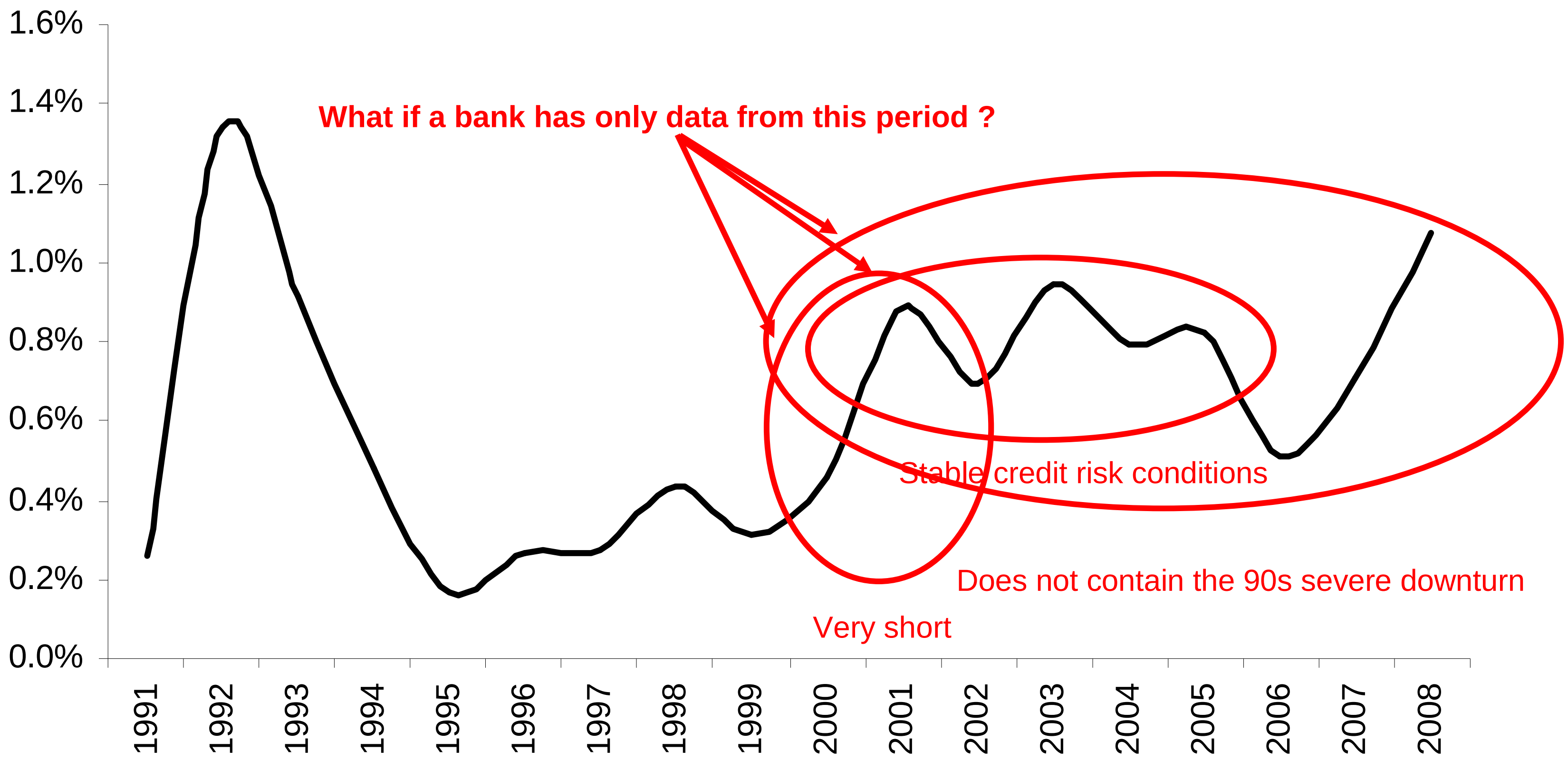
Corporate loan PD modelling using external data

Dimitrios Papanastasiou, Kah Guan Mok

The opinions expressed in this presentation are those of the authors and do not necessarily reflect views of Lloyds Banking Group.

Corporate lending banks usually have limited internal data covering a short period of the credit cycle. Therefore the use of an external dataset of adequate length becomes a necessity.

Medium to Large (>50m turnover) Companies Insolvency Rate



An external data set of sufficient length should be used, appropriately weighted to approximate the internal portfolio at hand

To illustrate the methodology, we select an external-market portfolio, consisting of medium to large UK companies, and a hypothetical internal portfolio as a risky subset of the market

LLOYDS BANKING GROUP

- Data
  - Companies House Data sourced from Jordan's database
  - ONS macroeconomic data
- External Portfolio
  - Medium to large UK companies
  - £50m turnover cut off point as a proxy of size classification
  - Insolvency related events to proxy Basel II definition of default
- Internal Portfolio
  - Constant number of companies over a period of 5 years, 2000-2004
  - The majority of companies are retained over subsequent years
  - Portfolio risk profile does not exhibit extreme movements
  - Riskier than external portfolio
  - Drawn from the external portfolio and consists of companies with Registration Numbers from 4501 to 5500
  - Same definition of default with external portfolio

Logistic regression is used as the basis to form the rating system. The final score is derived by calibrating to 2004 default experience

