

Impact of segmentation on the performance measures of LGD models

Lyn Thomas and Katarzyna Bijak

Credit Scoring and Credit Control XIV
26th-28th August 2015, Edinburgh

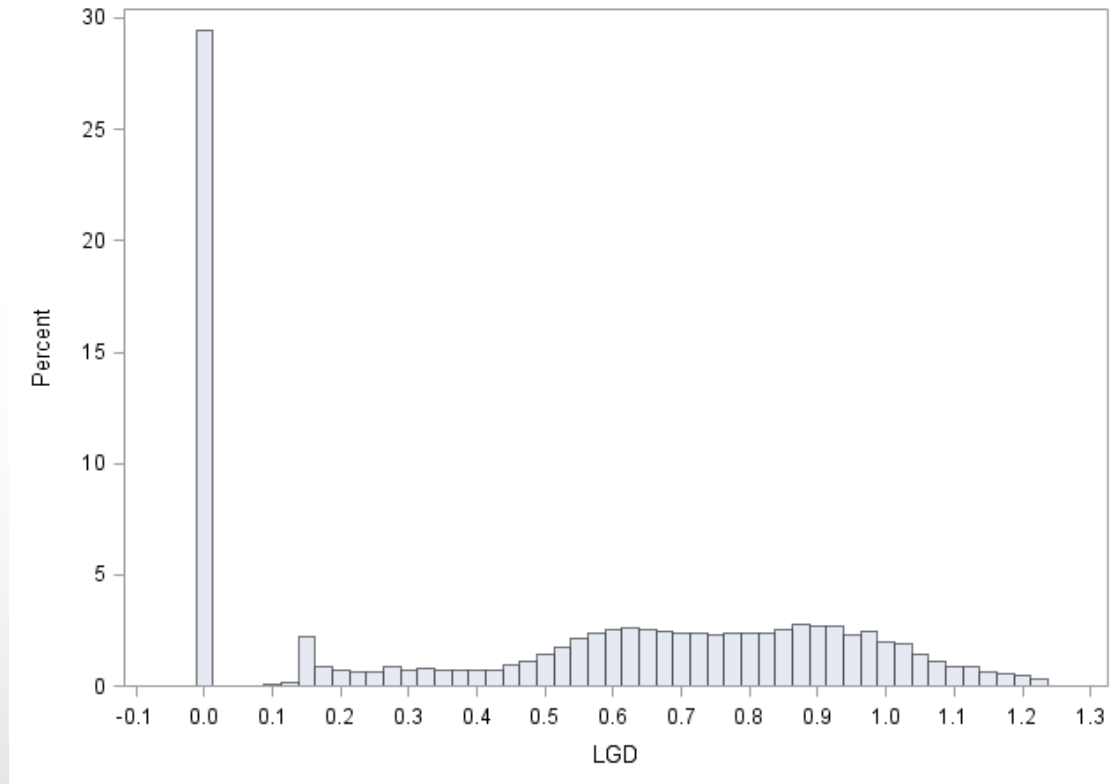
Outline

- Introduction
- LGD
- LGD models
- Segmentation
- Example
- Conclusions

LGD

- Loss Given Default (LGD)
 - The lender's loss on a loan due to the customer's default, i.e. failure to meet the credit commitment
- Recovery Rate (RR)
 - $LGD = 1 - RR$
- Basel II and III
 - Under the Advanced Internal Ratings-Based (AIRB) approach, lenders are allowed to use their own predictions of risk parameters, including LGD

LGD distribution example



LGD models

- Unsecured loans
 - One-stage models
 - Multi-stage approaches
 - Separation of 0s (+ Separation of 1s) + Prediction
- Mortgage loans
 - One-stage models
 - Two-stage approaches
 - Repossession model + Haircut model

LGD models

- Separation stage(s)
 - Logistic regression
 - Decision trees
- Prediction stage/one-stage models
 - Regression models
 - Tobit models
 - Survival analysis
 - Classification and Regression Trees (CART)
 - Other nonlinear models

Performance measures of LGD models

- Please come on Friday!
 - Parallel session 22B - Performance Assessment
 - 12:05 - 12:35

Performance of LGD models

- Poor performance
 - In particular for unsecured consumer credit
- For example, Pearson's correlation:
 - 0.38 - 0.51 (Yashkir and Yashkir, 2013)
 - 0.10 - 0.40 (Bellotti and Crook, 2008)
 - 0.08 - 0.27 (Bijak and Thomas, 2015)

Segmentation

- Poor performance should not be surprising
 - E.g. typical R-squared values when using regression for modelling outcomes over individual consumers are low
- Nevertheless, this is a bit embarrassing!
- ... but the Basel Accords only require LGD estimates at the portfolio segment level, where the lender has determined the appropriate segmentation of their portfolio

What happens at the segment level?

- Symbols
 - n – number of individual observations
 - K – number of segments
 - M – segment size ($n = KM$)
- Assumptions for calculations
 - Linear regression model
 - Independent and normally distributed residuals with zero mean and the same variance

What happens at the segment level?

- Example: MAE

$$\varepsilon_i \sim N(0, \sigma^2)$$

$$MAE_n = \frac{1}{n} \sum_{i=1}^n |\varepsilon_i| \approx \sigma \sqrt{\frac{2}{\pi}}$$

What happens at the segment level?

- Example: MAE

$$\sum \varepsilon_i \sim N(0, M\sigma^2)$$

$$\begin{aligned} MAE_K &= \frac{1}{K} \sum_{k=1}^K |\varepsilon_k| = \frac{1}{KM} \sum_{k=1}^K \left| \sum_{i=(k-1)M+1}^{kM} \varepsilon_i \right| \approx \frac{K\sigma}{KM} \sqrt{\frac{2M}{\pi}} \\ &= \sqrt{\frac{K}{n}} MAE_n \end{aligned}$$

What happens at the segment level?

- MSE

$$MSE_K \approx \frac{K}{n} MSE_n$$

- RMSE

$$RMSE_K \approx \sqrt{\frac{K}{n}} RMSE_n$$

What happens at the segment level?

- MAE

$$MAE_K \approx \sqrt{\frac{K}{n}} MAE_n$$

- RSE

$$RSE_K \approx 1 - \frac{n}{n - K + \frac{K}{1 - RSE_n}}$$

What happens at the segment level?

