The poor outside the lamplight: on the prevalence of poverty among population groups not included in household surveys

Matěj Bajgar, Petr Janský & Klára Kalíšková

Published online: 30 Nov 2018.

Submit your article to this journal ↗

View related articles ↗

View Crossmark data ↗

Citing articles: 2 View citing articles ↗

Routledge
Taylor & Francis Group

Check for updates

The poor outside the lamplight: on the prevalence of poverty among population groups not included in household surveys

Matěj Bajgar [a,b], Petr Janský [a,c] and Klára Kalíšková [a,d]

aCERGE-EI, Prague, Czech Republic; bOECD, Paris, France; cInstitute of Economic Studies, Faculty of Social Sciences, Charles University, Prague, Czech Republic; dUniversity of Economics, Prague, Czech Republic

ABSTRACT
Poverty rate calculations are often based on household surveys that exclude some of the groups most vulnerable to poverty such as the homeless. This paper documents the nature of the excluded groups and the quantitative importance of their omission for the Czech Republic. Our study combines European Union Statistics on Income and Living Conditions data for 2011 with information from the 2011 Population Census. Our preferred estimate indicates an increase in the poverty rate by almost one percentage point. The paper contributes to accurately identifying the most vulnerable members of the society, which is essential for setting effective social policies.

1. Introduction

How many poor people are there? And who are they? Obtaining good answers to these two questions is essential for any government that aims to set appropriate social policies to help those most in need, and to subsequently evaluate these policies. The standard source of information about the poor is household surveys. Although such surveys typically aim to be representative of the entire population, their sampling frame often excludes certain social groups. For instance, they usually exclude homeless people and people living in institutional homes. Even though such excluded groups represent a small fraction of the total population, they may account for a significant proportion of the poor, because they are likely to be some of the social groups most threatened by poverty. Identifying the poor solely from data that exclude these groups could, thus, result in an underestimation of the number of people below the poverty line and lead to a substantially distorted view of who the most vulnerable members of a given society are.

In the European Union, the only household survey suitable for studying poverty that is comparable across all member countries is the European Union Statistics on Income and Living Conditions (EU-SILC) survey. The EU-SILC data are used, among other purposes, for calculating the widely publicised at-risk-of-poverty rate, which is defined as the share of the population with an income below 60% of the median equivalised disposable income in a country. The main limitation of these data is that the survey

participants are sampled only from among people living in private dwellings. This means that the homeless and people living in other places, such as institutional homes and provisional houses, are excluded.

The aim of this paper is to evaluate, for the case of the Czech Republic, to what extent the exclusion of certain groups from EU-SILC distorts the available information about the number of poor people. In order to do so, it combines Czech EU-SILC data for 2011 with information from the 2011 Czech Population Census data, which, unlike EU-SILC, covers the whole population (but lacks some of the information needed for estimating income poverty). This combination should provide us with improved estimates of income poverty in the Czech Republic, a case of European and high-income countries. To that aim we develop a method that could be applied in any country with similar data, but the Czech Republic's results are not going to be directly applicable elsewhere because the number and characteristics of people not covered by surveys vary across countries (and how much is a question for further research). The case of the Czech Republic is also interesting in itself, since the levels of income inequality and poverty are among the lowest both in Europe (e.g. Decancq, Goedemé, Van Den Bosch, & Vanhille, 2013) and worldwide (e.g. Milanovic, 1999, 2011). Naturally one might wonder whether the low levels might perhaps be due to an exclusion of people with low incomes from the underlying survey, EU-SILC. We answer this question in our paper.

The core of our empirical analysis proceeds in two closely related steps. In the first step, we describe the social groups not represented in the EU-SILC, commenting on their characteristics as available from the Population Census data. In the second step, we estimate lower and upper bounds for the bias that the omission of these groups from the EU-SILC creates in the existing estimates of the at-risk-of-poverty rates for the Czech Republic.

In the Population Census data, we find that nearly 3% of the Czech population, or 300,000 people, live in non-private residences that are not covered by the EU-SILC. This diverse group includes, for example, people living in holiday homes, old people's homes, provisional housing and the homeless. They are on average older, more likely to be unemployed, single, widowed or divorced, and less likely to be well-educated or employed. The nature of these population groups living outside private residences means that most of them, although not all, are likely to be particularly vulnerable to poverty. The overall at-risk-of-poverty rate among people not living in private residences can, therefore, be expected to be substantially higher than in the overall population.

Building on these observations, we establish three sets of estimates for at-risk-of-poverty rates in the overall population that take into account people not living in private residences. Firstly, we combine information from EU-SILC with the Population Census data to impute poverty rates for those groups not directly captured by EU-SILC and we interpret this estimate as the lower bound for the actual poverty rates among these groups. The resulting overall poverty rate is only slightly higher than the benchmark estimate based on individuals living in private dwellings. Our second, upper-bound estimate is based on the extreme assumption that all people living outside private residences are at risk of poverty. This estimate leads to an overall poverty rate about three percentage points above the benchmark. A third, intermediate estimate makes specific assumptions about the poverty rates of individual subgroups outside private residences. This estimate leads to an overall poverty rate about one percentage point above the rate for people living in private residences, and we consider it likely the closest estimate to the actual size of the population at risk of poverty.

The rest of the paper is structured as follows. In Section 2 we briefly discuss relevant existing research. In the following section we describe our data and the procedure that we use to combine the EU-SILC and Population Census data. Two further sections describe the social groups excluded from the EU-SILC and present our estimates of the bounds for the bias from excluding them. The final section briefly discusses the implications of our analysis.

