Hasebe, Takuya; Vijverberg, Wim P. M.

Working Paper

A flexible sample selection model: A GTL-copula approach

IZA Discussion Papers, No. 7003

Provided in Cooperation with:
IZA – Institute of Labor Economics

Suggested Citation: Hasebe, Takuya; Vijverberg, Wim P. M. (2012) : A flexible sample selection model: A GTL-copula approach, IZA Discussion Papers, No. 7003, Institute for the Study of Labor (IZA), Bonn

This Version is available at:
https://hdl.handle.net/10419/67274

DISCUSSION PAPER SERIES

IZA DP No. 7003

A Flexible Sample Selection Model: A GTL-Copula Approach

Takuya Hasebe
Wim Vijverberg

November 2012

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

A Flexible Sample Selection Model: A GTL-Copula Approach

Takuya Hasebe
CUNY Graduate Center

Wim Vijverberg
CUNY Graduate Center
and IZA

Discussion Paper No. 7003
November 2012

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

IZA Discussion Paper No. 7003
November 2012

ABSTRACT

A Flexible Sample Selection Model: A GTL-Copula Approach*

In this paper, we propose a new approach to estimating sample selection models that combines Generalized Tukey Lambda (GTL) distributions with copulas. The GTL distribution is a versatile univariate distribution that permits a wide range of skewness and thick- or thin-tailed behavior in the data that it represents. Copulas help create versatile representations of bivariate distribution. The versatility arising from inserting GTL marginal distributions into copula-constructed bivariate distributions reduces the dependence of estimated parameters on distributional assumptions in applied research. A thorough Monte Carlo study illustrates that our proposed estimator performs well under normal and nonnormal settings, both with and without an instrument in the selection equation that fulfills the exclusion restriction that is often considered to be a requisite for implementation of sample selection models in empirical research. Five applications ranging from wages and health expenditures to speeding tickets and international disputes illustrate the value of the proposed GTL-copula estimator.

JEL Classification: C24, C35

Keywords: sample selection, copula, Generalized Tukey Lambda distribution

Corresponding author:

Wim P.M. Vijverberg
Department of Economics
City University of New York Graduate Center
365 Fifth Avenue
New York, NY 10016-4309
USA
E-mail: wvijverberg@gc.cuny.edu

* This research was supported, in part, by a grant of computer time from the City University of New York High Performance Computing Center under NSF Grants CNS-0855217 and CNS-0958379. The authors thank seminar participants at Kansai Labor Forum (Roudou Kenkyu Kai), Keio University, and the Melbourne Institute.

1 Introduction

Sample selection is an issue that arises frequently in empirical studies, especially with micro-data. Nonrandom data that are subject to sample selection yield estimates that may be biased and inconsistent, which harms inference about economic theory and guidance for policy making. Ever since the seminal work of Heckman (1974, 1979), sample selection has been an important issue in both theoretical and applied microeconometrics. The typical sample selection model consists of a selection equation and an outcome equation, where the outcome is observable only for the subsample of data. Lee (1978) extends the selection model to the case where outcome equations differ by regimes and a selection equation illustrate the sorting mechanism.

In the footsteps of these seminal works, most empirical applications follow a parametric approach where the model is estimated with the full information maximum likelihood method or a two-step (limited information maximum likelihood) method under the assumption of the joint normality. However, these estimators are criticized for their sensitivity to the distributional assumption. In general, violation of the normality assumption leads to inconsistency. For example, the Monte Carlo study by Zuehlke and Zeman (1991) shows that violation of the joint normal assumption results in biased and imprecise estimates. Newey (1999) shows that under certain assumptions, the two-step estimator can be consistent even when the distribution is misspecified.

The recent theoretical literature has paid more attention to semi-parametric or non-parametric approaches, which do not require any parametric distributional assumptions. The semiparametric approaches usually take a two-step estimation procedure, where the selection correction term is estimated semi- or non-parametrically: see, for example, Cosslett (1993), Ahn and Powell (1993), and Newey (2009). Das et al. (2003) develop a nonparametric approach, which allows an outcome equation to be non-parametric. While such estimators weaken the distributional assumptions, parametric estimation yields greater efficiency provided that the distribution is correctly specified.[1]

This paper proposes an alternative maximum likelihood estimator of the sample selection model, replacing the standard assumption of a joint normality with a more flexible distributional assumption. Early work on selection models that relaxes the normality assumption was done by Lee (1982, 1983). His approach was to transform nonnormal disturbances in the model into normal variates that are then assumed to be jointly normally distributed. As we will see, this is a special case of the copula approach that Smith (2003) applies to the sample selection model. Under the copula approach, a multivariate distribution is constructed from separately specified marginal distributions. Our proposed estimator continues along this line of research by inserting more flexible marginal distributions into the copula function.

[1] See Vella (1998) and Lee (2001) for a survey of the literature. In addition to the maximum likelihood and semiparametric approaches, the literature has added various other estimation methods. Golan et al. (2004) develop a generalized maximum entropy (GME) estimator, which may perform well with small samples. van Hassel (2011) develops a Bayesian inference of the selection model, where errors are both normally and non-normally distributed. Meijer and Wansbeek (2007) discuss estimation of the selection model within the Generalized Method of Moments (GMM) framework.

In particular, we model the marginal distributions with the Generalized Tukey lambda (GTL) distribution. First proposed by Pregibon (1980) and Freimer et al. (1988), the GTL distribution is very flexible, allowing varying degrees of asymmetry and tail thickness. For example, it nests the logistic and uniform distributions and approximately nests familiar distributions such as the normal, Student's t, Gumbel, and χ^2 distributions. Because of this versatility, the GTL is a good candidate distribution for modelling the unobservables in the selection and outcome equations. These GTLs are then inserted as marginal distributions into a copula in order to construct a highly versatile bivariate distribution that replaces the bivariate normality distribution that underlies the approach of Heckman (1979) and Lee (1978). The flexibility of this GTL-copula distribution effectively frees the sample selection model from any particular distributional assumption. Moreover, this flexibility is achieved with just a few additional parameters, which is both parsimonious and time-efficient relative to semi- or non-parametric approaches. The estimated parameters also indicate whether the distribution deviates from normality.

A new estimator has value if its strengths are exploited in applications with real data. We report on five applications on a wide variety of topics: wages of married women, wages of children in a lower-income country, health expenditures, speeding tickets, and international disputes. In all five, the joint normality assumption is rejected, and nine of the ten marginal densities are decidedly non-normal. Not surprisingly, the estimates of the selection and outcome equations differ significantly as well: in varying ways, the distributional misspecification changes the magnitude of the estimated economic effects and the interpretation of the estimated relationships. This sample of five applications is not really too self-selected: these were the only applications we examined.

The structure of this paper is as follows. The next section outlines the sample selection model and states the likelihood function in its general form. Section 3 lays out the copula approach with special reference to the sample selection model. This is followed in Section 4 by an introduction of the GTL distribution and a description of its attractive properties. In Section 5, we report on a Monte Carlo study of the proposed GTL-copula estimator: it performs well under both normal and nonnormal designs, whereas the standard estimator that assumes joint normality is subject to substantial bias if the distributional assumption is violated. In Section 6, we examine the relevance of the GTL-copula estimator in the five real-world applications that were mentioned above, in comparison with the familiar estimator of the joint normal sample selection model. Section 7 concludes the paper.

2 The Sample Selection Model

In this paper, we consider the simplest form of the sample selection model, which is sometimes referred to as a type 2 Tobit model (Amemiya, 1985). This model consists of two latent equations: a selection equation and an outcome equation. For observation i, $i = 1, \ldots, N$, the selection equation is

$$s_i = 1(z_i'\gamma + \nu_i > 0) \tag{1}$$

where $1(\cdot)$ is an indicator function, and the outcome equation is

$$y_i = x_i{}'\beta + \sigma\varepsilon_i, \tag{2}$$

where σ is a scale parameter.[2] (ν_i, ϵ_i) is independent of (z_i', x_i'), and (ν_i, ϵ_i) is identically and independently distributed across observations in the sample.

The outcome equation is of primary interest, but the outcome is observable only when $s_i = 1$. When ν_i and ε_i are not independent of each other, OLS estimation of equation (2) with the subsample for which $s_i = 1$ yields inconsistent estimates of β. This is the well-known selectivity bias problem (Heckman, 1979).

For expository simplicity, we focus on this model in the following discussions. However, it is straightforward to extend our proposed method to other variants of the selection model. For example, a different outcome might be observed depending on whether $s_i = 0$ or $s_i = 1$. This so-called switching regression model is also known as the Roy model or as a type 5 Tobit model (Roy, 1951; Amemiya, 1985).

Equations (1)-(2) may be estimated by the full information maximum likelihood method (Heckman, 1974). In a general form, the likelihood function of the sample selection model may be written as

$$L = \prod_{i=1}^{N} \left[\int_{-\infty}^{-z_i{}'\gamma} f_\nu(\nu) d\nu \right]^{s_i=0} \left[\int_{-z_i{}'\gamma}^{\infty} f_{\nu\varepsilon}(\nu, \varepsilon_i) d\nu \right]^{s_i=1}, \tag{3}$$

where f_ν is a univariate pdf of ν, and $f_{\nu\varepsilon}$ is a bivariate pdf of ν and ε. To implement maximum likelihood estimation, the functional forms of f_ν and $f_{\nu\varepsilon}$ must first be specified. The standard assumption is that ν_i and ε_i are jointly normally distributed. This leads to the following likelihood function:

$$L = \prod_{i=1}^{N} \left[\Phi(-z_i{}'\gamma)\right]^{s_i=0} \left[\sigma^{-1}\phi\left(\frac{y_i - x_i{}'\beta}{\sigma}\right) \Phi\left(\frac{z_i{}'\gamma + (\rho/\sigma)(y_i - x_i{}'\beta)}{\sqrt{1-\rho^2}}\right) \right]^{s_i=1} \tag{4}$$

where σ is now the standard deviation of ε and ρ is the coefficient of correlation between ε and ν. $\phi(\cdot)$ and $\Phi(\cdot)$ are the pdf and cdf of the standard normal distribution, respectively.

Although this maximum likelihood estimator (and the closely related two-step estimator of Heckman (1979) and Lee (1978)) is widely used in empirical applications, it is criticized for its relatively strong assumption of the normality. Generally, violation of the normality assumption results in inconsistency.

This paper relaxes the joint normality assumption while maintaining the parametric structure. Our proposed method follows and extends the copula approach suggested by Smith (2003).

[2] The scale parameter for the selection equation is set to 1 for identification.

3 The Copula Approach

A copula is a parametric representation of a joint distribution with given marginal distributions, thus permitting flexible dependence structures. For an introduction to copulas, see Nelsen (2006) and Trivedi and Zimmer (2007). Copulas have been widely used in the finance literature (e.g., see Cherubini et al. (2004) and references therein). Prokhorov and Schmidt (2009) discuss the estimation of panel data models with copulas to capture dependence over time. Cameron et al. (2004) use a copula to model two equations for count data; Zimmer and Trivedi (2006) add a binomial outcome equation to two of such equations. In this section, we briefly discuss the copula approach with particular reference to the sample selection model, drawing in particular on Smith (2003).

Let W_j be a continuous random variable with a marginal distribution $F_j = F_j(\omega_j) = Pr(W_j \leq \omega_j)$ for $j = 1, 2$. Define a joint distribution of these two random variables as $F_{12}(\omega_1, \omega_2) = Pr(W_1 \leq \omega_1, W_2 \leq \omega_2)$. A copula function $C(\cdot)$ couples the two marginal distributions together to generate the joint distribution:

$$F_{12}(\omega_1, \omega_2) = C(F_1, F_2; \theta),$$

where θ is a (vector of) parameter(s) that governs the degree of dependence between the random variables. The properties of the copula function are that (i) $C(F_1, 0; \theta) = C(0, F_2; \theta) = 0$, (ii) $C(F_1, 1; \theta) = F_1$ and $C(1, F_2; \theta) = F_2$, and (iii) it is 2-increasing. The third property is a technical expression that implies $\partial^2 C / \partial F_1 \partial F_2 \geq 0$, which in turn guarantees that the bivariate pdf is non-negative.[3] Note also $C(F_1, F_2; \theta) = C(F_2, F_1; \theta)$.

Given a joint cdf, the density function of $W_1 = \omega_1$ conditional on $W_2 \leq \omega_2$ is obtained simply by the chain rule:

$$\frac{\partial}{\partial \omega_1} F_{12}(\omega_1, \omega_2) = \frac{\partial}{\partial F_1} C(F_1, F_2; \theta) \times \frac{\partial F_1}{\partial \omega_1}, \tag{5}$$

and $\partial F_1 / \partial \omega_1$ is simply a univariate pdf, $f_1(\omega_1)$. $f_2(\omega_2 | W_1 \leq \omega_1)$ is similarly derived.

In order to rewrite the likelihood function of (3) in terms of a copula, note that the integral inside the second pair of brackets in (3) can be rewritten

$$\int_{-z_i'\gamma}^{\infty} f_{\nu\varepsilon}(\nu, \varepsilon_i) d\nu = \frac{\partial}{\partial \varepsilon} \left(F_\varepsilon(\varepsilon) - F_{\nu\varepsilon}(-z_i'\gamma, \varepsilon) \right) |_{\varepsilon = \varepsilon_i},$$

where $F_\varepsilon(\cdot)$ is a univariate cumulative distribution function (cdf) of ε and $F_{\nu\varepsilon}(\cdot)$ is a bivariate

[3] Let $F = C(F_1, F_2; \theta)$. The bivariate pdf is derived as $\partial^2 F / \partial \omega_1 \partial \omega_2 = (\partial^2 C / \partial F_1 \partial F_2) \times (\partial F_1 / \partial \omega_1) \times (\partial F_2 / \partial \omega_2)$. Furthermore, the third property has the following implication for continuous and discontinuous copula functions alike (Nelsen, 2006): for every $F_{1a} \leq F_{1b}$ and $F_{2a} \leq F_{2b}$, we have $C(F_{1b}, F_{2b}) - C(F_{1b}, F_{2a}) - C(F_{1a}, F_{2b}) + C(F_{1a}, F_{2a}) \geq 0$.

cdf of ν and ε. Using (5), the likelihood function is

$$L = \prod_{i=1}^{N} \left[F_\nu(-z_i'\gamma)\right]^{s_i=0} \left[\left(1 - \frac{\partial}{\partial F_\varepsilon} C(F_\nu(-z_i'\gamma), F_\varepsilon(\varepsilon_i); \theta)\right) \times f_\varepsilon(\varepsilon_i)\right]^{s_i=1},$$

since the integral inside the first pair of brackets of (3) is simply the cdf of ν, $F_\nu(\cdot)$.

Many different copula functions are available, each with its own characteristics. Here, we describe six of them that we will use later in this paper. One of the most frequently used copulas is the Gaussian copula:

$$C(F_1, F_2; \theta) = \Phi_2(\Phi^{-1}(F_1), \Phi^{-1}(F_2); \theta),$$

where $\Phi_2(\cdot)$ is a cumulative distribution function of a bivariate normal distribution with a coefficient of correlation θ, $-1 \leq \theta \leq 1$, which is a dependence parameter in the copula framework. If marginal distributions of W_1 and W_2 are normal, then the joint distribution is reduced to a bivariate normal distribution; if even only one of the marginal distributions is other than normal, it is not. In the context of sample selection models, this Gaussian copula appears as part of a two-step estimator in Lee (1982) and in the context of a FIML estimator in Lee (1983). The common selection model of Heckman (1974) uses a bivariate normal distribution, which is a Gaussian copula with normal marginals. Throughout this paper, we will call this the normal-Gaussian model.

Another example is the FGM (Farlie-Gumbel-Morgenstern) copula:

$$C(F_1, F_2; \theta) = F_1 F_2 (1 + \theta(1 - F_1)(1 - F_2)),$$

where θ is a dependence parameter, $-1 \leq \theta \leq 1$. Prieger (2002) inserts this FGM copula into a selection model of hospitalization duration. One of the attractive features of this copula is its mathematical simplicity, which facilitates computation since no integration is involved (unlike the Gaussian copula[4]).

Archimedean copulas belong to a family of copulas that are generically defined by a generator function $\varphi(\cdot)$ that is a continuous, convex and decreasing function with $\varphi(1) = 0$:

$$C(F_1, F_2; \theta) = \varphi^{-1}(\varphi(F_1) + \varphi(F_2)).$$

Table 1 lists four examples of such generator functions, each dependent on a single parameter θ that determines the degree of dependence, as will be discussed later on.[5] Archimedean

[4]If selection is not merely binomial but rather multinomial in unordered ways, the use of a Gaussian copula necessitates integration of a multivariate normal distribution, which becomes a tedious chore.

[5]A longer list of the family of Archimedean and non-Archimedean copulas is available in, for example, Nelsen (2006). Moreover, although this paper considers only copulas with one dependence parameter, there exist copulas with more than one dependence parameters. For example, a Student's t copula is given by $C(F_1, F_2; \theta_1, \theta_2) = t_2(t_{\theta_2}^{-1}(F_1), t_{\theta_2}^{-1}(F_2); \theta_1, \theta_2)$, where $t_2(\cdot; \theta_1, \theta_2)$ is a bivariate Student's t cdf with coefficient correlation θ_1 and θ_2 degrees of freedom, and $t_{\theta_2}(\cdot)$ is the inverse of univariate cdf of Student's t with θ_2 degrees of freedom.

Table 1: Archimedean copula and generator functions

Copula Name	$C(u_1, u_2; \theta)$	$\varphi(t)$
Clayton	$\left(F_1^{-\theta} + F_2^{-\theta} - 1\right)^{-1/\theta}$	$\theta^{-1}\left(t^{-\theta} - 1\right)$
Frank	$-\theta^{-1}\log\left(1 + \frac{(e^{-\theta F_1} - 1)(e^{-\theta F_2} - 1)}{(e^{-\theta} - 1)}\right)$	$-\log\left(\frac{e^{-\theta t} - 1}{e^{-\theta} - 1}\right)$
Gumbel	$\exp\left(-\left((-\log F_1)^\theta + (-\log F_2)^\theta\right)^{1/\theta}\right)$	$(-\log(t))^\theta$
Joe	$1 - \left[(\tilde{F}_1)^\theta + (\tilde{F}_2)^\theta - (\tilde{F}_1\tilde{F}_2)^\theta\right]^{1/\theta}$	$-\log\left(1 - (1-t)^\theta\right)$

For Joe, $\tilde{F}_j = 1 - F_j$ for $j = 1, 2$.

copulas have several attractive attributes (Smith, 2003). First, similar to the FGM copula, they do not require integration. Second, their mathematical structure facilitates the calculation of the likelihood function and its scores and Hessian. For example, derivatives of C follow straightforwardly with use of the rule for a derivative of an inverse function:

$$\frac{\partial}{\partial F_1} C(F_1, F_2; \theta) = \frac{\varphi'(F_1)}{\varphi'(C(F_1, F_2; \theta))},$$

where $\varphi'(\cdot)$ is a derivative of $\varphi(\cdot)$.

More importantly, Archimedean copulas are attractive since they exhibit various dependence structures. This is illustrated in Figure 1, which shows a contour plot of the bivariate pdf for each copula, where the marginal distributions are standard normal and the overall degree of dependence is the same for each case except FGM.[6] A Frank copula exhibits symmetric dependence in that the degree of dependence is the same in the lower and upper tails of a joint distribution.[7] In this aspect, the Frank copula is similar to the Gaussian and FGM copulas, but its dependence is weaker in the tails than the Gaussian one. In contrast, the Clayton copula is asymmetric with strong lower tail dependence but weaker upper tail dependence, and the Gumbel and Joe copulas exhibit strong upper tail but weaker dependence in a lower tail.

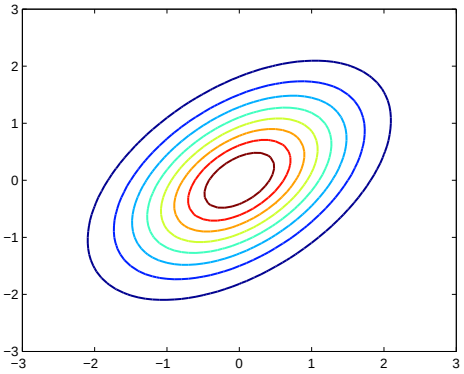
In application, the dependence structure is rarely known in advance, but the choice of the copula function does matter for the fit of the model. Copulas are not nested relative to each other. Thus, information criteria such as the Akaike Information Criterion (AIC) or the Bayesian Information Criterion (BIC) are useful in selecting the best-fitting copula.[8]

[6]For each copula, Kendall's τ equals 0.333, except for FGM where τ equals 0.2.

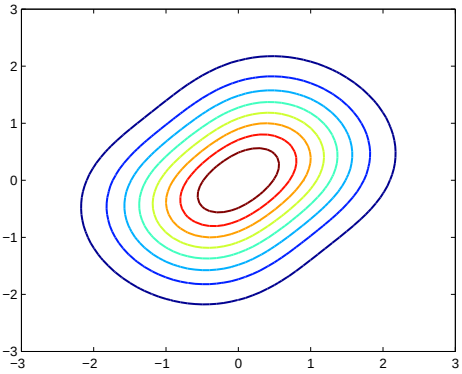
[7]Formally, a copula C is said to be radially symmetric if $C(F_1, F_2; \theta) = F_1 + F_2 - 1 + C(1 - F_1, 1 - F_2; \theta)$. However, a joint distribution generated with a radially symmetric copula is symmetric only if marginal distributions are symmetric as well.