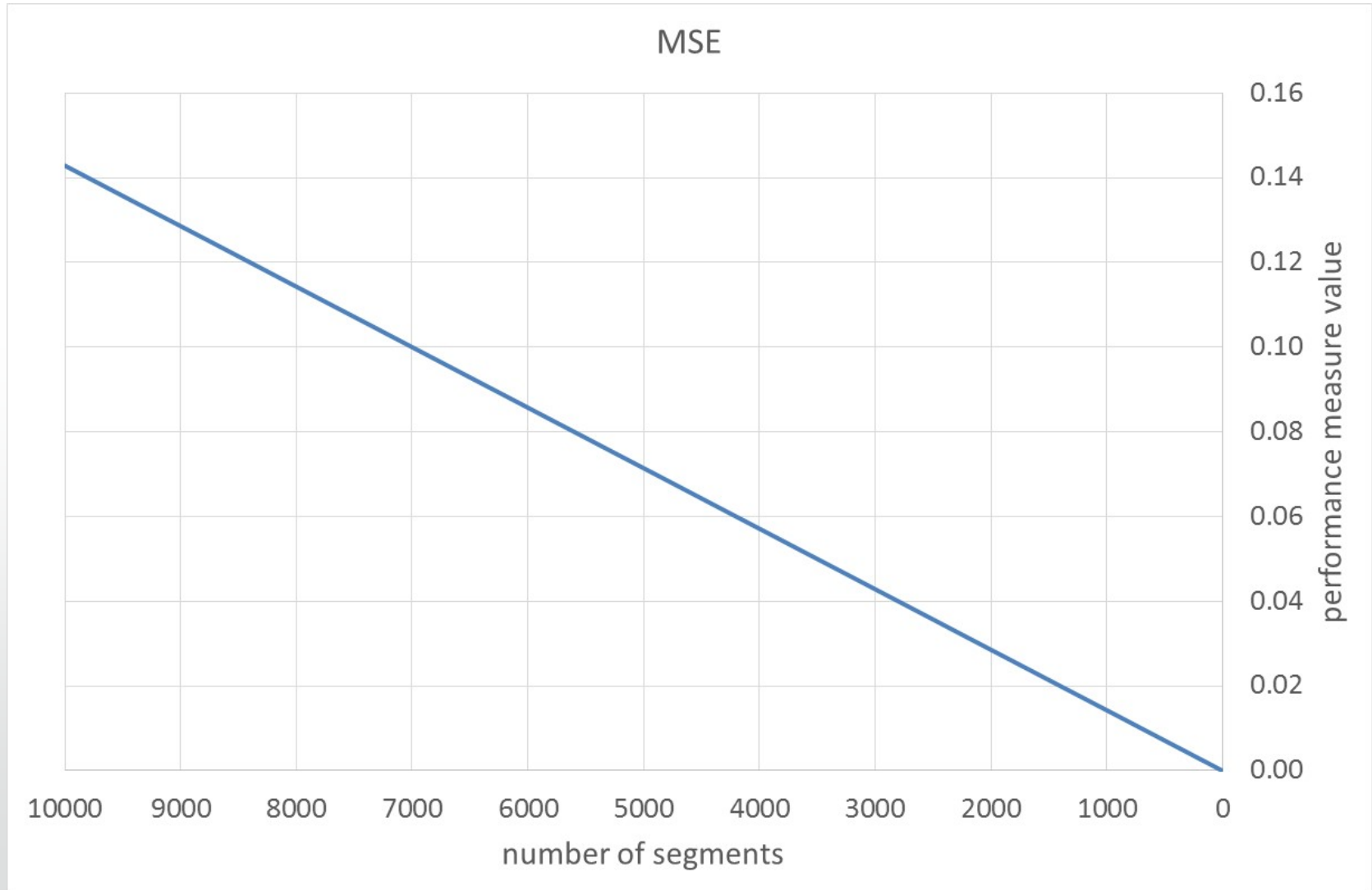
- R-squared

$$R_K^2 \approx \frac{n}{n - K + \frac{K}{R_n^2}}$$

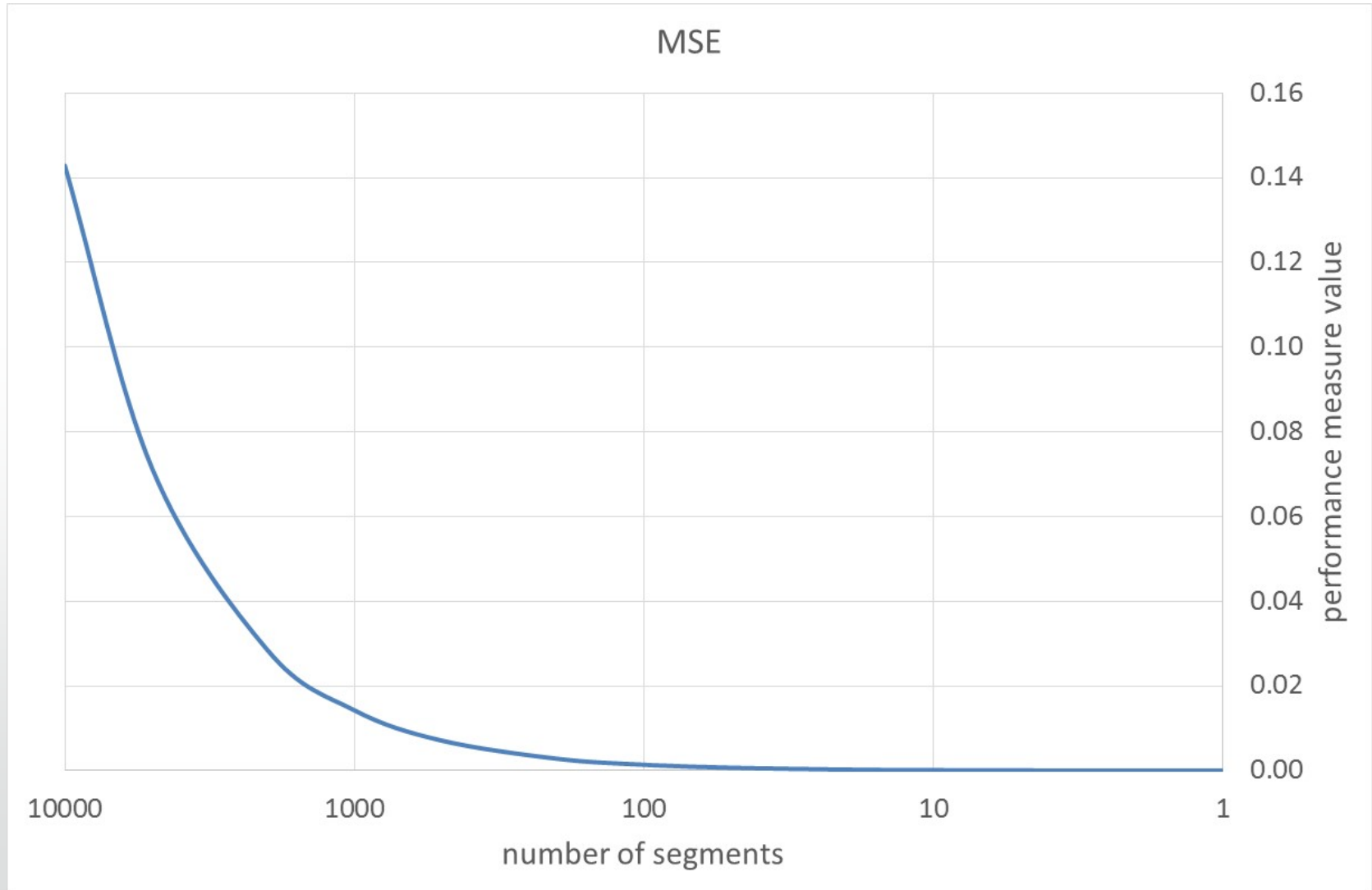
- Pearson's correlation

$$r_K \approx \sqrt{\frac{n}{n - K + \frac{K}{r_n^2}}}$$

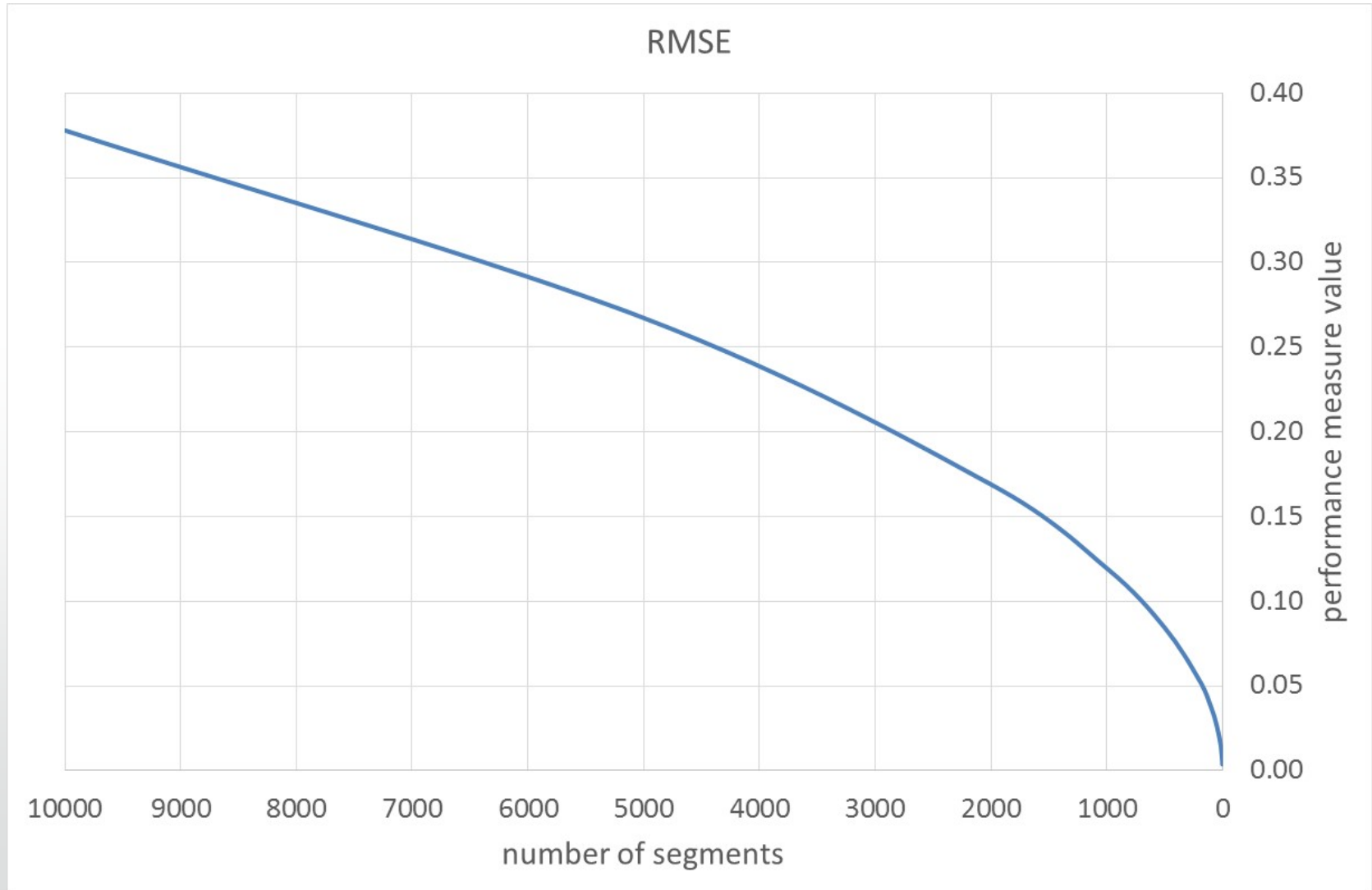
Impact of segmentation



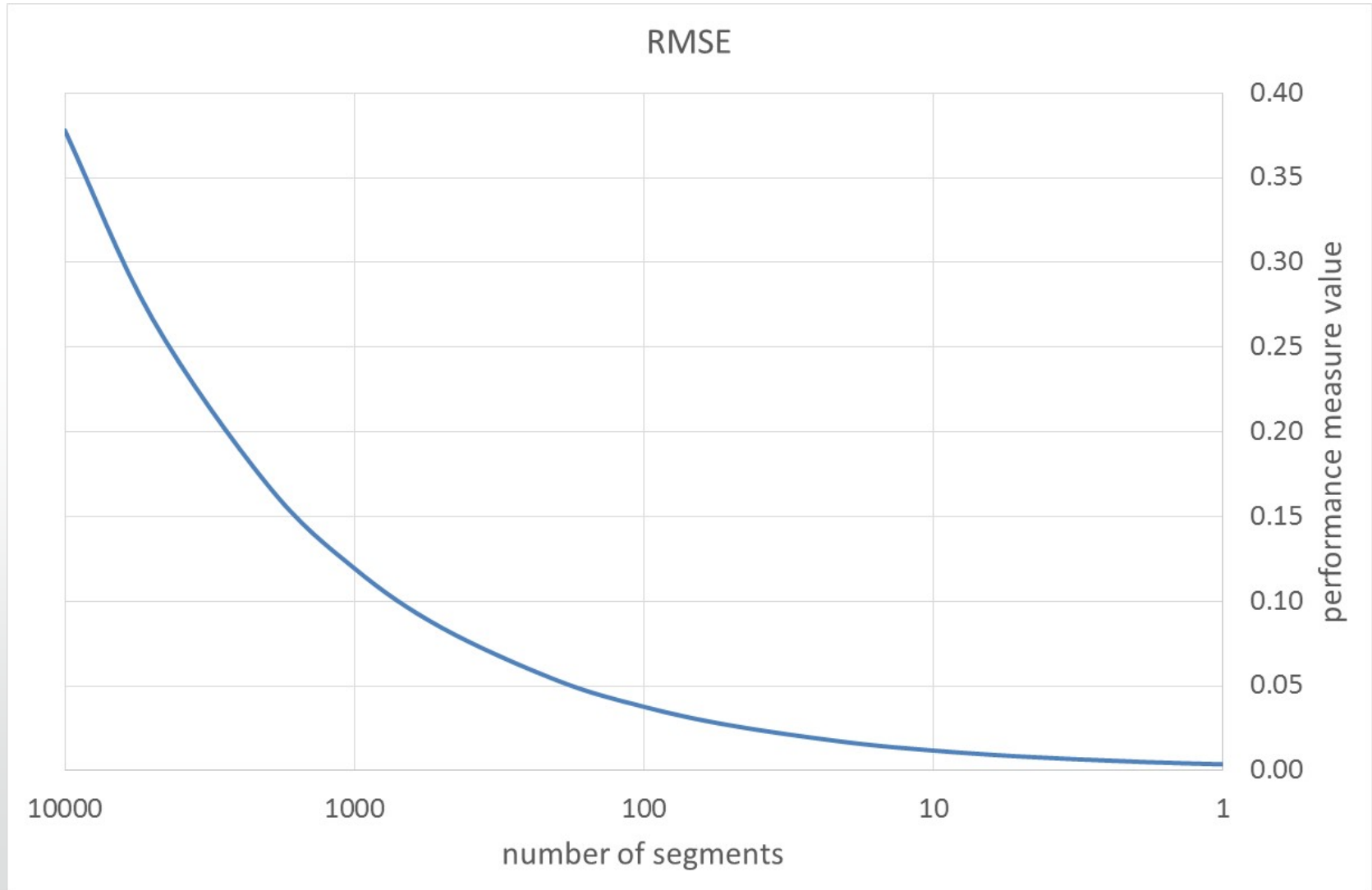
Impact of segmentation (log. scale)



