ESSAYS IN LABOR ECONOMICS AND ECONOMETRICS APPLICATIONS OF THE COPULA METHOD

by

Takuya Hasebe

A dissertation submitted to the Graduate Faculty in Economics in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York

2013

This manuscript has been read and accepted for the Graduate Faculty in Economics in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

Dr. Wim Vijverberg

Date — Chair of Examining Committee

Dr. Merih Uctum

Date — Executive Officer

Dr. Wim Vijverberg

Dr. David Jaeger

Dr. Michael Grossman

Dr. Partha Deb

Supervisory Committee

The City University of New York

Abstract

ESSAYS IN LABOR ECONOMICS AND ECONOMETRICS

APPLICATIONS OF THE COPULA METHOD

by

Takuya Hasebe

Adviser: Professor Wim Vijverberg

This dissertation mainly consists of three essays of original research. One element that these essays have in common is a copula method that generates joint distributions in flexible ways. Therefore, Chapter 1 describes the copula method as an introduction. Two essays, Chapter 2 and Chapter 3, are empirical research in the field of labor economics, in which the copula method is applied to construct econometrics models. One essay, Chapter 4, uses copulas in order to develop a new econometric technique.

Chapter 2 empirically investigates the difference in wage structures of permanent workers and temporary workers in the Netherlands. The findings are that starting wages of permanent workers are slightly lower than starting wages of temporary workers and that wages of permanent workers grow more rapidly than wages of temporary workers. These findings derive from an econometric model that is built on a distributional assumption using the copula method that relaxes the traditional model.

Chapter 3 empirically investigates the structure of adjustment costs of factors of production with a plant-level panel dataset from the Indonesian manufacturing sector. The copula method is applied in order to estimate the adjustment costs of labor and capital simultaneously and to differentiate the distribution assumption from a more standard approach used in previous studies. The estimates provide evidence of nonconvex and asymmetric adjustment costs of both labor and capital.

Chapter 4 proposes a new approach to estimating sample selection models that combines Generalized Tukey Lambda (GTL) distributions with the copula method. 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. 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 various data generating processes. Six applications illustrate the value of the proposed GTL-copula estimator.

Acknowledgments

This dissertation would not have been possible without the support and encouragement from the people around me. Above all, I would like to thank my supervisor, Dr. Wim Vijverberg, for his guidance, encouragement, understanding and patience all through my graduate study at City University of New York. His wisdom and knowledge have always inspired me. I would like to thank my dissertation committee members, Dr. David Jaeger, Dr. Michael Grossman, and Dr. Partha Deb, for their support and valuable comments on my research.

I would like to extend my gratitude to the people at the New York office of the National Bureau of Economic Research, especially, Christine Nagorski and Marinella Moscheni for their encouragement and assistance. I would also thank my fellow students in the Ph.D. Program in Economics for discussions on my research as well as my friends in New York, New Jersey, and Japan for being there for me.

I am also very grateful to my parents, Norio and Misako Hasebe, and my sisters, Fumi and Shiho Hasebe, who have always supported and believed in me.

Lastly, I would like to dedicate my dissertation to the late Dr. Robert Lipsey. I learned a lot from being a research assistant for him. I regret that I could not show him my completed dissertation.

Contents

Preface iv

List of Tables x

List of Figures xiii

1 Copula Method **1**

1.1 Introduction 1

1.2 Copula Functions 4

1.3 Dependence Structures 6

1.4 Dependence Measure 10

1.5 Copula Selection 13

1.6 Applications 14

1.6.1 Sample Selection Models with Continuous Outcome 14

1.6.2 Sample Selection Models with Binary Outcome 16

1.6.3 Sample Selection Models with Duration Outcome 18

1.7 Concluding Remarks 19

2 Effects of Contract Types on Starting Wage and Wage Growth **20**

2.1 Introduction 20

2.2 Institutional Setting 23

2.3 Related Literature 24

2.4 Data 27

2.5 Empirical Model 30

2.5.1 Starting Wage 30

2.5.2 Wage Growth 33

2.6 Results 35

2.7 Robustness Checks 43

2.8 Concluding Remarks 50

2.A The Estimation Procedure 52

2.B Additional Results 54

3 The Structure of Adjustment Costs of Factors of Production: Evidence from Indonesian Manufacturing Plants 57

3.1 Introduction 57

3.2 Theoretical Model 60

3.3 Estimation Strategy 63

3.3.1 Bivariate Normality Assumption 67

3.3.2 Alternative Distributional Assumption 68

3.4 Data 72

3.5 Results 74

3.5.1 Parameter Estimation Results 74

3.5.2 Model Predictions 83

3.6 Changes in Labor Market Regulations 88

3.7 Conclusion 94

3.A Data 96

3.B The Probabilities of Adjustments under Bivariate Normality 97

3.C Estimation with the Copula Method 99

3.D The Predictions from the Model 100

4 A Flexible Sample Selection Model: A GTL-Copula Approach
with Wim Vijverberg 104

4.1 Introduction 104
4.2 The Sample Selection Model 107
4.3 The Copula Approach 108
4.4 Specifying the Marginal Distributions: The Case for GTL 113
4.4.1 Marginal Distributions for the Copulas 113
4.4.2 The GTL Distribution 115
4.4.3 Econometric Issues 119
4.5 Monte Carlo Simulations 122
4.5.1 Varying the Dependence Structure 122
4.5.2 Varying the Marginal Distributions 127
4.5.3 A Bivariate χ^2 Distribution 129
4.5.4 Addressing the Exclusion Restriction 130
4.6 Applications 135
4.6.1 Wages of Married Women, Portugal 136
4.6.2 Wages of Married Women, USA 142
4.6.3 Wages of School-Aged Children, Mexico 146
4.6.4 Health Expenditures, USA 152
4.6.5 Speeding tickets, Massachusetts 155
4.6.6 International Disputes 161
4.7 Conclusion 165
4.A Comparison of normal-Gaussian and GTL-copula estimators: Additional Monte Carlo results 168
4.B Variable Definitions and Summary Statistics 172

Bibliography **178**

List of Tables

1.1 Archimedean copula and generator functions 5
1.2 Dependence parameter θ and Kendall's τ 11

2.1 Strictness of Protection and the Share of Temporary Employment in 2008 21
2.2 Descriptive Statistics 29
2.3 The Estimated Selection Equation 36
2.4 The Estimated Starting Wage Equation: Specification 2 38
2.5 The Estimated Starting Wage Differentials 39
2.6 The Estimated Current Wage Equation: Specification 2 41
2.7 Estimation Results from Different Measures of Wage 45
2.8 The Estimation Results by Gender 47
2.9 Starting Wage Differentials with and without Exclusion Restrictions . 49
2.10 The Estimated Result of Specification 1 55
2.11 The Estimated Result with the Copula Method 56

3.1 Summary Statistics 73
3.2 Frequencies of Adjustment Decisions 74
3.3 Estimation Result: Bivariate Normal 75
3.4 First-Step Log Likelihood Values under Various Copulas 78
3.5 Estimation Result: Gumbel Copula 79
3.6 Estimation Result: Gaussian Copula 80

3.7 Reparameterized Estimates of Adjustment Cost Parameters 82
3.8 The Sample and Predicted Means of H_{it}/L_{it-1} and I_{it}/K_{it-1} 84
3.9 The Structural Parameters Before and After the Year of 2003 90
3.10 Prediction Comparisons in 2003 93

4.1 Dependence parameter θ and Kendall's τ 110
4.2 Biases and standard deviations when DGPs use different copulas; $\tau = 0.333$ 124
4.3 Frequencies of selecting copulas under different DGPs for $\tau = 0.333$. 125
4.4 Biases and standard deviations when DGPs have nonnormal marginals 128
4.5 Simulation results when DGP is a bivariate χ^2 distribution 130
4.6 Bias and standard deviation of $\hat{\beta}_1$ with and without an instrument when DGPs use different copulas and normal marginals 132
4.7 Bias and standard deviation of $\hat{\beta}_1$ with and without an instrument when DGPs use nonnormal marginals and a Gaussian copula 133
4.8 GTL-copula model selection: log-likelihood values and Vuong tests . . 135
4.9 Estimation results: Log-wages of married women 138
4.10 Estimation results: Log Wage of Married Women using CPS-MORG data 145
4.11 Estimation results: Wages of school-aged children 150
4.12 Estimation result: Health expenditures 154
4.13 Estimation results: Speeding tickets (selection equation) 156
4.14 Estimation results: Speeding tickets (outcome equation) 157
4.15 Marginal effects on the probability of issuing a speeding ticket 160
4.16 Estimation result: Severity of interstate disputes 163
4.17 Marginal effects on the probability of an interstate dispute occurring . 165
4.18 Biases and standard deviations when DGPs use different copulas; $\tau = 0.2169$
4.19 Biases and standard deviations when DGPs use different copulas; $\tau = 0.5170$

4.20 Frequencies of selecting copulas under different DGPs for $\tau = 0.2$ . . 171
4.21 Frequencies of selecting copulas under different DGPs for $\tau = 0.5$ . . 171
4.B.1 Wages of married women, Portugal 172
4.B.2 Wages of married women, USA-CPS 173
4.B.3 Wages of school-aged children, Mexico 174
4.B.4 Health expenditure, USA 175
4.B.5 Speeding tickets, Massachusetts 176
4.B.6 Severity of interstate disputes 177

List of Figures

1.1 Contour plots of bivariate pdf for non-Archimedean copulas 7

1.2 Contour plots of bivariate pdf for Archimedean copulas 8

1.3 Scatter plots and conditional expectations for Copulas 9

1.4 Illustration of Positive and Negative Dependence 12

2.1 Predicted Probability $Pr(d_i)$: Student's t vs Normal 37

2.2 The Estimated Distributions of Wage at Different Time of Employment 42

2.3 Wage Growth over Months: Temporary Worker vs Permanent Worker 44

3.1 The Unconditional Means of H_{it}/L_{it-1} and I_{it}/K_{it-1} 87

3.2 Relative Size of Fixed and Convex Costs in Labor Adjustment 91

3.3 Relative Size of Fixed and Convex Costs in Capital Adjustment . . . 92

4.1 GTL distributions with different α and δ 116

4.2 Differences between normal-Gaussian and GTL-Joe: Wages of married women 139

4.3 Density of scaled disturbances of the wage equation 141

4.4 Log-wage profiles, married women in Portugal 141

4.5 Differences between normal-Gaussian and GTL-copula: Wages of Married Women-CPS 144

4.6 Predicted probability $Pr(s_i = 1)$: GTL-Gumbel vs normal-Gaussian . 146

4.7 Differences between normal-Gaussian and GTL-Gumbel: Predicted Wage Distribution . . . 147
4.8 Differences between normal-Gaussian and GTL-copula: Wages of school-aged children . . . 149
4.9 Kernel density plots of predicted log-wages . . . 151
4.10 Differences between normal-Gaussian and GTL-copula: Health expenditures . . . 153
4.11 Differences between normal-Gaussian and GTL-copula: Speeding tickets 159
4.12 Predicted probability $Pr(s_i = 1)$: GTL-nJoe vs normal-Gaussian . . . 160
4.13 Differences between normal-Gaussian and GTL-Clayton: Severity of interstate disputes . . . 162

1 Copula Method

1.1 Introduction

The main body of my dissertation consists of two chapters in labor economics, which conduct empirical studies, and one chapter in econometrics, which develops a new estimation technique. One thing that all the chapters have in common is the copula method. Therefore, in this chapter, I briefly describe the copula method. For an introduction to copulas, see Nelsen (2006) and Trivedi and Zimmer (2007a). Even though it can be extended to a multi-dimensional case and a discrete case, the following discussion is restricted to a two-dimensional continuous case such that there are two continuous random variables.

In short, a copula is a parametric representation of a joint distribution with given marginal distributions. 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. Let f_{12} be the bivariate pdf of ω_1 and ω_2. Using the chain rule, the bivariate pdf is derived as

$$f_{12}(\omega_1, \omega_2) = \frac{\partial^2 F_{12}(\omega_1, \omega_2)}{\partial \omega_1 \partial \omega_2} = \frac{\partial^2 C(F_1, F_2; \theta)}{\partial F_1 \partial F_2} \times \frac{\partial F_1}{\partial \omega_1} \times \frac{\partial F_2}{\partial \omega_2}$$

Note also $C(F_1, F_2; \theta) = C(F_2, F_1; \theta)$.

Using a copula function, the probabilities $Pr(W_1 > \omega_1, W_2 \leq \omega_2)$, $Pr(W_1 \leq \omega_1, W_2 > \omega_2)$, and $Pr(W_1 > \omega_1, W_2 > \omega_2)$ can be expressed, respectively,

$$\begin{aligned}
Pr(W_1 > \omega_1, W_2 \leq \omega_2) =& F_2 - C(F_1, F_2; \theta) \\
Pr(W_1 \leq \omega_1, W_2 > \omega_2) =& F_1 - C(F_1, F_2; \theta) \\
Pr(W_1 > \omega_1, W_2 > \omega_2) =& 1 - F_1 - F_2 + C(F_1, F_2; \theta)
\end{aligned}$$

The formulation of these probabilities is applicable in modeling a bivariate binary choice model (Winkelmann, 2011).

The joint distribution is said to be radially symmetric around 0 if $Pr(W_1 \leq \omega_1, W_2 \leq \omega_2) = Pr(W_1 > -\omega_1, W_2 > -\omega_2)$. The distribution is radially symmetric if and only if (i) the distribution is marginally symmetric such that $F_j(\omega_j) = 1 - F_j(-\omega_1)$ for $j = 1, 2$, and (ii) the copula is radially symmetric such that $C(1 - F_1, 1 - F_2; \theta) = 1 - F_1 - F_2 + C(F_1, F_2; \theta)$, (Nelsen, 2006, Prokhorov and Schmidt, 2009). For example, a bivariate standard normal distribution is a well-know example.

It is also useful to derive the probability $Pr(W_1 = \omega_1, W_2 \leq \omega_2)$.

$$
\begin{aligned}
Pr(W_1 = \omega_1, W_2 \leq \omega_2) &= \int_{-\infty}^{\omega_2} f_{12}(\omega_1, W_2) dW_2 \\
&= \frac{\partial}{\partial \omega_1} \left\{ \int_{-\infty}^{\omega_1} \int_{-\infty}^{\omega_2} f_{12}(W_1, W_2) dW_2 dW_1 \right\} \\
&= \frac{\partial}{\partial \omega_1} F_{12}(\omega_1, \omega_2) \\
&= \frac{\partial}{\partial F_1} C(F_1, F_2; \theta) \times \frac{\partial F_1}{\partial \omega_1}
\end{aligned}
$$

Similarly, we can find the probability $Pr(W_1 = \omega_1, W_2 > \omega_2)$ as

$$
\begin{aligned}
Pr(W_1 = \omega_1, W_2 > \omega_2) &= \int_{\omega_2}^{\infty} f_{12}(\omega_1, W_2) dW_2 \\
&= \frac{\partial}{\partial \omega_1} \left\{ \int_{-\infty}^{\omega_1} \int_{\omega_2}^{\infty} f_{12}(W_1, W_2) dW_2 dW_1 \right\} \\
&= \frac{\partial}{\partial \omega_1} \left\{ \int_{-\infty}^{\omega_1} \int_{-\infty}^{\infty} f_{12}(W_1, W_2) dW_2 dW_1 \right. \\
&\qquad \left. - \int_{-\infty}^{\omega_1} \int_{-\infty}^{\omega_2} f_{12}(W_1, W_2) dW_2 dW_1 \right\} \\
&= \frac{\partial}{\partial \omega_1} \left(F_1 - C(F_1, F_2; \theta) \right) \\
&= \left(1 - \frac{\partial}{\partial F_1} C(F_1, F_2; \theta) \right) \times \frac{\partial F_1}{\partial \omega_1}
\end{aligned}
$$

The formulation of these densities are applicable in modeling sample selection models with continuous outcome(s) (Smith, 2003, 2005). Chapters 2 and 4 are about the sample selection models. The former applies the technique to a particular empirical study, and the latter extend the technique to develop a new estimator.

1.2 Copula Functions

Many different copula functions are available, each with its own characteristics. One of the most frequently used copulas is the Gaussian copula:

$$\begin{aligned} C(F_1, F_2; \theta) &= \int_{-\infty}^{\Phi^{-1}(F_1)} \int_{-\infty}^{\Phi^{-1}(F_2)} \frac{1}{2\pi\sqrt{1-\theta^2}} \exp\left\{ \frac{-(u_1^2 - 2\theta u_1 u_2 + u_1^2)}{2(1-\theta^2)} \right\} du_1 du_1 \\ &= \Phi_2(\Phi^{-1}(F_1), \Phi^{-1}(F_2); \theta), \end{aligned}$$

where $\Phi(\cdot)$ is a cdf of a standard normal distribution and $\Phi_2(\cdot)$ is a cdf 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).

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, where one of marginal distributions is the distribution commonly used in duration (survival) study: for example, exponential, Weibull, Gamma distributions, and so on. One of the attractive features of this copula is its mathematical simplicity, which facilitates computation since no integration is involved (unlike the Gaussian copula). The Plackett copula is also

Table 1.1: Archimedean copula and generator functions

Copula Name	$C(F_1, F_2; \theta)$	$\varphi(t)$
AMH	$F_1F_2\left[1-\theta(1-F_1)(1-F_2)\right]^{-1}$	$\log\left(\frac{1-\theta(1-t)}{t}\right)$
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 $j = 1, 2$.

familiar (Plackett, 1965). It is

$$C(F_1, F_2; \theta) = \frac{r - \sqrt{r^2 - 4F_1F_2\theta(\theta-1)}}{2(\theta-1)},$$

where $r = 1 + (\theta - 1)(F_1 + F_2)$. Like the FGM copula, no integration is necessary to evaluate the joint distribution. As discussed later, the Plackett copula exhibits a wide range of degree of dependence while the FGM exhibits a limited degree of dependence.

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 1.1 lists five examples of such generator functions, each dependent on a single parameter θ that determines the degree of dependence, as will be discussed later on.[1] Archimedean copulas have several attractive attributes (Smith, 2003). First, similar to the FGM and Plackett copulas, they do not require integration. Second, their

[1] 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.

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:

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

for $j = 1, 2$, and

$$\frac{\partial^2}{\partial F_1 \partial F_2} C(F_1, F_2; \theta) = -\frac{\varphi'(F_1)\varphi'(F_2)\varphi''(C(F_1, F_2; \theta))}{[\varphi'(C(F_1, F_2; \theta))]^3},$$

where $\varphi'(\cdot)$ and $\varphi''(\cdot)$ are the first derivative and the second derivative of $\varphi(\cdot)$, respectively.

More importantly, Archimedean copulas are attractive since they exhibit various dependence structures.

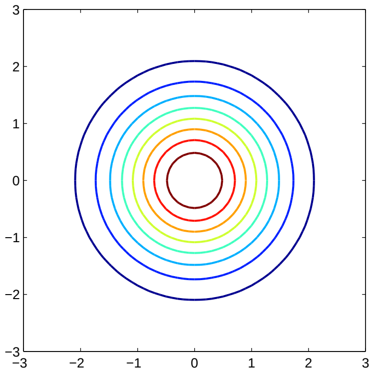
1.3 Dependence Structures

Different copulas exhibit different dependence structure. This is illustrated in Figures 1.1 and 1.2, 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 Product, FGM, and Plackett.[2] 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. In this aspect, the Frank copula is similar to the Gaussian, FGM, and Plackett copulas. These copulas are radially symmetric. However, the dependence of the Frank copula is weaker in the tails than the Gaussian one.

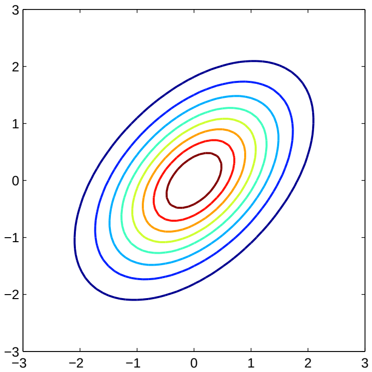
Other Archimedean copulas are not radially symmetric. The Clayton copula is

[2]For each copula, Kendall's τ equals 0.333. For Product, τ is 0. For FGM, τ equals 0.2. For Plackett, Spearman's ρ equals 0.333.

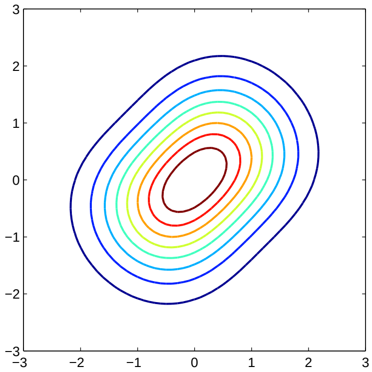
Figure 1.1: Contour plots of bivariate pdf for non-Archimedean copulas



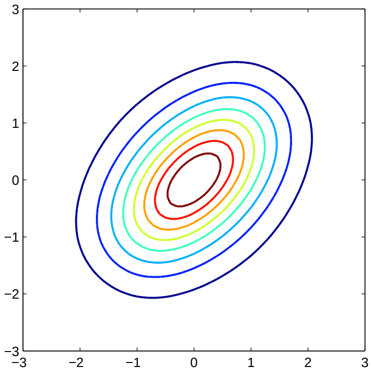
(a) Product copula: $\theta =$ N.A.



