Rank Mobility: A Functional Copula Approach

Costanza Naguib,* Patrick Gagliardini†

First version: August 2016. This version: July 2020

Abstract

In this paper we develop a novel semi-nonparametric panel copula model with external covariates for the study of wage rank dynamics. We focus on nonlinear dependence between the current and lagged worker's ranks in the wage residuals distribution, conditionally on individual characteristics. With respect to previous studies on relative wage mobility, we are able to relax restrictive parametric assumptions. We show consistency of the functional and finite-dimensional parameters estimators in a double asymptotics. We apply our model to US data and we find that relative mobility at the bottom of the distribution is high for workers with a college degree and some experience. On the contrary, less-educated individuals are likely to remain stuck at the bottom of the wage rank distribution year after year.

Keywords: Wage dynamics, rank, functional copula model, nonlinear autoregressive process, semi-nonparametric estimation

JEL codes: C14, J31

1 Introduction

[1] The aim of this paper is to specify and apply a novel semi-nonparametric model for individual earnings dynamics [2]. Our specific focus lies in *relative* mobility (see Shorrocks (1978), Fields and Ok (1999), 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 wage regression. The worker's percentile in

*University of Bern, WVI, Bern, Switzerland and Università della Svizzera italiana (USI), Faculty of Economics, Lugano, Switzerland.

†Università della Svizzera italiana (USI), Faculty of Economics, Lugano, Switzerland, and Swiss Finance Institute (SFI).

[1] We would like to thank Richard Blundell, Martin Browning, Robert Gary-Bobo, Christian Gourieroux, Bo Honoré, Thierry Kamionka, Rico Maggi, Fabrizio Mazzonna, Blaise Melly, Alain Monfort, Jean-Marc Robin, Sébastien Roux, Mirco Rubin, Stefan Sperlich, Mark Watson, and the participants to the Brown Bag seminar at USI Lugano on 9^{th} December 2015, the Annual Conference of the Swiss Society of Economics and Statistics on $9^{th} - 10^{th}$ June 2016 in Lugano, the ESEM conference in Geneva on $22^{th} - 26^{th}$ August 2016, the junior seminar at CREST, Paris, on 11^{th} April 2017, the Swiss Institute of Empirical Economic Research (SEW - St. Gallen) research seminar in Austria, on $5^{th} - 6^{th}$ June 2018, and the Econometric Society Winter Meeting in Rotterdam on $16^{th} - 17^{th}$ December 2019 for their insightful comments.

[2] In this paper the terms earnings and wages are used interchangeably; earnings stand for labor earnings.

this cross-sectional distribution is called wage residual rank. Low relative mobility implies that a worker may remain stuck for a long period in a given part of the wage distribution (even after having filtered out the effects of e.g. age and other individual characteristics) despite an overall increase in the wage level in the country. By definition, the cross-sectional distribution of individual ranks is the uniform distribution on the interval $[0, 1]$ - or the standard Gaussian distribution, when ranks are mapped into the real line by applying the $N(0, 1)$ quantile function. Therefore, the specification of dynamic models for ranks have to guarantee this distributional constraint. In this respect, copulas are a useful tool to model the joint distribution of the current and past individual ranks. In fact, copulas are joint distributions with uniform margins (see Joe (1997) and Nelsen (1999) for background and a review of parametric copula specifications). The (nonlinear) dependence between current and past ranks, conditional on individual characteristics, defines relative mobility. The main novelty of our paper compared to previous work is the use of a semi-nonparametric specification of the copula, instead of a parametric one, conditionally on external covariates. This flexible specification accommodates a rich variety of patterns of relative mobility as a function of the past position of the worker and her individual characteristics. The methodology allows us to investigate nonparametrically whether relative mobility is smaller in the lower part of the wage scale, i.e. there is a wage trap, and how this effect changes with e.g. education and work experience. Parametric models may lead to estimation biases since the patterns of mobility are rigidly implied by the chosen specification. While our empirical analysis focuses on the labor market, the modeling and estimation methodologies in this paper are of more general interest for studying positional mobility.

The contributions of this paper are manifold. On the methodological side, the first contribution consists in the specification of a novel copula family, which is parametrized by a function instead of a vector of unknown parameters. This copula specification is inspired by a nonlinear nonseparable autoregressive model with Gaussian invariant distribution. The copula functional parameter is the nonlinear autoregressive function of the past rank. To introduce external covariates in a parsimonious way, the copula functional parameter depends on a second argument, that is an index variable corresponding to a linear combination of the individual regressors. This yields a semi-nonparametric specification for the individual dynamics of the ranks conditionally on the observed characteristics in a panel framework. Second, based on this specification we define functional measures of relative mobility, which are the partial derivatives of the conditional expected or median rank w.r.t. the value of the past rank. The larger the absolute value of such derivatives, the smaller is the relative earnings mobility conditionally on covariates. The patterns of relative mobility are controlled by the shape of the functional parameter of the copula. Third, we define simple-to-implement estimators for the finite-dimensional and functional parameters of our model. In the first step, we obtain wage residuals from a standard fixed effect panel regression and 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 specification with the method of Sieves expanding on Hahn, Liao and Ridder (2018) in a panel framework. We show the consistency of our estimators in the double panel asymptotics with both the numbers of individuals N and time periods T tending to infinity. This double asymptotics in often used in the panel literature on bias correction for the incidental parameter problem [3]. Our estimation methodology also suffers from an incidental parameter bias when T is fixed, due to the individual fixed effects in the preliminary panel regression used to obtain the wage residuals. We conduct Monte Carlo experiments to show that this bias is small, for designs mimicking our empirical analysis with an unbalanced dataset of individuals with an average permanence in the sample of about 25 years.

Our fourth contribution is empirical in nature. We estimate our model on an unbalanced panel dataset of US workers from the PSID in the period from 1968 to 1997. We find a relatively high degree of positional persistence for workers with a low educational level in the bottom part of the wage residuals distribution. On the other hand, in the same period there was high positional mobility for those workers occupying a low position in the wage residual distribution but having a high educational level and/or several years of experience. The above-mentioned differences in the mobility patterns for workers with different characteristics are statistically significant. These results provide evidence for the existence of a wage trap but only for workers with low educational level.

This paper is not the first one to use copulas for studying individual earnings dynamics (we review this literature in Section 2.5). We build on the pioneering work of Bonhomme and Robin (2009) who use a copula model to specify the dynamics of the ranks of the transitory wage component. The joint distribution of the present and the past transitory components is modeled by the authors via the one-parameter Plackett copula. The copula parameter captures individual positional persistence conditionally on covariates, but the dependence between the current and past ranks is rigidly defined by the chosen parametric copula family. We expand on the work in Bonhomme and Robin (2009) by adopting a more flexible semi-nonparametric specification for the copula. Moreover, we allow marginal distributions to depend on individual explanatory variables. In the empirical application, we contrast the estimates from our model with those obtained with a parametric copula. We provide empirical evidence to demonstrate that our semi-nonparametric approach improves the understanding of the relative mobility patterns.

Copulas have been widely used for nonlinear time series modeling. In this context, copulas specify either

[3] See e.g. Hahn and Newey (2004), Fernandez-Val and Vella (2011), Hahn and Kuersteiner (2011), Fernandez-Val and Lee (2013), Fernandez-Val and Weidner (2016).

serial dependence in a univariate model (see e.g. Chen and Fan (2006 a) and Chen, Wu and Yi (2009) for efficient semi-parametric estimation, and Beare (2010) for the study of temporal mixing properties, in a copula Markov model), or contemporaneous dependence across innovations in a multivariate model (e.g., Chen and Fan (2006 b), Patton (2006), Härdle, Okhrin and Wang (2015), Chen, Huang and Yi (2019)), or both cross-sectional and serial dependence (e.g. Remillard, Papageorgiou and Soustra (2012)). Patton (2012) provides a review on copula models for economic time series. The setting of this paper is similar to the first case focusing on temporal dependence, albeit in a panel framework with independent individuals. The above references, as well as the vast majority of the literature, focus on parametric specifications for the copula, with nonparametric modeling of marginals. In this paper instead we consider a copula involving a functional parameter in order to capture flexible patterns of dependence between current and lagged endogenous variables. Chen, Koenker and Xiao (2009) examine the asymptotic properties of estimators for a copula-based quantile autoregressive model. The copula family is parametric, with parameter dependent on the quantile level. In the conclusions the authors mention semi-parametric modelling of the copula itself via the method of Sieves as a feasible strategy to expand the menu of the currently available copula models. Such semi-parametric estimation is performed in the present work. In fact, econometric analysis of copula densities with a functional parameter has received relatively scarce attention in the literature. Gagliardini and Gourieroux (2007) discusses efficient non parametric estimation in a framework without covariates, while Gagliardini and Gourieroux (2008) introduces a copula time series model for duration variables based on a proportional hazard specification.
The remaining of the paper is structured as follows. Section 2 introduces the model for the joint dynamics of the ranks and external covariates, and defines functional measures of relative mobility. Section 3 is devoted to the estimators and their asymptotic properties. Section 4 reports the results of the Monte Carlo simulations. Section 5 presents the dataset and the discussion of the estimation results on our sample of US workers. Section 6 concludes. Appendices A and B provide the regularity conditions and the proofs of the theoretical results, respectively. In the Supplementary Materials to the paper we present additional theoretical background on copulas and causality measures, and additional empirical results.

2 A semi-nonparametric model for ranks dynamics

In the wake of the previous literature on individual earnings dynamics, we consider a two-step modeling framework (e.g., Bonhomme and Robin (2009), Browning et al. (2010) among others). We start from the following specification for the log wage:

$$y_{it} = \alpha' X_{it} + \eta_i + \lambda_t + \varepsilon_{it} \tag{2.1}$$

for $i = 1, ..., N$ and $t = 1, ..., T$, where y_{it} is log annual wage of individual i in year t, X_{it} is a vector of observable individual characteristics, η_i is the individual effect, λ_t is the time effect and ε_{it} is the residual. This model resembles the classical income decomposition proposed by Lillard and Willis (1978) in their seminal paper on earnings mobility and later adopted by Geweke and Keane (2000); see Section 2.5 below for a detailed link with the literature. In equation (2.1), log earnings are expressed as the sum of different components. The first one, i.e. $\alpha' X_{it}$, in our empirical analysis is simply a polynomial of individual age, and hence can be regarded as deterministic (i.e. it follows a pre-determined trend). Then, the individual fixed effect η_i stands for worker specific, time-invariant unobservable characteristics, and the time effect λ_t captures year-specific aggregate shifts in the level of log wages due to business-cycle dynamics. Finally, the residual ε_{it} represents the yearly fluctuations around the individual life-cycle and the macro trends. We could adopt more sophisticated specifications for the wage decomposition in equation (2.1), e.g. introducing interactive fixed effects and heteroskedasticity in the error terms.

In the second step we focus on modeling the dynamics of the residual component $\varepsilon_{i,t}$. Given that our interest lies in relative mobility, we consider the individual ranks, i.e. the workers' positions in the cross-sectional distribution of wage residuals, and how they change over time. The uniform rank is defined as $F_{\varepsilon,t}(\varepsilon_{i,t})$, where $F_{\varepsilon,t}(\cdot)$ is the cross-sectional cumulative distribution function (cdf) of the residual component in year t, which we assume to be continuous. By construction, the uniform rank is uniformly distributed on $[0, 1]$ cross-sectionally at any date, and is interpretable as the individual percentile in the cross-sectional distribution of residuals. For modelling convenience, we transform this uniform rank into a Gaussian rank, i.e. a rank which is distributed as a standard normal, by applying the quantile function of the $N(0, 1)$ distribution. Hence, the Gaussian rank of the residual term is:

$$Z_{it} = \Phi^{-1}[F_{\varepsilon,t}(\varepsilon_{it})], \tag{2.2}$$

where Φ is the cdf of the standard normal distribution. In the remainder of this section we specify a semi-nonparametric model for the dynamics of the joint process (Z_{it}, X_{it}) of Gaussian ranks and individual observed characteristics. The fundamental difference between our approach and the vast majority of the existing literature on individual earnings dynamics (see Section 2.5) is that we model the dynamics of the ranks of the residuals instead of modelling the dynamics of the residuals directly (which would correspond to the study of absolute mobility instead).

2.1 The general framework

We start by introducing the assumptions which define a general framework for the joint dynamics of (Z_{it}, X_{it}). We assume independence and identical distribution (iid) across individuals[4].

Assumption 1. *The processes* $\{(Z_{it}, X_{it}), t \in \mathbb{N}\}$, *for* $i = 1, ..., N$, *are i.i.d. across individuals.*

Let us define $Z_t = (Z_{1,t}, ..., Z_{N,t})$ and $X_t = (X_{1,t}, ..., X_{N,t})$. The sample density is:

$$l(\underline{Z_T}, \underline{X_T}) = \prod_{i=1}^{N} l(\underline{Z_{i,T}}, \underline{X_{i,T}}) = \prod_{i=1}^{N} \prod_{t=1}^{T} l(Z_{i,t}, X_{i,t} | \underline{Z_{i,t-1}}, \underline{X_{i,t-1}})$$

where l stands for the (conditional) density of a random variable, $\underline{Z_{i,t-1}} = (Z_{i,t-1}, Z_{i,t-2}, Z_{i,t-3}, ...)$ and $\underline{Z_T} = (Z_T, Z_{T-1}, Z_{T-2}, ...)$.

We focus on modelling the conditional density $l(Z_{i,t}, X_{i,t} | \underline{Z_{i,t-1}}, \underline{X_{i,t-1}})$ for a generic individual i. Let us consider the following decomposition:

$$l(Z_{it}, X_{it} | \underline{Z_{i,t-1}}, \underline{X_{i,t-1}}) = l(X_{it} | \underline{Z_{i,t-1}}, \underline{X_{i,t-1}}) \cdot l(Z_{it} | \underline{Z_{i,t-1}}, \underline{X_{it}}). \tag{2.3}$$

The distribution of the process (Z_{it}, X_{it}) is thus characterized by two conditional densities, which are the transition density of the rank given the regressors history, namely $l(Z_{it} | \underline{X_{it}}, \underline{Z_{i,t-1}})$, and the transition density of the regressors given the past ranks, i.e.

$$l(X_{it} | \underline{X_{i,t-1}}, \underline{Z_{i,t-1}}). \tag{2.4}$$

We use the following Assumptions 2-4 on these conditional densities.

Assumption 2. *Process* (Z_{it}) *does not Granger cause process* (X_{it}), *for any individual* i.

Granger non-causality is equivalent to Sims non-causality (see e.g. Gourieroux and Monfort (1995) Property 1.2, see also Appendix C.2 for background discussion of non-causality concepts). Assumption 2 implies that the conditional density in (2.4) is such that:

$$l(X_{it} | \underline{X_{i,t-1}}, \underline{Z_{i,t-1}}) = l(X_{it} | \underline{X_{i,t-1}}). \tag{2.5}$$

Hence, under Assumption 2, the past ranks do not affect the regressor dynamics, i.e. the individual explanatory variables are exogenous.

[4]The independence across individual ranks implied by Assumption 1 might seem counterintuitive. We stress that this assumption concerns the theoretical ranks in (2.2), i.e. ranks computed with respect to an infinite population. Therefore, mechanical effects from ranks in finite populations are absent here.

Assumption 3. *Process* (X_{it}) *is Markovian of first-order with transition density* $l(X_{it}|X_{i,t-1})$*. We can write* $X_{it} = (X_{it}^{a'}, X_{it}^{b'})$ *and* X_{it}^{a} *is strictly stationary.*

The first-order Markov property implies that Equation (2.5) can be further rewritten as:

$$l(X_{it}|\underline{X_{i,t-1}}, \underline{Z_{i,t-1}}) = l(X_{it}|X_{i,t-1}). \tag{2.6}$$

We assume stationarity only for a subvector X_{it}^{a} of X_{it} which is used for modelling the conditional distribution of Z_{it} given X_{it} (see next subsection). The distribution of $X_{i,t}^{b}$ may be time-varying, as long as Assumptions A.1-A.9 in Appendix A are satisfied.

Assumption 4. *The rank dynamics is such that:* $l(Z_{it}|\underline{X_{it}}, \underline{Z_{i,t-1}}) = l(Z_{it}|X_{it}, X_{i,t-1}, Z_{i,t-1})$*.*

Assumption 4 implies that information about ranks and explanatory variables occurring before time $t-1$ is not relevant in determining the present rank Z_{it}.
From Equations (2.3), (2.6) and Assumption 4 we get:

$$l(Z_{it}, X_{it}|\underline{Z_{i,t-1}}, \underline{X_{i,t-1}}) = l(X_{it}|X_{i,t-1}) \cdot l(Z_{it}|Z_{i,t-1}, X_{it}, X_{i,t-1}). \tag{2.7}$$

As a consequence, the joint process (Z_{it}, X_{it}) is first-order Markov. Equation (2.7) yields functional restrictions on the specification of the transition density of such joint process in our model. The two transition densities of interest are

$$l(Z_{it}|Z_{i,t-1}, X_{it}, X_{i,t-1}) \tag{2.8}$$

and $l(X_{it}|X_{i,t-1})$. The latter is exogenously given, hence we will exclusively focus on the former one. In order for the model to be consistent with the interpretation of Z_{it} as a Gaussian rank, we need to ensure that the stationary density of Z_{it} is standard normal:

$$Z_{it} \sim N(0, 1). \tag{2.9}$$

In the following we prove that, under Assumptions 1-4 and additional constraints on transition density (2.8), the distributional property in (2.9) holds.

2.2 A specification based on copulas

In this subsection we introduce a nonparametric specification for transition density (2.8) based on a copula model. We take advantage of copulas to match the distributional restriction in (2.9). Given Assumption 1, for explanatory purpose we omit the subscript i in the following. Let $c(\cdot, \cdot; \rho)$ be a copula probability density function (pdf) that is indexed by the functional parameter $\rho = \rho(\cdot)$, which possibly depends on observable regressors. Let $g(\cdot|X^a)$ be a pdf of a real variable, for any value of X^a, and let $G(\cdot|X^a)$ be the corresponding cdf. We assume that g is such that:

$$\int g(Z|X^a)l(X^a)dX^a = \phi(Z), \tag{2.10}$$

where $\phi(\cdot) = \Phi'(\cdot)$ is the pdf of the standard normal distribution. Further, let us define:

$$l(Z_t|Z_{t-1}, X_t, X_{t-1}) = g(Z_t|X_t^a)c[G(Z_t|X_t^a), G(Z_{t-1}|X_{t-1}^a); \rho(\cdot, X_t)], \tag{2.11}$$

where $\rho(\cdot, X_t)$ is the functional copula parameter for given X_t. By a change of variable argument and the copula property $\int_0^1 c(u, v)du = 1, \forall v \in [0, 1]$, equation (2.11) defines a valid conditional density function. We come to the main result of this subsection.

Proposition 1. *Assume that the conditional distribution of ranks at date $t = 0$ is such that:*

$$Z_0|X_0 \sim g(\cdot|X_0^a), \tag{2.12}$$

and X_0^a is drawn from the stationary distribution of the regressors vector. Then, under Assumptions 1-4:

$$Z_t \sim N(0, 1) \tag{2.13}$$

for all $t \geq 1$.

The proof of this proposition is provided in Appendix B. Proposition 1 shows that, if the Gaussian rank process is initialized with a conditional distribution $g(\cdot|X_0^a)$ satisfying property (2.10), and the transition density is as in (2.11), then the condition of standard Gaussian marginal distribution for the rank process is met. In fact, the proof of Proposition 1 shows that $g(Z_t|X_t^a)$ is the conditional pdf of Z_t given X_t at any date t, and $c(\cdot, \cdot; \rho(\cdot, X_t))$ is the conditional copula pdf of (Z_t, Z_{t-1}) given X_t, X_{t-1}. Note that the conditional distribution of Z_t given X_t only depends on the stationary part of X_t, i.e. X_t^a, whereas the conditional copula function may depend on both the stationary and the nonstationary parts of X_t[5]. Proposition 1 allows us to introduce a model which is compatible with the condition of standard Gaussianity of the rank Z_t, required for model coherency, in a very general framework. It applies to any copula family, possibly with a functional parameter. This functional parameter, in turn, is allowed to depend on regressor X_t. Note that we do not specify the distribution of X_t, since we aim at deriving a result that holds for any process (X_t). The only requirement that we impose here is that exogenous process (X_t) is Markov and that its component X_t^a is stationary.

[5]It is worth underlying that the assumption of stationarity of the subset of explanatory variables X^a is less restrictive than it may seem. Indeed, as explained above, this assumption is only required to hold for those covariates used in the specification of the univariate conditional distribution of Z_{it}. On the contrary, there is no requirement of stationarity for variables entering into the joint copula distribution of the past and present ranks as long as Assumptions A.1-A.9 in Appendix A are satisfied to establish the large sample properties of estimators. This implies that the model at hand can explain how a variable such as age may influence individual mobility patterns.

2.3 The nonparametric family of autoregressive copulas

In the previous section we have shown how to specify a joint dynamics for rank Z_t and observable regressor X_t, by means of a generic copula function that can be indexed by a functional parameter. In this section we introduce a flexible nonparametric family of copula functions to be used in this setting. These copula functions are inspired by nonlinear first-order autoregressive processes. We first specify the model for the rank Z_t without the exogenous regressor X_t and then show in a second step that the regressor can be easily included as an argument of the copula functional parameter.

Let us consider the nonlinear autoregressive dynamics: [6]

$$Z_t = \Lambda(\rho(Z_{t-1}) + \omega_t) \tag{2.14}$$

where $\omega_t \sim IIN(0,1)$, function Λ is strictly monotonic increasing, and ρ is a function that expresses the dependence between the past and the present individual ranks. The variance of ω_t is normalized to 1. Indeed, a non unit variance can be absorbed into functions ρ and Λ. The larger is the partial derivative of the function $\rho(\cdot)$ with respect to the past rank, the higher the degree of positional persistence. Model (2.14) defines a member of the Generalized Accelerated Failure Time (GAFT) class considered in Ridder (1990) but in a dynamic framework. Also, function Λ^{-1} plays the role of the transformation function in a transformation model (see e.g. Horowitz (1996)). Our focus here lies in the study of the copula associated to this dynamic specification[7]. Equation (2.14) defines a time-homogeneous Markov process. We assume that functions ρ and Λ are such that (Z_t) is a strictly stationary process with unique invariant distribution. When function $\rho(z) = rz$ is linear with coefficient $r \in \mathbb{R}$, it is easily seen by setting $\Lambda(y) = \frac{1}{\sqrt{1+r^2}}y$ that the standard Gaussian distribution is an invariant distribution for process (Z_t). In fact, that case corresponds to a linear Gaussian autoregressive process, with autocorrelation coefficient equal to $\frac{r}{\sqrt{1+r^2}}$. Let us now consider the general case and derive the constraints to impose on functions Λ and ρ such that the invariant distribution of Markov process (Z_t) is $N(0,1)$.

