Are industrial companies in the Czech Republic able to predict the short-term future of the economy?

—

AMSE 2018

Veronika Ptáčková[1] Lubomír Štěpánek[(1), 2, 3] Vít Hanzal[(1)]

[1]Department of Economic Statistics
Faculty of Informatics and Statistics
University of Economics, Prague

[2]Institute of Biophysics and Informatics
First Faculty of Medicine
Charles University in Prague

[3]Department of Biomedical Informatics
Faculty of Biomedical Engineering
Czech Technical University in Prague

Content

Quick introduction

- short-term statistic and Business Tendency Survey
  - key indicators to assess and monitor the development of the economy
  - results are provided to government institutions, CB, financial institutions
- main question – how accurately can companies predict their future?

Data sources – short-term statistics in the industry

- precise *quantitative data*
- employment – the average number of employees of a company registered for a month
- sales – economic turnover of a company recorded for a month

Data sources – Business Tendency Survey

- employment
  - "The number of employees will ..." (in the next three months)
- sales
  - "The number of sales will ..." (in the next three months)
- respondents' answers, marked as $o_{t,i,c}$
  - decrease
  - stagnation (remain at the same level)
  - increase

  for given indicator $i \in \{\text{employment}, \text{sales}\}$, for given company c linked to months $t, t+1, t+2$
- thus $o_{t,i,c} \in \{\text{decrease}, \text{stagnation}, \text{increase}\}$
- *qualitative data*

Data characteristics

- period of 2003—2016
- monthly basis
- the first 400 companies ordered by length of consecutive responding chosen

Data transformation

- transformation of absolute values of the indicators to relative ones using the following metrics
  - *mean-to-first*, $MTF_{t,i,c}$
  - *last-to-first*, $LTF_{t,i,c}$
- so

$$MTF_{t,i,c} = \frac{\frac{1}{3}(x_{t,i,c} + x_{t+1,i,c} + x_{t+2,i,c})}{x_{t,i,c}}$$

$$LTF_{t,i,c} = \frac{x_{t+2,i,c}}{x_{t,i,c}}$$

where $x_{t,i,c}$ is an absolute value of indicator i of company c in month t

Data labeling

- categorization (labeling) of the computed relative values into three levels
- for $MTF_{t,i,c}$ is

$$MTF_{t,i,c}(k) = \begin{cases} \text{decrease}, & MTF_{t,i,c} \in \langle 0, \frac{1}{k}) \\ \text{stagnation}, & MTF_{t,i,c} \in \langle \frac{1}{k}, k) \\ \text{increase}, & MTF_{t,i,c} \in \langle k, +\infty) \end{cases}$$

  for given $k \geq 1$, given indicator $i \in \{\text{employment}, \text{sales}\}$, for given company c linked to months $t, t+1, t+2$
- analogously for $LTF_{t,i,c}$

Cartesian product of labeled metrics and company opinions

- for given $k \geq 1$, given indicator $i \in \{\text{employment}, \text{sales}\}$, given company c linked to months $t, t+1, t+2$ is

$$\begin{aligned} o_{t,i,c} &\in \{\text{decrease}, \text{stagnation}, \text{increase}\} \\ MTF_{t,i,c}(k) &\in \{\text{decrease}, \text{stagnation}, \text{increase}\} \\ LTF_{t,i,c}(k) &\in \{\text{decrease}, \text{stagnation}, \text{increase}\} \end{aligned}$$

- therefore we got tuples $[o_{t,i,c}, MTF_{t,i,c}(k)]$, $[o_{t,i,c}, LTF_{t,i,c}(k)]$ and

$$\begin{aligned} o_{t,i,c} \times MTF_{t,i,c}(k) &= \{\text{decrease}, \text{stagnation}, \text{increase}\} \times \\ &\times \{\text{decrease}, \text{stagnation}, \text{increase}\} \\ o_{t,i,c} \times LTF_{t,i,c}(k) &= \{\text{decrease}, \text{stagnation}, \text{increase}\} \times \\ &\times \{\text{decrease}, \text{stagnation}, \text{increase}\} \end{aligned}$$

