A Case Study of Bank Branch Performance Using Linear Mixed Models

Peggy Ng, Claudia Czado, Eike Brechmann and Jon Kerr

York University, Technische Universität München

August 24, 2010

Bank branch performance assessment

- Branches are the key contact point between customers and the central bank.
- Identify optimal branch network: number and location of branches. $\Rightarrow$ Assessment of branch performance.
- Commonly used: Data Envelope Analysis (DEA).
  - Non-parametric linear programming technique,
  - comparative ratio of inputs to outputs of each branch,
  - variables: in-branch (number of employees, operating expenses,...)
- Here: non-hierarchical linear mixed model.
  - Interaction effects,
  - variables: out-of-branch such as geographical and macroeconomic variables, in particular the influence of local socio-economic variables such as the wealth and local competition.

Linear mixed models

Definition

A linear mixed model is specified as follows:

$$Y_i = X_i\beta + Z_i u_i + \varepsilon_i$$

$$\left.\begin{array}{r} \varepsilon_i \sim \mathrm{N}(0, R_i) \\ u_i \sim \mathrm{N}(0, \Psi) \end{array}\right\} \text{independent}$$

where

$Y_i \in \mathbb{R}^{n_i}$ observations of group i

$X_i \in \mathbb{R}^{n_i \times p}$ design matrix for the fixed effects

$\beta \in \mathbb{R}^{p}$ fixed-effect coefficients

$Z_i \in \mathbb{R}^{n_i \times q}$ design matrix for the random effects

$u_i \in \mathbb{R}^{q}$ random-effect coefficients

$\varepsilon_i \in \mathbb{R}^{n_i}$ errors

$R_i \in \mathbb{R}^{n_i \times n_i}$ covariance matrix for the errors

$\Psi \in \mathbb{R}^{q \times q}$ covariance matrix for the random effects

Error structure

Often: $R_i = \sigma^2 I_{n_i}$.

Extensions:

- Let the dependent variable Y_{it} be time-dependent.

LMM with heterogeneous residual variances σ_t^2

$$\varepsilon_{it} \sim N(0, \sigma_t^2)$$

LMM with $ARMA(p, q)$-model as correlation structure

$$\varepsilon_{it} = \sum_{j=1}^{p} \phi_j \varepsilon_{i,t-j} + \sum_{j=1}^{q} \theta_j a_{t-j} + a_t$$

$\{a_t\}$ = zero mean white noise process with constant variance σ_a^2

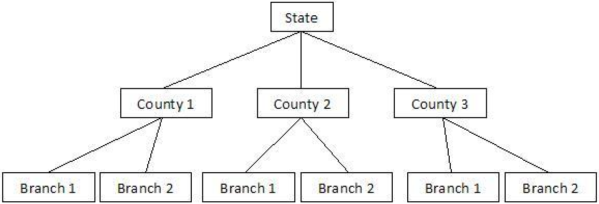
Estimation and testing

- Fixed effects:
  - Estimation usually using restricted maximum likelihood estimation (or standard maximum likelihood estimation).
  - Testing of the $H_0 : \beta_i = 0$ based on t-tests.
- Random effects:
  - Prediction using conditional expectations and estimated covariances.
- Testing with regard to covariance parameters based on likelihood ratio tests (LRT): null distribution is a mixture of χ^2 distributions.

Data

2988 branch-year records of a major US bank in the state of New York with multiple branches. 506 branches with observations over the period from 1994 to 2002.

- The data is clustered (branches within counties within state):



- The data is also longitudinal (observed over a period of 9 years).

$\Rightarrow$ Dependencies in the data.

$\Rightarrow$ Mixed model approach appropriate.

State variables

Constant over counties for each year, different outcomes for each year.

- *no.fail*: the number of branches that closed in NY during the year.
- *mshare*: the market share in NY.
- *branch.total*: the share of the number of branches in NY compared to the USA.
- *dep.total*: the share of the total deposits of the bank in NY compared to the USA.
- *av.dep*: the average deposit per branch in NY.

County variables

Constant over the branches within a county and changing for each year.

- *pop*: the population in the county (in 1000).
- *inc.pc*: the per capita income (in 1000).
- *unemp*: the unemployment rate in the county.

Branch variables

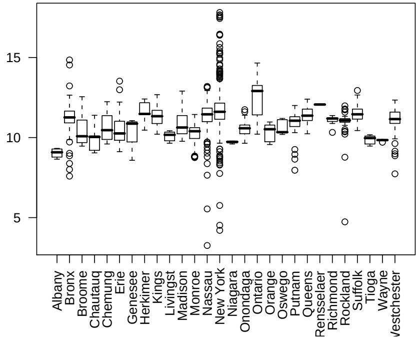
- *log.dep*: the total deposits in log form.
- *comp*: a measure of geographical competition of the branch (different for each year).
  - Scaled and standardized sum of the distances between a branch and all branches of other banks which have only one single branch or multiple branches respectively (0-100% competition).

Dependent variable

- Performance measure of total deposits of a branch.
  - One of the main business drivers of banks.
  - Easily collected and amenable for statistical analysis.

