SAS Modeling Tools To Assess Credit Risk

July 2011

All of us serving you™

usbank

Introduction

- **Presenter:** Kevin Hougas / kevin.hougas@usbank.com
- **Company:** U.S. Bank
- **Line of Business:** Credit Administration Risk Analytics
- **Key Responsibility:** Assessing Credit Risk

- Discussion Points
  - Overview of Credit Risk
  - Proc LOGISTIC
  - Proc REG
  - Proc ARIMA

Credit Risk

- There are 3 main analyses we perform for Credit Risk
  - 1.) Basel II Regulatory Capital
    - – Extreme loss event using **Stochastic** process
  - 2.) Economic Capital
    - – Extreme loss event using **Deterministic** process
  - 3.) Stress Testing
    - – Multiple loss events using macroeconomic scenarios

- In a nutshell…

  Produce loss estimates that are sensitive to multiple risk drivers

Methods Used to Produce Loss Estimates

- **Component Method** (Proc LOGISTIC, Proc REG)
  - Probability of Default (PD)
  - Loss Given Default (LGD)
  - Exposure at Default (EAD)

For any given borrower,

Loss = **PD** x **LGD** x **EAD**

- **Aggregated Method** (Proc ARIMA)
  - Work directly with observed losses as a percentage of outstanding balance

Loss % = **Loss** ÷ **Balance**

Model to Produce PD Estimates (Proc LOGISTIC)

- Probability of Default (PD) represents the likelihood that a borrower will default on their obligation over the next year.
- Proc LOGISTIC estimates the sensitivity of this likelihood to different risk drivers using logistic regression
  - Estimates from logistic regression are constrained between 0% and 100%.

- Data for specifying a logistic regression model can be seen in the table below.

Identification Variables			Modeling Variables		
			(Dependent Variable)	(Independent Variables)	
Borrower #	**Status Beginning of Year**	**Status End of Year**	**Default Indicator**	**Risk Rating**	**S&P 500**
1	Active	Active	0	2	1,000
2	Active	In Default	1	6	700
3	Active	Active	0	1	1,200
4	Active	Active	0	2	1,300
5	Active	Active	0	3	900

Sample SAS Code for Proc LOGISTIC

```
PROC LOGISTIC DATA=Modeling_Data DESCENDING;

CLASS Risk_Rating (PARAM=REF REF='8');

MODEL
default_indicator=

    Risk_Rating
    SP500;

output out=scoreseg p=phat lower=lcl upper=ucl;

RUN;
```

