

An Imputation Model for Dropouts in Unemployment Data

Petra Nilsson¹

Incomplete unemployment data is a fundamental problem when evaluating labour market policies in several countries. Many unemployment spells end for unknown reasons; in the Swedish Public Employment Service's register as many as 20 percent. This leads to an ambiguity regarding destination states (employment, unemployment, retired, etc.). According to complete combined administrative data, the employment rate among dropouts was close to 50 for the years 1992 to 2006, but from 2007 the employment rate has dropped to 40 or less. This article explores an imputation approach. We investigate imputation models estimated both on survey data from 2005/2006 and on complete combined administrative data from 2005/2006 and 2011/2012. The models are evaluated in terms of their ability to make correct predictions. The models have relatively high predictive power.

Key words: Follow-up study; multiple regression imputation; register data.

1. Introduction

Active labour market policies (ALMPs) have increasingly been promoted in the Organisation for Economic Co-operation and Development (OECD) countries and transition economies as a principal means of dealing with unemployment. Evaluation of ALMPs is important for future policy making and implementation, but incomplete information in unemployment data is a fundamental problem. Many unemployment spells end for unknown reasons, which leads to an ambiguity regarding the labour market state. In the Swedish Public Employment Service's register, the percentage of exits for unknown reasons is approximately 20 percent. When evaluating ALMPs, assumptions have to be made concerning whether these unemployment spells ended because of work or not. In the agency's performance reports only known exits to employment are presented, which means that the number of exits to employment is underestimated.

Similar data problems exist in several countries, and these countries have employed different methods of dealing with incomplete information in unemployment administrative data. For example, Wilke (2009) constructs bounds for unemployment duration in UK administrative unemployment data to describe the effect of missing information on interval information (length of unemployment) and destination states (employment, unemployment, retired, etc.). Arntz et al. (2007) conduct a similar exercise with German

¹ Research and Evaluation, Swedish Public Employment Service, SE-113 99 Stockholm, Email: petra.nilsson@arbetsformedlingen.se

Acknowledgments: The author thanks Anders Harkman, who came up with the idea and helped along the way, and Linda Wänström for her valuable contribution.

data. Our interest is in the destination states, and we choose to explore an imputation approach rather than constructing bounds.

Bring and Carling (2000) use data from a small survey in 1994 to estimate an imputation model that Statistics Sweden has used to compensate for missing destination states since 1994. This article extends the Bring and Carling (2000) methodology and uses more recent data. We estimate new imputation models based on both survey and register data. We use data from a larger survey conducted in 2005 and 2006 by the Swedish Public Employment Service. We also use data from the Swedish Longitudinal Integration Database for Health Insurance and Labour Market Studies (LISA), which includes information about gainful employment as of November each year. The imputation models are evaluated and the concordance between predicted values and survey/register data is studied. The predicted power of the new models is compared to the predicted power of Bring and Carling's model as well as to random imputation.

We also present the employment rate among dropouts over time based on LISA data. Another contribution is that the new imputation models based on survey data deal with nonresponse. Bring and Carling (2000) used only survey responses in their model and did not account for survey nonresponse. The nonresponse rate in the survey used in Bring and Carling (2000) is 20 percent. This refers both to unit and item nonresponse, since there was only one question in the survey.

2. Data

Survey data and register data are used to estimate imputation models for dropouts in unemployment data. Both survey data and register data contain measurement error. In register data there is unobserved/misreported information, since not all individual employment biographies are covered by the administrative process. Bound et al. (2001) discuss the causes of measurement errors in survey reports. The longer the recall period, the more difficult the reporting task and the less salient the event the more difficult it is to retrieve the information requested. Socially undesirable events tend to go unreported, while the opposite is true for socially desirable events. See for example Pyy-Martikainen and Rendtel (2009), who used Finnish linked survey and administrative data to analyse measurement error in survey data.

2.1. The Survey

The survey was conducted on twelve different measurement occasions between September 2005 and August 2006. Each measurement occasion includes exits for unknown reasons during one week, where the job seeker does not return to the Public Employment Service within 14 days. An unrestricted random selection of 300 periods of registration were made each measurement week. The total sample is therefore 3,600 periods of registration. We chose to include measurement weeks from as many different periods as possible during one year to take into account any seasonal effects.