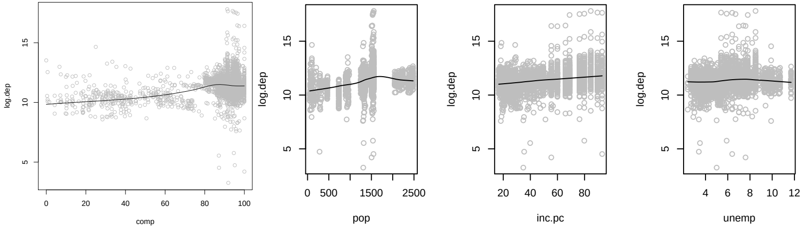
Aim

A model describing the (log-)deposits of branch i in county j in year t: $log.dep_{ijt}$.



Explorative data analysis

- Overall influence of *comp* is weakly positive.
- Weak positive influences of *pop* and *inc.pc*, no influence of *unemp*.



- No clear influence of any of the state variables.

$\Rightarrow$ Very variable effects.

Model formulation and fit: fixed and random effects

- Fixed effects for all branch, county and state variables and their interactions.
  - Select significant effects with t-tests at the 5% level.
- Random intercepts and slopes on the branch level b_{ij0}, b_{ij1}, as well as random intercepts on the county level b_j,
  - Distributions:

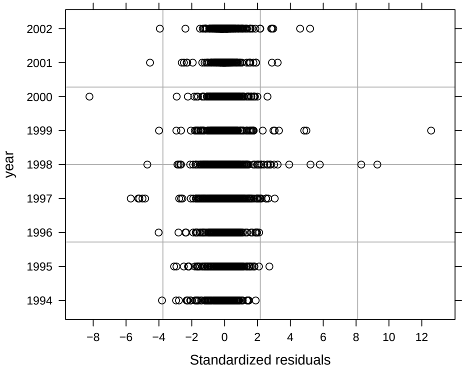
$$u_j \sim N(0, \psi_{00}^2)$$

$$u_{ij} = (u_{ij0}, u_{ij1})^T \sim N_2(0, \Psi) \text{ where } \Psi = \begin{pmatrix} \psi_0^2 & \psi_{01} \\ \psi_{01} & \psi_1^2 \end{pmatrix}.$$

  - Select significant effects with likelihood ratio tests at the 5% level.
- Non-hierarchical model, since random effects b_{ij0}, b_{ij1} are crossed with fixed effects of the county variables, e.g. pop_{jt}.

Variance structure of the errors

Since the number of observations and the values of *log.dep* in each year are varying, the within-group errors might be varying for each year, too.

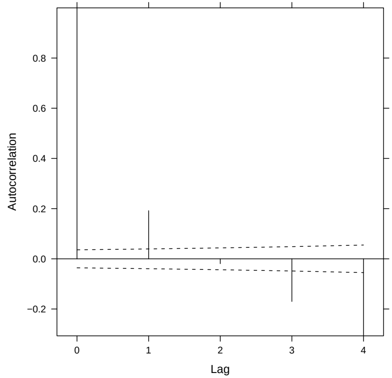


$H_0 : \sigma_t^2 = \sigma^2 \; \forall t$ rejected at the 5% level.

$\Rightarrow$ Include heterogeneous residual variances σ_t^2 for each year t.

Autocorrelation of the within-group errors

Since the observations are taken longitudinally on the same subjects, the within-group errors are probably autocorrelated.



$H_0 : \phi_1 = \theta_1 = 0$ rejected at the 5% level.
$\Rightarrow$ Include $ARMA(1,1)$-model as correlation structure.

Final model

- Fixed effects:

Variable	Estimate	Std. Error	p-value
Intercept	1.12 E+1	4.41 E−1	0.0000
pop	5.77 E−4	8.83 E−5	0.0000
inc.pc	−7.32 E−4	1.44 E−3	0.6121
unemp	−2.69 E−1	6.91 E−2	0.0001
no.fail	5.50 E−4	2.95 E−4	0.0624
mshare	−6.23 E+0	2.23 E+0	0.0054
branch.t	−3.90 E+0	1.80 E+0	0.0303
dep.t	3.44 E+0	1.68 E+0	0.0410
av.dep	1.70 E−6	2.87 E−7	0.0000

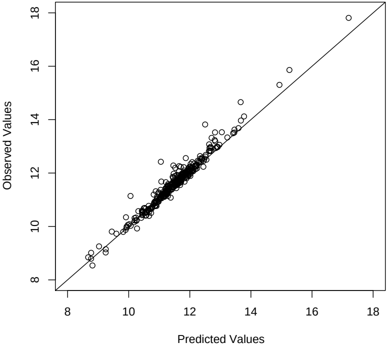
Interact.	Estimate	Std. Error	p-value
unemp× *no.fail*	−1.04 E−4	4.39 E−5	0.0184
unemp× *mshare*	1.37 E+0	3.67 E−1	0.0002
unemp× *branch.t*	1.05 E+0	2.82 E−1	0.0002
unemp× *dep.t*	−9.43 E−1	2.66 E−1	0.0004
inc.pc× *av.dep*	1.05 E−8	3.94 E−9	0.0078

- Random intercepts on the county level are not significant.
- Model diagnostics:
  - ✓ Residuals scatter around 0: 93.9% of all observations in [-2,2]-interval.
  - ✓ Assumption of normality appropriate for most years.
  - ✓ Zero mean and normality assumptions for random effects are plausible.