[8]AIC is defined as $-2\ln L + 2k$ and BIC as $-2\ln L + (\ln N)k$, where $\ln L$ is the maximized log likelihood and k is the number of the parameters in the model. The copula with the smallest information criterion is preferred. When marginal distributions are fixed across copulas, the selection based on the smallest

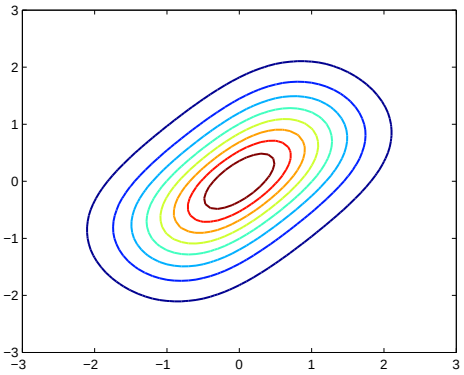
Figure 1: Contour plots of bivariate pdf for different copulas



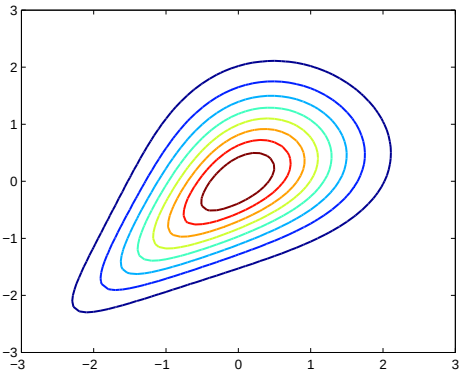
(a) Gaussian copula: $\theta = 0.5$



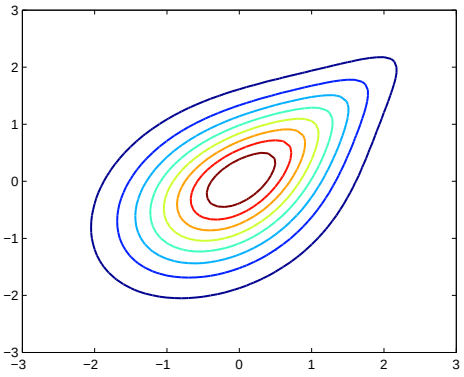
(b) FGM copula: $\theta = 0.8$



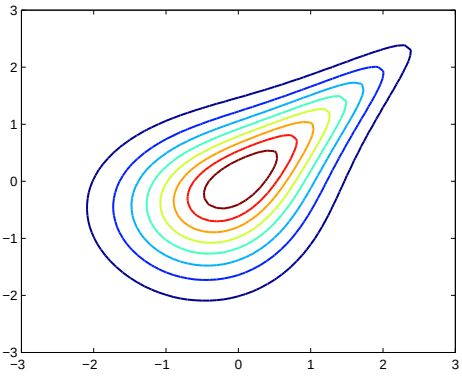
(c) Frank copula: $\theta = 3.3058$



(d) Clayton copula: $\theta = 1$



(e) Gumbel copula: $\theta = 1.5$



(f) Joe copula: $\theta = 1.9050$

Alternatively, Vuong's (1989) test can be used to weigh one copula against another.[9] To allow for potential copula misspecification, Trivedi and Zimmer (2007) recommend that the standard errors be estimated in robust sandwich form under the theory of quasi-maximum likelihood (White, 1982).

As the captions to Figure 1 suggest, the dependence parameter θ may govern the degree of dependence but it is not comparable across different copulas. A common measure of dependence is Kendall's τ, which may be computed for general copula functions as[10]

$$\tau = 4 \int_0^1 \int_0^1 C(F_1, F_2) dC(F_1, F_2) - 1,$$

and for Archimedean copulas as

$$\tau = 1 + 4 \int_0^1 \frac{\varphi(t)}{\varphi'(t)} dt. \tag{6}$$

In principle, τ ranges from -1 to 1. The lower and upper bounds correspond to perfect negative and positive dependence, respectively. A copula for which τ attains both bounds is called comprehensive. When $\tau = 0$, the two random variables are independent.[11] The copula corresponding to independence is the Product copula (also referred to as the Independence copula), $C(F_1, F_2) = F_1 F_2$. The Product copula can be expressed as a special (or limiting) case of each copula, achieved with a certain value of θ that we will denote as θ_{ind}; see Table 2.

If ν and ε are independent, the parameters of the outcome equation (β) may be estimated consistently and efficiently on the self-selected subsample. Therefore, testing for independence is practically important. If θ_{ind} falls in the interior of the range of θ, independence may be tested with Wald, Lagrangian Multiplier (LM), or Likelihood Ratio (LR) tests. Under the null of independence, the test statistic is distributed as $\chi^2(1)$, provided that the copula specification is treated as a "given." However, for an arbitrary copula, the model is estimated

information criteria is equivalent to choosing the copula attaining the largest value of the log likelihood function.

[9]For example, compare the Joe and Gumbel copula models. The Vuong test statistic V is calculated as

$$V = \frac{\sqrt{N}\bar{m}}{\sqrt{N^{-1}\sum_{i=1}^{N}(m_i - \bar{m})^2}} = \frac{N\bar{m}}{\sqrt{N-1}s_m},$$

where $m_i = \ln L_i^J - \ln L_i^G$, with $\ln L_i^J$ and $\ln L_i^G$ denoting the contribution of observation i to the log likelihood of the Joe and Gumbel models, respectively, and where $\bar{m} = N^{-1}\sum_{i=1}^{N} m_i$ and s_m is the sample standard deviation of m. V has an asymptotic standard normal distribution. At a 5% significance level, the Joe copula is preferred if V exceeds 1.96; the Gumbel copula is preferred if V is less than -1.96, and the test is inconclusive if V falls between these two critical values. Each pair of copula functions may thus be compared.

[10]Another common dependence measure is Spearman's ρ. See Nelsen (2006) for the definitions of Kendall's τ and Spearman's ρ.

[11]In contrast, if the familiar Pearson's coefficient of correlation equals 0, independence is not implied.

Table 2: Dependence parameter θ and Kendall's τ

Copula Name	Range of θ	θ_{ind}	Kendall's $\tau(\theta)$	Range of τ
Product	N.A.	N.A.	0	0
Gaussian	$-1 \leq \theta \leq 1$	0	$2\sin^{-1}(\theta)/\pi$	$-1 \leq \tau \leq 1$
FGM	$-1 \leq \theta \leq 1$	0	$2\theta/9$	$-2/9 \leq \tau \leq 2/9$
Archimedean Family				
Frank	$-\infty < \theta < \infty$	0	$1 - 4\left[1 - D_1(\theta)\right]/\theta$	$-1 < \tau < 1$
Clayton	$0 \leq \theta < \infty$	0	$\theta/(\theta+2)$	$0 \leq \tau < 1$
Gumbel	$1 \leq \theta < \infty$	1	$(\theta-1)/\theta$	$0 \leq \tau < 1$
Joe	$1 \leq \theta < \infty$	1	$\cdot$	$0 \leq \tau < 1$

Notes: θ_{ind} is the value of θ if independent. For Frank, $D_1(\theta)$ is a Debye function: $D_1(\theta) = \frac{1}{\theta}\int_0^\theta \frac{t}{e^t - 1}dt$. For Joe, there is no closed form. Equation (6) is evaluated numerically.

under the quasi-maximum likelihood principle, such that LR test is no longer distributed as $\chi^2(1)$ while Wald and LM tests are still valid with sandwich-type adjustments (White, 1982).[12]

Not all copulas are comprehensive as Table 2 shows. The Gaussian and Frank copulas are comprehensive, but the range of τ for the FGM copula is only $-2/9 \leq \tau \leq 2/9$, which indicates that it can accommodate only moderate degrees of dependence. Clayton, Gumbel and Joe copulas allow only positive dependence such that $0 \leq \tau \leq 1$. This seems restrictive, but a simple modification of the underlying model evades the restriction: specify $y_i = x_i'\beta + \sigma\varepsilon_i$ as in equation (2) but let $\varepsilon_i = -\varepsilon_i^*$ and define the copula with respect to (ε^*, ν) instead. This formulation does not change any other structure of the model but does allow for negative dependence between ε and ν even with these copulas: $-1 \leq \tau \leq 0$.[13] We will refer to these sign-switched copula implementations as nClayton (i.e., negative-Clayton), nGumbel and nJoe copulas.

For these three copulas, whether in regular or negative form, independence occurs at the boundary of the range of θ. In a such case, the test for independence is one-tailed rather than two-tailed, and the test statistic is distributed as a χ^2 mixture, namely $\chi^2_m = \frac{1}{2}\chi^2(0) + \frac{1}{2}\chi^2(1)$, (e.g., Gouriéroux et al. (1982)), where $\chi^2(0)$ is a mass at 0 with a probability of one—and the same caveats as above apply when the copula is selected arbitrarily or with a pretest

[12] A search for the best copula turns the quasi-ML estimator into a pretest estimator, which may cause deviations from these distributions.

[13] Equivalently, we can specify the selection equation (1) as $s_i^* = z_i'\gamma + \nu_i$ with $\nu = -\nu^*$ and a copula for (ε, ν^*). This kind of formulation is not uncommon in the literature; see, for example, Maddala (1983), Lee (1983) or Newey (1999). In a switching regression model that has two outcome equations, modifying one of the outcome equations is certainly preferable since it keeps the relation between the selection equation and the other outcome equation intact.

rather than a priori given.

In the sample selection model, the conditional expectation of the outcome given the selection status is often of interest. For example, the expectation of y_i conditional on $s_i = 1$ is

$$\begin{aligned} E(y_i|s_i = 1) &= x_i'\beta + \sigma E(\varepsilon_i|s_i = 1) \\ &= x_i'\beta + \sigma \int_{-\infty}^{\infty} \varepsilon f_{\varepsilon|\nu}(\varepsilon|\nu_i > -z_i'\gamma)d\varepsilon, \end{aligned} \tag{7}$$

where $f_{\varepsilon|\nu}$ is a conditional density of ε given ν. When the copula is Gaussian and the marginal distribution of ε_i is normal, then an analytical expression is available. Otherwise, the integral in equation (7) may be computed by Gaussian quadrature or simulation.[14]

4 Specifying the Marginal Distributions: The Case for GTL

4.1 Marginal Distributions for the Copulas

One of the main advantages of the copula approach is that it enables us to separate the specification of the marginal distributions of ε and ν from the specification of the dependence structure. In particular, there is no need to rely on marginal (or joint) normal distributions anymore, which has long been the traditional assumption. The recent literature is slowly realizing this possibility, but the marginal distributions that have been specified are not particularly flexible. Consider first the marginal distribution of ε in the outcome equation. Smith (2005) and Dancer et al. (2008) fix their marginals as normal distributions while considering several copulas. Genius and Strazzera (2008) consider normal, logistic, and Student's t distributions and, in their application, end up preferring the latter. As is well-known, the Student's t family contains the normal distribution as a limiting case when the degrees of freedom parameter goes to ∞, and it closely approximates the logistic distribution when the degrees of freedom parameter equals 8 or 9 (Albert and Chib, 1993; Mudholkar and George, 1978). More generally, the t distribution offers flexibility in capturing thicker (but not thinner) tails than normality. However, the t distribution is symmetric, which can be a drawback. Asymmetric distributions are available: Lee (1982) considers the χ^2 distribution with several degrees of freedom, and Yen et al. (2009) work with a generalized log-Burr distribution.[15] Other asymmetric distributions that might be useful are the gamma, skewed normal and skewed t distributions. Of all these, χ^2 and gamma are right-skewed only.

As for the marginal distribution of ν in the selection equation, most researchers assume normality because of the structure of the traditional Heckman model; infrequently, some specify a logistic distribution (underlying the logit discrete choice). Asymmetric alternatives are the extreme maximum value and extreme minimum value distributions that underlie the

[14] The analytical expression is also available with Student's t copula and ε is marginally distributed as the t distribution. Heckman et al. (2003) use Student's t copula even though they do not refer it to as a copula.

[15] The pdf of a (standardized) generalized log-Burr distribution is $f_\varepsilon(\varepsilon) = e^\varepsilon \left(1 + \frac{e^\varepsilon}{\kappa}\right)^{-\kappa-1}$, where κ is a shape parameter. With $\kappa = 1$, it is a logistic distribution, and as $\kappa \to \infty$, it is an extreme value distribution.

loglog and cloglog models. The scobit model of Nagler (1994) is developed from the Burr distribution. However, the asymmetry of these models is not flexible.[16]

Thus, the literature suggests several options that the applied researcher may choose from, but this menu is still not satisfactory. It is usually not known a priori whether a marginal distribution is symmetric or asymmetric. This is especially true of the selection equation because of its latent structure and its observed binary outcome. Furthermore, it is not practical to try out several marginal distributions. For example, if only three candidate marginal distributions are considered (say, one symmetric, one left-skewed, and one right-skewed), this already yields nine combinations to estimate for each copula, which itself must be selected optimally as well. This becomes a tedious, laborious task. Therefore, we propose using a flexible distribution for each margin: a Generalized Tukey lambda (GTL) distribution. This distribution allows symmetry or asymmetry and thick or thin tails. It nests a logistic distribution but also approximately nests other familiar distributions. With GTL distributions as marginals, the only remaining task is to choose a suitable copula. Let us therefore now examine the GTL distribution.

4.2 The GTL Distribution

The Generalized Tukey lambda distribution was first proposed as a link function by Pregibon (1980) in the context of a generalized linear model and was analyzed in detail by Freimer et al. (1988).[17] The random variable ϵ from the GTL distribution is given by a quantile function $Q(u)$,

$$\epsilon = Q(u) = \mu + \sigma \left(\frac{u^{\alpha-\delta} - 1}{\alpha - \delta} - \frac{(1-u)^{\alpha+\delta} - 1}{\alpha + \delta} \right), \tag{8}$$

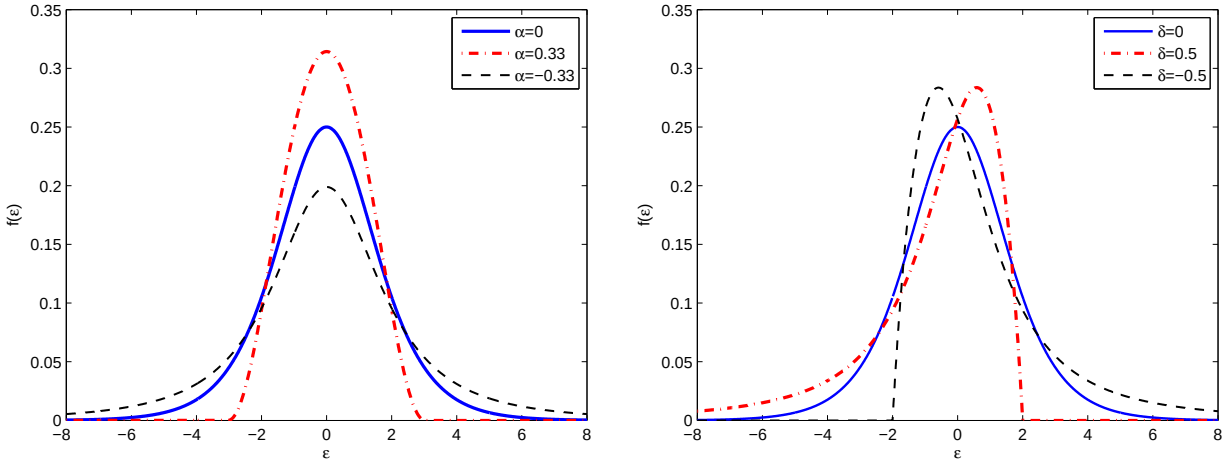
where $u \in [0, 1]$ and μ and σ are location and scale parameters, respectively, and α and δ are shape parameters. In the following discussions, we consider the canonical form such that $\mu = 0$ and $\sigma = 1$. The quantile function $Q(u)$ translates the quantile of u into a random variable ϵ. Therefore, the cdf of this distribution $F(\epsilon)$ is defined as

$$F(\epsilon) = u = Q^{-1}(\epsilon).$$

[16] Olsen (1980) assumes the uniform distribution for ν so that the selection equation is consistently estimated by a linear probability model.

[17] This distribution differs from the so-called Generalized Lambda Distribution (GLD) designed by Ramberg and Schmeiser (1974); see also Karian and Dudewicz (2011). Both the GTL and the GLD distributions are two-parameter extension of the one-parameter Tukey lambda distribution of Hastings et al. (1947) and Tukey (1960), but they differ in their formulation and in the parameter values that provide approximations to asymmetric distributions. The parameter space for GLD has gaps, which complicates MLE estimation. Moreover, while GTL and GLD generally approximate symmetric distributions in the same way, the logistic density is a special case of GTL but, due to a technicality, can only be closely approximated by the GLD.

Figure 2: GTL distributions with different α and δ



(a) δ is fixed at 0

(b) α is fixed at 0

Except for a few special values of α and δ, the function Q^{-1} does not have a closed-form expression and must therefore be evaluated numerically.[18] The pdf $f(\epsilon)$ is given by

$$f(\epsilon) = \frac{\partial F(\epsilon)}{\partial \epsilon} = \frac{\partial Q^{-1}(\epsilon)}{\partial \epsilon} = \frac{1}{\partial Q(u)/\partial u} = \frac{1}{u^{\alpha-\delta-1} + (1-u)^{\alpha+\delta-1}},$$

which is nonnegative for $u \in [0, 1]$.

The pdf of the GTL distribution exhibits a wide range of shapes for different values of α and δ. The parameter α controls thickness of tails: tails become thinner as α increases (Figure 2a). On the other hand, the parameter δ is related to the symmetry of the distribution. When $\delta < 0$ ($\delta > 0$), the distribution is right (left) skewed (Figure 2b).[19] For $\delta = 0$. the distribution is symmetric. The shape of the density does not even have to be bell-shaped as Figure 2 might suggest; for a suitable choice of α and δ, tails may end abruptly (have a positive value at the end of the range), and the density may be J-shaped or U-shaped.

The range of ϵ may be finite or infinite. The lower bound ϵ_L is finite and equals $-1/(\alpha-\delta)$ if $\alpha - \delta > 0$; otherwise, it is $-\infty$. The upper bound ϵ_U is $1/(\alpha+\delta)$ if $\alpha + \delta > 0$; otherwise, it is ∞. We define the density equal to zero if ϵ is outside of the finite range.

Although no analytical expression of the pdf exists, the moments are analytically calcu-

[18]For example, for $\alpha = 1$ and $\delta = 0$, Q is linear in u, so that ϵ has a uniform distribution. For $\alpha - \delta = \alpha + \delta = 0$, $Q(u) = \ln(u) - \ln(1-u)$ by L'Hôpital's rule, such that $F(\epsilon) = e^{\epsilon}/(1+e^{\epsilon})$ becomes the cdf of the logistic distribution. When $\alpha - \delta \to \infty$ and $\alpha + \delta = 0$, $u = 1 - e^{-\epsilon}$: this is the cdf of the exponential distribution.

[19]However, for large values of α, this link changes direction: a positive (negative) δ implies a minor degree of right (left) skewness (Freimer et al., 1988).

lated. For convenience in notation, define $\lambda_1 = \alpha - \delta$ and $\lambda_2 = \alpha + \delta$. The first and second moments are given by

$$E(\epsilon) = -1/(\lambda_1 + 1) + 1/(\lambda_2 + 1) = -2\delta/(\lambda_1 + 1)(\lambda_2 + 1)$$

$$E(\epsilon^2) = \frac{2}{(2\lambda_1 + 1)(\lambda_1 + 1)} + \frac{2}{(2\lambda_2 + 1)(\lambda_2 + 1)} - \frac{2}{\lambda_1 \lambda_2}\left(B(\lambda_1 + 1, \lambda_2 + 1) + \frac{\lambda_1\lambda_2 - 1}{(\lambda_1 + 1)(\lambda_2 + 1)}\right),$$

where $B(\cdot, \cdot)$ is the beta function. Clearly, the mean is zero if and only if $\delta = 0$. The variance, found as $Var(\epsilon) = E(\epsilon^2) - (E(\epsilon))^2$, varies with α and δ as well. Higher moments such as skewness and kurtosis may be similarly obtained, but the expressions are rather complicated.[20] For the k^{th} moment to exist, the condition that $\min(\lambda_1, \lambda_2) > -1/k$ must hold. That is, the mean exists only when $\lambda_1 > -1$ and $\lambda_2 > -1$, or equivalently, $-\alpha - 1 < \delta < \alpha + 1$; the variance (i.e., the second moment) exists when $\lambda_1 > -1/2$ and $\lambda_2 > -1/2$, or $-\alpha - 1/2 < \delta < \alpha + 1/2$; and so forth.

As shown above (footnote 18), GTL nests the logistic distribution by setting $(\alpha, \delta) = (0, 0)$ and the uniform distribution with $(\alpha, \delta) = (1, 0)$ or $(2, 0)$. It closely approximates a variety of other distributions with a suitable choice of α and δ: for example, the normal distribution with $(\alpha, \delta) = (0.1436, 0)$, a Student's $t(5)$ with $(\alpha, \delta) = (-0.0710, 0)$, and the Gumbel distribution[21] with $(\alpha, \delta) = (0.1422, -0.2290)$ (Freimer et al., 1988; Vijverberg and Vijverberg, 2012).[22]

The literature of statistical data analysis offers several methods to fit the GTL distribution to data. Ramberg et al. (1979) propose the method of moments using the first four moments. Öztürk and Dale (1985) discuss a least square estimation method, and King and MacGillivray (1999) fit data by the "starship" method, which is a computationally intensive grid-search. Su (2007) proposes an algorithm that combines a random grid search with maximum likelihood estimation. However, these studies are limited to a univariate data analysis.

Apart for data fitting exercises, the GTL distribution is not yet widely used, especially

[20] The k^{th} moment of ϵ is given by the following expression (Freimer et al., 1988; Su, 2007; Vijverberg and Vijverberg, 2012):

$$E(\epsilon^k) = \int_{e_L}^{e_U} \epsilon^k f(\epsilon) d\epsilon = \int_0^1 (Q(u))^k \, du = \int_0^1 \sum_{j=0}^{k} \binom{k}{j}^j (-1)^j \left(\frac{u^{\lambda_1} - 1}{\lambda_1}\right)^{k-j} \left(\frac{(1-u)^{\lambda_2} - 1}{\lambda_2}\right)^j du,$$

For the second equality, the fact that $\epsilon = Q(u)$ and $du/d\epsilon = f(\epsilon)$ is used, and the binomial theorem applies for the third equality. The solution of these integrals varies according to whether λ_1 and/or λ_2 equal 0; see Vijverberg and Vijverberg (2012, App.A).

[21] The Gumbel distribution should not be confused with the Gumbel copula.

[22] These comparisons may be determined by matching moments (Ramberg et al., 1979; Freimer et al., 1988). A better fit is achieved by minimizing the absolute difference in the densities (Vijverberg and Vijverberg, 2012). The absolute difference in the pdf is defined by $\int \left|\tilde{f}(\tilde{\epsilon}) - \phi(\tilde{\epsilon})\right| d\tilde{\epsilon}$, where $\tilde{\epsilon}$ is standardized by its mean and standard deviation and $\tilde{f}$ is the corresponding pdf. Alternatively, one might minimize the absolute difference between the cdfs or the largest difference between the cdfs or pdfs along the range of $\tilde{\epsilon}$ (Karian and Dudewicz, 2011; Vijverberg and Vijverberg, 2012).

outside the statistics literature. Pregibon (1980) was the first to develop the GTL formulation as a tool in a generalized linear model, which he applied to grouped data of a binary choice outcome (mortality of beetles). In the economics literature, in the context of discrete choices at the individual level, Koenker and Yoon (2009) explore maximum likelihood and Bayesian estimators. Vijverberg and Vijverberg (2012) provide a thorough examination of the discrete choice model with GTL disturbances, which they name the "pregibit" model, and discuss the link with other binary choice models such as the probit, logit, linear probability, loglog and cloglog models. The GTL distribution has also been used to estimate Lorenz curves in a study of income and wealth distributions (Sarabia, 1997) and to generate skewed random variables in Monte Carlo studies (e.g., Boero et al. (2004)). Vijverberg and Hasebe (2012) explore the GTL distribution as a way to model disturbances in a simple linear model and study the characteristics of the maximum likelihood estimators of this GTL regression model: the estimator is consistent and asymptotically normally distributed if $\max(\lambda_1, \lambda_2) < 1/2$. This property extends to the data-fitting estimator of Su (2007).

