Analysis of Default Data Using Hidden Markov Models

Mark H. Davis
mark.davis@imperial.ac.uk

Co-authors: Giacomo Giampieri & Martin Crowder

Department of Mathematics, Imperial College London

16 April 2004

Second International Conference on Credit Risk - Montréal - April 15-16, 2004

Summary

- Models of default interaction
- Hidden Markov model
- Statistical analysis of default data (Standard & Poor's database)
- Further developments

Default interaction modelling

- Latent variable models
- Mixture models
- Contagion models

(See Frey and McNeill)

Latent variable models

A latent variable model involves an underlying random vector X and a threshold vector d. Obligor i defaults if $X_i < d_i$. Dependence is introduced through the copula function for X. Examples:

- Merton-style models such as Moody's KMV.
- Li model: X_i is the default time for obligor i, so $d_i = T$ for defaults over a time interval $[0, T]$. Exponential marginals, Gaussian copula.

Mixture models

Here the default probability $p_i(\Psi)$ is assumed to depend on an 'economic factor' vector Ψ. Defaults are conditionally independent given Ψ. Generally no explicit use of copulas. Examples:

- Vasicek 1-factor model: $\Psi, \epsilon_i \sim N(0,1)$, obligor i defaults if $X_i < d_i$ where
$$X_i = \rho\Psi + \sqrt{1-\rho^2}\epsilon_i.$$

- CreditRisk+ model: $P[i \text{ defaults}|\Psi] = 1 - e^{-w_i'\Psi}$ where $\Psi = (\Psi_1, \ldots, \Psi_n)$, and Ψ_j are independent gamma-distributed r.v.s.

Contagion models

In these models, the event of default of obligor i directly affects the default probability (or hazard rate) of other obligors j.

- Davis-Lo, Jarrow-Yu: default of i pushes some or all of other obligors into a 'high risk' state.

- Giesecke-Weber, Neu-Kühn: local or global interactions based on models from statistical physics.

- Schellhorn-Cossin, Davis-Esparragoza: Queueing network models.

Distributions and Hazard Rates

The distinction between these model classes is not clear-cut. To see this, consider a default time $\tau \geq 0$ with density $f(t)$. The *survivor function* G and *distribution function* F are

$$P[\tau > t] = G(t) = 1 - F(t) = \int_t^\infty f(u)du.$$

The *hazard rate* is

$$h(t)dt = \frac{f(t)}{G(t)}dt \approx P[\tau \in]t, t + dt] | \tau > t],$$

and there is a 1-1 relation between h and G in that

$$G(t) = e^{-\int_0^t h(u)du}.$$

For two default times τ_1, τ_2 with joint density $f(t_1, t_2)$, define $\tau_{\min} = \min(\tau_1, \tau_2)$ and $\tau_{\max} = \max(\tau_1, \tau_2)$. The initial hazard rate is

$$h_0(t) = \frac{1}{G(t,t)} \left(\int_t^\infty f(u,t)du + \int_t^\infty f(t,v)dv \right)$$

where $G(s,t) = P[\tau_1 > s, \tau_2 > t]$. If τ_1 occurs first, the post-τ_1 conditional density of τ_2 is

$$\frac{f(\tau_1, t)}{\int_{\tau_1}^\infty f(\tau_1, v)dv},$$

so the hazard rate is

$$h_2(\tau_1, t) = \frac{f(\tau_1, t)}{\int_t^\infty f(\tau_1, v)dv}.$$

In summary, the hazard rate is a stochastic process

$$h(t) = h_0(t)\mathbf{1}_{(t<\tau_{\min})} + [h_1(\tau_2, t)\mathbf{1}_{(\tau_{\min}=\tau_2)} + h_2(\tau_1, t)\mathbf{1}_{\tau_{\min}=\tau_1}]\mathbf{1}_{[\tau_{\min},\tau_{\max}[}(t).$$

The hazard rate thus jumps at a default, but the way in which this happens is only indirectly specified.

Homework problem: formulate the general 2-default case as
(a) a latent variable model and
(b) a mixture model.

Uses of the models

1. **Pricing** (Risk-neutral measure) Requirement is primarily ease of calibration. Pricing = interpolation.

2. **Risk management** (Real-world measure) Requirement is statistical accuracy.

This paper

- Mixture model (no contagion)

- Real-world measure

Contagion models: a paradox

Suppose there are assets (bonds, loans, ...) $A_1, A_2, \ldots$.

Portfolio $\mathrm{A} = \{A_1, \ldots, A_n\}$.

Portfolio $\mathrm{B} = \{A_1, \ldots, A_{n+1}\}$.

Suppose asset A_{n+1} defaults and is removed from Portfolio B. In a contagion model, Portfolio B moves to an 'enhanced risk' state, while Portfolio A remains at normal risk. But these portfolios are the same!

Conclusion: a contagion model should either
(a) model the whole universe of assets, or
(b) include further variables representing the impact of defaults outside the portfolio.

A Hidden Markov Model (HMM)

This is a mixture model in which

- The *risk state* is an *exogenous* stochastic process.
- The risk state cannot be directly observed, but can be inferred through another stochastic process that produces the sequence of observations (default times).
- Our new model is a doubly stochastic model, in the sense that the the number of defaults in a given time period is binomially distributed, with an intensity which varies randomly, according to the exogenous risk state.

- In the simplest case, the *unobservable* underlying 'risk state' can take 2 values ('normal' and 'enhanced', as in Davis-Lo model). The hazard rate λ is enhanced by a factor $\kappa > 1$ during the high-risk periods.

- The hidden state is assumed to be specific to a particular region and sector.

- The dynamics follows a Hidden Markov Model, used in a wide range of signal processing applications (e.g., speech recognition).

Basics of HMM: Box and Ball Model

Box 1 : P_1[red], P_1[green]

Box 2 : P_2[red], P_2[green]

An unknown mechanism (e.g., a third person) first selects the initial box, and from this box a ball is chosen at random. The observer is told the color of this ball, but not the box from which it was selected. The process is repeated, and an observation sequence is created

$$\boldsymbol{O} = R, G, G, R, G, G, R, R, G, R, R, R, G \ldots$$

.

