EQUALITAS
Economics of Inequality and Poverty Analysis

EQUALITAS Working Paper No. 69

Personal Wellbeing and Intergenerational Mobility in Spain

Amaia Palencia - Esteban

Pedro Salas - Rojo

April, 2021

Wellbeing and Intergenerational Mobility in Spain*

Amaia Palencia-Esteban

Universidade de Vigo (Spain), ECOBAS and EQUALITAS

E-mail: apalencia@uvigo.es

Pedro Salas-Rojo

Universidad Complutense de Madrid (Spain), ICAE and EQUALITAS

E-mail: pedsalas@ucm.es

Abstract

This paper explores the relation between personal wellbeing - measured with life satisfaction - and intergenerational mobility in Spain (2017). We apply Statistical Learning techniques to overcome data limitations and estimate intergenerational income mobility, setting the ground for future research with incomplete databases. Then, by means of recently developed graphical tools and several econometric specifications, we find the relation between personal wellbeing and intergenerational income mobility to be non-significant. This result also applies to educational and occupational mobility. Confronting the comparison theory, improving or worsening one's fathers' socioeconomic status does not seem to have permanent effects on individuals' wellbeing. In line with the literature, other variables such as enjoying good health or being married are found to be positively associated with welfare.

Keywords: Wellbeing, Intergenerational Mobility, Life Satisfaction, Statistical Learning.

JEL Code: I14, I31, J62

---

* The research leading to these results has received funding from Ministerio de Ciencia, Innovación y Universidad (FPU grant 17/735) and Ministerio de Economía, Industria y Competitividad (FPI grant). The usual disclaimer applies.

1. Introduction

The study of wellbeing has received growing attention in the last decades, as social scientists have understood that the quality of life includes many factors beyond income or consumption capabilities (Stiglitz et al., 2009, p12). Moreover, the large economic downturn, insecurity, loneliness or social isolation caused during the COVID-19 pandemic have highlighted the importance of personal wellbeing. In this context, understanding the factors associated with higher welfare levels is crucial to design policy interventions aimed at improving and promoting the general life satisfaction and happiness. However, this is not an easy task. Being wellbeing a multidimensional variable, the literature has focused on many different aspects, such as gender, migration or employment status (EU, 2016), physical and mental health (Steptoe et al. 2015), the season of birth (Isen et al. 2017), height (Deaton and Arora, 2009) or the occupational/social status, educational level, and relevant aspects of childhood (Di Tella and MacCulloch, 2006; Hadjar and Samuel, 2015). Providing more evidence, this paper explores the connection between personal wellbeing and intergenerational mobility in Spain (2017).

Previous research on the relation between intergenerational mobility and wellbeing has led to inconclusive results. Some authors have found upward mobility to be associated with higher levels of subjective wellbeing, with downward mobility producing the opposite effect (Clark et al., 2008; Hadjar and Samuel, 2015; Zhao et al., 2017). In line with the so-called *comparison theory*, the permanent psychological effects derived from achieving a higher or lower socioeconomic status also apply to intergenerational mobility. While upward mobility generally implies the fulfillment of parental expectations, personal self-realization and higher consumption capabilities, downward mobility seems to have a strong and persistent negative psychological impact. However, further research has also found that these mobility effects are transitory and dissipate with time (Di Tella and MacCulloch, 2006; Guilbert and Paul, 2009). According to the *hedonic adaptation theory*, individuals adjust to their new status as they move along the socioeconomic ladder. Thus, wellbeing improvements derived from intergenerational mobility have non-lasting effects; they appear after the mobility takes place, vanishing afterwards. Finally, other authors (Zhang and de Graaf, 2016; Iveson and Deary, 2017) find intergenerational mobility and personal wellbeing to be unrelated, with no effects attributed neither to the short nor the long term.

The scarcity of modern data relating wellbeing and intergenerational mobility has generally restricted the analysis to country case-studies.[1] In this paper we take the data from the Centro de

[1] The literature has mainly considered Anglo-Saxon countries. For instance, Nicholaev and Burns (2014) studied the U.S., Guilbert and Paul (2009) Australia, Hadjar and Samuel (2015) Britain, and Iveson and Deary (2017) Scotland. Exceptionally, Clark et al. (2008) took data from Germany, and Zhang and de Graaf (2016) or Zhao et al. (2017) focused on China. Finally, some studies perform cross-country analyses, but

Investigaciones Sociológicas (CIS, 2017) and analyze the relation between personal wellbeing -measured with life satisfaction- and intergenerational income, educational and occupational mobility in Spain. Surprisingly, as far as we are aware, this is the first study performing this analysis in a Mediterranean country. The case of Spain is particularly interesting, as its different intergenerational mobility patterns make it suitable for analyzing whether the effects over wellbeing vary across the three mobility approaches considered. Indeed, while Spain holds a median position among the European countries regarding wellbeing levels (EU, 2016), its rank varies across different forms of mobility. Spain is the most educationally mobile country in the European Union, with 60% of its total population experiencing upward education mobility (EU, 2018; Cabrera et al., 2020), but it presents a relatively strong occupational persistence, and an intermediate intergenerational income correlation: income mobility is larger in the Nordic countries, but smaller in Italy and the US (Cervini-Pla, 2015).

Bearing this evidence in mind, considering Spain and three types of mobility allows us to test whether wellbeing is affected through different channels (Molina et al., 2011; Nikolaev and Burns, 2014). Although income is directly related to higher consumption and economic security, professional occupations are proxies of social class and recognition (Bukodi and Goldthorpe, 2011). Climbing the social ladder by working on a profession with higher social recognition than that of the father's may bring positive physiological rewards such as the fulfillment of expectations. Similarly, higher educational levels could provide greater life satisfaction through indirect channels like a healthier lifestyle or a more diverse leisure and cultural consumption (Torche, 2015; Michalos, 2017; Schuck and Steiber, 2018). The main objective of the paper is, precisely, to check whether any of these potential effects have a long-lasting impact on personal wellbeing.

An extra issue needs to be considered before studying these three channels. This paper uses data from the Centro de Investigaciones Sociológicas (CIS; 2017), which provides precise information on fathers' occupation and education. However, this database does not include information on fathers' incomes, hindering a direct approach to intergenerational income mobility. Indeed, the absence of valid fathers' income information is the norm in the intergenerational mobility analysis, and has traditionally hampered empirical approaches to the matter. Overcoming this limitation, we take data from three waves of the Encuesta de Presupuestos Familiares (1980/81, 1990/91 and 2000/01) to impute the fathers' income into the CIS database. In doing so, we apply the Statistical Learning techniques recently proposed in Bloise et al. (2020) to the canonical

they either use old data (1994-2001) from Eurostat (Molina et al., 2011) or restrict their analysis to young individuals and educational mobility (Schuck and Steiber, 2018).

imputation method "Two Samples Two Stages Least Squares (TSTSLS)" (Bjorklund and Jantti, 1997).

Our results show that the relation between wellbeing and intergenerational mobility is not significant. Contradicting the comparison theory, the potential effects derived from improving or worsening one's father's socioeconomic situation are not permanent in Spain. This conclusion is obtained from two different independent analyses. First, we apply the new graphical tools developed by Jenkins (2019a, 2020) and show that personal wellbeing is neither higher nor more unequally distributed among individuals who experience a given type of intergenerational mobility (upward or downward) with respect to the immobile. Indeed, its distribution is rather homogeneous across all mobility categories considered. Second, we deepen the analysis by regressing the individual measures of wellbeing against intergenerational mobility measures and several sociodemographic controls.

Both, our preferred specifications and the robustness checks provide ample evidence of the absence of a significant permanent relation between personal wellbeing and any mobility measure. We suggest that, if intergenerational mobility had an impact on welfare, it dissipates with time, just as proposed by the hedonic adaptation theory. Still, and in line with previous literature, we find that factors like enjoying good health or being married are positively connected to wellbeing. Furthermore, we find that belonging to higher income quantiles is positively related to higher life satisfaction, being the effects of the occupational and educational levels non-significant. Overall, it seems that once some basic necessities are covered, including good health and economic security (proxied by income), other immaterial aspects like social status do not have long-lasting effects over personal wellbeing.

The contribution of this article is three-fold. First, we contribute to the wellbeing debate by showing the absence of a permanent relation between life satisfaction and intergenerational mobility in Spain. Second, we use innovative Statistical Learning algorithms to palliate the incomplete and imperfect nature of the data and estimate intergenerational income mobility. These computing techniques, combined with the TSTSLS imputation method, allow us to include intergenerational income mobility in welfare analysis. Being this approach largely overlooked in the literature, mainly due to data availability, we set the ground for future empirical analysis. Finally, we apply the new analytical tools proposed by Jenkins (2019a, 2020), which provide a simple and intuitive framework that easies the comparison of categorical variables between population groups, and have not previously been implemented for analyzing wellbeing in Spain.

The reminder of the paper is structured as follows. Section 2 presents the main database, explains how wellbeing is measured and describes the remaining variables employed in the analysis, providing some theoretical background on how they might be related. Section 3 explains the

Statistical Learning methods and the auxiliary data used to compute intergenerational income mobility. Section 4 studies intergenerational income, educational and occupational mobility in Spain. Section 5 presents several graphical tools and econometric specifications relating various intergenerational mobility measures and subjective wellbeing. Section 6 concludes.

