

11th Comparative Analysis of Enterprise Data & COST Conference 2012  
26-28. 4. 2012, Federal Employment Agency, Nuremberg, Germany

# Focused information on skills demand using internet job search data

(with results for Slovak university graduates)

Ekonomický ústav SAV



Institute of Economic Research

Miroslav Štefánik  
miroslav.stefanik(at)savba.sk

# Motivation

- Skills surveys are expensive when realized properly
  - **Methodological problems** (low response rate, unclear definitions and categories, occupational variety ...)
  - **Realized surveys– often contradictory results** (looking at different sector- occupational segments, using different definitions)
- One web page dominantly covering the jobsearch market in Slovakia
- Availability of the data (low price, interest of the company)

# Classical vs. CV method

- Classical method
  - Content analysis of job advertisements
  - Counting the occurrence of a demanded skill (declared demand)
- CV method
  - Individuals upload CVs with declared skills (declared supply)
  - Employers search the CVs to fill a vacancy
  - Skills are visible for free, the display of contact details is charged
  - Counting the numbers of displays

# Classical method data:

- Employers are declaring the demand-employees choose
- Data contains:
  - **Basic information on the job vacancy**
    - *position name -string and categorized*
    - *location*
    - *type of contract ...)*
  - **Requirements on the applicant**
    - *level of required education*
    - *years of previous experience*
    - *required skills -string and categorized, ...*
  - **Information on the company**
    - *location*
    - *number of employees*
    - *business area*

# CV method data:

- Future employees are declaring the supply-employers choose
- Data contains:
  - gender
  - age
  - **achieved education** (university, faculty)
  - **experience**
  - **skills** (language skills, economic knowledge, office software s., advanced computer skills, ...)
  - **industry**, where the individual is looking for a job
  - **position**, which the individual is addressing
  - **region**, in which the individual is willing to enter a job
- **Number of displays**

# Representativeness

- Focused – not representative information
- Important to know the biases
  - Internet users
    - Age (young- up to 40)
    - Educational level (Upper -secondary and tertiary)
  - University graduates
    - Educational field
      - Technical and business faculties graduates are overrepresented
      - Artistic, medicine and pedagogical faculties graduates are underrepresented

# Representativeness-occupation

	Labour Force Survey		Advertisements		CVs	
	Frequency	Percent	Frequency	Percent	Frequency	Percent
Managers and Legislators	16144	<b>18,19</b>	16758	<b>16,65</b>	3125	<b>12,65</b>
Professionals	48714	<b>54,90</b>	42550	<b>42,26</b>	11155	<b>45,17</b>
Technicians and associate professionals	16715	<b>18,84</b>	21507	<b>21,36</b>	4282	<b>17,34</b>
Clerks	2723	<b>3,07</b>	11622	<b>11,54</b>	3886	<b>15,73</b>
Service workers	3035	<b>3,42</b>	6077	<b>6,04</b>	1817	<b>7,36</b>
Skilled agricultural workers	40	<b>0,05</b>	24	<b>0,02</b>	1	<b>0,00</b>
Craft and related trades workers	393	<b>0,44</b>	835	<b>0,83</b>	169	<b>0,68</b>
Operators and assemblers	371	<b>0,42</b>	286	<b>0,28</b>	18	<b>0,07</b>
Elementary occupations	593	<b>0,67</b>	1018	<b>1,01</b>	245	<b>0,99</b>
Total	88728	<b>100,00</b>	100677	<b>100,00</b>	24698	<b>100,00</b>

