CRC Credit Research Centre

CREDIT SCORING & CREDIT CONTROL XIII

28-30 AUGUST 2013

The Influence of Financial Crisis on UK SMEs Behaviour

University of Edinburgh, Business School
Presentation: Meng Ma
Supervisors:
Prof. Jake Ansell & Dr. Galina Andreeva

content

- Background
- Build panel models
  - Use of weight of evidence
  - Reduce independent variables
  - Select panel for four years
- Fixed effect or random effect
  - Failure of Fixed effect
- Influence from macro variables
  - One at a time
  - Set of best AIC

Background

- SMEs credit scoring: 4.9m SMEs in UK(BIS,2009), 2.3m covered by our data.
  - The 'Credit Crunch' :
    - aggregately default
    - public and supervisors' concern
  - Basel III: consistent internal rating system
- Lack of research in SMEs' behaviour:

  Challenge: missing category, performance variation between with not clear pattern, large data short time period
- Constant model during economic downside

Contributions

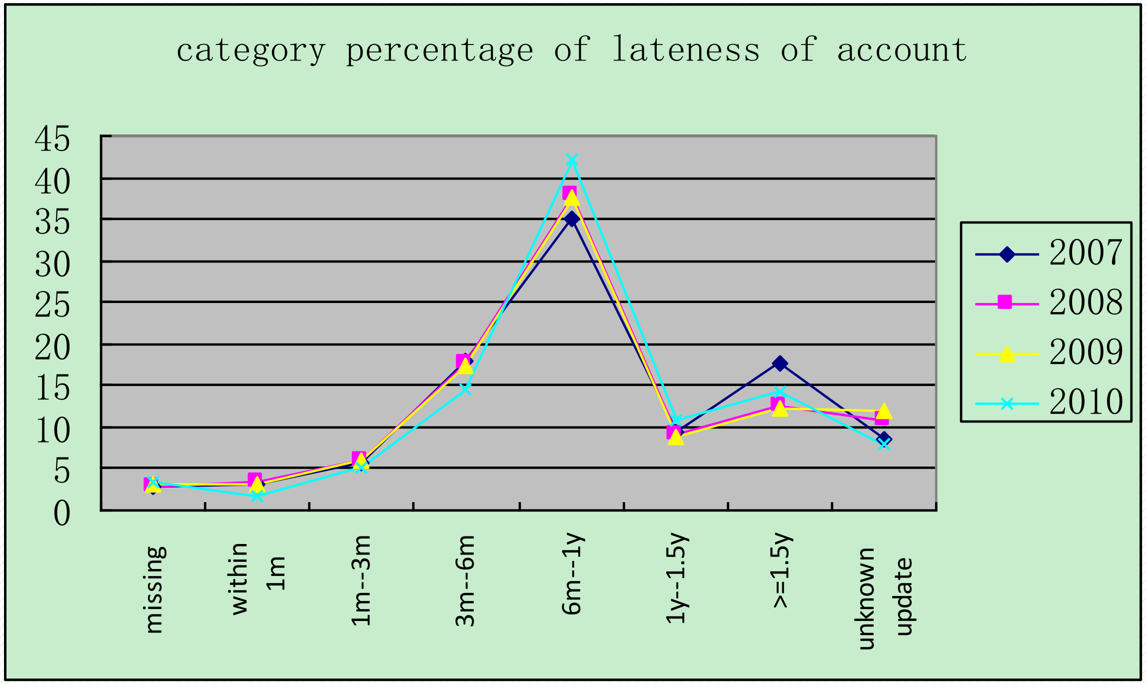
- Analysis of SMEs during ‘credit crunch’
- Muti-time-level modelling of SMEs default probability: using panel data for credit scoring
- Obtaining macroeconomic variables into SMEs modelling

