UNIVERSITÄT
BERN

Faculty of Business, Economics
and Social Sciences

Department of Economics

A Semi-nonparametric Copula Model for Earnings Mobility

Costanza Naguib, Patrick Gagliardini

23-02

January, 2023

DISCUSSION PAPERS

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

A Semi-nonparametric Copula Model for Earnings Mobility

Costanza Naguib[*], Patrick Gagliardini[†]

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. We show the asymptotic normality of the Sieve estimator for our preferred mobility measure, which is an irregular functional of both the finite- and infinite-dimensional parameters, in the double asymptotics with $N, T \to \infty$. We derive an analytical bias correction for the incidental parameters bias induced by the individual fixed-effects. 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, Sieve 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. 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

[*]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 a Co-Editor, two anonymous referees, 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.

within the cross-sectional distribution of the residuals from a wage regression. The worker's percentile in this cross-sectional distribution is called wage residual rank. In contrast with the previous literature, where fully parametric copulas have been used to model the joint distribution of the current and past individual ranks[2], we use a semi-nonparametric specification of the copula, 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 (nonlinear) dependence between current and past ranks, conditional on individual characteristics, defines relative mobility. While our empirical analysis focus on the labor market, the modeling and estimation methodologies in this paper are of more general interest for studying positional mobility.

The methodological contributions of this paper are manifold. The first one 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 uniform 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 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, in the second step we estimate the semi-nonparametric copula specification with the method of Sieves, see e.g. Wong and Severini (1991), Chen and Shen (1998), Ai and Chen (2003), Chen (2007). Because in the second step the likelihood function involves empirical ranks, our theoretical developments share some similarity with the results in Chen, Huang and Yi (2021) who apply Sieves estimation on GARCH filtered residuals. [3] Fourth, we develop a theory of asymptotic normality for some functionals of the finite- and infinite-dimensional parameters of interest, such as our preferred mobility measure, in a double panel asymptotics with both the numbers of individuals N and time periods T tending to infinity. As in Chen, Liao and Sun (2014)

[2] see Joe (1997) and Nelsen (1999) for a background and a review of parametric copula specifications.

[3] In a similar vein, Chen, Xiao and Wang (2020) estimate a parametric copula specification on filtered residuals of a model with nonstationary regressors.

and Chen and Liao (2014), the functionals in our paper are irregular, i.e. their directional derivative admits a Rietz representer with unbounded norm. Our estimation methodology suffers from an incidental parameters bias, due to the individual fixed effects in the preliminary panel regression used to obtain the wage residuals. [4] We characterize this asymptotic bias and we show how to perform an analytical bias correction. The present paper hence lies at the intersection of the literature on parametric bias correction in a panel framework and of that on semi-nonparametric Sieve estimation. Evidence from Monte Carlo simulations, for designs mimicking our empirical analysis with an unbalanced dataset of individuals with an average permanence in the sample of 15 years, shows that we can effectively correct up to 40-70% of the bias in the estimated mobility measure depending on the setting.

In an empirical illustration we estimate our semi-nonparametric copula 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 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 and/or scarce experience (proxied by age).

This paper is not the first one to use copulas for studying individual earnings dynamics. 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 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

[4] 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) for correcting the incidental parameter bias on a finite-dimensional parameter with double panel asymptotics.

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 (2021)), or both cross-sectional and serial dependence (e.g. Rémillard, 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. Appendix C concerns numerical implementation. In the Supplementary Materials to the paper we present additional theoretical as well as 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. 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 (including the constant), 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). 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 of log wage around the individual life-cycle and the macro trends. [5] In the second step we focus on modeling the dynamics of the residual component $\varepsilon_{i,t}$. For this purpose, we consider its rank, which is defined as

$$U_{i,t} = F_{\varepsilon,t}(\varepsilon_{i,t}), \tag{2.2}$$

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 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. In the remainder of this section we specify a semi-nonparametric model for the dynamics of the joint process (U_{it}, X_{it}) of ranks and individual observed characteristics. The fundamental difference between our approach and the vast majority of the existing literature on individual earnings dynamics 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 (U_{it}, X_{it}). We assume independence and identical distribution (iid) across individuals[6].

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

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

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

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

[5] 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.

[6] 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.

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

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

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

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

We use the following Assumptions 2-5 on these conditional densities.

Assumption 2. *Process (U_{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). Assumption 2 implies that the conditional density in (2.4) is such that:

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

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

Assumption 3. *Process (X_{it}) is Markovian of first-order with transition density $l(X_{it}|X_{i,t-1})$ and strictly stationary.*

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

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

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

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

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

As a consequence, the joint process (U_{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

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

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 U_{it} as a uniform rank, we need to ensure that the unconditional density of U_{it} is uniform on $[0, 1]$:

$$U_{it} \sim U(0, 1). \tag{2.9}$$

In the next subsection we prove that, under Assumptions 1-4 and additional constraints on transition density (2.8), the distributional property in (2.9) holds.
Finally, we need to impose an orthogonality condition between error terms and regressors to identify the parameter vector α in equation (2.1).

Assumption 5. *The joint distribution of covariates vector X_{it} and error term ε_{it} is such that*

$$E[(X_{it} - \bar{X}_{i\cdot})(\varepsilon_{it} - \bar{\varepsilon}_{i\cdot})] = 0, \;\; t = 1, ..., T, \;\; \forall T, \tag{2.10}$$

where $\bar{X}_{i\cdot} = \frac{1}{T}\sum_{t=1}^{T} X_{it}$ and similarly for $\bar{\varepsilon}_{i\cdot}$.

Assumption 5 involves regressors and errors in difference from their time means because of the individual effects in equation (2.1).

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 $g(\cdot|X)$ be a pdf of a random variable on the interval $[0, 1]$, for any value of X, and let $G(\cdot|X)$ be the corresponding cdf. We assume that g is such that:

$$\int g(U|X)l(X)dX = 1, \qquad \forall\, U \in (0, 1), \tag{2.11}$$

where $l(\cdot)$ denotes the stationary pdf of X_t. Further, let us define:

$$l(U_t|U_{t-1}, X_t, X_{t-1}) = g(U_t|X_t)c[G(U_t|X_t), G(U_{t-1}|X_{t-1}); \rho(\cdot, X_t)], \tag{2.12}$$

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 (2.12) defines a valid conditional density function. Note that equation (2.12) is an hypothesis of the model, as it is not directly derived from Assumptions 1-5 only. 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:*

$$U_0|X_0 \sim g(\cdot|X_0), \tag{2.13}$$

and X_0 is drawn from the stationary distribution of the regressors vector. Then, under Assumptions 1-5:

$$U_t \sim U(0,1) \tag{2.14}$$

for all $t \geq 1$.

The proof of this proposition is provided in Appendix B. Proposition 1 shows that, if the uniform rank process is initialized with a conditional distribution $g(\cdot|X_0)$ satisfying property (2.11), and the transition density is as in (2.12), then the condition of standard uniform marginal distribution for the rank process is met. In fact, the proof of Proposition 1 shows that $g(U_t|X_t)$ is the conditional pdf of U_t given X_t at any date t, and $c(\cdot,\cdot;\rho(\cdot,X_t))$ is the conditional copula pdf of (U_t, U_{t-1}) given X_t, X_{t-1}. Proposition 1 allows us to introduce a model which is compatible with the condition of uniform distribution for the rank U_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 stationary.

2.3 The nonparametric family of autoregressive copulas

In the previous section we have shown how to specify a joint dynamics for rank U_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 U_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: [7]

$$U_t = \Lambda[\rho(U_{t-1}) + \omega_t], \tag{2.15}$$

[7]Introducing further lags of the 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 ranks is uniform on $[0,1]$ - a condition which is needed for model coherence - in the presence of more than one lagged value of the ranks as arguments of the autoregressive function.

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. 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 (2.15) 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[8]. Equation (2.15) defines a time-homogeneous Markov process. Let us now derive the conditions on functions ρ and Λ, such that (U_t) is a strictly stationary process with unique invariant distribution, that is uniform on the interval $[0,1]$. The conditional cdf is:

$$P[U_t \leq u | U_{t-1} = v] = P[\omega_t \leq \Lambda^{-1}(u) - \rho(v)] = \Phi[\Lambda^{-1}(u) - \rho(v)], \tag{2.16}$$

where Φ denotes the cdf of standard Gaussian distribution. If we impose the property of uniform marginal distribution by integrating out U_{t-1}, we get:

$$u = P[U_t \leq u] = E[P(U_t \leq u | U_{t-1})] = \int_0^1 \Phi[\Lambda^{-1}(u) - \rho(v)] dv, \tag{2.17}$$

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

$$\Lambda(y) = \int_0^1 \Phi[y - \rho(v)] dv, \tag{2.18}$$

for $y \in \mathbb{R}$, i.e. function $\Lambda(\cdot)$ is univoquely determined by function $\rho(\cdot)$ under the condition of uniform margins for the Markov process (U_t). We summarize the result in the next Proposition.

Proposition 2. *For the Markov process defined by equation (2.15), the invariant distribution of U_t is uniform on $[0,1]$ if, and only if, the function $\Lambda(\cdot)$ is given by the following expression:*

$$\Lambda(y) = \int_0^1 \Phi[y - \rho(v)] dv, \tag{2.19}$$

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

Thus, the copula of (U_t, U_{t-1}) is completely characterized by function ρ. The copula is invariant to shifts of the functional parameter $\rho(\cdot) \to \rho(\cdot) + c$, for any constant c. We can normalize the parameter by imposing the restriction $\rho(u^*) = 0$ for a given $u^* \in (0,1)$.

[8] 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.

Let us now derive the copula pdf associated with the nonlinear autoregressive model (2.15). The nonlinear autoregressive copula is the joint distribution of U_t and U_{t-1}. The copula pdf is given by:

$$c(u, v; \rho(\cdot)) = \frac{\partial}{\partial u} P[U_t \leq u | U_{t-1} = v] = \frac{\phi[\Lambda^{-1}(u) - \rho(v)]}{\lambda[\Lambda^{-1}(u)]}, \tag{2.20}$$

for the arguments $u, v \in [0, 1]$, where $\lambda(y) = \int_0^1 \phi[y - \rho(v)]dv$ is the derivative of function Λ and $\phi = \Phi'$. This copula pdf is parameterized by function $\rho(\cdot)$. This copula family contains the Gaussian copula with correlation parameter $r/\sqrt{1 + r^2}$, for $r \geq 0$, when $\rho(v) = r\Phi^{-1}(v)$. [9] Hence, we get a nonparametric copula family that extends the parametric Gaussian copula model providing more flexibility. However, the copula family specified in equation (2.20) does not include any other commonly used parametric copula families besides the Gaussian copula.

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. This is possible since essentially any function $\rho(\cdot)$ is admissible as a parameter of the autoregressive copula. To cope with the curse of dimensionality in nonparametric estimation, we assume an index model specification for the effect of the observable characteristics. Hence, from (2.12) the conditional density of U_{it} given $U_{i,t-1}$, X_{it}, $X_{i,t-1}$ is:

$$l(U_{i,t}|U_{i,t-1}, X_{i,t}, X_{i,t-1}) = g(U_{i,t}|X'_{i,t}\beta_1) \cdot c[G(U_{i,t}|X'_{i,t}\beta_1), G(U_{i,t-1}|X'_{i,t-1}\beta_1), \rho(\cdot, X'_{i,t}\beta_2)] \tag{2.21}$$

where $g(\cdot|X'_{i,t}\beta_1)$ is the pdf of the distribution of the rank, conditional on the individual variables, $G(\cdot|X'_{i,t}\beta_1)$ is the corresponding cdf, $c[\cdot, \cdot, \rho(\cdot, X'_{it}\beta_2)]$ is the copula density in (2.20) with functional parameter $\rho(\cdot, X'_{i,t}\beta_2)$, and function $\Lambda(\cdot)$ replaced by

$$\Lambda(y; X'_{it}\beta_2) = \int_0^1 \Phi[y - \rho(v, X'_{it}\beta_2)]dv. \tag{2.22}$$

The model features two indexes, namely $W_{1,it} = X'_{it}\beta_1$ for the marginal univariate distribution, and $W_{2,it} = X'_{it}\beta_2$ for the copula functional parameter, where $\beta_1, \beta_2 \in \mathbb{R}^p$. 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, 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.

To implement the constraints on the conditional density g to be positive and integrate to 1 across u, we write the joint pdf of variables $U_{i,t}$ and $W_{1,it}$ as $h(u, w)^2$, for a square integrable function h of arguments

[9] Indeed, in this case we have $\Lambda(y) = \Phi\left(\frac{y}{\sqrt{1+r^2}}\right)$ and equation (2.15) becomes $\Phi^{-1}(U_t) = \frac{1}{\sqrt{1+r^2}}\left(r\Phi^{-1}(U_{t-1}) + \omega_t\right)$.

$u \in [0,1]$ and $w \in \mathbb{R}$. Then, we have $g(u|w) = h(u,w)^2 / \int_0^1 h(u,w)^2 du$. We impose the condition in (2.11), namely that the marginal distribution of $U_{i,t}$ is uniform on interval $[0,1]$, by the moment restrictions $\int_0^1 \int_{-\infty}^{\infty} u^l h(u,w)^2 du dw = \frac{1}{l+1}$ for all $l = 0, 1, ...$ Thus, the parametrized log conditional density becomes:

$$\log l(U_{i,t}|U_{i,t-1}, X_{i,t}, X_{i,t-1}; \theta) = \log \left(\frac{h(U_{i,t}, X'_{i,t}\beta_1)^2}{\int_0^1 h(u, X'_{i,t}\beta_1)^2 du} \right) + \log c \left(\frac{\int_0^{U_{i,t}} h(u, X'_{i,t}\beta_1)^2 du}{\int_0^1 h(u, X'_{i,t}\beta_1)^2 du}, \frac{\int_0^{U_{i,t-1}} h(u, X'_{i,t-1}\beta_1)^2 du}{\int_0^1 h(u, X'_{i,t-1}\beta_1)^2 du}, \rho(\cdot, X'_{i,t}\beta_2) \right), \quad (2.23)$$

where parameter $\theta = (\beta_1, \beta_2, h, \rho)$ contains both the finite-dimensional vectors $\beta_1, \beta_2 \subset \mathbb{R}^p$ and the infinite-dimensional parameters $h \in \mathcal{H}^h$ and $\rho \in \mathcal{H}^\rho$ that live in the following functions spaces:

$$\mathcal{H}^h = \left\{ h \in L^2([0,1] \times \mathbb{R}, 1) \, : \, \int_0^1 \int_{-\infty}^{\infty} u^l h(u,w)^2 du dw = \frac{1}{l+1}, \; l = 0, 1, ... \right\}, \quad (2.24)$$

and

$$\mathcal{H}^\rho = \left\{ \rho \in L^2([0,1] \times \mathbb{R}, q) \, : \, \rho \in H^s([0,1] \times \mathbb{R}), \; \rho(u^*, w) = 0, \; \forall w \in \mathbb{R} \right\}, \quad (2.25)$$

where $L^2([0,1] \times \mathbb{R}, q)$ is the space of square integrable functions on $[0,1] \times \mathbb{R}$ w.r.t. weight q, H^s is the Holder space of degree $s \geq 2$ (see Chen (2007)), and $u^* \in (0,1)$ is given. The true parameter values are denoted as $\theta_0 = (\beta_1^0, \beta_2^0, h^0, \rho^0) \in \Theta$, where the parameter set is $\Theta = B_1 \times B_2 \times \mathcal{H}^h \times \mathcal{H}^\rho$, and B_1 and B_2 are compact subsets of $\mathbb{R}^p$. For identification purpose, the coefficients in the indexes corresponding to the constants are set to 1.

The model defined by (2.21) admits a stochastic nonlinear autoregressive representation. Indeed, let us define the process:

$$Z_{it} = G(U_{it}|X'_{it}\beta_1^0) \quad (2.26)$$

The variables $Z_{i,t}, Z_{i,t-1}$ have the same copula as variables $U_{i,t}, U_{i,t-1}$, conditional on $X_{i,t}, X_{i,t-1}$, with uniform marginal distributions on $[0,1]$. Thus, the stochastic representation of our model is as follows:

$$Z_{it} = \Lambda[\rho(Z_{i,t-1}, X'_{it}\beta_2^0) + \omega_{it}; X'_{it}\beta_2^0] \quad (2.27)$$

where $\omega_{it} \sim IIN(0,1)$ is independent of $X_{i,t}$, and function $\Lambda(\cdot; X'_{it}\beta_2^0)$ is given in (2.22). This corresponds to a nonlinear autoreressive dynamics for Z_{it} driven by the exogenous process X_{it}. Then, from (2.26) the uniform ranks are $U_{it} = G^{-1}(Z_{it}|X'_{it}\beta_1^0)$, where inversion is w.r.t. the first argument.

Our semi-nonparametric approach has practical advantages compared to fully nonparametric estimation of the copula (Fermanian and Scaillet (2003)). In fact, in our model the copula is parametrized by a bivariate unknown function ρ related to the serial persistency of ranks. The fully nonparametric approach involves instead a trivariate conditional copula function that may be hard to interpret and estimate accurately.

Let us now establish explicitly the constraints imposed on the model parameter θ by the orthogonality condition in Assumption 5.

Proposition 3. *In the framework of model (2.23), Assumption 5 can be written as:*

$$\int_0^1 \int_{-\infty}^{\infty} \psi_t(u,w)h(u,w)^2 dudw = 0, \ \ \forall t = 1, ..., T, \tag{2.28}$$

where $\psi_t(u,w) = F_{\varepsilon,t}^{-1}(u)E(X_{i,t} - \bar{X}_{i\cdot}|W_{1,i,t} = w) - \frac{1}{T}\sum_{s=1}^T F_{\varepsilon,s}^{-1}(u)E(X_{i,t} - \bar{X}_{i\cdot}|W_{1,i,s} = w)$.

The condition (2.28) involves the parameter vector β_1 (via variable W_{1it}), function h, and the quantile $F_{\varepsilon,t}^{-1}(\cdot)$ of the cross-sectional error distribution, but not the copula vector and functional parameters β_2 and ρ. Because distribution $F_{\varepsilon,t}(\cdot)$ is unknown and has to be estimated, imposing restriction (2.28) on the Sieve space complicates substantially the derivation of the asymptotic properties of the estimator. For this reason we opt for an unconstrained estimator in Section 3.1 and test the empirical validity of restriction (2.28) in Section 5.

