0162 0.1758]	[-0.5711 -0.2186]	[1.238 2.12]
DTD_{trend}	0	0.173	-0.128	0.2861
		[0.08518 0.2609]	[-0.2968 0.04079]	[-1.945 2.517]
$CASH/TA_{level}$	0	-0.3956	1.313	2.29
		[-1.311 0.5199]	[0.3591 2.267]	[-0.00571 4.586]
$CASH/TA_{trend}$	0	-0.5451	0.5208	4.766
		[-1.09 -0.0001954]	[-1.291 2.333]	[3.144 6.387]
NI/TA_{level}	0	-3.209	0.7052	2.95
		[-4.067 -2.352]	[-1.005 2.415]	[1.002 4.899]
NI/TA_{trend}	0	-1.956	0.08383	0.4685
		[-2.958 -0.9546]	[-1.578 1.745]	[0.07104 0.8659]
$SIZE_{level}$	0	-0.2641	-0.4393	0.3757
		[-0.3401 -0.1881]	[-0.5175 -0.3611]	[0.253 0.4985]
$SIZE_{trend}$	0	-0.5617	1.339	7.218
		[-0.8221 -0.3012]	[0.4857 2.192]	[5.482 8.954]
M/B	0	-0.06624	0.1394	6.391
		[-0.08441 -0.04808]	[-0.0262 0.305]	[5.459 7.323]
SIGMA	0	2.761	-1.684	0.5247
		[1.902 3.62]	[-4.507 1.138]	[-0.9952 2.045]
Treasury Rate (current)	0	0.06362	0.3222	0.2044
		[-0.1514 0.2786]	[-0.3271 0.9715]	[-4.565 4.974]

Treasury Rate is the 3-month US Treasury rate, DTD is the distance-to-default, CASH/TA is cash and short-term investments over the total assets, NI/TA is the net income over the total assets, SIZE is log of firm's market equity value over the average market equity value of an S&P500 company, M/B is the market to book ratio, SIGMA is the 1-year idiosyncratic volatility. The subscript "level" denotes the average in the previous 12 months, and "trend" is the difference between current value and its previous 12-month average. The 90% confidence bands are reported in brackets.