Proposition 2. *For the Markov process defined by equation* (2.14)*, the invariant distribution of* Z_t *is standard normal if, and only if, the function* $\Lambda(\cdot)$ *is given by the following expression:*

$$\Lambda(y) = \Phi^{-1}\left[\int_{-\infty}^{\infty} \Phi(y - \rho(z))\phi(z)dz\right], \tag{2.15}$$

[6]Introducing further lags of the Gaussian ranks in the model would allow to have a better fit. However, with the increase in the number of lags, the problem of increasing dimensionality would arise. Moreover, it would not be straightforward to continue ensuring that the marginal distribution of the Gaussian ranks is standard normal - a condition which is needed for model coherence - in the presence of more than one lagged value of the Gaussian ranks as arguments of the autoregressive function.

[7]We could extend the specification to have a generic distribution for the error process ω_t. This would lead to a copula family characterized by two functional parameters.

for all $y \in \mathbb{R}$.

The proof is provided in Appendix B. Thus, function Λ is a functional of function ρ, and the copula of (Z_t, Z_{t-1}) is completely characterized by function ρ.

Let us now derive the copula pdf associated with the nonlinear autoregressive model (2.14). We have the conditional cdf and pdf $P(Z_t \leq z | Z_{t-1} = \zeta) = \Phi[\Lambda^{-1}(z) - \rho(\zeta)]$ and $l_{Z_t|Z_{t-1}}(z|\zeta) = \phi[\Lambda^{-1}(z) - \rho(\zeta)] \cdot \frac{1}{\lambda[\Lambda^{-1}(z)]}$, where λ is the first-order derivative of the function Λ. Therefore, we obtain the following joint pdf of Z_t and Z_{t-1}:

$$l_{Z_t,Z_{t-1}}(z,\zeta) = l_{Z_t|Z_{t-1}}(z|\zeta) \cdot l_{Z_{t-1}}(\zeta) = \phi[\Lambda^{-1}(z) - \rho(\zeta)] \cdot \frac{1}{\lambda[\Lambda^{-1}(z)]} \cdot \phi(\zeta), \quad z, \zeta \in \mathbb{R}.$$

We can now get the explicit copula density of Z_t and Z_{t-1}. The copula pdf is:

$$c(u, v; \rho(\cdot)) = \frac{l_{Z_t,Z_{t-1}}(\Phi^{-1}(u), \Phi^{-1}(v))}{l_{Z_{t-1}}(\Phi^{-1}(v)) \cdot l_{Z_t}(\Phi^{-1}(u))} = \frac{\phi[\Lambda^{-1}(\Phi^{-1}(u)) - \rho(\Phi^{-1}(v))]}{\phi(\Phi^{-1}(u))\lambda(\Lambda^{-1}(\Phi^{-1}(u)))}, \tag{2.16}$$

for the arguments $u, v \in [0, 1]$. This copula family is parametrized by the autoregressive function $\rho(\cdot)$. It contains the Gaussian copula as a special case, that corresponds to a linear autoregressive function $\rho(z) = rz$ for $r \in \mathbb{R}$, with Gaussian copula parameter $\frac{r}{\sqrt{1+r^2}}$. The copula $c(\cdot, \cdot; \rho(\cdot))$ is unchanged under a constant shift of the functional parameter $\rho(\cdot) \to \rho(\cdot) + c$, where c is constant. Therefore, we assume $\rho(0) = 0$ for identification purposes.

We combine the results from Subsections 2.2 and 2.3 to obtain our model. Specifically, first we reintroduce the individual index i. Second, we introduce the regressor vector $X_{i,t}$ in the copula functional parameter, which becomes $\rho(\cdot, X_{i,t})$. Note that this is possible since essentially any function $\rho(\cdot)$ is admissible as a parameter of the autoregressive copula. Hence, the distribution of Z_{it} given $Z_{i,t-1}$, X_{it}, $X_{i,t-1}$ is:

$$l(Z_{it}|Z_{i,t-1}, X_{it}, X_{i,t-1}) = g(Z_{it}|X_{it}^a)c[G(Z_{it}|X_{it}^a), G(Z_{i,t-1}|X_{i,t-1}^a); \rho(\cdot, X_{it})] \tag{2.17}$$

where $c(\cdot, \cdot; \rho(\cdot, X_{i,t}))$ is the autoregressive copula function defined in (2.16) with functional parameter $\rho(\cdot, X_{i,t})$, and $\Lambda(\cdot)$ replaced by

$$\Lambda(y; X_{it}) = \Phi^{-1}\left[\int_{-\infty}^{\infty} \Phi[y - \rho(z, X_{it})]\phi(z)dz\right], \tag{2.18}$$

and $g(\cdot|X_{it}^a)$ is a pdf that satisfies the condition in equation (2.10) for the stationary distribution of X_{it}^a. The functional parameter is normalized such that $\rho(0, X_{it}) = 0$, for any value of X_{it}. Note that, differently from what happened in the marginal distributions, there is no necessity for process (X_t) in the autoregressive function $\rho(\cdot, \cdot)$ to be stationary to ensure the distributional restriction $Z_{it} \sim N(0, 1)$.

The model defined by (2.17) admits a stochastic nonlinear autoregressive representation. Indeed, let us define the process:

$$\xi_{it} \equiv \Phi^{-1}[G(Z_{it}|X_{it}^a)]. \tag{2.19}$$

The variables $\xi_{i,t}, \xi_{i,t-1}$ have the same copula as variables $Z_{i,t}, Z_{i,t-1}$, conditional on $X_{i,t}, X_{i,t-1}$, with standard Gaussian marginal distributions. Thus, the stochastic representation of our model is as follows:

$$\xi_{it} = \Lambda[\rho(\xi_{i,t-1}; X_{it}) + \omega_{it}; X_{i,t}] \tag{2.20}$$

where $\omega_{it} \sim IIN(0,1)$ is independent of $X_{i,t}$, and function $\Lambda(\cdot; X_{it})$ is given in (2.18). This corresponds to a nonlinear autoreressive dynamics for ξ_{it} driven by the exogenous process X_{it}. Then, from (2.19) the Gaussian ranks are $Z_{it} = G^{-1}(\Phi(\xi_{it})|X_{it}^a)$.

Our semi-nonparametric approach has practical advantages compared to fully nonparametric estimation of the copula (Fermanian and Scaillet (2003)). In fact, when replacing regressor X_t with a linear index as in the next sections, in our model the copula is parametrized by a bivariate unknown function ρ. The fully nonparametric approach involves instead a trivariate conditional copula function that may be hard to estimate accurately.

2.4 A functional relative mobility measure

We now want to derive an adequate measure of positional mobility in our model. We define mobility as the partial derivative of the conditional expected rank with respect to the past rank:

$$m(Z_{i,t-1}; X_{i,t}, X_{i,t-1}) = \frac{\partial E(Z_{it}|Z_{i,t-1}, X_{it}, X_{i,t-1})}{\partial Z_{i,t-1}}. \tag{2.21}$$

Proposition 3. *In the framework of our model, the mobility measure takes the following explicit form:*

$$m(Z_{i,t-1}; X_{i,t}, X_{i,t-1}) = \int_{-\infty}^{\infty} \frac{\tilde{\lambda}[\tilde{\rho}(G(Z_{i,t-1}|X_{i,t-1}^a); X_{it}) + \omega; X_{it}]\phi(\omega)}{g\left[G^{-1}\left(\tilde{\Lambda}[\tilde{\rho}(G(Z_{i,t-1}|X_{i,t-1}^a); X_{it}) + \omega; X_{it}]|X_{it}^a\right)|X_{i,t}^a\right]} d\omega \times \tilde{\rho}'[G(Z_{i,t-1}|X_{i,t-1}^a); X_{it}]g(Z_{i,t-1}|X_{i,t-1}^a), \tag{2.22}$$

where $\tilde{\rho}(u; X_{it}) = \rho(\Phi^{-1}(u), X_{it})$, $\tilde{\Lambda}(k; X_{it}) = \int_0^1 \Phi(k - \tilde{\rho}(v; X_{it}))dv$, and $\tilde{\lambda} = \tilde{\Lambda}'$ and $\tilde{\rho}'$ are the derivatives of functions $\tilde{\Lambda}$ and $\tilde{\rho}$ w.r.t. their first argument.

This functional measure accounts for mobility in the different parts of the distribution of past ranks and for the effects of covariates. It depends on both functions g and ρ. Note that the expression in (2.22) provides rather an immobility measure, since the larger its (absolute) value is, the stronger the association between the past and the present ranks is.

In addition to the conditional expected rank and its partial derivative in (2.22), from the stochastic representation of the model we can also easily derive the conditional quantiles, which have the advantage of being computationally simpler to obtain, since they do not involve a numerical integral. The conditional quantiles are interesting quantities, since they 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} for percentile $u \in (0,1)$ can be derived from the following equation:

$$\begin{aligned} P(Z_{it} \leq z|Z_{i,t-1}, X_{it}, X_{i,t-1}) &= P(\gamma(Z_{i,t-1}, \omega_{it}, X_{it}, X_{i,t-1}) \leq z|Z_{i,t-1}, X_{it}, X_{i,t-1}) \\ &= \Phi\left[\gamma^{-1}(Z_{i,t-1}, z, X_{it}, X_{i,t-1})\right] = u, \end{aligned} \tag{2.23}$$

where $Z_{it} = \gamma(Z_{i,t-1}, \omega_{it}, X_{it}, X_{i,t-1}) = G^{-1}\left[\tilde{\Lambda}[\tilde{\rho}(G(Z_{i,t-1}|X^a_{i,t-1}); X_{it}) + \omega_{it}; X_{it}]|X^a_{it}\right]$ from (2.19)-(2.20) and γ^{-1} denotes the inverse of function γ with respect to the second argument. Hence,

$$\begin{aligned} Q_{Z,t}(u|Z_{i,t-1}, X_{it}, X_{i,t-1}) &= \gamma(Z_{i,t-1}, \Phi^{-1}(u), X_{it}, X_{i,t-1}) \\ &= G^{-1}\left[\tilde{\Lambda}[\tilde{\rho}(G(Z_{i,t-1}|X^a_{i,t-1}); X_{it}) + \Phi^{-1}(u); X_{it}]|X^a_{it}\right], \end{aligned} \tag{2.24}$$

for any percentile $u \in (0,1)$. 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:

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

$$= \frac{\tilde{\lambda}[\tilde{\rho}(G(Z_{i,t-1}|X^a_{i,t-1}); X_{it}) + \Phi^{-1}(u); X_{it}]\tilde{\rho}'[G(Z_{i,t-1}|X^a_{i,t-1}); X_{it}]g(Z_{i,t-1}|X^a_{i,t-1})}{g\left[G^{-1}\left(\tilde{\Lambda}[\tilde{\rho}(G(Z_{i,t-1}|X^a_{i,t-1}); X_{it}) + \Phi^{-1}(u); X_{it}]|X^a_{it}\right)|X^a_{i,t}\right]}. \tag{2.25}$$

Until now, we focused on mobility of the residual component ε_{it}. We complete our study of functional relative mobility measures by analyzing overall wage mobility. The Gaussian wage rank is defined as

$$r_{i,t} = \Phi^{-1}[F_{y,t}(y_{it})]$$

where $F_{y,t}(\cdot)$ is the cdf of the wage distribution in given year t. From the wage equation (2.1), the relationship between the wage rank and the residual component rank is given by:

$$r_{i,t} = \Phi^{-1}[F_{y,t}(\beta_0' X_{it} + \eta_i + \lambda_t + F^{-1}_{\varepsilon,t}(\Phi(Z_{i,t})))] \equiv b_t(Z_{i,t}, \eta_i, X_{i,t}). \tag{2.26}$$

The relationship between ranks $r_{i,t}$ and $Z_{i,t}$ depends on the individual fixed effect η_i, the observable regressors vector $X_{i,t}$ and the calendar time t (via the time fixed effect and cdfs $F_{y,t}$, $F_{\varepsilon,t}$). Equation (2.24) together with equation (2.26) allow us to easily obtain the conditional quantiles of the wage rank as a function of the past residual rank. The conditional quantile for percentile $u \in (0,1)$ is:

$$Q_{r,t}(u|Z_{i,t-1}, \eta_i, X_{i,t}, X_{i,t-1}) = b_t[Q_{Z,t}(u|Z_{i,t-1}, X_{i,t}, X_{i,t-1}), \eta_i, X_{i,t}]. \tag{2.27}$$

The partial derivative of the conditional quantiles w.r.t. $Z_{i,t-1}$ (or w.r.t. $r_{i,t-1}$) yield a wage immobility measure. In Section 5, we estimate such immobility measures on a dataset of US workers.

2.5 Link with the literature on individual earnings dynamics

In this section we relate our paper to the literature on individual earnings dynamics. A vast part of this literature builds on the decomposition of the log wage $y_{i,t} = P_{i,t} + T_{i,t}$ as the sum of a permanent and a transitory components, $P_{i,t}$ and $T_{i,t}$, respectively. The transitory component is typically modelled as either a white noise, or a moving average process. In early contributions, the permanent component is modeled as a fixed effect, i.e. $P_{i,t} = P_i$ (Lillard and Willis (1978), MaCurdy (1982), Abowd and Card (1989)). In Blundell, Pistaferri and Preston (2008), Hryshko (2012), Jensen and Shore (2015), Hu, Moffitt, and Sasaki (2019), among others, the permanent component follows a random walk, i.e. $P_{i,t} = P_{i,t-1} + I_{i,t}$, where $I_{i,t}$ is the innovation of the random walk. Further, Meghir and Pistaferri (2004) and Botosaru and Sasaki (2018) introduce conditional heteroschedasticity in the innovation of the random walk driving the permament component. If we neglect for a moment the observed characteristics $X_{i,t}$ and the time effect λ_t, we can interpret the empirical specification in equation (2.1) as a reduced form model, in which $\eta_i = P_{i,0}$ is the initial value of the permanent component, and $\varepsilon_{i,t} = \sum_{s=1}^{t} I_{i,s} + T_{i,t}$ is a superposition of the transitory component and the cumulated innovations of the permanent component. We depart from this literature in that, instead of modeling the dynamics of $\varepsilon_{i,t}$ through the structural components $P_{i,t}$ and $T_{i,t}$, we specify a semi-nonparametric model for the Gaussian ranks $Z_{i,t}$. This modeling choice is motivated by our focus on relative mobility as opposed to absolute mobility. Arellano, Blundell, and Bonhomme (2017) consider a non-separable nonlinear dynamics for the permanent component. They define a measure of nonlinear persistence as the derivative of the conditional quantile function of the permanent component $P_{i,t}$ w.r.t. the lagged value $P_{i,t-1}$. Our measure of relative mobility defined as the partial derivative of the conditional quantile of the Gaussian rank in (2.25) is a counterpart of their measure in our framework.

As already remarked in the Introduction, our empirical focus on relative wage mobility modeled via a copula makes our paper closer in spirit to 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 wage ranks. Relative mobility has been studied e.g. by Shorrocks (1978) and Cowell and Flachaire (2018), with a theoretical focus on the properties of the rank ordering, as well as Formby et al. (2004) and Van Kerm (2004). Other empirical work on relative mobility used transition matrices between deciles of the cross-sectional wage distribution, see e.g. Shorrocks (1978) [8].

Studies on individual labor earnings dynamics often distinguish between models with heterogeneous income profiles (HIPs) vs. restricted income profiles (RIPs). The first class of models allows for substantial unobserved heterogeneity in individual earnings dynamics, see e.g. Browning et al. (2010), Guvenen (2009). Browning et al. (2010) find that allowance for latent heterogeneity is empirically relevant and makes substantial difference to inferences of interest. On the other hand, for RIP specifications the coefficients of the dynamic models representing income dynamics are constant across individuals (for this reason, this type of model is also called "homogeneous income profiles"). In our specification, labor income profiles are homogeneous after controlling for observable characteristics, except for the individual fixed effect included in equation (2.1). Hence, our model falls substantially within the RIP approach, sharing this feature with the models in e.g. Bonhomme and Robin (2009), Hryshko (2012), Arellano, Blundell and Bonhomme (2017) and several other contributions in the literature reviewed above. Including unobserved heterogeneity in our semi-nonparametric copula specification is a challenging avenue for future research.

3 Estimation of the mobility model

Let us now discuss the estimation of the semi-nonparametric copula model of Section 2. To cope with the curse of dimensionality in nonparametric estimation, we assume an index model specification for the effect of the observable characteristics:

$$l(Z_{i,t}|Z_{i,t-1}, X_{i,t}, X_{i,t-1}) = g(Z_{i,t}; X^{a\prime}_{i,t}\beta^0_1) \cdot c[G(Z_{i,t}; X^{a\prime}_{i,t}\beta^0_1), G(Z_{i,t-1}; X^{a\prime}_{i,t-1}\beta^0_1), \rho(\cdot, X'_{i,t}\beta^0_2)]$$

where $G(\cdot; X^{a\prime}_{i,t}\beta^0_1)$ is the cdf of the distribution of the rank, conditional on the individual variables, $g(\cdot; X^{a\prime}_{i,t}\beta^0_1)$ is the corresponding pdf, $c[\cdot, \cdot, \rho(\cdot, X'_{it}\beta^0_2)]$ is the copula density in (2.16), and $\beta^0_1 \in \mathbb{R}^{p_1}$ and $\beta^0_2 \in \mathbb{R}^{p_2}$ are parameter vectors. The model features two indexes, namely $W_{1,it} = X^{a\prime}_{it}\beta^0_1$ for the marginal univariate distribution, and $W_{2,it} = X'_{it}\beta^0_2$ 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 worker's position in the cross-sectional distribution at a generic date. On the other hand,

[8] Individual rank mobility differs from earnings volatility, i.e. the variance or standard deviation of earnings, or the expectation of squared individual earnings changes, which has been studied, for example, by Gottschalk (1982, 1997), Meghir and Pistaferri (2004) and Jensen and Shore (2015). It also differs from aggregate or "macro" mobility, i.e. the average degree of wage mobility in a certain economy (Fields and Ok (1999)). This latter concept has been studied by e.g. Burkhauser and Poupore (1997), Maasoumi and Trede (2001), Moffitt and Gottschalk (2002), Baker and Solon (2003), Auten et al. (2013), and Kopczuk et al. (2010). Arellano, Blundell and Bonhomme (2017) propose relevant advances in the modellization of aggregate mobility, by studying nonlinear persistence of the earnings process. The focus of the authors lies in macro-persistence of income, i.e. evaluated at different (aggregate) percentiles.

the second index, called *mobility score*, accounts for the role of the same variables in determining the degree of mobility of the residual wage component.

In a first step we estimate the unobservable values of the Gaussian ranks $Z_{i,t}$ by means of the empirical Gaussian ranks. More specifically, from the standard fixed-effect estimation of the linear panel model (2.1) we get estimated residuals $\hat{\varepsilon}_{it}$, i.e. the estimated residual wage component. Then, we build the empirical Gaussian ranks:

$$\hat{Z}_{it} = \Phi^{-1}[\hat{F}_{\varepsilon,t}(\hat{\varepsilon}_{it})] \tag{3.1}$$

where $\hat{F}_{\varepsilon,t}(\varepsilon) = \frac{1}{n}\sum_{j=1}^{n} 1\{\hat{\varepsilon}_{jt} \leq \varepsilon\}$ is the empirical cross-sectional distribution of residuals in year t. In the asymptotic analysis performed in Section 3.3, we consider the case $T, N \to \infty$, as it is usual in the panel data literature on incidental parameters bias correction (see references in Section 3.3). When T and N both go to infinity, under the regularity conditions listed in Appendix A, the difference between $\hat{Z}_{it}$ and Z_{it} vanishes asymptotically.

We now turn to the estimation of the finite-dimensional and functional parameters in the ranks dynamics. We estimate the parameters of the marginal univariate distribution and those of the copula function sequentially. The alternative procedure consisting in estimating jointly the wage equation with individual fixed effects and the nonlinear rank dynamics is computationally challenging.

3.1 Univariate conditional rank distributions

Let us first consider the estimation of the univariate conditional distribution $G(\cdot; X_{i,t}^{a\prime}\beta_1^0)$. The use of a linear combination $X_{it}^{a\prime}\beta_1^0$ of the explanatory variables corresponds to a semiparametric single-index model. We first define an unconstrained nonparametric estimator of pdf g and then impose the constraint (2.10) by an information theoretic approach. We estimate the coefficient vector β_1^0 with a kernel single-index Maximum Likelihood approach applied on the empirical ranks:

$$\hat{\beta}_1 = \arg\max_{\beta_1 \in B_1} \sum_{i=1}^{N}\sum_{t=1}^{T} k_{it} \log\ \hat{g}_{-(i,t)}(\hat{Z}_{i,t}|\beta_1' X_{it}^a; \beta_1), \tag{3.2}$$

where $B_1 \subset \mathbb{R}^{p_1}$ is the parameter set, $\hat{g}_{-(i,t)}(z|w;\beta_1)$ is the leave-one-out conditional kernel density computed on the empirical ranks $\hat{Z}_{it}$ given $W_{1,it}(\beta_1) = w$, where $W_{1,it}(\beta_1) = X_{it}^{a\prime}\beta_1$. The leave-one-out conditional kernel density is defined by:

$$\hat{g}_{-(i,t)}(z|w;\beta_1) = \frac{\frac{1}{h_z}\sum_{s=1,s\neq t}^{T}\sum_{j=1,j\neq i}^{N} K\left(\frac{z-\hat{Z}_{jt}}{h_z}\right) K\left(\frac{w - W_{1,jt}(\beta_1)}{h_w}\right)}{\sum_{s=1,s\neq t}^{T}\sum_{j=1,j\neq i}^{N} K\left(\frac{w - W_{1,jt}(\beta_1)}{h_w}\right)} \tag{3.3}$$

where K is a kernel density that satisfies Assumption A.1 in Appendix A, and h_z and h_w are the bandwidths or smoothing parameters that satisfy Assumption A.2 in Appendix A. Moreover, k_{it} is the trimming term introduced by Rosemarin (2012)[9]. By plugging in the estimate $\hat{\beta}_1$ of β_1^0, we get an estimator of the pdf:

$$\breve{g}(z|x^a) = \frac{\frac{1}{H_z}\sum_{t=1}^{T}\sum_{i=1}^{N} K\left(\frac{z-\hat{Z}_{it}}{H_z}\right) K\left(\frac{\hat{\beta}_1'(x-X_{it}^a)}{H_w}\right)}{\sum_{t=1}^{T}\sum_{i=1}^{N} K\left(\frac{\hat{\beta}_1'(x-X_{it}^a)}{H_w}\right)}. \tag{3.4}$$

As in Rosemarin (2012) we adopt different bandwidths H_z, H_w in density estimator (3.4) since bandwidths h_z, h_w used for estimator $\hat{\beta}_1$ may necessitate undersmoothing for technical reasons (see Appendix A). Note that there are several approaches to estimate a single-index model (e.g. Powell et al. (1989), Härdle et al. (1993), Ichimura (1993), Horowitz and Härdle (1996), Li and Racine (2007)). Our choice is motivated by the existing literature, which is widely developed on the regression function, but is relatively more scarce on density function estimation.