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

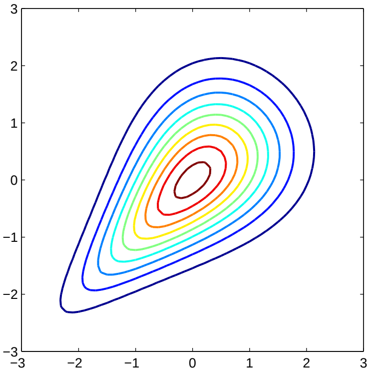


(c) FGM copula: $\theta = 0.8$

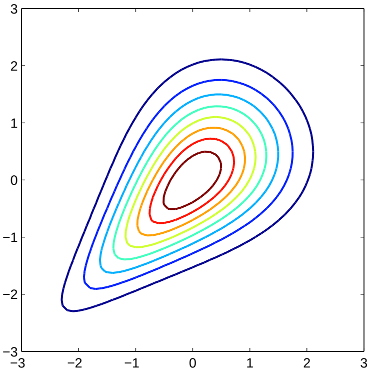


(d) Plackett copula: $\theta = 2.817$

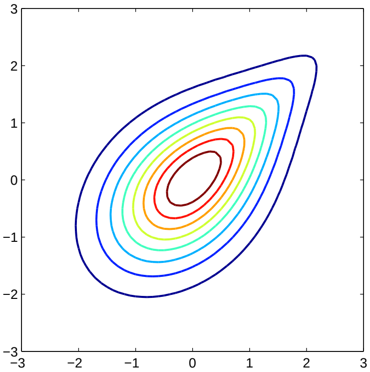
Figure 1.2: Contour plots of bivariate pdf for Archimedean copulas



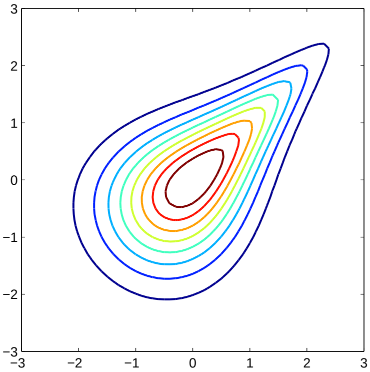
(a) AMH copula: $\theta = 0.9999$



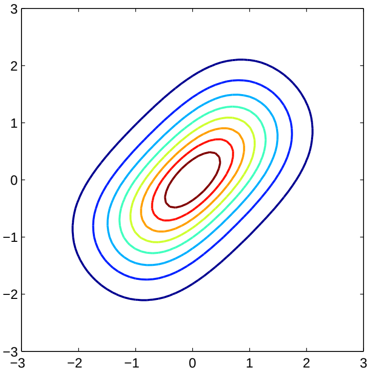
(b) Clayton copula: $\theta = 1$



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

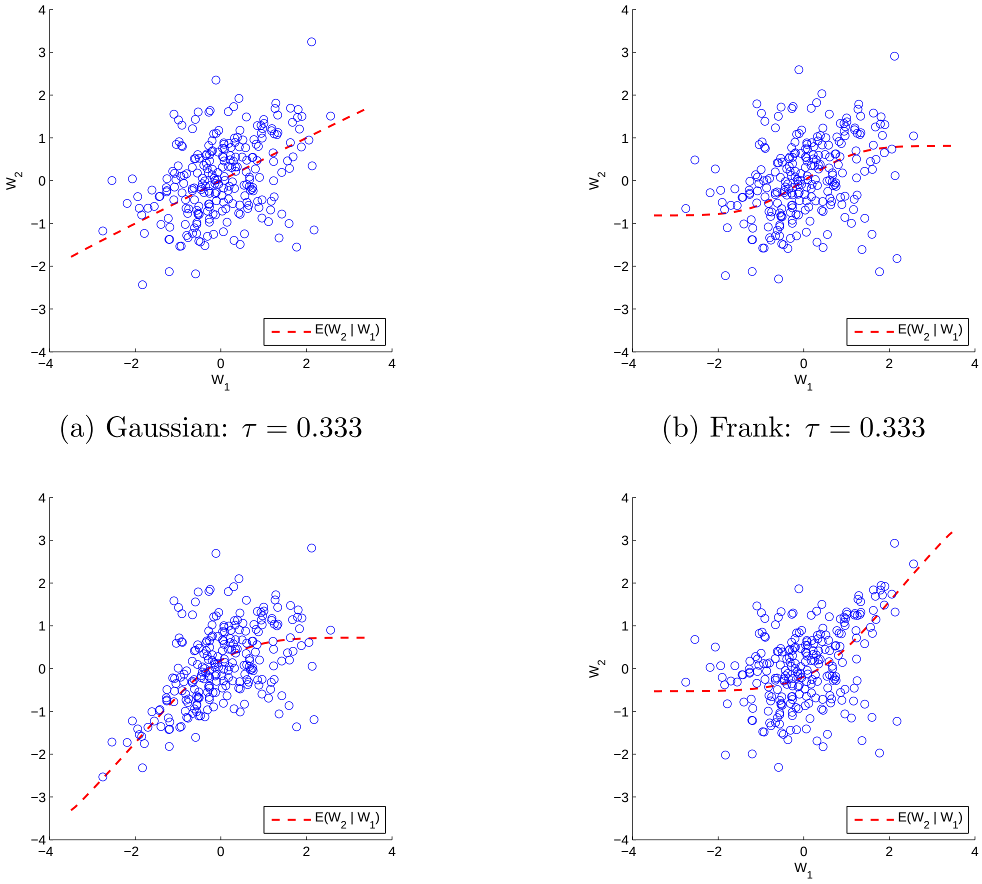


(d) Joe copula: $\theta = 1.9050$



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

Figure 1.3: Scatter plots and conditional expectations for Copulas



(a) Gaussian: $\tau = 0.333$

(b) Frank: $\tau = 0.333$

(c) Clayton: $\tau = 0.333$

(d) Joe: $\tau = 0.333$

asymmetric with strong lower tail dependence but weaker upper tail dependence, so is the AMH (Ali-Mikhail-Haq) copula. In contrast, the Gumbel and Joe copulas exhibit strong upper tail but weaker dependence in a lower tail.

To see the differences in tail dependence, I also compute the expectations of W_2 conditional on W_1, $E(W_2|W_1)$, for selected copulas, which are shown in Figure 1.3.[3] Although the scatter plots for the Gaussian and Frank copulas look alike, the conditional expectation $E(W_2|W_1)$ for the Gaussian copula is a linear function of W_1 over a

[3]250 pairs are drawn from each copula with standard normal as marginal distributions. To simulate draws from copulas, I adopt a conditional sampling method. The procedure is described in the appendix of Trivedi and Zimmer (2007a). For Gumbel and Joe copulas, the conditioning sampling is numerically iterated. The conditional expectations are numerically evaluated except for the Gaussian copula.

range of W_1 whereas the conditional expectation for the Frank is a nonlinear function at tails. The Clayton and Joe copulas show quite opposite patterns.

1.4 Dependence Measure

As the captions to Figures 1.1 and 1.2 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.[4] It is defined as, for independent pairs (W_1, W_2) and (W_1', W_2'),

$$\tau = Pr((W_1 - W_1')(W_2 - W_2') > 0) - Pr((W_1 - W_1')(W_2 - W_2') < 0). \tag{1.1}$$

With a given copula function, this measure can be computed as

$$\tau = 4 \int_0^1 \int_0^1 C(F_1, F_2) dC(F_1, F_2) - 1. \tag{1.2}$$

Furthermore, for Archimedean copulas, the expression can be simplified with a corresponding generator function,

$$\tau = 1 + 4 \int_0^1 \frac{\varphi(t)}{\varphi'(t)} dt. \tag{1.3}$$

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. In contrast, even if the familiar Pearson's coefficient of correlation equals 0, independence is not implied. The copula corresponding to independence is the Product copula (also referred to as the Independence copula), $C(F_1, F_2) = F_1 F_2$.

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

Table 1.2: 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	$2\theta/9$	$-2/9 \leq \tau \leq 2/9$
Plackett	$0 \leq \theta \leq \infty$	1	$\cdot$	$-1 \leq \tau 1$
Archimedean Family				
AMH	$-1 \leq \theta \leq 1$	0	$\left(\frac{3\theta-2}{\theta}\right)-\frac{2}{3}\left(1-\frac{1}{\theta}\right)^2 \ln(1-\theta)$	$-0.1817 \leq \tau < \frac{1}{3}$
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	$\cdot$	$0 \leq \tau < 1$

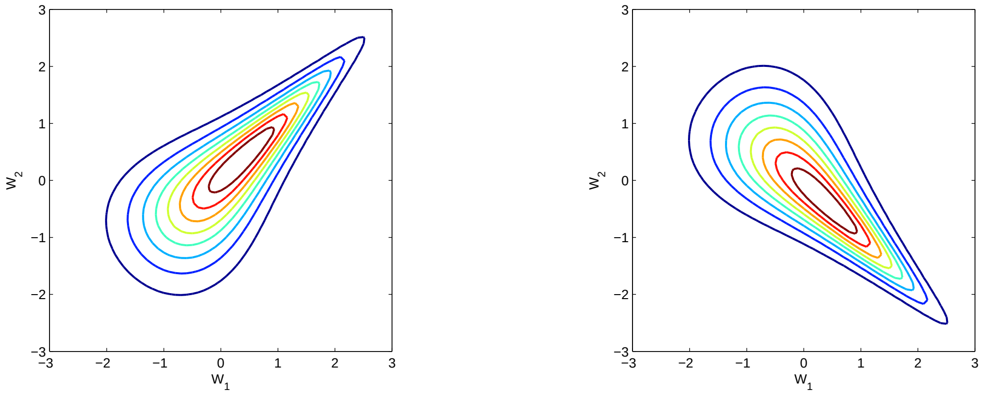
Notes: θ_{ind} is the value of θ if independent. For Frank, $D_1(\theta)$ is a Debye function: $D_1(\theta) = \frac{1}{\theta}\int_0^\theta \frac{t}{e^t-1}dt$. For Plackett Joe, there is no closed form. Equation (1.2) or Equation (1.3) is evaluated numerically.

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 1.2.

In applications, testing the independence is practically important. The fact that the independence corresponds to a specific value of the dependence parameter facilitates testing the independence. Conventional tests such as Wald, Lagrangian Multiplier (LM), and Likelihood Ratio (LR) tests can be applied with the null hypothesis: $H_0 : \theta = \theta_{ind}$. Under the null, the test statistic is distributed as $\chi^2(1)$, provided that the copula specification is treated as a "given."

Not all copulas are comprehensive as Table 1.2 shows. The Gaussian, Plackett, and Frank copulas are comprehensive, but the range of τ for the FGM and AMH copulas is limited, 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. Let (W_1, W_2) be random variables with

Figure 1.4: Illustration of Positive and Negative Dependence



(a) Joe copula: $\tau = 0.5$

(b) negative Joe copula: $\tau = -0.5$

a positive dependence τ, and let $W_2* = -W_2$. Then, (W_1, W_2^*) have a negative dependence, $-\tau$.[5] Figure 1.4 illustrates this modification with the Joe copula.

There is an additional remark on these three copulas. For these three copulas, whether in regular or negative form, independence occurs at the boundary of the range of θ. To gain the power in testing the independence, an appropriate alternative hypothesis is $H_A : \theta > \theta_{ind}$ instead of $H_A : \theta \neq \theta_{ind}$. 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.

[5]It is straightforward to show this. By the definition of Kendall's τ given by (1.1),

$$\begin{aligned} &Pr((W_1 - W_1^{'})(W_2^* - W_2^{*'}) > 0) - Pr((W_1 - W_1^{'})(W_2^* - W_2^{*'}) < 0) \\ &\quad = Pr(-(W_1 - W_1^{'})(W_2 - W_2^{'}) > 0) - Pr(-(W_1 - W_1^{'})(W_2 - W_2^{'}) < 0) \\ &\quad = -Pr((W_1 - W_1^{'})(W_2 - W_2^{'}) < 0) + Pr((W_1 - W_1^{'})(W_2 - W_2^{'}) > 0) \\ &\quad = -\tau \end{aligned}$$

1.5 Copula Selection

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.[6] When marginal distributions are fixed across copulas so that the number of parameters to estimate are equal, the selection based on the smallest information criteria is equivalent to choosing the copula attaining the largest value of the log likelihood function.

Alternatively, Vuong's (1989) test can be used to weigh one copula against another. A caveat of the Vuong's test only allows a pairwise comparison of two copulas. Chen and Fan (2005) propose the procedure for more than two candidate copulas, which is based on White (2000). Genest and Rivest (1993) and Genest et al. (2006) discuss the selection procedure based on Goodness-of-Fit tests. Alternatively, Huard et al. (2006) develop the Bayesian approach to the copula selection.

To allow for potential copula misspecification, Trivedi and Zimmer (2007a) recommend that the standard errors be estimated in robust sandwich form under the theory of quasi-maximum likelihood (White, 1982). Under the quasi-maximum likelihood, the LR test of the independence no longer has asymptotic χ^2 while the Wald and LM tests are still valid with sandwich form correction.

[6] 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.

1.6 Applications

The copula method has been widely used in the finance literature (e.g., see Cherubini et al. (2004) and references therein). It is, however, not widely used in other fields of econometrics despite its usefulness. At the end of this chapter, I illustrate the copula method in microeconometric models. All of the models are based on maximum likelihood and Stata routines I programmed are available.

1.6.1 Sample Selection Models with Continuous Outcome

One of the common econometric issues that arise in empirical studies, especially with micro-data, is a sample selection problem. Since the seminal work by Heckman (1974), sample selection models have been important econometric tools for applied researchers. A typical sample selection model consists of a selection equation and an outcome equation. where the outcome is observable only for the sub-sample of data. The model is sometimes referred to as a bivariate sample selection model or a Type 2 Tobit model. For observation i, $i = 1, ..., N$,

$$\begin{cases} s_i &= 1(z_i'\gamma + \nu_i > 0) \\ y_i &= x_i'\beta + \sigma\varepsilon_i, \end{cases}$$

where ν_i and ε_i are unobservable errors, and σ is a scaling parameter for ε_i.If the errors are independent, estimation of each equation separately causes no problem. Otherwise, however, it is necessary to take the dependence of the errors into consideration to obtain a consistent estimator. To do so, it is standard to make the assumption that ν_i and ε_i are jointly normally distributed. Instead, the copula method can be applied to allow more flexible dependence structures.

Using a copula function, the likelihood can be written as

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

When allowing a negative dependence using Clayton, Gumbel, and Joe copulas, the likelihood function can be simply written by replacing $y_i - x_i{}'\beta$ with $-(y_i - x_i{}'\beta)$. Smith (2003) proposes the copula approach to the sample selection model, and Dancer et al. (2008) and Genius and Strazzera (2008) also apply to the real data.

Another popular variant of the sample selection models is a so-called endogenous switching regression model, which is also known as a Roy model (Roy, 1951) and a Type 5 Tobit model (Maddala, 1983). The model consists of one selection equation and two outcome equations.

$$\begin{cases} s_i = & 1(z_i{}'\gamma + \nu_i > 0) \\ y_{0i} = & x_i{}'\beta_0 + \sigma_0\varepsilon_{0i}, \\ y_{1i} = & x_i{}'\beta_1 + \sigma_1\varepsilon_{1i}. \end{cases}$$

The outcome of interest is only observable in one of two possible regimes, and selection into one regime is stipulated by the selection equation. This model is frequently used in a study of policy evaluation. Treatment effects that are discussed in the literature are average treatment effect (ATE), average treatment effect on the treated (ATT), local average treatment effect (LATE), and marginal treatment effect (MTE). Heckman et al. (2003) show these treatment effect parameters can be derived from the endogenous switching regression model.

In this model, the copula method is applied to construct the dependence between ν_i and ε_{0i} and between ν_i and ε_{1i}. It is not possible to identify the dependence structure between ε_{0i} and ε_{1i} since y_{0i} and y_{1i} are not observable simultaneously for each observation. Smith (2005), however, shows that it is still possible to partially identify Kendall's τ, which is based on Vijverberg (1993).

The model can be simplified (restricted) so that only a constant term differs by regime. This model is called an endogenous dummy variable model. This is the standard model for endogenous treatment effects. This model can be formalized as

$$\begin{cases} d_i & = 1({z_i}'\gamma + \nu_i > 0) \\ y_i & = \delta d_i + {x_i}'\beta + \sigma\varepsilon_i. \end{cases}$$

The parameter δ measures the effect of the policy. Both endogenous switching regression model and endogenous dummy model are the specification used in Chapter 2.

Apart from flexible dependence structures, an advantage of the copula method is its flexibility of choosing marginal distributions. In Chapter 4, I propose using the Generalized Tukey Lambda (GTL) distribution, which is a versatile univariate distribution that permits a wide range of skewness and thickness, for marginal distributions in the sample selection models. Combined with the GTL distributions, the copula approach reduces the dependence of estimated parameters on distributional assumptions in applied research.

1.6.2 Sample Selection Models with Binary Outcome

Another application is a model with a binary-choice outcome, which is a common econometric modeling in empirical studies. The essence of the model structure is the same as the model with continuous outcomes. Researchers face endogeneity and selectivity issues. To overcome the issues, a joint distribution of error terms is modeled with the copula method.

One of the common bivariate binary choice models is

$$\begin{cases} d_i & = 1(z_i{}'\gamma + \nu_i > 0) \\ y_i & = 1(x_i{}'\beta + \varepsilon_i > 0), \end{cases}$$

The vector x_i may include d_i as an explanatory variable. In such case, d_i can be considered as an endogenous dummy variable. The literature refers to this model as a recursive model (Greene, 2008), of which structure is analogous to the endogenous dummy model. For identification, y_i cannot be included in z_i if x_i includes d_i (Wilde, 2000). If neither x_i includes d_i nor z_i includes y_i, the model is a seemingly unrelated regression (SUR) type specification. The joint estimation of two equations gain efficiency, compared to the separate estimation of each equation.

Using the copula method, the likelihood function is

$$\ln L = \prod_{i=1}^{N} \left[C(F_\nu, F_\varepsilon; \theta)\right]^{d_i=0, y_i=0} \times \left[F_\epsilon - C(F_\nu, F_\varepsilon; \theta)\right]^{d_i=1, y_i=0} \\ \times \left[F_\nu - C(F_\nu, F_\varepsilon; \theta)\right]^{d_i=0, y_i=1} \times \left[1 - F_\epsilon - F_\nu - C(F_\nu, F_\varepsilon; \theta)\right]^{d_i=1, y_i=1},$$

where $F_\nu(-z_i{}'\gamma)$ and $F_\varepsilon(-x_i{}'\beta)$. This likelihood can be simplified as

$$\ln L = \prod_{i=1}^{N} \left[y_i d_i - y_i(2d_i - 1)F_\nu - d_i(2y_i - 1)F_\varepsilon + (2y_i - 1)(2d_i - 1)C(F_\nu, F_\varepsilon; \theta)\right].$$

Winkelmann (2011) proposes the copula-based binary choice model. In his estimation, the standard normal distribution is specified as marginal distribution. As mentioned above, the copula method enables researchers to choose more flexible distribution as marginal distribution. The proposed method in Chapter 4 can be applied to this model as well.

As is the case with a continuous outcome, the selection problem may arise for a binary-choice model. I program Stata routines that implement a bivariate sample selection model and an endogenous switching regression model for a binary outcome as well as a continuous outcome. It is also useful to extend the model with other discrete choice outcomes such as ordered outcomes.

1.6.3 Sample Selection Models with Duration Outcome

Duration (or survival) analysis is also frequently applied in empirical studies of labor economics, health economics, and other applied microeconomics. As well as other analyses, the selection issue can arise in the duration analysis. For example, unemployment duration can be measured only for those who lose their jobs. Prieger (2002) applies the FGM copula to model hospitalization spells.

The likelihood function of the sample selection model with a duration outcome is quite similar to that of the standard sample selection model. Let y_i be a spell of interest and $d_i = 1(-z_i'\gamma)$ is a selection indicator. If $d_i = 1$, y_i is observed. Otherwise, it is not observable. Then,

$$\ln L = \prod_{i=1}^{N} [F_\nu(-z_i'\gamma)]^{s_i=0} \left[\left(1 - \frac{\partial}{\partial F_y} C\left(F_\nu(-z_i'\gamma), F_y\left(y_i|x_i\right); \theta\right) \right) \times f_y\left(y_i|x_i\right) \right]^{s_i=1}.$$

where $F_y(y_i|x_i)$ and $f_y(y_i|x_i)$ is a conditional cdf and pdf, respectively. For example, if an Exponential distribution, which is a common distribution in a duration analysis, is assumed, then,

$$F_y(y_i|x_i) = 1 - e^{-\lambda y} \qquad \text{and} \qquad f_y(y_i|x_i) = \lambda e^{-\lambda y},$$

where $\lambda = e^{-x_i'\beta}$. My Stata program supports distributions that are commonly used in the literature: Gamma, Log-logistic, Log-normal, and Weibull. All of these distributions are continuous distributions. Since the copula method is flexible in choosing marginal distributions, it is quite useful if the program also accommodates discrete duration distributions. It is one of my future research tasks.