Three basic problems:

1. Given the observation sequence $\boldsymbol{O}$ and the model Γ, compute (efficiently!) the probability $P[\boldsymbol{O}|\Gamma]$

2. Given the observation sequence $\boldsymbol{O}$ and the model Γ, find the 'optimal' hidden sequence ('boxes') which explains the observations

3. Find the model's parameters Γ that maximize $P[\boldsymbol{O}|\Gamma]$

These three problems are linked together under the HMM probabilistic framework, and can be solved by means of the *Forward-Backward procedure.*

Standard & Poor's database

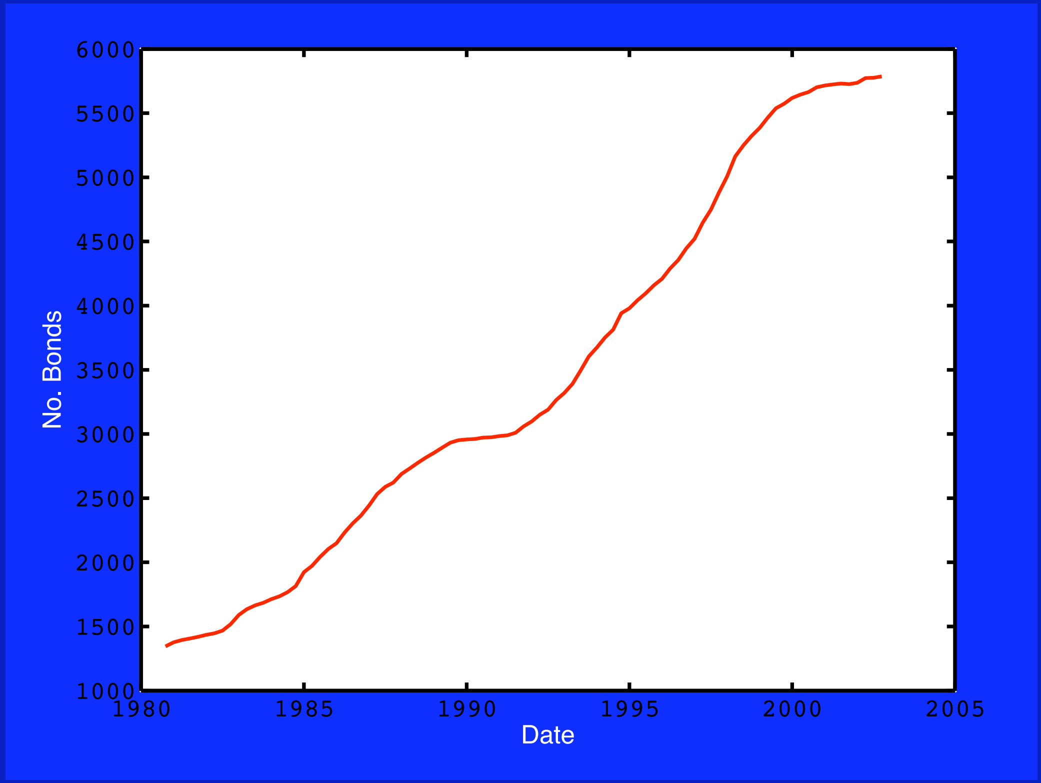
History of 9928 bonds divided in 13 sectors, including

- Automotive
- Consumer / service sector
- Energy and natural resources
- Health care / chemicals
- Leisure time / media
- Transportation

Subsectors and country are also specified.

Database provides date of first rating, and all rating transitions from 1/1/81 to 31/12/02. We have only considered US issues in the largest 4 sectors.

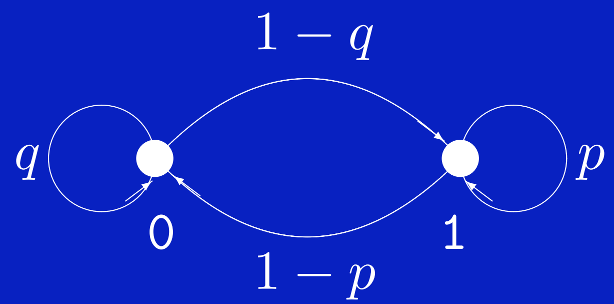
US Bond Population



Details of HMM

- Two states: 0 (normal risk), and 1 (enhanced risk).
- Discrete time (quarterly periods).
- Transition matrix for hidden state:

$$\boldsymbol{a} = \begin{pmatrix} q & 1-q \\ 1-p & p \end{pmatrix}$$



- In state 0, number m of observed defaults in each time step is distributed binomially, with parameter λ:

$$p_0(m) = \binom{N_s}{m} \lambda^m (1-\lambda)^{(N_s - m)}$$

where N_s is the number of issuers at beginning of period.

- In state 1, $p_1(m)$ is binomial with parameter $\kappa\lambda$, with $\kappa \geq 1$.

- *Observed sequence*: m_t = number of defaults in each quarter (~ 90 periods considered).

Estimation of parameters: Forward-Backward (aka Baum-Welch) algorithm

Model parameters are estimated by standard Maximum Likelihood techniques. Involves computation of two different probability terms:

- Forward path probability

$$\alpha_t(i) = P[m_1 m_2 \dots m_t, i]$$

– joint probability of having generated a partial observation sequence in the forward direction (i.e., from the start of the sequence) and having arrived at a certain state i at time t.

- Backward path probability

$$\beta_t(i) = P[m_{t+1} \dots m_T | i]$$

– probability of generating a partial observation sequence in the reverse direction (from the final time T), given that the state sequence starts from a certain state i at time t.

- The probability $\gamma_t(i) = P[i|\boldsymbol{O}]$ of being in a given state i at time t, given the whole observation sequence $\boldsymbol{O} = m_1, \dots, m_T$, is given by

$$\gamma_t(i) = \frac{\alpha_t(i)\beta_t(i)}{\sum_{j=0}^{1} \alpha_t(j)\beta_t(j)}$$

Iterative procedure

- We start from some initial guess for the model's parameters

$$\Gamma = (\lambda, \kappa, q, p)$$

- The parameters of the HMMs can be reestimated by using the probabilities α, β, γ.