Check of the predictive capability

Approach:

- The final model is estimated with the data of 1994 to 2001.
- The values of 2002 are predicted using this restricted model.

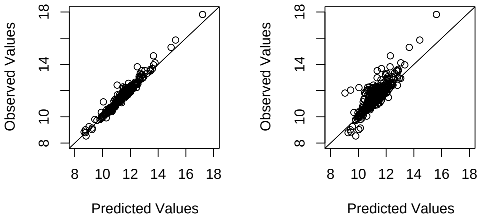


Points approximately lie on the line $y = x$. $\Rightarrow$ Quite good prediction.

Comparison to alternative models

Is the mixed model an improvement in the model fit compared to a linear model, to generalized least squares (GLS) model with heteroscedastic and correlated within-group errors (but no random effects) and to a hierarchical mixed model with the same error structure?

- The *AIC* of the linear mixed model is much smaller (531 vs. 7937 of the linear model and 3675 of the GLS model).
- The predictive capability of the hierarchical model is inferior to that of the non-hierarchical model.



⇒ Inclusion of random effects and flexible non-hierarchical structure significantly improve the fit.

Macroeconomic effects

- Positive effects of the market share, the average deposit per bank and the share of the number of branches in NY compared to the USA.
    1. ✓
    2. ✓
    3. No obvious explanation.
- Negative effects of the number of branches that closed during the year and the share of the total deposits in NY compared to the USA.
    1. Closures of branches $\leftrightarrow$ bad economic environment.
    2. No obvious explanation.

Geographical effects

- Positive effects of the county's population and the per capita income.
  1. population $\uparrow \Rightarrow$ deposits $\uparrow$
  2. income $\uparrow \Rightarrow$ deposits $\uparrow$
- Unclear effects of the unemployment rate and the local competition.
  1. Possible explanations:
     - Unemployed people have less cash flow (negative effect).
     - Unemployed people and people threatened by unemployment save more money because of the financial insecurity (positive effect).
  2. $\rightarrow$ next slide

Local competition

- No uniform influence of the geographical competition: *opposing trends*, if the competition increases.
    - Competition stimulates business.
    - More branches, same population $\Rightarrow$ less deposits.

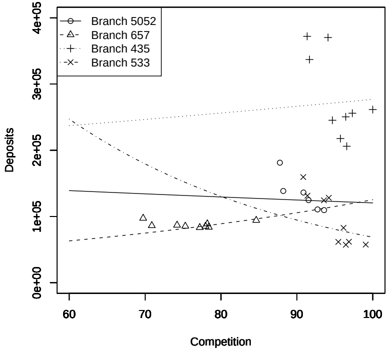
$\Rightarrow$ Classify branches as being in a "rural"/"urban", "rich"/"poor" area with low/high unemployment and compare average effect of competition.

- Stronger effect in rural areas: more likely to be influenced by marketing activities.
- Stronger effect in rich areas: choose banks more deliberately.
- Stronger effect in areas with high unemployment: more worried about money, susceptible to competing offers.

Branch-specific effects

- Some branches have more deposits than others if all other influences are disregarded: e.g. long-term customer loyalty or a particular good location in an area.
- Varying influence of the geographical competition on the deposits: see previous slide.

Influence of these branch-specific effects for four randomly chosen branches from Rockland (533), Suffolk (657), Nassau (5052) and New York (435):



Conclusion and outlook

- Regression analysis ...
  - summarizes information about *all* branches in the sample,
  - directly indicates causes of low performance, and
  - can be used to forecast deposits of new branches.

$\Rightarrow$ Easy evaluation of a single existing branch and of the potential of a new location.

- Outlook:
  - Use other performance variables such as fee income or the number of new deposit and/or lending accounts.
  - Include in-branch variables or competitive factors.

Bibliography I

N. K. Avkiran.
Models of retail performance for bank branches: predicting the level of key business drivers..
International Journal of Bank Marketing 15 (6), 224-237.

A. N. Berger and D. B. Humphrey.
Efficiency of financial institutions: International survey and directions for future research..
European Journal of Operational Research 98 (2), 175-212.

P. V. Boufounou.
Evaluating bank branch location and performance: A case study..
European Journal of Operational Research 87, 389-402.

Bibliography II

E. C. Brechmann, C. Czado and P. Ng.
Geographical and macroeconomic effects on bank branch deposits.
International Journal of Statistics and Management System, to appear, 2010.

J. C. Pinheiro and D. M. Bates.
Mixed-Effects Models in S and S-PLUS.
Springer, New York, 2nd edition, 2000.

B. T. West, K. B. Welch, and A. T. Galecki.
Linear Mixed Models: A Practical Guide Using Statistical Software.
Chapman & Hall/CRC, Boca Raton, 1st edition, 2007.