4.3 Econometric Issues

With the present paper, we are the first to utilize the flexibility of the GTL distribution in the context of sample selection models. The combination of the copula approach and the GTL marginal distributions creates a highly versatile bivariate distribution that essentially enables us to drop a priori assumptions regarding the shape of the marginal distributions. Since we retain the parametric structure of the model, we also achieve greater efficiency.[23] In the rest of this section, we discuss several estimation issues.

First of all, the GTL parameters α and δ can differ between the distributions of ν and ε because their shape may well differ. Thus, we add subscripts "ν" and "ε" to denote parameters associated with the distributions of ν and ε, respectively.

Second, following Vijverberg and Vijverberg (2012), the selection equation can be estimated in standardized form. Let μ_ν and σ_ν be the location and scale parameters (as in equation (8) of the GTL-distributed ν, which vary with the values of α_ν and δ_ν. Select μ_ν and σ_ν to be equal to the mean and standard deviation of ν, provided that they exist. Then, standardize ν with μ_ν and σ_ν: $\tilde{\nu} = (\nu - \mu_\nu)/\sigma_\nu$. This yields

$$F_\nu(-z_i{}'\gamma) = \int_{-\infty}^{-z_i{}'\gamma} f_\nu(\nu)d\nu = \int_{-\infty}^{-z_i{}'\tilde{\gamma}} f_{\tilde{\nu}}(\tilde{\nu})d\tilde{\nu} = F_{\tilde{\nu}}(-z_i{}'\tilde{\gamma}),$$

where $-z_i{}'\tilde{\gamma} = -(\mu_\nu + z_i{}'\gamma)/\sigma_\nu$ and $f_{\tilde{\nu}} = \sigma_\nu f_\nu$. Even though this standardization changes the estimated coefficients of z, the role of z in the selection mechanism is the same. Moreover, since the dependence between ν and ε is expressed only through the copula function, it does not affect the estimation of β in the outcome equation. The advantage of the standardization is to facilitate comparison of γ across different discrete choice models. But even during estimation, standardization is beneficial: it tends to speed up the optimization of the

[23] We leave formal comparisons of efficiency between our proposed estimator and semiparametric estimators as a topic for future research.

likelihood function relative to the unstandardized case (Vijverberg and Vijverberg, 2012). However, since these moments do not exist for all (α, δ) parameter values, standardization is not always feasible. In such cases, the median and interquartile range, which always exist, can be used instead of the mean and standard deviation.

Third, in regard to the outcome equation, let us ignore the selectivity issue for a moment and consider the outcome equation as a simple linear regression model: $y = x'\beta + \sigma_\varepsilon \varepsilon$. Consider the conditional expectation: $E(y|x) = x'\beta + \sigma_\varepsilon E(\varepsilon|x)$. Under the assumption that ε is independent of x, $E(\varepsilon|x) = E(\varepsilon)$. As shown above, the mean of a GTL$(\alpha_\varepsilon, \delta_\varepsilon)$-distributed ε is not zero unless $\delta_\varepsilon = 0$. Therefore, $x'\beta$ itself is not the conditional expectation of y since the intercept is shifted as a result of the non-zero mean of ε.

Furthermore, β is usually interpreted as the marginal effect of x on the conditional expectation of y; i.e., $\beta = \partial E(y|x)/\partial x$. Such an interpretation is valid, however, only when $E(\varepsilon)$ is defined. When ε is drawn from a GTL distribution with $\alpha_\varepsilon - \delta_\varepsilon \leq -1$ or $\alpha_\varepsilon + \delta_\varepsilon \leq -1$, $E(\varepsilon)$ cannot be defined, nor can $E(y|x)$, therefore. But a more general interpretation of β is still valid: as $x'\beta$ determines the location of the distribution of y conditional on x, β measures how much this location shifts as x changes.

Fourth, maximum likelihood estimation limits the parameter space of α_ε and δ_ε. To assure consistency and asymptotic normality, Vijverberg and Hasebe (2012) find that the feasible range of the shape parameters is restricted by the condition $\alpha_\varepsilon - 0.5 < \delta_\varepsilon < \alpha_\varepsilon + 0.5$ and consequently $\alpha_\varepsilon < 0.5$. As there is no lower bound on α_ε, it is not necessary to require that the moments of ε exist. As for (α_ν, δ_ν), there is no restriction, although Vijverberg and Vijverberg (2012) offer a consideration to impose a restriction $\alpha_\nu - 0.5 < \delta_\nu < \alpha_\nu + 0.5$ for reason of economic plausibility. Namely, at the left endpoints of the range of ν, the GTL density is tangent to the ν-axis only if $\alpha_\nu - \delta_\nu < 0.5$, has an angled positive slope if $\alpha_\nu - \delta_\nu = 0.5$, and is perfectly vertical if $0.5 < \alpha_\nu - \delta_\nu < 1$; and the tail is high if $\alpha_\nu - \delta_\nu \geq 1$. If $\alpha_\nu - \delta_\nu \geq 0.5$, the probability mass near the left endpoint is nonzero: extreme tail outcomes are not "rare." At the right endpoint, the tail exhibits the same shape depending on the magnitude of $\alpha_\nu + \delta_\nu$. Thus, the restriction $\alpha_\nu - 0.5 < \delta_\nu < \alpha_\nu + 0.5$ makes extreme tail outcomes rare, which is plausible from an economics perspective.

At these boundaries, the log-likelihood function is still continuous. We use mild penalty functions if the iterative parameter search either ends up, or derails, in the area beyond the bounds.

Fifth, as mentioned above briefly, we can reformulate the outcome equation in order to allow negative dependence with Clayton, Joe and Gumbel copulas: specify $y_i = x_i'\beta + \sigma_\varepsilon \varepsilon_i$, where $\varepsilon_i = -\varepsilon_i^*$ and define the copula with respect to ν and ε^*. If ε^* has a GTL distribution with parameters $(\alpha_\varepsilon^*, \delta_\varepsilon^*)$, ε has a GTL distribution with parameters $(\alpha_\varepsilon, \delta_\varepsilon) = (\alpha_\varepsilon^*, -\delta_\varepsilon^*)$. Moreover, if τ^* is the value of Kendall's measure of dependence between ε^* and ν, it can be easily shown that the dependence between ε and ν equals $\tau = -\tau^*$. Since, for comparability across models, we are interested in the distribution of (ε, ν), we will report values of τ and δ_ε rather than the values of τ^* and δ_ε^* that are actually estimated. Accordingly, the signs of the mean and skewness of ε are also switched relative to what is estimated for ε^*.

The next section examines the properties of the proposed estimator through a Monte

Carlo study. In Section 6, we apply the copula-GTL sample selection model to actual data. All the estimations in the following sections are implemented in STATA.[24]

5 Monte Carlo Simulations

It is difficult to anticipate how this new GTL-copula selection model compares with the familiar Heckman (normal-Gaussian) selection model or with a selection model that is based on a different bivariate distribution. Thus, we turn to a Monte Carlo study to shed light on the following questions: (i) how badly biased is the estimator if the assumed dependence structure is not correct, (ii) is the estimator able to detect the correct dependence structure, (iii) is the traditional assumption of normal marginals harmful if marginals are actually non-normal (and, in particular, are GTL densities), and (iv) are the parameters precisely estimated even without exclusion restrictions?

The basic structure of the data generating processes (DGPs) is as follows:

$$\begin{cases} s_i = & 1(\gamma_0 + \gamma_1 x_i + \gamma_2 z_i + \nu_i > 0) \\ y_i = & \beta_0 + \beta_1 x_i + \sigma_\varepsilon \varepsilon_i \end{cases}$$

where x_i is drawn from a standard normal distribution and z_i is from a uniform $U[0,1]$ distribution. z_i fulfils the exclusion restriction unless $\gamma_2 = 0$. The sample size is set at $N = 2,000$, and the number of replications is 500 for each of the following settings. For all the simulations, the parameters γ_1 and β_1 are fixed: $(\gamma_1, \beta_1) = (1, 1)$. The values of the other parameters vary with the research designs.

5.1 Varying the Dependence Structure

In our first research design, we draw ν_i and ε_i from each of the six copulas[25] that were described in Section 3 with standard normal marginals,[26] and we examine the consequences of estimating (i) a common normal-Gaussian selection model, (ii) a selection model with GTL marginals and the correct copula function, or (iii) an optimally selected general GTL-copula selection model.[27] Thus, for each DGP, we estimate the model with each of the copulas: both the correct copula and a series of erroneous ones that includes the Product (Independence) copula. In real-world applications, we do not know which copula is correct. For each replication, we select the best-fitting copula based on the largest value of the log likelihoods: in effect, this corresponds to selecting the best-fitting copula with the Akaike

[24] We use STATA's `ml d2` module. This program will be made available upon request.

[25] To simulate draws from copulas, we adopt a conditional sampling method. See the appendix of Trivedi and Zimmer (2007) for the procedure. For Gumbel and Joe copulas, the conditioning sampling is numerically iterated.

[26] The standard normal here is not the GTL-approximation of the standard normal.

[27] In this aspect, this study is similar to the simulation study by Winkelmann (2011), who examined a copula-based bivariate discrete choice model. To the best of our knowledge, there is no such simulation study for the continuous outcome models. Moreover, Winkelmann (2011) simulates data with the Gaussian, Frank, and Clayton; in this study, we also generate data with the FGM, Gumbel, and Joe copulas.

criterion since the number of parameters is the same across copula specifications. We call this the "Pretest" estimator.

We consider different degrees of dependence, specifically $\tau = 0.2, 0.333, 0.5$. If the DGP is jointly normal, these τ values correspond to correlation coefficients $\rho = 0.31, 0.5, 0.71$. In the case of the FGM copula, only $\tau = 0.2$ is considered since it does not allow for $\tau > 2/9$. The parameter vector is $(\gamma_0, \gamma_1, \gamma_2, \beta_0, \beta_1, \sigma_\varepsilon) = (0.5, 1, -1, 1, 1, 1)$. Given these settings, the portion of the sample with $s_i = 1$ is expected to be 50%: the outcome is observed for about one half of the sample.

Table 3 summarizes the results of the simulations. Specifically, the table reports the bias and standard deviation of the slope β_1 of the outcome equation and of an intercept β_0^* that adjusts for the nonzero $E[\varepsilon]$ such that it measures the predicted value of y for $x = \varepsilon = 0$ and thus is comparable across specifications.[28] The table also reports on the distributional parameters α_ε, δ_ε, α_ν, and δ_ν, as well as on Kendall's τ, which of course is derived from copula-specific dependence parameters θ. To save space, we report only the results from the case of $\tau = 0.333$; simulation results for $\tau = 0.2$ and $\tau = 0.5$ may be found in Appendix A.

The results show, first of all, that assuming the wrong dependence structure results in biased estimates. For example, as shown in the table, when the DGP copula is not Gaussian, the estimates obtained under a joint normality assumption are biased (first line of each panel). However, the bias is not as large as that under the assumption of independence (Product copula, second line of each panel in Table 3): assuming independence is more harmful than choosing a wrong dependence structure.

Second, as the third line of each panel of Table 3 shows, when the copula is correctly specified, the estimates are essentially unbiased. Third, it is unlikely that the true copula will be selected every time. Table 4 reports the relative frequencies of selecting each copula under different DGPs (distinguished by the DGP copula, with $\tau = 0.333$). The true copula is more likely to be selected, especially when dependence is stronger (see Appendix A)—but other copulas are occasionally erroneously preferred.

Fourth, as one might expect, there is a slight cost to not knowing the true copula: in the fourth line of each panel of Table 3, the "Pretest" copula estimator exhibits a slight bias and is less precise than the true copula estimator under each DGP. However, it still performs better than the joint normal estimator. This indicates that it is better to consider several dependence structures than to blindly assume a joint normal distribution.

Fifth, Table 4 also indicates that even when the true copula is not selected, a copula similar to the true one tends to be selected. For example, the Joe copula is the second most selected when the true copula is Gumbel, and vice versa. Moreover, the Joe copula is the second best after the Gumbel copula in terms of bias of $\hat{\beta}_1$ if the true copula is Gumbel (not reported in the table). The coefficient of correlation between the estimates of $\hat{\beta}_1$ from the Gumbel and Joe estimators is 0.92 when the true copula is Gumbel with $\tau = 0.333$, whereas the coefficient of correlation between the joint normal and Gumbel estimator is only 0.75. On the other hand, a copula with a dependence structure opposite that of the true copula is

[28] More specifically, $\beta_0^* = \beta_0 + \sigma_\varepsilon \left(\frac{1}{\alpha_\varepsilon + \delta_\varepsilon + 1} - \frac{1}{\alpha_\varepsilon - \delta_\varepsilon + 1} \right)$. Since the DGPs of Table 3 uses standard normal marginals, β_0^* in fact equals 1. For the estimator under the joint normality, $\hat{\beta}_0^* = \hat{\beta}_0$.

Table 3: Biases and standard deviations when DGPs use different copulas; $\tau = 0.333$

	β_0^*	β_1	α_ε	δ_ε	τ	α_ν	δ_ν
DGP Copula: Guassian							
Normal-Gaussian	0.006	-0.003			-0.007		
	(0.105)	(0.069)			(0.080)		
GTL-Product	0.397	-0.219	-0.003	-0.006		0.034	0.001
	(0.035)	(0.039)	(0.028)	(0.016)		(0.137)	(0.057)
GTL-Gaussian	0.006	-0.003	-0.001	-0.001	-0.007	0.033	0.002
	(0.107)	(0.070)	(0.032)	(0.019)	(0.083)	(0.134)	(0.055)
Pretest	0.014	-0.008	0.002	-0.006	-0.011	0.034	0.002
	(0.113)	(0.074)	(0.033)	(0.025)	(0.086)	(0.136)	(0.056)
DGP Copula: Frank							
Normal-Gaussian	0.038	-0.021			-0.046		
	(0.106)	(0.069)			(0.082)		
GTL-Product	0.381	-0.210	-0.033	0.016		0.034	0.001
	(0.035)	(0.038)	(0.028)	(0.016)		(0.137)	(0.057)
GTL-Frank	0.002	-0.001	0.000	0.000	-0.004	0.032	0.001
	(0.096)	(0.064)	(0.033)	(0.019)	(0.076)	(0.135)	(0.056)
Pretest	0.002	-0.003	-0.009	0.009	-0.009	0.029	-0.001
	(0.109)	(0.073)	(0.037)	(0.028)	(0.083)	(0.137)	(0.057)
DGP Copula: Clayton							
Normal-Gaussian	0.055	0.013			-0.065		
	(0.212)	(0.125)			(0.182)		
GTL-Product	0.363	-0.165	0.011	-0.037		0.032	0.001
	(0.032)	(0.036)	(0.029)	(0.017)		(0.136)	(0.057)
GTL-Clayton	0.010	-0.002	0.005	-0.004	-0.011	0.032	0.001
	(0.108)	(0.063)	(0.038)	(0.027)	(0.088)	(0.133)	(0.056)
Pretest	0.015	0.001	0.009	-0.010	-0.017	0.038	0.000
	(0.142)	(0.081)	(0.038)	(0.032)	(0.122)	(0.135)	(0.056)
DGP Copula: Gumbel							
Normal-Gaussian	0.024	-0.037			-0.020		
	(0.087)	(0.059)			(0.064)		
GTL-Product	0.410	-0.250	-0.006	0.012		0.033	0.001
	(0.038)	(0.040)	(0.030)	(0.016)		(0.137)	(0.057)
GTL-Gumbel	0.000	-0.001	-0.001	-0.001	-0.001	0.035	0.002
	(0.086)	(0.059)	(0.030)	(0.018)	(0.066)	(0.134)	(0.053)
Pretest	0.014	-0.008	-0.003	-0.002	-0.013	0.037	0.003
	(0.090)	(0.062)	(0.031)	(0.019)	(0.070)	(0.134)	(0.055)
DGP Copula: Joe							
Normal-Gaussian	0.014	-0.056			-0.008		
	(0.079)	(0.055)			(0.055)		
GTL-Product	0.427	-0.282	-0.004	0.027		0.034	0.001
	(0.040)	(0.041)	(0.031)	(0.016)		(0.137)	(0.057)
GTL-Joe	-0.004	0.001	-0.001	-0.001	0.003	0.035	0.002
	(0.067)	(0.051)	(0.029)	(0.017)	(0.048)	(0.131)	(0.052)
Pretest	-0.014	0.003	0.002	0.000	0.012	0.034	0.004
	(0.070)	(0.051)	(0.030)	(0.018)	(0.051)	(0.130)	(0.052)

Note: Bias and standard deviation (in parentheses) of the estimates are reported. For each DGP, both marginal distributions are standard normal.

rarely selected. For example, as seen in Figure 1, the Clayton copula exhibits the opposite dependence structure to Gumbel and Joe, and it is seldom chosen when the true copula is Gumbel or Joe. The opposite case is also true.

These results indicate that the true dependence structure may be captured well by re-

Table 4: Frequencies of selecting copulas under different DGPs for $\tau = 0.333$

Estimated	Copula of the DGP				
copula	Gaussian	Frank	Clayton	Gumbel	Joe
Gaussian	**0.574**	0.116	0.068	0.102	0.004
FGM	0.106	0.148	0.072	0.006	0.000
Frank	0.128	**0.594**	0.056	0.066	0.004
Clayton	0.060	0.074	**0.802**	0.002	0.000
Gumbel	0.118	0.058	0.000	**0.528**	0.170
Joe	0.014	0.010	0.002	0.296	**0.822**

Notes: Copula selection is based on the largest likelihood value. Boldface entries denote a correct selection of the copula function. Since τ exceeds 2/9, data cannot be simulated with the FGM copula. Thus, there is no column with a data generating process based on the FGM copula.

semblant copulas. This insight is important because even if the bivariate distribution differs from all of the copulas considered in this study, the estimated model may still adequately capture the true data generating process with one of the considered copulas since collectively these copulas are able to mimic diverse dependence structures.[29]

Sixth, the shape parameters of the GTL distribution for the outcome equation (α_ε and δ_ε) are estimated precisely with small biases.[30] Meanwhile, estimates of the shape parameters for the selection equation are somewhat less precise, with $\hat{\alpha}_\nu$ having larger standard deviations than $\hat{\delta}_\nu$; this result is consistent with the findings by Koenker and Yoon (2009) and Vijverberg and Vijverberg (2012). The difference in these two sets of parameters stems from the fact that the outcome y is a continuous variable that reflects tails in more detail than the discrete selection variable s can.

5.2 Varying the Marginal Distributions

The second question concerns the consequence of nonnormal marginal distributions. To see this, we employ a research design with a copula that is always Gaussian with $\tau = 0.333$, and we draw ν_i and ε_i from GTL distributions with, for simplicity, the same (α, δ). We force variations in the GTL tail properties by selecting six different combinations of α and δ: we consider thinner and thicker tails than normal, with $\alpha = 0.33$ and -0.33 respectively, and we examine the effect of asymmetry with δ varying from 0.15 (left-skewed) to 0 (symmetric) to -0.15 (right-skewed) for each value of α. When $\alpha = 0.33$, skewness equals -0.45 and 0.45 for $\delta = 0.15$ and $\delta = -0.15$, respectively, and kurtosis equals 2.34 for $\delta = 0$ and 2.72 for $\delta = 0.15$ or -0.15. When $\alpha = -0.33$, skewness equals 0 for $\delta = 0$ and is not defined for $\delta = 0.15$ or -0.15, and kurtosis does not exist for any value of δ.

[29] Of course, it is also better to have more copulas. The cost of taking more copulas into consideration is the additional time that it takes to maximize the added set of likelihood functions.

[30] The biases are evaluated by assuming that the true parameters are $\alpha_\varepsilon = 0.1436$ and $\delta_\varepsilon = 0$, representing the GTL approximation of the normal distribution.

In these settings, the disturbances are standardized such that their population mean is 0 and their population variance is 1, and the value of β_0^* is set at 1. All of this means that the values of β_0 and σ_ε change accordingly across the six DGPs. The parameter γ_2 equals -1, so that z_i satisfies the exclusion restriction. In each DGP, γ_0 is chosen such that approximately one half of the observations have observable outcomes.

Table 5 shows the simulation results from these settings. The deviations from normality affect the performance of the traditional joint normal estimator significantly. Especially, when the true distribution has thicker tails than the assumed normal distribution, the biases can be huge. The standard deviations are also so large that the estimator is not reliable: the assumption of bivariate normality makes the estimator vulnerable to outliers. However, interestingly, for the setting with $\alpha = -0.33$ and $\delta = 0.15$, the normal-Gaussian estimator seems to perform adequately—but τ is still estimated poorly. When tails are thinner, the bias of $\hat{\beta}_1$ is smaller. The bias of $\hat{\beta}_0^*$ tends to be larger when the distribution is skewed.

In contrast, our proposed estimator still performs well. As before, the shape parameters of the selection equation are somewhat imprecise, but the parameters of the outcome equation and τ are really well estimated. Under thick-tail distributions, the bias of $\hat{\beta}_1$ is essentially zero.

5.3 Addressing the Exclusion Restriction

The presence of a variable (i.e., an instrument) in the selection equation that is excluded from the outcome equation is crucial for semiparametric estimation. In parametric sample selection models, such an instrument is technically not necessary but practically highly recommended: with it, the slope of the inverse Mill's ratio in two-step estimation or the correlation coefficient in maximum likelihood estimation is more robustly identified; without it, the estimates of the entire model are sensitive to distributional misspecification (Vella, 1998; Puhani, 2000). Now, it should be noted that this recommendation is founded on evidence gathered from the standard Heckman model, one that employs, in the terminology of this paper, a Gaussian copula with normal marginals. The distributional assumption that underlies the GTL-copula proposed in this paper is much more flexible. As a result, it becomes more difficult to argue that the distribution is misspecified and to critique an application for not including an instrument. Instruments may be useful but are no longer virtually imperative. This is practically beneficial in empirical applications since, as is often the case with instrument variable estimation, it is difficult to find variables that satisfy this exclusion restriction. In this subsection, we explore several research designs to address these assertions.

