LGD Modelling for Mortgage Loans

August 2009

Mindy Leow, Dr Christophe Mues, Prof Lyn Thomas

School of Management
University of Southampton

Agenda

- Introduction & Current LGD Models
- Research Questions
- Data
- LGD Model
  - Probability of Repossession Model
  - Haircut Model
- Preliminary Conclusions
- Including Macroeconomic Variables
  - Probability of Repossession Model
  - Haircut Model
- Concluding Remarks

Introduction

- Why do we need LGD? Basel II in context
  - Under new Basel II capital framework (Pillar 1), calculation of minimum capital requirements done using one of two approaches: Standardized or Internal Ratings Based (IRB)
  - IRB approach further split into 2: Foundation or Advanced
  - Under IRB Advanced approach, need to develop models to estimate Probability of Default (PD), Loss Given Default (LGD), Expected Exposure at Default (EAD)
- LGD for Mortgage Lending

Current Mortgage LGD Models

- ► Repossession Model often only has Loan to Value ratio at estimated repossession date as explanatory variable
- ► LGD is derived using a combination of both models

(Lucas A, Basel II Problem Solving, Rhino Risk)

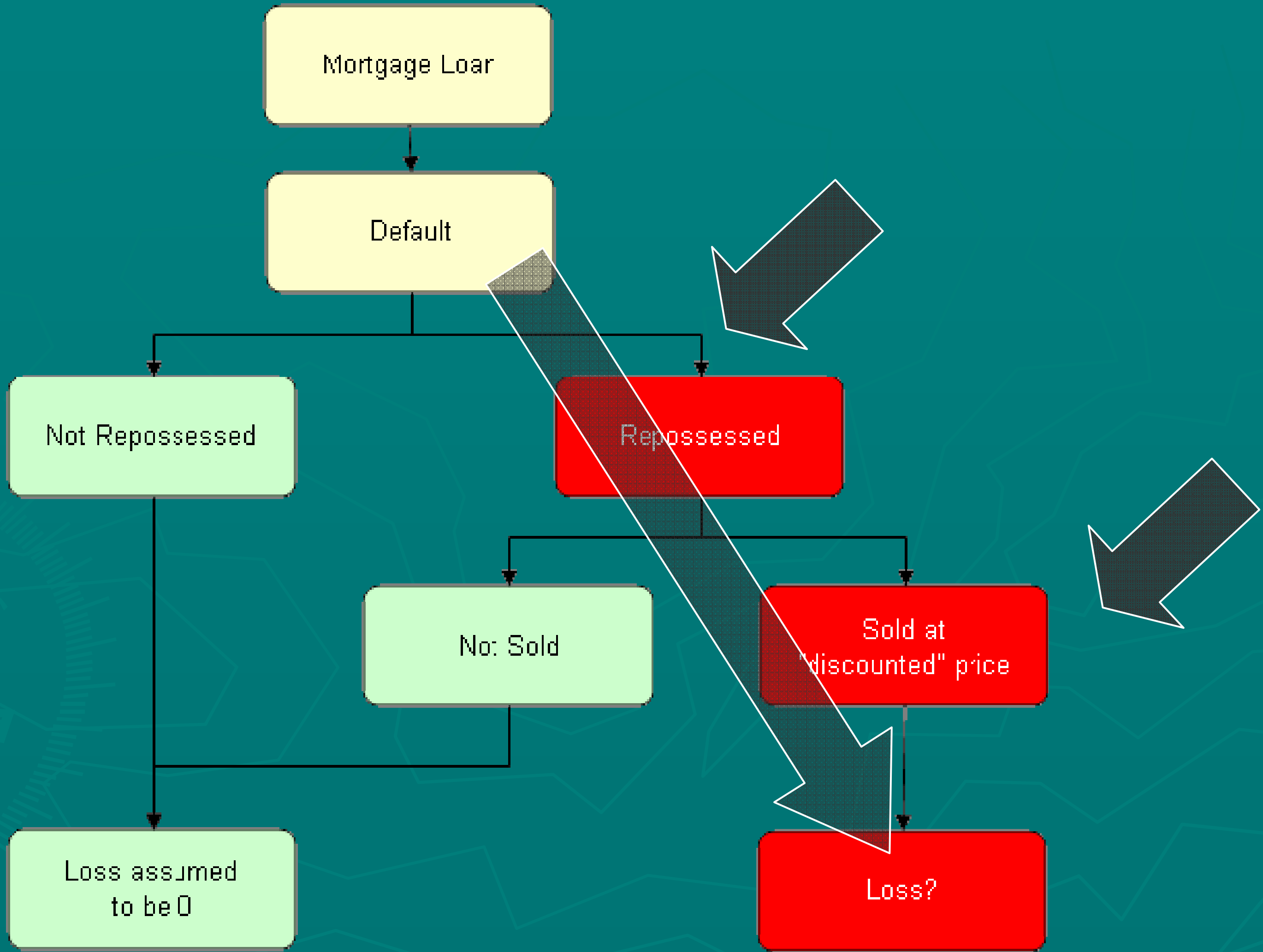
- ► Model LGD (Linear Regression) directly from characteristics of defaulted observations with high loan to value