2.4 A functional relative mobility measure

We now want to derive an adequate measure of positional mobility in our model. Let us write our dynamics as a function of ranks $U_{i,t}, U_{i,t-1}$. From equations (2.26) and (2.27) we have:

$$U_{it} = G^{-1}[\Lambda[\rho(G(U_{i,t-1}|X'_{i,t-1}\beta_1); X'_{it}\beta_2) + \omega_{it}; X'_{it}\beta_2] \, |X'_{it}\beta_1] \equiv a(\omega_{i,t}, U_{i,t-1}, X_{i,t}, X_{i,t}). \tag{2.29}$$

Function a is monotone increasing w.r.t. its first argument, that has a standard normal distribution. Thus, the conditional median of $U_{i,t}$ given $U_{i,t-1}, X_{it}, X_{i,t-1}$ is obtained as:

$$\begin{aligned} med(U_{it}|U_{i,t-1}; X_{it}, X_{i,t-1}) &= a(0, U_{i,t-1}, X_{i,t}, X_{i,t}) \\ &= G^{-1}[\Lambda[\rho(G(U_{i,t-1}|X'_{i,t-1}\beta_1); X'_{it}\beta_2); X'_{it}\beta_2] \, |X'_{it}\beta_1]. \end{aligned}$$

We define mobility as the partial derivative of the conditional median rank with respect to the past rank:

$$m(U_{i,t-1}; X_{i,t}, X_{i,t-1}) = \frac{\partial med(U_{it}|U_{i,t-1}, X_{it}, X_{i,t-1})}{\partial U_{i,t-1}}. \tag{2.30}$$

Hence, the mobility measure takes the following explicit form:

$$\begin{aligned} m(U_{i,t-1}; X_{i,t}, X_{i,t-1}) &= \frac{\lambda[\rho(G(U_{i,t-1}|X'_{i,t-1}\beta_1), X'_{it}\beta_2); X'_{it}\beta_2]}{g\{G^{-1}[\Lambda[\rho[G(U_{i,t-1}|X'_{i,t-1}\beta_1); X'_{it}\beta_2]; X'_{it}\beta_2] \, |X'_{it}\beta_1] \, |X'_{i,t}\beta_1\}} \\ &\quad \times \nabla_1 \rho[G(U_{i,t-1}|X'_{i,t-1}\beta_1); X'_{it}\beta_2] g(U_{i,t-1}|\beta'_1 X_{i,t-1}), \end{aligned}$$

where $\nabla_1 \rho$ denotes the partial derivative of function ρ w.r.t. its first argument.

Definition 1. *Let $\bar{u} \in (0,1)$ and $\bar{x} \in \mathbb{R}^p$ be given reference values for the percentile and the regressors. The functional $\gamma(\theta)$ is defined as the mobility measure evaluated for $U_{i,t-1} = \bar{u}$ and $X_{i,t} = X_{i,t-1} = \bar{x}$, i.e:*

$$\gamma(\theta) = \frac{\lambda[\rho(G(\bar{u}|\bar{x}'\beta_1), \bar{x}'\beta_2); \bar{x}'\beta_2]\nabla_1\rho[G(\bar{u}|\bar{x}'\beta_1); \bar{x}'\beta_2]g(\bar{u}|\beta_1'\bar{x})}{g\{G^{-1}[\Lambda[\rho[G(\bar{u}|\bar{x}'\beta_1); \bar{x}'\beta_2]; \bar{x}'\beta_2] \; |\bar{x}'\beta_1]|\bar{x}'\beta_1\}}, \tag{2.31}$$

where $g(u|\bar{x}'\beta_1) = h(u, \bar{x}'\beta_1)^2 / \int_0^1 h(u, \bar{x}'\beta_1)^2 du$ and $G(u|\bar{x}'\beta_1) = \int_0^u h(y, \bar{x}'\beta_1)^2 dy / \int_0^1 h(y, \bar{x}'\beta_1)^2 dy$.

This functional measure accounts for mobility in the different parts of the distribution of past ranks via argument $\bar{u}$ and for the effects of covariates. It depends on both functions h, ρ and vectors β_1, β_2. In fact, it involves the gradient $\nabla_1\rho(\cdot, \bar{x}'\beta_2)$ of the autoregressive function, and accounts also for the transformation function $\Lambda(\cdot, \bar{x}'\beta_2)$ in the rank dynamics and its derivative $\lambda(\cdot, \bar{x}'\beta_2)$, as well as the rank conditional distribution $g(\cdot|\bar{x}'\beta_1)$. Note that the expression in (2.31) 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 median rank and its partial derivative in (2.30), from the stochastic representation of the model we can also easily derive the other conditional quantiles. They 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 $Q_{U,t}(\tau|U_{i,t-1}, X_{it}, X_{i,t-1})$ of U_{it} for percentile $\tau \in (0,1)$ is given by

$$\begin{aligned} Q_{U,t}(\tau|U_{i,t-1}, X_{it}, X_{i,t-1}) &= a(\Phi^{-1}(\tau), U_{i,t-1}, X_{it}, X_{i,t-1}) \\ &= G^{-1}[\Lambda[\rho(G(U_{i,t-1}|X_{i,t-1}'\beta_1); X_{it}'\beta_2) + \Phi^{-1}(\tau); X_{it}'\beta_2] \; |X_{it}'\beta_1]. \end{aligned} \tag{2.32}$$

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

$$\begin{aligned} m^Q(U_{i,t-1}, \tau; X_{i,t}, X_{i,t-1}) &= \frac{\partial Q_{U,t}(\tau|U_{i,t-1}, X_{it}, X_{i,t-1})}{\partial U_{i,t-1}} \\ &= \frac{\lambda[\rho(G(U_{i,t-1}|X_{i,t-1}'\beta_1) + \Phi^{-1}(\tau), X_{it}'\beta_2); X_{it}'\beta_2]}{g\{G^{-1}[\Lambda[\rho[G(U_{i,t-1}|X_{i,t-1}'\beta_1); X_{it}'\beta_2] + \Phi^{-1}(\tau); X_{it}'\beta_2] \; |X_{it}'\beta_1] \; |X_{i,t}'\beta_1\}} \\ &\quad \times \nabla_1\rho[G(U_{i,t-1}|X_{i,t-1}'\beta_1); X_{it}'\beta_2]g(U_{i,t-1}|\beta_1'X_{i,t-1}). \end{aligned}$$

In Section 5, we estimate such immobility measures on a dataset of US workers. [10]

[10]We could also consider relative mobility measures based on the partial derivatives of the conditional expected ranks.

2.5 Link with the literature on individual earnings dynamics

In this section we relate our model 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 uniform ranks $U_{i,t}$. 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 rank 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

However, they involve numerical integration and their analysis is more complex. Moreover, while the analysis in this section focuses on the mobility of the residual component ε_{it}, in the Supplementary Materials we complete our study of functional relative mobility measures by analyzing overall wage mobility.

between deciles of the cross-sectional wage distribution, see e.g. Shorrocks (1978) [11].

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 Sieve semi-nonparametric estimation of the mobility model

In this section we estimate the semi-nonparametric copula model of Section 2. The estimation procedure is defined in two steps. In a first step we estimate the unobservable values of the ranks $U_{i,t}$ by means of the empirical ranks. More specifically, the estimated ranks are:

$$\hat{U}_{i,t} = \frac{1}{N-1} \sum_{j \neq i} \mathbb{I}(\hat{\varepsilon}_{j,t} \leq \hat{\varepsilon}_{i,t}), \tag{3.1}$$

where $\hat{\varepsilon}_{i,t} = y_{i,t} - \hat{\alpha}' X_{i,t} - \hat{\eta}_i - \hat{\lambda}_t$ is the residual in the wage equation (2.1) obtained from the fixed-effects least squares estimator. In (3.1) we use a leave-one-out procedure for technical reasons. The fixed effects estimator $\hat{\alpha}$ is consistent under Assumption 5 (see Lemma 1 a)). Then, in a second step, we estimate the parameters of the marginal distribution of the ranks and those of the copula function using the estimated ranks. The alternative procedure consisting in estimating jointly the wage equation with individual fixed effects and the nonlinear rank dynamics is computationally challenging and is not

[11] 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.

considered further in this paper. We estimate the rank distribution $g(\cdot|X_{i,t}'\beta_1) = \frac{h(\cdot,X_{i,t}'\beta_1)^2}{\int_0^1 h(u,X_{i,t}'\beta_1)^2du}$ and the copula pdf $c(\cdot,\cdot,\rho(\cdot,X_{i,t}'\beta_2))$ conditional on regressors via a simultaneous M-estimation of the parameter vectors β_1, β_2 and the functions $h(\cdot,\cdot)$ and $\rho(\cdot,\cdot)$ using 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 the unknown functions in the infinite-dimensional component of the parameter set by means of approximating spaces generated by a set of basis functions, whose dimensions grow with sample size. In our case, the estimation is performed via the following Sieve Maximum Likelihood procedure applied on the empirical ranks. The estimator of θ is:

$$\hat{\theta} = \arg\max_{\theta\in\Theta_{N,T}} \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\log l(\hat{U}_{i,t}|\hat{U}_{i,t-1}, X_{i,t}, X_{i,t-1};\theta), \tag{3.2}$$

where the log conditional density $\log l$ is defined in (2.23), the set $\Theta_{N,T} = B_1 \times B_2 \times \mathcal{H}_{N,T}^h \times \mathcal{H}_{N,T}^\rho$ is the Cartesian product of the compact sets B_1 and B_2 for the Euclidean parameter vectors and Sieve spaces $\mathcal{H}_{N,T}^h$ and $\mathcal{H}_{N,T}^\rho$ for approximating functions in $\mathcal{H}^h$ and $\mathcal{H}^\rho$. We use here tensor products of orthogonal series:

$$\mathcal{H}_{N,T}^h = \left\{ h(\cdot,\cdot) \ : \ h(u,w) = \sum_{j,k=0}^{m^h} \lambda_{j,k}\varphi_j(u)\psi_k(w), \ \sum_{j,k,r=0}^{m^h} \lambda_{j,k}\lambda_{r,k}\kappa_{j,r}^{(l)} = \frac{1}{l+1}, \ l = 0,1,...,m^h \right\},$$

for $\kappa_{j,r}^{(l)} = \int_0^1 u^l \varphi_j(u)\varphi_r(u)du$, and

$$\mathcal{H}_{N,T}^\rho = \left\{ \rho(\cdot,\cdot) \ : \ \rho(u,w) = \sum_{j,k=0}^{m^\rho} \mu_{j,k}\varphi_j(u)\tilde{\psi}_k(w), \ \sum_{j=0}^{m^\rho} \mu_{j,k}\varphi_j(u^*) = 0, \ k = 0,1,...,m^\rho \right\},$$

where $\{\varphi_j\}$, $\{\psi_k\}$, $\{\tilde{\psi}_k\}$ are complete orthonormal bases of $L^2[0,1]$, $L^2(\mathbb{R})$ and $L^2(\mathbb{R},q)$, respectively, $\lambda_{j,k}$ and $\mu_{j,k}$ are real coefficients, and integers $m^h = m_{N,T}^h$, $m^\rho = m_{N,T}^\rho$ grow with N, T (Chen (2007)). The linear and quadratic constraints on coefficients $\mu_{j,k}$ and $\lambda_{j,k}$ implement the restrictions on functions ρ and h in (2.24) and (2.25). In the Monte Carlo study and the empirical application, we use $\varphi_j(u) = \frac{1}{\sqrt{2^j j!}}H_j[\frac{\Phi^{-1}(u)}{\sqrt{2}}]$ and $\psi_j(w) = \frac{1}{\sqrt{2^j j!\sqrt{\pi}}}e^{-w^2/2}H_j(w)$ and $\tilde{\psi}_j(w) = \frac{1}{\sqrt{2^j j!\sqrt{\pi}}}H_j(w)$, where the $H_j(\cdot)$ are the Hermite polynomials, and $q(w) = e^{-w^2}$ (see Appendix C for details on implementation of the Sieve estimator). The estimator in (3.2) optimizes the log-likelihood function jointly with respect to the parameters in the copula specification and the rank marginal distribution (conditional on the regressors). An alternative approach not explored in this paper consists in considering a two-step Sieve estimator for (β_1, h) and (β_2, ρ), see e.g. Hahn, Liao and Ridder (2018).

Once the Sieve estimator $\hat{\theta}$ for the parameter of the rank dynamics is obtained from (3.2), we estimate the functional mobility measures of Section 2.4 by plug-in. The estimated functional based on the derivative of the median conditional rank (2.31) is given by:

$$\hat{\gamma} = \gamma(\hat{\theta}) = \frac{\hat{\lambda}[\hat{\rho}(\hat{G}(\bar{u}|\bar{x}'\hat{\beta}_1), \bar{x}'\hat{\beta}_2); \bar{x}'\hat{\beta}_2]\nabla_1\hat{\rho}[\hat{G}(\bar{u}|\bar{x}'\hat{\beta}_1); \bar{x}'\hat{\beta}_2]\hat{g}(\bar{u}|\hat{\beta}_1'\bar{x})}{\hat{g}\{\hat{G}^{-1}[\hat{\Lambda}[\hat{\rho}[\hat{G}(\bar{u}; \bar{x}'\hat{\beta}_1); \bar{x}'\hat{\beta}_2]; \bar{x}'\hat{\beta}_2] \, |\bar{x}'\hat{\beta}_1] \, |\bar{x}'\hat{\beta}_1\}}, \tag{3.3}$$

where $\hat{g}(u|\bar{x}'\hat{\beta}_1) = \hat{h}(u, \bar{x}'\hat{\beta}_1)^2 \Big/ \int_0^1 \hat{h}(u, \bar{x}'\hat{\beta}_1)^2 du$ and $\hat{\lambda}(y; \bar{x}'\hat{\beta}_2) = \int_0^1 \phi[y - \hat{\rho}(v; \bar{x}'\hat{\beta}_2)]dv$.
Note that the constraints implied by Proposition 3 are not imposed on estimator $\hat{\theta}$. Indeed, this would yield a Sieve space depending on estimated (unknown) quantities, namely $\hat{F}_{\varepsilon,t}$ and conditional expectations, making the derivations of the asymptotic distribution considerably more involved. Instead, the constraint (2.28) is empirically tested afterwards. The results of such test are reported at the end of Section 5 on the empirical results.

3.1 Asymptotic distribution

We establish the asymptotic normality of estimator $\hat{\gamma}$ in 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 $U_{i,t}$ with the empirical rank $\hat{U}_{i,t}$. However, the estimation error induced by the individual effects when computing the residuals $\hat{\varepsilon}_{i,t}$ yields a bias term in the asymptotic distribution of estimator $\hat{\gamma}$. We derive an explicit expression for this bias in Theorem 1 below.
In order to state our result, we introduce the following objects. The "tangent" space $\mathbb{V}$ is the linear space $\mathbb{V} = \mathbb{R}^p \times \mathbb{R}^p \times \mathbb{V}^h \times \mathbb{V}^\rho$, where $\mathbb{V}^h = \Big\{h \in L^2([0,1] \times \mathbb{R}) \, : \, \int_0^1 \int_{-\infty}^{\infty} u^l h(u,w) h^0(u,w) du dw = 0, \; l = 0, 1, ...\Big\}$ and $\mathbb{V}^\rho = \mathcal{H}^\rho$. It corresponds to the linear space of infinitesimal deviations from θ_0 belonging to Θ. This linear space is equipped with the scalar product

$$\langle v, v^* \rangle = v_{\beta_1}' v_{\beta_1}^* + v_{\beta_2}' v_{\beta_2}^* + \int_0^1 \int_{-\infty}^{\infty} v_h(u,w) v_h^*(u,w) du dw + \int_0^1 \int_{-\infty}^{\infty} v_\rho(u,w) v_\rho^*(u,w) q(w) du dw$$

for $v = (v_{\beta_1}, v_{\beta_2}, v_h, v_\rho) \in \mathbb{V}$ and $v^* = (v_{\beta_1}^*, v_{\beta_2}^*, v_h^*, v_\rho^*) \in \mathbb{V}$, which combines the Euclidean scalar product for the finite-dimensional part with the (weighted) L^2 scalar products for the infinite-dimensional components. The associated norm is $\|v\| = \sqrt{\langle v, v \rangle}$. For a real smooth functional $f(\theta)$, the directional derivative of f at $\theta = \theta_0$ is defined by $\frac{\partial f(\theta_0)}{\partial \theta}[v] = \lim_{\tau \to 0} \frac{f(\theta_0 + \tau v) - f(\theta_0)}{\tau}$ for $v \in \mathbb{V}$ (if the limit exists). The information operator $I_0 : \mathbb{V} \to \mathbb{V}$ is the self-adjoint operator defined by

$$\langle v, I_0 w\rangle = E\left[\frac{\partial \log l(U_{i,t}|U_{i,t-1}, X_{i,t}, X_{i,t-1};\theta_0)}{\partial\theta}[v]\frac{\partial \log l(U_{i,t}|U_{i,t-1}, X_{i,t}, X_{i,t-1};\theta_0)}{\partial\theta}[w]\right], \quad (3.4)$$

for $v, w \in \mathbb{V}$. Operator I_0 plays the role of the information matrix in our functional framework. It is assumed bounded, invertible with bounded inverse (see Assumption A.5 in Appendix A), which implies local identification of parameter θ_0 and the well-posedness of the estimation problem. The best approximation of θ_0 using the Sieve $\Theta_{N,T}$ is $\theta_{0,NT} = \arg\min_{\theta\in\Theta_{N,T}} \|\theta - \theta_0\|$. Let $\mathbb{V}_{N,T} = \mathbb{R}^p \times \mathbb{R}^p \times \mathbb{V}^h_{N,T} \times \mathbb{V}^\rho_{N,T}$, where the tangent spaces for the Sieves are the finite-dimensional linear spaces $\mathbb{V}^h_{N,T} = \left\{h(\cdot,\cdot) : h(u,w) = \sum_{j,k=0}^{m^h} \lambda_{j,k}\varphi_j(u)\psi_k(w),\ \sum_{j,k=0}^{m^h} \lambda_{j,k} b^{(l)}_{j,k} = 0,\ l = 0,1,...,m^h\right\}$, where $b^{(l)}_{j,k} = \int_0^1\int_{-\infty}^{\infty} u^l\varphi_j(u)\psi_k(w)h_{0,NT}(u,w)dudw$, and $\mathbb{V}^\rho_{N,T} = \mathcal{H}^\rho_{N,T}$. Finally, let $v_{N,T} \in \mathbb{V}_{N,T}$ be the Rietz representer of the directional derivative of functional $\gamma(\cdot)$ on the finite-dimensional linear space $\mathbb{V}_{N,T}$, i.e.

$$\frac{\partial\gamma(\theta_0)}{\partial\theta}[v] = \langle v_{N,T}, v\rangle, \ \ \forall v \in \mathbb{V}_{N,T}. \quad (3.5)$$

Vector $v_{N,T}$ is the counterpart of the gradient of the object of interest w.r.t. the parameter in the likelihood function. For a regular, resp. irregular, functional we have $\|v_{NT}\| = O(1)$, resp. $\|v_{NT}\| \to \infty$, as $N, T \to \infty$ (Chen, Liao and Sun (2014), Chen and Liao (2014)).