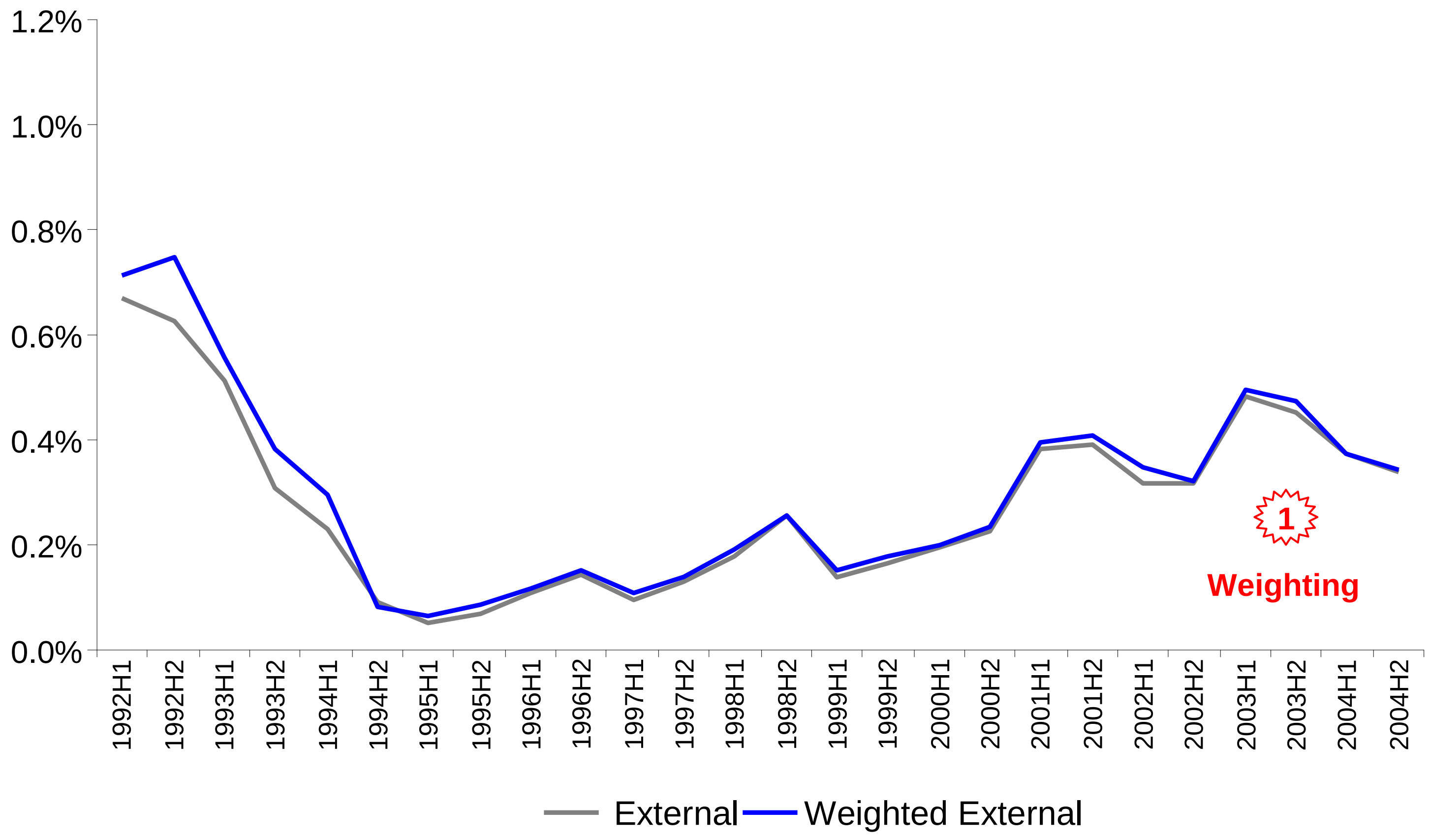
- PD scoring system:
  -
$$D_i \mid X_i \sim Bernoulli(PD_i)$$
$$Log\left(\frac{PD_i}{1-PD_i}\right) = Z_i = \mathbf{b} \cdot X_i$$
$$Z_i = \underset{(0.5264)}{1.4735} - \underset{(0.0679)}{0.3154}CR - \underset{(0.0137)}{0.0451}IC - \underset{(0.0281)}{0.1992}NW - \underset{(0.4261)}{2.5292}ET$$
  - Sommers' D: 35.2%
  - Calibration to 2004 default experience: $Z_i^{cal} = -\underset{(0.2356)}{2.0501} + \underset{(0.1413)}{0.5468}Z_i$
- Rating Grade Scale:

Rating Grade	Scores			PDs		
	Low	Mid Point	High	Low	Mid Point	High
1	-Inf	-5.03	-4.82	0%	0.65%	0.80%
2	-4.82	-4.59	-4.37	0.80%	1.01%	1.25%
3	-4.37	-4.15	-3.93	1.25%	1.55%	1.93%
4	-3.93	-3.71	-3.49	1.93%	2.39%	2.96%
5	-3.49	-3.27	-3.05	2.96%	3.67%	4.53%
6	-3.05	-2.83	-2.61	4.53%	5.58%	6.86%
7	-2.61	-2.39	-2.17	6.86%	8.41%	10.27%
8	-2.17	-1.95	-1.73	10.27%	12.48%	15.09%
9	-1.73	-1.51	-1.29	15.09%	18.13%	21.63%
10	-1.29	-1.07	-0.85	21.63%	25.59%	30%
11	-0.85	-0.41	Inf	30%	40%	100%

The implied historical time series for the internal portfolio default rates is derived by weighting the external dataset and adjusting for the difference in performance of the underlying rating model between the two datasets.