Source: Labour Force Survey and web jobsearch data

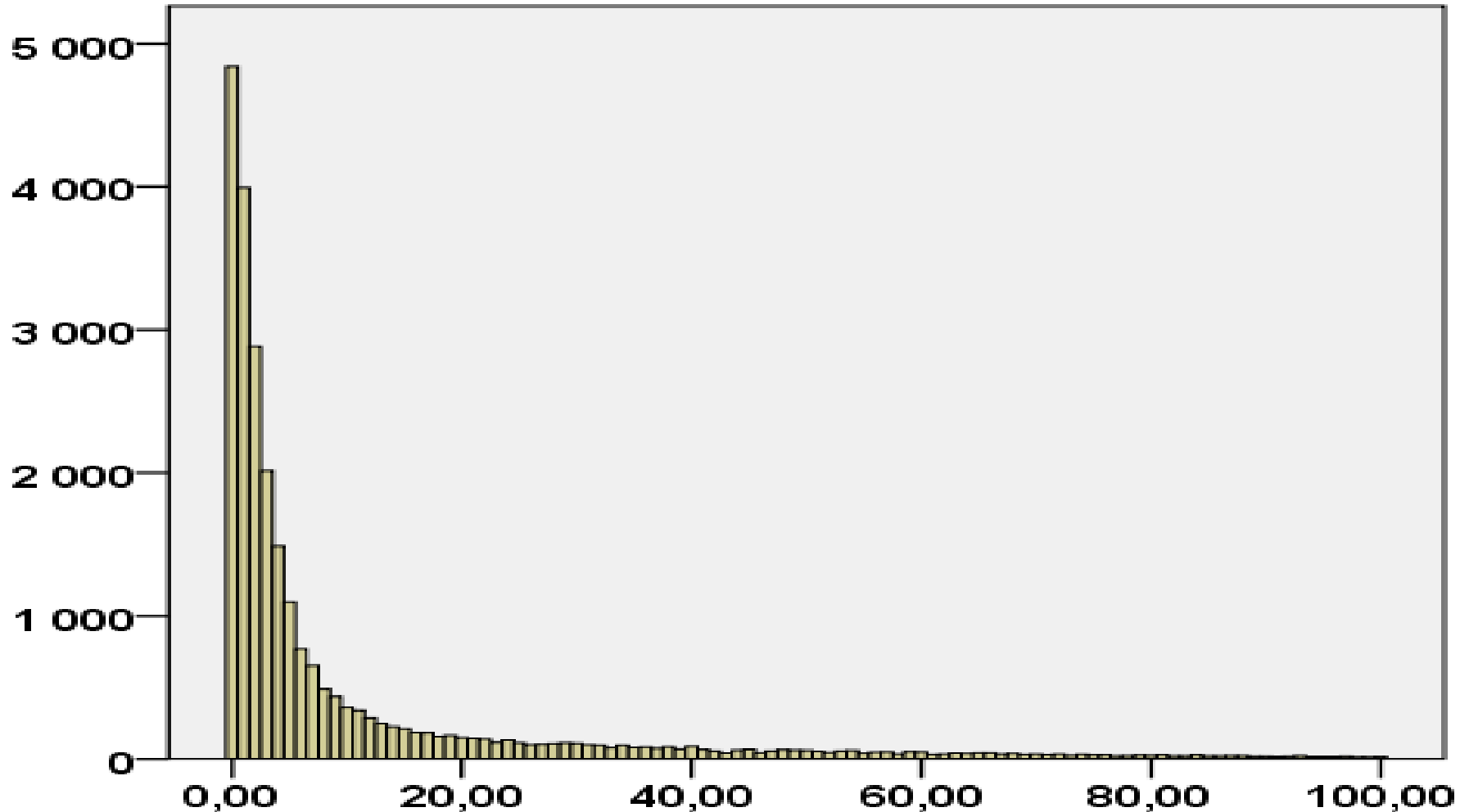
# Representativeness-economic sector

	Labour Force Survey		Advertisements		CVs	
	Frequency	Percent	Frequency	Percent	Frequency	Percent
Agruculture and mining (A-C)	1969	<b>2,28</b>	2780	<b>2,67</b>	356	<b>1,39</b>
Industry and utilities (DA-DJ +DN)	4490	<b>5,19</b>	3512	<b>3,37</b>	548	<b>2,14</b>
Electro-machinery and utilities (DK-DM +E)	6590	<b>7,62</b>	6345	<b>6,10</b>	2138	<b>8,34</b>
Construction (F)	2983	<b>3,45</b>	3431	<b>3,30</b>	748	<b>2,92</b>
Sales (G-H)	7120	<b>8,23</b>	18258	<b>17,54</b>	1798	<b>7,01</b>
Services (I-K)	16753	<b>19,37</b>	61846	<b>59,42</b>	17383	<b>67,81</b>
Public Services (L-O)	46567	<b>53,85</b>	7911	<b>7,60</b>	2663	<b>10,39</b>
Total	86472	<b>100,00</b>	104083	<b>100,00</b>	25634	<b>100,00</b>

Source: Labour Force Survey and web jobsearch data



# Distribution of number of displays (dependent variable $\leq 100$ )



# Descriptive statistics of number of displays

Displays		
N		25634
Mean		12,68312
Median		3
Mode		0
Std. Deviation		24,49471
Minimum		0
Maximum		311
Percentiles	10	0
	20	1
	30	1
	40	2
	50	3
	60	5
	70	8
	80	16
	90	39

# Data processing

- Filtering 3 occupational groups:
  - Programmers, Sales managers, Accountants
- Classical method data:
  - Frequencies of occurrences of skills in advertisements
- CV method data:
  - Generalized linear model
    - log link function
    - Poisson distribution
    - Number of days published-offset variable
  - Number of displays = skills + control variables

# CV method modelling

- Basic equation
- $D = \text{skill indexes} + \text{control variables}$

## Skill indexes:

language skills  
administrative and professional skills  
office software skills  
graphical software skills  
programming language skills  
database skills  
IT systems administration skills

## Control variables:

Gender  
Finished university  
Region  
Age  
Declared level of Slovak (5 Dummies)