As in the models outlined above, it is straightforward to extend the copula-based duration model to an endogenous switching regression model and an endogenous dummy model, of which Stata routines are available.

1.7 Concluding Remarks

The copula method is applicable not only in these models but also in many other models in microeconometrics. Cameron et al. (2004) proposes the copula method for bivariate count data. Deb et al. (2009) develops a bivariate hurdle model on panel data. Furthermore, even though the discussion of this chapter is restricted to a bivariate model, Zimmer and Trivedi (2006) develop the model with three interdependent outcomes. Prokhorov and Schmidt (2009) discuss the estimation of panel data models with copulas to capture dependence over time within individuals. Bhat and Sener (2009) apply the copula method to capture spatial dependence across observations for a binary-choice outcome.

In this dissertation, Chapters 2 and 3 apply the copula method in empirical studies, and Chapter 4 extend the sample selection models by utilizing the flexibility of the copula method. There is still a wide scope of microeconometric models where the copula method can potentially be applied. In my future research, I explore the copula method further, in both applied and theoretical works.

2 Effects of Contract Types on Starting Wage and Wage Growth

2.1 Introduction

One of the important tasks of the labor markets in industrial economies today is to balance out job stability and flexibility of the labor markets. Employment protection legislation (EPL) secures job stability of workers with permanent (indefinite-term or regular) contracts, but strict protection often creates rigidity. Labor markets in continental European countries with strict protection are characterized as rigid compared to countries with less strict protection such as the United States and the United Kingdom. High and persistent unemployment rates experienced in continental Europe are often ascribed to strict employment protection. To remedy rigidity and add flexibility to the labor markets, the labor market reforms in the European countries have made it easier to use temporary (fixed-term) employment contracts. While it is still costly for employers to dismiss workers on permanent contracts, employers are able to terminate the employment on temporary contracts easily when the contract comes to the end. The temporary contract has become a widely used form of employment, especially in the countries with stringent employment protection, as shown in Table 2.1.

The reforms on the use of temporary contracts have attracted attention from economists as well as policy-makers. The effects of temporary contract employment

Table 2.1: Strictness of Protection and the Share of Temporary Employment in 2008

	Strictness of Employment Protection[a]		Share of Temporary Employment (%)	
Country/Region	Regular Employment	Temporary Employment	All Workers	Young Workers[b]
France	2.47	3.63	14.89	52.46
Germany	3.00	1.25	14.72	56.75
Italy	1.77	2.00	13.32	43.27
Netherlands	2.72	1.19	18.17	45.17
Portugal	4.17	2.13	22.84	54.16
Spain	2.46	3.50	29.26	59.39
United Kingdom	1.12	0.38	5.43	11.99
Europe	-	-	14.6	38.7
Canada	1.25	0.25	12.29	27.16
Japan	1.87	1.00	13.62	25.98
United States[c]	0.17	0.25	4.21	8.07
OECD countries	2.11	1.77	11.95	25.06

Source: OECD statistics: `http://stats.oecd.org`

[a] The OECD indicator of employment protection on regular employment measures the procedures and costs involved in dismissing individuals employed on a regular basis. Protection on temporary employment quantifies regulation of fixed-term and temporary work agency contracts. It is scaled from 0 (least restrictive) to 6 (most restrictive). See `http://www.oecd.org/employment/emp/` for more information.

[b] Young workers are age of 15 to 24.

[c] The figures are for 2005 for United States.

have been studied extensively. Even though one of the main objectives of the labor market reforms is to reduce unemployment and render labor markets active, the effect of the use of temporary contracts may be perverse on the economy. Higher turnover rate brought by temporary contracts can lead to higher unemployment, lower productivity, and lower efficiency (Blanchard and Landier, 2002, Cahuc and Postel-Vinay, 2002).

It is also an important question whether temporary employment leads to secure permanent employment (stepping stone) or leads workers to be trapped in insecure employment (dead end) (Booth et al., 2002). Since the EPL makes it more costly to hire unproductive workers on a permanent contract, the use of a temporary contract is common in employing workers at an entry level, particularly, young workers whose productivity is relatively uncertain. The employment on a temporary contract is used as a probationary period and a screening device in employing young workers (Boockmann and Hagen, 2008). Table 2.1 shows that the share of temporary employment among young workers is twice or more as large as the share among all workers in every country. How the temporary contract impacts workers, especially young workers, is one of the important research questions.

Besides, the difference in earnings by contract type is of interest when investigating the effect of the temporary contract on individual workers. There is a large volume of literature on the wage differential, which is discussed more later. This paper adds new empirical evidence to the literature by examining the wage differentials using a dataset of new labor market entrants from the Netherlands.

Our contribution is threefold. First, we examine the wage differentials between permanent and temporary workers at the start of employment. Moreover, the dataset describes new entrants into the labor markets, as described later in Section 2.4. Therefore, our wage investigation is at the very beginning of employment career. Other studies in the literature often compare the wages at arbitrary time points. Workers become more heterogeneous at different times of their tenures and careers, which can make it difficult to interpret the empirical results. Second, we compare wage growth patterns between the two types of workers over a short period of time within the same employment, in addition to the starting wage differentials. The finding of a difference in short-term wage growth provide insights into the wage structures. Third, we relax the distributional assumption that is often made in the literature. The methodology used in this paper is not very new and it is easy to implement, but it has been rarely considered in the literature. As shown later, however, our methodology leads to considerably different resutls from the traditional approach.

This paper is organized as follows. The next section briefly describes the institutional settings of the labor market in the Netherlands. Section 2.3 provides a review on the related literature. Section 2.4 describes the dataset used in this study. Section 2.5 discusses the empirical methodology. Section 2.6 shows and discusses the estimation results, followed by the robustness checks in Section 2.7. Section 2.8 concludes the paper.

2.2 Institutional Setting

In this study, we use data from the Netherlands. Therefore, we briefly summarize the institutional settings of the labor market in the Netherlands.

As shown in Table 2.1, the employment protection on regular employment is as stringent in the Netherlands as in other continental European countries. When dismissing an employee on a permanent contract, an employer needs either (i) termination via a prior permission from the administrative authority, or (ii) judicial rescission of the contract through the court. In the procedure (i), after the permit is obtained, the employer should give the employee notice of dismissal prior to the statutory minimum notice period, which depends on the tenure of the employee. In this procedure, there is no statutory severance payment. On the other hand, the procedure (ii) requires the severance payment determined by the court, which also depends on the tenure and the age of the employee. The court decision becomes effective immediately. Even though the procedure (i) is financially less onerous because of no severance payment, it is administratively burdensome and it takes longer than the procedure (ii).

The use of temporary contracts is less restrictive in the Netherlands than in other European countries (Table 2.1). First of all, there is no restriction on reasons of the use of the contract, unlike other European countries such as Spain. The maximum number of successive contracts is three times, and a maximum duration of successive contract is three years. A fourth renewal of the contract or a renewal of the contract after three years will automatically be converted into a regular contract.[1] With relatively strict regulations on regular employment but less restriction on temporary employment, a temporary contract is a common form of employment in the Netherlands. It is widespread, especially, among young workers possibly because the contract is used as a probation device. Our dataset particularly contains workers who belong to this group.

2.3 Related Literature

Since the temporary contract became prevailing in European countries in 1980s, many researchers have investigated the effect of the contract on wages, both theoretically and empirically. In this section, I provide a brief summary of this literature.

If a labor market is perfect, the severance payment mandated by EPL and any other institutional firing costs are incorporated in the wage settings. By the amount of the severance payment, the wage of a permanent worker is cut, and then, the wage of a permanent worker is lower than the wage of a temporary worker (Lazear, 1990). The wage of a permanent worker could also be lower by compensating a temporary worker for the risk of unemployment. A worker would accept an insecure job position if the wage offer is higher than that of secure job. Even though these theories predict a positive wage differential for temporary workers, empirical studies usually find a positive wage differential for permanent workers.

Jimeno and Toharia (1993) find that temporary workers in Spain earn around 9 to 11% less than permanent workers in Spain. There are the empirical results of higher wages for permanent workers than for temporary workers from other countries as well: for example, Brown and Sessions (2003) and Booth et al. (2002) for Britain; Davia and Hernanz (2004) for Spain; de Graaf-Zijl (2012) for the Netherlands; Elia (2010) for Italy; and Hagen (2002) for West Germany. Brown and Sessions (2005)

[1] EPL is well-summarized at the website of ILO, `www.ilo.org/dyn/terminate/termmain.home`.

provide evidence from data collected in several industrial countries. Using quantile regressions, Mertens et al. (2007) also show that the wages of permanent workers are higher than the wages of temporary workers not only on average, but also across the distribution of wages.

The reason that permanent workers earn higher than temporary workers may be that workers who are likely to be permanent workers are different from those who are likely to take temporary contract jobs. Temporary contract employment is prevailing among youth, female, immigrants and the less skilled and educated (Kahn, 2007). Apart from the differences in these observable characteristics, which are controlled in empirical studies, there are several other theoretical reasons for the wage of permanent workers to be higher than the temporary workers' wages.

First, the wage differential may reflect the difference of bargaining power in wage negotiations. Since permanent workers are protected by high turnover costs, they are in a superior position to demand higher wages. On the other hand, temporary workers usually have weak bargaining power because they have little legal protection. Permanent workers as insiders have stakes only in their own wages but not in the wages of temporary workers who can be seen as outsiders (Bentolila and Dolado, 1994) or even as cheap competitors.

Second, temporary contract employment is often seen as a probationary period to screen a worker's productivity (Boockmann and Hagen, 2008). While highly productive workers are sorted into better paid and secure positions, less productive workers may be retained in temporary jobs for low wages. Repeating temporary jobs may be a bad signal of productivity.

Third, as permanent workers have low turnover rates and the employment relation is expected to last longer, both employers and the workers have incentives to invest in firm-specific human capital, which leads to wage growth. Meanwhile, temporary workers have little incentives to invest in human capital because of a short period of employment. Although training is not the only form of human capital investment, there indeed are some empirical studies showing that temporary workers are less likely to receive work-related training from firms: for example, Bentolila and Dolado (1994) and Booth et al. (2002) show empirical evidence from Britain and Albert et al. (2005) from Spain.

The theory of human capital has an implication for wage dynamics as well. As permanent workers accumulate human capital or skills, their wages increases over time. Temporary workers have limited opportunities to accumulate skills, and therefore, the wages remain stagnant. Having temporary contract can have prolonged effect. Booth et al. (2002) find that male workers who held temporary jobs at the beginning of their careers will experience lower wages than those who held permanent jobs throughout their work life.

Although there is evidence that temporary workers are less likely to receive training, this is not universal across countries. Using the European Community Household Survey, Arulampalam et al. (2004) find that in some countries such as Britain and Spain, a temporary contract is associated with a significantly lower probability of training, but this is not so in other countries such as Italy and the Netherlands. When temporary contracts are used for a probationary purpose, workers receive training. However, employers might share the benefits of training with temporary workers only when they pass the probationary period and their contracts are converted into permanent ones. Related to this idea, Amuedo-Dorantes and Serrano-Padial (2007) find that workers who had a temporary contract one year ago but remain in the same job gain considerable wage growth. As a matter of fact, almost all workers in our dataset, which will be explained shortly, receive training upon employment. Training activities can be considered as a part of the hiring process as discussed by Oi (1962). The way of passing the costs on to workers may differ by the contract type.[2]

[2]The costs of training may be different between permanent and temporary workers because of the difference in intensity of training. We do not have such the information on training intensity in

In this study, we examine the effect of the contract type on starting wages. The literature has not yet paid particular attention to the starting wage. We also investigate the wage growth patterns over a short period of time within the same job. Our findings reveal additional evidence on wage determination by the contract type.

2.4 Data

The dataset used in this study derives from the survey *Studies & Werk* by SEO Economic Research commissioned by Elsevier. Every year since 1996, SEO has surveyed a new cohort of new graduates from post-secondary schools in the Netherlands about their education, employment situation, and transition from school to work. On average, respondents answer the survey questionnaire after twenty months from graduation. Therefore, the dataset contains new entrants into the labor markets who have little, if any, work experience. Note also that these sample members have similar education attainment.

In the Netherlands, post-secondary education is divided into two levels: higher professional education (HBO) and university education (WO). HBO-education aims at preparing for specific professions and provides students with advanced vocational skills. WO-education is more academically oriented, and it is to some extent more intellectual than HBO-education. Among the cohort of 2009 in the data set, around 4,300 respondents obtained HBO degrees, and 3,700 obtained WO degrees. The figures are more or less the same across years of the survey.

This study paper uses the cohorts of the survey years from 2005 to 2009 but is restricted to those who started their jobs from 2003 to 2008. The samples exclude self-employed workers. Temporary workers include those who are employed through temporary work agencies. Note that a contract types of each worker is at the time of the survey. We are not able to identify the contract type at the beginning. Some

our dataset, and we do not explore this issue in this study.

workers may start their jobs as temporary workers at the beginning, and their contracts may be converted to permanent by the time of the survey. In the questionnaire, there is a retrospective question on gross monthly wages and hours of working at the beginning of current employment as well as wages and hours of work at the time of the survey. The hourly starting and current wages are computed from the monthly wages and hours of work and deflated by the monthly consumer price index.[3] Because of its nature of self-reported and retrospective wage information, the data are associated with some noise. To eliminate outliers, I discard the observations whose wages are in the top-1% or bottom-1% of the wage distribution. The length of employment is calculated from the years and months at the beginning of the current job and at the time of the survey and measured in months. Since a maximum duration of a temporary contract employment is three years, I exclude observations whose employment length is more than 36 months.

Table 2.2 shows that the descriptive statistics of the data. First of all, the share of permanent workers is smaller than that of temporary workers in the samples. Second, the unconditional means of starting wages seem comparable between temporary and permanent workers. The average wage of permanent workers is only slightly higher than the average wage of temporary workers. Even though the difference is small, the t test can reject the mean equality at the 1% level of significance. Third, the wage differential widens if current wages are compared. This may be because the wage growth is more faster for permanent workers than temporary workers because of the fact that the mean employment length of permanent workers is longer than temporary workers in the samples (15.1 months versus 10.6 months, respectively).

Table 2.2 also reports the descriptive statistics of individual/job characteristics that will be used as controls in the estimation. Male workers are more likely to hold permanent employment than female workers. Male and female workers tend to have

[3]The CPI data are from `www.cbs.nl`. The base year is 2006.

Table 2.2: Descriptive Statistics

	Definition	Whole Sample	TEMPORARY	PERMANENT
d	1 if permanent worker	0.423 (0.494)		
$\ln sw$	log of starting wage	2.544 (0.194)	2.537 (0.198)	2.551 (0.191)
$\ln cw$	log of current wage	2.621 (0.213)	2.593 (0.212)	2.656 (0.210)
m	months in job	12.516 (6.851)	10.630 (6.816)	15.093 (6.009)
AGE	age at the time of job start	24.687 (2.236)	24.918 (2.240)	24.371 (2.218)
GRADE	average grade at college	7.235 (0.557)	7.263 (0.569)	7.197 (0.536)
MALE	1 if male	0.433 (0.496)	0.404 (0.491)	0.473 (0.499)
WO	1 if WO degree	0.542 (0.498)	0.578 (0.494)	0.493 (0.500)
NATIVE	1 if native	0.946 (0.227)	0.943 (0.232)	0.949 (0.219)
DISABLED	1 if disabled	0.101 (0.300)	0.108 (0.311)	0.091 (0.288)
MEDIUM	1 if firm size is between 100 and 499 employees	0.227 (0.420)	0.227 (0.419)	0.227 (0.419)
LARGE	1 if firm size is 500 or more employees	0.401 (0.490)	0.399 (0.490)	0.404 (0.491)
HOME	1 if live at home with parent	0.150 (0.357)	0.144 (0.351)	0.159 (0.365)
CHILD	1 if have a child	0.031 (0.174)	0.029 (0.168)	0.034 (0.182)
Num. of Obs.		24,511	14,152	10,359
(Starting Wage)[a]		(12,785)	(6,281)	(6,504)
[Current Wage][b]		[22,318]	[12,623]	[9,695]

Standard deviations in parentheses.
[a] The number of observations whose starting wages are available.
[a] The number of observations whose current wages are available.

different employment prospects and seek job stability to different extents. For WO graduates, temporary contract jobs are more prevalent than for HBO graduates. By its vocational nature of HBO education, HBO graduates may be able to find a secured long term employment earlier than WO graduates, who may have to shop around jobs with temporary contracts.

Finally, note that not all of sample members respond to the questions on wages. Especially, the response rate of the starting wage is relatively low. The numbers of observations with observed wages are shown in Table 2.2. The respondents whose wages are missing but provide other information contribute to the likelihood function of the selection of employment contract type. The empirical strategy is outlined in the next section.

2.5 Empirical Model

2.5.1 Starting Wage

The main interest of this study is the effect of contract type on starting wages and wage growth over a short-period of time. First, we begin with the analysis of starting wages. We consider two specifications of the starting wage equation. In the first specification, the estimation equation of the log starting wage, $\ln sw_i$, for an individual i is

$$\ln sw_i = \delta d_i + \beta' x_i + \varepsilon_i, \tag{2.1}$$

where x_i is a vector of observable characteristics listed in Table 2.2 except HOME and CHILD,[4] and ε_i is an unobservable disturbance term. The vector x_i includes a constant term as well. The variable d_i is an indicator of type of contract: $d_i = 1$ if a permanent worker and $d_i = 0$ if a temporary worker. Therefore, the parameter δ measures the wage differential between a permanent worker and a temporary worker.

The selection of contract type by a worker may be presumably endogenous. We consider the following equation that stipulates the selection mechanism. The selection equation is a usual binary choice model: for individual i,

$$d_i = 1(\gamma' z_i + \nu_i > 0), \tag{2.2}$$

where $1(\cdot)$ is an indicator function, z_i is a set of observable characteristics, and ν_i is the disturbance term that affects the choice of contract-type. The vector z_i includes

[4] In addition to the variables, we also include the dummy variables for the year of starting a job and the industry of job.

all the variables listed in Table 2.2.[5] The problem of selectivity arises if the errors ν_i and ε_i are not independent of each other. It is well known that unless the errors are independent, OLS regression of the wage equation results in a biased and inconsistent estimator (Heckman, 1979). To estimate the equation (2.1) consistently, we estimate the starting wage equation (2.1) and the selection equation (2.2) jointly by allowing the dependence between ν_i and ε_i. To do so, it is standard to assume the joint normal distribution of ν_i and ε_i.[6] Then, the outcome equation (2.1) and the selection equation (2.2) are jointly estimated by maximum likelihood approach. The variables HOME and CHILD serve as exclusion restrictions since these variables are included in the selection equation but not in the wage equation.

Even though the assumption of the joint normality is commonly made in many of empirical studies, it is a relatively strong assumption. The violation of distributional assumption generally leads to an inconsistency of maximum likelihood estimator. The wage distribution conditional on observables may have thicker tails than the normal distribution assumes. To accommodate the possibility of thicker tails, we assume that the disturbance terms in both equations have a Student's t distribution. With smaller degrees of freedom, the t distribution has thicker tails than the normal distribution, and as the degrees of freedom approach infinity, the t distribution converges to the normal distribution. To allow the dependence between the errors marginally distributed as the t distribution, we follow the approach proposed by Lee (1982, 1983). The nonnormal disturbances ν_i and ε_i are transformed into normal variates, and then, the transformed variates are assumed to be jointly normally distributed.

[5] As in the wage equation, the dummies for year and industry are included.

[6] The joint distribution of the error terms is

$$\begin{pmatrix} \nu \\ \varepsilon \end{pmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho\sigma_\varepsilon \\ \rho\sigma_\varepsilon & \sigma_\varepsilon^2 \end{bmatrix} \right),$$

where σ_ε^2 is the variance of ε_i and ρ is the coefficient of correlation between ν_i and ε_i. The variance of ν_i is set to be one as identification.