The survey can be seen as a stratified sample with measurement weeks as strata. A stratification of a finite population $U = \{1, \dots, k, \dots, N\}$ means a partition of U in H subsets of the population (Lundström and Särndal 2001). The number of elements in stratum h is denoted N_h and the sample size in stratum h is denoted n_h . The probability that

a given element is included in the sampling, the inclusion probability, is given by

$$\pi_k = \frac{n_h}{N_h}. \quad (1)$$

Let

$$d_k = \frac{1}{\pi_k} \quad (2)$$

denote the design weight of element k .

Note that it is the period of registration and not the individual that constitutes an element in the survey. The population consists of unique periods of registration, but not of unique individuals, as some people occur in the data set multiple times. The outcome of registration periods for the same individual are probably correlated and the observations cannot be assumed to be independent. This problem is called correlated failure-time modelling or multiple spells modelling and is studied for example in the economic literature (Lancaster 1979; Heckman and Singer 1982).

The correlation structure is ignored in this article, which might lead to an underestimation of the variability of the imputation model. A very large percentage of registration periods concern unique individuals, however. In fact, 97.4 percent of registration periods that ended with deregistration for unknown reasons during the measurement weeks concern unique persons. Of the periods that constitute the sampling frame in the survey, periods where the individuals have not returned within 14 days, 98.5 percent concern unique individuals. In the sample, 99.8 percent concern unique individuals; three individuals are found twice.

The survey was conducted in the form of computer-aided telephone interviews by the Public Employment Service's interview unit. The interviews were conducted as close to deregistration from the Public Employment Service as possible, in order for the interviewed persons to be able to recall their work situation when they ceased to have contact with the Public Employment Service. Since there is a 14-day wait in order to exclude the return of job seekers to the Public Employment Service, the interviews were conducted within two to three weeks of deregistration.

In the survey, the individuals were asked about their current work situation ("What is your work situation today?"). The response options were the following:

1. Have work (full-time)
2. Have work (part-time)
3. Studying/in training
4. Participating in a labour market programme
5. Have started my own company
6. Long-term sick leave/sick leave/on parental leave
7. Unemployed/seeking work
8. Other

Individuals responding according to option 1, 2, or 5 are defined as having found work. The interviewers did not read out the response options to the question. When the interviewed person had difficulties giving an answer that fitted the response options, the interviewer helped by interpreting the purpose of the question.

Out of the total sample of 3,600 periods, 2,443 responded, giving a response rate of 68 percent. Nonresponse in the survey is thus 32 percent. Older persons, persons born outside Europe, persons with a low level of education, and persons without unemployment insurance are overrepresented in the nonresponse group. Probably, a lower extent of these persons had found work than those who responded in the survey (see, for example, [Bennmarker et al. 2007](#)).

There is a risk that estimates using data only from respondents will be biased. We therefore impute missing values. There are various methods of imputation; we use regression imputation as described in [Lundström and Särndal \(2001\)](#). Since both socioeconomic and employment-related explanatory variables can be linked to the individuals that drop out, missing data may be imputed using a logistic regression model that explains which categories of individuals have the greatest probability of having found work. [Rubin \(1996\)](#) recommends that an imputation model contain as many relevant variables as possible and the model used to impute the nonresponse in the survey includes many socioeconomic and employment-related explanatory variables; see [Appendix](#). Variables in [Bennmarker et al. \(2007\)](#) were considered.

We denote by y_k whether or not a period of registration k has ended because of work; $y_k = 1$ if period k ended because of work and $y_k = 0$ if period k did not end because of work. We assume that response or register values are obtained for the elements in a set denoted r . Regression imputation gives an imputed value for element k according to

$$\hat{y}_k = z_k' \hat{\beta} \quad (3)$$

where z_k is the value of the imputation vector for element k by $z_k = (z_{1k}, \dots, z_{jk}, \dots, z_{Jk})'$, a column vector with J explanatory variables, where z_{jk} is the value, for element k , of the j th explanatory variable and

$$\hat{\beta} = \left(\sum_r d_k z_k z_k' \right)^{-1} \sum_r d_k z_k y_k. \quad (4)$$

$\hat{\beta}$ is a vector of regression coefficients, resulting from the fit of a regression using the data (y_k, z_k) available for $k \in r$ and weighted with d_k .

