Dynamic Credit Rating Migration with a Prior Belief Structure

Jin-Chuan Duan*

(First Draft: October 20, 2016; This Draft: January 2, 2019)

Abstract

A credit rating system classifies obligors into K cohorts by assessing their credit risks. Any member in a static pool of obligors could later migrate to one of the K cohorts plus two additional possibilities – default and other corporate exit. Ideally, the probabilities defining this $(K+2) \times (K+2)$ migration system should reflect the phase of a credit cycle, and decline monotonically away from the diagonal, reflecting the fact that small moves in credit quality occur more frequently than big changes. We propose such a stochastic rating migration model constructed with a set of credit cycle indices to reflect market conditions while imposing an intuitive prior belief structure of declining migration probabilities off the diagonal. We estimate this rating migration model with a pseudo-Bayesian sequential Monte Carlo technique, and deduce a corresponding set of forward migration generators that can then be used to produce point-in-time (PIT) probabilities of default (PDs) for any forward starting time and prediction horizon. These forward PIT-PDs are naturally dynamic over time and reflective of a specific forward period of interest. We apply this rating migration model on the global corporate credit migration data reported by the S&P over the period of 2000-2015, and show how these PIT-PDs change through different phases of a credit cycle. This rating migration model also allows for examination of the stability of the implied through-the-cycle (TTC) PDs vis-a-vis the historical measure of TTC-PDs for different rating cohorts at different points of time.

Keywords: point-in-time, through-the-cycle, credit cycle, local momentum, forward, spot, default

---

*Duan is with the National University of Singapore (Business School, Risk Management Institute and Department of Economics). E-mail: bizdjc@nus.edu.sg. The author thanks many current and previous members of the Credit Research Initiative team at the Risk Management Institute of National University of Singapore for their able assistance. An earlier version of this paper was released under the title, "Dynamic Point-in-Time Probabilities of Default via Stochastic Rating Migration". The author also acknowledges the comments received from the participants of the Workshop on Bayesian Methods in Finance organized by ESSEC and Singapore Management University in December 2018.

1 Introduction

Credit relationships exist in many forms. From a corporate perspective, liabilities of a firm are credit exposures of its counterparties. Loans, bonds, account payables, insurance policy obligations, and derivatives exposures are some examples. Assets of a firm may also be subjected to credit risk. They may consist of liabilities of other firms reflected in account receivables and other credit claims on various obligors, for example, a bank's loans to its borrowers. Credit exposures are contingent claims with losses only occurring when some obligor defaults and the recovery is not full. Risk managing credit exposures and/or properly accounting for them would not be possible without deploying credit risk models. In the banking context, the Basel capital regulations has long made the modeling of bank credit risks an undertaking of great significance. The soon-to-be implemented IFRS9/CECL[1] financial reporting standard is likely to be an even broader and more impactful formal recognition of credit risk exposures of a firm.

Credit rating migration has been the subject of many studies; for example, Altman (1998) examined and compared the rating migrations reported by Moody's and S&P over 1970-1996. Credit rating migration has also been formally modeled by many. Among them, the CreditMetrics™ of JP Morgan (1997) is perhaps the best known earlier effort and followed by many since. The stochastic drivers in CreditMetrics™ are equity values of the corporates in the pool where credit rating migrations occur when an individual obligor's standardized equity return moves across different rating thresholds that are deduced under a standard normal distribution coupled with the expected default rates for different rating classes. Under such a model, credit risk correlations are naturally deduced from equity return correlations. The obvious drawback of such an approach rests with its overreliance on equity values which are subject to differences in leverage and liquidity, among other factors, critical to default. Quality aside, many obligors in typical credit portfolios are non-corporate or simply small and medium sized firms without traded stock prices, and naturally the applicability of CreditMetrics™ is limited. Bangia, *et al* (2002) added business cycle into their model by conditioning rating migration matrix on two regimes – expansion and contraction. Feng, *et al* (2008) described credit rating migration through a factor probit model where the driving stochastic factor is latent with a time dynamic, and the filtered latent factor path could then be used to reveal the credit cycle. Other approaches include Lando and Skdeberg (2002), Gagliardini and Gourieroux (2005), Mahlmann (2006), Frydman and Schuermann (2008), Kadam and Lenk (2008), and Marcucci and Quagliariello (2009), among others.

[1] The IFRS9 (International Financial Reporting Standard 9) is the International Accounting Standards Board's proposal to account for impairments of financial assets, which became effective in 2018 and has been implemented in stages in different jurisdictions. The CECL (Current Expected Credit Loss), on the other hand, started as a joint project of the Financial Accounting Standards Board of the US with the International Accounting Standards Board, but the US accounting body later decided to go a separate way with the CECL being scheduled to take effect in 2020.

Credit exposures are often complex because a credit relationship my involve multiple payments. In order to assess expected credit loss of a single debt instrument or a portfolio of debt obligations, one must have forward default probabilities at the time of evaluation, corresponding to different future periods, for that single obligor or all obligors in the portfolio. These forward probabilities are known as point-in-time (PIT) probabilities of default (PDs) and need to be coupled with suitable recovery rates to arrive at the final expected loss. Since the evaluation time can be at any phase of a credit cycle, it is essential to have these PIT-PDs reflective of the credit cycle. Obligors may have tendency to default together, and thus a good model should incorporate default correlations. Most credit rating agency models and internal credit risk models of financial institutions address an obligor's credit quality in isolation, i.e., marginal default likelihood, and slot obligors into rating cohorts, say, 10 categories. An obligor under such a rating system may migrate from one rating category to another. Recording the credit migration experience under a rating system generates a historical series of realized credit migration matrices. These credit migration matrices offer a wealth of information, but cannot be directly used to produce suitable forward PIT-PDs needed for the purpose of assessing expected credit loss unless a suitable dynamic model is developed. PIT-PDs are in sharp contrast with the through-the-cycle (TTC) ratings typically adopted by credit rating agencies, which emphasizes credit ratings being a smoothed quantity over a credit cycle.

We propose in this paper a new model that maps the historical time series of rating migration matrices into a forward-looking stochastic rating migration matrix for any future period of interest. These stochastic rating migration matrices serve as the device for generating forward PIT-PDs and forward-looking TTC-PDs. The idea is to link the realized default rates and other-exit rates of each rating cohort to a set of credit cycle indices, and these credit cycle indices are captured by a dynamic time series model that exhibits concurrently global mean revision and local momentum as proposed in Duan (2016). The exit rate from a static pool of obligors for reasons other than default/bankruptcy is a factor whose importance should not be understated. Take the S&P global corporate rating pool as an example, a corporate ceases to receive a rating due to at least two reasons. First, a corporate may disappear simply because of a merger. Second, a corporate with a low credit rating may opt out of rating because it no longer makes sense to pay for a service that explicitly reveals its poor credit quality. In the case of internal rating system of a bank, other exits may reflect a merger or simply a terminated borrowing relationship. The values of the these credit cycle indices at various points of time are in our model the means to reflect the phase of a credit cycle, and serve as the starting point for advancing the system forward into future periods.

We implement the credit rating migration model using a set of credit cycle indices (reflecting global and sectoral movements) similar to those of Duan and Miao (2015) where the PD and POE (probability of other exits) data used in constructing these indices are taken from the Credit Research Initiative (CRI) corporate PD database at the National University of Singapore. The credit rating migration data used in our demonstration are the S&P long-term global corporate issuer rating migration rates extracted from the European Securities and Markets Authority (ESMA) database. The empirical results show that the S&P credit rating migration can be sensibly captured by our model and used to generate informative PIT-PDs or even TTC-PDs if needed. The forward PIT-PDs clearly exhibit a term structure effect and are reflective of different phases of a credit cycle.

2 Credit cycle drivers and realized default/other-exit rates

Consider a rating/scoring system that classifies the extant non-default obligors into K cohorts with 1 being the highest credit quality and K the worst. In addition, defaulted obligors are put into Cohort $K+1$. Since some obligors may leave the pool for reasons other than default, we must create Cohort $K+2$ to accommodate other exits to ensure internal consistency in rating migrations. The other-exit category captures those obligors becoming unrated due to, say, a merger/acquisition, or a managerial decision to opt out of credit rating, or a termination of the extant lending relationship initiated either by the lender or borrower, depending on the nature of an obligor pool.

In practice, realized default/other-exit rates over, say one year, are typically compiled. So, time series of the realized rates for different cohorts are readily available. In the case of credit rating agencies such as S&P, Moody's, etc., these rates are released to the public. More often, such time series are guarded as proprietary information with access only granted to in-house analysts. As expected, these cohort-specific rates will evolve over time in a dynamic fashion reflecting different phases of a credit cycle. We denote the default and other-exit rates for a static pool of obligors, say, Cohort k, over τ periods from $(t-\tau)$ to t by $D_{k,t}^{(\tau)}$ and $O_{k,t}^{(\tau)}$, respectively, whose values are realized at time t. The credit cycle may be captured by some indices.

2.1 Credit cycle indices and their dynamics

We adopt the credit cycle drivers similar to the approach of Duan and Miao (2015), who used the CRI corporate PD database at the Risk Management Institute, National University of Singapore to generate monthly time series of median values of the one-month PDs and POEs (probabilities of other exits) for the global corporate and 10 industry sectors where the sectors are set according to the Bloomberg Industry Classification System.[2] One cannot

[2]The CRI PD and POEs are based on the forward-intensity model of Duan, *et al* (2012). For the technical details on how these PDs and POEs are computed, readers are referred to NUS-RMI Credit Research

meaningfully capture credit cycles without explicitly factoring in corporate exits for reasons other than defaults/bankruptcies. This becomes apparent by referring to Table 1 of Duan, *et al* (2012) where other corporate exit rates are shown to be about ten times the default/bankruptcy rates for US public firms.