Theorem 1. *Let the bias function $\mathcal{B}_T(v)$ be defined by:*

$$\begin{aligned}\mathcal{B}_T(v) \ &= \ \frac{1}{T}\sum_t E\left[\frac{\partial^2 \log l_{i,t}(\theta_0)}{\partial\theta\partial u}[v]H_t(\varepsilon_{i,t}) + \frac{\partial^2 \log l_{i,t}(\theta_0)}{\partial\theta\partial v}[v]H_{t-1}(\varepsilon_{i,t-1})\right] \\ &\quad -\frac{1}{T}\sum_{t,s} E\left[\left(\frac{\partial^2 \log l_{i,t}(\theta_0)}{\partial\theta\partial u}[v]f_t(\varepsilon_{i,t}) + \frac{\partial^2 \log l_{i,t}(\theta_0)}{\partial\theta\partial v}[v]f_{t-1}(\varepsilon_{i,t-1})\right)\varepsilon_{i,s}\right] \\ &\quad +\frac{\omega^2}{2T}\sum_t E\left[\Psi(\varepsilon_{i,t},\varepsilon_{i,t-1};\theta_0)[v]\right], \end{aligned} \quad (3.6)$$

for $v \in \mathbb{V}$, with $H_t(\varepsilon) = \sum_{s=1}^T E(\varepsilon_{i,s}|\varepsilon_{i,t} = \varepsilon)f_t(\varepsilon) + \frac{\omega^2}{2}f_t'(\varepsilon)$, $\omega^2 = \sum_{j=-\infty}^{\infty} \bar{E}[\varepsilon_{i,t}\varepsilon_{i,t-j}]$, where $\bar{E}(Z_t) = \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^T E(Z_t)$ denotes the long-run expectation, $f_t = F_t'$ is the pdf of the error term and $\sigma_t^2 = E(\varepsilon_{i,t}^2)$ its variance, and

$$\begin{aligned}\Psi(\varepsilon_{i,t},\varepsilon_{i,t-1};\theta_0)[v] \ &= \ \frac{\partial^2 \log l_{i,t}(\theta_0)}{\partial\theta\partial u}[v]f_t'(\varepsilon_{i,t}) + \frac{\partial^2 \log l_{i,t}(\theta_0)}{\partial\theta\partial v}[v]f_{t-1}'(\varepsilon_{i,t-1}) \\ &\quad +\frac{\partial^3 \log l_{i,t}(\theta_0)}{\partial\theta\partial u^2}[v][f_t(\varepsilon_{i,t})]^2 + \frac{\partial^3 \log l_{i,t}(\theta_0)}{\partial\theta\partial v^2}[v][f_{t-1}(\varepsilon_{i,t-1})]^2 \\ &\quad +2\frac{\partial^3 \log l_{i,t}(\theta_0)}{\partial\theta\partial u\partial v}[v]f_t(\varepsilon_{i,t})f_{t-1}(\varepsilon_{i,t-1}),\end{aligned}$$

where $l_{i,t}(\theta) = l(U_{i,t}|U_{i,t-1}, X_{i,t}, X_{i,t-1}; \theta)$. Further, let the long-run covariance operator $\Omega_0 : \mathbb{V} \to \mathbb{V}$ be defined by

$$\langle v, \Omega_0 w \rangle = \sum_{j=-\infty}^{\infty} E\left[\zeta_{i,t}[v]\zeta_{i,t-j}[w]\right], \quad v, w \in \mathbb{V}, \tag{3.7}$$

where:

$$\begin{aligned} \zeta_{i,t}[v] \quad = \quad & \frac{\partial \log l(U_{i,t}|U_{i,t-1}, X_{it}, X_{i,t-1}; \theta_0)}{\partial\theta}[v] \\ & + E\left[\frac{\partial^2 \log l_{j,t}(\theta_0)}{\partial\theta\partial u}[v]\{\mathbb{I}(U_{i,t} \le U_{j,t}) - U_{j,t}\}|U_{i,t}\right] \\ & + E\left[\frac{\partial^2 \log l_{j,t}(\theta_0)}{\partial\theta\partial v}[v]\{\mathbb{I}(U_{i,t-1} \le U_{j,t-1}) - U_{j,t-1}\}|U_{i,t-1}\right] \\ & - 2\varepsilon_{i,t}\bar{E}\left[\frac{\partial^2 \log l_{i,t}(\theta_0)}{\partial\theta\partial u}[v]f_t(\varepsilon_{i,t}) + \frac{\partial^2 \log l_{i,t}(\theta_0)}{\partial\theta\partial v}[v]f_{t-1}(\varepsilon_{i,t-1})\right], \end{aligned}$$

where $(U_{j,t}, U_{j,t-1}, X_{j,t}, X_{j,t-1})$ is an independent copy of $(U_{i,t}, U_{i,t-1}, X_{i,t}, X_{i,t-1})$. Then, as

$$N, T \to \infty, \ \textit{such that } T = o(N), \ N = o(T^3), \tag{3.8}$$

and under Assumptions A.1-A.10 in Appendix A, we have:

$$\sigma_{N,T}^{-1}\sqrt{NT}\left(\gamma(\hat{\theta}) - \gamma(\theta_0) - \frac{\partial\gamma(\theta_0)}{\partial\theta}[\theta_{0,NT} - \theta_0] - \frac{1}{T}\mathcal{B}_T(I_0^{-1}v_{N,T})\right) \ \Rightarrow \ N(0,1),$$

where $\sigma_{N,T}^2 = \langle v_{N,T}, I_0^{-1}\Omega_0 I_0^{-1} v_{N,T}\rangle$.

Under the condition:

$$\sigma_{N,T}^{-1}\sqrt{NT}\frac{\partial\gamma(\theta_0)}{\partial\theta}[\theta_{0,NT} - \theta_0] = o(1), \tag{3.9}$$

the bias from the Sieve approximation is asymptotically negligible, and we get:

$$\sigma_{N,T}^{-1}\sqrt{NT}\left(\gamma(\hat{\theta}) - \gamma(\theta_0) - \frac{1}{T}\mathcal{B}_T(I_0^{-1}v_{N,T})\right) \ \Rightarrow \ N(0,1).$$

The estimator $\gamma(\hat{\theta})$ needs a recentering by means of term $\frac{1}{T}\mathcal{B}_T(I_0^{-1}v_{N,T})$ at order $O(\frac{\|v_{N,T}\|}{T})$ to correct for the asymptotic bias induced by the individual fixed effects. Moreover, the convergence rate $\sqrt{NT/\sigma_{N,T}^2}$ is slower than the standard panel rate $\sqrt{NT}$ when $\sigma_{N,T} \to \infty$ for an irregular functional. Under the double asymptotics in (3.8) with N growing faster than T but less fast than T^3, some additional bias terms are asymptotically negligible.

The asymptotic variance and bias terms $\sigma_{NT}^2 = \langle \tilde{v}_{NT}, \Omega_0 \tilde{v}_{NT}\rangle$ and $\mathcal{B}_T(\tilde{v}_{NT})$ both involve the function $\tilde{v}_{NT} = I_0^{-1}v_{NT}$. In order to characterize explicitly these terms, it is useful to introduce matrix representations for linear operators. Being the linear spaces $\mathbb{V}_{NT}^h$ and $\mathbb{V}_{NT}^\rho$ finite-dimensional, by incorporating the linear constraints on the coefficients λ and μ we can write those spaces as the linear span of orthonormal basis functions Ψ_l^h, $l = 1, ..., M^h$ and Ψ_l^ρ, $l = 1, ..., M^\rho$, respectively, for $M^h = (m^h + 1)m^h$ and $M^\rho = (m^\rho + 1)m^\rho$ (see Appendix C for details). By combining these basis functions together with the standard Euclidean unit vectors in $\mathbb{R}^p$, we get a basis e_l, $l = 1, ..., M$, of linear space $\mathbb{V}_{N,T}$, with $M = 2p + M^h + M^\rho$, that is orthonormal w.r.t. the scalar product $\langle \cdot, \cdot \rangle$. Thus, the Rietz representer of $v_{NT} \in \mathbb{V}_{NT}$ can be written as $v_{NT} = e'\nu_{NT}$ where the elements of vector $\nu_{NT} \in \mathbb{R}^M$ are:

$$\nu_{NT,l} = \frac{\partial \gamma(\theta_0)}{\partial \theta}[e_l], \qquad l = 1, ..., M. \tag{3.10}$$

Moreover, we can approximate function $\tilde{v}_{NT}$ by its projection on $\mathbb{V}_{NT}$, i.e. $e'\tilde{\nu}_{NT}$, where vector $\tilde{\nu}_{NT}$ is given by $\tilde{\nu}_{NT} = I_{0,NT}^{-1}\nu_{NT}$ and the $M \times M$ matrix $I_{0,NT}$ has elements $\langle e_l, I_0 e_k \rangle$. Under the assumptions in Appendix A, this approximation has no effect on the asymptotic distribution. Moreover, let $\Omega_{0,NT}$ be the $M \times M$ matrix with elements $\langle e_k, \Omega_0 e_l \rangle$, and let b_{NT} be the $M \times 1$ vector with elements $\mathcal{B}_T(e_l)$. Then, the variance and bias terms

$$\sigma_{NT}^2 = \nu_{NT}' I_{0,NT}^{-1} \Omega_{0,NT} I_{0,NT}^{-1} \nu_{NT}, \qquad \mathcal{B}(\tilde{v}_{NT}) = b_{NT}' I_{0,NT}^{-1} \nu_{nT}, \tag{3.11}$$

are written in terms of the vectors ν_{NT}, b_{NT}, and the matrices $I_{0,NT}$ and $\Omega_{0,NT}$ representing the information and long run variance operators.

3.2 Results for the nonlinear autoregressive copula and mobility measure

Let us now specialize the results in Theorem 1 for the semi-nonparametric nonlinear autoregressive copula family considered in Section 2.3 and the relative mobility measure introduced in Definition 1. The directional derivatives of the log-likelihood function and the mobility measure are provided in the next theorem.

Theorem 2. *a) Let $u, u_{-1} \in [0, 1]$ be generic values of variables $U_{i,t}, U_{i,t-1}$, and let $x, x_{-1} \in \mathbb{R}^p$ be* *generic values of $X_{i,t}, X_{i,t-1}$. The directional derivative of the log conditional density w.r.t. θ is:*

$$
\begin{aligned}
&\frac{\partial \log l(u|u_{-1},x,x_{-1};\theta_0)}{\partial\theta}[v] \\
= \;& \frac{1}{\lambda_0[\Lambda_0^{-1}(z;w_2);w_2]}\left(\rho_0(z_{-1};w_2) - E[\rho_0(Z_{i,t-1};w_2)|Z_{i,t}=z, X_{i,t}=x, X_{i,t-1}=x_{-1}]\right) \\
&\qquad \times 2Cov\left(\mathbf{1}(U_{i,t}\leq u), \frac{v_h(U_{i,t},w_1)+\nabla_2 h^0(U_{i,t},w_1)x'v_{\beta_1}}{h^0(U_{i,t},w_1)}|X_{i,t}=x\right) \\
&+\left[\Lambda_0^{-1}(z;w_2)-\rho_0(z_{-1};w_2)\right]\nabla_1\rho_0(z_{-1};w_2) \\
&\qquad \times 2Cov\left(\mathbf{1}(U_{i,t-1}\leq u_{-1}), \frac{v_h(U_{i,t-1},w_{-1})+\nabla_2 h^0(U_{i,t-1},w_{-1})x_{-1}'v_{\beta_1}}{h^0(U_{i,t-1},w_{-1})}|X_{i,t-1}=x_{-1}\right) \\
&+\left[\Lambda_0^{-1}(z;w_2)-\rho_0(z_{-1};w_2)\right]\left(v_\rho(z_{-1},w_2)+\nabla_2\rho_0(z_{-1},w_2)x'v_{\beta_2}\right. \\
&\qquad \left. -E[v_\rho(Z_{i,t-1},w_2)+\nabla_2\rho_0(Z_{i,t-1},w_2)x'v_{\beta_2}|Z_{i,t}=z,X_{i,t}=x,X_{i,t-1}=x_{-1}]\right) \\
&+Cov\left(\rho_0(Z_{i,t-1};w_2), v_\rho(Z_{i,t-1},w_2)+\nabla_2\rho_0(Z_{i,t-1},w_2)x'v_{\beta_2}|Z_{i,t}=z,X_{i,t}=x,X_{i,t-1}=x_{-1}\right) \\
&+2\left(\frac{v_h(u,w_1)+\nabla_2 h^0(u,w_1)x'v_{\beta_1}}{h^0(u,w_1)} - E\left[\frac{v_h(U_{i,t},w_1)+\nabla_2 h^0(U_{i,t},w_1)x'v_{\beta_1}}{h^0(U_{i,t},w_1)}|X_{i,t}=x\right]\right),
\end{aligned}
$$

for $v=(v_{\beta_1},v_{\beta_2},v_h,v_\rho)\in\mathbb{V}$, where $z=G_0(u;w_1)$, $z_{-1}=G_0(u_{-1};w_{-1})$, and $w_1=x'\beta_1^0$, $w_2=x'\beta_2^0$, $w_{-1}=x_{-1}'\beta_1^0$.

b) The directional derivative of the mobility functional is:

$$
\begin{aligned}
\frac{\partial\gamma(\theta_0)}{\partial\theta}[v] = \gamma(\theta_0)&\left\{2\left(\frac{v_h(\bar u,\bar w_1)+\nabla_2 h^0(\bar u,\bar w_1)\bar x'v_{\beta_1}}{h^0(\bar u,\bar w_1)} - \frac{v_h(\zeta_1,\bar w_1)+\nabla_2 h^0(\zeta_1,\bar w_1)\bar x'v_{\beta_1}}{h^0(\zeta_1,\bar w_1)}\right)\right. \\
&+4\frac{\nabla_1 h^0(\zeta_1,\bar w_1)}{h^0(\zeta_1,\bar w_1)}\frac{1}{g_0(\zeta_1;\bar w_1)}Cov\left(\mathbf{1}(U_{i,t}\leq\zeta_1), \frac{v_h(U_{i,t},\bar w_1)+\nabla_2 h^0(U_{i,t},\bar w_1)\bar x'v_{\beta_1}}{h^0(U_{i,t},\bar w_1)}|X_{i,t}=\bar x\right) \\
&+2\left[E\left[\rho_0(Z_{i,t-1},\bar w_2)-\rho_0(\zeta_4,\bar w_2)|Z_{i,t}=\zeta_2,X_{i,t}=\bar x,X_{i,t-1}=\bar x\right] - 2\frac{\nabla_1 h^0(\zeta_1,\bar w_1)}{h^0(\zeta_1,\bar w_1)}\frac{\lambda_0(\zeta_3;\bar w_2)}{g_0(\zeta_1;\bar w_1)}\right. \\
&\qquad\left.+\frac{\nabla_1^2\rho_0(\zeta_4,\bar w_2)}{[\nabla_1\rho_0(\zeta_4,\bar w_2)]^2}\right]\nabla_1\rho_0(\zeta_4,\bar w_2)Cov\left(\mathbf{1}(U_{i,t}\leq\bar u), \frac{v_h(U_{i,t},\bar w_1)+\nabla_2 h^0(U_{i,t},\bar w_1)\bar x'v_{\beta_1}}{h^0(U_{i,t},\bar w_1)}|X_{i,t}=\bar x\right) \\
&+\left(E\left[\rho_0(Z_{i,t-1};\bar w_2)-\rho_0(\zeta_4,\bar w_2)|Z_{i,t}=\zeta_2,X_{i,t}=\bar x,X_{i,t-1}=\bar x\right] - 2\frac{\nabla_1 h^0(\zeta_1,\bar w_1)}{h^0(\zeta_1,\bar w_1)}\frac{\lambda_0(\zeta_3;\bar w_2)}{g_0(\zeta_1;\bar w_1)}\right) \\
\times&\left(v_\rho(\zeta_4,\bar w_2)+\nabla_2\rho_0(\zeta_4,\bar w_2)\bar x'v_{\beta_2} - E\left[v_\rho(Z_{i,t-1},\bar w_2)+\nabla_2\rho_0(Z_{i,t-1},\bar w_2)\bar x'v_{\beta_2}|Z_{i,t}=\zeta_2,X_{i,t}=\bar x,X_{i,t-1}=\bar x\right]\right) \\
&-Cov\left(\rho_0(Z_{i,t-1},\bar w_2)-\rho_0(\zeta_4,\bar w_2), v_\rho(Z_{i,t-1};\bar w_2)+\nabla_2\rho_0(Z_{i,t-1},\bar w_2)\bar x'v_{\beta_2}|Z_{i,t}=\zeta_2,X_{i,t}=\bar x,X_{i,t-1}=\bar x\right) \\
&\left.+\frac{\nabla_1 v_\rho(\zeta_4,\bar w_2)+\nabla_{1,2}^2\rho_0(\zeta_4,\bar w_2)\bar x'v_{\beta_2}}{\nabla_1\rho_0(\zeta_4,\bar w_2)}\right\},
\end{aligned}
$$

for $v=(v_{\beta_1},v_{\beta_2},v_h,v_\rho)\in\mathbb{V}$, where $\zeta_4=G_0(\bar u;\bar w_1)$, $\zeta_3=\rho_0(\zeta_4;\bar w_2)$, $\zeta_2=\Lambda_0(\zeta_3;\bar w_2)$ and $\zeta_1=G_0^{-1}(\zeta_2;\bar w_1)=med(U_{i,t}|U_{i,t-1}=\bar u,X_{i,t}=\bar x,X_{i,t-1}=\bar x)$ is the conditional median rank, and $\bar w_1=\bar x'\beta_1^0$, $\bar w_2=\bar x'\beta_2^0$.

In Theorem 2 the directional derivatives depend linearly on the "infinitesimal change" vectors $v_{\beta_1}, v_{\beta_2} \in \mathbb{R}^p$ for the Euclidean parameters as well as on the "infinitesimal change" functions $v_h \in \mathbb{V}^h$ and $v_\rho \in \mathbb{V}^\rho$ for the functional parameters. The latter dependence is via evaluation, differentiation and integration (i.e. conditional expectations and covariance) operators. By applying formulas (3.4), (3.6), (3.7) and (3.10) on the basis functions of $\mathbb{V}$, we get vectors ν_{NT}, b_{NT} and matrices $I_{0,NT}$, $\Omega_{0,NT}$. Then, by the equations in (3.11) we get the variance and bias terms.