Replicates of data containing regression-imputed values tend to have a lower degree of variance than data containing observed values y_k . We therefore add a randomly selected residual. Then the imputed value for element k is

$$\hat{y}_k = z_k' \hat{\beta} + e_k^* \quad (5)$$

where e_k^* is a randomly selected residual from the data set containing calculated residuals $\{e_k : k \in r\}$, where

$$e_k = y_k - z_k' \hat{\beta}. \quad (6)$$

Each missing value is imputed 20 times. The y_k values for the respondents are the same in all data sets, while the imputed values are different.

2.2. Register Data

In addition to survey data, we also use administrative records on gainful employment for this study when constructing imputation models. The LISA database is used to impute missing information about employment status to obtain complete combined administrative data sets. The administrative records in LISA are limited to the month of November so the combined administrative data sets are also limited to November. Information about gainful employment in LISA is used to impute missing destination states for dropouts in November each year.

The LISA database holds annual registers and includes all individuals 16 years of age and older registered in Sweden. It is available in spring the following year. The individuals are classified as employed if they are assumed to have worked for at least four hours during November. The estimation in LISA is model based where the correlation between several variables, for example information about payments from employers, is used for the classification. There are some misclassifications in the data compared to real working hours in November. The risk of misclassification is larger for persons who were working only parts of the year and for persons with a weaker connection to the labour market. Misclassifications are partly due to errors in the model but also due to incomplete information.

Imputation models for dropouts are estimated on complete combined administrative data from November 2005/2006 and from November 2011/2012. The years 2005/2006 are chosen to enable a comparison between a model based on administrative data and a model based on survey data for the same years. An imputation model is also estimated on administrative data from 2011/2012 to enable a comparison with later years.

Table 1 describes the combined administrative data sets for the years 2005/2006 and 2011/2012. Item nonresponse regarding employment status not filled in by LISA for 2005/2006 is 1.5 percent and for 2011/2012 it is two percent. Item nonresponse is more common for persons 55 years or older, for persons born outside of Europe, and for persons with a low level of education.

For the employment rate among dropouts over time we use all available data, which is November data for the years 1992 to 2012. For the years 1992 to 2006, the number of periods each November is approximately 12,000, while for the years 2007 to 2012 the number of periods each November has dropped to approximately 7,000 per year. The sharp decline from 12,000 to 7,000 is due to better administrative routines at the Swedish Public Employment Service. Fewer ended unemployment spells lack employment status.

Table 1. Description of register data November 2005/2006 and November 2011/2012 respectively.

	2005/2006 Complete data	Item nonresponse	2011/2012 Complete data	Item nonresponse
Number of observations	20,566	311	14,375	282
16–24 years (%)	45.7	15.8	43.8	20.9
55–66 years (%)	3.6	7.4	4.6	8.2
Born outside Europe (%)	17.2	24.4	29.5	44.0
Functional impairment (%)	3.6	2.9	6.6	3.6
Compulsory school (%)	26.2	52.8	29.5	44.3

Item nonresponse regarding employment in November not filled in by LISA is about one percent in the beginning of the period but close to two percent for later years.

3. The Employment Rate

Figure 1 shows the employment rate among dropouts in November each year according to combined administrative data. For the years 1992 to 2006 the employment rate is close to 50, but from 2007 the employment rate drops to 40 or less. The decline in the employment rate is probably due to the better administrative routines mentioned above.

The estimated employment rate is based on the 20 nonresponse imputed replicates of data, that is, 20 separate models are estimated. The different parameter estimates are then combined as described in Rubin (1987).