(Qi and Yang, 2009)

- ► Acknowledged combination of Repossession and Haircut Models as a methodology in estimation of LGD

(Somers and Whittaker, 2007)

Current Mortgage LGD Models



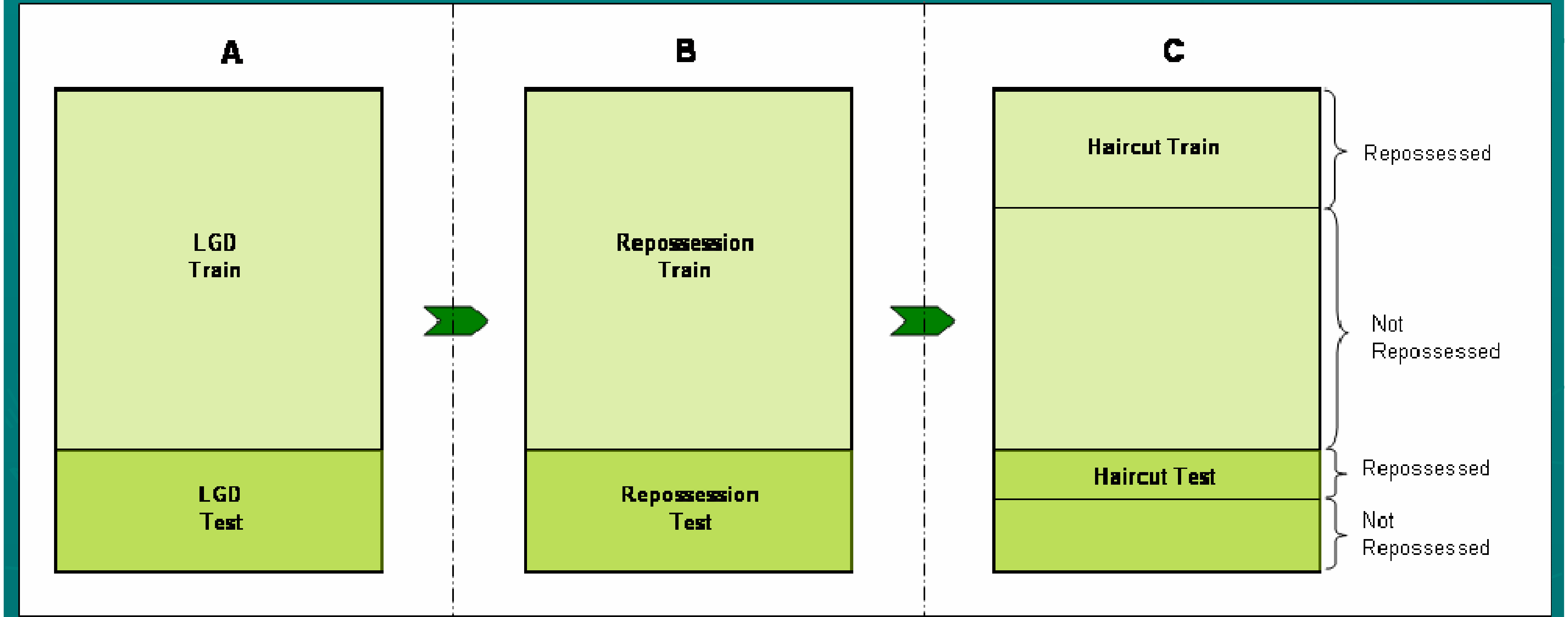
Research Objectives

- ► Evaluate added value of model with more than one variable (Loan to Value) in Probability of Repossession Model
- ► Validate approach of using combination of Repossession Model and Haircut Model to get estimate of LGD
- ► Explore possibility of improved model performance to be achieved by inclusion of macroeconomic variables