Sample Output from Proc LOGISTIC

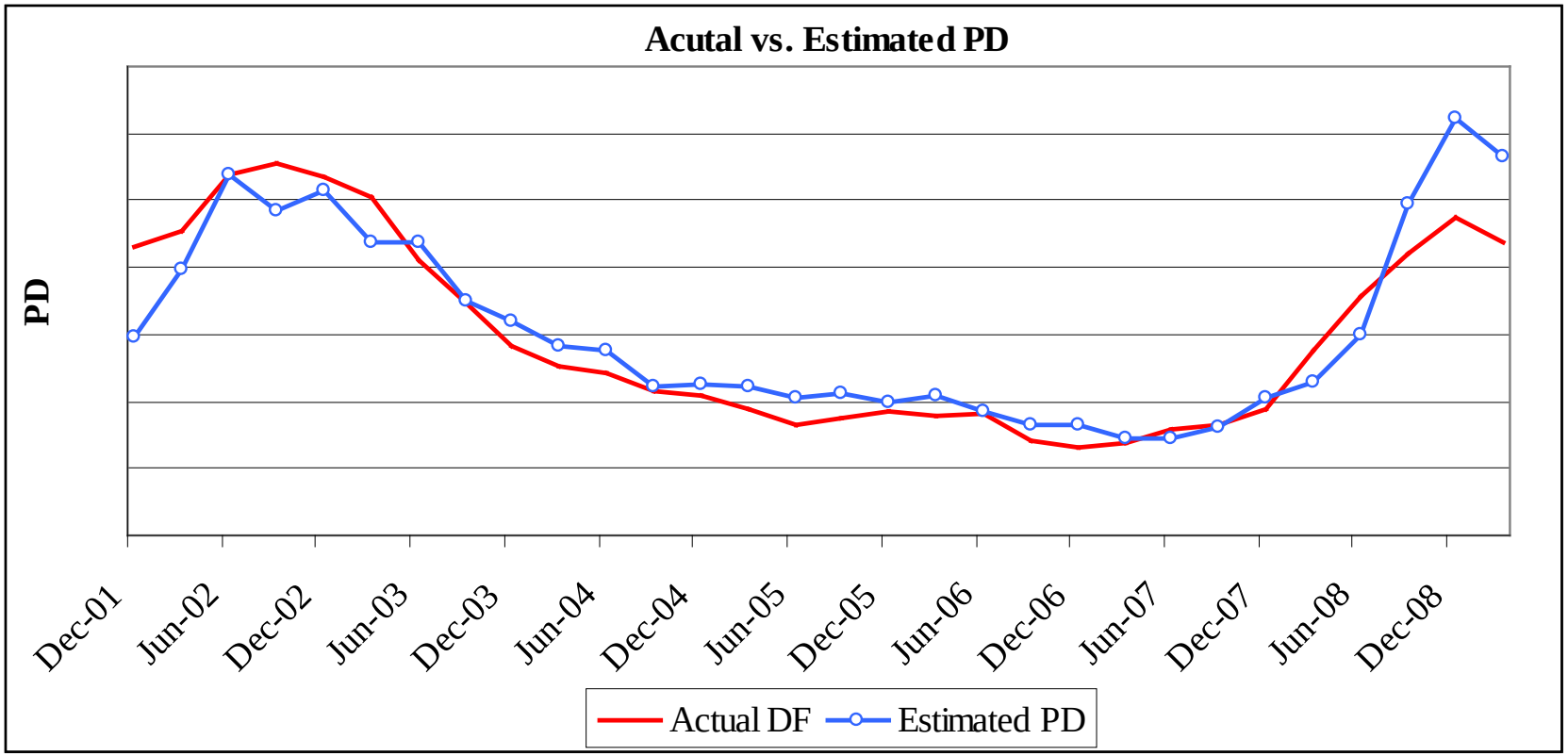
$$LOGIT = a_0 + a_1 X_1 + a_2 X_2 + \ldots a_n X_n$$

$$PD = \frac{1}{1 + \exp(-(LOGIT))}$$

Analysis of Maximum Likelihood Estimates						
Parameter		**DF**	**Estimate**	**Standard Error**	**Wald Chi-Square**	**Pr > ChiSq**
Intercept		1	0.490	0.089	31	<.0001
Risk Rating	**1**	1	-15.238	77.848	0	0.845
Risk Rating	**2**	1	-5.126	0.303	287	<.0001
Risk Rating	**3**	1	-4.462	0.145	947	<.0001
Risk Rating	**4**	1	-3.933	0.062	4,079	<.0001
Risk Rating	**5**	1	-3.152	0.039	6,682	<.0001
Risk Rating	**55**	1	-2.796	0.100	777	<.0001
Risk Rating	**6**	1	-2.089	0.054	1,485	<.0001
Risk Rating	**7**	1	-1.306	0.052	631	<.0001
S&P 500		1	-0.002	0.000	545	<.0001

Sample Output from Proc LOGISTIC

- Key Output Metric: AUC = 83.5%



Model to Produce LGD Estimates (Proc REG)

- Loss Given Default (LGD) is the % of remaining balance the bank will lose if the borrower defaults on their loan
- Proc REG estimates the sensitivity of LGD to different risk drivers using linear regression
  - Estimates from linear regression are not constrained.

- Data for specifying a linear regression model can be seen in the table below.