3.3 Analytical correction for the fixed-effects bias

We use Theorems 1 and 2 to obtain an analytical bias correction for the bias induced by the individual fixed effects. The bias corrected estimator of the relative mobility measure is:

$$\hat{\gamma}_B = \hat{\gamma} - \frac{1}{T}\hat{b}'_{NT}\hat{I}^{-1}_{NT}\hat{\nu}_{NT}. \tag{3.12}$$

The estimators $\hat{I}_{NT}$, $\hat{b}_{NT}$, $\hat{\nu}_{NT}$ of matrix $I_{0,NT}$ and vectors b_{NT}, ν_{NT} are obtained from formulas (3.4), (3.6) and (3.10) and the directional derivatives in Theorem 2 by replacing (i) errors $\varepsilon_{i,t}$ and ranks $U_{i,t}$ with their empirical analogues $\hat{\varepsilon}_{i,t}$ and $\hat{U}_{i,t}$, (ii) the true parameter $\theta_0 = (\beta_1^0, \beta_2^0, h^0, \rho^0)$ with the Sieve estimate $\hat{\theta}$ (including the conditional density of $Z_{i,t-1}$ given $Z_{i,t}$, $X_{i,t}$, $X_{i,t-1}$, and the conditional density of $U_{i,t}$ given $X_{i,t}$), and (iii) population expectations with sample averages. The bias adjustment involves numerical integration to evaluate e.g. the estimates of functions Λ, λ, and the conditional expectations of functions of $Z_{i,t-1}$ given $Z_{i,t}$, $X_{i,t}$, $X_{i,t-1}$. This is performed by either quadrature or Monte Carlo simulation. We get closed-form expressions when the Sieve estimate of function ρ used to compute the bias adjustment is restricted to correspond to a Gaussian copula. This Gaussian reference model to approximate the bias adjustment is akin to the one used to obtain the celebrated Silverman rule for the optimal bandwidth in kernel smoothing. Moreover, in order to avoid the cumbersome formulas of the higher-order derivatives of the log density w.r.t. copula arguments involved in function $\mathcal{B}_T[v]$ (see Theorem 1), we use Legendre polynomials approximations of the first-order derivatives and compute high-order derivatives from those approximations. We provide details on the numerical implementation of the bias adjustment in Appendix C.

With unbalanced panels and data missing-at-random, the bias adjustment in equation (3.12) becomes $\frac{1}{\bar{T}}\hat{b}'_{NT}\hat{I}^{-1}_{NT}\hat{\nu}_{NT}$ where $\frac{1}{\bar{T}} = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{T_i}$, i.e. $\bar{T}$ is the harmonic average of the individual time-series lengths T_i, and estimators $\hat{b}_{NT}$, $\hat{I}_{NT}$, $\hat{\nu}_{NT}$ are obtained by taking averages over the available sample.

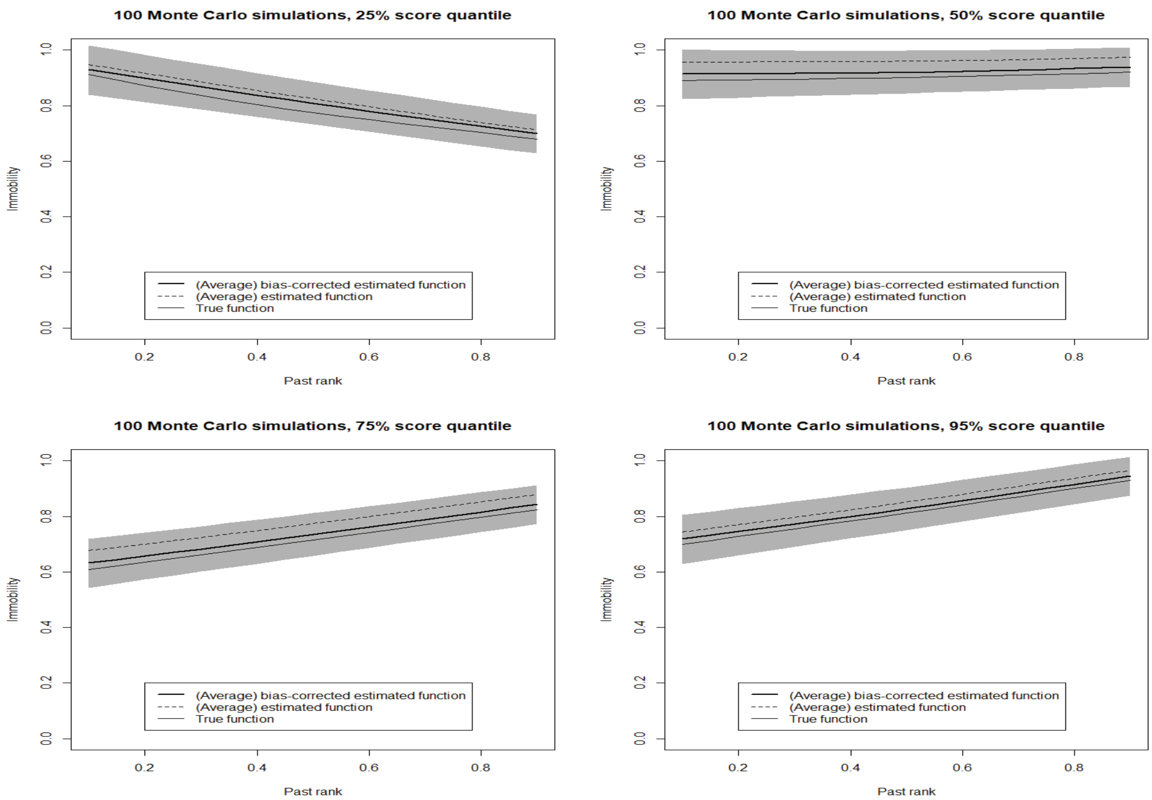
4 Monte Carlo simulations

In order to assess the finite sample properties and the numerical feasibility of our estimator, we perform a Monte Carlo analysis to compare the estimated mobility measures (both with and without the correction for the fixed-effect bias) 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 $n = 8429$ individuals with an average permanence in the sample of about 15 periods (years). Then, we randomly draw $U_{i,0}$, i.e. the initial rank, from the uniform distribution. By applying the semi-nonparametric model based on the nonlinear autoregressive copula from equations (2.20) and (2.21) with the estimated parameters obtained in the next section, starting from $U_{i,0}$ we simulate a trajectory of individual ranks U_{it}. Then, using the cdf F_ε corresponding to the estimate in the empirical analysis, we transform the simulated ranks U_{it} into simulated error terms ε_{it}. We 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 from the linear panel model (2.1) we are able to construct an unbalanced simulated panel of $y_{i,t}$ 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 fixed-effects bias in the framework of our empirical model, and the performance of the bias adjusted estimator. On this simulated sample we run the whole estimation procedure presented in Section 3 for both the functional and the finite-dimensional parameters, and obtain then the estimates for the mobility measure both with and without the adjustment to correct for the fixed-effects bias. For the latter, we use the analytical bias correction based on a Gaussian approximation for feasibility reasons. We repeat this procedure in 100 Monte Carlo replications.

In the following presentation of the Monte Carlo results, we focus on the bias of the relative mobility measure $\gamma = m(\bar{u}; \bar{x}, \bar{x})$, that is the primary object of interest in our empirical analysis. Simulation results for the copula functional parameter and the marginal rank distribution are available upon requests. In Figure 1 we plot the estimated mobility measure γ as a function of the past rank $\bar{u}$, for four choices of the individual characteristics vector $\bar{x}$ (which correspond to the 25%, 50%, 75% and 95% percentiles of the mobility score and of the marginal distribution score). We find that the bias of estimator $\hat{\gamma}$ is positive and moderate, but not negligible, in all the four cases considered. Our bias-adjusted estimator $\hat{\gamma}_B$ is able to correct on average for around 38% of the bias in the first case, and for about 67% of the bias in the last three cases. We also find that the true value of the mobility measure lies within the 95% point-wise confidence intervals that we constructed around the bias-adjusted estimate on the basis of the 100 Monte Carlo simulations.

To summarize, the main message from our Monte Carlo experiments is that the fixed-effects bias in the estimated mobility due to the small time series dimension of our panel is moderate, but not negligible, and the analytical bias-adjustment eliminates between 40% and 70% of this bias depending on the regressors' value.

Figure 1: Monte Carlo simulations - mobility measure



In each panel of this Figure, the thin line represents the true measure of mobility $m(\bar{u};\bar{x},\bar{x})$ vs the past rank $U_{i,t-1} = \bar{u}$, for given values of $X_{i,t} = X_{i,t-1} = \bar{x}$. The bold and dashed lines stand for the average estimated measure of mobility with, respectively without, bias adjustment. The shaded areas are the 95% pointwise confidence intervals, representing the variability of the Monte Carlo simulations. The four panels correspond to different values of the mobility score $W_{2,it}$ and of the marginal distribution score $W_{1,it}$, that are the 25%, 50%, 75% and 95% percentiles 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). 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. We consider for the analysis survey years from 1968 to 1997[12]. Our dataset contains 126'432 individual-year observations. They correspond to 8429 individuals, with an average permanence in the dataset of about 15 years. We correct the finite sample bias due to the estimated fixed effects in the preliminary wage regression. The Monte Carlo simulations presented in Section 4, which have been calibrated to mimic our dataset, confirm that the bias, once the correction is applied, is small[13]. 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, we exclude observations with wage equal to zero. Similarly to Bonhomme and Robin (2009), we use as explanatory variables age, age squared, and a qualitative variable representing the highest education level achieved by the individual. The education dummies are constructed on the basis of the variable "years spent in education". According to the US education system, the first dummy corresponds to 0-11 grades, the second dummy stands for high school or 12 grades and some nonacademic 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.

[12]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.

[13]An alternative that we do not explore in the present paper consists into performing a bias correction via jackknife, by 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).

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

Variable	Mean	Std. Dev.	Min.	Max.
Age	37.4830	11.2561	17	65
Elementary and middle school	0.2663	0.4420	0	1
High school	0.3565	0.4790	0	1
Some college	0.1989	0.3992	0	1
College degree or higher education	0.1783	0.3828	0	1
Log wage	9.2780	1.8641	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).

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, to take into account the potential presence of unobserved heterogeneity across workers[14]. 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 that log wage is concave in individual age. 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 ranks as in equation (3.1). Then, we use the empirical ranks $\hat{U}_{it}$ for the estimation of the semi-nonparametric dynamic copula specification defined in Section 2. We use the same regressors as above in the marginal rank distribution and the copula pdf (but rely on age at the begininng of the sample to avoid trending variables in the marginal distributions). 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 errors.

[14] 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.

In our empirical application, we do not account for the possible endogeneity of the unemployment patterns. We deal with missing observations in our unbalanced panel by making the missing-at-random assumption. 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[15]. 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. Details of the numerical implementation of the estimators, as well as detailed estimation results for the univariate distribution of the ranks conditional on covariates, and for the joint copula distribution of the ranks, are reported in the Supplementary Material, Appendix F. In Figure 2 we display the empirical counterpart of the mobility measure presented in Proposition 3. More precisely, we represent the estimate of $\gamma(\theta) = m(\overline{u}, \overline{x}, \overline{x})$ as a function of $\overline{u}$, keeping $\overline{x}$ fixed. In each of the panels of Figure 2, indeed, $\overline{x}$ is fixed and corresponds to a specific combination of explanatory variables (age and education), as explained in the caption.

In Figure 2, the y-axis records the value of our mobility measure. 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 2, 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),

[15] 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.

and in the bottom panel we have the case of an individual with relatively high scores (45-year-old worker with college degree).

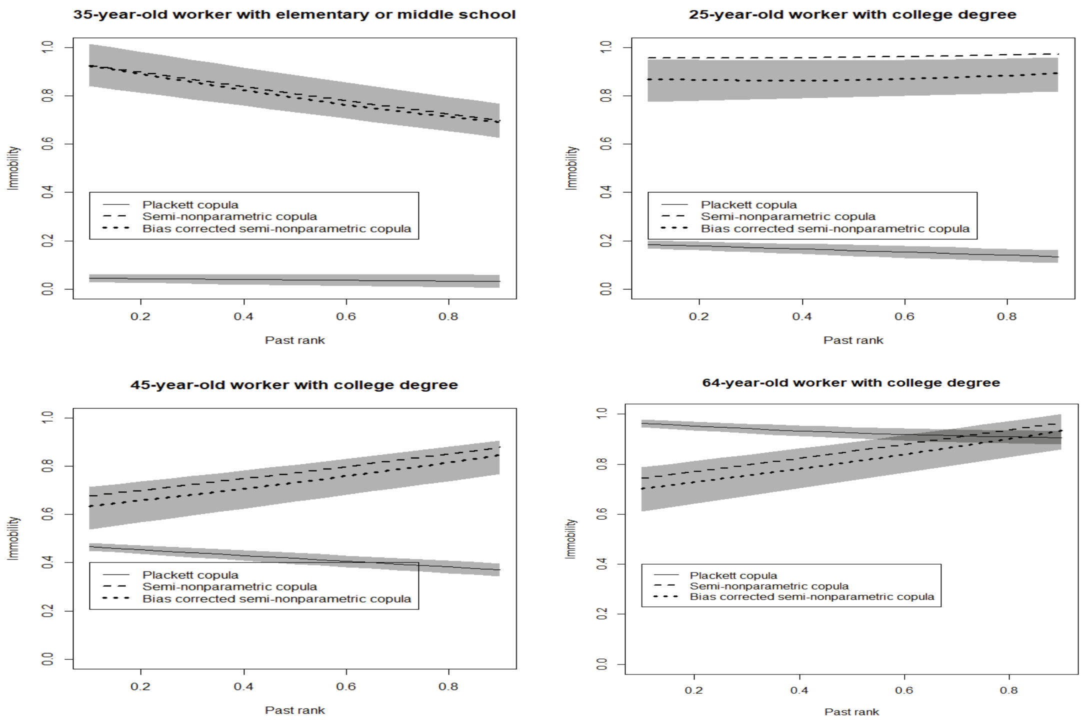
In Figure 2 we find that individuals with low values of the scores (35-year-old worker with elementary or middle school) exhibit lower mobility than their colleagues with higher scores (e.g. 45 or 64-year-old worker with college degree), in particular at the bottom of the rank distribution, i.e. they are subject to the low-pay trap. 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 2), we find that rank persistence is rather high (i.e. between 0.8 and 1) 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.

In the case of a worker with high or very high values of the scores (i.e. a 45 or a 64-year-old worker with college degree), 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 around 0.6-0.7 for this type of worker. This means that, if individuals characterized by high values of the scores are in a low position of the wage distribution, then they are 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[16]) only rarely overlap[17].

[16]Bootstrap methods have been found to be valid in the context of Sieve estimation (see Cheng and Huang (2010)).

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

Figure 2: 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 (3.3) using estimates from the semi-nonparametric copula model (dashed line). The bias-adjusted estimate is obtained with the analytical bias correction (short-dashed line). The chosen sets of individual characteristics correspond to different values of the marginal distribution and mobility scores, which are $W_{1,it} = -1.95$ and $W_{2,it} = -1.75$ in the upper left panel, $W_{1,it} = -0.68$ and $W_{2,it} = -0.23$ in the upper right panel, $W_{1,it} = 0.92$ and $W_{2,it} = 1.14$ in the bottom left panel and $W_{1,it} = 2.03$ and $W_{2,it} = 2.22$ in the bottom right 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 $B = 500$.

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 2 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). The Plackett copula cdf is $C(u, v; \tau) = \frac{1}{2\tau}(1 + \tau(u + v) - \sqrt{a(u, v)})$, where $a(u, v) = [1 + \tau(u + v)]^2 - 4uv\tau(\tau + 1)$ and $\tau \geq 0$. As in their paper, the parameter of the Plackett copula $\tau = \exp(X'_{i,t}\beta_2)$ is function of the individual characteristics. The marginal distribution of the rank is modeled as in our semi-nonparametric specification for comparison purpose. Our mobility measure computed with this parametric copula model is (see the Supplementary material):

$$m(\bar{u};\bar{x},\bar{x}) = \frac{\bar{\tau}}{2+\bar{\tau}} \frac{g(\bar{u}|\bar{x}'\beta_1)}{g\{G^{-1}[\frac{1+\bar{\tau}G(\bar{u}|\bar{x}'\beta_1)}{2+\bar{\tau}}|\bar{x}'\beta_1]\}}, \tag{5.2}$$

with $\bar{\tau} = \exp(\bar{x}'\beta_2)$. By contrasting the estimated mobility patterns obtained with the parametric, respectively with our semi-nonparametric model, we deduce that, by adopting a fully parametric copula 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 2, neither as a function of the past rank (the estimated mobility curves appear very flat), nor as a function of the individual characteristics. In fact, from equation (5.2) we see that, up to the effect of the marginal rank distribution (that in our estimate appears rather close to uniform, so that the second ratio in the RHS of (5.2) is close to 1) the mobility measure is independent of the past rank $\bar{u}$, and increasing in the mobility score, which explains the patterns in Figure 2. The Plackett copula 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 2.

Finally, note that, without applying the bias correction, in all the four cases (i.e. low score, middle score, high and very high score), we would overestimate the degree of association between the present and the past rank, i.e. we would sistematically underestimate wage mobility. Indeed, from the four panels of Figure 2, we notice that the bias-corrected (im-)mobility measure always lies below the uncorrected one. Note that the bias correction reported in the four panels of Figure 2 has been obtained by adopting a Gaussian approximation. However, the resulting estimated bias is rather close to the one that we obtain by adopting the full analytical bias correction. Indeed, the difference in the two estimated biases amounts, on average, to between 1% and 2% of our estimated mobility measure, in the four cases presented in Figure 2.

To conclude our empirical analysis, we present the results of testing the restriction (2.28) in Proposition 3. The test statistic is:

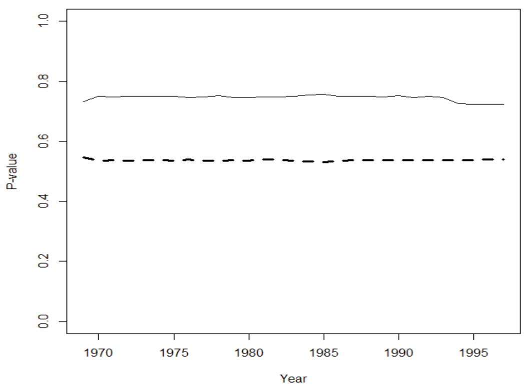
$$\hat{\xi}_t = \frac{\int_0^1 \int_{-\infty}^{\infty} \hat{\psi}_t(u,w)\hat{h}(u,w)^2 dudw}{\hat{\sigma}_t} \tag{5.3}$$

computed for each component of the vector of explanatory variables in the preliminary regression (i.e. age and age squared), standardized by the estimated standard deviation, $\hat{\sigma}_t$, obtained from 500 bootstrap replications[18]. The conditional expectations in $\hat{\psi}(u,w) = \hat{F}_{\varepsilon,t}^{-1}(u)\hat{E}(X_{it} - \bar{X}_{i\cdot}|W_{1,i,t} = w) -$

[18] Asymptotic normality of test statistics could be established by using arguments as in Theorem 1.