Suppose that \hat{Q}_i is an estimate of a scalar quantity of interest, obtained from a data set i , $i = 1, 2, \dots, m$ and \hat{W}_i is the variance associated with \hat{Q}_i . The overall estimate is the average of the individual estimates from the m complete replicates of data

$$\bar{Q} = \frac{1}{m} \sum_{i=1}^m \hat{Q}_i. \quad (7)$$

Assume that \bar{W} is the within-imputation variance, which is the mean value of the estimates from the m complete replicates of data

$$\bar{W} = \frac{1}{m} \sum_{i=1}^m \hat{W}_i \quad (8)$$

and B is the between-imputation variance

$$B = \frac{1}{m-1} \sum_{i=1}^m (\hat{Q}_i - \bar{Q})^2 \quad (9)$$

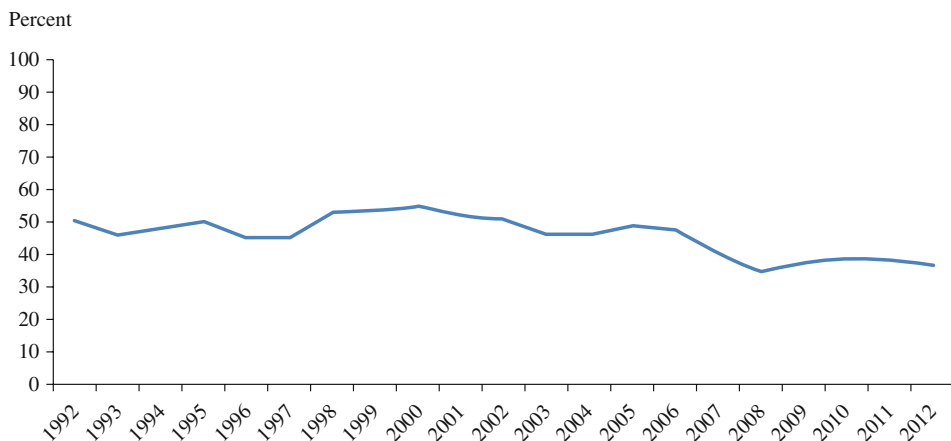


Fig. 1. Employment rate among dropouts in November 1992 to November 2012 according to combined administrative data.

Table 2. The estimated employment rate among dropouts based on the 2005/2006 survey.

	Estimated percentage	Stand. errors	95% Confidence Interval
Based both on the response group and the nonresponse group	47.3	1.1	(45.1; 49.4)
Based on the respondents	50.7	1.0	(48.6; 52.7)
Based on the nonresponse group (imputed values)	39.9	2.6	(34.8; 45.1)

then the total variance is

$$T = \bar{W} + (1 + \frac{1}{m})B. \quad (10)$$

[Table 2](#) shows the estimated employment rate among dropouts based on the survey from 2005/2006.

The estimated employment rate is 47.3 percent. A 95% confidence interval for this percentage is of the magnitude plus/minus two percentage points. The estimated employment rate based on the survey is close to the employment rate according to the combined administrative data for November 2005 and November 2006, which is 48.1 percent. [Table 2](#) also shows the estimated employment rate based only on those who responded to the survey and based only on the imputed values for the nonresponse group.

The estimated employment rate among dropouts in the 1994 survey is 44.7 percent ([Bring and Carling 2000](#)). The sample in the 1994 survey was drawn from the population of dropouts in January and February. In the combined administrative data for November 1994, the employment rate is 47.6 percent.

A factor that affects the comparability of survey data and the combined administrative data is that administrative data refers to the month of November while survey data refers to different periods during the year. The employment rate can be different depending on the season. [Table 3](#) displays the estimated employment rate per measurement week in the 2005/2006 survey. Standard errors and 95% confidence intervals are also shown.

Table 3. The estimated employment rate among dropouts in the 2005/2006 survey divided into measurement week.

Measurement week	Estimated employment rate	Stand. errors	95% Confidence Interval
Week 34 2005 (Aug)	44.3	3.1	(38.1; 50.5)
Week 35 2005 (Aug-Sep)	44.8	3.1	(38.7; 50.9)
Week 36 2005 (Sep)	43.0	3.1	(36.8; 49.1)
Week 37 2005 (Sep)	47.1	3.4	(40.3; 53.9)
Week 40 2005 (Oct)	46.6	3.4	(39.9; 53.4)
Week 3 2006 (Jan)	43.9	3.3	(37.4; 50.5)
Week 5 2006 (Jan-Feb)	44.8	3.3	(38.4; 51.3)
Week 9 2006 (Feb-Mar)	48.6	3.3	(42.1; 55.0)
Week 14 2006 (Apr)	52.4	3.2	(46.1; 58.8)
Week 22 2006 (May-Jun)	56.3	3.3	(49.9; 62.8)
Week 25 2006 (Jun)	54.5	3.6	(47.3; 61.7)
Week 31 2006 (Jul-Aug)	45.6	3.2	(39.2; 51.9)

The estimated employment rate varies from 43.0 to 56.3 between measurement weeks. The employment rate is lower at the beginning of the autumn and spring semester when the exit rate to education is high, and the employment rate is higher at the beginning of the summer when the exit rate to summer jobs is high.