Each pair of the credit cycle indices (PD and POE), being the global or one of the 10 industrial sectors, is assumed to follow a bivariate VAR(1) system with local momentum where the local momentum feature is motivated by the model of Duan (2016). As our later empirical results reveal, many of these credit cycle indices are indeed mean-reverting with local momentum but globally stationary. We use the first pair, i.e., the global PD and POE indices, to describe the bivariate system that is applicable to all other pairs. Note that the credit cycle indices are typically available on a higher-frequency, and thus the running index may be over subperiods of length s; for example, $s = 1/6$ when the credit cycle indices are on the monthly frequency whereas the rating migration data runs on the semiannual frequency.

Denote the pair of the global credit cycle indices, i.e., global median PD and POE, by $X_{0,t}^{(D)}$ and $X_{0,t}^{(O)}$. Let $X_{0,t}^{(D)*} = Logit(X_{0,t}^{(D)})$ and $X_{0,t}^{(O)*} = Logit(X_{0,t}^{(O)})$ where we apply the Logit transformation because PDs and POEs fall between 0 and 1. This pair of transformed global credit cycle indices is assumed to follow the following dynamics:

$$\begin{aligned}
\begin{bmatrix} \Delta X_{0,t}^{(D)*} \\ \Delta X_{0,t}^{(O)*} \end{bmatrix} &= \boldsymbol{\alpha}_0 + (\boldsymbol{\beta}_0 - \boldsymbol{I}_{2\times 2}) \begin{bmatrix} X_{0,t-s}^{(D)*} \\ X_{0,t-s}^{(O)*} \end{bmatrix} + \begin{bmatrix} \omega_0^{(D)} & 0 \\ 0 & \omega_0^{(O)} \end{bmatrix} \begin{bmatrix} \bar{X}_{0,t-s|n_0^{(D)}}^{(D)*} - X_{0,t-s}^{(D)*} \\ \bar{X}_{0,t-s|n_0^{(O)}}^{(O)*} - X_{0,t-s}^{(O)*} \end{bmatrix} \\
&\quad + \begin{bmatrix} \varepsilon_{0,t}^{(D)} \\ \varepsilon_{0,t}^{(O)} \end{bmatrix} \quad \text{for } t = 1s, 2s, \cdots \qquad (1) \\
\bar{X}_{0,t-s|n_0^{(D)}}^{(D)*} &= \frac{1}{n_0^{(D)}} \sum_{j=1}^{n_0^{(D)}} X_{0,t-js}^{(D)*} \\
\bar{X}_{0,t-s|n_0^{(O)}}^{(O)*} &= \frac{1}{n_0^{(O)}} \sum_{j=1}^{n_0^{(O)}} X_{0,t-js}^{(O)*}
\end{aligned}$$

Note that $\bar{X}_{0,t-s|n_0^{(D)}}^{(D)*}$ or $\bar{X}_{0,t-s|n_0^{(D)}}^{(D)*}$ is simply the average value over a moving window of length $n_0^{(D)}$ or $n_0^{(O)}$. According to Duan (2016), a negative (positive) value of $\omega_0^{(D)}$ implies that $\ln X_{0,t}^{(D)}$ exhibits the local momentum building (preserving) feature while being globally stationary.[3] Naturally, the same is true for $\omega_0^{(O)}$ and $\ln X_{0,t}^{(O)}$. $\boldsymbol{\alpha}_0$ is a 2-dimensional column

Initiative Technical Report Version: 2017 Update 1 in the reference list or see https://www.rmicri.org.

[3]Although we use simple moving average to define local momentum, the model of Duan (2016) allows for

vector and $\boldsymbol{\beta}_0$ is a 2×2 matrix. $\varepsilon_{0,t}^{(D)}$ and $\varepsilon_{0,t}^{(O)}$ are assumed to be two normal random variables with mean 0 and a covariance matrix $\boldsymbol{\Omega}_0$. In our later implementation, the transformed credit cycle indices are standardized to have mean zero and variance 1, and therefore $\boldsymbol{\alpha}_0$ is set to zero.

For the sectoral credit cycle indices, we again apply the Logit transformation, and then orthogonize them individually on their corresponding standardized transformed global credit cycle index. We denote the i-th sectoral pair of standardized, orthogonized and transformed sectorial credit cycle indices by $X_{i,t}^{(D)*}$ and $X_{i,t}^{(O)*}$.

Different from Duan and Miao (2015), we model the Logit-transformed median PDs and POEs instead of performing their specific nonlinear transformation. In Duan and Miao (2015), the transformed sectoral index pairs (PD and POE) are first orthogonized to the pair of transformed global indices and then sequentially orthogonized to other sectoral index pairs. For an easier interpretation of the sectoral indices, we only orthogonize each of the Logit-transformed sectoral index pairs on the global pair without performing further sequential orthogonizations. However, we allow correlated residuals across different pairs of sectoral indices because the orthogonization is only applied on the global pair.

The model in equation (1) for the global or sectoral index pairs can be straightforwardly estimated when $n_0^{(D)}$ and $n_0^{(O)}$ are known. But these moving window lengths are actually unknown and need to be estimated. Thus, we resort to the density-tempered SMC method as in, say, Del Moral, *et al* (2006) and Duan and Fulop (2015) to tackle the estimation task. However, this estimation is treated as a likelihood maximization problem rather than a Bayesian estimation with some prior belief. Our density-tempered SMC estimation is similar to that of Duan and Wang (2016) where the SMC procedure is started with an initialization sampler as opposed to a Bayesian prior distribution. The presence of the two discrete unknown parameters are unique to our problem, however. In our initialization sampler, $n_0^{(D)}$ and $n_0^{(O)}$ are treated as independent with equal probabilities from 2 to some number large enough. As the SMC algorithm progresses, $n_0^{(D)}$ and $n_0^{(O)}$ will settle on a small set of combinations with high likelihoods. We pick the $(n_0^{(D)}, n_0^{(O)})$ combination that yields the highest likelihood value. For statistical inference on other model parameters, we narrow our focus on the SMC subsample corresponding the chosen $(n_0^{(D)}, n_0^{(O)})$ combination and performs the analysis like Duan and Wang (2016), which invokes a result of Chernozhukov and Hong (2003) to justify the use of the SMC sample for asymptotic inference because the information equality holds when the correctly specified likelihood function is the target.[4] For the 10 sectoral pairs of credit cycle indices, we estimate the model in the same way.

any kind of weighted average. The parameter restriction needed for stationary of this bivariate system can also be established along the line of Duan (2016).

[4]The SMC sample size will be increased to a level at which at least 1,000 parameter values are obtained

Table 1: Parameter estimates for each pair of credit cycle indices as in equation (1) with standard errors in parentheses. α_1 and α_2 are set to zero because the credit cycle indices are demeaned. Only those with the 10% statistical significance are kept and the presented results are re-estimated.

Industry sector	$n^{(D)}$(months)	$w^{(D)}$	β_{11}	β_{12}	σ_1
Global	2	-0.6054	0.9590	-0.0251	0.1972
		(0.1095)	(0.0126)	(0.0127)	
Financial	8	-0.1033	0.9596		0.1824
		(0.0318)	(0.0108)		
Basic Material	3	-0.1854	0.9468		0.2726
		(0.0745)	(0.0151)		
Communications	3	-0.3523	0.9360		0.2858
		(0.0676)	(0.0173)		
Consumer Cyclical	3	-0.2389	0.8849	-0.0584	0.3598
		(0.0735)	(0.0242)	(0.0239)	
Consumer Non-cyclical	2		0.9364		0.2914
			(0.0165)		
Diversified	24	-0.0726	0.7997	-0.0435	0.3959
		(0.0187)	(0.0260)	(0.0232)	
Energy	4	-0.1953	0.9508		0.2860
		(0.0587)	(0.0169)		
Industrial	2		0.9149		0.4053
			(0.0238)		
Technology	3	-0.2826	0.9300		0.3161
		(0.0691)	(0.0187)		
Utilities	17	-0.0487	0.8981	-0.0572	0.3606
		(0.0291)	(0.0246)	(0.0216)	

Industry sector	$n^{(O)}$(months)	$w^{(O)}$	β_{21}	β_{22}	σ_2	ρ_{12}
Global	3	-0.2348		0.9771	0.1439	
		(0.0700)		(0.0085)		
Financial	23	-0.0426		1	0.2165	
		(0.0230)		(0.0151)		
Basic Material	7	-0.1593		0.9672	0.1887	
		(0.0336)		(0.0111)		
Communications	2			0.9463	0.3181	0.1309
				(0.0180)		(0.0583)
Consumer Cyclical	2		-0.0422	0.9265	0.3231	
			(0.0217)	(0.0211)		
Consumer Non-cyclical	2		0.0639	0.8928	0.3635	
			(0.0254)	(0.0251)		
Diversified	14	-0.0644		0.9092	0.3358	
		(0.0306)		(0.0227)		
Energy	2			0.9434	0.3321	
				(0.0194)		
Industrial	8	-0.0722	0.0318	0.9363	0.2861	
		(0.0386)	(0.0169)	(0.0182)		
Technology	2			0.9760	0.2254	0.1765
				(0.0128)		(0.0554)
Utilities	5	-0.1862		0.8974	0.3608	0.1186
		(0.0492)		(0.0228)		(0.0571)

The monthly time series of these 22 indices extracted from the CRI corporate PD database of the National University of Singapore in November 2017, and the data covers the sample period of December 1990 to December 2015, spanning a longer period than the S&P rating migration data from the ESMA database which are available on a semiannual frequency and, at the time of data extraction, spans the period from 2000 to 2015.

