Copula-Based Random Effects Models for Clustered Data

Santiago Pereda-Fernández*

Banca d'Italia

October 28, 2019

Abstract

In a binary choice panel data framework, probabilities of the outcomes of several individuals depend on the correlation of the unobserved heterogeneity. I propose a random effects estimator that models the correlation of the unobserved heterogeneity among individuals in the same cluster using a copula. I discuss the asymptotic efficiency of the estimator relative to standard random effects estimators, and to choose the copula I propose a specification test. The implementation of the estimator requires the numerical approximation of high-dimensional integrals, for which I propose an algorithm that works for Archimedean copulas that does not suffer from the curse of dimensionality. This method is illustrated with an application of labor supply in married couples, finding that about one half of the difference in probability of a woman being employed when her husband is also employed, relative to those whose husband is unemployed, is explained by correlation in the unobservables.

Keywords: Assortative mating, copula, female labor supply, high-dimensional integration, nonlinear panel data

JEL classification: C33, C35, J12, J22

*Banca d'Italia, Via Nazionale 91, 00184 Roma, Italy. I can be reached via email at santiago.pereda@.bancaditalia.it.

1 Introduction

Some empirical questions involve the outcomes of several individuals. In a panel data setting, this requires the estimation of the joint distribution of the unobserved heterogeneity. For example, the probability of a woman being employed conditional on the employment status of her husband, which is larger for married women. Part of this difference can be accounted by differences in the observed characteristics. However, the remaining fraction could be due to correlation in the unobservables.

This paper addresses the estimation of a binary choice panel data model when the unobserved heterogeneity is correlated across individuals in the same cluster (group). In this setting, available binary choice panel data estimators either ignore the correlation or do not account for the distribution of individual effects: in the first case the estimated probability is biased, and in the second case they only estimate some of its components, so it is not possible to estimate such probabilities.

There are three main contributions in this paper. First, I present the Copula-Based Random Effects estimator (CBRE) for clustered data. It is a binary choice panel data estimator that acknowledges the correlation of the unobservables for individuals in the same cluster and can consistently estimate the probability of joint and conditional events when there is such correlation. Second, I propose an algorithm for the numerical approximation of high-dimensional integrals with Archimedean copulas. Simulation results indicate that, for a similar precision level, it is faster than alternative simulation methods. Third, I estimate the correlation of the unobserved propensity to work of married couples in several countries in Europe, finding that this correlation and the covariates have a similar explanatory power.

I consider a setup in which outcomes are correlated for individuals in the same cluster, but they are independent across them. Clusters are present in many real world situations because agents often share the same environment. For example, test scores are correlated within classroom (Hanushek, 1971), which reflects the unobserved teacher's quality.

I model the marginal distributions and the copula of the random effects separately, which together characterize their joint distribution. This constitutes a flexible way of modeling the correlation, as it allows the combination of different copulas and marginals. Moreover, there is no loss of generality, as any continuous multivariate distribution can be expressed in terms of a multivariate copula and the marginals (Sklar, 1959). In fact, the standard Random Effects (RE) estimator is a particular case of the CBRE when the copula is independent.

I present the main results using a parametric copula. Because modeling the correlation of the unobserved heterogeneity is important, I adapt Schennach and Wilhelm (2017) test to this framework to select the most appropriate copula. This test compares the value of the likelihood with any two parametric copulas, and it can establish if one of them is statistically better than the other at fitting the data, even if neither is the true one. I also present a test of independence of the copula when the parameter lies at the boundary of the parameter space under the null hypothesis, which is the case for several parametric copulas.

This paper builds on the literature of binary choice panel data models (Chamberlain, 1984; Arellano and Bonhomme, 2011). The early focus was on the identification of the slope parameter when the unobserved heterogeneity is unrestricted. Prominent examples include the Conditional Fixed Effects Logit (Chamberlain, 1980), the Maximum Score Estimator (Manski, 1975, 1987), or variations of it (Lee, 1999). These estimators rely on parametric assumptions on the error term or that some regressors have unbounded support (Chamberlain, 2010). However, such estimators ignore the unobserved heterogeneity, so they cannot estimate joint probabilities on their own. An alternative approach would be to correct the bias arising from the incidental parameter problem (Hahn and Newey, 2004; Fernández-Val, 2009), which is feasible when the number of periods grows to infinity.

Finally, the random effects approach assumes that the unobserved heterogeneity has a known parametric distribution, frequently a probit with normally distributed individual effects (Butler and Moffitt, 1982). This approach has the advantage of jointly estimating the slope parameters and the distribution of the unobserved heterogeneity, allowing for the estimation of functions that depend on both, such as joint probabilities.

Regarding the second contribution, RE estimators need to numerically approximate an integral with respect to the distribution of the individual effects. If they are independent across individuals as in standard RE estimators, the integral is one-dimensional. However, when they are dependent the dimension of the integral equals the number of individuals in each cluster, making the estimator subject to the curse of dimensionality. Thus, much of the empirical analysis has been constrained to the usage of quadrature methods (Butler and Moffitt, 1982) for the integration when the latent variable has a low dimension (up to 4), or simulation-based methods (Geweke, 1989; Hajivassiliou et al., 1996), which are often restricted to the multivariate normal distribution. I propose an algorithm to approximate high-dimensional integrals numerically when the copula is Archimedean, and it can also be adapted to elliptical copulas. I compare the performance of the algorithm to that of Monte Carlo integration, and show that this algorithm overcomes the curse of dimensionality for the type of approximations considered.

To illustrate the applicability of the estimator, I estimate the correlation in the individual propensity to work of married couples for several European countries. There is evidence that people tend to marry to others with similar observed characteristics (Bruze, 2011; Charles et al., 2013), and I find that there is also a moderate degree of correlation in the unobservables that is heterogeneous across countries.

I decompose the difference in the probability of being employed for women married to either employed or unemployed husbands, into an endowment and a homophily effect. Both effects are statistically significant for every country, and the unobserved correlation that determines the homophily effect accounts for a larger share of the difference than the endowment effect in about half of them. In particular, it tends to be larger in Southern European countries and smaller in Northern ones. Moreover, I compute some counterfactual probabilities, finding that ignoring this correlation produces biases in the estimation of the probability that at least one member of the couple is employed in each period.

The rest of the paper is organized as follows. In Section 2 I describe the econometric framework and present the estimator, along with the specification tests. In Section 3 I discuss the efficiency of the CBRE estimator relative to alternative methods. Section 4 describes the algorithm used for the approximation of the multidimensional integral. In Section 5 I conduct a Monte Carlo analysis, and the study on labor supply in couples is shown in Section 6. Finally, Section 7 concludes. All proofs are shown in Appendix A.

2 Framework and Estimation

Consider the following binary panel data setup:

$$\begin{aligned} y_{igt} &= \mathbf{1}\left(y^*_{igt} \geq 0\right) \\ y^*_{igt} &= \eta_{ig} + x'_{igt}\beta_0 + \varepsilon_{igt} \end{aligned} \tag{1}$$

where the econometrician observes the dependent variable y_{igt} and the covariates x_{igt} for agent $i = 1, ..., N_g$ in cluster $g = 1, ..., G$ at time $t = 1, ..., T$, and $\mathbf{1}(\cdot)$ denotes the indicator function. Relative to the *iid* data framework, there exist clusters of size N_g and the individual effects η_{ig} can be correlated within clusters. Using the taxonomy in Manski (1993), they are correlated effects.

Equation 1 is modeled as a random effects model, so the distributions of the unobservables are parametric. Denote the distribution of ε_{igt} by F_ε and the marginal distribution of the individual effects by $F_\eta(\eta_{ig}|x_{ig};\sigma_0)$, where $x_{ig} \equiv (x_{ig1}, ..., x_{igT})'$, and $x_g \equiv \left(x_{1g}, ..., x_{N_g g}\right)'$. Moreover, denote the ranks of the individual effects by $u_{ig} \equiv F_\eta(\eta_{ig}|x_{ig};\sigma_0)$, such that $u_g \equiv \left(u_{1g}, ..., u_{N_g g}\right)'$. To stress the relation between the individual effects and the ranks, I henceforth denote the former by $\eta(u_{ig}|x_{ig};\sigma_0)$. The correlation of the individual effects is modeled with the copula $C(u_g|x_g;\rho_0)$ which, together with the marginals, characterizes their joint distribution. σ_0 and ρ_0 respectively denote the parameters of the marginal distribution and the copula of the individual effects. The latter determines the amount of correlation of these effects and typically nests the independence copula.

The cluster structure may also be reflected on the distribution of the covariates, which could be correlated across time and within clusters, but not across clusters. Formally,

Assumption 1. *u_g is iid for all $g = 1, ..., G$, x_g are iid $\forall g = 1, ..., G$, and ε_{igt} are iid for all $i = 1, ..., N_g, g = 1, ..., G, t = 1, ..., T$.*

Hence, I consider a static model in which there could be sorting on observables and unobservables, both of which would determine the outcome variable. On the other hand, I am ruling out time-varying individual effects and dynamic effects of both the dependent variable and the covariates. Note however, that it would be possible to consider a richer model that combines the within cluster dependence with individual serial correlation.

As usual in discrete choice models, the variance of ε_{igt} is fixed, and the slope parameters and the variance of the random effects reflect this normalization. Moreover, if x_{igt} includes the constant term or a set of dummies for every period, the mean of the random effect is normalized to zero. To keep notation compact, define $C_X(u_g; \rho_0) \equiv C(u_g|x_g; \rho_0)$. If the covariates are discrete, it is possible to use a different copula for each distinct value of the covariates, and if they are continuous, one could parameterize it. *E.g.*, for the Gaussian copula, one could let $\rho(x) = {}^{\exp(x'\rho)-1}/_{\exp(x'\rho)+1}$.

Let $\theta \equiv (\beta', \sigma', \rho')'$ and $z_{igt} \equiv \left(y_{igt}, x'_{igt}\right)'$. Denote vectors of stacked individual variables by the ig subscript, and vectors of stacked cluster-individual variables by the g subscript. The log-likelihood function is given by

$$\mathcal{L}(\theta) = \sum_{g=1}^{G} \log\left(\ell_g(z_g; \theta)\right) \equiv \sum_{g=1}^{G} \log\left(\int_{[0,1]^{N_g}} \prod_{i=1}^{N_g} P_{ig}(z_{ig}, u_{ig}; \mu)\, dC_X(u_g; \rho)\right) \qquad (2)$$

where $P_{ig}(z_{ig}, u_{ig}; \mu)$ is the individual contribution to the likelihood, *i.e.* $P_{ig}(z_{ig}, u_{ig}; \mu) \equiv \prod_{t=1}^{T}\left[1 - F_\varepsilon\left(-\left(\eta(u_{ig}|x_{ig}; \sigma) + x'_{igt}\beta\right)\right)\right]^{y_{igt}} F_\varepsilon\left(-\left(\eta(u_{ig}|x_{ig}; \sigma) + x'_{igt}\beta\right)\right)^{1-y_{igt}}$. The CBRE estimator is given by $\hat{\theta} = \arg\max_{\theta\in\Theta} \mathcal{L}(\theta)$, and it requires the integration of a product over a potentially large dimensional space. Nonparametric identification of all the components of Equation 2 cannot be attained under weak assumptions. This lack of identification is formally addressed in Appendix S1. The asymptotic distribution of the CBRE estimator is shown in Appendix S2. Relative to standard RE estimators, it requires the number of individuals in a cluster to be finite and some regularity conditions on the copula.

A particular case of this model was considered by Antweiler (2001), who modeled the unobserved heterogeneity as the sum of a group and an individual effect. Also, a series of papers assumed that all individuals in the same group had the same unobserved effect. Lin and Ng (2012) and Bonhomme and Manresa (2015) considered a linear panel data framework in which group membership is unknown, and they proposed estimators that account for it. Moreover, the latter allows the group effect to be time-varying. On the other hand, Bester and Hansen (2016) considered the estimation with group-effects in a nonlinear panel data framework with known group membership. These estimators require that the number of periods and individuals grow to infinity, either at the same or a different rate. This contrasts with the current setting, in which the number of periods is small and fixed, which limits the amount of unobserved heterogeneity that can be identified.

2.1 Estimation of Joint and Conditional Events

Let $\mathcal{S}$ denote the set of permutations of $y_g \equiv \left(y_{1g1}, ..., y_{1gT}, ..., y_{N_g gT}\right)$ in which an event of interest occurs. In the labor supply example, such event could be that at least one of the two partners is employed in every period, *i.e.* $\mathcal{S} = \{y_g : y_{1gt} + y_{2gt} \geq 1 \forall t\}$, where $y_{igt} = 1$ if individual i is employed at time t. The probability of such an event is given by

$$\mathbb{P}\left(y_g \in \mathcal{S} | x_g\right) = \sum_{b \in \mathcal{S}} \mathbb{P}\left(y_g = b | x_g\right) = \sum_{b \in \mathcal{S}} \int_{[0,1]^{N_g}} \prod_{i=1}^{N_g} \prod_{t=1}^{T} \mathbb{P}\left(b_{gt} | x_{igt}, u_{ig}\right) dC_X\left(u_g; \rho_0\right) \qquad (3)$$

where $\mathbb{P}\left(b_{gt} | x_{igt}, u_{ig}\right) = \left[1 - F_\varepsilon\left(-\left(\eta\left(u_{ig} | x_{ig}; \sigma\right) + x'_{igt}\beta_0\right)\right)\right]^{b_{gt}} F_\varepsilon\left(-\left(\eta\left(u_{ig} | x_{ig}; \sigma\right) + x'_{igt}\beta_0\right)\right)^{1-b_{gt}}$. To estimate the probability that the event of interest occurs, replace θ_0 by $\hat{\theta}$ and approximate the integral as shown in Section 4. The computation of a conditional event is straightforward given the estimates of the joint and marginal events: the probability of an event A given B

equals their joint probability divided by the marginal of the event B.

2.2 Choice of Copula and Specification Tests

CBRE can be implemented with any copula, including flexible ones such as mixture copulas (Trivedi et al., 2007) or Bernstein copulas (Sancetta and Satchell, 2004; see Appendix S7 for further details). However, such options are practical as long as the dimension of the clusters is small, and even then they may be slow to compute. This restricts the class of copulas to those that are computationally feasible, but in principle it is unknown which of them is the most appropriate. One possibility is to select the copula with a test.

Testing a parametric copula against another

Vuong (1989) proposed a test to choose between two parametric models by comparing their estimated likelihood. This test can be used to compare any pair of copulas and determine if either provides a statistically better fit than the other, even if neither is the true copula. However, it requires a pretest to determine if the variance of the difference of the likelihood is zero, *i.e.* if the test is in the degenerate. This can occur when both models are observationally equivalent, which could be the case in this paper, since the copula is not nonparametrically identified. To overcome this issue, Shi (2015) and Schennach and Wilhelm (2017) proposed each a test that is non degenerate and does not require pretesting. I use the latter, and I describe in Appendix S5 how it is adapted to the present setup. Because the copula models a vector of latent variables, existing nonparametric tests for copulas (*e.g.*, Prokhorov et al., 2019) cannot be used, since they require the variables to be observed.

Given the estimates with several copulas, one would choose the one that attains the maximum likelihood, and test it against any of the other copulas to determine if it is statistically better. Note that this test can also be used to test the marginal distribution of the individual effects (*e.g.*, normal versus Laplace distribution), or the distribution of the time-varying unobserved heterogeneity (*e.g.*, probit versus logit).

Testing for independence of the copula

The researcher may want to test the hypothesis of independence of the copula to justify using the standard RE estimator (*e.g.*, if the estimand of interest does not depend on the copula). For most parametric copulas, the independence case is a particular value of ρ, denoted by ρ^{ind}, so testing for independence amounts to testing $H_0 : \rho = \rho^{ind}$. If ρ^{ind} lies in the interior of the parameter space (*e.g.*, for the bivariate Gaussian copula, $H_0 : \rho = 0$, where $\rho \in [-1, 1]$), it is easy to test the null hypothesis using standard tests, such as a t-test.

A more complicated situation arises if ρ^{ind} lies on the boundary of the parameter space, *e.g.*, *Clayton* (0), *Gumbel* (1), or *Frank* (0). Self and Liang (1987) showed that in this case, the maximum likelihood estimator is still consistent, but not asymptotically normal. For expositional clarity, I focus on the case in which ρ is univariate. Let $W \sim \mathcal{N}(0, \sigma_W^2)$, where σ_W^2 is the element of the inverse of the information matrix that corresponds to ρ. The limiting distribution of $\sqrt{G}\left(\hat{\rho} - \rho^{ind}\right)$ is given by $W\mathbf{1}(W > 0)$. Hence, the asymptotic distribution of a Wald test of independence is a 50:50 mixture of a degenerate distribution at 0 and a χ_1^2 under the null hypothesis. Then, one would not accept the null hypothesis of independence if $\hat{\rho}$ is greater than the 95th percentile of this mixture distribution.

An important aspect to take into consideration in all cases is the possible error in the estimation of the asymptotic variance caused by the numerical approximation. If the estimated standard errors are excessively small, it would lead to a low rejection rate of the null hypothesis. The simulations in Skrainka and Judd (2011) indicate that this problem is particularly severe for simulation-based approximations.

3 Efficiency

Denote the vector of marginal parameters by $\mu \equiv (\beta', \sigma')'$. The RE estimator, denoted by $\tilde{\mu}$, is the maximizer of the following function:

$$\tilde{\mathcal{L}}(\mu) = \sum_{g=1}^{G} \sum_{i=1}^{N_g} \log\left(\ell_{ig}(z_{ig}; \mu)\right) \equiv \sum_{g=1}^{G} \sum_{i=1}^{N_g} \log\left(\int_0^1 \prod_{i=1}^{N_g} P_{ig}(z_{ig}, u_{ig}; \mu)\, du_{ig}\right) \tag{4}$$

Under the correct parametric assumptions and suitable regularity conditions, which are discussed in Appendix S2, RE and CBRE are consistent estimators of μ and asymptotically normal. If one is only interested in μ or functions that depend on it, such as Average Partial Effects (APE; see Appendix S3 for further details), one could use either estimator. RE is faster to implement, making it more appealing from a computational standpoint.

However, RE is generally inefficient. The Infeasible CBRE (ICBRE) estimator, which maximizes Equation 2 when the copula is fully known, is at least as efficient as RE. Moreover, combining the scores of RE and CBRE and using the optimal weighting matrix would yield a more efficient and feasible GMM estimator: the Augmented CBRE (ACBRE) estimator. The following Proposition establishes their relative efficiency:

Proposition 1. *Denote the asymptotic variances of RE, CBRE, ACBRE, and ICBRE by*

$$\Sigma_{\mu}^{RE} = D_{\mu\mu}^{RE,-1} \Omega_{\mu}^{RE} D_{\mu\mu}^{RE,-1} \tag{5}$$

$$\Sigma_{\theta}^{CBRE} = D_{\theta\theta}^{CBRE,-1} \tag{6}$$

$$\Sigma_{\theta}^{ACBRE} = \left(\begin{bmatrix} D_{\theta\theta}^{CBRE} \\ D_{\mu\theta}^{RE} \end{bmatrix}' \begin{bmatrix} \Omega_{\theta}^{CBRE} & \Omega_{\theta}^{CBRE,RE} \\ \Omega_{\theta}^{RE,CBRE} & \Omega_{\mu}^{RE} \end{bmatrix}^{-1} \begin{bmatrix} D_{\theta\theta}^{CBRE} \\ D_{\mu\theta}^{RE} \end{bmatrix} \right)^{-1} \tag{7}$$

$$\Sigma_{\mu}^{ICBRE} = D_{\mu\mu}^{CBRE,-1} \tag{8}$$

where $D_{\mu\mu}^{RE}$, $D_{\mu\theta}^{RE}$, $D_{\theta\theta}^{CBRE}$, Ω_{μ}^{RE}, Ω_{θ}^{CBRE}, $\Omega_{\theta}^{CBRE,RE}$ and $\Omega_{\theta}^{RE,CBRE}$ are defined in Appendix A.

1. *ACBRE is no less efficient than CBRE for the estimation of θ, and no less efficient than RE for the estimation of μ.*

2. *ICBRE is no less efficient than either CBRE or RE for the estimation of μ.*

3. *If $D_{\mu\mu}^{RE} = \Omega_{\mu}^{CBRE,RE}$, ACBRE and CBRE are equally efficient for the estimation of θ.*

4. *If the copula is independent, ICBRE, ACBRE and RE are equally efficient for the estimation of μ.*

5. If $D_{\mu\rho}^{CBRE} = 0$, CBRE and ICBRE are equally efficient for the estimation of μ.

Because $\tilde{\mathcal{L}}$ is a particular case of $\mathcal{L}$, ICBRE is generally more efficient than RE. Thus, the precision of the estimates of the marginal parameters can be improved thanks to the information provided by the copula, as is the case for ACBRE. Similarly, the efficiency of CBRE can be improved by adding the scores of the RE estimator, unless they are redundant, which also happens under independence. On the other hand, CBRE requires the estimation of ρ, so it is not always more efficient than RE. The simulation evidence in Section 5 suggests that the extra information coming from the copula is usually dominant, making CBRE slightly more efficient than RE, except when the copula is nearly independent.