4. Imputation Models for Dropouts

4.1. The Models

Imputation models for dropouts are estimated on survey data from 2005/2006 and on complete combined administrative November data from 2005/2006 and 2011/2012. The models are logistic regression models and the dependent variable is whether the person is employed or not. We estimate

$$\hat{P}(Y_k = 1|z_k) = \frac{1}{1 + \exp(-z_k\hat{B}')} \quad (11)$$

Age, country of birth, functional impairment, education, membership of an unemployment insurance fund, status prior to deregistration, and experience are used as explanatory variables. We want to include as many relevant socioeconomic and employment-related explanatory variables as possible to improve the models' predictive power, but at the same time we want to keep the models as simple as possible to use. Variables used in [Bring and Carling \(2000\)](#) and [Benmarker et al. \(2007\)](#) have been considered and variables with p -values smaller than 0.05 when estimated on administrative data are used in the final models.

Since survey data includes samples from different periods during a year, it is possible to include month of deregistration as an explanatory variable. Two different models are estimated on survey data; one with (Model 1) and one without (Model 2) month of deregistration.

For survey data we estimate the imputation models on the 20 nonresponse imputed replicates of data, that is, 20 separate models are estimated. The different parameter estimates are then combined as described in [Rubin \(1987\)](#). [Table 4](#) displays the imputation models based on survey data.

The employment rate is lower for, for example, older persons, persons with a low level of education, persons born outside of Europe, and persons with a functional impairment. Persons being registered as part-time or temporarily employed prior to deregistration have a higher employment rate than persons categorised as unemployed. Persons with many previous transitions to work and members of an unemployment insurance fund also have a higher employment rate. The alternative model (Model 2), where month of deregistration is included, shows that those deregistered in May, June, or July are employed to a higher extent.

[Table 5](#) shows imputation models estimated on complete combined administrative data for November 2005/2006 (Model 3) and November 2011/2012 (Model 4).

The interpretation of the estimates is basically the same as for [Table 4](#). One difference is that the estimate for the intercept and the estimate for individuals 16–24 years have

Table 4. Imputation models based on survey data 2005/2006, eleven and twelve covariates respectively.

	Model 1			Model 2		
	11 variab. Estimate	Stand. errors	p-value	12 variab. Estimate	Stand. errors	p-value
Intercept	0.14	0.14	0.33	0.07	0.14	0.61
16–24 years	–0.37	0.11	0.00	–0.37	0.11	0.00
55–66 years	–0.36	0.18	0.05	–0.37	0.18	0.05
Born outside Europe	–0.53	0.12	<.0001	–0.53	0.12	<.0001
Functional impairment	–0.37	0.31	0.23	–0.38	0.31	0.22
Compulsory school only	–0.30	0.12	0.02	–0.31	0.13	0.02
Member of an unemployment insurance fund	0.52	0.11	<.0001	0.52	0.11	<.0001
Work prior to deregistration	0.84	0.14	<.0001	0.85	0.14	<.0001
Other status prior to deregistration	–0.48	0.13	0.00	–0.50	0.13	<.0001
Number of periods of registration	–0.16	0.03	<.0001	–0.16	0.03	<.0001
Number of transitions into work	0.31	0.06	<.0001	0.31	0.06	<.0001
No experience	–0.15	0.12	0.21	–0.15	0.12	0.21
Deregistration in May, June or July				0.50	0.12	<.0001

Table 5. Imputation models based on complete combined administrative November data 2005/2006 and 2011/2012.

