UNIVERSITY OF
Southampton
School of Management

Monitoring relationship between score and odds in a propensity scorecard

Katarzyna Bijak

Credit Scoring and Credit Control XI
August 2009, Edinburgh

UNIVERSITY OF Southampton School of Management

Outline

- Propensity scoring
- Methodology
  - Kalman-filter-based monitoring
  - Relationship of score and log odds
- Empirical example
- Conclusions

Propensity scoring

- Propensity scorecards
  - similar to credit scorecards, allow forecasting which of the bank's customers will soon be interested in new credits
  - enable predicting customers' willingness to make application for new loans (credit propensity level)
  - are considered efficient tools to select customers for the bank's marketing campaigns (e.g. direct-mail ones)

UNIVERSITY OF Southampton
School of Management

Propensity scoring

- Customer's status and the odds
  - customers are divided into the **willing** and the **unwilling** to apply for new loans (their status in the outcome point)
  - the willing customers are defined as those who made application for new loans in a four-month outcome period
  - the odds are defined as a ratio of the unwilling to the willing among customers having a given score or a score coming from a given range (measure of credit propensity)

Methodology: Kalman-filter-based monitoring

- Scorecard monitoring based on the Kalman filtering
  - Whittaker J., Whitehead C., Somers M. (2007): A dynamic scorecard for monitoring baseline performance with application to tracking a mortgage portfolio, *J Opl Res Soc*
  - assume that the model parameters (constituting a process state estimated with the Kalman filter) change constantly and the monitoring samples provide their measurements
  - measurements are used to update the baseline model and the updated scorecard is the output of the Kalman filter

Methodology: Kalman-filter-based monitoring

- Kalman-filter-based monitoring – what’s new here?
  - a more general approach is used to monitor a scorecard
  - there are no assumptions on the specification of a scoring model, nor on the estimation method (a common problem in practice, while using some commercial software)
  - as the estimator features are unknown, the estimator covariance matrix is derived from the bootstrap

Methodology: Kalman-filter-based monitoring

- The model parameters β_t constitute the state of a process
- The **state equation** is a multi-dimensional random walk:

$$\beta_t = \beta_{t-1} + q_t$$

- … where covariance Q is a diagonal matrix:

$$\mathrm{var}(q_t) = Q = \sigma I$$

- … and a signal to noise ratio is assumed to be very low:

$$\sigma = 0.00001$$

Methodology: Kalman-filter-based monitoring

- The estimates $\hat{\beta}_t^m$ are treated as parameter measurements
- The **observation equation** has the following form:

$$\hat{\beta}_t^m = \beta_t + r_t$$

- ... where R_t is the estimator covariance matrix:

$$\text{var}(r_t) = R_t$$

- It is assumed that the estimator is unbiased and follows an asymptotic multivariate normal distribution:

$$\hat{\beta}_t^m \sim \text{N}(\beta_t, R_t)$$

Methodology: Kalman-filter-based monitoring

- The form of the estimator covariance matrix R_t is unknown
- It is not known whether the estimator is the most efficient one and therefore it would be unjustified to assume that it is the inverse Fisher information matrix
- An estimate of the matrix R_t is derived from the bootstrap:
  - a sample is chosen using proportional sampling with replacement
  - the parameters are estimated and then the estimates are collected
  - the collected estimates are used to compute the covariance matrix

UNIVERSITY OF Southampton
School of Management

Methodology: Kalman-filter-based monitoring

- The **time update equations** yield the *a priori* estimates
- The *a priori* estimates:

$$\hat{\beta}_t^- = \hat{\beta}_{t-1}$$

- The *a priori* error covariance matrix:

$$P_t^- = P_{t-1} + Q$$

- The Kalman gain:

$$K_t = P_t^- (P_t^- + R_t)^{-1}$$

UNIVERSITY OF Southampton
School of Management

Methodology: Kalman-filter-based monitoring

- Both the *a priori* estimates and the measurements (model parameter estimates based on the monitoring sample S_t) are used in the **measurement update equations** to obtain the *a posteriori* estimates

