Copula-Based Random Effects Models

Santiago Pereda-Fernández*

Banca d’Italia

September 17, 2015

PRELIMINARY, INCOMPLETE, PLEASE DO NOT CITE

Abstract

In a binary choice panel data framework, standard estimators usually assume that the unobserved heterogeneity is independent across individuals. Sorting, spillovers, or correlated confounders can create some correlation, and in those cases, these estimators cannot be used to estimate the probability of an event that depends on two or more individuals. In this paper I propose a random effects estimator that takes into account this correlation by using a parametric copula to model the dependence among the individuals in the same group. The likelihood function for each group is an integral whose dimension equal the size of the group, which requires potentially high-dimensional numerical integration. I propose an algorithm to approximate such integrals that works for a class of parametric copulas, overcoming the curse of dimensionality from which traditional methods like Monte Carlo integration suffer.

Keywords: Copula, High-dimensional integration, Nonlinear panel data

JEL classification: C33, C63

*Research Fellow at Banca d’Italia. Via Nazionale 91, 00184 Roma, Italy. The views presented in this paper do not necessarily reflect those of the Banca d’Italia. I can be reached via email at santiago.peredafernandez@esterni.bancaditalia.it
1 Introduction

Binary choice models are amongst the most frequently used estimators in the econometrician’s toolbox, and under the assumption that the unobservables of the individuals are mutually independent, these estimators can be used to compute the probability that a joint event occur.\footnote{I define a joint event as an event that involves the dependent variables of two or more individuals in the same group.} However, if the independence assumption is relaxed and the unobserved heterogeneity is allowed to be correlated across individuals, then it is not possible to identify the probability of an event that involves several individuals. In that situation, even if the marginal distribution of the dependent variable is identified, the joint distribution depends on the correlation structure of the unobservables, which in general is not known.

In a linear panel data setup, differencing the dependent variable out eliminates the need to control for any possible correlation it may display among individuals, whereas in nonlinear setup, it is not possible in general unless we make a particular parametric assumption.\footnote{Conditional fixed effects logit does not depend on the unobserved heterogeneity, although it does not identify the average partial effects or the probability of a joint event unconditionally.} Also, because of the incidental parameter problem, it is not wise to add individual dummies and then use them to estimate the correlation of the unobservables.

In this paper I propose an estimator of binary choice models that incorporates the correlation in the unobserved heterogeneity when there is a group structure in the data, \textit{i.e.} a number of individuals display a degree of correlation that can be caused by social interactions, correlated confounders, or sorting, but they are uncorrelated with individuals from other groups. In this setup, one possibility would be to model individuals’ behavior to derive the equilibrium conditions. However, neither unicity of the equilibrium, nor an easily implementable estimator is guaranteed, and if the model is misspecified, then the estimator may be inconsistent. The estimator I propose is less ambitious regarding uncovering the causality behind the correlation, but is easy to implement and acknowledges the existence of some correlation across individuals, which is necessary to estimate the probability of a joint event.
The estimator is an extension of the standard random effects estimators: under the assumption that the distribution of the unobserved heterogeneity is parametric and known up to a finite number of parameters, simply integrate out the probability of the data over this distribution, and maximize the likelihood function. The only difference with respect to standard random effects estimators is that this integral is now done at the group level rather than at the individual level. Hence, the integrals are multidimensional, with the dimension being the number of individuals in the group. Rather than proposing a joint multivariate distribution for the unobserved heterogeneity, I separately consider the marginal distribution and the copula, which together bring us the joint distribution. This provides a more flexible way of modeling the correlation and allows us to combine the same marginals with different copulas.

There are two contributions in this paper. First, I present the copula-based random effects estimator (CBRE), which is a generalization of the estimator used in Pereda-Fernández (2015), and it separately models the marginal distribution of the unobserved heterogeneity from its multivariate copula, which captures the correlation. Second, I propose an algorithm to numerically approximate high-dimensional integrals for a particular type of Archimedean copulas, including some of the most popular ones. I compare its performance to that of the common Monte Carlo integration, and show that this algorithm overcomes the curse of dimensionality for the type of problems considered in this paper.

