Multiperiod Corporate Default Prediction with the Partially-Conditioned Forward Intensity

Jin-Chuan Duan[1] and Andras Fulop[2]

(April 2013)

[1] RMI and Business School, National University of Singapore
[2] ESSEC Business School

The literature

- Corporate default/bankruptcy prediction literature is vast, but most do NOT deal with the censoring effect due to exits other than defaults/bankruptcies.

- Most papers do NOT address the natural dynamic setting of corporate default predictions where common risk factors and firm-specific attributes evolve over time.

- Bottom-up aggregation is possible only when one explicitly addresses the natural dynamic setting. My discussion focuses on doubly stochastic default prediction models that are capable of predicting single-name defaults and conducting portfolio credit risk analysis.

- Spot-intensity models, e.g. Duffie, Saita and Wang (2007, *J of Financial Economics*), Duffie, Eckner, Horel and Saita (2009, *J of Finance*):
  - A common spot-intensity function of covariates (common risk factors and firm-specific attributes) for all firms and at all time.
  - Specify a full joint time series dynamics for firm-specific attributes and common risk factors. Extremely high dimension! Ambitious, but bound to be quite ad hoc.
- Forward-intensity model, Duan, Sun, Wang (2012, *J of Econometrics*) (DSW)
  - Avoid specifying the dynamics of the covariates through the use of forward intensities.
  - A set of forward-intensity functions versus one universal spot-intensity function.
  - Easier implementation, but no real default correlations

This paper

- Extend the model of DSW (2012) to allow for default correlations by conditioning forward intensities on the future values of common risk factors (observed or latent).
- Need to specify a low-dimensional dynamic for the conditioning common risk factor(s), but still avoid specifying the joint dynamics for firm-specific attributes.
- Apply a pseudo-Bayesian sequential Monte Carlo technique to estimate the parameters (two sets: 44 and 40).
- Use recursive estimates for inference by self-normalized asymptotics.
- Provide empirical analysis of the model on a large US data sample of over 12,000 firms over 21 years.

The forward-intensity model of DSW (2012)

- Assume independence across firms' defaults and/or other exits conditional on the observed covariates:

  Forward default intensity

  $$f_{it}(\tau) = \exp\left(\alpha_0(\tau) + \alpha_1(\tau)x_{it,1} + \alpha_2(\tau)x_{it,2} + \cdots \alpha_k(\tau)x_{it,k}\right)$$

  Forward combined intensity

  $$\begin{aligned} g_{it}(\tau) &= \exp\left(\beta_0(\tau) + \beta_1(\tau)x_{it,1} + \beta_2(\tau)x_{it,2} + \cdots \beta_k(\tau)x_{it,k}\right) \\ &\quad + f_{it}(\tau) \end{aligned}$$

- Can derive a consistent PD curve for each company, but it is in essence a single-obligor PD model.
- The DSW model has been implemented on over 60,000 exchange-listed firms in 106 economies by the NUS-RMI Credit Research Initiative (non-profit credit ratings) to produce daily update PDs for these firms.

Our modification of the DSW (2012) model

- Keep the concept of conditional independence, but introduce common factors Z_t into the conditioning set.
- Conditional on future values of the common risk factors, defaults across firms are assumed to be independent:

$$\begin{aligned}
& \psi_{it}(\tau; \mathbf{Z}_u, u \leq t+\tau) \\
\equiv\; & -\frac{\ln E_t\left[\exp\left(-\int_t^{t+\tau}(\lambda_{is}+\phi_{is})ds\right)\middle| \mathbf{Z}_u, u \leq t+\tau\right]}{\tau} \\
& g_{it}(\tau; \mathbf{Z}_u, u \leq t+\tau) \\
\equiv\; & \psi_{it}(\tau; \mathbf{Z}_u, u \leq t+\tau) + \psi'_{it}(\tau; \mathbf{Z}_u, u \leq t+\tau)\tau \\
& f_{it}(\tau; \mathbf{Z}_u, u \leq t+\tau) \equiv e^{\psi_{it}(\tau; \mathbf{Z}_u, u \leq t+\tau)\tau} \times \\
& \lim_{\Delta t \to 0} \frac{E_t\left[\int_{t+\tau}^{t+\tau+\Delta t} \exp\left(-\int_t^s(\lambda_{iu}+\phi_{iu})du\right)\lambda_{is}ds \middle| \mathbf{Z}_u, u \leq t+\tau\right]}{\Delta t}.
\end{aligned}$$

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

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

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

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

Choose a particular type of partially-conditioned forward intensity functions.