Training sample

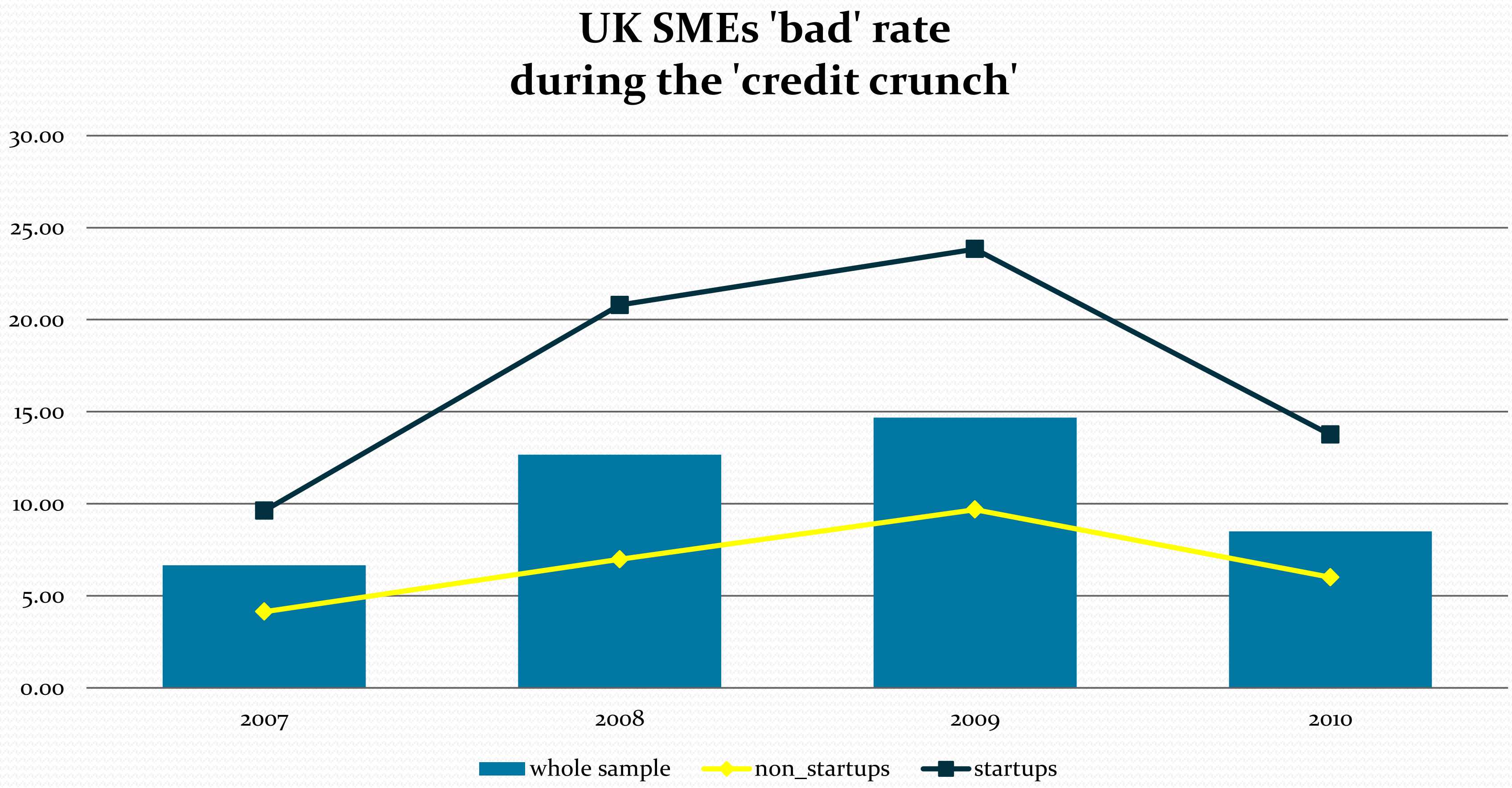
- Selected Training sample
  - Start-ups: younger than 36m
  - Non-start-ups: older than 36m

Size of the training sample				
year	2007	2008	2009	2010
Non-start-ups	111,021	103,072	94,039	83,697
Start-ups	94,742	72,034	51,350	39,773
total	205,763	175,106	145,389	123,470