LLOYDS BANKING GROUP

Implied Internal Historical Time Series of Default Rates



The transition matrix is assumed to follow a 1st order Markov process and events to be conditionally independent

- Transition matrix is assumed to evolve as a time inhomogeneous 1st order Markov process and can be decomposed as follows:

$$\mathbf{P}^{trans}_{t} = \begin{bmatrix} & TR & ABS \\ TR & \begin{bmatrix} P_{1,1,t} & ..... & P_{1,11,t} \\ .... & .... & .... \\ P_{11,1,t} & ..... & P_{11,11,t} \end{bmatrix} & \begin{bmatrix} PD_{1,t} \\ .... \\ PD_{11,t} \end{bmatrix} \\ & ABS & ABS \\ ABS & \begin{bmatrix} 0 & .... & 0 \end{bmatrix} & 1 \end{bmatrix}$$

- Rows should sum to unity for $\mathbf{P}^{trans}_{t}$ to be a transition matrix
- Both transitory (rating migrations) and absorbing (PDs) parts are assumed as conditionally independent events given the set of available information at time *t*

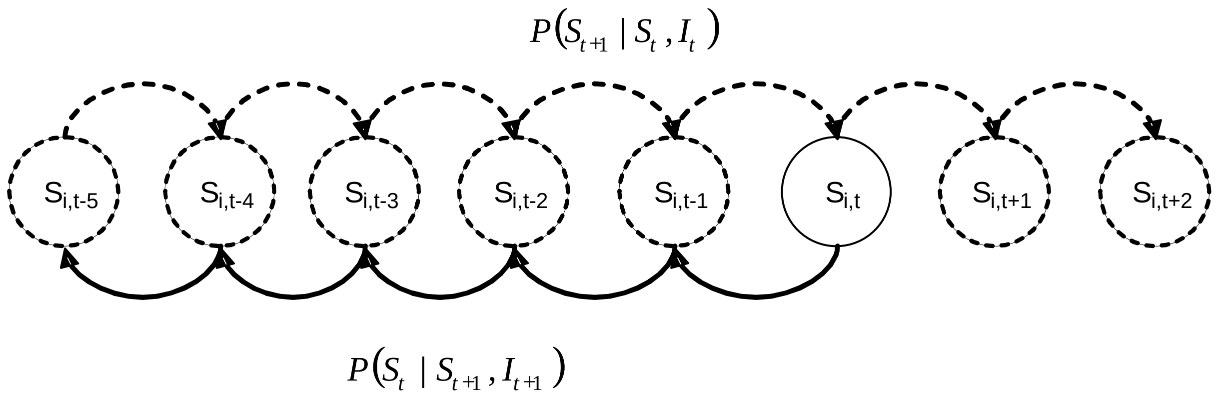
$$\Downarrow$$

$$T_{ij,t+1}\left(D_{i,t+1}\right) \mid S_t = i, I_{ij,t} \sim Binom\left(P^{trans}_{ij,t}\left(PD_{i,t}\right), N_{i,t}\right), \ \forall\, i \neq j$$

- Transitory and absorbing parts are modelled separately

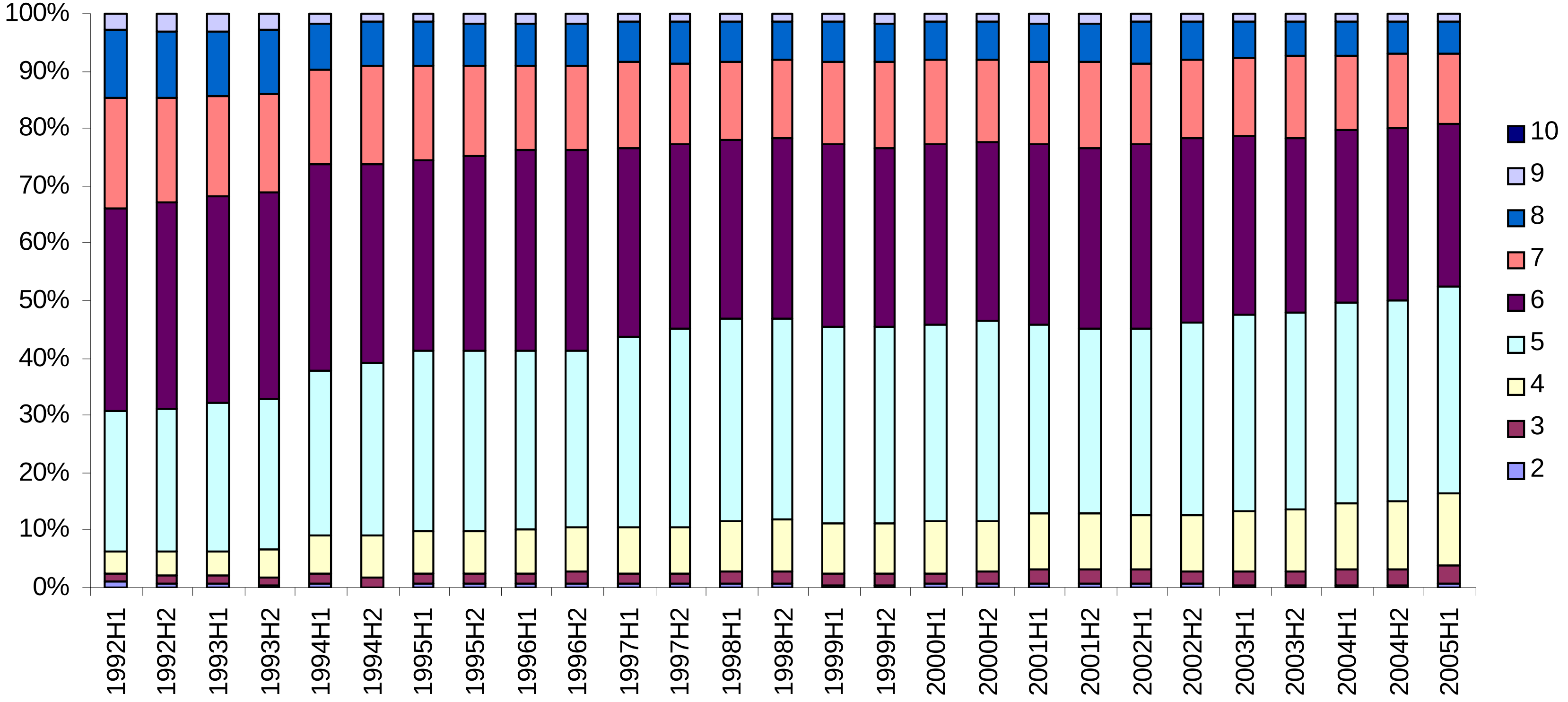
Migration dynamics are assumed the same between internal and external portfolios and weighting is performed by using the backward and forward transition probabilities calculated from the external dataset