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

Note: Default correlations come through future values of the common risk factors.

Discretized model

- The sample period $[0, T]$ is assumed to be divisible into $T/\Delta t$ periods. Let N be the total number of companies.
- 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}.
- Let $\mathbf{Z}_{t:t+j} = \{\mathbf{Z}_t, \mathbf{Z}_{t+\Delta t}, \cdots, \mathbf{Z}_{t+j\Delta t}\}$.

Partially-conditioned forward probabilities

$$
\begin{aligned}
& P_t(\tau_{Ci} > t + (j+1)\Delta t | \mathbf{Z}_{t:t+j}, \tau_{Ci} > t + j\Delta t) \\
= \;& e^{-g_{it}(j\Delta t; \mathbf{Z}_{t+j\Delta t})\Delta t}
\end{aligned}
$$

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

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

Computing default probabilities

- For any joint probability, one first comes up with the joint probability conditional on future values of the common risk factors by utilizing conditional independence, and then integrates over the common risk factors.

- For example,

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

Estimating the parameters

- Denote the unknown fixed model parameters by θ.
- The likelihood corresponding to time period $j\Delta t$ for the prediction interval $[j\Delta t, j\Delta t + \tau]$ can be written as

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

- Here $P_{i,j,\tau}(\theta; \mathbf{Z}_{j\Delta t:j\Delta t+\tau})$ is the individual likelihood for a given firm, conditional on the common risk factor.
- The expectation cannot be solved in a closed form, but can be computed by simulation.
- $\mathbf{Z}_t = (\mathbf{Y}_t, \mathbf{F}_t)$ where $\mathbf{F}_t$ is latent and its filtered value can be obtained by a particle filter.

Maximum pseudo likelihood

- The maximum sample pseudo likelihood is

$$\mathcal{L}_{\tau}(\theta; \tau_C, \tau_D, \mathbf{X}, \mathbf{Y}) = \prod_{j=0}^{T/\Delta t-1} L_{j,\min(T-j\Delta t,\tau)}(\theta)$$

- Due to the use of conditioning common risk factors, the decomposability property of DSW (2012) no longer applies. Parameters for different forward starting times must be estimated jointly.

A pseudo-Bayesian device for parameter estimation

- Lack of decomposability means that the number of parameters to be estimated jointly will be very large. The conventional gradient-based optimization methods does not work well.
- Consider the following pseudo-posterior distribution:

$$\gamma_t(\theta) \propto \prod_{j=0}^{t} L_{j,\min(T-j,\tau)}(\theta)\pi(\theta), \text{ for } t = 1, \cdots, T/\Delta t - 1 \tag{1}$$

- Apply the sequential batch-resampling routine of Chopin (2002).
- The only role of the prior, $\pi(\theta)$, is to provide the initial particle cloud from which the algorithm can start.

Details of the batch-resampling procedure

- For each t, this procedure yields a weighted sample of M points, $(\theta^{(i,t)}, w^{(i,t)}), i = 1, \ldots, M$.
- The weights can be easily updated by the computing only the pseudo-likelihood associated with additional data points when the particle set to represent the parameters remains unchanged.
- Without periodical resamplings (sometimes a rejuvenation sampling), parameter impoverishment or degeneracy will certainly occur. Thus, resampling is a must after advancing the system over a block of data over some period of time.
- Additional tempering steps are needed between $\gamma_t(\theta)$, $\gamma_{t+1}(\theta)$ for a smooth transition.
- The empirical distribution function will converge to $\gamma_t(\theta)$ as M increases.