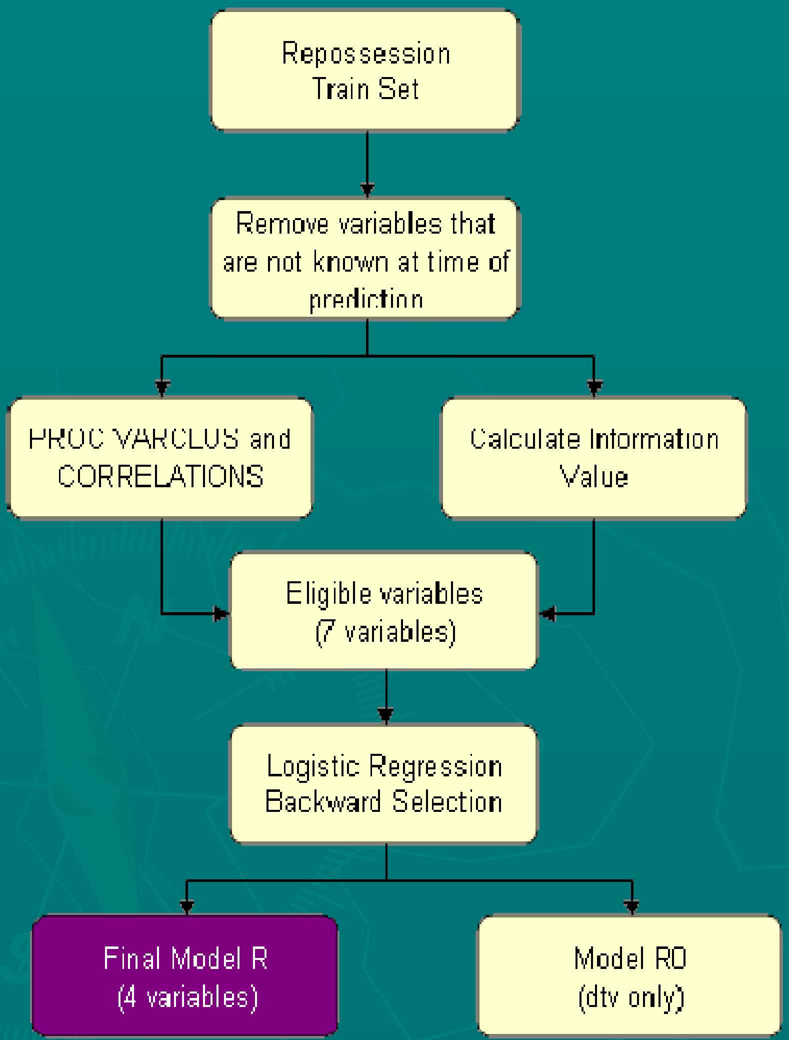
Data

- Source: major UK Bank
- All observations are defaulted mortgages, with information on subsequent repossession or otherwise
- Observations default between 1988 and 2001
- Observations are repossessed between 1989 and 2003

Training and Test Sets Split



Repossession Model Methodology



- Before development of Repossession Model, we
  - Remove variables not known at time of default
  - Identify any correlation between variables
  - Calculate information value of each variable
- Develop Logistic Regression Model, use backward selection to identify final variables to use in model
- Create Model R0, consisting of only DLTV (LTV at default)

Repossession Model Statistics

- According to the Delong Delong and Clarke-Pearson test, which assesses whether there are any significant differences between ROC of models, the 2 models are significantly different
- In terms of model performance statistics, see that the Test set of our Repossession Model manages to achieve an ROC of 0.75

Model	ROC	Cut-off	Specificity	Sensitivity	Accuracy
R Test Set	0.7553	0.4215	60.2858	77.3149	71.2448
R0 Test Set	0.7374	0.4359	58.6257	76.0075	69.8117

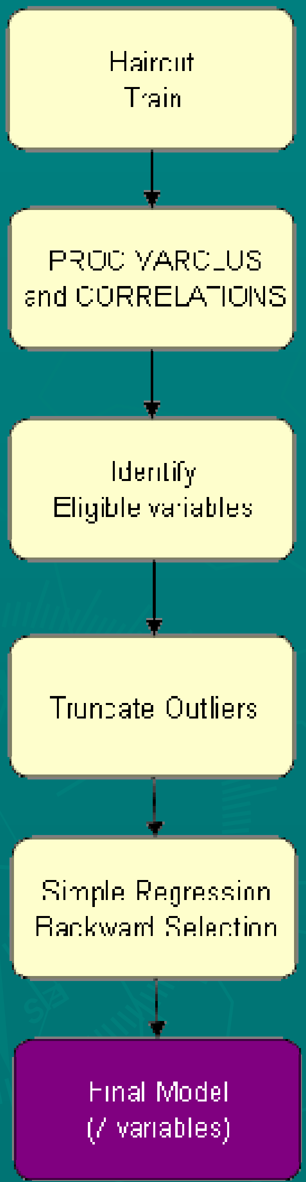
Table 1: Repossession Models Performance Statistics