- We assume both forward and backward migration chains are common across the two portfolios:
  - Rating Grades are homogeneous
  - Internal portfolio is a subset of the external
  - Internal portfolio is large enough to represent a significant proportion of the market
- Implied internal rating distribution time series:
  - Using the probabilities $P(S_t \mid S_{t+1}, I_{t+1})$ calculated from the external dataset, move back in time from the current internal portfolio distribution to get the historical rating distribution
  - Using the probabilities $P(S_{t+1} \mid S_t, I_t)$ calculated from the external dataset, move forward in time to get the forecasted rating distribution



The rating system represents a TTC discrimination of credit quality

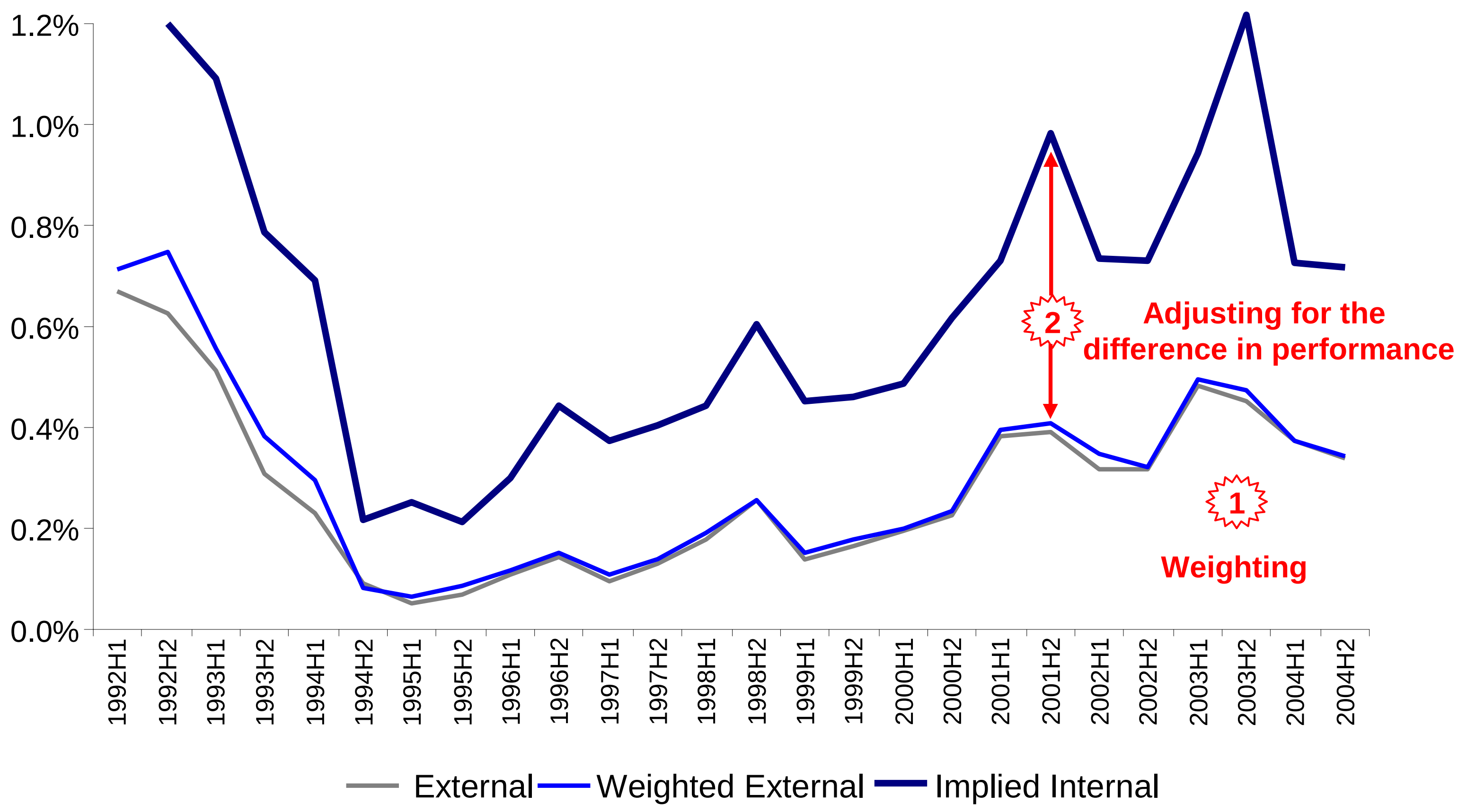
- The implied historical rating distribution:
  - Is relatively stable
  - Does not seem to be responsive to credit cycle (more TTC rating system)