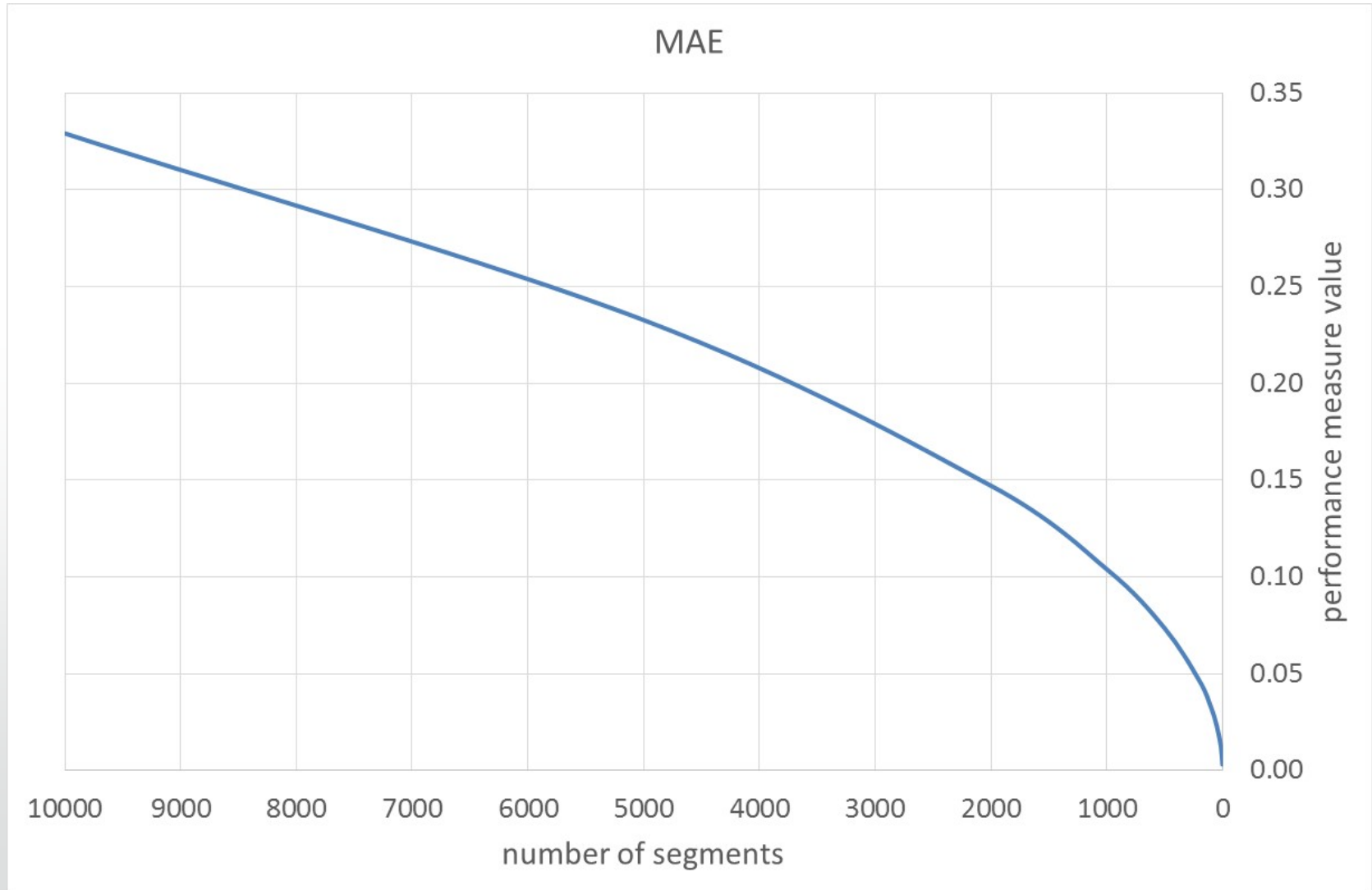
Impact of segmentation



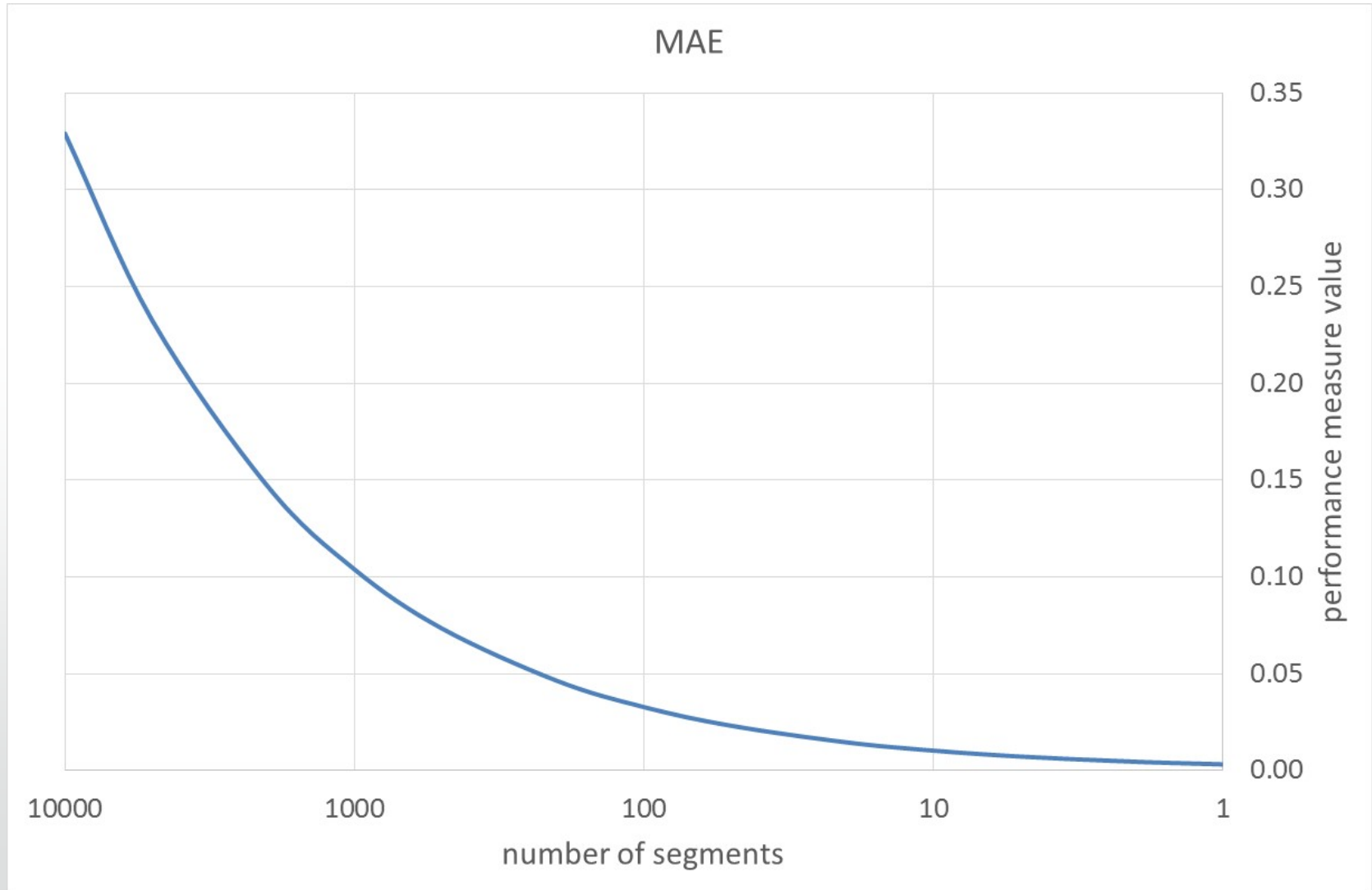
Impact of segmentation (log. scale)



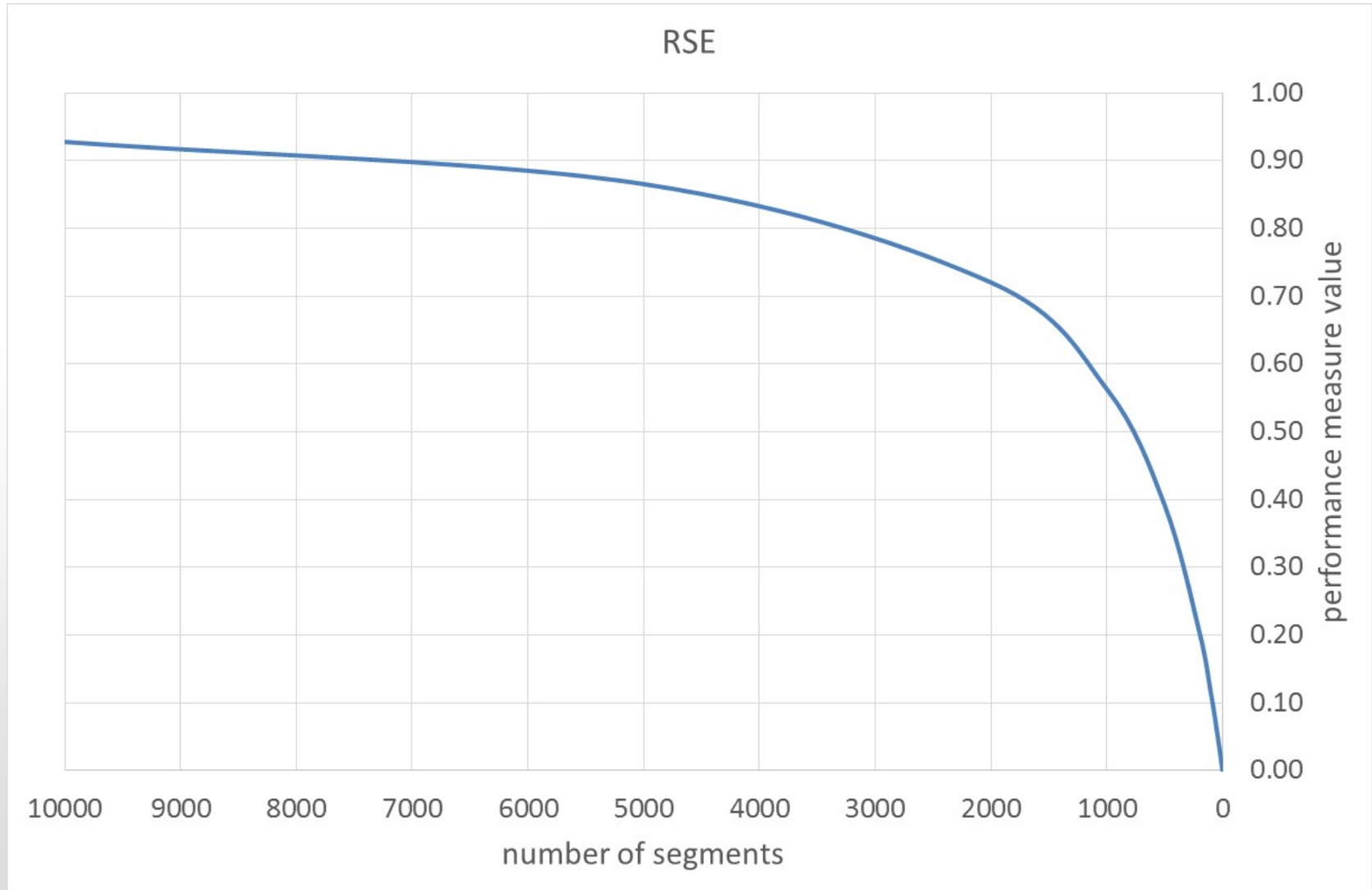
Impact of segmentation



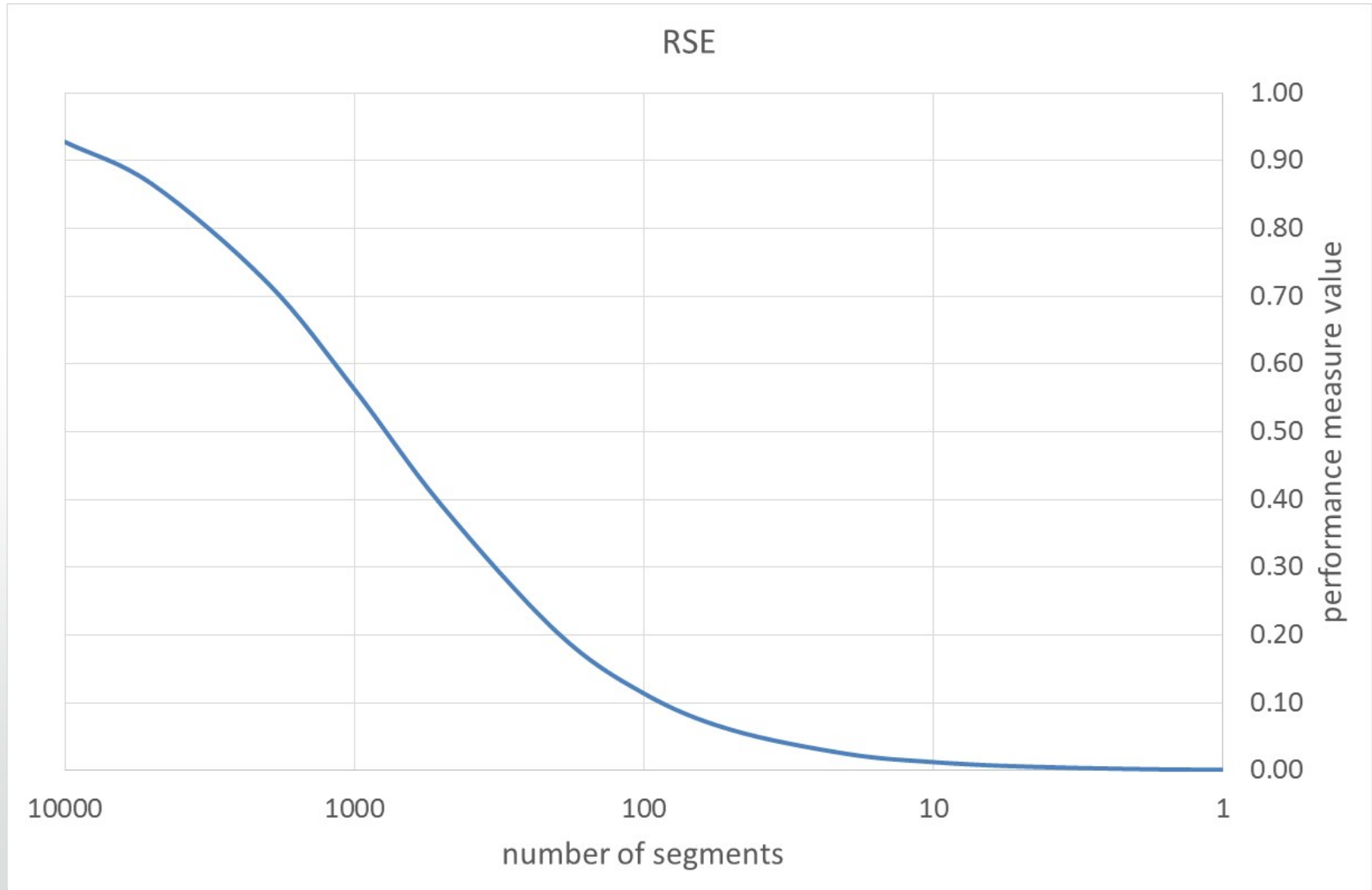
Impact of segmentation (log. scale)