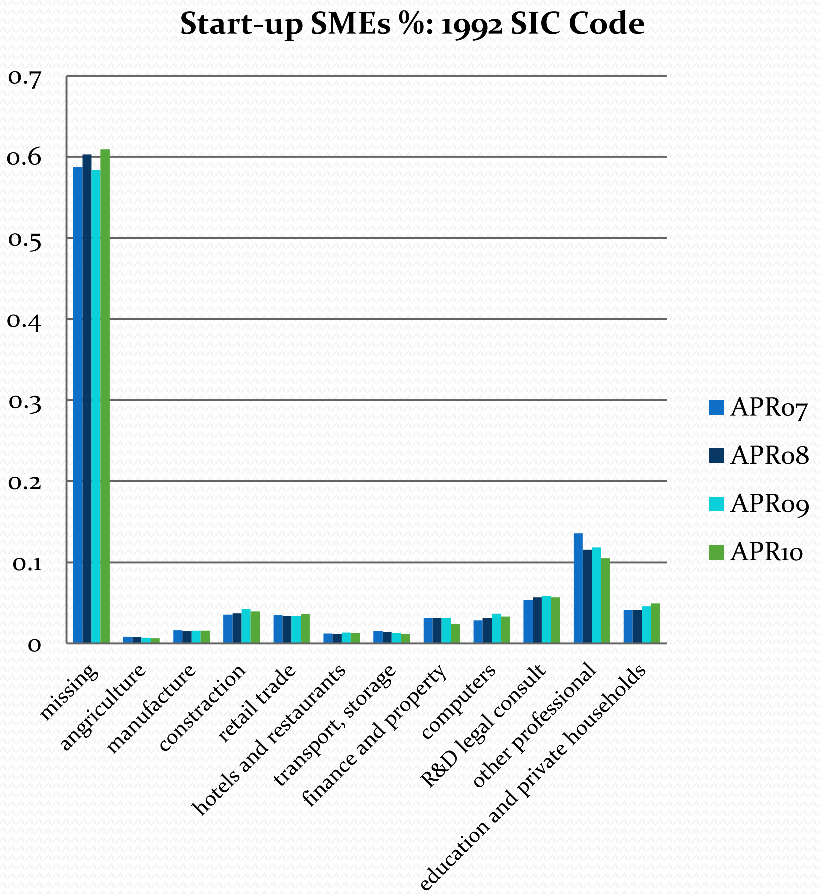
Lateness of Account

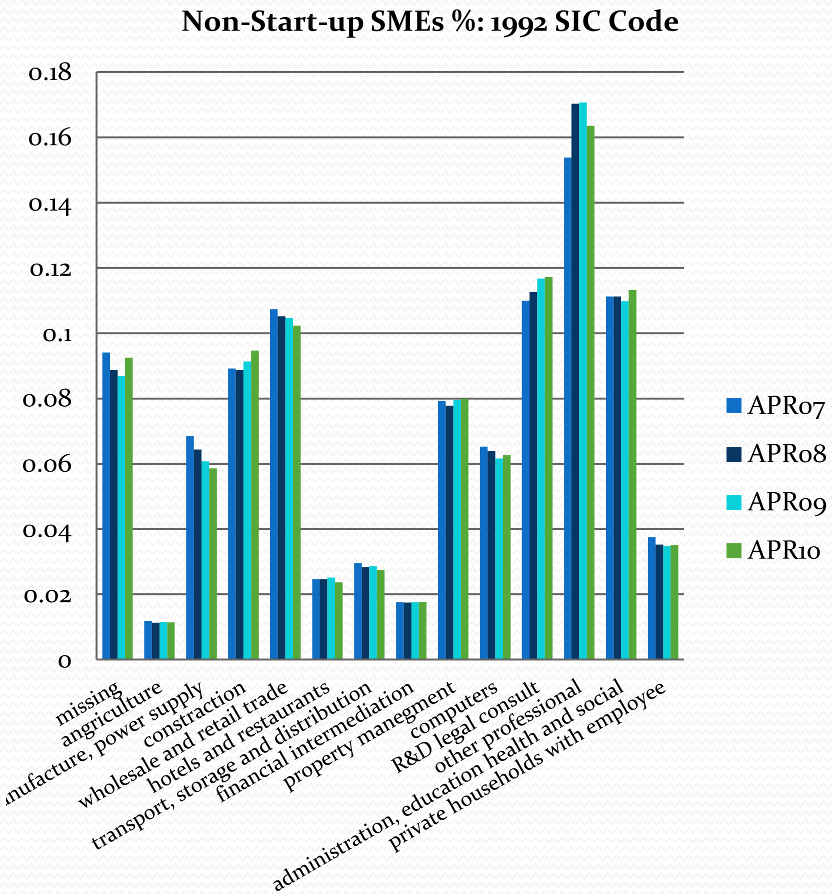


SMEs Performances

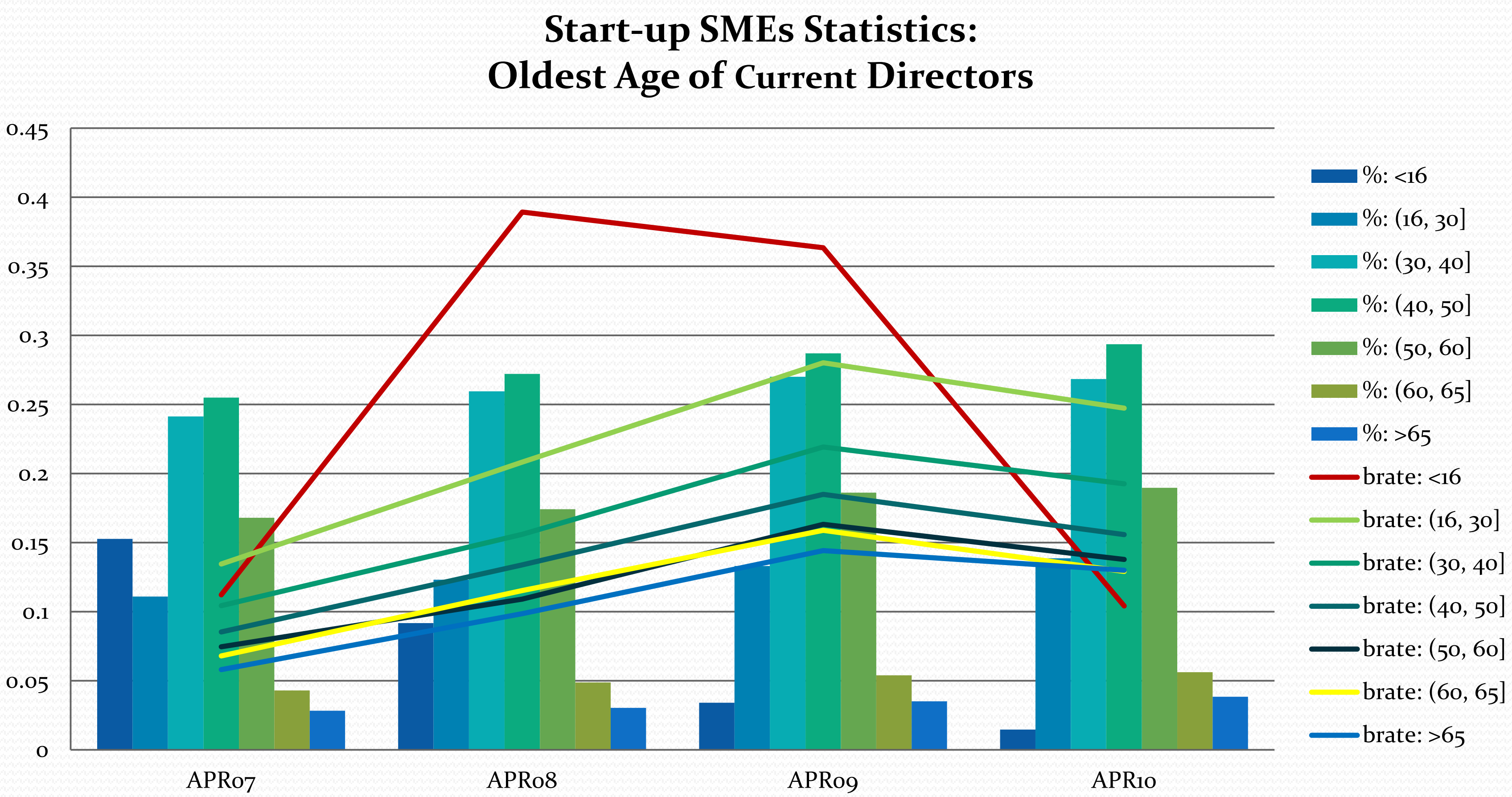


SME’s Categorical Performances

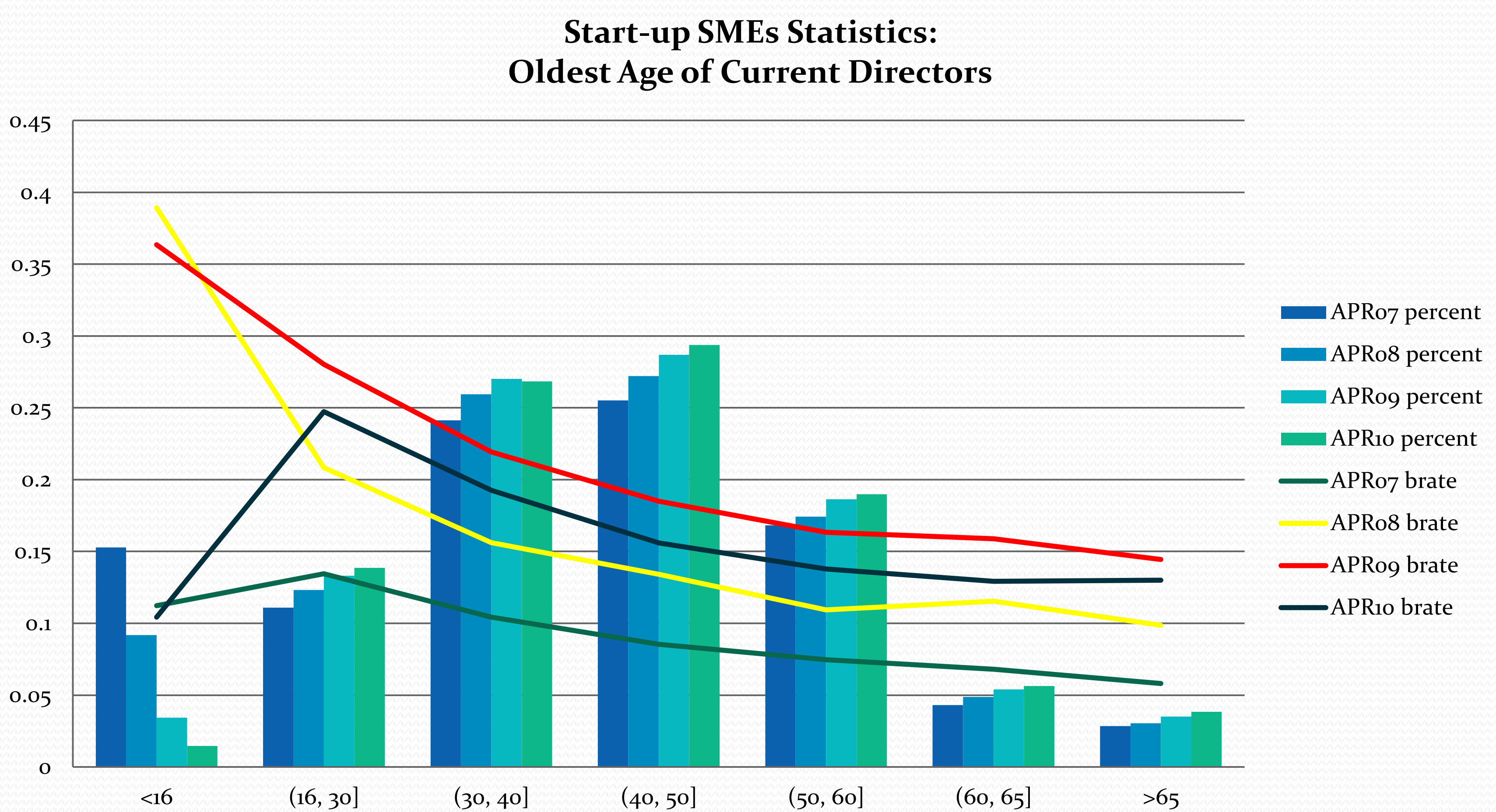




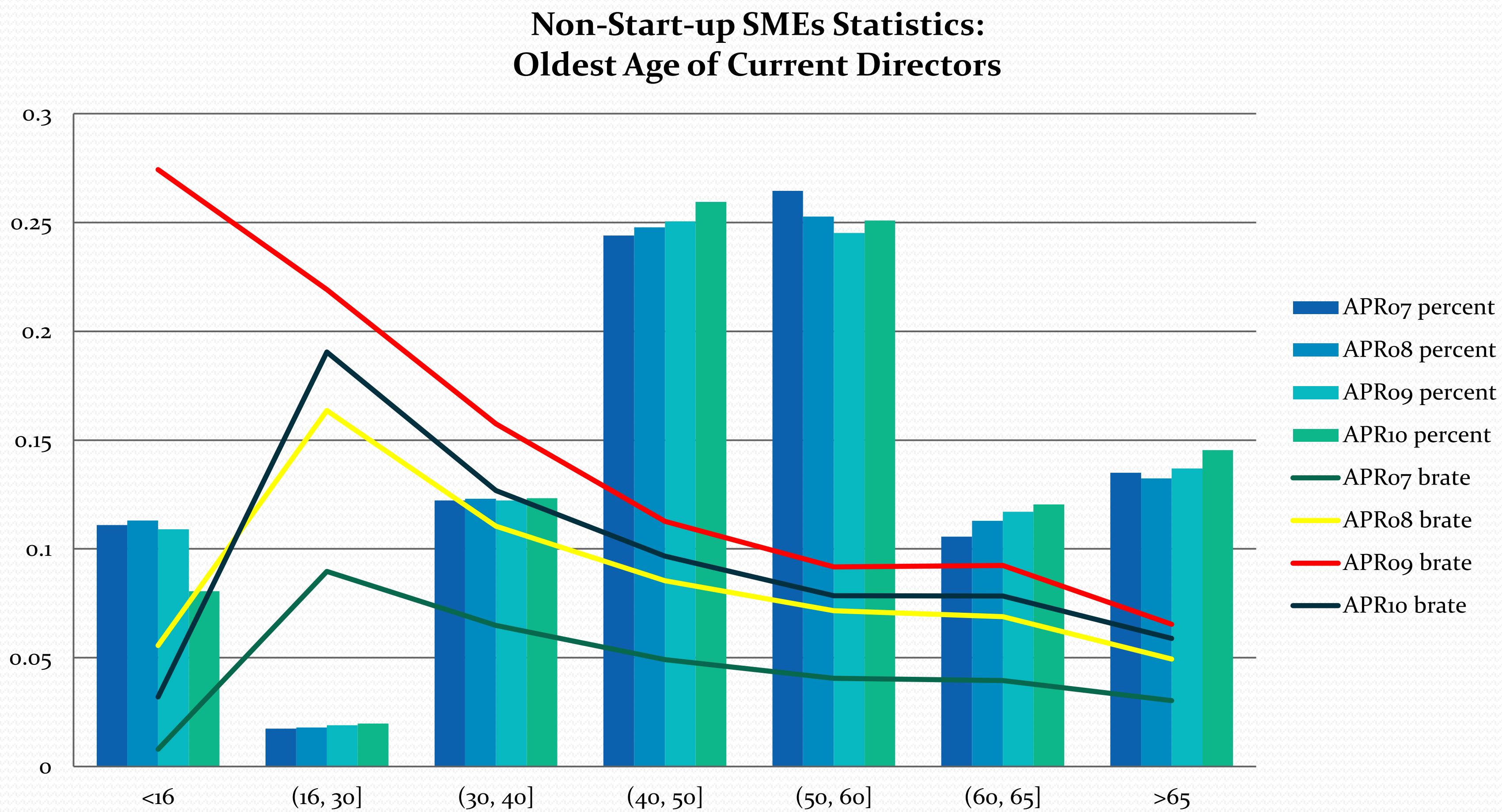
SME's Categorical Performances



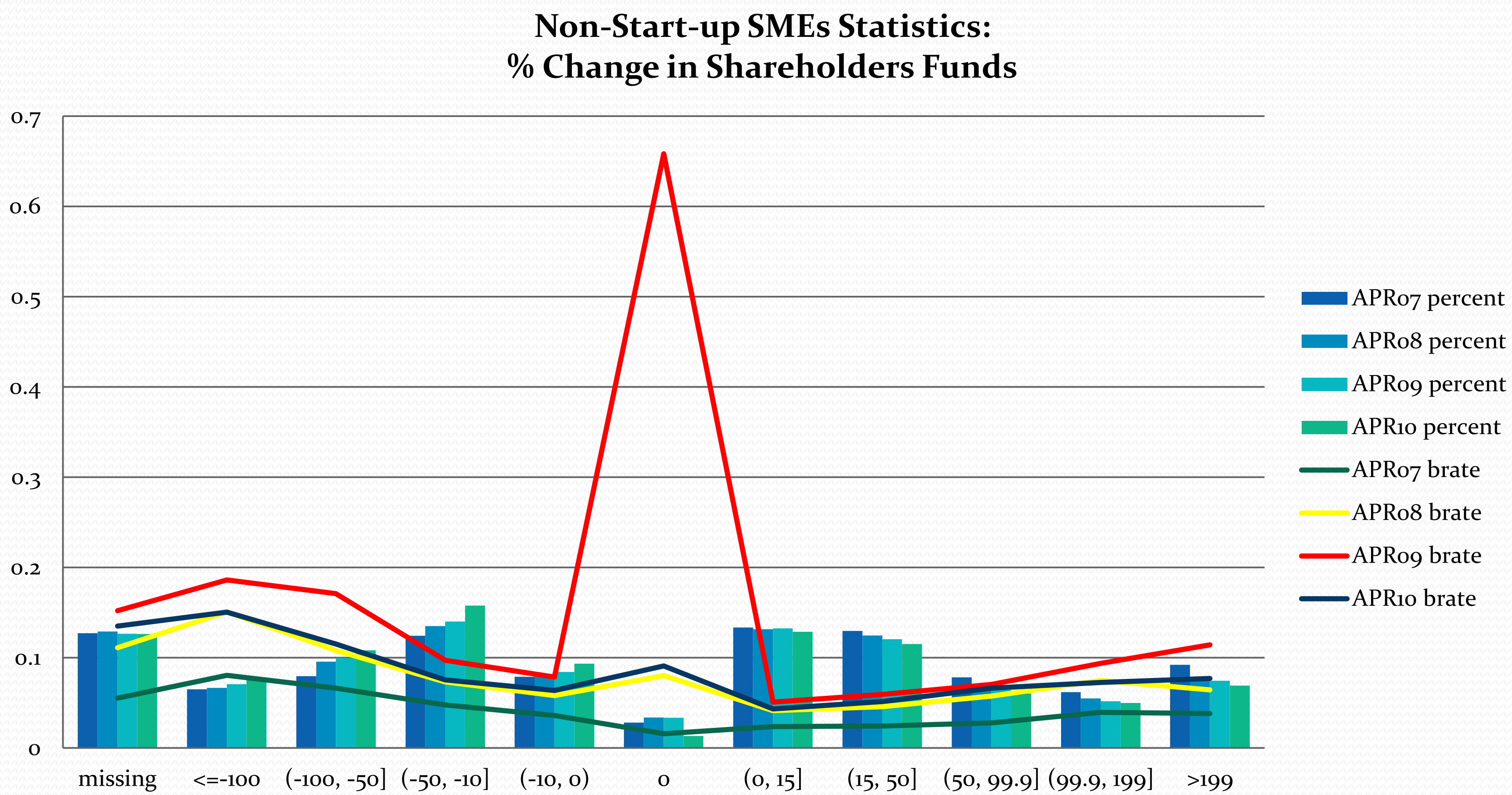
SME's Categorical Performances



SME’s Categorical Performances



SME's Categorical Performances



Use of WoE

- Weight of Evidence:

$$\text{WoE} = \ln\left(\frac{N_i}{P_i}\right) - \ln\left(\frac{\sum N_i}{\sum P_i}\right)$$