	Model 3			Model 4		
	2005/2006 Estimate	Stand. errors	p-value	2011/2012 Estimate	Stand. errors	p-value
Intercept	-0.50	0.05	<.0001	-0.67	0.06	<.0001
16-24 years	0.22	0.04	<.0001	0.42	0.05	<.0001
55-66 years	-0.43	0.09	<.0001	-0.48	0.10	<.0001
Born outside Europe	-0.50	0.05	<.0001	-0.20	0.05	<.0001
Functional impairment	-0.52	0.10	<.0001	-0.50	0.10	<.0001
Compulsory school only	-0.54	0.04	<.0001	-0.64	0.05	<.0001
Member of an unemployment insurance fund	1.22	0.04	<.0001	1.26	0.05	<.0001
Work prior to deregistration	1.66	0.05	<.0001	1.86	0.07	<.0001
Other status prior to deregistration	-0.24	0.05	<.0001	-0.45	0.07	<.0001
Number of periods of registration	-0.24	0.01	<.0001	-0.24	0.02	<.0001
Number of transitions into work	0.34	0.02	<.0001	0.40	0.03	<.0001
No experience	-0.41	0.04	<.0001	-0.64	0.06	<.0001

changed sign compared to the models based on survey data. The standard errors are about half of those from the models based on survey data (Table 4).

4.2. The Predictive Power of the Models

The predictive power of the imputation models can tell us whether or not the models can be used to impute missing values for those who leave the Public Employment Service for unknown reasons. By predictive power we mean the percentage of correct predictions.

We compare the predictive power of each model in Table 4 and Table 5 (Model 1–4) for both the survey data from 2005/2006 and the complete combined administrative data from November 2005/2006 and November 2011/2012. We also calculate the predictive power for random imputation and for Bring and Carling’s model based on survey data from 1994. In the cross validation, the imputation model has been estimated on 60 percent of the data and evaluated against the remaining 40 percent.

The imputation models estimate a probability between 0 and 1 that the individuals have found work. Imputation then requires a threshold, that is, at which predicted values the imputed value of having found work, $\hat{y}_k = 1$, or the imputed value of not having found work, $\hat{y}_k = 0$ should be classified. For each model, we select a threshold so that the imputation produces the employment rate observed in the dataset. For survey data the employment rate is 47.3, for the combined administrative data from November 2005/2006 it is 48.1 and for November 2011/2012 it is 37.4.

Table 6 shows the predictive power of each model for the different data sets. For survey data from 2005/2006, random imputation has the lowest predictive power, 50 percent correct predictions. The imputation model based on survey data from 1994 has 54 percent correct predictions. The imputation models investigated in this article have higher predictive power, 68 percent for all models. For the combined administrative data from November 2005/2006, the models based on combined administrative data have higher predictive power than the models based on survey data. The model based on combined administrative data from the same years (2005/2006) has the highest predictive power, 76 percent. For administrative data from 2011/2012, the models based on administrative

Table 6. Percent correct predictions.

	Model Survey 1994	Model 1 Survey 2005/2006 11 variab.	Model 2 Survey 2005/2006 12 variab.	Model 3 Register 2005/2006	Model 4 Register 2011/2012	Model Random
Data: Survey 2005/2006	54	68	68	68	68	50
Data: Combined administrative November data 2005/2006	58	74	74	76	75	50
Data: Combined administrative November data 2011/2012	61	72	72	74	74	53

data again have higher predictive power than the models based on survey data. Both models based on administrative data have the same predictive power, 74 percent.

5. Conclusions and Closing Discussion

Imputation can be used in evaluations using unemployment data as a means of dealing with missing information about destination state after a period of unemployment. It also can be used in the Public Employment Service's performance reports when the number of exits to employment is presented to avoid underestimation.

Two imputation models based on survey data and two models based on combined administrative data were investigated. The four models all have similar predictive power. The models based on administrative data have slightly higher predictive power than the models based on survey data.

The two imputation models using 2005/2006 survey data are based on more data and have a higher predictive power than the imputation model suggested in [Bring and Carling \(2000\)](#), which is estimated on a small 1994 sample. The new imputation models based on survey data from 2005/2006 and multiple imputation deal with nonresponse in a more satisfactory way. According to the new survey, the estimated employment rate among dropouts is 47 percent for 2005 and 2006, which is consistent with administrative November data for the same years.