2 Framework

Consider the following nonlinear panel data setup:

\[
\begin{align*}
    y_{ict} &= 1(y_{ict}^* \geq 0) \\
    y_{ict}^* &= \eta_{ic} + x_{ict}'\beta + \varepsilon_{ict}
\end{align*}
\]

(1)

for agent \(i = 1, \ldots, N_c\) in group \(c = 1, \ldots, C\) at time \(t = 1, \ldots, T\). The main departure of this framework from the usual one is that, for each group \(c\), the individual effects of the \(N_c\)
members of the group are correlated with each other, though they are independent of the individual effects of the members from other groups. This correlation can arise for a variety of reasons, such as spillovers and sorting.

**Example 1.** Suppose that students $i = 1, \ldots, N_c$ in classroom $c$ took an exam with $T$ questions, that were graded on a right/wrong basis. Then, the individual effect of all students can be correlated because of social interactions among students, or because they shared the same teacher. On the other hand, students from classroom $d$ neither interacted with students in classroom $c$, nor they had the same teacher, so there should be independence between classrooms. One quantity of interest for a principal would be the probability that all students or a subset of them get a passing grade. Because of the correlation in the unobserved heterogeneity, this probability is not the product of the probabilities for each single student. Therefore, this correlation needs to be estimated to consistently estimate the probability of all of them passing the exam.

**Example 2.** Consider a household in which we want to study the labor supply of the couple. During a period of crisis the proportion of unemployed workers increases. A marriage can act as an insurance mechanism: if at least one of the members of the couple has a job, then the household has some income, and if the proportion marriages with both partners unemployed is low, then (other things equal) so would be the between household inequality of income. There is ample evidence of sorting in marriages (Bruze, 2011; Charles et al., 2013), so it would reasonable to assume that there is some correlation between the unobservables of the husband and the wives. In order to analyze the impact of a crisis on labor income inequality, it is necessary to consistently estimate the probability that at least one member of the couple is employed for a number of periods of time, for which the estimation of the correlation in the couple’s individual effects is required.

To model this dependence, one possibility is to make a parametric assumption on the joint distribution of $\eta_c \equiv (\eta_{1c}, \ldots, \eta_{N_c})'$. Alternatively, it is possible to separately model the marginal distribution of each individual effect, $\eta_{ic}$, and the underlying correlation among
them, captured by the copula.$^3$ Copulas are modeled using the rank of the individual effects, i.e. $u_{ic} = F_{\eta}(\eta_{ic})$, which is a more convenient way to model the individual effects, since it is invariant to the marginal distribution of the effects.

To complete the model, assume that the distribution of the unobservables are given by $\varepsilon_{ict} \sim F_\varepsilon(\sigma_\varepsilon^2)$, $\eta_{ic} \sim F_\eta(\sigma_\eta^2)$, and the joint distribution of the individual effects in group $c$ is given by $u_c \sim C(\rho)$.$^4$ This model requires some normalization of the variances of the unobservables, as only their ratio is identified. Hence, normalize $\varepsilon_{ict}$ to have unit variance. Similarly, normalize the mean of both unobservables to zero. Then, the log-likelihood function is given by

$$L(\theta, \rho) = \sum_{c=1}^C \log \left( \int_{[0,1]^N} P_c(\theta) \, dC(u_c; \rho) \right) \equiv \sum_{c=1}^C l_c(\theta, \rho) \tag{2}$$

where $P_c(\theta) = \prod_{i=1}^N \prod_{t=1}^T [1 - F_\varepsilon(- (\eta_{ic} + x'_{ict}\beta))]^{y_{ict}} F_\varepsilon(- (\eta_{ic} + x'_{ict}\beta))^{1-y_{ict}}$ and $\eta_{ic} = F_{\eta}^{-1}(u_{ic})$. The CBRE estimator is given by $(\hat{\theta}', \hat{\rho}') = \arg\max_{\theta \in \Theta, \rho \in \mathbb{R}} L(\theta, \rho)$. Notice that this estimator requires the integration of a product over a potentially large dimensional space.