Identification Variables			Modeling Variables	
			(Dependent Variable)	(Independent Variable)
Date	**Total Balance at Default**	**Total Amount Recovered**	**Average LGD**	**HPI % Change**
Jan-02	10,000	1,000	90%	-20%
Feb-02	10,000	2,000	80%	-10%
Mar-02	10,000	3,000	70%	0%
Apr-02	10,000	4,000	60%	10%
May-02	10,000	5,000	50%	20%

Sample SAS Code for Proc REG

```
PROC REG DATA=Modeling_Data;

MODEL
LGD =

HPI_Change
;

output out=scoreseg p=phat;

run;
```

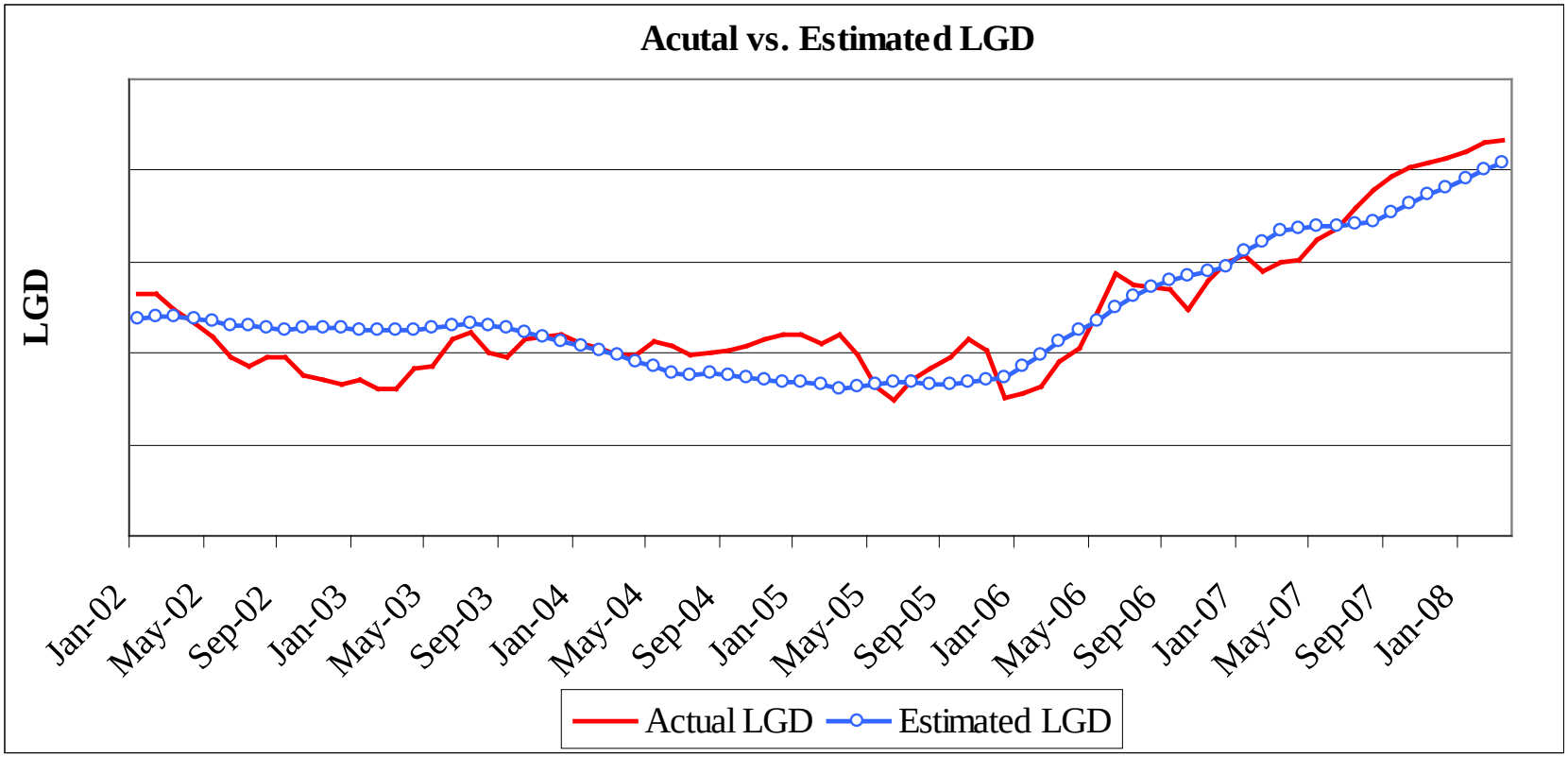
Sample Output from Proc REG

$$LGD = a_0 + a_1 X_1 + ... a_n X_n$$

Parameter Estimates					
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**
Intercept	1	0.9030	0.0026	346	<.0001
HPI_Change	1	-0.4092	0.0228	(18)	<.0001

Sample Output from Proc REG

- Key Output Metric: R-Square = 81.3%



Loss Estimate Using Component Method

PD - Derived from logistic regression model

LGD - Derived from linear regression model

EAD - Equal to outstanding balance at default

Loss = **PD** x **LGD** x **EAD**

Loss Estimates Using Aggregated Method (Proc ARIMA)

- Loss % is the ratio of observed losses to outstanding balance for a given quarter
- Proc ARIMA estimates the sensitivity of Loss % to different risk drivers using an Auto Regressive Integrated Moving Average model