More specifically, the joint distribution of ν_i and ε_i, $F(\nu_i, \varepsilon_i)$, is now expressed as

$$F(\nu_i, \varepsilon_i) = \Phi_2(\Phi^{-1}(F_\nu(\nu_i)), \Phi^{-1}(F_\varepsilon(\varepsilon_i)); \rho),$$

where $\Phi(\cdot)$ is the cumulative distribution function (cdf) of standard normal distribution, and $\Phi_2(\cdot)$ is the cdf of standard bivariate normal distribution with the coefficient of correlation ρ. $F_\nu(\cdot)$ and $F_\varepsilon(\cdot)$ are the cdf's of ν_i and ε_i, respectively, which are now assumed to be the t distributions with r_ν degrees of freedom and r_ε degrees of freedom, respectively. The degrees of freedom r_ν and r_ε are treated as parameters to be estimated along with other parameters of the model, and they are allowed to be different for each marginal distribution.[7]

Besides the different distributional assumptions, we also consider a different specification of the starting wage equation. Equation (2.1) (Specification 1, hereafter) assumes that the wage differential stems solely from the difference in the intercept, which is captured by δ. This is a common specification in previous empirical studies. However, it states that the economic returns to observable characteristics such as education are the same between the two types of workers. It is more plausible to assume that the returns are different by the type of contract. To allow the differences in the returns, we separate the outcome equation into two equations, based on the contract type:

$$\begin{cases} \ln sw_{0i} = \beta_0{}' x_i + \varepsilon_{0i} & \text{if } d_i = 0 \\ \ln sw_{1i} = \beta_1{}' x_i + \varepsilon_{1i} & \text{if } d_i = 1 \end{cases}. \tag{2.3}$$

For an individual i, the starting wage is either $\ln sw_{0i}$ or $\ln sw_{1i}$, depending on whether $d_i = 0$ or $d_i = 1$; thus, $\ln sw_i = (1 - d_i) \cdot \ln sw_{0i} + d_i \cdot \ln sw_{1i}$. As in Specification 1, we consider the endogeneity of the contract type by the selection equation (2.2).

[7]Instead of this approach, we might assume that (ν_i, ε_i) follows a bivariate t distribution. One limitation of the assumption of the bivariate t distribution is that the marginal distributions of the errors are the same. That is, the degrees of freedom of ν_i and ε_i are restricted to be the same. Our approach relaxes this restriction so that each disturbance has different degrees of freedom.

This model is known as an endogenously switching regression model.[8] We call this Specification 2.

In order to estimate this specification, we need to consider the joint distribution of the error terms in the two wage equations and the selection equation. We extend the previous bivariate distribution assumptions to the trivariate case: The error terms or normally transformed errors are distributed as a trivariate normal distribution[9]

The parameter of primary interest in this study is how much wages differ between temporary workers and permanent workers on average. This quantity is referred to as the average treatment effect in the policy evaluation literature. The average treatment effect is $ATE = E(\ln sw_1 - \ln sw_0)$. Following Heckman et al. (2003), it can be estimated with estimated coefficients as

$$\widehat{ATE} = (\widehat{\beta}_1 - \widehat{\beta}_0)'\bar{x}, \tag{2.4}$$

where $\bar{x}_i$ is the sample means of observable characteristics. If the quantity is positive, the starting wage is higher for permanent workers than temporary workers on average.

2.5.2 Wage Growth

The current wage equations are estimated in the similar manner to the starting wage equations. That is, we consider the two specifications of the equations, and two different distributional assumptions for each specification. In addition to the observable characteristics in the starting wage equation, we add the length of the job (measured

[8]It is also known as a Roy model (Roy, 1951) or a Type-5 Tobit model (Amemiya, 1985).

[9]The trivariate normal distribution is :

$$\begin{pmatrix} \nu \\ \varepsilon_0 \\ \varepsilon_1 \end{pmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho_0\sigma_{\varepsilon_0} & \rho_1\sigma_{\varepsilon_1} \\ \rho_0\sigma_{\varepsilon_0} & \sigma^2_{\varepsilon_0} & \rho_{01}\sigma_{\varepsilon_0}\sigma_{\varepsilon_1} \\ \rho_1\sigma_{\varepsilon_1} & \rho_{01}\sigma_{\varepsilon_0}\sigma_{\varepsilon_1} & \sigma^2_{\varepsilon_1} \end{bmatrix} \right),$$

where ρ_0 and ρ_1 are the coefficients of correlation between ν_i and ε_{0i} and ε_{1i}, respectively. Since only the starting wage of either permanent worker or temporary worker is observed for each observation, ρ_{01}, which is the coefficient of correlation between ε_{0i} and ε_{1i}, cannot be identified in this model.

in months) into the current wage equations. The coefficient on this variable measures the wage growth over time. We are particularly interested in whether and how wages grow differently between a permanent worker and a temporary worker.

For Specification 1, the current wage $\ln cw_i$ is

$$\ln cw_i = \mu m_i + \theta d_i \cdot m_i + \delta d_i + \beta' x_i + \xi_i, \tag{2.5}$$

where m_i is the tenure in months, and the coefficient on the interaction term, θ, measures how the wage growth varies between temporary and permanent workers. For Specification 2, the current wage for each type of contract is

$$\begin{cases} \ln cw_{0i} = \mu_0 m_i + \beta_0{}' x_{0i} + \xi_{0i} & \text{if } d_i = 0 \\ \ln cw_{1i} = \mu_1 m_i + \beta_1{}' x_{1i} + \xi_{1i} & \text{if } d_i = 1 \end{cases}. \tag{2.6}$$

Our interest in this specification is whether and to what degree the coefficients μ_0 and μ_1 are different.

As in the starting wage equation, we incorporate the selection mechanism by allowing the correlation between ξ_i and ν_i. Since the selection equation is already estimated jointly with the starting wage equation, we use those estimates. We also restrict the coefficients on observable characteristics as the estimates obtained from the starting wage estimation. In other words, at this stage of estimation, we estimate only the coefficients μ and θ for Specification 1 and μ_0 and μ_1 for Specification 2 of the wage equations.[10] For each specification of the current wage equation, we consider different distributional assumptions as for the starting wage equation. See Appendix 2.A for more detailed explanation of the estimation procedure.

Note that as mentioned in Section 2.4, we do not know exactly which type of

[10]Instead of this two-step estimation procedure, we may estimate all the equations of selection, starting wage, and current wages jointly. By not doing so, we do not take into account a possible correlation between the error disturbances of the starting wage and current wage equations, ε and ξ. Although ignoring the correlation loses efficiency, it does not affect consistency.

contract each worker held at the beginning. To be accurate, the starting wage differential that we estimate here measures the difference in starting wages between workers currently employed on permanent contract and workers currently employed on temporary contracts. Wage growth that we estimate may reflect a change in contract type that happened by the time of the survey.

2.6 Results

Using the data set in Section 2.4, we estimate the empirical models outlined in the preceding section. We present the estimation results of the selection equation, the starting wage equations, and the current wage equations, separately. As outlined in Section 2.5, we estimate two specifications of the wage equations. Here we report only the results from Specification 2. Appendix 2.B presents the results from Specification 1. Specification 1 is a nested (restricted) model of Specification 2. Under each of the distributional assumptions, both Wald and Likelihood Ratio (LR) tests can reject the restrictions at any conventional level of significance, which indicates that Specification 2 is preferred.[11] We will briefly discuss the evidence from Specification 1 later on.

Table 2.3 presents the estimation result of the selection equation (2.2). For the model under the assumption of the t distribution, the degree of freedom is a parameter to be estimated and is reported as r. The estimated value of 0.2987 implies that the distribution of ν has much thicker tails than the standard normal distribution, with even the first moment undefined. When the underlying distributions of a binary choice model are different, the coefficients are not directly comparable. To make the estimates from different distributions comparable, the coefficients reported in the table are standardized by a half of the interquartile range (IQR) of the distribution. For the standard normal distribution, the IQR is 1.349 while the IQR for the t distribu-

[11] Under the normal distribution, the Wald and LR test statistics are -141.97 and -163.21, respectively, with 24 degrees of freedom. Under the t distribution, the Wald and LR test statistics are -152.17 and -187.81, respectively, with 25 degrees of freedom.

Table 2.3: The Estimated Selection Equation[a]

	Normal Distribution		t Distribution	
	Coefficient[b]	Marginal Effect[c]	Coefficient[b]	Marginal Effect[c]
AGE	-0.1141 ***	-0.0281 ***	-0.0889 ***	-0.0316 ***
	(0.0069)	(0.0017)	(0.0173)	(0.0019)
GRADE	-0.0896 ***	-0.0221 ***	-0.0258 *	-0.0092
	(0.0230)	(0.0057)	(0.0155)	(0.0061)
MALE	0.2292 ***	0.0568 ***	0.1879 ***	0.0686 ***
	(0.0268)	(0.0066)	(0.0440)	(0.0100)
WO	-0.1529 ***	-0.0379 ***	0.0325	0.0115
	(0.0288)	(0.0072)	(0.0290)	(0.0093)
NATIVE	-0.0141	-0.0035	-0.0374	-0.0134
	(0.0550)	(0.0136)	(0.0386)	(0.0136)
DISABLED	-0.1344 ***	-0.0329 ***	-0.0526 **	-0.0185 **
	(0.0412)	(0.0100)	(0.0258)	(0.0089)
MEDIUM	0.0463	0.0114	0.0557 **	0.0197 **
	(0.0334)	(0.0082)	(0.0258)	(0.0085)
LARGE	0.0548 **	0.0135 *	0.1695 ***	0.0614 ***
	(0.0292)	(0.0072)	(0.0549)	(0.0143)
HOME	-0.2398 ***	-0.0583 ***	-0.1144 ***	-0.0393 ***
	(0.0380)	(0.0091)	(0.0359)	(0.0097)
CHILD	0.4797 ***	0.1198 ***	0.1702 ***	0.0609 ***
	(0.0727)	(0.0181)	(0.0665)	(0.0236)
CONSTANT	3.6845 ***		2.4364 ***	
	(0.2580)		(0.4304)	
r			0.2987	
			(0.0667)	

White (1982)-type robust standard error in parentheses. ***,**,* indicate statistical significance at 1%, 5%, and 10% levels, respectively.

[a] The selection equation also contains the dummies for regions of residence, industries, and years of starting job.

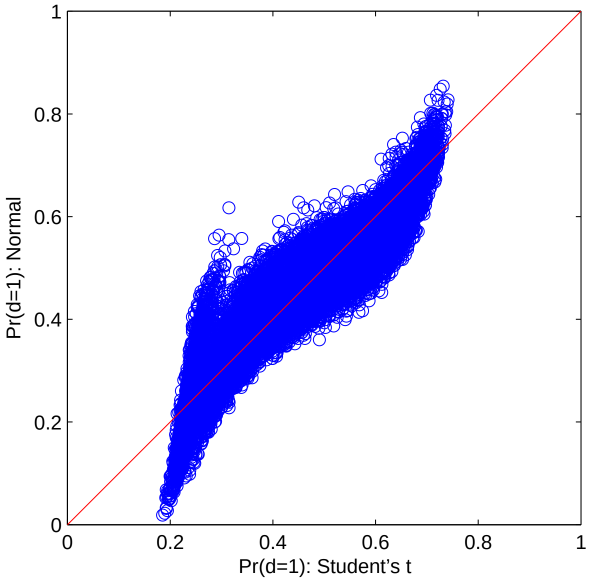
[b] The estimated coefficients are standardized by a half of the interquartile range of the underlying distribution.

[c] The marginal effects are shown in percentage points. See Footnote 12 for the computation of the marginal effects.

tion with 0.2987 degrees of freedom is 6.101. The standardized coefficients measure how much the latent variable that determines the selection outcome changes as the corresponding variables change by a half of IQR points. We also report the marginal effects on the probability of being a permanent worker, which can be easily compared and interpreted.[12]

[12] Marginal effects are evaluated for each observation and then averaged across all observations. The standard errors are computed by the Delta method. For continuous variables, the effects are computed by a calculus method. For discrete variables, the effects are the differences of the probabilities when the variable changes from 0 to 1, with other categories equal 0 if categorical

Figure 2.1: Predicted Probability $Pr(d_i)$: Student's t vs Normal



The considerable difference in the underlying distributions results in noticeable differences in the estimated coefficients. Under the normal distribution assumption, a WO degree decreases the probability of being a permanent worker statistically significantly. Under the t distribution assumption, on the other hand, the degree increases the probability though the effect is insignificant. Even for the variables with the same sign and significance levels, the magnitudes of the marginal effect considerably differ. For instance, the effect of CHILD under the normal distribution is almost twice as large as the effect under the t distribution. Besides the marginal effects, the predicted probabilities of being a permanent worker are also considerably different under the different distributional assumption. Figure 2.1 shows the scatter plot of the predicted probabilities. The mean absolute difference in the predicted probabilities is 4.14 percentage points, with the maximum difference of about 30 percentage points.

Table 2.4 reports the estimated starting wage equations of Specification 2. The results under each distributional assumption are presented. First of all, the estimated

variables.

Table 2.4: The Estimated Starting Wage Equation[a]: Specification 2

	Normal Distribution		t Distribution	
	TEMPORARY	PERMANENT	TEMPORARY	PERMANENT
AGE	0.0154***	0.0155 ***	0.0112***	0.0141***
	(0.0012)	(0.0013)	(0.0013)	(0.0013)
GRADE	0.0264***	0.0293 ***	0.0247***	0.0273***
	(0.0043)	(0.0038)	(0.0041)	(0.0036)
MALE	0.0248***	0.0282 ***	0.0365***	0.0299***
	(0.0047)	(0.0046)	(0.0048)	(0.0045)
WO	0.1093***	0.1344 ***	0.1119***	0.1353***
	(0.0051)	(0.0048)	(0.0049)	(0.0045)
NATIVE	0.0086	-0.0085	0.0068	-0.0101
	(0.0102)	(0.0099)	(0.0099)	(0.0092)
DISABLED	-0.0021	-0.0119	-0.0120*	-0.0137**
	(0.0069)	(0.0073)	(0.0069)	(0.0068)
MEDIUM	0.0248***	0.0053	0.0270***	0.0096**
	(0.0057)	(0.0058)	(0.0056)	(0.0055)
LARGE	0.0381***	0.0347 ***	0.0483***	0.0390***
	(0.0053)	(0.0049)	(0.0053)	(0.0047)
CONSTANT	1.8055***	1.8792 ***	2.0277***	1.9009***
	(0.0821)	(0.0512)	(0.0917)	(0.0472)
ρ	-0.2872***	-0.2418 ***	0.4179***	0.0129
	(0.0451)	(0.0509)	(0.0857)	(0.0559)
r			7.4228	6.4395
			(0.6443)	(0.4739)
σ	0.1755	0.1628	0.1553	0.1339
	(0.0022)	(0.0022)	(0.0051)	(0.0022)
$\ln L_1$	-10794.431		-10592.211	

White (1982)-type robust standard error in parentheses. ***,**,* indicate statistical significance at 1%, 5%, and 10% levels, respectively.

[a] The starting wage equation also contains the dummies for regions of residence, industries, and years of starting job.

degrees of freedom of the t distributions indicate that the error terms have much thicker tails than the normal distribution as in the selection. The Wald test shows that the difference is statistically significant. The information criteria, AIC and BIC, also indicate that the assumption of the t distribution is preferred to the normal distribution.[13]

[13]The AIC (BIC) under the normal distribution are 21738.861 (22346.877), respectively while the AIC and BIC under the t distribution are 21340.423 (21972.759), respectively. Although the t distribution becomes the normal distribution as the degree of freedom goes to infinity, these distributions are virtually the same when the degree of freedom exceeds 120. This threshold value makes it straightforward to test whether the distributions are equal. Clearly, the estimated degrees of freedom are significantly different from this threshold value.

Table 2.5: The Estimated Starting Wage Differentials

Naive[a]		Independence	Normal Distribution	t Distribution
0.0142***	Specification 1	0.0275***	0.1069***	-0.0157
(0.0034)		(0.0031)	(0.0126)	(0.0136)
	Specification 2	0.0265***	0.0983***	-0.0275
		(0.0074)	(0.0132)	(0.0176)

Robust Standard errors in parentheses. ***,**,* indicate statistical significance at 1%, 5%, and 10% levels, respectively.

[a] The naive estimator is the difference in sample means of each contract type workers.

As is in the case of the selection equation, the different assumption of the underlying distributions lead to the different estimation results. The most noticeable difference is the signs of the coefficients of correlation. The result under the normal distribution indicate significantly negative dependence between the errors of the starting wage equations. On the other hand, the result under the t distribution shows that while the dependence between the error of the permanent worker wage equation and the error of the selection equation is positive but not statistically significant, the error of the temporary worker wage equation and the error of the selection equation exhibit statistically significant and relatively strongly positive dependence. This may be because everything else equal, a highly ability worker is more likely to have a permanent contract, and even if the worker ends up to be a temporary worker, the worker tends to receive a higher wage.

The estimated coefficients under each of the distribution share the same signs. Even though the signs are the same, the significance levels are different for some variables and the estimated values are slightly different. The differences in the coefficients are more visible when we compare the average wage differentials as Equation 2.4, which is of primary interest of this study.

Table 2.5 summarizes the estimated starting wage differential. For comparison, we report the estimates from both specifications and both distributional assumptions. In Specification 1, the coefficient δ measures the wage differential. In addition, we also presents the results under the assumption of no selectivity into the contract type. We assume the independence of the error terms between the wage equations and the selection equation, which is equivalent OLS regression of each wage equation. The naive estimator of the wage differential is the difference in the sample means, and the independence cases control observable characteristics.

Each specification under the same distributional assumption yields similar results while different distribution assumptions lead to different results. When the selectivity is not considered, the starting wage is slightly but statistically significantly higher for a permanent worker than a temporary worker.[14] Once the selectivity is considered, the estimated wage differential rises under the normality assumption. The result says that upon the time of employment, permanent workers already earn more than temporary workers by approximately 10%. Interestingly, the wage differential becomes negative if we assume the t distribution. Even though it is only slightly lower than zero and not significantly different from zero based on a two-tail test, a one-tail test can reject the null hypothesis that the starting wage of permanent workers is greater than temporary workers at the 10% significance level.

The different distributional assumptions lead to different estimation results of the starting wage equations. As mentioned earlier, we can statistically discriminate these two distributional assumptions in favor of the t distribution. Therefore, we see the estimation results under the t distribution as our reliable findings.

Table 2.6 shows the estimation results of the current wage equations. The results reveal that the wage growth patterns are different across the two types of contracts, which is statistically significant based on the Wald test. As expected, the wage of permanent workers increases more rapidly than that of temporary workers. The

[14] In addition to what reported in Table 2.5, we also estimate the starting wage differential by the propensity score method (Rosenbaum and Rubin, 1983). The estimated value is 0.0286 with the standard error of 0.0034, which indicates statistical significance. The value is somewhat close to the independent case. This propensity score method is based on "selection on observables." See, for example, Imbens and Wooldridge (2009). Our estimation of the joint estimation of the selection and wage equations is based on "selection on unobservables."

Table 2.6: The Estimated Current Wage Equation: Specification 2

	Normal Distribution				t Distribution			
	TEMPORARY		PERMANENT		TEMPORARY		PERMANENT	
m	-0.0007		0.0046	***	-0.0015	***	0.0042	***
	(0.0005)		(0.0006)		(0.0005)		(0.0004)	
ρ	-0.6430	***	0.0217		-0.0951	**	0.2653	***
	(0.0320)		(0.0452)		(0.0477)		(0.0339)	
r					5.6564		4.3292	
					(0.3028)		(0.1758)	
σ	0.2225		0.1841		0.1614		0.1422	
	(0.0030)		(0.0019)		(0.0029)		(0.0021)	
$\ln L_2$	-10366.952				-9783.444			

Standard error in parentheses. ***,**,* indicate statistical significance at 1%, 5%, and 10% levels, respectively. The estimates are obtained, conditional on $\widehat{\gamma}$, $\widehat{\beta}_0$ and $\widehat{\beta}_1$ as reported in earlier tables. See Appendix 2.A for the computation of the standard errors.

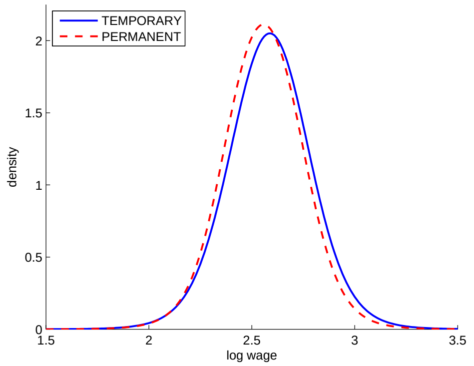
wage of temporary workers indeed decreases, which is statistically significant under the t distribution. Compared to the starting wage equations, the estimated scale parameters σ of the current wage equations are larger. The estimated values of r are also larger. These findings indicate that the current wage equations have larger variance than the starting wage equations.

Our findings say that a permanent worker will start with a lower wage than a temporary worker on average, but the wage grows faster for a permanent worker than a temporary worker. A worker will accept a lower starting wage in prospect of a faster wage growth (Connolly and Gottschalk, 2008). A permanent worker has stronger incentives to acquire firm-specific human capital than a temporary worker because of an expected longer employment relation. A temporary worker receives a higher starting wage, but the wage hardly grows because of little accumulation of skills.