We take into account the restriction (2.10) on the univariate conditional rank distribution by looking for the conditional pdf that minimizes an average Kullback-Leibler distance from the nonparametric estimator $\breve{g}(z|x^a)$ in (3.4), subject to the sample analogue of restriction (2.10). The estimator is the solution of the following constrained minimization problem:

$$\hat{g}(\cdot|\cdot) = \underset{g(\cdot|\cdot)}{\arg\min} \int_{\mathcal{X}_{N,T}} \int_{\mathcal{Z}_{N,T}} \log\left[\frac{g(z|x^a)}{\breve{g}(z|x^a)}\right] g(z|x^a)dz d\hat{F}_x(x^a)$$

$$s.t. \int_{\mathcal{X}_{N,T}} g(z|x^a)d\hat{F}_x(x^a) = \phi(z), \quad \forall z \in \mathcal{Z}_{N,T}, \tag{3.5}$$

$$\int_{\mathcal{Z}_{N,T}} g(z|X_{it}^a)dz = 1, \quad \forall i = 1,...,N, t = 1,...,T, \quad \text{such that} \quad X_{it}^a \in \mathcal{X}_{N,T} \tag{3.6}$$

where $\hat{F}_x$ denotes the empirical distribution of the regressor, namely $\int_{\mathcal{X}_{N,T}} g(z|x^a)d\hat{F}_x(x^a) = \frac{1}{NT}\sum_{i=1}^{n}\sum_{t=1}^{T} 1_{\mathcal{X}_{N,T}}(X_{it}^a)g(z|X_{it}^a)$. Here $\mathcal{Z}_{N,T}$ and $\mathcal{X}_{N,T}$ denote compact subsets of $\mathbb{R}$ and $\mathbb{R}^{p_1}$ that depend on sample sizes N, T. By solving this functional constrained minimization problem, we get the

[9]This trimming term is defined as follows: $k_{it} = \mathbb{I}_{it} \times \left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbb{I}_{it}\right)^{-1}$, where $\mathbb{I}_{it} = 1$ if $min\left\{\frac{1}{h_z}\sum_{s=1,s\neq t}^{T}\sum_{j=1,j\neq i}^{N} K\left(\frac{\hat{Z}_{jt}-z}{h_z}\right)K\left(\frac{W_{1,jt}(\beta_1)-w}{h_w}\right), \sum_{s=1,s\neq t}^{T}\sum_{j=1,j\neq i}^{N} K\left(\frac{W_{1,jt}(\beta_1)-w}{h_w}\right)\right\} > a_0 n^{-c}$ and zero otherwise, for some small constants $a_0, c > 0$. Thus, the indicator $\mathbb{I}_{it}$ is normalized to account for the actual number of observations considered in the computation of the log-likelihood.

estimator (see Appendix B.4 for the proof):

$$\hat{g}(z|x^a) = \frac{\breve{g}(z|x^a)e^{\mu(z)}}{\int_{\mathcal{Z}_{N,T}} \breve{g}(z|x^a)e^{\mu(z)}dz}, \tag{3.7}$$

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

$$e^{\mu(z)} = \phi(z) \left\{ \int_{\mathcal{X}_{N,T}} \breve{g}(z|x^a) \left[\int_{\mathcal{Z}_{N,T}} \breve{g}(z|x^a)e^{\mu(z)}dz \right]^{-1} d\hat{F}_x(x^a) \right\}^{-1}. \tag{3.8}$$

The constrained estimator $\hat{g}$ turns out to be the unconstrained estimator $\breve{g}$ multiplied by a tilting factor, which imposes the constraints. This approach is reminiscent of information theoretic approaches to estimate moment restriction models (Kitamura and Stutzer (1997), Imbens et al. (1998), Kitamura (2007)). The estimator of the cdf is obtained by integration:

$$\hat{G}(z_{it}|x^a_{it}) = \int_{l_{z,N,T}}^{z_{it}} \hat{g}(z|x^a_{it})dz, \tag{3.9}$$

where $l_{z,N,T}$ is the lower boundary of set $\mathcal{Z}_{N,T}$.

3.2 Estimation of the copula model

We now estimate the copula pdf $c(\cdot,\cdot,\rho(\cdot, X'_{i,t}\beta^0_2))$. We perform a simultaneous M-estimation of the parameter vector β^0_2 and of the autoregressive function $\rho(\cdot,\cdot)$ via the method of Sieves, in the spirit of e.g. Wong and Severini (1991) and Chen and Shen (1998) (see also Chen (2007) for a survey). The main idea of this method, which has been first developed by Grenander (1981) and Geman and Hwang (1982), is to estimate an unknown function by means of the approximating space generated by a set of basis functions. In our case, the estimation is performed via the following Sieve Maximum Likelihood procedure applied on the empirical ranks:

$$\hat{\theta} = \arg\max_{\theta \in \Theta_{N,T}} \sum_{i=1}^{N} \sum_{t=2}^{T} l(\hat{Y}_{it}, \hat{Y}_{i,t-1}, \theta, \hat{f}) \tag{3.10}$$

where:

$$l(\hat{Y}_{it}, \hat{Y}_{i,t-1}, \theta, f) \equiv \log c[\tilde{G}(\hat{Y}_{it}, f), \tilde{G}(\hat{Y}_{i,t-1}, f); \rho(\cdot, X'_{it}\beta_2)], \tag{3.11}$$

$\hat{Y}_{it} = (\hat{Z}_{it}, X^{a\prime}_{it})'$, $\tilde{G}(\hat{Y}_{it}, f) = \int_{-\infty}^{\hat{Z}_{it}} f(Z|X^a_{it})dZ$ is the conditional cdf of Z_{it} given $X^{a\prime}_{it}\beta^0_1$ evaluated at the empirical rank, and f is the functional parameter governing the univariate distribution, $\hat{f}(z|x^a) = \hat{g}(z|x^a)$ is the estimator defined by equation (3.7), $\theta = (\beta_2, \rho(\cdot,\cdot)) \in B_2 \times \mathbb{H} \equiv \Theta$, and $\Theta_{N,T} = B_2 \times \mathbb{H}_{N,T}$. Hence, the criterion function involves the first-step nonparametric estimate $\hat{f}$. Note that parameter

θ governing the copula includes both a finite-dimensional and an infinite-dimensional component. $\mathbb{H}$ is an infinite-dimensional space of bivariate functions and $\mathbb{H}_{N,T}$ is a bivariate Sieve space (made up by Hermite polynomials in our implementation) whose dimension depends on sample sizes N and T. More specifically, the approximation $\rho \in \mathbb{H}_{N,T}$ is

$$\rho(Z_{i,t-1}, W_{2,it}) = \sum_{k,l=0}^{m} \lambda_{k,l} H_k(W_{2,it}) H_l(Z_{i,t-1}) \tag{3.12}$$

where $\lambda_{k,l}$ are the coefficients of the polynomial basis used to approximate $\rho(\cdot)$. The number of polynomials used to approximate the autoregressive function depends on the dimension of the sample: $m = m(N,T)$ (Chen (2007)). By the linear independence of the Hermite polynomials, the normalization constraint $\rho(0, W_{2,it}) = 0$ for any value of $W_{2,it}$ holds under the set of linear constraints:

$$\sum_{l=0}^{m} \lambda_{k,l} H_l(0) = 0, \qquad k = 0, 1, ..., m, \tag{3.13}$$

which are imposed on the elements of $\mathbb{H}_{N,T}$. Once the estimators of the rank dynamics are obtained from (3.7) and (3.10), we estimate the functional mobility measures of Section 2.4 by plug-in.

3.3 Consistency of the estimators

We adopt the panel asymptotics with N, T going to infinity jointly. This double asymptotics is standard in the literature on bias correction for the incidental parameter problem (e.g. Hahn and Newey (2004), Fernandez-Val and Vella (2011), Hahn and Kuersteiner (2011), Fernandez-Val and Lee (2013), Fernandez-Val and Weidner (2016)). In our case $T \to \infty$ together with $N \to \infty$ implies a vanishing effect from estimating the true rank $Z_{i,t}$ with the empirical rank $\hat{Z}_{i,t}$. If T were fixed in the asymptotics, the estimation error induced by the individual effects when computing the residuals $\hat{\varepsilon}_{i,t}$ would yield a bias term at order $O(1/T)$. We leave for future research the derivation of a bias correction for such effect. In the Monte Carlo experiments in Section 4 we check that the bias is small in a setting that mimics our empirical analysis.

To show the convergence of our estimated univariate cdfs to the true ones, we suppose that Assumptions 1-4 in Section 2.1 and A.1-A.5 in Appendix A hold. We build on the results of Rosemarin (2012), who extended the work by Delecroix et al. (2003) on semi-nonparametric estimation of index models, and adapt them to a panel framework. In particular, we control the estimation error on the empirical ranks with Assumption A.5.

Proposition 4. *Under Assumptions 1-4 and A.1-A.5: (i) The estimator $\hat{\beta}_1$ defined in equation* (3.2) *is consistent:*

$$\hat{\beta}_1 \xrightarrow[N,T\to\infty]{p} \beta_1^0,$$

where $\xrightarrow[N,T\to\infty]{p}$ denotes convergence in probability as $N, T \to \infty$. (ii) The estimator $\hat{g}$ defined in (3.7) *satisfies:*

$$\sup_{(z,x^a)\in\mathbb{S}} |\hat{g}(z|x^a) - g(z|\beta_1^{0'} x^a)| \xrightarrow[N,T\to\infty]{p} 0,$$

where $\mathbb{S} \subset \mathbb{R}^{p_1+1}$ is a compact set introduced to control boundary effects.

Let us now consider the estimator of the copula parameter. We suppose that Assumptions A.6-A.9, which are reported in Appendix A, hold. In particular, we consider a norm $||\cdot||_\Theta$ on parameter set Θ which satisfies the requirements of Assumption A.6. The following proposition is obtained by extending Theorem 2.1 of Hahn, Liao and Ridder (2018), which proves the consistency of nonparametric two-step Sieve estimators, to a panel framework and controlling the estimation error on the empirical ranks by means of Assumption A.9.

Proposition 5. *If Assumptions 1-4 and A.1-A.9 hold, then the second-step Sieve M-estimator $\hat{\theta}$ defined in* (3.10) *is consistent:*

$$||\hat{\theta} - \theta_0||_\Theta \xrightarrow[N,T\to\infty]{p} 0.$$

The consistency of estimator $\hat{\theta}$ in Proposition 5 implies the consistency of both the finite-dimensional and infinite dimensional parts $\hat{\beta}_2$ and $\hat{\rho}(\cdot)$ in the Euclidean norm and the implied function space norm, respectively[10].

4 Monte Carlo simulations

In order to assess the finite sample properties and the numerical stability of our estimation algorithm, we perform 1000 Monte Carlo simulations and we compare the estimated copula autoregressive function and mobility measures with the true ones. The data generating process and the sample sizes used for the simulations have been calibrated on the data which are used for the empirical analysis in the next section. For each individual i, we randomly draw her/his history of the covariates $X_{i,t}$ from the PSID dataset described in Section 5.1. This yields an unbalanced panel with $n = 1320$ individuals with an

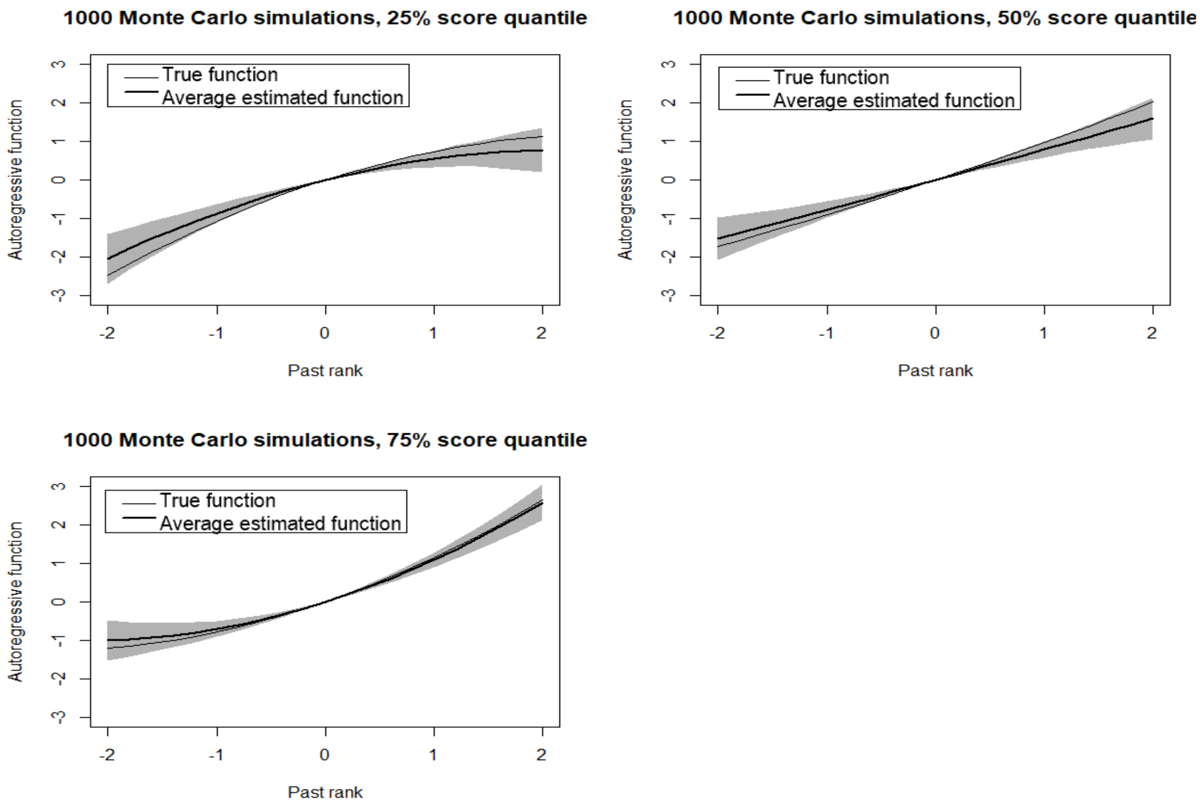
[10]The results in Hahn, Liao and Ridder (2018) might be used to establish asymptotic normality of vector-valued functionals of θ. Controlling for the effect of preliminary estimation of the ranks is not simple. Moreover, given the focus in our empirical application on the patterns of the mobility measures, rather than on finite-dimensional features of them, we do not investigate the asymptotic properties beyond consistency.

average permanence in the sample of about 25 periods (years). Then, we randomly draw $Z_{i,0}$, i.e. the initial Gaussian rank, from a standard normal distribution. By applying the autoregressive copula in equations (2.19)-(2.20) with the estimated parameters of the next section, starting from $Z_{i,0}$ we simulate a trajectory of individual Gaussian ranks Z_{it}. Now, using the pdf f_ε corresponding to the estimate in the empirical analysis, we transform the simulated Z_{it} into simulated ε_{it}. We then simulate the individual effects η_i by randomly drawing from a discrete distribution corresponding to the values of the η_i which are estimated in the empirical part. Similarly, for each year the time fixed effects λ_t are taken from the estimation of the linear two-way fixed effect model which is performed in Section 5. In this way we are able to construct an unbalanced simulated panel of y_{it} and $X_{i,t}$, whose cross-sectional and time-series dimensions, as well as the unbalanced panel properties, are similar to the ones of the sample used in our empirical analysis. This allows us to evaluate the size of the bias in the framework of our empirical model. On this simulated sample we run the whole estimation procedure presented in Section 3 for both the functional and the finite-dimensional parameters. We repeat this procedure 1000 times.

In the following analysis of the Monte Carlo results, we focus on the bias of the autoregressive function $\rho(\cdot;\cdot)$ and the relative mobility measure $m(\cdot)$, that are the primary interest of our empirical analysis. Simulation results for the marginal rank distribution could be easily obtained with the same procedure as well. The main message of the figures below is that the bias due to the small time series dimension of our panel is moderate for different values of both the mobility score and the past rank.

From Figure 1, we notice that the true function ρ lies within the 95% point-wise confidence intervals that we constructed on the basis of the 1000 Monte Carlo simulations. In particular, the bias is small in the top 75% mobility score quantile, especially for positive values of the past rank. The width of the confidence intervals themselves is relatively small, which endorses the accuracy of our estimation strategy for sample sizes similar as the ones used in the empirical analysis. The bias is positive when the past Gaussian rank is negative and vice versa, for all the values of the mobility score considered in our design. Further, the bias is smaller for central values of the past Gaussian rank, and larger for more extreme values of the past Gaussian rank.

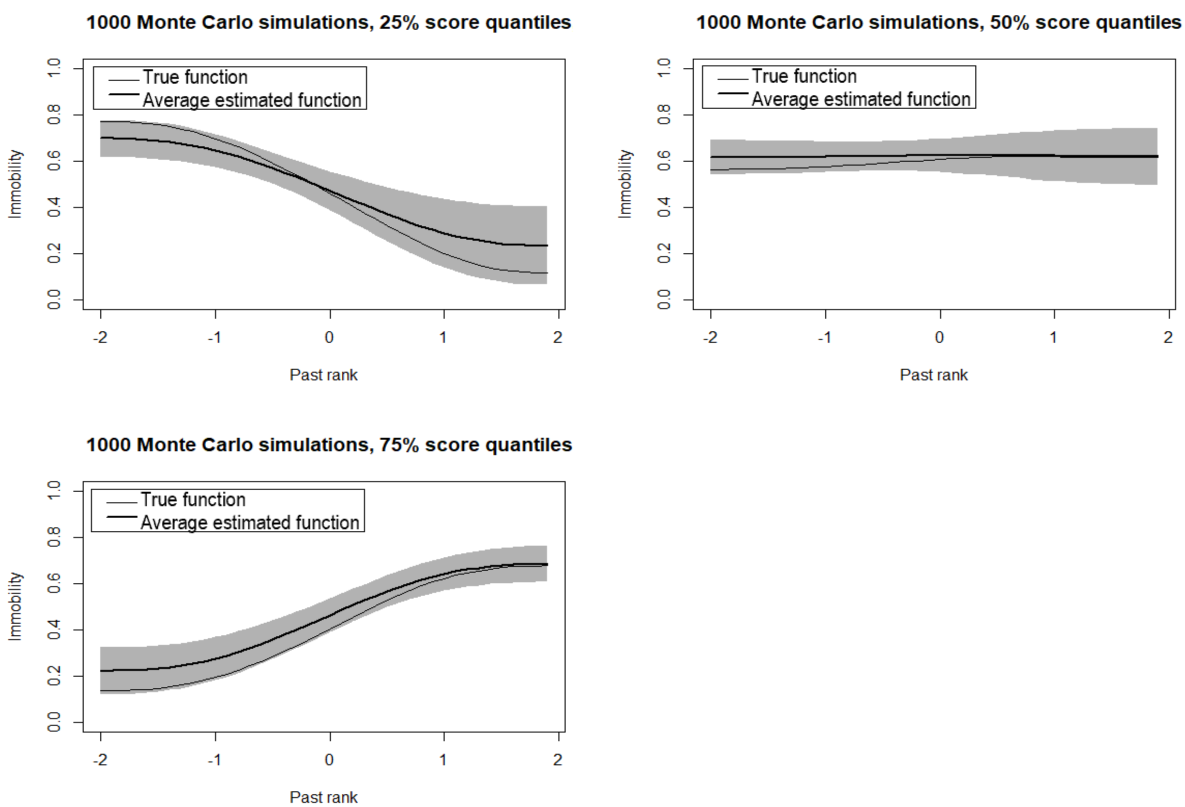
Figure 1: Monte Carlo simulations - functional copula parameter



In each panel of this Figure, the thin line represents the true autoregressive function $\rho(Z_{i,t-1}, W_{2,i,t})$ vs the past rank $Z_{i,t-1}$, whereas the bold line stands for the average estimated autoregressive function. The shaded areas are the 95% pointwise confidence intervals, representing the variability of the Monte Carlo simulations. The three panels correspond to different values of the mobility score $W_{2,it}$, that are the quartiles of its empirical distribution.

From Figure 2, we find evidence that the true value of the mobility measure (computed with (2.22) and the functions g and ρ used in the Monte Carlo design), too, lies within the 95% point-wise confidence intervals that we constructed on the basis of the 1000 Monte Carlo simulations. We find that the bias is moderate in all the three cases considered (i.e. 25%, 50% and 75% quartiles of the mobility score and of the marginal distribution score). In particular, the bias is smaller in the part of the past rank distribution in which it is more likely that an individual with a certain value of both the mobility and the marginal distribution score is. For example, when considering an individual at the 75% quantile of the mobility score and at the 75% quantile of the marginal distribution score, the true mobility measure and the average estimated mobility measure obtained in the Monte Carlo simulation almost overlap at the top of the past rank distribution. Overall, the Monte Carlo results corroborate the validity of our methodology in a framework similar to the one of the empirical analysis presented in the next section.

Figure 2: Monte Carlo simulations - mobility measure



In each panel of this Figure, the thin line represents the true measure of mobility $m(Z_{i,t-1}; X_{i,t}, X_{i,t-1})$ vs the past rank $Z_{i,t-1}$, whereas the bold line stands for the average estimated measure of mobility. The shaded areas are the 95% pointwise confidence intervals, representing the variability of the Monte Carlo simulations. The three panels correspond to different values of the mobility score $W_{2,it}$ and of the marginal distribution score $W_{1,it}$, that are the quartiles of their respective empirical distributions.

5 Empirical analysis

5.1 The data

We apply the methods developed in Sections 2 and 3 to the Panel Study of Income Dynamics (PSID). This is the most commonly used dataset in the literature on individual earnings dynamics. The wide literature which made use of this dataset includes, among others, Lillard and Willis (1978), MaCurdy (1982), Abowd and Card (1989), Blundell, Pistaferri, Preston (2008), Guvenen (2009), Arellano, Blundell and

Bonhomme (2017). The PSID is one of the few panel surveys in the world which does not rotate its sample, meaning that individuals can virtually continue to get surveyed for time spans as long as 50 years or more. From the PSID we extract an unbalanced sample of individuals who were observed for at least 20 years during the period 1968-1997 [11]. Our dataset contains 37'439 individual-year observations. The average permanence of each individual in the dataset is about 25 years, hence the finite sample bias due to the estimated fixed effects in the preliminary wage regression is expected to be moderate. The Monte Carlo simulations presented in Section 4, which have been calibrated to mimic our dataset, confirm that the bias is indeed very small[12]. A key assumption in our analysis is that observations are missing-at-random. To support this hypothesis, it is worth underlining that Fitzgerald et al. (1998), using data from the PSID, find that attrition does not substantially affect the coefficients in a regression of log labor income for male heads on a set of observable variables (age, age squared and education). Moreover, while Fitzgerald et al. (1998) find that attrition is affected by the volatility of individual incomes, the reported effects are small, so that they conclude that in the PSID "attrition is mostly noise", and hence it is unlikely that it can affect studies using dynamic measures of earnings as outcomes. Lillard and Panis (1998), also using data from the PSID, found no evidence of selective attrition for white males based on their permanent unobserved components of income. Further, Meghir and Pistaferri (2004) did not find a significant effect of attrition on their estimates.

