UNIVERSITÄT BERN

Faculty of Business, Economics and Social Sciences

Department of Economics

Financial Turmoil and Earnings Mobility

Costanza Naguib

22-08

January, 2022

DISCUSSION PAPERS

Schanzeneckstrasse 1
CH-3012 Bern, Switzerland
http://www.vwi.unibe.ch

Financial Turmoil and Earnings Mobility

Costanza Naguib

First version: December 2016. This version: January 2022

Abstract

We analyze how earnings dynamics changed in the US after the financial crisis of 2007-2009. Differently from most models for earnings mobility, we allow persistence patters to depend semi-nonparametrically on both the past individual position in the distribution and on a set of individual-level covariates. Allowing for more flexibility in the model yields a better fit to the data and permits us to uncover changes in earnings mobility patterns that would otherwise remain hidden. Indeed, at the aggregate level, we find no evidence of changes in individual positional persistence in any part of the earnings distribution after the crisis, both with the parametric and with the semi-nonparametric model. However, the semi-nonparametric copula allows us to uncover an increase in earnings mobility for 45-year-old workers with college degree after the crisis.

Key words: earnings dynamics, positional persistence, financial crisis, functional copula model, semi-nonparametric estimation
JEL codes: C14, J31

1 Introduction

[1] The aim of this paper is to provide evidence of the importance of adopting a flexible modellization for earnings dynamics. We apply both a parametric and a semi-nonparametric copula model for earnings to US data. We run estimations on two subsamples, before and after the financial crisis of 2007-2009, and we compare the mobility patterns found in the two periods with the two methods. We find that the semi-nonparametric model has better fit to the data both before and after the crisis. Since our application is simply based on a before and after identification, we refrain from claiming any causal effect of the financial crisis on recorded changes

[1] I would like to thank Patrick Gagliardini, Christian Gourieroux, Bo Honore, Thierry Kamionka, Igor Letina, Fabrizio Mazzonna, Blaise Melly, Rico Maggi, Alain Monfort, Cyril Monnet, Jean-Marc Robin, Sebastien Roux and Mark Watson for their insightful advice.

in earnings mobility patterns. We uniquely aim at describing mobility patterns before and after the crisis in the US with the parametric and the semi-nonparametric copula models. The aim of this paper, indeed, is to compare the performance of the two models and to provide insights into which type of information can only be obtained by applying a flexible methodology. To summarize, the contribution of this paper is twofold. First, we compare and contrast the performance of econometric methods characterized by different degrees of flexibility. Second, we show that the two-step semi-nonparametric copula model proposed by Naguib and Gagliardini (2020) excels at depicting mobility patterns, i.e. it allows for more flexibility and a better fit to the data.

The present paper is an extension of Bonhomme and Robin (2009). The major difference between that paper and ours is that Bonhomme and Robin (2009) consider a parametric copula family, namely the Plackett copula (Plackett (1965)), while our copula specification is semi-nonparametric. This choice is dictated by our interest in discovering nonparametrically the patterns of dependence between the current and past earnings ranks. The price to pay for this increase in flexibility describing earnings mobility patterns is that the inclusion of transitions into and out of unemployment into our model is computationally intractable[2]. Hence, differently from Bonhomme and Robin (2009), we decide to focus on individuals with a strong labor market attachment (i.e. those who were continuously employed during the analysis sample). In this balanced panel, our model allows to estimate the degree of relative earnings mobility virtually for any individual (i.e. conditional on the value of his/her covariates) at any point of the past rank distribution. This is an improvement over the fully parametric model.

The impact of the financial crisis in the US has been dramatic, with output in 2013 being 13% below its trend path before the crisis (Hall (2015)). Previous literature focused on the impact of the financial turmoil on earnings and employment levels (e.g. Eichhorst et al. (2010), Chodorow-Reich (2013), Dustmann et al. (2014), Christiano et al. (2015)) and on its influence on firms productivity and innovation (e.g. Huber (2018)). However, changes in earnings mobility patterns after the financial crisis have not been documented until now. Our focus lies in *relative* earnings mobility (Shorrocks (1978), Fields and Ok (1999), Formby et al. (2004),

[2]The estimation of the semi-nonparametric copula model implies the computation of several integrals by Monte Carlo simulation. Given the size of the dataset used in the present paper, this implies that a single estimation, which is performed with the software R, requires more than 16 hours working on a server with 16 cores. 500 of such estimates are run simultaneously in parallel in order to obtain our bootstrapped results. Adding a part on transitions into and out of unemployment to the semi-nonparametric copula model would further notably lengthen the required estimation time. Using more efficient estimation packages in R (e.g. nlm, optimx, or DEoptim) only marginally reduces the required computational time.

Bonhomme and Robin (2009), Cowell and Flachaire (2018)), i.e. the dynamics of the workers' positions within the cross-sectional distribution of the residuals from a preliminary earnings regression. The worker's percentile in this cross-sectional distribution is called residual earnings rank.

Relative or rank earnings mobility stands for how do the ranks or relative positions of individuals in the earnings distribution change over time. It is possible that in a given year all earnings rise (absolute earnings mobility), but the relative earnings, i.e. the individual positions in the distribution stay exactly the same (relative earnings mobility). Conversely, a worker's rank may change, even if his/her earnings do not, in case the earnings of the other workers change. Changes in the individual rank have been found to have a significant impact on individual well-being (see e.g. Luttmer (2005), Brown et al. (2008), Clark et al. (2009), Boes et al. (2010), Boyce et al. (2010)), health (Elstad et Al. (2006)), and social capital (Fischer and Torgler (2006))[3], even after controlling for the absolute level of earnings.

In the first step, we obtain earnings residuals from a panel regression with latent classes in the spirit of Bonhomme and Robin (2009) and we compute the corresponding empirical ranks as percentiles of the empirical cross-sectional distribution. Then, we estimate a semi-parametric single-index model for the conditional distribution of the rank at a given date conditionally on the individual characteristics. Finally, we estimate the semi-nonparametric copula model with the method of Sieves. As opposed to most of the existing literature on individual earnings dynamics, we model the dynamics of the ranks of the residuals, instead of modelling the dynamics of the residuals directly. This last option would correspond to the study of absolute earnings mobility (e.g. Hu, Moffitt and Sasaki (2019)). We use both a fully parametric copula (in the spirit of e.g. Bonhomme and Robin (2009)) and a semi-nonparametric copula[4] to estimate relative earnings mobility patterns. In the empirical application, we contrast the estimates from the two models and we provide empirical evidence that our semi-nonparametric approach improves the understanding of relative mobility patterns.

We aim at describing positional mobility patterns in the US before and after the financial disruption of 2007-2009. At the aggregate level, we find no indications of changes in earnings mobility after the crisis. This finding emerges both from descriptive statistics and from both our parametric and semi-nonparametric estimates. However, the semi-nonparametric model allows to estimate the degree of positional persistence (which is the inverse of positional earnings

[3]Social capital is a multidimensional concept introduced in the early '90s by Putnam (1993) which roughly includes compliance with social norms, trust among people and confidence in state institutions.

[4]This model is extensively described in Naguib and Gagliardini (2020).

mobility) for each individual at any point of the earnings distribution at a given point in time. By doing this, we uncover an increase in mobility at the bottom of the residual earnings rank distribution for a specific group of workers, namely, 45-year-old individuals with a college degree, after the crisis. The fully parametric method does not allow, in contrast, for such a precise subgroup analysis, as the shape of the mobility patterns strictly depends on the parametric family of copula chosen. Of course, with the parametric model as well, it is possible to divide the original sample into subsamples of indivduals with the same or similar characteristics (e.g. 35-45-year-old men with college degree) and then to estimate transition matrices in each of these subgroups. However, when using survey data like the Panel Study of Income Dynamics (for the US), in most cases sample size would be too small to perform such an analysis as many cells in the transition matrices will contain less than 50 observations. On the contrary, with the semi-nonparametric copula model such subsetting is not necessary, as the model allows for estimating earnings mobility patterns conditional on the individual covariates.
Finally, when comparing the prediction accuracy of the parametric and of the semi-nonparametric models, we find that the latter has a comparable performance to the former on a two-year time horizon, whereas it outperforms the former on a four-year time horizon. The remainder of the paper is structured as follows. Section 2 presents the semi-nonparametric econometric model. Section 3 is devoted to the description of the dataset and to some exploratory analysis. In Section 4, estimation results of the parametric and semi-nonparametric models are presented and discussed. Section 5 concludes. Appendix A presents the details of the estimation strategy for the semi-nonparametric copula model.

