

Kureková, Lucia Mýtna; Beblavý, Miroslav; Thum-Thysen, Anna

Article

Using online vacancies and web surveys to analyse the labour market: A methodological inquiry

IZA Journal of Labor Economics

Provided in Cooperation with:

IZA – Institute of Labor Economics

Suggested Citation: Kureková, Lucia Mýtna; Beblavý, Miroslav; Thum-Thysen, Anna (2015) : Using online vacancies and web surveys to analyse the labour market: A methodological inquiry, IZA Journal of Labor Economics, ISSN 2193-8997, Springer, Heidelberg, Vol. 4, pp. 1-20, <https://doi.org/10.1186/s40172-015-0034-4>

This Version is available at:

<https://hdl.handle.net/10419/152358>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<http://creativecommons.org/licenses/by/4.0/>

ORIGINAL ARTICLE

Open Access



Using online vacancies and web surveys to analyse the labour market: a methodological inquiry

Lucia Mýtna Kureková¹, Miroslav Beblavý² and Anna Thum-Thyssen^{3*}

* Correspondence:

anna.thum@ec.europa.eu

³Centre for European Policy Studies (CEPS) and European Commission (EC), CEPs: 1 Place du Congrès, 1000 Brussels, Belgium
Full list of author information is available at the end of the article

Abstract

This paper discusses methodological issues arising from the use of online job vacancy data and voluntary web-based surveys to analyse the labour market. We highlight the advantages and possible disadvantages of using online data and suggest strategies for overcoming selected methodological issues. We underline the difficulties in adjusting for representativeness of online job vacancies, but nevertheless argue that this rich source of data should be exploited.

JEL codes: E4, J2

Keywords: Internet; Job; Labour market; Methodology; Occupation; Online; Representativeness; Vacancy

1 Introduction

Internet-based data collection and research are growing research areas with a strong potential to deepen and widen our knowledge about various socio-economic issues. A particular area of increasing interest is the use of innovative data sources and analytical methods for the study of the labour market (Amuri and Marcucci 2010; Askitas and Zimmermann 2015). Many aspects of job search have been transformed due to the availability of online tools for job search, candidate search and job matching (European Commission and ECORYS 2012). Indeed Kuhn (2014) and Kuhn and Mansour (2014) provide evidence of the prominent role of online tools in job matching. This also presents new opportunities to collect and analyse web-based data about labour market demand and supply, which can enrich our micro-level understanding of pertinent issues such as skill and task requirements of employers, occupational change, wages and working conditions.

The debate over whether these new data sources should be used more extensively and to what purpose is still open. Two data sources that have been used most frequently for analysis are job vacancies posted online (via dedicated portals or the employers' websites) and voluntary web surveys (e.g. Wageindicator). One of the most prominent qualities of data collected or generated online is the large number of observations one can obtain. However, the potential value of such data sources goes beyond the sample size. Collection of vacancy data offline is costly and time consuming and online availability provides opportunities to access and analyse the content of job advertisements to better understand what employers require in a way which would not be

© 2015 Kureková et al. **Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

practically feasible using traditional newspaper sources of the past. Advantages such as time and cost effectiveness and the easy variability of survey questions also apply to voluntary web-based surveys (Wade and Parent 2001; Steinmetz and Tijdens 2009; Mang 2012). Voluntary web-based surveys collect information about wages and working conditions, which are problematic in representative surveys where people do not report their wages or where more detailed information about the working environment is absent.

Several methodological issues persist, however, which are predominantly related to the quality, reliability and representativeness of such data as well as generalisability of findings (Gosling et al. 2004; Steinmetz et al. 2009; de Stefánik 2012). These concerns are especially pertinent in research fields such as political science, economics and sociology, where the core quantitative analytical tools are based on representative¹ data and inferential statistical analysis. Despite these possible limitations, studies using online data and advancing research methods with new technologies have been published in leading social science journals, suggesting that the field is likely to expand rather than decline (Bellou 2015; Edelman 2012; Sappleton 2013; Taylor et al. 2014). Given the increasing reliance on Internet-based recruitment and the spreading access to the Internet across socio-economic groups and countries, it is highly likely that reliance on such data will grow. Indeed, Askitas and Zimmermann (2015) argue that as the Internet population becomes more and more equal to the total population, sampling may even become obsolete in cases where researchers would have access to the full data.

The aim of this paper is to address the methodological issues arising from the use of online vacancy and voluntary web-survey data to understand labour market developments. It reviews a selection of existing empirical works using online data from various disciplines in order to identify ways in which researchers and analysts have been dealing with such methodological issues. While methodological concerns are usually acknowledged in such papers, they are not given prominence there, and there is a lack of cross-cutting methodological work addressing these issues. The paper plugs this gap. It both highlights the advantages and possible disadvantages of using online job vacancies and proposes strategies for dealing with the issue of representativeness and generalisability of findings. The paper advances the debate on how some of the weaknesses could be corrected, for which types of research questions and research subfields they might be of a smaller concern, and which policy areas can be informed by analysis of these types of online data.

While we are aware of other types of Internet data that have been increasingly used for research of labour market and human resource issues — e.g. data about individual preferences, career histories, Google analytics data, social networking, nowcasting and forecasting, etc. (Reips 2006; Askitas 2009; Martínez-Torres et al. 2014; Vicente et al. 2015) — we prefer to focus on the vacancies and voluntary web surveys to enable in-depth discussion of methodological issues. The remainder of the paper is organised as follows: the second section discusses methodological issues arising in different forms of web-based data. The third section suggests solutions to selected important methodological issues, in particular representativeness. The fourth section concludes.

2 Methodological and research design questions arising from the use of web-based data

Internet-based data collections that focus on labour market research and analysis differ in their objectives and scope. In addition to more traditional activities such as posting

vacancies and CVs, there has been a proliferation of sites offering employer evaluations and salary comparisons in recent years.

First, there are projects whose explicit objective is the collection of data on selected aspects of the labour market, such as wages or employer evaluations based on web surveys. These data are gathered to enhance understanding of country-specific aspects of the labour market from the supply side (e.g. the WageIndicator project²). Commercially-oriented websites (e.g. Salary.com or Payscale) provide a similar service, but with less research-oriented focus. More ambitious services, such as Glassdoor or Vault, aim to integrate salary information into wider information about employers and working environments. Such information has already been used in research (Young and Case 2004; DeKay 2013).

The second type is online job portals that gather vacancies and CVs and serve as important platforms for labour market matching (Mang 2012). As the objective of online portals is not tied to research but rather to providing a platform on which demand and supply meet, data are seldom stored and used as an input to analyse labour market trends and developments. Nevertheless, it appears that engagement with online vacancy data has been more widespread in the US. For example, job openings activity has been measured by job advertisements in print as well as online to estimate the state of the labour market in policy environment in the form of a composite Help-Wanted index (Barnichon 2010). A useful potential source of online data in the European Union is the EURES website, which collects job vacancies across the EU countries in a standardised platform³. Its particular utility could potentially lie in enabling cross-country comparisons, which could inform policy-makers at the national and EU level (Tijdens et al. 2015; Kureková et al. 2015). In general, research engaged with online vacancy data has typically relied on private job portals (Backhaus 2004; Capiluppi and Baravalle 2010; Stefánik 2012; Kuhn and Shen 2013; Masso et al. 2014; Beblavý et al. 2015).