- 7 elaborated equations
- $D = \text{skills from one index} + \text{other skill indexes} + \text{control variables}$ 
  - 165 skills included

# Results-programmers

		% of adds		Wald Chi-Square
1	<b>English</b>	91,6	<b>Serbian</b>	17,045
2	<b>German</b>	17,8	<b>French</b>	8,012
3	<b>Java</b>	17,2	<b>SQL</b>	7,799
4	<b>SQL</b>	15,3	<b>German</b>	6,726
5	<b>C/C++</b>	12	<b>English</b>	6,183
6	<b>Java EE</b>	10,2	<b>Microsoft Dynamics NAV</b>	5,705
7	<b>C#</b>	9,3	<b>Autodesk 3ds Max</b>	5,702
8	<b>HTML</b>	8,9	<b>Python</b>	4,889
9	<b>Microsoft Excel</b>	8,3	<b>Human Resources</b>	4,331
10	<b>JavaScript</b>	7,7	<b>4GL</b>	4,32
11	<b>XML</b>	7,2	<b>Microsoft Powerpoint</b>	4,112

# Results- sales managers

		% of adds		Wald Chi-Square
1	<b>English</b>	67,3	<b>Fortran</b>	24,795
2	<b>Microsoft Excel</b>	61,5	<b>Internet (e-mail, www)</b>	12,926
3	<b>Microsoft Word</b>	59,6	<b>Microsoft Visual FoxPro</b>	10,139
4	<b>Microsoft Outlook</b>	40	<b>DirectX</b>	9,884
5	<b>Internet (e-mail, www)</b>	37,8	<b>LotusScript</b>	9,763
6	<b>Microsoft Powerpoint</b>	26	<b>Pro/ENGINEER</b>	8,814
7	<b>German</b>	18,8	<b>Human Resources</b>	8,084
8	<b>Microsoft Windows</b>	10,5	<b>Double-entry bookkeeping</b>	7,762
9	<b>Slovak</b>	8,4	<b>Microsoft Powerpoint</b>	6,34
10	<b>Business correspondence</b>	5,3	<b>Java</b>	6,058
11	<b>Invoicing</b>	5,1	<b>Windows server administration</b>	5,191
12	<b>Hungarian</b>	4,5	<b>LAN/WAN administration</b>	4,854
13	<b>Warehouse management</b>	2,6	<b>Client/server administration</b>	4,474
14	<b>OpenOffice</b>	2,2	<b>SIMATIC STEP 7</b>	4,425
15	<b>Russian</b>	1,7	<b>Lotus Notes</b>	4,031

# Selected results- accountants

		<b>% of adds</b>		<b>Wald Chi-Square</b>
1				
2	<b>English</b>	79,1	<b>POHODA</b>	18,474
3	<b>Microsoft Excel</b>	43,8	<b>English</b>	18,217
4	<b>Double-entry bookkeeping</b>	39,7	<b>Lotus Notes</b>	18,103
5	<b>Microsoft Word</b>	37,2	<b>Microsoft Powerpoint</b>	15,348
6	<b>German</b>	20,2	<b>French</b>	13,824
7	<b>Microsoft Outlook</b>	19	<b>Microsoft Visual FoxPro</b>	10,491
8	<b>Internet (e-mail, www)</b>	18,2	<b>German</b>	9,652
9	<b>Invoicing</b>	17,5	<b>Internet (e-mail, www)</b>	9,565
10	<b>Payroll accounting</b>	8,6	<b>Business correspondence</b>	9,226
11	<b>Cash register</b>	8,4	<b>SQL</b>	8,594
12	<b>Business correspondence</b>	5,7	<b>Invoicing</b>	7,69
13	<b>Single-entry bookkeeping</b>	5,6	<b>Adobe PageMaker</b>	7,589
14	<b>Microsoft Powerpoint</b>	5,2	<b>SAP</b>	6,623
15	<b>Warehouse management</b>	5,1	<b>Single-entry bookkeeping</b>	6,338

# Conclusions 1:

- Internet job-search data presents a source of information on skills demand, which is comparable to skills surveys
- There are 2 types of internet jobsearch data
  - Classical method data
  - CV data
- We should be aware of representativeness biases, when interpreting the results



# Conclusions 2:

- Generalization is problematic
- We are able to quantify the importance of skills for selected occupations
- Programmers:
  - Declared demand: English, Java, C++, SQL
  - Undeclared: Python, other languages, SQL
- Sales managers:
  - Declared demand: English, MS Office
  - Undeclared: Fortran, Fox Pro, Lotus Script
- Accountants:
  - Declared demand: English, Excel, Bookkeeping
  - Undeclared: POHODA, Lotus Notes, Fox Pro

# Thank you for your attention

Ekonomický ústav SAV

**EÚ SAV**

Institute of Economic Research

Miroslav Štefánik  
miroslav.stefanik(at)savba.sk