# Event History Analysis for Debt Collection Portfolios

Fanyin Zhou[1] Nick Heard [2] David Hand[1,2]

1. Institute for Mathematical Sciences, 2. Department of Mathematics
Imperial College London

Credit Scoring and Credit Control XI Conference

August 2009

## Consumer debt sales

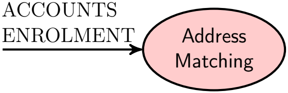
- **Debt type**: e.g. delinquent credit card payments and personal loans
- **Major players**:
  - Debt sellers: major banks and credit lenders
  - Debt buyers: debt purchase and collection companies
- **Transaction method**: closed tenders / public auctions
- **Contract types**: *one-off inventory sales* and *forward-flow agreements*
- **Portfolio composition**: *arrangement* and *general*
- **Portfolio price**: a fraction of debt face value

# Data sample

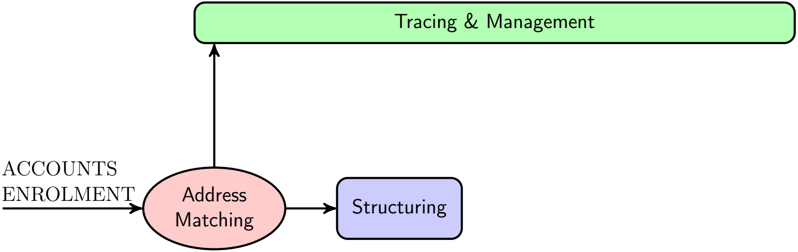
- 6,000$^+$ credit card accounts from 24 sequentially enrolled arrangement portfolios in the years 2002 and 2003.
- each account having
  - details of the customer (*account information*)
  - dates and amounts of payments made to the debt recovery company (*transaction details*)
  - records of all contacts made between account customer and the debt recovery company (*action records*)

  all up until Dec 2006.

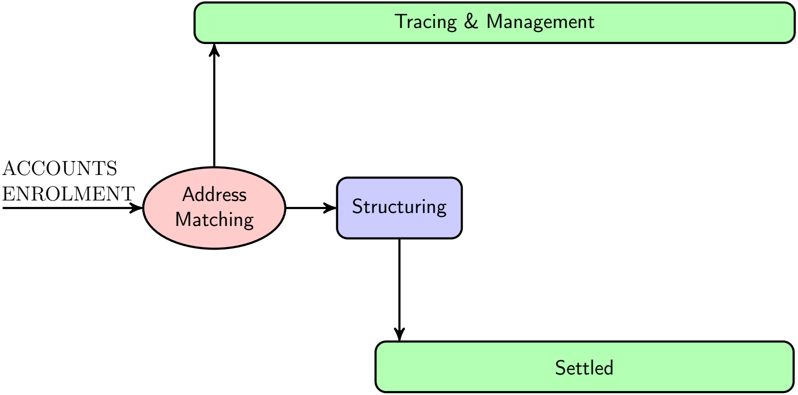
# A typical debt portfolio collection process



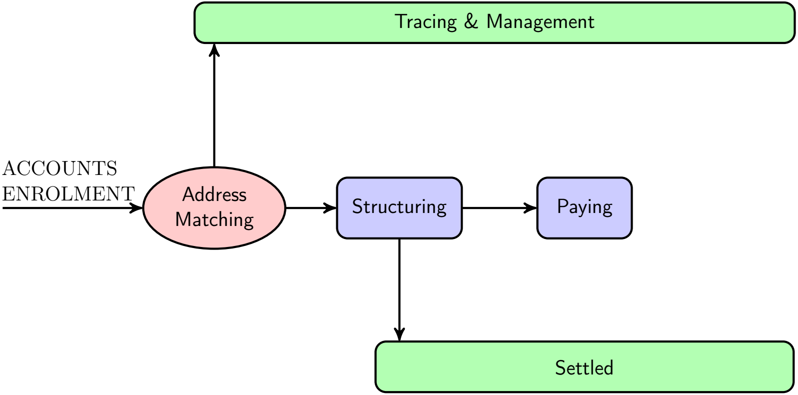
# A typical debt portfolio collection process