Table 1 contains the estimation results for the dynamics of these 22 credit cycle indices. As expected, all 22 credit cycle indices are highly autocorrelated as reflected in either β_{11} or β_{22}, whereas the cross correlations are weak as revealed by β_{12} or β_{21}. The local-momentum effect is clearly present in both global PD and POE indices and it is the local-momentum building type as defined in Duan (2016), meaning that the stochastic process likely continues its recent upward or downward trend, because $w^{(D)}$ and $w^{(O)}$ are significantly negative. For sectoral PD and POE indices, the results on local momentum are mixed with some estimates of w being statistically insignificant. Among the significant ones, some are negative and others are positive. When w is positive, it is considered to be the local-momentum preserving type according to Duan (2016), suggesting that the stochastic process likely hovers around its current level. Note that insignificant parameters at the 10% level are removed and the model re-estimated.

Although estimates for ρ_{12}'s are available from the pair-wise estimation, we do not use these correlations in the later application, because the residuals of the 22 series are allowed to be correlated beyond just the pairwise correlation under the model in equation (1). Next, we will elaborate on a practical way of estimating the overall correlation matrix of these 22 residuals while eliminating spurious correlations due to sampling errors.

We use a thresholding regularization procedure similar to Bickel and Levina (2008), Rothman, *et al* (2009), Cai and Liu (2011), and Duan and Miao (2015). First note that our task is simpler because the sample correlation matrix in our case is always positive semidefinite arising from the fact that missing data does not occur in these credit cycle indices. Thresholding is to apply a minimum magnitude, denoted by ρ_m, to correlations so that correlations with a smaller magnitude are set to zero. We identify the optimal ρ_m through cross-validation with L random splits of the sample into two subsets. For each

under the best $(n_0^{(D)}, n_0^{(O)})$ combination. This is achieved by starting with an initial SMC sample of 1,000 parameter values and later increasing the SMC sample size to the desired level by applying the k-fold duplication idea of Duan and Zhang (2016), which can avoid going through the density-tempering steps. The sample standard deviation for each parameter is then computed. Different from Duan and Wang (2016), however, we use the maximum likelihood estimator produced by the SMC procedure. To increase the precision of the SMC maximum likelihood estimate, we apply data cloning in the spirit of Lele, *et al* (2007) and Lele, *et al* (2010) to raise the power of the likelihood function (doubling each time) until the SMC maximum likelihood value is stabilized to the point where its log value can no longer be increased by more than 0.01.

random split, the data matrix, say, Z, is divided into the training set Z_1 and the validation set Z_2, where the sample sizes (T_1 and T_2) are determined by $T_2 = T/\ln(T)$ and $T_1 = T - T_2$. For the l-th split, $\hat{\Sigma}_1^{(\rho_m,l)}$ denotes the resulting correlation matrix after applying thresholding to the sample correlation matrix computed from the training dataset, whereas $\tilde{\Sigma}_2^{(l)}$ is the sample correlation matrix based on the validation data set. The best threshold value, ρ_m^*, is the solution to the following problem:

$$\rho_m^* = \arg\min_{\rho_m \geq 0} \; \frac{1}{L} \sum_{l=1}^{L} \left\| \hat{\Sigma}_1^{(\rho_m,l)} - \tilde{\Sigma}_2^{(l)} \right\|_F^2$$

where $|| \cdot ||_F$ stands for the Frobenius norm. We set $L = 10$ in our later implementation. The correlation matrix for the 22 residuals estimated with the thresholding technique is not presented in the paper to conserve space.

2.2 Linking default/other-exit rates to credit cycle drivers

Let $\boldsymbol{X}_t^{(D)}$ be the 11-dimensional row vector containing the PD-based credit cycle indices without the Logit transformation.[5] All sectoral credit cycle indices are again orthogonized on the global index. Similarly, $\boldsymbol{X}_t^{(O)}$ be the 11-dimensional row vector containing the POE-based credit cycle indices. We assume that one-period default and other-exit rates can be modeled by the following Tobit model: for $k = 1, 2, \cdots, K$ and $t = 1, 2, \cdots$,

$$\begin{aligned} D_{k,t}^{(1)} &= \begin{cases} D_{k,t}^{(1)*} & \text{if } 0 < D_{k,t}^{(1)*} < 1 \\ 0 & \text{if } D_{k,t}^{(1)*} \leq 0 \\ 1 & \text{if } D_{k,t}^{(1)*} \geq 1 \end{cases} \\ \text{where } D_{k,t}^{(1)*} &= \alpha_{D,k} + \left(\sum_{j=0}^{1/s-1} \boldsymbol{X}_{t-js}^{(D)} \right) \boldsymbol{\beta}_{D,k} + \epsilon_{D,k,t} \end{aligned} \tag{2}$$

and

$$\begin{aligned} O_{k,t}^{(1)} &= \begin{cases} O_{k,t}^{(1)*} & \text{if } 0 < O_{k,t}^{(1)*} < 1 \\ 0 & \text{if } O_{k,t}^{(1)*} \leq 0 \\ 1 & \text{if } O_{k,t}^{(1)*} \geq 1 \end{cases} \\ \text{where } O_{k,t}^{(1)*} &= \alpha_{O,k} + \left(\sum_{j=0}^{1/s-1} \boldsymbol{X}_{t-js}^{(O)} \right) \boldsymbol{\beta}_{O,k} + \epsilon_{O,k,t}. \end{aligned} \tag{3}$$

[5] We do not Logit-transform the credit cycle indices because the realized default rates used as the dependent variable in the Tobit regression cannot be likewise transformed due to the presence of zero.

Note that $\boldsymbol{\beta}_{D,k} = (\beta_{D,k,1}, \cdots, \beta_{D,k,p})'$ is the 11-dimensional regression coefficients, and $\epsilon_{D,k,t}$ is a normally distributed innovation term with mean 0 and standard deviation $\sigma_{D,k}$. $\boldsymbol{\beta}_{O,k}$ and $\epsilon_{O,k,t}$ are similarly defined. For the interpretability of $\boldsymbol{\beta}_{D,k}$ and $\boldsymbol{\beta}_{O,k}$ across different rating cohorts, we have standardized all explanatory variables in the Tobit regression to have mean 0 and variance 1. $\epsilon_{D,k,t}$ and $\epsilon_{O,k,t}$ are assumed to be independent over time, but may be contemporaneously correlated. The regressors are the sums over $1/s$ subperiods in order to accommodate a likely mixed-frequency situation in practice where, say, the credit risk cycle drivers are available monthly but the realized default and other-exit rates are semiannual so that $s = 1/6$. The default/other-exit rates realized over a 6-month period would not have been well captured by the credit risk cycle variables if one had only focussed on their period end values. We adopt a Tobit model because a high credit quality cohort often experiences zero realized default rates.[6] When a time series for a cohort contains too many zero default rates, there may create an identification problem. Merging into another cohort or finding a sensible proxy time series seems to be a sensible option. The latter is adopted in our implementation for the AAA, AA, A and BBB cohorts of the S&P ratings.

We extract realized default and other-exit rates from the ESMA database on the S&P long-term corporate issuer ratings, which at the time of extraction spans the period from 2000 to 2015. Data up to the semiannual frequency are available, and we consider seven rating cohorts (AAA, AA, A, BBB, BB, B, and CCC/CC/C) before default or other exits. The default rate is directly available from the ESMA database, but the other-exit rate is deduced by one minus the sum of the default rate and the migration rates for the seven rating cohorts.[7]

Semiannual default rates for each of the three rating cohorts (BB, B, and CCC/CC/C) are regressed on the 11 PD-based credit cycle indices where the indices are sampled monthly but averaged semiannually to match the frequency of the dependent variable. For the top four credit quality cohorts (AAA, AA, A and BBB), the time series of semiannual realized default rates are either all zeros or contain too few non-zeros, and thus running a Tobit reqression on these series would be meaningless. For these rating cohorts, we apply four proxy time series that are the median model PDs extracted from the CRI corporate PD database, where the comparable AAA, AA, A and BBB categories are determined by a PD-implied rating method which essentially determines a set of boundary PD values for different rating categories so as to map the CRI PDs for the global corporate pool of over 68,000 firms to the S&P historical average default rates.[8] We create these proxy series with the following

[6] Censoring above at 1 is not a practical concern, and so we have ignored it in implementation.

[7] The ESMA database reports separately the CCC, CC and C categories, but we follow the S&P 2014 report, "Annual Global Corporate Default Study and Rating Transitions," to combine these three categories into one.

[8] For technical details, please refer to the PDiR methodology in "Probability of Default implied Rating (PDiR) White Paper," 2018, the Credit Research Initiative, National University of Singapore.

three steps: (1) identify the group of obligors belonging to a rating cohort, say, AA six months prior to time t according to the PDiR methodology, (2) add up six one-month PDs leading to time t for each obligor, i.e., $t - \frac{5}{6}, \cdots, t - \frac{1}{6}, t$,[9] and (3) obtain the median PD value of the group and use it as a proxy value for the rating cohort.