2 A flexible semi-nonparametric model for earnings mobility

The aim of this section is to describe the key features of the analytic framework developed in Naguib and Gagliardini (2020). In particular, this section relies on Sections 2 and 3 in Naguib and Gagliardini (2020). For the theoretical details, as well as for the proofs of theorems and propositions, we refer the interested reader to the above-mentioned companion paper. Our main object of analysis is the panel of the individual residual earnings ranks, i.e. the individual positions in the cross-sectional distribution of the earnings residuals at a given date. In order to obtain the earnings residuals, we first regress log earnings on age and age squared, in order to disentangle the "deterministic" earnings component from the residual component. We also include a time fixed-effect, in order to take into account all the macro shifts, among them also the impact of inflation on earnings. The model reads:

$$Earnings_{i,t} = \alpha_1 Age_{it} + \alpha_2 Age_{it}^2 + \lambda_t + \eta_i + \varepsilon_{i,t} \tag{1}$$

where $Earnings_{i,t}$ stands for log real earnings, Age_{it} stands for individual age, λ_t stands for the time fixed-effect, and $\varepsilon_{i,t}$ is the error term. η_i is a strictly exogenous random effect, independent of the covariates (age) for all t. We assume that this latent variable η_i follows a discrete distribution with K support points, $\eta_1, \eta_2, ...\eta_K$, with respective probabilities $p_1, p_2, ..., p_K$. This approach mirrors the use of latent classes by Bonhomme and Robin (2009)[5]. From the residuals of this preliminary regression, we obtain the Gaussian earnings ranks via the following formula:

$$Z_{i,t} = \Phi^{-1}(F_t(\varepsilon_{i,t})) \tag{2}$$

where ε_{it} are the error terms from (1), η_i are the random effects from (1) and $F_t(\cdot)$ is the cross-sectional distribution of these error terms at each date t. Hence, $Z_{it} \sim N(0,1)$. In the empirical application, we replace the theoretical residuals ε_{it} with their empirical counterparts:

$$\hat{Resid}_{i,t} = Earnings_{i,t} - \hat{\alpha}_1 Age_{it} - \hat{\alpha}_2 Age_{it}^2 - \hat{\lambda}_t - \hat{\eta}_i.$$

Similarly, we replace $F_t(\varepsilon)$ with its estimated counterpart, i.e. the empirical cross-sectional distribution of the estimated residuals. This quantity corresponds to the (individual) percentile in the empirical cross-sectional distribution. We use the function Φ^{-1} to transform the uniform ranks into their standard Gaussian counterparts for ease of interpretation and mathematical modelling. Equation (1) is estimated separately for the two periods, i.e. before and after the crisis. We choose this approach because we deem unrealistic that the effect of age (which is a proxy for experience) on earnings, represented by coefficients α_1 and α_2, remained the same before and after the financial crisis. We do not account for the possible endogeneity of the unemployment patterns, but rather we focus on a sample of individuals with strong labor market attachment (i.e. males between 25 and 55 years old) who where continously employed in at least one of the two periods considered (i.e. pre-crisis and post-crisis). Further, in the wake of Bonhomme and Robin (2009) we assume that the coefficients for age and age squared in eq. (1) are the same for all individuals. This means that our model belongs to the family of the restricted income profiles (RIPs) specifications, sharing this feature with, e.g. Hryshko (2012), Arellano, Blundell and Bonhomme (2017)).

[5] Empirically, eq. (1) has been estimated with the STATA package gllamm. To construct the estimated residuals, the latent variable η_i is replaced by its empirical Bayes predictions ($\hat{\eta}_i$), which have been obtained via the STATA command gllapred.

We now need to model the dynamics of the Gaussian ranks Z_{it}. Throughout the present paper, we define relative earnings mobility as the dynamics of these objects Z_{it}. To start with, let us assume that the pair $(Z_{it}, Z_{i,t-1})$ has copula probability density function (pdf) $c(u, v)$ for arguments $u, v \in [0, 1], u = \Phi(Z_{it}), v = \Phi(Z_{i,t-1})$. A copula is a multivariate distribution with standard uniform marginals; it couples marginal distributions to obtain a joint one. Copulas are used to describe the dependence between random variables. Given that we are interested in the dependence structure between the present and the past residual earnings ranks, copulas are a natural choice. Copulas are useful devices, since they allow tackling the modelling of the joint and that of the marginal distributions as separate problems. For a comprehensive survey of copulas and their properties, we refer the reader to Joe (1997) and Nelsen (1999). To simplify the exposition, we abstract for a moment from the presence of individual covariates. We will introduce individual explanatory variables back later. Then, the conditional expected rank is:

$$\begin{aligned} E[Z_{it}|Z_{i,t-1}] &= \int_{-\infty}^{\infty} zc(\Phi(z), \Phi(Z_{i,t-1}))\phi(z)dz \\ &= \int_{0}^{1} \Phi^{-1}(u)c(u, \Phi(Z_{i,t-1}))du, \end{aligned} \quad (3)$$

where ϕ is the pdf of the $N(0, 1)$ distribution. Our mobility measure for the residual earnings component reads:

$$\frac{\partial E[Z_{it}|Z_{i,t-1}]}{\partial Z_{i,t-1}} = \int_{0}^{1} \Phi^{-1}(u) \frac{\partial c}{\partial v}(u, \Phi(Z_{i,t-1}))du \cdot \phi(Z_{i,t-1}), \quad (4)$$

i.e. the derivative of the conditional expected rank with respect to the past rank. Note that the expression in (4) provides rather an immobility measure, since the larger its value is, the closer the association between the past and the present rank in the distribution is. We select the quantity in (4) as our main measure of rank persistence, which is the inverse of rank mobility. Indeed, the quantity reported in (4) represents how strong the association between the present and the past rank is, depending on the value of the past rank itself. We are interested in studying how this association changes across the cross-sectional rank distribution. This (im-)mobility measure is close in spirit to the nonlinear persistence measure proposed by Arellano, Blundell, and Bonhomme (2017). The easiest way to measure relative mobility in this framework is to use a fully parametric copula in eq. (4). In the following, we resort to the one-parameter Plackett copula, because it has been used by Bonhomme and Robin (2009). The Plackett copula bivariate cumulative distribution function (cdf) is (Plackett (1965)):

$$C(u, v) \equiv C(u, v; \tau) = \frac{1}{2\tau} \left[1 + \tau(u + v) - a(u, v; \tau)^{1/2}\right], \quad (5)$$

where $a(u, v; \tau) = [1 + \tau(u + v)]^2 - 4\tau(1 + \tau)uv$ and $u, v \in [0, 1], u = \Phi(Z_{it}), v = \Phi(Z_{i,t-1})$. The copula parameter τ is a function of individual covariates (age and education). In the empirical part, we first estimate earnings mobility patterns by means of equation (4), by plugging in the copula density corresponding to the bivariate Plackett copula. Second, we replace this fully parametric copula density with a more flexible, semi-nonparametric specification, which is defined in the reminder of the present Section.

Let us relax the assumption that the copula function is fully parametric. In the wake of Naguib and Gagliardini (2020), we construct a new family of copulas, in which the finite-dimensional parameter is replaced by a functional parameter, which in turn is allowed to depend on covariates. As shown by Naguib and Gagliardini (2020), a flexible nonparametric family of copula functions can be written as follows. Let us consider the nonlinear autoregressive dynamics:

$$Z_{i,t} = \Lambda(\rho(Z_{i,t-1}) + \epsilon_{i,t}) \tag{6}$$

where by hypothesis $\epsilon_{i,t} \sim IIN(0, 1)$, Λ is a strictly monotonic increasing function and ρ is a function that expresses the dependence between the past and the present individual ranks. The larger is the value of the partial derivative of the function $\rho(\cdot)$ with respect to the past rank, the higher is the degree of individual positional persistence.[6]. Under the condition that $\Lambda(k)$ is such that

$$\varLambda(k) = \Phi^{-1}\left[\int_{-\infty}^{\infty} \Phi(k - \rho(Z_{i,t-1}))\phi(Z_{i,t-1})dZ_{i,t-1}\right], \tag{7}$$

the invariant distribution of Markov process $(Z_{i,t})$ is $N(0, 1)$. Following Naguib and Gagliardini (2020), the copula density of $Z_{i,t}$ and $Z_{i,t-1}$ reads:

$$c(u, v; \rho(\cdot)) = \frac{\phi[\varLambda^{-1}(\Phi^{-1}(u)) - \rho(\Phi^{-1}(v))]}{\phi(\Phi^{-1}(u))\lambda(\varLambda^{-1}(\Phi^{-1}(u)))}, \tag{8}$$

for the arguments $u, v \in [0, 1], u = \Phi(Z_{it}), v = \Phi(Z_{i,t-1})$. This copula family is parametrized by the autoregressive function $\rho(\cdot)$. We now introduce the vector of regressors, $X_{i,t}$ in the functional parameter indexing the copula, which then becomes $\rho(\cdot, X_{i,t})$. This is admissible because essentially any function $\rho(\cdot)$ can be used to parametrize the autoregressive copula. Hence, the copula pdf will be a function of ρ, Λ and $X_{i,t}$. The only requirement on the exogenous process

[6] We make here the hypothesis that the quantity expressed by equation (4) is bounded between zero and one. This make sense, since this measure essentially represent a sort of correlation between past and present residual earnings ranks. When we estimate quantity (4) in Section 4, we notice that it ranges between zero and one, i.e. our hypothesis is verified empirically in the data.