In the wake of Bonhomme and Robin (2009), we drop observations for students, retirees and self-employed workers and we only include observations relative to full-time employees (i.e. working more than 1200 hours a year) aged between 15 and 64, in order to limit the role of variations in the intensive margin of labor supply on wage dynamics (Bonhomme and Robin (2009), Hu, Moffitt and Sasaki (2019)). Moreover, as usual in the wage mobility literature (e.g. Buchinsky and Hunt (1999)) we exclude observations with wage equal to zero. We are aware that there may be differences in the functioning of the labor market in different states of the US. However, due to feasibility constraints and following the previous literature (Buchinsky and Hunt (1999), Auten et al. (2013), Kim (2013)) we will consider the US labor market as a whole. 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", which is recorded in the PSID. According to the US education system, the first dummy corresponds to 0-11 grades, the second dummy stands for high school or 12 grades and some nonacademic

[11] The choice of the time span analyzed here is due to the features of the PSID data. Indeed, the structure of the survey changed in 1997, becoming bi-annual instead of annual.

[12] An alternative that we do not explore in the present paper consists into performing a bias correction via jackknife, estimating our model on the full sample and then on one half of the same sample and combining the estimates, in the spirit of Dhaene and Jochmans (2015).

training, the third one represents college dropout, whereas the fourth one stands for college degree or advanced/professional degree. We argue that these dummy variables are exogenous, i.e. that they are not influenced by the individual position in the wage distribution. Indeed, we only consider education that takes place before labor market entry. We do not include among the explanatory variables any form of on-the-job training, due to its potential endogeneity. This implies that, in our sample, education is time-invariant.

Table 1: Descriptive statistics, PSID data, 1968-1997

Variable	Mean	Std. Dev.	Min.	Max.
Age	40.1511	10.1335	17	65
Elementary and middle school	0.2315	0.4218	0	1
High school	0.342	0.4744	0	1
Some college	0.1994	0.3996	0	1
College degree or higher education	0.2272	0.419	0	1
Log wage	9.8449	0.7684	5.0106	12.8992

This table reports some descriptive statistics for age, education and (log)wage, for our pooled data for the period 1968-1997. Log wage stands for the natural logarithm of annual wage (expressed in US dollars).

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 error. Thus, an interpretation of our estimated distribution of the residuals is that it represents a mixture of transitory shocks and measurement error. Table 1 above reports the descriptive statistics for our main variables of interests, i.e. wage, age and education dummies. As in Section 2, we use a linear panel regression to separate the deterministic wage component from the fluctuations around its trend. The model in equation (2.1) reads:

$$Wage_{i,t} = \alpha_0 + \alpha_1 Age_{it} + \alpha_2 Age_{it}^2 + \eta_i + \lambda_t + \varepsilon_{i,t} \tag{5.1}$$

where $Wage_{i,t}$ stands for log earnings, η_i represents the individual fixed effect, λ_t the time fixed effect, and $\varepsilon_{i,t}$ the residual wage component. The estimation is performed via the panel data fixed-effect technique, in order to adequately take into account the potential presence of unobserved heterogeneity across workers[13]. We include in the model a time fixed-effect in order to take into account all the macroeconomic shocks on wages, among them the impact of inflation on wages. As usual in the literature, we find

[13] An alternative would consist in applying a heterogeneous income profile (HIP) model, i.e. to allow wage to grow differently with age across individuals. A test of the HIP models vs the restricted income profile (RIP) ones, i.e. models in which wage increases with age uniformly across all individuals, as in equation (5.1), lies beyond the scope of the present paper. For a review and an empirical test of HIP models, we refer to Guvenen (2009). Also we could include interaction effects between age and education level.

evidence of a positive relationship between wage and age and a negative relationship between wage and age squared, i.e. log wage is concave in individual age. Indeed, the estimated coefficient α_1 is positive and statistically significant at the 99% confidence level, whereas the estimated coefficient α_2 is negative and statistically significant at the same confidence level. From the estimation of the linear panel model (5.1) we get estimated residuals $\hat{\varepsilon}_{it}$, i.e. the estimated residual wage component, which we use to build the empirical Gaussian ranks as in equation (3.1). Then, we use the empirical Gaussian ranks $\hat{Z}_{it}$ for the estimation of the semi-nonparametric dynamics copula specification defined in Section 2.

In our empirical application, we do not account for the possible endogeneity of the unemployment patterns. The estimation method could be extended to include a parametric specification for unemployment dynamics as in Bonhomme and Robin (2009). However, this parametric specification does not pair well with our semi-nonparametric approach and additionally it would require a single-step estimation approach, that is considerably more computationally demanding than the two-step procedure that we display in Section 5.2[14]. Moreover, Bonhomme and Robin (2009) analyzed the case of France, "where unemployment rates are chronically above or around 10%" in the sample they study (1990-2000), and hence ignoring transitions into and out of unemployment could cause a relevant selection bias (Bonhomme and Robin (2009), p.68). On the contrary, we use PSID data for estimation and, in the period considered (1968-1997), the unemployment rate for males in the US was on average below 7%; hence, we are confident that the selection bias due to unemployment is moderate and that the missing-at-random assumption is not unrealistic for our data. For these reasons, we abstain from an in-depth analysis of the possible endogeneity of unemployment patterns in this paper.

5.2 Estimation results

We apply the estimation procedure to the sample of US data described in Section 5.1. The bandwidths for the semi-parametric single-index estimator of Section 3.1 are computed as: $h_z = H_z = \sigma_z(NT)^{-1/6}$ and $h_w = H_w = \sigma_w(NT)^{-1/6}$, where σ_z and $\sigma_w = \sigma_w(\beta_1)$ are sample standard deviations of variables Z_{it} and $W_{1,it}(\beta_1)$, for any given β_1, following the Silverman rule for bivariate data (Silverman (1986)). Details of the numerical implementation of the estimators, in particular regarding the numerical one-dimensional integration, are reported in the Supplementary Material, Appendix D.3. Let us start with the estimates of the univariate distribution of the ranks conditional on covariates. The estimated marginal

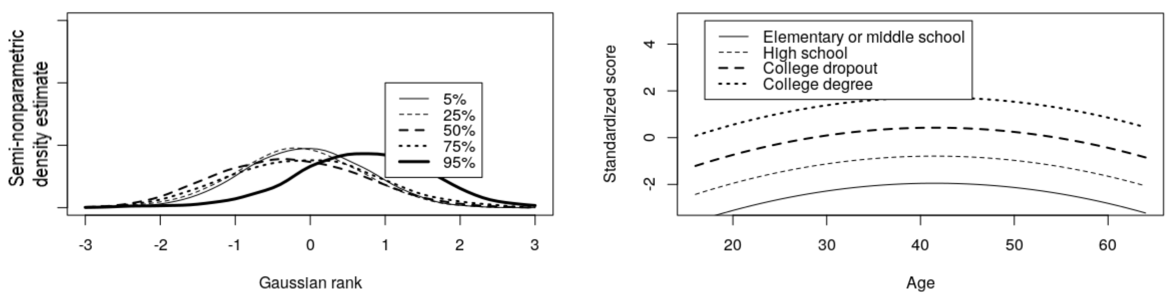
[14]Such a joint model of earnings, employment, job changes, wage rates and work hours has been recently proposed by Altonji et al. (2013). The authors estimate their full model via indirect inference. However, it is not straightforward to adapt the fully parametric joint model proposed by Altonji et al. (2013) to our semi-nonparametric framework. Further, the indirect inference method applied by the authors is already rather computationally intensive, since it is simulation-based. For these reasons, the extension of our model to take into account unemployment dynamics lies beyond the scope of the present paper.

distribution score governing the univariate rank distribution takes the following form:

$$\hat{W}_{1,it} = 0.1Age_{it} - 0.0012Age_{it}^2 + 0.5551Edu2_{it} + 1.1373Edu3_{it} + 1.7552Edu4_{it}.$$

Note that for identification purposes we fix the coefficient of age at the value 0.1 and we omit the dummy variable of the lowest education level, $Edu1_{it}$. As it can be deduced from the right panel of Figure 3, where we report the estimated marginal distribution score $\hat{W}_{1,it}$ as a function of age for different education levels, $\hat{W}_{1,it}$ is increasing in the education level and has an inverse-U relationship with age, reaching a maximum around the age of 40.

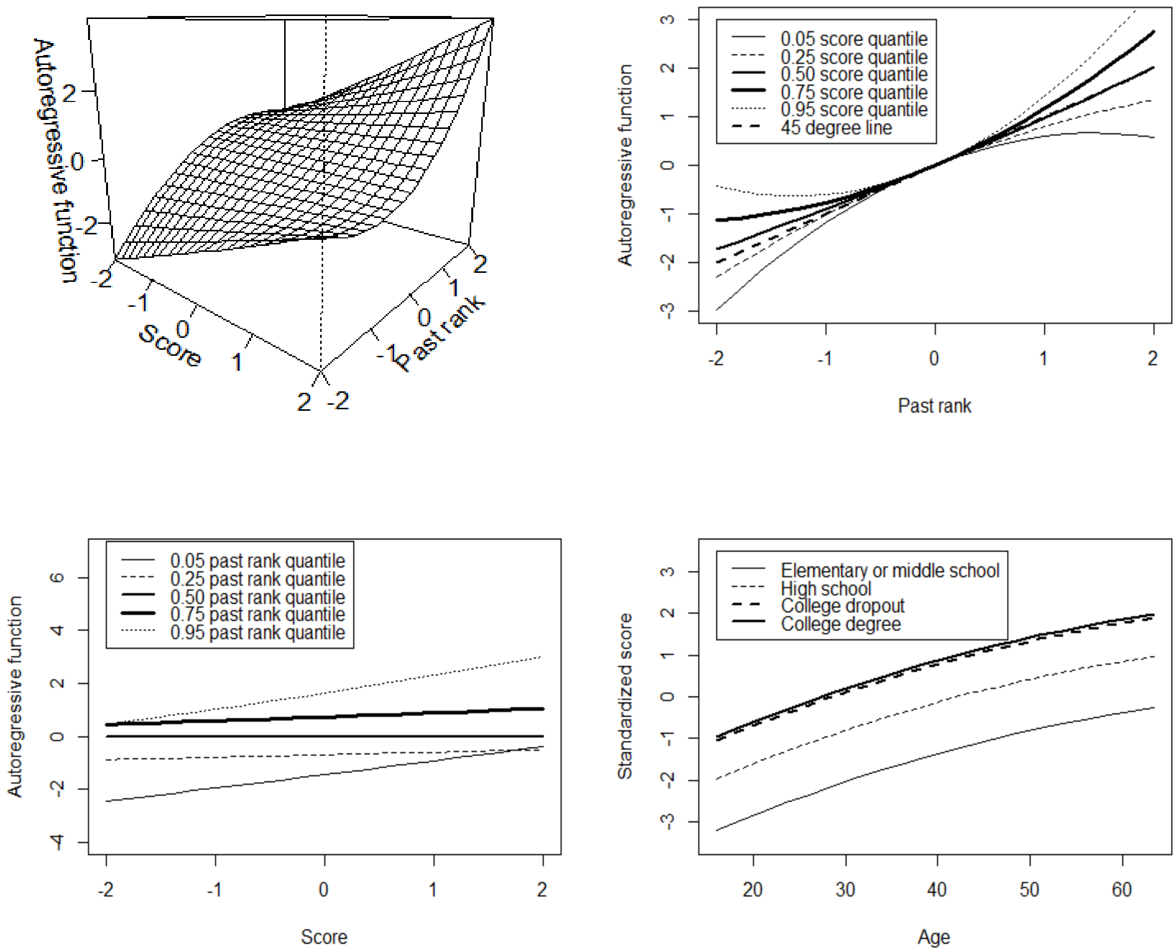
Figure 3: Estimation results of the marginal distributions of ranks



In the left panel of this figure we display the estimated pdf $\hat{g}(\cdot|X_{it}^a)$ of the marginal distribution of the rank for different values of the score $\hat{W}_{1,it} = X_{it}^{a'}\hat{\beta}_1$ corresponding to quantiles 5%, 25%, 50%, 75%, 95% of the score distribution. In the right panel, we report the value of the standardized marginal distribution score $\hat{W}_{1,it}$ as a function of age for different education levels.

Further, from the left panel of Figure 3, where we display the estimated pdf $\hat{g}(\cdot|X_{it}^a)$ for different values of $\hat{W}_{1,it}$, we notice that the explanatory variables have a remarkable influence on the marginal distribution of the ranks. More specifically, a higher value of the marginal distribution score is associated with a shift of the cross-sectional distribution of the residual wage rank towards larger values. For small (resp. large) values of the marginal distribution score, the conditional pdf $\hat{g}(\cdot|X_{it}^a)$ tends to be right (resp. left) skewed. These findings show the importance of including observable regressors in the specification of the marginal ranks distribution in addition to having them in the rank dynamics.

Figure 4: Estimation results for the copula distribution of ranks



The upper-left panel of this figure shows the three-dimensional representation of the estimated autoregressive function $\rho(Z_{i,t-1}, W_{2,it})$ in terms of past rank $Z_{i,t-1}$ and mobility score $W_{2,it}$. In the upper-right panel we display the value of the estimated function $\rho(Z_{i,t-1}, W_{2,it})$, as a function of past rank $Z_{i,t-1}$, for different values of the mobility score $W_{2,it}$, which correspond to quantiles 5%, 25%, 50%, 75%, and 95% of its sample distribution. In the lower-left panel we display the value of the estimated autoregressive function $\rho(Z_{i,t-1}, W_{2,it})$ as a function of the mobility score $W_{2,it}$, for different values of the past rank, which correspond to percentiles 5%, 25%, 50%, 75%, and 95% of the standard Gaussian distribution. In the lower-right panel, we display the value of the mobility score as a function of age for different education levels.

Let us now examine the estimation results of the semi-nonparametric copula model. The estimated mobility score reads[15]:

$$\hat{W}_{2,it} = 0.1Age_{it} - 0.0006Age_{it}^2 + 1.1007Edu2_{it} + 1.9183Edu3_{it} + 1.9972Edu4_{it}.$$

As before, for identification purposes in the mobility score we fix the value of the coefficient of age to 0.1 and we omit the dummy variable standing for the lowest education level. As shown in the bottom right panel of Figure 4, the mobility score is increasing for all the education levels, and is also monotonically increasing in age. There is a non-negligible difference in the values of the mobility scores between elementary school and high school and also between high school and college (uniformly across age).

The estimate of the copula autoregressive function using a tensor product Sieve of Hermite polynomials up to order 2 is the following[16]:

$$\begin{aligned}\hat{\rho}(Z_{i,t-1}, W_{2,it}) = \;& 0.0418 + 0.2116W_{2,it} + 0.9409Z_{i,t-1} + 0.0366W_{2,it}Z_{i,t-1} \\ & 0.0054(W_{2,it}^2 - 1) + 0.0418(Z_{i,t-1}^2 - 1) + 0.2116(Z_{i,t-1}^2 - 1)W_{2,it} \\ & 0.0047(W_{2,it}^2 - 1)Z_{i,t-1} + 0.0054(W_{2,it}^2 - 1)(Z_{i,t-1}^2 - 1).\end{aligned} \tag{5.2}$$

In Figure 4, we display the estimated autoregressive function (5.2). For all the possible values of the past rank, the autoregressive function is non-decreasing in the mobility score (see the lower left panel of Figure 4). Hence, for fixed $Z_{i,t-1}$ (and given value of the marginal distribution score), almost any conditional quantile of Z_{it} is non-decreasing in the mobility score.

This effect is more evident for values of the past rank at the boundaries of the support, whereas in the middle the autoregressive function appears almost flat, i.e. relatively less sensitive to the value of the mobility score.[17] We notice that there are relevant differences in the slope of the autoregressive function ρ w.r.t. the past rank, both for a fixed value of the mobility score in different parts of the rank distribution, and for different values of the mobility score (see the upper right panel of Figure 4). In particular, the autoregressive function exhibits substantial nonlinearities, with its slope and concavity/convexity changing across the past rank distribution and the different values of the score. This diversity in patterns

[15]We have checked that, in our sample, simple aggregate measures of positional mobility do not feature strong variation across time (see Appendix D.2 of the Supplementary Material). For this reason and the sake of simplicity we do not include macroeconomic variables in the mobility score $W_{2,it}$. However, it would be interesting to use our model to explore the link between relative earnings mobility and the business cycle more generally. Further, it would be surely interesting to investigate the impact of race on wage mobility. However, in order to be able to directly compare our results with those obtained by Bonhomme and Robin (2009), we include the same explanatory variables and restrain from an analysis of racial differences in wage dynamics.

[16]Note that the mobility score $W_{2,it}$ is centered and standardized. Further, in order to ensure identification constraints (3.13) of the autoregressive function, we impose the following constraints on the coefficients of the Hermite polynomials: $\lambda_{k,0} = \lambda_{k,2}$ for $k = 0, 1, 2$.

[17]Note that for $Z_{i,t-1} = 0$ the autoregressive function is constant w.r.t. the mobility score because of the normalization restriction $\rho(0, X_{it}) = 0$, for all X_{it}.

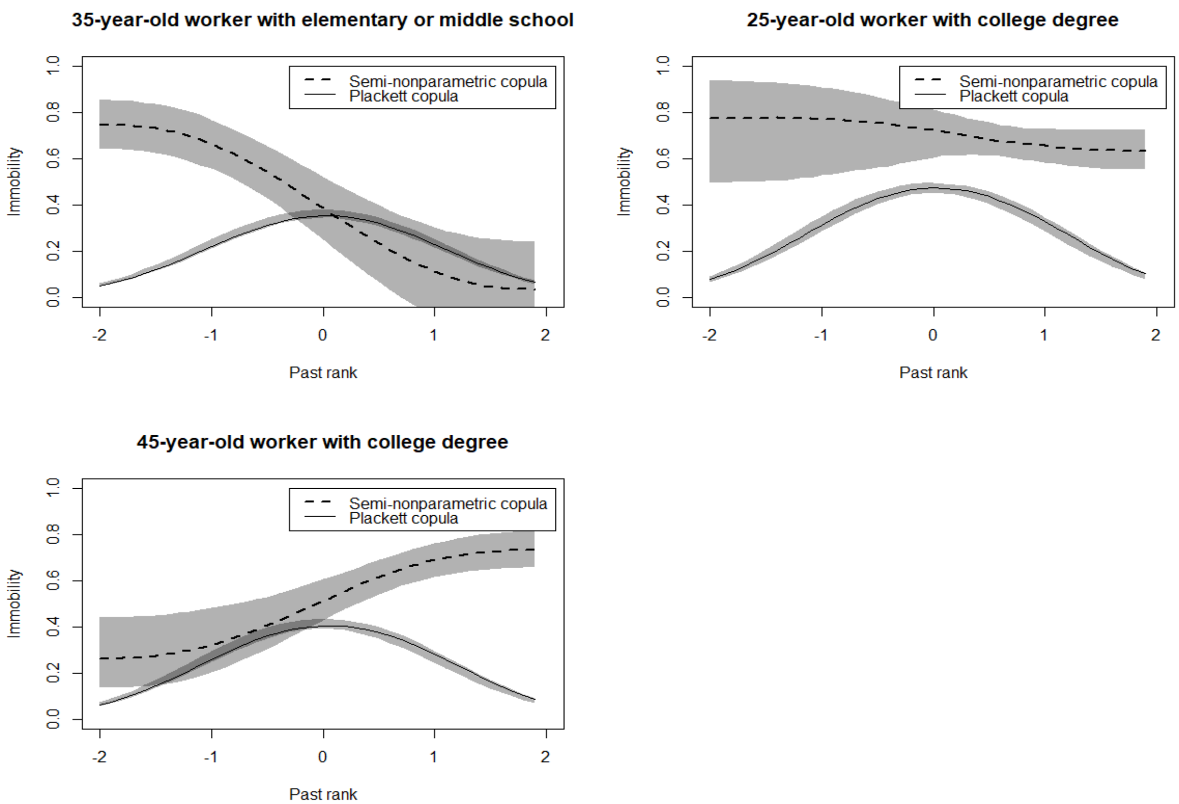
of the autoregressive function across the range of past rank and covariate values, speaks in favour of a flexible semi-nonparametric estimation of the rank dynamics as proposed in our paper.

Let us now discuss in more detail the patterns of the copula autoregressive function. Workers with low mobility scores, for example, are characterized by a relatively steeper autoregressive function than their colleagues with higher scores at the bottom part of the rank distribution, whereas the opposite is true at the top of the rank distribution. This finding suggests that, for workers with low mobility scores, the association between the present and the past rank of the transitory wage component is stronger if they are at the bottom of the distribution (low-wage trap). The contrary holds for individuals with a high mobility scores, i.e. they present a higher degree of wage mobility in the bottom part of the distribution than at the top of it. In general, the slope of the autoregressive function is rather high (i.e. higher than 1) for all the score values, for a past rank close to zero. This means that there is a rather high degree of persistence in the rank dynamics, at least for the central values of the past rank, and then even more for some mobility score values. However, due to the presence of the nonlinear transformation Λ in the ranks dynamics, the slope of the autoregressive function cannot be directly interpreted as a measure of mobility. Therefore, in Figure 5 we represent mobility, as measured by the empirical counterpart of the r.h.s. of equation (2.22), for different age and education groups. Note that, in Figure 5, the y-axis records the value of our mobility measure, as defined by equation (2.22). The higher this quantity is, the stronger is the association between the present and the past rank and hence the lower the degree of positional mobility is. This is the reason why the y-axis is labelled as "immobility".

In Figure 5, in the upper left panel, we consider the case of a worker with low values of both the mobility and marginal distribution scores (i.e. a 35-year old individual with elementary or middle school), in the upper right panel we have a case in which both scores take an intermediate value (i.e. a 25-year-old worker who completed college), and in the bottom panel we have the case of an individual with relatively high scores (45-year-old worker with college degree).