In the following subsections we will discuss methodological issues arising in web-based survey data and in online job portals respectively. The literature referred to is summarised in Table 1 below.

2.1 Web-based data based on survey data

Research on the use of web-based voluntary surveys appears to be methodologically more advanced. This type of data has already developed a relatively strong following in the field of psychology since mid-1990s (Gosling et al. 2004; Reips 2006; Reips and Buffardi 2012).

Representativeness however is an issue often discussed in various methodological studies of web-based survey data. For an extensive collection of both methodological and content-based literature on web survey data we refer the reader to the network Webdataset⁴ or the online bibliography of Web Survey Methodology⁵ and highlight some recent comprehensive studies in the following. Based on an extensive review of existing studies and surveys, Couper (2000) for instance points out that while web-based survey data has the advantage of easy access and potentially wide coverage, it is more difficult to determine the quality of the survey, and as the number of surveys grows it becomes more difficult to conduct high quality surveys as participants may become wary. Indeed, he points to coverage and sampling errors as important issues to

Table 1 Overview of referenced literature concerning methodological or content research based on web surveys and job portal data (Section 2)

Authors	Type of paper	Data	Methods	Results
Askitas and Zimmermann (2015)	Methodological	Literature and database survey	Review	As the internet population is becoming more and more equal to the total population, sampling may even become obsolete in cases where researchers would have access to the full data. On the other hand privacy issues can be a threat to the use of internet data.
Barnichon (2010)	Methodological	Job advertisements in print and online	Elaboration of composite index	Calculates a composite Help-Wanted index used in the US in policy and research to estimate the state of the labour market.
Couper (2000)	Methodological	Literature and database survey	Review	Web-based survey data has the advantage of easy access and potentially wide coverage, it is more difficult to determine the quality of the survey and as the number of surveys grows it is becoming more difficult to conduct high quality surveys as participants may become wary.
Dillman et al. (2014)	Methodological	-	Methodological guide	Guide to conducting web surveys making use of comparisons to more traditional survey methods to show challenges arising across various types of survey methods such as non-response and sample selection.
Gosling et al. (2004)	Methodological	Internet sample and traditional sample	Systematic comparison	Web surveys can be characterised by certain 'preconceptions' related to the quality of the data, such as the lack of demographic diversity of Internet samples, social isolation and depression of Internet users, anonymity of participants, lack of motivation of respondents inter alia
Guzi and Kahanec (2014)	Methodological/ Content research	WagelIndicator	ILO standards described in Anker (2011)	Computation of living wages data
de Leeuw (2005, 2010)	Methodological	Literature and database survey	Review	Discusses the advantages and disadvantages of mixed survey methods based on written, telephone and internet surveys.
Mang (2012)	Methodological	Literature and database survey	Review	Job portals serve as an important platform for labour market matching.
de Pedraza et al. (2010)	Methodological	WagelIndicator data and Labour Force Survey in Spain	Statistical analysis of labour market issues	Some parameters moved in the right direction after weighting, but other methods need to be considered, including propensity score adjustment or a priori measures, such as marketing and targeting.

Table 1 Overview of referenced literature concerning methodological or content research based on web surveys and job portal data (Section 2) (Continued)

Reips (2006)	Methodological	Literature review	Categorisation of web-based data in psychology	In the field of psychological assessment, methodology is likely to develop and adapt to accommodate benefits of web-based techniques of data collection
Reips (2012)	Methodological	Literature and database survey	Review	He concludes that while there are still doubts about whether this relatively new data source should be widely used, it is already used in numerous studies.
Reips and Buffardi (2012)	Methodological	Social networking webpages	Review of Internet-Based Research Methods and application to social networking data	Internet data can be used for research both with migrants and on the psychological impressions others have of migrants.
Schonlau et al. (2009)	Methodological	US Health and Retirement study		Adjustments help to correct variances in sample means but differences between a general sample and an internet sample remain large on a number of parameters.
Steinmetz and Tijdens(2009)	Methodological	WagIndicator data and German SOEP data	Implementation of different weighting techniques to correct web-based data	Different correction techniques do not fundamentally improve the representativeness of web survey data, but more research is needed to identify why this is the case and suggest strategies for using web-based survey data.
Steinmetz et al. (2009)	Methodological	German and Dutch WagIndicator data	Implementation of different weighting techniques to correct web-based data	The impact of the applied weights (post-stratification and propensity score adjustment) is very limited - using balancing variables does not make websurvey data more comparable to the general population.
Steinmetz et al. (2013)	Methodological	WagIndicator data: US, Argentinean, Brazilian, Dutch, German and Spanish	Descriptive analysis, correlation analysis using post-stratification weighting.	The impact of post-stratification weights is rather limited, but more appropriate in the case of descriptive analysis.
Stern et al. (2014)	Methodological	Literature review	-	Analyse the current state of survey methodology and point towards growing opportunities but also challenges as well as future developments that need to be observed.
Backhaus (2004)	Content research	Job advertisements at Monster.com	Content analysis of text	Recruitment tactics of firm vary across different industries; firms typically focus on company branding than on employee benefits.
Beblavý et al. (2015)	Content research	Private Slovak job portal	Content analysis of text, simple statistical methods	Employers in Slovakia are fairly demanding even in formally low-skilled jobs requiring a wide set of skills.
	Content research	Collective Agreements	Automated text analysis	The minority of agreements include specific wage levels

Table 1 Overview of referenced literature concerning methodological or content research based on web surveys and job portal data (Section 2) (Continued)

Besamusca and Tijdens (2015)		Database of the WagelIndicator		while clauses on social security, working hours and work-family arrangements are included in the majority of cases.
de Bustillo and de Pedraza (2010)	Content research	WagelIndicator	Logistic regression	Age, wages, education as well as the type of contract determine job insecurity.
Capiluppi and Baravalle (2010)	Content research	Job advertisements for IT staff at Monster.com	Content analysis of text	A mismatch between requirements of UK industry and offer of educational and training institutions exists.
DeKay (2013)	Content research	Glassdoor	Content analysis of text, structured comparison	Testing and largely confirming validity of Frederick Herzberg's Motivator-Hygiene theory
Guzi and de Pedraza (2015)	Content research	WagelIndicator	Ordinary least squares	Certain job characteristics going beyond pay such as job insecurity determine job satisfaction and other subjective well-being indicators.
Huang et al. (2009)	Content research	Scholarly articles, practitioner literature and online job ads in IT profession	Content analysis of text, structured comparison	Findings suggest that the online advertisements list a wider mix of skills, while practitioner literature tends to focus heavily on technical skills.
Kennan et al. (2006)	Content research	Online job advertisements in Australia and US	Content analysis of text	The study finds differences in the relative importance of skills across the two countries, and also variations over time. The results are presented as generalised for the librarian profession under study.
Kuhn and Shen (2013)	Content research	Chinese private job portal	Web-crawler, complex statistical methods	Explicit gender discrimination exists in the Chinese labour market.
Kureková et al. (2015)	Content research	EURES	Web crawling and automated text analysis	Non-cognitive skills matter in the service sector in a selected number of countries.
Masso et al. (2014)	Content research	Online job portal from Estonia	Complex statistical methods	The authors did not find any positive effect on occupational mobility of return migrants, but this could be due to a short duration of migration or the character of home labour market.
Maurer-Fazio and Lei (2015)	Content research	China's Internet job board and own online survey	Descriptive statistics and multivariate probit	Candidates with faces considered being unattractive need to send 33% more applications than those considered being attractive.
Maurer-Fazio (2012)	Content research	China's Internet job board and field experiment	Multivariate and multinomial probit	Significant differences in the call-back rates based on ethnicity as well as heterogeneity in call-back rates across ethnic groups.
Kureková and Zilinčíková (2015)	Content research	Profesia.sk	Logistic regression	Low educated workers and student workers do not compete for jobs but rather