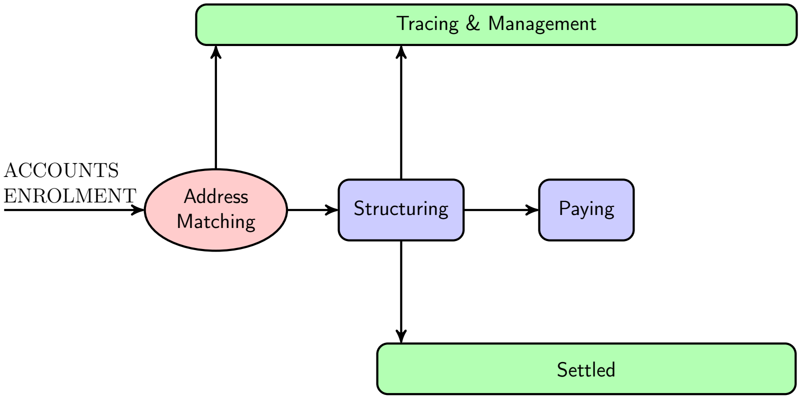
# A typical debt portfolio collection process



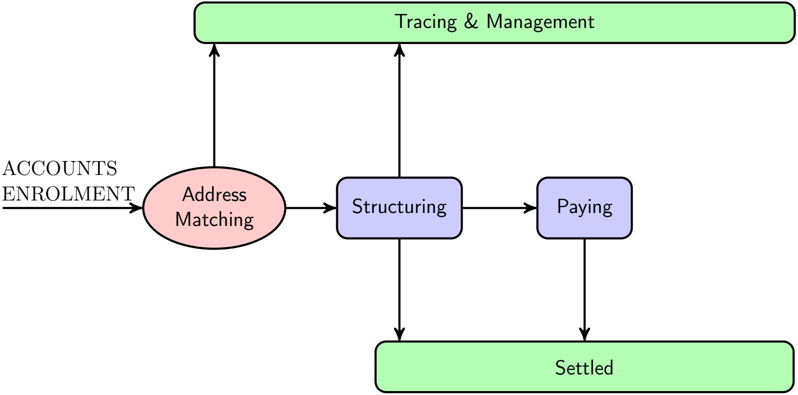
# A typical debt portfolio collection process



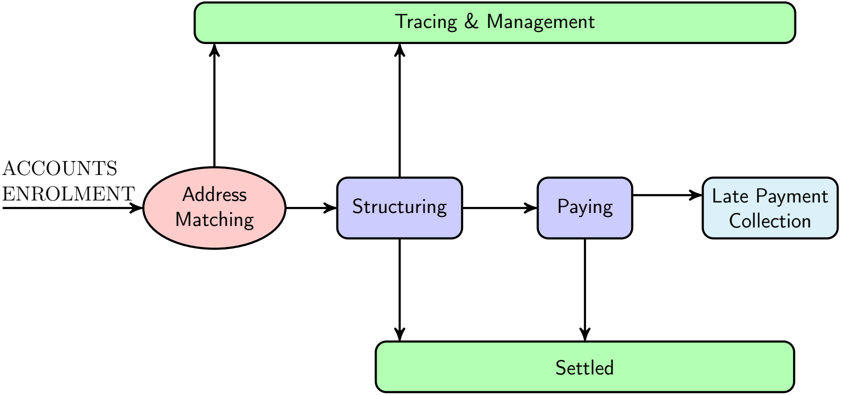
# A typical debt portfolio collection process



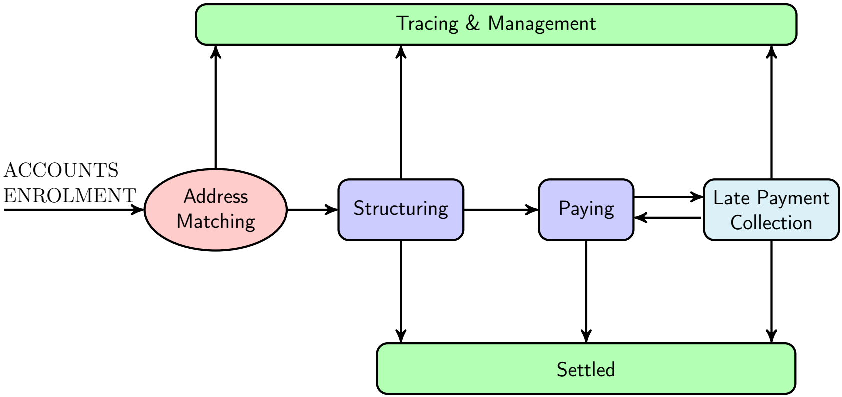
# A typical debt portfolio collection process