In Figure 5 we find confirmation of our previous insights about wage mobility patterns for individuals with different mobility scores. Indeed, we find that individuals with low values of the scores (35-year-old worker with elementary or middle school) exhibit lower mobility than their colleagues with higher scores, in particular at the bottom of the rank distribution, i.e. they are subject to the low-pay trap when they are at the bottom of the distribution. On the other hand, the degree of rank mobility at the top of the rank distribution is higher for less-educated individuals than for their higher-educated colleagues. This suggests that the former ones have a higher risk of falling downwards in the rank distribution, even if their current position is high. In the case of intermediate values of the scores (i.e. the upper right panel of Figure 5), we find that rank persistence is rather high (i.e. between 0.7 and 0.8) across the whole distribution of the past rank, and the degree of persistence recorded by those workers at the bottom part of the rank distribution is clearly higher than that of individuals with higher values of the scores. Finally, in the case of a worker with high values of the scores (i.e. a 45-year-old worker with college degree), the mobility pattern is relevantly different from the first two cases described above. Indeed, in this third case, mobility is highest at the bottom end of the distribution and small elsewhere. The association between the present and the past rank at the bottom of the rank distribution is close to zero (i.e. 0.2) for this type of worker, so there is no low-pay trap in this case.

Figure 5: Mobility patterns for different age and education levels



We display the immobility measure as a function of the past rank for three combinations of age and education levels. The immobility measure is computed according to formula (2.22) using estimates from the semi-nonparametric copula model (dashed line). The chosen sets of individual characteristics correspond to different values of the marginal distribution and mobility scores, which are $W_{1,it} = -1.15$ and $W_{2,it} = -1.69$ in the upper left panel, $W_{1,it} = 1.08$ and $W_{2,it} = -0.19$ in the upper right panel, and $W_{1,it} = 1.57$ and $W_{2,it} = 1.16$ in the bottom panel. Both index values have been standardized. The solid line in each panel represents the mobility function estimated using a Plackett copula model for the ranks dynamics. The shaded areas correspond to the 95% pointwise confidence intervals and have been obtained by nonparametric bootstrap, with number of replications $B = 500$.

This means that, if individuals characterized by high values of the scores are in a low position of the wage distribution, then they are highly likely to improve their position in the following period. These differences in the mobility patterns for individuals with different characteristics are statistically significant. The 95% pointwise confidence intervals for different individuals (computed by nonparametric bootstrap) only rarely overlap[18]. Further, these confidence bands have different widths across the rank distribution for a worker with certain characteristics, in particular they are wider in the part of the past rank distribution in which it is more unlikely that the given individual is.

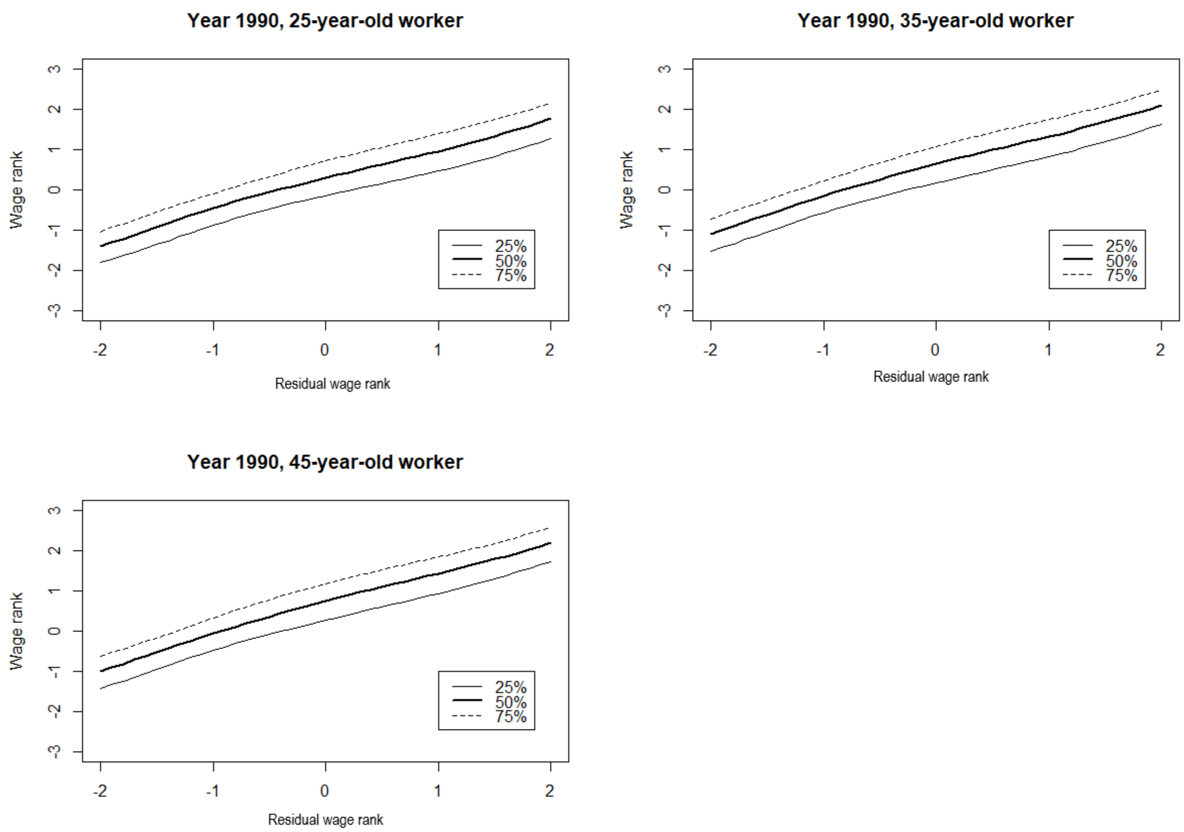
The quantity reported in Figure 5 is the slope of the conditional rank expectation function represented in the left panels of Figure 8 in the Supplementary Material. The right panels of Figure 8 report the conditional quantiles of the present rank as function of the past rank, computed according to the empirical analogue of equation (2.24). We notice that the conditional median is close to the conditional expectation represented in the left panels of the same Figure 8, for all the three cases considered.

Our result that relative mobility is increasing in education is somehow consistent with the estimation results obtained by Bonhomme and Robin (2009). However, we find that this is only true at the bottom of the past rank distribution. In Figure 5 we superimpose our mobility results to those obtained by estimating a fully parametric Plackett copula model on our data, in the spirit of Bonhomme and Robin (2009). As in their paper, the parameter of the Plackett copula is function of the individual characteristics. From the comparison of the estimated mobility patterns obtained with the parametric, respectively with our semi-nonparametric model, we deduce that, by adopting a fully parametric model, the pattern of rank mobility is a priori determined by the copula family. On the contrary, adopting a more flexible specification as in our case, it is possible to obtain different mobility patterns, according to the different individual characteristics. The Plackett copula is not able to capture well the mobility patterns emerging from the panels of Figure 5, neither as a function of the past rank, nor as a function of the individual characteristics. Indeed, the Plackett density function does not change its functional shape depending on its parameter, and this explains the large differences between the fully parametric model and our semi-nonparametric specification in describing mobility patterns, as it is apparent in Figure 5. In particular, from the two upper panels of Figure 5, we deduce that estimating rank mobility patterns via a fully parametric copula in our framework leads to a severe overestimation of mobility at the bottom of the rank distribution for workers with low and intermediate values of the mobility score. Indeed, by estimating the Plackett copula on our data we find that the degree of rank persistence at the bottom of the distribution is close to zero for any possible combination of age and educational level. On the contrary, by using the semi-nonparametric copula model presented in Section 2 and 3, we find evidence of a substantial amount

[18] A more formal test should be based on the difference between the estimates.

of rank persistence at the bottom of the past rank distribution for individuals with either low experience[19] or a low educational level.

Figure 6: Wage rank as a function of the residual wage rank



We display the relationship between the wage rank and the rank of the residual wage component, defined by the function $b_t(Z_{it}, \eta_i, X_{it})$ in equation (2.25) for individuals of different ages. In each panel in this Figure, the dashed line stands for a value of the individual effect η_i equal to the 75% percentile of its distribution, the bold line corresponds to the median value of η_i, and the continuous line stands for η_i equal to the 25% percentile of its distribution.

To conclude the empirical analysis, let us now consider wage mobility including both the deterministic component and the fixed effects. In Figure 6, we plot function b_t defined in (2.26) linking wage rank and residual wage rank for year 1990, for different values of age and of the fixed effect η. The function b_t is not very sensitive to the observable characteristics and does not vary substantially over time. It is close to a straight line and its level is mostly driven by the value of η. The observed linearity of function b_t is due to the fact that distributions $F_{\varepsilon,t}(\cdot)$ and $F_{y,t}(\cdot)$ are close the Gaussian (see Figure 7 in the Supplementary

[19]We consider age as a proxy for experience here.

Material). Of course, it is possible to draw function $b_t(\cdot)$ for any combination of age and individual effect. In particular, it is worth underlying that, in Figure 6, wage ranks are more stable than transitory ranks, in the sense that the former only ranges between approximately -1.5 and 1.5, when we move the latter from -2 to -2. This is due to the fact that we have drawn function $b_t(\cdot)$ for fixed values of the individual effect η_i and individual characteristics X_{it}, hence, all the remaining variability only depends on "luck". We further notice that the effect of η_i on the function $b_t(\cdot)$ is essentially linear.

In Figure 9 in the Supplementary Material, we display the conditional median and the lower and upper quartiles for year 1990, for different combinations of individual characteristics reflecting those chosen in Figure 5, and different levels for the individual effect (median and upper quartile). We use the empirical analog of (2.26) with $u = 0.5$, and 0.25, 0.75, respectively. The patterns in Figure 9 are similar to those in the right panels in Figure 8. This is due to the fact that the function $b_t(\cdot)$, represented in Figure 6, is essentially linear in the different cases. The slope of the conditional median represented in Figure 9 stands for immobility, defined as persistence of wage rank with respect to a shock on the past residual wage rank. This means that, for example, an individual with characteristics as represented in the lower panels of Figure 9 is relatively insensitive to transitory shocks when she is in the bottom part of the residual rank distribution. On the contrary, there is a positive association between a shock in the past residual wage component and the present wage rank in the upper part of the transitory rank distribution[20].

6 Concluding remarks

As explained by Schumpeter (1955), the wage scale can be compared to a hotel, in which there are both luxury rooms and cheaper rooms. At any point in time, all the rooms are occupied by some guest; however, they do not always occupy the same room. The study of wage rank mobility aims at understanding how frequently do guests "switch rooms" and which are the factors underlying this process. In this paper we presented a flexible model for the wage rank dynamics. We developed a new family of semi-nonparametric copulas which are well-suited to describe the dynamics of earning ranks. We proposed consistent estimators for both the marginal rank distribution, and the copula distribution of present and past ranks, conditional on covariates, and we provide Monte Carlo evidence that the finite sample bias due to the incidental parameter problem is moderate in our framework. This novel semi-nonparametric copula model allows for greater flexibility than a fully parametric one, and in particular allows to estimate the nonlinear autoregressive function which links the present and the past Gaussian ranks, conditional on some individual characteristics.

[20]We can easily recover a figure in which the past wage rank $r_{i,t-1}$ is on the x-axis, and the conditional quantiles of the present wage rank $r_{i,t}$ are on the y-axis, by means of representation $Z_{i,t-1} = b_{t-1}^{-1}(r_{i,t-1}, \eta_i, X_{i,t-1})$.

From the empirical application we get evidence that, in the US labor market, there is a rather high degree of mobility at the bottom of the distribution for workers with a high educational level and some experience (i.e. with high scores). On the contrary, we find that workers who are either at the beginning of their career or who have a low educational level (i.e. with low scores) are subject to the risk of being stuck in the so called low-wage trap.

Our semi-nonparametric model can be easily used to simulate wage trajectories. By simulating wage trajectories we would be able to compute the present values of individual earnings in the medium and in the long term, and to compute from these values the evolution of some summary inequality indices over time. This constitutes scope for future research. A limitation of the present work is that we rely on the missing-at-random assumption for our unbalanced panel data. However, according to Fitzgerald et al. (1998), Lillard and Panis (1998) and Meghir and Pistaferri (2004), this assumption is quite realistic for PSID data in the period considered. Moreover, in the present paper we only considered age and the education level achieved by the individual as explanatory variables. As a future outlook, our semi-nonparametric model may be extended in the direction of including other variables such as gender, or the presence of a migration background of the worker. Further, in a companion paper (Naguib and Gagliardini (2019)) we investigate the efficacy of an analytical bias correction for a two-step semi-nonparametric estimator like the one developed in the present work. This aspect, too, represents an avenue for future research.

Appendix A: Assumptions

In this Appendix we list the regularity conditions used to establish the large sample properties of our estimators.

Assumption A.1. *(i) Function $K(\cdot)$ is a compactly supported kernel in $\mathbb{R}$ and it is three times differentiable with bounded derivatives.*
(ii) Further, kernel function $K(\cdot)$ has order p, i.e. $\int u^j K(u)du = 0 \ \ for j = 1, ..., p-1$, and $\int u^p K(u)du \neq 0$, for an integer $p \geq 2$.

Assumption A.2. *(i) The bandwidths $h_z, h_w > 0$ satisfy $h_z, h_w = o(1)$ and $(NT)^{2-\delta} h_z h_w^5 \to \infty$ for an arbitrarily small constant $\delta > 0$, as $N, T \to \infty$.*
(ii) For the trimming factor $k_{i,t}$ in (3.2) we require that a_0, $c > 0$ and $(NT)^c (h_z^2 + h_w^2) = o(1)$ and $(NT)^{1-2c} h_z h_w \to \infty$.
(iii) The bandwidths $H_z, H_w > 0$ satisfy $H_z, H_w = O((NT)^{-1/6})$, $\frac{\log(NT)}{NTH_zH_w} = o(1)$, $NTH_zH_w(h_z^{2p} + h_w^{2p}) = o(1)$ and $\frac{H_zH_w}{h_z h_w^3} = o((NT)^{1-\delta})$, for $\delta > 0$.

Assumption A.3. *Let $\mathbb{S}$ be a compact subset of $\mathbb{R}^{p_1+1}$ and, for $\delta > 0$ small, let $\mathbb{S}_\delta \subset \mathbb{R}^2$ be the subset of points with a distance from set $\{(z, \beta_1' x^a) : \ (z, x^a) \in \mathbb{S}, \ \beta_1 \in B_1\}$ smaller than δ.*
(i) For all $\beta_1 \in B_1$, the bivariate vector $(Z_{it}, \beta_1' X_{it}^a)$ has probability density $g_{Z,\beta_1' X^a}(z, w)$ with respect to Lebesgue measure on set $\mathbb{S}_\delta$, such that $\inf_{(z,w)\in\mathbb{S}_\delta} g_{Z,\beta_1' X^a}(z, w) > 0$. Moreover, $g_{Z,\beta_1' X^a}(z, w)$ and $E(X_{it}^a | Z_{it} = z, \beta_1' X_{it}^a = w)$ and $E(X_{it}^a X_{it}^{a'} | Z_{it} = z, \beta_1' X_{it}^a = w)$ are $(2+p)$-times continuously differentiable with respect to $(z, w) \in \mathbb{S}_\delta$. Further, the joint density of $(Z_{i,1}, \beta_1' X_{i,1}^a)$ and $(Z_{i,t}, \beta_1' X_{i,t}^a)$ is bounded, for t large enough.
(ii) For all $\beta_1 \in B_1$, let $E_{\mathbb{S}}[\cdot]$ be the conditional expectation given $(Z_{it}, X_{it}^a) \in \mathbb{S}$. Then, function $E_{\mathbb{S}}[\log g(Z_{it} | \beta_1' X_{it}^a)]$ is finite and has a unique global maximum β_1^0 that lies in the interior of B_1.
(iii) $q' \Omega(\beta_1^0) q > 0$ for any vector $q \in \mathbb{R}^{p_1}$ s.t. $q \perp \beta_1^0$, where $\Omega(\beta_1^0) = E_{\mathbb{S}}[-\nabla^2_{\beta_1\beta_1} \log g(Z_{it} | \beta_1^{0'} X_{it}^a)]$ and $\nabla^2_{\beta_1\beta_1}$ denotes the matrix of second-order partial derivatives with respect to β_1.

Assumption A.4. *The mixing coefficient*

$$\alpha_h^i = \sup \{ |P(A \cap B) - P(A)P(B)| : A \in \sigma\{Y_{i,s}, \quad s < t\}, B \in \sigma\{Y_{i,s}, \quad s \geq t + h\} \}$$

where $\sigma\{Y_{i,s}, \quad s < t\}$ is the sigma field generated by variables $Y_{i,s} = (Z_{i,s}, X_{i,s}^{a\prime})', s < t$, satisfies: $\alpha_h^i \leq C \bar{\rho}^h$ for constants $C > 0$ and $\bar{\rho} \in (0, 1)$, and all i and h.

Assumption A.5. *(i) Let* $\mathcal{L}_{NT}(\beta_1) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} k_{i,t}\log\tilde{g}_{-(i,t)}(Z_{i,t}|\beta_1'X_{i,t}^a;\beta_1)$ *where* $\tilde{g}_{-(i,t)}(z|w;\beta_1) = \frac{\frac{1}{h_z}\sum_{s=1,s\neq t}^{T}\sum_{j=1,j\neq i}^{N}K(\frac{z-Z_{j,s}}{h_z})K(\frac{w-W_{1,js}(\beta_1)}{h_w})}{\sum_{s=1,s\neq t}^{T}\sum_{j=1,j\neq i}^{N}K(\frac{w-W_{1,js}(\beta_1)}{h_w})}$*, and let* $\hat{\mathcal{L}}_{NT}(\beta_1)$ *be defined analogously replacing* $Z_{i,t}$ *with* $\hat{Z}_{i,t}$*. Then we have* $\sup_{\beta_1\in B_1}|\hat{\mathcal{L}}_{N,T}(\beta_1)-\mathcal{L}_{N,T}(\beta_1)| = o_p(1)$*. (ii) We have* $\sup_{\beta_1\in B_1}|\mathcal{L}_{N,T}(\beta_1)-\mathcal{L}(\beta_1)| = o_p(1)$*, where* $\mathcal{L}(\beta_1) = E_{\mathbb{S}}[\log g(Z_{it}|\beta_1'X_{it}^a)]$*.*
(iii) Let $\tilde{g}(z|\beta_1'x^a) = \frac{\frac{1}{H_z}\sum_{i=1}^{N}\sum_{t=1}^{T}K(\frac{z-Z_{i,t}}{H_z})K(\frac{\beta_1'(x^a-X_{i,t}^a)}{H_w})}{\sum_{i=1}^{N}\sum_{t=1}^{T}K(\frac{\beta_1'(x^a-X_{i,t}^a)}{H_w})}$ *and let* $\breve{g}(z|\beta_1'x^a)$ *be defined analogously replacing* $Z_{i,t}$ *with* $\hat{Z}_{i,t}$*. Then* $\sup_{(z,x^a)\in\mathbb{S},\beta_1\in B_1}|\breve{g}(z|\beta_1'x^a)-\tilde{g}(z|\beta_1'x^a)| = o_p(1)$ *and* $\sup_{(z,x^a)\in\mathbb{S}}|\tilde{g}(z|\hat{\beta}_1'x^a) - \tilde{g}(z|\tilde{\beta}_1'x^a)| = o_p(1)$ *if* $\hat{\beta}_1-\tilde{\beta}_1 = o_p(1)$*.*

Assumption A.6. *(i) The expectation* $E[l(\tilde{Y}_{it},\theta_0,f_0)]$ *exists in* $\mathbb{R}$*, where*

$$l(\tilde{Y}_{it},\theta_0,f_0) = \log c[\tilde{G}(Y_{it},f_0),\tilde{G}(Y_{i,t-1},f_0);\rho_0(\cdot,X_{it}'\beta_2^0)]$$

and $\tilde{Y_{it}} = (Y_{it},Y_{i,t-1},X_{i,t}')'$ *and* $\tilde{G}(Y_{i,t},f) = \int_{-\infty}^{Z_{i,t}}f(z|X_{i,t}^a)dz$*.*
(ii) For all $\epsilon>0$*, there exists some non-increasing positive sequence* $c_{N,T}(\epsilon)$ *such that, for all* $N,T\geq 1$*,*

$$E[l(\tilde{Y}_{it},\theta_0,f_0)] - \sup_{\{\theta\in\Theta_{N,T}:||\theta-\theta_0||_\Theta\geq\epsilon\}}E[l(\tilde{Y}_{it},\theta,f_0)] \geq c_{N,T}(\epsilon) \quad \text{(a.1)}$$

where $\lim_{N,T\to\infty}c_{N,T}(\epsilon) > 0$*,* $\forall\epsilon>0$*, and* $||\cdot||_\Theta$ *is a metric defined on* Θ*, or some metric space containing* Θ*.*

Assumption A.7. $\theta_0\in\Theta$*, and let* $\Theta_{N,T}\equiv\Theta_m$ *with* $m = m(N,T)\in\mathbb{N}$*, such that* $\Theta_m\subset\Theta_{m+1}\subset\Theta$ *for all* $m\geq 1$ *and there exists some* $\theta_m\in\Theta_m$ *such that:*

$$|E[l(\tilde{Y}_{i,t},\theta_m,f_0) - l(\tilde{Y}_{i,t},\theta_0,f_0)]| = O(\eta_{2,N,T})$$

where $\eta_{2,N,T}$ *is some positive non-increasing sequence.*

Assumption A.8. *(i)* $\sup_{\theta\in\Theta_m,f\in\mathcal{N}_{f,N,T}}|\mu_{N,T}l(Y,\theta,f)]| = O_p(\eta_{0,N,T})$ *where* $\eta_{0,N,T}$ *is some finite positive non-increasing sequence converging to zero,* $\mu_{N,T}[l(Y,\theta,f)] = \frac{1}{N(T-1)}\sum_{i=1}^{N}\sum_{t=2}^{T}\{l(\tilde{Y}_{i,t},\theta,f) - E[l(\tilde{Y}_{i,t},\theta,f)]\}$ *and* $\mathcal{N}_{f,N,T} = \{f\ :||f-f_0||_{\mathbb{F}}\leq\delta_{f,N,T}\}$ *with* $\delta_{f,N,T} = o(1)$*, where* $||f-f_0||_{\mathbb{F}} = \sup_{(z,x)\in\mathbb{S}}|f(z|x)-f_0(z|x)|$*.*
(ii) There is a finite positive non-increasing sequence $\eta_{1,N,T}$ *going to zero such that*

$$\sup_{\theta\in\Theta_m,f\in\mathcal{N}_{f,N,T}}|E[l(\tilde{Y}_{i,t},\theta,f) - l(\tilde{Y}_{i,t},\theta,f_0)]| = O(\eta_{1,N,T}). \quad \text{(a.2)}$$

Assumption A.9. *(i) It holds* $\sup_{\theta \in \Theta_m} |\log c(u_1, v_1; \rho(\cdot, X'\beta_2)) - \log c(u_2, v_2; \rho(\cdot, X'\beta_2))| \leq C_{NT}\left(|u_1 - u_2| + |v_1 - v_2|\right)$ *for a constant* C_{NT} *and any* $u_1, u_2, v_1, v_2 \in [0,1]$ *and* X.
(ii) $\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}|\hat{Z}_{i,t} - Z_{i,t}| = O_p(\tilde{\eta}_{NT})$*, for a positive non-increasing sequence* $\tilde{\eta}_{NT}$ *tending to zero as* $N, T \to \infty$*, and* $C_{NT}\tilde{\eta}_{NT} = o(1)$.