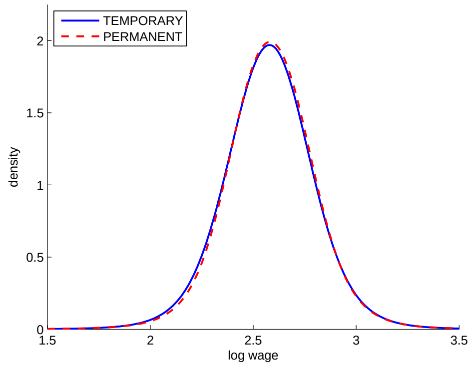
From a firm's perspective, this wage scheme is also reasonable. A firm needs to recoup the costs that it incurs during a hiring process, which can be seen as a form of investment. A firm reaps the benefits by not raising the wage for a temporary worker during a short employment period, whereas it reaps and shares the benefits with

Figure 2.2: The Estimated Distributions of Wage at Different Time of Employment

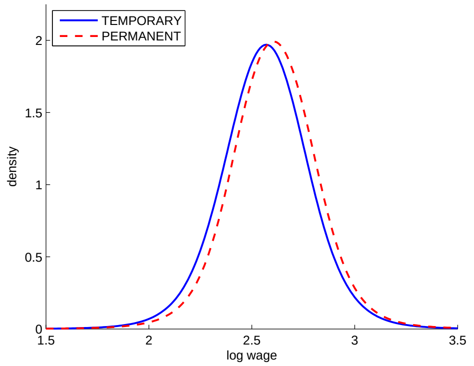
(a) At the Beginning



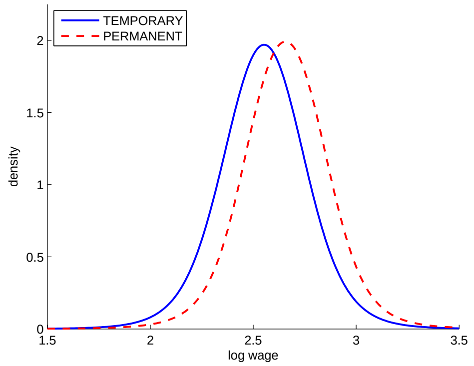
(b) At 6 Months



(c) At 12 Months



(d) At 24 Months



permanent workers gradually by raising wages. Sharing the benefits will reduce the likelihood of separation of employees and will build a longer employment relationship.

A back-of-the-envelope calculation says that the wage of a permanent worker catches up with the wage of a temporary worker in approximately 5 months, and in 24 months since the time of employment, a permanent worker receives about a 10% higher wage than a temporary worker. The cumulative wage earnings of a permanent worker exceed those of a temporary worker after about 11 months. Figure 2.2 also presents the comparisons of wage distributions between temporary workers and permanent workers at different time points of employment.[15] As we can see from

[15]The wage distributions are obtained by calculating the (log) wage densities at different points. In turn, the density is computed using the estimated parameters and coefficients. For each contract

Figure 2.2a, at the time of employment, the wage distribution of a permanent worker is slightly below that of a temporary worker. At six months, the two distributions almost overlap with each other. At a year, the wage distribution of a permanent worker slightly exceeds that of a temporary worker, and the gap widens in two years. Our finding of the lower starting wages of permanent worker does not conflict with the findings of the empirical literature. The higher wages of permanent workers found in the previous studies might reflect the wage growth patterns we find in this study.

2.7 Robustness Checks

In this section, we check robustness of the results in the previous section. Unless stated otherwise, we estimate the model of Specification 2 under the t distribution, which is superior to other models.

First, we examine the specification of the current wage equations. The benchmark model (2.6) assumes a linear effect of months in jobs: the wage increases (decreases) proportionally to the months in job. However, nominal wages usually do not change as frequently as monthly. As a matter of fact, reported current wages and starting wages are equal for a large portion of the observations. In such a case, the changes in real wages are due to the changes in the monthly price index. Wages may grow in a more complicated way than the linearity simply presumes. To allow a nonlinear

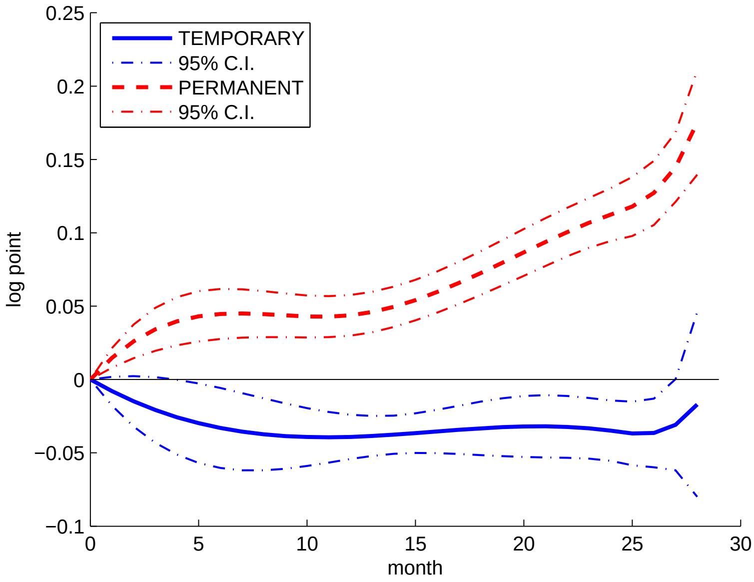
---

type, d, the density of wage, w_d, is

$$\widehat{f(w_d)} = \int \widehat{f}_d(w_d|x)dF_X(x) \approx N^{-1}\sum_{i=1}^{N}\widehat{f}_d(w_d|x_i)$$

where $w_d = \widehat{\beta}_d' x_i$ at the time of job start and $w_d = (\widehat{\mu}_d m + \widehat{\beta}_d' x_i)$ at m months. $\widehat{f}_d(\cdot)$ is the pdf of the t distribution with estimated degree of freedom and scale parameter σ, which are reported in Table 2.4 at the time of job start and in Table 2.6 at later time points. We evaluated the density for every 0.01 point from $w_d = 1.5$ to $w_d = 3.5$.

Figure 2.3: Wage Growth over Months:
Temporary Worker vs Permanent Worker



effect of job tenure, we consider the following specification: for $d = 0$ and 1,

$$\ln cw_{di} = g_d(m_i) + \beta_d{}' x_i + \xi_{di},$$

where $g_d(m_i)$ is a cubic b-spline function of m_i with knots at $m_i = 12$ and $m_i = 24$.[16]

The estimation results are presented graphically in Figure 2.3. To see how differently wage grows between a temporary worker and a permanent worker, we set the starting wages equal. Figure 2.3 shows the wage growth patterns are not linear as we suspect. While the (real) wage of a temporary worker declines at the beginning and stay the same later on, the wage of a permanent worker rises gradually, in particular after the first 12 months. Even though this figure shows nonlinear wage

[16]More specifically, this spline function is expressed as

$$g_d(m_i) = \mu_{d1} m_i + \mu_{d2} m_i^2 + \mu_{d3} m_i^3 + \mu_{d4} m_i^3 \times 1(m_i > 12) + \mu_{d5} m_i^3 \times 1(m_i > 24),$$

for $d = 0$ and 1, and where $1(\cdot)$ is an indicator function. Although this choice of the knots is arbitrary, we also tried different points of knot and different number of knots. The results are similar to the presented result.

Table 2.7: Estimation Results from Different Measures of Wage

	Real Term		Nominal Term	
	Hourly	Monthly	Hourly	Monthly
Starting Wage Differentials				
	-0.0275	-0.0331 **	-0.0242	-0.0354 **
	(0.0176)	(0.0146)	(0.0170)	(0.0146)
Wage Growth Coefficients				
TEMPORARY	-0.0015 ***	0.0004	-0.0005	0.0013 *
	(0.0005)	(0.0007)	(0.0005)	(0.0007)
PERMANENT	0.0042 ***	0.0050 ***	0.0050 ***	0.0060 ***
	(0.0004)	(0.0004)	(0.0004)	(0.0004)

Standard error in parentheses. ***,**,* indicate statistical significance at 1%, 5%, and 10% levels, respectively.

growth patterns, this result still shares an implication with the previous result of the simple linear specification. The wage of a permanent worker increases while that of a temporary worker slightly declines. These growths are statistically significant as the 95% confidence intervals indicate.[17]

This exercise then raises a question on the measurement of the wage. In the estimation so far, the wage is measured in real hourly wage, which is computed by dividing nominal monthly wage by hours of work per month and the monthly price index. Apart from a change in a nominal wage, a real hourly wage changes over time due to a change in the price index and a change in the hours of work. Different wage measures may yield different wage patterns. Furthermore, in the samples, on average permanent workers have held their jobs longer than temporary workers. Therefore, the starting wages of permanent workers may be discounted more than temporary workers. To investigate this measurement issue, we estimate with four different measures of wage: hourly or monthly and in real term or in nominal term.

Table 2.7 summarizes the results of this exercise. The first column corresponds to the result that is previously shown using the real hourly wage. The different

[17]The 95% confidence intervals are based on the standard errors computed by the Delta method.

wage measures yield different results, in particular, about the wage growth coefficient of a temporary worker. Although the wage growth is significantly negative for a temporary worker when measured in real term per hour, it is significantly positive when the nominal monthly wage is used. From the comparisons between the results from hourly and monthly wages, we can also see that changes in hours of work matter for the wage growth patterns as well.

Despite slightly different results, all the results share the important finding. The starting wage of permanent workers is lower but their wage growth is faster than that of temporary workers. For each measure of wage, we can reject the null of the equality of wage growth coefficients of temporary and permanent workers. The starting wage differentials are negative in all the measures and at least marginally significant at a one-tail test.

The next thing to check is the gender differences in starting wage differentials and wage growth patterns. As shown in Table 2.3 and Table 2.4, females are less likely to be permanent workers and receive lower wages than males. As Booth et al. (2002) argue, female workers and male workers may have different career prospects and have different attitudes toward temporary contracts. They also show temporary contracts affect wages differently by gender.

Therefore, we re-estimate the model separately by gender, and the results are presented in Table 2.8. The LR and Wald tests can reject the null hypothesis that all the parameters in the model are equal between genders. As obvious in Table 2.8, the determinants of the contract type are different by gender. While AGE matters for female workers, it does not for male workers. The WO degree have opposite effects on the selection of contract type by gender.

Besides these differences, what is interesting for us is the effect of having children. It is plausible to think that a female worker with children will choose to work at a flexible contract. The result, however, shows that having children is significantly

Table 2.8: The Estimation Results by Gender

	FEMALE			MALE		
	Selection Eq.[a] ME.[b]	Starting Wage Eq.[a] TEMP.	PERM.	Selection Eq.[a] M.E.[b]	Starting Wage Eq.[a] TEMP.	PERM.
AGE	-0.036***	0.011***	0.017***	-0.021	0.012***	0.013***
	(0.005)	(0.002)	(0.002)	(0.016)	(0.002)	(0.002)
GRADE	-0.022***	0.022***	0.022***	0.001	0.028***	0.031***
	(0.003)	(0.006)	(0.005)	(0.001)	(0.006)	(0.005)
WO	-0.052***	0.101***	0.132***	0.053***	0.112***	0.139***
	(0.009)	(0.007)	(0.007)	(0.016)	(0.008)	(0.006)
NATIVE	-0.030	-0.006	-0.008	-0.004	0.023*	-0.008
	(0.018)	(0.012)	(0.012)	(0.016)	(0.017)	(0.014)
DISABLED	-0.029**	-0.015	-0.013	-0.026***	-0.008	-0.014
	(0.013)	(0.009)	(0.010)	(0.010)	(0.010)	(0.010)
MEDIUM	0.004	0.024***	0.005	0.022**	0.027***	0.016**
	(0.011)	(0.007)	(0.008)	(0.010)	(0.009)	(0.008)
LARGE	0.001	0.037***	0.037***	0.066***	0.048***	0.042***
	(0.011)	(0.007)	(0.007)	(0.016)	(0.008)	(0.007)
HOME	-0.037***			-0.031***		
	(0.013)			(0.011)		
CHILD	0.124***			0.050***		
	(0.029)			(0.020)		
CONSTANT		1.967***	1.867***		2.044***	1.938***
		(0.109)	(0.070)		(0.129)	(0.064)
ρ		0.543***	-0.069		0.102	0.049
		(0.082)	(0.076)		(0.118)	(0.016)
r	1.342	8.519	6.903	0.253	6.290	6.069
	(0.628)	(1.159)	(0.751)	(0.095)	(0.702)	(0.668)
σ		0.167	0.141		0.140	0.126
		(0.007)	(0.003)		(0.004)	(0.003)
$\ln L_1$		-6243.9689			-4086.4057	
		Current Wage Equation			Current Wage Equation	
m		0.000	0.004***		-0.003	0.005***
		(0.001)	(0.001)		(0.002)	(0.001)
ρ		0.006	0.183***		-0.385***	0.278***
		(0.063)	(0.196)		(0.085)	(0.057)
r		5.754	4.223		5.708	4.531
		(0.394)	(0.231)		(1.199)	(0.293)
σ		0.163	0.143		0.166	0.139
		(0.003)	(0.003)		(0.017)	(0.003)
$\ln L_2$		-5753.131			-3738.942	
Num. of Obs.		13,893			10,618	

Standard error in parentheses. ***,**,* indicate statistical significance at 1%, 5%, and 10% levels, respectively.

[a] The selection equation and starting wage equations also contains the dummies for regions, industries, and years of starting jobs.

[b] M.E. stands for marginal effect. See the footnote of Table 2.3 for the computation of marginal effect and its standard error.

associated with higher probability of being a permanent worker. Moreover, the effect is bigger for a female worker than a male worker. A possible explanation of this finding is that female workers with stability of employment can choose to bear a child whereas females with instability jobs find it more difficult to bear a child.[18] This, then, raises a concern about the exogeneity of this variable in the selection equation.

Likewise, there is a potential endogeneity of the variable HOME. We argue that a worker living with parents find it affordable to shop around for jobs until the worker finds a right job by accepting a temporary contract job. Instead, the worker with a temporary job may be forced to secure her stability by living with parents until she finds a secure job.

To avoid this endogeneity problem of the variables used as exclusion restrictions, we estimate the model without these variables. That is, our estimation is now without exclusion restrictions. However, it raises another econometric concern. Even though in theory the parametric sample selection model can be identified without exclusion restrictions, the estimator tends to perform poorly. This is mainly due to the collinearity between regressors and the selectivity correction term called the inverse Mill's ratio. Although our estimation is based on the maximum likelihood rather than the two-step estimation using the inverse Mill's ratio, the problem may also be serious in the maximum likelihood estimation (Leung and Yu, 1996). See also Puhani (2000) for the survey of this issue. To see how serious this issue is in our approach with the t distribution, we conduct simple Monte Carlo simulations. The simulation results indicate that the issue is not as severe as in the case of the normality assumption.[19]

[18] In the Netherlands, pregnancy is a prohibited to dismiss workers. Pregnant workers are entitled to pregnancy and maternity leave, and they cannot be dismissed during a maternity leave and within the first six weeks after a maternity leave. When the contract of workers expires during pregnancy, the workers will receive pregnancy and maternity benefits during 16 weeks from the social security agency.

[19] The simulation setting is as follows. For 5,000 observations, we generate the selection equation and two potential outcome equations.

$$d_i = 1(-0.5 + x_i - z_i + \nu_i > 0), \qquad y_{0i} = 1 + 0.5x_i + \varepsilon_{0i}, \qquad y_{1i} = 1.5 + 1x_i + \varepsilon_{1i},$$

where x_i is drawn from a uniform distribution from 0 to 1, and z_i is a binary variable from a Bernoulli distribution with p=0.5. In this setting, the average treatment effect is 1. For the normal distribution case, we generate error terms from a joint normal distribution with $\rho_0 = 0.25$ and $\rho_1 = 0.33$. For the t distribution case, we generate errors so that the normally transformed errors are jointly normally distributed with the same ρ_0 and ρ_1 as above, but these errors ν, ε_0, and ε_1 are marginally distributed as the t distribution with degrees of freedom equals to 1, 6 and 5, respectively.

Table 2.9: Starting Wage Differentials with and without Exclusion Restrictions

	Normal Distribution		t Distribution	
Exclusion Restrictions	With	Without	With	Without
Whole Sample	0.0983***	-0.0789***	-0.0275	-0.0469***
	(0.0132)	(0.0103)	(0.0173)	(0.0166)
FEMALE	0.1007***	-0.0499***	-0.0189	-0.0326*
	(0.0314)	(0.0148)	(0.0198)	(0.0188)
MALE	0.1020***	-0.0989***	0.0035	-0.0144
	(0.0162)	(0.0133)	(0.0247)	(0.0267)

Standard error in parentheses. ***,**,* indicate statistical significance at 1%, 5%, and 10% levels, respectively.

Thus, we can estimate the model without exclusion restrictions without too much worries.

Table 2.9 summarizes the results of the estimation with and without exclusion restrictions. For comparison, we present the results from the normal distribution. Under the normal distribution, the estimated starting wage differentials with and without exclusion restrictions are quite different. The signs are completely different and the magnitudes are relatively large in each case. On the other hand, under the t distribution, whether to include the exclusion restrictions has at best a moderate impact on the estimated starting wage differentials. As the results obtained previously, the average wage differentials are negative, which are significant for the whole sample and the sample of females. In addition to the starting wage differentials, we obtain the almost same results of the wage growth coefficients without exclusion restrictions though not reported here.

At the end, we extend our empirical methodology. Our approach, which follows

ε_0 and ε_1 are rescaled so that the variance equals 1, and ν is rescaled by the ratio of the IQR of the normal distribution to the IQR of the t distribution with 1 degree of freedom (Cauchy distribution). For each distribution, we consider the estimation with and without exclusion restriction z_i. From 1,000 replications, we compute mean absolute error (MAE) and root mean squared error (RMSE) of the ATE estimator. In the case of the normal distribution, MAE and RMSE with z_i are 0.060 and 0.077, respectively, but MAE and RMSE jump up to 0.393 and 0.492, respectively, without z_i: these measures increase by about 550%. On the other hand, in the case of the t distribution, MAE and RMSE are 0.054 and 0.070 with z_i while 0.082 and 0.103 without z_i: about 50% increases.

Lee (1982, 1983), relaxes the joint normality assumption of the traditional estimator by allowing the marginal distributions to be nonnormal. However, this method is yet limited in that the transformed errors are assumed to be jointly normal. It implicitly assumes that the errors are linearly dependent. The dependence between the errors can be more complicated. To allow the flexibility of the dependence patterns, we consider the copula-based sample selection model proposed by Smith (2003, 2005). As a matter of fact, our approach is a special case of the copula-based approach.

Even with the copula-based approach, we obtain results that are similar to what we have found thus far. The starting wage of permanent workers is lower than the wage of temporary workers on average by about 4.14%, which is statistically significant at the 1% level. The wage growth pattern of each contract worker are also found to be very similar to the previous results. The results are shown in Table 2.11 in the appendix.[20]

Through the exercises in this section, we confirm the robustness of our main finding. The starting wage is lower for a permanent worker on average than for a temporary worker, but the wage of a permanent worker grows at a more faster rate than the wage of a temporary worker.

2.8 Concluding Remarks

In this study, we examine the differences in the wage structures between permanent and temporary workers. Using a dataset of new labor market entrants in the Nether-

[20]Readers who are interested in the copula method are referred to Trivedi and Zimmer (2007a). The copula approach specifies copula functions for the dependence of the errors, each of which exhibits a unique dependence pattern. We need to choose best fitting copulas among several copulas: More specifically, we consider Gaussian, FGM, Frank, Clayton, Gumbel, and Joe. For the dependence between the error of the selection equation and the error of each of the starting wage equations, the Joe copulas are best-fitting, based on the maximized likelihood value. The Vuong test (Vuong, 1989) also indicates the Joe copula model is statistically preferred to the model with the Gaussian copula, which is our previous specification. For the dependence between the error of the selection equation and the error of each of the current wage equations, the Gaussian copulas are best-fitting. This result is based on the estimation with the exclusion restrictions. Without the exclusion restrictions, we obtain the wage differential of -0.0507, with the standard error of 0.0109.

lands, we find that while permanent workers receive lower wages than temporary workers at the starting point of employment by approximately 3 to 5%, and the wage growth for permanent workers is faster than for temporary workers. After about 6 months since the starting jobs, the wages of permanent workers exceed the wages of temporary workers. The findings are robust to several alternative specifications such as using different wage measures and excluding the exclusion restrictions.

This paper also highlights the importance of the distributional assumption in the estimation. The standard estimation under the joint normality yields quite different results from our approach that relaxes the distributional assumption. The econometric literature has developed even more flexible approaches in both parametric and semiparametric ways. We leave the task to confirm our findings in this paper with these recent techniques for the future research.

Appendices

2.A The Estimation Procedure

This appendix illustrates the estimation strategy, which consists of two steps.

The first-step estimation is the maximum likelihood that estimates the selection equation and the starting wage equation jointly. For notational simplicity, let β be the vector of all the coefficients of the wage starting equation and η be other nuisance parameters. That is, the vector β includes δ and β if Specification 1 and β_0 and β_1 if Specification 2. The vector η includes the scaling parameter(s) σ and the coefficient(s) of correlation ρ, and it also includes the degrees of freedom parameter(s) if the marginal distributions are assumed to be the t distribution. Let $\ln L_1$ be the log likelihood function for this estimation.