These designs are built around four different specifications. The main structure of the simulated model is the same as the previous subsections. Specification 1 is the model of the previous subsections and is the benchmark: the DGP contains one variable that satisfies the restriction (z with a slope $\gamma_2 = -1$), and z is indeed included in the estimated model. Specification 2 uses the same DGP as Specification 1, but now the variable z is omitted from the estimated model. In Specifications 3 and 4, z is not included in DGP; that is, $\gamma_2 = 0$. In Specification 3, z is correctly omitted from the estimated model, whereas in Specification

Table 5: Biases and standard deviations when DGPs have nonnormal marginals

	β_0^*	β_1	α_ε	δ_ε	τ	α_ν	δ_ν
DGP:$\alpha = 0.33$ and $\delta = 0.15$							
Normal-Gaussian	0.111	-0.049			-0.094		
	(0.113)	(0.070)			(0.093)		
GTL-Product	0.308	-0.075	-0.005	0.027		0.076	0.017
	(0.032)	(0.034)	(0.034)	(0.018)		(0.184)	(0.072)
GTL-Gaussian	0.002	-0.002	0.014	0.001	-0.004	0.074	0.016
	(0.065)	(0.025)	(0.030)	(0.018)	(0.057)	(0.177)	(0.068)
Pretest	0.005	-0.005	0.019	0.000	-0.005	0.076	0.020
	(0.074)	(0.027)	(0.033)	(0.027)	(0.061)	(0.199)	(0.081)
$\alpha = 0.33$ and $\delta = 0$							
Normal-Gaussian	0.051	-0.027			-0.043		
	(0.098)	(0.062)			(0.074)		
GTL-Product	0.374	-0.173	-0.023	0.034		0.059	-0.001
	(0.034)	(0.037)	(0.028)	(0.015)		(0.152)	(0.058)
GTL-Gaussian	0.005	-0.003	0.013	-0.001	-0.005	0.059	0.000
	(0.083)	(0.047)	(0.031)	(0.019)	(0.068)	(0.149)	(0.056)
Pretest	0.021	-0.013	0.015	-0.004	-0.016	0.060	0.001
	(0.092)	(0.053)	(0.032)	(0.020)	(0.075)	(0.151)	(0.057)
$\alpha = 0.33$ and $\delta = -0.15$							
Normal-Gaussian	-0.085	0.038			0.049		
	(0.162)	(0.099)			(0.114)		
GTL-Product	0.396	-0.201	-0.044	0.054		0.072	-0.010
	(0.039)	(0.036)	(0.029)	(0.016)		(0.254)	(0.135)
GTL-Gaussian	-0.001	0.002	0.016	-0.006	-0.001	0.062	-0.015
	(0.096)	(0.058)	(0.035)	(0.022)	(0.075)	(0.173)	(0.075)
Pretest	0.022	-0.013	0.014	-0.008	-0.017	0.064	-0.013
	(0.106)	(0.064)	(0.039)	(0.024)	(0.084)	(0.173)	(0.073)
$\alpha = -0.33$ and $\delta = 0.15$							
Normal-Gaussian	0.031	-0.017			-0.153		
	(0.033)	(0.029)			(0.082)		
GTL-Product	0.070	-0.030	0.026	-0.051		0.009	0.003
	(0.015)	(0.011)	(0.040)	(0.025)		(0.105)	(0.047)
GTL-Gaussian	-0.003	0.000	0.000	0.002	0.005	0.010	0.004
	(0.025)	(0.012)	(0.043)	(0.029)	(0.056)	(0.102)	(0.045)
Pretest	-0.002	0.001	0.004	-0.001	0.008	0.010	0.003
	(0.030)	(0.012)	(0.047)	(0.033)	(0.062)	(0.103)	(0.046)
$\alpha = -0.33$ and $\delta = 0$							
Normal-Gaussian	-0.207	0.139			0.129		
	(0.205)	(0.154)			(0.153)		
GTL-Product	0.264	-0.120	0.024	-0.086		0.003	0.001
	(0.036)	(0.025)	(0.042)	(0.025)		(0.126)	(0.052)
GTL-Gaussian	-0.003	0.000	0.000	0.001	0.001	0.002	0.000
	(0.063)	(0.032)	(0.044)	(0.032)	(0.059)	(0.122)	(0.051)
Pretest	-0.004	0.003	0.003	-0.004	0.010	0.036	0.011
	(0.069)	(0.035)	(0.050)	(0.038)	(0.071)	(0.533)	(0.419)
$\alpha = -0.33$ and $\delta = -0.15$							
Normal-Gaussian	-0.234	0.182			0.239		
	(0.271)	(0.232)			(0.184)		
GTL-Product	0.086	-0.041	-0.004	-0.049		0.011	-0.002
	(0.025)	(0.011)	(0.043)	(0.026)		(0.104)	(0.047)
GTL-Gaussian	0.001	0.000	0.002	0.000	0.003	0.010	-0.002
	(0.021)	(0.013)	(0.042)	(0.026)	(0.048)	(0.104)	(0.046)
Pretest	0.001	0.000	0.003	-0.002	0.008	0.010	-0.002
	(0.022)	(0.013)	(0.043)	(0.027)	(0.057)	(0.104)	(0.046)

Note: See Table 3. For each DGP, the copula is Gaussian with $\tau = 0.333$ and marginals that both are GTL with the specified shape parameters.

4, z is included in order to satisfy the exclusion restriction even though it is irrelevant.

The econometric model of Specifications 1, 3 and 4 may be estimated consistently with the GTL-copula estimator. Under Specification 2, the estimator is biased and inconsistent because of the omission of z. However, the flexibility of GTL is expected to be helpful: z is effectively absorbed in the disturbance of the selection equation, such that the estimated GTL distribution may absorb much of the effect of the omission of z. The normal-Gaussian estimator cannot accommodate such an adjustment (unless z is normally distributed). Part of this advantage of the GTL-copula estimator comes from an artifact of our specifications: z is uncorrelated with x. Correlation would impart additional bias to both the GTL-copula and the normal-Gaussian estimator because x is no longer exogenous in the selection equation. However, we do not consider such complications and instead focus on the purest implementation of the exclusion restriction, a purely uncorrelated z.

The disturbances in these DGPs are generated with all of the GTL-copula distributions considered in the previous subsections, but we fit the model only with a true copula in order to save time. For each DGP, the parameter γ_0 is selected such that about a half of observations have observable outcomes.

Tables 6 and 7 show the results of the simulations. To save space, the tables report only the bias and standard deviation of $\hat{\beta}_1$, the slope of x in the outcome equation, which is of main interest in most applications. Table 6 repeats the distributional assumptions of Table 3; thus, the column for Specification 1 repeats the results for $\hat{\beta}_1$ in first and third lines of each panel in Table 3. Table 7 is similarly linked with Table 5.

In the first panel of Table 6 where the DGP uses a joint normal distribution to generate disturbances, the normal-Gaussian estimator yields slightly smaller biases than the GTL-copula estimator: the structure imposed by (correctly assumed) joint normality limits the bias in $\hat{\beta}_1$, or, stated otherwise, the unneeded flexibility of the GTL-copula estimator makes $\hat{\beta}_1$ a little wilder. The absolute difference between β_1 and the median of the sampling distribution of $\hat{\beta}_1$ is actually less than 0.005 for both estimators and always smaller for the GTL-copula estimator. In the other panels of Table 6, the bias of the normal-Gaussian estimator mainly reflects the violation of the distributional assumption; the omitted variable bias that is added in Specification 2 is only minor. The rise in bias in Specifications 3 and 4 as compared with Specification 1 illustrates the conclusion in the literature that the presence of an instrument in the selection equation benefits the normal-Gaussian estimator.

By comparison, the GTL-copula estimator evidences very little bias, even when the instrument is erroneously omitted (Specification 2) or when the selection and outcome equations depend on the same explanatory variables (Specification 3). An instrument does improve the estimator (Specification 1) but is not mandatory. Moreover, its standard deviation tends to be lower than the normal-Gaussian estimator.

Table 7 reinforces all of these conclusions. Moreover, as the GTL distribution changes from left-skewed to right-skewed, the bias of the normal-Gaussian estimator becomes more positive, whether tails are thin (in first three panels) or thick (in the last three panels). All the while, the GTL-copula estimator is virtually unbiased.

In sum, the results of the Monte Carlo study show that the assumption of the joint

Table 6: Bias and standard deviation of $\hat{\beta}_1$ with and without an instrument when DGPs use different copulas and normal marginals

z in DGP/in model:	Specification (1) Yes / Yes	Specification (2) Yes / No	Specification (3) No / No	Specification (4) No / Yes
DGP Copula: Gaussian				
Normal-Gaussian	-0.0033	-0.0213	-0.0108	-0.0132
	(0.0687)	(0.1104)	(0.0933)	(0.0989)
GTL-Gaussian	-0.0035	-0.0228	-0.0193	-0.0218
	(0.0700)	(0.1165)	(0.1150)	(0.1193)
DGP Copula: Frank				
Normal-Gaussian	-0.0207	-0.0655	-0.0645	-0.0635
	(0.0691)	(0.1507)	(0.1478)	(0.1470)
GTL-Frank	-0.0010	-0.0144	-0.0079	-0.0085
	(0.0636)	(0.1026)	(0.0872)	(0.0885)
DGP Copula: Clayton				
Normal-Gaussian	0.0129	-0.0721	-0.0862	-0.0865
	(0.1253)	(0.2198)	(0.2195)	(0.2192)
GTL-Clayton	-0.0024	-0.0045	-0.0047	-0.0048
	(0.0629)	(0.0748)	(0.0745)	(0.0750)
DGP Copula: Gumbel				
Normal-Gaussian	-0.0374	-0.0566	-0.0537	-0.0537
	(0.0592)	(0.0672)	(0.0647)	(0.0652)
GTL-Gumbel	-0.0011	-0.0093	-0.0028	-0.0028
	(0.0590)	(0.0762)	(0.0679)	(0.0682)
DGP Copula: Joe				
Normal-Gaussian	-0.0560	-0.0384	-0.0379	-0.0379
	(0.0553)	(0.0651)	(0.0643)	(0.0650)
GTL-Joe	0.0013	-0.0063	-0.0034	-0.0035
	(0.0507)	(0.0675)	(0.0641)	(0.0644)

Note: In each cell, the parenthesized value represents the standard deviation of $\hat{\beta}_1$.

normality is restrictive and results in a biased and imprecise normal-Gaussian estimator if the assumption is violated, especially if the selection equation does not contain an instrument. Our proposed GTL-copula estimator performs very well, even if instruments are absent from the selection equation. The GTL-copula approach requires a choice from a menu of copula functions because the researcher is typically ignorant of the true dependence structure. This choice results in a less precise "pretest" estimator, but our approach is better than assuming joint normality all the time. We are leaving a comparison of our proposed estimator with semiparametric and nonparametric estimators for future research, but we do expect that for plausible joint distributions our estimator is both statistically and computationally efficient.

Table 7: Bias and standard deviation of $\hat{\beta}_1$ with and without an instrument when DGPs use nonnormal marginals and a Gaussian copula

z in DGP/in model:	Specification (1) Yes / Yes	Specification (2) Yes / No	Specification (3) No / No	Specification (4) No / Yes
DGP: $\alpha = 0.33$ and $\delta = 0.15$				
Normal-Gaussian	-0.0494	-0.1255	-0.1008	-0.1023
	(0.0703)	(0.2113)	(0.1739)	(0.1769)
GTL-Gaussian	-0.0017	-0.0033	-0.0020	-0.0012
	(0.0248)	(0.0381)	(0.0251)	(0.0322)
DGP: $\alpha = 0.33$ and $\delta = 0$				
Normal-Gaussian	-0.0272	-0.0709	-0.0613	-0.0655
	(0.0620)	(0.1347)	(0.1210)	(0.1293)
GTL-Gaussian	-0.0034	-0.0061	-0.0050	-0.0052
	(0.0473)	(0.0541)	(0.0528)	(0.0528)
DGP: $\alpha = 0.33$ and $\delta = -0.15$				
Normal-Gaussian	0.0375	0.0508	0.0476	0.0471
	(0.0991)	(0.1576)	(0.1462)	(0.1473)
GTL-Gaussian	0.0018	0.0021	0.0026	0.0018
	(0.0579)	(0.0683)	(0.0687)	(0.0688)
DGP: $\alpha = -0.33$ and $\delta = 0.15$				
Normal-Gaussian	-0.0172	-0.0244	-0.0211	-0.0211
	(0.0292)	(0.0571)	(0.0331)	(0.0331)
GTL-Gaussian	0.0001	0.0011	0.0004	0.0004
	(0.0116)	(0.0131)	(0.0125)	(0.0125)
DGP: $\alpha = -0.33$ and $\delta = 0$				
Normal-Gaussian	0.1395	0.1862	0.1746	0.1744
	(0.1540)	(0.1812)	(0.1739)	(0.1740)
GTL-Gaussian	0.0002	0.0024	0.0019	0.0019
	(0.0323)	(0.0351)	(0.0339)	(0.0341)
GDP: $\alpha = -0.33$ and $\delta = -0.15$				
Normal-Gaussian	0.1819	0.3085	0.2424	0.2423
	(0.2323)	(0.2341)	(0.2668)	(0.2668)
GTL-Gaussian	0.0000	0.0043	0.0004	0.0004
	(0.0125)	(0.0143)	(0.0131)	(0.0131)

Note: In each cell, the parenthesized value represents the standard deviation of $\hat{\beta}_1$.

6 Applications

Greater flexibility is desirable only if it has practical relevance. Thus, we turn to five applications of a varied nature that all yield substantial changes in the estimation results: wages of married women in Portugal (subject to labor force participation), wages of school-aged children in Mexico (subject to not attending school and actually working), health expenditures in the US (subject to having nonzero health expenditures), fines for speeding in Massachusetts (subject to being given a ticket when speeding), and the intensity of international conflicts (subject to a conflict existing between a pair of countries).

Table 8: GTL-copula model selection: log-likelihood values and Vuong tests

Estimation method	Applications				
	Wages of married women	Wages of school-aged children	Health expenditures	Speeding tickets	International disputes
Normal-Gaussian [a]	-2487.94	-4927.80	-10170.11	-45173.91	-9209.59
Copula [b]					
Product	-2317.81	-4473.11	-10135.11	-10402.06	-9095.19
Gaussian	-2307.33	-4468.23	*-10132.26*	-10397.16	-9092.53
FGM	-2312.97	-4471.31	-10135.05	-10305.82	-9075.48
Frank	-2307.70	-4470.39	-10135.11	-10382.39	*-9071.47*
Clayton	-2311.63	-4473.11	**-10132.24**	... [c]	**-9063.70**
Gumbel	*-2300.68*	-4473.11	-10134.55	... [c]	-9095.19
Joe	**-2295.18**	-4473.11	-10134.71	... [c]	-9095.19
nClayton	-2317.81	-4471.52	-10135.11	-5630.22	-9095.19
nGumbel	-2317.81	*-4461.85*	-10135.11	*-4320.31*	-9095.19
nJoe	-2317.81	**-4461.63**	-10136.83	**-4098.00**	-9095.19
Vuong tests [d]					
Second best	2.50	0.44	0.01	13.30	3.10
GTL-Gaussian	3.26	1.57	0.01	24.12	8.60
Normal-Gaussian	10.29	11.73	4.30	76.04	4.08
Number of observations					
Selection	2,339	15,526	5,574	68,357	149,004
Outcome	1,400	1,657	4,281	31.674	972
Share with outcome (%)	59.9	10.7	76.8	46.3	0.7

Boldface and italized entries denote the best and second-best model specification, respectively.

[a] The normal-Gaussian estimator corresponds to the traditional Heckman (1974) maximum likelihood estimator that assumes joint normality.

[b] All copula models use different GTL marginals for the selection and outcome equations.

[c] The maximum likelihood estimation routine did not converge.

[d] Vuong tests are relative to the best model. Vuong test statistics are distributed standard normal; critical values are 1.645 (10 percent), 1.96 (5 percent), 2.58 (1 percent).

Table 8 gives an overview of these five applications. The sample size varies greatly, as does the share of the observations for which an outcome is observed. This variation builds a picture of how the GTL-copula estimator performs with real-life data. The GTL-copula estimator dominates the common Heckman (normal-Gaussian) estimator that presumes joint normality, sometimes by a wide margin. Different copulas are preferred in different applications. Thus, experimentation with a variety of copulas is recommended. In fact, as already highlighted in Figure 1, the shapes of the copulas vary so much that a few specifications fail to converge, particularly when the copula limits the range of dependence (Clayton, Gumbel, Joe). We interpret this to mean that the data simply do not conform to the structure that such a copula imposes: usually, when the copula's "sign" is switched, the iterative search proceeds more smoothly.

The table shows three Vuong tests, each of them comparing with the best model. Here,

we find that the rejection of the joint-normal selection model happens sometimes because neither the marginals nor the Gaussian correlation structure conform to a normal distribution (columns 1, 4 and 5). Other times, the Gaussian copula is adequate but the marginals are nonnormal (columns 2 and 3). Note also that a higher log-likelihood value does not always translate into a larger Vuong test value: in column 5, the individual contributions to the log-likelihood value vary much more between the GTL-Clayton and normal-Gaussian models than between GTL-Clayton and the GTL-Gaussian or GTL-Frank models. Such variation lowers the Vuong test value and reduces its power.

For all applications, detailed definitions and descriptive statistics of all variables are provided in Appendix B. For better readability, we have adjusted the acronyms of the variables from what the authors used in their papers.

6.1 Wages of Married Women, Portugal

The first application is a study of wages among married women in Portugal (Martins, 2001).[31] This topic is a classical example of the sample selection model (Heckman, 1974): market wages are not observed for women who do not work. Estimation of the wage equation only with the subsample of working women results in selectivity bias.

Martins (2001) estimates the model by both semiparametric and maximum likelihood (normal-Gaussian) methods; we compare our GTL-copula estimator with the normal-Gaussian estimator. Genius and Strazzera (2008) use the same data and estimate the model by a copula-based maximum likelihood method that assumes a logistic distribution for the selection equation and a t distribution for the wage (outcome) equation.

We use the same specification as Martins (2001, Table 1). The covariates for the selection equation are the number of children younger than 18 and younger than 3 living in the family, years of schooling, age and age squared, and the log of the husband's monthly wage. The explanatory variables in the wage equation are years of schooling, Mincerian potential experience (PEXP, defined as age minus years of schooling minus 6, divided by 10) in linear and squared form (PEXP2), and interactions of the potential experience variables with the number of children younger than 18. Out of 2,339 observations, 1,400 married women (about 60%) participate in the labor market reported their wages.

Among all specifications, the Joe copula attains the largest log likelihood value (Table 8). The selection of the Joe copula is consistent with Genius and Strazzera (2008) and further supported with the Vuong test: the Joe copula is preferred over the Gumbel copula, the second-largest log likelihood value, with a test statistic of 2.50.

Table 9 summarizes the results of the normal-Gaussian and Joe-GTL estimators. In order to make the selection equation comparable across estimators, the equation is standardized for each copula; we standardize by the median and the interquartile range since with $(\hat{\alpha}_\nu, \hat{\delta}_\nu) = (-0.524, -0.080)$, the standard deviation of the disturbance ν of the selection equation is not defined.[32] The estimated GTL distribution of ν has thicker tails than the standard normal, as is shown by the dashed curves in Figure 3a and 3b. δ_ν is close to 0 and statistically not

[31] The data are available online at `http://qed.econ.queensu.ca/jae/2001-v16.1/martins/`.

[32] For the standard normal distribution, the median is 0 and the interquartile range is 1.349.

Table 9: Estimation results: Log-wages of married women

	Normal-Gaussian		GTL-Joe	
Variables	Coeff.	(S.E.) [a]	Coeff.	(S.E.) [a]
Selection equation: Labor force participation [b]				
CHILDY18	-0.088	(0.022)	-0.081	(0.021)
CHILDY3	-0.061	(0.053)	-0.041	(0.049)
lnHUSBW	-0.076	(0.059)	-0.123	(0.059)
YRSCH	0.111	(0.008)	0.146	(0.019)
AGE	0.602	(0.195)	0.649	(0.229)
AGE2	-0.092	(0.025)	-0.094	(0.030)
constant	-0.440	(0.728)	-0.294	(0.700)
α_ν			-0.524	(0.363)
δ_ν			-0.080	(0.176)
Outcome equation: Log of hourly wages				
YRSCH	0.114	(0.004)	0.132	(0.003)
PEXP	0.134	(0.080)	0.326	(0.057)
PEXP2	-0.002	(0.017)	-0.043	(0.011)
PEXP × CHILDY18	0.035	(0.019)	0.008	(0.011)
PEXP2 × CHILDY18	-0.012	(0.006)	-0.005	(0.003)
constant	4.472	(0.100)	4.191	(0.073)
α_ε			-0.278	(0.033)
δ_ε			0.174	(0.021)
σ	0.555	(0.018)	0.200	(0.013)
θ	0.346	(0.063)	2.892	(0.336)
τ	0.225		0.505	
$\ln L$	-2487.94		-2295.18	
AIC	5005.88		4628.37	

[a] White-robust standard errors in parentheses.

[b] The selection equation is standardized by the median and interquartile range.

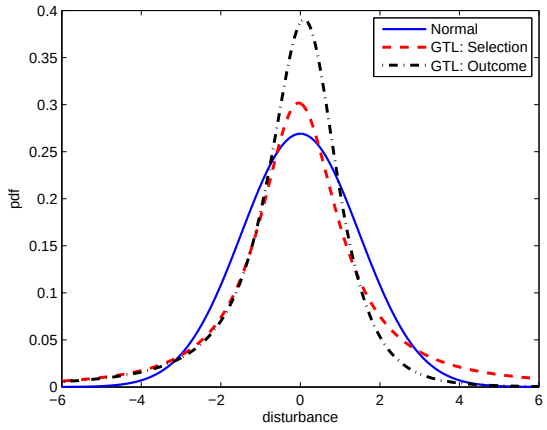
significantly different from it, implying that the distribution is essentially symmetric. α_ν is estimated imprecisely as well. In fact, we fail to reject the null hypotheses of $(\alpha_\nu, \delta_\nu) = (0, 0)$ and even $(\alpha_\nu, \delta_\nu) = (0.1436, 0)$ at the 10% level of significance:[33] our GTL estimates are not inconsistent with ν being distributed logistically (Genius and Strazzera, 2008) or even normally (Martins, 2001).

The estimated coefficients of the selection equation are mostly similar between the two estimators except the husband's wage: the GTL-Joe estimate is statistically significant and is much larger than the normal-Gaussian estimate, which is insignificant.