Another estimator more efficient than RE would use the scores with respect to μ for each group member separately. This requires that all groups have the same size and it follows because the individual scores could be correlated: using their sum is less efficient than using all of them individually, akin to the QMLE and IQMLE estimators considered by Prokhorov and Schmidt (2009). The following proposition formalizes this argument:

Proposition 2. *Denote the set of individual marginal scores by*

$$s_g(\mu) \equiv \begin{bmatrix} \nabla_\mu \log\left(\ell_{1g}\left(z_{1g};\mu\right)\right) \\ \ldots \\ \nabla_\mu \log\left(\ell_{\overline{N}g}\left(z_{\overline{N}g};\mu\right)\right) \end{bmatrix} \tag{9}$$

for $g = 1, ..., G$. Define the Improved RE (IRE) estimator as the GMM estimator based on $\mathbb{E}\left[s_g(\mu_0)\right] = 0$ that uses the optimal weighting matrix. IRE is no less efficient than RE, and under independence they are equally efficient.

4 Implementation Algorithm

Typically, integrals like those in Equation 2 do not have a closed form solution. Thus, solving for $\hat{\theta}$ in Equation 2 requires the numerical evaluation of the score and the Hessian (see Appendix S4). Simulation methods like Monte Carlo tend to perform slowly when the dimension of the integrals is large, and have been outperformed by some recent advances in high dimensional numerical integration methods (Heiss and Winschel, 2008; Skrainka and Judd, 2011), although they are still subject to the curse of dimensionality. I propose an algorithm to numerically approximate a class of integrals when the copula is Archimedean.

By Corollary 2.2 in Marshall and Olkin (1988), an Archimedean copula is given by

$$C(u) = \int_{\mathbb{R}_+^N} \exp\left(-\sum_{i=1}^{N} \zeta_i \phi_i^{-1}(u_i)\right) dB(\zeta) \tag{10}$$

where B is the cdf of ζ, and ϕ_i is the Laplace transform of the marginal distributions of B. For some of the most common Archimedean copulas, ζ is unidimensional and $\phi_i = \phi \forall i$. Consider the following integral:

$$\mathcal{I} = \int_{[0,1]^N} \prod_{i=1}^{N} \ell_i(u_i)\, dC(u) = \int_{\mathbb{R}_+} \prod_{i=1}^{N} \left[\int_0^1 \ell_i(u_i)\, dF^{\zeta}(u_i)\right] dB(\zeta) \tag{11}$$

where $F^{\zeta}(u_i) = \exp\left(-\zeta\phi^{-1}(u_i)\right)$. The N-dimensional integral can be expressed as the integral of the product of N independent integrals, reducing the dimensionality from N to 2. Intuitively, the condition that ζ is unidimensional and $\phi_i = \phi \forall i$ implies that the elements of the copula are *exchangeable*, *i.e.* $C(u_1, u_2) = C(u_2, u_1)$, and therefore the correlation between any two elements of the copula is always the same. This seems reasonable in a context in which all individuals within a cluster are affected by a common factor, *e.g.*, all students in a classroom are affected by teacher's quality. This condition is more likely to hold when one conditions on the covariates. To see this, consider the following example. Let $y_i^* = x_i + \varepsilon_i$ for $i = 1, 2$, where the marginal distribution of x_i is F_X, the copula between x_1 and x_2 is given by C_X, the marginal distribution of ε_i is F_ε, and ε_1 and ε_2 are independent. Then, $\mathbb{P}(Y_1^* \leq y_1^*, Y_2^* \leq y_2^* | x_1, x_2) = F_\varepsilon(y_1^* - x_1) F_\varepsilon(y_2^* - x_2)$, so the copula between Y_1^* and Y_2^* conditional on X_1 and X_2 is exchangeable. On the other hand, $\mathbb{P}(Y_1^* \leq y_1^*, Y_2^* \leq y_2^*) = \int C_X(F_X(y_1^* - \varepsilon_1), F_X(y_2^* - \varepsilon_2))\, dF_\varepsilon(\varepsilon_1)\, dF_\varepsilon(\varepsilon_2)$. If C_X is not exchangeable, then neither is the copula between Y_1^* and Y_2^*. Taking this into consideration, the proposed algorithm to approximate the integral is given by

1. Compute a grid of values of ζ, given by $\zeta_j = B^{-1}\left(\frac{j}{N_1+1}\right)$, $\forall j = 1, ..., N_1$.

2. Compute a grid of values of u $\forall j$, given by $u_{jh} = \phi\left(-\frac{1}{\zeta_j}\log\left(\frac{h}{N_2+1}\right)\right)$, $\forall h = 1, ..., N_2$.

3. Approximate the integral by $\hat{\mathcal{I}} = \frac{1}{N_1}\sum_{j=1}^{N_1}\prod_{i=1}^{N}\left[\frac{1}{N_2}\sum_{h=1}^{N_2}\ell_i\left(u_{jh}\right)\right]$.

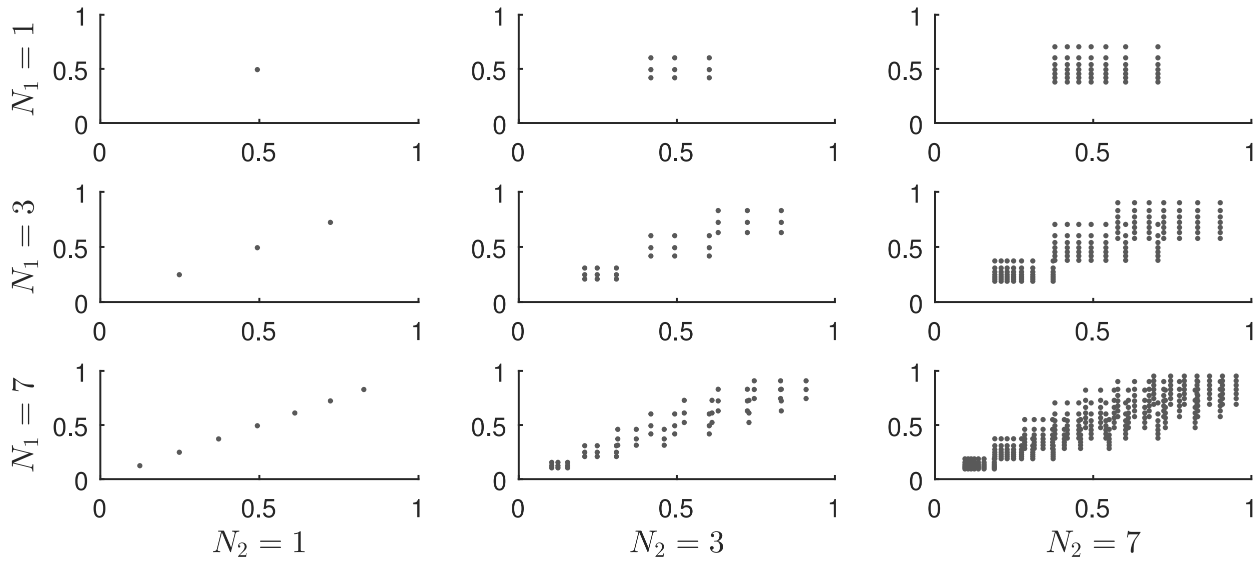
To understand how the algorithm works, consider the integration of a function m with respect to a distribution F: $\int m(x)\,dF(x)$. This integral is approximated by evaluating the function m at Q evenly spaced quantiles and taking the average across them: $\frac{1}{Q+1}\sum_{q=1}^{Q} m(x_q)$, where $x_q = F^{-1}\left(\frac{q}{Q+1}\right)$ is the qth quantile of the function F.

The algorithm uses this approximation twice, and Figure 1 shows how the selection of the points used for integration is done in practice. For a fixed value of ζ, there are N_2 different values for each u_i as shown in the upper graphs. These points split the unit interval into $N_2 + 1$ intervals that have the same probability of occurring, conditional on ζ. Hence, the inner integral of each dimension j is approximated by $\frac{1}{N_2}\sum_{h=1}^{N_2}\ell_i\left(u_{jh}\right)$. The symmetry of the copula means that the points u_{jh} are indeed the same for each dimension, so there is no need to compute a different number of points of support for each dimension. Then, to approximate the outer integral, repeat the previous operation for the N_1 values of ζ_j and calculate the average across them. This reasoning can also be applied to approximate integrals with Elliptical copulas, and a variation of this algorithm is presented in Appendix S6.

Graphically, the number of squares increases as one moves from the upper to the lower graphs (Figure 1). As $N_1, N_2 \to \infty$, the unit square is covered by more points and $\hat{\mathcal{I}} \to \mathcal{I}$. For higher dimensions the intuition remains the same, and for each value of ζ_j there is a hypercube composed of N_2^d points.

This algorithm can be used for the joint maximization of $\mathcal{L}(\theta)$ with respect to all parameters. However, if the number of parameters is large, the algorithm could be combined with other strategies to obtain the CBRE estimator or an approximation of it. For example,

Figure 1: Algorithm grid to approximate the integral



one could maximize $\mathcal{L}(\theta)$ with respect to μ and ρ iteratively until convergence is achieved; alternatively one could estimate μ by RE, and maximize $\mathcal{L}(\hat{\mu}_{RE}, \rho)$ with respect to ρ. The latter could be also used as the initial condition for the CBRE estimator.

I compare the performance in terms of speed and precision of the algorithm relative to Monte Carlo simulation, by approximating the integral $\mathcal{I} \equiv \int_{[0,1]^d} \prod_{j=1}^{d} \sqrt{u_j} dC(u_1, ..., u_d; \rho)$ with a $Clayton(4)$ copula and dimension $d = \{2, 3\}$, which has no closed form solution. The number of draws of the Monte Carlo equals $N_1 N_2^d$, *i.e.* the total number of points evaluated by the algorithm. The results are shown in Table 1.

Even when $d = 2$, the algorithm proposed in this paper is several orders of magnitude faster than the traditional Monte Carlo. Moreover, for a given number of points, their performance is similar, and the approximation is within two standard deviations of the Monte Carlo. The algorithm consistently reports a number inferior to the mean across repetitions of the Monte Carlo. However, increasing the number of points at which the integral is evaluated is not as costly as for the Monte Carlo, resulting in a more accurate approximation for a given amount of computational time. When $d = 3$, the algorithm loses some accuracy with respect to the Monte Carlo, but time gains are large enough to allow for an increase in the number of integration points of the algorithm and still outperform the

Table 1: Implementation algorithm & Monte Carlo comparison

$N_1 = N_2$		9	19	49	99
			$d = 2$		
$\hat{\mathcal{I}}$	Algorithm	0.4933	0.4934	0.4936	0.4937
	Monte Carlo	0.4942	0.4944	0.4940	0.4940
S.D.	Algorithm	-	-	-	-
	Monte Carlo	0.0112	0.0033	0.0008	0.0003
Time	Algorithm	0.0001	0.0003	0.0011	0.0032
	Monte Carlo	0.0027	0.0036	0.0569	0.4776
			$d = 3$		
$\hat{\mathcal{I}}$	Algorithm	0.3786	0.3827	0.3854	0.3863
	Monte Carlo	0.3868	0.3873	0.3873	0.3873
S.D.	Algorithm	-	-	-	-
	Monte Carlo	0.0039	0.0007	0.0001	0.0000
Time	Algorithm	0.0001	0.0003	0.0011	0.0035
	Monte Carlo	0.0252	0.0828	3.3281	58.8875

Notes: $\hat{\mathcal{I}}$ is the approximated value of the integral for the proposed algorithm, and the mean value across repetitions for the Monte Carlo simulations. The sampling algorithm for the Monte Carlo is the one proposed by Marshall and Olkin (1988). The true value $\mathcal{I}$ has no closed-form expression.

Monte Carlo. Note that the consistency of the estimator requires increasing the number of grid points as the sample size increases. This is common to simulation methods, although the approximation error is more severe for the latter (Skrainka and Judd, 2011).

Table 2 shows the performance of the algorithm in high dimensions: its accuracy decreases with the dimensionality of the problem, as reflected in the changes in the approximation when the number of points used to evaluate the integral is increased. However, for given N_1 and N_2, computational time remains unchanged despite the increase of the dimensionality. This suggests that for applications with moderately large clusters, values of N_1 and N_2 between 20 and 50 should suffice. Moreover, the simulations in Section 5 suggest that keeping $N_1 = N_2$ is crucial to obtain consistent estimates of the copula parameters.

Table 2: Performance in high dimensions

$N_1 = N_2$	9	19	49	99	199
			$d = 2$		
$\hat{\mathcal{I}}$	0.4933	0.4934	0.4936	0.4937	0.4938
Time	0.0001	0.0003	0.0011	0.0032	0.0108
			$d = 3$		
$\hat{\mathcal{I}}$	0.3786	0.3827	0.3854	0.3863	0.3868
Time	0.0001	0.0003	0.0011	0.0035	0.0107
			$d = 5$		
$\hat{\mathcal{I}}$	0.2452	0.2534	0.2584	0.2602	0.2610
Time	0.0001	0.0003	0.0011	0.0035	0.0109
			$d = 10$		
$\hat{\mathcal{I}}$	0.1076	0.1176	0.1238	0.1260	0.1271
Time	0.0001	0.0003	0.0011	0.0033	0.0109
			$d = 50$		
$\hat{\mathcal{I}}$	0.0017	0.0033	0.0049	0.0057	0.0062
Time	0.0001	0.0003	0.0011	0.0033	0.0108

Notes: $\hat{\mathcal{I}}$ is the approximated value of the integral for the proposed algorithm. The true value $\mathcal{I}$ has no closed-form expression.

5 Monte Carlo

The finite sample performance of the estimator is shown in a Monte Carlo exercise with the following data generating process: $y_{igt} = \mathbf{1}\left(\eta_{ig} + \gamma_t + x'_{igt}\beta + \varepsilon_{igt} > 0\right)$, where ε_{igt} is logistically distributed, $\eta_{ig} \sim \mathcal{N}(0, \sigma_0^2)$, $u_g \sim Clayton(\rho_0)$, $\gamma_0 = (-1.5, -1, -0.5, 0)'$, $x_{igt} \sim U(0,1)$, $\beta_0 = 1$, $\sigma_0 = 3$, and $\rho_0 = 8$ for $t = 1,2,3,4$, $i = 1,...,10$, $g = 1,...,G$, where $G = 100, 200, 500$.

Table 3 shows the estimates of the parameters. As argued in Section 3, both CBRE and RE consistently estimate the time effects and β. The CBRE estimates of σ have a slightly upward approximation bias that decreases as the number of points used for the approximation increases, though it is of the same magnitude regardless of the sample size. RE is implemented using Gaussian quadrature, greatly reducing the approximation bias. Setting $N_1 \neq N_2$ results in a substantial bias of the correlation parameter: if $N_1 > N_2$, then the parameter is downward biased, whereas if $N_1 < N_2$, the bias is positive. In terms of efficiency, CBRE displays a slightly smaller variance.

Table 3 also shows the estimates of the joint probability that $y_{1gt} = y_{2gt} = 1$ (P_1), and the conditional probabilities that $y_{1gt} = 1$ conditional on either $y_{2gt} = 1$ or $y_{2gt} = 0$ (P_2 and P_3, respectively). CBRE estimates the three probabilities consistently, but RE is biased because it does not account for the correlation in the unobserved heterogeneity. Notice also that the approximation bias of is small CBRE displays little sensitivity to the choice of N_1 and N_2. On the other hand, both CBRE and RE consistently estimate the unconditional probability that y_{1gt} equals 1 (P_4), although the latter is less efficient.

The standard errors of CBRE and the cluster-robust ones of RE correctly estimate the asymptotic variance of the estimator across repetitions (Table 4). This is true both when the individual effects are correlated and when they are independent. The only exception is the standard error of the copula parameter in the independence case, which follows because it lies at the boundary of the parameter space, making the asymptotic distribution non-normal. On the other hand, the naive standard errors of RE are biased when the copula shows dependence. Table 4 also shows that the efficiency gain on the marginal parameters of CBRE vanishes when the copula is independent.

Finally, Table 5 shows the performance of the estimator under misspecification. In particular, I consider a logit with a Frank copula, and a probit with a Clayton copula. When the true copula is a Clayton, RE tends to have a worse fit than the correctly specified estimators, as reflected by the independence tests. Using a misspecified copula results in a slightly worse fit, but in most cases their fit is statistically equivalent. Changing the logit to a probit does not reduce the fit by much, in line with previous results in the literature (Chambers and Cox, 1967; Amemiya, 1981). Indeed, when there is positive correlation, the specification tests tend to accept the null of equal fit for probit with the correctly specified copula more often than for the logit with the incorrect copula. Hence, the test has little power against the false hypothesis of equal fit unless the sample size is large. In all cases

Table 3: Monte Carlo Results

N	1000					2000					5000					
	CBRE				RE	CBRE				RE	CBRE				RE	θ_0
N_1	10	10	50	50		10	10	50	50		10	10	50	50		
N_2	10	50	10	50		10	50	10	50		10	50	10	50		
γ_1	-1.48	-1.52	-1.47	-1.50	-1.51	-1.47	-1.51	-1.45	-1.49	-1.50	-1.47	-1.50	-1.45	-1.48	-1.50	-1.5
	(0.32)	(0.32)	(0.33)	(0.33)	(0.34)	(0.22)	(0.22)	(0.23)	(0.23)	(0.24)	(0.14)	(0.14)	(0.14)	(0.14)	(0.15)	
γ_2	-0.98	-1.02	-0.96	-1.00	-1.00	-0.97	-1.01	-0.96	-0.99	-1.01	-0.97	-1.00	-0.95	-0.98	-1.00	-1
	(0.32)	(0.32)	(0.33)	(0.33)	(0.34)	(0.22)	(0.22)	(0.23)	(0.23)	(0.24)	(0.14)	(0.14)	(0.14)	(0.14)	(0.15)	
γ_3	-0.47	-0.51	-0.46	-0.49	-0.50	-0.47	-0.51	-0.46	-0.49	-0.50	-0.47	-0.50	-0.45	-0.49	-0.50	-0.5
	(0.33)	(0.32)	(0.33)	(0.33)	(0.34)	(0.22)	(0.22)	(0.23)	(0.23)	(0.24)	(0.14)	(0.14)	(0.14)	(0.14)	(0.15)	
γ_4	0.03	-0.01	0.04	0.01	0.00	0.03	-0.01	0.04	0.01	0.00	0.03	0.00	0.05	0.01	0.00	0
	(0.32)	(0.32)	(0.33)	(0.32)	(0.33)	(0.22)	(0.22)	(0.23)	(0.23)	(0.24)	(0.14)	(0.14)	(0.14)	(0.14)	(0.15)	
β	0.99	0.99	0.99	0.99	0.99	1.00	1.00	1.00	1.00	1.00	1.01	1.01	1.01	1.01	1.00	1
	(0.18)	(0.18)	(0.17)	(0.17)	(0.19)	(0.13)	(0.13)	(0.13)	(0.13)	(0.13)	(0.08)	(0.08)	(0.08)	(0.08)	(0.09)	
σ	3.56	3.52	3.17	3.12	3.00	3.55	3.51	3.16	3.12	2.99	3.55	3.51	3.17	3.12	2.99	3
	(0.26)	(0.26)	(0.24)	(0.24)	(0.27)	(0.19)	(0.19)	(0.18)	(0.18)	(0.20)	(0.12)	(0.12)	(0.11)	(0.11)	(0.12)	
ρ	7.61	9.21	6.57	8.00	-	6.96	8.53	6.20	7.63	-	6.88	8.43	6.12	7.53	-	8
	(13.03)	(12.86)	(6.83)	(5.55)		(1.01)	(1.20)	(0.94)	(1.13)		(0.60)	(0.72)	(0.56)	(0.67)		
$\mathcal{L}$	-1658	-1658	-1658	-1658	-2137	-3327	-3327	-3327	-3326	-4287	-8324	-8323	-8323	-8322	-10729	
P_1	0.269	0.267	0.266	0.265	0.150	0.270	0.269	0.267	0.266	0.150	0.271	0.269	0.268	0.267	0.151	0.264
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	
P_2	0.692	0.691	0.685	0.685	0.386	0.693	0.692	0.687	0.686	0.387	0.694	0.693	0.688	0.687	0.388	0.684
	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	
P_3	0.196	0.195	0.199	0.199	0.386	0.197	0.196	0.200	0.199	0.387	0.196	0.196	0.200	0.199	0.388	0.199
	(0.02)	(0.02)	(0.02)	(0.02)	(0.04)	(0.01)	(0.01)	(0.01)	(0.01)	(0.03)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	
P_4	0.389	0.387	0.388	0.388	0.386	0.390	0.389	0.390	0.390	0.387	0.391	0.389	0.391	0.391	0.388	0.386
	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	

Notes: Mean estimates of the parameters across 1000 repetitions, standard deviations across repetitions in parentheses, $\mathcal{L}$ denotes the maximized value of the likelihood function, and θ_0 the true value of the parameters. P_1 denotes the average probability across periods that $y_{1gt} = y_{2gt} = 1$, P_2 denotes the average probability across periods that $y_{1gt} = 1$, conditional on $y_{2gt} = 1$, P_3 denotes the average probability across periods that $y_{1gt} = 1$, conditional on $y_{2gt} = 0$, and P_4 denotes the average probability across periods that $y_{1gt} = 1$. The RE estimates were calculated by approximating the integral with a Gauss-Hermite quadrature with 20 points.