Parameter estimate

- Denote the pseudo-posterior mean of the parameter estimates by $\hat{\theta}_t$:

$$\hat{\theta}_t = \frac{1}{\sum_{i=1}^{M} w^{(i,t)}} \sum_{i=1}^{M} w^{(i,t)} \theta^{(i,t)}$$

- $\gamma_t(\theta)$ is not a true posterior because the likelihood function in equation (1) is not a true likelihood function. Thus, it cannot directly provide valid inference.

How to conduct valid inference?

- We can turn to the results of Chernozhukov and Hong (2003) to give a classical interpretation to the simulation output.
- Let the pseudo-score matrix be $s_T(\theta) = \nabla_\theta l_T(\theta)$ and the negative of the scaled Hessian be $J_T(\theta) = -\nabla_{\theta\theta'} l_T(\theta)/(T/\Delta t)$.
- Theorem 1 of Chernozhukov and Hong (2003) states that for large T, $\gamma_{T/\Delta t-1}(\theta)$ is approximately a normal density with the random mean parameter: $\theta_0 + J_T(\theta_0)^{-1} s_T(\theta_0)/(T/\Delta t)$. This implies that
  - Pseudo posterior mean $\hat{\theta}_t$ provides a consistent estimate for θ_0
  - Scaled estimation error can be characterized by

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

Using the recursive estimates for inference I

- Access to the recursive estimates $\hat{\theta}_t$ provides us an easy way to construct confidence set using the self-normalized approach of Shao (2010). Hence, no need to estimate asymptotic variance.
- Assume that the functional CLT applies to the scaled pseudo-score:

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

- Define a norming matrix

$$\hat{C}_T = \frac{1}{(T/\Delta t)^2}\sum_{l=0}^{T/\Delta t-1} l^2(\hat{\theta}_l - \hat{\theta}_{T/\Delta t-1})(\hat{\theta}_l - \hat{\theta}_{T/\Delta t-1})'$$

Using the recursive estimates for inference II

- Then we can form the following asymptotically pivotal statistic

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

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. (Note that the nuisance scale matrix S disappears from this quadratic form!)

Using the recursive estimates for inference III

- 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:

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

where $\hat{\delta}_{i,T}$ is the i^{th} diagonal element of $\hat{C}_T$.

Theoretical recursion for filtering frailty

- 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 simulate 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 following theoretical recursion:

$$
\begin{aligned}
& f(F_{j\Delta t}|\mathcal{D}_{j\Delta t}) \\
\propto \quad & f(F_{j\Delta t}|F_{(j-1)\Delta t}) \left(\prod_{i=1}^{N} P_{i,j,0}(Y_{(j-1)\Delta t}, F_{(j-1)\Delta t}) \right) \\
& \times f(F_{(j-1)\Delta t}|\mathcal{D}_{(j-1)\Delta t})
\end{aligned}
$$

Sequential importance sampling I

- Assume that we have M particles representing $f(F_{(j-1)\Delta t}|\mathcal{D}_{(j-1)\Delta t})$, denoted by $F^{(m)}_{(j-1)\Delta t}$.
- 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 of Malik and Pitt (2011)
  3. Sample from the transition density

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

Sequential importance sampling II

- The resulting particle cloud, $F_{(j)\Delta t}^{(m)}$ is approximately distributed according to $f(F_{j\Delta t}|\mathcal{F}_{j\Delta t})$.
- To obtain the expectations over the path of the common variables $\tilde{Z}_{j\Delta t:j\Delta t+\tau}$, we simulate M paths.
- To decrease Monte Carlo noise, we use the same random numbers across calls at different parameter sets.