2. Related literature

The existing research focused on estimating poverty and inequality in the EU was reviewed thoroughly by Decancq et al. (2013) and provides references to relevant research. The share of the population at risk of poverty, which defines the poverty line as 60% of the country-specific median equivalised disposable income, is one of the most frequently used indicators of relative poverty in the EU.[1] Based on this measure in 2015, the Czech Republic has the lowest percentage of inhabitants at risk of poverty in the EU, 9.7%. This measure has been criticised for its numerous drawbacks, including its arbitrary poverty line setting (60% of the median) and its exclusive focus on only one dimension of poverty (income) (Decancq et al., 2013) as well as the failure to always provide consistent documentation and proper sample design variables in the EU-SILC dataset (Goedemé, 2013). Also, the indicator has been discussed critically with regard to the Czech Republic by Mysíková and Večerník (2016). Still, we use the poverty measure in this study because of its frequent use in other research and its central role in setting policy targets.[2] Also, importantly, our focus here is on the population groups that are not included in the EU-SILC (the data on which the measure is based) rather than on the measures of income poverty themselves.

Within the EU, the Czech Republic is an interesting case study since it has persistently had one of the lowest income inequality and relative poverty rates. Income poverty in the Czech Republic has been studied using various sources of data in the past, as reviewed recently by Bartošová and Želinský (2013), Mysíková and Večerník (2016) and Večerník (2011), and we provide only a brief overview of some of the most important contributions. While Rabušic (1998) focuses on the alleged poverty of the elderly, Večerník (2004) answers the question of who is actually poor. Sirovátka and Mareš (2006) investigate the pattern of poverty and social exclusion in the Czech Republic and the impact of social policy on this pattern. In a more recent contribution built on the research of Higgins and Lustig (2016) and Lustig and Higgins (2013), Janský, Kalíšková, and Münich (2016) analyse income inequality and relative poverty in the Czech Republic. They combine EU-SILC data with a household budget survey in a similar way to Sutherland, Taylor, and Gomulka (2002) and analyse the impact of tax and transfer systems on poverty and inequality in the Czech Republic. They find that market income without pensions is not very egalitarian (the Gini coefficient is as high as 0.46), while the inclusion of pensions in market income decrease income inequality substantially (the Gini coefficient of market income with pensions included is 0.33). They further find (p. 200) that the narrowly defined tax–benefit system (i.e. direct taxes together with social benefits, without pensions) does decrease the average poverty gap and inequality as measured by a Gini coefficient, but does not, somewhat surprisingly, change the poverty rate measure which we use in this paper. They find that further accounting for

indirect taxes increases inequality. However, the existing research does not include all the people living in the Czech Republic, only those covered by the EU-SILC survey. This is a problem not specific to the Czech Republic, which has been recognised in both sociological and economic literature.

There is a recently developed sociological literature on poverty estimation using household and population census data. A book edited by Tourangeau, Edwards, Johnson, Wolter, and Bates (2014) reviews this research and sheds new light on hard-to-survey populations, including those likely to be poor, and on how the difficulty of surveying them might affect poverty estimates. Their approach is quite broad and covers various types of hard-to-survey populations, from hard-to-sample and hard-to-identify to hard-to-find and hard-to-interview; we refer to them for an overview of these issues. Nicaise and Schockaert take a similar approach in Chapter 26 of that book to ours in this paper, with the aim of examining the hard-to-reach among the poor in Europe, and draw lessons from Eurostat's EU-SILC survey in Belgium. They also discuss the issue of underestimating the poverty rate due to excluding people from the sample framework. Also for Belgium, Schokaert, Morissens, Cincinnato, and Nicaise (2012) discuss the questions of selection bias and a priori exclusion of some groups from the EU-SILC survey. Interestingly, they run a survey to find that the poverty rate is 72% for people without home or residence and 96% for undocumented people in contrast with the EU-SILC estimate of overall poverty of 15%. Along similar lines, Carr-Hill (2017) discusses citizen volunteer surveys as a solution to the fact that household surveys omit by design several groups that constitute a substantial proportion of the poorest of the poor. Another relevant contribution of sociological and social policy literature is to look at the use of administrative data for a construction of some variables in EU-SILC in some countries, with Lohmann (2011) finding a substantial impact on results of shares of working and non-working poor. Also eastern European sociologists have been contributing to this debate – for example, Siptár, Tésits, and Alpek (2016) examine the characteristics of poverty's regional segregation in one Hungarian county, while Mysíková and Večerník (2018) question the often perceived egalitarian nature of the Czech Republic.

Both sociological and economic research shows that the problem with household surveys not covering the whole population is relevant for both developed and developing countries (Chavez & Samman, 2015). Ravallion (2003) compares the results achieved by surveys to those using national accounts. Several studies also combine survey and population census data in the context of measuring poverty. Hentschel, Lanjouw, Lanjouw, and Poggi (2000) use the case of Ecuadorian poverty maps to demonstrate how sample survey data can be combined with population census data to yield predicted poverty rates for the population covered by the population census. Alderman, Babita, Demombynes, Makhatha, and Özler (2002) combine population census and survey data to map poverty in South Africa and argue that research should make better use of population census data. In a related strand of literature, Tarozzi and Deaton (2009) document that recent years have seen widespread use of small-area maps based on census data enriched by relationships estimated from household surveys that predict variables, such as income, not covered by the census (which is the case of the Czech Republic's census as well). Tarozzi and Deaton (2009) use the 2000 Population Census for Mexico and advise caution that if population census data are to be useful in predicting

poverty in small areas there needs to be a degree of spatial homogeneity, which is unlikely to be satisfied in practice. Although our approach is different and we are not focusing on poverty in small areas as Tarozzi and Deaton (2009) did, we are also going to be cautious and, for example, provide lower and upper bounds for our estimates.

We build on the existing research to develop an empirical approach for combining the household survey and population census data, with the aim of providing a more complex empirical analysis of poverty in the Czech Republic. As far as we know, this is the first time such research questions have been addressed for the Czech Republic or any other transition or post-communist economy, and we believe that our approach is also relevant for other countries in the region.

3. Describing people living outside private residences

We start by outlining the descriptive statistics of the Population Census data. Table 1 shows a breakdown of the Czech population by housing arrangement. The main population groups not represented in the EU-SILC, each accounting for more than a tenth of the uncovered group, are people living in provisional housing, in holiday or mobile homes and in old people's homes. There is also a substantial group of people (65,000) whose residence is unknown (and who are mostly foreigners).