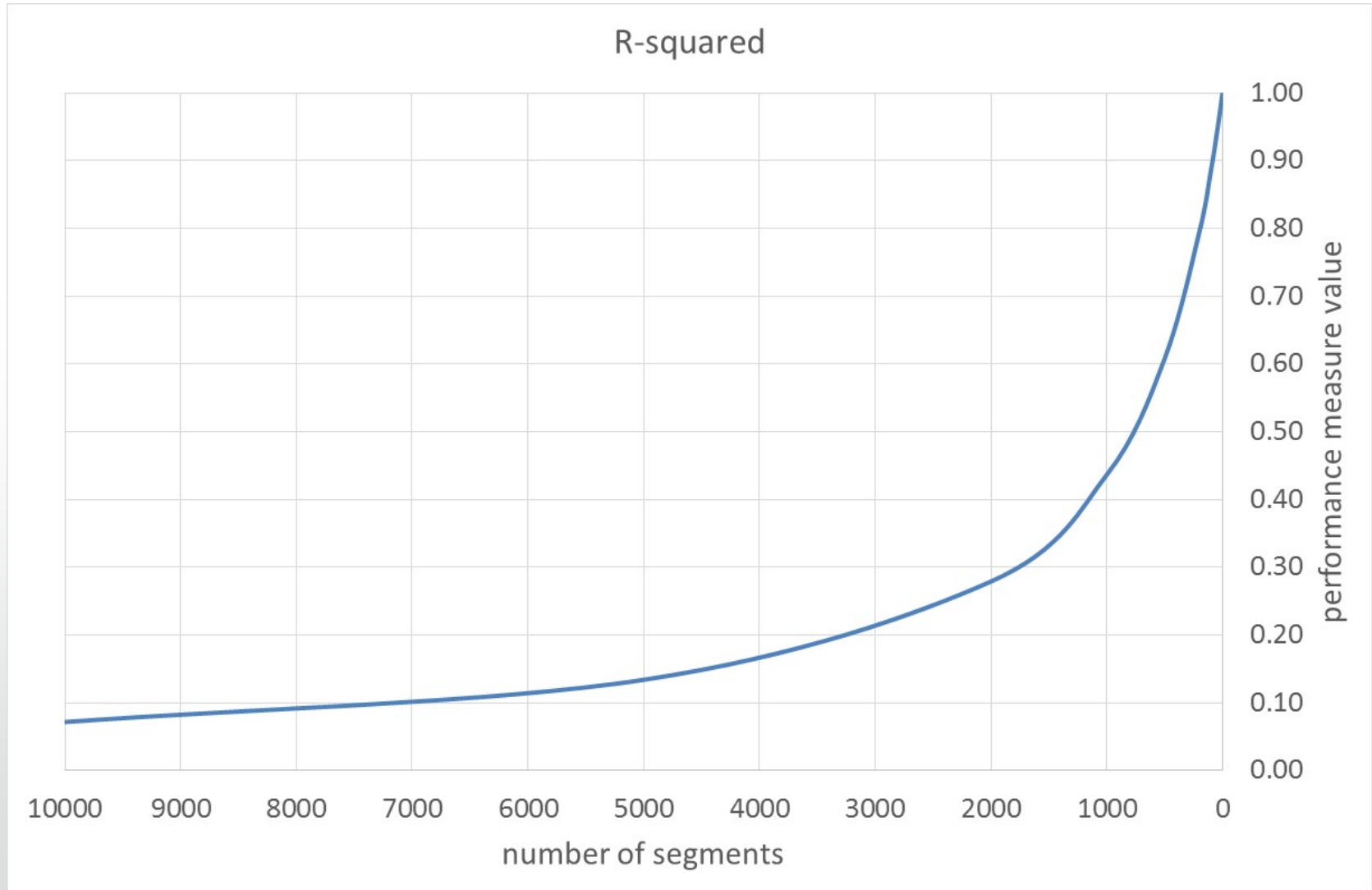
Impact of segmentation



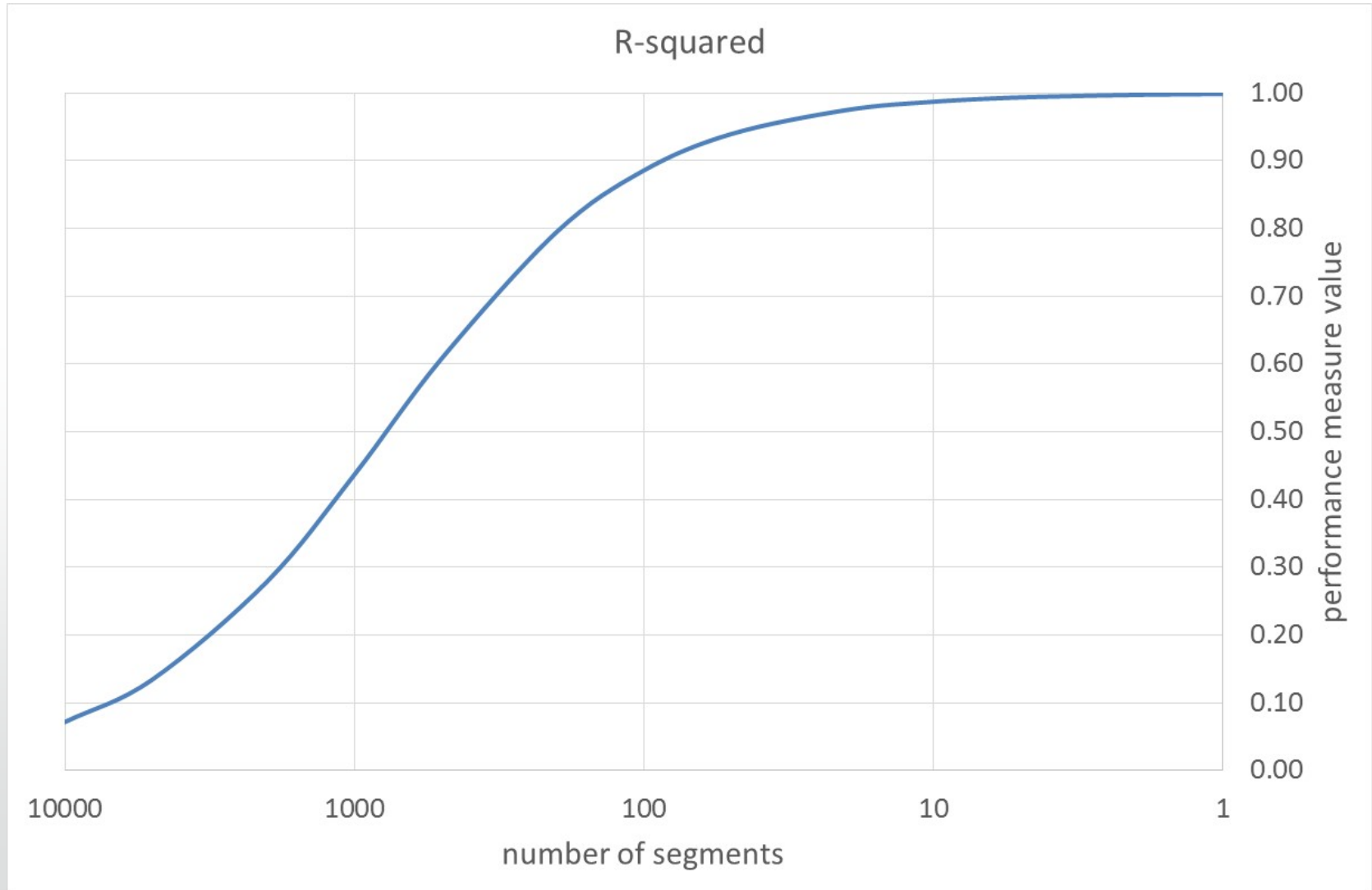
Impact of segmentation (log. scale)