$(X_{i,t})$ is that it is stationary and Markov. The (im-)mobility measure associated with the semi-nonparametric family of copulas presented above is defined as follows. It is an adaptation of (4) to the flexible model for the joint dynamics of the present and past residual earnings ranks:

$$\frac{\partial E(Z_{it}|Z_{i,t-1}, X_{it}, X_{i,t-1})}{\partial Z_{i,t-1}} = \int_{-\infty}^{\infty} \frac{1}{g\left[G^{-1}\left(\tilde{\Lambda}[\tilde{\rho}(G(Z_{i,t-1}|X_{i,t-1}); X_{it}) + \epsilon; X_{it}]|X_{it}\right)\right]} \cdot \tilde{\lambda}[\tilde{\rho}(G(Z_{i,t-1}|X_{i,t-1})|X_{it}) + \epsilon|X_{it}]\phi(\epsilon)d\epsilon \cdot \tilde{\rho}'(G(Z_{i,t-1}|X_{i,t-1})|X_{it})g(Z_{i,t-1}|X_{i,t-1}). \quad (9)$$

where $\tilde{\rho}(v; X_{it}) = \rho(\Phi^{-1}(v), X_{it})$ and $\tilde{\Lambda}(k; X_{it}) = \int_0^1 \Phi(k - \tilde{\rho}(v; X_{it}))dv$, $\tilde{\lambda} = \tilde{\Lambda}'$ and $\tilde{\rho}'$ is the derivative of function $\tilde{\rho}$ w.r.t. its first argument. In addition to the mobility measure in (9), we can also easily derive the conditional quantiles, which provide us with further information on the conditional distribution of the present rank. Moreover, their partial derivatives with respect to the past rank yield additional measures of rank (im-)mobility. The conditional quantile $Q_{Z,t}(u|Z_{i,t-1}, X_{it}, X_{i,t-1})$ of Z_{it} given $Z_{i,t-1}, X_{it}, X_{i,t-1}$ for percentile $u \in (0, 1)$ is given by:

$$Q_{Z,t}(u|Z_{i,t-1}, X_{it}, X_{i,t-1}) = G^{-1}\left[\tilde{\Lambda}[\tilde{\rho}(G(Z_{i,t-1}|X_{i,t-1}); X_{it}) + \Phi^{-1}(u); X_{it}]|X_{it}\right]. \quad (10)$$

In particular, for $u = 0.5$ we get the conditional median. The relative mobility measure based on quantiles is given by the partial derivative w.r.t. the past rank:

$$m^Q(Z_{i,t-1}, u; X_{i,t}, X_{i,t-1}) = \frac{\partial Q_{Z,t}(u|Z_{i,t-1}, X_{it}, X_{i,t-1})}{\partial Z_{i,t-1}}. \quad (11)$$

The estimation procedure adopted consists in three steps (see Appendix A for further details). In the first one, we estimate the preliminary regression (1) and we construct the empirical ranks $\hat{Z}_{it}$ on the basis of the residuals. In the second step, we estimate the conditional marginal distributions of the present and of the past Gaussian ranks, via a kind of Maximum Likelihood procedure. In order to tackle the curse of dimensionality due to the number of regressors, we adopt an index approach, i.e. we summarize all the individual covariates by their weighted sum, which is called marginal distribution score or index, and then we estimate simultaneously the weights of this sum and the marginal rank distributions conditioned on this index. Then, in the third and final step we estimate the autoregressive function indexing the copula pdf by adopting a Sieve Maximum Likelihood approach. This autoregressive function $\rho(\cdot;\cdot)$ admits two arguments: the individual past Gaussian rank in the distribution and a second index or score, i.e. a weighted sum of individual explanatory variables, whose weights are estimated simultaneously with function ρ itself. We estimate the function ρ non-parametrically with the method of Sieves.

In particular, in our implementation this autoregressive function is approximated via a bivariate Hermite polynomial basis of degree two. To summarize, the copula function presented in (8) takes here the following form:

$$c(\cdot,\cdot,X_{i,t}) = c(\cdot,\cdot,\rho(\cdot,X'_{i,t}\beta)), \tag{12}$$

where we call the index $W_{it} = \beta' X_{it}$ "mobility score".

3 Data and exploratory analysis

The dataset used for analysis has been constructed by appending together waves from 2003 to 2017 of the Panel Study of Income Dynamics (PSID), which is biannual since 1997[7]. In particular, family files for survey years from 2003 to 2017, which correspond to data years from 2002 to 2016, are appended together. In the PSID, a data year is always equal to the corresponding survey year minus one. In the following, we will exclusively refer to data years. Only household heads are kept for the analysis. We assume that panel attrition is random (Fitzgerald and Gottschalk (1998), Lillard and Panis (1998), Meghir and Pistaferri (2004)). However, transitions into and out zero earnings state are most likely not random. We consider data years 2008 and 2010 as affected by the crisis and hence excluded from our analysis. In the main estimations, we focus on a pre-crisis (2002-2006) and on a post-crisis (2012-2016) period. We restrict our sample to male individuals aged between 25 and 55 in each data year. In this way we have a sample of individuals with a strong labor market attachment.

[7] The PSID began in 1968 with two samples: the Survey of Economic Opportunity (SEO) sample focused on low income families, while the Survey Research Center (SRC) sample interviewed a nationally representative selection of families. Members of these households became PSID sample members and were surveyed annually until 1997 (each yearly survey is called a "wave"), after which they were surveyed biannually. Furthermore, all lineal descendants of original sample members become sample members themselves and were independently followed and surveyed once they started their own families. Due to budgetary constraints, in 1997 the PSID dropped approximately 25% of its sample households, with reductions made mainly to the SEO subsample. We limit our sample to the SEO and SRC samples, eliminating individuals from the Immigrant and Latino surveys due to limited data availability.

Table 1: Data description, by calendar year, 25-55-year-old males. In the unbalanced panel, only individuals who recorded at least one non-zero earnings year in at least one of the two periods 2002-2006 and 2012-2016 are included.

t	2002	2004	2006	2012	2014	2016
Age (bal. panel)	38.28	40.22	42.25	36.33	38.39	40.37
Mean log real earnings (bal. panel)	9.92	9.93	9.97	9.81	9.93	10.01
Log earnings variance (bal panel)	0.59	0.53	0.55	0.71	0.63	0.55
Age (unbal. panel)	39.83	39.49	39.44	38.08	38.15	38
Mean log real earnings (unbal. panel)	9.85	9.82	9.82	9.74	9.79	9.83
Log earnings variance (unbal. panel)	0.70	0.68	0.70	0.89	0.82	0.71
Share of zero earnings (unbal. panel)	3.58%	3.26%	3.2%	3.02%	3.18%	3.28%
N. of individuals in balanced panel	1724	1724	1724	1888	1888	1888
N. of individuals in unbalanced panel	2930	2918	2940	3184	3141	3354

The final (unbalanced) dataset thus obtained, excluding 2008 and 2010, includes 18'467 individual-year observations, whereas the average sample size per year is $n = 3078$. Missing years of data (i.e. non reports) are treated as missing-at-random and dropped. On the other hand, reported zero earnings are treated as individuals being unemployed. These observations are included in the descriptive statistics if the individual reports at least one non-zero earnings year in at least one of the subperiods of the analysis. However, these observations are excluded from the main model estimations (Section 4). Note that a person has to be inactive for an entire year to be considered as unemployed. Workers that are inactive part of the year are still counted as active. Our main variable of interest is individual annual real earnings, i.e. total real earnings from wages and salaries of all the jobs of the individual. Self-employed individuals are dropped. Table 1 reports summary statistics for the main variables before and after the crisis. Note that in Table 1 the share of zero earnings is rather low, since it only includes individuals with at least one non-zero earnings year in at least one of the two periods 2002-2006 and 2012-2016. We notice that in the post-crisis period the mean of log annual real earnings is slightly higher than its pre-crisis level in the balanced panel, whereas it is slightly lower than its pre-crisis level in the unbalanced panel. On the other hand, log earnings variance slighly increased after the crisis both in the balanced and in the unbalanced panel. Survey data like the PSID are often contaminated with errors (Bound, Brown, and Mathiowetz (2001)). In the absence of additional information, it is not possible to disentangle the residual terms from classical measurement errors. Classical measurement error will influence the ranking of individuals and the earnings mobility results in the direction of less persistence. We follow Arellano, Bonhomme and Blundell (2017) in disregarding measurement error in the empirical analysis. Similarly to Bonhomme and Robin

(2009), we use as explanatory variables age, age squared, and a qualitative variable representing the highest education level achieved by the individual. The education dummies are constructed on the basis of the variable "years spent in education". According to the US education system, the first dummy corresponds to 1-11 years of education (i.e. less than High School), the second dummy stands for 12 years (High School graduate), the third one represents some College (13-15 years), and the last one stands for College graduate and above (16 or more years of education). We argue that these dummy variables are exogenous, i.e. they are not influenced by the individual position in the earnings distribution. Indeed, we only consider education that takes place before labor market entry. Age is included in all three stages of estimation, whereas education is included in the second and in the third stage. This approach closely resembles the one followed by Bonhomme and Robin (2009)[8].