  For example, new q and p are the diagonal components of the matrix

$$\bar{a}_{ij} = \frac{\text{expected number of transitions from state } i \text{ to state } j}{\text{expected number of transitions from state } i} = F[\alpha, \beta, \gamma]$$

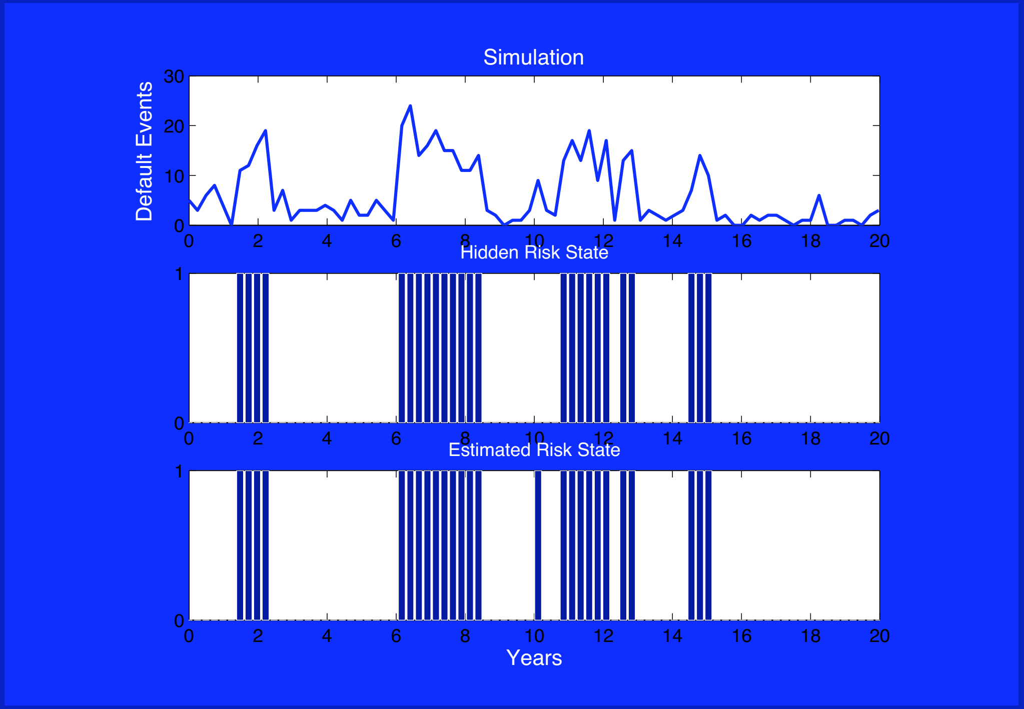
- It can be shown that the new model $\bar{\Gamma}$ is more likely than the original one Γ, in the sense that

$$P[\boldsymbol{O}|\bar{\Gamma}] \geq P[\boldsymbol{O}|\Gamma]$$

- Before applying this procedure to real data, we have run a simulation. We have considered a HMM model with the following characteristics:
  - $N = 1000$ bonds, issued at time $t = 0$
  - $T = 20$ years of observations, at 90 days intervals
  - $\lambda = 0.004$
  - $\kappa = 5$
  - $q = p = 0.9$

Results of the simulation

Starting from the initial guess $\lambda = 0.001, \kappa = 2, q = p = 0.5$, after only five iterations we find $\lambda = 0.0038, \kappa = 4.6727, q = 0.9075, p = 0.9043$.