Confusion matrix

- if we fix indicator i, then we can sum up all products $o_{t,i,c} \times MTF_{t,i,c}(k)$ for all times t and companies c and finally get *confusion matrix*

$$\boldsymbol{C}_i(k) = \sum_t \sum_c o_{t,i,c} \times MTF_{t,i,c}(k) = \{n_{jl}\}_{j,l}$$

where $n_{jl} \in \{0, 1, 2, \ldots\}$ for each

$$j, l \in \{\text{decrease}, \text{stagnation}, \text{increase}\}$$

is a count of all companies c during the time t which predicted j-th value (of the set $\{\text{decrease}, \text{stagnation}, \text{increase}\}$) of indicator i

- analogously for $o_{t,i,c} \times LTF_{t,i,c}(k)$

Confusion matrix & accuracy

- confusion matrix $\boldsymbol{C}_i(k)$ for indicator $i \in \{\text{employment}, \text{sales}\}$ and metrics $MTF_{t,i,c}(k)$ or $LTF_{t,i,c}(k)$ follows the form

		predicted values		
		decrease	stagnation	increase
true values	decrease	$n_{1,1}$	$n_{1,2}$	$n_{1,3}$
	stagnation	$n_{2,1}$	$n_{2,2}$	$n_{2,3}$
	increase	$n_{3,1}$	$n_{3,2}$	$n_{3,3}$

- for a given k is

$$\text{accuracy}(k) = \frac{\operatorname{tr} \boldsymbol{C}_i(k)}{\sum \boldsymbol{C}_i(k)} = \frac{\sum_{j=1}^{3} n_{jj}}{\sum_{j=1}^{3} \sum_{l=1}^{3} n_{jl}}$$