  - ARIMA requires a time series data source
  - Can be used to address serial correlation and specify lag structures

- Data for specifying an ARIMA model can be seen in the table below.

Identification Variables			Modeling Variables	
			(Dependent Variable)	(Independent Variable)
Date	**Balance**	**Loss Amount**	**Loss %**	**Unemployment**
Jan-02	10,000	1,000	10%	6%
Apr-02	10,000	2,000	20%	7%
Jul-02	10,000	3,000	30%	8%
Oct-02	10,000	4,000	40%	9%
Jan-03	10,000	5,000	50%	10%

Sample SAS Code for Proc ARIMA

```
ODS Graphics On;

proc arima data=Modeling_Data;

    identify var=Loss_Rate crosscorr=(Unemp);
    estimate method=ML plot;

run;

estimate p=2 q=0
    input=(

        2$Unemp
    )

method=ML plot;
forecast lead=12 id=date interval=quarter out=scoreseg printall;

quit;

run;
```

Sample Output from Proc ARIMA

Crosscorrelations to Unemployment				
Lag	Covariance	Correlation	-1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1	
0	0.00013	94%	| . |********************|	
1	0.00012	88%	| . |******************|	
2	0.00010	78%	| . |****************|	
3	0.00009	65%	| . |************|	
4	0.00007	51%	| . |**********|	
5	0.00005	36%	| . |*******|	
6	0.00003	20%	| . |****|	
7	0.00001	7%	| . |*|	