I discuss in section 4 how to approximate this integral for a particular family of copulas, the Archimedean copulas.$^5$ The asymptotic properties of the estimator and the estimation of the asymptotic variance are standard.$^6$ Regarding the average partial effect with respect to variable $x_{ictj}$, $APE(x_{ictj})$, notice that they depend only on the marginal distribution of $\eta_{ic}$, so there is no need to know the copula in order to identify them, and they can be computed the standard way.

---

$^3$Sklar (1959) showed that any multivariate cdf can be written in terms of a copula whose arguments are the marginal distributions, i.e. $\mathbb{P}(X_1 \leq x_1, \ldots, X_d \leq x_d) = C(F_1(x_1), \ldots, F_d(x_d))$.

$^4$Extending this model to allow for the existence of correlated random effects is straightforward, simply let the marginal distribution of $\eta_{ic}$ depend on $x_{ic} \equiv (x_{ict1}, \ldots, x_{ictT})'$. Consequently, the conditional ranks would be given by $u_{ic} = F_{\eta}(\eta_{ic}|x_{ic})$. Similarly, one could allow the copula to depend on $x_c$.

$^5$The score, which also requires a similar integration and is necessary to compute the estimator, is shown in appendix A.

$^6$See, for instance Newey and McFadden (1994).
2.1 Estimation of Joint Events

As argued in the introduction, if the econometrician is interested in identifying the probability of an event that involves several agents in the same group, then identification of the copula is necessary. Let the set $S$ denote all the permutations of $y_c \equiv (y_{1cT}, \ldots, y_{1cT}, \ldots, y_{NcT})$ that satisfy a condition $C$. In the first example, $S = \left\{ y_c : \sum_{t=1}^{T} y_{ict} \geq g^p \forall i \right\}$, where $g^p$ is the passing grade, and $y_{ict} = 1$ if student $i$ correctly answered question $t$. In the second example, $S = \left\{ y_c : y_{ict} + y_{2ct} \geq 1 \forall t \right\}$, where $y_{ict} = 1$ if individual $i$ is employed at time $t$.

In general, the probability of such events is given by

$$P (y_c \in S) = \sum_{d \in S} P (y_c = d)$$

$$= \sum_{d \in S} \int_{[0,1]^N_c} \prod_{i=1}^{N_c} \prod_{t=1}^{T} P (d_{ict}|x_{ict}, \eta_{ic}) \, dC (u_c; \rho)$$

(3)

where $P (d_{ict}|x_{ict}, \eta_{ic}) = \left[ 1 - F_\varepsilon (- (\eta_{ic} + x'_{ict}\beta)) \right]^{d_{ict}} F_\varepsilon (- (\eta_{ic} + x'_{ict}\beta))^{1-d_{ict}}$. Estimation of the probability of the set $S$ is straightforward: simply replace the parameters by $(\hat{\theta}', \hat{\rho})$ and approximate the integral as shown in section 4.

3 Comparison with Standard Random Effects Methods

The likelihood function given by equation 2 gives us an estimator of $\theta$. However, if one is not interested in the estimation of $\rho$, an alternative estimator of $\theta$ would simply ignore the within group correlation of the individual effects, and instead assume that they are independent, arriving to the following likelihood function:

$$\tilde{L} (\theta) = \sum_{c=1}^{C} \sum_{i=1}^{N_c} \log \left( \int_0^1 \tilde{P}_{ic} (\theta) \, du_{ic} \right)$$

(4)

where $\tilde{P}_{ic} (\theta) \equiv \prod_{t=1}^{T} [1 - F_\varepsilon (- (\eta_{ic} + x'_{ict}\beta))]^{y_{ict}} F_\varepsilon (- (\eta_{ic} + x'_{ict}\beta))^{1-y_{ict}}$. Notice that if