Delecroix et al. (2003) present the regularity conditions under which a semi-parametric single-index model yields a consistent and asymptotically normal estimator in a cross-sectional framework. Their work has been extended by Rosemarin (2012). Assumptions A.1-A.4 build on the conditions used in Rosemarin (2012) and imply the consistency of our semi-parametric single-index estimators $\hat{\beta}_1$ and $\hat{g}(\cdot)$ in a panel data setting, if the true ranks $Z_{i,t}$ were observed. Specifically, Assumption A.1 imposes standard conditions on the kernel function. Assumption A.2 concerns the conditions on the bandwidths h_z and h_w used in the criterion to estimate the index vector $\hat{\beta}_1$, and the bandwidths H_z and H_w used for the kernel density estimator yielding $\hat{g}$. Different bandwidths may be required since the technical conditions on h_z and h_w in Assumption A.2 (i) and (ii) may imply undersmoothing for density estimation (see Rosemarin (2012)). Assumption A.3 (i) includes some requirements for the continuity and differentiability of expectations and density functions. Assumption A.3 (ii) is a global identifiability condition for β_1^0 and $E_{\mathbb{S}}[\log g(Z_{it}|\beta_1' X_{it}^a)]$ is the limit value of the criterion which is maximized in the sample to estimate $\hat{\beta}_1$. Matrix $\Omega(\beta_1^0)$ in Assumption A.3 (iii) is a sort of Hessian matrix and Assumption A.3 (iii) itself is a common requirement in the dimension reduction literature (see Hall and Yao (2005)). Assumption A.4 imposes conditions on the strong mixing coefficient uniformly across individuals. Assumption A.5 is a high-level condition to ensure that replacing the true ranks with the empirical ones does not have an effect asymptotically. It requires in particular that the difference between the criterion function of the semi-parametric single-index estimator computed with either the empirical, or the true ranks, is uniformly negligible.

Assumptions A.6-A.8 are adapted to our panel framework from Hahn, Liao and Ridder (2018) and are completed with Assumption A.9 to cope with estimation error on empirical ranks. They are needed to ensure consistency of the second-step Sieve Maximum Likelihood estimator $\hat{\theta}$. As underlined by Hahn, Liao and Ridder (2018), Assumption A.6 essentially requires that the second-step Sieve M estimation problem is well-posed. Intuitively, our second-step estimation is a well-posed problem, since, as it appears from equations (2.16) and (3.10)-(3.11), the estimation criterion does not only depend on the approximating function $\rho(\cdot)$ via an integral of $\rho(\cdot)$, but also directly, i.e. via the values of ρ at sample points. Verification of Assumption A.6 from more primitive conditions is not simple since the estimation problem is nonlinear. Assumption A.7 is a condition on the accuracy of the approximation of Θ by the sequence of Sieve spaces, it imposes a bound on the approximation error and requires that the Sieve space is expanding. According to Hahn, Liao and Ridder (2018), it is essentially condition b of Lemma A.2 in Chen and Pouzo (2012). Hahn, Liao and Ridder (2018) also point out that Assumption A.7 is implied by conditions 3.2 and 3.3 in Chen (2007). Assumption A.8 is similar to condition 3.5 of Theorem 3.1 in Chen (2007) and the first part of Condition d of Lemma A.2 in Chen and Pouzo (2012); it requires uniform convergence of the empirical process $\mu_{N,T}[l(Y, \theta, f)]$. Assumption A.9 (i) requires Lipschitz continuity of the log copula density function, uniformly w.r.t. the functional parameter. Assumption A.9 (ii) yields an average error rate for estimating the empirical ranks. These conditions are used to show that replacing true ranks with empirical ranks has a vanishing effect asymptotically for estimator $\hat{\theta}$.

Appendix B: Proofs of Propositions

B.1 Proof of Proposition 1

The proof consists in two steps. In Step 1, we show the following implication. For any $t \geq 1$, if:

$$l(Z_{t-1}|X_{t-1}) = g(Z_{t-1}|X_{t-1}^a) \tag{b.1}$$

then:

$$l(Z_t|X_t) = g(Z_t|X_t^a), \tag{b.2}$$

where $g(Z|X^a)$ is the conditional pdf such that (2.10) holds. To show this, suppose that equation (b.1) holds. Then we have:

$$\begin{aligned} l(Z_t|X_t) &= \int\int l(Z_t|Z_{t-1}, X_{t-1}, X_t) l(Z_{t-1}, X_{t-1}|X_t) dZ_{t-1} dX_{t-1} \\ &= \int\int l(Z_t|Z_{t-1}, X_{t-1}, X_t) l(Z_{t-1}|X_{t-1}, X_t) l(X_{t-1}|X_t) dZ_{t-1} dX_{t-1}. \end{aligned}$$

Then, by Lemma 6 in Appendix C of the Supplementary material, the absence of Granger causality in Assumption 2 implies that $l(Z_{t-1}|X_t, X_{t-1}) = l(Z_{t-1}|X_{t-1})$. Then:

$$l(Z_t|X_t) = \int\int l(Z_t|Z_{t-1}, X_{t-1}, X_t) l(Z_{t-1}|X_{t-1}) l(X_{t-1}|X_t) dZ_{t-1} dX_{t-1}.$$

Now, by replacing the definition of the conditional density in (2.11), and using (b.1), we get:

$$\begin{aligned} l(Z_t|X_t) = \int\int g(Z_t|X_t^a) c[G(Z_t|X_t^a), G(Z_{t-1}|X_{t-1}^a); \rho(\cdot, X_t)] \cdot \\ g(Z_{t-1}|X_{t-1}^a) \cdot l(X_{t-1}|X_t) dZ_{t-1} dX_{t-1}. \end{aligned} \tag{b.3}$$

Hence, after a change of variable from Z_{t-1} to $v = G(Z_{t-1}|X_{t-1}^a)$, and using $\int c(u,v)dv = 1, \forall u$, we obtain:

$$\begin{aligned} l(Z_t|X_t) &= \int\int g(Z_t|X_t^a)c[G(Z_t|X_t^a), v; \rho(\cdot, X_t)]l(X_{t-1}|X_t)dvdX_{t-1} \\ &= \int g(Z_t|X_t^a)l(X_{t-1}|X_t)dX_{t-1} = g(Z_t|X_t^a) \end{aligned}$$

which yields (b.2).

In Step 2 of the proof, we use the initial condition in (2.12) and use repeatedly the implication derived in Step 1, to get $l(Z_t|X_t) = g(Z_t|X_t^a)$, at any $t \geq 1$. Then, by integrating out the explanatory variables vector X_t and using the property in (2.10) we get:

$$l(Z_t) = \int l(Z_t|X_t)l(X_t)dX_t = \int g(Z_t|X_t^a)l(X_t)dX_t = \int g(Z_t|X_t^a)l(X_t^a)dX_t^a = \phi(Z_t),$$

which yields (2.13).

B.2 Proof of Proposition 2

By the law of iterated expectations we have:

$$\begin{aligned} P(Z_t \leq z) &= E[P(Z_t \leq z|Z_{t-1})] = E\left[P\{\Lambda(\rho(Z_{t-1}) + \omega_t) \leq z|Z_{t-1}\}\right] \\ &= E\left[P\{\omega_t \leq \Lambda^{-1}(z) - \rho(Z_{t-1})|Z_{t-1}\}\right] = \int_{-\infty}^{\infty} \Phi[\Lambda^{-1}(z) - \rho(Z_{t-1})]\phi(Z_{t-1})dZ_{t-1}, \end{aligned}$$

where Φ is the cdf of the standard normal distribution of ω_t, and ϕ is the pdf of the standard normal distribution of Z_{t-1}. Therefore, the following condition must hold: $\int_{-\infty}^{\infty} \Phi[\Lambda^{-1}(z) - \rho(Z_{t-1})]\phi(Z_{t-1})dZ_{t-1} = \Phi(z)$. By applying a change of variable: $\Lambda^{-1}(z) = y \Leftrightarrow z = \Lambda(y)$, we get the restriction for $\Lambda(y)$: $\Phi(\Lambda(y)) = \int_{-\infty}^{\infty} \Phi(y - \rho(Z_{t-1}))\phi(Z_{t-1})dZ_{t-1}$, i.e. equation (2.15).

B.3 Proof of Proposition 3

Let us first write the conditional expectation of the rank of the residual component:

$$\begin{aligned} E(Z_{it}|Z_{i,t-1}, X_{it}, X_{i,t-1}) &= E[G^{-1}(U_{it}|X_{it}^a)|Z_{i,t-1}, X_{it}, X_{i,t-1}] \\ &= E[G^{-1}(\tilde{\Lambda}[\tilde{\rho}(U_{i,t-1}; X_{it}) + \omega_{it}; X_{it}]|X_{it}^a)|Z_{i,t-1}, X_{it}, X_{i,t-1}] \\ &= \int_{-\infty}^{\infty} G^{-1}\left(\tilde{\Lambda}[\tilde{\rho}(G(Z_{i,t-1}|X_{i,t-1}^a); X_{it}) + \omega; X_{it}]|X_{it}^a\right)\phi(\omega)d\omega, \end{aligned} \quad \text{(b.4)}$$

where $U_{it} \equiv G(Z_{i,t}|X_{i,t}^a) = \Phi(\xi_{i,t})$ follows the stochastic representation $U_{it} = \tilde{\Lambda}[\tilde{\rho}(U_{i,t-1}; X_{it}) + \omega_{it}; X_{it}]$ from (2.20), with $\tilde{\rho}(u; X_{it}) = \rho(\Phi^{-1}(u), X_{it})$ and $\tilde{\Lambda}(k; X_{it}) = \int_0^1 \Phi(k - \tilde{\rho}(v; X_{it}))dv$. Then, by computing the derivative w.r.t. $Z_{i,t-1}$ we get equation (2.22).

B.4 Proof of equation (3.7)

We solve the functional constrained minimization problem (3.5)-(3.6) by Lagrangian methods. The Lagrangian function is:

$$\mathcal{L} = \int_{\mathcal{X}_{N,T}} \int_{\mathcal{Z}_{N,T}} \left\{ \log \left[\frac{g(z|x^a)}{\breve{g}(z|x^a)} \right] g(z|x^a) - \mu(z) g(z|x^a) - \nu(x^a) g(z|x^a) \right\} dz d\hat{F}_x(x^a) \quad \text{(b.5)}$$

where $\mu(\cdot)$ is the functional Lagrange parameter for constraint (3.5) and $\nu(X_{i,t}^a)$ for $i = 1, ..., N$, $t = 1, ..., T$ are the scalar Lagrange multipliers for constraints (3.6). The first-order condition for $g(z|x^a)$ is:

$$1 + \log \left[\frac{g(z|x^a)}{\breve{g}(z|x^a)} \right] - \mu(z) - \nu(x^a) = 0, \quad \text{(b.6)}$$

for all $z \in \mathcal{Z}_{N,T}$ and all $x^a \in \mathcal{X}_{N,T}$ in the sample, which yields

$$g(z|x^a) = \breve{g}(z|x^a) \exp\{\mu(z) + \nu(x^a) - 1\}. \quad \text{(b.7)}$$

We find the Lagrange multipliers $\nu(X_{i,t}^a)$ by imposing the constraints (3.6):

$$1 = \int_{\mathcal{Z}_{N,T}} g(z|x^a) dz = \int \breve{g}(z|x^a) e^{\mu(z)} dz \exp(\nu(x^a) - 1),$$

hence,

$$\exp(\nu(x^a) - 1) = \frac{1}{\int \breve{g}(z|x^a) e^{\mu(z)} dz}.$$

Therefore, from (b.7) we get:

$$g(z|x^a) = \frac{\breve{g}(z|x^a) e^{\mu(z)}}{\int_{\mathcal{Z}_{N,T}} \breve{g}(z|x^a) e^{\mu(z)} dz}. \quad \text{(b.8)}$$

The functional Lagrange multiplier $\mu(z)$ must be such that the conditional density in (b.8) satisfies the constraint in (3.5). Thus

$$\phi(z) = \int_{\mathcal{X}_{N,T}} g(z|x^a) d\hat{F}_x(x^a) = \int_{\mathcal{X}_{N,T}} \breve{g}(z|x^a) \left\{ \int_{\mathcal{Z}_{N,T}} \breve{g}(z|x^a) e^{\mu(z)} dz \right\}^{-1} d\hat{F}_x(x^a) e^{\mu(z)}$$

which yields equation (3.8).

B.4 Proof of Proposition 4

Part (i)

We first show that the estimation error on empirical ranks is asymptotically negligible, i.e.:

$$\hat{\beta}_1 - \tilde{\beta}_1 = o_p(1), \quad \text{(b.9)}$$

where $\hat{\beta}_1 = \arg\max_{\beta_1 \in B_1} \hat{\mathcal{L}}_{NT}(\beta_1)$ and $\tilde{\beta}_1 = \arg\max_{\beta_1 \in B_1} \mathcal{L}_{NT}(\beta_1)$, with $\hat{\mathcal{L}}_{NT}(\beta_1)$ and $\mathcal{L}_{NT}(\beta_1)$ as defined in Assumption A.5. Let $\epsilon > 0$ be given. We have:

$$P\left(\|\hat{\beta}_1 - \tilde{\beta}_1\| > \epsilon\right) \leq P\left(\sup_{\beta_1 \in B_1 : \|\beta_1 - \tilde{\beta}_1\| > \epsilon} \mathcal{L}_{NT}(\beta_1) \geq \mathcal{L}_{NT}(\hat{\beta}_1)\right). \tag{b.10}$$

From Assumptions A.3 (ii) and A.5 (ii), and the consistency of $\tilde{\beta}_1$ (see (b.11) below), we have:

$$\begin{aligned} & \sup_{\beta_1 \in B_1 : \|\beta_1 - \tilde{\beta}_1\| > \epsilon} \mathcal{L}_{NT}(\beta_1) \leq \sup_{\beta_1 \in B_1 : \|\beta_1 - \beta_1^0\| \geq \epsilon/2} \mathcal{L}_{NT}(\beta_1) \leq \sup_{\beta_1 \in B_1 : \|\beta_1 - \beta_1^0\| \geq \epsilon/2} \mathcal{L}(\beta_1) + \eta + o_p(1) \\ \leq \; & \mathcal{L}(\beta_1^0) - 3\eta + o_p(1) \leq \mathcal{L}_{NT}(\tilde{\beta}_1) - 2\eta + o_p(1) \leq \mathcal{L}_{NT}(\tilde{\beta}_1) - \eta, \end{aligned}$$

w.p.a. 1, for a $\eta > 0$. Moreover, we have $\mathcal{L}_{NT}(\hat{\beta}_1) \geq \hat{\mathcal{L}}_{NT}(\hat{\beta}_1) + o_p(1) \geq \hat{\mathcal{L}}_{NT}(\tilde{\beta}_1) + o_p(1) \geq \mathcal{L}_{NT}(\tilde{\beta}_1) + o_p(1)$, where the first and the third inequalities are implied by Assumption A.5 (i). Therefore,

$$P\left(\sup_{\beta_1 \in B_1 : \|\beta_1 - \tilde{\beta}_1\| > \epsilon} \mathcal{L}_{NT}(\beta_1) \geq \mathcal{L}_{NT}(\hat{\beta}_1)\right) \leq P\left(-\eta \geq o_p(1)\right) + o(1) = o(1).$$

Then, from (b.10) we get $P\left(\|\hat{\beta}_1 - \tilde{\beta}_1\| > \epsilon\right) = o(1)$, for any $\epsilon > 0$, which yields (b.9).
We now show that

$$\|\tilde{\beta}_1 - \beta_1^0\| = o_p(1). \tag{b.11}$$

For this purpose, we apply Theorem 1 in Rosemarin (2012). In our model, under Assumptions 1-4 and A.1-A.4, the hypotheses of Rosemarin (2012) hold. In particular, Assumption A1 in Rosemarin (2012), which requires the data to be strong mixing, is implied by our Assumption 1, i.e. that the processes (Z_{it}, X_{it}^a) are iid across individuals, and by Assumption A.4, i.e. the process (Z_{it}, X_{it}^a) of individual i is mixing. Indeed, we can re-name the observations as follows, so that our data are a particular case of time-series with weak serial dependence. We re-arrange the TN observations in a unique vector, such that the first T observations correspond to the first individual, the observations from $T+1$ to $2T$ correspond to the second individual, and so on, until the $NT - th$ observation, corresponding to period T for individual N. More specifically, observations are re-indexed as follows: $Z_j = Z_{i,t} \Leftrightarrow j = T(i-1) + t$, for $j = 1, ..., NT$. And similarly for X^a. In this way, by Assumption 1 we ensure that there is dependence at most for T lags in our observations. The condition of strict stationarity and mixing in Assumption A1 in Rosemarin (2012) can be relaxed to heterogeneous mixing: $\sup_j \alpha_{j,h} \leq Ah^{-\eta}$ where $\eta > 3$, $\alpha_{j,h}$ is the mixing coefficient for the sigma algebras of observations with indices smaller than j and larger than $j + h$. In our case $\sup_j \alpha_{j,h} \leq \sup_i \alpha_{i,h} \leq C\rho^h$, and hence we have heterogeneous strong mixing with $\eta = \infty$.

Further, our Assumptions A.2 (i)-(ii) on the bandwidths imply Assumptions A3 and A5 in Rosemarin (2012). Similarly, our Assumption A.1 on the kernel function implies Assumption A2 in Rosemarin (2012). Moreover, our Assumption A.3 (i) implies Assumption A4 in Rosemarin (2012), and our Assumption A.3 (ii) implies Assumption A6 in Rosemarin (2012). Then, by Theorem 1 in Rosemarin (2012) we get equation (b.11). Equations (b.9) and (b.11) yield Part (i).

Part (ii)

We first show that replacing the unobserved ranks $Z_{i,t}$ with the estimated ranks $\hat{Z}_{i,t}$ has a negligible effect asymptotically under our assumptions, i.e.:

$$\sup_{(z,x^a)\in\mathbb{S}} |\breve{g}(z|x^a) - \tilde{g}(z|x^a)| = o_p(1), \tag{b.12}$$

where $\breve{g}(z|x^a) = \breve{g}(z|\hat{\beta}_1' x^a)$ and $\tilde{g}(z|x^a) = \tilde{g}(z|\tilde{\beta}_1' x^a)$, with $\breve{g}(z|\beta_1' x^a)$ and $\tilde{g}(z|\beta_1' x^a)$ defined as in Assumption A.5 (iii). By the triangular inequality we have:

$$\begin{aligned}\sup_{(z,x^a)\in\mathbb{S}} |\breve{g}(z|x^a) - \tilde{g}(z|x^a)| \quad &\leq \quad \sup_{(z,x^a)\in\mathbb{S}} |\breve{g}(z|\hat{\beta}_1' x^a) - \tilde{g}(z|\hat{\beta}_1' x^a)| \\ &\quad + \sup_{(z,x^a)\in\mathbb{S}} |\tilde{g}(z|\hat{\beta}_1' x^a) - \tilde{g}(z|\tilde{\beta}_1' x^a)|.\end{aligned}$$

The first term in the RHS is bounded by $\sup_{(z,x^a)\in\mathbb{S},\beta_1\in B_1} |\hat{g}(z|\beta_1' x^a) - \tilde{g}(z|\beta_1' x^a)| = o_p(1)$ by Assumption A.5 (iii). The second term in the RHS is $o_p(1)$ by Assumption A.5 (iii) and bound (b.9). This yields (b.12).

Next we show the following probability bound:

$$\sup_{(z,x^a)\in\mathbb{S}} |\tilde{g}(z|x^a) - g(z|\beta_1^{0'} x^a)| = O_p\left(\left(\frac{\log(NT)}{NTH_zH_w}\right)^{1/2}\right) = o_p(1). \tag{b.13}$$

This bound follows directly from Theorem 3 in Rosemarin (2012). With respect to the proof of part (i) of our Proposition 4, here some additional assumptions are required. In particular, we also need Assumptions A7-A10 in Rosemarin (2012). Our Assumption A.1 (ii) implies Assumption A7 in Rosemarin (2012). Assumption A.2 (i) implies Assumption A8 in Rosemarin (2012). Assumption A.3 (i) implies Assumption A9 and Assumption A.3 (iii) implies Assumption A10 in Rosemarin (2012). Moreover, our Assumption A.2 (iii) yield the conditions on the bandwidths H_z, H_w for Theorem 3 in Rosemarin (2012). We conclude the proof by noting that:

$$\sup_{(z,x^a)\in\mathbb{S}} |\hat{g}(z|x^a) - \breve{g}(z|x^a)| = o_p(1). \tag{b.14}$$

This bound can be shown by using the fact that the true conditional pdf $g(z|\beta_1^{0\prime}x^a)$ satisfies the constraints: $\int g(z|\beta_1^{0\prime}x^a)dF(x^a) = \phi(z), \quad \forall z$, and $\int g(z|\beta_1^{0\prime}x^a)dz = 1, \quad \forall x^a$, and if sets $\mathcal{Z}_{N,T}$, $\mathcal{X}_{N,T}$ grow with N, T fast enough. Combining (b.12), (b.13) and (b.14), Part (ii) follows.

B.5 Proof of Proposition 5

The proof builds on the results on two-step Sieve estimators in Hahn, Liao and Ridder (2018) and expands their arguments to account for (i) the preliminary replacement of the unobserved ranks $Z_{i,t}$ with the estimated ranks $\hat{Z}_{i,t}$, and (ii) the panel framework with double asymptotics $N, T \to \infty$ instead of the cross-sectional setting.