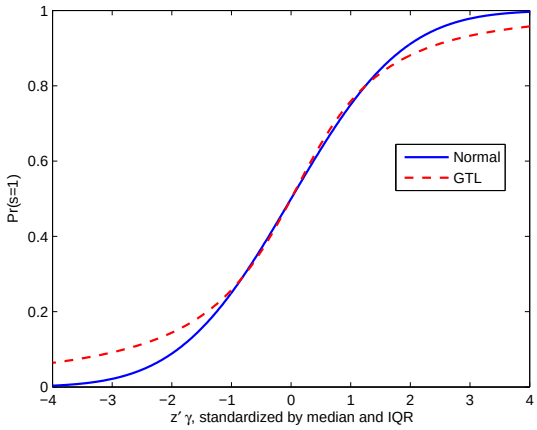
On the other hand, the shape parameters of the distribution of ε of the wage equation are statistically significant. The estimated value of α_ε imply that the distribution has thick tails. This finding is consistent with Genius and Strazzera (2008), who estimate the degree of

[33] Wald test statistics are 2.53 and 4.41, respectively, with a 10% $\chi^2(2)$ critical value of 4.61.

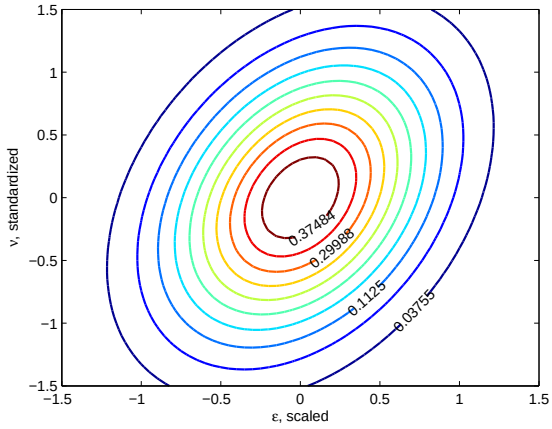
Figure 3: Differences between normal-Gaussian and GTL-Joe: Wages of married women



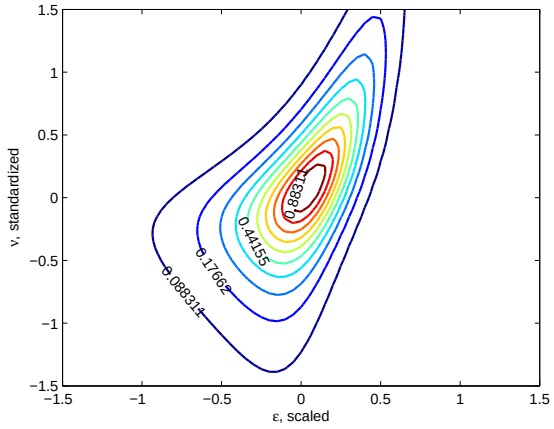
(a) PDF comparison



(b) $Pr(s_i = 1)$: GTL versus probit

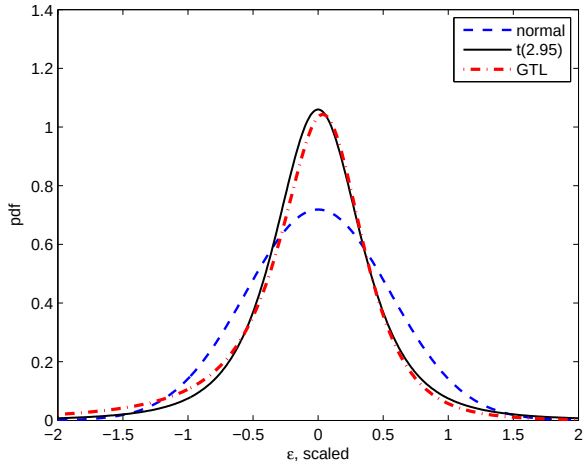


(c) Contour plot: normal-Gaussian



(d) Contour plot: GTL-Joe

Figure 4: Density of scaled disturbances of the wage equation



freedom parameter of their assumed t distribution to be 2.95 and thus find thick tails as well.[34] At the same time, with $\hat{\delta}_\varepsilon = 0.174$, the estimated GTL distribution of ε is not only heavy-tailed but also left-skewed.[35] In principle, this skewness could matter for the estimated wage equation (e.g., see footnote 37 below). To illustrate, Figure 4 plots three estimated densities of ε: normal, t with 2.95 degrees of freedom, and GTL with $(\hat{\alpha}_\varepsilon, \hat{\delta}_\varepsilon) = (-0.278, 0.174)$. Each density is scaled by the estimated scale parameter $\hat{\sigma}_\varepsilon$. Obviously, the normal density appears unsuitable for these data. The t and GTL densities are much closer to each other, but the GTL density is left-skewed. Thus, under GTL, low-wage outliers are more common than high-wage outliers and are therefore not allowed to influence the location of the regression line as much as under the t or normal distributional assumption.

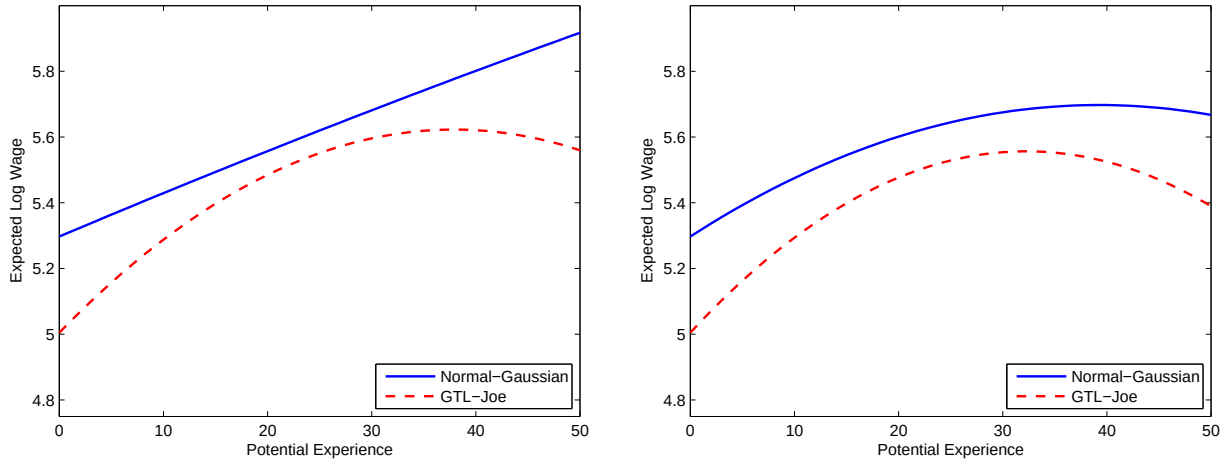
The difference in the underlying distributional assumption is reflected in the estimated coefficients of the wage equation. The rate of return to schooling rises by 1.8 percentage points, and the estimated wage profiles are altered. Specifically, the GTL-Joe slopes of PEXP and PEXP2 are larger in absolute values than the normal-Gaussian slopes and they become statistically significant; at the same time, the slopes of PEXP and PEXP2 interacted with the number of children less than 18 are smaller in magnitude and are no longer statistically significant. Figure 5 plots expected log-wages as a function of potential experience, evaluated at the average years of schooling, 7.24.[36] Figure 5a shows that for a woman without a child

[34] We also replicate the result by Genius and Strazzera (2008) but suppress it to save space. If we estimate a restricted model with $\delta_\varepsilon = 0$, the estimated α_ε is -0.227. Indeed, a GTL distribution with $(\alpha, \delta) = (-0.227, 0)$ approximates the t distribution with 2.95 degrees of freedom very closely.

[35] Even though we cannot define skewness (and kurtosis) with these estimates, the positive value of $\hat{\delta}_\varepsilon$ indicates the distribution is left-skewed. One measure of asymmetry is $S = (Q_{75} - Q_{50})/(Q_{50} - Q_{25})$, where Q_k is the k-th quantile of the distribution: $S < (>)1$ denotes a left (right) skew. For our estimated $(\alpha_\varepsilon, \delta_\varepsilon)$, this equals 0.83.

[36] For the GTL-Joe estimate, we must account for the fact that $\sigma_\varepsilon E(\varepsilon) = -0.1417 \neq 0$, given the estimates of α_ε and δ_ε; see equation (2).

Figure 5: Log-wage profiles, married women in Portugal



(a) Log-wage profile for CHILD18 = 0

(b) Log-wage profile for CHILD18 = 2

in the household the normal-Gaussian estimate of the wage profile is straight, while the GTL-Joe wage profile exhibits curvature. Having two children adds curvature to the wage profiles (Figure 5b), but only slightly so for the GTL-Joe estimates.[37]

Finally, the estimators imply a different degree of dependence between ε and ν. As estimated by the GTL-Joe model, τ is more than twice as large as that of the normal-Gaussian model. Contour plots of the joint density in Figures 3c and 3d are strikingly different. The Joe copula exhibits strong right tail dependence but weak left tail dependence. It indicates that those who are more likely to participate in the labor market tend to earn higher wages, conditional upon observable variables, but wages show more variation for those who are less likely to participate. As a consequence, the estimated GTL-copula model imposes a greater selectivity correction on the wage equation. This is well expressed in the normal-Gaussian two-step estimation method by the familiar inverse Mill's ratio that is added to the wage equation for labor market participants (Heckman, 1979). Accordingly, in Figure 5, the expected (unconditional) GTL-Joe wage profile lies substantially below the normal-Gaussian profile.

6.2 Wages of School-Aged Children, Mexico

The sample selection model is widely used. Its estimates not only inform on the role of explanatory variables, but also permit prediction of the outcome variable among those for whom outcomes are not observed. Attanasio et al. (2012) use such predicted outcomes in

[37] The wage profile estimated by Genius and Strazzera (2008) is steeper and more sharply curved for women without children (the slope estimates of PEXP and PEXP2 are 0.379 and −0.055, respectively), and the profile is unchanged when the number of children increases.

a structural model that evaluates the effect of a social experiment, PROGRESA, on school attendance in Mexico. Specifically, the wage a child could earn when (s)he does not attend school plays a role in the educational choice made by or for the child. To predict each child's (actually, boy's) potential wage, the authors estimate the wage equation with an inverse Mill's ratio that corrects for sample selectivity. We compare the joint-normality (i.e., normal-Gaussian) maximum likelihood version of this selection model with our GTL-copula estimator.[38]

The selection mechanism is actually of a dual nature: the out-of-school wage is observed only if the child does not attend school *and* is at work earning a wage. However, we will treat this as a single indicator. Our formulation of the selection and outcome equations follows the specification of Attanasio et al. (2012). The selection equation has a long list of explanatory variables that describe the child, his parents, the community's schools, and the PROGRESA program; see the table below. The explanatory variables in the outcome equation are the community-level male agricultural wage (in log form), age, years of schooling, and a dummy denoting residence in a community with a PROGRESA grant program. PROGRESA may have an indirect, general-equilibrium impact on children's wages. The sample contains 15,526 children, of whom 1,657 (about 11%) are not in school and work for a wage.[39]

Estimates are reported in Table 10. The best copula is nJoe, based on the log likelihood values (Table 8). However, a Vuong test of GTL-nJoe against GTL-nGumbel, which attains the second largest log likelihood value, is inconclusive. Even in comparison with GTL-Gaussian, the Vuong test is not significantly in favor of GTL-nJoe. Therefore, we also report the results of the Gaussian copula for comparison.

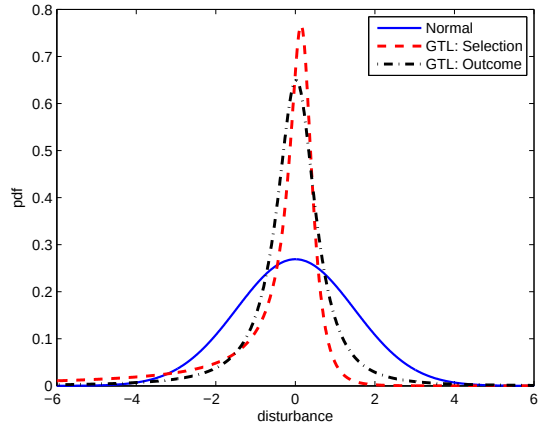
The disturbance of the selection equation has thick tails and is left-skewed. We can easily reject the hypothesis of logistic ($(\alpha_\nu, \delta_\nu) = (0, 0)$) or normal ($(\alpha_\nu, \delta_\nu) = (0.1436, 0)$) disturbances. The outcome equation is right-skewed, however, with thick tails as well. Figure 6a demonstrates the large difference of these GTL densities with the normal density (standardized by the interquartile range). The cdf that yields selection probabilities (Figure 6b) is therefore quite different as well. According to Table 10, the normal-Gaussian model finds weak positive dependence that is statistically insignificantly different from independence. The negative Joe and Gaussian copula models find weak negative dependence that is statistically still significant. Summarizing these various elements, the contour plots in Figure 6 bear out the fact that the GTL-nJoe distribution differs greatly from the joint normality that underlies the Heckman model: the contours no longer have the familiar elliptical shape.

Interestingly, the GTL-nJoe and GTL-Gaussian contour plots are visually very similar (Figures 6d and 6e). This is due to weak dependence that the estimated τ indicates. With weak dependence, the difference in the Joe and Gaussian copula functions is not visible. The

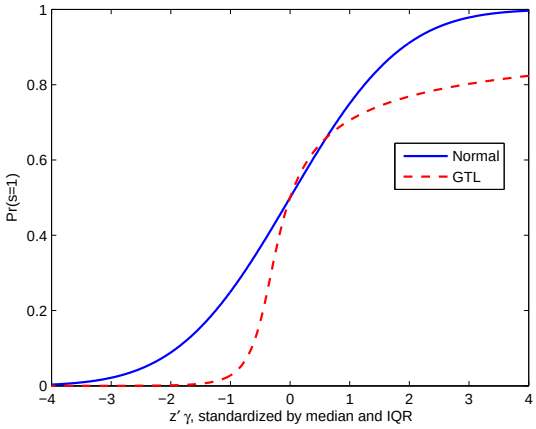
[38]The normal-Gaussian estimator produces almost identical estimates and predictions as the two-step procedure that Attanasio et al. (2012) use.

[39]The data are available online at `http://restud.oxfordjournals.org/content/79/1.toc`. We were able to replicate equation (9) of Attanasio et al. (2012). However, their selection indicator was whether the child is in school, and the logwage equation was estimated over all children who had a wage, whether they attended school or not. From the discussion in the paper, this apparently was not the intent of the analysis of this wage information.

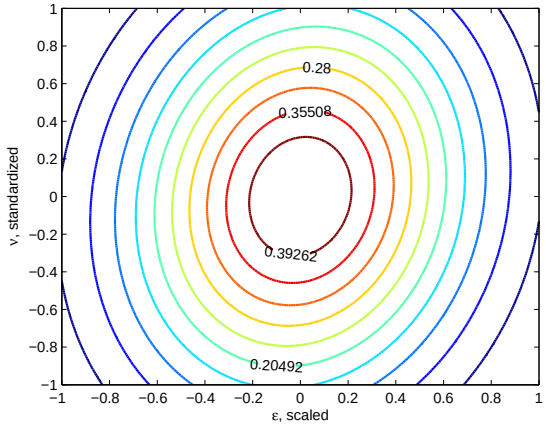
Figure 6: Differences between normal-Gaussian and GTL-copula: Wages of school-aged children



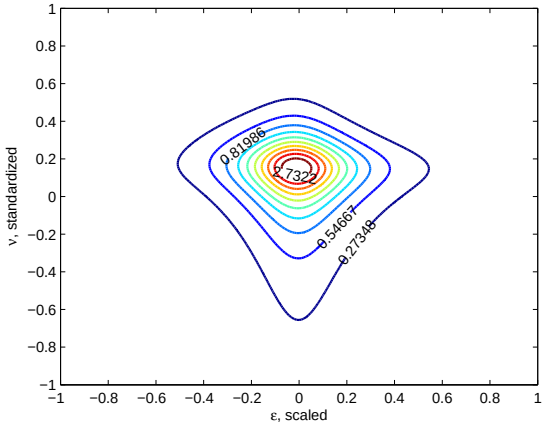
(a) PDF comparison



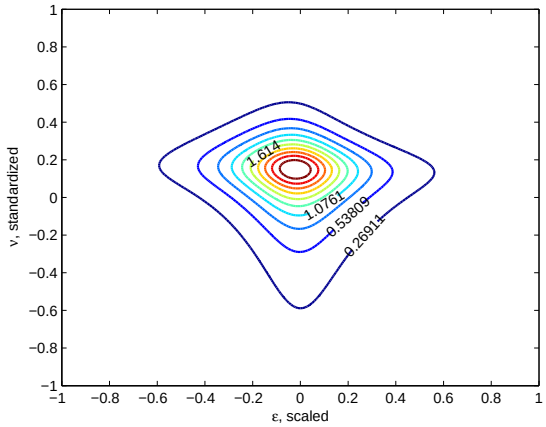
(b) $Pr(s_i = 1)$: GTL versus probit



(c) Contour plot: Normal-Gaussian



(d) Contour plot: GTL-nJoe



(e) Contour plot: GTL-Gaussian

Table 10: Estimation results: Wages of school-aged children

Variables	Joint Normal Coeff.	(S.E.) [a]	Negative Joe - GTL Coeff.	(S.E.) [a]	Gaussian - GTL Coeff.	(S.E.) [a]
The Selection Equation [b]						
PROGRESA/eligible	-0.016	(0.052)	-0.010	(0.019)	-0.010	(0.018)
PROGRESA/ineligible	-0.145	(0.053)	-0.045	(0.037)	-0.044	(0.037)
Control/ineligible	-0.123	(0.055)	-0.037	(0.035)	-0.033	(0.032)
ORDER	0.015	(0.010)	0.006	(0.005)	0.006	(0.005)
AGE	0.348	(0.129)	0.407	(0.409)	0.360	(0.375)
AGE2	-0.002	(0.004)	-0.010	(0.011)	-0.009	(0.010)
GRANT	-0.027	(0.024)	-0.006	(0.009)	-0.005	(0.009)
FATHERhome	0.066	(0.062)	0.018	(0.025)	0.018	(0.024)
MOTHERhome	-0.164	(0.063)	-0.062	(0.049)	-0.060	(0.049)
INDIGENOUS	-0.061	(0.041)	-0.024	(0.019)	-0.025	(0.020)
DISTANCE97	0.034	(0.035)	0.008	(0.016)	0.009	(0.016)
DISTANCE98	-0.017	(0.037)	-0.002	(0.015)	-0.003	(0.015)
PRIMARY97	0.135	(0.197)	0.020	(0.079)	0.035	(0.080)
PRIMARY98	-0.150	(0.138)	-0.036	(0.059)	-0.047	(0.062)
SECONDARY97	-0.280	(0.128)	-0.050	(0.064)	-0.039	(0.059)
SECONDARY98	0.226	(0.123)	0.028	(0.056)	0.019	(0.052)
α_s			-1.268	(0.480)	-1.299	(0.493)
δ_s			1.087	(0.460)	1.131	(0.474)
The Wage Equation						
lnMAWAGE	0.862	(0.045)	0.881	(0.032)	0.881	(0.032)
AGE	0.062	(0.036)	0.006	(0.006)	0.005	(0.007)
YRSCH	0.014	(0.006)	0.005	(0.003)	0.005	(0.003)
PROGRESA	0.062	(0.028)	0.030	(0.015)	0.033	(0.015)
constant	-1.039	(0.699)	-0.002	(0.100)	0.022	(0.119)
α_1			-0.430	(0.030)	-0.514	(0.045)
δ_1			-0.080	(0.027)	-0.118	(0.055)
σ	0.502	(0.022)	0.097	(0.006)	0.101	(0.007)
ρ or θ	0.107	(0.213)	1.105	(0.033)	-0.200	(0.073)
τ	0.068		-0.057		-0.128	
$\ln L$	-4927.80		-4461.63		-4468.23	
AIC	9935.60		9011.27		9024.46	

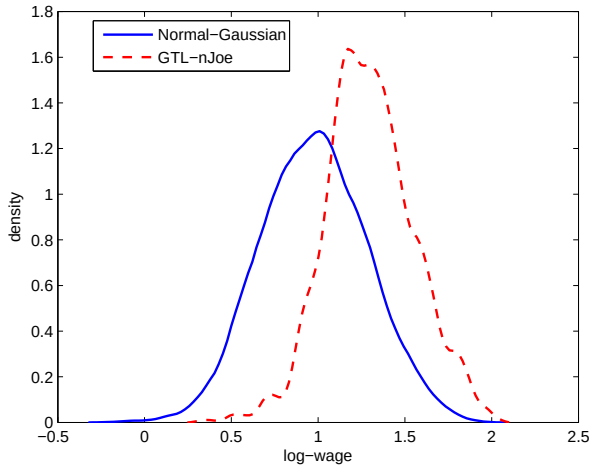
[a] Standard errors are clustered at a community level.

[b] The selection equation is standardized by the median and interquartile range. The selection equation also contains a long list of dummy variables for father's education, mother's education, and state of residence, as well as an intercept.

coefficients in the outcome equation are only different by the third decimal place (except constant terms). Therefore, it is plausible that we are not able to discriminate the two models statistically.

According to the normal-Gaussian estimates, child wages are mostly related to the com-

Figure 7: Kernel density plots of predicted log-wages



munity's male agricultural wage and for the rest vary with age, education and PROGRESA status. The GTL-nJoe model refutes these age and education effects, and the PROGRESA effect is reduced by half. Child wages move in tandem with male wages.[40]

As mentioned above, the purpose of the sample selection estimation in Attanasio et al. (2012) is to predict a child's potential wage out of school, and this predicted wage is used in subsequent estimation to evaluate the impact of PROGRESA on school participation. The differences in the parameter estimates result in differences in the predicted values. As can be seen from the kernel densities in Figure 7, the predicted log-wages are very different: the GTL-nJoe predictions are on average 0.321 higher than the normal-Gaussian predictions. This is not merely a proportional shift: since the individual differences have a standard deviation of 0.158, the predicted wages also vary relative to each other. It is conceivable that these wage differences lead to different results in the subsequent analysis, but re-estimation of the structural model developed by Attanasio et al. (2012) is beyond the scope of this paper.

6.3 Health Expenditures, USA

Our third example concerns health expenditures, using data from Deb and Trivedi (2002) that were originally drawn from the RAND Health Insurance Experiment.[41] Out of the sample of 5,574 observations, 76.8% report positive expenditures. In this application, the selection is whether a person reported nonzero medical expenditures; the outcome variable

[40]The slope estimates of the selection equation are also generally smaller, but these slopes are not the only thing that determines the magnitude of the impact on the discrete outcome. We found any given marginal effect to be roughly similar when the estimated GTL-nJoe slope is about 40 percent of the normal-Gaussian slope. This has to do with the fact that the mode of the estimated GTL density function is much taller (Figure 6a).