2. Data

The data comes from the module "Social Inequality and Social Mobility in Spain", a survey conducted by the Centro de Investigaciones Sociológicas (CIS) in 2017 based on the design explained in Betancort et al. (2019).[2] This survey is the latest database available in Spain where the respondents are retrospectively asked about their fathers' information during their adolescence. From the 2500 interviews originally performed following a stratified multi-stage sampling procedure, the CIS reports 2482 valid observations that are representative for the Spanish population by age and gender. To exclusively include individuals participating in the labor market, while also following the intergenerational mobility literature, the sample is restricted to individuals aged between 30 and 60 years. Once we apply these restrictions, we are left with a final sample of 1151 observations.[3]

This section presents the variables employed in our analysis. First, we focus on the dependent variable, explaining how life satisfaction is used to measure wellbeing. Then, we describe the variables used to estimate intergenerational mobility - income, educational level and occupation of the individuals and their fathers –, and relate them to the main theories proposed to explain their relation to life satisfaction. Finally, we introduce the social and demographic controls used to account for the remaining factors that, according to the literature, affect wellbeing.

2.1 Dependent variables.

The literature usually proxies wellbeing by self-reported life satisfaction (Molina et al., 2011; Iveson and Deary, 2017; Shuck and Steiber, 2018; Mahler and Ramos, 2019). Despite we acknowledge certain levels of subjectivity in this variable, ample evidence demonstrates that it provides meaningful, reliable and valid information about individuals' wellbeing. In general, life

---

[2] The CIS is a dependent entity of the Spanish Ministry of Presidency whose main task consists on improving the scientific knowledge of the Spanish society. The database, the questionnaire and details of the sample design are available at:
http://www.cis.es/cis/opencm/ES/2_bancodatos/estudios/ver.jsp?estudio=14350

[3] The literature has repeatedly aware about employing surveys with missing observations, as they might reduce the efficiency of estimators and lead to wrong interpretations of the results (Zhong, 2010; Kline and Santos, 2013; Chen and Fu, 2015). Following the advices of these authors, we fill the missing observations with a multiple imputation method, based on Markov Chain Monte Carlo (Spade, 2020). The results with the non-imputed dataset do not meaningfully vary, and are available upon request.

satisfaction is related to long-term factors used by the individuals to make judgements about the quality of their lives. Among those factors, the literature highlights positive psychological aspects such as the self-fulfillment of personal ambitions or expectances, enjoying good health and shape, or the benefits derived from stable social interactions, including a successful marriage. In this paper, we measure life satisfaction by collecting the answers to the following question: *On a discrete scale from 0 (completely unsatisfied) to 10 (completely satisfied), do you consider yourself to be satisfied with your life?.* Figure 1 shows its distribution: the mean reported life satisfaction reaches a value of 7.54 over 10, the standard deviation being of 1.72 points.

[INSERT HERE FIGURE 1]

In addition, we test the robustness of our results by repeating the whole analysis with the self-reported happiness, which also ranges between 0 (completely unhappy) and 10 (completely happy). The long lasting effects that happiness has over personal wellbeing remain unclear, as it is usually related to external shocks such as unemployment, divorces or deaths of relatives (Oswald and Powdthavee, 2008), so we employ life satisfaction as our preferred dependent variable.[4]

2.2 Intergenerational mobility variables.

As discussed in the introduction, the literature has found inconclusive results on the relation between personal wellbeing and intergenerational mobility. On the one side, some authors claim that both factors are non-connected (Zhang and de Graaf, 2016; Iveson and Deary, 2017), so other aspects like health, consumption capabilities and future prospects are to be blamed for the different welfare levels achieved by the individuals. On the other side, other authors find the relation to be significant but disagree on whether the effects of mobility over wellbeing are permanent or transitory. In this stream of the literature, Clark and Oswald (1996) and Miles and Rossi (2007) provide evidence supporting the comparison theory: individuals always bear the psychological rewards/punishments of improving/worsening their parents' social or economic position. According to this theory, individuals maintain a certain reference group during their lives, which in many cases is composed by their own parents. Achievements and events in life are put in perspective, balancing them with the permanent status of the comparison group. In this context, Guilbert and Paul (2009) show that the psychological self-punishment derived from downward mobility is stronger than the reward obtained from upward mobility. In contrast, the hedonic adaptation theory defended by Di Tella and MacCulloch (2006) proposes that wellbeing

[4] Results employing self-reported happiness as dependent variable remain largely unchanged, and are available upon request.

effects disappear as the individuals get used to their newly achieved status. For instance, in the case of upward mobility, these authors find that despite higher levels of consumption capabilities increase marginal utility, its positive psychological effects dissipate relatively soon.

Measuring intergenerational income mobility requires, by definition, two variables: one that collects fathers' income and another reporting income of the children. Unfortunately, the CIS (2017) database does not include information on fathers' incomes. Overcoming this limitation, we follow the mainstream intergenerational mobility literature (Olivetti and Paserman, 2015; Jerrim et al., 2016) and implement the "Two-Sample Two-Stage Least Squares" (TSTSLS) methodology to impute fathers' incomes from previous surveys (Bjorklund and Jantti, 1997).[5] This technique, despite not being difficult to apply, requires a separated explanation, which is assessed in section 3. Consequently, for now, we only present the main statistics of income reported in the CIS database, this is, children's income.

The literature on intergenerational income mobility has traditionally focused on personal income to analyze the transmission of opportunities from fathers to sons and daughters (Jantti and Jenkins, 2013). Here we propose that individuals' household income should be the variable considered when relating intergenerational income mobility to life satisfaction. Using household income does not only control for assortative mating, but also collects more complete information about consumption and saving capacities within the family unit. Relying on personal income would not account for these aspects of household economies, and could lead to biased results in the context of personal wellbeing. Thereby, in this article we use household income, which includes all sources of income perceived by the household where the respondent lives, net of taxes and transferences.[6] Still, for comparative reasons, we divide the household income by the squared root of the household size, as this is the scale of equivalence method commonly used for inequality studies in Spain (see Cabrera et al., 2020).

Table 1 shows the summary statistics of household per capita adjusted income by age groups. The relative similitude of the mean values and the standard deviations goes in line with the permanent income hypothesis. Regarding income inequality, the Gini coefficient associated with the household income distribution gets to 0.301 points, close to the 0.315 Gini that Ayala (2016) calculated for Spain in 2014.

[INSERT HERE TABLE 1]

[5] Particularly, we employ three waves of the Encuesta de Presupuestos Familiares (EPF): 1980/81, 1990/91 and 2000/01. The details and statistics proving the validity of this imputation are presented in Section 3.
[6] All monetary units used in the paper are adjusted to 2017 euros.

Regarding intergenerational occupational mobility, we focus on the occupational change between the respondents and their fathers.[7] The CIS employs the ISCO-08 classification, providing occupations disaggregated up to the 3-digit level. However, occupational mobility matrixes require an aggregated dimension; the occupational groups should reflect certain professional status so that moving from one group to another may imply substantial changes in a person's wellbeing. With this aim, and following Cabrera et al. (2020), we use the ISCO-08 skills classification to create four different categories that represent diverse skill levels and constitute different class status, working conditions and wage levels. First, we have unqualified workers (ISCO-08=9); second, semi-qualified and qualified laborers (ISCO-08=4-8); third, technicians and support professionals (ISCO-08=3); fourth, managers and professionals (ISCO-08=1-2*)*.

Table 2 presents the summary statistics of the occupational distribution of the respondents and their fathers. The structural changes of the Spanish labor market are evident: while 34.2% of respondents have high skill occupations (ISCO-08=1-3), this ratio only reaches 22.9% for the fathers. Still, the main difference lies in the semi-qualified and qualified workers, as their proportion is much smaller for the respondents than for their fathers.

[INSERT HERE TABLE 2]

Finally, intergenerational educational mobility summarizes how the education of the respondents relates to their fathers'. In the CIS database, the educational categories are defined following the ISCED classification (UNESCO, 2012) but, again, we follow Cabrera et al. (2020) and recode the levels of studies to create four groups. First, those with zero or primary education (ISCED=0-1); second, those with lower-secondary education (ISCED=2); third, those with upper secondary education and post-secondary (ISCED=3-4); forth, those with tertiary education (ISCED=5-8).

[INSERT HERE TABLE 3]

Table 3 presents the summary statistics of the educational distribution of the respondents and their fathers, reflecting the expansion of the compulsory secondary education in Spain. While the share of individuals with just primary education is reduced by more than a half, those with post-secondary education double their proportion. Indeed, upper secondary education presents the biggest improvement: 29.9% of the respondents hold tertiary degrees when only 9.8% of the fathers does.

[7] Considering mothers' occupation would substantially reduce the sample size due to the late incorporation of women into the Spanish labor market, as a big share of the respondents' mothers carried household informal works.

2.3 Control variables.

We also include several control variables that account for social and demographic aspects that the literature has traditionally found to be related to personal wellbeing. In particular, we consider respondents' age and age squared to control for life cycle and its non linearities, gender (binary), health status, marital status (being married or not) and having children (binary). Table 4 shows the summary statistics. The mean age is located at around 45 years, the sample being evenly distributed across men and women. Regarding the self-assessed health status, on average, individuals situate themselves at 3.9 on a scale from 1 (very bad) to 5 (very good). Finally, while only 60% of our sample is married, individuals have, on average, 1.27 kids.

[INSERT HERE TABLE 4]

Finally, in the econometric analysis, and depending on the type of intergenerational mobility considered, the regressions also include the respondents' current income quintile, occupational group or educational level as controls. This is, when we include educational mobility as an independent variable, we also use the highest level of education achieved by the respondent as a regressor, substituting it by the occupational status when studying the effect of occupational mobility. Hence, the coefficients capturing the impact that different types of intergenerational mobility have on wellbeing are not capturing the effects of the occupational, educational or income group to which the respondent belongs.