Table 4: Monte Carlo Results

Positive correlation

N	1000				2000				5000			
	CBRE		RE		CBRE		RE		CBRE		RE	
	$\overline{se}\left(\hat{\theta}\right)$	$Var\left(\hat{\theta}\right)$	$\overline{se}\left(\hat{\theta}\right)$	$Var\left(\hat{\theta}\right)$	$\overline{se}\left(\hat{\theta}\right)$	$Var\left(\hat{\theta}\right)$	$\overline{se}\left(\hat{\theta}\right)$	$Var\left(\hat{\theta}\right)$	$\overline{se}\left(\hat{\theta}\right)$	$Var\left(\hat{\theta}\right)$	$\overline{se}\left(\hat{\theta}\right)$	$Var\left(\hat{\theta}\right)$
γ_1	0.326	0.322	0.173	0.335	0.229	0.222	0.122	0.235	0.144	0.143	0.077	0.145
			[0.335]				[0.237]				[0.150]	
γ_2	0.325	0.325	0.169	0.339	0.228	0.224	0.119	0.239	0.143	0.143	0.075	0.145
			[0.335]				[0.236]				[0.150]	
γ_3	0.324	0.315	0.166	0.329	0.227	0.221	0.117	0.238	0.143	0.143	0.074	0.146
			[0.334]				[0.236]				[0.149]	
γ_4	0.324	0.322	0.165	0.333	0.227	0.224	0.116	0.235	0.143	0.141	0.073	0.143
			[0.331]				[0.234]				[0.148]	
β	0.182	0.173	0.188	0.184	0.125	0.127	0.133	0.137	0.078	0.081	0.084	0.087
			[0.186]				[0.132]				[0.084]	
σ	0.249	0.257	0.147	0.282	0.174	0.181	0.104	0.202	0.109	0.107	0.066	0.122
			[0.280]				[0.198]				[0.125]	
ρ	1.692	1.754	-	-	1.107	1.146	-	-	0.669	0.681	-	-

Independence

N	1000				2000				5000			
	CBRE		RE		CBRE		RE		CBRE		RE	
	$\overline{se}\left(\hat{\theta}\right)$	$Var\left(\hat{\theta}\right)$	$\overline{se}\left(\hat{\theta}\right)$	$Var\left(\hat{\theta}\right)$	$\overline{se}\left(\hat{\theta}\right)$	$Var\left(\hat{\theta}\right)$	$\overline{se}\left(\hat{\theta}\right)$	$Var\left(\hat{\theta}\right)$	$\overline{se}\left(\hat{\theta}\right)$	$Var\left(\hat{\theta}\right)$	$\overline{se}\left(\hat{\theta}\right)$	$Var\left(\hat{\theta}\right)$
γ_1	0.185	0.178	0.173	0.178	0.127	0.123	0.122	0.122	0.079	0.075	0.077	0.075
			[0.172]				[0.122]				[0.077]	
γ_2	0.180	0.173	0.169	0.173	0.123	0.119	0.119	0.119	0.077	0.073	0.075	0.073
			[0.169]				[0.120]				[0.076]	
γ_3	0.177	0.169	0.166	0.168	0.122	0.119	0.117	0.119	0.076	0.075	0.074	0.075
			[0.166]				[0.118]				[0.074]	
γ_4	0.176	0.169	0.165	0.169	0.121	0.121	0.116	0.120	0.075	0.071	0.073	0.071
			[0.165]				[0.117]				[0.074]	
β	0.196	0.192	0.189	0.191	0.136	0.134	0.133	0.134	0.084	0.083	0.084	0.083
			[0.186]				[0.132]				[0.084]	
σ	0.160	0.158	0.148	0.151	0.111	0.110	0.104	0.104	0.069	0.069	0.066	0.066
			[0.147]				[0.104]				[0.066]	
ρ	0.033	0.017	-	-	0.022	0.012	-	-	0.013	0.007	-	-

Notes: $Var\left(\hat{\theta}\right)$ and $\overline{se}\left(\hat{\theta}\right)$ respectively denote the variance of the estimates and their mean standard errors across 1000 repetitions. Cluster-robust standard errors of the RE estimates in brackets; RE estimates were calculated by approximating the integral with a Gauss-Hermite quadrature with 20 points.

the APE is correctly estimated, and the bias on the estimated probabilities is large only for the RE estimator. When the true copula is independent, all three estimators have a similar performance, and the specification tests tend to accept the null hypothesis of equal fit.

6 Labor Supply in Married Couples

I apply the presented methodology to study how assortative mating affects labor supply in married couples. I use the 2012 wave of the EU-SILC (European Union Statistics on Income and Living Conditions) dataset, which follows 279,115 individuals during the 2009-2012 period. I keep the sub-population of married couples, in which both individuals were aged 21-65 during the whole period, totaling 28,246 individuals.

I estimate six different specifications combining the logit and the probit with three different parametric copulas: Clayton, Frank, and Gaussian. For each of these specifications, I compute the CBRE estimator with correlated random effects, reporting the one with the lowest Akaike Information Criterion (AIC). To relax the parametric assumption of the copula, I also compute CBRE with a Bernstein copula of orders 2-4. These estimators are then compared to RE. Additionally, I also consider all the preceding estimators with standard random effects. The estimates with the Bernstein copula and the RE estimates are presented in Appendix S8, both with standard and correlated random effects. Additionally, I report the Bayesian Information Criterion (BIC) and the 10-fold cross validated likelihood.

All three parametric copulas, as well as the Bernstein copula of order 2 satisfy the exchangeability condition. On the other hand, Bernstein copulas of higher order are not necessarily symmetric. In all specifications, the marginal distribution of the random effects is normal. The dependent variable takes value one if agent i worked during year t and zero otherwise. The covariates included are gender, number of children smaller than 5 years old, age, level of education, and total household non-labor income, along with yearly dummies.

The estimates are qualitatively similar in all countries (Table 6): married females work

Table 5: Monte Carlo Tests

Positive correlation

N	1000				2000				5000				True value
	CBRE			RE	CBRE			RE	CBRE			RE	
	Logit	Logit	Probit	Logit	Logit	Logit	Probit	Logit	Logit	Logit	Probit	Logit	
	Clayton	Frank	Clayton	-	Clayton	Frank	Clayton	-	Clayton	Frank	Clayton	-	
APE	0.19	0.20	0.20	0.20	0.19	0.20	0.20	0.20	0.19	0.20	0.20	0.20	0.20
	(0.03)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	
P_1	0.27	0.25	0.26	0.15	0.27	0.25	0.26	0.15	0.27	0.25	0.27	0.15	0.29
P_2	0.69	0.68	0.68	0.39	0.69	0.68	0.68	0.39	0.69	0.68	0.68	0.39	0.77
P_3	0.20	0.19	0.20	0.39	0.20	0.19	0.20	0.39	0.20	0.19	0.20	0.39	0.14
H_0	-	92.8	94.1	0.1	-	89.2	92.5	0.0	-	77.0	86.4	0.0	
H_1	-	6.8	4.7	99.9	-	10.5	6.6	100.0	-	22.9	13.4	100.0	
H_2	-	0.4	1.2	-	-	0.3	0.9	-	-	0.1	0.2	-	

Independence

N	1000				2000				5000				True value
	CBRE			RE	CBRE			RE	CBRE			RE	
	Logit	Logit	Probit	Logit	Logit	Logit	Probit	Logit	Logit	Logit	Probit	Logit	
	Clayton	Frank	Clayton	-	Clayton	Frank	Clayton	-	Clayton	Frank	Clayton	-	
APE	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20
	(0.03)	(0.03)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	
P_1	0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.14
P_2	0.39	0.39	0.39	0.39	0.39	0.39	0.39	0.39	0.39	0.39	0.39	0.39	0.37
P_3	0.39	0.39	0.39	0.39	0.39	0.39	0.39	0.39	0.39	0.39	0.39	0.39	0.37
H_0	-	99.7	94.4	98.1	-	99.8	93.8	97.5	-	100.0	94.0	97.0	
H_1	-	0.1	3.5	1.9	-	0.0	4.6	2.5	-	0.0	4.7	3.0	
H_2	-	0.2	2.1	-	-	0.2	1.6	-	-	0.0	1.3	-	

Notes: Mean estimates of the average partial effects (APE) across 1000 repetitions, standard deviations across repetitions in parentheses. P_1 denotes the average probability across periods that $y_{1gt} = y_{2gt} = 1$, P_2 denotes the average probability across periods that $y_{1gt} = 1$, conditional on $y_{2gt} = 1$, and P_3 denotes the average probability across periods that $y_{1gt} = 1$, conditional on $y_{2gt} = 0$. For the independence test, H_0 denotes the percentage of cases in which the null hypothesis of independence was accepted, and H_1 denotes the percentage of cases in which it was rejected. For the specification tests, H_0 denotes the percentage of cases in which the null hypothesis of equal fit was accepted, H_1 denotes the percentage of cases in which the alternative hypothesis that the Clayton-logit estimator provides a better fit was accepted, and H_2 denotes the percentage of cases in which the alternative hypothesis that the alternative estimator provides a better fit was accepted. For all tests, the confidence level is 95%. The RE estimates were calculated by approximating the integral with a Gauss-Hermite quadrature with 20 points.

with a smaller probability than their husbands, the probability to work decreases with age, and increasing the level of education is correlated with an increase in the probability to work (with the exceptions of Greece and the United Kingdom, where workers with primary or no education have a larger probability of working than those with secondary education, although it is not statistically significant). The number of children smaller than 5 years old typically has a negative effect, particularly on women. However, this effect is not always significantly different from zero. In some countries the effect is positive, though never significant. On the other hand, the coefficient of non-labor income is not significant in all countries.

The standard deviation of the distribution of the time-invariant unobserved propensity to work, η_{ic}, is substantially large and significant in all countries, highlighting its relevance to explain the probability of being employed. Moreover, the estimated copula is statistically different from the independence copula at the 95% confidence level in all countries in the sample, despite the moderate sample size in some of them. The correlation of the individual effects, measured by the Kendall's τ correlation coefficient, is smaller in Northern European countries (Norway, Finland or Poland), with the exception of Denmark; it is larger in Southern European ones (Belgium, Spain or Portugal), with the exception of Italy.

The selected model varies by country. To test whether they are statistically better than the alternatives, I report pairwise Schennach and Wilhelm tests in the first six rows of Table 7. With the exception of the UK, the logit provided a better fit than the probit, and only in four countries the logit is significantly better than the probit when the copula is the same (Denmark, France, the Netherlands and Norway). Regarding the copula, the best fit is achieved with the Clayton in four countries (Denmark, Finland, Hungary and Norway), the Gaussian in another four countries (Greece, the Netherlands, Portugal and Spain) and the Frank in the rest. The test accepts the hypothesis that the selected copula is statistically better than one of the alternative copulas only in one country, France. Hence, for this sample size, CBRE improves the fit of RE from a statistical standpoint, but the copula choice is a secondary concern in most cases.

Table 6: Parametric copula CBRE estimates

	AT	BE	BG	CZ	DK	EL	ES	FI	FR	HU	IT	NL	NO	PL	PT	UK
FE	-1.86	-3.40	-0.86	-1.62	-1.04	-4.99	-4.26	-0.70	-1.62	-0.58	-4.07	-5.24	-1.99	-2.51	-2.88	-0.61
	(0.53)	(0.85)	(0.48)	(0.36)	(0.37)	(0.62)	(0.33)	(0.41)	(0.32)	(0.33)	(0.31)	(0.70)	(0.35)	(0.40)	(0.53)	(0.29)
C5	-0.77	0.11	-3.67	0.12	-0.60	1.01	0.48	-2.10	-0.31	-0.19	0.87	0.30	-1.23	0.69	-0.07	0.87
	(0.91)	(0.78)	(2.98)	(1.73)	(1.15)	(0.67)	(0.38)	(0.67)	(0.55)	(0.51)	(0.46)	(3.58)	(0.64)	(0.51)	(1.32)	(1.22)
C5*FE	-0.83	-0.41	2.77	-4.21	0.67	-1.27	-1.41	-0.46	-1.27	-2.57	-1.90	-1.72	0.53	-1.81	0.01	-2.36
	(1.17)	(1.35)	(4.18)	(1.85)	(1.68)	(0.92)	(0.54)	(0.89)	(0.67)	(0.54)	(0.59)	(3.66)	(0.75)	(0.61)	(1.55)	(1.32)
AGE	-0.27	-0.45	-0.18	-0.28	-0.10	-0.22	-0.13	-0.18	-0.43	-0.21	-0.20	-0.39	-0.10	-0.35	-0.27	-0.06
	(0.04)	(0.07)	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)	(0.05)	(0.02)	(0.03)	(0.03)	(0.02)
SE	0.94	1.39	1.68	2.53	0.30	-0.09	2.03	1.84	1.19	1.69	2.51	2.16	1.99	2.21	1.68	-0.03
	(0.41)	(0.45)	(0.61)	(0.59)	(0.43)	(0.30)	(0.29)	(0.51)	(0.26)	(0.29)	(0.29)	(0.64)	(0.39)	(0.67)	(0.48)	(0.21)
TE	1.82	2.51	4.40	4.46	1.92	3.86	3.86	2.71	4.15	3.65	4.74	4.32	3.27	5.89	4.32	0.45
	(0.61)	(0.48)	(0.80)	(0.78)	(0.50)	(0.55)	(0.31)	(0.58)	(0.39)	(0.42)	(0.46)	(0.63)	(0.42)	(0.79)	(0.81)	(0.24)
IN	0.00	-0.01	0.10	0.04	0.00	0.03	-0.01	0.00	0.00	-0.01	0.00	0.03	0.00	-0.01	0.04	0.00
	(0.01)	(0.01)	(0.66)	(0.05)	(0.01)	(0.05)	(0.03)	(0.02)	(0.00)	(0.10)	(0.02)	(0.02)	(0.01)	(0.18)	(0.11)	(0.02)
$\overline{C5}$	-1.42	-3.78	4.11	-1.50	-1.48	-1.21	-1.21	0.27	-2.46	1.23	-0.20	-3.06	0.58	-1.99	-0.60	-0.76
	(1.49)	(2.36)	(3.43)	(1.87)	(1.55)	(1.31)	(0.74)	(1.22)	(0.89)	(1.13)	(0.79)	(4.14)	(1.15)	(1.09)	(1.96)	(1.32)
$\overline{C5*FE}$	-4.21	-3.15	-9.09	-5.61	0.07	-1.39	0.98	-5.74	-4.98	-7.65	-1.93	-1.28	-2.73	-4.98	-2.49	0.14
	(1.69)	(2.37)	(4.78)	(2.10)	(1.94)	(1.46)	(0.87)	(1.50)	(1.09)	(1.35)	(0.93)	(4.12)	(1.24)	(1.22)	(2.15)	(1.44)
$\overline{IN}$	0.04	-0.01	-0.58	0.04	-0.01	-0.04	0.09	-0.01	0.01	0.09	0.11	-0.03	0.00	-0.03	0.12	-0.06
	(0.02)	(0.06)	(1.41)	(0.15)	(0.01)	(0.08)	(0.07)	(0.05)	(0.01)	(0.62)	(0.04)	(0.03)	(0.01)	(0.09)	(0.22)	(0.05)
$\hat{\sigma}$	4.65	7.20	5.13	4.67	3.46	5.76	4.76	4.09	6.45	5.01	4.77	6.64	4.10	6.10	5.53	2.50
	(0.36)	(0.79)	(0.43)	(0.28)	(0.29)	(0.41)	(0.21)	(0.31)	(0.26)	(0.26)	(0.21)	(0.49)	(0.26)	(0.32)	(0.40)	(0.22)
$\hat{\rho}$	1.80	2.29	1.39	2.01	0.55	0.35	0.42	0.26	1.47	0.41	0.79	0.33	0.19	0.78	0.48	1.24
	[0.00]	[0.00]	[0.02]	[0.00]	[0.00]	[0.00]	[0.00]	[0.01]	[0.00]	[0.00]	[0.00]	[0.01]	[0.01]	[0.00]	[0.00]	[0.03]
Model	Log	Log	Log	Log	Log	Log	Log	Log	Log	Log	Log	Log	Log	Log	Log	Pro
Copula	Fra	Fra	Fra	Fra	Cla	Gau	Gau	Cla	Fra	Cla	Fra	Gau	Cla	Fra	Gau	Fra
τ	0.19	0.24	0.15	0.21	0.22	0.23	0.27	0.11	0.16	0.17	0.09	0.21	0.09	0.09	0.32	0.14
N	764	604	780	1404	696	930	1906	882	3370	1422	2186	1360	1364	1848	814	640

Notes: Standard errors in parentheses for all coefficients except for $\hat{\rho}$, for which I report the p-values in square brackets. FE, C5, AGE, SE, TE, and IN respectively denote female, number of children smaller than 5 years old, age, secondary education, tertiary education, and non-labor income (expressed in thousands of euros); the $\bar{\cdot}$ symbol is used to denote the coefficient of the correlated random effect; Model denotes the best fitting binary choice model: logit (Log) or probit (Pro); Copula denotes the best fitting copula: Clayton (Cla), Frank (Fra), or Gaussian (Gau); τ denotes Kendall's τ correlation coefficient; the contour plots of the copula estimates are shown in Figure 3; N is the sample size of each country. The estimates shown are those of Austria, Belgium, Bulgaria, Czech Republic, Denmark, Greece, Spain, Finland, Hungary, Italy, Netherlands, Norway, Poland, Portugal, and the United Kingdom. The estimates of the remaining countries (Cyprus, Estonia, Iceland, Lithuania, Luxemburg, Latvia, Malta, and Slovenia) are available upon request.

Table 7: Tests and counterfactuals

	AT	BE	BG	CZ	DK	EL	ES	FI	FR	HU	IT	NL	NO	PL	PT	UK
SW_1	79.3	54.8	9.7	6.5	-	23.0	30.6	-	1.3	-	6.6	24.4	-	36.1	19.5	6.7
SW_2	-	-	-	-	39.3	27.7	28.8	70.0	-	19.0	-	31.2	48.5	-	15.4	21.9
SW_3	76.5	50.1	9.7	8.1	38.9	-	-	52.8	4.2	20.8	13.8	-	47.1	65.3	-	10.5
SW_4	62.7	53.0	3.4	6.7	1.0	22.0	7.3	46.9	0.2	8.6	1.8	1.5	1.3	80.8	9.2	6.1
SW_5	15.0	75.6	5.0	11.6	19.3	33.1	5.9	60.6	1.1	14.4	4.6	2.7	9.4	80.4	6.0	-
SW_6	39.6	55.1	2.0	8.3	8.7	40.0	5.3	42.8	0.4	14.1	2.3	1.4	3.4	83.4	10.5	9.1
PW_ρ	72.9	67.4	77.1	74.5	86.7	53.1	55.3	76.4	73.3	65.0	54.5	77.3	82.9	63.8	65.0	77.6
	(1.8)	(3.1)	(1.8)	(2.5)	(1.4)	(1.9)	(1.3)	(1.6)	(0.9)	(1.4)	(1.3)	(1.3)	(1.2)	(1.4)	(1.9)	(2.0)
$CP1_\rho$	76.6	74.7	81.0	77.1	89.0	56.9	58.6	79.0	78.9	70.7	58.6	78.7	84.2	68.7	70.3	79.1
	(1.7)	(2.6)	(1.6)	(2.1)	(1.3)	(2.0)	(1.3)	(1.6)	(0.8)	(1.4)	(1.3)	(1.3)	(1.1)	(1.3)	(1.8)	(1.9)
$CP0_\rho$	51.8	39.7	61.7	54.2	66.6	38.9	40.7	63.0	49.9	49.4	37.8	59.6	67.3	47.7	45.8	67.7
	(2.9)	(3.4)	(2.6)	(3.6)	(2.5)	(2.2)	(1.4)	(2.2)	(1.3)	(1.7)	(1.4)	(2.1)	(1.7)	(1.7)	(2.3)	(2.6)
$\Delta p\left(y_{2ct}\|x_g;\hat{\theta}_{CBRE}\right)$	24.9	35.0	19.3	22.9	22.4	17.9	17.9	16.0	29.0	21.3	20.7	19.1	16.8	21.0	24.5	11.4
	(2.0)	(1.4)	(1.4)	(2.1)	(1.7)	(1.5)	(0.5)	(1.4)	(0.8)	(1.2)	(0.9)	(1.4)	(1.0)	(1.0)	(1.2)	(1.5)
Endowment effect	12.1	13.1	8.8	8.2	3.6	12.5	9.4	7.3	19.5	8.3	14.5	14.6	6.7	16.1	13.6	1.8
	(3.8)	(1.1)	(1.9)	(1.8)	(0.7)	(1.2)	(0.9)	(2.6)	(1.1)	(1.7)	(1.1)	(1.1)	(1.3)	(1.0)	(1.0)	(0.8)
Homophily effect	12.8	21.9	10.5	14.7	18.8	5.4	8.5	8.7	9.5	13.1	6.2	4.5	10.1	4.9	10.9	9.5
	(4.3)	(1.8)	(2.4)	(2.7)	(1.8)	(1.9)	(1.1)	(3.0)	(1.4)	(2.1)	(1.4)	(1.8)	(1.7)	(1.4)	(1.5)	(1.6)
P_ρ	79.9	70.3	77.5	84.8	89.8	65.3	69.0	82.9	75.8	67.9	67.3	90.6	92.2	67.5	69.3	85.6
	(5.4)	(2.8)	(3.8)	(1.6)	(1.7)	(2.4)	(1.5)	(1.8)	(2.4)	(1.6)	(4.5)	(1.2)	(1.2)	(4.6)	(2.1)	(5.2)
P_I	83.7	72.9	81.0	87.8	93.7	65.4	69.4	85.0	78.1	68.6	68.6	91.4	94.2	68.8	69.9	89.1
	(1.9)	(3.9)	(2.1)	(2.5)	(1.5)	(2.6)	(1.6)	(1.7)	(1.0)	(1.8)	(1.7)	(1.0)	(0.9)	(1.7)	(2.3)	(2.3)
Δ_P	-3.7	-2.7	-3.5	-3.0	-3.9	-0.1	-0.4	-2.1	-2.4	-0.7	-1.3	-0.8	-1.9	-1.4	-0.7	-3.5
	(5.0)	(1.4)	(4.0)	(3.3)	(0.7)	(0.2)	(0.2)	(0.6)	(2.1)	(0.3)	(3.5)	(0.6)	(0.7)	(3.6)	(0.3)	(5.1)
N	764	604	780	1404	696	930	1906	882	3370	1422	2186	1360	1364	1848	814	640