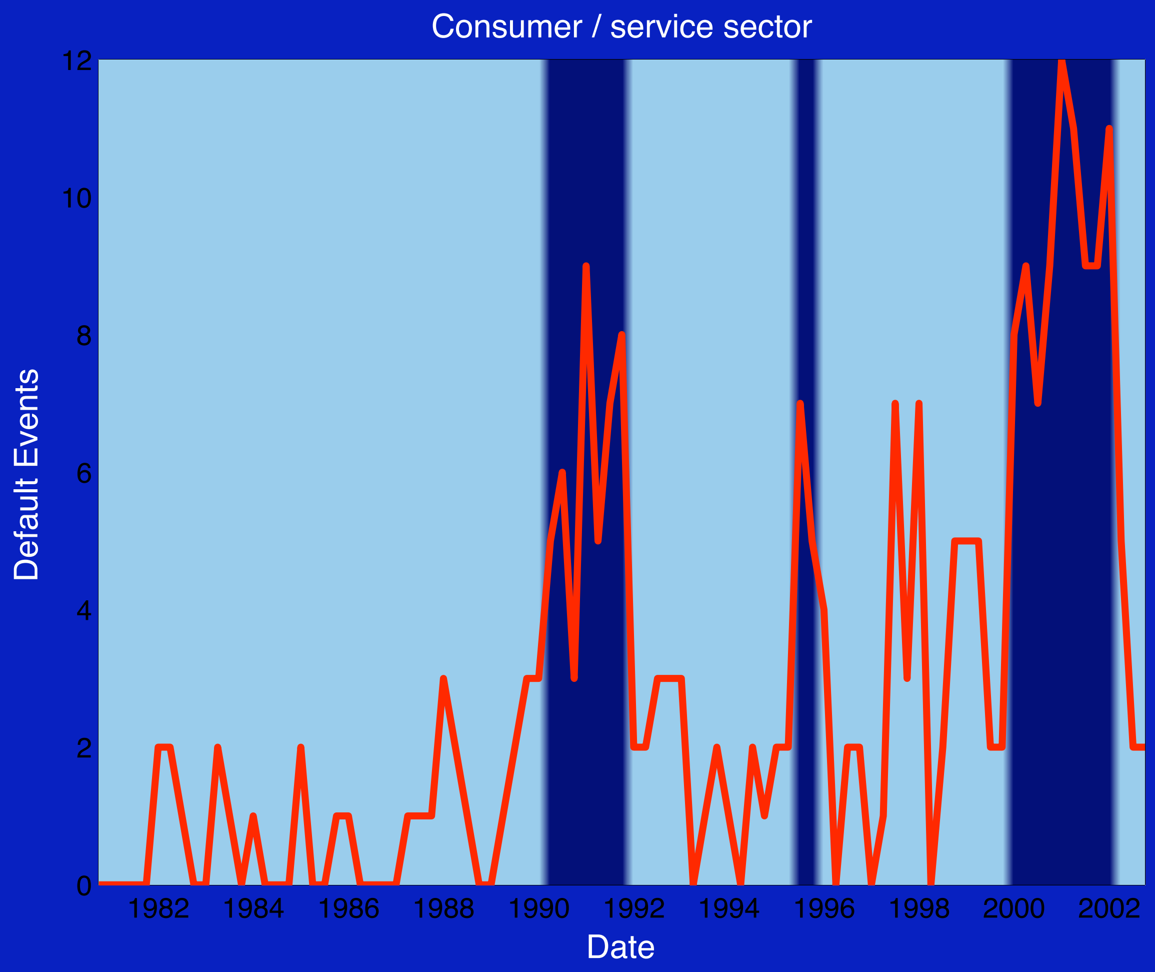
Results from S&P's CreditPro database

- We have considered the actual time sequence of default events, as provided by Standard & Poor's for the following US sectors: consumer, energy, media, and transport.
- The database provides also the time when each bond was first rated. These times define a deterministic birth process, which we have included in our HMM model.
- Default events were grouped in quarterly periods, and the Forward-Backward procedure was applied.

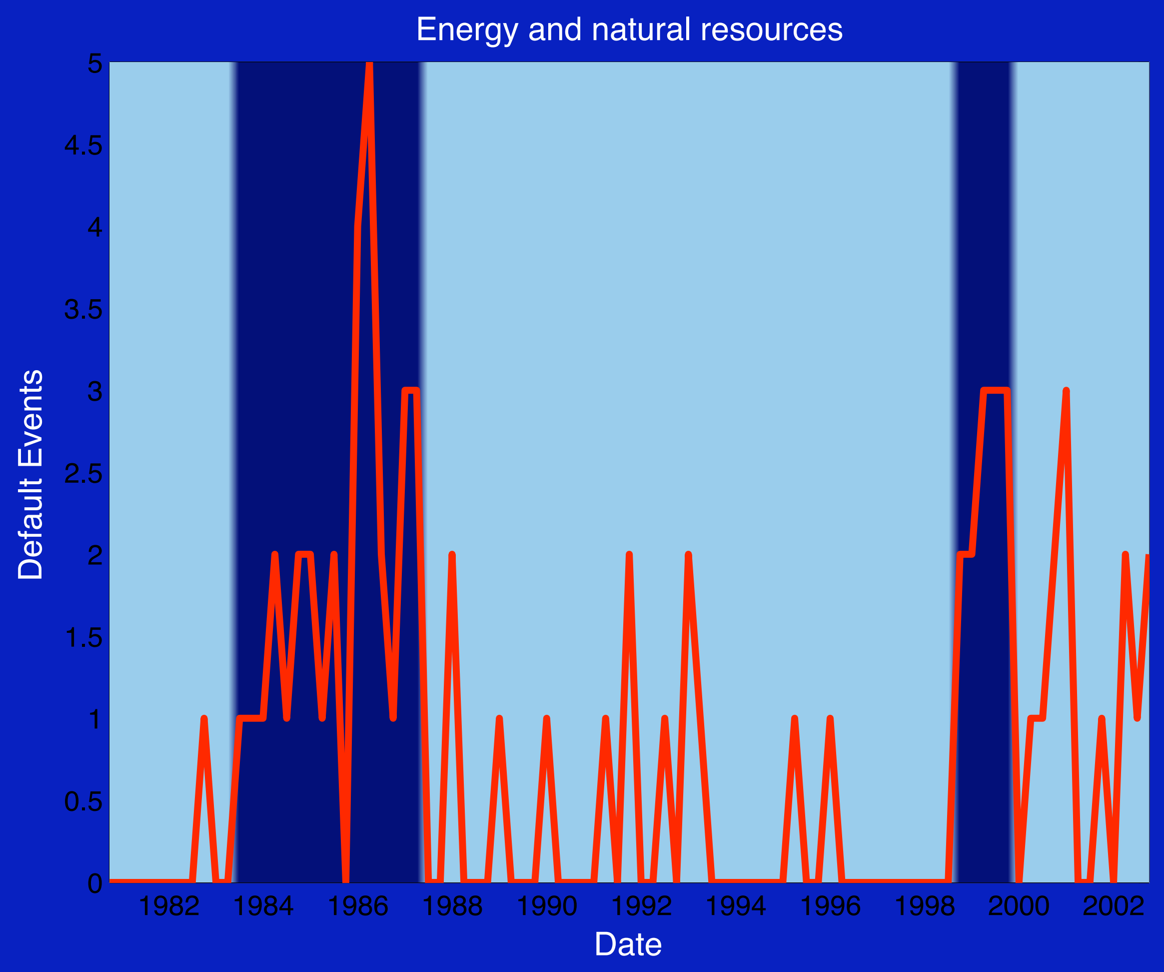
Results from S&P's database - US issuers

Sector	N_{tot}	N_{def}	λ	κ	q	p
consumer	1041	251	0.0026	6.1	0.95	0.81
energy	420	71	0.0014	7.1	0.95	0.88
media	650	133	0.0027	7.2	0.96	0.83
transport	281	59	0.0025	8.9	0.97	0.78