In Table 2, we report the share of individuals with zero annual earnings in each data year, by their previous data year earnings quintile, for our sample of 25-55-year-old males. Note that in Table 2, as in Table 1, we only include individuals who record at least one non-zero earnings year in at least one of the two periods: 2002-2006 and 2012-2016. From this Table, we deduce that zero earnings in one or two years of each subsample are more common for individuals who where already at the bottom quintile of the earnings distribution (the share ranges between around 8 and 13%). In the other quintiles, the share of individuals with zero earnings is moderate, ranging from less than 1% to around 4% in the years considered in our sample. For most of the analyses presented in the following, as well as for all the estimations presented in Section 4, we restrict our attention to a sub-sample of individuals who have a strong labor market attachment, i.e. males aged between 25 and 55 year old who were continously employed in at least one of the two sample periods (2002-2006 and 2012-2016). Note that in all the analyses on earnings dynamics presented in the present paper, we exclude data years 2008 and 2010, since it is essentially when the financial crisis took place.

[8]Note that in principle the same variables can be included in all the three stages of estimation. Indeed, in the preliminary earnings regression we estimate the role of a certain variable in determining the raw earnings level. On the other hand, in the second step we estimate the role of the same variable in determining the individual rank in the empirical cross-sectional distribution of the residual earnings. Finally, in the third step we estimate its role in determining the individual degree of rank persistence or immobility.

Table 2: Percentage of zero earnings by previous data year earnings quintile, PSID data, males 25-55-year-old, by data year. Only individuals who recorded at least one non-zero earnings year in at least one of the two periods 2002-2006 and 2012-2016 are included. Zero earners in the past year are included in the bottom quintile.

Percentage of zero earnings in each year in the unbalanced sample

Past earn. quintile	2004	2006	2014	2016
1	8.23%	12.44%	9.68%	12.68%
2	3.56%	2.83%	2.59%	4.04%
3	2.47%	0.80%	1.43%	1.62%
4	1.72%	1.30%	1.71%	0.37%
5	1.14%	1.90%	0.61%	0.63%

From the upper panels of Figure 1, we notice that the dispersion of total annual real earnings is slightly higher after the crisis than before it, both in the balanced and in the unbalanced sample. From the bottom panels of Figure 1 we find that inequality, as measured by the Gini coefficient, increased after the crisis, peaking in 2012, and then slowly coming back to its pre-crisis level. This is true both in the balanced and in the unbalanced sample. A two-sample Kolmogorov-Smirnov test for equality of distribution functions shows that the earnings distribution in our sample significantly changed before and after the crisis. This holds both in the balanced and in the unbalanced sample. The fact that the distribution of earnings changed over time implies that, in the framework of our analysis, absolute and relative mobility are not the same thing. For this test, we take 2002-2006 as the pre-crisis period and 2012-2016 as the post crisis period. The result of the test is reported in Table 3.

Table 3: Kolmogorov-Smirnov test

	Group 1: 2002-2006	Group 2: 2012-2016	P-value
Balanced sample	Obs. 5172	Obs. 5664	0.0000
Unbalanced sample	Obs. 8788	Obs. 9679	0.0000

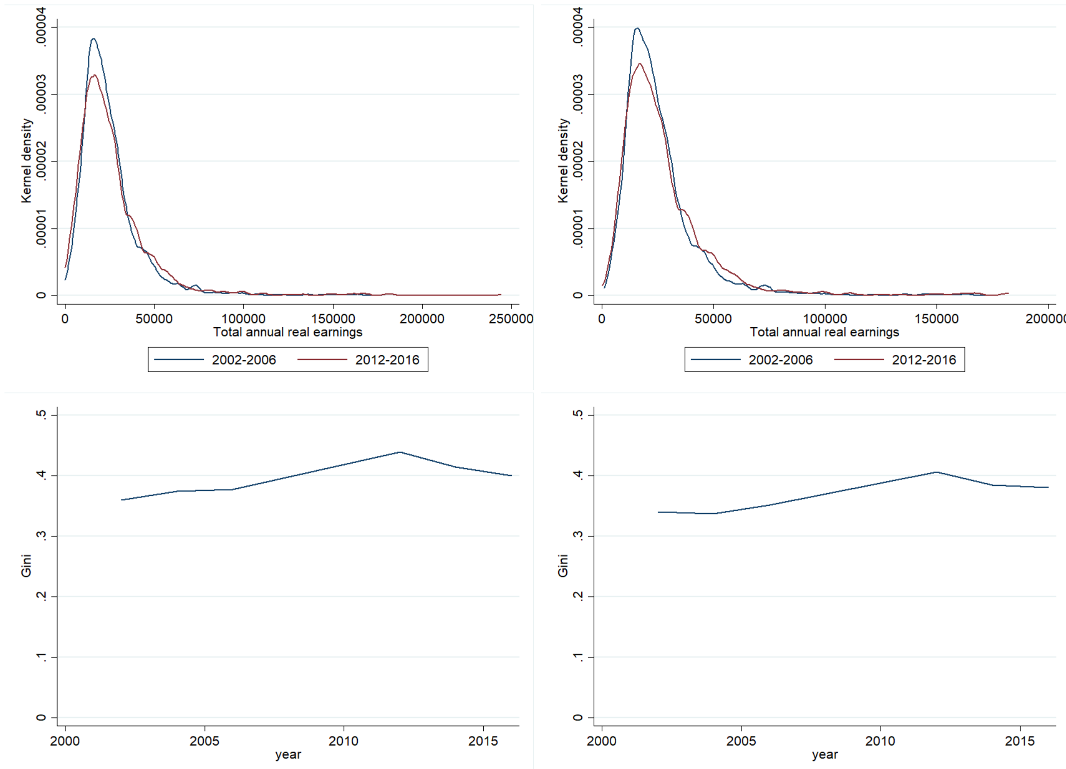

Figure 1: Upper panels: kernel density estimation of total annual real earnings, by data year, PSID data, 25-55-year-old males, top 2% has been trimmed for readability, unbalanced sample (left) and balaced sample (right). Right panel: Gini coefficient of total annual real earnings, by data year, PSID data, 25-55-year-old males, unbalanced sample (left) and balaced sample (right).

Figure 2 shows the evolution of residual earnings percentiles over the years, respectively in the unbalanced panel (left) and in the balanced panel (right). These percentiles have been constructed in the basis of the residuals from equation (1): $\hat{Resid}_{i,t} = Earnings_{i,t} - \hat{\alpha}_1 Age_{it} - \hat{\alpha}_2 Age_{it}^2 - \hat{\lambda}_t - \hat{\eta}_i$. Note that the unbalanced panel does not include individuals who recorded zero earnings every year in at least one of the two periods 2002-2006 and 2012-2016. From Figure 2, we find more variability in the percentiles after the crisis (i.e. starting from data year 2010), both in the balanced and in the unbalanced sample, especially at the bottom of the residual earnings distribution, Indeed, after 2010 the distance from the 5th and the 10th percentile notably widens. This means that, after the financial crisis, the (residual) part of earnings which was not explained deterministically by age and by the time shift was larger than before the crisis. This is particularly true for the 5% residual earnings percentile, which witnessed a large drop in 2012. This means that people staying in their same relative positions (percentiles) at the bottom of the residual earnings distribution before and after the crisis have seen the absolute value of their residual earning decrease. Absolute wage mobility has moved in the direction of increasing residual earnings inequality starting from 2010 onward, as the distance between the bottom 5% percentile and the bottom 10% percentile has widened, while the distance between, e.g. the 20% and the 40% or the 40% and the 60% percentiles has remained almost unchanged. These findings hold both in the balanced and in the unbalanced sample. As all the following analyses are based on the balanced panel, percentiles computed on the balanced panel will be used to construct transition matrices.

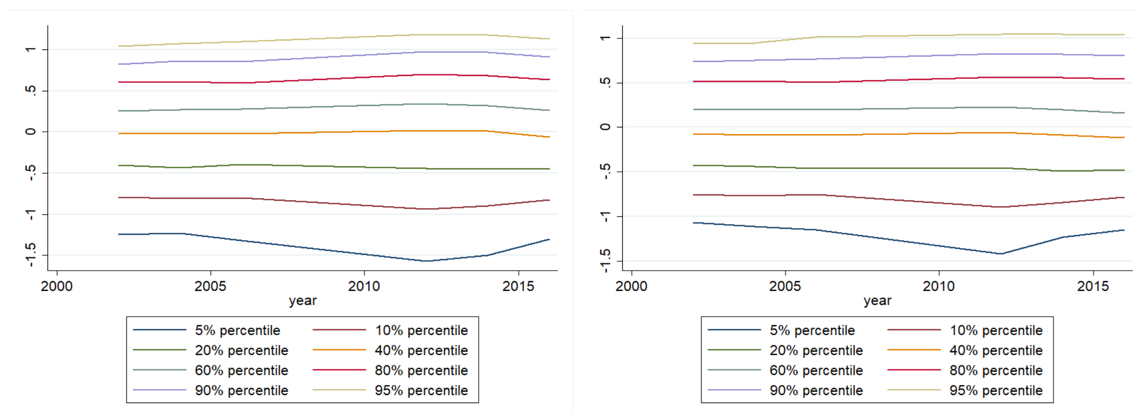

Figure 2: Evolution of residual earnings quintiles, by data year, unbalanced panel (left) and balanced panel (right). These percentiles have been constructed in the basis of the residuals from equation (1): $\hat{Resid}_{i,t} = Earnings_{i,t} - \hat{\alpha}_1 Age_{it} - \hat{\alpha}_2 Age_{it}^2 - \hat{\lambda}_t - \hat{\eta}_i$. Note that the unbalanced panel does not include individuals who recorded zero earnings every year in at least one of the two periods 2002-2006 and 2012-2016.