Table 1 Overview of referenced literature concerning methodological or content research based on web surveys and job portal data (Section 2) (Continued)

				provide employers with different skill sets in a complementary way.
Shen and Kuhn (2013)	Content research	Chinese private job portal	Web-crawler, complex statistical methods	Findings point to difficulties in labour market integration of Chinese college-educated youth
Stefánik (2012)	Content research/Methodological	CVs and vacancies from private Slovak job portal and national LFS data	Content analysis and distributional frequencies	Analysis of online data can produce valuable information on skill demand for some particular groups in labour market.
Tijdens et al. (2015)	Content research	WageIndicator and EURES	Web crawling and descriptive statistics	Mismatches can be identified for about one fourth of the examined occupations on the demand side and one third on the supply side.
Young and Case (2004)	Content research	Voluntary web survey	Descriptive statistics	Examination of implementation levels of Internet use policies, management training, and clinical rehabilitation and their level of perceived effectiveness to deter employee Internet abuse.

take into account in inference based on this type of data, but also non-response can be a threat.

Dillman et al. (2014) provide a guide to conducting web surveys making use of comparisons to more traditional survey methods to show challenges arising across various types of survey methods such as non-response and sample selection. Stern et al. (2014) analyse the current state of survey methodology and point towards growing opportunities but also challenges as well as future developments that need to be observed. Reips (2012) provides an overview of web-based data collection and draws the reader's attention to challenges as well as advantages. He concludes that while there are still doubts about whether this relatively new data source should be widely used, it is already used in numerous studies. Like Askitas and Zimmermann (2015) he points out that privacy issues may be an obstacle in the future. De Leeuw (2005, 2013) discusses the advantages and disadvantages of mixed survey methods based on written, telephone and internet surveys.

In the area of labour market research in particular, both methodological and content-research studies related to the WageIndicator project (see above) are relatively advanced with respect to testing data representativeness and ways to improve its usability for generalisable research and also allowing comparisons across countries. The key concern behind the use of these data is the sampling error that can occur due to the fact that respondents are not randomly selected from a representative sampling frame, but the target population rather forms a 'convenience sample' of self-selected individuals⁶. This is closely related to the coverage error. The socio-economic and demographic characteristics of the pool of Internet users differs in important respects from those of the general population, e.g. the elderly or people with lower levels of education typically

have lower access to the Internet. Research results based on web-based data are therefore unable to provide inferences for these categories of people, or only with limited precision and reliability.

An additional source of error is the non-response bias, which appears if non-respondents differ from the survey respondents in important characteristics, again making the inferences based on such data imprecise with respect to the overall target population. Compared to online vacancy data (which we discuss below), an important advantage of web-based individual voluntary surveys is the fact that the sample population's characteristics can typically be ascertained. In developing countries, where representative surveys might not exist, web-based surveys can provide a valuable and unrivalled source for understanding these labour markets. Critiques of web-based data should also keep in mind that even probability-based samples face problems of self-selection, non-response or coverage.

Censuses or representative labour market surveys can be used to compare the characteristics structure of web survey respondents and potentially make adjustments. Among main weighting techniques that have been suggested to correct for the biases inherent in web surveys are post-stratification weighting and propensity score adjustment (Steinmetz et al. 2009). Post-stratification weighting can be applied to correct mainly for demographic differences in the data. Propensity score adjustment can also be applied to correct for socio-demographic as well as attitudinal or behavioural differences related to an individual's decision to take the survey. In WageIndicator data, results of these corrections found that different correction techniques do not fundamentally improve the representativeness of web survey data (Steinmetz et al. 2009; Steinmetz and Tijdens 2009). At times unweighted samples can bring more consistent results between web-based survey data and representative samples than adjusted data samples. Comparisons also revealed that types of biases differ across countries and are often related to the overall income inequality in the country, and the strength of biases can vary across different variables (Steinmetz and Tijdens 2009; de Steinmetz et al. 2013). Schonlau et al. (2009) used the US Health and Retirement study to compare an Internet sample and general sample and found that adjustments helped to correct variances in sample means but that differences remained large on a number of parameters. It hence appears that the type of required correction of a web-based survey needs to be tested specifically for a particular sample.

The WageIndicator data has already been used in numerous research projects. Guzi and de Pedraza (2015) use weighted WageIndicator data to study life- and job satisfaction, and based on an Ordinary Least Squares (OLS) analysis, they confirm the importance of certain job characteristics going beyond pay such as job insecurity for job satisfaction and other subjective well-being indicators. De Bustillo and de Pedraza (2010) conduct a logistic regression analysis to study job insecurity and find that age, wages, education as well as the type of contract determine job insecurity. Guzi and Kahanec (2014) make use of WageIndicator to determine the so-called living wage based on a set of assumptions and methodological standards outlined in Anker (2011). Besamusca and Tijdens (2015) use the Collective Agreements Database from the WageIndicator to retrieve data on collective bargaining, and based on automated text analysis and descriptive statistics, they find that the minority of agreements include specific wage levels, while clauses on social security, working hours and work-family arrangements are included in the majority of cases. Finally, Tijdens et al. (2015) study

skill mismatches by concepts from employing descriptive statistics and comparing the distributional characteristics of requirements and attainments using both vacancy data (EURES) and data of jobholders from the WageIndicator web-survey. They find that mismatches can be identified for about one fourth of the examined occupations on the demand side and one third on the supply side.

2.2 Web-based data from online portals

Job advertisement research using online data sources is quite recent, but it has been preceded by studies based on printed job advertisements. These typically test large sociological theories and concepts, such as class, merit selection or characteristics of modern industrial societies (Jackson 2001; Jackson et al. 2005; Jackson 2007). No concern over a possible selection bias or the representativeness of overall labour market demand is expressed in the studies and the findings are discussed in a generalisable form. For example, Jackson (2007) performed content analysis of advertisements from national and local newspapers with a high circulation, and the sample was chosen to be representative of the range of occupations in the occupational sectors. Findings were generalised to inform new trends in how occupational positions are allocated on the basis of meritocratic criteria, questioning classes as an appropriate unit of sociological research in modern industrial societies. Dörfler and van de Werfhorst (2009) analyse the Austrian labour market and test the merit selection hypothesis over time. Printed job advertisements enable retrospective historical analysis, which is not possible with surveys, which typically gather data at a particular point in time. The authors suggest that the problem of non-response, typical for survey research, is not present, although other biases might exist, such as underrepresentation of certain skill levels. They apply inferential statistical methods (multivariate regression analysis) to test inductively formulated hypotheses. They find that employers require a wider set of skills over time, including social and personal skills.