Real-time updating

- In practice the model need to be updated periodically as new data arrive and/or some old data get revised, i.e., T is increased to $T + \Delta t$.
- Assume that from the previous run up to T we have a weighted set of particles $(\theta^{(i,T/\Delta t-1)}, w^{(i,T/\Delta t-1)})$ representing the pseudo-posterior $\gamma^{(T)}_{T/\Delta t-1}(\theta)$.
- Next, set $\theta^{(i,T/\Delta t)} = \theta^{(i,T/\Delta t-1)}$ and reweight by

$$w^{(i,T/\Delta t)} = w^{(i,T/\Delta t-1)} \times \frac{\gamma^{(T+\Delta t)}_{T/\Delta t}(\theta^{(i,T/\Delta t)})}{\gamma^{(T)}_{T/\Delta t-1}(\theta^{(i,T/\Delta t)})}$$

- The weighted set $(\theta^{(i,T/\Delta t)}, w^{(i,T/\Delta t)})$ represents the new pseudo-posterior $\gamma^{(T+\Delta t)}_{T/\Delta t}(\theta)$.
- If the weights are too uneven, intermediate tempered densities can be constructed and resample-move steps can

Data of DSW (2012)

- Sample period: 1991-2011, monthly.
- Database:
  - Compustat
  - CRSP
  - Credit Research Initiative database
- 12,268 U.S. public companies (including financial firms) totaling 1,104,963 firm-month observations.

Year	Active Firms	Defaults	(%)	Other Exit	(%)
1991	4012	32	0.80%	257	6.41%
1992	4009	28	0.70%	325	8.11%
1993	4195	25	0.60%	206	4.91%
1994	4433	24	0.54%	273	6.16%
1995	5069	19	0.37%	393	7.75%
1996	5462	20	0.37%	463	8.48%
1997	5649	44	0.78%	560	9.91%
1998	5703	64	1.12%	753	13.20%
1999	5422	77	1.42%	738	13.61%
2000	5082	104	2.05%	616	12.12%
2001	4902	160	3.26%	577	11.77%
2002	4666	81	1.74%	397	8.51%
2003	4330	61	1.41%	368	8.50%
2004	4070	25	0.61%	302	7.42%
2005	3915	24	0.61%	291	7.43%
2006	3848	15	0.39%	279	7.25%
2007	3767	19	0.50%	352	9.34%
2008	3676	59	1.61%	285	7.75%
2009	3586	67	1.87%	244	6.80%
2010	3396	25	0.74%	242	7.13%
2011	3224	21	0.65%	226	7.01%

Covariates of this paper

- 3-month treasury rate
- frailty factor
- Distance to default
- Cash and short-term investments/Total assets
- Net income/Total assets
- Relative size
- Market to book ratio
- Idiosyncratic volatility

Note: see Duan and Wang (2012, *Global Credit Review*) for estimating DTDs of non-financial and financial firms.

Parameter smoothing and restrictions

- DSW (2012) demonstrated that their parameter estimates can be modeled by the Nelson-Siegel (1987) type of term structure function:

$$
\begin{aligned}
& h(\tau; \varrho_0, \varrho_1, \varrho_2, d) \\
= \quad & \varrho_0 + \varrho_1 \frac{1 - \exp(-\tau/d)}{\tau/d} + \varrho_2 \left[\frac{1 - \exp(-\tau/d)}{\tau/d} - \exp(-\tau/d) \right]
\end{aligned}
$$

  We further impose $d > 0$ for all parameter functions, and set ϱ_0 to 0 for all current values of the stochastic covariates so that when $\tau \to \infty$, their impacts will vanish. The future values of the partial conditioning variables are NOT subject to the same limiting requirement.
- When there are 1 intercept, 12 covariates and 1 partial conditioning common risk factor, the total number of default parameters becomes $12 \times 3 + 2 \times 4 = 44$.

Macro drivers and the frailty Factor

- We assume that the macro variables follow a first-order VAR

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

- Here we only use the US treasury rate. The coefficient of trailing S&P index return, used in DSW (2012), was found to be unstable so that it is not used in the analysis.