---

7The most common estimators are random effects logit and random effects probit. See Wooldridge (2010) for further details.
we denote by $\rho^\text{ind}$ the value of $\rho$ that makes the copula independent, then $\tilde{L}(\theta) = L(\theta, \rho^\text{ind})$, and hence equation 4 is a particular case of equation 2. This implies that $L(\hat{\theta}, \hat{\rho}) \geq L(\tilde{\theta}, \rho^\text{ind})$, where $\tilde{\theta} \equiv \arg\max_{\theta \in \Theta} \tilde{L}(\theta)$, so in general each estimator yields a different estimate, and the likelihood for the CBRE estimator cannot be smaller than the likelihood of the RE estimator. In any case, in the presence of a group structure, computation of the standard errors requires acknowledging the correlation, which the likelihood of the CBRE estimator does by default, unlike the RE estimator. For the latter, a correction is required to reflect the actual variability of the estimator.

4 Implementation

As in any discrete choice random effects model, implementing the estimator requires the integration of an integral that cannot be done analytically, and hence it is necessary to approximate those integrals using numerical methods. As equations 7 to 9 show, the Jacobian depends on the copula of the individual effects, which implies that if the size of the groups is large, so is the dimensionality of the integrals. Simulation methods like Monte Carlo tend to perform slowly in this kind of setup, and some recent advances in high dimensional numerical integration would outperform standard simulation methods. However, these methods do not overcome the curse of dimensionality. In this paper I propose an algorithm to numerically approximate the integrals. These algorithm works for a class of Archimedean copulas, and it overcomes the curse of dimensionality. In particular, even when the dimensionality of the integral increases, there exists a transformation of the integral into a two-dimensional one.

By Corollary 2.2 in Marshall and Olkin (1988), an Archimedean copula can be written as

$$C(u) = \int_{[0,1]^N} \exp \left( - \sum_{i=1}^N \theta_i \phi^{-1}_i(u_i) \right) dG(\theta)$$

---

8The independent copula satisfies $C(u_c) = \prod_{i=1}^{N_c} u_{ic}$.

9See, for instance, Heiss and Wentschel (2008) or Skrainka and Judd (2011).

10A copula is Archimedean if $C(u_1, ..., u_d; \rho) = \phi \left( \sum_{i=1}^d \phi^{-1}(u_i; \rho) \right)$, where $\phi$ is the so called generator function. This family includes some of the most popular copulas, such as Clayton, Frank, or Gumbel.
where $G$ is the cdf of $\theta$, and $\phi_i$ is the Laplace transform of the marginal distributions of $G$. Some of the most common copulas have $\theta$ unidimensional and $\phi_i = \phi \forall i$. For example, if the copula is a Clayton ($\rho$), then $G$ is the cdf of a $\Gamma (\rho, 1)$, and $\phi$, also known as the generator of the copula, equals $\phi(\cdot; \rho) = (1 + \cdot)^{-\rho}$. Now consider the following integral:

\[
I = \int_{[0,1]^N} \prod_{i=1}^N \ell_i(u_i) \, dC(u)
= \int_0^1 \prod_{i=1}^N \left[ \int_0^1 \ell_i(u_i) \, dF^\theta(u_i) \right] \, dG(\theta) \tag{6}
\]

where $F^\theta(u_i) = \exp(-\theta \phi^{-1}(u_i))$. The originally integral of dimension $N$ equals the integral of the product of $N$ independent integrals, reducing the dimensionality from $N$ to 2. Hence, an algorithm to approximate the integral is given by

1. Compute a grid of values of $\theta$, given by $\theta_j = G^{-1}\left(\frac{j}{N_1+1}\right)$, $\forall j = 1, ..., N_1$.
2. Compute a grid of values of $u$ $\forall j$, given by $u_{jh} = \phi\left(-\frac{1}{\theta_j} \log\left(\frac{h}{N_2+1}\right)\right)$, $\forall h = 1, ..., N_2$.
3. Approximate the integral by $\hat{I} = \frac{1}{N_1} \sum_{j=1}^{N_1} \prod_{i=1}^N \left[ \frac{1}{N_2} \sum_{h=1}^{N_2} \ell_i(u_{jh}) \right]$.