The implied historical time series for the internal portfolio default rates is derived by weighting the external dataset and adjusting for the difference in performance of the underlying rating model between the two datasets.

LLOYDS BANKING GROUP

Implied Internal Historical Time Series of Default Rates



The scoring model's discriminatory power is different between the two portfolios and an adjustment is needed to align the default rates. The adjustment is performed on the log odds scale

- The default rates between the two portfolios should be the same, if:
  - Rating Grades are homogeneous
  - (Conditionally) independent defaults
  - Same discriminatory power of the rating system for both portfolios
- Rating system performs better for the internal portfolio:
  - Rating system is built based on the internal portfolio and it better discriminates default risk for companies that belong to this portfolio
  - Score homogeneity does not imply default risk homogeneity for each rating grade
- Adjustment to align the default rates:
  - Based on overlapping period (2000-2004)
  - Intercept and slope adjustment in log odds scale
  - Adjustments for grades 4-8 (grades with defaults for the internal portfolio):

$$\begin{bmatrix} LogOdds_4^{Int} \\ LogOdds_5^{Int} \\ LogOdds_6^{Int} \\ LogOdds_7^{Int} \\ LogOdds_8^{Int} \end{bmatrix} = \begin{bmatrix} -5.32323 \\ 5.467902 \\ -0.76372 \\ -1.40794 \\ -2.70974 \end{bmatrix} + \begin{bmatrix} 0.004149 \\ 1.728268 \\ 0.720195 \\ 0.550106 \\ 0.262389 \end{bmatrix} \begin{bmatrix} LogOdds_4^{Ext} \\ LogOdds_5^{Ext} \\ LogOdds_6^{Ext} \\ LogOdds_7^{Ext} \\ LogOdds_8^{Ext} \end{bmatrix}'$$

To take account of the complicated nature of the default process, a unobserved frailty factor is introduced to capture correlation that is not attributed to the macroeconomic environment

LLOYDS BANKING GROUP

- Conditionally on the information up to time *t*, defaults are independent.
- The information set consists of both observed and unobserved factors
- Default is not linked directly to observed macroeconomic variables and a latent factor is introduced to take account of the residual correlation
- Model specification:

$$D_{k,t} \mid X_t, u_t \sim Binomial\left(PD_{k,t}, N_{k,t}\right)$$

$$Log\left(\frac{PD_{k,t}}{1-PD_{k,t}}\right) = a_k + \mathbf{b}_k \cdot X_t + u_t$$

$$u_t = \varphi \cdot u_{t-1} + e_t \quad , \; e_t \sim N\left(0, \sigma^2\right)$$