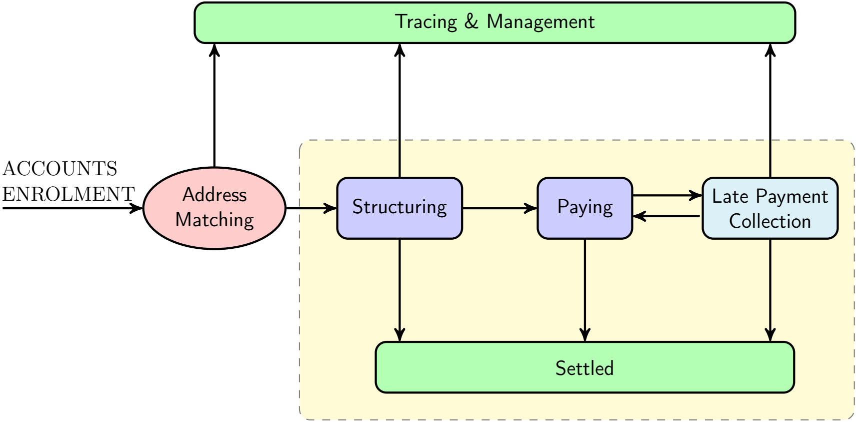
# A typical debt portfolio collection process



# A typical debt portfolio collection process



# A typical debt portfolio collection process



## Multi-state models : Formulation

**State structure** specifies the states and the possible transitions between states.

For a given data set,

- The state structure is **NOT** unique;
- Selecting a good state structure makes the data analysis more approachable.

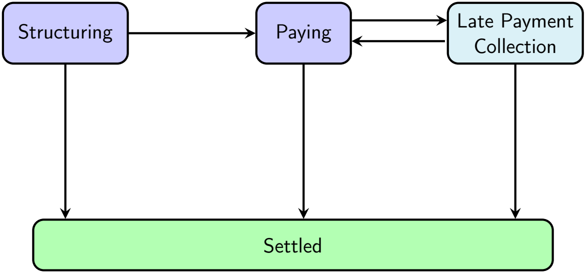
## Multi-state models : Formulation

**State structure** specifies the states and the possible transitions between states.

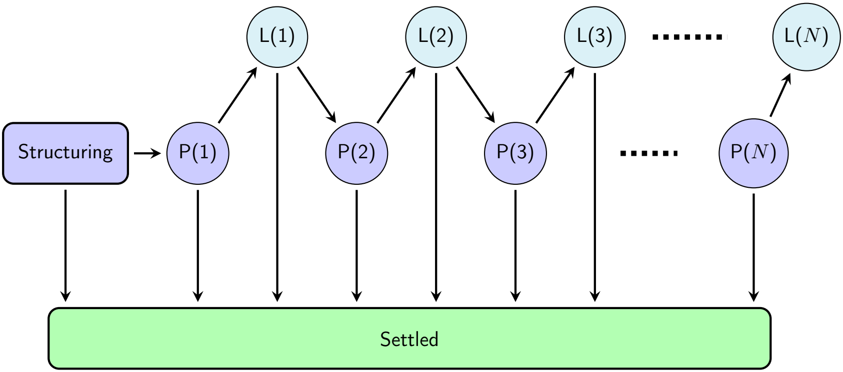
For a given data set,

- The state structure is **NOT** unique;
- Selecting a good state structure makes the data analysis more approachable.

# The initial formulation



# The *unfolded* formulation



# Possible factors

For each *Paying state* $P(i)$ or *Late-payment-collection state* $L(i)$ $(i = 1, \ldots, N)$, we have a list of possible factors to be tested in the regression models:

- **Background variables**: Age, Gender, Balance, Debt grade, Type of credit card, etc.
- **Performance variables**: Times of earlier transitions, Number of contacts made in earlier states, etc.

## Regression models

For each $P(i)$ or $L(i)$ $(i = 1, \ldots, N)$ state , we have a **competing risks model**:

The risk of proceeding to next $P$ or $L$ state
VS.
The risk of settlement.