Probability of Repossession Model Parameters

Variable	Relationship to Probability of Repossession	Explanation
LTV at start	+	If large proportion of loan is tied up in security, likelihood of repossession increases
Number of Months in Arrears	+	Loan with large number of months in arrears indicates inability to keep up with payments, so likelihood of repossession increases
Time on Book	-	Older loans imply that more of the loan is repaid which decreases likelihood of repossession
Security	-	Lower range properties are more likely to be repossessed in the case of default

Table 2: Probability of Repossession Parameter Estimate Signs

Haircut Model Methodology & Statistics



- ► Similar to the Repossession Model, we identify any correlations between variables, before truncating outliers
- ► Develop a simple Linear Regression, and use backward selection to select final variables

Model	MSE	MAE	R-sq
Test Set	0.0448	0.1489	0.1194

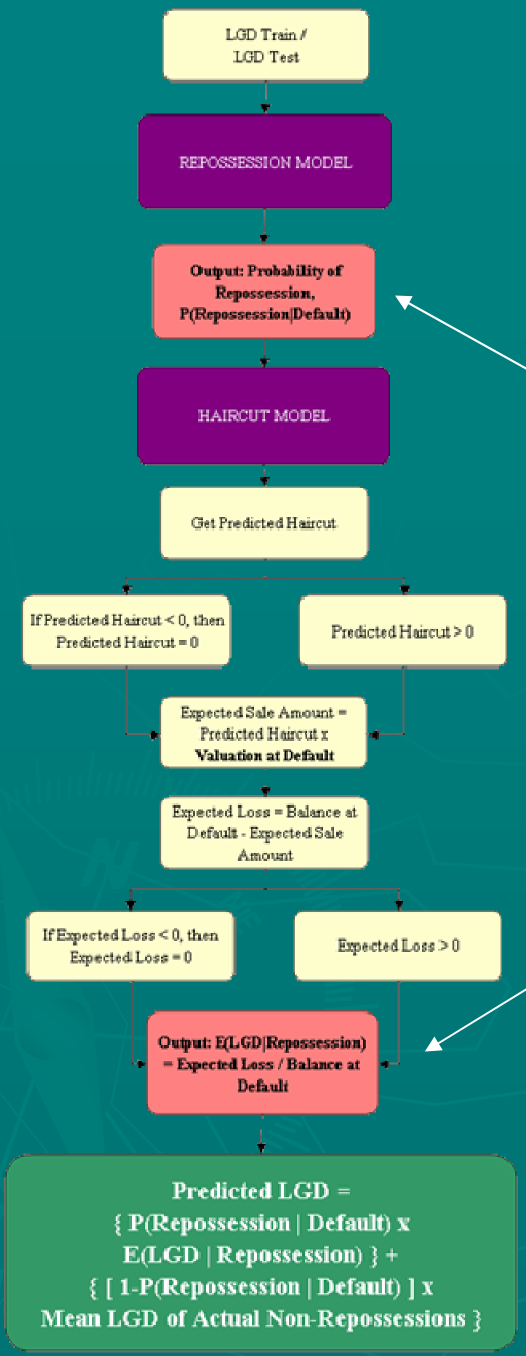
Table 3: Haircut Model Performance Statistics

Haircut Model Parameters

Variable	Relation to Haircut (sale price / valuation at default)	Explanation
LTV at start	+	Could be due to policy decisions taken by the bank. Due to the large loan the bank has committed towards the property, when the account does go into default and subsequent repossession, the bank is reluctant to let go the repossessed property unless it is able to fetch a price close to the current property valuation.
Ratio of valuation of security at default to average property valuation in that region	-	negative sign we get can be explained by the strong negative relationship that is observed in the higher end of the value-to-average ratio spectrum.
Time on book (in years)	+	Older loans imply greater uncertainty and error in estimation of value of security at default, so higher Haircut is possible
Security	+	Haircut tends to be higher for higher-end properties
Age group of property	+	Haircut tends to be higher for new properties
Region		-

Table 4: Haircut Model Parameter Estimate Signs

LGD Methodology



For example,

An account goes into Default.