$\frac{1}{T}\sum_{s=1}^{T}\hat{F}_{\varepsilon,s}^{-1}(u)\hat{E}(X_{i,t}-\bar{X}_{i\cdot}|W_{1,i,s}=w)$ have been approximated via Sieve estimation with three polynomial terms, while $\hat{F}_{\varepsilon,t}^{-1}(u)$ is the empirical quantile of residuals. Since the bias at order $\frac{1}{T}$ proves to be moderate in our results (see e.g. Figure 2), we neglect bias correction for the test. We compute the p-values of the statistics using the asymptotic standard Gaussian distribution. From Figure 3, we deduce that the p-values are large at each sample date. The test cannot reject the null of the validity of constraints (2.28) corresponding to Assumption 5 at usual significance levels. Hence, even if we do not theoretically impose Assumption 5 in our estimation strategy, we can show empirically that it is not violated. As a robustness check, in Appendix F in the Supplementary Material we show the same quantities, but obtained, respectively, with a two-term and with a four-term Sieve polynomial. The findings do not change, in the sense that we still never reject the null hypothesis.

Figure 3: P-values for the test statistics obtained with Sieve estimation with three terms



We display the p-values for the test statistic in (5.3) for each year in our sample, for the two covariates that are included in the preliminary regression (2.1): age and age squared.

6 Concluding remarks

In this paper we estimate a flexible model for the wage rank dynamics. We develop a new family of semi-nonparametric copulas which are well-suited to describe the dynamics of earning ranks conditional on regressors. This novel semi-nonparametric copula model allows for greater flexibility than a fully parametric one, and in particular allows to investigate relative mobility as a function of past wage rank and individual characteristics. We propose consistent estimators for both the marginal rank distribution, and the functional parameter which characterizes the copula distribution of present and past ranks, conditional on covariates. We show the asymptotic normality of our Sieve estimator for the rank mobility functional when N and T grow to infinity, with an adjustment for the incidental parameter bias induced by the individual fixed effects in the wage equation. We provide Monte Carlo evidence that the analytical bias adjustment works well in our setting.

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. On the contrary, we find that workers who are either at the beginning of their career or who have a low educational level 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.

Appendix A: Assumptions

In this Appendix we list the regularity conditions used to establish the large sample properties of our estimators. We start with introducing the necessary notation. Let us define the pseudo-norm $\|\cdot\|_*$ by $\|\theta-\theta_0\|_*^2 = E\left[\left(\frac{\partial \log l_{i,t}(\theta_0)}{\partial\theta}[\theta-\theta_0]\right)^2\right] = \langle\theta-\theta_0, I_0(\theta-\theta_0)\rangle$. The Kullback-Leibler discrepancy $K(\theta_0,\theta) = E[\log l_{i,t}(\theta_0) - \log l_{i,t}(\theta)]$, where $l_{i,t}(\theta) := l(U_{i,t}|U_{i,t-1}, X_{i,t}, X_{i,t-1};\theta)$, is minimized for $\theta=\theta_0$. Let us define $\tilde{\theta}_{N,T}(\cdot) = \underset{\theta\in\Theta_{N,T}}{\arg\max}\, \frac{1}{T}\sum_t E[\log \hat{l}_{i,t}(\theta)]$ and $K_{NT}(\tilde{\theta}_{N,T},\theta) := \frac{1}{T}\sum_t E[\log \hat{l}_{i,t}(\tilde{\theta}_{N,T}) - \log \hat{l}_{i,t}(\theta)]$, where $\hat{l}_{i,t}(\theta) := l(\hat{U}_{i,t}|\hat{U}_{i,t-1}, X_{i,t}, X_{i,t-1};\theta)$, and similarly $\bar{\theta}_{N,T} = \underset{\theta\in\Theta_{N,T}}{\arg\max}\, \frac{1}{T}\sum_{t=1}^T E[\log \tilde{l}_{i,t}(\theta)]$, where $\tilde{l}_{i,t}(\theta) = l(\tilde{U}_{i,t}|\tilde{U}_{i,t-1}, X_{i,t}, X_{i,t-1};\theta)$, for $\tilde{U}_{i,t} := U_{i,t} + \frac{1}{T}H_t(\varepsilon_{i,t}) - f_t(\varepsilon_{i,t})\bar{\varepsilon}_i + \frac{1}{2}f_t'(\varepsilon_{i,t})\bar{\varepsilon}_i^2$ and $H_{\varepsilon,t}(\varepsilon) = \sum_{s=1}^T E[\varepsilon_{i,s}|\varepsilon_{i,t} = \varepsilon]f_{\varepsilon,t}(\varepsilon) + \frac{\omega^2}{2}f'_{\varepsilon,t}(\varepsilon)$. Functions $\tilde{\theta}_{N,T}$ and $\bar{\theta}_{N,T}$ correspond to the counterparts of true parameter function θ_0 computed with the estimated ranks $\hat{U}_{i,t}$, and with their large-N limits $\tilde{U}_{i,t}$, respectively. Finally, let $\mu_{N,T}(g(z)) = \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T(g(z_{i,t}) - E[g(z_{i,t})])$ for a function $g(\cdot)$.

Assumption A.1. *Process $(U_{i,t}, X_{i,t})$, for t varying, is geometrically β-mixing, for any i.*

Assumption A.2. *We have $c_1\|\theta-\theta_0\|_*^2 \leq K(\theta_0,\theta) \leq c_2\|\theta-\theta_0\|_*^2$, for any $\theta\in\Theta$ and some constants $c_2 \geq c_1 > 0$.*

Assumption A.3. *The Sieve space $\Theta_{N,T}$ is such that: (i) $\|\theta_0 - \pi_{N,T}\theta_0\|_* = o((NT)^{-1/4})$ with $\pi_{N,T}\theta_0 \in \Theta_{N,T}$, and (ii) the bracketing metric entropy of set $\mathcal{F}_{N,T} = \{\log\frac{l(\cdot|\cdot,\cdot,\cdot;\theta)}{l(\cdot|\cdot,\cdot,\cdot;\theta_0)} : \theta\in\Theta_{N,T}\}$ is $H(\epsilon,\mathcal{F}_{N,T}) \leq Ak_{N,T}\log(k_{N,T}/\epsilon)$, where A is a constant and (iii) the number of parameters $k_{N,T} = m_{N,T}^\rho(m_{N,T}^\rho + 1) + m_{N,T}^h(m_{N,T}^h + 1) + 2p$ indexing the set $\Theta_{N,T}$ is such that $k_{N,T}\log(NT) = o(\sqrt{NT})$.*

Assumption A.4. *(i) We have*

$$\sup_{\theta \in \Theta_{N,T}: \|\theta-\theta_0\|_* \leq \epsilon} V[l_{i,t}(\theta_0) - l_{i,t}(\theta)] \leq A_1 \epsilon^2,$$

for a constant $A_1 > 0$. *(iv) We have for* $s \in (0, 2)$*:*

$$\sup_{\theta \in \Theta_{N,T}: \|\theta-\theta_0\|_* \leq \epsilon} |l_{i,t}(\theta_0) - l_{i,t}(\theta)| \leq \epsilon^s W_{N,T,i,t},$$

where $\sup_{N,T \geq 1} E[W_{N,T,i,t}^{\gamma}] \leq A_3$*, for* $\gamma > 2$.

Assumption A.5. *The information operator* $I_0 : \mathbb{V} \to \mathbb{V}$ *is bounded, invertible, with bounded inverse.*

Assumption A.6. *We have* $\gamma(\theta) = \gamma(\theta_0) + \frac{\partial \gamma(\theta_0)}{\partial \theta}[\theta - \theta_0] + O\left(\|\theta - \theta_0\|_*^2\right)$*, for any* θ *in a small neighborhood of* θ_0 *such that* $\|\theta - \theta_0\|_* = O(\delta_{NT})$*, where* $\delta_{N,T} = (NT)^{-1/4}$.

Assumption A.7. *Let* $\hat{r}_{i,t}[\theta - \tilde{\theta}_{N,T}] = \log \hat{l}_{i,t}(\theta) - \log \hat{l}_{i,t}(\tilde{\theta}_{N,T}) - \frac{\partial \log \hat{l}_{i,t}(\tilde{\theta}_{N,T})}{\partial \theta}[\theta - \tilde{\theta}_{N,T}]$ *be the remainder term in the Taylor expansion of* $\log \hat{l}_{i,t}(\theta)$ *around* $\tilde{\theta}_{N,T}$*. Then we have:*

$$(i) \quad \sup_{\theta \in \Theta_{N,T}: \|\theta-\tilde{\theta}_{N,T}\|_* \leq \delta_{N,T}} \mu_{N,T}\left(\hat{r}_{i,t}[\theta - \tilde{\theta}_{N,T}] - \hat{r}_{i,t}[\pi_{N,T}\theta^*(\theta, \varepsilon_{N,T}) - \tilde{\theta}_{N,T}]\right) = O_p(\varepsilon_{N,T}^2),$$

$$(ii) \quad \sup_{\theta \in \Theta_{N,T}: \|\theta-\tilde{\theta}_{N,T}\|_* \leq \delta_{N,T}} \left[K_{N,T}(\tilde{\theta}_{N,T}, \theta) - \frac{1}{2}\|\theta - \tilde{\theta}_{N,T}\|_{N,T}^2\right] = O(\varepsilon_{N,T}^2),$$

$$(iii) \quad \sup_{\theta \in \Theta_{N,T}: \|\theta-\tilde{\theta}_{N,T}\|_* \leq \delta_{N,T}} \|\theta^*(\theta, \varepsilon_{N,T}) - \pi_{N,T}\theta^*(\theta, \varepsilon_{N,T})\|_* = O(\delta_{N,T}^{-1}\varepsilon_{N,T}^2),$$

$$(iv) \quad \sup_{\theta \in \Theta_{N,T}: \|\theta-\tilde{\theta}_{N,T}\|_* \leq \delta_{N,T}} \mu_{N,T}\left(\frac{\partial \log \hat{l}_{i,t}(\tilde{\theta}_{N,T})}{\partial \theta}[\theta^*(\theta, \varepsilon_{N,T}) - \pi_{N,T}\theta^*(\theta, \varepsilon_{N,T})]\right) = O_p(\varepsilon_{N,T}^2),$$

where $\|\theta - \tilde{\theta}_{N,T}\|_{N,T}^2 := \langle \theta - \tilde{\theta}_{N,T}, \tilde{I}_{N,T}(\theta - \tilde{\theta}_{N,T})\rangle$ *and* $\tilde{I}_{N,T}$ *is a bounded and self-adjoint operator, invertible with bounded inverse, for any* N, T *large enough,* $\theta^*(\theta, \varepsilon_{N,T}) = \theta + \varepsilon_{N,T}\tilde{I}_{N,T}^{-1}u^*$*, with* $u^* = \pm v_{N,T}$ *and* $\varepsilon_{N,T} = o((NT)^{-1/2})$.

Assumption A.8. *For a* $\bar{\beta} > 2$ *we have:* $E\left[\left|\frac{\partial^{1+\alpha_u+\alpha_v} \log l_{i,t}(\theta_0)}{\partial \theta \partial u^{\alpha_u} \partial v^{\alpha_v}}[\tilde{v}_{N,T}]\right|^{\bar{\beta}}\right] = O(1)$*, for* $\alpha_u, \alpha_v = 0, 1, 2$*, where* $\tilde{v}_{N,T} = \tilde{I}_{N,T}^{-1} v_{N,T}$.

Assumption A.9. *The distribution* $F_t(\cdot)$ *is continuous and: (i) the density* $f_t(\cdot) = F_t'(\cdot)$ *is differentiable, with bounded derivative, (ii) the errors* $\varepsilon_{i,t} = F_t^{-1}(U_{i,t})$ *are such that* $E(\varepsilon_{i,t}) = 0$*,* $E[|\varepsilon_{i,t}|^{2\bar{\beta}}] \leq C$*, for a constant* C.

Assumption A.10. *The zero-mean variables* $\tilde{\zeta}_{i,t} := \tilde{\xi}_{i,t}[\tilde{v}_{N,T}] + \tilde{\nu}_{i,t}[\tilde{v}_{N,T}]\Delta_{i,t} + \tilde{\eta}_{i,t}[\tilde{v}_{N,T}]\Delta_{i,t-1}$ *satisfy a CLT, i.e.*

$$\tilde{\sigma}_{N,T}^{-1} \frac{1}{\sqrt{NT}} \sum_i \sum_t \tilde{\zeta}_{i,t} \Rightarrow N(0,1)$$

as $N, T \to \infty$ *with* $\tilde{\sigma}_{N,T}^2 = V[\frac{1}{\sqrt{NT}} \sum_i \sum_t \tilde{\zeta}_{i,t}]$, *where* $\tilde{\xi}_{i,t}[\tilde{v}_{N,T}] := \frac{\partial \log l(\tilde{U}_{i,t}|\tilde{U}_{i,t-1}, X_{i,t}, X_{i,t-1}; \bar{\theta}_{N,T})}{\partial \theta}[\tilde{v}_{N,T}]$, $\tilde{\nu}_{i,t}[\tilde{v}_{N,T}] := \frac{\partial^2 \log l(\tilde{U}_{i,t}|\tilde{U}_{i,t-1}, X_{i,t}, X_{i,t-1}; \bar{\theta}_{N,T})}{\partial \theta \partial u}[\tilde{v}_{N,T}]$ *and* $\tilde{\eta}_{i,t}[\tilde{v}_{N,T}] := \frac{\partial^2 \log l(\tilde{U}_{i,t}|\tilde{U}_{i,t-1}, X_{i,t}, X_{i,t-1}; \bar{\theta}_{N,T})}{\partial \theta \partial v}[\tilde{v}_{N,T}]$, *and* $\Delta_{i,t} := \frac{1}{N-1} \sum_{j \neq i} (\mathbb{I}(U_{j,t} \leq U_{i,t}) - U_{i,t}) - \left(\frac{1}{N-1} \sum_{j \neq i} \varepsilon_{j,t} \right) f_t(\varepsilon_{i,t})$.

Appendix B: Proofs of Propositions and Theorems

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:

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

then:

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

where $g(U|X)$ is a conditional pdf such that (2.11) holds. To show this, suppose that equation (b.1) holds. Then we have:

$$\begin{aligned} l(U_t|X_t) &= \int \int l(U_t|U_{t-1}, X_{t-1}, X_t) l(U_{t-1}, X_{t-1}|X_t) dU_{t-1} dX_{t-1} \\ &= \int \int l(U_t|U_{t-1}, X_{t-1}, X_t) l(U_{t-1}|X_{t-1}, X_t) l(X_{t-1}|X_t) dU_{t-1} dX_{t-1}. \end{aligned}$$

Moreover, we have:

$$l(U_{t-1}|X_{t-1}, X_t) = \frac{l(U_{t-1}, X_{t-1}, X_t)}{l(X_{t-1}, X_t)} = \frac{l(X_t|U_{t-1}, X_{t-1}) l(U_{t-1}, X_{t-1})}{l(X_t|X_{t-1}) l(X_{t-1})} = \frac{l(U_{t-1}, X_{t-1})}{l(X_{t-1})} = l(U_{t-1}|X_{t-1}),$$

where the third equality is an implication of the absence of Granger causality in Assumption 2. Then:

$$l(U_t|X_t) = \int \int l(U_t|Z_{t-1}, X_{t-1}, X_t) l(U_{t-1}|X_{t-1}) l(X_{t-1}|X_t) dU_{t-1} dX_{t-1}.$$

Now, by replacing the definition of the conditional density in (2.12), and using (b.1), we get:

$$\begin{aligned} l(U_t|X_t) = \int \int & g(U_t|X_t) c[G(U_t|X_t), G(U_{t-1}|X_{t-1}); \rho(\cdot, X_t)] \\ & \cdot g(U_{t-1}|X_{t-1}) l(X_{t-1}|X_t) dU_{t-1} dX_{t-1}. \end{aligned} \tag{b.3}$$

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

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

which yields (b.2).

In Step 2 of the proof, we use the initial condition in (2.13) and use repeatedly the implication derived in Step 1, to get $l(U_t|X_t) = g(U_t|X_t)$, at any $t \geq 1$. Then, by integrating out the explanatory variables vector X_t and using the property in (2.11) we get:

$$l(U_t) = \int l(U_t|X_t)l(X_t)dX_t = \int g(U_t|X_t)l(X_t)dX_t = 1,$$

which yields (2.14).

B.2 Proof of Proposition 2

From equation (2.17) it follows that equation (2.18) is both sufficient and necessary for the uniform distribution $U(0,1)$ to be an invariant distribution for process (U_t).

B.3 Proof of Proposition 3

The condition in Assumption 5 is:

$$E[\varepsilon_{i,t}(X_{i,t} - \bar{X}_{i,\cdot})] = \frac{1}{T}\sum_{s=1}^{T} E[\varepsilon_{i,s}(X_{i,t} - \bar{X}_{i,\cdot})], \tag{b.4}$$

for all $t = 1, ..., T$. With $\varepsilon_{i,s} = F_{\varepsilon,s}^{-1}(U_{i,s})$ and $X_i = (X_{i,1}, ..., X_{i,T})'$, we can rewrite the expectation on the RHS as:

$$\begin{aligned} E[\varepsilon_{i,s}(X_{i,t} - \bar{X}_{i,\cdot})] &= \int\int F_{\varepsilon,s}^{-1}(u_{i,s})(X_{i,t} - \bar{X}_{i,\cdot})l(u_{i,s}|X_i)l(X_i)du_{i,s}dX_i \\ &= \int\int F_{\varepsilon,s}^{-1}(u_{i,s})(X_{i,t} - \bar{X}_{i,\cdot})g(u_{i,s}|X_{i,s})l(X_i)du_{i,s}dX_i \\ &= \int\int F_{\varepsilon,s}^{-1}(u_{i,s})E(X_{i,t} - \bar{X}_{i,\cdot}|X_{i,s})g(u_{i,s}|X_{i,s})l(X_{i,s})du_{i,s}dX_{i,s}, \end{aligned}$$

where we use Assumption 2 and the specification (2.23) in the second equality. Now we use that $g(u_{i,s}|X_{i,s}) = g(u_{i,s}|W_{1,i,s})$ by the index model, where $W_{1,i,s} = \beta_1' X_{i,s}$, which implies that

$$\int E(X_{i,t} - \bar{X}_{i,\cdot}|X_{i,s})g(u_{i,s}|X_{i,s})l(X_{i,s})dX_{i,s} = \int E(X_{i,t} - \bar{X}_{i,\cdot}|W_{1,i,s})g(u_{i,s}|W_{1,i,s})l(W_{1,i,s})dW_{1,i,s}$$

using the law of the iterated expectation. We get:

$$\begin{aligned} E[\varepsilon_{i,s}(X_{i,t} - \bar{X}_{i,\cdot})] &= \int\int F_{\varepsilon,s}^{-1}(u_{i,s}) E(X_{i,t} - \bar{X}_{i,\cdot} | W_{1,i,s}) g(u_{i,s}, W_{1,i,s}) du_{i,s} dW_{1,i,s} \\ &= \int\int F_{\varepsilon,s}^{-1}(u) E(X_{i,t} - \bar{X}_{i,\cdot} | W_{1,i,s} = w) g(u, w) du dw. \end{aligned}$$

By a similar argument:

$$E[\varepsilon_{i,t}(X_{i,t} - \bar{X}_{i,\cdot})] = \int\int F_{\varepsilon,t}^{-1}(u) E(X_{i,t} - \bar{X}_{i,\cdot} | W_{1,i,t} = w) g(u, w) du dw.$$

With the definition $\psi_t(u,w) = F_{\varepsilon,t}^{-1}(u) E(X_{i,t} - \bar{X}_{i,\cdot} | W_{1,i,t} = w) - \frac{1}{T}\sum_{s=1}^{T} F_{\varepsilon,s}^{-1}(u) E(X_{i,t} - \bar{X}_{i,\cdot} | W_{1,i,s} = w)$, condition (b.4) becomes $\int\int \psi_t(u,w) g(u,w) du dw = 0$, for all t. Using $g(u,w) = h(u,w)^2$ the conclusion follows.