The parameters to estimate in this likelihood estimation are $(\gamma', \beta', \eta')'$. The likelihood function takes these parameters as its arguments and assuming independence across observations: $\ln L_1 = \sum_{i=1}^{N} \ln L_{1i}(\gamma, \beta, \eta)$, where $\ln L_{1i}(\gamma, \beta, \eta)$ is the contribution to the log likelihood by an observation i. For instance, under Specification 2 with the joint normality,

$$\begin{aligned} \ln L_{1i} =& (1-d_i)\left[-\ln\sigma_0 + \ln\phi\left(\varepsilon_{0i}/\sigma_0\right) + \ln\Phi\left(\frac{-z_i{}'\gamma - \rho_0(\varepsilon_{0i}/\sigma_0)}{\sqrt{1-\rho_0^2}}\right)\right] \\ &+ d_i\left[-\ln\sigma_1 + \ln\phi\left(\varepsilon_{1i}/\sigma_1\right) + \ln\Phi\left(\frac{z_i{}'\gamma + \rho_1(\varepsilon_{1i}/\sigma_1)}{\sqrt{1-\rho_1^2}}\right)\right], \end{aligned}$$

If the t distribution is assumed, we replace $\varepsilon_{di}/\sigma_{di}$ with $\varepsilon^*_{di} = \Phi^{-1}(F_t(\varepsilon_{di}/\sigma_{di}))$ for $d = 0$ and 1, and we replace $z_i{}'\gamma$ with $\Phi^{-1}(F_t(z_i{}'\gamma))$, where $F_t(\cdot)$ is the cdf of the t distribution, of which degrees of freedom are different for each error term. If the starting wage is missing for observation i, the contribution is $\ln L_{1i} = \ln\Phi((2d_i - 1)z_i{}'\gamma)$ or $\ln F_t((2d_i - 1)z_i{}'\gamma)$

Define $\partial \ln L_{1i}(\gamma, \beta, \eta)/\partial\gamma = S_{\gamma i}$, $\partial \ln L_{1i}(\gamma, \beta, \eta)/\partial\beta = S_{\beta i}$, and $\partial \ln L_{1i}(\gamma, \beta, \eta)/\partial\eta = S_{\eta i}$. The first-step estimator of those parameters can be seen as the solution of the first-order conditions: $\sum_{i=1}^{N}({S_{\gamma i}}', {S_{\beta i}}', {S_{\eta i}}')' = 0$.

Likewise, let $\ln L_2$ be the likelihood function for the estimation of the current wage equation. In order to allow the selectivity, the likelihood function also consists of the selection equation. Then, the likelihood function is $\ln L_2 = \sum_{i=1}^{N} \ln L_{2i}(\mu, \kappa, \gamma, \beta)$. The vector μ includes the coefficients on months in jobs: μ and θ in Specification 1 and μ_0 and μ_1 in Specification 2. κ is the vector of nuisance parameters that appear in this step of the estimation such as the coefficient(s) of correlation between ν and ξ and degrees of freedom parameter(s) if ξ is assumed to be t distribution. Since we already estimated γ and β, the parameters to estimate at this step are μ and κ, which are obtained as the solution of the first-order condition: $\sum_{i=1}^{N}(S'_{\mu i}, S'_{\kappa i})' = 0$, where $S_{\mu i} = \partial \ln L_{2i}(\mu, \kappa, \widehat{\gamma}, \widehat{\beta})/\partial\mu$ and $S_{\kappa i} = \partial \ln L_{2i}(\mu, \kappa, \widehat{\gamma}, \widehat{\beta})/\partial\kappa$. Notice that the estimated values of γ and β, which are obtained in the first-step estimation are substituted.

Let $\lambda_1 = (\gamma', \beta', \eta')'$ be the vector of the parameter estimated in the first step. Likewise, let $\lambda_2 = (\mu', \kappa')'$ be the vector of the parameter estimated in the second step. We can obtain the asymptotic variance of λ_1 from the maximum likelihood principle or the quasi-maximum likelihood principle (White, 1982). For the inference for λ_2, however, we need to take into account the fact that γ and β are estimated at the first step. The following derivation of the asymptotic variance of λ_2 is based on Newey (1984) and Cameron and Trivedi (2005). Let $V(\lambda_2)$ be the asymptotic variance of λ_2:

$$\begin{aligned} V(\lambda_2) =& H_{22}^{-1} G_{22} H_{22}^{-1} + H_{22}^{-1} H_{21} \left[H_{11}^{-1} G_{11} H_{11}^{-1} \right] H_{21}^{'} H_{22}^{-1} \\ & - H_{22}^{-1} \left[H_{21} H_{11}^{-1} G_{12} + G_{21} H_{11}^{-1} H_{21}^{'} \right] H_{22}^{-1}, \end{aligned}$$

where

$$H_{11} = \lim N^{-1} \sum_{i=1}^{N} E\left(\frac{\partial^2 \ln L_{1i}}{\partial \lambda_1 \partial \lambda_1^{'}}\right) \qquad H_{22} = \lim N^{-1} \sum_{i=1}^{N} E\left(\frac{\partial^2 \ln L_{2i}}{\partial \lambda_2 \partial \lambda_2^{'}}\right)$$

$$H_{21} = \lim N^{-1} \sum_{i=1}^{N} E\left(\frac{\partial^2 \ln L_{2i}}{\partial \lambda_2 \partial \lambda_1^{'}}\right)$$

and

$$G_{11} = \lim N^{-1} \sum_{i=1}^{N} E\left(S_{\lambda_1 i} S_{\lambda_1 i}^{'}\right) \qquad G_{12} = \lim N^{-1} \sum_{i=1}^{N} E\left(S_{\lambda_1 i} S_{\lambda_2 i}^{'}\right)$$

$$G_{21} = \lim N^{-1} \sum_{i=1}^{N} E\left(S_{\lambda_2 i} S_{\lambda_1 i}^{'}\right) \qquad G_{22} = \lim N^{-1} \sum_{i=1}^{N} E\left(S_{\lambda_2 i} S_{\lambda_2 i}^{'}\right).$$

$S_{\lambda_1 i}$ and $S_{\lambda_2 i}$ are defined as $S_{\lambda_1 i} = (S_{\gamma i}^{'}, S_{\beta i}^{'})'$ and $S_{\lambda_2 i} = (S_{\mu i}^{'}, S_{\kappa i}^{'})'$. We estimate the asymptotic variance by replacing population moments with sample counterparts.

2.B Additional Results

Table 2.10: The Estimated Result of Specification 1

	Normal Distribution		t Distribution	
	Selection Eq.[a,b]	Starting Wage Eq.[a]	Selection Eq.[a,b]	Starting Wage Eq.[a]
d		0.1069 ***		-0.0157
		(0.0126)		(0.0136)
AGE	-0.0281 ***	0.0158 ***	-0.0315 ***	0.0132 ***
	(0.0017)	(0.0009)	(0.0019)	(0.0009)
GRADE	-0.0221 ***	0.0286 ***	-0.0086	0.0266 ***
	(0.0057)	(0.0029)	(0.0060)	(0.0027)
MALE	0.0568 ***	0.0267 ***	0.0694 ***	0.0333 ***
	(0.0066)	(0.0033)	(0.0097)	(0.0033)
WO	-0.0378 ***	0.1237 ***	0.0116	0.1255 ***
	(0.0072)	(0.0035)	(0.0090)	(0.0033)
NATIVE	-0.0037	0.0005	-0.0143	-0.0011
	(0.0136)	(0.0071)	(0.0134)	(0.0068)
DISABLED	-0.0330 ***	-0.0057	-0.0190 **	-0.0123 **
	(0.0100)	(0.0051)	(0.0089)	(0.0049)
MEDIUM	0.0111	0.0149 ***	0.0203 **	0.0174 ***
	(0.0082)	(0.0041)	(0.0085)	(0.0039)
LARGE	0.0133 *	0.0355 ***	0.0618 **	0.0427 ***
	(0.0072)	(0.0036)	(0.0139)	(0.0035)
HOME	-0.0592 ***		-0.0407 ***	
	(0.0091)		(0.0095)	
CHILD	0.1199 ***		0.0696 ***	
	(0.0181)		(0.0228)	
CONSTANT		1.7725 ***		1.9201 ***
		(0.0424)		(0.0406)
ρ		-0.2869 ***		0.1539 ***
		(0.0428)		(0.0520)
r			0.2982	6.9665
			(0.0630)	(0.3872)
σ		0.1707		0.1426
		(0.0018)		(0.0018)
$\ln L_1$	-10876.034		-10302.597	
		Current Wage Eq.		Current Wage Eq.
m		0.0021 ***		0.0014 ***
		(0.0004)		(0.0002)
$d \times m$		0.0057 ***		0.0056 ***
		(0.0007)		(0.0002)
ρ		-0.3620 ***		0.0376 *
		(0.0397)		(0.0226)
r				5.3147
				(0.0453)
σ		0.2026		0.1562
		(0.0014)		(0.0007)
$\ln L_2$	-10810.631		-10233.899	

[a] The selection equation and starting wage equation also contain the dummies for regions of residence, industries, and years of starting job.
[b] Marginal effects are reported.

Table 2.11: The Estimated Result with the Copula Method

	Starting Wage Equation[a]		Selection Equation[a]	
	TEMPORARY	PERMANENT	Coefficient[b]	Marginal Effect[c]
AGE	0.0134 ***	0.0111 ***	-0.0912 ***	-0.0316 ***
	(0.0011)	(0.0012)	(0.0149)	(0.0018)
GRADE	0.0254 ***	0.0267 ***	-0.0246	-0.0085
	(0.0041)	(0.0036)	(0.0159)	(0.0059)
MALE	0.0312 ***	0.0370 ***	0.1939 ***	0.0691 ***
	(0.0044)	(0.0044)	(0.0379)	(0.0088)
WO	0.1111 ***	0.1383 ***	0.0329	0.0114
	(0.0048)	(0.0046)	(0.0264)	(0.0084)
NATIVE	0.0075	-0.0123	-0.0401	-0.0140
	(0.0099)	(0.0092)	(0.0389)	(0.0135)
HANDICAP	-0.0102	-0.0164 **	-0.0544 **	-0.0186 **
	(0.0068)	(0.0068)	(0.0263)	(0.0089)
MEDIUM	0.0255 ***	0.0124 **	0.0601 **	0.0207 **
	(0.0055)	(0.0055)	(0.0256)	(0.0084)
LARGE	0.0430 ***	0.0472 ***	0.1796 ***	0.0635 ***
	(0.0050)	(0.0047)	(0.0494)	(0.0127)
HOME			-0.1104 ***	-0.0371 ***
			(0.0320)	(0.0090)
CHILD			0.1911 ***	0.0666 ***
			(0.0638)	(0.0215)
CONSTANT	1.9460	1.9523	2.4855	
	(0.0832)	(0.0473)	(0.3778)	
Kendall's τ[d]	0.1630	0.1438		
r	7.0525	6.3524	0.3049	
	(0.6163)	(0.4134)	(0.0571)	
σ	0.1600	0.1255		
	(0.0037)	(0.0021)		
$\ln L_1$	-10570.794			
	Current Wage Eq.		**Starting Wage Differential**	
m	-0.0013 **	0.0045 ***	0.0414 ***	
	(0.0005)	(0.0004)	(0.0112)	
Kendall's τ[d]	-0.1416 ***	0.2796 ***		
	(0.0247)	(0.0191)		
r	5.5580	4.3900		
	(0.2931)	(0.1817)		
σ	0.1623	0.1471		
	(0.0027)	(0.0025)		
$\ln L_2$	-9710.753			

[a] The selection equation and starting wage equation also contain the dummies for regions of residence, industries, and years of starting job.

[b,c] See the notes of Table 2.3

[d] The Kendall's τ measures the dependence of the error of the selection equation and the error of the corresponding wage equation. It ranges from -1 to 1, and the closer to 0, the weaker the dependence.

3 The Structure of Adjustment Costs of Factors of Production: Evidence from Indonesian Manufacturing Plants

3.1 Introduction

In this paper, I investigate the structure of adjustment costs of factors of production using a plant-level panel dataset from Indonesian manufacturing sectors. The structure of the adjustment costs of employment has been a central interest of the literature on dynamic labor demand. Different specifications of the structure generate different predictions for the dynamic path of labor demand in response to shocks.

In the early stages of the literature, it was standard to assume convex adjustment costs. Under this assumption, employment is adjusted smoothly and gradually toward a new target level. Even though it is a reasonable approximation at the industry or economy level, it is not consistent with observations at a firm- or plant-level, which show lumpy adjustment at one period and inaction over several periods. Since the seminal work by Hamermesh (1989), the literature has shifted to nonconvex functional forms of the costs. Using micro data, the literature has accumulated evidence of nonconvex adjustment costs (Caballero et al., 1997, Nilsen et al., 2007, Cooper and Willis, 2009). Besides nonconvexity, there are several pieces of empirical evidence that the adjustment cost function exhibits asymmetry so that a firm responds to negative and positive economic shocks in different ways (Jaramillo et al., 1993, Pfann and Palm, 1993).

However, in the analysis of employment adjustment, it is often assumed that adjustments of other factors are frictionless. Alternatively, it is assumed that other factors, in particular capital, are fixed. These assumptions make employment adjustment decisions independent of adjustment of other factors. These assumptions, however, result in biased estimates if, apart from labor, other factors are costly to adjust as well. If two factors are costly to adjust, which factor is more flexible is determined endogenously based on the parameters of the adjustment costs structure (Dixit, 1997).[1]

Indeed, there is a huge literature on adjustment costs of capital. The adjustment cost of capital is often discussed in connection with Tobin's Q (Tobin, 1969) as in Hayashi (1982). Parallel to the literature on labor adjustment, the empirical literature of capital adjustment has come to emphasize firm/plant-level behaviors (Fazzari et al., 1988). Nonconvexity is important at a micro level, which is similar in labor adjustment costs (Caballero et al., 1995, Nilsen and Schiantarelli, 2003, Cooper and Haltiwanger, 2006). However, as the labor adjustment literature does, it is commonly assumed that capital is the only factor that is costly to adjust. To estimate the structure of the adjustment costs, it is important to consider the adjustment process of all factors together.

There are a few recent studies that estimate the adjustment costs of labor and capital together, using micro level datasets.[2] Bloom (2009) models and estimates nonconvex adjustment costs of capital and labor, using firm-level data from U.S. Compustat. The findings are that the model with nonconvex adjustment costs of both inputs fits the best, followed in order of fit by a model with only capital adjustment

[1] Dixit (1997) analyzes the case with linear adjustment costs of two inputs. Eberly and Mieghem (1997) extend the analysis to a multi-factor case.

[2] Using aggregate data, the interrelated adjustment costs are examined under the assumption of quadratic forms. For example, (Shapiro, 1986) finds insignificant coefficients on interaction terms among changes in factors.

costs, one with convex costs of both inputs, one with labor adjustment costs only, and one with no adjustment costs of both inputs. All of this suggests that it is important to incorporate capital adjustment costs into a model estimating labor adjustment costs. Eslava et al. (2010) also empirically examine the adjustment costs of both inputs at the same time, by expanding the “gap” approach by Caballero et al. (1997). Their findings also show that firm’s adjustments of one factor depend on adjustments of the other. Asphjell et al. (2010) develop an empirical model based on maximum likelihood estimation. Using Norwegian manufacturing plant-level panel data, they find that adjustment costs of both factors are interrelated.

In this study, I follow and modify the methodology by Asphjell et al. (2010), which is computationally more tractable than the other two studies, and apply it to a plant-level panel dataset from Indonesian manufacturing sectors. My contribution to the literature is twofold. First, there are relatively few empirical works that estimate the adjustment costs of labor and capital at the same time. I add new evidence to this literature. Even though my estimation strategy is mostly based on Asphjell et al. (2010), I relax a distributional assumption imposed by them, using the copula method. The finding indicates that the distributional assumption is important for the parameter estimation.

Second, I employ data from a developing country. The empirical evidence on the adjustment costs of labor and capital is concentrated in manufacturing sectors in developed countries. There are only a few pieces of empirical evidence from developing countries.[3] It is reasonable to ask whether the adjustment cost structures in developing countries are the same as in developed countries. I use a dataset from Indonesia to investigate the structure of the adjustment costs of factors. This study provides empirical evidence that, in line with the previous studies from developed countries, adjustment costs are asymmetric and nonconvex in Indonesia .

[3]For example, Eslava et al. (2010) using the dataset of Columbian manufacturing plants.

In Indonesia, employment protection regulations are strict, and labor markets are characterized as rigid: Indonesia was ranked the worst among Southeast Asian countries and 103th out of 136 countries in the world by the rigidity of employment index by the World Bank (2005).[4] With high severance pay rates and restrictions on the use of temporary (fixed-term) contracts, it is difficult for firms to adjust their employment level. Moreover, in 2003, Indonesia introduced a law that tightened labor market regulations, which were already strict by international and regional standards. This study investigates the effect of this law on the structural parameters of the adjustment costs. The result shows that the law has significant impacts on the parameter estimates. In addition, I calculate two model predictions for 2003: one with the parameter estimates after the law and the other with the parameter estimates before the law, which can be considered as counterfactual as if the law had not been enacted. The comparison of these predictions shows that in 2003, the probability that a plant decreases its level of employment is 10 percentage points lower than it would have been if the law had not been enacted.

The structure of this chapter is as follows. Section 3.2 illustrates the theoretical model of this paper, and Section 3.3 explains the estimation strategy. Following the data description in Section 3.4, I show the estimation results in Section 3.5. Finally, Section 3.6 examines the effects of the law in 2003 on the estimated parameters of the adjustment costs structure. Section 3.7 concludes the paper.

3.2 Theoretical Model

This section illustrates the theoretical model in this study. The model is based on the theoretical framework by Abel and Eberly (1994).

[4] Southeast Asian countries are Cambodia, Indonesia, Lao, Malaysia, Philippines, Singapore, Thailand, and Vietnam. The rigidity of employment index is the average of three subindexes: difficulty of hiring index, rigidity of hours index, and difficulty of firing index. See www.doingbusiness.org.

A theoretical model of dynamic factor demand is developed from a dynamic optimization problem by a representative production unit (a plant in this study). Assuming that a plant uses two factors of production, capital K_t and labor L_t, at each time period t, the plant's objective function at time period t is:

$$V_t = E_t \sum_{s=0}^{\infty} \beta^{t+s} \left[F\left(K_{t+s}, L_{t+s}, \Lambda_{t+s}\right) - w_{t+s}L_{t+s} - G(H_{t+s}, I_{t+s}, L_{t+s-1}, K_{t+s-1})\right], \tag{3.1}$$

where β is a discount factor. $F(\cdot)$ is a production function, and $G(\cdot)$ is an adjustment cost function, which is of interest in this study. Notice that there is no payment to capital at each period. Instead, $G(\cdot)$ includes the purchase/acquisition cost of capital. I_t denotes gross investment (disinvestment if negative) at time period t. H_t denotes changes in labor L_t at time period t. Assuming perfect competition in input markets, a plant unit takes the wage rate w_t as given. Λ_t represents a demand shock or a technology shock and is the source of randomness in the model.

The specification of the adjustment cost function $G(\cdot)$ plays an important role in determining optimal paths of factor demands. As mentioned earlier, the literature has accumulated pieces of evidence that the adjustment costs of factors are asymmetric and nonconvex at a plant-level. Accordingly, I specify:

$$\begin{aligned} G(H_t, I_t, L_{t-1}, K_{t-1}) = {} & \mathrm{I}(H_t > 0)\left[a_L^+ + b_L^+ H_t + \frac{c_L^+}{2}\left(\frac{H_t}{L_{t-1}}\right)^2 L_{t-1}\right] \\ & + \mathrm{I}(H_t < 0)\left[a_L^- - b_L^- H_t + \frac{c_L^-}{2}\left(\frac{H_t}{L_{t-1}}\right)^2 L_{t-1}\right] \\ & + \mathrm{I}(I_t > 0)\left[a_K^+ + b_K^+ I_t + \frac{c_K^+}{2}\left(\frac{I_t}{K_{t-1}}\right)^2 K_{t-1}\right] \\ & + \mathrm{I}(I_t < 0)\left[a_K^- - b_K^- I_t + \frac{c_K^-}{2}\left(\frac{I_t}{K_{t-1}}\right)^2 K_{t-1}\right], \end{aligned}$$

where $\mathrm{I}(\cdot)$ is the indicator function, which is 1 if its argument is true. The presence of the fixed components a_L^+, a_L^-, a_K^+, and a_K^-, all of which are assumed to be positive, makes the adjustment costs nonconvex.[5]

Capital K_t and Labor L_t evolve according to

$$K_t = (1 - \delta_K)K_{t-1} + I_t \qquad \text{and} \qquad L_t = (1 - \delta_L)L_{t-1} + H_t,$$

where δ_K is a depreciation rate and δ_L is a voluntary quit rate.