Repossession Model predicts Probability of Repossession = 0.774

The Haircut Model predicts Haircut, which gives Expected LGD = 0.417

Predicted LGD

= (0.774 x 0.417) + [(1-0.774) x 0]

=0.323

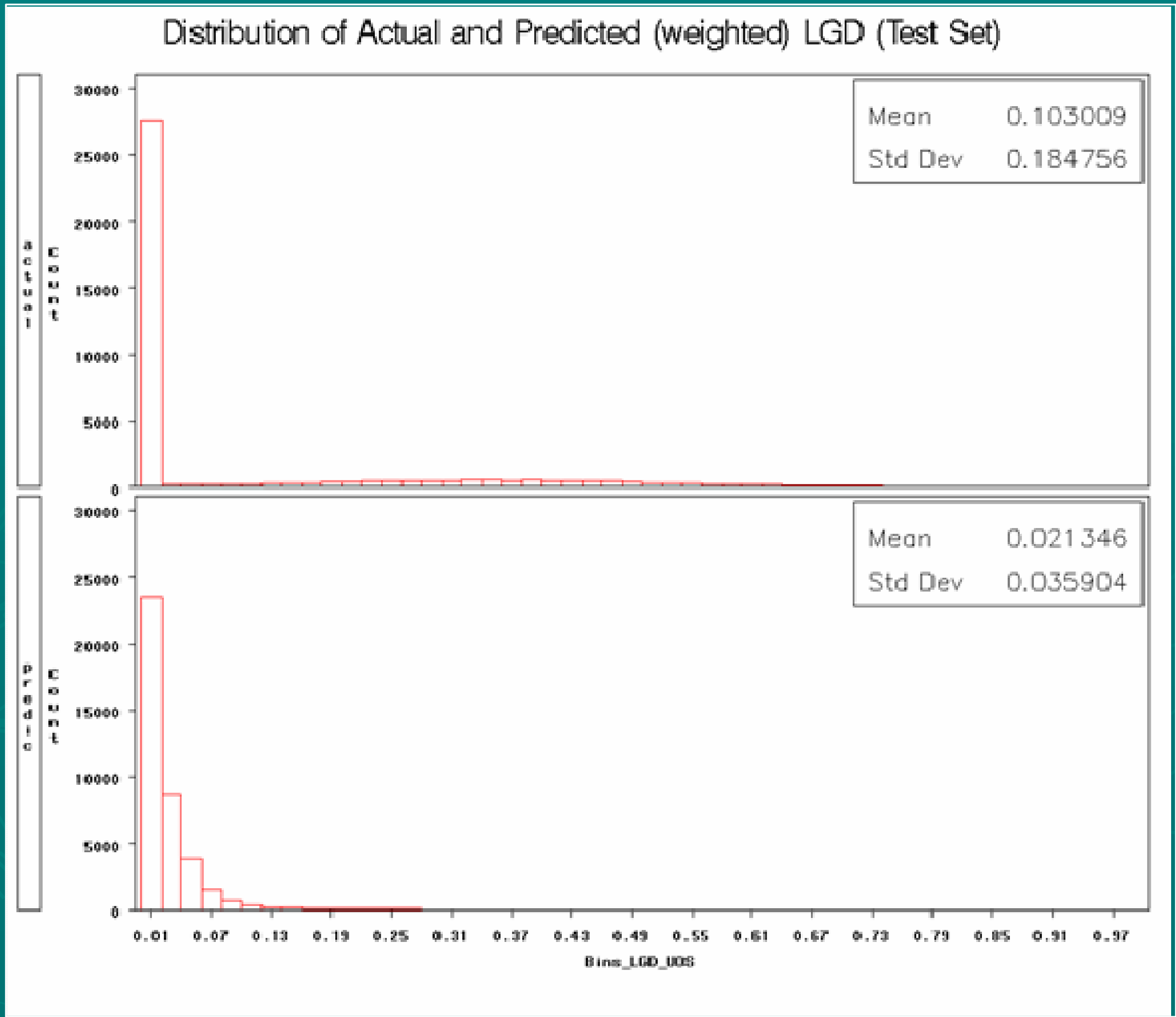
LGD Model Performance

Method, Dataset	R-sq	MSE	MAE
Single Stage Test	0.2133	0.0269	0.1209
2-Stage Test	0.2961	0.0240	0.0973