For semiannual other-exit rates, a similar regression on the 11 POE-based credit cycle indices is run for each cohort. Since the other-exit rate series for any cohort always contains many non-zeros, there is no need to find a proxy time series for any rating cohort.

The Tobit regression is estimated with the adaptive Lasso penalty to avoid over-fitting due to too many regressors. We choose to incorporate the adaptive Lasso penalty by Zou (2006) into the Tobit regression because the adaptive Lasso retains the oracle property. When coupled with with the convexity formulation of Tobit model by Olsen (1978), the estimation remains to be a convex minimization problem for which a unique global solution exists. The suitable level of penalty, i.e., the tuning parameter, is determined by the BIC criterion.

The Tobit model estimation results are summarized in Table 2. Evidently, realized default rates in all cohorts (or their proxy values for the AAA, AA, A and BBB cohorts explained earlier) are positively related to the global credit cycle PD index as reflected by the results in Panel A. For some cohorts, the results show that they are also related to some sectoral credit cycle PD indices. Note that some coefficients are negative because the sectoral credit cycle indices have been orthogonalized to the global credit cycle index. The explanatory powers indicated by R^2 of the Tobit regression are computed from the residual variances using the two expected values under the Tobit model with and without regressors. The lowest R^2 is about 57% corresponding to the CCC/CC/C category, whereas the largest explanatory power is at 85.6% for the B rating cohort. These results are quite intuitive, and suggest that the Tobit model works fairly well in relating the default experience to the credit cycle indices.

The Tobit regression results for the other-exit rates show a quite different pattern. The AA cohort's other-exit rate responds to the global and all sectoral POE credit cycle indices with a very high R^2 (a bit over 60%), but other cohorts, particularly those riskier categories, do not respond to POE credit cycle indices at all. We contend that this is expected because the reason for not getting a S&P rating for better rated cohorts has more to do with merger/acquisition activities. For firms below the investment grade, there is no incentive to pay for a credit rating that actually reveals one's poor credit quality.

[9] The defaulters during the six-month period are assigned a PD of 1 and those experienced other exits are given a PD of 0.

Table 2: Summary of the Tobit regressions for default and other-exit rates of different rating cohorts (with the adaptive Lasso regularization and the BIC selection of the tuning parameter)

Panel A. Defaults

Industry sector	AAA	AA	A	BBB	BB	B	CCC/CC/C
Constant	2.01×10^{-6}	3.25×10^{-5}	0.0001	0.0006	0.0015	0.0159	0.1653
Global	5.85×10^{-7}	5.52×10^{-6}	2.42×10^{-5}	0.0001	0.0007	0.0131	0.0518
Financial	5.19×10^{-7}				-0.0003	0.0092	
Basic Material		-2.00×10^{-6}				0.0027	
Communications		-1.62×10^{-6}			0.0034	0.0133	
Consumer Cyclical		2.73×10^{-6}	1.97×10^{-5}			0.0135	
Consumer Non-cyclical					0.0017		
Diversified			2.13×10^{-5}		0.0022	0.0137	0.0147
Energy			-1.36×10^{-5}				
Industrial			-8.17×10^{-6}				
Technology					-0.0030		
Utilities	2.53×10^{-7}	8.11×10^{-6}	4.14×10^{-5}	0.0001	-0.0012		0.0136
R^2	0.6215	0.8079	0.7985	0.6181	0.6858	0.8560	0.5706

Panel B. Other Exits

Industry sector	AAA	AA	A	BBB	BB	B	CCC/CC/C
Constant	0.0032	0.0208	0.0250	0.0342	0.0473	0.0509	0.0604
Global		0.0054					
Basic Material	-0.0077	-0.0031					
Consumer Cyclical		0.0066					
Diversified	0.0082	0.0055	0.0052				
Energy		0.0050					
R^2	0.3760	0.6012	0.2222	0.0000	0.0000	0.0000	0.0000

Note: Only present the coefficients chosen by the adaptive Lasso.

3 Dynamic Point-in-Time rating migration matrices

The rating migration among the $(K + 2)$ categories from time t to $t + \tau$ is driven by a $(K + 2) \times (K + 2)$ time-dependent transition probability matrix, which is a conditional expectation at time t of future migrations driven by stochastic one-period default and other-exit rates, which are in large part determined by by how they react to the credit cycle indices, $\boldsymbol{X}_t^{(D)}$ and $\boldsymbol{X}_t^{(O)}$, over the period of t to $t + \tau$. Let $\tilde{D}_{j,i}^{(1)} = \max\left(0, D_{j,i}^{(1)} - \eta_j\right)$ if $D_{j,i}^{(1)}$ is a proxy time series, and otherwise $\tilde{D}_{j,i}^{(1)} = D_{j,i}^{(1)}$. Inevitably, one would have to use proxy time series to replace the realized default rates for some high credit quality cohorts, because their realized default rates are typically zeros. In our empirical implementation on the S&P rating migration, top four rating categories use proxy time series. Applying the Tobit model to generate default rates will naturally cause upward bias in these cases. Therefore, an adjustment parameter, η_j, is introduced into $\tilde{D}_{j,i}^{(1)}$ whenever a proxy time series is used. This newly introduced parameter η_j is set equal to the difference between the sample means of the proxy and realized default rate series. One naturally expects a positive η_j because a proxy series is used only when the realized default rates are mostly zeros.

Our specific construction is

$$\boldsymbol{S}_t(\tau) = E_t\left(\prod_{i=t+1}^{t+\tau} \boldsymbol{R}_{i-1,i}\right) \tag{4}$$

where $\boldsymbol{R}_{i-1,i}$ is defined as

$$\begin{bmatrix} q_{11}(D_{1,i}^{(1)}, O_{1,i}^{(1)}) & q_{12}(D_{1,i}^{(1)}, O_{1,i}^{(1)}) & \cdots & q_{1K}(D_{1,i}^{(1)}, O_{1,i}^{(1)}) & \frac{\tilde{D}_{1,i}^{(1)}}{\max\left(1,\tilde{D}_{1,i}^{(1)}+O_{1,i}^{(1)}\right)} & \frac{O_{1,i}^{(1)}}{\max\left(1,\tilde{D}_{1,i}^{(1)}+O_{1,i}^{(1)}\right)} \\ q_{21}(D_{2,i}^{(1)}, O_{2,i}^{(1)}) & q_{22}(D_{2,i}^{(1)}, O_{2,i}^{(1)}) & \cdots & q_{2K}(D_{2,i}^{(1)}, O_{2,i}^{(1)}) & \frac{\tilde{D}_{2,i}^{(1)}}{\max\left(1,\tilde{D}_{2,i}^{(1)}+O_{2,i}^{(1)}\right)} & \frac{O_{2,i}^{(1)}}{\max\left(1,\tilde{D}_{2,i}^{(1)}+O_{2,i}^{(1)}\right)} \\ \vdots & \vdots & \cdots & \vdots & \vdots & \vdots \\ q_{K1}(D_{K,i}^{(1)}, O_{K,i}^{(1)}) & q_{22}(D_{K,i}^{(1)}, O_{K,i}^{(1)}) & \cdots & q_{KK}(D_{K,i}^{(1)}, O_{K,i}^{(1)}) & \frac{\tilde{D}_{K,i}^{(1)}}{\max\left(1,\tilde{D}_{K,i}^{(1)}+O_{K,i}^{(1)}\right)} & \frac{O_{K,i}^{(1)}}{\max\left(1,\tilde{D}_{K,i}^{(1)}+O_{K,i}^{(1)}\right)} \\ 0 & 0 & \cdots & 0 & 1 & 0 \\ 0 & 0 & \cdots & 0 & 0 & 1 \end{bmatrix} \tag{5}$$

The default and other exit rates in the last two columns are adjusted in a way to ensure the sum of the two entries not exceeding 1. Without the adjustment, $\boldsymbol{R}_{i-1,i}$ need not be a legitimate stochastic migration matrix because the Tobit model can generate with a positive probability default and other exit rates with their sum exceeding 1. The entries in the second to the last row of $\boldsymbol{R}_{i-1,i}$, i.e., corresponding to the category of defaulters, are fairly obvious, but the last row requires explanation. When tracking a static pool of obligors over some horizon of interest, there are no new entries into any one of the first K cohorts, and thus those transition probabilities should equal zeros by definition. This is precisely the migration matrix needed for determining the PIT-PDs facing the existing obligors. This is also the migration matrix appropriate for generating expected default rates, over multiple periods, faced by a fixed set of obligors in any of the K cohorts. The migration matrix can in turn be used in parameter estimation to center observed realized default rates at their corresponding theoretical default rates.