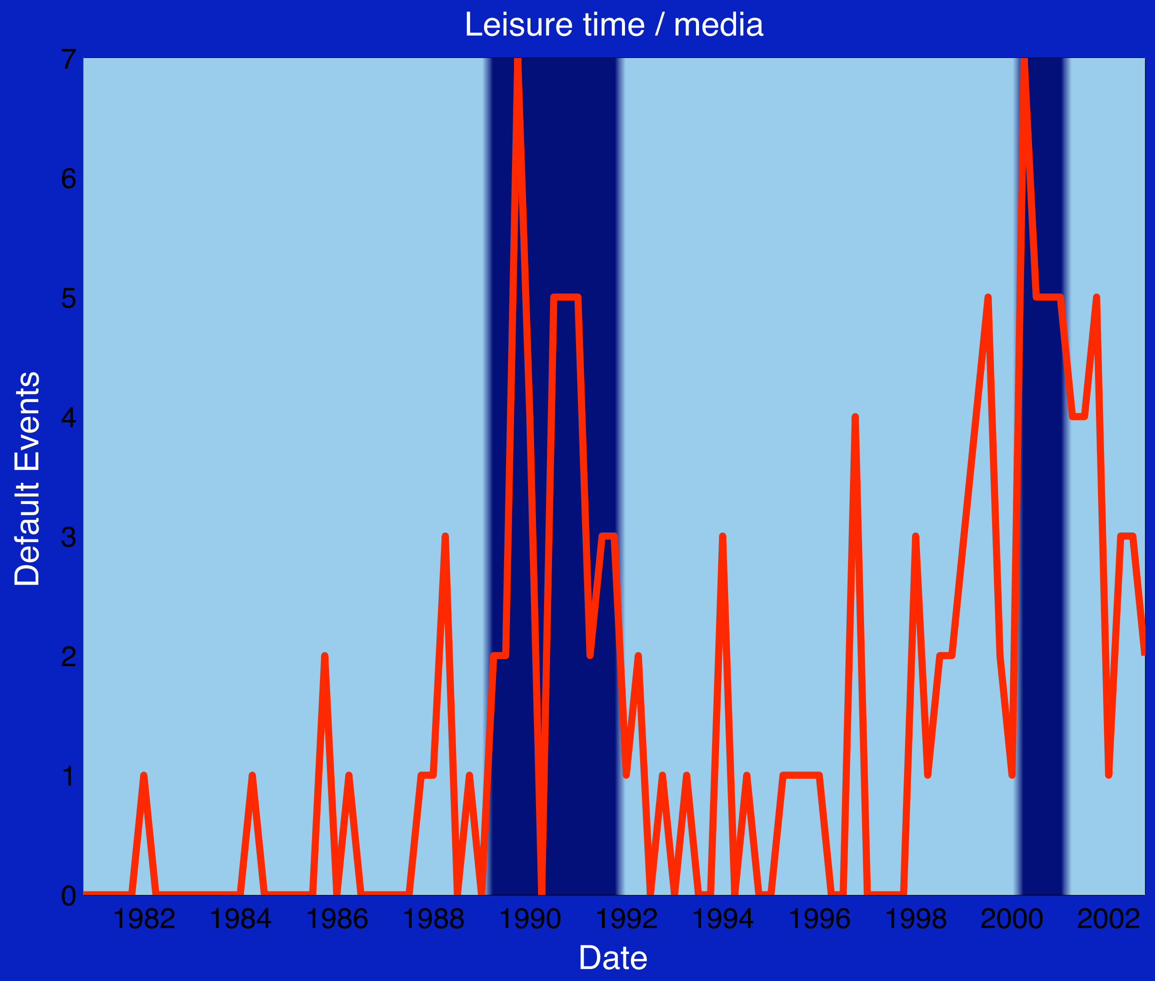
US Consumer sector



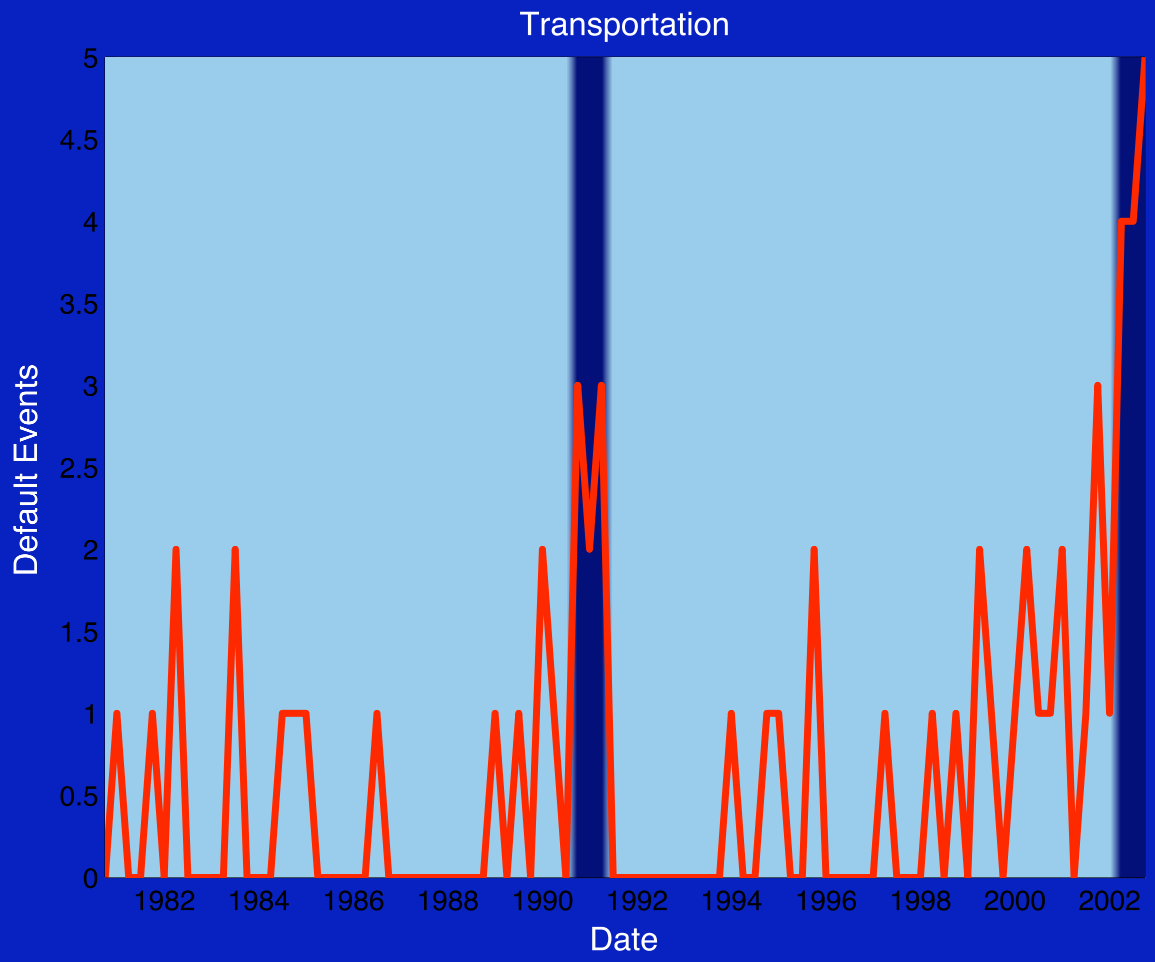
US Energy sector



US Media sector



US Transport sector



Parametric Bootstrap Technique

The covariance matrices for the four parameters found with the ML estimator can be estimated via a parametric bootstrap technique. We focus on the US consumer sector. The bootstrap technique is as follows