Let us define

$$
\begin{aligned}
Q(\theta, f) &= E[l(Y_{i,t}, Y_{i,t-1}, X_{i,t}, \theta, f)], \\
Q_{NT}(\theta, f) &= \frac{1}{N(T-1)} \sum_{i=1}^{N} \sum_{t=2}^{T} l(Y_{i,t}, Y_{i,t-1}, X_{i,t}, \theta, f), \\
\hat{Q}_{NT}(\theta, f) &= \frac{1}{N(T-1)} \sum_{i=1}^{N} \sum_{t=2}^{T} l(\hat{Y}_{i,t}, \hat{Y}_{i,t-1}, X_{i,t}, \theta, f),
\end{aligned}
$$

where $Y_{i,t} = (Z_{i,t}, X_{i,t}^{a\prime})'$ and $\hat{Y}_{i,t} = (\hat{Z}_{i,t}, X_{i,t}^{a\prime})'$ and function l is defined in Assumption A.6. Then, we can write the Sieve ML estimator in (3.10) as

$$
\hat{\theta} = \underset{\theta \in \Theta_m}{\arg\max}\, \hat{Q}_{NT}(\theta, \hat{f}),
$$

where Θ_m with $m = m(N, T)$ is the Sieve parameter space. The true parameter value $\theta_0 \in \Theta$ is the unique maximizer of $Q(\theta, f_0)$ under Assumption A.6 (ii). Note further that $Q_{NT}(\theta, \hat{f})$ would correspond to the empirical Sieve ML criterion if the true ranks were observed. From Proposition 4 (ii) we have a vanishing neighborhood $\mathcal{N}_{f,NT} = \{f \ : \|f - f_0\|_{\mathbb{F}} \leq \delta_{f,NT}\}$, where $\delta_{f,NT} = o(1)$ and $\|f - f_0\|_{\mathbb{F}} = \sup_{(z,x^a)\in\mathbb{S}} |f(z|x^a) - f_0(z|x^a)|$, such that $\hat{f} \in \mathcal{N}_{f,NT}$ w.p.a. 1.

We use the next Lemma, which corresponds to Theorem 5.1 in Hahn, Liao and Ridder (2018) written at a level of generality that covers our framework.

Lemma 1. *Assume that:*

(i) For all $\epsilon > 0$, there exists some non-increasing positive sequence $c_m(\epsilon)$ such that, for all $m \geq 1$,

$$
Q(\theta_0, f_0) - \sup_{\{\theta \in \Theta_m : \|\theta - \theta_0\|_{\Theta} \geq \epsilon\}} Q(\theta, f_0) \geq c_m(\epsilon),
$$

where $\lim_{m \to \infty} c_m(\epsilon) > 0$, $\forall \epsilon > 0$.
(ii) $\theta_0 \in \Theta$, *and* $\Theta_m \subset \Theta_{m+1} \subset \Theta$ *for all* $m \geq 1$, *and there exists* $\theta_m \in \Theta_m$ *such that:*

$$|Q(\theta_m, f_0) - Q(\theta_0, f_0)| = O(\eta_{1,NT}),$$

where $\eta_{1,NT}$ *is a positive sequence tending to zero with* $N, T \to \infty$.
(iii) $\sup_{\theta \in \Theta_m,\ f \in \mathcal{N}_{f,NT}} |\hat{Q}_{NT}(\theta, f) - Q(\theta, f)| = O_p(\eta_{0,NT})$, *where* $\eta_{0,NT}$ *is a finite positive non-increasing sequence tending to zero as* $N, T \to \infty$.
(iv) There is a finite positive non-increasing sequence $\eta_{1,NT}$ *converging to zero such that*

$$\sup_{\theta \in \Theta_m, f \in \mathcal{N}_{f,NT}} |Q(\theta, f) - Q(\theta, f_0)| = O(\eta_{1,NT}).$$

Then $\|\hat{\theta} - \theta_0\|_\Theta \xrightarrow[N,T \to \infty]{p} 0$.

We show that the four conditions in Lemma 1 are implied by our assumptions. Conditions (i) and (ii) are implied by Assumptions A.6 and A.7, respectively. To prove condition (iii) in Lemma 1, we first notice that

$$\sup_{\theta \in \Theta_m,\ f \in \mathcal{N}_{f,NT}} |Q_{NT}(\theta, f) - Q(\theta, f)| = O_p(\eta_{0,NT}), \tag{b.15}$$

from Assumption A.8 (i). Further,

$$\begin{aligned} & \hat{Q}_{NT}(\theta, f) - Q_{NT}(\theta, f) \\ = \; & \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} [\log c(\hat{U}_{i,t}, \hat{U}_{i,t-1}; \rho(\cdot, X'_{i,t}\beta_2)) - \log c(\tilde{U}_{i,t}, \tilde{U}_{i,t-1}; \rho(\cdot, X'_{i,t}\beta_2))], \end{aligned}$$

where $\hat{U}_{i,t} = \int_{-\infty}^{\hat{Z}_{i,t}} f(z|X^a_{i,t})dz$ and $\tilde{U}_{i,t} = \int_{-\infty}^{Z_{i,t}} f(z|X^a_{i,t})dz$. Moreover, $\|\hat{U}_{i,t} - U_{i,t}\| \leq C\|\hat{Z}_{i,t} - Z_{i,t}\|$ for a constant C and any $f \in \mathcal{N}_{f,NT}$, from Assumption A.3 (i). Then, from Assumption A.9 we have:

$$\sup_{\theta \in \Theta_m, f \in \mathcal{N}_{f,NT}} |\hat{Q}_{NT}(\theta, f) - Q_{NT}(\theta, f)| = O_p(C_{NT}\tilde{\eta}_{NT}). \tag{b.16}$$

Thus, from (b.15), (b.16) and the triangular inequality, condition (iii) in Lemma 1 is satisfied. Finally, condition (iv) is implied by Assumption A.8 (ii).

References

1. Abowd, J. M., & Card, D. (1989). On the covariance structure of earnings and hours changes. Econometrica, 411-445.

2. Altonji, J. G., Smith, A. A., & Vidangos, I. (2013). Modeling earnings dynamics. Econometrica, 81(4), 1395-1454.

3. Arellano, M., Blundell, R., & Bonhomme, S. (2017). Earnings and consumption dynamics: A nonlinear panel data framework. Econometrica, 85(3), 693-734.

4. Auten, G., Gee, G., & Turner, N. (2013). Income inequality, mobility, and turnover at the top in the US, 1987–2010. The American Economic Review, 103(3), 168-172.

5. Baker, M., & Solon, G. (2003). Earnings dynamics and inequality among Canadian men, 1976–1992: Evidence from longitudinal income tax records. Journal of Labor Economics, 21(2), 289-321.

6. Beare, B. K. (2010). Copulas and temporal dependence. Econometrica, 78(1), 395-410.

7. Blundell, R., Pistaferri, L., & Preston, I. (2008). Consumption inequality and partial insurance. American Economic Review, 98(5), 1887-1921.

8. Bonhomme, S., & 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.

9. Bonhomme, S., & Robin, J. M. (2010). Generalized non-parametric deconvolution with an application to earnings dynamics. The Review of Economic Studies, 77(2), 491-533.

10. Botosaru, I., & Sasaki, Y. (2018). Nonparametric heteroskedasticity in persistent panel processes: An application to earnings dynamics. Journal of Econometrics, 203(2), 283-296.

11. Bound, J., Brown, C., & Mathiowetz, N. (2001). Measurement error in survey data. Handbook of Econometrics, 5, 3705-3843.

12. Browning, M., Ejrnaes, M., & Alvarez, J. (2010). Modelling income processes with lots of heterogeneity. The Review of Economic Studies, 77(4), 1353-1381.

13. Buchinsky, M., and Hunt, J. (1999). Wage mobility in the United States. Review of Economics and Statistics, 81(3), 351-368.

14. Burkhauser, R. V., & Poupore, J. G. (1997). A cross-national comparison of permanent inequality in the United States and Germany. Review of Economics and Statistics, 79(1), 10-17.

15. Chen, X. (2007). Large sample Sieve estimation of semi-nonparametric models. Handbook of Econometrics, 6, 5549-5632.

16. Chen, X., & Fan, Y. (2006a). Estimation of copula-based semiparametric time series models. Journal of Econometrics, 130(2), 307-335.

17. Chen, X., & Fan, Y. (2006b). Estimation and model selection of semiparametric copula-based multivariate dynamic models under copula misspecification. Journal of Econometrics, 135(1-2), 125-154.

18. Chen, X., Huang, Z., & Yi, Y. (2019). Efficient estimation of multivariate semi-nonparametric GARCH filtered copula models. Working Paper.

19. Chen, X., Koenker, R., & Xiao, Z. (2009). Copula-based nonlinear quantile autoregression. The Econometrics Journal, 12, 50-67.

20. Chen, X., & Pouzo, D. (2012). Estimation of nonparametric conditional moment models with possibly nonsmooth generalized residuals. Econometrica, 80(1), 277-321.

21. Chen, X., & Shen, X. (1998). Sieve extremum estimates for weakly dependent data. Econometrica, 289-314.

22. Chen, X., Wu, W. B., & Yi, Y. (2009). Efficient estimation of copula-based semiparametric Markov models. The Annals of Statistics, 37(6B), 4214-4253.

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

24. Dhaene, G., & Jochmans, K. (2015). Split-panel jackknife estimation of fixed-effect models. The Review of Economic Studies, 82(3), 991-1030.

25. Delecroix, M., Härdle, W., & Hristache, M. (2003). Efficient estimation in conditional single-index regression. Journal of Multivariate Analysis, 86(2), 213-226.

26. Fermanian, J. D., & Scaillet, O. (2003). Nonparametric estimation of copulas for time series. Journal of Risk, 5, 25-54.

27. Fernandez-Val, I., & Lee, J. (2013). Panel data models with nonadditive unobserved heterogeneity: Estimation and inference. Quantitative Economics, 4(3), 453-481.

28. Fernandez-Val, I., & Vella, F. (2011). Bias corrections for two-step fixed effects panel data estimators. Journal of Econometrics, 163(2), 144-162.

29. Fernandez-Val, I., & Weidner, M. (2016). Individual and time effects in nonlinear panel models with large N, T. Journal of Econometrics, 192(1), 291-312.

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

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

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

33. Gagliardini, P., & Gourieroux, C. (2007). An efficient nonparametric estimator for models with nonlinear dependence. Journal of Econometrics, 137(1), 189-229.

34. Gagliardini, P., & Gourieroux, C. (2008). Duration time-series models with proportional hazard. Journal of Time Series Analysis, 29(1), 74-124.

35. Geman, S., & Hwang, C. R. (1982). Nonparametric maximum likelihood estimation by the method of Sieves. The Annals of Statistics, 401-414.

36. Geweke, J., & Keane, M. (2000). An empirical analysis of earnings dynamics among men in the PSID: 1968–1989. Journal of Econometrics, 96(2), 293-356.

37. Gourieroux, C., & Monfort, A. (1995). Statistics and Econometric Models (Vol. 1). Cambridge University Press.

38. Gottschalk, P. (1982). Earnings mobility: Permanent change or transitory fluctuations?. The Review of Economics and Statistics, 450-456.

39. Gottschalk, P. (1997). Inequality, income growth, and mobility: The basic facts. The Journal of Economic Perspectives, 11(2), 21-40.

40. Granger, C. W. (1969). Investigating causal relations by econometric models and cross-spectral methods. Econometrica, 37(3), 424-438.

41. Grenander, U. (1981). Abstract inference. New York: Wiley.

42. Guvenen, F. (2009). An empirical investigation of labor income processes. Review of Economic dynamics, 12(1), 58-79.

43. Hahn, J., Liao, Z., & Ridder, G. (2018). Nonparametric two-step sieve M estimation and inference. Econometric Theory, 34(6), 1281-1324.

44. Hahn, J., & Kuersteiner, G. (2011). Bias reduction for dynamic nonlinear panel models with fixed effects. Econometric Theory, 27(6), 1152-1191.

45. Hahn, J., & Newey, W. (2004). Jackknife and analytical bias reduction for nonlinear panel models. Econometrica, 72(4), 1295-1319.

46. Hall, P., & Yao, Q. (2005). Approximating conditional distribution functions using dimension reduction. Annals of Statistics, 33(3), 1404-1421.

47. Härdle, W. K., Okhrin, O., & Wang, W. (2015). Hidden Markov structures for dynamic copulae. Econometric Theory, 31(5), 981-1015.

48. Härdle, W., Hall, P., & Ichimura, H. (1993). Optimal smoothing in single-index models. Annals of Statistics, 21(1), 157-178.

49. Horowitz, J. L. (1996). Semiparametric estimation of a regression model with an unknown transformation of the dependent variable. Econometrica, 64(1), 103-137.

50. Horowitz, J. L., & Härdle, W. (1996). Direct semiparametric estimation of single-index models with discrete covariates. Journal of the American Statistical Association, 91(436), 1632-1640.

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

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

53. Ichimura, H. (1993). Semiparametric least squares (SLS) and weighted SLS estimation of single-index models. Journal of Econometrics, 58(1-2), 71-120.

54. Imbens, G., Spady, R. H., & Johnson, P. (1998). Information theoretic approaches to inference in moment condition models. Econometrica, 66(2), 333-358.

55. Jensen, S. T., & Shore, S. H. (2015). Changes in the distribution of earnings volatility. Journal of Human Resources, 50(3), 811-836.

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

57. Kim, S. (2013). Wage mobility of foreign-born workers in the United States. Journal of Human Resources, 48(3), 628-658.

58. Kitamura, Y. (2007). Empirical likelihood methods in econometrics: Theory and practice. Advances in Economics and Econometrics, Theory and Applications, Ninth World Congress of the Econometric Society, Vol. 3, ed. by R. Blundell, W. Newey, and T. Persson. Cambridge, MA: Cambridge University Press, 174–237.

59. Kitamura, Y., & Stutzer, M. (1997). An information theoretic alternative to generalized method of moments estimation. Econometrica, 65, 861–874.

60. Kopczuk, W., Saez, E., & Song, J. (2010). Earnings inequality and mobility in the United States: Evidence from social security data since 1937. The Quarterly Journal of Economics, 125(1), 91-128.

61. Li, Q., & Racine, J. S. (2007). Nonparametric econometrics: Theory and practice. Princeton University Press.

62. 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.

63. Lillard, L. A., & Willis, R. J. (1978). Dynamic aspects of earning mobility. Econometrica. 985-1012.

64. Maasoumi, E., & Trede, M. (2001). Comparing income mobility in Germany and the United States using generalized entropy mobility measures. Review of Economics and Statistics, 83(3), 551-559.

65. MaCurdy, T. E. (1982). The use of time series processes to model the error structure of earnings in a longitudinal data analysis. Journal of Econometrics, 18(1), 83-114.

66. Moffitt, R. A., & Gottschalk, P. (2002). Trends in the transitory variance of earnings in the United States. The Economic Journal, 112(478), C68-C73.

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

68. Naguib, C. & Gagliardini, P. (2019). Analytical bias correction for a fixed effects panel model with functional parameter. Working paper.

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

70. Patton, A. J. (2006). Estimation of multivariate models for time series of possibly different lengths. Journal of Applied Econometrics, 21(2), 147-173.

71. Patton, A. J. (2012). A review of copula models for economic time series. Journal of Multivariate Analysis, 110, 4-18.

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

73. Powell, J. L., Stock, J. H., & Stoker, T. M. (1989). Semiparametric estimation of index coefficients. Econometrica, 57(6), 1403-1430.

74. Rémillard, B., Papageorgiou, N., & Soustra, F. (2012). Copula-based semiparametric models for multivariate time series. Journal of Multivariate Analysis, 110, 30-42.

75. Ridder, G. (1990). The non-parametric identification of generalized accelerated failure-time models. Review of Economic Studies, 57(2), 167-181.

76. Rosemarin, R. (2012). Semiparametric estimation of single-index conditional densities for dependent data. Working paper.

77. Schumpeter, J. A. (1955). Imperialism and social classes: Two essays (Vol. 4). Ludwig von Mises Institute.

78. Shorrocks, A. F. (1978). The measurement of mobility. Econometrica, 46(5), 1013-1024.

79. Silverman, B. W. (1986). Density estimation for statistics and data analysis. Chapman & Hall.

80. Sims, C. A. (1972). Money, income, and causality. The American Economic Review, 62(4), 540-552.

81. Sklar, A. (1959). Fonctions de repartition a n dimensions et leurs marges. Inst. Statist. Univ. Paris Publ., 8, 229-231.

82. Van Kerm, P. (2004). What lies behind income mobility? Reranking and distributional change in Belgium, Western Germany and the USA. Economica, 71(282), 223-239.

83. Wong, W. H., & Severini, T. A. (1991). On Maximum Likelihood estimation in infinite dimensional parameter spaces. Annals of Statistics, 19(2), 603-632.

Supplementary material to: "Wage Mobility: A Functional Copula Approach", by C. Naguib and P. Gagliardini.

Appendix C: Additional theoretical background

In this appendix we provide a short review of some theoretical concepts used in the paper. We start with copulas.

C.1 Copulas

Joe (1997) and Nelsen (1999) provide extensive surveys on copula theory. A copula function couples marginal distributions to get a joint distribution. By definition, a joint cumulative distribution function C on the square $[0,1]^2$ with uniform marginal distributions is a copula. Thus, a function $C:[0,1]^2 \to [0,1]$ is a copula if the following conditions hold:

$$C(0,v) = C(u,0) = 0, \forall u, v \in [0,1],$$

$$C(1,v) = v, C(u,1) = u, \forall u, v \in [0,1],$$

and additionally for any rectangle $R = [u_1, u_2] \times [v_1, v_2] \subset [0,1]^2$:

$$\int_R \int C(du, dv) = C(u_2, v_2) - C(u_1, v_2) - C(u_2, v_1) + C(u_1, v_1) \geq 0.$$

When the distribution C is continuous, the associated density is called the copula density:

$$c(u,v) = \frac{\partial^2 C(u,v)}{\partial u \partial v},\ u, v \in [0,1].$$

C.1.1 Sklar's theorem (Sklar (1959))

Let $F(x,y)$ be the bivariate cdf of random variables X and Y, with marginal cdf's equal to F_x and F_y, respectively. Let then $U = F_x(X)$ and $V = F_y(Y)$. Variables U and V have uniform marginal distributions on $[0,1]$. The joint cdf of variables U and V is given by:

$$P(U \leq u, V \leq v) = P(X \leq F_x^{-1}(U), Y \leq F_y^{-1}(V)) = F(F_x^{-1}(u), F_y^{-1}(v)),\ \forall u, v \in [0,1]$$

Then there exists a copula such that:

$$C(u,v) = F(F_x^{-1}(u), F_y^{-1}(v)), \forall u, v \in [0,1].$$

i.e. $F(x,y) = C[F_x(x), F_Y(y)],\ \forall x, y.$

This copula is unique if F is a continuous distribution. Assuming differentiability conditions, it is possible to obtain the following conditional cdf: $P(U \leq u \mid V = v) = \partial C(u,v)/\partial v$. The conditional density function is obtained by deriving once more with respect to u. The copula cdf is limited from above and from below by the so-called Frechet-Hoeffding bounds. The lower bound is: $C_L(u,v) = max(u+v-1;0)$, while the upper bound is: $C_U(u,v) = min(u,v)$. The lower bound C_L models perfect negative dependence, while the upper bound C_U models perfect positive dependence. In the case of independence we have $C_I(u,v) = u \cdot v$. For a review of the many parametric families of copulas we refer to Joe (1997) and Nelsen (1999). Here we focus on one of them, namely the Gaussian copula.

C.1.2 The Gaussian copula

A Gaussian copula is associated with a bivariate Gaussian distribution with zero expected values, unitary variances and correlation equal to ρ. For $-1 \leq \rho \leq 1$, the Gaussian copula is defined as:

$$C(u,v;\rho) = \Phi_2(\Phi^{-1}(u), \Phi^{-1}(v); \rho)$$

where Φ is the $N(0,1)$ cdf, Φ^{-1} is the quantile function (the inverse of Φ) and $\Phi_2(\cdot,\cdot;\rho)$ is the bivariate standard normal cdf with correlation parameter equal to ρ. The Gaussian copula density is defined as:

$$c(u,v;\rho) = \frac{\phi_2(\Phi^{-1}(u), \Phi^{-1}(v); \rho)}{\phi(\Phi^{-1}(u))\phi(\Phi^{-1}(v))}$$

where ϕ is the standard Gaussian density and $\phi_2(x,y;\rho)$ is the density of the bivariate standard normal distribution with correlation coefficient ρ. More explicitly,

$$c(u,v;\rho) = \frac{1}{\sqrt{1-\rho^2}} \cdot \exp\left(\frac{x^2+y^2}{2} + \frac{2\rho xy - x^2 - y^2}{2(1-\rho^2)}\right)$$

with $x = \Phi^{-1}(u)$ and $y = \Phi^{-1}(v)$. It is known that the Gaussian copula $C(\cdot,\cdot;\rho)$ equals the Frechet-Hoeffding bound C_U for $\rho = 1$, and C_L for $\rho = -1$, while it equals the independence copula for $\rho = 0$ (Joe 1997 and Nelsen 1999). The bivariate Gaussian copula features zero tail dependence.

C.2 Non-causality

In this section we review some causality concepts and related results that are used in the paper.

C.2.1 Granger (1969) non-causality

Let us consider two stochastic processes, (Z_t) and (X_t). If the linear predictor of the current value of X, given its own past and the past of Z, does not depend on the latter, then there is absence of Granger causality from Z to X. This definition, which has been originally formulated for the linear predictor case, can be generalized to density functions. Process (Z_t) is said not to cause process (X_t) in the Granger sense if the following holds:

$$l(X_t|\underline{Z_{t-1}}, \underline{X_{t-1}}) = l(X_t|\underline{X_{t-1}}) \tag{c.1}$$

where $\underline{Z_{t-1}} = (Z_{t-1}, Z_{t-2}, ...)$ and $\underline{X_{t-1}} = (X_{t-1}, X_{t-2}, ...)$, and $l(\cdot)$ represents the conditional density.

C.2.2 Sims (1972) non-causality

Let us consider two stochastic processes, (Z_t) and (X_t). Process Z does not cause process X in the Sims definition if the linear predictor of Z_t, based on the entire history $..., X_{t-1}, X_t, X_{t+1}, ...$ is identical to the linear predictor of Z_t based on $X_t, X_{t-1}, ...$ alone. In this case, too, a generalization of the above-presented definition is possible. We say that X does not Sims-cause Z if the following holds:

$$l(Z_t|\underline{Z_{t-1}}, \underline{X_T}) = l(Z_t|\underline{Z_{t-1}}, \underline{X_t}) \tag{c.2}$$

for any $T \geq t$ where $\underline{X_T} = (X_1, ..., X_t, ..., X_T)$ and $\underline{X_t} = (X_1, ..., X_t)$.

C.2.3 Some useful lemmas

We provide here some lemmas that are used in the proof of Proposition 1. In the remainder of this Section, we assume that X_t is strictly stationary, i.e. that $X_t = X_t^a$.

Lemma 2. *If there is absence of Granger causality from Z to X, then the following holds:*

$$l(\overline{X^{t+1}}|\underline{Z_t}, \underline{X_t}) = l(\overline{X^{t+1}}|\underline{X_t})$$

where $\overline{X^{t+1}} = (X_{t+1}, ..., X_T)$.