In order to define $q_{jk}(D_{j,i}^{(1)}, O_{j,i}^{(1)})$ for $j = 1, \cdots, K$ and $k = 1, \cdots, K$, we let $\boldsymbol{Q}_{i-1,i}$ denote them, which is the top-left $k \times k$ submatrix of $\boldsymbol{R}_{i-1,i}$. The following rather general specification is adopted:

$$\boldsymbol{Q}_{i-1,i} = diag\left(\boldsymbol{N}_i\right)\left\{\boldsymbol{A} \cdot \dot{\exp}\left[diag\left(\boldsymbol{D}_i^{(1)*}\right)\boldsymbol{B}\right]\right\} \tag{6}$$

where $diag(\cdot)$ denotes the operator of making a column vector into a diagonal matrix, "$\cdot$" is an element-by-element multiplication of two matrices (i.e., the Hadamard product), $\dot{\exp}(\cdot)$ stands for an element-by-element exponentiation, $\boldsymbol{D}_i^{(1)*} = \left[D_{1,i}^{(1)}/\bar{D}_1^{(1)}, D_{2,i}^{(1)}/\bar{D}_2^{(1)}, \cdots, D_{K,i}^{(1)}/\bar{D}_K^{(1)}\right]'$ with $\bar{D}_j^{(1)}$ denoting the historical sample average of $D_{j,i}^{(1)} \geq 0$ for cohort j, which is used to scale default rates to make them comparable in magnitude across cohorts. $\boldsymbol{N}_i$ is a K-dimensional column vector with element j equal to $\frac{1-\left[\left(\tilde{D}_{j,i}^{(1)}+O_{j,i}^{(1)}\right)/\max\left(1,\tilde{D}_{j,i}^{(1)}+O_{j,i}^{(1)}\right)\right]}{RowSum_j\left\{\boldsymbol{A}\cdot\dot{\exp}\left[diag\left(\boldsymbol{D}_i^{(1)*}\right)\boldsymbol{B}\right]\right\}}$, and finally $\boldsymbol{A}$ and $\boldsymbol{B}$ are as follows:

$$\boldsymbol{A} = \begin{bmatrix} 1 & a_{12} & \cdots & a_{1K} \\ a_{21} & 1 & \cdots & a_{2K} \\ \vdots & \vdots & \cdots & \vdots \\ a_{K1} & a_{K2} & \cdots & 1 \end{bmatrix} \qquad \boldsymbol{B} = \begin{bmatrix} 0 & b_{12} & \cdots & b_{1K} \\ -b_{21} & 0 & \cdots & b_{2K} \\ \vdots & \vdots & \cdots & \vdots \\ -b_{K1} & -b_{K2} & \cdots & 0 \end{bmatrix}.$$

The prior belief structure on $\boldsymbol{A}$ and $\boldsymbol{B}$ is captured by an indicator function $\mathbf{1}_{(\boldsymbol{A},\boldsymbol{B})}$ with its value equal to 1 to indicate that the prior belief is satisfied by $\boldsymbol{A}$ and $\boldsymbol{B}$ and 0 otherwise. Specifically, the prior belief structure on $\boldsymbol{A}$ is $a_{ij} \geq a_{ik} \geq 0$ for $k > j \geq i$ and $0 \leq a_{ik} \leq a_{ij}$ for $k < j \leq i$ because the migration probabilities should not be negative and their magnitudes should decline away from the diagonal. The prior belief structure on $\boldsymbol{B}$ is $b_{ij} \geq b_{ik} \geq 0$ for $k > j \geq i$ and $0 \leq b_{ik} \leq b_{ij}$ for $k < j \leq i$ to reflect the intuition that a higher (lower) default rate is expected to be associated with higher migration rates to lower (higher) rating categories so that parameter values to the right (left) of the diagonal need to be positive (negative). The diagonals of $\boldsymbol{A}$ and $\boldsymbol{B}$ are with the assigned values for the identification purpose, because an arbitrary diagonal in $\boldsymbol{A}$ could yield an exactly same outcome through a compensating adjustment in $\boldsymbol{N}_i$. Likewise, the diagonal of $\boldsymbol{B}$ is set to zero because adding an arbitrary constant to any row would yield the same outcome again by a compensating adjustment in $\boldsymbol{N}_i$. Note that when a_{jk} is set to 0, its corresponding value for b_{jk} naturally becomes irrelevant and can be set to 0 for convenience.

It is straightforward to verify that all $q_{jk}(D_{j,i}^{(1)}, O_{j,i}^{(1)})$ in the above formulation are non-negative and $\sum_{k=1}^{K} q_{jk}(D_{j,i}^{(1)}, O_{j,i}^{(1)}) + \frac{\tilde{D}_{j,i}^{(1)}+O_{j,i}^{(1)}}{\max\left(1,\tilde{D}_{j,i}^{(1)}+O_{j,i}^{(1)}\right)} = 1$ for $j = 1, 2, \cdots, K$, and therefore $\boldsymbol{R}_{i-1,i}$ is a legitimate stochastic rating migration matrix that responds to changes in the realized default rates for different cohorts. The above specification is motivated by a combination of definition and intuition. The to-be-realized default rate of a cohort is, by definition, the fraction of the obligors in this cohort that jump to default over the coming period. Similarly, this applies to the to-be-realized other-exit rate. For migration to other cohorts, the default rates over the coming period are not directly tied to their migration probabilities, but are likely informative; for example, a higher (lower) to-be-realized default rate intuitively suggests that a higher fraction of obligors in a cohort are expected to migrate to next cohort indicated by a lower (higher) credit quality. One would expect a higher migration probability in either direction for moving one cohort than two, which can be reflected in the magnitude of b_{jk} being decreasing further away from the diagonal. The above stochastic migration matrix contains at most $2(K-1)^2$ unknown parameters in $\boldsymbol{A}$ and $\boldsymbol{B}$, and they are subject to such a prior belief structure.

When the modeling situation warrants, migration possibilities can be reduced to, say, at most two cohorts over one period, which can be easily accomplished by setting to zero all entries beyond two levels off the diagonals of $\boldsymbol{A}$ in either direction. Since users may want to limit the number of parameters immediately off the diagonal of $\boldsymbol{A}$ and $\boldsymbol{B}$, we define the left and right index limits: $k_l(j) = \min(j - k^*, 1)$ and $k_r(j) = \max(j + k^*, K)$ for row j where k^* is the maximum number allowed. When $k^* = 2$ for example, estimation is to use the following $\boldsymbol{A}$ and $\boldsymbol{B}$:

$$\boldsymbol{A} = \begin{bmatrix} 1 & a_{12} & a_{13} & 0 & 0 & 0 & 0 \\ a_{21} & 1 & a_{23} & a_{24} & 0 & 0 & 0 \\ a_{31} & a_{32} & 1 & a_{34} & a_{35} & 0 & 0 \\ 0 & a_{42} & a_{43} & 1 & a_{45} & a_{46} & 0 \\ 0 & 0 & a_{53} & a_{54} & 1 & a_{56} & a_{57} \\ 0 & 0 & 0 & a_{64} & a_{65} & 1 & a_{67} \\ 0 & 0 & 0 & 0 & a_{75} & a_{76} & 1 \end{bmatrix} \quad \boldsymbol{B} = \begin{bmatrix} 0 & b_{12} & b_{13} & 0 & 0 & 0 & 0 \\ -b_{21} & 0 & b_{23} & b_{24} & 0 & 0 & 0 \\ -b_{31} & -b_{32} & 0 & b_{34} & b_{35} & 0 & 0 \\ 0 & -b_{42} & -b_{43} & 0 & b_{45} & b_{46} & 0 \\ 0 & 0 & -b_{53} & -b_{54} & 0 & b_{56} & b_{57} \\ 0 & 0 & 0 & -b_{64} & -b_{65} & 0 & b_{67} \\ 0 & 0 & 0 & 0 & -b_{75} & -b_{76} & 0 \end{bmatrix}$$

Dynamic $\boldsymbol{S}_t(\tau)$ defined in equation (4) can be computed concurrently by Monte Carlo simulations for all cohorts using the values at time t for $\boldsymbol{X}_t^{(D)}$, $\boldsymbol{X}_t^{(O)}$, $D_{k,t}^{(1)}$ and $O_{k,t}^{(1)}$ for $k = 1, 2, \cdots, K$. The system defined in equations (1)-(3) can be used to generate future paths of $\boldsymbol{X}_t^{(D)}$ and $\boldsymbol{X}_t^{(O)}$, and then $D_{k,t}^{(1)}$ and $O_{k,t}^{(1)}$ for $k = 1, 2, \cdots, K$, which in turn determines $\boldsymbol{R}_{i-1,i}$ for $i = t+1, t+2, \cdots, t+\tau$.

Note that future values of $\boldsymbol{X}_t^{(D)}$ and $\boldsymbol{X}_t^{(O)}$ are simulated at a higher frequency, say, monthly, whereas those of $D_{k,t}^{(1)}$ and $O_{k,t}^{(1)}$ for $k = 1, 2, \cdots, K$ are generated, say, semiannually. Repeat the simulation, say, 100 times and compute the average of the stochastic matrix, $\prod_{i=t+1}^{t+\tau} \boldsymbol{R}_{i-1,i}$, for each of targeted τ to arrive an Monte Carlo estimate of $\boldsymbol{S}_t(\tau)$.[10]

Note that when a proxy series is used, the value of η has already been determined based on the difference between sample means of the proxy and realized default rate series. One can utilize recorded realized rating migration matrix over horizon $[t, t+\tau_i]$ denoted by $\boldsymbol{M}_t(\tau_i)$ for $i = 1, 2, \cdots, m$. $\boldsymbol{M}_t(\tau_i)$ should be understood as a $K \times (K+2)$ matrix, recording migration rates from K rating cohorts to $K+2$ outcomes (K rating cohorts plus the default and other-exit categories). Perform a nonlinear least squares estimation by pairing $\boldsymbol{M}_t(\tau_i)$ with the corresponding $K \times (K+2)$ submatrix of $\boldsymbol{S}_t(\tau_i)$ over the sample period. If only some entries of $\boldsymbol{M}_t(\tau_i)$ are available, say, realized default rate (i.e., the $(K+1)$-th column of $\boldsymbol{M}_t(\tau_i)$), estimation can just focus on comparing the values in this column. In the following exposition, we assume the whole matrix is available.