- Variable selection: 79 variables, 15 selected for non-start-ups and 12 selected for Start-ups

	Non-start-ups	start-ups
Number of variable selected by at least **one** year's logistic regression	41	37
Number of variable selected by at least **two** year's logistic regression	24	24
Number of variable selected by at least **three** year's logistic regression	13	11
Number of variable selected by all **four** year's logistic regression	10	5

Panel data models

- Model building:
  - Logit panel model: out comes fall between 0 and 1
  - Using same variables as logistic regression
  - Omitted variables:
    - Soft information
    - Business cycles
  - Fixed effect or random effect for firm specific omitted variables

Logit panel model

- Logit panel data model:

$$\Pr(y_{it} = 1 | X_{it} = x_{it}) = \frac{\exp(\beta x_{it} + \alpha_i + \delta_t + u_{it})}{1 + \exp(\beta x_{it} + \alpha_i + \delta_t + u_{it})}$$

- Fixed Effects:

$$\Pr\left(y_i \,\middle|\, \sum_{t=1}^{T_i} y_{it} = k_{1i} \right) = \frac{e^{\sum_{t=1}^{T_i} y_{it} x_{it} \beta}}{\sum_{d \in S_i} e^{\sum_{t=1}^{T_i} y_{it} x_{it} \beta}} = f_i(T_i, k_{1i})$$

  - Small Variation Within independent variables
  - Customers with no change in dependent variables are moved out of estimation: 170,354 observations removed for non-start-ups and only use the left 33.9% for estimation
  - Hausman: both estimators should be **consistent**
  - Assume: soft information is independent of existing variables

Logit panel model

- Random effects: $c_i \sim N(0, \sigma_c^2)$
  - Assuming $\Pr(y_{i1}, y_{i2}, \cdots, y_{in} | \mathrm{x}_{i1}, \mathrm{x}_{i2}, \cdots, \mathrm{x}_{in}) = \int_{\infty} \frac{e^{-c_i^2} / 2\sigma_c^2}{\sqrt{2\pi}\sigma_c} \left\{ \prod_{t=i}^{n_i} F(y_{it}, \mathrm{x}_{it}\beta + c_i) \right\} dc_i$