3.1 The transition matrix approach to earnings mobility

In the following, transition matrices between residual earnings quintiles before and after the financial crisis are presented. Transition matrices are one of the most commonly used methods to assess the degree of earnings mobility within an economy (see e.g. Shorrocks (1978), Fields and Ok (1999), Dickens (2000), Bonhomme and Robin (2009)). When naming data pairs, we always use the latter year. For example, mobility in 2006 means mobility between 2004 and 2006. We compute the quintiles of the residual earnings distribution and present 2-year and 4-year transitions between quintiles. Residual earnings quintiles are re-computed in each data year. Tables 4-5 are constructed on the basis of the balanced panel, i.e. we only include 25-

55 year old males who where continosly employed in at least one of the two sample periods (2002-2006 and 2012-2016). This reduces our sample to 59% of its original size before the crisis and to 58.5% of it after the crisis. These transition matrices are constructed on the basis of the residuals from equation (1):

$$\hat{Resid}_{i,t} = Earnings_{i,t} - \hat{\alpha}_1 Age_{it} - \hat{\alpha}_2 Age^2_{it} - \hat{\lambda}_t - \hat{\eta}_i \quad (13)$$

Note that these transition matrices do not need to be symmetric. Indeed, rows stand for past quintiles, whereas columns stand for present quintiles. The row totals of each matrix are equal to one. Hence, for example the cell in the second row, first column of the upper part of Table 4, means that, among all individuals who were in the second residual earning quintile in period $t-1$ in the pre-crisis subsample (2002-2006), 19.83% fell in the first (i.e. bottom) residual earnings quintile in period t. Conversely, the cell in the first row, second column of the same Table signifies that, among all individuals who were in the bottom quintile at time $t-1$ during the pre-crisis period, 19.8% ended up in the second quintile in period t. In Tables 4-5, an individual is recorded as a stayer if he/she is recorded to be in the same residual earnings quintile in period $t-k$ and in period t, and a mover otherwise. The percentages of stayers are displayed in bold along the main diagonal of the transition matrix. We report this result for $k = 1, 2$[9].

From the transition matrices reported in Tables 4-5, we deduce that: (i) the degree of two-year positional mobility in the earnings distribution is not so high, with around 55%-60% (i.e. the average of the numbers reported on the main diagonal of the matrices) of individuals remaining in the same residual earnings quintile after two years. (ii) We observe more positional persistence at the top and at the bottom of the distribution, i.e. in the first and in the fifth residual earning quintiles. Persistence takes hence a U-shape, i.e. it is higher at the bottom and at the upper quintiles, while it is lower in the middle of the distribution. This is true both before and after the crisis and on all the time horizons considered. This finding, i.e. more stability at the extremes of the distribution, is consistent with those of, e.g. Cardoso (2006), Pavlopoulos et al. (2007), Bonhomme and Robin (2009), Germandt (2009)). (iii) We find no relevent change in persistence in any part of the residual earnings distribution after the crisis, in any of the two time horizons considered.

[9] Recall that PSID data are collected every second year since data year 1996. Hence, in the following we show transition matrices over a two- and and four-year time horizon.

Table 4: Empirical 2-year transition matrices, balanced panel. Present quintile is on the columns, past quintile is on the rows. Row total is 1. Transitions are computed on the basis of the residuals defined in equation (13).

Pre-crisis sample (2002-2006), transitions 2002-2004 and 2004-2006					
	destination				
origin	1	2	3	4	5
1	**0.6951**	0.198	0.0592	0.0318	0.0159
2	0.1983	**0.5543**	0.1838	0.0449	0.0188
3	0.0684	0.1907	**0.5386**	0.1732	0.0291
4	0.0173	0.0462	0.1934	**0.583**	0.1602
5	0.0204	0.0117	0.0248	0.1693	**0.7737**
Post-crisis sample (2012-2016), transitions 2012-2014 and 2014-2016					
	destination				
origin	1	2	3	4	5
1	**0.6772**	0.2134	0.0725	0.029	0.0079
2	0.2258	**0.5046**	0.1939	0.0571	0.0186
3	0.0715	0.2013	**0.5086**	0.1907	0.0278
4	0.0238	0.0623	0.1987	**0.5841**	0.1311
5	0.0093	0.0106	0.0265	0.1393	**0.8143**

Table 5: Empirical 4-year transition matrices, balanced panel. Present quintile is on the columns, past quintile is on the rows. Row total is 1. Transitions are computed on the basis of the residuals defined in equation (13).

Pre-crisis sample (2002-2006), transitions 2002-2006					
	destination				
origin	1	2	3	4	5
1	**0.6329**	0.2399	0.0751	0.0347	0.0173
2	0.2334	**0.4669**	0.2161	0.0605	0.0231
3	0.0877	0.2427	**0.4357**	0.1901	0.0439
4	0.0318	0.0462	0.2254	**0.5202**	0.1763
5	0.0117	0.0087	0.0466	0.1953	**0.7376**
Post-crisis sample (2012-2016), transitions 2012-2016					
	destination				
origin	1	2	3	4	5
1	**0.6005**	0.2407	0.0979	0.0476	0.0132
2	0.2725	**0.4101**	0.209	0.0847	0.0238
3	0.0902	0.2626	**0.4058**	0.1989	0.0424
4	0.0423	0.0582	0.2407	**0.4921**	0.1667
5	0.0053	0.0186	0.0451	0.1777	**0.7533**

In the following Section we compare the empirical transition matrices presented here with those estimated via the two copula models presented in Section 2, i.e. the parametric and the semi-nonparametric ones.

4 Results and discussion

In this Section, we report and compare the estimation results obtained via the parametric and the semi-nonparametric copula models. All the estimates presented in this Section have been obtained on the balanced panel described in Section 3.

First, we estimate the parameter τ of the Plackett copula, which is a linear combination of individual explanatory variables, separately for the pre-crisis and for the post-crisis period, via Maximum Likelihood. On the basis of the estimated parameter for the Plackett copula, we can predict the estimated $\tilde{Z}_{it}$ for each individual in both our pre-crisis and post-crisis samples, given the value of $Z_{i,t-1}$ and those of the individual covariates, according to the following formula, which is the empirical counterpart of equation (3):

$$\tilde{Z}_{i,t} = E[Z_{it}|Z_{i,t-1}] = \int_0^1 \Phi^{-1}(u)\hat{c}(u, \Phi(Z_{i,t-1}))du \tag{14}$$

where $\Phi^{-1}(\cdot)$ is the standard normal quintile function, $\Phi(\cdot)$ is the standard normal cdf and $\hat{c}(\cdot,\cdot) = c(\cdot,\cdot,\hat{\tau})$, i.e. the Plackett copula density, whose parameter depends on the individual covariates and has been estimated by maximum likelihood. The integral is computed by Monte Carlo simulation with 1000 simulations. This allows us to construct estimated transition matrices between $Z_{i,t-k}$ and $\tilde{Z}_{it}$ for $k = 1, 2$. The row totals of each matrix are equal to one.

From Tables 6-7, we deduce that the fully parametric model has a rather good fit to the data. For example, the predicted percentage of stayers at the bottom of the residual earnings distribution over a 2-year horizon is around 0.68 before the crisis and 0.69 after the crisis. In the data (Tables 4-5) these percentages computed on the basis of actual transitions are, respectively, 0.7 and 0.68. Further, the predicted shares of stayers (i.e. individuals on the main diagonals) in the middle quintiles range between 0.47 and 0.57, whereas from actual transitions these shares range between 0.5 and 0.58. The prediction accuracy of the fully parametric model appears as quite good also over a 4-year horizon. Indeed, from Table 7 we deduce that the predicted share of stayers in the top residual earnings quintile is around 0.76-0.78, both before and after the crisis. This is not very different from the share of 0.74-0.75 that we find in the data (Table 5).

Table 6: Parametric copula, 2-year transition matrices, balanced panel. Present quintile is on the columns, past quintile is on the rows. Row total is 1. Transitions between past ranks $Z_{i,t-k}$ constructed as $Z_{i,t} = \Phi^{-1}(F_t(\hat{Resid}_{i,t}))$, where $\hat{Resid}_{i,t}$ is defined in eq. (13), and present ranks $\tilde{Z}_{i,t}$ predicted according to eq. (14): $\tilde{Z}_{i,t} = \int_0^1 \Phi^{-1}(u)\hat{c}(u, \Phi(Z_{i,t-1}))du$ are shown.