[41]The data are available online at `http://cameron.econ.ucdavis.edu/mmabook/mmaprograms.html`.

is that person's annual individual medical expenditures in logarithmic form.

The explanatory variables include health insurance variables (coinsurance rate, a dummy "IDP" denoting a deductible plan, an annual participation incentive payment "API", and maximum medical deductible expenditures "MMDE"), health status variables (number of chronic diseases and dummies for physical limitation and health status), family income, and demographic characteristics such as age, gender, race, and household composition.[42] The database does not provide any variable that could be used as an instrument in the selection equation. Therefore, the model does not satisfy exclusion restrictions and motivates the discussion of the exclusion restriction in Section 5.3.

Table 11 reports the estimation results; demographic variables are suppressed to save space. In this application, the optimal copula is Clayton, but the maximized likelihood value of the Clayton copula is only slightly larger than that of the Gaussian copula, and Vuong's test does not discriminate the estimators significantly.

The shape parameters of the selection equation are statistically significantly different from zero; a hypothesis of a normal distribution with $(\alpha_\nu, \delta_\nu) = (0.1436, 0)$ or a logistic distribution with $(\alpha_\nu, \delta_\nu) = (0, 0)$ is easily rejected. Instead, the underlying distribution is heavily left-skewed, with an implied skewness of -1.22 and a kurtosis of 5.78.

As for the outcome equation, its distribution is nearly symmetric—$\hat{\delta}_\varepsilon$ is small and statistically insignificant and skewness is only 0.10—and has slightly heavier tails than the normal distribution—kurtosis is 4.02. The estimated shape parameters would not permit rejection of the logistic distribution, but the distribution does differ significantly from normality. The GTL-Clayton slopes estimates in the outcome equation are smaller in absolute value that the normal-Gaussian estimates with the exception of lnMDE: we find that health expenditures are less sensitive to coinsurance, incentive payments, health status, family income, and demographic variables (with the exception of age, which has the same effect). By implication, health expenditures may well be less sensitive to a health policy reform than traditional model estimates suggest.

Both sets of estimates indicate positive dependence between the disturbances in the two equations. However they do differ: the GTL-Clayton estimate of the dependence of 0.24 is only half as large as the normal-Gaussian dependence of 0.53.

6.4 Speeding tickets, Massachusetts

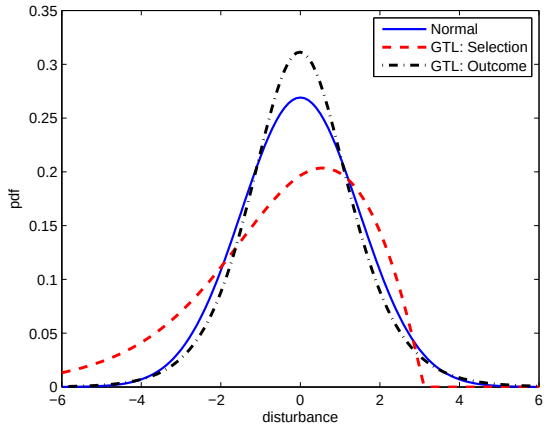
The sample selection model is also common in fields other than labor and health economics. As a case in point, Makowsky and Stratmann (2009) examine the determinants of traffic citations and fines for speeding, using a database that consists of all speeding-related stops in Massachusetts from April 1, 2001 through May 31, 2001.[43]

A traffic stop results in either a ticket or a warning. When a ticket is issued, a driver has to pay a fine. Whether a police officer issues a ticket or gives a warning is at the officer's discretion. If a ticket is issued, state law provides a formula for the amount of the fine: \$50

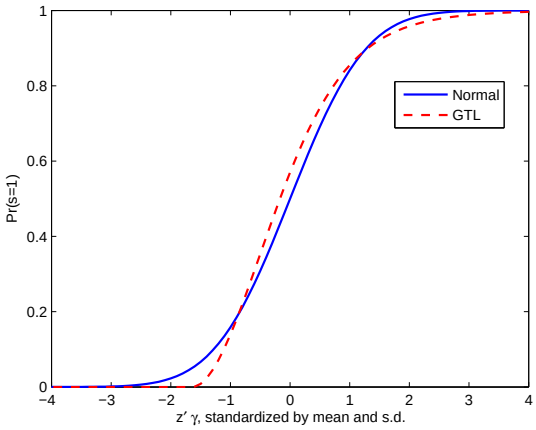
[42] The variables are defined in Table B.3; see also Cameron and Trivedi (2005, Table 20.4).

[43] The data are available online at `http://www.aeaweb.org/issue.php?journal=AER&volume=99&issue=1`.

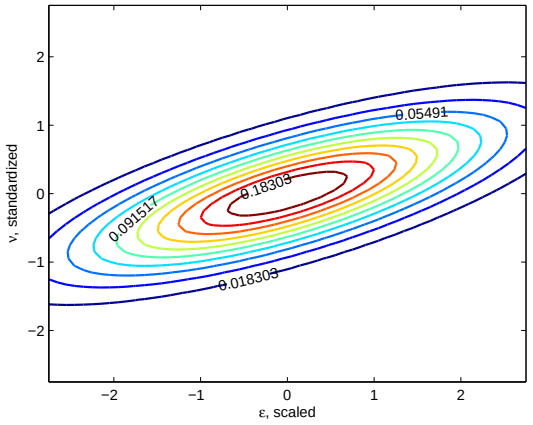
Figure 8: Differences between normal-Gaussian and GTL-copula: Health expenditures



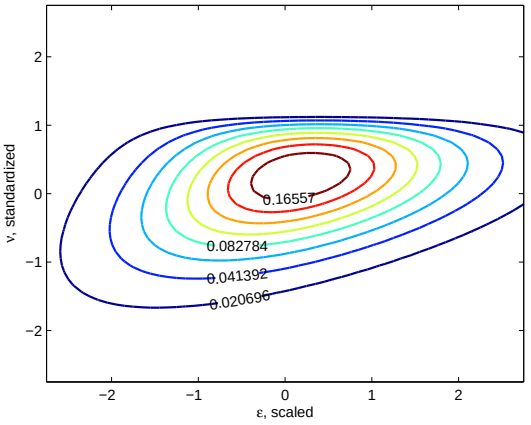
(a) PDF comparison



(b) $Pr(s_i = 1)$: GTL versus Probit



(c) Contour plot: Normal-Gaussian



(d) Contour plot: GTL-Clayton

Table 11: Estimation result: Health expenditures

	Normal-Gaussian		GTL-Clayton	
Variables	Coef.	(S.E.) [a]	Coef.	(S.E.) [a]
Selection equation: incurring health expenditures [b]				
lnCOINRATE	-0.1068	(0.0277)	-0.1468	(0.0322)
IDP	-0.1088	(0.0550)	-0.0849	(0.1012)
lnAPI	0.0295	(0.0087)	0.0350	(0.0102)
lnMMDE	0.0007	(0.0160)	0.0098	(0.0185)
PHYSLIM	0.2848	(0.0729)	0.3618	(0.0937)
NDISEASE	0.0211	(0.0035)	0.0276	(0.0043)
HEALTHgood	0.0577	(0.0428)	0.0453	(0.0499)
HEALTHfair	0.2237	(0.0822)	0.2347	(0.0968)
HEALTHpoor	0.7984	(0.2286)	0.8446	(0.3035)
lnFAMINC	0.0553	(0.0166)	0.0501	(0.0196)
α_ν			0.2190	(0.1074)
δ_ν			0.2592	(0.1047)
Outcome equation: Log of health expenditures				
lnCOINRATE	-0.0760	(0.0351)	-0.0476	(0.0339)
IDP	-0.1497	(0.0693)	-0.1203	(0.0648)
lnAPI	0.0149	(0.0104)	0.0085	(0.0098)
lnMMDE	-0.0235	(0.0197)	-0.0274	(0.0180)
PHYSLIM	0.3549	(0.0784)	0.2967	(0.0755)
NDISEASE	0.0286	(0.0038)	0.0247	(0.0037)
HEALTHgood	0.1559	(0.0523)	0.1323	(0.0469)
HEALTHfair	0.4451	(0.1001)	0.4130	(0.0929)
HEALTHpoor	0.9986	(0.2109)	0.7554	(0.2092)
lnFAMINC	0.1214	(0.0222)	0.0955	(0.0201)
α_ε			0.0150	(0.0244)
δ_ε			-0.0121	(0.0261)
σ	1.5701	(0.0303)	0.8592	(0.0313)
θ	0.7356	(0.0366)	0.6190	(0.3063)
τ	0.5262		0.2364	
$\ln L$	-10170.11		-10132.24	
AIC	20416.22		20348.49	

The selection and outcome equations also contain demographic variables; see Table B.3.

[a] White-robust standard errors.

[b] The selection equation is standardized by mean and standard deviation.

+ \$10 × (speed - (speed limit + 10)). Makowsky and Stratmann (2009, p.513) discuss the political economy hypothesis and the opportunity-cost hypothesis of officer behavior. The former relates the officers' decision to "the fiscal condition of the government that employs them and to whether the driver is a potential voter in local elections," and the latter predicts that "officers have a higher likelihood of issuing a ticket and issuing a larger fine amount when the opportunity cost for contesting the ticket is higher for drivers."

In this application, the selection indicator is whether a ticket is issued. The outcome

Table 12: Estimation results: Speeding tickets (selection equation)

Variables	Normal-Gaussian Coef.	(S.E.) [a]	GTL-nJoe Coef.	(S.E.) [a]	GTL-Gaussian Coef.	(S.E.) [a]
Selection Equation: Issuing a ticket [b]						
lnMPHOVER	1.2231	(0.0758)	1.3974	(0.1002)	1.6417	(0.1795)
CDL	-0.2182	(0.0256)	-0.1751	(0.0307)	-0.2506	(0.0334)
OUTTOWN	0.1958	(0.0277)	0.1576	(0.0406)	0.2130	(0.0370)
OUTSTATE	0.1602	(0.0376)	0.1111	(0.0360)	0.1798	(0.0456)
BLACK	-0.0044	(0.0423)	-0.0625	(0.0410)	-0.0176	(0.0529)
HISPANIC	0.2196	(0.0414)	0.1378	(0.0365)	0.2363	(0.0525)
FEMALE	-0.6602	(0.0810)	-0.4249	(0.0823)	-0.6251	(0.1014)
lnAGE	-0.3069	(0.0187)	-0.1772	(0.0240)	-0.3201	(0.0250)
FEMALE × lnAGE	0.1456	(0.0232)	0.0887	(0.0229)	0.1314	(0.0292)
lnDISTCOURT	0.0093	(0.0172)	0.0226	(0.0129)	0.0071	(0.0181)
lnPVALUEPC	-0.3238	(0.0846)	-0.2234	(0.0815)	-0.3516	(0.1087)
OR	-0.0007	(0.1476)	-0.0776	(0.2020)	-0.0543	(0.2117)
OR × OUTTOWN	0.5792	(0.2380)	0.3680	(0.2159)	0.6932	(0.3642)
OR × lnDISTCOURT	-0.0109	(0.0601)	0.0007	(0.0379)	-0.0320	(0.0680)
SP	-1.3561	(1.2928)	-1.1961	(1.0324)	-1.0851	(1.6202)
SP × OUTTOWN	-0.0519	(0.0462)	-0.0632	(0.0429)	-0.0619	(0.0565)
SP × lnDISTCOURT	0.0672	(0.0236)	0.0335	(0.0163)	0.0797	(0.0316)
SP × lnPVALUEPC	0.1815	(0.1160)	0.1564	(0.0915)	0.1707	(0.1431)
SP × OR	-0.2771	(0.2231)	-0.2249	(0.1549)	-0.2575	(0.2925)
constant	1.0289	(0.9946)	-1.0468	(0.8880)	0.1530	(1.4814)
α_ν			-0.8769	(0.2327)	-0.6821	(0.2126)
δ_ν			-0.4288	(0.0627)	-0.2623	(0.0803)

[a] Clustered standard error at municipality level.

[b] The selection equation is standardized by the median and interquartile range.

variable is the amount of fine (in logarithmic form), which is only observed when a ticket is issued. About 46% of the 68,357 stops resulted in a speeding ticket with an average fine of $122. The explanatory variables include the excess speed of the driver ("MPHOVER", in log form), driver characteristics (residence, race, ethnicity, gender, age, and the distance to court), and measures of the fiscal condition of a municipality (a dummy "OR" whether a municipality rejected a tax increase via an override referendum applicable to the operating budget of the 2001 fiscal year; property value per capita; and a dummy "SP" whether the traffic stop was made by a state police officer, who may have different incentives than a local police officer). The regression model includes several interactions as well, as indicated in the tables of results below. Finally, the selection equation also includes a dummy variable "CDL", denoting a commercial driver's license, which fulfills the exclusion restriction.

Tables 12 and 13 report the estimation results of the selection equation and the outcome equation, respectively. In this application, the best copula is nJoe. The likelihood values are strikingly different between the normal-Gaussian estimator and the GTL-nJoe estimator.

Table 13: Estimation results: Speeding tickets (outcome equation)

Variables	Normal-Gaussian Coef.	(S.E.) [a]	GTL-nJoe Coef.	(S.E.) [a]	GTL-Gaussian Coef.	(S.E.) [a]
Outcome equation: Amount of fine						
lnMPHOVER	0.9523	(0.0145)	1.1686	(0.0011)	1.2338	(0.0063)
OUTTOWN	0.0271	(0.0123)	-0.0002	(0.0001)	0.0001	(0.0005)
BLACK	-0.0228	(0.0100)	0.0001	(0.0001)	-0.0009	(0.0005)
HISPANIC	0.0363	(0.0108)	-0.0001	(0.0001)	-0.0002	(0.0007)
FEMALE	-0.0936	(0.0391)	0.0007	(0.0004)	0.0002	(0.0016)
lnAGE	-0.0214	(0.0086)	0.0004	(0.0001)	0.0013	(0.0009)
FEMALE × lnAGE	0.0163	(0.0109)	-0.0002	(0.0001)	0.0000	(0.0005)
lnDISTCOURT	0.0263	(0.0035)	-0.0001	(0.0000)	-0.0001	(0.0002)
lnPVALUEPC	-0.0371	(0.0272)	0.0003	(0.0002)	-0.0008	(0.0010)
OR	0.0220	(0.0744)	0.0003	(0.0003)	0.0059	(0.0021)
OR × OUTTOWN	0.0569	(0.0649)	-0.0009	(0.0004)	-0.0081	(0.0021)
OR × lnDISTCOURT	0.0007	(0.0107)	0.0000	(0.0001)	0.0003	(0.0004)
SP	-0.1530	(0.3404)	0.0007	(0.0021)	-0.0017	(0.0147)
SP × OUTTOWN	0.0143	(0.0196)	0.0000	(0.0001)	-0.0001	(0.0007)
SP × lnDISTCOURT	0.0094	(0.0044)	0.0002	(0.0001)	0.0007	(0.0004)
SP × lnPVALUEPC	0.0201	(0.0308)	-0.0002	(0.0002)	-0.0001	(0.0013)
SP × OR	-0.0548	(0.0317)	0.0002	(0.0003)	-0.0019	(0.0016)
constant	2.3312	(0.3048)	1.6608	(0.0026)	1.4807	(0.0195)
α_ε			-2.1298	(0.0525)	-1.8155	(0.2145)
δ_ε			1.1202	(0.0557)	1.3718	(0.2057)
σ	0.3361	(0.0079)	0.0004	(0.0000)	0.0030	(0.0006)
ρ or θ	0.3411	(0.0373)	4.0190	(0.8066)	0.0553	(0.1506)
τ	0.2216		-0.6151		0.0352	
$\ln L$	-45173.91		-4098.00		-10397.16	
AIC	90427.82		8283.99		20882.31	

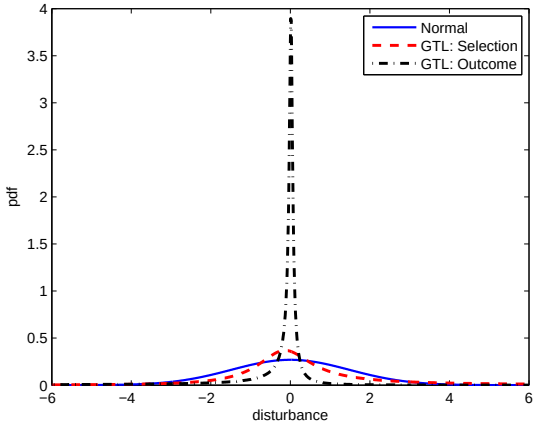
[a] Clustered standard error at municipality level.

The difference between the GTL-nJoe and GTL-Gaussian estimates is also large. Indeed, the Vuong test can statistically discriminate between the three models in favor of nJoe (Table 8).

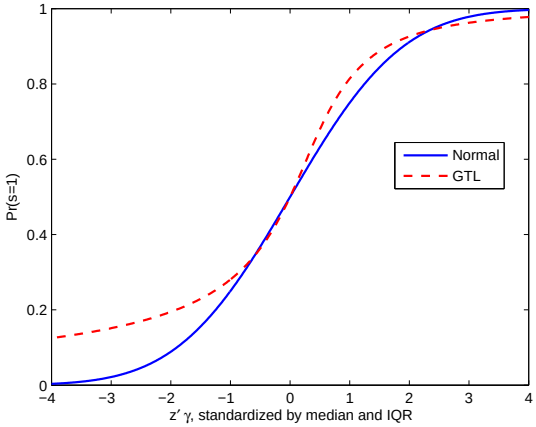
The implied degrees of dependence between the error terms are very different across the estimators. The normal-Gaussian normal estimator yields a moderately positive dependence, whereas the GTL-nJoe estimator exhibits a strong negative dependence; both are statistically highly significant. Figures 9c and 9d shows the contour plots implied by these estimators. As an intermediate form, the GTL-Gaussian estimator yields a small and insignificant degree of dependence.

The values of $\hat{\alpha_\nu}$ and $\hat{\delta_\nu}$ indicate that the disturbances driving the selection equation are considerably skewed and heavy-tailed. Even the first moment of this GTL distribution cannot be defined. Accordingly, the selection equation is standardized by median and interquartile

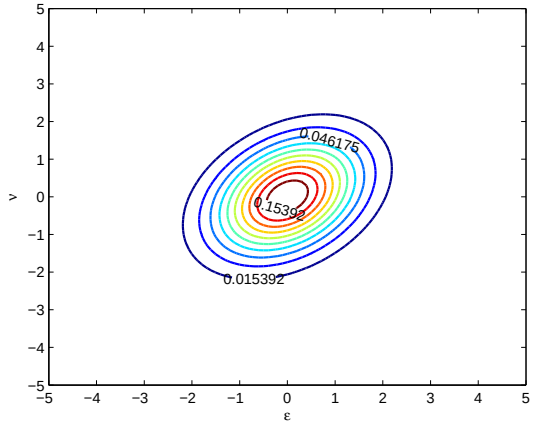
Figure 9: Differences between normal-Gaussian and GTL-copula: Speeding tickets



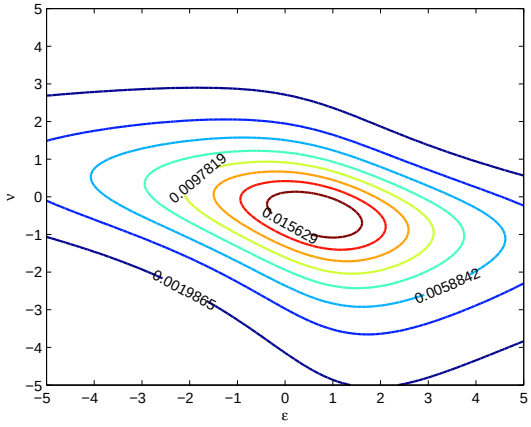
(a) PDF comparison



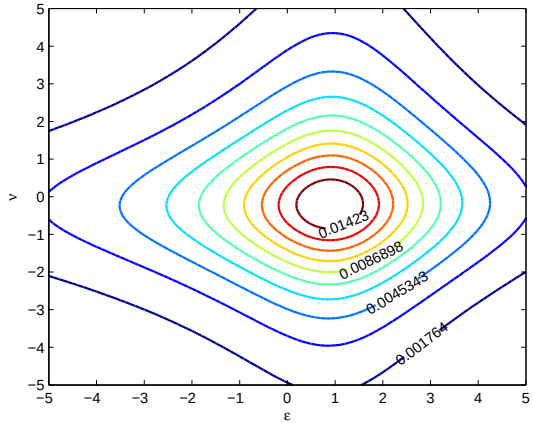
(b) $Pr(s_i = 1)$: GTL versus probit



(c) Contour Plot: Normal-Gaussian



(d) Contour Plot: GTL-nJoe



(e) Contour Plot: GTL-Gaussian

Table 14: Marginal effects on the probability of issuing a speeding ticket

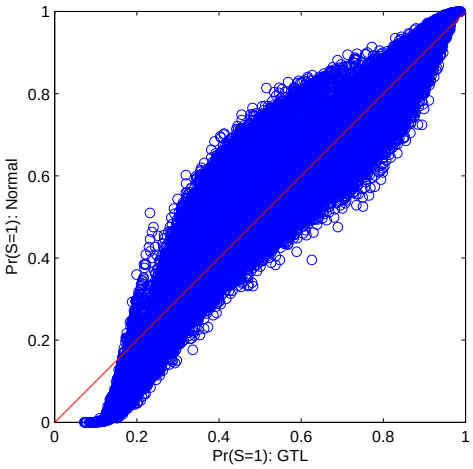
	Normal-Gaussian		GTL-nJoe		GTL-Gaussian	
Variables	Estimate	(S.E.) [a]	Estimate	(S.E.) [a]	Estimate	(S.E.) [a]
lnMPHOVER	0.489	(0.025)	0.508	(0.096)	0.510	(0.186)
OUTTOWN [b]	0.078	(0.010)	0.053	(0.009)	0.065	(0.013)
OR [b]	0.147	(0.046)	0.066	(0.029)	0.123	(0.057)
FEMALE [b]	-0.061	(0.005)	-0.042	(0.006)	-0.052	(0.009)
lnPVALUEPC	-0.129	(0.033)	-0.081	(0.034)	-0.109	(0.032)

Marginal effects are first evaluated for each observation, considering the interaction terms as well, and then averaged across all observations.

[a] Standard errors are computed by the Delta method.

[b] For a dummy variable d_i, the marginal effect for each observation is calculated as $Pr(s_i = 1|d_i = 1) - Pr(s_i = 1|d_i = 0)$.