When a plant adjusts employment optimally, the growth rate of employment can be shown as

$$\frac{H_t^*}{L_{t-1}} = \begin{cases} \frac{1}{c_L^+}\left(q_t^L - b_L^+\right) & \text{if } H_t > 0 \\ \frac{1}{c_L^-}\left(q_t^L + b_L^-\right) & \text{if } H_t < 0 \end{cases}, \tag{3.2}$$

where q_t^L is the shadow price of labor. It represents the present discounted value of the marginal product of labor net of wage costs and future adjustment decisions. Assuming c_L^+ and c_L^- are positive, the optimal growth rate of employment is an increasing (nondecreasing) function of q_t^L.

Similarly, an optimal investment rate is

$$\frac{I_t^*}{K_{t-1}} = \begin{cases} \frac{1}{c_K^+}\left(q_t^K - b_K^+\right) & \text{if } I_t > 0 \\ \frac{1}{c_K^-}\left(q_t^K + b_K^-\right) & \text{if } I_t < 0 \end{cases}, \tag{3.3}$$

where q_t^K is the shadow price of capital.

Because of nonconvexity, a plant does not adjust the factors of production at every period. Rather, it makes adjustments only when the benefit of adjustment exceeds

[5]Even though Asphjell et al. (2010) assumes that the adjustment costs of labor and capital interact with each other, I do not assume the interaction. Instead, I assume the asymmetry of the adjustment costs, which they do not assume. With the asymmetry, the estimation of the interacted adjustment costs becomes much more complicated. For simplicity, I ignore the interaction and leave it as a future task.

the cost of adjustment. For example, a plant increases its employment level when

$$q_t^L H_t^* > a_L^+ + b_L^+ H_t^* + \frac{c_L^+}{2}\left(\frac{H_t^*}{L_{t-1}}\right)^2 L_{t-1},$$

where H_t^* is the optimal level of hiring given by equation (3.2). This leads to the following condition: $q_t^L > \sqrt{\frac{2a_L^+ c_L^+}{L_{t-1}}} + b_L^+$ or $q_t^L < -\sqrt{\frac{2a_L^+ c_L^+}{L_{t-1}}} + b_L^+$. Given that H_{it}^*/L_{it-1} is an increasing function of q_t^L, H_{it}^*/L_{it-1} is positive only when

$$q_t^L > \sqrt{\frac{2a_L^+ c_L^+}{L_{t-1}}} + b_L^+.$$

With the same reasoning, it can be derived that a plant decreases its level of employment when

$$q_t^L < -\sqrt{\frac{2a_L^- c_L^-}{L_{t-1}}} - b_L^-.$$

When q_t^L is between the two thresholds, a plant keeps its employment level intact.

Similarly, gross investment and disinvestment, respectively, occur when

$$q_t^K > \sqrt{\frac{2a_K^+ c_K^+}{K_{t-1}}} + b_K^+ \qquad \text{and} \qquad q_t^K < -\sqrt{\frac{2a_K^- c_K^-}{K_{t-1}}} - b_K^-.$$

Otherwise, a plant makes neither investment nor disinvestment, and its capital stock decreases only due to depreciation. In total, there are 9 decision states: decrease, inaction, and increase for each of the two factors of production.

3.3 Estimation Strategy

The estimation of the structural parameters needs two steps. First, the decision of whether to adjust its factors is estimated by maximum likelihood (ML) method. Then, the amounts of hiring and investment are estimated by OLS.

Since the shadow prices of labor and capital are not directly observable, these are approximated by a linear function of observable variables:

$$q_{it}^L = \gamma_0^L + {\gamma_L}' Z_{it}^L + \varepsilon_{it}^L$$
$$q_{it}^K = \gamma_0^K + {\gamma_K}' Z_{it}^K + \varepsilon_{it}^K \tag{3.4}$$

where Z_{it}^L and Z_{it}^K are vectors of observable variables, and the γ's are parameters. Notice that now I introduce the subscript i to represent a plant and stochastic terms ε_{it}^L and ε_{it}^K. The stochastic terms ε_{it}^L and ε_{it}^K have zero means. Assume that these terms are independent across plants and over time, but these terms are not independent of each other contemporaneously. The distributions of ε_{it}^L and ε_{it}^K are scale families of certain distributions with the scale parameters σ_L and σ_K, respectively. I will specify the distributional assumption of ε_{it}^L and ε_{it}^K later in this section.

Following Nilsen et al. (2007) and Letterie and Pfann (2007), Z_{it}^L and Z_{it}^K here include the lagged values of the ratio of output to employment Y_{it-1}/L_{it-1} and of the ratio of output to capital Y_{it-1}/K_{it-1}, and the lagged wage rate w_{it-1}. In order to allow nonlinearity, Z_{it}^L and Z_{it}^K include the square terms of these variables. The shadow values can be expressed as the expected discounted value of future marginal productivity of a corresponding factor (net of wage for labor) and future adjustment decisions. The variables are used as proxies of marginal productivities of factors. To reduce concerns about endogeneity, lagged values are used instead of current values. Z_{it}^L and Z_{it}^K also include year dummies, industry dummies (at 2-digit level of ISIC 3.1), and dummies for type of ownerships (domestic, foreign, or government-owned). In the estimation, the set of variables included in Z_{it}^L and Z_{it}^K are the same all the time. Therefore, I let $Z_{it} = Z_{it}^L = Z_{it}^K$ hereafter.

For notational simplicity in the following discussion, define

$$A^{+}_{Jit} = \sqrt{\frac{2a^{+}_{J}c^{+}_{J}}{J_{it-1}}} - (\gamma^{J}_{0} - b^{+}_{J}) - \gamma_{J}{}' Z_{it}$$
$$A^{-}_{Jit} = -\sqrt{\frac{2a^{-}_{J}c^{-}_{J}}{J_{it-1}}} - (\gamma^{J}_{0} + b^{-}_{J}) - \gamma_{J}{}' Z_{it},$$

for $J = L, K$. From now on, let $\widetilde{x}$ stand for a standardized value of x divided by the corresponding scale parameter, where x can be the parameters or the variables in the model. For example, A^{-}_{Lit} is standardized by dividing by σ_L: $\widetilde{A}^{-}_{Lit} \equiv A^{-}_{Lit}/\sigma_L$. Likewise, $\widetilde{A}^{-}_{Kit} \equiv A^{-}_{Kit}/\sigma_K$.

Combining the approximations of q^{L}_{it} and q^{K}_{it} with the adjustment decisions derived above, I obtain the following log likelihood function:

$$\begin{aligned}
\log L = &\sum_{t=1}^{T} \sum_{i \in \Omega^{-,-}_{it}} \log(Pr(\Omega^{-,-}_{it})) + \sum_{t=1}^{T} \sum_{i \in \Omega^{-,0}_{it}} \log(Pr(\Omega^{-,0}_{it})) + \sum_{t=1}^{T} \sum_{i \in \Omega^{-,+}_{it}} \log(Pr(\Omega^{-,+}_{it})) \\
&+ \sum_{t=1}^{T} \sum_{i \in \Omega^{0,-}_{it}} \log(Pr(\Omega^{0,-}_{it}) + \sum_{t=1}^{T} \sum_{i \in \Omega^{0,0}_{it}} \log(Pr(\Omega^{0,0}_{it}) + \sum_{t=1}^{T} \sum_{i \in \Omega^{0,+}_{it}} \log(Pr(\Omega^{0,+}_{it})) \\
&+ \sum_{t=1}^{T} \sum_{i \in \Omega^{+,-}_{it}} \log(Pr(\Omega^{+,-}_{it})) + \sum_{t=1}^{T} \sum_{i \in \Omega^{+,0}_{it}} \log(Pr(\Omega^{+,0}_{it})) + \sum_{t=1}^{T} \sum_{i \in \Omega^{+,+}_{it}} \log(Pr(\Omega^{+,+}_{it})),
\end{aligned} \tag{3.5}$$

where $\Omega^{\cdot,\cdot}_{it}$ denotes the state where a plant adjusts its employment and its capital stock. The sign of superscripts indicate negative, zero, and positive adjustments, with the first corresponding to labor and the second to capital. For example, $\Omega^{-,-}_{it}$ indicates that plant i decreases its employment and capital stock at time t. The specific functional form of $Pr(\Omega^{\cdot,\cdot}_{it})$ depends on the distributional assumption of ε^{L}_{it} and ε^{K}_{it}.

This first-step ML estimation recovers $\tilde{c}^{+}_{L}\tilde{a}^{+}_{L}$, $\tilde{c}^{-}_{L}\tilde{a}^{-}_{L}$, $\tilde{c}^{+}_{K}\tilde{a}^{+}_{K}$, $\tilde{c}^{-}_{K}\tilde{a}^{-}_{K}$, $\tilde{\gamma}^{L}_{0} + \tilde{b}^{-}_{L}$, $\tilde{\gamma}^{L}_{0} - \tilde{b}^{+}_{L}$, $\tilde{\gamma}^{K}_{0} + \tilde{b}^{-}_{K}$, $\tilde{\gamma}^{K}_{0} - \tilde{b}^{+}_{K}$, $\tilde{\gamma}_{L}$, and $\tilde{\gamma}_{K}$ as well as parameters belonging to the joint distribution of $\tilde{\varepsilon}_{it}^L$ and $\tilde{\varepsilon}_{it}^K$. Note that as usual in probit or logit estimation, the structural parameters can be estimated only proportional to the scale parameters of the stochastic terms, σ_L and σ_K.

In the second step, the factor adjustment equations, (3.2) and (3.3), are estimated by OLS. Notice that approximating q_{it}^L and q_{it}^K by (3.4) introduces the stochastic terms into the equations (3.2) and (3.3), and notice also that these equations are conditional on the decisions of adjustment. Therefore, the estimation equations should include the means of ε_{it}^L and ε_{it}^K conditional on the decisions of adjustment:

$$\frac{H_{it}}{L_{it-1}} = \begin{cases} \dfrac{1}{\tilde{c}_L^+}\left(\widehat{\tilde{\gamma}_0^L - \tilde{b}_L^+} + \hat{\tilde{\gamma}}_L{}' Z_{it} + \hat{\lambda}_{Lit}^+\right) + \eta_{Lit}^+ & \text{if } H_{it} > 0 \\ \dfrac{1}{\tilde{c}_L^-}\left(\widehat{\tilde{\gamma}_0^L + \tilde{b}_L^-} + \hat{\tilde{\gamma}}_L{}' Z_{it} + \hat{\lambda}_{Lit}^-\right) + \eta_{Lit}^- & \text{if } H_{it} < 0 \end{cases}, \tag{3.6}$$

for employment adjustment, and for capital adjustment:

$$\frac{I_{it}}{K_{it-1}} = \begin{cases} \dfrac{1}{\tilde{c}_K^+}\left(\widehat{\tilde{\gamma}_0^K - \tilde{b}_K^+} + \hat{\tilde{\gamma}}_K{}' Z_{it} + \hat{\lambda}_{Kit}^+\right) + \eta_{Kit}^+ & \text{if } I_{it} > 0 \\ \dfrac{1}{\tilde{c}_K^-}\left(\widehat{\tilde{\gamma}_0^K + \tilde{b}_K^-} + \hat{\tilde{\gamma}}_K{}' Z_{it} + \hat{\lambda}_{Kit}^-\right) + \eta_{Kit}^- & \text{if } I_{it} < 0 \end{cases}. \tag{3.7}$$

The terms η_{Lit}^+, η_{Lit}^-, η_{Kit}^+, and η_{Kit}^- are mean-zero errors, which are independent of ε_{it}^L, ε_{it}^K, and Z_{it} but can be heteroskedastic. The terms λ_{Lit}^+, λ_{Lit}^-, λ_{Kit}^+, and λ_{Kit}^- are the expected values of $\tilde{\varepsilon}_{it}^L$ and $\tilde{\varepsilon}_{it}^K$ conditional on the adjustment decisions.[6] That is, $\lambda_{Jit}^+ \equiv E[\tilde{\varepsilon}_{it}^J | \tilde{\varepsilon}_{it}^J > \widetilde{A}_{Jit}^+]$ and $\lambda_{Jit}^- \equiv E[\tilde{\varepsilon}_{it}^J | \tilde{\varepsilon}_{it}^J < \widetilde{A}_{Jit}^-]$ for $J = L, K$. Hats over the parameters denote that those parameters are the estimated values from the first-step ML estimation, and hats over λ's denote that these conditional expected values are computed with estimated first-step parameters inserted.[7]

From these OLS regressions, the parameters $\tilde{c}_L^+$, $\tilde{c}_L^-$, $\tilde{c}_K^+$, and $\tilde{c}_K^-$ can be identified

[6] If ε_{it}^J is normally distributed, λ takes a form of the familiar inverse Mill's ratio. Under a different distributional assumption, this conditional mean takes an different expression, as shown later on.

[7] With the assumption of mean-zero η's, each OLS contains only one explanatory variable, which is computed from the first-step ML estimation.

(again, up to scale of σ_L and σ_K). Using the estimates from the first step ML estimation, the parameters $\tilde{a}_L^+$, $\tilde{a}_L^-$, $\tilde{a}_K^+$, and $\tilde{a}_K^-$ can be identified (up to scale of σ_L and σ_K). The linear components of the adjustment cost function, $\widetilde{b}_L^+$, $\widetilde{b}_L^-$, $\widetilde{b}_K^+$, and $\widetilde{b}_K^+$ cannot be separated from the constant terms $\widetilde{\gamma}_0^L$ and $\widetilde{\gamma}_0^K$.

There are a few notes with regard to this estimation procedure. First, the statistical inference in this model is not straightforward since the estimation of the structural parameters is based on the two-step procedure. For the inference, I estimate the bootstrap intervals of the estimated parameters.

Second, this panel data estimation model assumes that there is no plant specific random effect. Plant specific effects add considerable complexity to the estimation procedure. This study assumes that there is no such effect, and I leave this issue for future work.

Third, in order to implement this estimation, I should make an assumption on the distribution of ε_{it}^L and ε_{it}^K. In the previous literature, it is common to assume that the stochastic terms are normally distributed as Nilsen et al. (2007) and Asphjell et al. (2010) do. However, there is no prior information on the distribution of the stochastic terms. In this study, I consider alternative assumptions about the distribution as well as the assumption of normality.

3.3.1 Bivariate Normality Assumption

As a baseline assumption, first I assume bivariate normality of ε_{it}^L and ε_{it}^K. These stochastic terms are jointly normally distributed with zero means, variances σ_L^2 and σ_K^2, and a correlation coefficient ρ.

The distribution assumption specifies the expressions of the probabilities in the log likelihood function (3.5) in the first-step ML estimation. Under the assumption of a bivariate normal distribution of ε_{it}^L and ε_{it}^K, for example, the probability of the state where a plant decreases its employment and capital stock, $Pr(\Omega_{it}^{-,-})$, can be specified as: $Pr(\Omega_{it}^{-,-}) = \Phi_2(\widetilde{A}_{Lit}^{-}, \widetilde{A}_{Kit}^{-}, \rho)$, where $\Phi_2(\cdot, \cdot, \rho)$ is the cdf of a bivariate standard normal distribution with the correlation coefficient ρ. To save space, the expression for the probabilities of other states are suppressed here. See Appendix 3.B.

The assumption of a bivariate normality implies that each marginal distribution of ε_{it}^{L} and ε_{it}^{K} also is normal. Therefore, the truncated means can be expressed as $\lambda_{Jit}^{+} = E[\tilde{\varepsilon}_{it}^{J} | \tilde{\varepsilon}_{it}^{J} > \widetilde{A}_{Jit}^{+}] = \frac{\phi(\widetilde{A}_{Jit}^{+})}{\Phi(-\widetilde{A}_{Jit}^{+})}$, and $\lambda_{Jit}^{-} = E[\tilde{\varepsilon}_{it}^{J} | \tilde{\varepsilon}_{it}^{J} < \widetilde{A}_{Jit}^{-}] = -\frac{\phi(\widetilde{A}_{Jit}^{-})}{\Phi(\widetilde{A}_{Jit}^{-})}$ for $J = L$, and K. These terms are known as inverse Mills ratio's.

With the specified expressions of the log likelihood and the truncated means, I am now able to implement the estimation to obtain the structural parameters in the model (proportional to the scale parameters). This estimation procedure is based on the assumption of a bivariate normal distribution. However, it is well-known that misspecification of a distribution in ML estimation generally leads to inconsistent estimation. Even though the normality assumption is common in the previous studies, it is not necessarily a correct assumption. It would be good to estimate the model under an alternative distributional assumption.

3.3.2 Alternative Distributional Assumption

Now I discuss the alternative way to estimate the model by making different distributional assumptions. The basic estimation procedure is the same as in the case of a bivariate normality. First, the adjustment decisions are jointly estimated by maximum likelihood. Then, the employment and investment rates are estimated by OLS.

At this time, I assume that marginal distributions of ε_{it}^{L} and ε_{it}^{K} are t distributions, as opposed to the previous assumption of the normality of these stochastic variables. Assume that ε_{it}^{L} has a t distribution with degrees of freedom ν_L and a scale parameter σ_L.[8] I also assume that ε_{it}^{K} has a t distribution with degrees of freedom ν_K and a scale

[8]Here, it is not appropriate to call σ_L^2 (σ_L) a variance (standard deviation) because, theoretically,

parameter σ_K. The degrees of freedom of the two t distributions ν_L and ν_K are not necessarily the same. A t distribution has thicker tails than a normal distribution. Extreme values are more likely to happen under the assumption of t distribution than a normal distribution, so that a large adjustment can be fitted better under a t distribution than a normal distribution.

For the second-step OLS in (3.6) and (3.7), it is necessary to derive the expressions of the truncated moments of a standardized random variable $\tilde{\varepsilon}_{it}^L$ and $\tilde{\varepsilon}_{it}^K$: $\lambda_{Jit}^+ = E[\tilde{\varepsilon}_{it}^J | \tilde{\varepsilon}_{it}^J > \widetilde{A}_{Jit}^+]$ and $\lambda_{Jit}^- = E[\tilde{\varepsilon}_{it}^J | \tilde{\varepsilon}_{it}^J < \widetilde{A}_{Jit}^-]$ for $J = L$ and K. According to Kim (2008),

$$\lambda_{Jit}^+ = \frac{\nu_J^{\nu_J/2}\Gamma((v_J-1)/2)(\widetilde{B}_{Jit}^+)^{-(\nu_J-1)/2}}{2(1-F_{\nu_J}(\widetilde{A}_{Jit}^+))\Gamma(\nu_J/2)\Gamma(1/2)},$$

where $\nu_J > 1$, $\widetilde{B}_{Jit}^+ = (\widetilde{A}_{Jit}^+)^2 + \nu_J$. F_{ν_J} is the cdf of a t distribution with degrees of freedom ν_J, and $\Gamma(\cdot)$ is the gamma function. Similarly,

$$\lambda_{Jit}^- = \frac{\nu_J^{\nu_J/2}\Gamma((v_J-1)/2)(-\widetilde{B}_{Jit}^-)^{-(\nu_J-1)/2}}{2F_{\nu_J}(\widetilde{A}_{Jit}^-)\Gamma(\nu_J/2)\Gamma(1/2)},$$

where $\widetilde{B}_{Jit}^- = (\widetilde{A}_{Jit}^-)^2 + \nu_J$. The second-step OLS regressions include these truncated means, which are calculated using the estimates from the first-step ML estimation.

To implement the first-step ML estimation, I need to specify a joint distribution of ε_{it}^L and ε_{it}^K in the log likelihood function (3.5). Given the marginal distributions, the joint distribution of ε_{it}^L and ε_{it}^K is constructed with a copula method. The copula method is a useful technique to generate a joint distribution parametrically. With a copula function, a joint cumulative distribution can be easily constructed from two (or more) marginal distributions of random variables. See Trivedi and Zimmer (2007b) for the basic foundations of the copula method and illustrations of its applications.

One of the interesting properties of the copula method, which is relevant to this

---

it is not variance (standard deviation) of ε_{it}^L. Instead, it should be called a scale parameter. The variance of a t distribution with degrees of freedom ν with a scale parameter σ is $\sigma^2 \dfrac{\nu}{\nu-2}$ for $\nu > 2$.

study, is that the copula method can accommodate a relatively flexible dependence structure in the tails of a joint distribution. The pdf of a bivariate normal distribution has an elliptic contour, which indicates that tail dependences are symmetric across the two tails. In contrast, some copula functions can generate an asymmetric joint distribution so that there is stronger dependence at one tail than the other. A joint distribution may show strong dependence at negative extreme values of two random variables but weak at positive extremes, or vice versa. Furthermore, the copula can create a joint distribution with thicker (or thinner) tails relative to a bivariate normal distribution.

Flexible tail properties are interesting in the context of the model in this study. Randomness in the model comes from Λ_t in equation (3.1). It represents either technology shocks or demand shocks, or both. However, these shocks have different implications. Demand shocks seem to make ε_{it}^L and ε_{it}^K move together while technology shocks, if not neutral, induce substitutions between L and K, which implies ε_{it}^L and ε_{it}^K tend to move in opposite directions. The model does not have these prior distinctions of Λ_t. Rather, it can have composite effects, and the ways of composition can differ between a positive tail and a negative tail. Therefore, it is advantageous to have a flexible dependence structure between ε_{it}^L and ε_{it}^K.