To understand how the algorithm works, consider the integration of a function $g$ with respect to a distribution $F$: $\int_0^1 g(x) \, dF(x)$. One possibility is to split the unit interval into $Q+1$ intervals of equal probability using $Q$ evenly spaced quantiles: $0 < \frac{1}{Q+1} < ... < \frac{Q}{Q+1} < 1$. Then we can approximate the integral by evaluating the function $g$ at these quantiles and taking the average across quantiles: $\frac{1}{Q+1} \sum_{q=1}^Q g(x_q)$, where $x_q = F^{-1}\left(\frac{q}{Q+1}\right)$ is the $q$th 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. Suppose that we want to do an integration using a bivariate Clayton copula. Fix $\theta$ in equation 6, we have $N_2$ 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 $\theta$. Thus, it is possible to approximate the interior integral of each dimension $j$ as $\frac{1}{N_2+1} \sum_{h=1}^{N_2} \ell_i(u_{jh})$. 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, and therefore the product of the interior integrals is straightforward. Then, only the approximation of the exterior integral remains, for which we intuitively take the average across squares. In figure 1, as we move from the upper to the lower graphs, the number of squares increases. And as \( N_1, N_2 \to \infty \), the unit square is covered by more points and \( \hat{I} \to I \). For higher order dimensions, the intuition is the same, except for the fact that, for each \( \theta \), rather than a square composed of \( N_2^2 \) points, we have a hypercube of \( N_2^d \) points.

![Figure 1: Support of the integral](image)

4.1 Numerical Performance of the Algorithm

To show the performance of the algorithm, I numerically approximate one integral using this method and compare it to the traditional Monte Carlo. I consider the following integral \( \int_{[0,1]^d} \prod_{j=1}^d \sqrt{u_j} dC(u_1, \ldots, u_d; \rho) \) for a Clayton (4) copula with dimension \( d = \{2, 3\} \). For the algorithm used in this paper I set \( N_1 = N_2 = \{9, 19, 49, 99\} \). The number of draws in the Monte Carlo equals the number of points evaluated by the algorithm, i.e. \( N_1 N_2^d \) draws, and the number of repetitions was 100. The sampling algorithm for the Monte Carlo is the one

\footnote{The statistical program used for this exercise was MATLAB.}
proposed by Marshall and Olkin (1988). Table 1 shows the mean value of both methods, the standard deviation of the Monte Carlo across repetitions (the algorithm provided in this paper yields always the same result.), and the average time spent per repetition.

<table>
<thead>
<tr>
<th></th>
<th>$d = 2$</th>
<th>$d = 3$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$N_1 = N_2 = 9$</td>
<td>$N_1 = N_2 = 19$</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>Mean</td>
</tr>
<tr>
<td></td>
<td>New method</td>
<td>Monte Carlo</td>
</tr>
<tr>
<td></td>
<td>0.4933</td>
<td>0.4942</td>
</tr>
<tr>
<td></td>
<td>0.4944</td>
<td>0.4940</td>
</tr>
<tr>
<td></td>
<td>0.4934</td>
<td>0.4940</td>
</tr>
<tr>
<td></td>
<td>0.4936</td>
<td>0.4940</td>
</tr>
<tr>
<td></td>
<td>0.4937</td>
<td>0.4940</td>
</tr>
<tr>
<td></td>
<td>S.D.</td>
<td>S.D.</td>
</tr>
<tr>
<td></td>
<td>New method</td>
<td>Monte Carlo</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>0.0112</td>
</tr>
<tr>
<td></td>
<td>0.0001</td>
<td>0.0033</td>
</tr>
<tr>
<td></td>
<td>0.0027</td>
<td>0.0036</td>
</tr>
<tr>
<td></td>
<td>0.0036</td>
<td>0.0011</td>
</tr>
<tr>
<td></td>
<td>0.0569</td>
<td>0.0001</td>
</tr>
<tr>
<td></td>
<td>0.4776</td>
<td>0.0003</td>
</tr>
<tr>
<td></td>
<td>Time</td>
<td>Time</td>
</tr>
<tr>
<td></td>
<td>New method</td>
<td>Monte Carlo</td>
</tr>
<tr>
<td></td>
<td>0.0001</td>
<td>0.0003</td>
</tr>
<tr>
<td></td>
<td>0.5580</td>
<td>0.0828</td>
</tr>
<tr>
<td></td>
<td>0.0001</td>
<td>0.0003</td>
</tr>
<tr>
<td></td>
<td>0.0036</td>
<td>0.0011</td>
</tr>
<tr>
<td></td>
<td>0.4776</td>
<td>0.0003</td>
</tr>
</tbody>
</table>