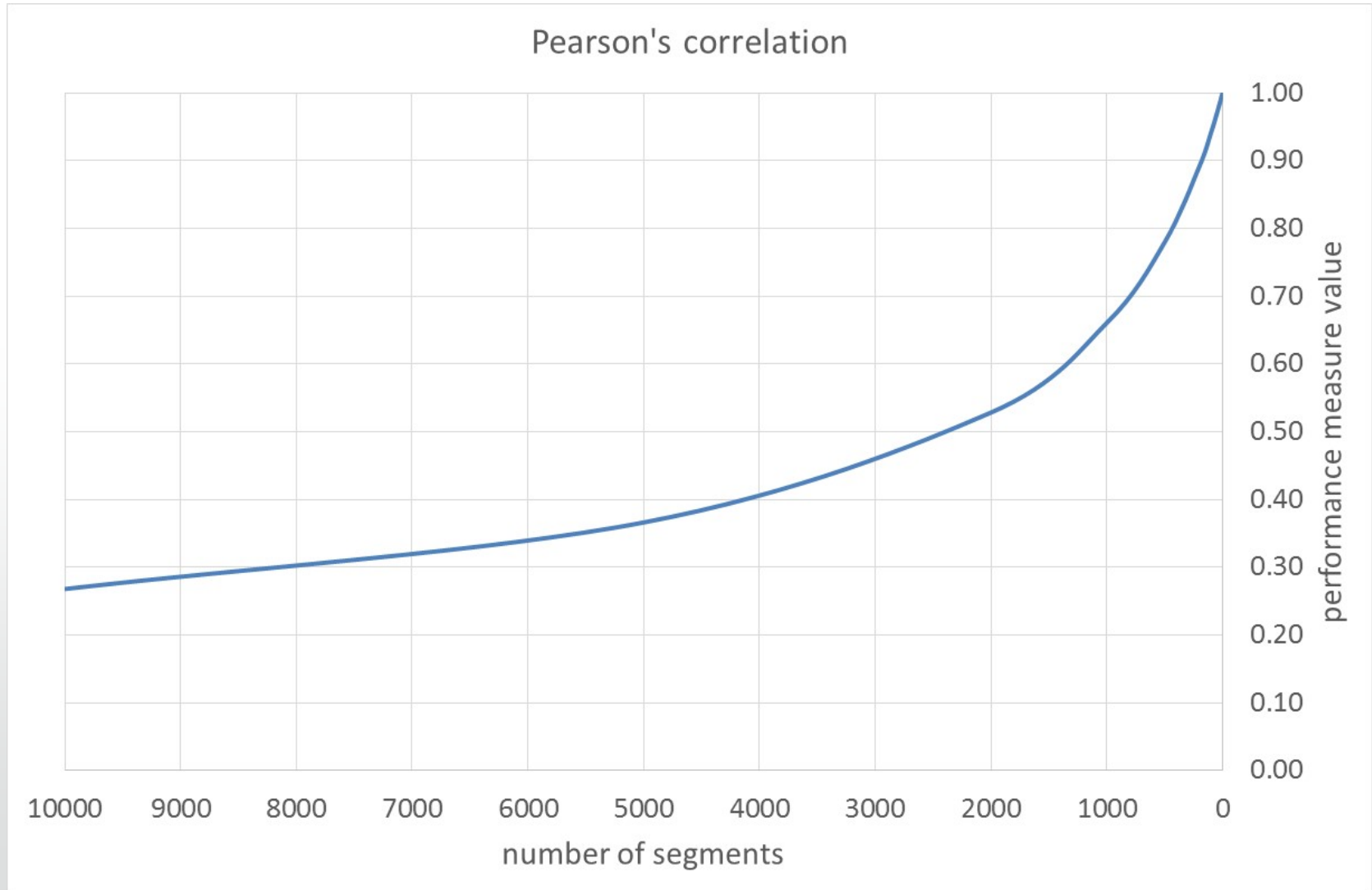
Impact of segmentation



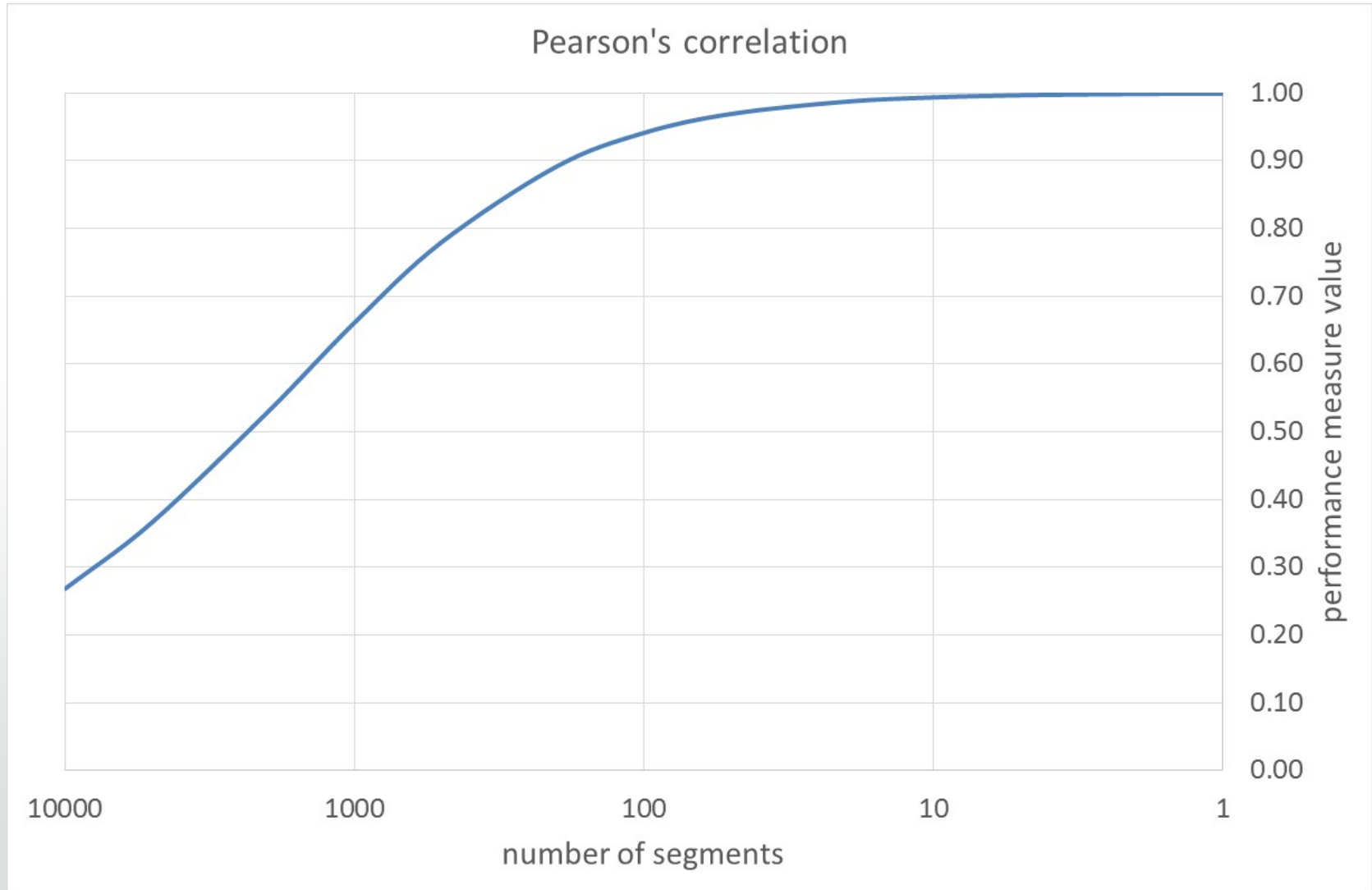
Impact of segmentation (log. scale)