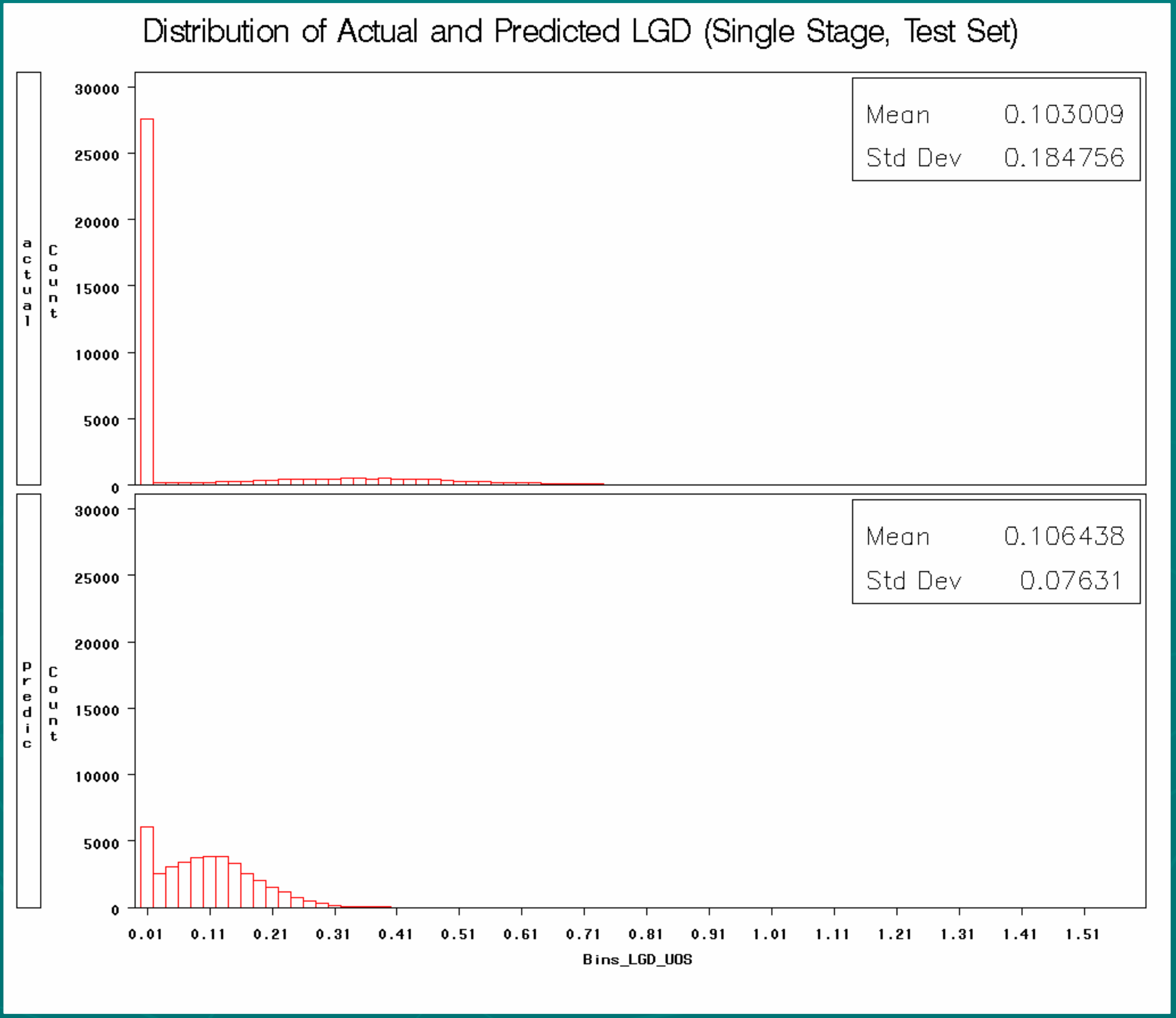
Table 5: Performance Statistics of Single and 2-Stage LGD Models

- Recall single stage model: LGD directly modelled from characteristics of defaulted observations
- Although single stage model achieves similar values of MSE and MAE, R-square is much worse
- Also, Single stage model unable to model distribution of LGD
- Hence confirming the need for a 2 stage model