Tables 2A–C report the basic demographic characteristics of these groups classified by housing. People not living in private residences seem on average to be older and more likely to be single, widowed or divorced. Of this group, 36% do not have high school education, which is a much higher share than among the rest of the population (19%). People not living in private residences are also less likely to be employed or self-employed and more likely to be unemployed (9% compared to 5% for the rest of the population), fulfilling domestic tasks (7% compared to 2% of people in private residences) and retired (37% in comparison to 23% among the rest of the population). People not living in private residences reside relatively more in large cities, including Prague (18% compared to 12% of people in private residences), and they have a disproportionally higher representation in four other regions from Prague westwards.

Table 1. Breakdown of people living in the Czech Republic by housing.

Housing arrangement	Number of people	Share (%)
Private residence	10,144,961	97
Non-private residence	291,599	3
Holiday/mobile home	36,405	12
Provisional housing	49,242	17
Institutional home	194,456	67
Old people's homes	35,010	18
Lodging houses	28,143	14
Homes for people with disabilities	15,015	8
Prisons	14,455	7
Student housing	11,446	6
Children and young people's homes	3992	2
Refugee centres	2735	1
Hospitals	7612	4
Hotels	10,756	6
Monasteries	723	0
Unknown	64,569	33
Homeless	11,496	4

Source: Authors on the basis of the Czech Census 2011 data provided to the authors by the CSO.

Table 2A. Demographic characteristics of people living in the Czech Republic, grouped by housing.

	Male	Age 0–14	Age 15–29	Age 30–44	Age 45–59	Age 60–74	Age 75–96	Single	Married	Widowed	Divorced
Private residence	49	15	19	24	20	16	7	40	43	7	10
Not in private residence	58	58	6	24	26	20	11	14	47	24	13
Holiday/ mobile home	53	13	14	24	24	21	4	34	45	5	17
Provisional housing	60	10	26	34	20	7	3	50	31	3	16
Institutional homes	57	4	25	24	18	10	20	49	19	18	14
	27	0	0	1	4	18	77	13	10	64	14
	71	4	27	36	28	5	1	49	32	3	17
	51	5	14	23	25	17	16	77	3	13	7
	96	0	32	49	17	2	0	62	10	1	28
	53	0	95	4	1	0	0	97	3	0	1
	57	47	45	4	3	1	0	95	3	0	2
	55	28	17	22	21	8	5	62	10	6	23
	42	3	7	10	15	24	42	23	23	36	18
	57	4	33	29	24	8	3	55	27	5	14
	16	1	6	26	17	9	41	91	4	2	2
	63	5	26	31	22	8	8	49	29	9	15
Homeless	79	4	21	33	33	10	1	49	10	3	39

Source: Authors on the basis of the CSO data.

They are also, perhaps by definition, mostly classified as living in other types of household than a family or an individual.

In the discussion that follows, we briefly look at some of these characteristics for the selected groups from Tables 2A–C. For many of these groups, their demographic characteristics point to a substantially higher risk of poverty than in the overall population. The 11,000 homeless people reported in the Population Census data have very distinct demographic characteristics in comparison to the rest of the population. They are predominantly male (79%), neither very young nor very old, and very often single (49%) or divorced (39%), which is the highest divorce rate among the studied groups. Homeless people tend to have a poor education (32% did not complete high school and 49% completed high school without the official high school-leaving diploma). Their unemployment rate (42%) is the highest in the studied groups: less than a third of people are employed or self-employed. According to those homeless people reported in the Population Census, the homeless population is distributed proportionally to the overall population across regions with one exception – their share in one region (Moravskoslezský) is almost twice as high.

Not surprisingly, people living in old people's homes are old (77% older than 75 years of age), often widowed (64%), retired (99%) and mostly female (73%). People living in lodging houses are quite often of working age (64% between 30 and 59 years of age), male (71%) and unemployed (17%). People living in homes for people with disabilities tend to be of all ages, but are very often retired (79; likely due to their disabilities) and did not complete their high school education (83%). The prison population is dominated by men (96%); they are usually middle-aged and a third of them are employed. Students living in student housing (classified as living there rather than, for example, with parents if that is where they usually reside and spend their everyday leisure time) reside where there are large universities, especially in Prague (50% of all of them). Students living in

Table 2B. Demographic characteristics of people living in the Czech Republic, grouped by housing.

	No high school	High school without diploma	High school with diploma	Higher education	Employee	Self-employed	Unemployed	Student	Retired	Fulfilling domestic tasks	Other non-active
Private residence	19	35	32	15	40	6	5	7	23	2	17
Not in private residence	36	36	32	23	9	26	4	9	7	37	7
Holiday/mobile home	16	39	32	12	36	8	7	5	26	3	16
Provisional housing	23	34	30	14	40	9	11	7	12	6	15
Institutional homes	44	29	20	7	20	2	6	8	48	9	8
53	29	14	4	1	0	0	0	99	0	0	0
34	38	22	7	47	6	17	3	11	8	9	9
83	12	4	1	4	0	1	2	79	6	9	9
43	40	14	3	32	0	6	0	3	58	0	0
1	1	64	34	17	1	1	79	0	0	1	1
71	9	15	4	6	0	1	39	1	1	52	52
47	38	13	3	10	1	23	3	19	3	42	42
34	36	22	8	10	0	5	1	78	1	4	4
24	37	31	9	55	4	9	7	15	1	9	9
21	20	35	25	39	0	2	2	52	1	4	4
38	32	22	8	29	5	11	6	32	5	11	11
Homeless	32	49	16	3	26	5	42	1	15	1	10

Source: Authors on the basis of the CSO data.

Table 2C. Demographic characteristics of people living in the Czech Republic, grouped by housing.