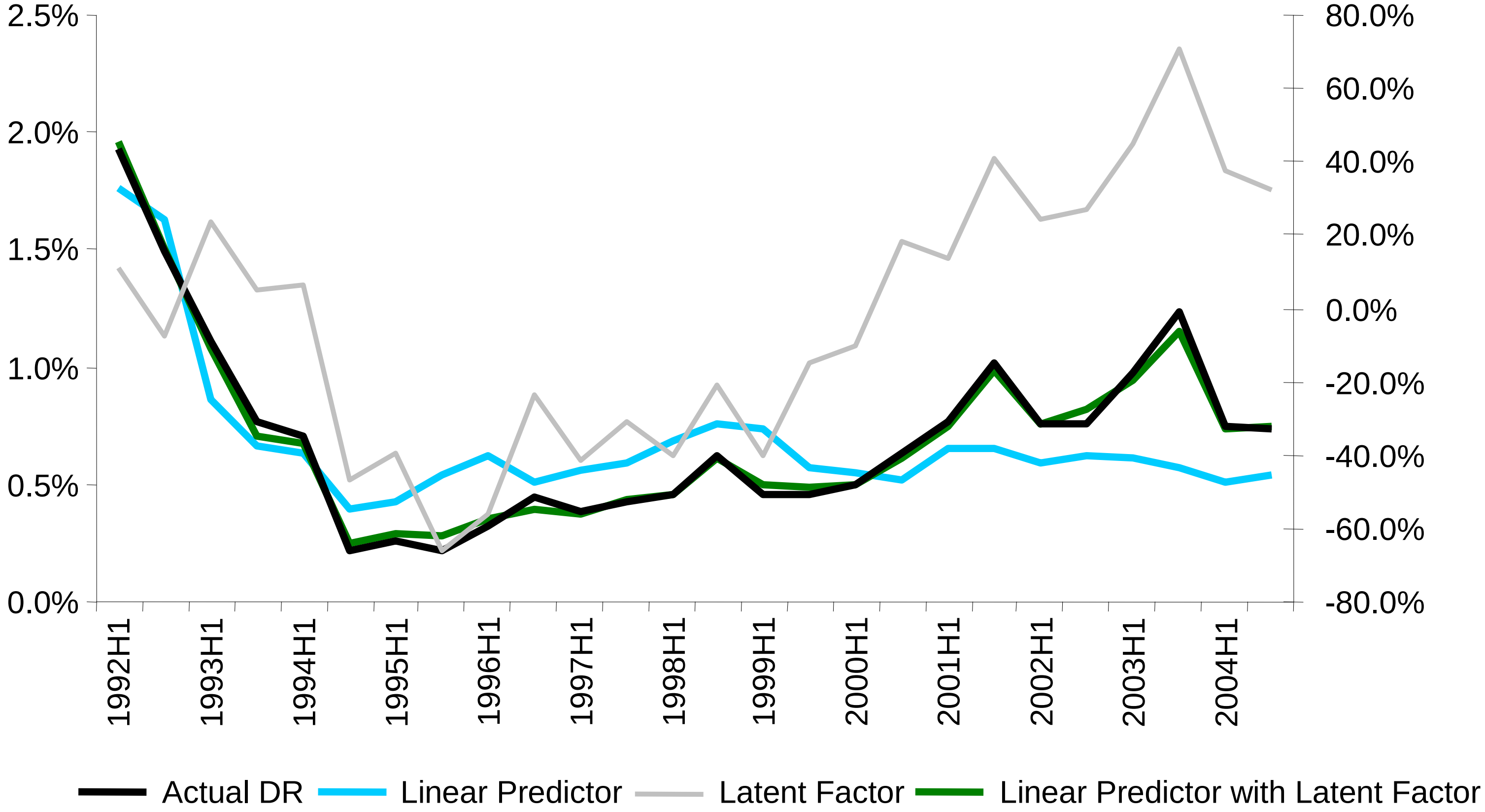
- Model estimates:

$$\begin{bmatrix} PD_{4,t} \\ PD_{5,t} \\ PD_{6,t} \\ PD_{7,t} \\ PD_{8,t} \\ PD_{9,t} \end{bmatrix} = F\left(\begin{bmatrix} \underset{(0.4906)}{-5.8163} \\ \underset{(0.2135)}{-4.2848} \\ \underset{(0.2130)}{-4.5254} \\ \underset{(0.1974)}{-4.2434} \\ \underset{(0.2006)}{-4.0643} \\ \underset{(0.2861)}{-3.9467} \end{bmatrix} + \begin{bmatrix} \underset{(22.8838)}{-81.4683} & \underset{(26.1877)}{-99.4897} \\ \underset{(4.4073)}{-47.7586} & \underset{(8.2913)}{42.1413} \\ \underset{(3.9124)}{-18.6253} & \underset{(6.7903)}{15.3668} \\ 0 & \underset{(5.8277)}{10.1551} \\ 0 & \underset{(5.8277)}{10.1551} \\ \underset{(9.5717)}{-33.5506} & \underset{(5.8277)}{10.1551} \end{bmatrix} \cdot \begin{bmatrix} \text{Trailing GDP Growth}_{t\text{-}1} \\ \text{Short - Long Spread}_{t\text{-}1} \end{bmatrix} + u_t\right)$$

$$u_t = \underset{(0.1758)}{0.7351} \cdot u_{t-1} + e_t \quad e_t \sim N\left(0, \underset{(0.1071)}{0.1708}\right)$$

PD forecasting model consists of equally important observed and unobserved components, each capturing different credit cycle dynamics

- Score and macro-variables explain the early 90s downturn but not the increase in default rates in 2000-2003
- The increased DR over the period 2000-2003 can be attributed to unobserved heterogeneity captured by the latent factor



An aggregate upgrade and an aggregate downgrade factors drive all transitions. The aggregate factors are linked to the same macroeconomic variables that drive defaults

- Transitions are conditionally independent and are driven by an aggregate upgrade and an aggregate downgrade factors
- Model specification:

$$T_{ij,t+1} \mid S_t = i, f_t \sim Binom\left(P_{ij,t}^{trans}, N_{i,t}\right), \ \forall\, i \neq j$$

$$Log\left(\frac{P_{ij,t}^{trans}}{1-P_{ij,t}^{trans}}\right) = \lambda_{ij} + \beta_{ij} \cdot f_t \qquad f_t = \begin{cases} f_t^{down}, \text{if } i < j \\ f_t^{up}, \text{if } i > j \end{cases}$$

$$P_{ii,t}^{trans} = 1 - \sum_{i<j} P_{ij,t}^{trans} - \sum_{i>j} P_{ij,t}^{trans} - \sum_{i} PD_{i,t}$$