Studies using online job advertisements differ from those relying on printed job adverts typically in the number of ads they analysed and, relatedly, in the techniques they are able to employ. A leading private internet recruitment site – Monster.com – has been used as a source of online job vacancy data in various studies. Capiluppi and Baravalle (2010) developed a ‘web spider’ to download vacancies from the Monster.com website and then analysed the skills required of IT personnel in the UK using content analysis. They found a mismatch between requirements of UK industry and offer of educational and training institutions. Backhaus (2004) analysed job advertisements from Monster.com from the perspective of employers. Using content analysis she studied corporate descriptions in job adverts to understand aspects of company branding and marketing in human resources, pioneering this type of research. She finds that firms focus more on presenting company characteristics than employee advancement and that differences exist across firms in different industries in their recruitment tactics.

Kuhn and Shen (2013) study gender discrimination in the recruitment process in the Chinese labour market. Data were collected by a web-crawler from the third-largest online job portal in China. This led to a sample size of over a million job ads from the late 2000s, subsequently merged with firm data, suitable for sophisticated empirical analysis using advanced statistical methods. The results revealed high levels of gender

preference, although vacancies for highly skilled positions were less discriminatory. They acknowledge that their sample of job ads is not representative of the overall population of jobs in the Chinese labour market, using the 2005 census in China as a comparator dataset. While acknowledging these limitations, authors present general findings about aspects of explicit gender discrimination in the Chinese labour market with implications for policy-making in developing countries more generally.

Shen and Kuhn (2013) analyse job applications submitted in response to a selected number of job ads online in Chinese urban areas to study what effect over-qualification has on labour market entrants. Representativeness is not considered, and only the bias related to duration of vacancy posting is addressed by additional analysis. Sophisticated statistical tools are employed, and the findings are generalised for Chinese urban youth. Other studies analysing discrimination on the Chinese labour market based on field data, online job market data and their own internet survey include Maurer-Fazio (2012) and Maurer-Fazio and Lei (2015). The latter study examines the role of gender and facial attractiveness on China's Internet job board labour market. The authors find that candidates with faces considered being unattractive need to send 33% more applications than those considered being attractive. Based on a field experiment the former study examines the chances of job market candidates from ethnic minorities on China's Internet job board. The author finds significant differences in the call-back rates based on ethnicity as well as heterogeneity in call-back rates across ethnic groups.

Štefánik (2012) studied online data from a private job portal in Slovakia, Profesia.sk, analysing both vacancies and CV data. His studies concentrated on the labour market segment of the highly skilled and examined the matching of demand and supply of university graduates in the belief that they would be relatively well represented among the job applicants due to characteristics of Internet users and would therefore offer a representative data sample. His representativeness test was based on comparing the structure of portal vacancies and CVs to the structure of the whole population based on the national Labour Force Survey. Beblavý et al. (2015) use the same private online job portal data to study demand for formal qualifications and other skills in a wide range of low- and medium-skilled occupations by means of content analysis and simple statistical methods. They consider their findings generalisable for the Slovak context due to the portal's dominant market share and very high reputation of the portal among employers and employees. They find that employers in Slovakia are fairly demanding even in formally low-skilled jobs requiring a wide set of skills. Kureková and Žilincíková (2015) use a population of vacancies from the Profesia.sk portal to study characteristics of student labour market and test the crowding-out theory. They perform logistic regression and find that low educated workers and student workers do not compete for jobs but rather provide employers with different skill sets in a complementary way.

Online data have also been used in studies about migration where data collection is problematic. Masso et al. (2014) use data from an online job portal in Estonia and study the effect of foreign work experience on occupational mobility of return migrants using multivariate regression techniques. The authors did not find any positive effect on occupational mobility of return migrants, but this could be due to a short duration of migration or the character of home labour market.

Wider research attention has been given to IT-related professions, which have been on the rise in the past decades. Wade and Parent (2001) study the relationship between job

skills and performance of webmasters looking at job vacancy data and organising a targeted web-based survey of webmasters to determine the required skill mix and the degree to which subjective assessments of the possession of skills affect job performance in this occupation. They employ multivariate analysis techniques. The authors highlight the usefulness of researching online job ads in identifying the mix of skills sought after, especially in new professions, and in building profiles of positions, which can serve as valuable input to student counselling services or curricula development. They acknowledge that the coverage bias can be remedied by complementing their methodology with structured interviews with employers or recruiters.

Huang et al. (2009) examine technical, humanistic and business IT skills across three genres of text: scholarly articles, practitioner literature and online job ads. Findings suggest that the online advertisements list a wider mix of skills, while practitioner literature tends to focus heavily on technical skills. In order to construct reliable profiles of job positions, the study finds that it might be useful to review a wider range of sources.

Comparative studies based on job ads data are rather scarce. An exception is the work of Kennan et al. (2006), who study changing workplace demands for information specialists, looking specifically at the librarian profession to determine the required skills in Australia and the US. The study combines data from printed ads and online ads and by means of cluster analysis identifies 'skill clusters'. The study finds differences in the relative importance of skills across the two countries, and also variations over time. The results are presented as generalised for the librarian profession under study. Kureková et al. (2015) pioneer the usage of the European-wide publicly administered job-vacancy portal EURES. They perform content analysis and carry out a comparative study of employers' skill demand in small European economies. They find that the mix of skills called for is very diverse across the countries, implying that there is no universal set of requirements and also that domestic institutions and structures strongly affect how demand is formulated. The authors argue that EURES data are a well-suited source for comparative analysis due to their standardised platform and relatively wide usage across European countries (see also Ackers 2012).

To sum up, existing research using online or printed job vacancies has been characterised by a single-country focus and has grown mainly through the usage of data from private online job portals rather than publicly collected data (e.g. public employment services data). Various types of questions have been investigated, ranging from testing or enriching established social science concepts and theories (gender discrimination, social stratification, merit selection, expansion of service sector, company branding, crowding-out theory or migration) to relatively narrow and focused questions related to a particular sector or industry (IT sector, librarian profession). Interestingly, methodological concerns are acknowledged, but they are not given prominence in the reviewed studies. In the following section, we synthesise and critique the various approaches taken by different authors in attempting to deal with representativeness issues related to the usage of online job vacancy data for analytical and policy-making purposes.

3 Strategies for overcoming selected methodological issues

3.1 Existing approaches to increasing representativeness

A key challenge in using online job vacancies is ascertaining whether the set of online job vacancies is a representative sample of all job vacancies in a specified economy⁷.

Even if it should be considered finite⁸, the population of job vacancies at any given moment in time is not easily counted nor is its structure easy to determine. In only a handful of countries are employers obliged to report all job vacancies⁹.

And even if such reporting of vacancies (typically to the labour office) is mandatory, much hiring takes place internally or through informal means and networks. Jobs are reallocated in the labour market through many mechanisms, some of which do not entail a formal 'vacancy' announcement: people are reallocated internally, or tasks are split, restructured or partially switched. This is likely to affect certain aspects of the labour market more than others. For example, large international firms often first recruit from internally available candidates before announcing a vacancy in an open labour market. In smaller towns or villages where labour market participants have closer links and relations, job vacancies might first be offered to candidates known personally to the employer. Recruitment means and strategies might have sectoral and occupational specificities and can vary across countries (Teichler 2009; Keep and James 2010).