Even when the dimensionality of the problem is low, the algorithm proposed in this paper is several orders of magnitude faster than the traditional Monte Carlo, and it increases substantially when the dimension of the integral increases. Regarding the accuracy, when the dimension is 2, for a given number of points, the performance is quite similar, and the approximation is within two standard deviations of the Monte Carlo. Even though the new algorithm consistently reports a number inferior to the mean across repetitions of the Monte Carlo, 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 when spending the same time to get the answer. When $d = 3$, the algorithm I propose loses some accuracy with respect to the Monte Carlo, but the time gains are so large, that again for a given amount of time, the new algorithm is more accurate. Table 2 digs deeper into the issue of the dimensionality for the algorithm, showing that its accuracy decreases with the dimensionality of the problem, as seen by the changes in the approximation when we increase the number of points used to
evaluate the integral. Computational time, however, remains practically unchanged despite
the increase of the dimensionality.

<table>
<thead>
<tr>
<th>Table 2: Performance in high dimensions</th>
</tr>
</thead>
<tbody>
<tr>
<td>( N_1 = N_2 = 9 ) \hspace{1cm} ( N_1 = N_2 = 19 ) \hspace{1cm} ( N_1 = N_2 = 49 ) \hspace{1cm} ( N_1 = N_2 = 99 ) \hspace{1cm} ( N_1 = N_2 = 199 )</td>
</tr>
<tr>
<td>( d = 2 )</td>
</tr>
<tr>
<td>Mean \hspace{1cm} 0.4933 \hspace{1cm} 0.4934 \hspace{1cm} 0.4936 \hspace{1cm} 0.4937 \hspace{1cm} 0.4938</td>
</tr>
<tr>
<td>Time \hspace{1cm} 0.0001 \hspace{1cm} 0.0003 \hspace{1cm} 0.0011 \hspace{1cm} 0.0032 \hspace{1cm} 0.0108</td>
</tr>
<tr>
<td>( d = 3 )</td>
</tr>
<tr>
<td>Mean \hspace{1cm} 0.3786 \hspace{1cm} 0.3827 \hspace{1cm} 0.3854 \hspace{1cm} 0.3863 \hspace{1cm} 0.3868</td>
</tr>
<tr>
<td>Time \hspace{1cm} 0.0001 \hspace{1cm} 0.0003 \hspace{1cm} 0.0011 \hspace{1cm} 0.0035 \hspace{1cm} 0.0107</td>
</tr>
<tr>
<td>( d = 5 )</td>
</tr>
<tr>
<td>Mean \hspace{1cm} 0.2452 \hspace{1cm} 0.2534 \hspace{1cm} 0.2584 \hspace{1cm} 0.2602 \hspace{1cm} 0.2610</td>
</tr>
<tr>
<td>Time \hspace{1cm} 0.0001 \hspace{1cm} 0.0003 \hspace{1cm} 0.0011 \hspace{1cm} 0.0035 \hspace{1cm} 0.0109</td>
</tr>
<tr>
<td>( d = 10 )</td>
</tr>
<tr>
<td>Mean \hspace{1cm} 0.1076 \hspace{1cm} 0.1176 \hspace{1cm} 0.1238 \hspace{1cm} 0.1260 \hspace{1cm} 0.1271</td>
</tr>
<tr>
<td>Time \hspace{1cm} 0.0001 \hspace{1cm} 0.0003 \hspace{1cm} 0.0011 \hspace{1cm} 0.0033 \hspace{1cm} 0.0109</td>
</tr>
<tr>
<td>( d = 50 )</td>
</tr>
<tr>
<td>Mean \hspace{1cm} 0.0017 \hspace{1cm} 0.0033 \hspace{1cm} 0.0049 \hspace{1cm} 0.0057 \hspace{1cm} 0.0062</td>
</tr>
<tr>
<td>Time \hspace{1cm} 0.0001 \hspace{1cm} 0.0003 \hspace{1cm} 0.0011 \hspace{1cm} 0.0033 \hspace{1cm} 0.0108</td>
</tr>
</tbody>
</table>