- The *a posteriori* estimates (updated scorecard):

$$\hat{\beta}_t = \hat{\beta}_t^- + K_t(\hat{\beta}_t^m - \hat{\beta}_t^-)$$

- The *a posteriori* error covariance matrix:

$$P_t = (I - K_t)P_t^-$$

UNIVERSITY OF Southampton
School of Management

Methodology: Kalman-filter-based monitoring

- The initial *a posteriori* estimates are assumed to be the baseline model parameter estimates (based on sample S_0):

$$\hat{\beta}_0 = \hat{\beta}_0^b$$

- The initial *a posteriori* error covariance matrix should be such that the influence of those estimates is limited:

$$P_0 = 10000 I$$

- The updated scorecard depends more on the estimates from the monitoring sample than on the baseline model

Methodology: relationship of score and log odds

- Relationship between the score and the log odds
  - it is a common practice that once a scorecard has been developed, an additional linear model is estimated
  - the linear relationship is used to determine the customer's credit propensity level according to the scorecard
  - the relationship is often used to scale the scorecard (when an institution wants several scorecards to be consistent)
  - based on that relationship a cutoff can be determined

Methodology: relationship of score and log odds

- The linear model for the **baseline scorecard**:

$$\ln(\hat{o}_i) = \hat{a}_0^b + \hat{b}_0^b \cdot s_i^b$$

- … where s_i^b is a score coming from that scorecard
- The parameters are estimated on the baseline sample S_o
- The above relationship is used to determine the customer’s credit propensity according to the baseline scorecard:
  - a customer *i* with the baseline score s_i^b is willing to apply for new loans at the level corresponding to the odds $\hat{o}_i$ (**point estimation**)

UNIVERSITY OF Southampton
School of Management

Methodology: relationship of score and log odds

- The model is estimated based on the mid-points and the log odds of *m* score ranges into which the score range is divided

- The log odds confidence interval (**interval estimation**):

$$(l_i^l, l_i^u) = \left(\ln(\hat{o}_i) - t_* S_i, \ln(\hat{o}_i) + t_* S_i\right)$$

- E.g. the 90% confidence interval is derived from:

$$\mathrm{P}\left\{\ln(\hat{o}_i) - t_* S_i < \ln(o_i) < \ln(\hat{o}_i) + t_* S_i\right\} = 0.9$$

- … where t_* is the Student's distribution value with *m* – 2 degrees of freedom and S_i is the *ex ante* forecast error

Methodology: relationship of score and log odds

- The linear model for the **updated scorecard**:

$$\ln(\hat{o}_i) = \hat{a}_t + \hat{b}_t \cdot s_i$$

- … where s_i is a score coming from that scorecard
- The parameters are estimated on the monitoring sample S_t
- The above relationship is used to determine the customer’s credit propensity according to the updated scorecard:
  – a customer *i* with the updated score s_i is willing to apply for new loans at the level corresponding to the odds $\hat{o}_i$ (**point estimation**)

Methodology: relationship of score and log odds

- Log odds vs. score relationship monitoring
  - the baseline linear model → the 90% confidence interval of the log odds gained using the customer's baseline score
  - the updated linear model → the log odds (point) estimate determined on the basis of the customer's updated score
  - provided that the baseline scorecard is still up-to-date, the log odds (point) estimate should lie within the interval:

    For what % of customers in the monitoring sample it is not true?

UNIVERSITY OF Southampton
School of Management

Empirical example

- Data samples
  - propensity scorecard developed on the credit bureau data
  - the baseline sample: the observation point is Sep-05 while the outcome point is Jan-06 → the baseline scorecard
  - 11 monitoring samples: the observation points are: Oct-05, Nov-05, ..., Aug-06 while the outcome points are: Feb-06, Mar-06, ..., Dec-06 → measurements for the Kalman filter
  - in the baseline and monitoring samples the odds are ca 4