One difference between survey data and combined administrative data is that administrative data refers to the month of November, while survey data refers to twelve different measurement occasions during 2005 and 2006. We have no information about the predictive power of the investigated imputation models for all dropouts in unemployment data. We therefore cannot say which model is the best. Probably it does not matter a great deal which model is used. One suggestion is to use the imputation model based on the combined administrative data from November 2011/2012, which is the model based on the latest available data.

Appendix

In the model used to impute survey data the explanatory variables are:

- female
- 16–24 years
- 35–44 years
- 45–66 years
- born in the Nordic Countries
- born abroad
- functional impairment
- compulsory school only
- higher education ≤ 2 years
- higher education > 2 years
- experience in professions applied for
- seeking only full-time work
- seeking work beyond commuting distance
- member of an unemployment insurance fund

participating in the activity guarantee
 status work prior to deregistration
 other status prior to deregistration (not work or unemployment)
 forest county
 other counties (not forest or major-city region)
 number of periods of registration the previous five years
 number of transitions into work the previous five years
 Deregistration in the months May to July
 16–24 years and member of an unemployment insurance fund
 35–44 years and member of an unemployment insurance fund
 45–66 years and member of an unemployment insurance fund
 16–24 years and experience
 35–44 years and experience
 45–66 years and experience

6. References

- Arntz, M., S. Lo, and R. Wilke. 2007. "Bounds Analysis of Competing Risks: A Nonparametric Evaluation of the Effect of Unemployment Benefits on Migration in Germany." ZEW - Centre for European Economic Research Discussion Paper No. 07-049. Doi: <http://dx.doi.org/10.2139/ssrn.1010286>.
- Benmarker, H., K. Carling, and A. Forslund. 2007. *Vem blir långtidsarbetslös?* Report 2007:29. Uppsala: Institute for Labour Market Policy Evaluation (IFAU). Available at: <http://www.ifau.se/globalassets/pdf/se/2007/r07-20.pdf> (accessed June 1, 2016).
- Bound, J., C. Brown, and N. Mathiowetz. 2001. "Measurement error in survey data." In *Handbook of Econometrics*, vol. 5, edited by J. Heckman and E. Leamer, 3705–3833. Amsterdam: Elsevier. Available at: <http://www.psc.isr.umich.edu/pubs/pdf/r00-450.pdf> (accessed June 1, 2016).
- Bring, J. and K. Carling. 2000. "Attrition and Misclassification of Drop-outs in the Analysis of Unemployment Duration." *Journal of Official Statistics* 16: 321–330. Available at: <http://www.jos.nu/Articles/abstract.asp?article=164321> (accessed June 1, 2016).
- Heckman, J. and B. Singer. 1982. "Population Heterogeneity in Demographic Models." In *Multidimensional Mathematical Demography*, edited by K. Land and A. Rogers, 567–599. New York: Academic Press. Available at: <http://www.popline.org/node/410098> (accessed June 1, 2016).
- Lancaster, T. 1979. "Econometric Methods for the Duration of Unemployment." *Econometrica* 47: 939–956. Doi: <http://dx.doi.org/10.2307/1914140>.
- Lundström, S. and C-E. Särndal. 2001. *Estimation in the Presence of Nonresponse and Frame Imperfections*. Örebro: SCB-Tryck.
- Pyy-Martikainen, M. and U. Rendtel. 2009. "Measurement Errors in Retrospective Reports of Event Histories. A Validation Study with Finnish Register Data." *Survey Research Methods* 3: 139–155. Doi: <http://dx.doi.org/10.18148/srm/2009.v3i3.2372>.
- Rubin, D.B. 1987. *Multiple Imputation for Nonresponse in Surveys*. New York: John Wiley.

- Rubin, D.B. 1996. "Multiple Imputation After 18+ Years (with discussion)." *Journal of the American Statistical Association* 91: 473–489. Doi: <http://dx.doi.org/10.1080/01621459.1996.10476908>.
- Wilke, R. 2009. "Unemployment Duration in the United Kingdom: An Incomplete Data Approach." Doi: <http://dx.doi.org/10.2139/ssrn.1348019>.

Received January 2015

Revised April 2016

Accepted April 2016