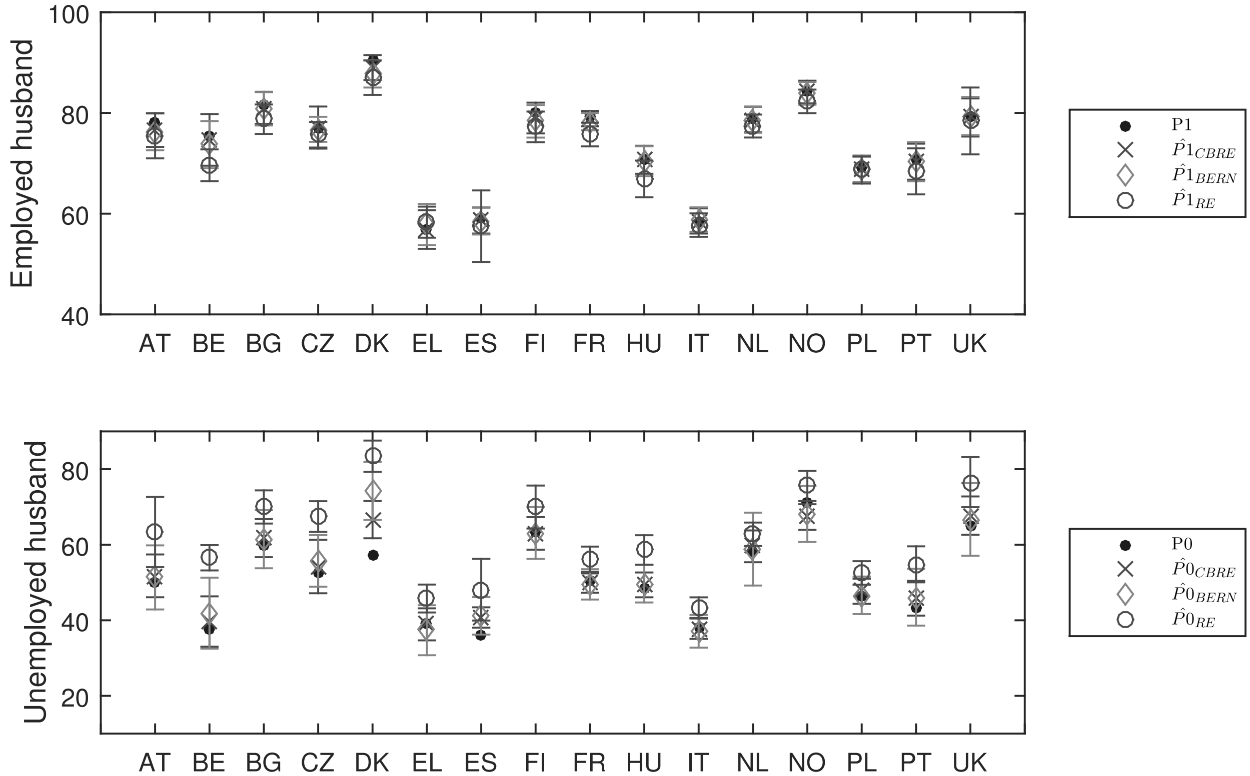
Notes: SW_i denotes the p-value (in %) of the ith Schennach and Wilhelm test of the best fitting model against the remaining alternatives considered: logit (tests 1-3), probit (tests 4-6), Clayton copula (tests 1 and 4), Frank copula (tests 2 and 5), and Gaussian copula (tests 3 and 6). Standard errors of the estimated probabilities were computed using the delta method. PW_ρ denotes the unconditional probability that a wife is employed, $CP1_\rho$ and $CP0_\rho$ respectively denote the probability that a wife is employed conditional on her husband being employed, and conditional on her husband being unemployed; $\Delta p\left(y_{2gt}|x_g;\hat{\theta}_{CBRE}\right)$ and the homophily and endowment effects are defined in the text; P_ρ and P_I respectively denote the probability (in %) that at least one member of the couple was employed in every period when the parameter of the copula is the estimated one and when the copula is independent, and Δ_P denotes the difference between the two; N is the sample size of each country. The estimates shown are those of Austria, Belgium, Bulgaria, Czech Republic, Denmark, Greece, Spain, Finland, Hungary, Italy, Netherlands, Norway, Poland, Portugal, and the United Kingdom. The estimates of the remaining countries (Cyprus, Estonia, Iceland, Lithuania, Luxemburg, Latvia, Malta, and Slovenia) are available upon request.

To get a sense of the importance of this correlation, I use both CBRE and RE to predict the mean probability of a female being employed conditional on the employment status of their husband. The fit provided by RE to the actual proportions observed in the data is much worse than the fit of CBRE (Figure 2), both when the copula used is parametric and when it uses the more flexible Bernstein copula. These two probabilities may differ because of assortative mating in couples. On the one hand, the observed characteristics of both groups of women are different (endowment effect): those married to employed husbands are on average younger, have more young children, and a higher probability of having a university degree. On the other hand, it is also possible that their unobserved individual effects are larger than those of women whose husbands are unemployed (homophily effect): employed men have a higher than average unobserved propensity to work, which is positively correlated with the propensity of their wives. CBRE can account for both types of differences, whereas RE only takes differences in the observed characteristics into account.

A policy maker interested in increasing the employability of women could benefit from knowing the size of these two effects. For example, a way to reduce inequality across households would be to increase the proportion of employed women whose husbands are unemployed. The introduction of a monetary incentive to work for women with a low level of education, given that there is a larger proportion among those married to an unemployed husband, would be more effective if the endowment effect is large. However, if the homophily effect dominates, the effect of such policy would be limited. Moreover, to model female labor supply, it is important to model the participation, which is substantially different depending on which effect is dominant. Formally, the aforementioned difference in probability, given the observed covariates x_c and the parameter vector θ, can be written as

$$\begin{aligned}\Delta p\left(y_{2gt}|x_g;\theta\right) \equiv & \frac{1}{\sum_{g=1}^{G}\mathbf{1}\left(y_{1gt}=1\right)}\sum_{g=1}^{G}\mathbf{1}\left(y_{1gt}=1\right)\mathbb{P}\left(y_{2gt}=1|y_{1gt}=1,x_g;\theta\right)\\ & -\frac{1}{\sum_{g=1}^{G}\mathbf{1}\left(y_{1gt}=0\right)}\sum_{g=1}^{G}\mathbf{1}\left(y_{1gt}=0\right)\mathbb{P}\left(y_{2gt}=1|y_{1gt}=0,x_g;\theta\right)\end{aligned}$$

Figure 2: Conditional probability of being employed for females



Pj, Pj_{CBRE}, Pj_{BERN}, and Pj_{RE} respectively denote the observed percentage of women who are employed, and the estimated analogue using CBRE with a parametric copula, CBRE with a Bernstein copula, and RE, conditional on the employment status of their husbands ($j = 1$ if employed, $j = 0$ if unemployed).

The decomposition is given by:

$$\Delta p\left(y_{2gt}|x_g;\hat{\theta}_{CBRE}\right) = \Delta p\left(y_{2gt}|x_g;\hat{\theta}_{CBRE}\right) - \Delta p\left(y_{2gt}|x_g;\hat{\theta}_{RE}\right) + \Delta p\left(y_{2gt}|x_g;\hat{\theta}_{RE}\right) \tag{12}$$

The first term in Equation 12 is the homophily effect, and is expressed as the difference in the probability of changing the CBRE estimates by the RE estimates. On the other hand, the second term is the endowment effect. The results show a large difference in the estimated probability of a woman being employed depending on the employment status of her husband: in some countries the difference is small because the probability is low for both groups (Greece, Italy and Spain); in others it is small because the probability is always relatively high (Bulgaria, Finland, Norway, the Netherlands, and the UK), and in a third group of countries it is large because the probability is relatively low when the husband is unemployed, and relatively high when he is employed (Austria, Belgium, Czech Republic, Denmark, France, Hungary, Poland, and Portugal).

Both effects explain an important share of the difference in every country shown in Table 7. In particular, the endowment effect explains more than half of the difference in seven countries (France, Greece, Italy, the Netherlands, Poland, Portugal, and Spain), most of which have small female employment rates when the husband is unemployed. Excluding the Netherlands, the homophily effect is dominant in the second group of countries, as well as in many of the third group. In other words, there is a strong correlation between the probability of women being employed, conditional on the husband being employed, with the relative size of the homophily effect.

Finally, I compute the probability that at least one member of the couple was employed in every period, to which I refer as a *working household*, and then I change the estimated copula by the independence copula, obtaining the counterfactual probability when there is no homophily. The estimates are shown in Table 7, and consistently with the large differences in the labor market across countries, the probability of observing a working household has a lot of variation: countries with a low unemployment rate, such as Denmark, the Netherlands, or Norway, have a high probability, whereas countries with high unemployment rate, such as Greece, Spain, or Portugal, have a low probability.

In the counterfactual scenario, this probability would be statistically larger in six countries, reducing the proportion of non-working households. Notice that in those countries in which the selected copula was Frank, the estimates were not significant, whereas when the selected copula was either Clayton or Gaussian, the estimates were significant. Moreover, with the Bernstein copula, the estimates were significant in 14 countries. This evidence suggests that the Frank copula is not accurately estimated. However, this increase is larger in those countries with an already high probability: the largest one would be in Denmark, followed by Austria and Bulgaria. On the other hand, countries with a relatively low probability of having working households (Greece and Spain) would only have a slightly higher probability in the counterfactual scenario. Lastly, the probability change would also be small in the Netherlands, which has the highest probability of having working households.

Some remarks are in order. These results cannot be extrapolated to the whole population, as the characteristics of married couples in working age, both observable and unobservable, differ from those of singles. Moreover, the coefficient ρ does not have a causal interpretation in this example: marrying someone with a higher propensity to work does not imply that the own propensity is changed in any direction, nor marrying people at random would lead to the counterfactual scenario if there are such spillovers inside the marriage. The goal of this exercise is to isolate the contribution of the correlation in the unobserved propensity to work inside couples to the probability of having working households.

7 Conclusion

In this paper I present the CBRE estimator, a random effects estimator for binary choice panel data in which the unobserved heterogeneity of individuals in the same cluster is correlated, and can be used to consistently estimate the probability of joint and conditional events. I study the efficiency of this estimator relative to RE and I consider two types of hypothesis tests: a specification test to select the most appropriate copula, and an independence test when the correlation parameter lies on the boundary of the parameter space.

This paper focuses on a fixed T panel, in which the distribution of individual effects is not nonparametrically identified. Increasing the number of periods would allow to tighten the identification result, and if T grew to infinity at the same rate as the sample size, it would be possible to use bias-correcting methods to jointly estimate the slope parameters and the individual effects, along the lines of Fernández-Val (2009). Future research could investigate the properties of such estimators when the individual effects have the correlation structure considered in this paper, although these methods would require a moderate to large number of time periods.

The computation of the estimator requires the numerical approximation of potentially high-dimensional integrals. To overcome this issue, I propose an algorithm that approximates such integrals for Archimedean copulas. This algorithm does not suffer from the curse of dimensionality, unlike traditional simulation methods such as Monte Carlo integration.

I illustrate the use of the estimator with an empirical application of labor supply of married couples. The results indicate that the unobserved propensity to be employed between the two members of a couple is positively correlated. I use these estimates to decompose the difference in the probability of being employed for women married to either employed or unemployed husbands, into an endowment effect and a homophily effect. Both effects have an explanatory power of similar magnitude, and ignoring the unobserved correlation leads to a substantial underestimation of the impact of assortative mating on the labor supply of married women. Finally, the endowment effect tends to dominate in Southern European countries, while the opposite is true for Northern European ones.

Acknowledgements

I would like to thank the associate editor, two anonymous referees, Lorenzo Burlon, Lorenzo Cappellari, Domenico Depalo, Simone Emiliozzi, Iván Fernández-Val, Giuseppe Ilardi, George Kapetanios, Giovanni Mellace, Marco Savegnago, Laura Sigalotti, Jeffrey Wooldridge, Paolo Zacchia, and seminar participants at Banca d'Italia, EC Joint Research Center, the 2015 EWM of the Econometric Society, the 3rd Annual Conference of the IAAE, the 7th ICEEE Annual Conference, the IPDC 2017 Conference, and the International Workshop on Causal Inference, Program Evaluation and External Validity. All remaining errors are my own. This paper is based on data from Eurostat, EU Statistics on Income and Living Conditions. The views presented in this paper do not necessarily reflect those of the Banca d'Italia, and the

responsibility for all conclusions drawn from the data lies entirely with the author.

References

Amemiya, T. (1981). Qualitative response models: A survey. *Journal of economic literature 19*(4), 1483–1536.

Antweiler, W. (2001). Nested random effects estimation in unbalanced panel data. *Journal of Econometrics 101*(2), 295–313.

Arellano, M. (2003). Discrete choices with panel data. *Investigaciones Económicas XXVII*(3), 423–458.

Arellano, M. and S. Bonhomme (2011). Nonlinear panel data analysis. *Annual Review of Economics 3*(1), 395–424.

Bester, C. A. and C. B. Hansen (2016). Grouped effects estimators in fixed effects models. *Journal of Econometrics 190*(1), 197–208.

Bonhomme, S. (2012). Functional differencing. *Econometrica 80*(4), 1337–1385.

Bonhomme, S. and E. Manresa (2015). Grouped patterns of heterogeneity in panel data. *Econometrica 83*(3), 1147–1184.

Breusch, T., H. Qian, P. Schmidt, and D. Wyhowski (1999). Redundancy of moment conditions. *Journal of econometrics 91*(1), 89–111.

Bruze, G. (2011). Marriage choices of movie stars: Does spouse's education matter? *Journal of Human Capital 5*(1), 1–28.

Butler, J. S. and R. Moffitt (1982). A computationally efficient quadrature procedure for the one-factor multinomial probit model. *Econometrica: Journal of the Econometric Society 50*(3), 761–764.

Cambanis, S., S. Huang, and G. Simons (1981). On the theory of elliptically contoured distributions. *Journal of Multivariate Analysis 11*(3), 368–385.

Chamberlain, G. (1980). Analysis of covariance with qualitative data. *The Review of Economic Studies 47*(1), 225–238.

Chamberlain, G. (1984). Panel data. *Handbook of econometrics 2*, 1247–1318.

Chamberlain, G. (2010). Binary response models for panel data: Identification and information. *Econometrica 78*(1), 159–168.

Chambers, E. A. and D. R. Cox (1967). Discrimination between alternative binary response models. *Biometrika 54*(3-4), 573–578.

Charles, K. K., E. Hurst, and A. Killewald (2013). Marital sorting and parental wealth. *Demography 50*(1), 51–70.

Chernozhukov, V., I. Fernández-Val, J. Hahn, and W. Newey (2013). Average and quantile effects in nonseparable panel models. *Econometrica 81*(2), 535–580.

Embrechts, P., F. Lindskog, and A. McNeil (2001). Modelling dependence with copulas. Technical report, Institut Fédéral de Technologie de Zurich, Zurich.

Escanciano, J. C. (2016). Semiparametric identification and fisher information. Unpublished working paper.

Fernández-Val, I. (2009). Fixed effects estimation of structural parameters and marginal effects in panel probit models. *Journal of Econometrics 150*(1), 71–85.

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

Geweke, J. (1989). Bayesian inference in econometric models using monte carlo integration. *Econometrica: Journal of the Econometric Society 57*(6), 1317–1339.

Graham, B. S. and J. L. Powell (2012). Identification and estimation of average partial effects in "irregular" correlated random coefficient panel data models. *Econometrica 80*(5), 2105–2152.

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

Hajivassiliou, V., D. McFadden, and P. Ruud (1996). Simulation of multivariate normal rectangle probabilities and their derivatives theoretical and computational results. *Journal of econometrics 72*(1), 85–134.

Hanushek, E. (1971). Teacher characteristics and gains in student achievement: Estimation using micro data. *The American Economic Review 61*(2), 280–288.

Heiss, F. and V. Winschel (2008). Likelihood approximation by numerical integration on sparse grids. *Journal of Econometrics 144*(1), 62–80.

Lee, M.-J. (1999). A root-n consistent semiparametric estimator for related-effect binary response panel data. *Econometrica 67*(2), 427–433.

Lin, C.-C. and S. Ng (2012). Estimation of panel data models with parameter heterogeneity when group membership is unknown. *Journal of Econometric Methods 1*(1), 42–55.

Manski, C. F. (1975). Maximum score estimation of the stochastic utility model of choice. *Journal of Econometrics 3*(3), 205–228.

Manski, C. F. (1987). Semiparametric analysis of random effects linear models from binary panel data. *Econometrica: Journal of the Econometric Society 55*(2), 357–362.

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *The Review of Economic Studies 60*(3), 531–542.

Marshall, A. W. and I. Olkin (1988). Families of multivariate distributions. *Journal of the American Statistical Association 83*(403), 834–841.

Newey, W. K. and D. McFadden (1994). Large sample estimation and hypothesis testing. *Handbook of econometrics 4*, 2111–2245.

Prokhorov, A., U. Schepsmeier, and Y. Zhu (2019). Generalized information matrix tests for copulas. *Econometric Reviews 38*(9), 1024–1054.

Prokhorov, A. and P. Schmidt (2009). Likelihood-based estimation in a panel setting: robustness, redundancy and validity of copulas. *Journal of Econometrics 153*(1), 93–104.

Sancetta, A. and S. Satchell (2004). The bernstein copula and its applications to modeling and approximations of multivariate distributions. *Econometric theory 20*(03), 535–562.

Schennach, S. M. and D. Wilhelm (2017). A simple parametric model selection test. *Journal of the American Statistical Association 112*(520), 1663–1674.

Self, S. G. and K.-Y. Liang (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association 82*(398), 605–610.

Shi, X. (2015). A nondegenerate vuong test. *Quantitative Economics 6*(1), 85–121.

Sklar, M. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Istitut de Statistique de l'Universitè de Paris 8*, 229–231.

Skrainka, B. S. and K. L. Judd (2011). High performance quadrature rules: How numerical integration affects a popular model of product differentiation. Technical report, SSRN.

Trivedi, P. K., D. M. Zimmer, et al. (2007). *Copula modeling: an introduction for practitioners*. Now Publishers, Inc.

Vuong, Q. H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica: Journal of the Econometric Society 57*(2), 307–333.

Appendix

A Mathematical proofs

A.1 Proof to Proposition 1

Define

$$
\begin{aligned}
D_{\mu\mu}^{RE} &\equiv -\mathbb{E}\left[\nabla_{\mu\mu}\sum_{i=1}^{N_g}\log\left(\ell_{ig}\left(z_{ig};\mu_0\right)\right)\right] \\
D_{\mu\theta}^{RE} &\equiv \left[D_{\mu\mu}^{RE},0\right] \\
D_{\theta\theta}^{CBRE} &\equiv -\mathbb{E}\left[\nabla_{\theta\theta}\log\left(\ell_g\left(z_g;\mu_0\right)\right)\right] \\
\Omega_{\mu}^{RE} &\equiv \mathbb{E}\left[\nabla_{\mu}\sum_{i=1}^{N_g}\log\left(\ell_{ig}\left(z_{ig};\mu_0\right)\right)\nabla_{\mu}\sum_{i=1}^{N_g}\log\left(\ell_{ig}\left(z_{ig};\mu_0\right)\right)'\right] \\
\Omega_{\theta}^{CBRE} &\equiv \mathbb{E}\left[\nabla_{\theta}\log\left(\ell_g\left(z_g;\mu_0\right)\right)\nabla_{\theta}\log\left(\ell_g\left(z_g;\mu_0\right)\right)'\right] \\
\Omega_{\theta}^{CBRE,RE} &\equiv \mathbb{E}\left[\nabla_{\theta}\log\left(\ell_g\left(z_g;\mu_0\right)\right)\nabla_{\mu}\sum_{i=1}^{N_g}\log\left(\ell_{ig}\left(z_{ig};\mu_0\right)\right)'\right] \\
\Omega_{\theta}^{RE,CBRE} &\equiv \Omega_{\theta}^{CBRE,RE'}
\end{aligned}
$$

Result 1 is a well known result, so its proof is omitted.