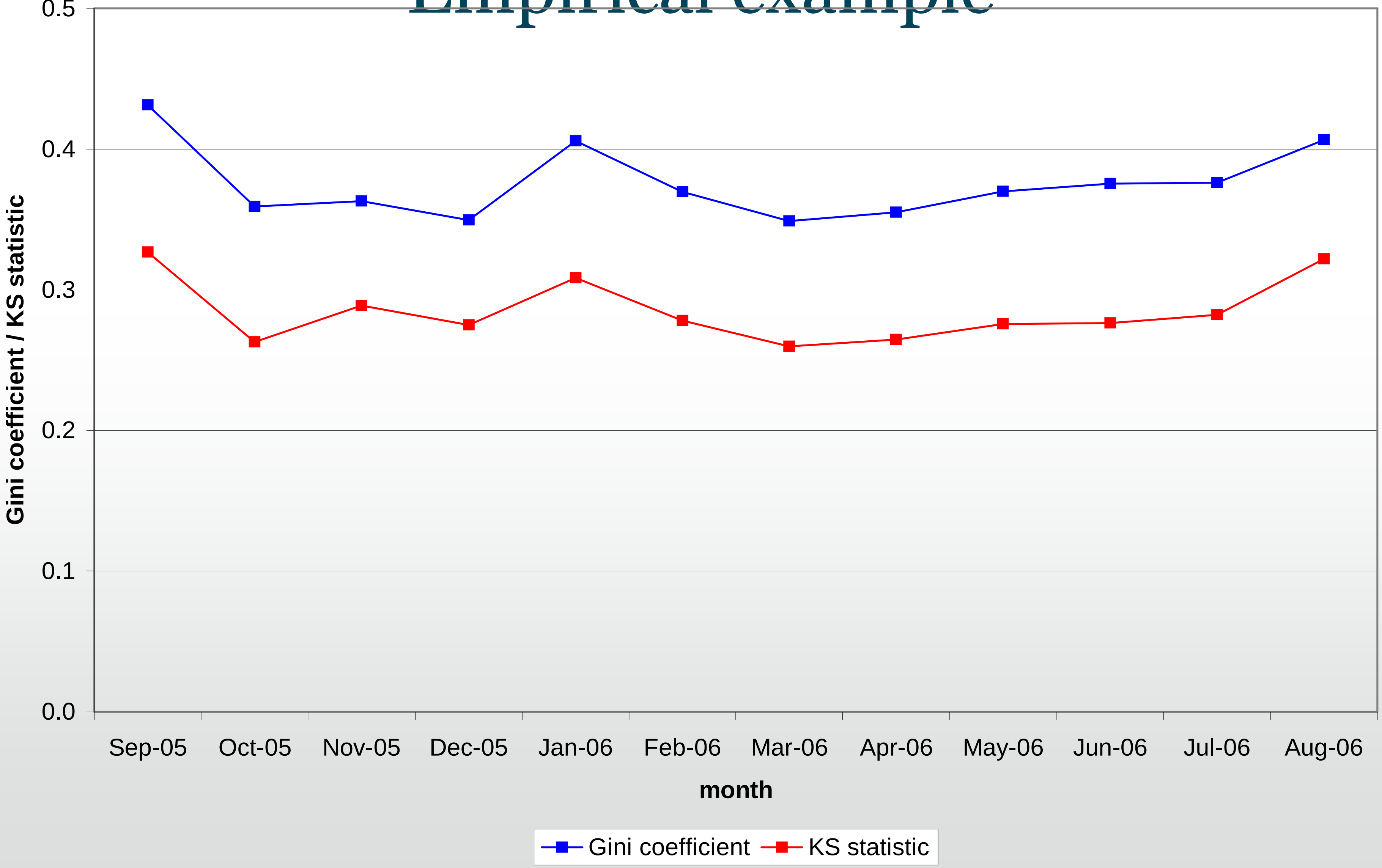
UNIVERSITY OF Southampton
School of Management

Empirical example

- The scorecard has 9 characteristics, such as e.g.:
  - number of credit inquiries within the last 12 months
  - number of loans granted within the last 12 months
  - number of past loans
  - time since the last credit inquiry
- The baseline scorecard discriminatory power:
  - the Gini coefficient = 0.43
  - the KS statistic = 0.33
- The baseline scorecard stability confirmed on a test dataset