From this perspective, even if we collect all online reported job vacancies, there is a share of vacancies that are never publically advertised, and will therefore fall outside our population sample. On the other hand, vacancies tend to occur where there is either an unfulfilled demand or where employers for some reason prefer to have a selection process. Therefore, if we are interested to know which types of jobs employers find difficult to fill through internal or informal search channels, then online vacancy data can be highly useful. Due to the difficulties in identifying the structure of the *population* of vacancies, however, we argue that weighting as an adjustment technique, which has been tried in improving the representativeness of web-based voluntary surveys is hardly possible in the context of projects based on online job vacancy data.

The reviewed studies have — implicitly or explicitly — adopted rather different approaches to deal with the problem of (non-)representativeness of job vacancy data. Some researchers have used representative data describing the labour market structure, such as the Labour Force Survey (LFS), and judged the coverage of online vacancies based on the sectoral and occupational structure of LFS data (Jackson 2007; Štefánik 2012). We find this approach problematic, however, for a number of reasons. Foremost, the Labour Force Survey is not a straightforward measure of the structure of the demand side of the labour market but rather includes supply, demand and job matches. We therefore do not see the LFS as a suitable proxy for the demand side. Furthermore, current demand is subject to seasonal trends and reflects developments in a particular sector that might not match the existing structure but rather reflects future trends in a particular labour market segment. For example, vacancies in the IT industry in many countries are due to the recent rapid expansion of this sector, which is overrepresented among vacancies, and their share might not reflect the actual share of the industry in a national structure of employers and/or employees.

A more promising direction is the one taken by Van Ours and Ridder (1992), who tried to achieve representative results in their study of the duration of vacancies in order to determine aspects of employers' search strategy by selecting a 5% random sample from all establishments in the Netherlands. They faced attrition in the process of receiving responses from the firms they approached with a two-stage questionnaire (the first stage identified firms that were hiring, the second stage investigated aspects of the search strategy and the characteristics of the hired person). Although the authors adopted a rigorous sampling

framework, they were not able to deal fully with all representativeness problems due to non-responses and other forms of data-collection problems encountered. Moreover, such a research approach is obviously also expensive and time-consuming. The availability of online vacancies could potentially simplify some steps of their research process, but this needs to be evaluated and weighed against particular research questions and objectives.

Alternative approaches have been adopted to (partially) address the representativeness issue, building on the diversification of data sources. First, a number of studies decided to focus on the segment of the labour market where the coverage bias is likely to be less of a problem¹⁰. Examples include focusing on graduates' CVs and vacancies targeting this labour market segment or on sectors and professions that are by definition characterised by widespread access to the Internet (IT, librarians, webmasters, etc.). Second, diversification of data sources has been used in many studies. In addition to online job ads, other types of data sources can be analysed in parallel, such as practitioner literature or administrative data. A sample of vacancies based on administrative data can be created. Eurostat (2010) gives an overview of several such examples from European countries. For instance, Martikainen and Eurostat (2010) uses the Finnish Business Register to create a representative sample of job vacancies by weighting the observations in an appropriate way and using a suitable estimation technique that takes potential flaws into account. Data collection was done through computer-aided interviews and through the Internet. Around 10,000 establishments out of 150,000 were drawn by stratified sampling per year. Another example of an alternative approach takes the form of complementing content analysis of job advertisements with interviews with HR managers or recruiters. Third, some scope to correct possible biases might exist if online vacancies data can be linked to firm characteristics, which can then be controlled in quantitative analysis. Fourth, market coverage and technical advancement of the online job portal(s) in a given country need to be assessed in each country. In countries where a dominant portal exists, collected vacancies might be the best available source. These alternatives need to be weighed against the costs of collecting data by other means. Using job advertisements from an established portal and interpreting the results with caution to avoid potential biases can be a valid and acceptable choice.

3.2 Statistical methods to address sample design problems

Another way to address the problem of representativeness of online job vacancy data *as an accurate sample* of job vacancy data is to employ statistical tools that account for sample design problems. A common type of data used for analysing employers' preferences is survey data (see Colombo 2009), and statistical tools have been developed to address estimation problems arising from the survey design¹¹. We argue that some of these tools and potentially their variations can be suitably used to address problems arising from the way the respective online data has been collected¹².

These tools can be situated in the field of statistical analysis with missing data, which is a useful field of literature to understand how to estimate population parameters in the most accurate way concerning samples that are likely to not be random. Data can be missing either because of an accidental omission due to a variable unrelated to the variable with the missing values ('missing at random') or because of a specific reason that is related to the variable with the missing value ('not missing at random'). In our case, a job vacancy could be missing in the set of online job vacancies either because it

was not posted online or because it was not advertised at all. Both reasons would point to considering our data as containing not missing at random (NMAR) values. In fact, in our case the missing values are due to the sampling procedure: online job vacancies are a sample of the population of all job vacancies in which those values are missing that have not been posted online. If the sampling procedure is under the control of the statistician, it can be addressed. However, if it is not under his or her control, the assumptions made about the mechanism leading to the missing data need to be made clear (Little and Rubin 1987). We demonstrate below how a model could be constructed that accounts for over- or underrepresentation that may arise from the selection mechanism.

Little and Rubin (1987, 2002) provide a comprehensive overview of statistical analysis with missing data and distinguish four main methodological schools in addressing missing-value issues in statistics. We focus on a less frequently used approach¹³: the model-based approach. This method builds on the idea that the estimation of a population mean from a sample is a similar problem to the *prediction* of a population mean (Royall 1992). Several statistical works present the main features, limitations and advantages of the model-based approach (Valliant et al. 2000; Chambers and Skinner 2003; Longford 2005; Aitkin and Aitkin 2011). The idea is to predict — through an underlying probability model¹⁴ — quantities based on data that include unobserved values. We fill in the unknown parts of a data distribution by using our knowledge of the subject matter to construct a model of how that missing data could be determined. Such a model could take the form of a density of the variable in question, conditional on *i*) a set of other variables representing information used in the survey design and *ii*) a set of parameters. We argue that it could be a useful method in the context of analysing online job vacancy data, where adjustments need to be made to allow for variables correlated with over- or under-representation of certain types of observation units.

For the purpose of understanding employers' preferences with the help of job vacancies, the information we need from job vacancies are certain characteristics about the jobs that are advertised, such as the distribution of the need for a certain skill. It is these characteristics of occupations that are being adjusted for, such as required skills, rather than the number of each occupation *per se*. While we might expect the characteristics of a particular type of job to vary somewhat with establishment characteristics such as size, it seems reasonable to suggest that the core aspects of a given occupation are likely to remain constant across types of establishments, and it is the latter conditional relationship that we are trying to compensate for.