For result 2, note that by the optimality property of MLE, and since RE is a particular case of ICBRE when the copula is independent, the asymptotic variance of RE cannot be smaller than that of ICBRE. Regarding CBRE, note that the asymptotic variance of μ is given by $\Sigma_{\mu}^{CBRE} = \left(D_{\mu\mu}^{CBRE} - D_{\mu\rho}^{CBRE}D_{\rho\rho}^{CBRE,-1}D_{\rho\mu}^{CBRE}\right)^{-1}$. $\Sigma_{\mu}^{ICBRE,-1} - \Sigma_{\mu}^{CBRE,-1} = D_{\mu\rho}^{CBRE}D_{\rho\rho}^{CBRE,-1}D_{\rho\mu}^{CBRE}$ is a positive semidefinite quadratic form. Therefore, $\Sigma_{\mu}^{CBRE} - \Sigma_{\mu}^{ICBRE}$ is positive semidefinite and the result follows.

By Theorem 1 in Breusch et al. (1999), the scores from RE are redundant for the estimation of θ if there exists some matrix A such that $D_{\theta\theta}^{CBRE} = \Omega_{\theta}^{CBRE}A$ and $D_{\mu\theta}^{RE} = \Omega_{\theta}^{RE,CBRE}A$. By the information equality, $D_{\theta\theta}^{CBRE} = \Omega_{\theta}^{CBRE}$, so A is the identity matrix. Therefore, if $D_{\mu\mu}^{RE} = \Omega_{\mu}^{RE,CBRE}$, result 3 follows.

To show result 4, note that under independence,

$$
\begin{aligned}
& D_{\mu\mu}^{RE,-1} = D_{\mu\mu}^{CBRE,-1} \\
& \Leftrightarrow \mathbb{E}\left[\nabla_{\mu\mu}\sum_{i=1}^{N_g}\log\left(\ell_{ig}\left(z_{ig};\mu_0\right)\right) - \nabla_{\mu\mu}\log\left(\ell_g\left(z_g;\mu_0\right)\right)\right] = 0 \\
& \Leftrightarrow \mathbb{E}\left[\int\left(\nabla_{\mu\mu}P_g\left(z_g,u_g;\mu_0\right) - \nabla_\mu P_g\left(z_g,u_g;\mu_0\right)\nabla_\mu P_g\left(z_g,u_g;\mu_0\right)'\right)\cdot\right. \\
& \left.\left(\frac{1}{\int P_g\left(z_g,u_g;\mu_0\right)\prod_{j=1}^{N_g}du_{jg}} - \frac{c_X\left(u_g;\rho_0\right)}{\int P_g\left(z_g,u_g;\mu_0\right)c_X\left(u_g;\rho_0\right)\prod_{j=1}^{N_g}du_{jg}}\right)\prod_{i=1}^{N_g}du_{ig}\right] = 0 \\
& \Leftrightarrow c_X\left(u_g;\rho_0\right) = 1 \\
& \Leftrightarrow C_X\left(u_g;\rho_0\right) = \prod_{i=1}^{N_g}u_{ig}
\end{aligned}
$$

where $P_g\left(z_g,u_g;\mu_0\right) \equiv \prod_{i=1}^{N_g}P_{ig}\left(z_{ig},u_{ig};\mu_0\right)$. Moreover, $D_{\mu\mu}^{RE} = \Omega_\mu^{RE}$, so $\Sigma_\mu^{RE} = D_{\mu\mu}^{RE,-1} = \Sigma_\mu^{ICBRE}$. Moreover, by Theorem 8 in Breusch et al. (1999), the scores of CBRE are partially redundant for μ if there exists some matrix R such that $D_{\theta\mu}^{CBRE} - \Omega_\theta^{CBRE,RE}\Omega_\mu^{RE,-1}D_{\mu\mu}^{RE} = D_{\theta\rho}^{CBRE}R$. Under independence, $D_{\mu\mu}^{RE} = \Omega_\mu^{RE}$, $D_{\mu\mu}^{CBRE} = \Omega_\mu^{CBRE,RE} = D_{\mu\mu}^{RE}$, and $D_{\rho\mu}^{CBRE} = \Omega_\rho^{CBRE,RE}$, so $D_{\theta\rho}^{CBRE}R = 0$, which is satisfied if $R = 0$. Hence, the asymptotic variance of ACBRE and RE is the same.

Finally, for result 5, if $D_{\mu\rho}^{CBRE} = 0$, then $\Sigma_\mu^{CBRE} = D_{\mu\mu}^{CBRE,-1} = \Sigma_\mu^{ICBRE}$.

A.2 Proof to Proposition 2

Define $D_\mu^{IRE} \equiv -\mathbb{E}\left[s_g\left(\mu_0\right)\right]$ and $\Omega_\mu^{IRE} \equiv \left[s_g\left(\mu_0\right)s_g\left(\mu_0\right)\right]$. The asymptotic variances of IRE and RE are given by $\Sigma_\mu^{IRE} = \left(D_{\mu\mu}^{IRE}\Omega_\mu^{IRE,-1}D_{\mu\mu}^{IRE'}\right)^{-1}$ and $\Sigma_\mu^{RE} = \left(AD_{\mu\mu}^{IRE}\left(A\Omega_\mu^{IRE}A'\right)^{-1}D_{\mu\mu}^{IRE'}A'\right)^{-1}$, where $A = \iota_{\overline{N}}' \otimes I_k$, $\iota_{\overline{N}}$ is an $\overline{N}$ dimensional vector of ones, I_k is the identity matrix of dimension k, and $k = dim\left(\mu\right)$. This setup is equivalent to the case considered in Theorem 1 in Prokhorov and Schmidt (2009), so the first result of the Proposition follows.

Under independence, Ω_μ^{IRE} is a block diagonal matrix in which the ith block equals $\mathbb{E}\left[\nabla_\mu\log\left(\ell_{ig}\left(z_{ig};\mu_0\right)\right)\nabla_\mu\log\left(\ell_{ig}\left(z_{ig};\mu_0\right)\right)'\right]$. This, together with the information inequality implies $\Sigma_\mu^{IRE} = \left(D_{\mu\mu}^{RE}\Omega_\mu^{RE,-1}D_{\mu\mu}^{RE'}\right)^{-1} = D_{\mu\mu}^{RE,-1} = \Sigma_\mu^{RE}$.

Supplementary Material

S1 Partial Identification

As pointed out by Arellano (2003), identification in a binary choice panel setup is fragile, and it usually hinges on assumptions that are either not satisfied or impossible to verify. Such as when the model is a logit (Chamberlain, 1984, 2010), if a regressor has unbounded support, in which case β_0 is point-identified (Manski, 1987), or if the support of this distribution is finite (Bonhomme, 2012). Chernozhukov et al. (2013) showed that when the covariates have discrete support, the marginal distribution of the individual effects is not identified. The following lemma extends this result to the lack of identification of their copula:

Lemma 1. *Assume that the distribution of X_{igt} is discrete with finite support, and let $\mathbb{P}(Y_g|X_g) = \int_{\mathcal{Y}} P_g(u_g;\mu)\, dC_X(u_g;\rho)$, where $P_g(u_g;\mu)$ is defined as in the main text and is a measurable function of u_g for each $\mu \in M$, and $\mathcal{Y}$ denotes the support of $\eta(u_g|x_g;\sigma_0)$. Then, for each β, every marginal distribution $F_\eta(\eta_{ig}|x_{ig};\sigma)$ on the support of η_{ig}, and every copula $C(u_g;\rho)$ on $[0,1]^N$, there exists a discrete distribution $F_\eta^{k,N,T}$ with no more than 2^{NT} support points such that $\int_{\mathcal{Y}} P_g(z_g, u_g;\mu)\, dC_X(u_g;\rho) = \int_{\mathcal{Y}} P_g(z_g, u_g;\mu)\, dF_\eta^{k,N,T}(\eta_g)$.*

Proof. By the definition of the copula, there exists a multivariate distribution $\tilde{F}_\eta$ such that $C_X(u_g;\rho) = \tilde{F}_\eta(u_g|x_g;\sigma,\rho)$. For each $k = 1,...,K$ of the possible values that the vector $(X_{11},...,X_{NT})$ can take, there are $J = 2^{NT}$ distinct values that the vector $(Y_{11},...,Y_{NT})$ can take. Apply lemma 7 in Chernozhukov et al. (2013) to $\int_{\mathcal{Y}} P_g(z_g,\eta_g;\beta)\, d\tilde{F}_\eta(\eta_g;\sigma,\rho)$ to obtain the desired result. □

S2 Asymptotic Distribution

To derive the asymptotic properties of the estimator, I consider the following assumptions:

Assumption 2. $\theta \neq \theta_0 \Rightarrow \ell_g(z_g;\theta) \neq \ell_g(z_g;\theta_0)$.

Assumption 3. $\theta \in int\Theta$, *where* Θ *is compact.*

Assumption 4. $\ell_{ig}(z_{ig};\mu)$ *is continuous for all* $\theta \in \Theta$.

Assumption 5. $\mathbb{E}\left[\sup_{\theta\in\Theta}\left|\log\left(\ell_{ig}(z_{ig};\mu)\right)\right|\right] < \infty$.

Assumption 6. $\ell_{ig}(z_{ig};\mu)$ *is twice continuously differentiable with respect to* θ; $\ell_{ig}(z_{ig};\mu) > 0$ *in a neighborhood* $\mathcal{N}$ *of* θ_0.

Assumption 7. $\int \sup_{\theta\in\mathcal{N}} \left\|\nabla_\mu \ell_{ig}(z_{ig};\mu)\right\| dz_{ig} < \infty$, $\int \sup_{\theta\in\mathcal{N}} \left\|\nabla_{\mu\mu} \ell_{ig}(z_{ig};\mu)\right\| dz_{ig} < \infty$.

Assumption 8. $\Sigma_\theta^{CBRE} \equiv \mathbb{E}\left[\nabla_\theta \log\left(\ell_g(z_g;\theta)\right) \nabla_\theta \log\left(\ell_g(z_g;\theta)\right)'\right]$ *exists and is nonsingular.*

Assumption 9. $\mathbb{E}\left[\sup_{\theta\in\mathcal{N}}\left\|\nabla_{\theta\theta}\log\left(\ell_g(z_g;\theta)\right)\right\|\right] < \infty$.

Assumption 10. *The copula has pdf* $c_X(u_g;\rho)$ *which is twice continuously differentiable in* ρ. *Moreover,* $\int_{[0,1]^{N_g}} \left|\nabla_\rho c_X(u_g;\rho)\right| \prod_{i=1}^{N_g} du_{ig} < \infty$ *and* $\int_{[0,1]^{N_g}} \left|\nabla_{\rho\rho} c_X(u_g;\rho)\right| \prod_{i=1}^{N_g} du_{ig} < \infty$.

Assumption 11. $\int \sup_{\theta\in\mathcal{N}} \left\|\nabla_\rho \int_{[0,1]^{N_g}} c_X(u_g;\rho) \prod_{i=1}^{N_g} du_{ig}\right\| dz_g < \infty$ *and* $\int \sup_{\theta\in\mathcal{N}} \left\|\int_{[0,1]^{N_g}} \nabla_{\rho\rho} c_X(u_g;\rho) \prod_{i=1}^{N_g} du_{ig}\right\| dz_g < \infty$.

Assumption 12. *Cluster size is either predetermined, or it is drawn from a distribution with bounded support, independently of all other variables:* $N_g \sim F_N(n)$ $n \in \left\{1,...,\overline{N}\right\}$, *for some* $\overline{N} \in \mathbb{N}$.

Assumption 2 is the identification condition. It is hard to verify, and a necessary condition is that the number of parameters of the distribution of the random effects (σ,ρ) is not too large relative to the number of time periods and individuals in a cluster. To be more specific, and assuming that the results are conditional on $X = x$, there are $2^{N_g T}$ distinct results of the outcome variable. This implies that there are $2^{N_g T} - 1$ probabilities that vary freely. Consider the matrix in which each column contains the derivatives of these probabilities with respect to each parameter of the distribution of random effects. Assumption 2 is satisfied when the rank of this matrix equals the number of parameters, so it cannot be larger than $2^{N_g T} - 1$. Moreover, if the probabilities satisfy an exchangeability condition, *i.e.* if $\mathbb{P}(Y_1 = y_1, Y_0 = y_0) = \mathbb{P}(Y_1 = y_0, Y_0 = y_1)$ and similarly when the cluster dimension is higher, then the number of probabilities that vary freely is smaller. Specifically, it would be equal to $\sum_{i_{1,1}=0}^{1} ... \sum_{i_{N_g,T}=0}^{1} \mathbf{1}\left(i_{1,1} \leq i_{1,2} \leq ... \leq i_{1,T} \leq ... \leq i_{N_g,1} \leq ... \leq i_{N_g,T}\right)$. Consequently, the maximum number of parameters is at most this number.

Assumptions 1 to 9 mimic the assumptions in theorems 2.5 and 3.3 in Newey and McFadden (1994). With some small modifications, these assumptions work for standard RE estimators. In other words, they allow us to extend any RE estimator to have the cluster dependence described in Section 2. It would be possible to relax some of these assumptions, but it could result in non-standard properties. For example, if Assumption 3 allowed the true value of the parameter lies at the boundary of the parameter space, the asymptotic distribution of $\hat{\theta}$ could be a mixture.

Assumptions 10 and 11 impose smoothness restrictions on the copula, as well as bounds on some functionals of its derivative with respect to the copula parameter. It covers the independence case in which its density equals one everywhere, but not the perfect correlation case in which the copula has no proper density. Assumption 12 limits cluster size to $\overline{N}$, ruling out the possibility that the size of a group grows to infinity as the sample size grows. This assumption is required to bound the likelihood function, and it should be satisfied in most applications. Regarding its independence with respect to all other variables, it could be relaxed at the cost of complicating the analysis.

The following proposition establishes the asymptotic distribution of the CBRE estimator:

Proposition 3. *Under Assumptions 1 to 12, the CBRE estimator $\hat{\theta}$ is a consistent estimator for θ_0 and its asymptotic distribution is given by $\sqrt{G}\left(\hat{\theta} - \theta_0\right) \xrightarrow{d} \mathcal{N}\left(0, \Sigma_{\theta}^{CBRE}\right)$.*

Proof. The proposition is shown by checking that Assumptions 1 to 12 satisfy the assumptions in theorems 2.5 and 3.3 in Newey and McFadden (1994). Rather than considering the data *iid* at the individual level, I do it at the cluster level. Begin with the consistency result.

By Assumptions 10 and 12,

$$\ell_g\left(z_g; \theta\right) = \int_{[0,1]^{N_g}} \prod_{i=1}^{N_g} P_{ig}\left(z_{ig}, u_{ig}; \mu\right) dC_X\left(u_g; \rho\right) < \int_{[0,1]^{N_g}} dC_X\left(u_g; \rho\right) = 1$$

So $\ell_g\left(z_{ig}; \theta\right)$ is well defined and finite. By Assumptions 4 and 10, for any sequence

$\theta_n : \theta_n \to \theta$, $\prod_{i=1}^{N_g} P_{ig}(z_{ig}, u_{ig}; \mu_n) c_X(u_g; \rho_n) \to \prod_{i=1}^{N_g} P_{ig}(z_{ig}, u_{ig}; \mu) c_X(u_g; \rho)$ for almost every u_g. Thus, by the dominated convergence theorem, $\ell_g(z_g; \theta_n) \to \ell_g(z_g; \theta)$, so $\ell_g(z_g; \theta)$ is continuous with respect to θ. By a similar argument, $-\infty < \log(\ell_g(z_g; \theta)) < 0$. To get the lower bound, notice that $\ell_g(z_g; \theta) > 0 \Leftrightarrow \exists u_g : \prod_{i=1}^{N_g} P_{ig}(z_{ig}, u_{ig}; \mu) c_X(u_g; \rho) > 0$. By Assumption 5, $\exists u_{ig}$ be such that $P_{ig}\left(z_{ig}, u_{ig}^*\right) > 0 \forall i$. By Assumption 10, the marginals must integrate to 1, and the copula is continuous, so it has a proper pdf. Hence, $\exists u_{1c}^*$: $c_X\left(u_{1g}^*, u_{2g}, ..., u_{N_g g}\right) > 0 \forall u_{2g}, ..., u_{N_g g}$. Therefore, $\mathbb{E}\left[\sup_{\theta \in \Theta} \left|\log(\ell_g(z_g; \theta))\right|\right] < \infty$. These two results, together with Assumptions 1 to 3, verify the conditions in theorem 2.5 in Newey and McFadden (1994) and hence $\hat{\theta} \xrightarrow{P} \theta_0$.

By Assumptions 10, and 12,

$$\begin{aligned}
\nabla_\mu \ell_g(z_g; \theta) &= \sum_{i=1}^{N_g} \int_{[0,1]^{N_g}} \nabla_\mu P_{ig}(z_{ig}, u_{ig}; \mu) \prod_{j \neq i} P_{jg}(z_{jg}, u_{jg}; \mu)\, dC_X(u_g; \rho) \\
&= \sum_{i=1}^{N_g} \int_0^1 \nabla_\mu P_{ig}(z_{ig}, u_{ig}; \mu) \left[\int_{[0,1]^{N_g - 1}} \prod_{j \neq i} P_{jg}(z_{jg}, u_{jg}; \mu)\, dC_X(u_{-ig} | u_{ig}; \rho)\right] du_{ig}
\end{aligned}$$

for all $\theta \in \mathcal{N}$, where u_{-ig} denotes the set $\{u_{jg}\}_{j \neq i}$. Note that the term in square brackets is a probability, and hence it takes values on the unit interval. Hence, by Assumption 7,

$$\int \sup_{\theta \in \mathcal{N}} \|\nabla_\mu \ell_g(z_g; \theta)\|\, dz_g < \overline{N} \max_{i=1,...,N_g} \int \sup_{\theta \in \mathcal{N}} \|\nabla_\mu \ell_{ig}(z_{ig}; \mu)\|\, dz_{ig} < \infty$$

By Assumptions 10 and 12

$$\begin{aligned}
\nabla_\rho \ell_g(z_g; \theta) &= \int_{[0,1]^{N_g}} \prod_{i=1}^{N_g} P_{ig}(z_{ig}, u_{ig}; \mu) \nabla_\rho c_X(u_g; \rho)\, du_{ig} \\
&\leq \int_{[0,1]^{N_g}} \prod_{i=1}^{N_g} P_{ig}(z_{ig}, u_{ig}; \mu) \left|\nabla_\rho c_X(u_g; \rho)\right| du_{ig} \\
&\leq \int_{[0,1]^{N_g}} \left|\nabla_\rho c_X(u_g; \rho)\right| \prod_{i=1}^{N_g} du_{ig} < \infty
\end{aligned}$$

This, together with assumption 11 implies $\int \sup_{\theta \in \mathcal{N}} \|\nabla_\rho \ell_g(z_g; \theta)\|\, dz_g < \infty$.

By a parallel argument, the second derivatives can be bounded. Consequently, it follows that $\int \sup_{\theta\in\mathcal{N}} \|\nabla_\theta \ell_g(z_g;\theta)\| \, dz_g < \infty$ and $\int \sup_{\theta\in\mathcal{N}} \|\nabla_{\theta\theta} \ell_g(z_g;\theta)\| \, dz_g < \infty$. Taking this result, together with Assumptions 1, 3, 6, and 8, and 9, the conditions in Theorem 3.3 in Newey and McFadden (1994) are verified and $\sqrt{G}\left(\hat{\theta}-\theta_0\right) \xrightarrow{d} \mathcal{N}\left(0, \Sigma_\theta^{CBRE}\right)$.

□

S3 Estimation of Average Partial Effects

Frequently, the researcher is interested in the estimation of the APE rather than the slope coefficients. The APE is defined as the marginal effect that increasing the jth regressor $x_{igt,j}$ would have on the probability of the dependent variable being equal to one, averaged over the whole population. Mathematically,

$$APE\left(x_{igt,j}\right) \equiv \int_{\mathbb{R}} \frac{\partial}{\partial x_{igt,j}} \mathbb{P}\left(y_{igt}=1|x_{igt},\eta_{ig}\right) dF_\eta\left(\eta_{ig}|x_{ig};\sigma_0\right) \tag{13}$$

Since it just depends on the marginal distribution of η_{ig}, there is no need to know the copula to identify them, and it can be computed using the sample analogue of Equation 13. It is worth highlighting that the APE depend on the parametric assumptions. See, for instance, Graham and Powell (2012), Chernozhukov et al. (2013), Fernández-Val and Lee (2013), or Escanciano (2016) for discussions on the identification and estimation of APE in this and other related frameworks.