Figure 10: Predicted probability $Pr(s_i = 1)$: GTL-nJoe vs normal-Gaussian



range in order to make the model comparable. Thus we find that, under GTL-nJoe, a higher speed figures more prominently in the chance of receiving a speeding ticker, that the effect of Hispanic ethnicity diminishes (as does the advantage of younger women), and that property values matter less. The many interactions obscure the effect of the distributional (normal-Gaussian) misspecification. Thus, Table 14 reports the marginal effect of a few selected variables: indeed, while the signs of marginal effects are the same across the estimators, the magnitudes differ. The political economy hypothesis as measured by property values and the OR dummy finds less support.

Altogether, the distributional misspecification changes the predicted probability of receiving a ticket (Figure 10): for example, for drivers who have a 60% chance of getting a speeding ticket for their offense under the normal-Gaussian model, the GTL-nJoe model assigns anywhere between a 40% and an 80% chance.

The difference in the outcome equation is even more striking. The disturbances are sharply peaked and have thick tails (Figure 9a). Even though the coefficients on some of variables are still statistically significant, no variable other than lnMPHOVER is economically significant any longer. The results indicate that the amount of fine is not varying at the discretion of officer in response to observable factors, in contrast with the findings by Makowsky and Stratmann (2009), but unobservable factors can occasionally cause major deviations from the fine that state law prescribes.

The estimated values of the coefficient on lnMPHOVER also have different implications. For the normal-Gaussian estimator, the coefficient on lnMPHOVER is less than 1, indicating that the fine is inelastic with respect to the severity of the speeding violation (miles over the speed limit). On the other hand, both GTL-copula estimators indicates an elasticity greater than 1: the fine is elastic. This is more intuitive: a more severe speeding violation draws an increasingly severe penalty; this also corresponds with the prescription in state law.[44]

6.5 International Disputes

The last application is in the area of political science and concerns the occurrence and severity of international disputes. Sweeney (2003) hypothesizes that a dispute between a pair of states with a greater degree of interest similarity is less likely to escalate to a severe level. There may also be a significant interaction effect between the degree of interest similarity and the balance of military capability. On the one hand, even if interests are dissimilar, balanced military capabilities may still prevent disputes from escalating. On the other, if interests are dissimilar and military capabilities are unbalanced, the weaker state may submit to the stronger state without putting up a fight. Sweeney's database consists of 149,004 country pairs between 1886 and 1992 with 972 disputes between them (0.65%).[45]

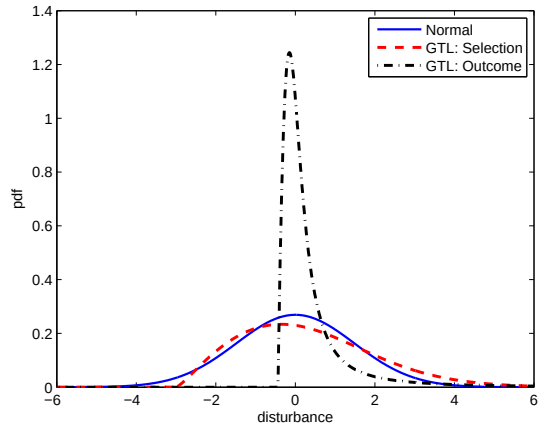
Dispute severity is measured by an index that combines the level of hostility and the number of fatalities. States express their interest similarity by being involved in similar alliances. Military balance is expressed as the ratio of the military capacity of the strongest member of a pair of states over their total capability. Other control variables include democracy, economic interdependence (measured by mutual trade flows relative to gross domestic product), common membership in international governmental organizations, geography (country contiguity and distance), aspects of the dispute (about territory and the number of states involved in the dispute), and one dummy whether one of the states is labeled a major power and another whether both states are labeled a major power.[46]

[44]Simple algebra with the formula for the amount of fine reveals that the elasticity exceeds 1 as long as the driver exceeded the speed limit by 5 miles. The elasticity is not constant, though. Furthermore, when mph over speed limit is less than 5, the elasticity is negative. However, in the entire sample, only 1% of the stopped drivers were going less than five miles over speed limit; one fifth of them received a ticket.

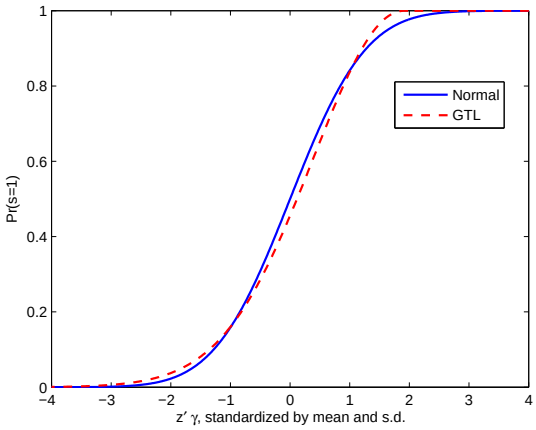
[45]The data are available online at `http://jcr.sagepub.com/content/47/6.toc`. For more detail on the variables, see also Oneal and Russett (1999).

[46]Sweeney (2003, Table 1) reports only one dummy variable, namely whether both are a major power. However, his Table 1 can only be replicated with this set of two dummy variables. This is the specification we follow.

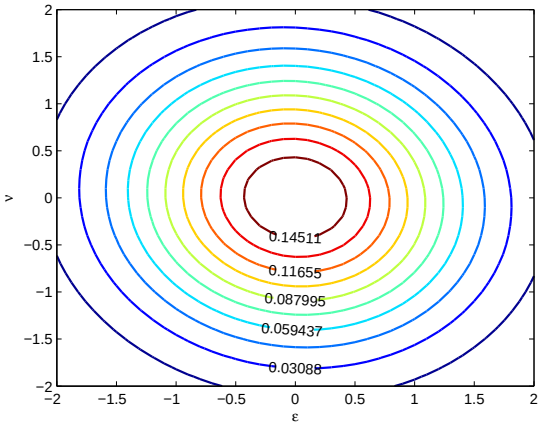
Figure 11: Differences between normal-Gaussian and GTL-Clayton: Severity of interstate disputes



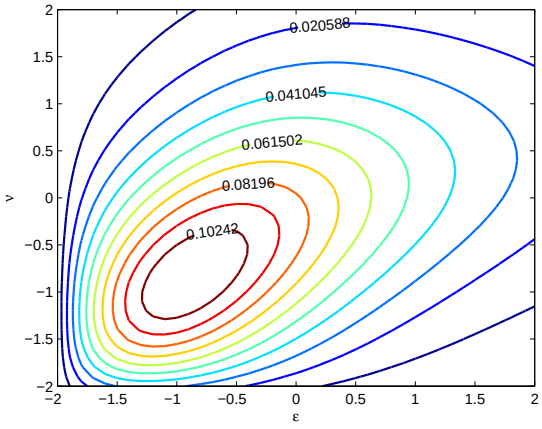
(a) PDF comparison



(b) $Pr(s_i = 1)$: GTL versus probit



(c) Contour Plot: Normal-Gaussian



(d) Contour Plot: GTL-Clayton

Table 15: Estimation result: Severity of interstate disputes

Variables	Normal-Gaussian Coef.	(S.E.) [a]	GTL-Clayton Coef.	(S.E.) [a]
Selection equation: Occurrence of interstate dispute [b]				
lnCAPRATIO	-0.569	(0.078)	-0.752	(0.122)
DEMOCRACY	-0.026	(0.003)	-0.036	(0.005)
DEPENDENCE	-15.386	(3.023)	-19.460	(4.101)
COMMON IGO	0.009	(0.001)	0.012	(0.002)
ALLIES	-0.182	(0.042)	-0.258	(0.061)
MAJOR POWERS	0.725	(0.035)	0.952	(0.067)
CONTIGUOUS	0.932	(0.038)	1.267	(0.113)
lnDISTANCE	-0.166	(0.016)	-0.214	(0.021)
PEACE	-0.097	(0.005)	-0.131	(0.012)
sp: PEACE-10	0.098	(0.005)	0.133	(0.013)
constant	-1.559	(0.135)	-1.808	(0.195)
α_ν			0.259	(0.033)
δ_ν			-0.176	(0.035)
Outcome equation: Severity of interstate dispute				
lnCAPRATIO	130.575	(70.53)	1.846	(16.07)
INT SIMILARITY	1.821	(31.21)	-13.535	(6.823)
lnCAPRATIO × INT SIMILARITY	-152.243	(78.29)	-5.704	(18.47)
DEMOCRACY	0.546	(0.318)	-0.044	(0.082)
DEPENDENCE	-1318.071	(294.9)	-30.810	(79.66)
COMMON IGO	-0.168	(0.105)	0.048	(0.024)
MAJOR v MAJOR	12.193	(6.437)	-1.413	(1.328)
CONTIGUOUS	8.854	(5.330)	0.987	(1.010)
lnDISTANCE	-1.035	(1.812)	-0.045	(0.499)
TERRITORY	12.100	(4.261)	-0.162	(1.037)
ACTORS	3.791	(0.477)	0.108	(0.094)
constant	66.516	(32.69)	26.483	(7.931)
α_ε			-0.173	(0.019)
δ_ε			-0.670	(0.019)
σ	47.787	(0.950)	6.985	(0.535)
ρ or θ	-0.054	(0.084)	3.111	(0.500)
τ	-0.034		0.609	
$\ln L$	-9209.59		-9063.70	
AIC	18469.18		18185.40	

[a] White-robust standard errors.

[b] The selection equation is standardized by mean and standard deviation.

To account for duration dependence (Beck et al., 1998), Sweeney (2003) adds to the selection equation the number of years since the last dispute (PEACE) in the form of four cubic spline variables. These variables make the matrix of explanatory variables in the selection equation highly multicollinear: the condition number equals 888.2 where a value of 20 is supposed to raise a red flag. This multicollinearity interferes with the iterative search

Table 16: Marginal effects on the probability of an interstate dispute occurring

	Normal-Gaussian				GTL-Clayton			
	5-pct	Mean	Median	95-pct	5-pct	Mean	Median	95-pct
lnCAPRATIO	-0.749	-0.152	-0.031	-0.003	-1.325	-0.363	-0.195	-0.058
DEMOCRACY	-1.170	-0.237	-0.049	-0.005	-2.135	-0.586	-0.315	-0.094
DEPENDENCE	-0.496	-0.101	-0.021	-0.002	-0.839	-0.230	-0.124	-0.037
COMMON IGO	0.004	0.172	0.035	0.848	0.068	0.425	0.229	1.550
ALLIES	-1.158	-0.229	-0.042	-0.004	-2.221	-0.587	-0.303	-0.086
CONTIGUOUS	0.205	2.201	1.191	7.533	1.376	5.078	3.880	12.632
lnDISTANCE	-0.999	-0.203	-0.042	-0.005	-1.728	-0.474	-0.255	-0.076
MAJOR POWERS	0.087	1.395	0.492	5.613	0.731	3.120	1.989	9.256
PEACE [a]	-1.227	-0.227	-0.065	-0.006	-1.803	-0.478	-0.288	-0.079
sp: PEACE-10 [b]	0.000	0.001	0.000	0.005	0.001	0.003	0.002	0.010

Effects shown are in percentage points. Effects are calculated for each observation separately and then summarized across the sample. Continuous explanatory variables change by one standard deviation. Dummy variables change from 0 to 1. Peacetime year variables change by one year.

[a] Summary statistics are computed over observations with at most 10 years since the last dispute.

[b] Summary statistics are computed over observations with more than 10 years since the last dispute. The variable PEACE changes simultaneously.

for the maximum likelihood estimate of GTL-copula models. We opt for a simpler linear spline with a knot at 10 years; this lowers the condition number to 38.3 and yields much smoother convergence.

Estimation results are presented in Table 15 and Figure 11. The marginal distribution of the selection disturbances is nonnormal mostly because it lacks a left tail (Figure 11a): skewness equals 0.66 and kurtosis 3.35. In other words, states initiate a dispute for important unobserved reasons (a large positive value of ν) but it never happens that they refrain from disputes for important unobserved reasons (a large negative ν). The disturbances of the outcome equation are strongly right-skewed with no left tail and a heavy right tail; skewness and kurtosis values cannot even be computed for this $(\hat{\alpha}_\varepsilon, \hat{\delta}_\varepsilon)$. Whereas the normal-Gaussian model shows virtually no correlation between the selection and outcome equations ($\hat{\rho} = -0.054$), the GTL-Clayton model estimates a strong positive dependence $\tau = 0.609$, which indeed is plausible: when a strong unobserved factor (a positive ν) causes states to initiate a dispute, the dispute is more likely to become severe (a positive ε).

The estimated slopes of the selection equation of the GTL-Clayton model are all a bit larger than those of the normal-Gaussian model. The explanatory variables indeed have a larger effect: the probability of a dispute occurring ranges from 0 to 0.44 with the normal-Gaussian model and from 0 to 0.57 with the GTL-Clayton model. The marginal effects of the explanatory variables are accordingly much stronger, more than doubling on average; see Table 16.

Even more striking is the sensitivity of the estimated outcome equation to the distributional assumption. Whereas the normal-Gaussian model counts six variables with t-statistics

above 1.85 (including lnCAPRATIO and its interaction with INT SIMILARITY), the GTL-Clayton model has only two statistically significant determinants: interest similarity by itself (in a negative direction in confirmation of Sweeney's hypothesis) and the number of common IGOs (in an implausible positive direction). Military capacity appears to be irrelevant after all; the number of actors and the volume of mutual trade flows that were the most significant variables are no longer significant contributors to the explanation of dispute severity; disputes over territory are no more severe than disputes for other reasons, nor are disputes between two major powers. These magnitudes of these changes may be surprising, but the matrix of explanatory variables is highly colinear (condition number of 89.9). When multicollinearity is present, slight changes in the specification often lead to large changes in estimated slopes.

7 Conclusion

In this paper, we propose a new maximum likelihood estimator for the sample selection model. We relax the assumption of a bivariate normal distribution by means of GTL marginal distributions and copula functions that tie the marginals together into a bivariate distribution. While we still make a distributional assumption, it is a weak assumption, such that our proposed estimator essentially does not impose any particular shape on the distribution. The GTL distribution allows thick or thin tails and left-skewed or right-skewed shapes; the collective set of copulas accommodates diverse dependence structures between two random variables. Together, they create a highly versatile bivariate distribution, which include the traditional joint normal estimator as a special case. In line with the terminology of the GTL-copula estimator, we term the traditional estimator the normal-Gaussian estimator, as it combines normal marginal distributions with a Gaussian copula.

The Monte Carlo study shows that the proposed estimator performs well under both normal and non-normal settings, whereas the normal-Gaussian estimator performs poorly when the distributional assumption is violated. A particularly valuable insight is that, unlike the traditional estimator, the GTL-copula estimator is much less dependent on the presence of an instrument in the selection equation that fulfills the exclusion restriction. Thus, no longer should it be considered problematic that the selection equation contains the same explanatory variables as the outcome equation.

The applications to real data also show economically significant differences between the traditional estimator and our proposed estimator. For example, the amount of fine for speeding violations proves to be determined only by the driver's excessive speed—unlike the estimates generated by the normal-Gaussian model that shows variations in fine by age, gender, ethnicity, and out-of-town residential location. The GTL-copula-estimated effect of a government grant program on wages of school-age children in Mexico proves to be only half as large as the effect estimated with the traditional normal-Gaussian estimator. The wage profile of Portuguese women is more sharply curved, with a curvature that varies less with the presence of children in the household. Moreover, the husband's wage is found to be more important for the wife's participation in the labor force. Health expenditures in the U.S. prove to be less sensitive to coinsurance, incentive payments, health status, family income,

and most demographic variables. In the application regarding international disputes, the GTL-copula estimator changes the list of factors that increase the severity of these disputes. In particular, military balance is no longer a relevant factor, which is a hotly debated issue in the international relations literature.

The applications offered samples of size 2,339 to nearly 150,000 observations; the subsamples for which an outcome was observed consisted of anywhere between 0.7% and 76.8% of the total sample. Together, these applications illustrate how this highly nonlinear GTL-copula estimator performs in applied research. In our experience, out of ten copula functions (as in Table 8), there are often a few that prove to be incompatible with the data and therefore present difficulties during the iterative search for the maximum likelihood estimate. This should not be considered an undesirable feature of the GTL-copula approach since different copula functions have different features. Indeed, it is the wide range of features that makes the GTL-copula approach so flexible. We did find that the GTL-copula estimator has more difficulty dealing with highly colinear sets of explanatory variables. We speculate that this happens because GTL link functions must be numerically inverted; slight numerical inaccuracies become magnified when the direction of search derives from an ill-conditioned hessian matrix.

The GTL-copula estimator that this paper proposes in the context of the standard sample selection model (with one selection equation and one outcome equation, i.e., the type-2 Tobit model) can be straightforwardly extended to other types of sample selection models. For example, the Roy selection model has one selection equation that separates observations into two states, with one outcome equation for each state, akin to Lee (1978). In preliminary work, we examined three applications of the Roy model with the GTL-copula estimator and found evidence similar to the findings reported in this paper for the typical sample selection model. The model may also be expanded with more than two states or with a dual or higher-dimensional selection mechanism. Thus, we add a highly flexible and practical estimator to the literature.

Appendices

A Comparison of normal-Gaussian and GTL-copula estimators: Additional Monte Carlo results

In this appendix, we present Monte Carlo results that parallel those of Tables 3 and 4 with a reduced ($\tau = 0.2$) and a strengthened ($\tau = 0.5$) level of dependence.

Table A.1: Biases and standard deviations when DGPs use different copulas; $\tau = 0.2$

	β_0^*	β_1	α_ε	δ_ε	τ	α_ν	δ_ν
DGP Copula: Guassian							
Normal-Gaussian	0.012	-0.006			-0.011		
	(0.135)	(0.083)			(0.103)		
GTL-Product	0.244	-0.135	-0.002	-0.001		0.034	0.001
	(0.037)	(0.039)	(0.028)	(0.017)		(0.137)	(0.057)
GTL-Gaussian	0.012	-0.007	-0.001	0.000	-0.012	0.034	0.002
	(0.140)	(0.086)	(0.030)	(0.018)	(0.108)	(0.137)	(0.057)
Pretest	0.007	0.002	0.001	-0.002	-0.002	0.035	0.002
	(0.158)	(0.104)	(0.032)	(0.023)	(0.106)	(0.139)	(0.057)
DGP Copula: FGM							
Normal-Gaussian	0.018	-0.010			-0.023		
	(0.147)	(0.089)			(0.112)		
GTL-Product	0.236	-0.129	-0.013	0.012		0.034	0.001
	(0.036)	(0.038)	(0.029)	(0.017)		(0.137)	(0.057)
GTL-FGM	0.032	-0.017	-0.002	0.002	-0.028	0.031	0.001
	(0.103)	(0.068)	(0.030)	(0.018)	(0.079)	(0.136)	(0.057)
Pretest	-0.022	0.012	-0.002	0.008	0.016	0.039	0.001
	(0.143)	(0.088)	(0.034)	(0.025)	(0.099)	(0.212)	(0.057)
DGP Copula: Frank							
Normal-Gaussian	0.094	-0.022			-0.091		
	(0.196)	(0.115)			(0.162)		
GTL-Product	0.234	-0.129	-0.014	0.014		0.034	0.001
	(0.036)	(0.039)	(0.029)	(0.017)		(0.137)	(0.057)
GTL-Frank	0.007	-0.001	0.001	0.000	-0.010	0.032	0.001
	(0.123)	(0.066)	(0.034)	(0.024)	(0.100)	(0.137)	(0.057)
Pretest	0.033	-0.006	0.006	-0.008	-0.034	0.037	0.000
	(0.193)	(0.106)	(0.035)	(0.027)	(0.161)	(0.138)	(0.057)
DGP Copula: Clayton							
Normal-Gaussian	0.036	-0.019			-0.040		
	(0.137)	(0.084)			(0.106)		
GTL-Product	0.224	-0.097	0.005	-0.018		0.033	0.001
	(0.034)	(0.037)	(0.029)	(0.017)		(0.137)	(0.057)
GTL-Clayton	0.008	-0.004	0.002	0.001	-0.009	0.033	0.001
	(0.125)	(0.079)	(0.032)	(0.019)	(0.100)	(0.137)	(0.057)
Pretest	-0.011	0.005	-0.004	0.008	0.003	0.032	0.000
	(0.125)	(0.080)	(0.033)	(0.025)	(0.097)	(0.138)	(0.057)
DGP Copula: Gumbel							
Normal-Gaussian	0.001	-0.019			0.003		
	(0.105)	(0.068)			(0.077)		
GTL-Product	0.258	-0.161	0.002	0.007		0.034	0.001
	(0.039)	(0.041)	(0.029)	(0.016)		(0.137)	(0.057)
GTL-Gumbel	0.011	-0.008	-0.001	-0.001	-0.009	0.036	0.003
	(0.110)	(0.075)	(0.029)	(0.017)	(0.083)	(0.137)	(0.055)
Pretest 0.010	-0.006	-0.002	-0.001	-0.009	0.036	0.002	
	(0.098)	(0.068)	(0.030)	(0.018)	(0.074)	(0.138)	(0.056)
DGP Copula: Joe							
Normal-Gaussian	-0.016	-0.026			0.023		
	(0.094)	(0.063)			(0.067)		
GTL-Product	0.273	-0.186	0.011	0.013		0.034	0.001
	(0.041)	(0.042)	(0.030)	(0.016)		(0.137)	(0.057)
GTL-Joe	0.004	-0.004	-0.002	-0.001	-0.003	0.036	0.003
	(0.080)	(0.060)	(0.028)	(0.017)	(0.058)	(0.135)	(0.054)
Pretest	-0.017	0.003	0.002	-0.001	0.016	0.036	0.004
	(0.080)	(0.059)	(0.030)	(0.017)	(0.059)	(0.135)	(0.055)

Note: See Table 3.