- We assume that the latent frailty F_t is one dimensional and follows an AR(1) process:

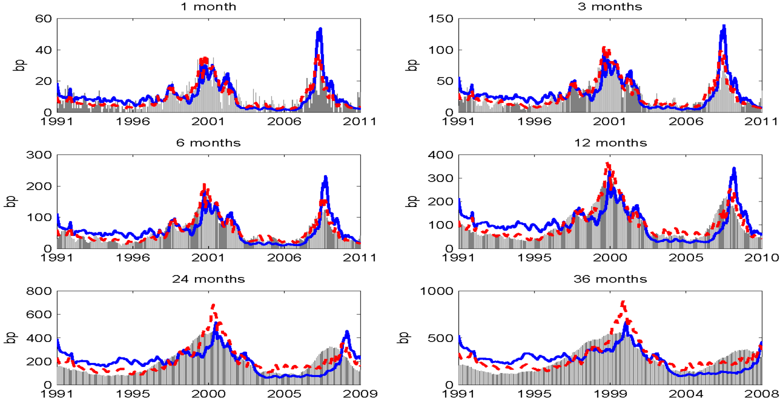
$$F_{(j+1)\Delta t} = cF_{j\Delta t} + \epsilon_{j+1}$$

Model specifications investigated

- No correlation among the forward intensities (DSW).
- The current value of the frailty factor enters into the forward intensities (DSW-F).
- Conditioning depends both on the current and future values of the frailty factor (PC-F).
- Conditioning depends on the current value of the frailty factor and the future value of the US treasury rate (PC-M).

Aggregate # of defaults

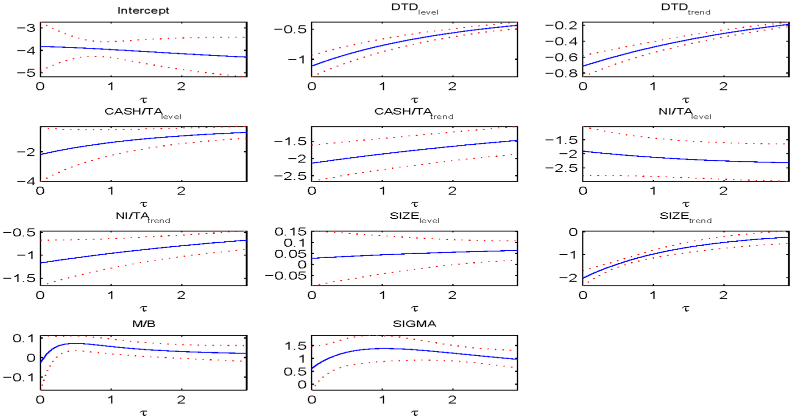
Figure 1: Aggregate default rate predictions of the DSW and PC-F models



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.

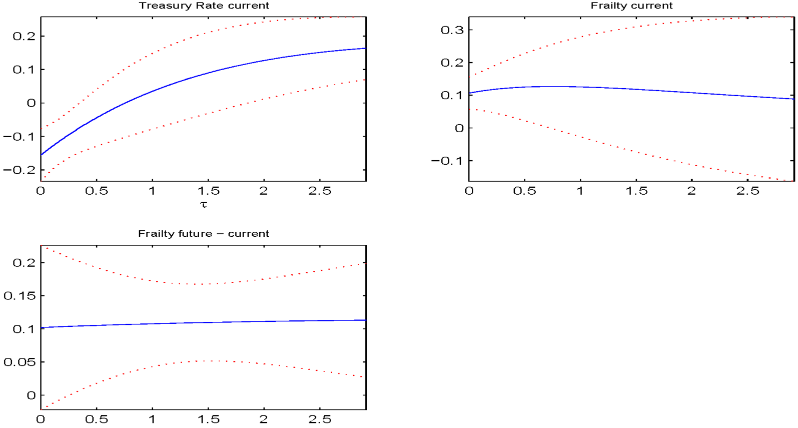
Parameter term structure (firm-specific attributes