S4 Score and Hessian

Let F_{igt} and f_{igt} be shorthand for $F_\varepsilon\left(-\left(\eta\left(u_{ig}|x_{ig};\sigma\right)+x'_{igt}\beta\right)\right)$ and $f_\varepsilon\left(-\left(\eta\left(u_{ig}|x_{ig};\sigma\right)+x'_{igt}\beta\right)\right)$, denote the quantile function of $\eta\left(u_{ig}|x_{ig};\sigma\right)$ by $Q_\eta\left(u|x;\sigma\right) \equiv F_\eta^{-1}\left(u|x;\sigma\right)$ and by $q_\eta\left(u|x;\sigma\right)$

its derivative with respect to σ. Then, the score is given by

$$\frac{\partial\mathcal{L}(\theta)}{\partial\beta}=\sum_{g=1}^{G}\frac{\int_{[0,1]^{N_g}}P_g(z_g,u_g;\mu)\sum_{i=1}^{N_g}\sum_{t=1}^{T}\frac{f_{igt}}{F_{igt}(1-F_{igt})}(y_{igt}-(1-F_{igt}))\,x_{igt}dC_X(u_g;\rho)}{\int_{[0,1]^{N_g}}P_g(z_g,u_g;\mu)\,dC_X(u_g;\rho)} \tag{14}$$

$$\frac{\partial\mathcal{L}(\theta)}{\partial\sigma}=\sum_{g=1}^{G}\frac{\int_{[0,1]^{N_g}}P_g(z_g,u_g;\mu)\sum_{i=1}^{N_g}\sum_{t=1}^{T}\frac{f_{igt}}{F_{igt}(1-F_{igt})}(y_{igt}-(1-F_{igt}))\,q_\eta(u_{ig}|x_{ig};\sigma)\,dC_X(u_g;\rho)}{\int_{[0,1]^{N_g}}P_g(z_g,u_g;\mu)\,dC_X(u_g;\rho)} \tag{15}$$

$$\frac{\partial\mathcal{L}(\theta)}{\partial\rho}=\sum_{g=1}^{G}\frac{\int_{[0,1]^{N_g}}P_g(z_g,u_g;\mu)\,\frac{\partial c_X(u_g;\rho)}{\partial\rho}\prod_{i=1}^{N_g}du_{ig}}{\int_{[0,1]^{N_g}}P_g(z_g,u_g;\mu)\,dC_X(u_g;\rho)} \tag{16}$$

Note that if η belongs to a scale family of distributions, *i.e.* if $\eta=\sigma\tilde{\eta}_{ig}$, where $\tilde{\eta}_{ig}\sim F_\eta(1)$, then $Q_\eta(u_{ig})=\sigma\tilde{\eta}_{ig}$, and thus $q_\eta(u_{ig};\sigma)=\tilde{\eta}_{ig}$. It is immediate to approximate Equations 14 and 15 using the proposed algorithm presented in this paper. Regarding Equation 16, it is more convenient to numerically evaluate the derivative, *i.e.* $\frac{\partial\mathcal{L}(\theta)}{\partial\rho}\approx\frac{\mathcal{L}(\mu,\rho+\varepsilon)-\mathcal{L}(\mu,\rho)}{\epsilon}$. Finally, the Hessian is estimated by

$$\hat{H}\left(\hat{\theta}\right)=\frac{1}{G}\sum_{g=1}^{G}\frac{\partial\log\left(\hat{\ell}_g\left(z_g;\hat{\theta}\right)\right)}{\partial\theta}\frac{\partial\log\left(\hat{\ell}_g\left(z_g;\hat{\theta}\right)\right)}{\partial\theta'}$$

S5 Schennach and Wilhelm Test for Copulas

Consider two different parametric copulas, $C_{X,1}(u_g;\rho_1)$ and $C_{X,2}(u_g;\rho_2)$, where both ρ_1 and ρ_2 belong to the interior of their respective parameter spaces. Denote their respective likelihoods by $\ell_{1,g}(z_g;\theta_1)$ and $\ell_{2,g}(z_g;\theta_2)$, where $\theta_1\equiv(\mu',\rho_1')'$ and $\theta_2\equiv(\mu',\rho_2')'$ and let

$\omega_g\left(\hat{\epsilon}_g\right) = 1 + \hat{\epsilon}_G \mathbf{1}$ (g is even). $\hat{\epsilon}_G$ is chosen as suggested in Schennach and Wilhelm (2017):

$$\hat{\epsilon}_G = \left(\frac{\hat{C}_{SD}}{\hat{C}^*_{PL}}\right)^{1/3} N^{-1/6} \left(\ln \ln N\right)^{1/3}$$

where the constants $\hat{C}_{SD}$ and $\hat{C}^*_{PL}$ are defined in Schennach and Wilhelm (2017). Note that the choice of $\hat{\epsilon}_G$ faces a trade-off between the power and the size of the test, depending on the true model. Then, the test statistic is given by:

$$\tilde{t}_G = \frac{1}{\sqrt{G}\hat{\hat{\sigma}}} \sum_{g=1}^{G} \omega_g\left(\hat{\epsilon}_G\right) \log\left(\ell_{1,g}\left(z_g; \hat{\theta}_1\right)\right) - \omega_g\left(\hat{\epsilon}_G\right) \log\left(\ell_{2,g}\left(z_g; \hat{\theta}_2\right)\right)$$

where

$$\begin{aligned}\hat{\hat{\sigma}}^2 \equiv & \frac{1}{G}\sum_{g=1}^{G}\left[\omega_g\left(\hat{\epsilon}_G\right) \log\left(\ell_{1,g}\left(z_g; \hat{\theta}_1\right)\right) - \omega_g\left(\hat{\epsilon}_G\right) \log\left(\ell_{2,g}\left(z_g; \hat{\theta}_2\right)\right)\right]^2 \\ & - \left[\frac{1}{G}\sum_{g=1}^{G} \omega_g\left(\hat{\epsilon}_G\right) \log\left(\ell_{1,g}\left(z_g; \hat{\theta}_1\right)\right) - \omega_g\left(\hat{\epsilon}_G\right) \log\left(\ell_{2,g}\left(z_g; \hat{\theta}_2\right)\right)\right]^2\end{aligned}$$

This test has a limiting normal distribution, uniformly over the subset of distributions that satisfy the null hypothesis of equal fit (Theorem 2 in Schennach and Wilhelm (2017)). Given a size α for the test, the null hypothesis is not rejected if the test is between the $\alpha/2$ and $1 - \alpha/2$ quantiles of the normal distribution, whereas the alternative hypothesis that C_1 is better than C_2 and the other way around are not rejected if the test is below the $\alpha/2$ or above the $1 - \alpha/2$ quantiles, respectively

S6 Elliptical Copulas

Elliptical copulas (Cambanis et al., 1981) constitute one of the major parametric families of copulas, including two of the most widely used copulas, the Gaussian and the *t*. The algorithm proposed in Section 4 cannot be adapted to these copulas if some restrictions on the correlation structure are imposed.

Let R denote the correlation matrix of a d-variate normal distribution and Φ_R its cdf. The Gaussian copula with correlation R is given by $C(u;R) = \Phi_R\left(\Phi^{-1}(u_1), ..., \Phi^{-1}(u_d)\right)$. If R is positive definite, then it is possible to obtain the Cholesky decomposition, denoted by A. It is possible to express the Gaussian copula in terms of d independent normal distributions and the coefficients of A (Embrechts et al., 2001), where the (i,j) element is denoted by a_{ij}. Hence, it is possible to rewrite the integral that is required to evaluate the likelihood as

$$\mathcal{I} = \int_{[0,1]^d} \prod_{i=1}^{d} \ell_i(u_i)\, dC(u;R) = \int_{[0,1]^d} \prod_{i=1}^{d} \ell_i\left(\Phi\left(\sum_{j=1}^{i} a_{ji}\Phi^{-1}(v_j)\right)\right) \prod_{j=1}^{d} dv_j$$

The likelihood can be decomposed into d independent random variables. However, the dimensionality of the integral is not reduced as it was the case with the Archimedean copulas.

A similar reformulation of the integral for the t (or any other elliptical) copula is possible: denote by $t_{\nu,R}$ the cdf of the d-variate t distribution with ν degrees of freedom and correlation matrix R, then the t copula is given by $C(u;\nu,R) = t_{\nu,R}\left(t_\nu^{-1}(u_1), ..., t_\nu^{-1}(u_d)\right)$. Again, if R is positive definite, and following Embrechts et al. (2001), the copula can be written in terms of d independent normal variables and a χ^2 with ν degrees of freedom, and its cdf is denoted by F_ν. The integral $\mathcal{I}$ is then given by

$$\mathcal{I} = \int_{[0,1]^d} \prod_{i=1}^{d} \ell_i(u_i)\, dC(u;\nu,R) = \int_{[0,1]^{d+1}} \prod_{i=1}^{d} \ell_i\left(t_\nu\left(\frac{\sqrt{\nu}}{\sqrt{F_\nu^{-1}(w)}} \sum_{j=1}^{i} a_{ji}\Phi^{-1}(v_j)\right)\right) \prod_{j=1}^{d} dv_j dw$$

With respect to the Gaussian copula, the only remarkable difference is the inclusion of the χ^2, which results in an increase of the dimension of the integral from d to $d+1$.

If one is willing to adopt a symmetric correlation among the elements of the copula, *i.e* if all the off-diagonal elements of R were equal to ρ, then it would be possible to obtain a reduction of the dimensionality of the integral similar to the one attained for the Archimedean copulas. To see this, notice that by the properties of the normal distribution, it is possible to obtain a d-variate normal distribution with covariance function $R = (1-\rho) I_d + \rho \iota_d \iota_d'$, where ι_d is a vector of ones. Each element is the sum of two independent random normals, one specific to each dimension, and one common to all, with weights $\sqrt{1-\rho}$ and $\sqrt{\rho}$. Hence, when the copula is Gaussian, the integral $\mathcal{I}$ can be rewritten as

$$\mathcal{I} = \int_0^1 \prod_{i=1}^d \left[\int_0^1 \ell_i \left(\Phi \left(\sqrt{\rho}\Phi^{-1}(z) + \sqrt{1-\rho}\Phi^{-1}(v_i) \right) \right) dv_i \right] dz$$

For the t copula a similar decomposition is feasible, but the dimensionality of the resulting integral is 3, because of the χ^2 distribution:

$$\mathcal{I} = \int_{[0,1]^2} \prod_{i=1}^d \left[\int_0^1 \ell_i \left(t_\nu \left(\frac{\sqrt{\nu}}{\sqrt{F_\nu^{-1}(w)}} \left(\sqrt{\rho}\Phi^{-1}(z) + \sqrt{1-\rho}\Phi^{-1}(v_i) \right) \right) \right) dv_i \right] dzdw$$

S7 Bernstein Copulas

Bernstein copulas are a family of multivariate copulas introduced by Sancetta and Satchell (2004). The M-th degree Bernstein polynomial is given by $P_{v,M}(u) = \binom{M}{v} u^v (1-u)^{M-v}$ for $0 \leq v \leq M \in \mathbb{N}$ and $u \in [0,1]$. Define the map $C_B : [0,1]^d \to [0,1]$ as

$$C_B(u_1, ..., u_d) = \sum_{v_1=0}^{M_1} ... \sum_{v_d=0}^{M_d} \alpha \left(\frac{v_1}{M_1}, ..., \frac{v_d}{M_d} \right) P_{v_1,M_1}(u_1) ... P_{v_d,M_d}(u_d)$$

where $\alpha\left(\frac{v_1}{M_1}, ..., \frac{v_d}{M_d}\right)$ are some constants $\forall v_j = 0, ..., M_j, j = 1, ..., d$. C_B is a Bernstein copula if it satisfies (Theorem 1 in Sancetta and Satchell (2004))

$$\sum_{l_1=0}^{1} ... \sum_{l_d=0}^{1} (-1)^{l_1+...+l_d} \alpha \left(\frac{v_1+l_1}{M_1}, ..., \frac{v_d+l_d}{M_d} \right) \geq 0$$

$$\min\left(0, \frac{v_1}{M_1} + ... + \frac{v_d}{M_d} - (d-1)\right) \leq \alpha \left(\frac{v_1}{M_1}, ..., \frac{v_d}{M_d} \right) \leq \min \left(\frac{v_1}{M_1}, ..., \frac{v_d}{M_d} \right)$$

Moreover, the α parameters have to satisfy the doubly stochastic matrix condition, *i.e.* $\sum_{v_1}^{M} M\beta\left(\frac{v_1}{M}, ..., \frac{v_d}{M}\right) = 1 \forall j = 1, ..., d$, where $\beta\left(\frac{v_1}{M}, ..., \frac{v_d}{M}\right) \equiv (M+1)^d \Delta_{1,...,m} \alpha\left(\frac{v_1}{M+1}, ..., \frac{v_d}{M+1}\right)$.

This copula has a well-defined density that can be expressed in terms of Bernstein polynomials. Its coefficients are a function of the α coefficients. It is well-known that Bernstein polynomials can uniformly approximate any continuous function on $[0,1]^d$ if the degree of the polynomial is high enough. Consequently, one can use the Bernstein copula to approximate any continuous multivariate copula. Sancetta and Satchell (2004) showed that the Bernstein copula and its approximand converge to an arbitrary limit at a different speed. Hence, although it can capture increasing dependence as one moves to the tails, it is not the appropriate choice to model copulas with extreme tail behavior.

The total number of parameters to estimate a d-variate Bernstein copula with polynomials of degree M equals M^d, and hence is subject to the curse of dimensionality, making it impractical to work with it when the dimension of the data is large. However, implementation when the dimension is small is relatively straightforward, since it just depends on a finite and small number of parameters. This, coupled with the flexibility of the copula makes it an attractive choice when the underlying copula is unknown.

S8 Extra Results

Figure 3: Copula contour of Table 6 $\hat{\rho}$ estimates

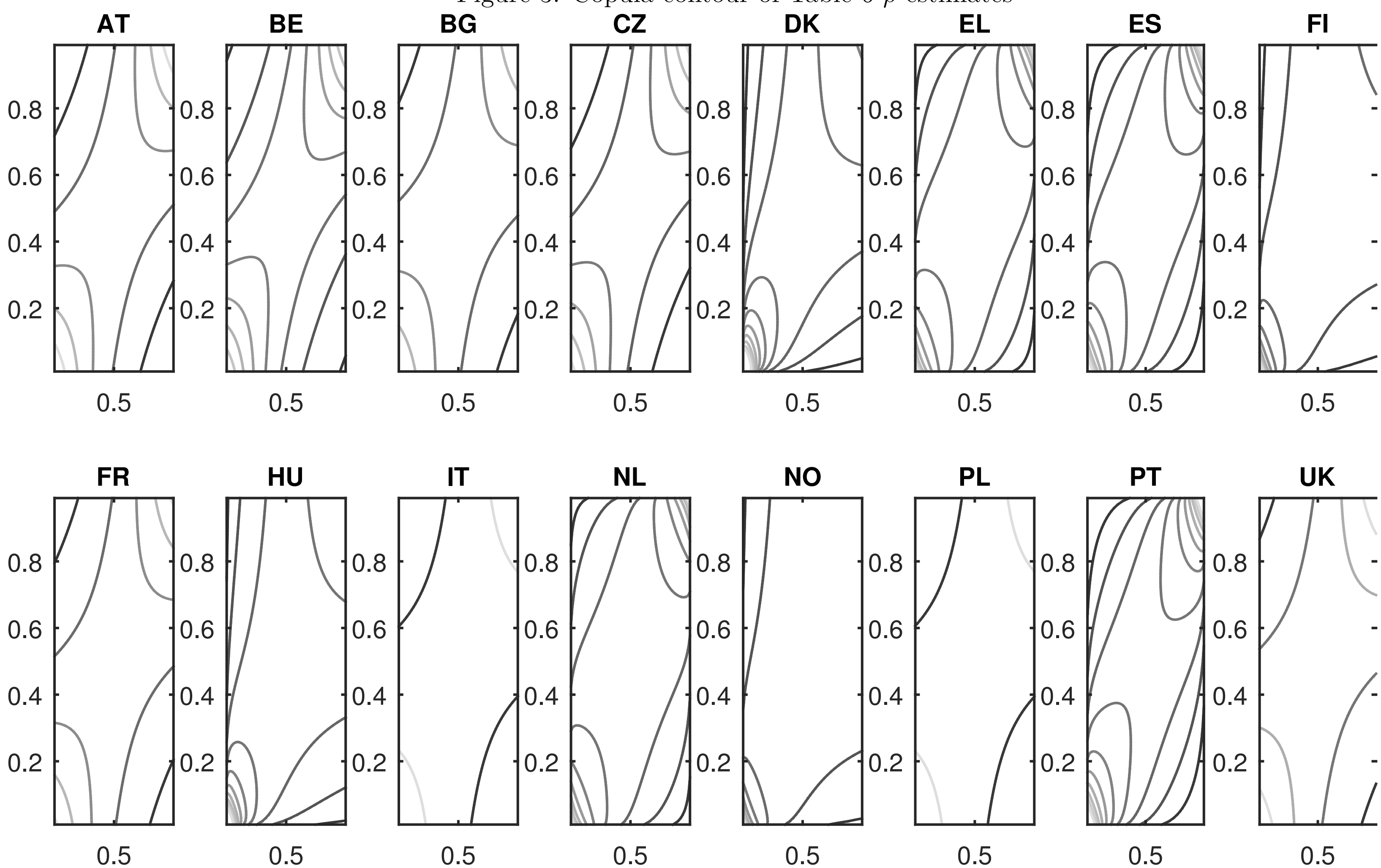


Table 8: Difference in the covariates between women married to employed and unemployed husbands

	AT	BE	BG	CZ	DK	EL	ES	FI	FR	HU	IT	NL	NO	PL	PT	UK
C5	0.12	0.12	0.03	0.16	0.01	0.17	0.08	0.06	0.15	0.13	0.18	0.15	0.08	0.15	0.10	0.13
	(0.03)	(0.03)	(0.02)	(0.02)	(0.04)	(0.03)	(0.02)	(0.03)	(0.01)	(0.02)	(0.02)	(0.03)	(0.03)	(0.02)	(0.02)	(0.04)
AGE	-8.19	-7.69	-5.60	-9.68	-3.50	-8.36	-3.31	-5.65	-9.30	-7.46	-7.42	-8.63	-4.49	-8.97	-6.90	-5.54
	(0.77)	(0.82)	(0.69)	(0.62)	(0.95)	(0.62)	(0.50)	(0.73)	(0.34)	(0.51)	(0.42)	(0.78)	(0.82)	(0.44)	(0.64)	(1.02)
SE	0.07	0.07	-0.06	0.05	-0.07	0.05	0.08	-0.02	0.02	0.08	0.13	0.00	-0.11	-0.07	0.04	-0.12
	(0.04)	(0.04)	(0.04)	(0.03)	(0.06)	(0.04)	(0.02)	(0.04)	(0.02)	(0.03)	(0.02)	(0.05)	(0.04)	(0.02)	(0.03)	(0.05)
TE	0.09	0.20	0.17	0.11	0.09	0.09	0.17	0.20	0.25	0.15	0.12	0.09	0.30	0.14	0.11	0.11
	(0.03)	(0.05)	(0.03)	(0.02)	(0.05)	(0.03)	(0.02)	(0.04)	(0.02)	(0.02)	(0.01)	(0.04)	(0.04)	(0.02)	(0.02)	(0.05)
IN	0.87	-1.11	-0.05	0.18	0.49	0.48	0.15	0.92	1.86	-0.03	0.27	0.84	3.40	0.19	0.18	-0.49
	(0.69)	(1.23)	(0.05)	(0.07)	(2.20)	(0.36)	(0.16)	(0.92)	(1.30)	(0.04)	(0.21)	(0.46)	(1.61)	(0.09)	(0.15)	(0.68)

Notes: Standard errors in parentheses. C5, AGE, SE, TE, and IN respectively denote female, number of children smaller than 5 years old, age, secondary education, tertiary education, and non-labor income (expressed in thousands of euros). The estimates shown are those of Austria, Belgium, Bulgaria, Czech Republic, Denmark, Greece, Spain, Finland, Hungary, Italy, Netherlands, Norway, Poland, Portugal, and the United Kingdom. The estimates of the remaining countries (Cyprus, Estonia, Iceland, Lithuania, Luxemburg, Latvia, Malta, and Slovenia) are available upon request.