B.4 Consistency rate of the Sieve ML estimator $\hat{\theta}$

As a preliminary result in view of proving the asymptotic normality of estimator $\hat{\gamma}$, we first establish the convergence rates of estimator $\hat{\alpha}$ in Euclidean norm, and of estimator $\hat{\theta}$ in norm $\|\cdot\|_*$. The convergence rate $\|\tilde{\theta} - \theta_0\|_* = o_p((NT)^{-1/4})$ for the unfeasible Sieve estimator defined by $\tilde{\theta} = \arg\max_{\theta \in \Theta_{N,T}} \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} \log l(U_{i,t}|U_{i,t-1}, X_{i,t}, X_{i,t-1}; \theta)$ based on the true ranks follows from Theorem 1 in Chen and Shen (1998) and Assumptions A.1-A.4. To cope with the estimation error induced by replacing $U_{i,t}$ with the empirical ranks $\hat{U}_{i,t}$, we follow a similar approach as in the proof of Theorem 3.1 in Ai and Chen (2003).

Lemma 1. *Under Assumptions 1-5 and A.1-A.4 we have: a)* $|\hat{\alpha} - \alpha_0| = O_p((NT)^{-1/2})$ *and b)* $\|\hat{\theta} - \theta_0\|_* = o_p((NT)^{-1/4})$.

B.5 Proof of Theorem 1

From Lemma 1 b) we have $\|\hat{\theta} - \theta_0\|_* \leq \delta_{N,T}$ w.p.a. 1, with $\delta_{N,T} = (NT)^{-1/4}$. Thus, from Assumption A.6, for the plug-in estimator $\hat{\gamma} = \gamma(\hat{\theta})$ of parameter $\gamma_0 = \gamma(\theta_0)$ we have the second-order Taylor expansion:

$$\gamma(\hat{\theta}) = \gamma(\theta_0) + \frac{\partial \gamma(\theta_0)}{\partial \theta}[\hat{\theta} - \theta_0] + O_p(\|\hat{\theta} - \theta_0\|_*^2). \tag{b.5}$$

Let us first isolate the bias term that arises from the Sieve approximation. On the finite-dimensional linear space $\mathbb{V}_{NT}$, the Rietz representer $v_{NT} \in \mathbb{V}_{NT}$ w.r.t. $\langle \cdot, \cdot \rangle$ is such that

$$\frac{\partial \gamma(\theta_0)}{\partial \theta}[v] = \langle v_{NT}, v \rangle, \qquad \forall v \in \mathbb{V}_{NT}.$$

Thus, the second-order expansion (b.5) yields:

$$\begin{aligned}\gamma(\hat{\theta}) &= \gamma(\theta_0) + \frac{\partial\gamma(\theta_0)}{\partial\theta}[\hat{\theta} - \theta_{0,NT}] + \frac{\partial\gamma(\theta_0)}{\partial\theta}[\theta_{0,NT} - \theta_0] + O_p(\|\hat{\theta} - \theta_0\|_*^2) \\ &= \gamma(\theta_0) + \langle v_{NT}, \hat{\theta} - \theta_{0,NT}\rangle + \frac{\partial\gamma(\theta_0)}{\partial\theta}[\theta_{0,NT} - \theta_0] + O_p(\|\hat{\theta} - \theta_0\|_*^2),\end{aligned}$$

where $\theta_{0,NT} = \underset{\theta\in\Theta_{N,T}}{\arg\min}\|\theta-\theta_0\|$. Now, we have $\langle v_{NT}, \hat{\theta}-\theta_{0,NT}\rangle = \langle v_{NT}, \hat{\theta}-\theta_0\rangle$ because $\langle v_{NT}, \theta_{0,NT}-\theta_0\rangle = 0$ for any $v_{NT} \in \mathbb{V}_{N,T}$. Hence we have

$$\sqrt{NT}\left(\gamma(\hat{\theta}) - \gamma(\theta_0) - \frac{\partial\gamma(\theta_0)}{\partial\theta}[\theta_{0,NT} - \theta_0]\right) = \sqrt{NT}\langle v_{NT}, \hat{\theta} - \theta_0\rangle + O_p(\sqrt{NT}\|\hat{\theta} - \theta_0\|_*^2).$$

By Lemma 1 b), the second-order term on the RHS is $o_p(1)$, and we get:

$$\sqrt{NT}\left(\gamma(\hat{\theta}) - \gamma(\theta_0) - \frac{\partial\gamma(\theta_0)}{\partial\theta}[\theta_{0,NT} - \theta_0]\right) = \sqrt{NT}\langle v_{NT}, \hat{\theta} - \theta_0\rangle + o_p(1). \tag{b.6}$$

Let us now single-out the bias contribution from the estimation of the ranks. In Lemma 2 we derive an asymptotic expansion for the rank estimator $\hat{U}_{i,t}$ and plug it into the expectation $E[\log l(\hat{U}_{i,t}|\hat{U}_{i,t-1}, X_{i,t}, X_{i,t-1}; \theta)]$ to get an asymptotic expansion for $\tilde{\theta}_{N,T} - \theta_0$ in Lemma 3.

Lemma 2. *We have the asymptotic expansion of the rank estimate as $N, T \to \infty$:*

$$\hat{U}_{i,t} = U_{i,t} + \Delta_{i,t} + \frac{1}{T}H_t(\varepsilon_{i,t}) - f_t(\varepsilon_{i,t})\bar{\varepsilon}_i + \frac{1}{2}f_t'(\varepsilon_{i,t})\bar{\varepsilon}_i^2 + \frac{1}{2N}\sigma_t^2 f_t'(\varepsilon_{i,t}) + \frac{1}{2}f_t'(\varepsilon_{i,t})\bar{\varepsilon}_{\cdot t}^2, \tag{b.7}$$

up to negligible terms.

Lemma 3. *As $N, T \to \infty$ we have:*

$$\langle v_{N,T}, \tilde{\theta}_{N,T} - \theta_0\rangle = \frac{1}{T}\mathcal{B}_T(\tilde{v}_{NT}) + O(1/T^2 + 1/N) + o(1/\sqrt{NT}), \tag{b.8}$$

where function $\mathcal{B}_T(\cdot)$ is defined in Theorem 1, and $\tilde{v}_{N,T} = \tilde{I}_{N,T}^{-1}v_{NT}$.

The remainder term in (b.8) is $o(1/\sqrt{NT})$ if $N, T \to \infty$ such that $T \ll N \ll T^3$.
In Lemma 4 we write the scalar product $\langle v_{N,T}, \hat{\theta} - \tilde{\theta}_{N,T}\rangle$ in terms of a sample average of the directional derivative of the log-likelihood.

Lemma 4. *Under Assumption A.7 and regularity conditions:*

$$\sqrt{NT}\langle v_{N,T}, \hat{\theta} - \tilde{\theta}_{N,T}\rangle = \frac{1}{\sqrt{NT}}\sum_i\sum_t \frac{\partial \log l(\hat{U}_{i,t}|\hat{U}_{i,t-1}, X_{it}, X_{i,t-1}; \tilde{\theta}_{N,T})}{\partial\theta}[\tilde{v}_{N,T}] + o_p(1).$$

By equations (b.6) and (b.8), Lemma 4, and condition $T \ll N \ll T^3$, we get the expansion:

$$\begin{aligned}
&\sqrt{NT}\left(\gamma(\hat{\theta}) - \gamma(\theta_0) - \frac{\partial\gamma(\theta_0)}{\partial\theta}[\theta_{0,NT} - \theta_0] - \frac{1}{T}\mathcal{B}_T(\tilde{v}_{N,T})\right) \\
&= \frac{1}{\sqrt{NT}}\sum_i\sum_t \frac{\partial \log l(\hat{U}_{i,t}|\hat{U}_{i,t-1}, X_{it}, X_{i,t-1}; \tilde{\theta}_{N,T})}{\partial\theta}[\tilde{v}_{N,T}] + o_p(1). \qquad \text{(b.9)}
\end{aligned}$$

Let us now show that $\frac{1}{\sqrt{NT}}\sum_i\sum_t \frac{\partial \log l(\hat{U}_{i,t}|\hat{U}_{i,t-1}, X_{it}, X_{i,t-1}; \tilde{\theta}_{N,T})}{\partial\theta}[\tilde{v}_{N,T}]$ is asymptotically standard normal after an appropriate rescaling.

Lemma 5. *We have:*

$$\begin{aligned}
&\frac{1}{\sqrt{NT}}\sum_i\sum_t \frac{\partial \log l(\hat{U}_{i,t}|\hat{U}_{i,t-1}, X_{it}, X_{i,t-1}; \tilde{\theta}_{N,T})}{\partial\theta}[\tilde{v}_{N,T}] \\
= \;& \frac{1}{\sqrt{NT}}\sum_i\sum_t \left(\tilde{\xi}_{i,t}[\tilde{v}_{N,T}] + \tilde{\nu}_{i,t}[\tilde{v}_{N,T}]\Delta_{i,t} + \tilde{\eta}_{i,t}[\tilde{v}_{N,T}]\Delta_{i,t-1}\right) + o_p(1),
\end{aligned}$$

where $\tilde{\xi}_{i,t}[\tilde{v}_{N,T}] := \frac{\partial \log l(\tilde{U}_{i,t}|\tilde{U}_{i,t-1}, X_{i,t}, X_{i,t-1}; \bar{\theta}_{N,T})}{\partial\theta}[\tilde{v}_{N,T}]$, $\tilde{\nu}_{i,t}[\tilde{v}_{N,T}] := \frac{\partial^2 \log l(\tilde{U}_{i,t}|\tilde{U}_{i,t-1}, X_{i,t}, X_{i,t-1}; \bar{\theta}_{N,T})}{\partial\theta\partial u}[\tilde{v}_{N,T}]$ *and* $\tilde{\eta}_{i,t}[\tilde{v}_{N,T}] := \frac{\partial^2 \log l(\tilde{U}_{i,t}|\tilde{U}_{i,t-1}, X_{i,t}, X_{i,t-1}; \bar{\theta}_{N,T})}{\partial\theta\partial v}[\tilde{v}_{N,T}]$.

Lemma 6. *Under Assumption A.10, we have as $N, T \to \infty$:*

$$\sigma_{N,T}^{-1}\frac{1}{\sqrt{NT}}\sum_i\sum_t \left(\tilde{\xi}_{i,t}[\tilde{v}_{N,T}] + \tilde{\nu}_{i,t}[\tilde{v}_{N,T}]\Delta_{i,t} + \tilde{\eta}_{i,t}[\tilde{v}_{N,T}]\Delta_{i,t-1}\right) \;\Rightarrow\; N(0,1),$$

where the asymptotic variance $\sigma_{N,T}^2$ is as defined in Theorem 1.

By equation (b.9) and Lemmas 5 and 6, the conclusion follows.

B.6 Proof of Theorem 2

a) We first derive the directional derivative of the log transition density. From (2.23) we have:

$$\log l(u|u_{-1}, x, x_{-1}; \theta) = \log c\left[\frac{\int_0^u h(y, x'\beta_1)^2 dy}{\int_0^1 h(y, x'\beta_1)^2 dy}, \frac{\int_0^{u_{-1}} h(y, x_{-1}'\beta_1)^2 dy}{\int_0^1 h(y, x_{-1}'\beta_1)^2 dy}; \rho(\cdot, x'\beta_2)\right] + \log\left(\frac{h(u, x'\beta_1)^2}{\int_0^1 h(y, x'\beta_1)^2 dy}\right),$$

for $\theta = (\beta_1, \beta_2, h, \rho)$, where the log density of the nonlinear autoregressive copula is:

$$\log c(z_t, z_{t-1}; \varrho(\cdot)) = \log\phi[\Lambda^{-1}(z_t) - \varrho(z_{t-1})] - \log\lambda[\Lambda^{-1}(z_t)], \qquad \text{(b.10)}$$

parameterized by the univariate function $\varrho(\cdot)$, with $\Lambda(y) = \int_0^1 \Phi(y - \varrho(v))dv$ and $\lambda(y) = \int_0^1 \phi(y - \varrho(v))dv$. The directional derivative of the log transition density is

$$\frac{\partial \log l(u|u_{-1}, x, x_{-1}; \theta_0)}{\partial \theta}[v] = \frac{\partial \log c(Z(\theta_0), Z_{-1}(\theta_0); \rho_0(\cdot, x'\beta_2^0))}{\partial z_t} \frac{\partial Z(\theta_0)}{\partial \theta}[v]$$
$$+ \frac{\partial \log c(Z(\theta_0), Z_{-1}(\theta_0); \rho_0(\cdot, x'\beta_2^0))}{\partial z_{t-1}} \frac{\partial Z_{-1}(\theta_0)}{\partial \theta}[v]$$
$$+ \frac{\partial \log c(Z(\theta_0), Z_{-1}(\theta_0); \rho_0(\cdot, x'\beta_2^0))}{\partial \varrho}[v_\rho(\cdot, x'\beta_2^0) + \nabla_2 \rho_0(\cdot, x'\beta_2^0)x'v_{\beta_2}]$$
$$+2\frac{v_h(u, x'\beta_1^0) + \nabla_2 h^0(u, x'\beta_1^0)x'v_{\beta_1}}{h^0(u, x'\beta_1^0)} - 2\frac{\int_0^1 h^0(y, x'\beta_1^0)[v_h(y, x'\beta_1) + \nabla_2 h^0(y, x'\beta_1^0)x'v_{\beta_1}]dy}{\int_0^1 h^0(y, x'\beta_1^0)^2 dy}, \quad \text{(b.11)}$$

for $v = (v_{\beta_1}, v_{\beta_2}, v_h, v_\rho) \in \mathbb{V}$, where $Z(\theta) := \int_0^u h(y, x'\beta_1)^2 dy / \int_0^1 h(y, x'\beta_1)^2 dy$ and

$$\begin{aligned} \frac{\partial Z(\theta_0)}{\partial \theta}[v] &= 2\frac{\int_0^1 h^0(y, x'\beta_1^0)[\mathbf{1}(y \leq u) - Z(\theta_0)][v_h(y, x'\beta_1^0) + \nabla_2 h^0(y, x'\beta_1^0)x'v_{\beta_1}]dy}{\int_0^1 h^0(y, x'\beta_1^0)^2 dy} \\ &= 2Cov\left(\mathbf{1}(U_{i,t} \leq u), \frac{v_h(U_{i,t}, x'\beta_1^0) + \nabla_2 h^0(U_{i,t}, x'\beta_1^0)x'v_{\beta_1}}{h^0(U_{i,t}, x'\beta_1^0)}|X_{i,t} = x\right), \quad \text{(b.12)} \end{aligned}$$

and similarly for $Z_{-1}(\theta) = \int_0^{u_{-1}} h(y, x'\beta_1)^2 dy / \int_0^1 h(y, x'\beta_1)^2 dy$ and its directional derivative.

Lemma 7. *The partial derivatives w.r.t. the copula arguments are:*

$$\frac{\partial \log c(z_t, z_{t-1}; \varrho_0)}{\partial z_t} = \frac{1}{\lambda_0[\Lambda_0^{-1}(z_t)]} (\varrho_0(z_{t-1}) - E[\varrho_0(Z_{t-1})|Z_t = z_t]]), \quad \text{(b.13)}$$

and

$$\frac{\partial \log c(z_t, z_{t-1}; \varrho_0)}{\partial z_{t-1}} = [\Lambda_0^{-1}(z_t) - \varrho_0(z_{t-1})]\varrho_0'(z_{t-1}),$$

where the conditional expectation in (b.13) is w.r.t. variables (Z_t, Z_{t-1}) which are uniformly distributed on $[0, 1]$ with copula $c(\cdot, \cdot; \varrho_0)$. The directional derivative w.r.t. the functional copula parameter is:

$$\begin{aligned} \frac{\partial \log c(z_t, z_{t-1}; \varrho_0)}{\partial \varrho}[v] &= [\Lambda_0^{-1}(z_t) - \varrho_0(z_{t-1})] (v(z_{t-1}) - E[v(Z_{t-1})|Z_t = z_t)]) \\ &\quad +Cov (\varrho_0(Z_{t-1}), v(Z_{t-1})|Z_t = z_t). \end{aligned} \quad \text{(b.14)}$$

By using equations (b.11) and (b.12), Lemma 7, the equalities $Z(\theta_0) = G_0(u|w_1) = z$ and $Z_{-1}(\theta_0) = z_{-1}$, and the fact that $Z_{i,t}, Z_{i,t-1}$ have copula $c(\cdot, \cdot; \rho_0(\cdot; X_{i,t}'\beta_2^0))$ conditional on $X_{i,t}, X_{i,t-1}$, the conclusion follows.