- The aggregate factors are modelled as follows:

$$T_t^{down} \mid S_t, X_t^{down} \sim Binom\left(P_t^{down}, N_t\right)$$

$$Log\left(\frac{P_t^{Down}}{1-P_t^{Down}}\right) = f_t^{down} = \underset{(0.03609)}{-3.3949} + \underset{(0.02229)}{0.9256} \cdot 1_{July-December} + \underset{(1.0441)}{14.9021} GDP_t$$

$$T_t^{up} \mid S_t, X_t^{up} \sim Binom\left(P_t^{up}, N_t\right)$$

$$Log\left(\frac{P_t^{Up}}{1-P_t^{Up}}\right) = f_t^{up} = -\underset{(0.03516)}{3.28} + \underset{(0.02217)}{0.9274} \cdot 1_{July-December} + \underset{(1.0263)}{9.9325} GDP_t - \underset{(0.9622)}{8.5121} Spread_t$$

2 stress testing scenarios: a) A severe downturn at the end of the stress testing period, b) A severe downturn in 2 years and a subsequent partial economic recovery

- Stress Scenario 1:
  - Realized GDP Growth and Short-Long Spread over the period 2005-2009
  - Benign economic conditions over the period 2005-2007
  - Severe downturn over the period 2008-2009
- Stress Scenario 2:
  - Based on forecasts for both series, using the simple AR(1) models:

$$TrGDP_t = \underset{(0.05543)}{0.979} \cdot TrGDP_{t-1} + \varepsilon_t^{GDP}, \quad \varepsilon_t^{GDP} \sim N\left(0, 0.0161^2\right)$$

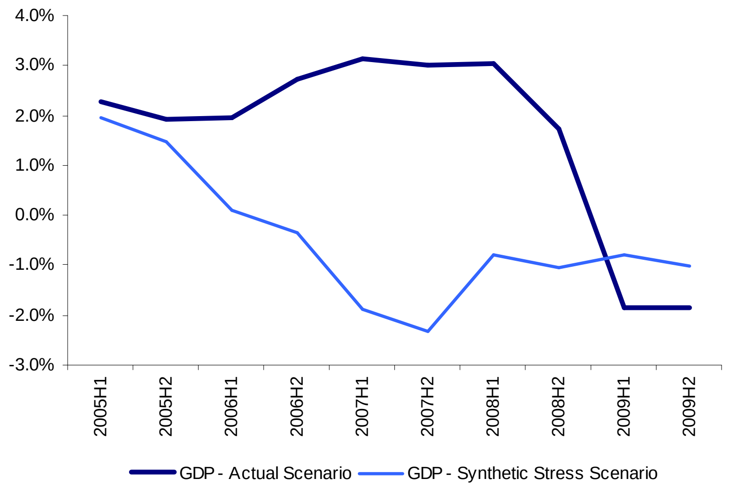
$$Spread_t = \underset{(0.14611)}{0.65563} \cdot Spread_{t-1} + \varepsilon_t^{Spread}, \quad \varepsilon_t^{Spread} \sim N\left(0, 0.012899^2\right)$$

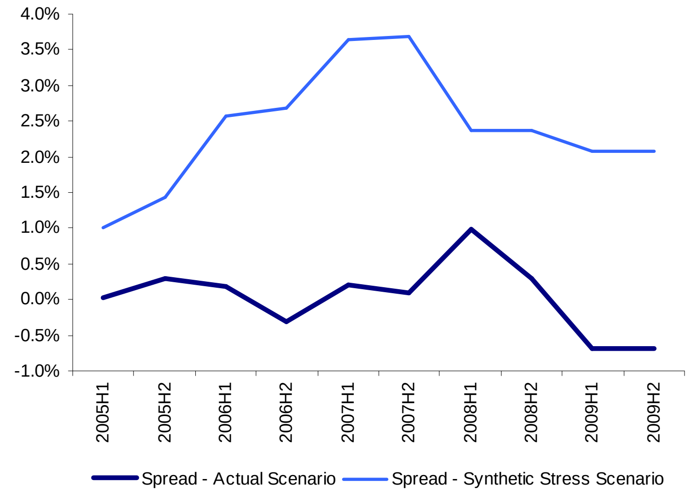
  - Severe economic downturn in the medium term (2-3 years) and subsequent partial economic recovery, by using the following percentiles of the forecasting distributions:

Percentile	2005	2006	2007	2008	2009
TrGDP	20%	5%	1%	10%	15%
Spread	80%	95%	99%	90%	85%

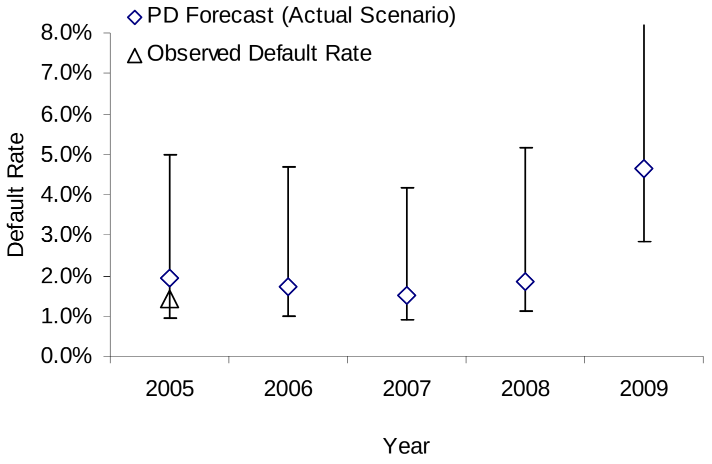
2 stress testing scenarios: a) A severe downturn at the end of the stress testing period, b) A severe downturn in 2 years and a subsequent partial economic recovery