	Household: individual	Household: family	Household: other	Municipality size: < 2000	Municipality size: 2000–49,999	Municipality size: 50,000+	Region: Praha	Region: Středočeský	Region: Jihočeský	Region: Plzeňský	Region: Karlovarský	Region: Ústecký	Region: Liberecký	Region: Královehradecký
Private residence	14	63	24	27	43	30	12	12	6	5	3	8	4	5
Not in private residence	11	11	77	23	38	40	19	14	5	7	4	8	4	5
Holiday/mobile home	26	49	25	53	30	17	7	28	4	12	6	9	7	4
Provisional housing	48	21	31	29	36	35	17	14	8	6	4	8	4	5
Institutional homes				16	38	45	22	11	4	6	4	9	3	5
Old people's homes		4	96	18	54	27	5	12	7	5	2	10	3	6
Lodging houses		4	96	10	38	53	24	16	3	6	5	3	4	3
Homes for people with disabilities		1	99	48	39	13	2	12	6	9	3	13	1	4
Prisons				23	41	36								
Student housing		0	100	0	3	97	51	1	2	2	1	1	6	1
Children's and young people's homes		1	99	23	45	32	7	12	7	5	2	15	7	6
Refugee centres		5	95	4	45	50	13	8	6	7	2	21	2	4
Hospitals		1	99	16	48	36	11	16	2	10	6	10	3	3
Hotels		2	98	7	22	71	52	7	5	2	2	6	1	3
Monasteries		0	100	16	25	59								
Unknown		3	97	14	37	49	29	11	3	8	3	10	2	5
Homeless				7	51	42	11	8	5	3	2	9	2	4

Source: Authors on the basis of the CSO data.

student housing and individuals in children's and young people's homes are, naturally, predominantly young, single and economically non-active. According to the Population Census data, there are almost 3000 people living in refugee centres (located not evenly across the country) and they are on average young, single or widowed, without much education (47% without high school education) and economically non-active.

Almost 8000 people are reported as residents in hospitals; these are mostly older, retired women. People living in hotels, mostly in Prague, tend to be younger, employed men. With fewer than 1000 people reported to be living in monasteries, this group is interesting in having the largest share of women (84%) and, not surprisingly, one of the largest share of singles. A relatively large group of 65,000 people has no residency reported in the Population Census, and is labelled as unknown. People in this group are more likely to be male (63%), of working age, single (49%) or divorced (15%), with low education (most either have no high school education or high school without a diploma) and either retired (32%) or unemployed (11%).

4. Data and methodology

4.1. *Data*

We draw on two principal sources of data: the EU-SILC survey and the latest national Population Census, both from the Czech Statistical Office (CSO). Specifically, we use EU-SILC 2012 for the Czech Republic, which includes information on incomes for the year 2011. The EU-SILC data are representative of all people living in private residences (or households) as opposed to collective households or institutional households that are for instance: hospitals, old people's homes or prisons. It contains detailed information that can be used to estimate various measures of poverty. The Czech Population Census was undertaken in 2011. The crucial characteristic of this Population Census for our analysis is that it attempted to collect information on almost all people living in the Czech Republic, including those not covered by the EU-SILC survey. According to the CSO (2012), the 2011 Population Census included, for the first time, information on homeless people who make use of social services and are willing to communicate (the homeless living in the wild or unwilling to communicate are still, unfortunately, not included, as might not be some other groups of people, about which however the CSO does not provide more information, but they might, we speculate, include foreigners not registered by the authorities). While we include the homeless in the analysis, it is not the focus of our paper in contrast with a recent study by Kuchařová and Peychlová (2016), who also made use of the Population Census. While the Population Census does not contain the variables necessary for measuring the risk of poverty because it does not include any information about income, it contains a range of demographic and labour market participation characteristics which enable us to match the poverty rates calculated from EU-SILC to the Population Census data.

For the purpose of our analysis, we split the Czech population by housing arrangement. This seems to be the most natural split, first because housing information is available in both the Population Census and the EU-SILC data, and second because at least for the groups that we focus on, such as people living in old people's homes, prisoners and the homeless, their housing arrangements are defining characteristics of

their living conditions and social status. We underline here that the EU-SILC data only cover private residences. According to the Population Census, nearly 3% of the Czech population, or 300,000 people, live in non-private residences and are, thus, not covered by the EU-SILC survey.

4.2. *Imputing poverty rates in the Population Census data*

To benefit concurrently from the better coverage of the Population Census and the detailed information about income in the EU-SILC data, we use the EU-SILC data as a basis for imputing poverty rates into the Population Census data. We carry out the imputation on the basis of eight variables that are important demographic characteristics and for which there is available information in both the EU-SILC and the Population Census data: gender, age, marital status, educational attainment, labour market status, type of household, municipality size and NUTS 3 region.

We begin by recoding the demographic variables used for linking the two datasets to the lowest common denominators, so that both datasets include the same categories. Table 3 summarises our aggregation of the categories, and shows the demographic categories used in the imputation described in detail below. For each variable it shows two types of categories – one more detailed and one more aggregate. Table 3 also shows the share of the total population falling in each category for both datasets (with the shares for SILC being based on population weights available in the data). The distribution of people across

Table 3. Making variables comparable between the SILC and Population Census data.