Finding lowest $k \geq 1$ maximizing accuracy

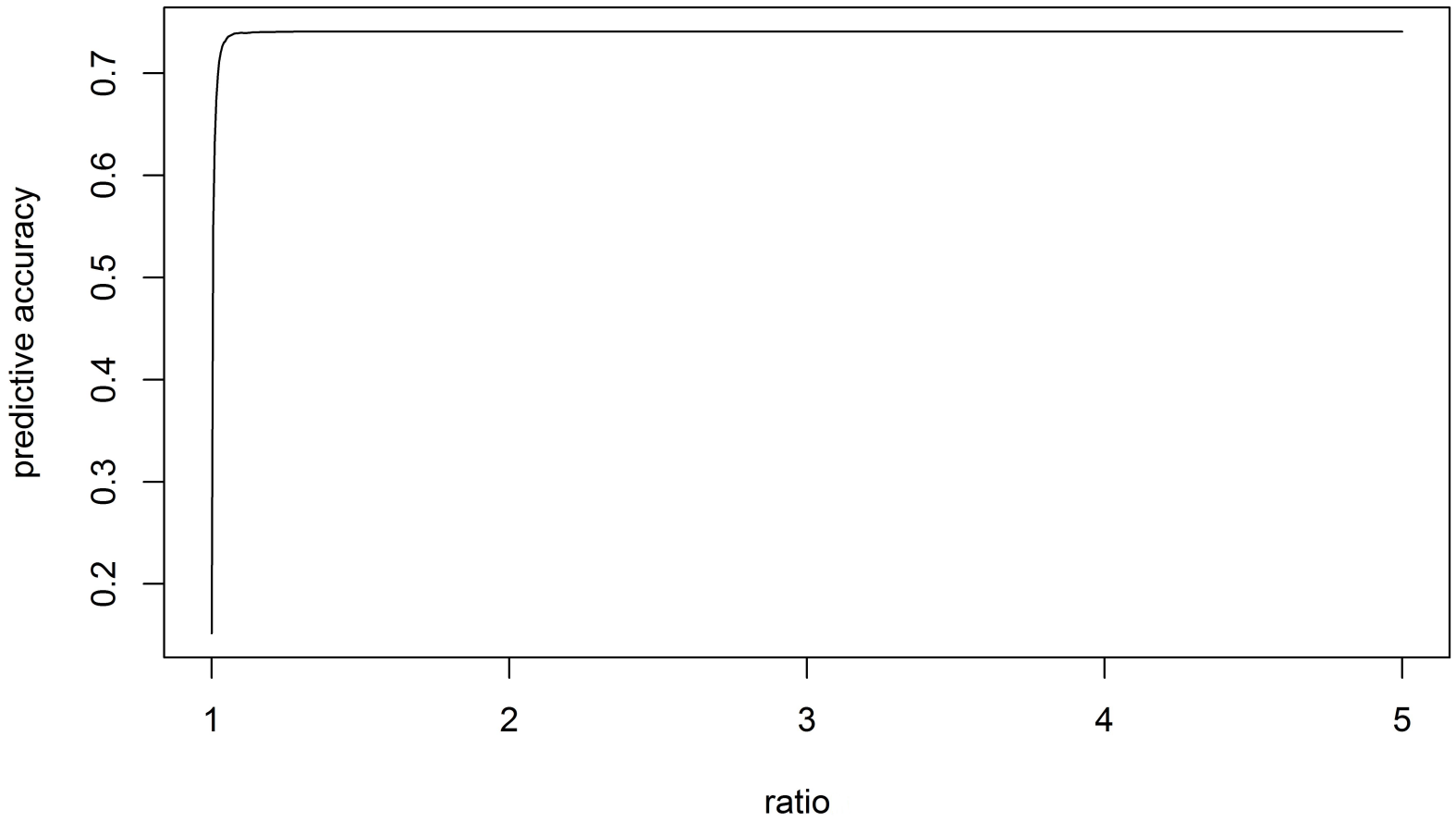
- we can reformulate the task as finding the lowest $k \geq 1$ "fuzzy" maximizing $\text{accuracy}(k) = \frac{\text{tr}\, \boldsymbol{C}_i(k)}{\sum \boldsymbol{C}_i(k)}$ for a given indicator $i \in \{\text{employment}, \text{sales}\}$ and metrics $MTF_{t,i,c}(k)$ or $LTF_{t,i,c}(k)$, $k \in \mathbb{R}$
- formally

$$\arg\min_k \left\{ \text{fuzzy max} \frac{\text{tr}\, \boldsymbol{C}_i(k)}{\sum \boldsymbol{C}_i(k)} \right\} \qquad \text{s. t.}$$