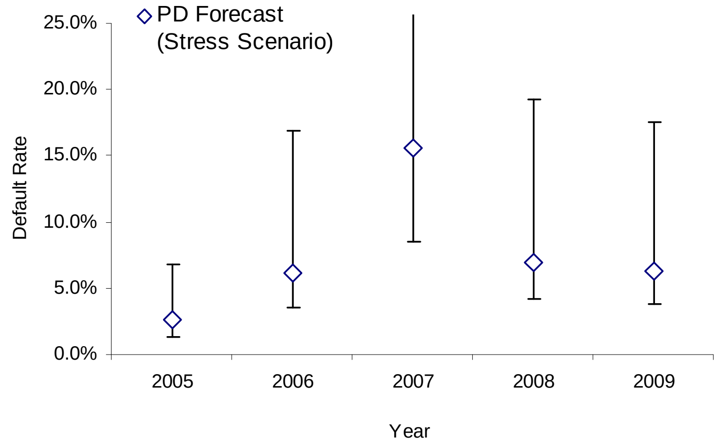
- Stressed Trailing GDP Growth and Short Long Spread:





- Portfolio PD Under Stressed Scenarios:

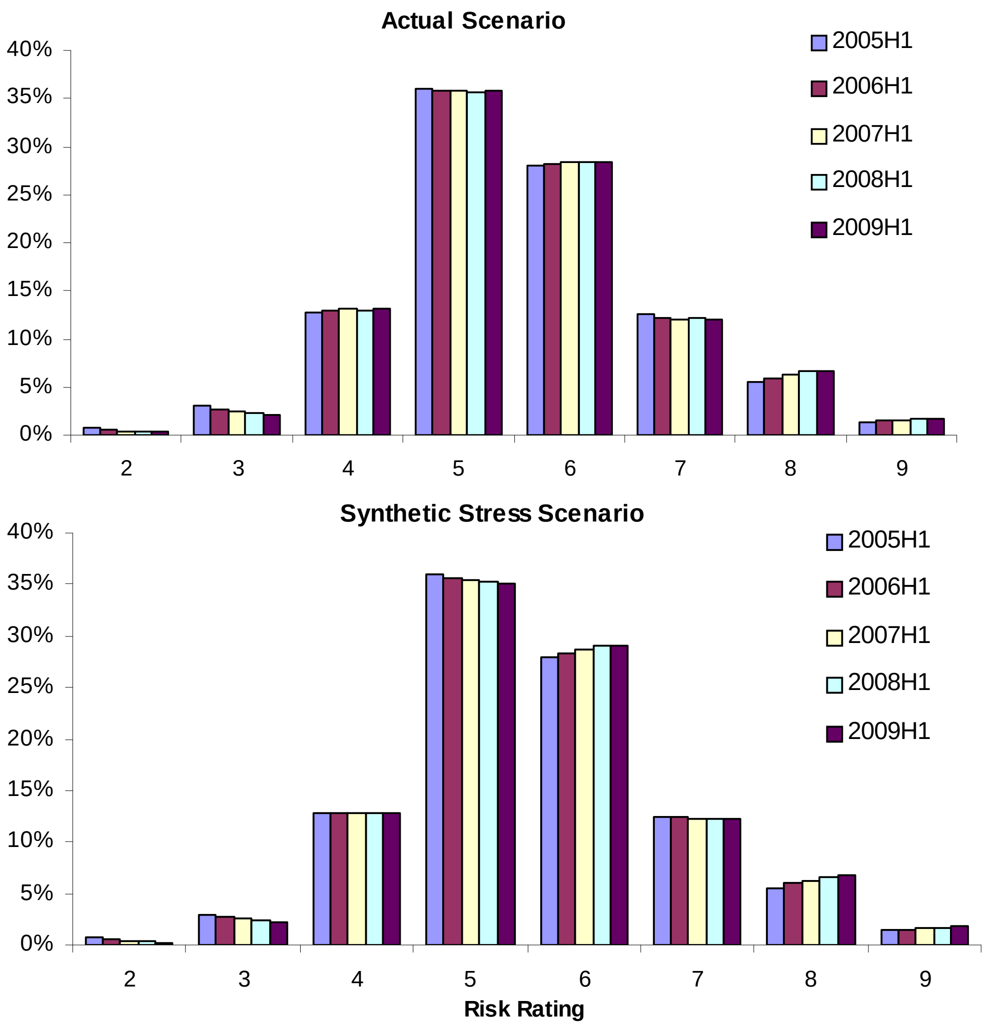




2 stress testing scenarios: a) A severe downturn at the end of the stress testing period, b) A severe downturn in 2 years and a subsequent partial economic recovery

LLOYDS BANKING GROUP

- Stress Rating Grade distributions:



Conclusion

- Summary
  - Practical, statistical solution to the lack of internal data when a suitable external data source is available
  - Obtain a historical view of the current portfolio composition and forecasts of how it is most likely to behave in the future
  - Decompose rating system into PD and rating migrations and separately model each part
  - Link PD and rating migrations to the same underlying drivers, simplifying stress testing
  - Inclusion of an unobserved factor to take account of the additional correlation, not captured from observed macro variables
  - Can be used for forward calibration, business planning, stress testing, capital, and loss provisioning
- Extensions:
  - Include portfolio evolution scenarios
  - Allow for different drivers of PD and rating migration
  - Include latent factors for rating migrations