- Simulate a large (e.g., $N = 50$) number of realization of the fitted HMM.
- For each generated realization, estimate the four parameters using Baum-Welch.
- Each result is stored in a vector θ_i, and the covariance matrix estimator is then given by $(\hat{\theta} \equiv (1/N) \sum_{i=1}^{N} \theta_i)$

$$\boldsymbol{C} = \frac{1}{N-1} \sum_{i=1}^{N} \left(\theta_i - \hat{\theta}\right)' \cdot \left(\theta_i - \hat{\theta}\right)$$

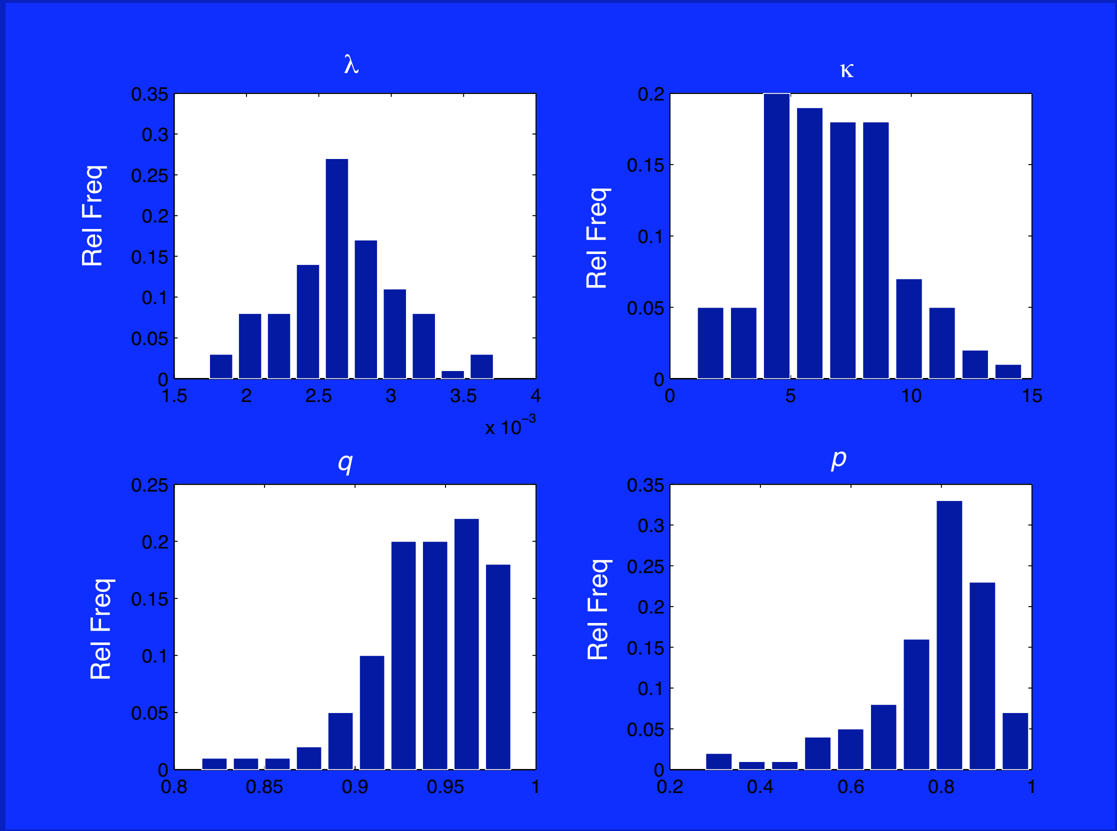
For the US consumer case we find:

$$\mathbf{C} = \begin{pmatrix} \text{8.8E-8} & \text{-2.0E-4} & \text{1.9E-6} & \text{2.4E-6} \\ \text{-2.0E-4} & \text{2.4E0} & \text{9.4E-3} & \text{3.5E-2} \\ \text{1.9E-6} & \text{9.4E-3} & \text{1.8E-3} & \text{-6.2E-4} \\ \text{2.4E-6} & \text{3.5E-2} & \text{-6.2E-4} & \text{3.7E-2} \end{pmatrix}$$

Assuming gaussian distribution around the mean, the square root of each diagonal component gives standard deviation of parameter, thus

$$\begin{aligned} \lambda &= 0.0019 \pm 0.0003 \\ \kappa &= 6.2 \pm 1.5 \\ q &= 0.93 \pm 0.04 \\ p &= 0.80 \pm 0.19 \end{aligned}$$

Frequency distribution of estimated parameters



Test of the Binomial distribution

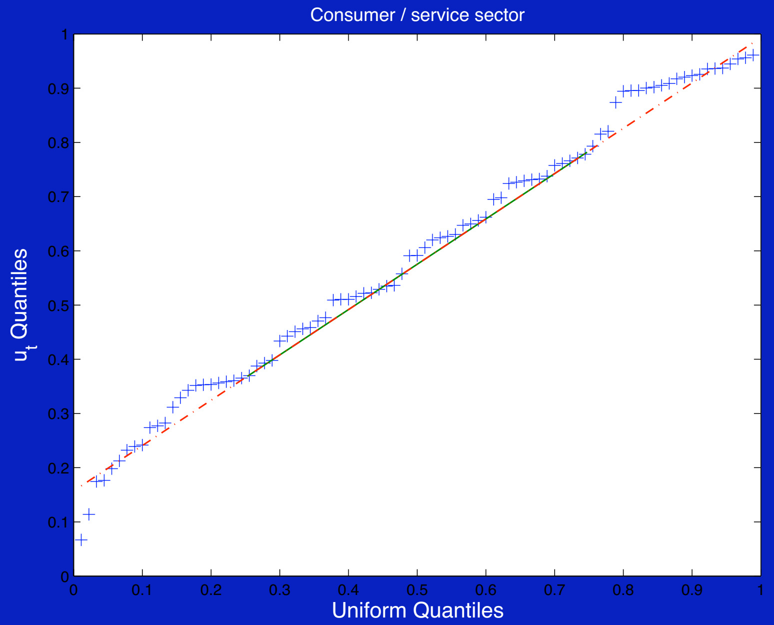
According to the model, at each time step the number m of defaults is distributed according to the binomial p.d.f. $p_0(m)$ or $p_1(m)$, depending on the risk state. We can now ask whether this is supported by the data.