b) To compute the directional derivative of $\gamma(\theta)$, we write $\gamma(\theta) = \gamma_1(\theta)\gamma_2(\theta)\gamma_3(\theta)$ where

$$\begin{aligned} \gamma_1(\theta) &= \frac{g(\bar{u}; \bar{x}'\beta_1)}{g[G^{-1}[\Lambda[\rho[G(\bar{u}; \bar{x}'\beta_1); \bar{x}'\beta_2]; \bar{x}'\beta_2]; \bar{x}'\beta_1]; \bar{x}'\beta_1]} \equiv \psi(\zeta_1(\theta); \theta)^2, \\ \gamma_2(\theta) &= \lambda[\rho[G(\bar{u}; \bar{x}'\beta_1); \bar{x}'\beta_2]; \bar{x}'\beta_2] \equiv \lambda(\zeta_3(\theta); \bar{x}'\beta_2), \\ \gamma_3(\theta) &= \nabla_1 \rho[G(\bar{u}; \bar{x}'\beta_1); \bar{x}'\beta_2] \equiv \nabla_1 \rho(\zeta_4(\theta); \bar{x}'\beta_2), \end{aligned}$$

with the notation $\psi(y;\theta) = \frac{h(\bar{u};\bar{x}'\beta_1)}{h(y;\bar{x}'\beta_1)}$, $\zeta_1(\theta) = G^{-1}(\zeta_2(\theta);\bar{x}'\beta_1)$, $\zeta_2(\theta) = \Lambda(\zeta_3(\theta);\bar{x}'\beta_2)$, $\zeta_3(\theta) = \rho(\zeta_4(\theta);\bar{x}'\beta_2)$, $\zeta_4(\theta) = G(\bar{u};\bar{x}'\beta_1) = \frac{\int_0^{\bar{u}} h(u,\bar{x}'\beta_1)^2 du}{\int_0^1 h(u,\bar{x}'\beta_1)^2 du}$. By the product rule:

$$\frac{\partial\gamma(\theta)}{\partial\theta}[v] \quad = \quad \gamma(\theta)\sum_{i=1}^{3}\frac{1}{\gamma_i(\theta)}\frac{\partial\gamma_i(\theta)}{\partial\theta}[v], \tag{b.15}$$

where $v = (v_{\beta_1}, v_{\beta_2}, v_h, v_\rho) \in \mathbb{V}$. We compute now separately the three directional derivatives on the RHS:

$$\begin{aligned}
\frac{\partial\gamma_1(\theta)}{\partial\theta}[v] &= 2\psi(\zeta_1(\theta);\theta)\left(\frac{\partial\psi(y;\theta)}{\partial\theta}[v]\Big|_{y=\zeta_1(\theta)} + \nabla_1\psi(\zeta_1(\theta);\theta)\frac{\partial\zeta_1(\theta)}{\partial\theta}[v]\right) \\
&= 2\gamma_1(\theta)\left(\frac{v_h(\bar{u},\bar{x}'\beta_1) + \nabla_2 h(\bar{u},\bar{x}'\beta_1)\bar{x}'v_{\beta_1}}{h(\bar{u},\bar{x}'\beta_1)} - \frac{v_h(\zeta_1(\theta);\bar{x}'\beta_1) + \nabla_2 h(\zeta_1(\theta),\bar{x}'\beta_1)\bar{x}'v_{\beta_1}}{h(\zeta_1(\theta),\bar{x}'\beta_1)}\right. \\
&\quad \left. -\frac{\nabla_1 h(\zeta_1(\theta);\bar{x}'\beta_1)}{h(\zeta_1(\theta),\bar{x}'\beta_1)}\frac{\partial\zeta_1(\theta)}{\partial\theta}[v]\right),
\end{aligned}$$

$$\begin{aligned}
\frac{\partial\gamma_2(\theta)}{\partial\theta}[v] &= \nabla_1\lambda(\zeta_3(\theta);\bar{x}'\beta_2)\frac{\partial\zeta_3(\theta)}{\partial\theta}[v] + \frac{\partial\lambda(y;\bar{x}'\beta_2)}{\partial\theta}[v]\Big|_{y=\zeta_3(\theta)} \\
&= \int_0^1 \phi'[\zeta_3(\theta) - \rho(z;\bar{x}'\beta_2)]dz\frac{\partial\zeta_3(\theta)}{\partial\theta}[v] \\
&\quad -\int_0^1 \phi'[\zeta_3(\theta) - \rho(z;\bar{x}'\beta_2)][v_\rho(z;\bar{x}'\beta_2) + \nabla_2\rho(z;\bar{x}'\beta_2)\bar{x}'v_{\beta_2}]dz \\
&= \gamma_2(\theta)\Big\{E\left[\rho(Z_{i,t-1};\bar{x}'\beta_2) - \rho(\zeta_4(\theta),\bar{x}'\beta_2)|Z_{i,t} = \zeta_2(\theta), X_{i,t} = \bar{x}, X_{i,t-1} = \bar{x}\right]\frac{\partial\zeta_3(\theta)}{\partial\theta}[v]
\end{aligned}$$

$$-E\left[\left(\rho(Z_{i,t-1};\bar{x}'\beta_2) - \rho(\zeta_4(\theta),\bar{x}'\beta_2)\right)\left(v_\rho(Z_{i,t-1};\bar{x}'\beta_2) + \nabla_2\rho(Z_{i,t-1};\bar{x}'\beta_2)\bar{x}'v_{\beta_2}\right)|Z_{i,t} = \zeta_2(\theta), X_{i,t} = \bar{x}, X_{i,t-1} = \bar{x}\right]\Big\},$$

where we use $\frac{1}{\lambda(\zeta_3(\theta);\bar{x}'\beta_2)}\int_0^1 \phi[\zeta_3(\theta) - \rho(z;\bar{x}'\beta_2)]\psi(z)dz = E[\psi(Z_{i,t-1})|Z_{i,t} = \zeta_2(\theta), X_{i,t} = \bar{x}, X_{i,t-1} = \bar{x}]$, and:

$$\frac{\partial\gamma_3(\theta)}{\partial\theta}[v] \quad = \quad \nabla_1^2\rho(\zeta_4(\theta);\bar{x}'\beta_2)\frac{\partial\zeta_4(\theta)}{\partial\theta}[v] + \nabla_1 v_\rho(\zeta_4(\theta);\bar{x}'\beta_2) + \nabla_{1,2}^2\rho(\zeta_4(\theta);\bar{x}'\beta_2)\bar{x}'v_{\beta_2}.$$

The directional derivatives of $\zeta_1(\theta), ..., \zeta_4(\theta)$ are:

$$\begin{aligned}
\frac{\partial \zeta_1(\theta)}{\partial \theta}[v] &= \frac{1}{g[G^{-1}(\zeta_2(\theta);\bar{x}'\beta_1);\bar{x}'\beta_1]}\frac{\partial \zeta_2(\theta)}{\partial \theta}[v] + \frac{\partial G^{-1}(z;\bar{x}'\beta_1)}{\partial \theta}[v]\Big|_{z=\zeta_2(\theta)} \\
&= \frac{1}{g[\zeta_1(\theta);\bar{x}'\beta_1]}\left(\frac{\partial \zeta_2(\theta)}{\partial \theta}[v] - \frac{\partial G(y;\bar{x}'\beta_1)}{\partial \theta}[v]\Big|_{y=\zeta_1(\theta)}\right) \\
&= \frac{1}{g[\zeta_1(\theta);\bar{x}'\beta_1]}\left(\frac{\partial \zeta_2(\theta)}{\partial \theta}[v]\right. \\
&\quad \left. -2\frac{\int_0^1 h(u,\bar{x}'\beta_1)[\mathbf{1}(u \le \zeta_1(\theta)) - G(\zeta_1(\theta);\bar{x}'\beta_1)][v_h(u,\bar{x}'\beta_1) + \nabla_2 h(u,\bar{x}'\beta_1)\bar{x}'v_{\beta_1}]du}{\int_0^1 h(u,\bar{x}'\beta_1)^2 du}\right) \\
&= \frac{1}{g[\zeta_1(\theta);\bar{x}'\beta_1]}\left[\frac{\partial \zeta_2(\theta)}{\partial \theta}[v]\right. \\
&\quad \left. -2Cov\left(\mathbf{1}(U_{i,t} \le \zeta_1(\theta)), \frac{v_h(U_{i,t},\bar{x}'\beta_1) + \nabla_2 h(U_{i,t},\bar{x}'\beta_1)\bar{x}'v_{\beta_1}}{h(U_{i,t},\bar{x}'\beta_1)}|X_{i,t} = \bar{x}\right)\right],
\end{aligned}$$

where we use $\frac{\partial G^{-1}(z;\bar{x}'\beta_1)}{\partial \theta}[v] = -\frac{1}{g[G^{-1}(z;\bar{x}'\beta_1);\bar{x}'\beta_1]}\ \frac{\partial G(y;\bar{x}'\beta_1)}{\partial \theta}[v]\Big|_{y=G^{-1}(z;\bar{x}'\beta_1)}$, and a calculation similar to (b.12),

$$\begin{aligned}
\frac{\partial \zeta_2(\theta)}{\partial \theta}[v] &= \lambda(\zeta_3(\theta);\bar{x}'\beta_2)\frac{\partial \zeta_3(\theta)}{\partial \theta}[v] + \frac{\partial \Lambda(y;\bar{x}'\beta_2)}{\partial \theta}[v]\Big|_{y=\zeta_3(\theta)} \\
&= \lambda(\zeta_3(\theta);\bar{x}'\beta_2)\frac{\partial \zeta_3(\theta)}{\partial \theta}[v] - \int_0^1 \phi[\zeta_3(\theta) - \rho(z;\bar{x}'\beta_2)][v_\rho(z,\bar{x}'\beta_2) + \nabla_2\rho(z,\bar{x}'\beta_2)\bar{x}'v_{\beta_2}]dz \\
&= \lambda(\zeta_3(\theta);\bar{x}'\beta_2)\left(\frac{\partial \zeta_3(\theta)}{\partial \theta}[v]\right. \\
&\quad \left. -E\left[v_\rho(Z_{i,t-1};\bar{x}'\beta_2) + \nabla_2\rho(Z_{i,t-1};\bar{x}'\beta_2)\bar{x}'v_{\beta_2}|Z_{i,t} = \zeta_2(\theta), X_{i,t} = \bar{x}, X_{i,t-1} = \bar{x}\right]\right),
\end{aligned}$$

$$\frac{\partial \zeta_3(\theta)}{\partial \theta}[v] = \nabla_1\rho(\zeta_4(\theta),\bar{x}'\beta_2)\frac{\partial \zeta_4(\theta)}{\partial \theta}[v] + v_\rho(\zeta_4(\theta);\bar{x}'\beta_2) + \nabla_2\rho(\zeta_4(\theta),\bar{x}'\beta_2)\bar{x}'v_{\beta_2},$$

and

$$\begin{aligned}
\frac{\partial \zeta_4(\theta)}{\partial \theta}[v] &= 2\frac{\int_0^1 h(u,\bar{x}'\beta_1)[\mathbf{1}(u \le \bar{u}) - G(\bar{u};\bar{x}'\beta_1)][v_h(u,\bar{x}'\beta_1) + \nabla_2 h(u,\bar{x}'\beta_1)\bar{x}'v_{\beta_1}]du}{\int_0^1 h(u,\bar{x}'\beta_1)^2 du} \\
&= 2Cov\left(\mathbf{1}(U_{i,t} \le \bar{u}), \frac{v_h(U_{i,t},\bar{x}'\beta_1) + \nabla_2 h(U_{i,t},\bar{x}'\beta_1)\bar{x}'v_{\beta_1}}{h(U_{i,t},\bar{x}'\beta_1)}|X_{i,t} = \bar{x}\right).
\end{aligned}$$

By combining these results we have:

$$\frac{1}{\gamma_1(\theta)}\frac{\partial\gamma_1(\theta)}{\partial\theta}[v] = 2\left(\frac{v_h(\bar{u},\bar{x}'\beta_1)+\nabla_2 h(\bar{u},\bar{x}'\beta_1)\bar{x}'v_{\beta_1}}{h(\bar{u},\bar{x}'\beta_1)} - \frac{v_h(\zeta_1(\theta),\bar{x}'\beta_1)+\nabla_2 h(\zeta_1(\theta),\bar{x}'\beta_1)\bar{x}'v_{\beta_1}}{h(\zeta_1(\theta),\bar{x}'\beta_1)}\right)$$
$$-2\frac{\nabla_1 h(\zeta_1(\theta),\bar{x}'\beta_1)}{h(\zeta_1(\theta),\bar{x}'\beta_1)}\frac{\lambda(\zeta_3(\theta);\bar{x}'\beta_2)}{g(\zeta_1(\theta),\bar{x}'\beta_1)}\big(v_\rho(\zeta_4(\theta),\bar{x}'\beta_2)+\nabla_2\rho(\zeta_4(\theta),\bar{x}'\beta_2)\bar{x}'v_{\beta_2}$$
$$-E\left[v_\rho(Z_{i,t-1},\bar{x}'\beta_2)+\nabla_2\rho(Z_{i,t-1},\bar{x}'\beta_2)\bar{x}'v_{\beta_2}|Z_{i,t}=\zeta_2(\theta),X_{i,t}=\bar{x},X_{i,t-1}=\bar{x}\right]\big)$$
$$+4\frac{\nabla_1 h(\zeta_1(\theta),\bar{x}'\beta_1)}{h(\zeta_1(\theta),\bar{x}'\beta_1)}\frac{1}{g(\zeta_1(\theta),\bar{x}'\beta_1)}Cov\left(\mathbf{1}(U_{i,t}\le\zeta_1(\theta)),\frac{v_h(U_{i,t},\bar{x}'\beta_1)+\nabla_2 h(U_{i,t},\bar{x}'\beta_1)\bar{x}'v_{\beta_1}}{h(U_{i,t},\bar{x}'\beta_1)}|X_{i,t}=\bar{x}\right)$$
$$-4\frac{\nabla_1 h(\zeta_1(\theta),\bar{x}'\beta_1)}{h(\zeta_1(\theta),\bar{x}'\beta_1)}\frac{\lambda(\zeta_3(\theta);\bar{x}'\beta_2)}{g(\zeta_1(\theta),\bar{x}'\beta_1)}\nabla_1\rho(\zeta_4(\theta),\bar{x}'\beta_2)$$
$$\times Cov\left(\mathbf{1}(U_{i,t}\le\bar{u}),\frac{v_h(U_{i,t},\bar{x}'\beta_1)+\nabla_2 h(U_{i,t},\bar{x}'\beta_1)\bar{x}'v_{\beta_1}}{h(U_{i,t},\bar{x}'\beta_1)}|X_{i,t}=\bar{x}\right),$$

$$\frac{1}{\gamma_2(\theta)}\frac{\partial\gamma_2(\theta)}{\partial\theta}[v] = E\left[\rho(Z_{i,t-1};\bar{x}'\beta_2)-\rho(\zeta_4(\theta),\bar{x}'\beta_2)|Z_{i,t}=\zeta_2(\theta),X_{i,t}=\bar{x},X_{i,t-1}=\bar{x}\right]$$
$$\times\big(v_\rho(\zeta_4(\theta);\bar{x}'\beta_2)+\nabla_2\rho(\zeta_4(\theta),\bar{x}'\beta_2)\bar{x}'v_{\beta_2}$$
$$-E\left[v_\rho(Z_{i,t-1};\bar{x}'\beta_2)+\nabla_2\rho(Z_{i,t-1},\bar{x}'\beta_2)\bar{x}'v_{\beta_2}|Z_{i,t}=\zeta_2(\theta),X_{i,t}=\bar{x},X_{i,t-1}=\bar{x}\right]\big)$$
$$-Cov\big(\rho(Z_{i,t-1};\bar{x}'\beta_2)-\rho(\zeta_4(\theta),\bar{x}'\beta_2),v_\rho(Z_{i,t-1};\bar{x}'\beta_2)+\nabla_2\rho(Z_{i,t-1};\bar{x}'\beta_2)\bar{x}'v_{\beta_2}|Z_{i,t}=\zeta_2(\theta),X_{i,t}=\bar{x},X_{i,t-1}=\bar{x}\big)$$
$$+2E\left[\rho(Z_{i,t-1};\bar{x}'\beta_2)-\rho(\zeta_4(\theta),\bar{x}'\beta_2)|Z_{i,t}=\zeta_2(\theta),X_{i,t}=\bar{x},X_{i,t-1}=\bar{x}\right]\nabla_1\rho(\zeta_4(\theta),\bar{x}'\beta_2)$$
$$\times Cov\left(\mathbf{1}(U_{i,t}\le\bar{u}),\frac{v_h(U_{i,t},\bar{x}'\beta_1)+\nabla_2 h(U_{i,t},\bar{x}'\beta_1)\bar{x}'v_{\beta_1}}{h(U_{i,t},\bar{x}'\beta_1)}|X_{i,t}=\bar{x}\right),$$

and:

$$\frac{1}{\gamma_3(\theta)}\frac{\partial\gamma_3(\theta)}{\partial\theta}[v] = 2\frac{\nabla_1^2\rho(\zeta_4(\theta);\bar{x}'\beta_2)}{\nabla_1\rho(\zeta_4(\theta);\bar{x}'\beta_2)}Cov\left(\mathbf{1}(U_{i,t}\le\bar{u}),\frac{v_h(U_{i,t},\bar{x}'\beta_1)+\nabla_2 h(U_{i,t},\bar{x}'\beta_1)\bar{x}'v_{\beta_1}}{h(U_{i,t},\bar{x}'\beta_1)}|X_{i,t}=\bar{x}\right)$$
$$+\frac{\nabla_1 v_\rho(\zeta_4(\theta);\bar{x}'\beta_2)+\nabla_{1,2}^2\rho(\zeta_4(\theta);\bar{x}'\beta_2)\bar{x}'v_{\beta_2}}{\nabla_1\rho(\zeta_4(\theta);\bar{x}'\beta_2)}.$$

By setting $\theta=\theta_0$ and using equation (b.15), the conclusion follows.

Appendix C: Numerical implementation

In this Appendix we provide details about the implementation of our Sieve estimator, the characterization of the bases of the “tangent spaces" $\mathbb{V}^h_{NT}$, $\mathbb{V}^\rho_{NT}$, and the analytical bias correction.

C.1 Sieve spaces based on tensor Hermite polynomials

The Hermite polynomials are defined by $H_k(x) = (-1)^k e^{x^2}\frac{d^k}{dx^k}e^{-x^2}$, for real argument x and integer $k = 0, 1, ...$ We use the Hermite polynomials to define functions $\varphi_k(u) = \frac{1}{\sqrt{2^k k!}}H_k[\frac{\Phi^{-1}(u)}{\sqrt{2}}]$, which build a complete orthonormal basis of $L^2[0,1]$ w.r.t. scalar product $\int_0^1 f(u)g(u)du$. Moreover, functions $\psi_k(w) = \frac{1}{\sqrt{2^k k! \sqrt{\pi}}} e^{-w^2/2} H_k(w)$ build a complete orthonormal basis of $L^2(\mathbb{R})$ w.r.t. scalar product $\int_{-\infty}^{\infty} f(w)g(w)dw$, and functions $\tilde{\psi}_k(w) = \frac{1}{\sqrt{2^k k! \sqrt{\pi}}} H_k(w)$ build a complete orthonormal basis of $L^2(\mathbb{R}, q)$ w.r.t. scalar product $\int_{-\infty}^{\infty} f(w)g(w)q(w)dw$, for $q(w) = e^{-w^2}$.

For the sake of concreteness, we focus on the case with $m^h = m^\rho = 2$ and $u^* = 0.5$, as in our empirical analysis and Monte Carlo study. Then, we have $\varphi_0(u) = 1$, $\varphi_1(u) = \Phi^{-1}(u)$, $\varphi_2(u) = \frac{1}{\sqrt{2}}([\Phi^{-1}(u)]^2 - 1)$; $\psi_0(w) = \frac{1}{\pi^{1/4}} e^{-w^2/2}$, $\psi_1(w) = \frac{\sqrt{2}}{\pi^{1/4}} w e^{-w^2/2}$, $\psi_2(w) = \frac{1}{\sqrt{2}\pi^{1/4}}(2w^2 - 1)e^{-w^2/2}$; and $\tilde{\psi}_0(w) = \frac{1}{\pi^{1/4}}$, $\tilde{\psi}_1(w) = \frac{\sqrt{2}}{\pi^{1/4}} w$, $\tilde{\psi}_2(w) = \frac{1}{\sqrt{2}\pi^{1/4}}(2w^2 - 1)$. We define $\varphi(u) = (\varphi_0(u), \varphi_1(u), \varphi_2(u))'$, $\psi(w) = (\psi_0(w), \psi_1(w), \psi_2(w))'$ and $\tilde{\psi}(w) = (\tilde{\psi}_0(w), \tilde{\psi}_1(w), \tilde{\psi}_2(w))'$. By reworking the constraints on the coefficients, the Sieve spaces $\mathcal{H}^h_{N,T}$, $\mathcal{H}^\rho_{N,T}$ can be rewritten as follows.