Proof. We have:

$$\begin{aligned} l(\overline{X^{t+1}}|\underline{Z_t}, \underline{X_t}) &= l(X_{t+1}, X_{t+2}, ..., X_T|\underline{Z_t}, \underline{X_t}) \\ &= l(X_T|X_{t+1}, ..., X_{T-1}, \underline{Z_t}, \underline{X_t})l(X_{t+1}, ..., X_{T-1}|\underline{Z_t}, \underline{X_t}) \\ &= l(X_T|\underline{Z_t}, \underline{X_{T-1}})l(X_{t+1}, ..., X_{T-1}|\underline{Z_t}, \underline{X_t}). \end{aligned} \tag{c.3}$$

The first term in the right hand side of equation (c.3) is equal to $l(X_T|\underline{X_{T-1}})$, independent of $\underline{Z_t}$, due to the assumption of Granger non-causality. The same argument can be applied recursively to show that $l(X_{t+1}, ..., X_{T-1}|\underline{Z_t}, \underline{X_t})$ does not depend on $\underline{Z_t}$. Therefore, we have proved that:

$$l(\overline{X^{t+1}}|\underline{Z_t}, \underline{X_t}) = l(\overline{X^{t+1}}|\underline{X_t}),$$

which yields the conclusion.

□

Lemma 3. *Granger non-causality implies Sims non-causality (in the sense of equations* (c.1) *and* (c.2)*).*

It is known that the converse statement also holds (see e.g. Gourieroux and Monfort (1995)). Therefore, the two non-causality concepts are equivalent. For completeness, we present here a proof of Lemma 3, which is similar to that presented in Gourieroux and Monfort (1995).

Proof. We assume that Z_t does not cause X_t in the definition of Granger and we want to show that this implies that Z_t does not cause X_t in the Sims sense. Let us consider the conditional density of Z_t given its own past history, $\underline{Z_{t-1}}$ and the whole history of X, denoted by $\underline{X_T}$. The following decomposition is always valid:

$$l(Z_t|\underline{Z_{t-1}}, \underline{X_T}) = \frac{l(\underline{Z_t}, \underline{X_T})}{l(\underline{Z_{t-1}}, \underline{X_T})} = \frac{l(\underline{Z_t}, \underline{X_t})l(\overline{X^{t+1}}|\underline{Z_t}, \underline{X_t})}{l(\underline{Z_{t-1}}, \underline{X_t})l(\overline{X^{t+1}}|\underline{Z_{t-1}}, \underline{X_t})}, \tag{c.4}$$

where $\overline{X^{t+1}} = (X_{t+1}, ..., X_T)$.
Using Lemma 2, equation (c.4) yields:

$$l(Z_t|\underline{Z_{t-1}}, \underline{X_T}) = \frac{l(\underline{Z_t}, \underline{X_t})l(\overline{X^{t+1}}|\underline{X_t})}{l(\underline{Z_{t-1}}, \underline{X_t})l(\overline{X^{t+1}}|\underline{X_t})} = \frac{l(\underline{Z_t}, \underline{X_t})}{l(\underline{Z_{t-1}}, \underline{X_t})} = l(Z_t|\underline{Z_{t-1}}, \underline{X_t}), \tag{c.5}$$

i.e. that Z does not cause X in the Sims sense. □

Lemma 4. *If (Z_t) is a first-order Markovian process conditional on (X_t), namely $l(Z_t|\underline{Z_{t-1}}, \underline{X_T}) = l(Z_t|Z_{t-1}, \underline{X_T})$, and if there is absence of Sims-causality from Z to X, then the following holds:*

$$l(Z_t|\underline{Z_{t-1}}, \underline{X_t}) = l(Z_t|Z_{t-1}, \underline{X_t})$$

Proof. From the assumption of Sims non-causality from Z to X we can write the following:

$$l(Z_t \mid \underline{Z_{t-1}}, \underline{X_t}) = l(Z_t \mid \underline{Z_{t-1}}, \underline{X_T}). \tag{c.6}$$

From the assumption of first-order markovianity of Z_t conditional on (X_t) we know that:

$$l(Z_t|\underline{Z_{t-1}}, \underline{X_T}) = l(Z_t|Z_{t-1}, \underline{X_T}). \tag{c.7}$$

This last conditional density can be rewritten as it follows:

$$l(Z_t|Z_{t-1}, \underline{X_T}) = \frac{l(Z_t, Z_{t-1}, \underline{X_T})}{l(Z_{t-1}, \underline{X_T})} = \frac{l(Z_t, Z_{t-1}, \underline{X_t})l(\overline{X^{t+1}}|Z_t, Z_{t-1}, \underline{X_t})}{l(Z_{t-1}, \underline{X_t})l(\overline{X^{t+1}}|Z_{t-1}, \underline{X_t})}.$$

From Lemma 3 we know that Sims non-causality implies Granger non-causality (the two concepts are equivalent). Therefore, from the assumption of Granger non-causality from Z to X and Lemma 2 we obtain:

$$l(Z_t|Z_{t-1}, \underline{X_T}) = \frac{l(Z_t, Z_{t-1}, \underline{X_t})l(\overline{X^{t+1}}|\underline{X_t})}{l(Z_{t-1}, \underline{X_t})l(\overline{X^{t+1}}|\underline{X_t})} = \frac{l(Z_t, Z_{t-1}, \underline{X_t})}{l(Z_{t-1}, \underline{X_t})} = l(Z_t|Z_{t-1}, \underline{X_t}) \tag{c.8}$$

From (c.6), (6.19) and (c.8) we get:

$$l(Z_t|\underline{Z_{t-1}}, \underline{X_t}) = l(Z_t|Z_{t-1}, \underline{X_T}) = l(Z_t|Z_{t-1}, \underline{X_t}).$$

□

Lemma 5. *If there is absence of Granger causality from Z to X, then the following holds:*

$$l(Z_t|X_{t-p}, ..., X_{t+q}) = l(Z_t|X_{t-p}, ..., X_t),$$

for any integers $p, q \geq 0$.

Proof. We have

$$\begin{aligned} l(Z_t|X_{t-p}, ..., X_{t+q}) &= \frac{l(Z_t, X_{t-p}, ..., X_{t+q})}{l(X_{t-p}, ..., X_{t+q})} \\ &= \frac{l(X_{t+1}, ..., X_{t+q}|Z_t, X_t, ..., X_{t-p})l(Z_t, X_t, ..., X_{t-p})}{l(X_{t+1}, ..., X_{t+q}|X_t, ..., X_{t-p})l(X_t, ..., X_{t-p})}. \end{aligned}$$

From the assumption of Granger non-causality and Lemma 2, we obtain the following result:

$$l(Z_t|X_{t-p}, ..., X_{t+q}) = \frac{l(Z_t, X_t, ..., X_{t-p})}{l(X_t, ..., X_{t-p})} = l(Z_t|X_t, ..., X_{t-p}).$$

□

Lemma 6. *If there is absence of Granger causality from Z to X, then:*

$$l(Z_{t-1}|X_{t-1}, X_t) = l(Z_{t-1}|X_{t-1}).$$

Proof. It follows from Lemma 5 with $p = 0, q = 1$, and stationarity.

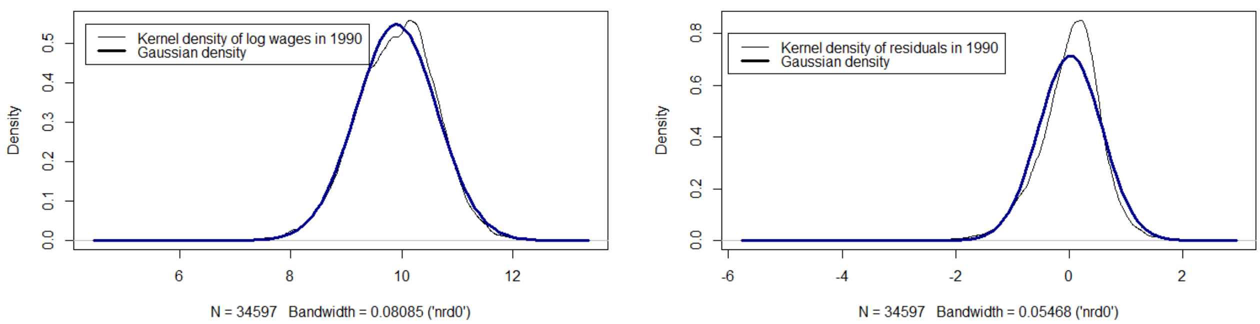
□

Appendix D: Additional data analysis

This Appendix provides additional results supporting our empirical analysis on the PSID dataset.

D.1 Kernel densities of real wages and transitory component

Figure 7: Density of the log wages and of the residual wage component in 1990



The left panel of this figure displays the estimated kernel density of the log wages in year 1990 in our sample. This density is compared to the Gaussian density with the same mean and standard deviation. The right panel of this figure displays the estimated kernel density of the residual wage component in year 1990 in our sample. This density is compared to a Gaussian density with the same mean and standard deviation.

D.2 A simplified model of ranks dynamics

In this subsection we present a simplified model for the Gaussian rank dynamics. Compared to the semi-nonparametric copula model introduced in Section 2 of the paper, the model of this subsection does not include observable regressors and does not impose the structural restriction that the cross-sectional distribution of Gaussian ranks is $N(0,1)$ by definition. The goal of our analysis is to show the empirical relevance of the more sophisticated specification displayed in Section 2-3 to uncover the patterns of wage mobility. We consider a regression model, in which linear, quadratic and cubic terms in the past rank are present:

$$Z_{i,t} = a_0 + aZ_{i,t-1} + bZ_{i,t-1}^2 + cZ_{i,t-1}^3 + e_{i,t}. \tag{d.1}$$

We choose to estimate a polynomial rank regression of third degree in order to allow for asymmetry in the relationship between the present and the past ranks, given that such asymmetry emerges from the data. From the polynomial regression reported in Table 2, we deduce that the coefficient of the linear term in equation (d.1) is rather high (0.93), thus suggesting a substantial degree of positional immobility from one year to the following one. The coefficients of the quadratic term and that of the cubic past rank term are statistically significant at a 99% confidence level, thus providing preliminary evidence supporting the hypothesis of a nonlinear relationship between the present and the past Gaussian ranks. The nonlinear patterns are more carefully highlighted by the semi-nonparametric specification of Sections 2 and 3.

Table 2: Third-degree polynomial regression

	Full sample (1968-1997)
constant	-.0181476***
	(.0030926)
$Z_{i,t-1}$	.9289741***
	(.0042103)
$Z^2_{i,t-1}$	.0335625***
	(.0018814)
$Z^3_{i,t-1}$	-.0197652***
	(.0011985)
n. of obs.	37'439

This table reports the estimated coefficients of the third-degree polynomial regression in (d.1). This regression has been run on pooled PSID data for the period 1968-1997.

It is worth underlying that he empirical analysis conducted in Section 5 of the paper with our semi-nonparametric copula model including observable regressors shows that the patterns of mobility strongly vary with the individual characteristics. Hence, the analysis with the simplified model (d.1) can be misleading, highlighting the empirical relevance of the fully-fledged specification developed in Sections 2 and 3.

In unreported results, we regress the cross-sectional correlation between the present and the past ranks over the annual GDP growth rate[21], in order to check for the stability over time of the relationship between the present and the past rank, and we find the the estimated coefficient is very small and not statistically significant. Hence, we deduce that the correlation between the present and the past ranks exhibits no cyclical behavior. We also tried to regress our correlation on the annual unemployment rate in the US and on the share of immigrants on the total population[22]. In both cases, the estimated association between the two variables was not statistically significant (results not reported for brevity). This convinced us not to include macroeconomic factors among the explanatory variables that we deem at the origin of the rank dynamics[23].

[21]Data source: World Bank Indicators

[22]Data source: OECD

[23]Note that, in all the three linear regressions described here, the dependent variable, i.e. the correlation between the present

D.3 Numerical implementation of the semi-nonparametric estimators

The computation of the proposed estimators requires numerical evaluation of some univariate integrals. The integral in equation (3.9) is obtained by Gaussian quadrature, whereas the integrals in equations (2.18) and (2.22) are computed by simulation, with the number of simulations equal to $S = 2000$. The fix-point problem defining function $\mu(\cdot)$ in equation (3.7) is solved iteratively, where the integral over z is computed by Gaussian quadrature. The choice of the technique to be used for the numerical evaluation of those integrals has been driven by considerations of numerical efficiency.

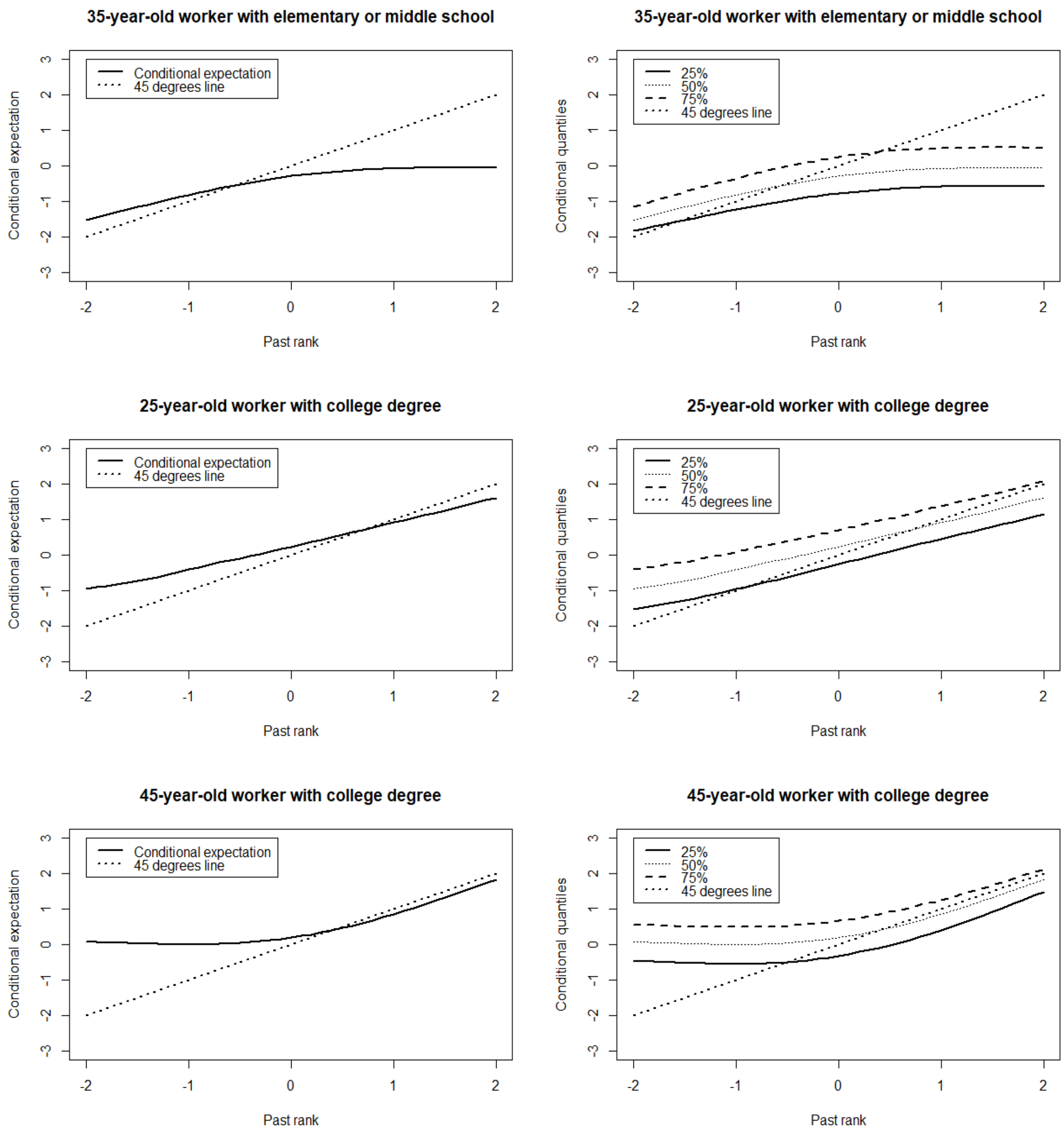
As far as the nonparametric estimation of the marginal rank distributions is concerned, the kernel K is the standard Gaussian pdf. The bandwidths in equation (3.3) have been selected as follows: $h_z = \sigma_z(NT)^{-1/6}$ and $h_w = \sigma_w(NT)^{-1/6}$, where σ_z and $\sigma_w = \sigma_w(\beta_1)$ are sample standard deviations of variables Z_{it} and $W_{1,it}(\beta_1)$, for any given β_1, which corresponds to the Silverman rule for bivariate data. Moreover, we set $H_z = h_z$ and $H_w = h_w$ in (3.4). For the Sieve estimation of our copula model, the polynomial basis used to approximate the functional parameter $\rho(\cdot,\cdot)$ in (3.12) is a Hermite basis of degree $m = 2$. We also used other orders for the Hermite polynomial basis, e.g. $m = 3, 4$ or 5, and we found that the estimation results did not change significantly. Using larger numbers of polynomials in the basis increases the estimators variance.

---

and the past rank ρ has been transformed by means of the inverse logistic function, $\log(\rho/(1-\rho))$, so that it has an unbounded support.

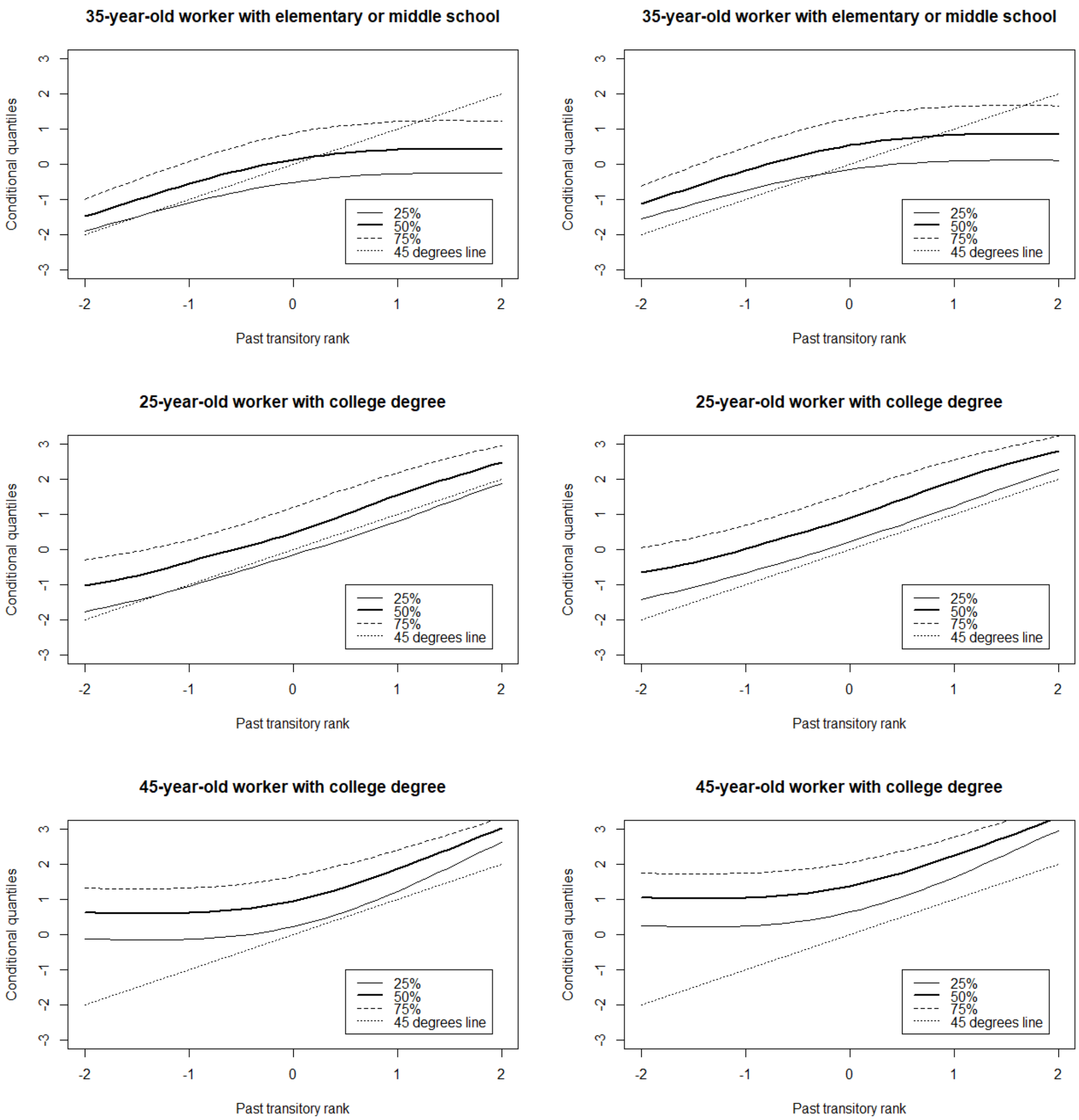
D.4 Additional estimation results

Figure 8: Conditional expectations and conditional quantiles of the residual wage component



In the left panels we display the conditional expected rank as a function of the past rank for different age and education levels. In the right panels we report three conditional quantiles (median, lower and upper quartiles) as a function of the past rank for the same sets of individual characteristics. The conditional quantiles are obtained from the empirical version of (2.24) with $u = 0.25, 0.50, 0.75$. The chosen sets of individual characteristics correspond to different values of the marginal distribution and mobility scores, which are $W_{1,it} = -1.15$ and $W_{2,it} = -1.69$ in the upper panels, $W_{1,it} = 1.08$ and $W_{2,it} = -0.19$ in the middle panels, and $W_{1,it} = 1.57$ and $W_{2,it} = 1.16$ in the bottom panels. Both index values have been standardized.

Figure 9: Conditional wage quantiles



In this figure we display the conditional quartiles of the wage ranks $Q_{r,t}(u) = b_t[Q_{Z,t}(u|Z_{i,t-1}, X_{it}, X_{i,t-1}), \eta_i, X_{it}]$ as a funtion of the past residual rank $Z_{i,t-1}$ for year $t = 1990$. The chosen sets of individual characteristics correspond to different values of the marginal distribution and mobility scores, which are $W_{1,it} = -1.15$ and $W_{2,it} = -1.69$ in the upper panels, $W_{1,it} = 1.08$ and $W_{2,it} = -0.19$ in the middle panels, and $W_{1,it} = 1.57$ and $W_{2,it} = 1.16$ in the bottom panels. Both index values have been standardized. In each panel, the curves correspond to median (50%), and lower and upper quartile (25% and 75%). In the left panels, the individual effect η_i is equal to the median value of the distribution, whereas in the right panels it is equal to the 75% quantile.