Pre-crisis sample (2002-2006), transitions 2002-2004, 2004-2006					
	destination				
origin	1	2	3	4	5
1	**0.6754**	0.2	0.0609	0.0464	0.0174
2	0.2104	**0.5447**	0.1758	0.049	0.0202
3	0.0671	0.1778	**0.5685**	0.1574	0.0292
4	0.0233	0.064	0.1773	**0.5698**	0.1657
5	0.0262	0.0087	0.0174	0.1802	**0.7674**

Post-crisis sample (2012-2016), transitions 2012-2014, 2014-2016					
	destination				
origin	1	2	3	4	5
1	**0.6905**	0.2037	0.0767	0.0212	0.0079
2	0.2175	**0.5119**	0.2228	0.0424	0.0053
3	0.0637	0.2069	**0.4721**	0.2255	0.0318
4	0.0212	0.061	0.2069	**0.5438**	0.1671
5	0.008	0.0159	0.0239	0.1671	**0.7851**

Table 7: Parametric copula, 4-year transition matrices, balanced panel. Present quintile is on the columns, past quintile is on the rows. Row total is 1. Transitions between past ranks $Z_{i,t-k}$ constructed as $Z_{i,t} = \Phi^{-1}(F_t(\hat{Resid}_{i,t}))$, where $\hat{Resid}_{i,t}$ is defined in eq. (13), and present ranks $\tilde{Z}_{i,t}$ predicted according to eq. (14): $\tilde{Z}_{i,t} = \int_0^1 \Phi^{-1}(u)\hat{c}(u, \Phi(Z_{i,t-1}))du$ are shown.

Pre-crisis sample (2002-2006), transitions 2002-2006					
	destination				
origin	1	2	3	4	5
1	**0.6647**	0.2052	0.0665	0.0462	0.0173
2	0.219	**0.5418**	0.17	0.049	0.0202
3	0.0674	0.1789	**0.563**	0.1584	0.0323
4	0.0231	0.0548	0.1844	**0.5677**	0.17
5	0.0263	0.0146	0.0175	0.1784	**0.7632**

Post-crisis sample (2012-2016), transitions 2012-2016					
	destination				
origin	1	2	3	4	5
1	**0.6878**	0.2063	0.0767	0.0212	0.0079
2	0.2196	**0.5106**	0.2196	0.045	0.0053
3	0.0638	0.2074	**0.4707**	0.2261	0.0319
4	0.0185	0.0582	0.2063	**0.5423**	0.1746
5	0.0106	0.016	0.0293	0.1649	**0.7793**

Let us now discuss the estimation results of the semi-nonparametric copula model. On the basis of our estimated semi-nonparametric copula model, we are able to predict the present rank, $\tilde{Z}_{it}$ given the values of $Z_{i,t-1}$ and those of the individual covariates:

$$\tilde{Z}_{it} = \hat{\Lambda}(\hat{\rho}(Z_{i,t-1}, X'_{i,t}\hat{\beta}_2) \tag{15}$$

where $\hat{\Lambda}(k)$ is the estimated function $\Lambda(k)$ which appears in eq. (6), and $\hat{\rho}(Z_{i,t-1})$ is the emprical counterpart of function $\rho(Z_{i,t-1})$ in eq. (6), too. Then, similarly to what we did above, we estimate transition matrices between $\tilde{Z}_{it}$ and $Z_{i,t-1}$, over different time horizons. From Tables 8-9, we observe that the semi-nonparametric copula model has a good fit to the data. Indeed, for example the predicted shares of stayers at the bottom of the residual earnings distribution over a 2-year horizon are 0.68 (pre-crisis) and 0.7 (post-crisis), whereas the actual ones (reported in Tables 4-5) are 0.7 and 0.68.

Table 8: Semi-nonparametric copula, 2-year transition matrices, balanced panel. Present quintile is on the columns, past quintile is on the rows. Row total is 1. Transitions between past ranks $Z_{i,t-k}$ constructed as $Z_{i,t} = \Phi^{-1}(F_t(\hat{Resid}_{i,t}))$, where $\hat{Resid}$ is defined as in eq. (13), and present ranks $\tilde{Z}_{i,t}$ predicted according to eq. (15): $\tilde{Z}_{it} = \hat{\Lambda}(\hat{\rho}(Z_{i,t-1}, X'_{i,t}\hat{\beta}_2)$ are shown.

Pre-crisis sample (2002-2006), transitions 2002-2004, 2004-2006					
	destination				
origin	1	2	3	4	5
1	**0.6783**	0.1942	0.0667	0.0406	0.0203
2	0.2	**0.5507**	0.1826	0.0522	0.0145
3	0.064	0.2093	**0.5262**	0.1686	0.032
4	0.0406	0.0435	0.1942	**0.5594**	0.1623
5	0.0174	0.0087	0.0233	0.1802	**0.7703**

Post-crisis sample (2012-2016), transitions 2012-2014, 2014-2016					
	destination				
origin	1	2	3	4	5
1	**0.6984**	0.2011	0.0714	0.0212	0.0079
2	0.2255	**0.5305**	0.1804	0.0557	0.008
3	0.0582	0.1984	**0.5265**	0.1772	0.0397
4	0.0133	0.0505	0.1941	**0.5798**	0.1622
5	0.0106	0.0159	0.0265	0.1671	**0.7798**

Table 9: Semi-nonparametric copula, 4-year transition matrices, balanced panel. Present quintile is on the columns, past quintile is on the rows. Row total is 1. Transitions between past ranks $Z_{i,t-k}$ constructed as $Z_{i,t} = \Phi^{-1}(F_t(\hat{Resid}_{i,t}))$, where $\hat{Resid}$ is defined as in eq. (13), and present ranks $\tilde{Z}_{i,t}$ predicted according to eq. (15): $\tilde{Z}_{it} = \hat{\Lambda}(\hat{\rho}(Z_{i,t-1}, X'_{i,t}\hat{\beta}_2)$ are shown.

Pre-crisis sample (2002-2006), transitions 2002-2006					
	destination				
origin	1	2	3	4	5
1	**0.659**	0.1965	0.0694	0.0549	0.0202
2	0.2248	**0.5187**	0.1844	0.0519	0.0202
3	0.0587	0.2346	**0.4839**	0.1848	0.0381
4	0.0288	0.0461	0.2305	**0.4986**	0.196
5	0.0263	0.0088	0.0263	0.2105	**0.7281**
Post-crisis sample (2012-2016), transitions 2012-2016					
	destination				
origin	1	2	3	4	5
1	**0.6825**	0.2143	0.0714	0.0238	0.0079
2	0.2354	**0.5**	0.2063	0.045	0.0132
3	0.0638	0.2181	**0.4761**	0.2021	0.0399
4	0.0132	0.0503	0.2222	**0.5423**	0.172
5	0.0106	0.0133	0.0239	0.1862	**0.766**

Further the predicted percentages of stayers at the top of the distribution over the same horizon are 0.77 (pre-crisis) and 0.78 (post-crisis) vs the actual recorded values of, respectively, 0.77 and 0.81. The percentage of stayers over a 2-year horizon ranges between 0.53 and 0.78, which is also in line with actual transitions recorded in the data. On a 4-year horizon, the predicted share of stayers at the bottom of the distribution is 0.66 (pre-crisis) and 0.68 (post-crisis), whereas in the data we find, respectively, 0.63 and 0.60. Finally, in Figure 3 we report an alternative measure of mobility, which is based on quantiles (see eq. (10) in Section 2.1). The slope of the conditional quantiles of the present rank, which are depicted in Figure 3, provides an alternative measure of rank immobility. From Figure 3, our previous findings are confirmed: there are no relevant differences in the mobility patterns before and after the crisis. However, we find that 45-year-old workers with college degree become more mobile at the bottom of the rank distribution after the crisis. Further, we notice that the degree of mobility at the bottom of the distribution is higher for young and highly educated workers (i.e. 25-year-old with college degree), both before and after the crisis. This is apparent from the two middle panels of Figure 3, in which the slope of the conditional quintile is almost flat for a 25-year-old worker with college degree at the bottom of the residual earnings rank distribution. This suggests a low degree of association between the past and the present Gaussian ranks.