# Regression models

Regression models for the sub-distribution hazard:

- Cox regression model:

$$h_k(t|\mathbf{X}) = h_{k,0}(t)\exp(\beta_k^T\mathbf{X}) \qquad k = 1,2$$

- Cox regression model with time-dependent covariates:

$$h_k(t|\mathbf{X}(t)) = h_{k,0}(t)\exp(\beta_k^T\mathbf{X}(t))$$

- Aalen Additive regression model:

$$h_k(t|\mathbf{X}(t)) = Y(t)(\alpha_k(t)^T\mathbf{X}(t))$$

# Regression models

Regression models for the sub-distribution hazard:

- Cox regression model:

$$h_k(t|\mathbf{X}) = h_{k,0}(t)\exp(\beta_k^T \mathbf{X}) \qquad k = 1, 2$$

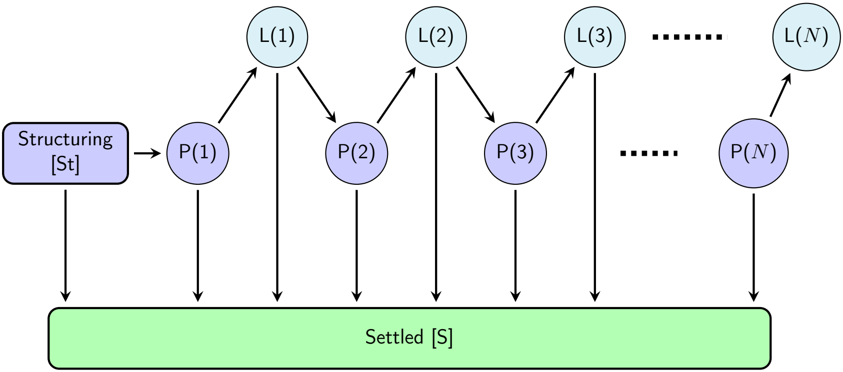
- Cox regression model with time-dependent covariates:

$$h_k(t|\mathbf{X}(t)) = h_{k,0}(t)\exp(\beta_k^T \mathbf{X}(t))$$

- Aalen Additive regression model:

$$h_k(t|\mathbf{X}(t)) = Y(t)(\alpha_k(t)^T \mathbf{X}(t))$$

# The *unfolded* model



# Results: Stepwise variable selection

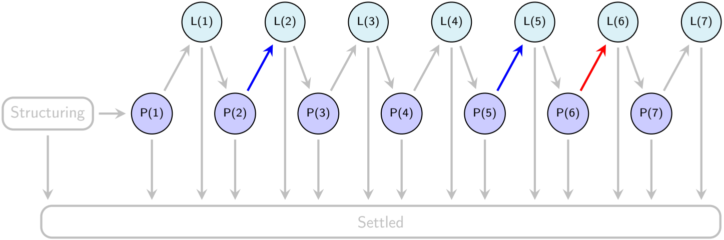
| | P1L1 | P2L2 | P3L3 | P4L4 | P5L5 | P6L6 |
|---|---|---|---|---|---|---|
| StP1 time | 0.09 | 0.10 | | | | |
| P1L1 time | | -0.24 | -0.11 | | | |
| L1P2 time | | 0.38 | | | | |
| P2L2 time | | | -0.34 | -0.16 | | -0.38 |
| L2P3 time | | | | | | |
| P3L3 time | | | | -0.31 | | |
| L3P4 time | | | | | | |
| P4L4 time | | | | | -0.24 | |
| L4P5 time | | | | | | |
| P5L5 time | | | | | | -0.004 |
| L5P6 time | | | | | | |

| | L1P2 | L2P3 | L3P4 | L4P5 | L5P6 | L6P7 |
|---|---|---|---|---|---|---|
| StP1 time | -0.19 | | | | | -1.47 |
| P1L1 time | 0.05 | | 0.10 | | | |
| L1P2 time | | -0.30 | -0.40 | | | |
| P2L2 time | | 0.17 | | | | |
| L2P3 time | | | -0.32 | | | |
| P3L3 time | | | 0.12 | 0.33 | | 1.00 |
| L3P4 time | | | | -0.80 | | 3.43 |
| P4L4 time | | | | 0.21 | | |
| L4P5 time | | | | | | |
| P5L5 time | | | | | 1.00 | |
| L5P6 time | | | | | | |
| P6L6 time | | | | | | |