3. Methods

Computing intergenerational income mobility requires income information from two cohorts or generations. However, fathers' income is not available in the CIS (2017) nor in any other modern Spanish database that also includes wellbeing variables. This lack of data is prevalent in many developed economies and hinders the study of intergenerational income mobility and its associated factors. Overcoming this limitation, Bjorklund and Jäntti (1997) proposed the TSTSLS methodology based on a two-sample instrumental variable estimator (Angrist and Krueger, 1992). This technique has been repeatedly used in the intergenerational income mobility literature, such as in Cervini-Pla (2015) for Spain, Barbieri et al. (2019) for Italy or Bloise et al. (2020) for the US and South Africa.

The TSTSLS estimation method requires two different samples. The main sample must include data on individuals' current income and fathers' socioeconomic variables like their educational level or occupation. However, as the main sample (the CIS in our case) lacks information on fathers' income, a secondary or auxiliary sample must be employed. Precisely, this sample comes from an earlier survey that contains the same information - income and socioeconomic variables-but for previous cohorts. The main idea of this procedure consists on considering individuals in the auxiliary sample as pseudo-parents, estimating their income conditioned on the selected common set of socioeconomic factors. The resulting fitted income values are then imputed into the main sample by matching the fathers' and pseudo-fathers' socioeconomic information that is present in both surveys.

Formally, consider equation (1):

$$y_i^s = \alpha + \beta y_i^f + \varepsilon_i \tag{1}$$

Where y_i^s is the logarithm of the sons' permanent individual income, y_i^f is the logarithm of fathers' permanent earnings, α is the mean income of sons' and ε_i is an error term that collects individual's income not explained by the fathers'. As the CIS dataset does not include y_i^f, we use the auxiliary sample to estimate the following equation:

$$y_i^{pf} = \varphi + \gamma z_i^{pf} + \delta_i \tag{2}$$

Where y_i^{pf} is individual income of the pseudo-fathers in the auxiliary sample and z_i^{pf} is a vector of time-invariant socioeconomic factors used to predict income. Finally, δ_i is the component of pseudo-fathers' income not explained by the control socioeconomic factors. Equation (2) is estimated by OLS, and then used to predict fathers' income: $\hat{y}_i^{pf} = \hat{\gamma} z_i^{pf}$.

This method poses an extra problem. The vector of estimated coefficients ($\hat{\gamma}$) is estimated with imperfect and incomplete data, as fathers' occupation and educational level are the only variables we have to match both samples and use as regressors in equation (2). Since the exclusion of relevant socioeconomic controls makes the imputation highly dependent on data quality, the resulting fitted values are probably biased.[8] To improve the accuracy of our imputations, we follow Bloise et al. (2020) and apply Statistical Learning methods to increase the precision of our intergenerational income mobility estimates.

Formally, the best possible imputation is obtained when we reduce at a minimum the squared difference between $\hat{y}_i^{pf}$ and y_i^f:

$$min\left\{E\left[\left(y_i^f - \hat{y}_i^{pf}\right)^2\right]\right\} = min\left\{E\left[\left(y_i^f - f\left(z_i^{pf}\right)\right)^2\right]\right\} \tag{3}$$

The expected squared error of equation (3) can be decomposed into three different elements:

$$E\left[\left(y_i^f - \hat{y}_i^{pf}\right)^2\right] = var\left(\hat{f}\left(z_i^{pf}\right)\right) + (bias)^2 + var(\delta_i) \tag{4}$$

8 For a complete formal explanation, see Nybom and Stuhler (2016) or Bloise et al. (2020).

The first term on the left, $var\left(\hat{f}\left(z_i^{pf}\right)\right)$, is the error coming from the sensibility of equation (2) to the random noise in the auxiliary sample. The second term is the bias of the model, which quantifies the error generated by the selection of the variables in the data generation process. The last term is an irreducible error that captures the smallest possible error we must cope with when predicting y_i^{pf}.

By definition, a trade-off exists in equation (4). Very complex models, such as those including all occupational and educational categories as dummies in vector z_i^{pf}, diminish the bias term but increase the variance, leading to a potential over fitting. On the contrary, too simple models that use highly aggregated variables as controls diminish the variance component at the expense of increasing the bias term. Solving this tension, same as Bloise et al. (2020), we estimate equation (2) with the regularization term first introduced by Zou and Hastie (2005). This statistical learning method consists on adding up an extra term to the classical least-square regression, so that the estimated coefficients are obtained by minimizing equation (5):

$$\sum_{i=1}^{n}\left(y_i^{pf}-\sum_{j=1}^{k}\rho_j z_{j,i}^{pf}\right)^2+\lambda\left(\alpha\sum_{j=1}^{k}\left|\rho_j\right|+(1-\alpha)\sum_{j=1}^{k}\rho_j^2\right) \tag{5}$$

The left-hand side term is a canonical OLS element, with all potential k regressors included in the dummy vector $z_{k,i}^{pf}$. The right-hand side element is a regularization term that penalizes over fitting by shrinking some of the estimated coefficients towards zero. The main idea of the algorithm lies on including as much information as possible and, simultaneously, eliminating the coefficients that do not provide meaningful information to minimize equation (3).

Summing up, if equation (5) was estimated including a high number of covariates in vector z, in our case all educational and occupational categories, the regularization term would shrink many ρ_j coefficients to zero, optimizing the predictive capacity and avoiding overfitting. Thus, given the available data, estimating the mincerian equation (3) in this way provides the most accurate possible prediction of fathers' income.

3.1 Auxiliary database

The data for the auxiliary sample comes from the Household Budget Survey (Encuesta de Presupuestos Familiares, EPF) conducted by the Spanish National Institute of Statistics (Instituto Nacional de Estadística, INE). First implemented in 1973, this survey is representative of the Spanish population and collects information on incomes, expenses and a wide range of socioeconomic characteristics of the Spanish Households. The INE carried two other waves in 1980 and 1990 before changing its design to a panel structure in1997.

Even though the literature tends to consider a single wave as the auxiliary sample, we use three waves (1980/81, 1990/91, 2000).[9] Recall that we have restricted the main sample by keeping individuals aged between 30 and 60 years. The CIS data was collected in 2017 and the respondents are retrospectively asked about the fathers' information when they were 16. Thus, if we only used the 1980-81 wave to impute fathers' income, the imputation for younger individuals of the CIS might be biased, as they were not even born in that year. Second, the Spanish economy experienced relevant structural changes during the 80s and 90s, which in turn affected the occupational and wage structures. Since using a single wave would overlook those changes, we use several waves to correct and control for those effects.

Table 5 presents the correspondence between respondents' age in the main sample (CIS) and the auxiliary waves employed to impute their respective fathers' income. Younger cohorts (those aged between 30 and 35) receive their fathers' income imputation from the EPF 2000. Considering that the data was collected in 2017, those who are aged 32 (the median point between the age range 30-35) were 16 years old when the 2000/01 EPF wave was collected. Similarly, middle-aged (36-45 years) and older cohorts (45-60 years) receive, respectively, their fathers' income imputations from the EPF 1990-91 and 1980-81.

[INSERT HERE TABLE 5]

3.2 Imputation.

Once we establish the correspondence between the main sample and the three auxiliary samples, we apply the TSTSLS methodology. Following equation (5), we use the EPF to regress the pseudo-fathers' socioeconomic factors captured in vector $z_{j,i}^{pf}$ (the educational level and occupation) against their reported incomes. The resulting fitted income values, $\hat{y}_i^{pf}$, are then imputed into the CIS by matching fathers' and pseudo-fathers socioeconomic information, which is present in both surveys.[10] However, equation (5) is not only defined with the usual parameters (ρ_j), but also includes a regularization term with other two undefined extra parameters: λ and α.

[9] The INE has traditionally carried out two types of EPFs: the structural or basic ones every eight or ten years (our 1980-81 and 1990-91 surveys) and, since 1997, the quarterly ones (our 2000 survey). For this last surveys and for each year, the INE also provides an annual longitudinal database collecting the corresponding four quarter, but relevant variables like education and occupation are too aggregated. Thus, we use the four quarterly databases, which offers more disaggregated classifications.

[10] Occupational classifications have suffered several updates. Thus, we convert CNO-79 and CNO-94 (EPF's classifications) into CNO-2011 (CIS's classification) using the correspondence tables published by the INE. Occupation is reported at a two-digits level. We also recode the educational categories into six levels: illiterates, primary, secondary (first stage), secondary (second stage), professional formation and tertiary education. Detailed information is available upon request.

The values of λ and α should not be arbitrarily selected. Indeed, if their setting was left to the researchers' criteria, they could easily affect the quality of the imputation by implicitly leading to the exclusion of more or less regressors, artificially shrinking their coefficients towards zero. Avoiding exogenous alterations on the imputation, the proposed algorithms compute all possible tunings and combinations of λ and α to finally select the one that delivers the smallest Mean Squared Errors. To make their tuning completely transparent, the interested reader may find further information in the Technical Appendix.

Table 6 presents the summary statistics of the income imputations performed for each EPF wave. While the mean imputed income is similar in the three waves, the dispersion is reduced over time, in line with the declining inequality described in Ayala (2016). Once we estimate the vector of income for each wave, and taking Table 5 as a reference, we impute those values to each cohort in the CIS database by matching the occupation and educational level of the fathers and pseudo-fathers. This last step completes the imputation, as fathers have received their correspondent imputed income in the main sample.