In order to infer the social-skill needs of a nurse from a sample consisting of online vacancies — which we believe to under-represent certain parts of the population of vacancies — we would construct a conditional distribution of skill needs given a certain number of covariates. How do we choose the variables to include in the set of covariates? We might believe that only nurse vacancies in hospitals with more than 10 employees are advertised. We might know from a set of interviews with doctors or hospital directors that the skill needs for a hospital with less than 10 employees differ from those of larger hospitals: in small hospitals more social skills are needed due to higher interaction with the patients. In that case we would not be able to make correct inferences on social-skill needs of all vacancies based on a sample excluding skill needs for the smallest hospitals. We could then first assess the extent of the bias by retrieving the number of small hospitals compared to

larger hospitals via external data such as an establishment panel, which contains a representative sample of companies. We might find out that the percentage of small hospitals is 15%. Our skill needs based on the online sample would consequently exclude 15% of the hospitals, in which social-skill needs are highest. We can think of the sample distribution of social-skill needs as truncated. We can now establish the trend of nurses' social-skill needs given the hospital size. In our case it would be a downward sloping curve in hospital size. To build the model that we will use to predict the social-skill needs of nurses, we would extrapolate this trend to the smallest hospital size and add a further upward sloping bit to the curve at the end of the smallest size.

Now we may ask ourselves whether hospital size is the only variable on which the online sample is underrepresented or overrepresented. To select these variables, we would need to study very carefully which vacancies we believe are placed online and which ones are not. Such variables could be region, company size, "openness of the country/ratio between tradable and non-tradable goods" (to account for whether the labour market is likely to be open for publishing vacancies on the globally available internet) or "importance of social capital in the country/occupational field" (to account for job vacancies that are not posted online because they are filled by connections). We could include all variables in our model and select the best-fitting model via model selection criteria such as the Akaike Information Criterion (AIC), which is a statistical tool to assess the quality of a statistical model.

Longford (2005) acknowledges that this approach is based on un-testable assumptions and therefore proposes to conduct a sensitivity analysis. His understanding of a sensitivity analysis is a tool to measure the impact of an assumption on the result of the analysis. He suggests in the case of not missing at random values to test sensitivity of results in one or several directions since the vast array of possible directions is not feasible to account for (Longford 2005). The way to implement such a sensitivity analysis in our case would be to simulate our model for cases with more or less large companies and see how the distribution of skill needs changes.

The estimation would be conducted based on maximum likelihood or Bayesian methods. The likelihood function would be set up, and it would take a certain functional form based on the assumptions about the error terms. In complex cases, the function might not be numerically tractable, in which case one would use numerical integration methods based on Bayesian or frequentist statistical methods¹⁵.

The limitation with this approach would be that i) we do not know how much of the missing data in the population we do not predict since we do not know the population size, and ii) we cannot predict data for which there is no information to base a model on. This problem, however, also needs to be addressed for wage assessments that do not include the black market segment.

4 Conclusions, implications and suggestions

Research based on new sources of data and innovative research methods related to the spread of the Internet has been on the rise. These trends have also affected possibilities of conducting research on the labour market with respect to both its supply and demand side. This paper has critically reviewed the existing literature to summarise the various ways in which researchers have been dealing with methodological issues related to web-based sources of new data and specifically the key problem of data representativeness and generalisability of their findings. We also suggest statistical methods for

dealing with missing data as a tool to estimate population parameters in the most accurate way, when the samples concerned are likely not to be random, such as the case of online job vacancies.

We find the current debate to be flawed by its failure to acknowledge that every exercise in data collection — including the census — has its limitations. Interestingly, in most studies analysing printed or online job advertisements, representativeness issues are not widely discussed and the findings and conclusions are presented in a generalising manner. We propose that rather than dismissing out of hand research efforts using online job advertisement and other types of web-based data due to weaknesses of data representativeness, a debate should be launched on how these weaknesses can be compensated for and for which types of research questions and fields they might be of a smaller concern.

We highlight that adjusting for representativeness is a particularly formidable task with respect to online job vacancy data. This is due to the fact that the population of job vacancies and its structure is practically unknown. For this reason, adjustment methods, such as weighting, that have been tested as a means of improving web-based voluntary surveys cannot be used as a technique for adjusting online job vacancy data.

Based on the review and synthesis of existing studies and broader statistical approaches to sample correction, we would also like to offer more general recommendations with respect to the usage of online job vacancies for future research. First, the representativeness and reliability of the data source used need to be evaluated at the country level. Dominant market share can be considered as an adequate source of data that can lead to reliable and transferable research results. Second, representativeness and reliability need to be assessed vis-à-vis a particular research focus. The use of a data segment or sub-sample that can be considered (more) representative can address certain aspects of coverage and sampling errors. Examples we encountered focused on professions or labour market segments that are highly exposed to the Internet (web-designers, graduate labour market), where these biases are expected to be less pronounced. Third, depending on the particular research focus, online job ads could be coupled with other sources of vacancy data or text describing analysed professions. Fourth, more sophisticated statistical methods anchored in the literature on missing data, such as the model-based approach to the correction of data not missing at random, could also be used to adjust biases stemming from the structure of online vacancy data.

The good news for this methodological debate is that Internet-based job searching is likely to become an increasingly more prominent tool for job matching, which promises to improve the coverage of the population of workers and firms that engage in it. Recent studies evaluating the quality of online job searching and matching already find a positive impact (European Commission and ECORYS 2012; Mang 2012; Kuhn 2014). Compared to traditional employment channels (newspapers, friends and agencies), online job portals are able to provide a wider range of choice as well as increasingly more advanced tools to evaluate the suitability of a job or a job candidate. An especially positive impact was found for workers with interruptions in their employment history, distancing them from the labour market, such as mothers. This has important implications with respect to attempts to re-activate disadvantaged groups in various labour markets. Labour market policy can be enriched by a better understanding of what employers need, which in turn can be used to inform job counselling services,

second-chance education and training, as well as integration of disadvantaged jobseekers in the labour market (Keep and James 2010; Kureková et al. 2015). Other public policy areas could benefit from the usage of web-based data about labour markets. Examples include social policy, in particular aspects of labour market discrimination, or the education and training sector which could incorporate online job vacancy information into curricula development. The motivation to pursue innovative sources of data and analytical methods and to improve the reliability of the results is not only academic, but also driven by the practical benefits they may yield.

Endnotes

¹A representative sample is one that accurately reflects the characteristics of the underlying population.

²Wageindicator project collects data about wages and working conditions through web-based platforms in a range of countries across the world. For more see: <http://www.wageindicator.org/>.

³Created in 1993, EURES is a European ‘job mobility’ portal operated jointly by the European Commission and the Public Employment Services of the EEA Member States. Its purpose is to provide information, advice and recruitment/placement (job-matching) services for the benefit of workers and employers as well as any citizen wishing to benefit from the principle of the free movement of persons.

⁴<http://webdatanet.cbs.dk/>

⁵<http://www.websm.org/c/85/Bibliography/?preid=0>

⁶Probabilistic web-based surveys are also conducted where a proper sampling frame exists, which allows for drawing a probability-based random sample from a population in which every individual has the same probability of being selected. Examples include email requests, mixed-mode surveys or pre-recruited access panels of Internet users (Steinmetz et al. 2009).

⁷Throughout the following text we assume that we can process all online job vacancies available, but this is obviously not the case for large countries. For simplicity purposes we do not consider this case however in this theoretical study.

⁸A finite population is one for which it is possible to count its individuals or units.

⁹Denmark is an example of a country in which the public authorities are legally obliged to report all job vacancies online.

¹⁰This technique is in a way similar to stratified sampling, in which random samples are drawn from different categories that are pre-defined by the researcher. Eurostat (2010) presents this methodology in the context of job vacancies.

¹¹With the rise of behavioural economics, data from experiments are also gaining in importance. But since this is a new field and we are interested in established methods to study sample design issues, we will not further review data gathered through experiments.