LGD Two-Stage Model Performance



Single Stage Model Performance



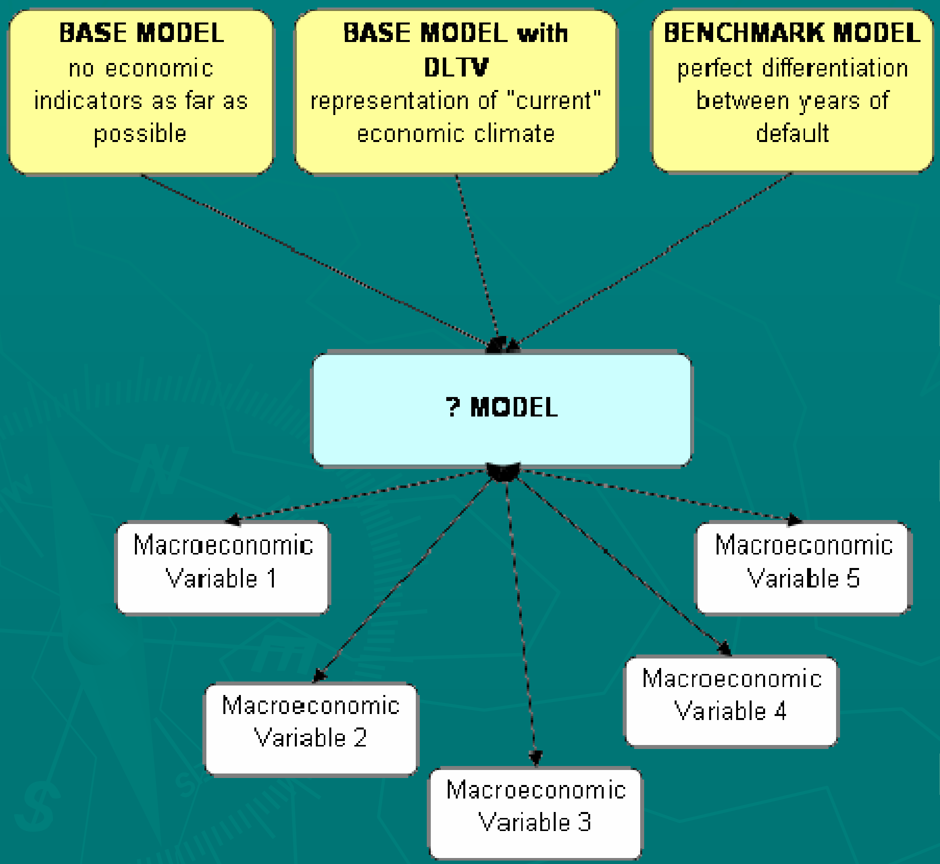
Preliminary Conclusions

- Probability of Repossession Model benefits from inclusion of variables on top of just DLTV
- Single-stage model that directly models LGD is unable to accurately reflect distribution of LGD, thus validating the essential combination of the Repossession Model and Haircut Model to predict LGD

Investigating Effect of Macroeconomic Variables on Predictive Performance

- So far, deliberately kept economic variables out of analysis as far as possible
- Encouraging literature on impact of macroeconomic variables on corporate LGD
  - Recoveries affected by when on economic cycle default happened (Frye, 2000a, 2000b)
  - Predictive variables of recovery include industry and macroeconomic conditions (Gupton & Stein, 2002, 2005)
  - Recovery models benefit statistically from inclusion of variable which represents the macroeconomy (Altman et al, 2005)

Including Macroeconomic Variables: Methodology



- Decide on "best" starting model before including macroeconomic variables, separately and independently
- Variables taken at 2 time points – start and default
- For each macroeconomic variable, compare improvement to models (if any)
- Repeat for both component models

Macroeconomic Variables Considered

Macroeconomic Variable	Source	Time Unit	Definition
Net Lending Growth	ONS	Quarterly	Total consumer credit, net lending, seasonally adjusted, quarter on (previous) quarter percentage change
Disposable Income Growth	ONS	Quarterly	Real households' disposable income per head, seasonally adjusted, (constant 2003 prices), quarter on (previous) quarter percentage change
GDP Growth	ONS	Quarterly	Gross Domestic Product, seasonally adjusted, (constant 2003 prices), quarter on (previous) quarter percentage change
Purchasing Power Growth	ONS	Annually	Internal purchasing power of the pound (based on Retail Prices Index), not seasonally adjusted, (constant 2003 prices), year on year percentage change
Unemployment Rate	ONS	Monthly	Unemployment rate, UK, All aged 16 and over, percentage, seasonally adjusted
Saving Ratio	ONS	Quarterly	Household saving ratio, seasonally adjusted
Interest Rate	BOE	Monthly	Bank of England interest rate, mean over the month
House Price Index Growth	Halifax	Quarterly	All houses, all buyers, non seasonally adjusted, quarter on (previous) quarter percentage change