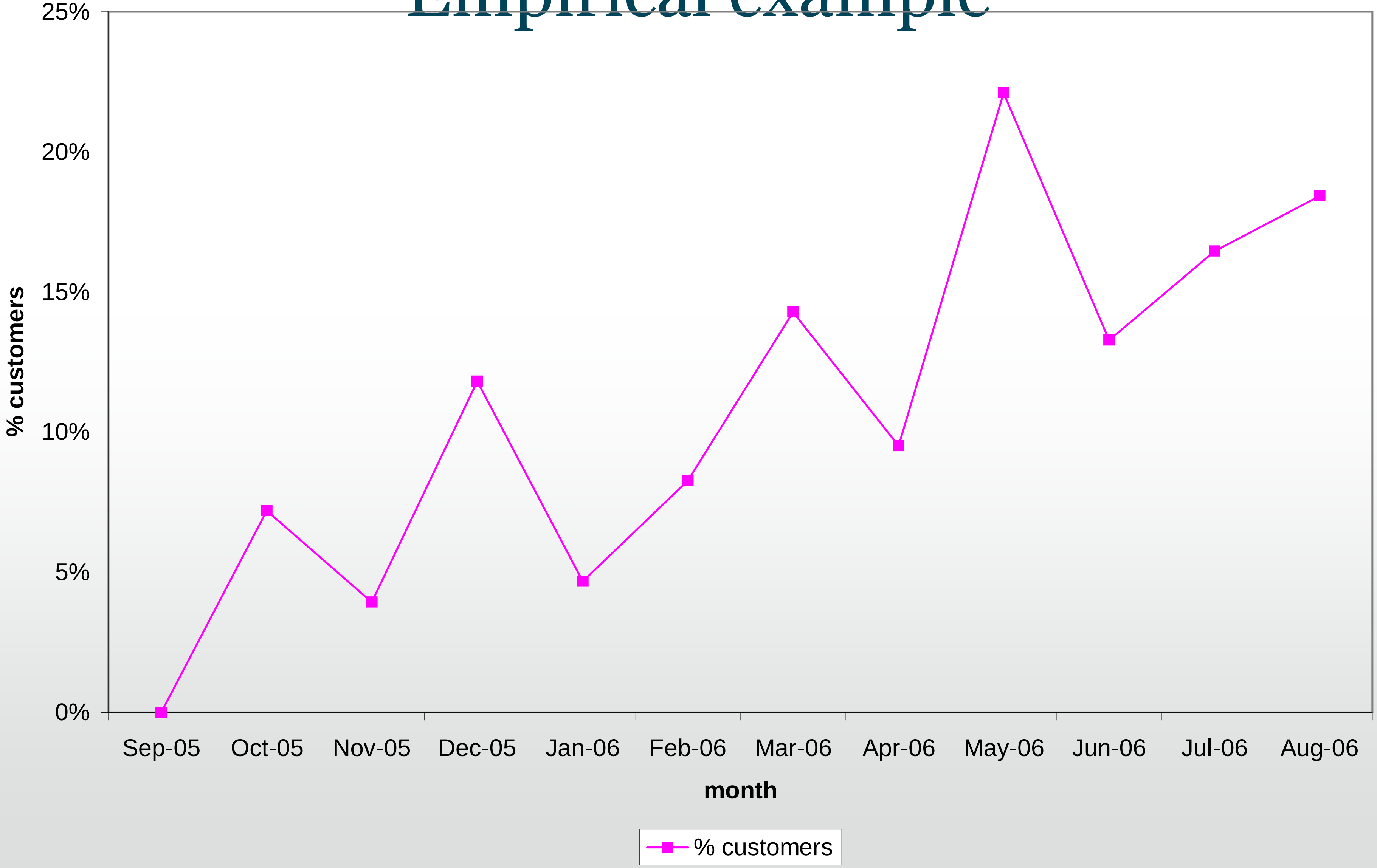
UNIVERSITY OF Southampton
School of Management

Empirical example



UNIVERSITY OF Southampton
School of Management

Empirical example



Empirical example

- Results
  - the Gini coefficient and KS statistic remain rather stable
  - the % of customers whose updated scores do not lie within the interval, increases: the credit propensity level is either under- or overestimated for the increasing % of customers
  - the scorecard maintains its ability to separate the willing from the unwilling as well as to rank customers according to their willingness to apply for new loans, but it becomes less and less up-to-date in terms of assigning the odds

Conclusions

- It is up to the decision makers, what is the maximum value of the analyzed percentage that can still be accepted (e.g. 10%, since the 90% confidence intervals are used). Once that value is exceeded, a new model should be developed
- As a result of using not only the current monitoring sample but – through the Kalman filter – all previous ones as well, possible local disturbance should have limited influence on the monitoring results and decisions made on their basis
- Further modifications of the methodology could especially distinguish between under- and overestimation of the odds

UNIVERSITY OF Southampton
School of Management

Thank you!