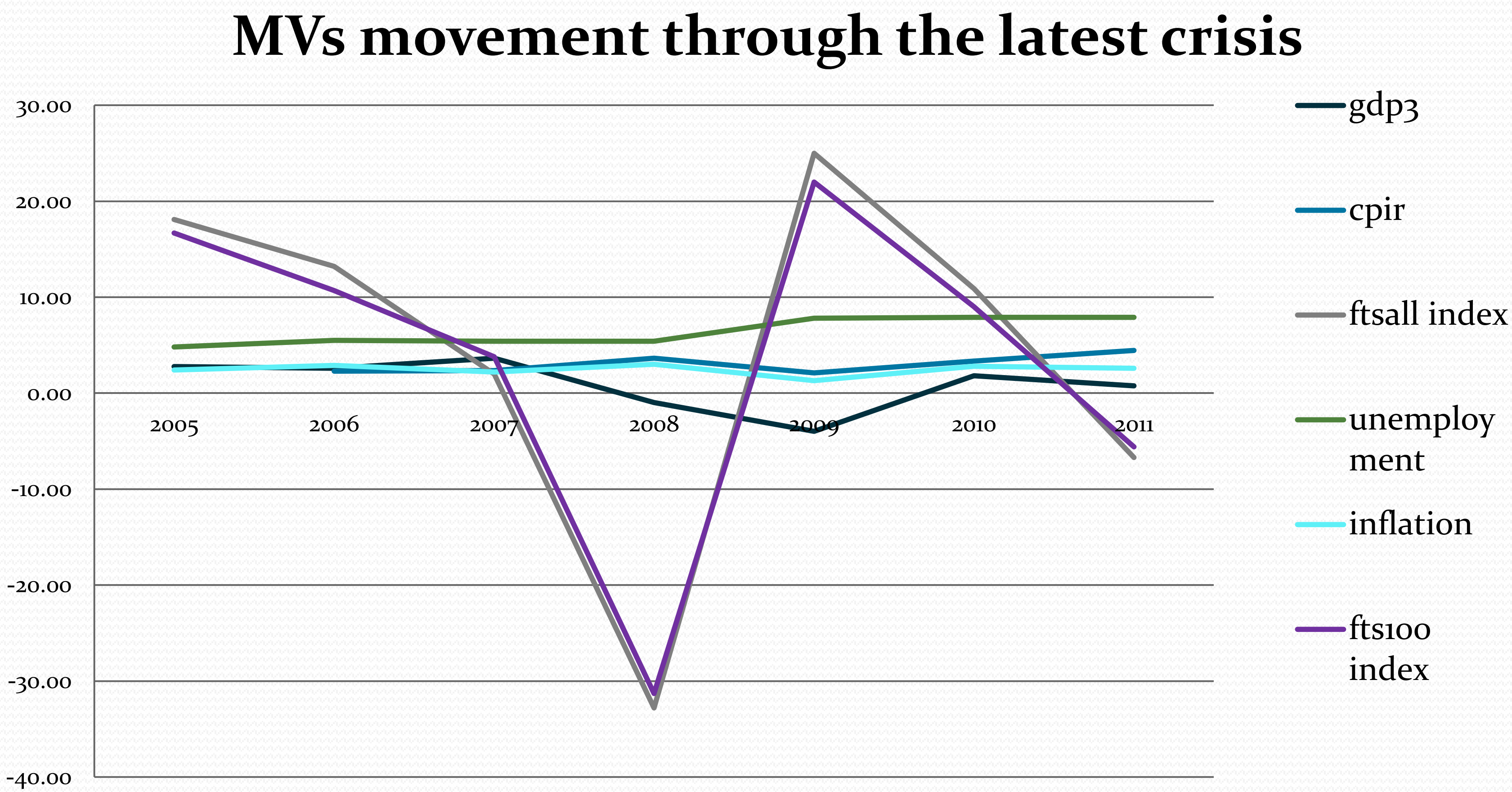
then

$$F(y, z) = \begin{cases} \frac{1}{1+e^{-z}}, & z \neq 0 \\ \frac{1}{1+e^{z}}, & z = 0 \end{cases}$$

$$l_i = \int_{\infty} \frac{e^{-c_i^2} / 2\sigma_c^2}{\sqrt{2\pi}\sigma_c} \left\{ \prod_{t=i}^{n_i} F(y_{it}, \mathrm{x}_{it}\beta + c_i) \right\} dc_i = \int_{\infty} g(y_{it}, \mathrm{x}_{it}, c_i) dc_i$$

Which will be estimated by M-point Gauss-Hermite quadrature

Adding Macroeconomic Variables (MVs)