$$k \geq 1 \qquad (\dagger)$$

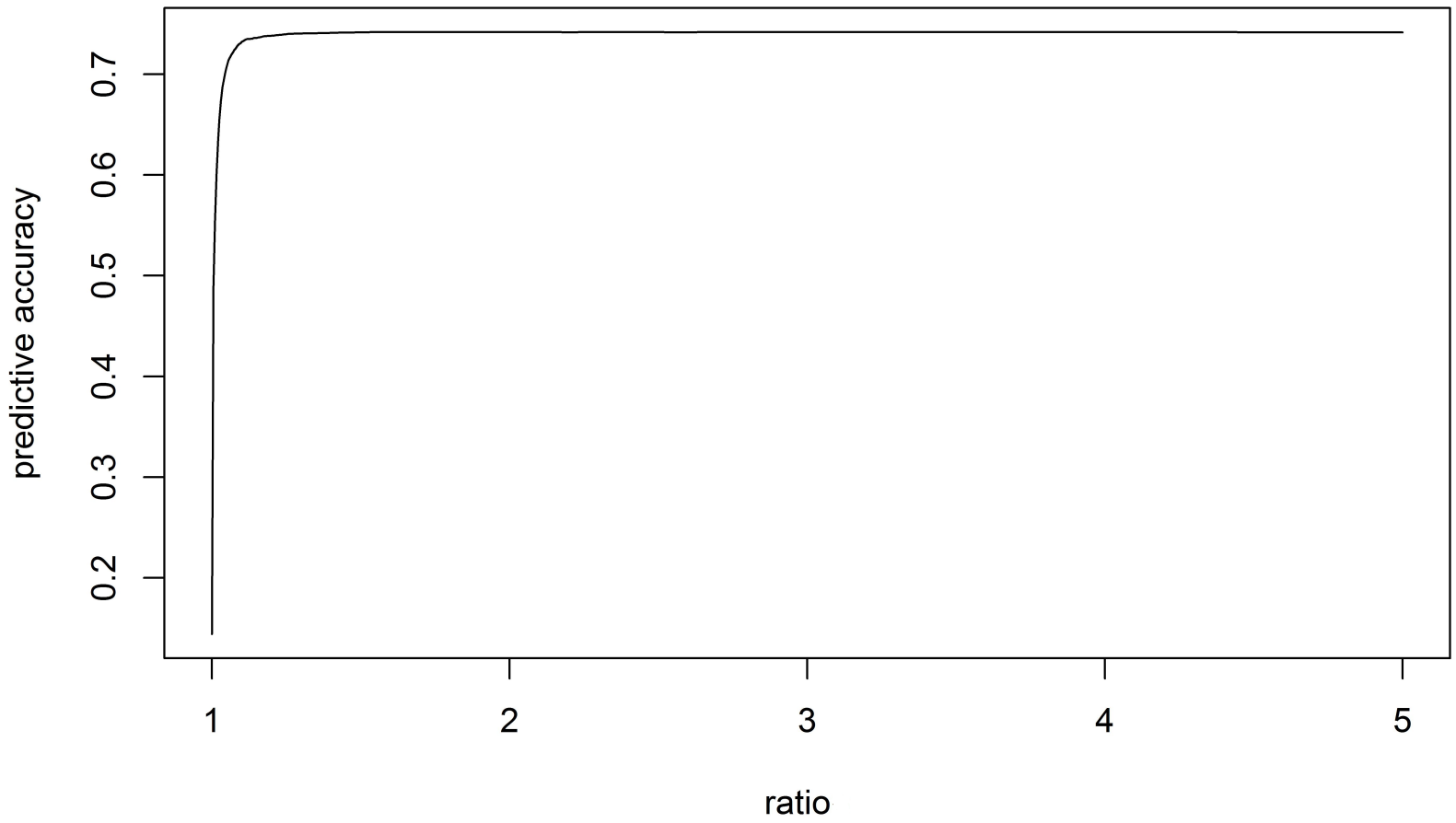
Results for employment and $MTF_{t,i,c}(k)$

- $k \approx 1.1$ (approx.), $\text{accuracy}(k) \approx 0.741$ (approx.)