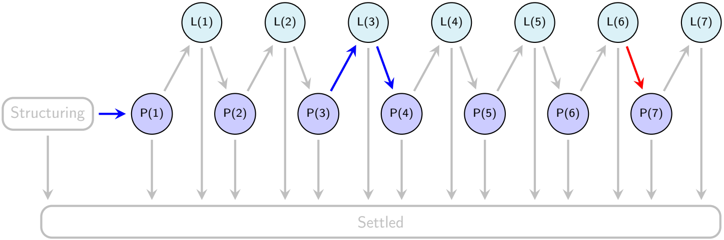
| | P1L1 | P2L2 | P3L3 | P4L4 | P5L5 | P6L6 |
|---|---|---|---|---|---|---|
| # actions in St | 0.11 | 0.06 | | | -0.16 | |
| # actions in P1 | | 0.09 | | | | |
| # actions in L1 | | -0.26 | | -0.21 | | -0.59 |
| # actions in P2 | | | 0.11 | | | |
| # actions in L2 | | | | | | |
| # actions in P3 | | | | | | |
| # actions in L3 | | | | | | |
| # actions in P4 | | | | | | |
| # actions in L4 | | | | | | |
| # actions in P5 | | | | | | |
| # actions in L5 | | | | | | |

| | L1P2 | L2P3 | L3P4 | L4P5 | L5P6 | L6P7 |
|---|---|---|---|---|---|---|
| £ leaving-St payment | -0.08 | | | | | |
| £ leaving-L1 payment | | -0.22 | | | -1.09 | |
| £ leaving-L2 payment | | | -0.29 | | 1.70 | |
| £ leaving-L3 payment | | | | -0.25 | | |
| £ leaving-L4 payment | | | | | | |
| £ leaving-L5 payment | | | | | | |

| | P1L1 | P2L2 | P3L3 | P4L4 | P5L5 | P6L6 |
|---|---|---|---|---|---|---|
| StP1 time | 0.09 | 0.10 | | | | |
| P1L1 time | | -0.24 | -0.11 | | | |
| L1P2 time | | 0.38 | | | | |
| P2L2 time | | | -0.34 | -0.16 | | -0.38 |
| L2P3 time | | | | | | |
| P3L3 time | | | | -0.31 | | |
| L3P4 time | | | | | | |
| P4L4 time | | | | | -0.24 | |
| L4P5 time | | | | | | |
| P5L5 time | | | | | | -0.004 |
| L5P6 time | | | | | | |



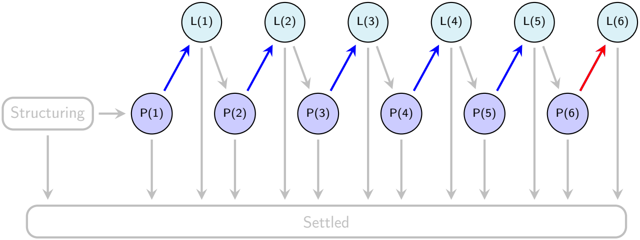
| | L1P2 | L2P3 | L3P4 | L4P5 | L5P6 | L6P7 |
|---|---|---|---|---|---|---|
| StP1 time | -0.19 | | | | | -1.47 |
| P1L1 time | 0.05 | | 0.10 | | | |
| L1P2 time | | -0.30 | -0.40 | | | |
| P2L2 time | | 0.17 | | | | |
| L2P3 time | | | -0.32 | | | |
| P3L3 time | | | 0.12 | 0.33 | | 1.00 |
| L3P4 time | | | | -0.80 | | 3.43 |
| P4L4 time | | | | 0.21 | | |
| L4P5 time | | | | | | |
| P5L5 time | | | | | 1.00 | |
| L5P6 time | | | | | | |
| P6L6 time | | | | | | |



# Tailored stepwise variable selection

To facilitate the interpretation of covariate effects, we

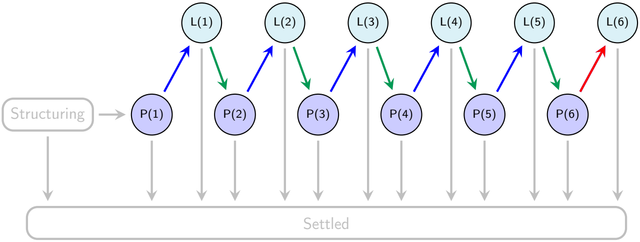
- only allow the $p$th lag of covariate $x$ to be considered in the selection procedure when lags $1, 2, \ldots, p-1$ are also included in the model.