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



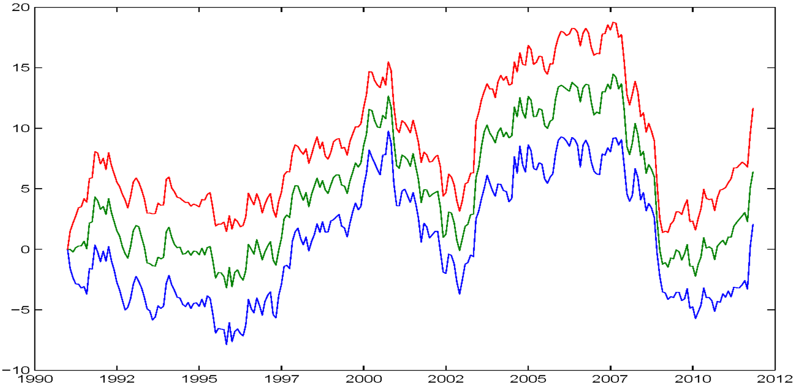
Parameter term structure (common factors

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



Frailty factor time series

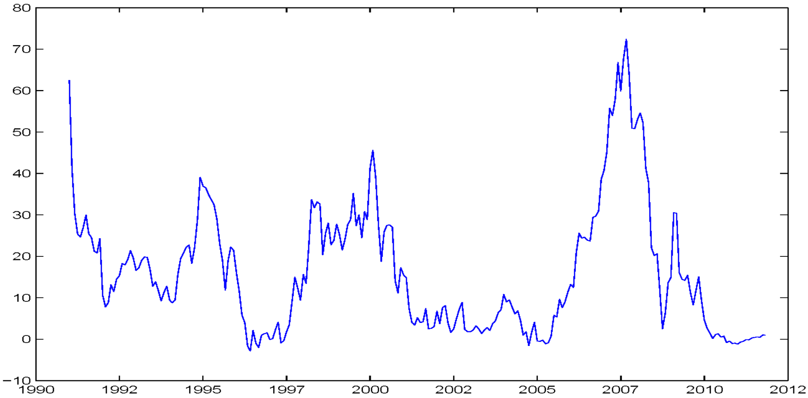
Figure 4: Estimates of the frailty factor time series under the PC-F model



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.

Log-pseudo-likelihood difference

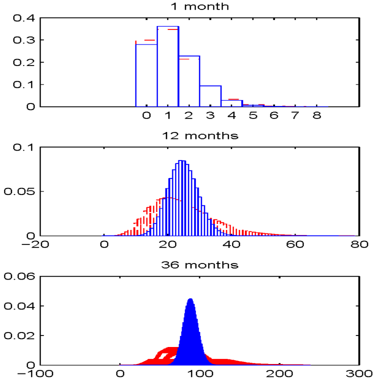
Figure 5: Log-pseudo-likelihood differences between the PC-F and DSW models



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

Portfolio default distributions at two time points

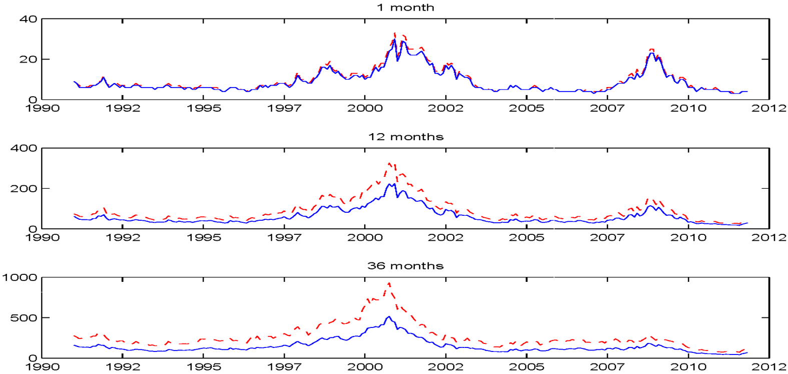
Figure 6: Portfolio default distributions implied by the PC-F model with and without default correlations under two market conditions



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.

Portfolio default distributions (99% time series)

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.

Accuracy ratios

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%

Parameter estimates (default)

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

Parameter estimates (other exits)

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]