Census	Paper (more detailed)	Paper (less detailed)	Share in census	Share in SILC
Gender	Male	Any	0.48	0.49
Age	0–14	0–14	0.15	0.15
	15–29	15–59	0.18	0.19
	30–44		0.23	0.23
	45–59		0.20	0.20
	60–74	60+	0.17	0.17
	75+		0.07	0.06
Marital status	Single	Any	0.40	0.39
	Married		0.43	0.44
	Widowed		0.07	0.09
	Divorced		0.10	0.08
Education	Elementary	Any	0.16	0.13
	High school w/o 'maturita'		0.29	0.30
	High school w 'maturita'		0.27	0.29
	Higher		0.13	0.13
	Undefined (children)		0.15	0.15
Labour market status	Employee	Working	0.40	0.36
	Self-employed		0.06	0.08
	Retired	Not working	0.05	0.05
	Unemployed		0.07	0.07
	In household		0.23	0.24
	Student (age> 14)		0.02	0.03
	Other non-active		0.17	0.16
Type of household	Individual	Any	0.13	0.11
	Complete family		0.63	0.56
	Other		0.23	0.33
Size of village/town/city	< 2000	Any	0.27	0.27
	2000–49,999		0.43	0.41
	50,000+		0.30	0.32
Region	14 NUTS 3 regions	Any	–	–

Source: Authors on the basis of the SILC and Census data from the CSO.

categories is very similar between the two datasets, indicating that the SILC sample is broadly representative of the population (with certain qualifications including, of course, the one which is the focus of this paper).

Next, we develop an approach for imputing the poverty rates observed in the EU-SILC for people observed in the Population Census. One option would be to estimate a regression of the poverty indicator observed in the EU-SILC on the variables observed in both datasets and potentially also their interactions, and then use the estimates to obtain out-of-sample predicted values for poverty rates in the Population Census data. While the regression approach would have its advantages, the drawback would be that, unless we included the (huge) full set of interactions, the regression would impose strong restrictions on how different variables are related to poverty for different population groups. For example, if there were no interactions at all, the regression would assume that the effect of being divorced on the likelihood of being in poverty is the same for a 35-year-old male with a high school degree as for a 50-year-old woman without a high school diploma. If, on the other hand, we wanted to allow for a full set of interactions then the number of dummies on the right-hand side of the regression would become exceedingly large. As an alternative between all and no interactions, we could identify the most important interactions and include only those, which, although feasible, would require us to make decisions about where the cut-off point is for the importance of the interactions. Instead, therefore, we take a non-parametric approach, where we calculate the at-risk-of-poverty rates for many subgroups of the EU-SILC data (using cross-sectional weights) and impute these rates for the corresponding subgroups in the Population Census data.

Given that the EU-SILC dataset contains only about 20,000 observations, an important challenge in this approach is the appropriate choice of the size of the cells used for the imputation. The more detailed the cells for which we estimate the poverty rates are, the more precise definition of demographic characteristics is used and thus we pair to more similar people in the Population Census data. But when the cells become very small, the estimates of cell-specific poverty rates obtained from the EU-SILC data will become less precise, because only a few individuals are used to estimate the at-risk-of-poverty rate making the estimates unreliable. We address this issue by only considering cells with at least 50 observations in the EU-SILC. This threshold is obviously arbitrary and we chose 50 because it seems to us that this is close to the smallest cell size for which the standard error in the estimation of the mean could still be deemed acceptable (for example, for a group with a true poverty rate of 0.1, a sample size of 50 leads to a standard error in the estimation of the poverty rate of around 0.04).

We proceed in eight steps to calculate the at-risk-of-poverty rate for many subgroups of the EU-SILC data and impute these rates to the corresponding subgroups in the Population Census data. In each step, we match cells in the EU-SILC to those in the Population Census using all or some of the demographic characteristics, and we impute the poverty rates for those cells that correspond to at least 50 observations in the EU-SILC. The earlier steps use a more precise definition of demographic variables and thus allow a more precise match on observables, while the later steps use less precise categorisations of people based on their demographics in order to obtain cells with at least 50 observations for groups which do not yet have an imputed poverty rate. The variables and categories used in individual steps are as follows.

- **Step 1** (3.5% of people in the census matched in this step). We calculate the at-risk-of-poverty rates in the EU-SILC data for the subgroups defined based on the more detailed categories of all eight demographic variables (two categories by gender, six age categories, four categories by marital status, four education categories, seven labour market status categories, three categories by type of household, three categories by the size of municipality, and 14 regional categories, see Table 3). We then match these at-risk-of-poverty rates to the corresponding subgroups in the census data if the number of observations in the given subgroup in the EU-SILC data is at least 50. As this match is the most demanding one, only 3.5% of people in the Population Census belong to cells which meet the threshold of 50 individuals in SILC and have poverty rates imputed in this step.
- **Step 2** (40% of people in the census matched in this step). We calculate the at-risk-of-poverty rates for subgroups in the EU-SILC data, where subgroups are created based on the more detailed definition of categories for seven demographic variables and less detailed definition for one demographic variable. This step is repeated eight times, using the less detailed definition for one of the eight variables in every repetition. We then match these subgroups to the corresponding subgroups in the Population Census data and assign poverty rates if the number of observations in the EU-SILC data is at least 50. An additional 40% of the census population have poverty rates imputed in this step. Removing region from the match criteria is responsible for virtually all the additional matches; relaxing the other criteria (one at a time) leads to matches for only about 0.5% of census observations.
- **Step 3** (33% of people in the census matched in this step). We create subgroups using the less detailed categories for region and one other demographic variable, while using the more detailed categories for the remaining six variables. We repeat this step seven times, taking one of the seven demographic variables (excluding region) at a time and using the less detailed categories for this variable. The rest of the procedure is the same as in the previous step. This step imputes poverty for an additional 33% of people in the census. The criteria, relaxing which (together with the region criterion) leads to most of these matches are the size of the village/town/city, the labour market status and the type of household.
- **Step 4** (13% of people in the census matched in this step). We create subgroups based on less detailed categories for region, municipality size and one other demographic variable (one at a time, repeating this step six times for each variable), while we use the more detailed categories for the remaining five variables. This leads to imputed poverty rates for an additional 13% of people. Half of this is due to relaxing the region and municipality size in combination with the marital status.
- **Step 5** (7% of people in the census matched in this step). We create subgroups using the less detailed categories for region, municipality size, marital status and one of the other demographic variables (again just one at a time, repeating this step five times for each of them), while we still use the more detailed categories for the four remaining variables. As a result, 7% of the census population have poverty rates matched in this step.
- **Step 6** (3% of people in the census matched in this step). The subgroups are created based on the less detailed categories of region, municipality size, marital

status, labour market status and one other variable (one at a time making four repetitions of this step), while the remaining three variables still use the more detailed categories. Of the census population, 3% have poverty imputed in this step.