[INSERT HERE TABLE 6]

4. Intergenerational Mobility in Spain

Thanks to the income imputation, we can now proceed to study intergenerational income mobility in Spain by means of a transition matrix. Table 7 tabulates the quantiles of fathers' imputed income (rows) against the quintiles of household adjusted income (columns), and reports intergenerational persistence (observations that remain in the main diagonal), upward (observations situated above the main diagonal) and downward mobility (those below the main diagonal).

[INSERT HERE TABLE 7]

Around 36% of our sample experienced upward mobility, while a similar proportion suffered downward mobility. However, when we measure relative mobility, same as Cervini-Pla (2015), we find a strong persistence at the tails of both distributions. Up to 33.63% (76/226) of the richest fathers have children that stay in the same quintile, whereas only 13.27% (31/226) have descendants in the lowest quintile. On the contrary, while one third of low income fathers (71/234) have low income children, only 13.25% (31/234) of them have children in the fifth quintile.

Income is not the only factor affecting life satisfaction, so we get a more comprehensive picture by addressing other types of mobility. As explained, sociologists have traditionally used professional occupations as a proxy for social class, as apart from representing a professional status or skill level, it also reflects social recognition. In our context, individuals are expected to experience psychological benefits from climbing the social ladder and accessing a profession with more social recognition than that of the father. On the contrary, intergenerational occupational downward mobility could be related to lower levels of life satisfaction as a result of the failure to meet previously-determined expectations.

Following this logic, we measure occupational mobility. Table 8 tabulates the professional status of the fathers (rows) against that of the children (columns). In line with the literature, we find a strong intergenerational occupational persistence (EU, 2018). Around 50.4% of our sample has a job with a qualification requirement similar to that of their fathers, while 27.1% experience upward mobility and 22.5% downward mobility. Moreover, when we analyze relative mobility, we find that fathers' position largely conditions the occupation of his descendent. From fathers who participated in the highest occupational level, 37.3% (59/158) have descendants in that same category, but only 5.1% (8/158) have children in the less qualified jobs. By contrast, while just 13.6% (9/66) of low qualified fathers have descendants who reach the highest occupational level, around one third (20/66) of their children remain in the lowest category.

[INSERT HERE TABLE 8]

Finally, we study intergenerational educational mobility, as the literature has shown that having access to a wider variety of forms of cultural consumption and leisure are also related to higher personal welfare (Torche, 2015; Schuck and Steiber, 2018). Table 9 presents the transition matrix of education, showing that around half of the sample (50.8%) experienced upward mobility, while only 7.6% have less education than their fathers. Indeed, absolute mobility ratios are encouraging, but, once again, the relative mobility analysis highlights the strong persistence of the educational levels between generations and the unequal opportunities that hide behind these results. While 67.9% (95/140) of highly educated fathers have kids with the same educational level, this ratio descends to a mere 18.1% (139/770) when we consider fathers with primary or lower studies and highly educated children. Clearly, upward educational mobility has not been homogeneously distributed among the Spanish population.

[INSERT HERE TABLE 9]

5. Results

This section delves deeper into the relation between personal wellbeing and intergenerational mobility. First, we apply the new graphical tools presented in Jenkins (2019a, 2020) to analyze the bivariate connection between life satisfaction and intergenerational mobility. After that, we get a deeper understanding by running several regressions and robustness checks that control for several sociodemographic factors.

Being interested on whether personal wellbeing is higher for individuals who experience a given type of intergenerational mobility, a direct approach would be to graphically check for distributional dominances. However, as explained in Jenkins (2019a), distributional comparisons for ordinal variables cannot be undertaken with the same methods commonly applied to cardinal variables, as the mean is not order-preserving under scale changes. The analytical tools proposed by the author adapt the dominance checks traditionally used for comparisons of income distributions to ordinal variables, such as life satisfaction and other self-reported wellbeing measures.

Since these techniques are based on Cumulative Distribution Functions (CDFs), we compute the CDFs of life satisfaction for each of the three different groups we have created in the previous section: immobile, upward and downward moving individuals. These Figures are separately computed for income, occupational and educational mobility to detect whether further differences exist between the three mobility types.

The graphical analysis proposed by Jenkins (2019a, 2020) allows ranking two distributions without making strong assumptions about the nature of the social welfare function The author distinguishes two types of dominances, each one having a different interpretation. On the one hand, the F-dominance checks whether the variable of interest, in our case wellbeing, is higher (or lower) for a certain group. Formally, following the First-Order Stochastic Dominance criteria, wellbeing distribution A F-dominates wellbeing distribution B if and only if the CDF_A lies nowhere above the CDF_B. On the other hand, the S-dominance shows which distribution is more unequal, understanding inequality in this context as having a greater spread away from the median. Formally, distribution A S-dominates B (A is more evenly distributed) if two conditions are fulfilled. First, both distributions share the same median value m. Second, for all categories $k<m$, the CDF_B is nowhere below the CDF_A, while for all categories $k \geq m$, the CDF_B is nowhere above the CDF_A. In other words, distribution A has more of density mass concentrated in the extremes. Note that, by definition, F-dominance and S-dominance are incompatible: the CDFs cannot cross for F-dominance, but they must cross once for S-dominance.

Focusing on F-dominance results, we study whether wellbeing is higher or lower for a certain group. Figure 2 shows the CDFs of life satisfaction for each of the three categories of intergenerational income mobility.[11] We find that upward mobility F-dominates immobility, as the former distribution is always below the latter, meaning that those who experience upward income mobility present higher levels of life satisfaction than the immobile. However, no F-

[11] Life satisfaction is rescaled (+ 1) so that the information used for the dominance checks can be created with "ineqord" stata command (Jenkins, 2019b).

dominance is found between upward and downward income mobility, as their curves cross at the right tail of the distribution.

[INSERT FIGURE 2 HERE]

[INSERT FIGURE 3 HERE]

[INSERT FIGURE 4 HERE]

Similarly, Figures 3 and 4 respectively display the CDFs of life satisfaction for the different categories of intergenerational occupational and educational mobility. No F-dominances are found. The relation between the groups is not so clear now, as the CDFs cross at different points and mislead the analysis.

Now we turn to the S-dominance, and check whether some groups have more equal or unequal wellbeing distributions. First, we calculate the median life satisfaction. It equals 9 (remember it is rescaled) in all mobility groups except for the educationally downward moving individuals, whose median is 8. Second, for the groups with the same median value, we check whether the conditions for S-dominance are met: i.e., their CDFs only cross at the median and one group is more clustered around the median (less concentrated in the extremes). Coming back to Figures 2, 3 and 4, only upward educational mobility S-dominate educational immobility. Life satisfaction is more evenly distributed among individuals who climb the educational ladder than the immobile. We find no S-dominances for the other mobility groups. Either their CDFs cross more than once, or they cross at a point different from the median, or no CDF shows more concentration in the extremes. We can simply conclude that none of the remaining groups has a more equal welfare distribution. All in all, the graphical analysis suggests that, in general, personal wellbeing is neither higher nor more unequally distributed for a given type of intergenerational mobility. To providing further evidence, we perform a more precise econometric analysis by running several OLS regressions and controlling for other factors that might also affect personal wellbeing.[12] It is necessary to remark that although we cannot make strong causal relations, we can still interpret and explain the significant (and non-significant) relations among these variables.

All Tables presented below include four different models, each one accounting for a different potential effect of intergenerational mobility on personal wellbeing. In particular, Model 1 includes mobility as a discrete variable to measure its direct impact, where -1 represents downward mobility, 0 immobility and 1 upward mobility. Model 2 considers the intensity of the

[12] Despite being ordered probit regressions the most accurate model, the large number of categories that are included in the dependent variable (defined from 0 to 10) make the interpretation of the results quite cumbersome. Thus, our analysis is based on traditional OLS regression, but we have checked that the results are robust to using ordered probit regressions.

mobility experienced, this is, the magnitude of the movement. This measure is constructed with the number of ladders ascended or descended between generations. For instance, education is classified in four levels, so respondents who achieve the highest educational level and have a low educated father are assigned a value of 3 (they ascend three steps in the educational ladder), while those whose fathers attended upper secondary education but are low educated receive a -2 (they descend two steps). Model 2 assumes that the effect of intensity is linear in mobility, this is, homogeneous and independent from the number of steps climbed or descended. Model 3 broadens the analysis and accounts for potential non-linearities by squaring the intensity variable defined for Model 2. Finally, Model 4 includes two dummies, one for upward and the other for downward mobility, the immobility status being the omitted category. This final model is used to disentangle whether the effects found in Models 1, 2 or 3, if significant, are caused by those who improve or worsen their situation.

Table 10 relates life satisfaction and intergenerational income mobility, showing that their relation is never significant. Contrary to the comparison theory, intergenerational income mobility is not statistically related to wellbeing. If mobility had a short-term impact on life satisfaction, as suggested by the hedonic adaptation theory, our results show that the aggregate effect is non-lasting and dissipates with time. The four different specifications confirm this result, which is similar to that found in Zhang and de Graaf (2016) and Iveson and Deary (2017).

[INSERT HERE TABLE 10]

The coefficients, signs and significance of the remaining control variables are robust among the different specifications and present no surprising results. Gender, age and having kids are never relevant for life satisfaction. On the contrary, and in line with the literature, having good health is one of the variables that contribute the most to personal wellbeing (Deaton and Arora, 2009; Iveson and Deary, 2017). In general, healthy people are less likely to report sadness, physical pain, stress and anger, which are emotions related to lower levels of life satisfaction. Same as Hamermesh (2020), the effect of being married over life satisfaction is significant and positive, probably due to the emotional stability derived from marriage. Finally, individuals in the third and fourth income quantiles, followed by those at the fifth and second quantiles, experience higher levels of life satisfaction than those located in the first quantile.