Lemma 8. *For $m^h = m^\rho = 2$ and $u^* = 0.5$, the Sieve spaces defined in Section 3 become:*

$$
\begin{aligned}
\mathcal{H}^h_{N,T} &= \left\{ h(\cdot,\cdot) \ : \ \lambda'[\varphi(u) \otimes \psi(w)], \ \lambda \in \mathbb{R}^9, \ s.t. \ \lambda'[\kappa^{(l)} \otimes I_3]\lambda = \frac{1}{1+l}, \ l = 0,1,2 \right\}, \\
\mathcal{H}^\rho_{N,T} &= \left\{ \rho(\cdot,\cdot) \ : \ \mu' A^{\rho\prime}[\varphi(u) \otimes \tilde{\psi}(w)], \ \mu \in \mathbb{R}^6 \right\},
\end{aligned}
$$

where symmetric matrices $\kappa^{(l)} := \int_0^1 u^l \varphi(u)\varphi(u)' du$ for $l = 0,1,2$ are given by:

$$
\kappa^{(0)} = I_3, \qquad \kappa^{(1)} = \frac{1}{2} \begin{pmatrix} 1 & \frac{1}{\sqrt{\pi}} & 0 \\ & 1 & \frac{3}{2\sqrt{2\pi}} \\ & & 1 \end{pmatrix}, \qquad \kappa^{(2)} = \frac{1}{3} \begin{pmatrix} 1 & \frac{3}{2\sqrt{\pi}} & \frac{\sqrt{3}}{2\sqrt{2\pi}} \\ & 1 + \frac{\sqrt{3}}{2\pi} & \frac{9}{4\sqrt{2\pi}} \\ & & 1 - \frac{7}{4\sqrt{3}\pi} \end{pmatrix}, \quad \text{(b.16)}
$$

and the 9×6 matrix A^ρ is given by $A^\rho = \begin{pmatrix} \frac{1}{\sqrt{3}} I_3 & 0_{3\times 3} \\ 0_{3\times 3} & I_3 \\ \frac{\sqrt{2}}{\sqrt{3}} I_3 & 0_{3\times 3} \end{pmatrix}$.

C.2 Numerical computation of the Sieve estimator $\hat{\theta}$

We compute the estimator $\hat{\theta}$ in (3.2) using the parameterization of the Sieve spaces in Lemma 8. The log-likelihood function involves numerical integrals to evaluate the function Λ in (2.22) and the marginal distribution. These numerical integrals are evaluated by simulations. Our choice $m^h = m^\rho = 2$ for the polynomial degrees in the Sieve spaces yields a good performance in the Monte Carlo experiments. We have verified that the estimation results are similar for other choices of m^h, m^ρ near our quadratic approximation.

C.3 Orthonormal basis functions for tangent spaces $\mathbb{V}^h_{N,T}$ and $\mathbb{V}^\rho_{N,T}$

The Sieve $\mathcal{H}^\rho_{N,T}$ is a linear space. Thus, the tangent space $\mathbb{V}^\rho_{N,T}$ coincides with the Sieve space $\mathcal{H}^\rho_{N,T}$ itself, and from Lemma 8 it is spanned by the six functions Ψ^ρ_l, $l = 1, ..., 6$ that are components of the vector $\Psi^\rho(u, v) = A^{\rho\prime}[\varphi(u) \otimes \tilde{\psi}(w)]$. Because $A^{\rho\prime} A^\rho = I_6$ and the orthonormal properties of functions φ_k and $\tilde{\psi}_k$, it follows that functions Ψ^ρ_l, $l = 1, ..., 6$ are orthonormal w.r.t. scalar product $\langle v_\rho, v^*_\rho\rangle = \int_0^1 \int_{-\infty}^\infty v_\rho(u, w)v^*_\rho(u, w)q(w)dudw$.
The Sieve $\mathcal{H}^h_{N,T}$ is not a linear space due to the constraints on the coefficient vector λ, and the tangent space $\mathbb{V}^h_{N,T}$ at $h^0_{N,T}$ is derived next by linearization.

Lemma 9. *The linear space $\mathbb{V}^h_{N,T}$ is spanned by the six functions Ψ^h_l, $l = 1, ..., 6$ that are components of the vector $\Psi^h(u, v) = A^{h\prime}[\varphi(u) \otimes \psi(w)]$, where the columns of the 9×6 matrix A^h are an orthonormal basis of the orthogonal complement of the range of matrix $B = [\lambda^0_{N,T} \ : \ (\kappa^{(1)} \otimes I_3)\lambda^0_{N,T} \ : \ (\kappa^{(2)} \otimes I_3)\lambda^0_{N,T}]$, where $\lambda^0_{N,T} \in \mathbb{R}^9$ is the coefficient vector of $h^0_{N,T}(u, w) = \lambda^{0\prime}_{N,T}[\varphi(u) \otimes \psi(w)]$.* [19]

Functions Ψ^h_l, $l = 1, ..., 6$ are orthonormal w.r.t. scalar product $\langle v_h, v^*_h\rangle = \int_0^1 \int_{-\infty}^\infty v_h(u, w)v^*_h(u, w)dudw$.

C.4 Closed-form analysis and Gaussian reference model for bias adjustment

To apply Theorem 2 for implementing an analytical bias correction as discussed in Section 3.3, we need certain conditional expectations involving the distribution of $U_{i,t}$ given $X_{i,t}$ for $h \in \mathcal{H}^h_{N,T}$ (specifically, for the Sieve estimate of h_0). They are provided in the next lemma.

Lemma 10. *For $h \in \mathcal{H}^h_{N,T}$ with $h = \lambda'[\varphi \otimes \psi]$, it holds:*

$$G(u, w_1) = E[1(U_{i,t} \le u)|X_{i,t} = x] = \frac{\lambda'[K(u) \otimes (\psi(w_1)\psi(w_1)')]\lambda}{\lambda'[I_3 \otimes (\psi(w_1)\psi(w_1)')]\lambda}, \tag{b.17}$$

$$E\left[1(U_{i,t} \le u)\frac{\nabla_2 h(U_{i,t}, w_1)}{h(U_{i,t}, w_1)}|X_{i,t} = x\right] = \frac{\lambda'[K(u) \otimes (\nabla\psi(w_1)\psi(w_1)')]\lambda}{\lambda'[I_3 \otimes (\psi(w_1)\psi(w_1)')]\lambda}, \tag{b.18}$$

$$E\left[1(U_{i,t} \le u)\frac{\Psi^h(U_{i,t}, w_1)}{h(U_{i,t}, w_1)}|X_{i,t} = x\right] = \frac{A^{h\prime}[K(u) \otimes (\psi(w_1)\psi(w_1)')]\lambda}{\lambda'[I_3 \otimes (\psi(w_1)\psi(w_1)')]\lambda}, \tag{b.19}$$

where

$$K(u) := \int_0^u \varphi(y)\varphi(y)'dy = uI_3 - \phi[\Phi^{-1}(u)] \begin{pmatrix} 0 & 1 & \frac{1}{\sqrt{2}}\Phi^{-1}(u) \\ & \Phi^{-1}(u) & \frac{1}{\sqrt{2}}([\Phi^{-1}(u)]^2 + 1) \\ & & -\frac{1}{2}[\Phi^{-1}(u)]([\Phi^{-1}(u)]^2 + 1) \end{pmatrix}. \tag{b.20}$$

To obtain simple closed-form expressions for other quantities needed for bias correction, we use the following approximations.

[19] For instance, we can take $A^h = \tilde{A}(\tilde{A}'\tilde{A})^{-1/2}$, where $\tilde{A}$ is the matrix built by the first 6 columns of $I_9 - B(B'B)^{-1}B$.

C.4.1 Legendre polynomial approximation of higher-order derivatives

The vector b_{NT} has elements $\mathcal{B}_T(e_l)$, for $l = 1, ..., M$. The latter can be estimated as

$$\begin{aligned} \tilde{\mathcal{B}}_T(e_l) &= \frac{1}{NT}\sum_i \sum_t \left[\frac{\partial^2 \log l_{i,t}(\theta_0)}{\partial\theta\partial u}[e_l]H_t(\varepsilon_{i,t}) + \frac{\partial^2 \log l_{i,t}(\theta_0)}{\partial\theta\partial v}[e_l]H_{t-1}(\varepsilon_{i,t-1})\right] \\ &\quad -\frac{1}{NT}\sum_i \sum_{t,s} \left[\frac{\partial^2 \log l_{i,t}(\theta_0)}{\partial\theta\partial u}[e_l]f_t(\varepsilon_{i,t}) + \frac{\partial^2 \log l_{i,t}(\theta_0)}{\partial\theta\partial v}[e_l]f_{t-1}(\varepsilon_{i,t-1})\right]\varepsilon_{i,s} \\ &\quad +\frac{\omega^2}{2NT}\sum_i \sum_t \Psi(\varepsilon_{i,t}, \varepsilon_{i,t-1}; \theta_0)[e_l]. \end{aligned}$$

We can dispense of the cumbursome expressions of the first- and second-order derivatives of $\frac{\partial^2 \log l(u|v, X_{i,t}, X_{i,t-1};\theta_0)}{\partial\theta}[e_l]$ w.r.t. the copula arguments u and v by deploying a polynomial approximation. Using the Legendre polynomials $L_k(x) = (2^k k!)^{-1}\frac{d^k}{dx^k}[(x^2-1)^k]$ for argument $x \in [-1, 1]$ and integer $k = 0, 1, ...$, we define the polynomials $P_k(u) = \sqrt{2k+1}L_k(2u-1)$, that are orthonormal functions on $[0, 1]$ w.r.t. the standard L^2 scalar product. We use the series expansion with orthogonal polynomials:

$$\frac{\partial \log l_{i,t}(\theta_0)}{\partial\theta}[e_l] = \sum_{k,\ell=1}^{L} \alpha_{i,t,k,\ell}P_k(U_{i,t})P_\ell(U_{i,t-1}), \tag{b.21}$$

for a large integer L, where $\alpha_{i,t,j,k} = \int_0^1\int_0^1 \frac{\partial \log l(u|u_{-1}, X_{i,t}, X_{i,t-1})}{\partial\theta}[e_l]P_k(u)P_\ell(u_{-1})du du_{-1}$, (for expository purpose we omit the dependence of the coefficients α on index l). Then we get:

$$\begin{aligned} \tilde{\mathcal{B}}_T(e_l) &= \frac{1}{NT}\sum_i\sum_t\sum_{k,\ell}\alpha_{i,t,k,\ell}\left[P_k'(U_{i,t})P_\ell(U_{i,t-1})H_t(\varepsilon_{i,t}) + P_k(U_{i,t})P_\ell'(U_{i,t-1})H_{t-1}(\varepsilon_{i,t-1})\right] \\ &\quad -\frac{1}{NT}\sum_i\sum_{t,s}\sum_{k,\ell}\alpha_{i,t,k,\ell}\left[P_k'(U_{i,t})P_\ell(U_{i,t-1})f_t(\varepsilon_{i,t}) + P_k(U_{i,t})P_\ell'(U_{i,t-1})f_{t-1}(\varepsilon_{i,t-1})\right]\varepsilon_{i,s} \\ &\quad +\frac{\omega^2}{2NT}\sum_i\sum_t\sum_{k,\ell}\alpha_{i,t,k,\ell}\Big[P_k''(U_{i,t})P_\ell(U_{i,t-1})[f_t(\varepsilon_{i,t})]^2 + P_k(U_{i,t})P_\ell''(U_{i,t-1})[f_{t-1}(\varepsilon_{i,t-1})]^2 \\ &\quad +2P_k'(U_{i,t})P_\ell'(U_{i,t-1})f_t(\varepsilon_{i,t})f_{t-1}(\varepsilon_{i,t-1})\Big]. \end{aligned} \tag{b.22}$$

By interchanging summation and integration, the bivariate numerical integrals on the RHS of (b.22) can be computed once for all i, t, k, ℓ (e.g. by Monte Carlo simulation). A feasible estimator involves using residuals $\hat{\varepsilon}_{i,t}$ and consistent (nonparametric) estimators of ω^2, f_t, H_t.

C.4.2 Gaussian reference model

For $\rho = \mu' A^{\rho\prime}[\varphi(\cdot)\otimes\tilde{\psi}(\cdot)] \in \mathcal{H}^\rho_{N,T}$ with $\mu \in \mathbb{R}^6$ and $\mu_1 = \mu_2 = \mu_3 = 0$, we have $\rho(u, w) = r(w)\Phi^{-1}(u)$ with $r(w) = \mu_4\tilde{\psi}_0(w) + \mu_5\tilde{\psi}_1(w) + \mu_6\tilde{\psi}_2(w) = \frac{1}{\pi^{1/4}}(\mu_4 + \sqrt{2}\mu_5 w + \frac{\mu_6}{\sqrt{2}}(2w^2-1))$. Then from (2.22) we have (see the proof of Lemma 11):

$$\Lambda(y,w) \quad = \quad \Phi\left(\frac{y}{\sqrt{1+r(w)^2}}\right), \tag{b.23}$$

From (2.27) we get:

$$\Phi^{-1}(Z_{i,t}) = \frac{r(X_{i,t}'\beta_2^0)}{\sqrt{1+r(X_{i,t}'\beta_2^0)^2}}\Phi^{-1}(Z_{i,t-1}) + \frac{1}{\sqrt{1+r(X_{i,t}'\beta_2^0)^2}}\omega_{i,t},$$

with $\omega_{i,t} \sim IIN(0,1)$. Thus, $Z_{i,t}, Z_{i,t-1}$ conditionally on $X_{i,t}, X_{i,t-1}$ have a Gaussian copula with correlation parameter $\frac{r(X_{i,t}'\beta_2^0)}{\sqrt{1+r(X_{i,t}'\beta_2^0)^2}}$. The Gaussian copula allows to get closed-form expressions for the conditional expectations in Theorem 2, that are provided in the next lemma.

Lemma 11. *For $\rho = \mu' A^{\rho\prime}[\varphi(\cdot) \otimes \tilde{\psi}(\cdot)] \in \mathcal{H}_{N,T}^{\rho}$ with $\mu \in \mathbb{R}^6$ and $\mu_1 = \mu_2 = \mu_3 = 0$, we have:*

$$\begin{aligned}
E[\rho(Z_{i,t-1}; w_2)|\Omega(z,x,x_{-1})] &= \frac{r(w_2)^2}{\sqrt{1+r(w_2)^2}}\Phi^{-1}(z), && \text{(b.24)}\\
E[\nabla_2\rho(Z_{i,t-1}; w_2)|\Omega(z,x,x_{-1})] &= \frac{r(w_2)\nabla r(w_2)}{\sqrt{1+r(w_2)^2}}\Phi^{-1}(z), && \text{(b.25)}\\
Cov(\rho(Z_{i,t-1}; w_2), \nabla_2\rho(Z_{i,t-1}; w_2)|\Omega(z,x,x_{-1})] &= \frac{r(w_2)\nabla r(w_2)}{1+r(w_2)^2}, && \text{(b.26)}\\
E[\Psi^{\rho}(Z_{i,t-1}; w_2)|\Omega(z,x,x_{-1})] &= A^{\rho\prime}[(D(w_2)\varphi(z)) \otimes \tilde{\psi}(w_2)], && \text{(b.27)}\\
Cov(\rho(Z_{i,t-1}; w_2), \Psi^{\rho}(Z_{i,t-1}; w_2)|\Omega(z,x,x_{-1})] &= A^{\rho\prime}[(F(w_2)\varphi(z)) \otimes \tilde{\psi}(w_2)], && \text{(b.28)}
\end{aligned}$$

where the conditioning set is $\Omega(z,x,x_{-1}) := \{Z_{i,t} = z, X_{i,t} = x, X_{i,t-1} = x_{-1}\}$, and $D(w) = diag(1, \frac{r(w)}{\sqrt{1+r(w)^2}}, \frac{r(w)^2}{1+r(w)^2})$ and $F(w) = \frac{r(w)}{1+r(w)^2}\begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & \sqrt{2}\frac{r(w)}{\sqrt{1+r(w)^2}} & 0 \end{pmatrix}$.

Let us finally consider a reference model in which the conditional distribution of $U_{i,t}$ given $X_{i,t}$ is independent of $X_{i,t}$ and uniform, i.e. $G(u,w) = u$, and the cross-sectional density f_t is Gaussian with variance σ_t^2. Then, the errors $\varepsilon_{i,t}$ themselves have a Gaussian dynamics conditionally on the regressors. Then, we have

$$\sum_{s=1}^{T} E(\varepsilon_{i,s}|\varepsilon_{i,t} = \varepsilon) \simeq \sum_{s} \frac{\sigma_{st}}{\sigma_t^2}\varepsilon \simeq \frac{\omega^2}{\sigma_t^2}\varepsilon,$$

for large T. Then, by using $f_t'(\varepsilon) = -\frac{1}{\sigma_t^2}\varepsilon f_t(\varepsilon)$, we get $H_t(\varepsilon) = 0$ asymptotically. Hence, we can omit the terms in (b.22) involving the H_t.

References

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

2. Ai, C., & Chen, X. (2003). Efficient estimation of models with conditional moment restrictions containing unknown functions. Econometrica, 71(6), 1795-1843.

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

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

5. 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.

6. 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.

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

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

9. 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.

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. 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.

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

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

16. 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.

17. Chen, X., Huang, Z., & Yi, Y. (2021). Efficient estimation of multivariate semi-nonparametric GARCH filtered copula models. Journal of Econometrics, 222(1), 484-501.

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

19. Chen, X., Liao, Z., & Sun, Y. (2014). Sieve inference on possibly misspecified semi-nonparametric time series models. Journal of Econometrics, 178, 639-658.

20. Chen, X., & Liao, Z. (2014). Sieve M inference on irregular parameters. Journal of Econometrics, 182(1), 70-86.

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

22. Chen, X., Xiao, Z., & Wang, B. (2020). Copula-based time series with filtered nonstationarity. Journal of Econometrics, 228, 127-155.

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

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

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

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

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

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

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

35. 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.

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

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

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

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

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

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

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

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

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

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

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

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

48. Hu, Y., Moffitt, R., & Sasaki, Y. (2019). Semiparametric estimation of the canonical permanent-transitory model of earnings dynamics. Quantitative Economics, 10 (4), 1495-1536.

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

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

51. 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.

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

53. 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.

54. 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.

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

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

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

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

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

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

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

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

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

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

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

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

67. 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.

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