Adding Macroeconomic Variables

- Figlewski(2010) MVs' influence on US cooperations' default: three group of macroeconomic variables (direction of economy, general economic conditions, financial market.)
- Previous research on credit card scoring using MVs
- lags
  - First: no lag
  - Averaged macroeconomic variables:

$$X_t = \frac{\sum_{K=0}^{2} \delta^K X_{t-k}}{\sum_{K=0}^{2} \delta^K}$$

$where\ \delta$=0.88, and $k = 0, 1, 2.$

Influence of MVs

segments		Start-ups			Non-start-ups		
lags		no lag		Averaged	no lag		Averaged
	MV	one at a time	Best AIC set	Best AIC set	one at a time	Best AIC set	Best AIC set
Direction of economy	GDP growth rate	+	+	+	+	+	+
General economic conditions	CPI rate	-			-		
	unemployment rate	-	-	-	-	-	-
Financial market condition	FTSE100 index changing rate	+	+	+	-	+	+
	FTSEall share index changing rate	+			-		

Correlation within MVs

NON-START-UPS					
	GDP growth rate	CPI changing rate	Unemployment rate	FTSE all share annual return	FTSE100 annual return
GDP growth rate	1				
CPI changing rate	0.075	1			
Unemployment rate	-0.4393	-0.1853	1		
FTSE all share annual return	-0.1489	-0.7448	0.7627	1	
FTSE100 annual return	-0.0974	-0.7744	0.7193	0.9977	1
START-UPS					
	GDP growth rate	CPI changing rate	Unemployment rate	FTSE all share annual return	FTSE100 annual return
GDP growth rate	1				
CPI changing rate	-0.0967	1			
Unemployment rate	-0.4943	-0.1642	1		
FTSE all share annual return	-0.069	-0.775	0.7199	1	
FTSE100 annual return	-0.0084	-0.8073	0.6691	0.9973	1

Pooled Estimator V.S. Panel Estimators

- **Likelihood ratio test:** whether RE estimator is different from pooled estimators:

$$\rho = \frac{\sigma_v^2}{\sigma_v^2 + \sigma_\epsilon^2}$$

denote panel level variance as σ_v^2, σ_ε^2 is the normal residual.

- If ρ is not significantly different from zero, then we cannot reject the hypothesis that panel estimators is the same as pooled estimators
- Logit panel models **only** using same variable as logistic regression: no difference with pooled estimator
- Logit panel models using same variable as logistic regression **and time influence(dummy or MVs)**: different from pooled estimator

Model Comparison

- Training sample: ROC
  - 2007-2010, 10% of original data

non-start-ups					
	Test Result Variable(s)	AUROC			
		2007	2008	2009	2010
logistic regression	pd_log	.819	.855	.894	.859
panel model	pd_pan	.802	.843	.888	.850
panel model with year dummy	pd_pandum	.819	.855	.893	.860
panel model with selected no lagged MV (highest AIC in each category)	pd_L0B3	.819	.855	.893	.860
panel model with selected one year lagged MV (highest AIC in each category)	pd_L1B3	.819	.855	.893	.860
panel model with selected averaged MV (highest AIC in each category)	pd_AVB3	.819	.855	.893	.860
panel model with no lagged GDP_growth rate	pd_L0GDP3	.804	.855	.889	.852
panel model with one year lagged GDP_growth rate	pd_L1GDP3	.805	.855	.889	.853
panel model with averaged GDP_growth rate	pd_AVGDP3	.819	.855	.893	.860

Model Comparison

- Training sample: ROC
  - 2007-2010, 10% of original data

start-ups						
	Test Result Variable(s)	AUROC				
		2007	2008	2009	2010
logistic regression	pd_log	.777	.827	.878	.832
panel model	pd_pan	.784	.829	.880	.837
panel model with year dummy	pd_pandum	.784	.829	.881	.837
panel model with selected no lagged MV (highest AIC in each category)	pd_L0B3	.794	.838	.886	.847
panel model with selected one year lagged MV (highest AIC in each category)	pd_L1B3	.794	.838	.886	.847
panel model with selected averaged MV (highest AIC in each category)	pd_AVB3	.794	.838	.886	.847
panel model with no lagged GDP_growth rate	pd_L0GDP3	.778	.836	.885	.839
panel model with one year lagged GDP_growth rate	pd_L1GDP3	.773	.835	.883	.837
panel model with averaged GDP_growth rate	pd_AVGDP3	.794	.837	.886	.847

Model Comparison

- Holdout Sample: ROC
  - 2010 original data exclude training sample

	Area Under the Curve			
	non		st	
logistic regression	pd_log	.837	pd_log	.753
panel model	pd_pan	.828	pd_pan	.757
panel model with year dummy	pd_pandum	.843	pd_pandum	.769
panel model with selected no lagged MV (highest AIC in each category)	pd_L0B3	.843	pd_L0B3	.758
panel model with selected one year lagged MV (highest AIC in each category)	pd_L1B3	.843	pd_L1B3	.758
panel model with selected averaged MV (highest AIC in each category)	pd_AVB3	.843	pd_AVB3	.758
panel model with no lagged GDP_growth rate	pd_L0GDP3	.833	pd_L0GDP3	.759
panel model with one year lagged GDP_growth rate	pd_L1GDP3	.832	pd_L1GDP3	.758
panel model with averaged GDP_growth rate	pd_AVGDP3	.842	pd_AVGDP3	.758

Thank you!

Q&A