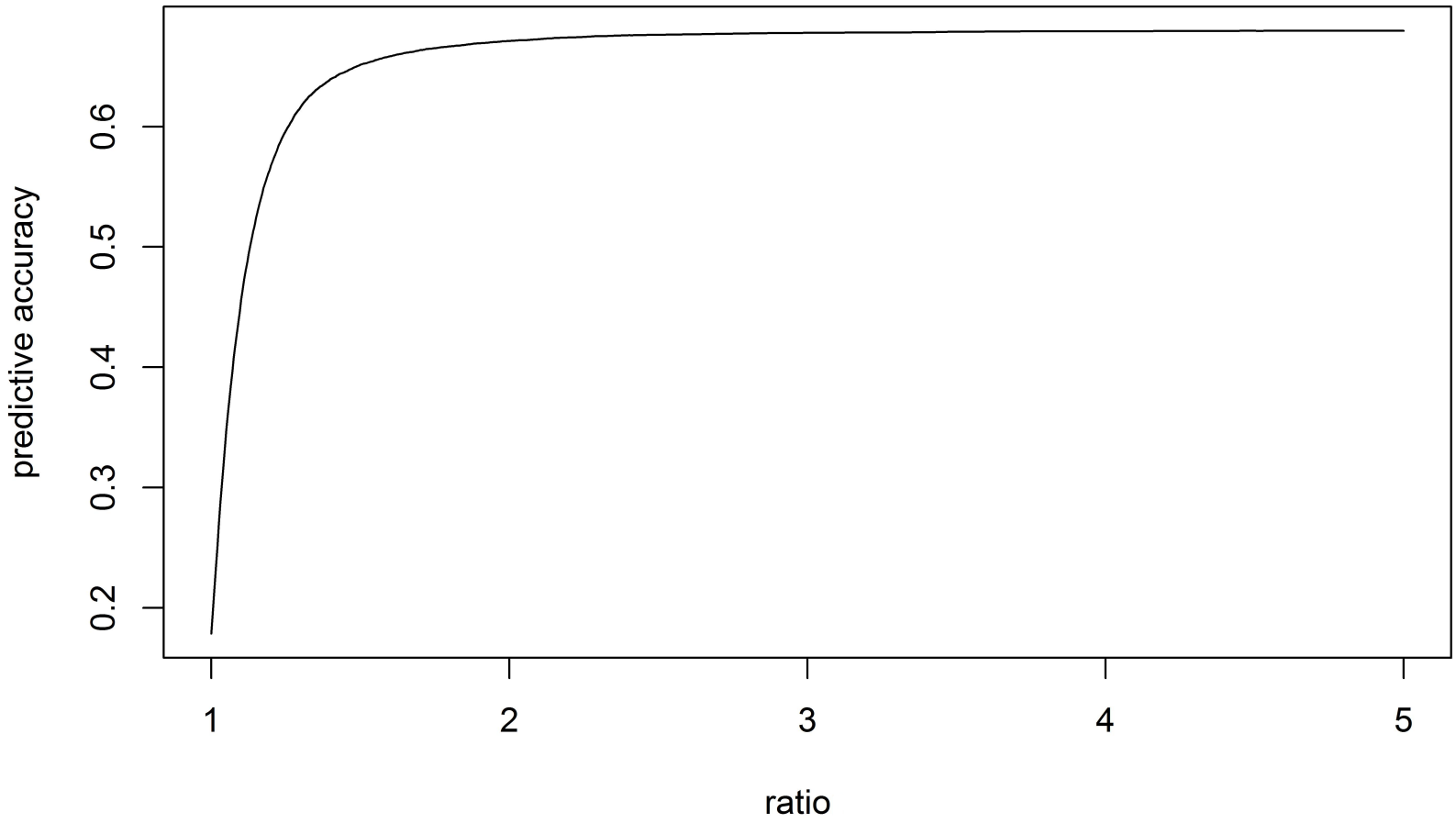
Results for employment and $LTF_{t,i,c}(k)$

- $k \approx 1.2$ (approx.), $\text{accuracy}(k) \approx 0.742$ (approx.)