Tables 11 and 12 respectively analyze the effect of educational and occupational mobility on life satisfaction, reinforcing the robustness of our previous results. No mobility measure has a significant relation with life satisfaction, and the control factors generally maintain their sign, size and significance, with one remarkable exception: no category of occupation nor education is significantly related to life satisfaction. Only the income quantile has an effect on personal wellbeing, as found in Table 10. It seems that once individuals cover some basic necessities, including consumption, economic security, health or the emotional stability obtained from marriage, other factors related to social status (proxied by the occupational and the educational level), are not directly related to personal welfare.

[INSERT HERE TABLE 11]

[INSERT HERE TABLE 12]

6. Conclusions

This paper explores the relation between personal wellbeing and intergenerational mobility in Spain (2017). First, we apply innovative Statistical Learning techniques to overcome data limitations and estimate intergenerational income mobility. Second, we broaden the analysis by also considering intergenerational educational and occupational mobility. Third, we employ the analytic tools recently proposed by Jenkins (2019a) to make distributional comparisons of our welfare variable (life satisfaction) and different mobility groups (immobility, upward and downward mobility). Finally, we run several OLS regressions to control for the classical sociodemographic factors affecting wellbeing.

The literature has traditionally focused on exploring factors that have a long-lasting effect on life satisfaction, such as, gender, the migration status or health. In this context, we find that no measure nor approach to intergenerational mobility is permanently associated with higher nor lower levels of personal wellbeing. This finding contradicts the comparison theory, which proposes that individuals permanently cope with the emotional benefits or punishments derived from intergenerational mobility. In line with the literature, we find other variables like enjoying good health and being married, to be positively associated with higher wellbeing levels. All in all, it seems that once individuals cover some basic necessities, including economic and emotional stability or good health, other immaterial aspects like intergenerational mobility are not important in defining their general welfare.

Disentangling and understanding the effects that different factors have on personal wellbeing is necessary to promote policies aimed at improving general welfare. So far, studies relating intergenerational income mobility and personal wellbeing have been scarce, probably due to the lack of valid data. Being the imputation methods we employ easily applicable to other countries with incomplete databases, future research should go beyond the traditional analysis of Anglo-Saxon countries. Many questions remain open: is the relation between intergenerational mobility and wellbeing in other Mediterranean countries similar to Spain? Are there any short-term effects? Has the non-persistent relation changed through time? These questions are left for further research.

References

Angrist, J. D. and Krueger, A. B. (1992) “The Effect of Age at School Entry on Educational Attainment: An Application of Instrumental Variables with Moments from Two Samples”, *Journal of the American Statistical Association*, 87, 328-336.

Ayala, L. (2016). *La Desigualdad En España: Fuentes, Tendencias y Comparaciones Internacionales*. Madrid: FEDEA.

Barbieri, T., Bloise, F. and Raitano, M. (2019) “Intergenerational Earnings Inequality: New Evidence from Italy”, *Review of Income and Wealth*, https://doi.org/10.1111/roiw.12420

Betancort, M., Darias, S., Cabrera, L., Nieves, C., Marrero, G.A., Rodríguez, J.G., and Sanchez, D. (2019): “Inequality of Opportunity in an Outermost Region: The Case of Canary Islands”. *Island Studies Journal*, 14: 23-42.

Bjorklund, A. and Jantti, M. (1997) “Intertenerational Income Mobility in Sweden Compared to the United States”, *American Economic Review*, 87, 1009-1018.

Bloise, F., Brunori, P. and Piraino, P. (2020) “Estimating Intergenerational Income Mobility on Sub-Optimal Data: A Machine Learning Approach”, ECINEQ Working Paper Series, no 526.

Bukodi, E. and Goldthorpe, J. H. (2011) “Class Origins, Education and Occupational Attainment in Britain: Secular Trends or Cohort-Specific Effects?”, *European Societies*, 3, 347-375.

Cabrera, L., Marrero, G. A., Rodríguez, J. G. and Salas-Rojo, P. (2020) “Inequality of Opportunity in Spain: New Insights from New Data”, *forthcoming in Review of Public Economics*.

Cervini-Pla, M. (2015) “Intergenerational Earnings and Income Mobility in Spain”, *Review of Income and Wealth*, 61, 812-828.

Chen, Y. and Fu, D. (2015) “Measuring Income Inequality Using Survey Data: the Case of China”, *Journal of Economic Inequality*, 13, 299-307.

Clark, A. E., Dienere, E., Georgellis, Y. and Lucas, R. E. (2008) “Lags and Leads in Life Satisfaction: A Test of the Baseline Hypothesis”, *The Economic Journal*, 118, 222-243.

Clark, A. E. and Oswald, A. J. (1996) “Satisfaction and Comparison Income”, *Journal of Public Economics*, 61, 359-381.

Deaton, A. and Arora, R. (2009) “Life at the Top: the Benefits of Height”, *Economics and Human Biology*, 7, 133-126.

Di Tella, R. and MacCulloch, R. (2006) “Some Uses of Happiness Data in Economics”, *Journal of Economic Perspectives*, 20, 25-46.

European Union (2016) “Analytical Report on Subjective Well-Being”, Statistical Working Papers Eurostat, no KS-TC-16-005.

European Union (2018) “Inter-generational Mobility in Social Status, Education and Occupation”, *Special Eurobarometer*, 471, Summary (Chaper IV).

Guilbert, D. and Paul, S. (2009) “Income and Happiness: An Analysis of Adaptation and Comparison Income Effects”, *HILDA Survey Research Conference*, 16-17.

Iveson, M. H. and Deary, I. J. (2017) “Intergenerational Social Mobility and Subjective Wellbeing in Later Life”, *Social Science and Medicine*, 188, 11-20.

Isen, A., Rossin-Slater, M. and Walker, R. (2017) "Relationship Between Seasons of Birth, Temperature Exposure and Later Life Wellbeing", *Proceedings of the National Academy of Sciences*, 114, 13447-13452.

Hadjar, A. and Samuel, R. (2015) "Does Upward Social Mobility Increase Life Satisfaction? A Longitudinal Analysis Using British and Swiss Panel Data", *Research in Social Stratification and Mobility*, 39, 48-58.

Hamermesh, D. (2020) "Life Satisfaction, Loneliness and Togetherness, with and Application to Covid 19 Lock-downs", NBER Working Paper, no 27018 (revised version forthcoming in Review of Economics of the Household).

Jantti M. and Jenkins, S. P. (2013) "Income Mobility", Handbook of Income Distribution, 2, 807-935.

Jenkins, S. P. (2019a) "Inequality Comparisons with Ordinal Data", IZA Discussion Paper, no 12811.

Jenkins, S. P. (2019b) "INEQORD: Stata module to calculate indices of inequality and polarization for ordinal data", Statistical Software Components S458723, Boston College Department of Economics.

Jenkins, S. P. (2020) "Comparing Distributions of Ordinal Data", IZA Discussion Paper, no 13057.

Jerrim, J., Choi, A. and Simancas, R. (2016) "Two-Sample Two-Stage Least Squares Estimates of Earnings Mobility: How Consistent Are They?", *Survey Research Methods*, 10, 85-102.

Kline, P. and Santos, A. (2013) "Sensitivity to Missing Data Assumptions: Theory and an Evaluation of the U.S. Wage Structure", *Quantitative Economics*, 4, 231-367.

Mahler, D. G. and Ramos, X. (2019) "Equality of Opportunity in Four Measures of Well-Being", *Review of Income and Wealth*, 65, 228-255.

Michalos, A. C. (2017) "Education, Happiness and Wellbeing", *Connecting the Quality of Life Theory to Health, Well-being and Education*, Springer, New York.

Miles, D. and Rossi, M. (2007) "Learning About One's Relative Position and Subjective Wellbeing", *Applied Economics*, 39, 1711-1718.

Molina, J. A., Navarro, M. and Walker, I. (2011) "Intergenerational Well Being Mobility in Europe", *Kyklos*, 64, 253-270.

Nikolaev, B. and Burns, A. (2014) "Intergenerational Mobility and Subjective Well-Being, Evidence from the General Social Survey", *Journal of Behavioral and Experimental Economics*, 53, 82-96.

Nybom, M. and Stuhler, J. (2016) "Heterogeneous Income Profiles and Lifecycle Bias in Intergenerational Mobility Estimation", *Journal of Human Resources*, 51, 239-268.

Olivetti, C. and Paserman, D. (2015) "In the Name of the Son (and the Daughter): Intergenerational Mobility in the United States, 1850-1940". *American Economic Review*, 105, 2695 – 2724.

Oswald, A. J. and Powdthavee, N. (2008) "Death, Happiness and the Calculation of Compensatory Damages", *The Journal of Legal Studies*, 37, 217-251.

Schuck, B. and Steiber, N. (2018) “Does Intergenerational Educational Mobility Shape the Well Being of Young Europeans? Evidence from the European Social Survey”, *Social Indicators Research*, 139, 1237-1255.

Spade, D. (2020) “Markov Chain Monte Carlo Methods: Theory and Practice”, *Handbook of Statistics*, 43, 1-66.

Stiglitz, J., Sen, A. and Fitoussi, J. (2009) “The Measurement of Economic Performance and Social Progress Revisited: Reflections and Overview. Available at http://www.stiglitz-sen-fitoussi.fr/ documents/overview-eng.pdf.