Table A.2: Biases and standard deviations when DGPs use different copulas; $\tau = 0.5$

	β_0^*	β_1	α_ε	δ_ε	τ	α_ν	δ_ν
DGP Copula: Guassian							
Normal-Gaussian	0.000	0.000			-0.002		
	(0.074)	(0.053)			(0.053)		
GTL-Product	0.564	-0.317	-0.007	-0.021	-0.500	0.034	0.001
	(0.033)	(0.037)	(0.028)	(0.016)		(0.137)	(0.057)
GTL-Gaussian	0.001	-0.001	0.000	-0.002	-0.002	0.031	0.002
	(0.079)	(0.055)	(0.034)	(0.022)	(0.057)	(0.128)	(0.053)
Pretest	0.005	-0.003	0.007	-0.009	-0.003	0.034	0.003
	(0.081)	(0.057)	(0.040)	(0.031)	(0.058)	(0.131)	(0.054)
DGP Copula: Frank							
Normal-Gaussian	0.032	-0.018			-0.045		
	(0.075)	(0.053)			(0.057)		
GTL-Product	0.545	-0.302	-0.060	0.005		0.033	0.001
	(0.032)	(0.036)	(0.028)	(0.016)		(0.137)	(0.057)
GTL-Frank	-0.001	0.001	0.000	-0.001	0.000	0.030	0.001
	(0.068)	(0.049)	(0.034)	(0.020)	(0.050)	(0.132)	(0.055)
Pretest	-0.010	0.004	-0.011	0.008	0.002	0.026	-0.003
	(0.075)	(0.055)	(0.042)	(0.031)	(0.054)	(0.135)	(0.057)
DGP Copula: Clayton							
Normal-Gaussian	-0.062	0.087			0.034		
	(0.104)	(0.071)			(0.086)		
GTL-Product	0.524	-0.260	0.015	-0.063		0.034	0.001
	(0.030)	(0.035)	(0.028)	(0.017)		(0.137)	(0.057)
GTL-Clayton	0.003	0.001	0.008	-0.007	-0.002	0.027	0.002
	(0.078)	(0.054)	(0.042)	(0.028)	(0.060)	(0.126)	(0.055)
Pretest	-0.001	0.006	0.011	-0.011	0.001	0.033	0.000
	(0.079)	(0.057)	(0.043)	(0.033)	(0.061)	(0.129)	(0.055)
DGP Copula: Gumbel							
Normal-Gaussian	0.039	-0.046			-0.038		
	(0.070)	(0.050)			(0.050)		
GTL-Product	0.572	-0.340	-0.026	0.005		0.034	0.001
	(0.034)	(0.038)	(0.030)	(0.016)		(0.137)	(0.057)
GTL-Gumbel	-0.002	0.000	-0.001	-0.001	0.000	0.033	0.002
	(0.067)	(0.048)	(0.033)	(0.019)	(0.049)	(0.127)	(0.051)
Pretest	0.008	-0.004	-0.001	-0.004	-0.008	0.037	0.002
	(0.070)	(0.050)	(0.034)	(0.024)	(0.053)	(0.128)	(0.053)
DGP Copula: Joe							
Normal-Gaussian	0.043	-0.077			-0.046		
	(0.066)	(0.047)			(0.046)		
GTL-Product	0.583	-0.370	-0.048	0.032		0.034	0.001
	(0.035)	(0.038)	(0.031)	(0.015)		(0.137)	(0.057)
GTL-Joe	-0.002	0.000	-0.001	-0.001	0.001	0.033	0.002
	(0.055)	(0.042)	(0.031)	(0.018)	(0.039)	(0.124)	(0.049)
Pretest	-0.005	0.002	0.000	0.002	0.006	0.031	0.003
	(0.070)	(0.066)	(0.033)	(0.020)	(0.048)	(0.124)	(0.050)

Note: See Table 3.

Table A.3: Frequencies of selecting copulas under different DGPs for $\tau = 0.2$

Estimated	Copula of the DGP					
copula	Gaussian	FGM	Frank	Clayton	Gumbel	Joe
Gaussian	**0.350**	0.150	0.132	0.136	0.100	0.028
FGM	0.156	**0.430**	0.242	0.182	0.022	0.008
Frank	0.128	0.218	**0.356**	0.112	0.080	0.018
Clayton	0.148	0.138	0.130	**0.512**	0.016	0.000
Gumbel	0.144	0.054	0.100	0.036	**0.362**	0.262
Joe	0.078	0.012	0.040	0.022	0.420	**0.684**

Notes: See Table 4.

Table A.4: Frequencies of selecting copulas under different DGPs for $\tau = 0.5$

Estimated	Copula of the DGP				
copula	Gaussian	Frank	Clayton	Gumbel	Joe
Gaussian	**0.754**	0.110	0.038	0.074	0.002
FGM	0.002	0.006	0.000	0.000	0.000
Frank	0.130	**0.804**	0.022	0.034	0.002
Clayton	0.022	0.024	**0.940**	0.000	0.000
Gumbel	0.088	0.052	0.000	**0.698**	0.078
Joe	0.004	0.004	0.000	0.194	**0.920**

Notes: See Table 4.

B Variable Definitions and Summary Statistics

Table B.1: Wages of married women, Portugal [a]

Number of Obs.		Whole Sample 2,339		Selected Sample 1,400	
Variables	Definition	Mean	Std. dev.	Mean	Std. dev.
Selection	1 if wage is observed	0.599	0.490		
Outcome	log of hourly wage (in escudos)			5.832	0.660
CHILDY18	the number of children younger than 18 living in the family	1.622	1.096	1.524	0.997
CHILDY3	the number of children younger than 3	0.197	0.438	0.206	0.433
lnHUSBW	log of husband's monthly wage (in escudos)	11.196	0.378	11.222	0.382
YRSCH	years of schooling	7.237	3.771	8.335	4.043
AGE	age in years, divided by 10	3.842	0.941	3.715	0.880
AGE2	AGE squared	15.647	7.447	14.576	6.859
PEXP	potential experience, age - years of schooling - 6, divided by 10	2.518	1.063	2.282	0.996
PEXP2	PEXP squared	7.472	5.737	6.197	5.059
PEXP×CHILDY18	PEXP × CHILDY18	4.509	3.697	3.947	3.212
PEXP2×CHILDY18	PEXP2 × CHILDY18	12.818	13.598	10.405	11.454

[a] Source: Derived from the dataset used in Martins (2001); variable names have been slightly changed.

Table B.2: Wages of school-aged children, Mexico [a]

Number of Obs.		Whole Sample 15,526		Selected Sample 1,657	
Variables	Definition	Mean	S.D.	Mean	S.D.
Selection	1 if out of school and reporting a wage	0.107	0.309		
Outcome	log of hourly wage			1.254	0.564
PROGRESA/eligible	1 if in PROGRESA community and eligible	0.527	0.499	0.496	0.500
PROGRESA/ineligible	1 if in PROGRESA community but ineligible	0.097	0.296	0.104	0.305
Control/eligible	1 if in Control community but eligible (reference catogory)	0.312	0.463	0.323	0.468
Control/ineligible	1 if in Control community and ineligible [b]	0.063	0.243	0.077	0.266
ORDER	order of birth	2.522	1.280	3.176	1.552
AGE	age in years	12.904	2.556	15.801	1.315
AGE2	age squared	173.042	66.481	251.395	39.308
GRANT	ratio of grant to household income	0.665	0.814	0.671	0.893
FATHERhome	1 if father presents in household	0.855	0.352	0.814	0.389
MOTHERhome	1 if mother presents in household	0.881	0.323	0.820	0.385
INDIGENOUS	1 if speak indigenous language	0.297	0.457	0.262	0.440
DISTANCE97	distance to secondary school in 1997	2.169	2.139	2.194	2.018
DISTANCE98	distance to secondary school in 1998	2.100	2.096	2.119	1.989
PRIMARY97	1 if primary school existed in community in 1997	0.970	0.170	0.974	0.159
PRIMARY98	1 if primary school existed in community in 1998	0.991	0.097	0.989	0.106
SECONDARY97	1 if secondary school existed in community in 1997	0.275	0.446	0.259	0.438
SECONDARY98	1 if secondary school existed in community in 1998	0.279	0.448	0.264	0.441
lnMAWAGE	log of community-averaged male agriculture wage	1.290	0.286	1.316	0.290
YRSCH	completed years of education	5.075	2.258	5.931	2.260
PROGRESA	1 if in PROGRESA community	0.624	0.484	0.600	0.490

[a] Source: Derived from the dataset used in Attanasio et al. (2012); variable names have been slightly changed. The selection equation also contains dummy variables for father's education, mother's education, and state of residence.

[b] This definition is different from the data description provided at `http://restud.oxfordjournals.org/content/79/1.toc`.

Table B.3: Health expenditure, USA [a]

Number of Obs.		Whole Sample 5,574		Selected Sample 4,281	
Variables	Definition	Mean	S.D.	Mean	S.D.
Selection	1 if medical expenditures > 0	0.768	0.422		
Outcome	log of annual medical expenditures, constant dollars, excluding dental and outpatient mental expenditures			4.069	1.499
lnCOINRATE	ln(coinsurance rate+1) with $0 \leq$ rate ≤ 100	2.421	2.044	2.254	2.044
IDP	1 if individual deductible plan	0.262	0.440	0.245	0.430
lnAPI	ln(annual participation incentive payment) or 0 if no payment	4.727	2.681	4.740	2.676
lnMMDE	ln(maximum medical deductible expenditure) if IDP=1 and MMDE>1 or 0 otherwise.	4.065	3.451	3.855	3.515
PHYSLIM	1 if physical limitation	0.124	0.323	0.140	0.342
NDISEASE	number of chronic diseases	11.205	6.789	11.795	7.033
HEALTHgood	1 if good health	0.365	0.481	0.366	0.482
HEALTHfair	1 if fair health	0.078	0.269	0.080	0.272
HEALTHpoor	1 if poor health	0.016	0.124	0.018	0.134
lnFAMINC	log of family income (in dollars)	8.697	1.221	8.778	1.091
lnFAMSIZE	log of family size	1.241	0.540	1.222	0.532
YRSCHHEAD	education of household head (in years)	11.947	2.837	12.117	2.824
AGE	age in years	25.576	16.730	26.431	17.121
FEMALE	1 if female	0.518	0.500	0.544	0.498
CHILD	1 if age is less than 18	0.405	0.491	0.382	0.486
GIRL	FEMALE × CHILD	0.196	0.397	0.184	0.387
BLACK	1 if black	0.186	0.386	0.138	0.341

[a] Source: Derived from the dataset used in Cameron and Trivedi (2005); variable names have been slightly changed.

Table B.4: Speeding tickets, Massachusetts [a]

Number of Obs.		Whole Sample 68,306		Selected Sample 31,642	
Variables	Definition	Mean	Std. dev.	Mean	Std. dev.
Selection	1 if ticket issued	0.463	0.499		
Outcome	log of fine amount (in dollars)			4.707	0.438
lnMPHOVER	log of mile per hour over speed limit	2.666	0.328	2.783	0.333
CDL	1 if commercial driver	0.030	0.169	0.023	0.149
OUTTOWN	1 if out of town driver	0.773	0.419	0.848	0.359
OUTSTATE	1 if out of state driver	0.155	0.362	0.222	0.415
BLACK	1 if a driver is black	0.045	0.206	0.051	0.219
HISPANIC	1 if a driver is hispanic	0.035	0.183	0.047	0.211
FEMALE	1 if a driver is female	0.390	0.488	0.332	0.471
lnAGE	log of age	3.498	0.376	3.442	0.366
FEMALE × lnAGE	FEMALE × lnAGE	1.368	1.726	0.332	0.471
lnDISTCOURT	log of distance to court (in miles)	2.624	1.211	2.886	1.298
lnPVALUEPC	log property value per capita	11.258	0.499	11.165	0.499
OR	1 if a tax increase rejected via override referendum	0.020	0.139	0.026	0.160
OR × OUTTOWN	OR × OUTTOWN	0.018	0.134	0.025	0.157
OR × lnDISTCOURT	OR × lnDISTCOURT	0.055	0.427	0.074	0.494
SP	1 if an officer is state police	0.269	0.444	0.445	0.497
SP × lnDISTCOURT	SP × lnDISTCOURT	0.886	1.606	1.508	5.540
SP × lnPVALUEPC	SP × lnPVALUEPC	3.003	4.952	4.951	0.077
SP × OR	SP × OR	0.003	0.056	0.006	0.077

[a] Source: Derived from the dataset used in Makowsky and Stratmann (2009); variable names have been slightly changed.

Table B.5: Severity of interstate disputes [a]

Number of Obs.		Whole Sample 149,004		Selected Sample 972	
Variables	Definition	Mean	S.D.	Mean	S.D.
Selection	1 if an interstate dispute occurs	0.007	0.081		
Outcome	severity of dispute			70.751	51.779
lnCAPRATIO	log of strong-state military capability divided by the total dyad capability	-0.210	0.196	-0.252	0.204
INT SIMILARITY	similarity of revealed preferences between the two states			0.892	0.093
DEMOCRACY	democracy level of the least democratic state in the dyad	-3.320	6.630	-4.894	5.197
DEPENDENCE	economic interdependence: ratio of total dyad trade divided by gross domestic product	0.001	0.005	0.003	0.006
COMMON IGO	number of joint international governmental organization (IGO) memberships	25.256	14.251	28.214	15.883
ALLIES	1 if formally allied	0.133	0.340	0.258	0.438
MAJOR POWER	1 if one of the states is a major power	0.175	0.380	0.521	0.500
MAJOR v MAJOR	1 if both states are major powers			0.130	0.336
CONTIGUOUS	1 if states are contiguous	0.080	0.271	0.687	0.464
lnDISTANCE	log of distance	8.042	0.897	6.969	1.077
PEACE	time since last dispute	20.939	20.788	9.188	15.328
sp: PEACE-10	linear spline term, = PEACE - 10 if PEACE > 10, $= 0$ otherwise	13.128	19.068	5.184	12.418
TERRITORY	1 if the dispute was over territorial issues			0.258	0.438
ACTORS	number of states in the dispute			3.789	4.528

[a] Source: Derived from the dataset used in Sweeney (2003); variable names have been slightly changed.

References

Ahn, H., Powell, J. L., 1993. Semiparametric estimation of censored selection models with a nonparametric selection mechanism. Journal of Econometrics 58 (1), 3–29.

Albert, J. H., Chib, S., 1993. Bayesian analysis of binary and polychotomous response data. Journal of the American Statistical Association 88 (422), 669–679.

Amemiya, T., 1985. Advanced econometrics. Harvard University Press, Cambridge, Mass.

Attanasio, O. P., Meghir, C., Santiago, A., 1 2012. Education choices in mexico: Using a structural model and a randomized experiment to evaluate progresa. Review of Economic Studies 79, 37–66.

Beck, N., Katz, J. N., Tucker, R., 10 1998. Taking time seriously: Time-series-cross-section analysis with a binary dependent variable. American Journal of Political Science 42 (4), 1260–1288.

Boero, G., Smith, J., Wallis, K. F., 2004. The sensitivity of chi-squared goodness-of-fit tests to the partitioning of data. Econometric Reviews 23 (4), 341–370.

Cameron, A. C., Trivedi, P. K., 2005. Microeconometrics: Methods and Applications. Cambridge University Press, New York, NY.

Cameron, C., Trivedi, P. K., Zimmer, D. M., 2004. Modelling the differences in counted outcomes using bivariate copula models with application to mismeasured counts. Econometrics Journal 7 (2), 566–584.

Cherubini, U., Luciano, E., Vecchiato, W., 2004. Copula methods in finance. John Wiley & Sons, Hoboken, NJ.

Cosslett, S. R., 1993. Semiparametric estimation of a regression model with sample selectivity. In: Barnett, W. A., Powell, J., Tauchen, G. (Eds.), Nonparametric and Semiparametric Methods in Econometrics and Statistics. Cambridge University Press, pp. 175–197.

Dancer, D., Rammohan, A., Smith, M. D., 2008. Infant mortality and child nutrition in Bangladesh. Health Economics 17 (9), 1015–1035.

Das, M., Newey, W. K., Vella, F., 2003. Nonparametric estimation of sample selection models. The Review of Economic Studies 70 (1), 33–58.

Deb, P., Trivedi, P. K., 2002. The structure of demand for health care: latent class versus two-part models. Journal of Health Economics 21 (4), 601–625.

Freimer, M., Kollia, G., Mudholkar, G. S., Lin, C. T., 1988. A study of the generalized tukey lambda family. Communications in Statistics - Theory and Methods 17 (10), 3547–3567.

Genius, M., Strazzera, E., 2008. Applying the copula approach to sample selection modelling. Applied Economics 40, 1443–1455.

Golan, A., Moretti, E., Perloff, J. M., 2004. A small-sample estimator for the sample-selection model. Econometric Reviews 23 (1), 71–91.

Gouriéroux, C., Holly, A., Monfort, A., 1982. Likelihood ratio test, Wald test, and Kuhn-Tucker test in linear models with inequality constraints on the regression parameters. Econometrica 50 (1), 63–80.

Hastings, C., Mosteller, F., Tukey, J. W., Winsor, C. P., 9 1947. Low moments for small samples: A comparative study of order statistics. Annals of Mathematical Statistics 18 (3), 413–426.

Heckman, J., 1974. Shadow prices, market wages, and labor supply. Econometrica 42 (4), 679–694.

Heckman, J., Tobias, J. L., Vytlacil, E., 2003. Simple estimators for treatment parameters in a latent-variable framework. The Review of Economics and Statistics 85 (3), 748–755.

Heckman, J. J., 1979. Sample selection bias as a specification error. Econometrica 47 (1), 153–161.

Karian, Z. A., Dudewicz, E. J., 2011. Handbook of Fitting Statistical Distributions with R. CRC Press., Boca Raton, FL.

King, R. A. R., MacGillivray, H. L., 1999. A starship estimation method for the generalized λ distributions. Australian & New Zealand Journal of Statistics 41 (3), 353.

Koenker, R., Yoon, J., 2009. Parametric links for binary choice models: A Fisherian – Bayesian colloquy. Journal of Econometrics 152 (2), 120–130.

Lee, L.-F., June 1978. Unionism and wage rates: A simultaneous equations model with qualitative and limited dependent variables. International Economic Review 19 (2), 415–433.

Lee, L.-F., 1982. Some approaches to the correction of selectivity bias. The Review of Economic Studies 49 (157), 355–372.

Lee, L.-F., 1983. Generalized econometric models with selectivity. Econometrica 51 (2), 507–512.

Lee, L.-F., 2001. Self-selection. In: Baltagi, B. H. (Ed.), A Companion to Theoretical Econometrics. Wiley-Blackwell, pp. 383–409.

Maddala, G., 1983. Limited Dependent and Qualitative Variables in Econometrics. Cambridge University Press, Cambridge, U.K.

Makowsky, M. D., Stratmann, T., 2009. Political economy at any speed: What determines traffic citations? American Economic Review 99 (1), 509–527.

Martins, M. F. O., 2001. Parametric and semiparametric estimation of sample selection models: an empirical application to the female labour force in Portugal. Journal of Applied Econometrics 16 (1), 23–39.

Meijer, E., Wansbeek, T., 2007. The sample selection model from a method of moments perspective. Econometric Review 26 (1), 25–51.

Mudholkar, G. S., George, E. O., 1978. A remark on the shape of the logistic distribution. Biometrika 65 (3), 667–668.

Nagler, J., 1994. Scobit: An alternative estimator to logit and probit. American Journal of Political Science 38 (1), 230–255.

Nelsen, R. B., 2006. An introduction to copulas, 2nd Edition. Springer, New York.

Newey, W. K., 1999. Consistency of two-step sample selection estimator despite misspecification of distribution. Economics Letter 63 (2), 129–132.

Newey, W. K., 2009. Two-step series estimation of sample selection models. Econometrics Journal 12, S217–S229.

Olsen, R. J., 1980. A least squares correction for selectivity bias. Econometrica 48 (7), 1815–1820.

Oneal, J. R., Russett, B., 10 1999. The kantian peace: The pacific benefits of democracy, interdependence, and international organizations, 1885-1992. World Politics 52 (1), 1–37.

Öztürk, A., Dale, R. F., 1985. Least squares estimation of the parameters of the generalized lambda distribution. Technometrics 27 (1), 81–84.

Pregibon, D., 1980. Goodness of link tests for generalized linear models. Journal of the Royal Statistical Society. Series C (Applied Statistics) 29 (1), pp. 15–23, 14.

Prieger, J. E., 2002. A flexible parametric selection model for non-normal data with application to health care usage. Journal of Applied Econometrics 17 (4), 367–392.

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

Puhani, P., 2000. The Heckman correction for sample selection and its critique. Journal of Economic Surveys 14 (1), 53–68.

Ramberg, J. S., Schmeiser, B. W., 1974. An approximate method for generating asymmetric random variables. Communications of the ACM 17 (2), 78–82.

Ramberg, J. S., Tadikamalla, P. R., Dudewicz, E. J., Mykytka, E. F., 1979. A probability distribution and its uses in fitting data. Technometrics 21 (2), 201–214.

Roy, A. D., 1951. Some thoughts on the distribution of earnings. Oxford Economic Papers 3 (2), 135–146.

Sarabia, J.-M., 1997. A hierarchy of Lorenz curves based on the Generalized Tukey's lambda distribution. Econometric Reviews 16 (3), 305–320.

Smith, M. D., 2003. Modelling sample selection using Archimedean copulas. Econometrics Journal 6 (1), 99–123.

Smith, M. D., 2005. Using copulas to model switching regimes with an application to child labour. Economic Record 81, S47–S57.

Su, S., 2007. Numerical maximum log likelihood estimation for generalized lambda distributions. Computational Statistics & Data Analysis 51 (8), 3983–3998.

Sweeney, K. J., 12 2003. The severity of interstates disputes: Are dyadic capability preponderances really more pacific? Journal of Conflict Resolution 47 (6), 728–750.

Trivedi, P. K., Zimmer, D. M., 2007. Copula modeling: An introduction for practioners. Foundations and Trends in Econometrics 1 (1), 1–111.

Tukey, J. W., 1960. The practical relationship between the common transformation of percentages of count and of amount. Tech. Rep. 36, Princeton University.

van Hassel, M., 2011. Bayesian inference in a sample selection model. Journal of Econometrics 165 (2), 221–232.

Vella, F., 1998. Estimating models with sample selection bias: A survey. The Journal of human resources 33 (1), 127–169.

Vijverberg, C.-P. C., Vijverberg, W. P., 2012. Pregibit: A family of discrete choice models. Discussion Paper 6359, IZA.

Vijverberg, W. P., Hasebe, T., 2012. GTL regression: A linear model with skewed and thick-tailed disturbances, forthcoming.

Vuong, Q. H., 1989. Likelihood ratio tests for model selection and non-nested hypotheses. Econometrica 57 (2), 307–333.

White, H., 1982. Maximum likelihood estimation of misspecified models. Econometrica 50 (1), 1–25.

Winkelmann, R., 2011. Copula bivariate probit models: with application to medical expenditures. Health Economics.

Yen, S. T., Yuan, Y., Liu, X., 2009. Alcohol consumption by men in China: A non-Gaussian censored system approach. China Economic Review 20 (2), 162–173.

Zimmer, D. M., Trivedi, P. K., 2006. Using trivariate copulas to model sample selection and treatment effects: Application to family health care demand. Journal of Business & Economic Statistics 24 (1), 63–76.

Zuehlke, T. W., Zeman, A. R., 1991. A comparison of two-stage estimators of censored regression models. The Review of Economics and Statistics 73 (1), 185–188.