Table 9: RE estimates

	AT	BE	BG	CZ	DK	EL	ES	FI	FR	HU	IT	NL	NO	PL	PT	UK
FE	-1.63	-1.51	-0.96	-1.63	-0.98	-4.36	-4.41	-0.82	-2.35	-0.46	-4.03	-5.07	-2.09	-2.33	-3.40	-0.42
	(0.60)	(0.50)	(0.60)	(0.65)	(0.32)	(0.44)	(0.53)	(0.43)	(0.43)	(0.64)	(0.29)	(0.44)	(0.39)	(0.34)	(0.84)	(0.42)
C5	-0.79	-0.20	-3.65	0.45	-0.68	0.98	0.49	-2.18	-0.29	-0.19	0.87	0.23	-1.23	0.70	-0.15	0.80
	(1.00)	(1.59)	(1.87)	(0.71)	(1.10)	(0.94)	(0.47)	(0.91)	(0.48)	(0.68)	(0.58)	(0.43)	(0.82)	(0.62)	(1.10)	(0.49)
C5*FE	-0.78	-0.13	2.50	-4.50	0.79	-1.27	-1.42	-0.39	-1.23	-2.49	-1.89	-1.79	0.52	-1.88	0.19	-2.27
	(1.48)	(2.15)	(1.74)	(2.81)	(1.62)	(1.02)	(0.56)	(2.23)	(0.73)	(1.44)	(0.72)	(1.42)	(1.01)	(1.03)	(1.36)	(0.62)
AGE	-0.29	-0.33	-0.17	-0.27	-0.11	-0.20	-0.15	-0.19	-0.42	-0.21	-0.20	-0.36	-0.10	-0.36	-0.26	-0.07
	(0.05)	(0.03)	(0.05)	(0.03)	(0.03)	(0.03)	(0.04)	(0.06)	(0.03)	(0.04)	(0.02)	(0.03)	(0.03)	(0.02)	(0.07)	(0.04)
SE	1.16	1.51	1.71	2.68	0.50	-0.45	1.69	2.00	1.34	2.11	2.48	2.39	2.08	2.08	1.52	0.03
	(0.61)	(0.46)	(0.55)	(1.24)	(0.67)	(0.51)	(0.47)	(0.96)	(0.36)	(0.40)	(0.32)	(0.55)	(0.60)	(0.43)	(0.60)	(0.29)
TE	2.33	3.14	4.81	4.54	2.36	3.58	3.62	3.04	4.05	4.11	4.69	4.38	3.42	6.02	4.80	0.57
	(1.16)	(0.68)	(0.67)	(1.47)	(0.78)	(0.43)	(0.59)	(0.90)	(0.39)	(1.26)	(0.34)	(0.76)	(0.65)	(0.84)	(3.15)	(0.35)
IN	0.00	-0.01	0.04	0.05	0.00	0.03	-0.01	0.00	0.00	-0.01	0.00	0.03	0.00	-0.01	0.03	0.00
	(0.01)	(0.01)	(0.16)	(0.07)	(0.00)	(0.05)	(0.03)	(0.01)	(0.00)	(0.09)	(0.02)	(0.01)	(0.00)	(0.02)	(0.40)	(0.01)
$\overline{C5}$	-2.09	-0.88	4.21	-1.42	-1.76	-0.28	-1.64	0.50	-3.38	0.45	-0.42	-0.09	0.82	-2.61	-1.71	-1.09
	(1.83)	(1.77)	(2.42)	(1.36)	(1.94)	(1.09)	(0.95)	(2.27)	(0.87)	(0.88)	(1.18)	(1.58)	(1.19)	(1.26)	(1.72)	(2.02)
$\overline{C5 * FE}$	-4.01	-4.35	-7.08	-5.56	-0.25	-0.32	1.53	-5.70	-3.92	-6.51	-1.81	-2.61	-3.00	-4.48	-0.87	0.10
	(2.09)	(2.38)	(2.11)	(2.94)	(3.09)	(1.30)	(0.93)	(2.80)	(1.01)	(2.19)	(1.20)	(2.06)	(1.40)	(1.85)	(3.42)	(2.42)
$\overline{IN}$	0.04	0.01	-0.47	0.05	-0.01	0.00	0.12	0.00	0.01	-0.02	0.12	-0.02	0.00	0.01	-0.02	-0.07
	(0.01)	(0.02)	(0.39)	(0.26)	(0.01)	(0.05)	(0.23)	(0.03)	(0.00)	(0.28)	(0.04)	(0.01)	(0.00)	(0.04)	(0.97)	(0.04)
$\hat{\sigma}$	4.25	6.95	4.98	4.16	3.31	5.50	4.46	3.81	5.56	4.52	4.48	6.00	3.89	5.40	5.10	2.52
	(0.53)	(0.68)	(0.48)	(0.32)	(0.33)	(0.40)	(0.23)	(0.49)	(0.22)	(0.63)	(0.21)	(0.42)	(0.33)	(0.31)	(0.86)	(0.21)
PW	73.5	66.9	76.9	74.7	86.6	55.7	55.8	76.0	71.8	64.7	55.0	76.3	81.8	64.9	65.5	78.2
	(2.4)	(1.6)	(1.5)	(1.3)	(1.8)	(1.6)	(3.7)	(1.6)	(1.2)	(1.7)	(1.2)	(1.2)	(1.2)	(1.4)	(2.3)	(3.4)
$CP1$	75.4	69.6	78.8	75.7	87.0	58.3	57.5	77.2	75.7	66.9	57.7	77.4	82.3	68.7	68.4	78.4
	(2.3)	(1.6)	(1.5)	(1.3)	(1.7)	(1.6)	(3.6)	(1.5)	(1.2)	(1.9)	(1.2)	(1.2)	(1.2)	(1.4)	(2.3)	(3.4)
$CP0$	63.4	56.6	70.0	67.5	83.4	45.8	48.1	70.0	56.2	58.6	43.3	62.8	75.6	52.5	54.8	76.6
	(4.7)	(1.7)	(2.2)	(2.1)	(2.1)	(1.9)	(4.2)	(2.9)	(1.7)	(2.0)	(1.4)	(1.6)	(2.0)	(1.6)	(2.4)	(3.4)
N	764	604	780	1404	696	930	1906	882	3370	1422	2186	1360	1364	1848	814	640

Notes: Standard errors in parentheses. FE, C5, AGE, SE, TE, and IN respectively denote female, number of children smaller than 5 years old, age, secondary education, tertiary education, and non-labor income (expressed in thousands of euros); the symbol is used to denote the coefficient of the correlated random effect; PW denotes the unconditional probability that a wife is employed, $CP1$ and $CP0$ respectively denote the probability that a wife is employed conditional on her husband being employed, and conditional on her husband being unemployed; the selected model (logit/probit) for each country is the same as in 6; N is the sample size of each country. The estimates shown are those of Austria, Belgium, Bulgaria, Czech Republic, Denmark, Greece, Spain, Finland, Hungary, Italy, Netherlands, Norway, Poland, Portugal, and the United Kingdom. The estimates of the remaining countries (Cyprus, Estonia, Iceland, Lithuania, Luxemburg, Latvia, Malta, and Slovenia) are available upon request.

Table 10: Bernstein copula CBRE estimates

	AT	BE	BG	CZ	DK	EL	ES	FI	FR	HU	IT	NL	NO	PL	PT	UK
FE	-1.97	-3.42	-0.93	-1.56	-1.20	-4.96	-4.36	-0.76	-1.74	-0.50	-4.14	-5.33	-2.09	-2.57	-2.95	-1.10
	(0.55)	(0.90)	(0.48)	(0.37)	(0.44)	(0.61)	(0.33)	(0.43)	(0.33)	(0.33)	(0.31)	(0.69)	(0.37)	(0.40)	(0.54)	(0.52)
C5	-0.79	0.11	-3.65	0.36	-0.68	1.00	0.49	-2.13	-0.35	-0.20	0.89	0.49	-1.22	0.71	-0.10	1.52
	(0.92)	(0.79)	(2.94)	(1.77)	(1.16)	(0.67)	(0.38)	(0.67)	(0.55)	(0.51)	(0.46)	(3.81)	(0.64)	(0.51)	(1.32)	(2.21)
C5*FE	-0.84	-0.39	2.57	-4.54	0.83	-1.25	-1.42	-0.47	-1.24	-2.58	-1.90	-1.93	0.51	-1.84	0.05	-4.21
	(1.17)	(1.35)	(4.11)	(1.88)	(1.70)	(0.91)	(0.54)	(0.89)	(0.67)	(0.54)	(0.59)	(3.88)	(0.76)	(0.61)	(1.55)	(2.40)
AGE	-0.30	-0.47	-0.20	-0.28	-0.14	-0.25	-0.15	-0.21	-0.43	-0.23	-0.21	-0.41	-0.12	-0.36	-0.29	-0.12
	(0.04)	(0.07)	(0.03)	(0.02)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.05)	(0.02)	(0.03)	(0.03)	(0.04)
SE	0.95	1.38	1.73	2.67	0.44	-0.16	2.02	1.88	1.18	1.85	2.54	2.18	2.00	2.25	1.71	-0.04
	(0.41)	(0.45)	(0.61)	(0.63)	(0.43)	(0.31)	(0.29)	(0.53)	(0.26)	(0.28)	(0.29)	(0.65)	(0.40)	(0.69)	(0.49)	(0.38)
TE	1.87	2.52	4.46	4.77	2.25	3.85	3.87	2.73	4.00	3.85	4.79	4.33	3.33	5.86	4.46	0.83
	(0.62)	(0.48)	(0.80)	(0.81)	(0.50)	(0.55)	(0.31)	(0.60)	(0.39)	(0.42)	(0.46)	(0.64)	(0.43)	(0.80)	(0.83)	(0.43)
IN	0.00	-0.01	0.04	0.04	0.00	0.03	-0.01	0.00	0.00	-0.01	0.00	0.03	0.00	-0.01	0.03	0.00
	(0.01)	(0.01)	(0.65)	(0.05)	(0.01)	(0.05)	(0.03)	(0.02)	(0.00)	(0.10)	(0.02)	(0.02)	(0.01)	(0.18)	(0.11)	(0.04)
$\overline{C5}$	-1.58	-3.94	3.60	-2.49	-2.13	-1.27	-1.53	-0.07	-2.67	1.29	-0.50	-3.46	0.06	-2.29	-0.98	-1.41
	(1.52)	(2.36)	(3.46)	(1.85)	(1.61)	(1.31)	(0.74)	(1.17)	(0.83)	(1.16)	(0.79)	(4.34)	(1.13)	(1.07)	(1.96)	(2.40)
$\overline{C5 * FE}$	-4.50	-3.26	-8.63	-4.98	-0.06	-1.43	0.99	-5.90	-4.64	-8.07	-1.88	-1.25	-2.73	-4.96	-2.46	0.09
	(1.74)	(2.48)	(4.82)	(2.13)	(2.04)	(1.46)	(0.88)	(1.51)	(1.07)	(1.41)	(0.93)	(4.37)	(1.25)	(1.20)	(2.18)	(2.63)
$\overline{IN}$	0.05	0.00	-0.48	0.01	0.00	-0.03	0.10	0.00	0.01	0.07	0.12	-0.03	0.00	-0.05	0.11	-0.11
	(0.03)	(0.06)	(1.41)	(0.13)	(0.01)	(0.08)	(0.07)	(0.05)	(0.01)	(0.58)	(0.04)	(0.03)	(0.01)	(0.12)	(0.22)	(0.09)
$\hat{\sigma}$	4.82	7.40	5.22	4.92	3.57	5.76	4.78	4.19	6.69	4.98	4.82	6.81	4.15	6.24	5.55	4.57
	(0.36)	(0.78)	(0.38)	(0.29)	(0.29)	(0.41)	(0.21)	(0.31)	(0.26)	(0.25)	(0.20)	(0.50)	(0.25)	(0.32)	(0.39)	(0.40)
Model	Log	Log	Log	Log	Log	Log	Log	Log	Log	Log	Log	Log	Log	Log	Log	Log
Order	2	2	2	2	2	3	2	2	2	2	2	2	2	2	2	2
τ	0.21	0.22	0.16	0.21	0.22	0.16	0.11	0.16	0.17	0.20	0.10	0.08	0.16	0.11	0.14	0.17
N	764	604	780	1404	696	930	1906	882	3370	1422	2186	1360	1364	1848	814	640

Notes: Standard errors in parentheses. FE, C5, AGE, SE, TE, and IN respectively denote female, number of children smaller than 5 years old, age, secondary education, tertiary education, and non-labor income (expressed in thousands of euros); the symbol is used to denote the coefficient of the correlated random effect; the contour plots of the copula estimates are shown in Figure 4; Model denotes the best fitting binary choice model: logit (Log) or probit (Pro); N is the sample size of each country. The estimates shown are those of Austria, Belgium, Bulgaria, Czech Republic, Denmark, Greece, Spain, Finland, Hungary, Italy, Netherlands, Norway, Poland, Portugal, and the United Kingdom. The estimates of the remaining countries (Cyprus, Estonia, Iceland, Lithuania, Luxemburg, Latvia, Malta, and Slovenia) are available upon request.

Figure 4: Contour of estimated Bernstein copulas

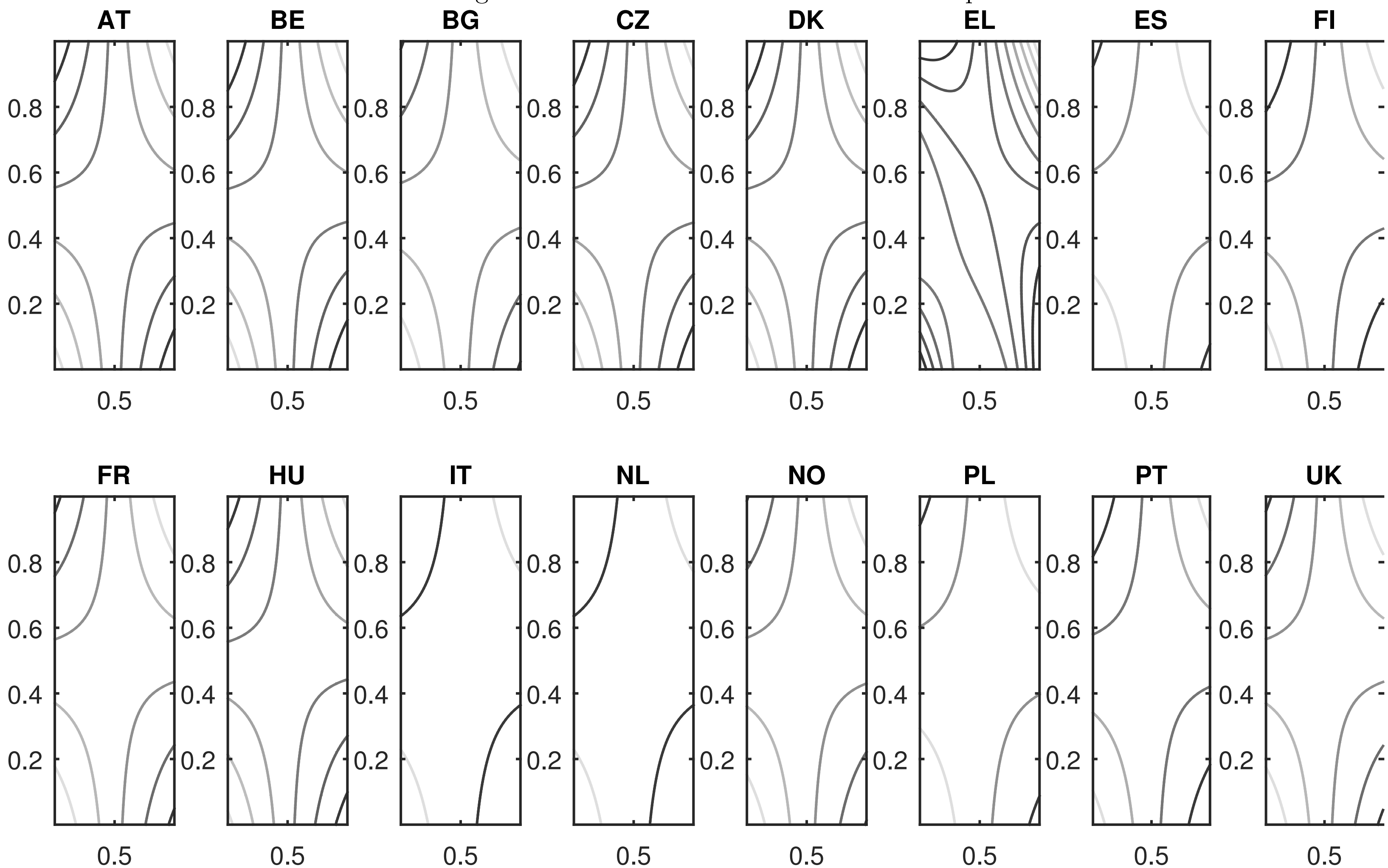


Table 11: Counterfactuals with Bernstein copulas

	AT	BE	BG	CZ	DK	EL	ES	FI	FR	HU	IT	NL	NO	PL	PT	UK
PW_ρ	72.6	67.0	76.9	74.3	86.4	54.2	55.3	75.9	72.8	65.1	54.6	77.2	82.7	63.7	65.0	77.8
	(1.9)	(2.3)	(1.8)	(1.3)	(1.4)	(1.9)	(1.3)	(1.6)	(0.9)	(1.4)	(1.2)	(1.3)	(1.2)	(1.3)	(1.9)	(2.0)
$CP1_\rho$	76.3	73.7	80.8	76.7	87.8	57.9	58.6	78.3	78.3	70.4	58.8	78.7	83.9	68.9	70.3	79.4
	(1.9)	(2.4)	(1.7)	(1.3)	(1.4)	(2.1)	(1.4)	(1.7)	(0.9)	(1.5)	(1.2)	(1.3)	(1.2)	(1.3)	(2.0)	(1.9)
$CP0_\rho$	51.4	41.9	61.5	55.7	74.2	37.4	41.2	63.1	49.5	49.7	37.1	58.9	68.1	46.6	46.1	66.7
	(4.3)	(4.8)	(3.9)	(3.5)	(3.9)	(3.4)	(2.5)	(3.5)	(2.0)	(2.5)	(2.2)	(4.9)	(3.8)	(2.5)	(3.8)	(4.9)
$\Delta p\left(y_{2ct}\|x_g;\hat{\theta}_{CBRE}\right)$	24.9	31.8	19.3	21.0	13.5	20.5	17.4	15.2	28.8	20.7	21.7	19.8	15.7	22.3	24.2	12.7
	(4.3)	(5.0)	(3.9)	(3.5)	(3.9)	(3.8)	(2.6)	(3.5)	(2.1)	(2.7)	(2.3)	(4.9)	(3.7)	(2.7)	(4.0)	(4.8)
Endowment effect	12.1	13.1	8.8	8.2	3.6	12.5	9.4	7.3	19.5	8.3	14.5	14.6	6.7	16.1	13.6	1.8
	(3.8)	(1.1)	(1.9)	(1.9)	(0.8)	(1.2)	(0.9)	(2.7)	(1.1)	(1.6)	(1.0)	(1.0)	(1.4)	(1.0)	(1.0)	(0.7)
Homophily effect	12.8	18.8	10.6	12.8	10.0	7.9	8.0	8.0	9.3	12.5	7.3	5.2	9.0	6.2	10.6	10.9
	(5.7)	(5.1)	(4.3)	(4.0)	(4.0)	(3.9)	(2.8)	(4.4)	(2.4)	(3.2)	(2.5)	(5.0)	(4.0)	(2.9)	(4.1)	(4.9)
P_ρ	77.6	65.2	75.2	82.6	90.1	64.5	66.2	80.8	72.5	62.2	65.0	89.9	91.9	64.8	65.3	84.7
	(2.6)	(3.4)	(2.9)	(1.8)	(1.7)	(3.3)	(1.9)	(2.4)	(1.4)	(2.3)	(1.8)	(1.6)	(1.2)	(1.9)	(2.9)	(2.9)
P_I	83.0	72.3	80.2	86.9	93.3	66.0	69.5	84.5	77.5	68.8	68.0	91.2	94.1	68.3	70.0	88.7
	(1.9)	(2.7)	(2.0)	(1.2)	(1.0)	(2.4)	(1.5)	(1.6)	(1.0)	(1.8)	(1.5)	(1.0)	(0.7)	(1.5)	(2.2)	(1.8)
Δ_P	-5.4	-7.1	-5.0	-4.3	-3.2	-1.5	-3.3	-3.8	-5.0	-6.6	-2.9	-1.4	-2.2	-3.5	-4.7	-4.1
	(1.6)	(2.1)	(1.8)	(1.1)	(1.3)	(2.1)	(1.0)	(1.5)	(0.8)	(1.3)	(0.9)	(1.1)	(0.9)	(1.1)	(1.7)	(1.8)
N	764	604	780	1404	696	930	1906	882	3370	1422	2186	1360	1364	1848	814	640

Notes: Standard errors of the estimated probabilities were computed using the delta method. PW_ρ denotes the unconditional probability that a wife is employed, $CP1_\rho$ and $CP0_\rho$ respectively denote the probability that a wife is employed conditional on her husband being employed, and conditional on her husband being unemployed; $\Delta p\left(y_{2gt}|x_g;\hat{\theta}_{CBRE}\right)$ and the homophily and endowment effects are defined in the text; P_ρ and P_I respectively denote the probability (in %) that at least one member of the couple was employed in every period when the parameter of the copula is the estimated one and when the copula is independent, and Δ_P denotes the difference between the two; N is the sample size of each country. The estimates shown are those of Austria, Belgium, Bulgaria, Czech Republic, Denmark, Greece, Spain, Finland, Hungary, Italy, Netherlands, Norway, Poland, Portugal, and the United Kingdom. The estimates of the remaining countries (Cyprus, Estonia, Iceland, Lithuania, Luxemburg, Latvia, Malta, and Slovenia) are available upon request.