Steptoe, A., Deaton, A. and Stone, A. A. (2015) “Subjective Wellbeing, Health and Ageing”, *The Lancet*, 385, 640-648.

Torche, F. (2015) “Analyses of Intergenerational Mobility: An Interdisciplinary Review”, *The ANNALS of The American Academy of Political and Social Science*, 657, 37-62.

Varian, H. R. (2014) “Big Data: New Tricks for Econometrics”, *Journal of Economic Perspectives*, 28, 3-28.

Zhao, Y., Li, Y., Heath, A. and Shryane, N. (2017) “Inter and Intra Generational Social Mobility Effects on Subjective Well-Being, Evidence from Mainland China”, *Research in Social Stratification and Mobility*, 48, 54-66.

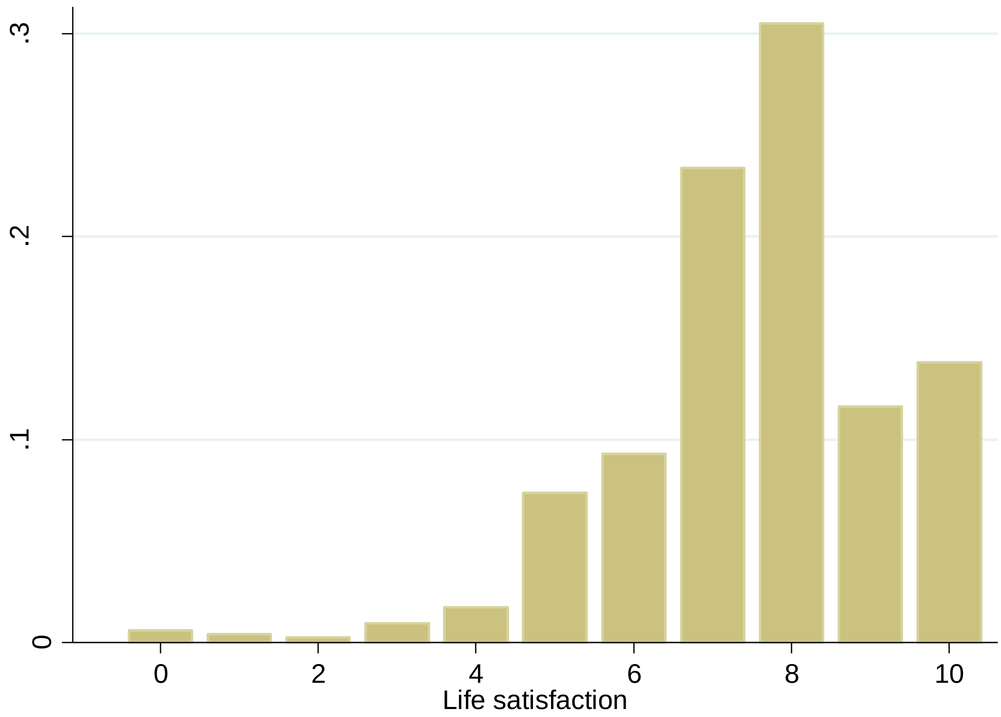
Zhang, E. and De Graaf, N. D. (2016) “Frustrated Achievers or Satisfied Losers? Inter and Intergenerational Social Mobility and Happiness in China”, *Sociological Science*, 3, 779-800.

Zhong, H. (2010) “The Impact of Missing Data in the Estimation of Concentration Index: a Potential Source of Bias”, *European Journal of Health Economics*, 11, 255-266.

Zou, H. and Hastie, T. (2005) “Regularization and Variable Selection Via the Elastic Net”, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67, 301-320.

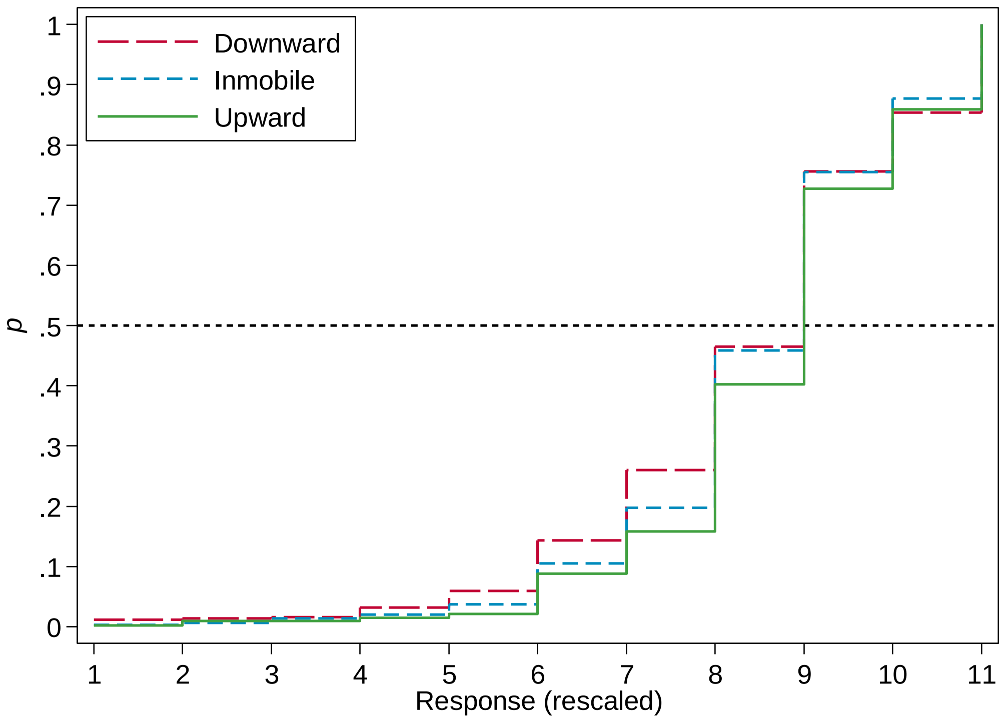
Figures

Figure 1: Density of life satisfaction



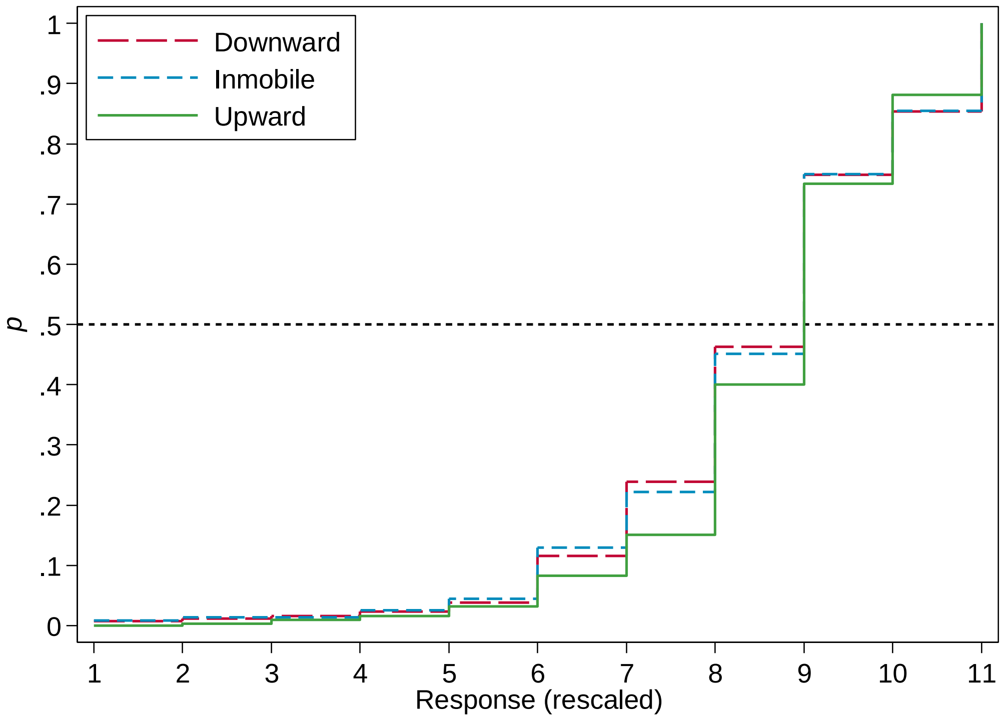
Note: Source: CIS 2017

Figure 2: Cumulative Density Function of Life Satisfaction by Income Mobility



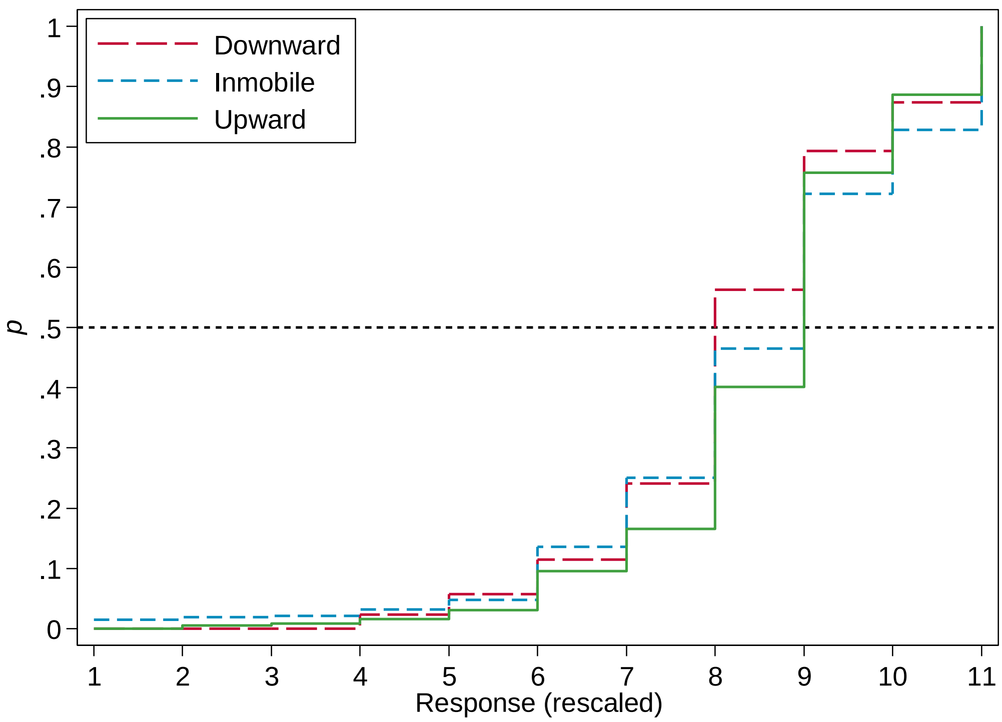
Note: Source: CIS 2017

Figure 3: Cumulative Density Function of Life Satisfaction by Occupation Mobility