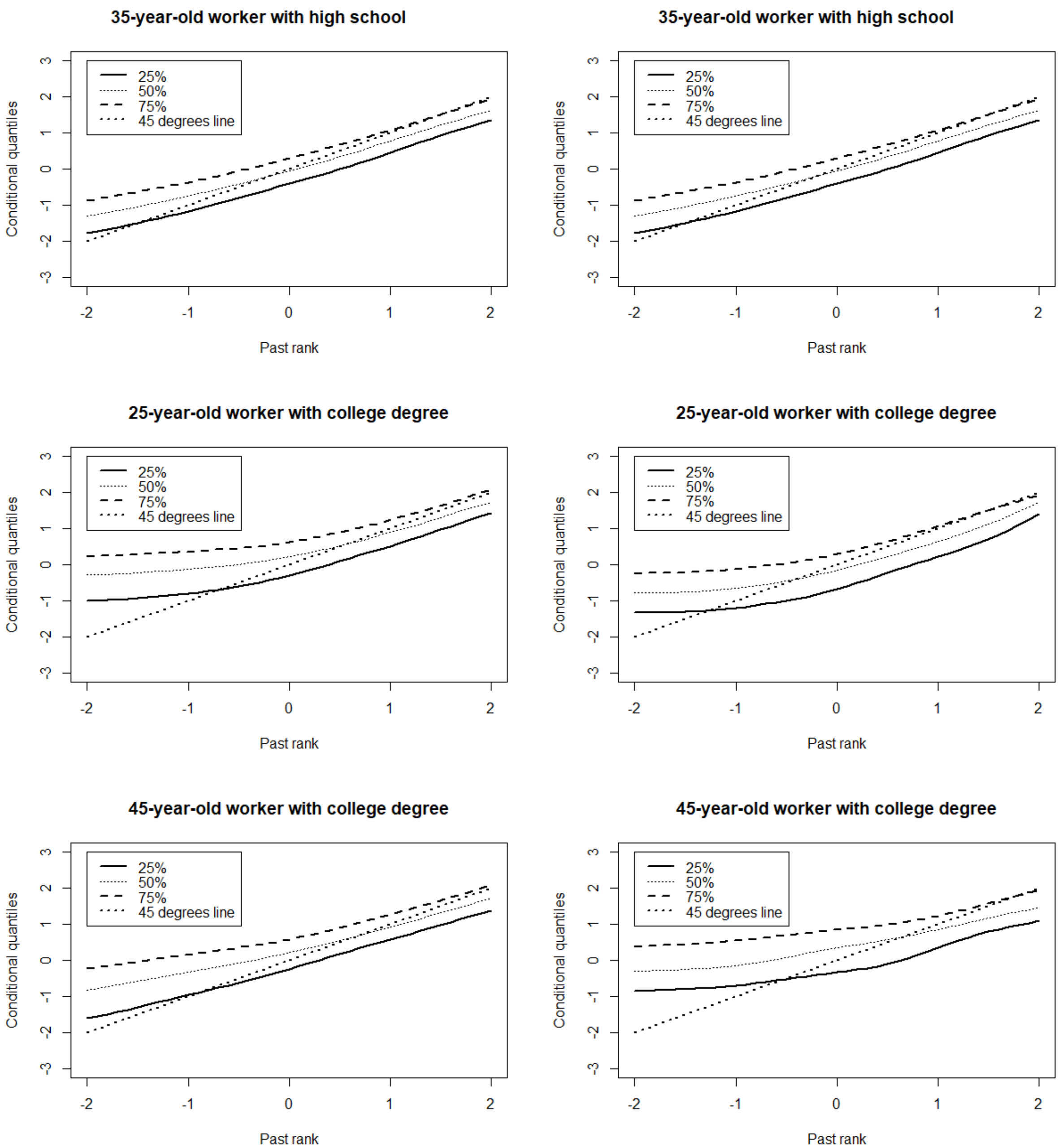

Figure 3: Conditional quantiles of the residual earnings component. In this figure we report three conditional quantiles (median, lower and upper quartiles) as a function of the past rank for the same sets of individual characteristics as in Figure 4, respectively before the crisis (left panels and after the crisis (right panels). The conditional quantiles are estimated using equation (10) for workers at the first, second and third quartiles of the present year residual rank distribution ($u = 0.25, 0.50, 0.75$).

In order to adequately compare the performance of the different models, in Table 10 we report a sort of (Chi-square) aggregate measure of prediction accuracy. This measure is computed as follows:

$$PredAccuracy_m = \sum_{i=1}^{5}\sum_{j=1}^{5}(P_{ijm} - A_{ij})^2/A_{ij} \quad (16)$$

where ij stands for each cell of a transition matrix (i stands for the row and j for the column) and m represents the model used to produce the predictions, i.e. fully parametric Plackett copula or the semi-nonparametric copula. A_{ij} is the actual transition probability observed in cell ij and and P_{ijm} is the transition probability predicted with model m for the same cell ij. From Table 10, we deduce that, on the two-year time horizon, the semi-nonparametric copula is equally good or slightly worse than the fully parametric Plackett copula (0.04 vs 0.03 pre-crisis and 0.05 for both models after the crisis). However, the semi-nonparametric copula notably outperforms the fully parametric one on the four-year horizon (0.08 vs 0.16 pre-crisis and 0.16 vs 0.18 post crisis).

Table 10: Prediction accuracy of different parametric and semiparametric models

Distance from actual transitions	Before crisis		After crisis	
	2-year	4-year	2-year	4-year
Plackett	0.03	0.16	0.05	0.18
Semi-nonpar copula	0.04	0.08	0.05	0.16

5 Concluding remarks

In this paper we analyzed relative earnings mobility before and after the Great Recession in the United States. At the aggregate level (i.e. residual earnings quintiles), we find no evidence of changes in the earnings mobility patterns in any part of the residual earnings distribution after the financial crisis. This is confirmed by both actual transitions recorded in the data and estimation results obtained with a fully parametric and a semi-nonparametric copula. However, the semi-nonparametric model has the advantage of allowing us to estimate the degree of relative earnings mobility for virtually each individual in our sample. Thanks to the model's flexibility, we are able to compute an alternative mobility measure, based on quantiles, and in this way to uncover the fact that mobility increased at the bottom of the distribution for 45-year-old workers with a college degree after the crisis. Such a feature of individual mobility is likely to remain undiscovered when only relying on a fully parametric method.

Further, the semi-nonparametric copula has a better fit to the data as shown in Table 10. When using the more flexible copula model, we decided not to model zero earnings years, since it would have been computationally not tractable, as explained in the Introduction. This is why the focus of the present paper lies on workers with high labor market attachment. Finding a way to include the modelling of zero years into the functional copula model represents an avenue for further research. The use of our semi-nonparametric copula model to generate simulated earnings histories, thus creating a link between absolute and relative mobility, also constitutes scope for future research.

References

1. Arellano, M., Blundell, R., & Bonhomme, S. (2017). Earnings and consumption dynamics: a nonlinear panel data framework. Econometrica, 85(3), 693-734.
2. Baily, M. N.,& Bosworth, B. P. (2014). US manufacturing: Understanding its past and its potential future. Journal of Economic Perspectives, 28(1), 3-26.
3. Blair, P. Q., & Deming, D. J. (2020, May). Structural Increases in Demand for Skill after the Great Recession. In AEA Papers and Proceedings (Vol. 110, pp. 362-65).
4. Boes, S., Staub, K., & Winkelmann, R. (2010). Relative status and satisfaction. Economics Letters, 109(3), 168-170.
5. Bonhomme, S., and Robin, J. M. (2009). Assessing the equalizing force of mobility using short panels: France, 1990 - 2000. The Review of Economic Studies, 76(1), 63-92.
6. Bound, J., Brown, C., & Mathiowetz, N. (2001). Measurement error in survey data. Handbook of Econometrics, 5, 3705-3843.
7. Boyce, C. J., Brown, G. D., & Moore, S. C. (2010). Money and happiness: Rank of income, not income, affects life satisfaction. Psychological Science, 21(4), 471-475.
8. Brown, G. D., Gardner, J., Oswald, A. J., & Qian, J. (2008). Does Wage Rank Affect Employees' Well-being?. Industrial Relations: A Journal of Economy and Society, 47(3), 355-389.
9. Cardoso, A. R. (2006). Wage mobility: do institutions make a difference?. Labour economics, 13(3), 387-404.

10. Chen, X. (2007). Large sample sieve estimation of semi-nonparametric models. Handbook of econometrics, 6, 5549-5632.

11. Chodorow-Reich, G. (2013). The employment effects of credit market disruptions: Firm-level evidence from the 2008-9 financial crisis. The Quarterly Journal of Economics, 129(1), 1-59.

12. Christiano, L. J., Eichenbaum, M. S., & Trabandt, M. (2015). Understanding the great recession. American Economic Journal: Macroeconomics, 7(1), 110-67.

13. Clark, A. E., Kristensen, N., & Westergard-Nielsen, N. (2009). Job Satisfaction and Co-worker Wages: Status or Signal?. The Economic Journal, 119(536), 430-447.

14. Cowell, F. A., & Flachaire, E. (2018). Measuring mobility. Quantitative Economics, 9(2), 865-901.

15. Dickens, R. 2000. Caught in a trap? Wage mobility in Great Britain: 1975-1994. Economica, 67, 477-197.

16. Dustmann, C., Fitzenberger, B., Schoenberg, U., & Spitz-Oener, A. (2014). From sick man of Europe to economic superstar: Germany's resurgent economy. The Journal of Economic Perspectives, 28(1), 167-188.

17. Elstad, J. I., Dahl, E., & Hofoss, D. (2006). Associations between relative income and mortality in Norway: a register-based study. The European Journal of Public Health, 16(6), 640-644.

18. Fields, G. S., & Ok, E. A. (1999). The measurement of income mobility: An introduction to the literature. Springer Netherlands, 557-598.

19. Fischer, J. A., & Torgler, B. (2006). The effect of relative income position on social capital. Economics Bulletin, 26(4), 1-20.