Estimation is to maximize the following pseudo-likelihood function:

$$\mathcal{L}(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{\psi}(1), \boldsymbol{\psi}(2), \cdots, \boldsymbol{\psi}(m); \mathcal{D}_{1:T}) = \frac{\mathbf{1}_{(\boldsymbol{A},\boldsymbol{B})}}{\prod_{i=1}^{m} \prod_{t=1}^{T-\tau_i} \prod_{k=1}^{K} \prod_{l=1}^{K+1} \sqrt{2\pi}\psi_{kl}(i)} \times \exp\left\{ -\sum_{i=1}^{m} \sum_{t=1}^{T-\tau_i} \sum_{k=1}^{K} \sum_{l=1}^{K+1} \frac{\left(\boldsymbol{M}_t^{(k,l)}(\tau_i) - \boldsymbol{S}_t^{(k,l)}(\tau_i; \boldsymbol{A}, \boldsymbol{B})\right)^2}{2\psi_{kl}^2(i)} \right\} \tag{7}$$

subject to $\psi_{kl}(i) > 0$ for $i = 1, 2, \cdots, m,\ k = 1, 2, \cdots, K,$ and $l = 1, 2, \cdots, K+1$

In the above, $\mathcal{D}_{1:T}$ denotes the data set from time 1 to T, which comprises $\boldsymbol{M}_t(\tau_i)$ for $t = 1, 2, \cdots, T$ and $i = 1, 2, \cdots, m$. $\boldsymbol{A}$ and $\boldsymbol{B}$ are added to $\boldsymbol{S}_t(\tau_i)$ to emphasize dependency on $\boldsymbol{A}$ and $\boldsymbol{B}$. The superscript (k, l) is used to indicate an element of $\boldsymbol{S}_t(\tau_i; \boldsymbol{A}, \boldsymbol{B})$ and $\boldsymbol{M}_t(\tau_i)$. Note that we only run the index up to $K+1$ because by definition any row sum of either $\boldsymbol{S}_t(\tau_i; \boldsymbol{A}, \boldsymbol{B})$ or $\boldsymbol{M}_t(\tau_i)$ equals 1, which naturally removes one degree of freedom. The nonlinear regression errors can potentially be captured by $\boldsymbol{\psi}(i)$, a $K \times (K+1)$ matrix of parameters for horizon τ_i. However, that could be excessively general for practical usage, and we thus adopt a more restricted but sensible form in the later application.

[10]We found that 100 simulated paths sufficient for the estimation of $\boldsymbol{A}$ and $\boldsymbol{B}$ using the S&P credit migration data, but one may need to use more simulation paths in order to obtain smoother estimates of $\boldsymbol{S}_t(\tau)$ for certain applications.

Although $\boldsymbol{S}_t(\tau_i; \boldsymbol{A}, \boldsymbol{B})$ is computed by Monte Carlo simulation, the likelihood function can be differentiable in parameters in $\boldsymbol{A}$ and $\boldsymbol{B}$ if common random numbers are used in evaluating the pseudo-likelihood function when the parameter value is varied. A gradient-based optimization method can in principle be deployed to perform this nonlinear least squares estimation. Due to a large number of parameters and the prior belief structure, we again resort to the density-tempered SMC method to tackle the estimation task. When $\boldsymbol{A}$ and $\boldsymbol{B}$ are known, the optimal $\boldsymbol{\psi}$ can be analytically solved. Hence, the SMC algorithm is implemented to take advantage of this feature. The statistical inference in this case is more complicated, because the pseudo-likelihood function in equation (7) is created by making predictions at a time point over several overlapping horizons. In short, the information equality as described in Chernozhukov and Hong (2003) no longer holds to justify a direct use of the SMC sample to perform inference. Thus, we take the SMC sample means as the pseudo maximum likelihood estimates and deploy the self-normalized approach of Shao (2010) to generate their standard errors.

The semiannual time series of $\boldsymbol{M}_t(\tau_i)$ for the S&P ratings are extracted from the ESMA database for six horizons (τ equals 6 months, 1 year, 2 years, 3 years, 4 years and 5 years) and seven rating cohorts. We adopt the following restricted error structure, which recognizes that the diagonal values are supposed to be larger than others, the off-diagonal values become smaller further away from the diagonal, and the default rates in column eight are cohort-specific except for the top three rating categories.

$$\boldsymbol{\psi}(\tau)_{7\times 8} = \begin{pmatrix} \psi_{11}(\tau) & \psi_{12}(\tau) & \psi_{13}(\tau) & \psi_{13}(\tau) & \psi_{13}(\tau) & \psi_{13}(\tau) & \psi_{13}(\tau) & \psi_{18}(\tau) \\ \psi_{12}(\tau) & \psi_{11}(\tau) & \psi_{12}(\tau) & \psi_{13}(\tau) & \psi_{13}(\tau) & \psi_{13}(\tau) & \psi_{13}(\tau) & \psi_{18}(\tau) \\ \psi_{13}(\tau) & \psi_{12}(\tau) & \psi_{11}(\tau) & \psi_{12}(\tau) & \psi_{13}(\tau) & \psi_{13}(\tau) & \psi_{13}(\tau) & \psi_{18}(\tau) \\ \psi_{13}(\tau) & \psi_{13}(\tau) & \psi_{12}(\tau) & \psi_{11}(\tau) & \psi_{12}(\tau) & \psi_{13}(\tau) & \psi_{13}(\tau) & \psi_{48}(\tau) \\ \psi_{13}(\tau) & \psi_{13}(\tau) & \psi_{13}(\tau) & \psi_{12}(\tau) & \psi_{11}(\tau) & \psi_{12}(\tau) & \psi_{13}(\tau) & \psi_{58}(\tau) \\ \psi_{13}(\tau) & \psi_{13}(\tau) & \psi_{13}(\tau) & \psi_{13}(\tau) & \psi_{12}(\tau) & \psi_{11}(\tau) & \psi_{12}(\tau) & \psi_{68}(\tau) \\ \psi_{13}(\tau) & \psi_{13}(\tau) & \psi_{13}(\tau) & \psi_{13}(\tau) & \psi_{13}(\tau) & \psi_{12}(\tau) & \psi_{11}(\tau) & \psi_{78}(\tau) \end{pmatrix}$$

With seven rating cohorts, the total number of free parameters in $\boldsymbol{A}$ and $\boldsymbol{B}$ equals 44. Adding the 8 parameters in $\boldsymbol{\psi}$ for each of six horizons (restricting AAA, AA, and A cohorts to share the same regression error on column eight) yields a grand total of 92 unknown parameters.

The estimated coefficient matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ based on the semiannual frequency are presented in Table 3 with their standard errors in parentheses. These estimates obey the prior belief structure in a natural way, meaning that immediately off-diagonal elements are larger in magnitude, suggesting migration to the nearest cohort is not only more likely ($\boldsymbol{A}$ matrix) but also more sensitive to the realized default rate ($\boldsymbol{B}$ matrix), and none of them hitting the boundary conditions implied by the prior belief structure. Also presented in

Table 3: Estimated parameters for the dynamic point-in-time rating migration matrices

$\boldsymbol{A}$	c1	c2	c3	c4	c5	c6	c7
r1	1	0.1131	0.0003	0	0	0	0
		(0.0019)	(0.0009)				
r2	0.0046	1	0.0472	0.0031	0	0	0
	(0.0004)		(0.0032)	(0.0002)			
r3	0.0013	0.0015	1	0.0192	0.0030	0	0
	(0.0001)	(0.0001)		(0.0025)	(0.0001)		
r4	0	0.0006	0.0187	1	0.0106	0.0028	0
		(0.0001)	(0.0011)		(0.0024)	(0.0003)	
r5	0	0	0.000770194	0.0276	1	0.0419	0.0033
			(0.0001)	(0.0011)		(0.0014)	(0.0003)
r6	0	0	0	0.0016	0.0246	1	0.0290
				(0.0001)	(0.0014)		(0.0044)
r7	0	0	0	0	0.0011	0.1066	1
					(0.0003)	(0.0089)	

$\boldsymbol{B}$	c1	c2	c3	c4	c5	c6	c7
r1	0	0.3226	0.2676	0	0	0	0
		(0.0550)	(0.0512)				
r2	-0.1807	0	0.0569	0.0064	0	0	0
	(0.0185)		(0.0448)	(0.0037)			
r3	-0.1123	-4.5374	0	0.3650	0.2835	0	0
	(0.0332)	(0.1852)		(0.1157)	(0.0876)		
r4	0	-0.1294	-0.2478	0	0.4276	0.0199	0
		(0.1134)	(0.3025)		(0.0534)	(0.0627)	
r5	0	0	-0.0569	-0.3216	0	0.1172	0.0063
			(0.0489)	(0.0992)		(0.0309)	(0.0249)
r6	0	0	0	-0.0171	-0.8305	0	0.3469
				(0.0146)	(0.0963)		(0.0397)
r7	0	0	0	0	-0.6088	-0.6393	0
					(0.1610)	(0.1590)	

$\boldsymbol{\psi}(\tau)$	ψ_{11}	ψ_{12}	ψ_{13}	ψ_{18}	ψ_{48}	ψ_{58}	ψ_{68}	ψ_{78}
τ = 0.5y	0.0623	0.0408	0.0021	0.0004	0.0011	0.0023	0.0162	0.0676
	(0.0022)	(0.0018)	(0.0001)	(0.0000)	(0.0001)	(0.0001)	(0.0006)	(0.0033)
τ = 1y	0.0809	0.0533	0.0039	0.0007	0.0024	0.0080	0.0399	0.0807
	(0.0029)	(0.0025)	(0.0001)	(0.0000)	(0.0001)	(0.0005)	(0.0021)	(0.0031)
τ = 2y	0.0947	0.0617	0.0102	0.0018	0.0065	0.0224	0.0729	0.1032
	(0.0052)	(0.0041)	(0.0005)	(0.0001)	(0.0005)	(0.0008)	(0.0037)	(0.0026)
τ = 3y	0.0941	0.0650	0.0127	0.0034	0.0103	0.0329	0.0901	0.1101
	(0.0072)	(0.0059)	(0.0008)	(0.0003)	(0.0008)	(0.0009)	(0.0053)	(0.0029)
τ = 4y	0.0894	0.0653	0.0152	0.0048	0.0123	0.0361	0.0948	0.1166
	(0.0064)	(0.0064)	(0.0008)	(0.0005)	(0.0010)	(0.0013)	(0.0058)	(0.0048)
τ = 5y	0.0784	0.0643	0.0187	0.0059	0.0130	0.0375	0.0942	0.1227
	(0.0064)	(0.0068)	(0.0012)	(0.0008)	(0.0010)	(0.0022)	(0.0071)	(0.0064)

Table 3 is $\boldsymbol{\psi}$ matrix. The estimates are consistent with the intuition that the model error becomes larger for a migration over a longer period, i.e., increasing τ.