- Compute the conditional distribution of the observed values, given all preceding observations

$$F_t(m) = P(m_t \leq m | m_s, s < t)$$

- If the m_t were continuous random variables, the transformed variates $u_t = F_t(m)$ would form a random sample from the uniform distribution on $[0, 1]$.

- A quantile-quantile plot of the ordered u_t vs. uniform quantiles should follow a linear trend.



- The plot follows the expected line fairly well except at the lower end (effect of the discrete-continuous approximation near $m = 0$).

Global effects

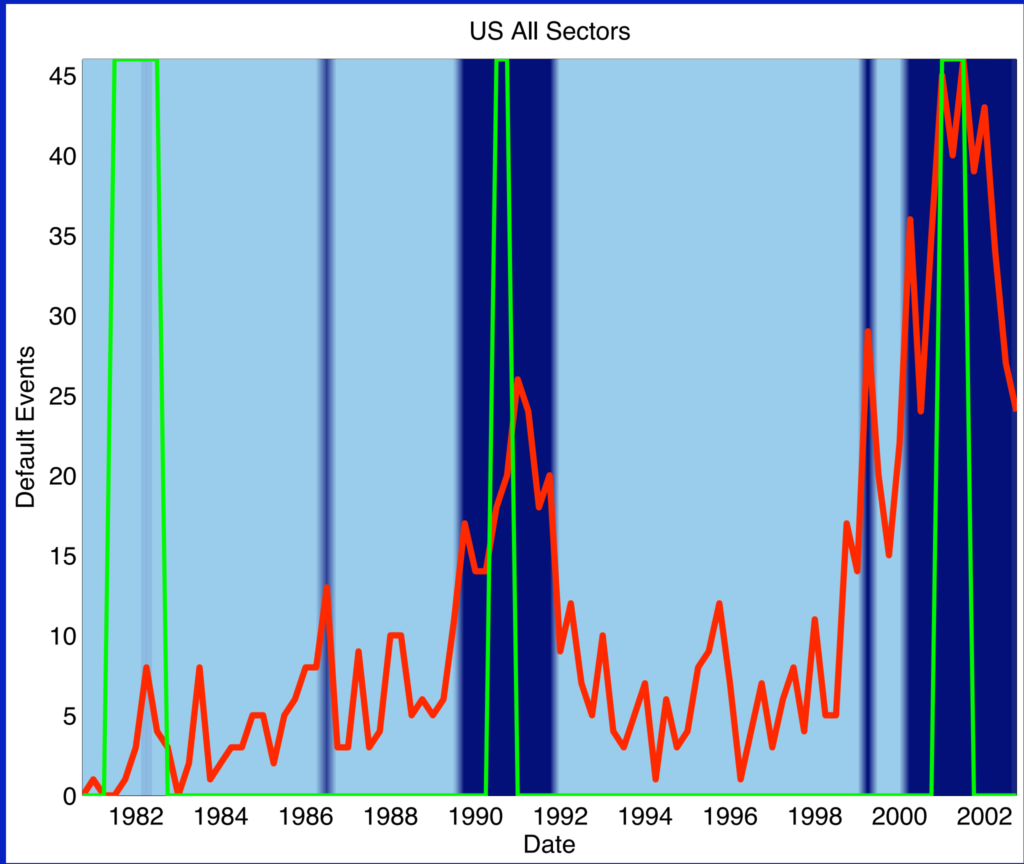
- Any portfolio should consider risks associated with macroeconomic factors along with sector-specific risk factors.

- Our model allows, in principle, disentangling the two risk components.

- We have applied our HMM to the whole databases (US issuers only). Total issues: 6775. Total defaults 1013. Baum-Welch parameter estimates:

$$
\begin{aligned}
\lambda &= 0.0017 \\
\kappa &= 4.9 \\
q &= 0.94 \\
p &= 0.85
\end{aligned}
$$

Correlation with US Business Cycle



Note: Recession periods from NBER website (http://www.nber.org/cycles/cyclesmain.html)

Global vs. Sector-specific risk

- One may be able to disentangle, a posteriori, the default risks specific to a particular sector from those associated with the global economy.

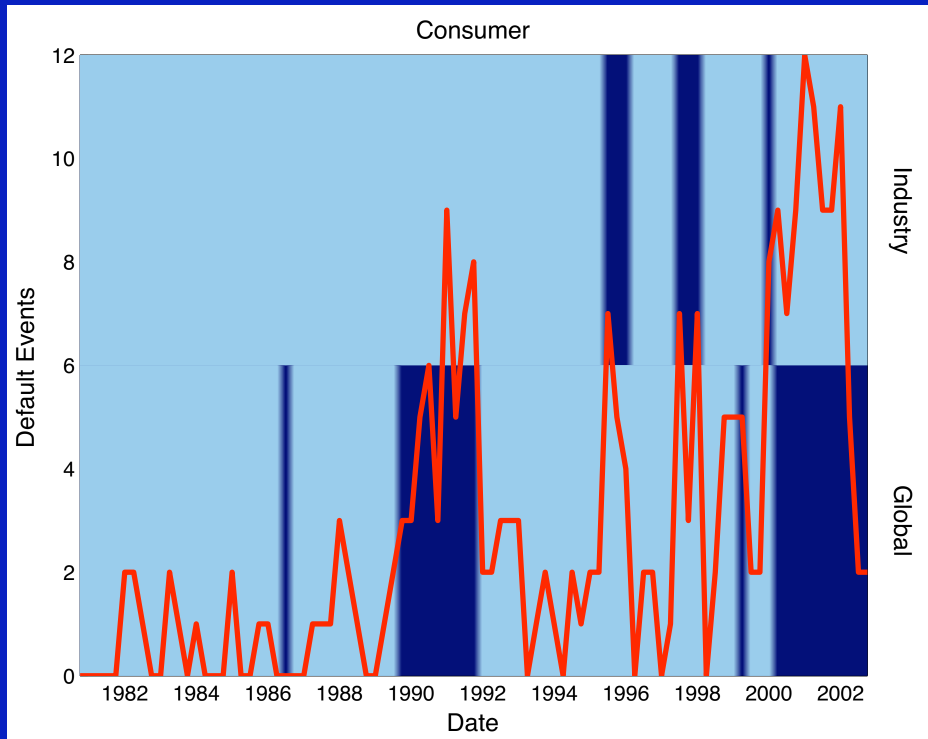
- The two hidden sequences are, by definition, uncorrelated.