20. Fitzgerald, J., & Gottschalk, P. (1998). An analysis of sample attrition in panel data. Journal of Human Resources, 33(2), 251-299.

21. Formby, J. P., Smith, W. J., & Zheng, B. (2004). Mobility measurement, transition matrices and statistical inference. Journal of Econometrics, 120(1), 181-205.

22. Gernandt, J. (2009). Decreasing wage mobility in Germany. ZEW-Centre for European Economic Research Discussion Paper, (09-044).

23. Hall, R. E. (2015). Quantifying the lasting harm to the US economy from the financial crisis. NBER Macroeconomics Annual, 29(1), 71-128.

24. Hryshko, D. (2012). Labor income profiles are not heterogeneous: Evidence from income growth rates. Quantitative Economics, 3(2), 177-209.

25. Hu, Y., Moffitt, R., & Sasaki, Y. (2019). Semiparametric Estimation of the Canonical Permanent-Transitory Model of Earnings Dynamics. Quantitative Economics. Forthcoming.

26. Huber, K. (2018). Disentangling the effects of a banking crisis: evidence from German firms and counties. American Economic Review, 108(3), 868-98.

27. Joe, H. (1997). Multivariate models and multivariate dependence concepts. CRC Press

28. Li, A., Wallace, M., & Hyde, A. (2019). Degrees of inequality: The Great Recession and the college earnings premium in US metropolitan areas. Social science research, 84, 102342.

29. Lillard, L. A., & Panis, C. W. (1998). Panel attrition from the Panel Study of Income Dynamics: Household income, marital status, and mortality. Journal of Human Resources, 437-457.

30. Luttmer, E. F. (2005). Neighbors as negatives: Relative earnings and well-being. The Quarterly journal of economics, 120(3), 963-1002.

31. Meghir, C., & Pistaferri, L. (2004). Income variance dynamics and heterogeneity. Econometrica, 72(1), 1-32.

32. Naguib, C., Gagliardini, P. (2020). Wage Mobility: A Functional Copula Approach, SSRN working paper.

33. Nelsen, R. B. (1999). An introduction to copulas (Vol. 139). Springer Science and Business Media.

34. Plackett, R. L. (1965). A class of bivariate distributions. Journal of the American Statistical Association, 60(310), 516-522.

35. Putnam, R. D. (1993). The prosperous community: social capital and public life. The american prospect, (13).

36. Pavlopoulos, D., Fouarge, D., Muffels, R., & Vermunt, J. (2007). Job mobility and wage mobility of high-and low-paid workers. Schmollers Jahrbuch, 127(1).

37. Shorrocks, A. F. (1978). The measurement of mobility. Econometrica: Journal of the Econometric Society, 1013-1024.

Appendix A Estimation procedure

In this Appendix we provide details on the estimation strategy used to obtain the results presented in Section 4. This Appendix relies on Section 3 of Naguib and Gagliardini (2020); for further details we refer the interested reader to the companion paper. We estimate a copula model in which mobility is represented by a function of the past rank and of some explanatory variables. The model features two indexes, namely $V_{i,t} = X'_{it}\beta_1^0$ for the marginal distribution and $W_{i,t} = X'_{it}\beta_2^0$ for the copula functional parameter. The first index, which we call "marginal distribution score", accounts for the role of the individual characteristics in determining the position in the cross-sectional distribution at any given date. On the other hand, the second index, called "mobility score", accounts for the role of the same variables in determining the degree of mobility of the transitory earnings component. As explanatory variables, in both indices we include age and a qualitative variable representing the highest education level achieved by the individual. All these variables are strictly exogenous and stationary (age has been de-trended). We estimate the parameters of the marginal distribution and those of the copula function sequentially. Let $G(Z_{i,t}|X'_{i,t}\beta_1)$ be the distribution of the rank, conditional on the individual variables, and $c[\cdot,\cdot,\rho(\cdot)]$ the copula density. The univariate conditional distributions are estimated via a kernel single-index model. We estimate the coefficients β_1 with a Maximum Likelihood approach:

$$argmax_{\beta_1} \sum_t \sum_i log\, \tilde{g}_{-(it)}(\hat{Z}_{i,t}|\beta_1' X_{it}) \mathbf{I}_{\{Z_{it}, X_{it}\} \in S}, \tag{17}$$

where $\tilde{g}_{-(it)}(z|w)$ is the estimated conditional kernel density, $w_{jt}(\beta_1) = \beta_1' X_{jt}$, $\hat{Z}_{jt}$ represents the estimated Gaussian rank and K is a standard Gaussian pdf:

$$\tilde{g}_{-(it)}(z|w;\beta_1) = \frac{\frac{1}{h_z}\sum_t \sum_{j,j\neq i} K\left(\frac{z-\hat{Z}_{jt}}{h_z}\right) K\left(\frac{w-w_{jt}(\beta_1)}{h_w}\right)}{\sum_t \sum_{j,j\neq i} K\left(\frac{w-w_{jt}(\beta_1)}{h_w}\right)}. \tag{18}$$

With this procedure, we obtain an estimator of β_1 that is consistent under some regularity conditions (see Naguib and Gagliardini (2020)). By plugging in the estimate $\hat{\beta}_1$, we get an estimator of the pdf:

$$\tilde{g}(z|x) = \frac{\frac{1}{h_z}\sum_{t=1}^{T}\sum_{i=1}^{N} K\left(\frac{z-\hat{Z}_{it}}{h_z}\right) K\left(\frac{\hat{\beta}_1'(x-X_{it})}{h_w}\right)}{\sum_{t=1}^{T}\sum_{i=1}^{N} K\left(\frac{\hat{\beta}_1'(x-X_{it}^a)}{h_w}\right)}. \tag{19}$$

Further, we take into account the restriction that the univariate conditional distribution must be standard normal, and the constraint that the conditional density must integrate to 1 as follows:

$$\hat{g}(z|x) = \frac{\tilde{g}(z|x)e^{\mu(z)}}{\int_{\mathcal{Z}_N} \tilde{g}(z|x)e^{\mu(z)}}, \tag{20}$$

where $\mu(\cdot)$ is such that

$$e^{\mu(z)} = \phi(z) \left\{ \int_{\mathcal{X}_N} \tilde{g}(z|x) \left[\int_{\mathcal{Z}_N} \tilde{g}(z|x)e^{\mu(z)}dz \right]^{-1} dF_x(x) \right\}^{-1}. \tag{21}$$

The proof is provided in Naguib and Gagliardini (2020). The desired estimated cdf can be easily recovered by integration (for the details of the numerical implementation, see Naguib and Gagliardini (2020)). We now estimate the copula function, $c(\cdot, \cdot, \rho(\cdot, X'_{i,t}\beta_2))$ via the following Sieve Maximum Likelihood procedure:

$$\max_{\theta \in \Theta_N} \sum_{i=1}^{N} l(Y_i, \theta) \tag{22}$$

$$l(Y_i, \theta) \equiv \sum_{t=1}^{T} logc[\hat{U}_{i,t}, \hat{U}_{i,t-1}, \rho(Z_{i,t-1}, X'_{i,t}\beta_2)] \tag{23}$$

where $Y_i = (Z_{it}, X_{it}, t = 1, ..., T)$. We have $\theta = (\beta_2, \rho) \in B_2 \times \mathbb{H} \equiv \Theta$, and $\Theta_N = B_2 \times \mathbb{H}_N$ where $\mathbb{H}$ is an infinite-dimensional space of bivariate functions and $\mathbb{H}_N$ is a bivariate Sieve space (made up by Hermite polynomials in our implementation) whose dimension depends on the sample size N. Moreover, $\hat{U}_{i,t} = \hat{G}(\hat{Z}_{i,t}|X'_{i,t}\hat{\beta}_1)$, $\hat{U}_{i,t-1} = \hat{G}(\hat{Z}_{i,t-1}|X'_{i,t-1}\hat{\beta}_1)$ are the estimated (empirical) cdfs of the present and of the past Gaussian ranks, conditional on the individual explanatory variables, and $c(u, v; \rho(\cdot, W_t))$ is the copula density presented in equation (8) of Section 2. We estimate the function $\rho(Z_{i,t-1}, X'_{i,t}\beta_2)$ non-parametrically via the method of Sieves. More specifically, the approximation $\rho \in \mathbb{H}_N$ is:

$$\rho(Z_{i,t-1}, W_{i,t}) \approx \sum_{k,l=0}^{m} \lambda_{k,l} H_k(W_{i,t}) H_l(Z_{i,t-1}) \tag{24}$$

where the $\lambda_{h,l}$ are the coefficients of the polynomial basis used to approximate $\rho(\cdot)$. The number of Hermite polynomials used to approximate the autoregressive function depends on the dimension of the sample: $m = m(N)$ (Chen (2007)). Convergence of this estimator has been proved in Naguib and Gagliardini (2020) for $N \rightarrow \infty$ and $T \rightarrow \infty$. The normalization constraint $\rho(0, W_{i,t}) = 0$ for any value of $W_{i,t}$ holds under the linear constraint

$$\sum_{l=0}^{m} \lambda_{k,l} H_l(0) = 0, \quad k = 0, 1, ..., n. \tag{25}$$