4 Forward Migration Generators, and Point-in-Time and Through-the-Cycle PDs

Recall that $\boldsymbol{S}_t(\tau_i)$ is the spot PIT τ-period rating migration matrix at time t, and its $(K+1)$-th column provides the corresponding spot PIT-PDs for different cohorts. Our model estimated with rating migration matrices measured over horizons from 6 months to 5 years only provide values for $\boldsymbol{S}_t(\tau_i)$ at $i = 1, 2, \cdots$, with each τ_i increased by 6 months. Because $\boldsymbol{S}_t(\tau_i)$ constitutes a set of legitimate rating migration matrices, there exists an obvious way to interpolate/extrapolate through creating a set of corresponding forward migration generators at time t. Denote these forward generator matrices by $\{\boldsymbol{F}_t(\tau_0), \boldsymbol{F}_t(\tau_1), \cdots\}$ where $\tau_0 = 0$. We can define the extended $\boldsymbol{S}_t(\tau)$ for all horizons in continuous time with these forward migration generators as follows: for $\tau \geq 0$ and $\tau_m \leq \tau < \tau_{m+1}$,

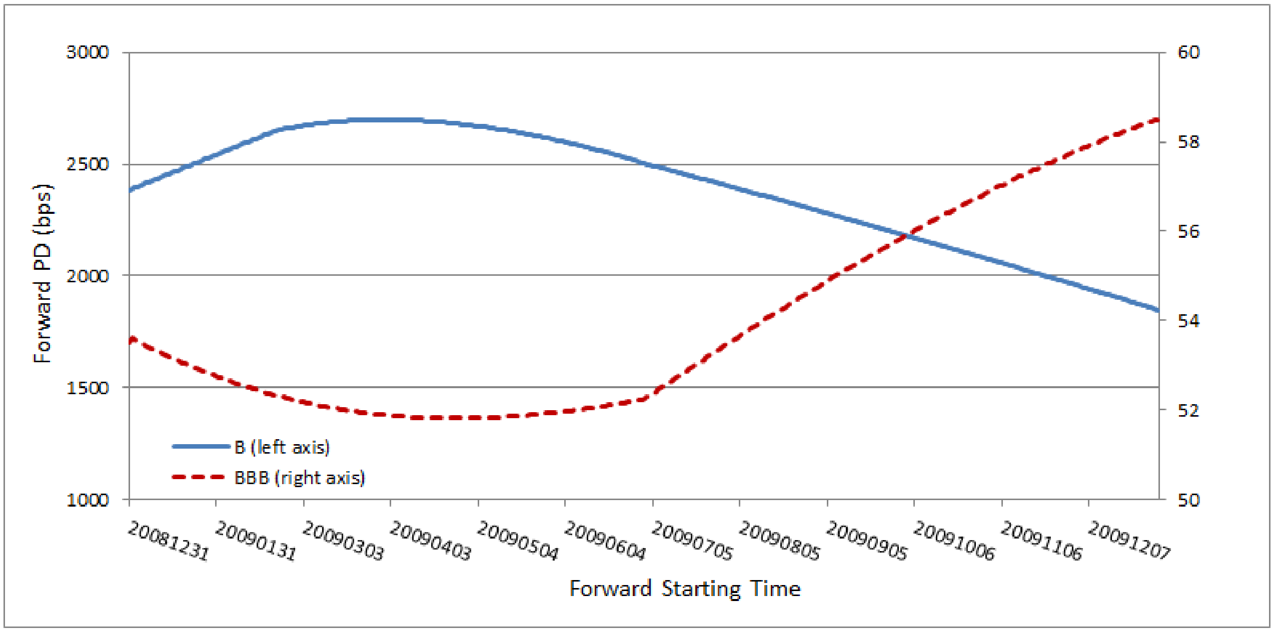
$$\boldsymbol{S}_t(\tau) = \left(\prod_{i=1}^{m} \boldsymbol{exp}_M \left[(\tau_i - \tau_{i-1}) \boldsymbol{F}_t(\tau_{i-1}) \right] \right) \boldsymbol{exp}_M \left[(\tau - \tau_m) \boldsymbol{F}_t(\tau_m) \right] \tag{8}$$

In the above, $\boldsymbol{exp}_M(\cdot)$ denotes the matrix exponential operator. Since $\boldsymbol{F}_t(\tau_i)$'s are not symmetric matrices, their multiplication is not commutative and the order of the multiplication must be respected. The forward generators are easily solvable sequentially over $\tau_1, \tau_2, \cdots$ with the matrix logarithm operator built in programming languages such as Matlab, Julia, etc. With these forward migration generators, one in essence turns the originally discrete-time model at a semiannual frequency, i.e., $\boldsymbol{S}_t(\tau_i), i = 1, 2, \cdots$, into a continuous-time model, with which complex debt tenor structures facing financial institutions can be conveniently handled.

The forward PIT-PD for an obligor in rating cohort k at time t and forward-starting at time $t + \tau$ with a prediction horizon of s, denoted by $f_t(\tau; s, k)$, should be understood as the conditional PD by presuming that the obligor has survived the period of t to $t + \tau$. Note that any defaultable obligor's forward PD would eventually approach zero by definition if the survival probability were not factored in. Recall that $\boldsymbol{S}_t^{(k,i)}(\tau)$ denotes the (k, i) element of matrix $\boldsymbol{S}_t(\tau)$. The following is a simple expression for any forward PIT-PD applicable to all rating cohorts:

$$f_t(\tau; s, k) = \frac{\boldsymbol{S}_t^{(k,K+1)}(\tau + s) - \boldsymbol{S}_t^{(k,K+1)}(\tau)}{\sum_{i=1}^{K} \boldsymbol{S}_t^{(k,i)}(\tau)} \tag{9}$$

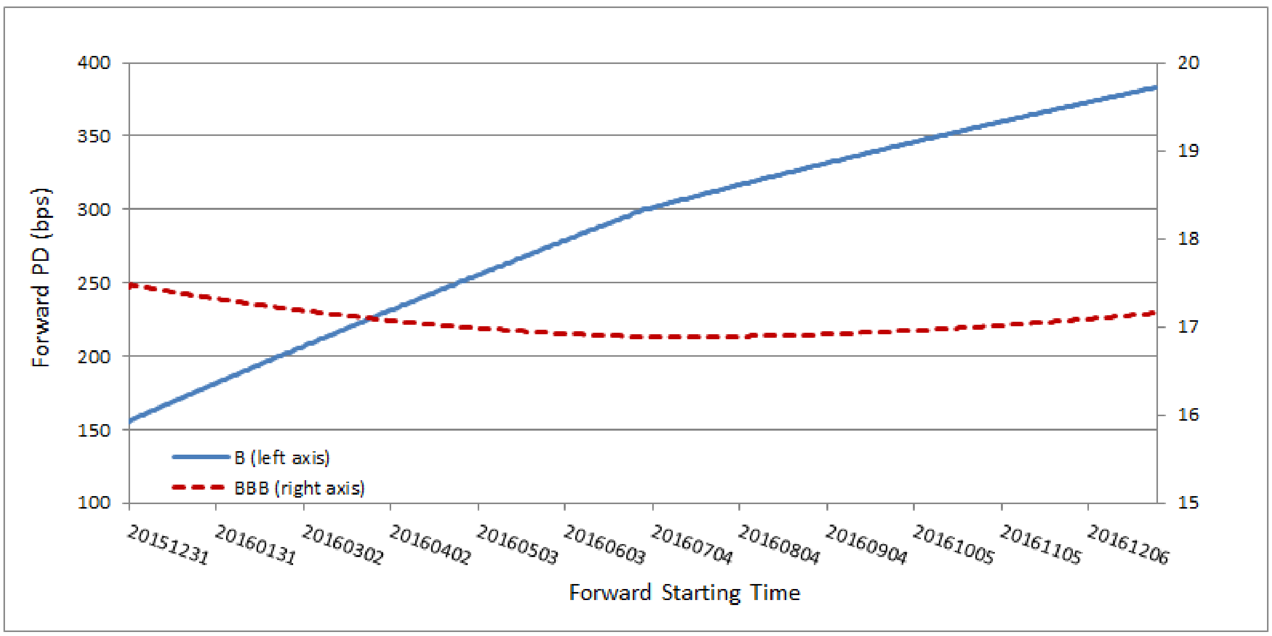
Figure 1: Forward PIT 1-year PDs at the end of December 2008 for BBB and B cohorts for various forward starting times



The forward PIT-PDs implied by our model exhibit a term structure effect much like forward interest rates, and this can be clearly seen in Figures 1 and 2 where 6-month PDs for the BBB and B rating cohorts for a range of forward starting times up to five years are plotted for two time points (December 2008 and December 2015). Note that the forward 1-year PDs for the B rating are referenced to the left vertical axis whereas those for the BBB rating use the right one. Not at all surprising is to find the forward PDs both either rating category are much higher for December 2008 right after the Lehman Brothers' bankruptcy vis-a-vis December 2015 after the credit market has calmed down. The forward PD term structures are complicated with a hump shape for the B rating in December 2008, suggesting further heightening of credit risk for some time to come before eventually calming down, but for the BBB rating, the opposite is true. However in December 2015, the forward PD curve for the B rating has turned from hump-shaped into upward sloping all the way to five years, pointing to higher future credit risks.