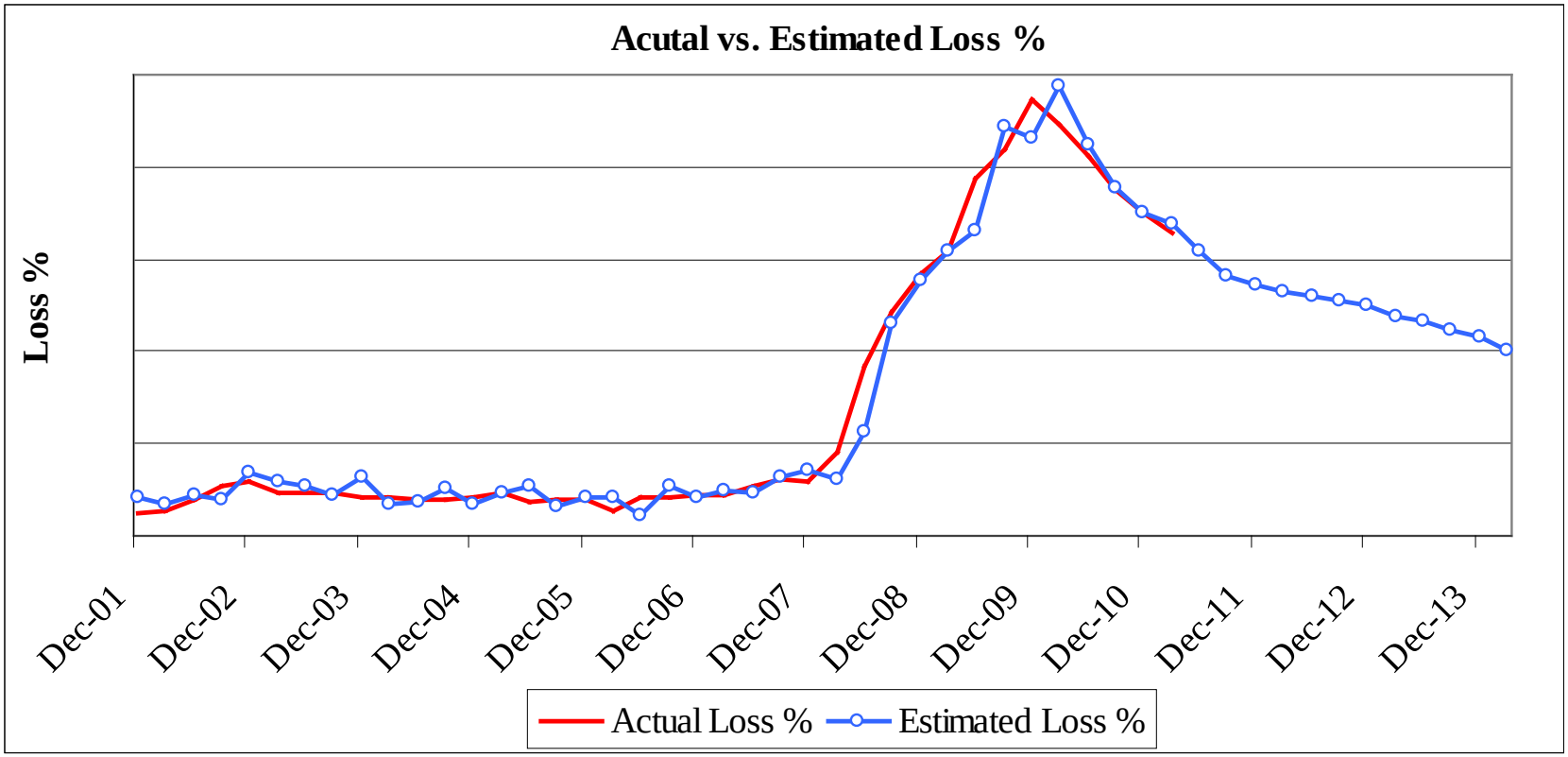
$$Loss\_Rate = a_0 + a_1 X_1 + ...a_n X_n + Z_t$$

$$Z_t = \theta_1 Z_{t-1} + ...\theta_n Z_{t-n} + \varepsilon_t$$

Maximum Likelihood Estimation							
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag	Variable	Shift
MU	-0.0004185	0.0056068	-0.07	0.9405	0	Residential_mortgages	0
AR1,1	1.47421	0.15021	9.81	<.0001	1	Residential_mortgages	0
AR1,2	-0.51639	0.14749	-3.5	0.0005	2	Residential_mortgages	0
Unemp	0.12586	0.06794	1.85	0.064	0	Unemp	2

Sample Output from Proc ARIMA

- Forecast assumes Unemployment falls to 7.0%
- Key Output Metric: AIC = -414, White Noise = 20%



Review

- Proc LOGISTIC
  - Estimates the likelihood of Yes or No events
- Proc REG
  - Estimates the relationship between a continuous dependent variable and multiple independent variables
- Proc ARIMA
  - Forecasts time series data while accounting for serial correlation and lag structure

Questions?