Table 6: Macroeconomic Variables and Definitions

Probability of Repossession Model Revisited

Model	Additional Variable		ROC (Test)
Base	-		0.7553
Base + Years	Yr_def (dummys)		0.7635
Base + DLTV	DLTV		0.773
Base + DLTV + Years	Yr_def (dummys)		0.7863
Model	**Additional Variable**	**Model Sign**	**ROC (Test)**
AT START			
Base + DLTV + MV 1	Net Lending Growth	+	insignificant
Base + DLTV + MV 2	Disposable Income Growth	+	insignificant
Base + DLTV + MV 3	GDP Growth	+	insignificant
Base + DLTV + MV 4	Purchasing Power Growth	-	insignificant
Base + DLTV + MV 5	Unemployment Rate	-	insignificant
Base + DLTV + MV 6	Saving Ratio	+	0.7731
Base + DLTV + MV 7	Interest Rate	+	insignificant
Base + DLTV + MV 8	House Price Index Growth	-	0.7731
AT DEFAULT			
Base + DLTV + MV 9	Net Lending Growth	+	0.7737
Base + DLTV + MV 10	Disposable Income Growth	+	0.7731
Base + DLTV + MV 11	GDP Growth	-	0.7738
Base + DLTV + MV 12	Purchasing Power Growth	-	0.7764
Base + DLTV + MV 13	Unemployment Rate	-	0.7839
Base + DLTV + MV 14	Saving Ratio	-	LTV p-value >0.01
Base + DLTV + MV 15	Interest Rate	+	0.7755
Base + DLTV + MV 16	House Price Index Growth	+	insignificant

Table 7: Performance of Repossession Model with Macroeconomic Variables

Haircut Model Revisited

Model	Additional Variable	Model Sign	R-Square (Test)
Base	-		0.1192
Base + Years	Yr_def (dummys)		0.1542
Base + DLTV	DLTV		0.1267
Base + DLTV + Years	Yr_def (dummys)		0.1564
Model	**Additional Variable**	**Model Sign**	**R-Square (Test)**
AT START			
Base + DLTV + MV 1	Net Lending Growth	-	insignificant
Base + DLTV + MV 2	Disposable Income Growth	-	insignificant
Base + DLTV + MV 3	GDP Growth	-	insignificant
Base + DLTV + MV 4	Purchasing Power Growth	+	0.1317
Base + DLTV + MV 5	Unemployment Rate	-	0.1285
Base + DLTV + MV 6	Saving Ratio	+	0.1262
Base + DLTV + MV 7	Interest Rate	-	0.1352
Base + DLTV + MV 8	House Price Index Growth	+	0.1283
AT DEFAULT			
Base + DLTV + MV 9	Net Lending Growth	-	0.1274
Base + DLTV + MV 10	Disposable Income Growth	+	insignificant
Base + DLTV + MV 11	GDP Growth	+	0.1328
Base + DLTV + MV 12	Purchasing Power Growth	+	TOB p-value >0.01
Base + DLTV + MV 13	Unemployment Rate	+	insignificant
Base + DLTV + MV 14	Saving Ratio	+	0.1266
Base + DLTV + MV 15	Interest Rate	-	0.1478
Base + DLTV + MV 16	House Price Index Growth	+	0.1357

Table 8: Performance of Haircut Model with Macroeconomic Variables

Two-Stage LGD Revisited

- Both component models benefit from inclusion of DTV

Method, Dataset	R-sq	MSE	MAE
Single Stage	0.2133	0.0269	0.1209
2-Stage	0.2961	0.0240	0.0973
2-Stage, DLTV	0.2962	0.0240	0.0981
2-Stage, DLTV, MV	0.3166	0.0233	0.0954

Table 9: Performance of all LGD Models

- Although quite a number of macroeconomic variables turn out to be significant in component models
- They do not add much predictive power to LGD Model

Concluding Remarks (I)

- Although macroeconomic variables have gained significance in corporate LGD models, they do not seem to have the same level of importance in retail LGD models.
- Both component models benefit from inclusion of DLTV, although not as large as expected because HPI (the leading macroeconomic variable in the housing market) is already unavoidably embedded in both the Haircut Model and the calculation of mortgage loan LGD
- Macroeconomic variables are statistically significant but they do not seem to give much further improvement to predictive performance

Concluding Remarks (II)

- Again, because the HPI is already involved in calculation of mortgage loan LGD and DLTV, the improvement in predicted LGD that is derived from the inclusion of macroeconomic variables is not as large as expected. This result is similar to that of Bruche and Gonazalez-Aguado (2009).
- Current work on survival analysis model with time-dependent macroeconomic variables, looking to predict number of months taken to go from default to repossession and/or close, which will be used mainly for stress-testing purposes

Thank you ☺