- **Step 7** (0.5% of people in the census matched in this step). The subgroups are defined based on the less detailed categories of region, municipality size, marital status, labour market status, type of household and one of the remaining variables (one at a time making three repetitions of this step), while the remaining two variables use the more detailed categories. This steps imputes poverty for most of the remaining, who amount to about 0.5% of the census population.
- **Step 8** (0.04% of people in the census matched in this step). We define subgroups based on the less detailed categories of region, municipality size, marital status, labour market status, type of household, education and one of the remaining two variables (one at a time making two repetitions of this step), while the one remaining variable uses the more detailed categories. The small number of people who were not matched in the previous steps are matched in this round.

These eight steps impute poverty for the entire population observed in the Population Census. Three quarters of people in the census have poverty imputed by step 3 and 90% by step 4. Thus, a vast majority of imputations are based on a match in most of the more detailed categories (left column of Table 3), only relaxing the region requirement and one or two of these categories. The advantage of this eight-step procedure is that, for each population group, it uses the cell size that provides the most precise match while having enough observations in EU-SILC to enable a meaningful estimation of the poverty rate for the group in question. The disadvantage of the procedure is that it contains a certain degree of arbitrariness in terms of the order in which different variables have their categories aggregated. However, experimenting with alternative variable orders has revealed that they have only a minor effect on the results.

In following the eight steps to calculate the at-risk-of-poverty rate for many subgroups of the EU-SILC data and to impute these rates to the corresponding subgroups in the Population Census data, we make a number of assumptions that we at least briefly discuss here. In inputting poverty rates from people in the EU-SILC data to people covered only by the Population Census data, we assume that people in these two groups are similar in a number of aspects. We assume that the composition of households, equivalence scales and the related economies of scale are similarly relevant in the two groups. The same is assumed about gifts and any other transfers from family members or other people. Since our focus here is on income-based indicators of poverty, we neglect the differences between non-cash incomes across the two groups. We are aware of these limitations, but the available data do not provide much space to consider them empirically. Further, we assume that the probability of being at risk of poverty is independent on the type of housing when controlling for observable characteristics. This is a strong assumption, but one that we need to make in order to use our imputation procedure. This assumption is less likely to be satisfied for some groups of people by type of residence than for others. Nevertheless, we address these limitations in the following section by the introduction of two alternative estimates of the poverty rates for people outside private residencies, which relax these assumptions for some groups.

4.3. *Estimates of poverty*

We believe that the straightforward application of this imputation procedure produces a lower-bound estimate of poverty among people not living in private residences. The matching procedure assumes that these people have, conditional on the eight matching characteristics, the same likelihood of income poverty as those living in private residences. In reality, many of the subgroups outside private residences (e.g. the homeless) are disproportionately more at risk of poverty because of the very reason for which they are not in private residences. For this reason, we supplement the results of the imputation with two further alternative approaches. The first alternative approach attempts to provide the upper bound for the poverty rates. To construct this upper bound estimate, we assume that all people living outside private residences are at risk of poverty.

The second alternative approach aims to obtain intermediate estimates between the lower and upper bounds; these represent what we believe to be the most realistic estimate. For these intermediate estimates we make a number of assumptions. First, we assume that the homeless, people living in refugee centres, and those in children's and young people's homes are all living below the poverty threshold (i.e. a poverty rate of 100% for these selected groups). Second, we assume that people living in prisons, old people's homes, homes for people with disabilities and unknown residences, a realistic value of the poverty rate is an average of the lower bound (imputation) and upper bound (100% poverty rate assumption) estimates. For the remaining housing groups, we assume that the matching procedure provides a realistic poverty estimate of the respective housing groups. The three approaches including their assumed poverty rates for all the groups are specified and described, together with the results, in Table 4.

5. Estimates of at-risk-of-poverty rates

We now turn to applying the imputation procedure described above to the Czech Republic's EU-SILC and Population Census data. For the first time, we derive estimates of the poverty rate for people covered only in the Population Census data. This results in adjusted overall poverty estimates, i.e. the share of the population at-risk-of-poverty for people living in both private and non-private residences.

The poverty rate for individuals in private residences estimated by our imputation procedure is 8.4% (Table 4). This is lower than the corresponding official figures published by Eurostat for 2012 (9.6%). The difference reflects the different weighting used and potential imprecisions driven by our imputation procedure. In these estimates we use the Population Census data population weights to construct the poverty measure for both people living in and outside private residences, whereas the Eurostat uses the sampling weights contained in the EU-SILC data. Importantly, these differences in the level of the poverty rate for people in private residences should not affect the main results of our analysis. This is because our focus is on the difference between the poverty rates with and without people outside private dwellings, when calculated using the same procedure and the same population weights.

Table 4 shows the results for the lower bound (imputation), upper bound (100% poverty rate assumption) and intermediate (different assumptions for different groups) estimates of poverty rate for all groups defined by types of residence. The lower bound poverty estimate

Table 4. Breakdown of people living in the Czech Republic by housing (poverty rate and number of people below the poverty line).