¹²We are thankful to Mr. Nicholas Sofroniou, Expert at CEDEFOP, Thessaloniki, for this idea.

¹³Little and Rubin (1987) also add procedures based on completely recorded units (analysing only observations with complete data), as well as design-based weighted estimation and imputation-based procedures (the missing values are filled in) to their taxonomy of methods with partially missing data.

¹⁴A probability model is a mathematical representation of the probability of the occurrence of an event, where an event can be the realisation of a random variable.

¹⁵Bayesian methods can also be used with a frequentist view; see for instance Carneiro et al. (2003).

Competing interests

The IZA Journal of Labor Economics is committed to the IZA Guiding Principles of Research Integrity. The authors declare that they have observed these principles.

Author's information

Lucia Mýtna Kureková

Slovak Governance Institute (SGI), Central European University (CEU), Institute for the Study of Labour (IZA) and Central European Labour Studies Institute (CELSI)

Miroslav Beblavý

Center for European Policy Studies (CEPS)

Anna Thum-Thyssen

Center for European Policy Studies (CEPS) and European Commission (EC)

Acknowledgements

This research was funded by the European Commission (project name NEUJOBS) through the Seventh Framework Programme (FP7), grant Agreement No. 266833. The authors wish to thank for highly valuable comments to participants of the Skillsnet Methodological Conference at CEDEFOP in Thessaloniki in October 2013, an anonymous referee and in particular to Nick Sofroniou. All errors remain our own.

Responsible editor: Klaus F Zimmermann.

Funding

The research was funded by the European Commission (project name NEUJOBS) through the Seventh Framework Programme (FP7), grant Agreement No. 266833.

Author details

¹Slovak Governance Institute, Bratislava, Central European University, Budapest, IZA, Bonn and CELSI, SGI, Štúrova 3, 811 03, Bratislava, Slovakia. ²CEPS, 1 Place du Congrès, 1000 Brussels, Belgium. ³Centre for European Policy Studies (CEPS) and European Commission (EC), CEPS: 1 Place du Congrès, 1000 Brussels, Belgium.

Received: 31 March 2015 Accepted: 8 August 2015

Published online: 25 September 2015

References

- Ackers D (2012) The Experience of EURES. Improving Access to Labour Market Information for Migrants and Employers. Paper presented at High Level Conference, European Commission DG Employment Social Affairs and Inclusion Unit C4, 6 November 2012. www.labourmigration.eu/events/document/163?format=raw
- Aitkin M, Aitkin I (2011) Statistical Modeling of the National Assessment of Educational Progress. Springer, New York Dordrecht Heidelberg London
- Amuri FD, Marcucci J (2010) "Google it!" Forecasting the US unemployment rate with a Google job search index. Fondazione Eni Enrico Mattei: Global challenges, Nota di lavoro
- Anker R (2011) Estimating a living wage: A methodological review. ILO Conditions of Work and Employment Series 29
- Askitas N, Zimmermann K (2015) The internet as a data source for advancement in social sciences. Int J of Manpow 36(1):2–12
- Askitas N (2009) Zimmermann K (2009) Google econometrics and unemployment forecasting. App Econ Quart 55(2):107–120
- Backhaus KB (2004) An Exploration of Corporate Recruitment Descriptions on Monster.com. J Bus Commun 41:115–136. doi:10.1177/0021943603259585
- Barnichon R (2010) Building a composite help-wanted index. Econ Lett 109:175–178
- Beblavý M, Kureková LM, Háta C (2015) The surprisingly exclusive nature of medium- and low-skilled jobs: Evidence from a Slovak job portal. Personnel Review (in press)
- Bellou A (2015) The impact of Internet diffusion on marriage rates: Evidence from the Broadband Market. J Popul Econ 28:265–297. doi:10.1007/s00148-014-0527-7
- Besamusca J, Tijdens K (2015) Comparing collective bargaining agreements for developing countries. Int J Manpow 36:86–102. doi:10.1108/IJM-12-2014-0262
- Capiluppi A, Baravalle A (2010) Matching Demand and Offer in On-line Provision: a Longitudinal Study of Monster.com. In: WSE 2010 Proceedings the 12th IEEE International Symposium on Web Systems Evolution (WSE 2010), Timisoara, 17–18 September 2010. <http://roar.uel.ac.uk/995/>
- Carneiro PM, Cunha F, Heckman JJ (2003) Interpreting the Evidence of Family Influence on Child Development. In: The Economics of Early Childhood Development: Lessons. The Federal Reserve Bank of Minneapolis, Minneapolis
- Chambers RL, Skinner CJ (2003) Analysis of Survey Data. John Wiley & Sons, New York. <https://books.google.be/books?isbn=0470864397>
- Colombo E (2009) Measuring Skill Needs through Employers' Surveys: Problems and Methods - Colombo Presentation.pdf. Paper presented at the Agora conference, CEDEFOP, Thessaloniki, 11 June, 2009
- Couper MP (2000) Review: Web surveys: A review of issues and approaches. Public Opin Q 64(4):464–494
- de Bustillo RM, de Pedraza P (2010) Determinants of job insecurity in five European countries. Eur J Ind Relat 16:5–20