A credit cycle is defined as over N periods, which may, for example, be regarded as 10 years. The TTC τ-period rating migration matrix at time t can be understood as $\bar{\boldsymbol{S}}_t(\tau, N) = \frac{1}{N}\sum_{i=1}^{N} E_t[\boldsymbol{S}_{t+i-1}(\tau)]$, which is an average over time of expected future spot PIT migration matrices. The $(K+1)$-th column of $\bar{\boldsymbol{S}}_t(\tau, N)$ provides the corresponding TTC-PDs for

Figure 2: Forward PIT 1-year PDs at the end of December 2015 for BBB and B cohorts for various forward starting times



different cohorts. The TTC-PDs are time-varying to reflect the phase of a credit cycle, but the variations are expected to be smaller due to averaging over the credit cycle. Without a theoretical model, the TTC-PDs are often approximated by the total count of one-year transition from one particular cohort to another divided the total sum of the obligors in the starting cohort over, say, 10 years. Alternatively, one can simply ignore the variation in the number of obligors over differ rating cohorts and over time to just adopt the average τ-period realized default rate computed over, say, 10 years.

We consider a theoretical TTC 1-year credit rating migration based on a 10-year cycle; that is $\bar{\boldsymbol{S}}_t(1, 10)$ where ten non-overlapping 1-year rating migration matrices are averaged to correspond to a 10-year cycle. This theoretical TTC 1-year migration matrix at an evaluation time is forward-looking and computable to any degree of accuracy (increasing the number of simulation paths). We compare them with the backward-looking TTC 1-year migration matrix computed per usual, which is the total count of 1-year transition from one particular cohort to another divided the total sum of the obligors for that starting cohort over the 10-year period immediately prior to the evaluation time.

Table 4 provides a comparison of TTC (10-year cycle) 1-year PDs for all seven S&P rating cohorts under the forward-looking model and backward-looking historical average in December 2008 and December 2015, respectively. Evidently from these results, the forward-looking and backward-looking TTC-PDs are quite different particularly for the lower credit quality categories, and the differences are more pronounced in December 2008 when the credit cycle was at around its peak vs. December 2015 when credit risk had subsided. In order to see the effect due to the length of a credit cycle, we conduct the same comparison except for shortening the presumed cycle length to 5 years. The results in Table 4 show clearly that the differences between the forward-looking and backward-looking TTC-PDs become much larger, a result that can be anticipated because the averaging effect becomes weaker when the cycle length is shortened.

References

[1] Altman, E.I., 1998, The Importance and Subtlety of Credit Rating Migration, *Journal of Banking and Finance* 22, 1231-1247.

[2] Bangia, A., F.X. Diebold, A. Kronimus, C. Schagen and T. Schuermann, 2002, Ratings Migration and the Business Cycle, with Application to Credit Portfolio Stress Testing, *Journal of Banking and Finance* 26, 445-474.

[3] Bickel, P.J. and E. Levina, 2008, Covariance Regularization by Tresholding, *The Annals of Statistics* 36(6), 2577-2604.

Table 4: Forward-looking TTC 1-year PDs vs. historic average TTC 1-year PDs where the PDs are in basis points.

	Forward-looking	Historical	Forward-looking	Historical
Rating	December 2008 (10-year cycle)		December 2015 (10-year cycle)	
AAA	0.02	0.00	0.02	0.00
AA	0.27	2.83	0.23	3.29
A	5.12	10.12	4.70	3.35
BBB	14.86	22.53	13.00	10.37
BB	97.59	106.87	58.52	29.06
B	695.70	430.97	388.73	257.65
CCC/CC/C	3243.05	2541.64	2609.38	2616.82
Rating	December 2008 (5-year cycle)		December 2015 (5-year cycle)	
AAA	0.03	0.00	0.02	0.00
AA	0.33	5.64	0.26	0.00
A	5.72	6.81	4.92	0.00
BBB	17.30	7.26	13.76	1.19
BB	136.28	42.67	61.87	12.05
B	1001.29	176.34	411.86	177.35
CCC/CC/C	4072.00	1629.88	2872.05	2459.97

[4] Cai, T. and W. Liu, 2011, Adaptive Thresholding for Sparse Covariance Matrix Estimation, *Journal of the American Statistical Association* 106(494), 672-684.

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

[6] Del Moral, P., A. Doucet, and A. Jasra, 2006, Sequential Monte Carlo Samplers, *Journal of the Royal Statistical Society: Series B*, 68(3), 411-436.

[7] Duan, J.C., 2016, Local Momentum Autoregression and the Modeling of Interest Rate Term Structure, *Journal of Econometrics* 194(2), 349-359.

[8] Duan, J.C., and A. Fulop, 2015, Density-Tempered Marginalized Sequential Monte Carlo Samplers, *Journal of Business and Economic Statistics* 33(2), 192-202.

[9] Duan, J.C., and W. Miao, 2016, Default Correlations and Large-Portfolio Credit Analysis, *Journal of Business and Economic Statistics* 34(4), 536-546.

[10] Duan, J.C., J. Sun and T. Wang, 2012, Multiperiod Corporate Default Prediction – A Forward Intensity Approach, *Journal of Econometrics* 170(1), 191-209.

[11] Duan, J.C., and W.T. Wang, 2016, Estimating Distance-to-Default with a Sector-Specific Liability Adjustment via Sequential Monte Carlo, National University of Singapore working paper (to appear as a book chapter in "Applied Quantitative Finance" eds. C. Chen, W. Hardle and L. Overbeck, Springer).

[12] Duan, J.C., and C. Zhang, 2016, Non-Gaussian Bridge Sampling with an Application, National University of Singapore working paper.

[13] Feng, D., C. Gourieroux and J. Jasiak, 2008, The Ordered Qualitative Model for Credit Rating Transitions, *Journal of Empirical Finance* 15, 111130.

[14] Frydman, H. and T. Schuermann, 2008, Credit Rating Dynamics and Markov Mixture Models, *Journal of Banking and Finance* 32, 10621075.

[15] Gagliardini, P. and C. Gourieroux, 2005, Stochastic Migration Models with Applications to Corporate Risk, Journal of Financial Econometrics 3(2), 188-226.

[16] Marcucci, J. and M. Quagliariello, 2009, Asymmetric Effects of the Business Cycle on Bank Credit Risk, *Journal of Banking and Finance* 33, 16241635.

[17] JP Morgan, 1997, CreditMetrics™ technical document, J.P. Morgan, New York.

[18] Kadam, A and P. Lenk, 2008, Bayesian Inference for Issuer Heterogeneity in Credit Ratings Migration, *Journal of Banking and Finance* 32, 22672274.

[19] Lando, D and T.M. Skdeberg, 2002, Analyzing Rating Transitions and Rating Drift with Continuous Observations, *Journal of Banking and Finance* 26, 423444.

[20] Lele S.R., B. Dennis, and F. Lutscher, 2007, Data Cloning: Easy Maximum Likelihood Estimation for Complex Ecological Models using Bayesian Markov chain Monte Carlo Methods, *Ecology Letters* 10(7), 551563.

[21] Lele S.R., K. Nadeem, and B. Schmuland, 2010, Estimability and Likelihood Inference for Generalized Linear Mixed Models using Data Cloning. *Journal of the American Statistical Assocication* 105(492), 16171625.

[22] Mahlmann, T., 2006, Estimation of Rating Class Transition Probabilities with Incomplete Data, *Journal of Banking and Finance* 30, 32353256.

[23] NUS-CRI Staff, 2017, NUS-RMI Credit Research Initiative Technical Report Version: 2017 Update 1, the Credit Research Iniatiative, National University of Singapore.

[24] NUS-CRI Staff, 2018, Probability of Default implied Rating (PDiR) White Paper, the Credit Research Iniatiative, National University of Singapore.

[25] Olsen, R.J., 1978, Note on the Uniqueness of the Maximum Likelihood Estimator for the Tobit Model, *Econometrica* 46(5), 1211-1215.

[26] Rothman, A.J., E. Levina and J. Zhu, 2009, Generalized Thresholding of Large Covariance Matrices, *Journal of the American Statistical Association* 104(485), 177-186.

[27] 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.

[28] Zou, H., 2006, The Adaptive Lasso and Its Oracle Properties, *Journal of the American Statistical Association*, 101(476), 1418-1429.