All	Lower bound		Upper bound		Intermediate	
	Rate (%) 9	Number 887,108	Rate (%) 11	Number 1,143,776	Rate (%) 9	Number 958,516
Private residence	8.4	852,177	8	852,177	8	852,177
Not in private residence	12	34,991	100	291,599	36	106,339
Holiday/mobile home	9.3	3386	100	36,405	9	3386
Provisional housing	11	5417	100	49,242	11	5417
Institutional homes	13	25,279	100	194,456	44	86,042
Old people's homes	7	2311	100	35,010	53	18,660
Lodging houses	13	3659	100	28,143	13	3659
Homes for people with disabilities	18	2703	100	15,015	59	8859
Prisons	22	3180	100	14,455	61	8818
Student housing	9	1019	100	11,446	9	1019
Children's and young people's homes	21	838	100	3992	100	3992
Refugee centres	22	602	100	2735	100	2735
Hospitals	10	738	100	7612	10	738
Hotels	10	1033	100	10,756	10	1033
Monasteries	7	48	100	723	7	48
Unknown	13	8394	100	64,569	57	36,481
Homeless	20	2299	100	11,496	100	11,496

Source: Authors on the basis of SILC and Census data from the CSO.

results in a poverty rate ranging from the same level as the average for people living in private residences (for those living in mobile/holiday homes and student housing) to almost three times that poverty rate (for the prisoners, children and young people living in children's and young people's homes, people living in refugee centres and the homeless, with an imputed poverty rate of 20–22%). The only group of people living in non-private residences whose estimated poverty rate is below the average for people in private residences are those living in monasteries (whose imputed poverty rate is 6.6%). The resulting lower bound at-risk-of-poverty rate for people living in non-private residences is 12% and the share of the population at-risk-of-poverty is 8.5%, only about 0.1 percentage points higher than the level estimated when only people living in private residences are included.

As we discussed in the Methodology section, some people who live outside private residences are disproportionately more at risk of poverty because of the very reason for which they are there, which is something the imputation procedure cannot capture. Homeless people likely represent the strongest example of this bias in the imputation procedure: the poverty rate implied by the imputation for the homeless is 20% (Table 4), but Barták (2011) reports that 70% of the homeless in his survey have an income below CZK5000 (the poverty rate for individuals in the Czech Republic is just above CZK10,000). We attempt to adjust for this likely bias with our upper bound and intermediate poverty estimates.

The upper bound poverty rate, based on the assumption that all people outside private residences are at risk of poverty, is 11.0%; this is 2.5 percentage points above the lower bound estimate. This is certainly an overestimate of the true poverty rate, since some groups outside private residences include a large share of people who may be reasonably well-off. For example, some people with high incomes were likely in hospital or living in an old people's home at the time of the Population Census data collection. Therefore, an intermediate estimate of the poverty rate, based on different assumptions for individual subgroups, will give us a result that is likely closer to reality. The intermediate estimate suggests that 36.5% of people living in non-private residences are at risk of poverty, which results in an intermediate estimate of the poverty rate of 9.2% for the whole population. The overall poverty rate for people living in private and non-private residences is thus about 0.8 percentage points higher than the estimate produced based on the EU-SILC data alone.

6. Conclusion

The share of people living below the poverty line in the Czech Republic is very low in international comparison. Estimates using the EU-SILC data indicate that less than 10% of the population live under the poverty threshold. However, the EU-SILC data exclude individuals who do not live in private residences. In this article we used the Population Census data, which covers the entire population, to adjust the poverty estimate by including people who do not live in private residences.

We establish three estimates. The first estimate relies directly on the poverty rates imputed using the EU-SILC data, and results in an overall poverty rate which is only about 0.1 percentage point higher than the benchmark estimate for individuals in private residences. We interpret this estimate as the lower bound for the actual poverty

rate. The second estimate makes the extreme assumption that all people outside private residences are at risk of poverty. This approach leads to an overall poverty rate about 2.5 percentage points higher than the benchmark, and we take it as the upper bound for the actual poverty rate. The third estimate uses further assumptions about where the different subgroups outside private residences lie between the lower and upper bound estimates. This approach leads to an overall poverty rate about 0.8 percentage points above the benchmark estimate. We believe this estimate is likely the closest to the actual prevalence of poverty.

Three important lessons emerge. First, accounting for people outside private residences leads to an increase in the overall at-risk-of-poverty rate and that increase is non-negligible but also not dramatic. This implies that estimates based purely on EU-SILC data provide reasonably accurate aggregate information despite not taking into account the situations of individuals outside private residences, at least in the Czech Republic. Second, our intermediate estimate implies that 11% of people at risk of poverty do not live in private residences (while the lower and upper bound estimates imply 4% and 25%, respectively). This is not a small number, and it shows that while the aggregate poverty rate is not greatly underestimated, people living outside private residences should not be forgotten when designing social policies and programmes (such as social housing, policies dealing with homelessness, etc.). Third, we show that the key groups that are not captured by the EU-SILC data and are disproportionally at risk of poverty are people living in institutional homes (in particular old people, the disabled, people in lodging houses and prisoners) and the homeless.

Finally, we discuss a few questions for further research. First of all, a natural extension of our paper lies in the application of the procedure we have used to other countries and also to different measures than only the at-risk-of-poverty rate. More complex and comparative research in this area should shed new light on the relative poverty rates in European countries and the potential biases in the standard estimates of poverty, and how these biases differ across countries. However, achieving such a comparison may be complicated by the insufficient quality and limited coverage of – as well as difficult to access – population census data in many countries. Further research should also focus on the non-income dimensions of poverty, such as social exclusion, which might be of even higher prevalence and importance in the population groups that are left out of the standard household surveys.

Notes

1. The at-risk-of-poverty indicator is defined in the following way. People at risk of poverty are those living in a household with an equivalised disposable income below the risk-of-poverty threshold, which is set at 60% of the national median equivalised disposable income (after social transfers). The equivalised income is calculated by dividing the total household income by its size, as determined by applying the following weights: 1.0 to the first adult, 0.5 to each other household member aged 14 or over, and 0.3 to each household member aged under 14.
2. As one of the five headline target indicators, the Europe 2020 strategy aims to reduce poverty by lifting at least 20 million people out of the risk of poverty or social exclusion by the year 2020. There are three components of this indicator: people living in households with very low work intensity; people at risk of poverty after social transfers; people severely

materially deprived. The second of these three components is the at-risk-of-poverty rate that we exclusively focus on in this paper.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

Matěj Bajgar 🔘 http://orcid.org/0000-0002-3780-7970
Petr Janský 🔘 http://orcid.org/0000-0001-6499-9923
Klára Kalíšková 🔘 http://orcid.org/0000-0002-4762-086X

References

Alderman, H., Babita, M., Demombynes, G., Makhatha, N., & Özler, B. (2002). How low can you go? Combining census and survey data for mapping poverty in South Africa. *Journal of African Economies*, *11*(2), 169–200.