## Tailored stepwise variable selection

To facilitate the interpretation of covariate effects, we

- only allow the $p$th lag of covariate $x$ to be considered in the selection procedure when lags $1, 2, \ldots, p-1$ are also included in the model.



# Results: Tailored stepwise variable selection

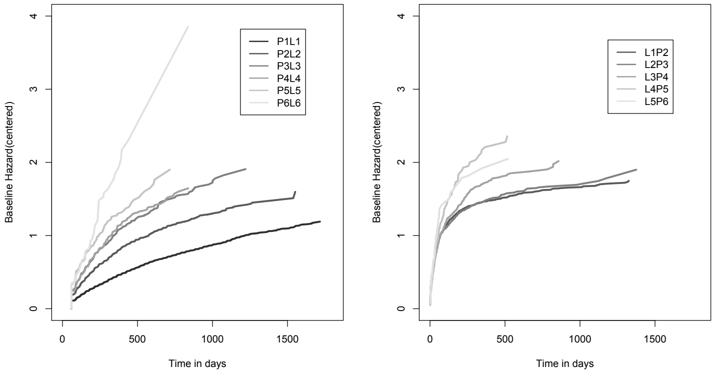
| | P1L1 | P2L2 | P3L3 | P4L4 | P5L5 | P6L6 |
|---|---|---|---|---|---|---|
| StP1 time | 0.09 | 0.10 | | | | |
| P1L1 time | | -0.24 | -0.16 | | | |
| L1P2 time | | 0.38 | | | | |
| P2L2 time | | | -0.42 | -0.15 | | |
| L2P3 time | | | | | | |
| P3L3 time | | | | -0.34 | | |
| L3P4 time | | | | | | |
| P4L4 time | | | | | -0.27 | |
| L4P5 time | | | | | | |
| P5L5 time | | | | | | -0.003 |
| L5P6 time | | | | | | |

| | L1P2 | L2P3 | L3P4 | L4P5 | L5P6 | L6P7 |
|---|---|---|---|---|---|---|
| StP1 time | -0.19 | | -0.12 | | | |
| P1L1 time | 0.05 | | | | | |
| L1P2 time | | -0.30 | -0.42 | | | |
| P2L2 time | | 0.17 | | | | |
| L2P3 time | | | -0.35 | | | |
| P3L3 time | | | 0.11 | | | |
| L3P4 time | | | | -0.52 | | |
| P4L4 time | | | | | | |
| L4P5 time | | | | | | |
| P5L5 time | | | | | 0.002 | |
| L5P6 time | | | | | | |
| P6L6 time | | | | | | |

| | P1L1 | P2L2 | P3L3 | P4L4 | P5L5 | P6L6 |
|---|---|---|---|---|---|---|
| # actions in St | 0.11 | 0.06 | | | | |
| # actions in P1 | | 0.09 | | | | |
| # actions in L1 | | -0.26 | | | | |
| # actions in P2 | | | 0.12 | | | |
| # actions in L2 | | | | | | |
| # actions in P3 | | | | | | |
| # actions in L3 | | | | | | |
| # actions in P4 | | | | | | |
| # actions in L4 | | | | | | |
| # actions in P5 | | | | | | |
| # actions in L5 | | | | | | |

| | L1P2 | L2P3 | L3P4 | L4P5 | L5P6 | L6P7 |
|---|---|---|---|---|---|---|
| £ leaving-St payment | -0.08 | | | | | |
| £ leaving-L1 payment | | -0.22 | | | | |
| £ leaving-L2 payment | | | -0.29 | | | |
| £ leaving-L3 payment | | | | -0.40 | | |
| £ leaving-L4 payment | | | | | | |
| £ leaving-L5 payment | | | | | | |

# Baseline hazards



# Summary

In conclusion, we

- proposed a multi-state framework for the debt collection process,
- explored a state structure which allows us to add performance variables into regression models, and
- implemented a tailored variable selection algorithm to achieve improved interpretability of regression results.

# Thank you!