Impact of segmentation



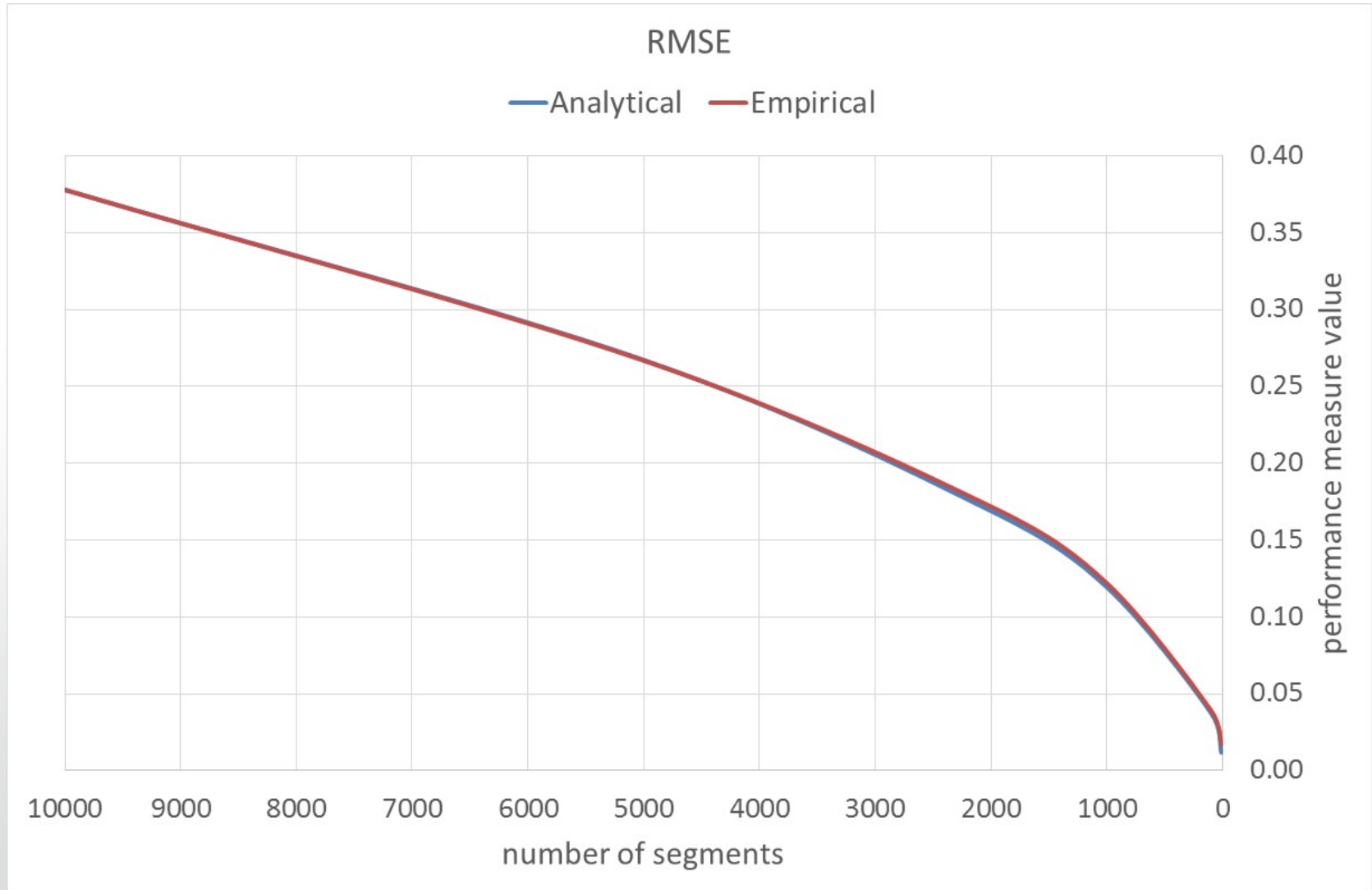
Impact of segmentation (log. scale)