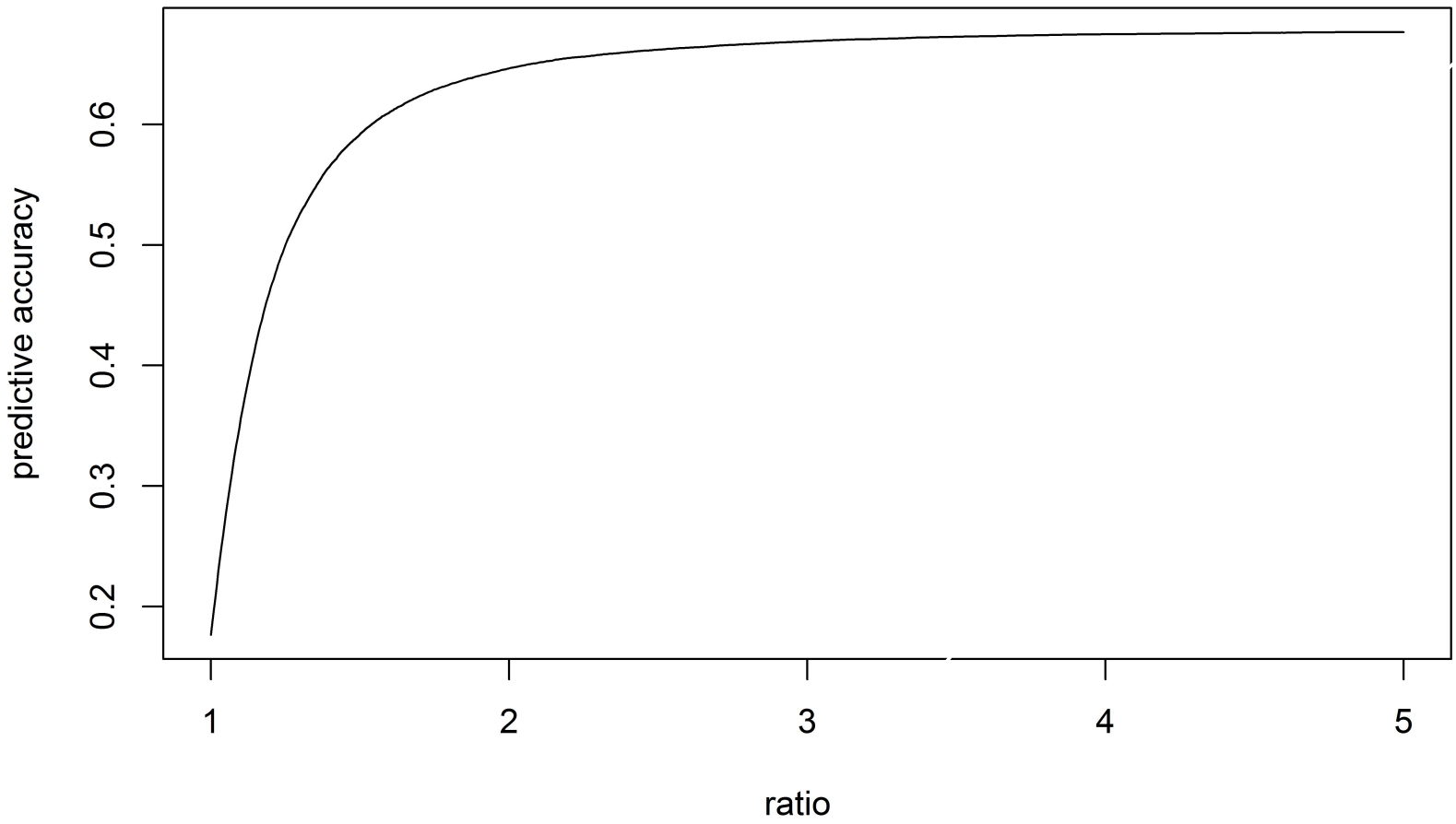
Results for sales and $MTF_{t,i,c}(k)$

- $k \approx 1.6$ (approx.), $\text{accuracy}(k) \approx 0.680$ (approx.)



Results for sales and $LTF_{t,i,c}(k)$

- $k \approx 1.7$ (approx.), $\text{accuracy}(k) \approx 0.677$ (approx.)



Recap

- metrics give similar ratios
- employment seems to be more sensitive
- calculated ratios — recommendation for the respondents
- future research
  - finding ratios for trade, construction and selected services
  - survey on survey

Thank you for your attention!

veronika.ptackova@vse.cz
lubomir.stepanek@{lf1.cuni, fbmi.cvut}.cz