- DeKay SH (2013) Peering Through Glassdoor.com: What Social Media Can Tell Us About Employee Satisfaction and In: Genest CM (ed) Conference on Corporate Communication 2013: Abstracts of Conference Proceedings. Conference on Corporate Communication, Baruch College/CUNY, New York, 4–7 June 2013. http://www.corporatecomm.org/pdf/Abstracts_Proceedings_CCI_CCC_2013.pdf#page=60
- de Leeuw ED (2005) To mix or not to mix data collection modes in surveys. *J Off Stat* 21(2):233–255
- de Leeuw ED (2013) Mixed-mode surveys and the Internet. *Survey Practice* 3:6
- de Pedraza P, Tijdens K, de Bustillo RM, Steinmetz S (2010) A Spanish Continuous Volunteer Web Survey: Sample Bias, Weighting and Efficiency/Una encuesta voluntaria y continua en la red en España: sesgo, ponderación y eficiencia. *Reis* 131:109–130.
- Dillman DA, Smyth JD, Melani L (2014) Internet, Phone, Mail, and Mixed-Mode Surveys: The Tailored Design Method, 4th edn. Wiley, New York
- Dörfler L, van de Werfhorst HG (2009) Employers' demand for qualifications and skills. *Eur Soc* 11:697–721. doi:10.1080/14616690802474374
- Edelman B (2012) Using Internet Data for Economic Research. *J Econ Perspect* 26:189–206. doi:10.2307/41495310
- European Commission, ECORYS (2012) European Vacancy and Recruitment Report 2012. Publications Office of the EU, Luxembourg
- Eurostat (2010) 1st and 2nd International Workshops on Methodologies for Job Vacancy Statistics - Proceedings. Publications Office of the EU, Luxembourg
- Gosling SD, Vazire S, Srivastava S, John OP (2004) Should we trust web-based studies. *Am Psychol* 59:93–104
- Guzi M, de Pedraza GP (2015) A web survey analysis of subjective well-being. *Int J Manpow* 36:48–67
- Guzi M, Kahanec M (2014) Wagelindicator Living Wages Methodological Note. WageIndicator Foundation, Amsterdam
- Huang H, Kvasny L, Joshi KD et al (2009) Synthesizing IT job skills identified in academic studies, practitioner publications and job ads. In: Proceedings of the special interest group on management information system's 47th annual conference on Computer personnel research, Limerick, Ireland, May 28–30
- Jackson M (2001) Non-Meritocratic Job Requirements and the Reproduction of Class Inequality: An Investigation. *Work Employ Soc* 15:619–630. doi:10.1177/09500170122119020
- Jackson M (2007) How far merit selection? Social stratification and the labour market. *Br J Sociol* 58:367–390. doi:10.1111/j.1468-4446.2007.00156.x
- Jackson M, Goldthorpe J, Mills C (2005) Education, Employers and Class Mobility. *Res Soc Stratif Mobil* 23:3–33. doi:10.1016/S0276-5624(05)23001-9
- Keep E, James S (2010) Recruitment and Selection – the Great Neglected Topic. SKOPE Research Paper No. 88.
- Kennan MA, Cole F, Willard P et al (2006) Changing workplace demands: what job ads tell us. *Aslib Proc* 58:179–196. doi:10.1108/00012530610677228
- Kuhn PJ (2014) The internet as a labor market matchmaker. In: IZA World of Labour: 18
- Kuhn P, Mansour H (2014) Is Internet Job Search Still Ineffective? *Econ J* 124:1213–1233
- Kuhn P, Shen K (2013) Gender Discrimination in Job Ads: Evidence from China. *Q J Econ* 128:287–336. doi:10.1093/qje/qjs046
- Kureková LM, Beblavý M, Haita C, Thum A-E (2015) Employers' skill preferences across Europe: between cognitive and non-cognitive skills. *J Educ Work* 0:1–26. doi:10.1080/13639080.2015.1024641
- Kureková LM, Zilinčíková Z (2015) Low-Skilled Jobs and Student Jobs: Employers' Preferences in Slovakia and the Czech Republic. IZA DP No. 9145.
- Little RJ, Rubin DB (1987) Statistical analysis with missing data. John Wiley & Sons, New York
- Little RJ, Rubin DB (2002) Statistical analysis with missing data, 2nd edn. John Wiley & Sons, New York
- Longford NT (2005) Missing data and small-area estimation: Modern analytical equipment for the survey statistician. Springer, New York Dordrecht Heidelberg London
- Mang C (2012) Online Job Search and Matching Quality. Ifo Working Paper No. 147
- Martikainen J (2010) Weighting and estimation methods: JVS estimation in Finland by Horowitz-Thomson-Type estimator. In: Eurostat (ed). Publications Office of the EU, Luxembourg
- Martínez-Torres MR, Toral SL, Fornara N (2014) Big data and virtual communities: methodological issues. University of Amsterdam, Amsterdam
- Masso J, Eamets R, Mötsmees P (2014) Temporary migrants and occupational mobility: evidence from the case of Estonia. *Int J Manpow* 35:753–775. doi:10.1108/IJM-06-2013-0138
- Maurer-Fazio M (2012) Ethnic discrimination in China's internet job board labor market. *IZA J Migr* 1:1–24. doi:10.1186/2193-9039-1-12
- Maurer-Fazio M, Lei L (2015) "As rare as a panda": How facial attractiveness, gender, and occupation affect interview callbacks at Chinese firms. *Int J Manpow* 36:68–85. doi:10.1108/IJM-12-2014-0258
- Reips U-D (2006) Web-based methods. In: Eid M, Diener E (eds) Handbook of multimethod measurement in psychology. American Psychological Association, Washington DC, pp 73–85
- Reips U-D (2012) Using the Internet to collect data. In: Cooper HE, Camic PM, Long DL et al (eds) APA handbook of research methods in psychology, Vol 2: Research designs: Quantitative, qualitative, neuropsychological, and biological, pp 1–20
- Reips U-D, Buffardi LE (2012) Studying migrants with the help of the Internet: methods from psychology. *J Ethn Migr Stud* 38:1405–1424
- Royall RM (1992) The model based (prediction) approach to finite population sampling theory. In: Ghosh M, Pathak PK (eds) Current issues in statistical inference: essays in honor of D. Basu. Institute of Mathematical Statistics, Hayward, CA
- Sappler N (2013) Advancing Research Methods with New Technologies, 1st edn. IGI Global, Hershey, PA
- Schonlau M, Van Soest A, Kapteyn A, Couper M (2009) Selection bias in web surveys and the use of propensity scores. *Sociol Methods Res* 37:291–318
- Shen K, Kuhn P (2013) Do Chinese employers avoid hiring overqualified workers? Evidence from an internet job board. In: Giulietti C, Tatsiramos K, Zimmermann K (eds) Labor Market Issues in China. Emerald Group Publishing, Bingley, pp 1–30
- Štefánik M (2012) Internet job search data as a possible source of information on skills demand (with results for Slovak university graduates). In: CEDEFOP (ed) Building on skills forecasts — Comparing methods and applications. Publications Office of the European Union, Luxembourg

- Steinmetz S, Raess D, Tijdens K, de Pedraza P (2013) Chapter 6: Measuring Wages Worldwide: Exploring the Potentials and Constraints of Volunteer Web Surveys. In: Sappleton N (ed) Advancing Research Methods with New Technologies, 1st edn. IGI Global, Hershey, PA
- Steinmetz S, Tijdens K (2009) Can weighting improve the representativeness of volunteer online panels? Insights from the German Wage indicator data. CM News 5:7–11
- Steinmetz S, Tijdens K, Pedraza P (2009) Comparing different weighting procedures for volunteer web surveys. In: AIAS Working Paper 09–76
- Stern MJ, Bilgen I, Dillman DA (2014) The State of Survey Methodology Challenges, Dilemmas, and New Frontiers in the Era of the Tailored Design. Field Methods 26:284–301. doi:10.1177/1525822X13519561
- Taylor L, Schroeder R, Meyer E (2014) Emerging practices and perspectives on Big Data analysis in economics: Bigger and better or more of the same? Big Data Soc 1:2053951714536877. doi:10.1177/2053951714536877
- Teichler U (2009) Higher education and the world of work. Sense Publishers, Rotterdam
- Tijdens K, Beblavý M, Thum-Thysen A (2015) Do educational requirements in vacancies match the educational attainments of job holders? An analysis of web-based data for 279 occupations in the Czech Republic. In: GRID Working Paper
- Valliant R, Dorfman AH, Royall RM (2000) Finite population sampling and inference: a prediction approach. John Wiley & Sons, New York
- Van Ours J, Ridder G (1992) Vacancies and the recruitment of new employees. J Labor Econ 10(2):138–155
- Vicente MR, López-Menéndez AJ, Pérez R (2015) Forecasting unemployment with internet search data: Does it help to improve predictions when job destruction is skyrocketing? Technol Forecast Soc Change 92:132–139. doi:10.1016/j.techfore.2014.12.005
- Wade MR, Parent M (2001) Relationships between Job Skills and Performance: A Study of Webmasters. J Manag Inf Syst 18:71–96. doi:10.2307/40398554
- Young KS, Case CJ (2004) Internet abuse in the workplace: new trends in risk management. Cyberpsychol Behav 7:105–111

Submit your manuscript to a SpringerOpen® journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com