Figure 2 compares graphically in a two dimensional setup the algorithm to the Monte
Carlo: rather than selecting points with a probability equal to the pdf of the copula, the
algorithm systematically selects a grid that depends on the copula, which changes for different
values of \( \rho \). As \( N_1 \) and \( N_2 \) increase, the grid gets larger and larger, eventually covering the
whole hyperplane.

References


*Demography* 50(1), 51–70.
The number of points used in each of the three simulations was $5^3$, $10^3$ and $15^3$.


Appendix

A Jacobian and Hessian

Let $F_{ict}$ and $f_{ict}$ be shorthand for $F_{\varepsilon}(-(\eta_{ic} + x'_{ict} \beta))$ and $f_{\varepsilon}(-(\eta_{ic} + x'_{ict} \beta))$, denote the quantile function of $\eta_{ic}$ by $Q_{\eta}(u) \equiv F_{\eta}^{-1}(u)$ and by $q_{\eta}(u; \sigma_{\eta})$ its derivative with respect to $\sigma_{\eta}$. Then, the score is given by

$$\frac{\partial L(\theta, \rho)}{\partial \beta} = \sum_{c=1}^{N_c} \int_{[0,1]} N_c P_c(\theta) \sum_{i=1}^{N_c} \sum_{t=1}^{T} \frac{f_{\varepsilon}(y_{ict} - (1 - F_{\varepsilon}))}{F_{\varepsilon}(1 - F_{\varepsilon})} x_{ict} dC(u_c; \rho)$$

(7)

$$\frac{\partial L(\theta, \rho)}{\partial \sigma_{\eta}} = \sum_{c=1}^{N_c} \int_{[0,1]} N_c P_c(\theta) \sum_{i=1}^{N_c} \sum_{t=1}^{T} \frac{f_{\varepsilon}(y_{ict} - (1 - F_{\varepsilon})) q_{\eta}(u_{ict}; \sigma_{\eta})}{F_{\varepsilon}(1 - F_{\varepsilon})} dC(u_c; \rho)$$

(8)

$$\frac{\partial L(\theta, \rho)}{\partial \rho} = \sum_{c=1}^{N_c} \int_{[0,1]} N_c \frac{\partial^{N_c+1} c(u_c; \rho)}{\prod_{i=1}^{N_c} du_{ic}} \prod_{i=1}^{N_c} d u_{ic}$$

(9)

where $c(u_c; \rho)$ is the pdf of the copula. Thus, it is immediate to approximate equations 7 and 8 using the numerical integration presented in this paper. Regarding equation 9, it is more convenient to numerically evaluate the derivative, i.e. $\frac{\partial L(\theta, \rho)}{\partial \rho} \approx \frac{L(\theta, \rho+\epsilon) - L(\theta, \rho)}{\epsilon}$. Finally, the Hessian is easily computed by

$$\hat{H}(\hat{\theta}, \hat{\rho}) = \frac{1}{C} \sum_{c=1}^{C} \frac{\partial^{2} L(\theta', \rho')}{\partial (\theta', \rho')^{T}} \frac{\partial^{2} L(\theta', \rho')}{\partial (\theta', \rho')^{T}}$$

12 If $\eta_{ic}$ belongs to a scale family of distributions, i.e. if $\eta_{ic} = \sigma_{\eta} \tilde{\eta}_{ic}$, where $\tilde{\eta}_{ic} \sim F_{\eta}(1)$, then $Q_{\eta}(u_{ic}) = \sigma_{\eta} \tilde{\eta}_{ic}$, and thus $q_{\eta}(u_{ic}; \sigma_{\eta})$ simplifies to $q_{\eta}(u_{ic}; 1) = \tilde{\eta}_{ic}$.