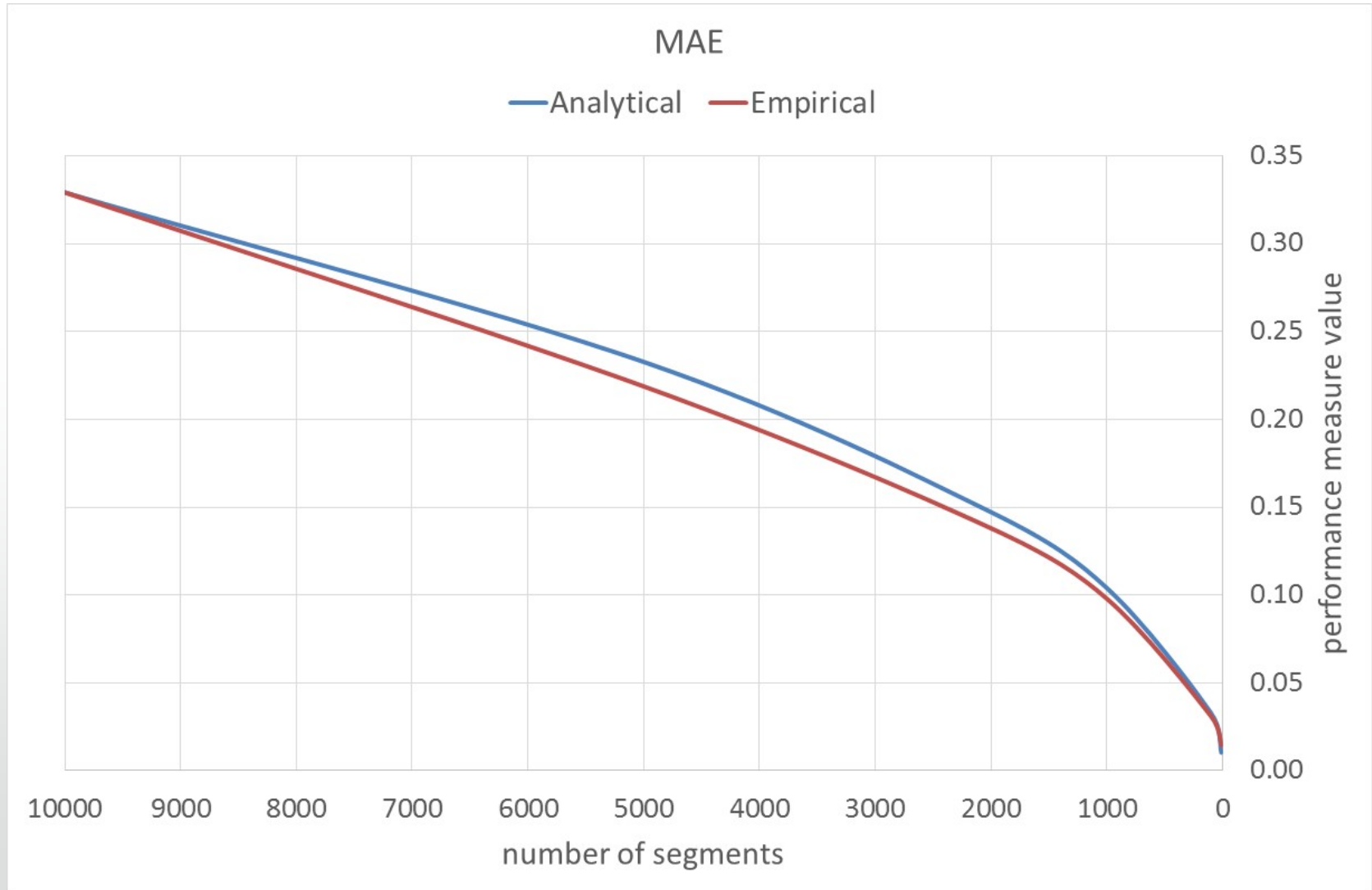
Analytical vs empirical results



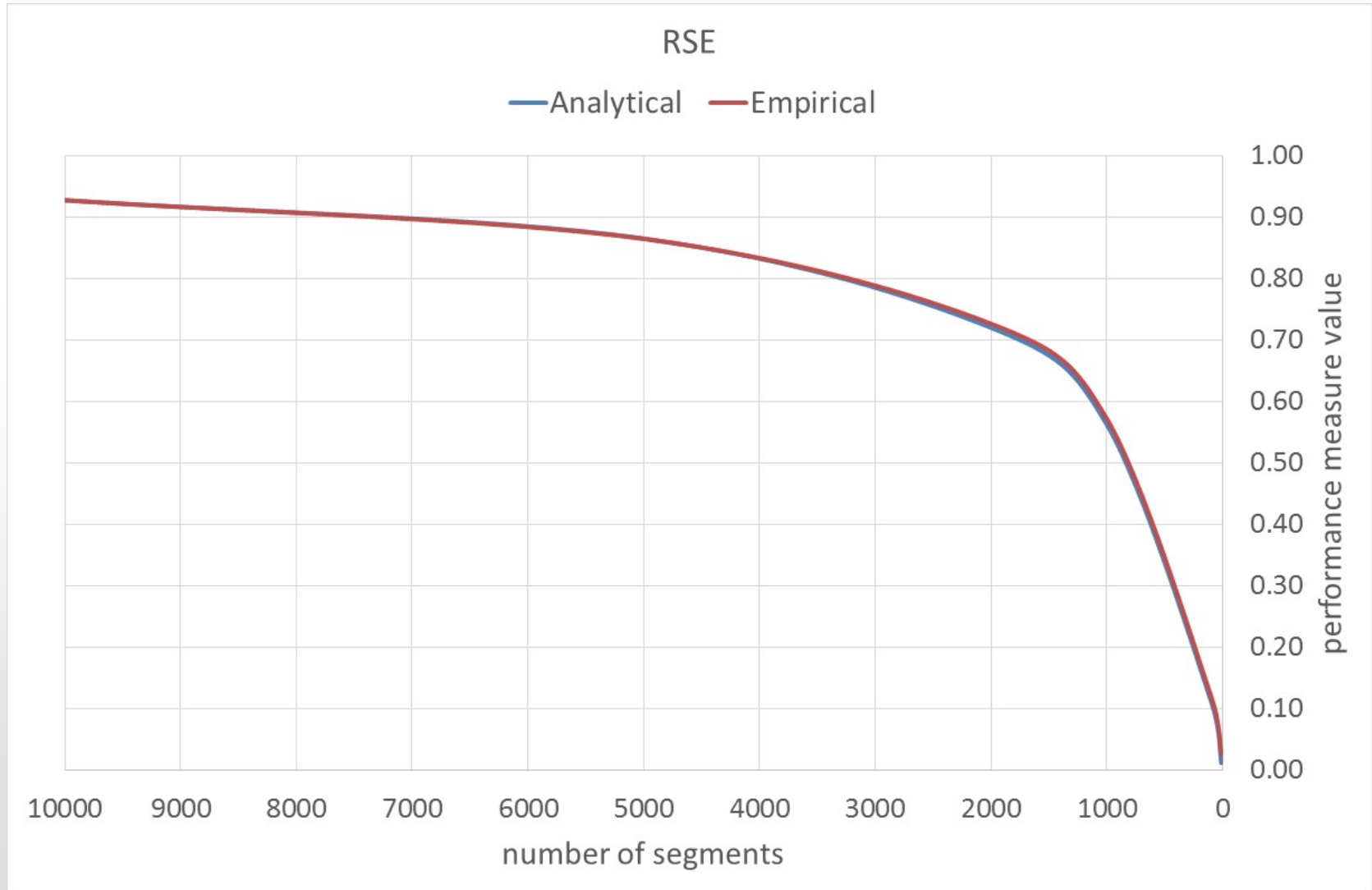
Analytical vs empirical results



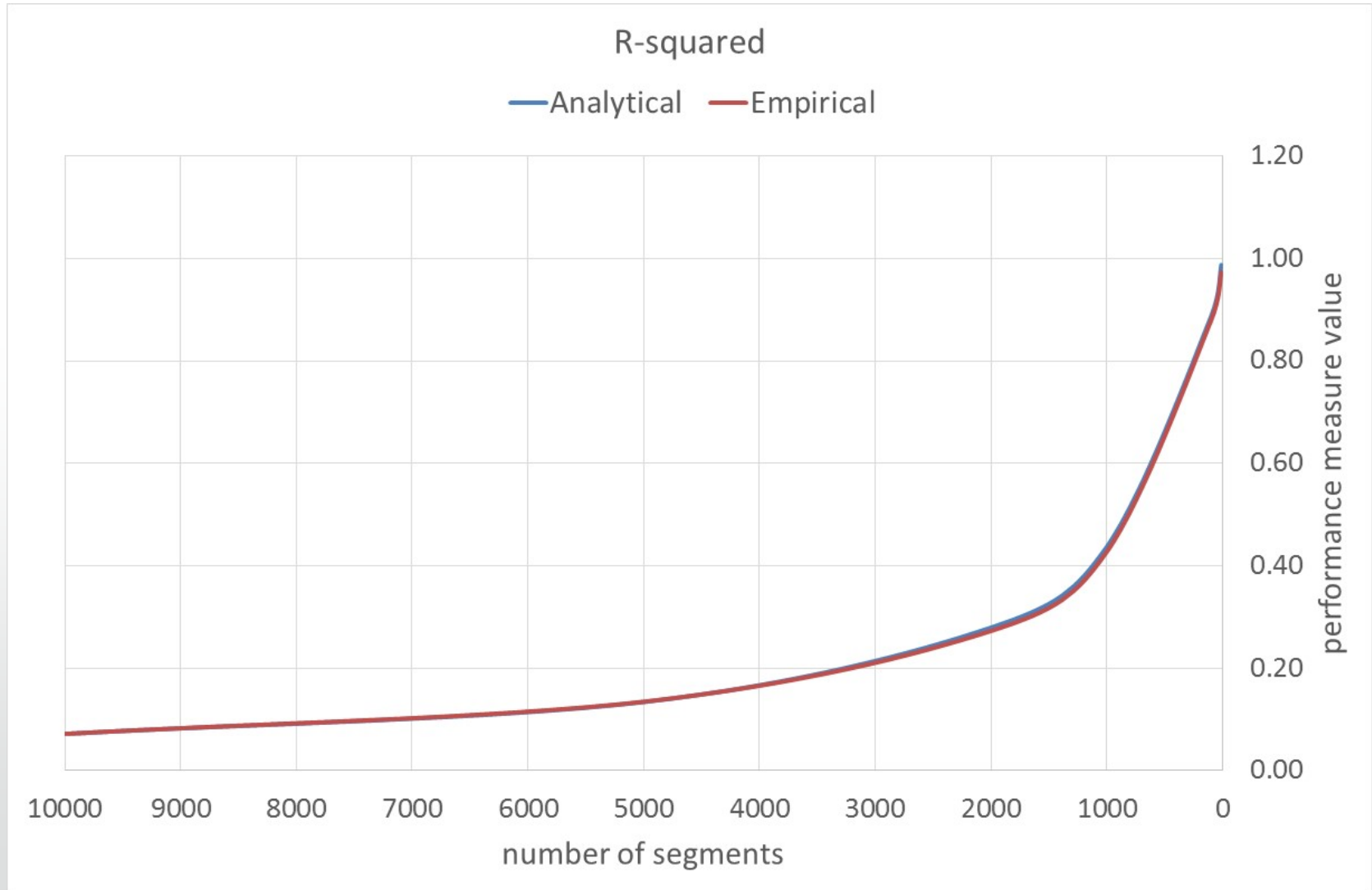
Analytical vs empirical results



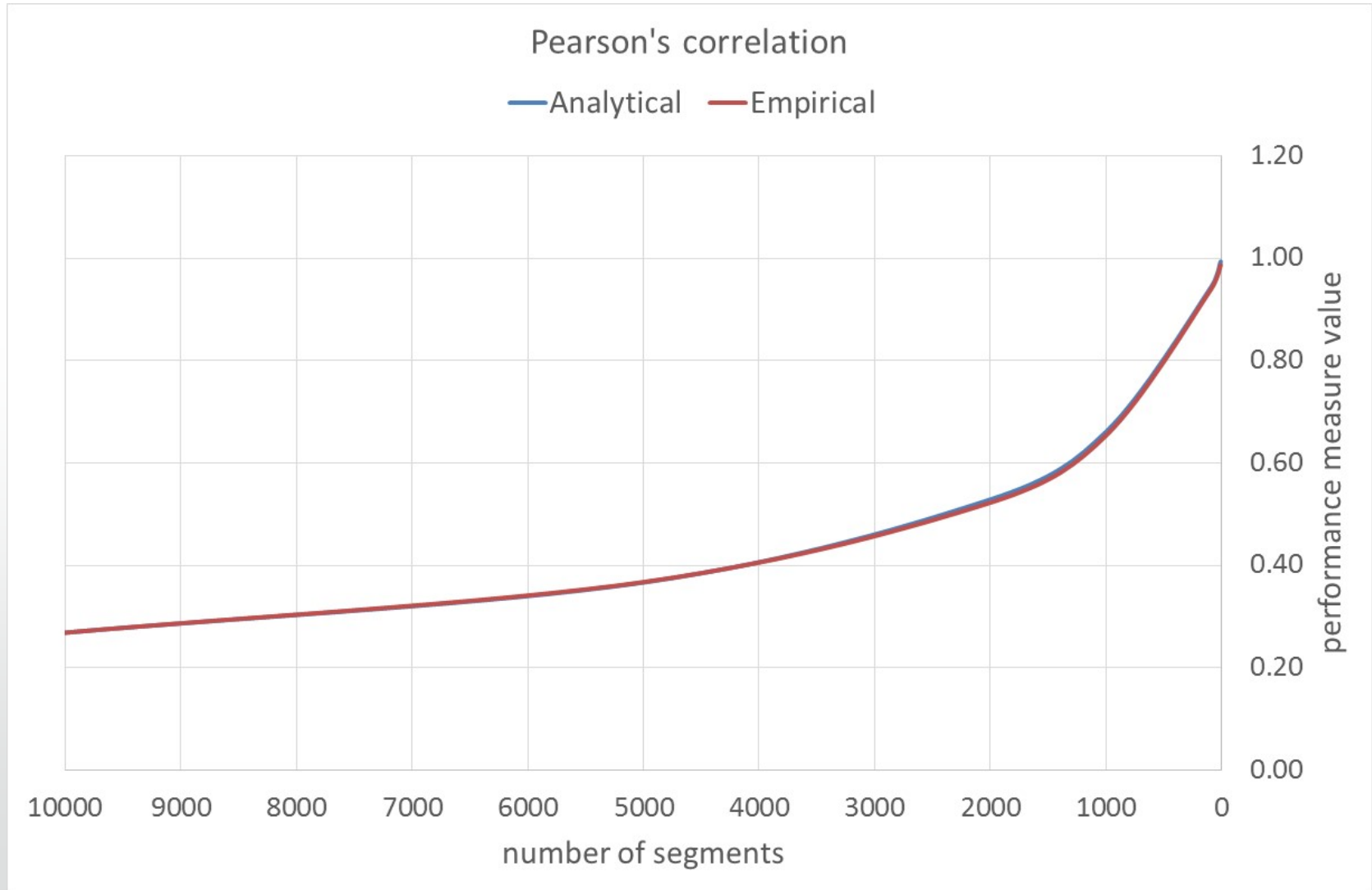
Analytical vs empirical results



Analytical vs empirical results



Analytical vs empirical results



Conclusions

- Poor model performance at the loan level
- For a number of measures the segmentation leads to a significant improvement in their values
- Is this true for all performance measure?
- Do LGD models perform better than we thought?

Thank you!

Empirical results