Note: Source: CIS 2017

Figure 4: Cumulative Density Function of Life Satisfaction by Education Mobility



Note: Source: CIS 2017

Tables

Table 1: Summary Statistics of the Household per capita adjusted income, by cohorts.

Age	Mean household adjusted income	Sd. of the household adjusted income
30-40	23,777.26	13,442.23
41-50	24,974.19	13,149.09
51-60	22,498.83	13,858.61

Note: Sd. Stands for Standard Deviation. All values in €2017. Source: CIS 2017

Table 2: Summary Statistics of the occupation of the respondents and their fathers

	Respondents	Fathers
ISCO-08=9, Unqualified workers (1)	12.16%	5.73%
ISCO-08=4-8, Semi-qualified and qualified workers (2)	53.61%	71.42%
ISCO-08=3, Technicians and support professionals (3)	14.16%	9.12%
ISCO-08=1-2, Managers and professionals (4)	20.07%	13.73%
Mean	2.42	2.30
Standard Deviation	0.94	0.78

Note: Source: CIS 2017

Table 3: Summary Statistics of the education of the respondents and their fathers.

	Respondents	Fathers
ISCED=0-1, Primary education (1)	29.63%	66.90%
ISCED=2, Low secondary education (2)	13.12%	11.12%
ISCED=3-4, Upper secondary education (3)	29.89%	9.82%
ISCED=5-8, Post-secondary (4)	27.37%	12.16%
Mean	2.54	1.67
Standard Deviation	1.18	1.07

Note: Source: CIS 2017

Table 4: Summary Statistics of the control variables.

	Mean	Standard Deviation
Age	45.05	8.33
Gender (Men=1)	0.50	0.50
Health Status	3.91	0.81
Civil Status (Married=1)	0.60	0.49
Kids (Have kids=1)	1.27	0.45

Note: Health Status ranges from 1 (very bad) to 5 (very good). Source: CIS 2017

Table 5: Relation between the age of the respondent and wave employed to impute fathers' income.

	Group 1	Group 2	Group 3
Age of the respondent in the main sample	30-35	36-45	45-60
EPF wave	2000	1990-91	1980-81
Year when the age-median observation had 16	2000	1992	1980

Note: Source: CIS 2017

Table 6: Summary statistics of the imputed fathers' income for the three EPF waves.

Wave	1980-1981	1990-1991	2000
Number of observations	14,987	15,567	15,567
Log(λ*)	-7.2844	-6.5427	-5.8132
Mean income imputed with α=1	23,205.84	24,333.60	24,236.88
Sd of income imputed with α=1	7,829.52	6,046.32	5,462.52

Note: Sd stands for Standard Deviation. All monetary values in €2017Source: EPF 1980/81, 1990/91, 2000.

Table 7: Intergenerational income transition matrix.

		Household per capita adjusted income					
		Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5	Total
Fathers income	Quintile 1	71	51	51	30	31	234
	Quintile 2	59	49	41	49	32	230
	Quintile 3	46	42	54	40	45	227
	Quintile 4	43	49	50	44	48	234
	Quintile 5	31	24	49	46	76	226
	Total	250	215	245	209	232	1,151

Note: Source: CIS 2017

Table 8. Transition matrix of occupation.

		Respondents occupation				
		ISCO-08=9, Unqualified workers	ISCO-08=4-8, Semi-qualified and qualified workers	ISCO-08=3, Technicians and support professionals	ISCO-08=1-2, Managers and professionals	Total
Fathers occupation	ISCO-08=9, Unqualified workers	20	31	6	9	66
	ISCO-08=4-8, Semi-qualified and qualified workers	105	478	103	136	822
	ISCO-08=3, Technicians and support professionals	7	48	23	27	105
	ISCO-08=1-2, Managers and professionals	8	60	31	59	158
	Total	140	617	163	231	1,151

Note: Source: CIS 2017

Table 9. Transition matrix of education.

		Education of the respondents				
		ISCED=0-1, Primary education	ISCED=2, Low secondary education	ISCED=3-4, Upper secondary education	ISCED=5-8, Post-secondary education	Total
Education of the fathers	ISCED=0-1, Primary education	314	104	213	139	770
	ISCED=2, Low secondary education	17	21	48	42	128
	ISCED=3-4, Upper secondary education	8	17	49	39	113
	ISCED=5-8, Post-secondary education	2	9	34	95	140
	Total	341	151	344	315	1,151

Note: Source: CIS 2017

Table 10: Life Satisfaction and Income Mobility

VARIABLES	(1) Model 1	(2) Model 2	(3) Model 3	(4) Model 4
Mobility	0.040			
	(0.066)			
Mob. Intensity		0.022		
		(0.034)		
Mob. Intensity (squared)			-0.003	
			(0.011)	
Up Mob.				-0.053
				(0.122)
Down Mob.				-0.134
				(0.128)
Sex	0.032	0.032	0.032	0.029
	(0.094)	(0.094)	(0.094)	(0.094)
Age	-0.095	-0.095	-0.096	-0.094
	(0.061)	(0.061)	(0.061)	(0.061)
Age squared	0.001	0.001	0.001	0.001
	(0.001)	(0.001)	(0.001)	(0.001)
Health	0.550***	0.550***	0.549***	0.551***
	(0.068)	(0.069)	(0.069)	(0.068)
Married	0.566***	0.566***	0.573***	0.571***
	(0.118)	(0.118)	(0.117)	(0.118)
Kids	-0.054	-0.054	-0.053	-0.050
	(0.122)	(0.122)	(0.122)	(0.123)
Income quantile 2	0.354**	0.351**	0.364**	0.358**
	(0.169)	(0.172)	(0.174)	(0.170)
Income quantile 3	0.518***	0.510***	0.537***	0.523***
	(0.152)	(0.157)	(0.155)	(0.153)
Income quantile 4	0.509***	0.495***	0.545***	0.514***
	(0.169)	(0.177)	(0.160)	(0.169)
Income quantile 5	0.437**	0.421**	0.493***	0.432**
	(0.178)	(0.195)	(0.157)	(0.178)
Constant	7.405***	7.426***	7.397***	7.441***
	(1.350)	(1.349)	(1.351)	(1.351)
Observations	1,151	1,151	1,151	1,151
R-squared	0.147	0.147	0.147	0.148

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: Source: CIS 2017

Table 11: Life Satisfaction and Occupational Mobility

VARIABLES	(1) Model 1	(2) Model 2	(3) Model 3	(4) Model 4
Mobility	-0.053			
	(0.088)			
Mob. Intensity		-0.031		
		(0.059)		
Mob. Intensity (squared)			-0.004	
			(0.031)	
Up Mob.				0.063
				(0.141)
Down Mob.				0.153
				(0.141)
Sex	0.049	0.049	0.048	0.050
	(0.093)	(0.093)	(0.093)	(0.093)
Age	-0.086	-0.086	-0.084	-0.090
	(0.061)	(0.061)	(0.061)	(0.061)
Age squared	0.001	0.001	0.001	0.001
	(0.001)	(0.001)	(0.001)	(0.001)
Health	0.573***	0.573***	0.575***	0.573***
	(0.069)	(0.069)	(0.069)	(0.069)
Married	0.612***	0.611***	0.605***	0.610***
	(0.121)	(0.121)	(0.119)	(0.120)
Kids	-0.029	-0.029	-0.030	-0.032
	(0.122)	(0.122)	(0.122)	(0.122)
Qualified occup.	0.006	-0.008	-0.036	0.069
	(0.185)	(0.179)	(0.175)	(0.199)
Technicians	-0.001	-0.025	-0.074	-0.012
	(0.226)	(0.215)	(0.196)	(0.226)
Directives and professionals	0.319	0.309	0.239	0.318
	(0.234)	(0.235)	(0.187)	(0.235)
Constant	7.320***	7.332***	7.338***	7.325***
	(1.377)	(1.375)	(1.373)	(1.377)
Observations	1,151	1,151	1,151	1,151
R-squared	0.136	0.136	0.136	0.137