- The parameters κ and κ' determine the enhancement factor for the intensity associated with each of the two risk factors, respectively.

First example: US Consumer sector

The Baum-Welch algorithm applied to the modified model gives

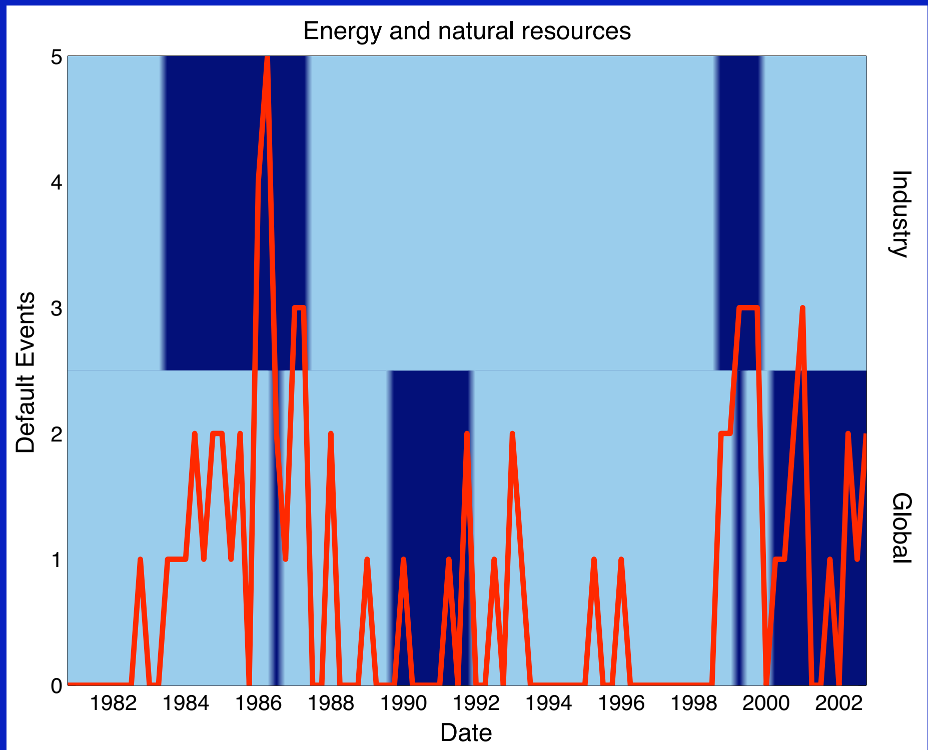
$\lambda = 0.0025, \kappa = 5.2, q = 0.95, p = 0.22, \kappa' = 4.5.$



Second example: US Energy sector

In this case we find

$\lambda = 0.0014, \kappa = 7.2, q = 0.95, p = 0.88, \kappa' = 1.0$



Extended HMM

- We are now developing a more sophisticated model which consider interaction effects in rating, as rating transitions are included in S&P's database.

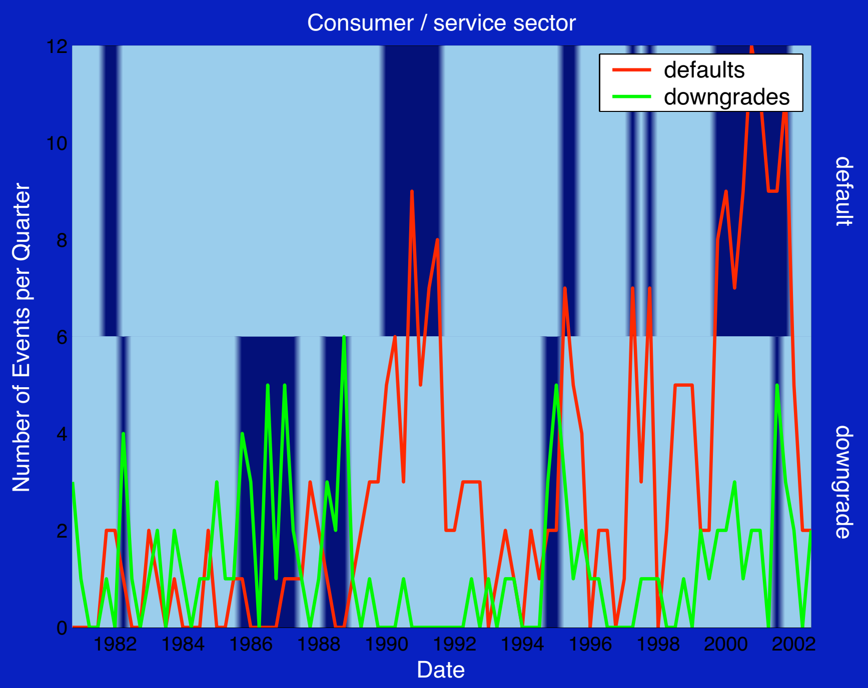
- We have divided all ratings into 3 classes:
  - Investment Grade (BBB- and above)
  - Speculative Grade (BB+ and below)
  - Defaulted (D)

- The underlying idea is that an enhanced risk state not only increases the chances of default, as in the previous model, but also affects the possibility of being downgraded. Coherently, the probability of upgrading is expected to be lower during an enhanced risk period.

- We have found that at least 2 hidden risk states are needed, since periods of high downgrade probability do not necessarily coincide with periods of high default probability.

Model's Result for US Consumer sector

The hidden sequences in the background represent the risk state for defaults (top half) and for downgrades (bottom half).



Conclusions

- Simple 2-state HMM seems surprisingly successful in explaining default interaction.
- We have a tool for disentangling global and industry-specific factors.
- Clear from the data that the business cycle is the key factor determining default interaction.
- The business cycle effect varies widely across different industry sectors.
- A more sophisticated model also considers interaction effects in rating.