Barták, M. (2011). *Bezdomovství v ČR: Zdravotní stav bezdomovců a jeho determinanty*. Ústí nad Labem: Univerzita JE Purkyně v Ústí nad Labem.

Bartošová, J., & Želinský, T. (2013). The extent of poverty in the Czech and Slovak Republics 15 years after the split. *Post-Communist Economies*, *25*(1), 119–131.

Carr-Hill, R. (2017). Improving population and poverty estimates with citizen surveys: evidence from East Africa. *World Development*, *93*, 249–259.

Chavez, C. V., & Samman, E. (2015). *Exclusion in household surveys: Causes, impacts and ways forward* (ODI Research reports and studies). The Overseas Development Institute. Retrieved from https://www.odi.org/publications/8824-exclusion-household-surveys-causes-impacts-ways-forward

CSO. (2012). *Výsledky sčítání bezdomovců 2011*. Retrieved from https://www.czso.cz/csu/sldb/vysledky_scitani_bezdomovcu

Decancq, K., Goedemé, T., Van Den Bosch, K., & Vanhille, J. (2013). The evolution of poverty in the European Union: Concepts, measurement and data. *For Better For Worse, For Richer For Poorer: Labour Market Participation, Social Redistribution and Income Poverty in the EU*. Retrieved from http://webhost.ua.ac.be/csb/ImPRovE/Working%20Papers/ImPRovE%20WP%201301_1.pdf

Goedemé, T. (2013). How much Confidence can we have in EU-SILC? Complex sample designs and the standard error of the Europe 2020 poverty indicators. *Social Indicators Research*, *110*(1), 89–110.

Hentschel, J., Lanjouw, J. O., Lanjouw, P., & Poggi, J. (2000). Combining census and survey data to trace the spatial dimensions of poverty: a case study of Ecuador. *The World Bank Economic Review*, *14*(1), 147–165.

Higgins, S., & Lustig, N. (2016). Can a poverty-reducing and progressive tax and transfer system hurt the poor? *Journal of Development Economics*, *122*(SupplementC), 63–75.

Janský, P., Kalíšková, K., & Münich, D. (2016). Does the Czech tax and benefit system contribute to one of Europe's lowest levels of relative income poverty and inequality? *Eastern European Economics*, *54*(3), 191–207.

Kuchařová, V., & Peychlová, J. K. (2016). Velikost a struktura skupin osob bez domova a osob vyloučených z bydlení. *VÚPSV*. Retrieved from http://www.vupsv.cz/?p=news&id=382&site=default

Lohmann, H. (2011). Comparability of EU-SILC survey and register data: The relationship among employment, earnings and poverty. *Journal of European Social Policy*, *21*(1), 37–54.

Lustig, N., & Higgins, S. (2013). *Commitment to equity assessment (CEQ): estimating the incidence of social spending, subsidies and taxes. handbook*. CEQ Working Paper No. Retrieved from http://www.commitmentoequity.org/publications_files/Methodology/CEQWPNo1%20Handbook%20Jan%202013.pdfRetrieved from

Milanovic, B. (1999). Explaining the increase in inequality during transition. *The Economics of Transition*, *7*(2), 299–341.

Milanovic, B. (2011). *Worlds apart: measuring international and global inequality*. Princeton, NJ: Princeton University Press.

Mysíková, M., & Večerník, J. (2016). Poverty in the Czech Republic: a critical look at EU indicators. *Mimeo*. doi:10.13140/RG.2.1.2043.7524

Mysíková, M., & Večerník, J. (2018). Personal Earnings Inequality and Polarization: The Czech Republic in Comparison with Austria and Poland. *Eastern European Economics*, *56*(1), 57–80.

Rabušic, L. (1998). The poverty of the Czech elderly–myth or reality? *Czech Sociological Review*, *6*(1), 5–24.

Ravallion, M. (2003). Measuring aggregate welfare in developing countries: how well do national accounts and surveys agree? *Review of Economics and Statistics*, *85*(3), 645–652.

Schokaert, I., Morissens, A., Cincinnato, S., & Nicaise, I. (2012). *Armoede tussen de plooien: aanvullingen en correcties op eu-silc voor verborgen groepen armen*. Leuven: HIVA-KULeuven.

Siptár, D., Tésits, R., & Alpek, L. (2016). Cultural and regional characterictics of poverty segregations. *Eastern European Countryside*, *22*(1), 107–134.

Sirovátka, T., & Mareš, P. (2006). Poverty, social exclusion and social policy in the Czech Republic. *Social Policy & Administration*, *40*(3), 288–303.

Sutherland, H., Taylor, R., & Gomulka, J. (2002). Combining household income and expenditure data in policy simulations. *Review of Income and Wealth*, *48*(4), 517–536.

Tarozzi, A., & Deaton, A. (2009). Using census and survey data to estimate poverty and inequality for small areas. *Review of Economics and Statistics*, *91*(4), 773–792.

Tourangeau, R., Edwards, B., Johnson, T. P., Wolter, K. M., & Bates, N. (2014). *Hard-to-Survey Populations*. Cambridge: Cambridge University Press. Retrieved from http://ebooks.cambridge.org/ref/id/CBO9781139381635.

Večerník, J. (2004). Who is poor in the Czech Republic? The changing structure and faces of poverty after 1989. *Sociologickỳ Časopis/Czech Sociological Review*, *40*(06), 807–833.

Večerník, J. (2011). Empirickỳ vỳzkum chudoby v českých zemích ve třech historických obdobích. *Data a Vỳzkum-SDA Info (Data and Research-SDA Info)*, *5*(2), 133–146.