Table 12: Akaike Information Criterion across specifications

			AT	BE	BG	CZ	DK	EL	ES	FI	FR	HU	IT	NL	NO	PL	PT	UK
RE	-	Lo	1932.8	1385.0	2014.7	1278.7	3068.2	1506.7	1381.9	2629.4	5461.7	2283.3	7507.3	4096.1	707.4	6303.5	2104.4	5632.3
RE	-	Pr	1933.7	1388.8	2017.5	1277.3	3073.1	1510.9	1383.3	2629.2	5463.4	2286.1	7522.0	4098.7	708.8	6306.6	2102.2	5637.4
CRE	-	Lo	1909.5	1372.3	2012.4	1281.8	3023.1	1509.5	1378.1	2634.8	5457.4	2264.8	7418.1	4052.0	711.4	6290.9	2107.2	5611.1
CRE	-	Pr	1910.5	1370.9	2015.7	1280.0	3020.1	1513.6	1377.4	2634.1	5461.0	2264.9	7444.2	4049.8	712.5	6293.0	2104.4	5615.6
CBRE	Cl	Lo	1915.4	1365.5	2016.2	1280.2	3057.5	1483.9	1385.6	2630.8	5443.0	2278.2	7440.4	4065.8	710.6	6300.2	2109.5	5611.8
CBRE	Fr	Lo	1915.3	1359.7	2012.6	1281.0	3054.6	1487.6	1385.6	2629.0	5442.9	2279.8	7438.5	4066.9	709.9	6294.9	2108.8	5608.1
CBRE	Ga	Lo	1915.8	1361.6	2014.5	1280.6	3056.7	1485.9	1385.6	2628.4	5441.0	2280.4	7437.2	4066.9	710.1	6296.5	2109.0	5609.1
CBRE	Cl	Pr	1918.5	1365.9	2018.9	1280.8	3064.2	1488.2	1386.2	2631.2	5446.9	2281.4	7456.2	4069.0	710.8	6304.2	2109.4	5616.4
CBRE	Fr	Pr	1918.7	1360.1	2015.4	1281.3	3061.5	1492.8	1386.2	2629.2	5446.7	2283.2	7454.2	4070.3	710.0	6301.7	2108.6	5612.8
CBRE	Ga	Pr	1919.2	1362.0	2017.2	1281.1	3063.6	1490.4	1386.2	2628.7	5444.9	2283.7	7452.7	4070.3	710.3	6300.5	2108.9	5613.7
CBCRE	Cl	Lo	1896.1	1357.9	2011.1	1280.9	3014.8	1487.8	1378.0	2630.3	5444.7	2257.7	7368.4	4017.4	713.8	6288.5	2112.3	5595.2
CBCRE	Fr	Lo	1895.4	1352.8	2008.1	1281.4	3012.7	1491.7	1378.1	2628.6	5444.6	2258.4	7360.1	4017.7	712.9	6283.7	2111.7	5591.6
CBCRE	Ga	Lo	1895.7	1354.7	2009.6	1281.0	3013.7	1489.6	1378.0	2628.2	5442.8	2258.9	7363.5	4017.5	713.4	6285.1	2111.8	5592.9
CBCRE	Cl	Pr	1898.4	1358.3	2013.4	1281.5	3014.8	1492.0	1377.9	2630.8	5448.6	2259.0	7389.0	4018.6	714.1	6292.3	2112.3	5599.6
CBCRE	Fr	Pr	1898.0	1353.2	2010.2	1282.0	3013.0	1496.0	1377.9	2629.0	5448.6	2259.7	7380.2	4018.8	713.1	6287.6	2111.6	5596.5
CBCRE	Ga	Pr	1898.1	1355.3	2011.9	1281.6	3013.7	1493.9	1377.9	2628.6	5446.7	2260.2	7383.5	4018.6	713.6	6288.8	2111.7	5597.3
CBRE	B2	Lo	1916.6	1361.6	2013.2	1280.8	3064.0	1492.9	1385.7	2627.7	5443.2	2282.4	7440.2	4069.7	709.4	6293.5	2108.9	5606.2
CBRE	B3	Lo	1920.9	1365.8	2017.5	1285.9	3068.4	1495.0	1390.6	2623.8	5445.9	2283.9	7443.1	4072.2	715.1	6298.2	2113.1	5608.9
CBRE	B4	Lo	1930.7	1375.4	2026.4	1295.0	3077.5	1503.3	1399.5	2632.6	5451.4	2291.3	7450.6	4080.8	724.6	6303.8	2122.4	5614.5
CBRE	B2	Pr	1920.7	1362.2	2016.5	1281.3	3071.4	1498.1	1386.7	2628.5	5449.0	2287.0	7456.8	4074.1	709.8	6300.0	2109.4	5611.7
CBRE	B3	Pr	1926.1	1366.3	2021.5	1286.4	3075.5	1500.5	1391.8	2624.7	5452.2	2288.5	7458.1	4076.5	715.7	6304.2	2113.9	5614.2
CBRE	B4	Pr	1934.8	1375.9	2029.6	1295.8	3084.5	1509.1	1401.1	2633.7	5457.5	2295.6	7465.1	4083.3	725.2	6309.7	2123.1	5619.6
CBCRE	B2	Lo	1896.7	1354.8	2008.6	1281.4	3024.4	1496.3	1378.1	2627.7	5445.2	2261.4	7364.2	4021.1	712.5	6282.7	2111.9	5590.9
CBCRE	B3	Lo	1901.7	1359.6	2012.9	1286.7	3029.8	1498.5	1382.8	2624.3	5447.9	2264.4	7366.1	4023.7	718.2	6287.1	2115.9	5593.1
CBCRE	B4	Lo	1911.1	1369.0	2021.8	1296.0	3038.7	1507.1	1391.6	2633.0	5453.6	2272.5	7373.9	4031.9	727.6	6292.7	2125.0	5600.7
CBCRE	B2	Pr	1900.1	1355.4	2011.6	1282.2	3025.4	1501.9	1378.4	2628.8	5451.2	2264.1	7385.8	4023.6	713.2	6289.5	2112.4	5596.6
CBCRE	B3	Pr	1905.6	1360.3	2016.7	1287.5	3030.9	1504.4	1383.1	2625.4	5454.5	2266.8	7388.0	4026.0	719.0	6293.3	2116.8	5598.8
CBCRE	B4	Pr	1914.3	1369.7	2025.1	1297.0	3039.7	1513.1	1391.9	2634.4	5460.0	2275.1	7394.7	4034.0	728.5	6298.2	2125.8	5606.5
N			764	604	780	568	1404	696	498	930	1906	882	3370	1422	386	2186	774	2294

Notes: CBCRE and CRE respectively denote the CBRE and RE estimators with correlated random effects; Lo and Pr stand for logit and probit; Cl, Fr, Ga, and Bx stand for independent, Clayton, Frank, Gaussian, and Bernstein copula of order x; N is the sample size of each country. The estimates shown are those of Austria, Belgium, Bulgaria, Czech Republic, Denmark, Greece, Spain, Finland, Hungary, Italy, Netherlands, Norway, Poland, Portugal, and the United Kingdom. The estimates of the remaining countries (Cyprus, Estonia, Iceland, Lithuania, Luxemburg, Latvia, Malta, and Slovenia) are available upon request.

Table 13: Bayesian Information Criterion across specifications

			AT	BE	BG	CZ	DK	EL	ES	FI	FR	HU	IT	NL	NO	PL	PT	UK
RE	-	Lo	2005.1	1454.5	2087.2	1347.5	3147.8	1577.8	1449.1	2704.1	5544.9	2357.3	7597.4	4175.8	771.5	6388.4	2176.9	5717.8
RE	-	Pr	2006.0	1458.3	2090.1	1346.1	3152.7	1582.1	1450.5	2703.9	5546.6	2360.2	7612.1	4178.4	773.0	6391.5	2174.7	5722.9
CRE	-	Lo	1999.9	1459.2	2103.1	1367.7	3122.6	1598.4	1462.0	2728.1	5561.5	2357.3	7530.7	4151.7	791.5	6397.1	2197.7	5718.0
CRE	-	Pr	2000.9	1457.8	2106.3	1365.9	3119.6	1602.6	1461.3	2727.4	5565.1	2357.4	7556.8	4149.5	792.6	6399.2	2195.0	5722.4
CBRE	Cl	Lo	1993.8	1440.8	2094.8	1354.6	3143.7	1561.0	1458.4	2711.7	5533.2	2358.4	7538.0	4152.2	780.0	6392.1	2188.0	5704.4
CBRE	Fr	Lo	1993.7	1435.0	2091.2	1355.5	3140.8	1564.7	1458.3	2709.8	5533.2	2360.0	7536.1	4153.3	779.3	6386.8	2187.3	5700.7
CBRE	Ga	Lo	1994.1	1436.9	2093.1	1355.1	3142.9	1563.0	1458.4	2709.3	5531.2	2360.6	7534.8	4153.3	779.6	6388.5	2187.4	5701.7
CBRE	Cl	Pr	1996.9	1441.2	2097.5	1355.3	3150.5	1565.3	1459.0	2712.1	5537.1	2361.5	7553.9	4155.4	780.3	6396.2	2187.9	5709.0
CBRE	Fr	Pr	1997.0	1435.3	2094.0	1355.8	3147.7	1569.9	1458.9	2710.1	5536.9	2363.4	7551.9	4156.7	779.4	6393.7	2187.1	5705.4
CBRE	Ga	Pr	1997.5	1437.2	2095.7	1355.6	3149.8	1567.5	1459.0	2709.6	5535.1	2363.9	7550.3	4156.7	779.8	6392.5	2187.4	5706.3
CBCRE	Cl	Lo	1992.5	1450.5	2107.8	1372.6	3120.9	1582.8	1467.6	2729.8	5555.8	2356.4	7488.5	4123.7	799.2	6401.7	2208.9	5709.2
CBCRE	Fr	Lo	1991.8	1445.5	2104.8	1373.0	3118.8	1586.6	1467.6	2728.1	5555.6	2357.1	7480.3	4124.0	798.4	6396.9	2208.3	5705.6
CBCRE	Ga	Lo	1992.1	1447.3	2106.3	1372.7	3119.8	1584.5	1467.5	2727.7	5553.8	2357.6	7483.7	4123.8	798.8	6398.3	2208.4	5706.9
CBCRE	Cl	Pr	1994.8	1450.9	2110.2	1373.2	3120.9	1586.9	1467.4	2730.3	5559.7	2357.7	7509.2	4124.9	799.6	6405.6	2208.9	5713.6
CBCRE	Fr	Pr	1994.4	1445.9	2106.9	1373.6	3119.1	1590.9	1467.4	2728.5	5559.6	2358.4	7500.3	4125.1	798.6	6400.8	2208.2	5710.5
CBCRE	Ga	Pr	1994.5	1448.0	2108.6	1373.3	3119.9	1588.9	1467.4	2728.2	5557.7	2358.9	7503.6	4125.0	799.1	6402.0	2208.4	5711.3
CBRE	B2	Lo	1994.9	1436.9	2091.8	1355.3	3150.3	1570.0	1458.5	2708.6	5533.4	2362.6	7537.8	4156.1	778.8	6385.5	2187.4	5698.8
CBRE	B3	Lo	2017.3	1458.5	2114.2	1377.5	3174.5	1589.9	1480.1	2723.3	5556.9	2382.6	7563.2	4178.6	800.6	6411.4	2209.7	5722.9
CBRE	B4	Lo	2057.2	1497.0	2153.3	1415.3	3216.8	1627.9	1517.1	2763.3	5597.2	2420.8	7608.3	4220.4	836.8	6452.4	2249.2	5764.2
CBRE	B2	Pr	1999.0	1437.4	2095.1	1355.8	3157.6	1575.2	1459.4	2709.4	5539.2	2367.2	7554.5	4160.5	779.3	6391.9	2187.8	5704.3
CBRE	B3	Pr	2022.5	1459.0	2118.2	1378.0	3181.6	1595.4	1481.4	2724.3	5563.2	2387.2	7578.3	4182.8	801.2	6417.5	2210.5	5728.2
CBRE	B4	Pr	2061.3	1497.5	2156.6	1416.1	3223.8	1633.7	1518.6	2764.3	5603.2	2425.1	7622.7	4222.9	837.4	6458.3	2249.9	5769.2
CBCRE	B2	Lo	1993.1	1447.4	2105.3	1373.0	3130.5	1591.2	1467.6	2727.3	5556.3	2360.1	7484.3	4127.4	797.9	6395.9	2208.5	5704.8
CBCRE	B3	Lo	2016.2	1469.7	2127.7	1395.5	3155.8	1611.2	1489.2	2742.5	5579.7	2381.6	7508.8	4149.9	819.7	6421.5	2230.6	5728.5
CBCRE	B4	Lo	2055.7	1508.0	2166.9	1433.5	3197.9	1649.5	1525.9	2782.3	5620.1	2420.6	7554.1	4191.5	855.8	6462.5	2269.9	5771.7
CBCRE	B2	Pr	1996.5	1448.1	2108.3	1373.9	3131.5	1596.8	1467.9	2728.4	5562.3	2362.8	7506.0	4129.9	798.6	6402.7	2209.0	5710.6
CBCRE	B3	Pr	2020.1	1470.3	2131.5	1396.3	3156.9	1617.1	1489.4	2743.6	5586.3	2384.0	7530.7	4152.3	820.5	6427.7	2231.5	5734.1
CBCRE	B4	Pr	2058.9	1508.7	2170.1	1434.5	3198.9	1655.5	1526.3	2783.7	5626.6	2423.1	7574.9	4193.5	856.7	6468.1	2270.7	5777.5
N			764	604	780	568	1404	696	498	930	1906	882	3370	1422	386	2186	774	2294

Notes: CBCRE and CRE respectively denote the CBRE and RE estimators with correlated random effects; Lo and Pr stand for logit and probit; Cl, Fr, Ga, and Bx stand for independent, Clayton, Frank, Gaussian, and Bernstein copula of order x; N is the sample size of each country. The estimates shown are those of Austria, Belgium, Bulgaria, Czech Republic, Denmark, Greece, Spain, Finland, Hungary, Italy, Netherlands, Norway, Poland, Portugal, and the United Kingdom. The estimates of the remaining countries (Cyprus, Estonia, Iceland, Lithuania, Luxemburg, Latvia, Malta, and Slovenia) are available upon request.

Table 14: 10-fold cross validated log-likelihood value across specifications

			AT	BE	BG	CZ	DK	EL	ES	FI	FR	HU	IT	NL	NO	PL	PT	UK
RE	-	Lo	-974.3	-719.5	-1011.7	-644.0	-1542.8	-756.5	-691.3	-1316.7	-2732.6	-1147.6	-3762.6	-2059.0	-354.2	-3153.0	-1054.2	-2827.6
RE	-	Pr	-973.1	-743.5	-1014.0	-642.9	-1543.5	-758.1	-692.2	-1316.7	-2732.8	-1149.3	-3770.5	-2057.5	-357.3	-3155.2	-1049.6	-2826.6
CRE	-	Lo	-986.1	-699.7	-1008.7	-654.4	-1522.0	-765.3	-696.2	-1329.3	-2730.9	-1138.4	-3718.4	-2031.1	-358.9	-3150.8	-1067.1	-2817.4
CRE	-	Pr	-988.7	-699.6	-1010.4	-651.4	-1515.2	-766.2	-694.6	-1326.3	-2732.0	-1137.7	-3738.5	-2028.1	-360.9	-3150.1	-1080.1	-2820.4
CBRE	Cl	Lo	-960.2	-701.7	-1008.4	-641.2	-1533.0	-744.2	-691.9	-1317.0	-2722.9	-1144.0	-3725.4	-2037.5	-355.8	-3151.5	-1054.4	-2805.9
CBRE	Fr	Lo	-959.6	-696.9	-1005.8	-642.1	-1527.8	-739.8	-691.3	-1313.6	-2723.5	-1145.2	-3724.1	-2036.6	-356.4	-3147.8	-1053.3	-2802.7
CBRE	Ga	Lo	-960.8	-698.4	-1007.4	-641.7	-1532.7	-745.0	-691.5	-1315.6	-2722.1	-1146.2	-3723.6	-2037.7	-356.3	-3149.7	-1054.1	-2804.5
CBRE	Cl	Pr	-962.0	-703.5	-1009.8	-641.8	-1537.0	-746.1	-692.9	-1317.3	-2724.9	-1145.7	-3734.9	-2039.5	-357.8	-3153.9	-1054.5	-2808.2
CBRE	Fr	Pr	-962.2	-696.9	-1007.4	-642.1	-1533.1	-736.3	-692.2	-1309.8	-2720.5	-1145.8	-3733.3	-2039.2	-356.5	-3147.9	-1053.6	-2804.2
CBRE	Ga	Pr	-962.7	-701.3	-1009.0	-642.1	-1536.8	-747.1	-692.5	-1315.9	-2724.2	-1148.0	-3733.3	-2040.0	-357.4	-3152.0	-1053.9	-2806.6
CBCRE	Cl	Lo	-970.8	-697.9	-1004.8	-645.2	-1511.3	-751.6	-690.1	-1316.8	-2723.6	-1133.8	-3689.3	-2012.7	-358.7	-3146.6	-1065.1	-2803.0
CBCRE	Fr	Lo	-970.6	-694.5	-1003.3	-645.8	-1499.4	-750.8	-689.6	-1315.0	-2724.1	-1133.0	-3683.9	-2012.2	-359.4	-3144.2	-1063.3	-2799.4
CBCRE	Ga	Lo	-969.8	-695.7	-1004.2	-645.0	-1511.2	-751.8	-690.0	-1315.4	-2723.0	-1135.0	-3686.3	-2012.6	-358.9	-3144.9	-1064.5	-2800.3
CBCRE	Cl	Pr	-979.5	-700.1	-1006.5	-645.7	-1511.4	-753.7	-690.4	-1316.3	-2725.8	-1134.5	-3704.2	-2013.8	-359.9	-3148.7	-1072.3	-2805.1
CBCRE	Fr	Pr	-979.3	-697.5	-1004.0	-646.6	-1499.7	-743.3	-689.6	-1315.3	-2725.7	-1133.6	-3699.7	-2012.9	-359.6	-3145.5	-1073.4	-2801.7
CBCRE	Ga	Pr	-979.6	-698.2	-1005.6	-645.9	-1511.0	-753.9	-690.2	-1316.3	-2725.1	-1135.7	-3700.8	-2013.2	-360.5	-3146.9	-1072.3	-2802.5
CBRE	B2	Lo	-960.6	-697.5	-1006.4	-641.9	-1539.8	-746.4	-691.3	-1315.3	-2723.2	-1147.0	-3725.4	-2038.5	-355.7	-3148.1	-1053.9	-2802.8
CBRE	B3	Lo	-960.2	-697.3	-1005.0	-642.5	-1539.3	-745.2	-692.6	-1310.4	-2721.7	-1143.1	-3722.5	-2036.4	-355.9	-3149.4	-1052.8	-2802.1
CBRE	B4	Lo	-960.2	-698.7	-1004.9	-642.6	-1538.7	-745.3	-691.2	-1311.0	-2718.3	-1142.2	-3722.6	-2036.0	-355.9	-3145.4	-1053.5	-2799.4
CBRE	B2	Pr	-963.1	-699.6	-1008.2	-641.8	-1543.5	-748.0	-692.7	-1315.5	-2725.6	-1148.6	-3733.9	-2040.3	-355.6	-3150.5	-1053.8	-2804.8
CBRE	B3	Pr	-964.2	-699.1	-1007.0	-642.8	-1542.8	-746.9	-694.3	-1310.7	-2723.9	-1145.5	-3731.5	-2038.0	-356.3	-3151.8	-1053.3	-2803.9
CBRE	B4	Pr	-962.3	-700.9	-1005.9	-642.8	-1542.3	-746.9	-693.3	-1311.7	-2720.0	-1143.5	-3728.7	-2036.4	-356.3	-3147.7	-1053.6	-2800.8
CBCRE	B2	Lo	-972.6	-695.6	-1003.9	-645.1	-1523.4	-753.4	-690.1	-1315.8	-2724.2	-1136.4	-3685.5	-2014.4	-358.1	-3143.9	-1064.5	-2799.6
CBCRE	B3	Lo	-972.2	-695.5	-1002.2	-646.3	-1523.5	-752.6	-691.1	-1310.9	-2722.9	-1134.8	-3682.6	-2012.2	-358.7	-3145.3	-1062.3	-2798.7
CBCRE	B4	Lo	-972.9	-696.6	-1002.2	-646.4	-1522.0	-753.1	-689.4	-1311.2	-2720.0	-1135.2	-3683.3	-2011.1	-358.6	-3142.0	-1063.6	-2797.8
CBCRE	B2	Pr	-1636.7	-698.0	-1004.3	-645.4	-1521.8	-756.3	-690.4	-1316.6	-2727.7	-1136.5	-3698.5	-2015.5	-358.8	-3145.2	-1074.5	-2800.7
CBCRE	B3	Pr	-1636.7	-697.8	-1004.2	-646.9	-1522.4	-755.7	-690.5	-1311.4	-2726.3	-1136.9	-3696.4	-2013.3	-359.1	-3146.4	-1079.1	-2800.0
CBCRE	B4	Pr	-1634.7	-698.5	-1002.9	-646.9	-1520.8	-756.8	-689.7	-1312.6	-2722.9	-1135.3	-3694.7	-2012.4	-359.5	-3142.8	-1079.1	-2800.1
N			764	604	780	568	1404	696	498	930	1906	882	3370	1422	386	2186	774	2294

Notes: CBCRE and CRE respectively denote the CBRE and RE estimators with correlated random effects; Lo and Pr stand for logit and probit; Cl, Fr, Ga, and Bx stand for independent, Clayton, Frank, Gaussian, and Bernstein copula of order x; N is the sample size of each country. The estimates shown are those of Austria, Belgium, Bulgaria, Czech Republic, Denmark, Greece, Spain, Finland, Hungary, Italy, Netherlands, Norway, Poland, Portugal, and the United Kingdom. The estimates of the remaining countries (Cyprus, Estonia, Iceland, Lithuania, Luxemburg, Latvia, Malta, and Slovenia) are available upon request.