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: Source: CIS 2017

Table 12: Life Satisfaction and Educational Mobility

VARIABLES	(1) Model 1	(2) Model 2	(3) Model 3	(4) Model 4
Mobility	0.108			
	(0.082)			
Mob. Intensity		0.043		
		(0.043)		
Mob. Intensity (squared)			0.010	
			(0.017)	
Up Mob.				0.059
				(0.128)
Down Mob.				-0.190
				(0.188)
Sex	0.055	0.054	0.054	0.057
	(0.094)	(0.094)	(0.094)	(0.094)
Age	-0.091	-0.089	-0.089	-0.091
	(0.062)	(0.062)	(0.062)	(0.062)
Age squared	0.001	0.001	0.001	0.001
	(0.001)	(0.001)	(0.001)	(0.001)
Health	0.574***	0.575***	0.575***	0.574***
	(0.069)	(0.069)	(0.069)	(0.069)
Married	0.609***	0.611***	0.614***	0.610***
	(0.119)	(0.119)	(0.119)	(0.119)
Kids	-0.035	-0.028	-0.025	-0.034
	(0.122)	(0.122)	(0.122)	(0.122)
Compulsory Secondary	-0.041	-0.002	0.016	-0.000
	(0.197)	(0.188)	(0.185)	(0.217)
Post-compulsory Secondary	0.004	0.023	0.059	0.043
	(0.144)	(0.144)	(0.137)	(0.161)
Tertiary	0.100	0.104	0.140	0.127
	(0.148)	(0.156)	(0.152)	(0.157)
Constant	7.426***	7.378***	7.347***	7.425***
	(1.392)	(1.393)	(1.394)	(1.392)
Observations	1,151	1,151	1,151	1,151
R-squared	0.135	0.134	0.134	0.135

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: Source: CIS 2017

Technical Appendix

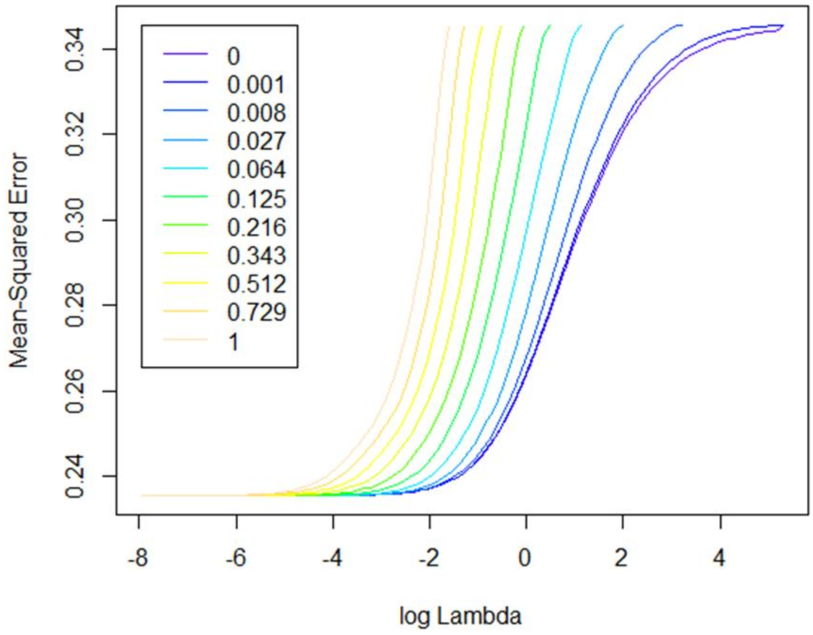
As said, the values of λ and α in equation (5) should not be arbitrarily selected. The former (λ), controls the importance of the regularization term and is equal or higher than zero. The latter (α), is the elastic net regulator obtained from a linear combination of two standard Statistical learning techniques, and its possible values range between 0 and 1. In particular, equation (5) is equivalent to the Least Absolute Shrinkage and Selection Operator (LASSO) when α=1, but it is equal to a ridge regression when α=0 (see Varian, 2014). Different combinations of these parameters might lead to different imputations, leading to non-robust results. Solving this problem, the proposed algorithms compute all possible tunings and combinations of λ and α to finally select the one that delivers the smallest Mean Squared Errors (MSE).

To make the tuning as transparent as possible, we plot the relation between the MSEs and several values of λ and α. This way, we show that our tuning provides the smallest possible associated MSEs. Figures TA1 to TA3 correspond to the imputations performed on the EPF 1980/81, EPF 1990/91 and EPF 2000, respectively.

Figure TA1 should be interpreted as follows. The MSE produced by equation (5) is stable at less than 0.24 for any α value lower than log (λ) ≈ -5. At that point, the MSE associated with α = 1 starts rising, while the MSEs associated with the rest of possible α values remain constant. Clearly, for values of α smaller than 1, the associated MSEs sequentially take off from log (λ) ≈ -5 onwards, until the MSE produced by the ridge regression (α = 0) rises at log (λ) ≈ -3. Figures TA2 and TA3 are interpreted similarly, with diverse MSEs associated with different parameter settings.

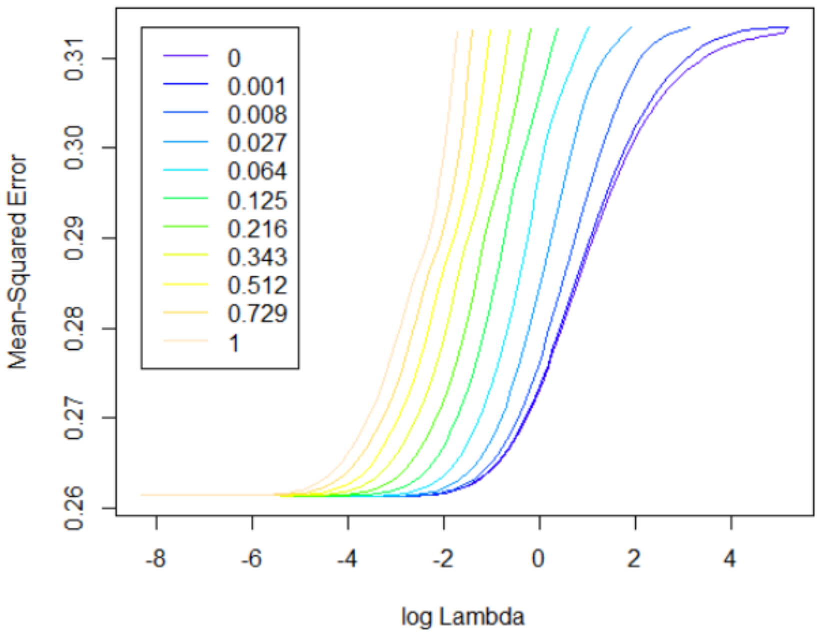
As the algorithm searches for the values of the parameters associated with the lowest stable MSE, it provides the combination of λ and α that leads to the most accurate imputation. For instance, in Figure TA1, although it cannot be graphically distinguished, this occurs when log(λ) equals -7.2844. At that point, no matter the value of α we select, the MSE is constant and has the lowest possible value. For any other combination, the associated MSE is higher (for higher λ) or stable (for lower λ). Similarly, in Figure TA2, the optimal log(λ) equals -6.5427, and in Figure TA3, it reaches -5.8132. Then, these values conform the parameter tuning, as summarized in Table 6, being the selected α equal to 1, as the LASSO regression has been used more often in the literature. However, results for α equal to 0 (and the same lambda values) are available upon request, varying the resulting imputations in around 20€.

Figure TA1: Mean squared error provoked by the regularization term in an OLS regression



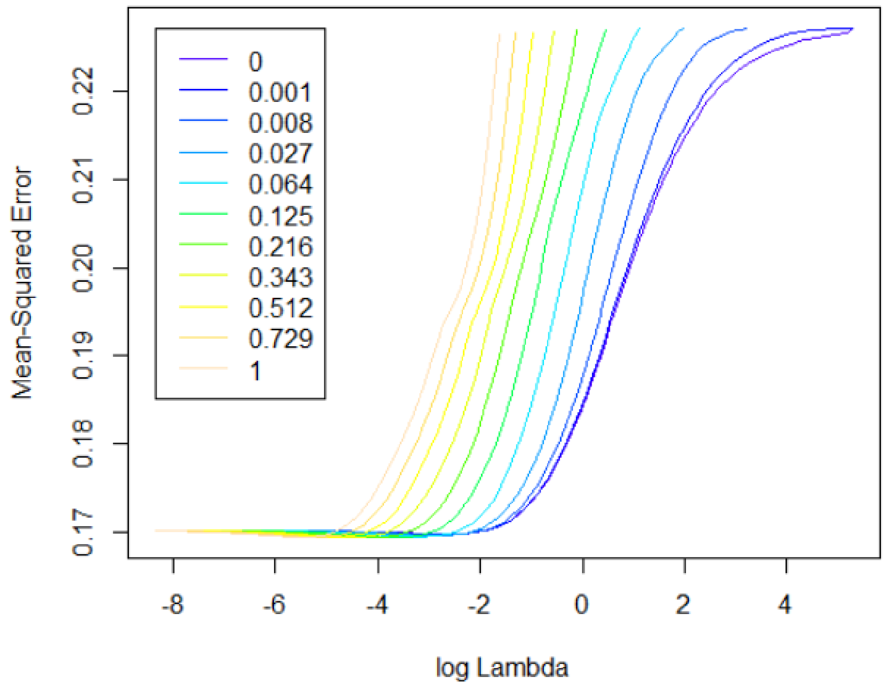
Note: Source: EPF 1980/81.

Figure TA2: Mean squared error provoked by the regularization term in an OLS regression.



Note: Source: EPF 1990/91.

Figure TA3: Mean squared error provoked by the regularization term in an OLS regression



Note: Source: EPF 2000.