There are a variety of copula functions. In this study, I consider the class of Archimedean copulas, which have mathematical properties that are relatively easy to deal with. Specifically, I consider the families of AMH, Clayton, Frank, Gumbel, and Joe, each of which has a unique tail dependence.[9]

Instead of estimating the entire model by each of these copula functions, I conduct the first-step ML estimation using each copula listed above, and I implement the second-step OLS with the best-fitting copula model. Each copula function is not

[9]Smith (2003) describes the copula method, paying special attentions to Archimedean copulas. He also illustrates the applications to sample selection models, where a multivariate distribution plays an important role.

nested with one another. Therefore, the selection is usually based on information criteria such as AIC and BIC. However, it is equivalent to a selection based on the maximized log likelihood since each estimation has the same number of parameters. For marginal distributions, I assume t distributions for ε_{it}^L and ε_{it}^K with degrees of freedom, $v_L = 3$ and $v_K = 2.1$, and fix this assumption across copula functions.[10] I select the copula with the largest maximized log likelihood as the best-fitting copula, and I proceed to the second-step OLS regressions using the estimates from the selected copula. See Appendix 3.C for more details on the ML estimation with the copula method.

In addition to the selected copula, I also estimate the entire model using a Gaussian copula in order to see whether results change because of differences in marginal distributions and/or dependence structures. A Gaussian copula is also referred to as Bivariate Normal Copula and has a symmetric tail dependence. However, the Gaussian copula is different from a bivariate normal distribution in that it does not assume normality of margins. If marginal distributions are normal, a joint distribution by the Gaussian copula is indeed a bivariate normal distribution. As stated above, marginal distributions are now assumed to be t distributions in this study. Therefore, the difference from the model under the bivariate normality is due to the difference in marginal distributions.

The empirical models outlined in this section are estimated using a dataset from Indonesia. The next section describes the dataset, and the estimation results are presented in the following section.

[10]It is ideal to estimate the degrees of freedom as well as other parameters. However, my attempt encountered a convergence problem. Moreover, as the number of estimated parameters increases, the estimation procedure takes longer. It is worrisome since the inference is based on bootstrap so that the entire estimation procedure needs to be repeated many times. Preliminarily, I estimated with several values of degrees of freedom. The models attained larger values of log likelihood when the degrees of freedom are smaller. The results are in favor of smaller degrees of freedom. With degrees of freedom $\nu_K = 2.1$, a variance of $\tilde{\varepsilon}_{it}^K$ can be defined so that the estimates from different distributional assumptions can be compared. It will be explained in greater detail later on.

3.4 Data

To estimate the model described in the previous section, I use the dataset from Statistics Indonesia (Badan Pusat Statistik). The data source is the Annual Manufacturing Survey (Survei Tahunan Perusahaan Industri) from 1997 to 2005. It is an annual survey of manufacturing establishments. Establishments with more than twenty employees are required to fill out the survey form. The fact that the sample contains only establishments with more than twenty employees results in a sample selection issue. There are some establishments that go in and out of the sample. These observations are not missing at random but rather systematically.

Employment L_{it} is the average number of employees per working day in year t. A change in employment H_{it} is measured as the difference between L_{it} and L_{it-1}. Note that this measure of H_{it} can only identify a net flow of employment. A gross flow of employment cannot be measured since the data do not contain voluntary quits. Therefore, it is not appropriate to call H_{it} "hiring" (or "firing") in this case.[11]

Wage rate w_{it} is calculated by dividing total compensation to workers, which includes overtime payment and other bonuses, by the number of employees, L_{it}. Output Y_{it} is measured as the reported figure of value added, sales minus expenditures on raw materials, energy, and fuel. These nominal values are deflated to 1993 prices by the wholesale price index.

Capital stock K_{it} is measured by the perpetual inventory method. Gross investment (disinvestment) I_{it} is measured by purchases net of sales of fixed capital: buildings, machinery, vehicles, and other capital goods. Land is excluded from fixed capital in this study. Gross investment is deflated by appropriate price indexes.

To reduce noise in the data, I remove outliers: the observations with Y_t/L_t, w_t, or Y_t/K_t above the 99 percentile or below the 1 percentile. I also discard the obser-

[11] Put it in another way, δ_L is assumed to be zero.

Table 3.1: Summary Statistics

Variables	Mean	Standard Deviation	Median	Min	Max
H_t/L_{t-1}	0.0088	(0.2180)	0	-0.9685	1.6092
I_t/K_{t-1}	0.0429	(0.1598)	0	-0.9650	1.8145
Y_{t-1}/L_{t-1}	7.0140	(12.2251)	3.0908	0.0010	149.9719
w_{t-1}	1.9175	(1.6092)	1.4949	0.0612	13.8322
Y_{t-1}/K_{t-1}	2.2225	(4.2957)	0.9949	0.0001	64.8688
L_{t-1}	213.3298	(665.6407)	51	20	34,980
K_{t-1}	4827.36	(3,6312.55)	194.0392	0.0997	2,804,275

Note: The monetary variables w_{t-1} and K_{t-1} are in 1,000,000 Indonesian Rupiah in 1993 price. The exchange rate against US dollar in 1993 was approximately 2,100 Indonesia Rupiah per US dollar.

vations if the employment rate H_t/L_{t-1} or the investment rate I_t/K_{t-1} exceeds the 99 percentile. Finally, I restrict the sample only to plants with eight usable observations. So, the data are balanced panel data of 60,440 observations (8 time periods of 7,555 plants) in total.[12] When estimating with an unbalanced panel of all usable observations, the results are similar to those from the balanced panel. Therefore, I keep using the balanced panel so that it reduces the time for bootstrapping to some extent. See Appendix 3.A for more details on data cleaning and sample construction.

Table 3.1 shows the summary statistics of key variables. The means of H_{it}/L_{it-1} and I_{it}/K_{it-1} are almost zero, which are mainly due to high frequencies of regimes of no change in employment and zero investment. Table 3.2 shows frequencies of nine different decision regimes. Three decision regimes regarding labor adjustment are rather evenly observed. The relative frequency of employment expansion is lower than the relative frequencies of other two decision regimes by about 5%. On the other hand, the frequencies of three decision regimes of capital adjustment are not even. The regime of zero investment accounts for about three quarters of the sample. The frequency of disinvestment is remarkably low and raises concerns about identification.

The means of H_{it}/L_{it-1} and I_{it}/K_{it-1} are reported in the next section with the

[12]The estimation strategy in this study drops the first time period, the year of 1997.

Table 3.2: Frequencies of Adjustment Decisions

		I_t < 0	$= 0$	> 0	
	< 0	508 (0.84)	14,876 (24.61)	5,509 (9.11)	20,893 (34.57)
H_t	$= 0$	275 (0.45)	18,105 (29.96)	3,225 (5.34)	21,605 (35.75)
	> 0	312 (0.52)	12,073 (19.98)	5,557 (9.19)	17,942 (29.69)
		1,095 (1.81)	45,054 (74.54)	14,291 (23.64)	60,440 (100.00)

Note: The figures in brackets represent relative frequencies.

predictions from the model. The next section presents the estimation results of the model using this Indonesian dataset.

3.5 Results

This section shows the estimation results of the model. The predictions from the model are also presented.

3.5.1 Parameter Estimation Results

Table 3.3 reports the estimation results from the baseline assumption of bivariate normality. The upper half of the table shows the estimates of $\tilde{\gamma}$, the parameters to approximate the shadow values q_{it}^L and q_{it}^K. All the variables used to approximate the shadow values are statistically significant, including the square terms among these variables. The significance of the squared terms indicates that the variables are nonlinearly associated with the shadow values.

The lagged value of the output-labor ratio Y_t/L_t is positively associated with the shadow value of labor, and thus, with the probability of employment expansion for

Table 3.3: Estimation Result: Bivariate Normal

A. The Estimates of $\tilde{\gamma}$ [a]	Labor		Capital	
Y_{t-1}/L_{t-1}	0.01018		0.00566	
	(0.00098)	[10.35]	(0.00114)	[4.95]
$(Y_{t-1}/L_{t-1})^2$	-0.00007		-0.00003	
	(0.00001)	[7.10]	(0.00001)	[2.64]
w_{t-1}	0.05998		0.10261	
	(0.00877)	[6.84]	(0.01021)	[10.05]
${w_{t-1}}^2$	-0.00445		-0.00413	
	(0.00092)	[4.87]	(0.00105)	[3.92]
Y_{t-1}/K_{t-1}	-0.01149		-0.03672	
	(0.00219)	[5.24]	(0.00326)	[11.24]
$(Y_{t-1}/K_{t-1})^2$	0.00019		0.00056	
	(0.00006)	[3.21]	(0.00008)	[6.54]

Correlation Coefficient ρ	0.07487	(0.00607)	[12.33]
Kendall's τ	0.04771		
Log Likelihood	-101758		
Number of Observations	60440		

B. The Estimates of the Structural Parameters [b]

	Labor			Capital		
$\tilde{a}^+$	0.40667	[0.26930	0.58846]	0.35259	[0.18153	0.56779]
$\tilde{a}^-$	0.63085	[0.55899	0.71559]	1.87040	[1.30719	2.55417]
$\tilde{c}^+$	2.29560	[2.14847	2.44760]	4.99242	[4.74007	5.45760]
$\tilde{c}^-$	7.35974	[7.12500	7.60973]	5.96579	[5.25323	6.78781]
$\tilde{b}^+ + \tilde{b}^-$	0.71766	[0.67687	0.75451]	2.28557	[2.21849	2.36146]

[a] Industry dummies (23 industries in total), ownership dummies, and year dummies are included in Z_{it}, but not reported here to save space. The values in round brackets and square brackets are asymptotic standard errors and t-ratios in absolute value, respectively.

[b] The values in square brackets are the lower and upper bounds of the bootstrap 95% confidence intervals. The intervals are based on 500 bootstrap resamples clustered by plants. The estimates of the structural parameters are ratios to σ_L and σ_K.

most of the observations. Even though it exhibits concavity (the negative sign of the squared term), a critical point where the effect changes from positive to negative, which is about 72.71, is much greater than the median of the variable, 3.09. Likewise, the wage is positively associated with the probability of an increase in employment for most of the observations. The lagged value of wage in the labor equation has a positive sign. Even though the squared term is negative, the effect of wage becomes negative when it exceeds 6.74. The median of wage w_{t-1} is 1.50. In contrast, the lagged value of the ratio of output to capital stock Y_t/K_t has a negative effect on the probability of investment for most of observations. The same relations between the shadow value of capital and those variables are found. The shadow values can be related to those variables in more complex matters. It may be more preferable to approximate the shadow values nonparametrically.

The error terms ε_{it}^L and ε_{it}^K are positively correlated as the positive value of the estimated correlation coefficient shows. The correlation is statistically significant. Even though it is not reported in the table, a likelihood ratio test also rejects the hypothesis that there is no correlation between these error terms.[13] The significance of the correlation coefficient implies that the adjustment model should be estimated jointly. Nevertheless, the correlation coefficient is remarkably small. The error terms ε_{it}^L and ε_{it}^K are only weakly correlated. The same weak correlation is also found by Asphjell et al. (2010).

The main purpose of this study is to estimate the structure of the adjustment cost function. The lower panel of Table 3.3 reports the estimates of the structural parameters. The structural parameters are recovered from the two-step procedures. From the first-step ML estimation, the combinations of the structural parameters are obtained: the multiplications of fixed and quadratic components. Then, each of the expansion and contraction equations for labor (3.6) and for capital (3.7) is estimated by OLS in order to obtain (the inverses of) quadratic components: $\tilde{c}_L^+$, $\tilde{c}_L^-$, $\tilde{c}_K^+$, and $\tilde{c}_K^-$.[14] With these OLS estimates and the results from the first step estimation, the

[13] The log likelihood of the unrestricted model is reported in Table 3.3. The log likelihood of the restricted model is the sum of the log likelihoods of the separated estimation of labor and capital adjustments: -65,480 for labor and -36,352 for capital. The likelihood ratio test has a limiting χ^2 distribution with one degree of freedom. The test statistic is 148, and its p-value is 0.00. The hypothesis can be rejected at any conventional level of significance.

[14] In total, there are four OLS estimations, and each OLS is estimated without a constant term under the assumption of mean-zero η's. The results of the OLS estimations are suppressed to save space. All the coefficients are statistically significant, and R^2 ranges from 0.199 to 0.486.

fixed components $\tilde{a}_L^+$, $\tilde{a}_L^-$, $\tilde{a}_K^+$, and $\tilde{a}_K^-$ are identified. The linear components $\tilde{b}_L^+$, $\tilde{b}_L^-$, $\tilde{b}_K^+$, and $\tilde{b}_K^-$ cannot be separated from $\tilde{\gamma}_L$ and $\tilde{\gamma}_K$. Only $\tilde{b}_L^+ + \tilde{b}_L^-$ and $\tilde{b}_K^+ + \tilde{b}_K^-$ can be obtained, which are reported in Table 3.3 to give the idea of their sizes. Note that the parameters are estimated only relative to the scale parameters, σ_L and σ_K.

For statistical inferences, the bootstrap 95% confidence intervals are reported. Resampling is replicated 500 times. To reduce the concern about a plant specific effect, the bootstrap resampling is clustered by plants. The lower and upper bounds of the bootstrap 95% confidence interval are 2.5 and 97.5 percentiles, respectively, of boostrap estimates of each parameter. The confidence intervals for all the structural parameters do not contain zero. Therefore, the structural parameters, α's, are statistically significantly different from zero. It is evidence for nonconvexities of the adjustment costs for both labor and capital in line with the evidence from other studies with microdata.

The estimated adjustment cost functions of labor and capital exhibit asymmetry. The point estimate of $\tilde{a}_L^-$ is larger than that of $\tilde{a}_L^+$. The confidence interval for each parameter does not include the point estimate of the other parameter. The bootstrap confidence interval of the difference $\widehat{\widetilde{a}}_L^+ - \widehat{\widetilde{a}}_L^-$ is also estimated even though it is not reported. I can reject the hypothesis that $\widetilde{a}_L^+ = \widetilde{a}_L^-$ at the 5% significance level since the 95% confidence interval does not contain zero although I cannot at the 1% significance level. Similarly, $\widehat{\widetilde{c}}_L^-$ exceeds $\widehat{\widetilde{c}}_L^+$. The difference is statistically significant even at the 1% significant level, based on the 99% bootstrap confidence interval. These findings indicate that a reduction of employment level is more costly than an expansion of employment level. Considering strict employment protection in Indonesia, it is reasonable to think that it is more difficult to fire than hire.

The adjustment costs of capital show the same pattern. The components for contraction, $\widehat{\widetilde{a}}_K^-$ and $\widehat{\widetilde{c}}_K^-$, are larger than the components for expansion, $\widehat{\widetilde{a}}_K^+$ and $\widehat{\widetilde{c}}_K^+$. The 99% confidence interval for $\widehat{\widetilde{a}}_K^+ - \widehat{\widetilde{a}}_K^-$ does not contain zero, and the 95% confidence

Table 3.4: First-Step Log Likelihood Values under Various Copulas

Copula	AMH	Clayton	Frank	Gumbel	Joe	Gaussian
Log Likelihood	-101224	-101263	-101214	-101101	-101365	-101208

Note: See Appendix 3.C for the first-step ML estimation with a copula method

interval for $\widetilde{\widetilde{c}}_K^+ - \widetilde{\widetilde{c}}_K^-$ does not contain zero. The results imply that disinvestment is more costly than investment. It may be that the market for secondhand capital goods is not well developed in Indonesia.

There is no directly comparable evidence of asymmetric costs from developed countries. Previous empirical studies of capital adjustment tend to assume a symmetric adjustment cost function. There is sometimes a data issue on disinvestment so that the studies can only look at nonnegative investment. For example, Letterie and Pfann (2007) and Polder and Verick (2004) use datasets from the Netherlands, in which the information on disinvestment is missing. On the other hand, there are several pieces of empirical evidence on a nonconvex capital adjustment cost (Caballero et al., 1995, Cooper and Haltiwanger, 2006). The results in this study are consistent with these findings.

Now I discuss the results from the estimation under the alternative distribution assumption. The marginal distributions of the error terms ε_{it}^L and ε_{it}^K are t distributions with degrees of freedom $v_L = 3$ and $v_K = 2.1$. The joint distribution is constructed using the copula method. As mentioned above, the entire estimation procedure is implemented only for the Archimedean copula attaining the largest log likelihood value in the first-step ML estimation.

Table 3.4 shows the maximized values of the log likelihood functions in the first-step ML estimation. Among the Archimedean copulas considered here, the Gumbel copula attained the largest log likelihood value. Therefore, I choose the Gumbel copula as the best-fitting copula and proceed to the second-step OLS regressions using

Table 3.5: Estimation Result: Gumbel Copula

A. The Estimates of $\tilde{\gamma}$ [a]	Labor		Capital	
Y_{t-1}/L_{t-1}	0.01163		0.00051	
	(0.00112)	[10.41]	(0.00157)	[0.11]
$(Y_{t-1}/L_{t-1})^2$	-0.00009		0.00001	
	(0.00001)	[7.18]	(0.00002)	[0.57]
w_{t-1}	0.07750		0.1582	
	(0.01018)	[7.61]	(0.01469)	[10.77]
${w_{t-1}}^2$	-0.00593		-0.00639	
	(0.00106)	[5.58]	(0.00144)	[4.44]
Y_{t-1}/K_{t-1}	-0.01303		-0.04690	
	(0.00245)	[5.32]	(0.00577)	[8.12]
$(Y_{t-1}/K_{t-1})^2$	0.00022		0.00062	
	(0.00007)	[3.26]	(0.00016)	[3.85]
Dependence Parameter θ	1.06522	(0.00382)		
Kendall's τ	0.06122			
Log Likelihood	-101101			
Number of Observations	60440			

B. The Estimates of the Structural Parameters [b]

	Labor			Capital		
$\tilde{a}^+$	0.30055	[0.20558	0.43336]	1.31109	[0.91732	1.74814]
$\tilde{a}^-$	0.56575	[0.50608	0.64273]	6.59896	[4.47123	9.08197]
$\tilde{c}^+$	4.28611	[4.10423	4.46573]	11.69586	[10.92279	12.45721]
$\tilde{c}^-$	10.40325	[10.14817	10.70708]	61.03106	[53.12401	70.67687]
$\tilde{b}^+ + \tilde{b}^-$	0.81188	[0.76927	0.86040]	3.52051	[3.24138	3.86146]

See notes in Table 3.3.

the estimates from the first-step ML estimation using the Gumbel copula. The Gumbel copula allows only positive dependence and exhibits strong right tail dependence and weak left tail dependence (Trivedi and Zimmer, 2007b).[15] That is, the result that the Gumbel copula is fitting best implies that ε_{it}^L and ε_{it}^K are strongly correlated at positive values but less correlated at negative values. For comparison, I also implement the entire estimation using the Gaussian copula.

Tables 3.5 and 3.6 show the estimation results using Gumbel and Gaussian cop-

[15]See Appendix 3.C for the functional form of the Gumbel copula.

Table 3.6: Estimation Result: Gaussian Copula

A. The Estimates of $\tilde{\gamma}$ [a]	Labor		Capital	
Y_{t-1}/L_{t-1}	0.01159		0.00037	
	(0.00112)	[10.38]	(0.00157)	[0.24]
$(Y_{t-1}/L_{t-1})^2$	-0.00009		0.00001	
	(0.00001)	[7.16]	(0.00002)	[0.65]
w_{t-1}	0.07407		0.15499	
	(0.01018)	[7.28]	(0.01465)	[10.57]
${w_{t-1}}^2$	-0.0056		-0.00608	
	(0.00106)	[5.28]	(0.00144)	[4.24]
Y_{t-1}/K_{t-1}	-0.01282		-0.04662	
	(0.00245)	[5.23]	(0.00579)	[8.05]
$(Y_{t-1}/K_{t-1})^2$	0.00021		0.00061	
	(0.00007)	[3.21]	(0.00016)	[3.73]
Dependence Parameter θ	0.07221	(0.00608)		
Kendall's τ	0.04601			
Log Likelihood	-101208			
Number of Observations	60440			

B. The Estimates of the Structural Parameters [b]

	Labor			Capital		
$\tilde{a}^+$	0.27834	[0.18772	0.40045]	1.33163	[0.93839	1.77052]
$\tilde{a}^-$	0.55734	[0.49940	0.63367]	6.58129	[4.45585	9.07430]
$\tilde{c}^+$	4.31704	[4.13371	4.50253]	11.73277	[10.94405	12.48949]
$\tilde{c}^-$	10.37269	[10.11636	10.67567]	61.02205	[53.11470	70.60107]
$\tilde{b}^+ + \tilde{b}^-$	0.80791	[0.76550	0.85604]	3.52055	[3.24081	3.85860]

See notes in Table 3.3.

ulas, respectively. The tables report the estimates of dependence parameters θ. The dependence parameter θ governs the degree of dependence. The parameter cannot, however, be directly compared across copulas. For the Gumbel copula, underlying random variables are independent when $\theta = 1$. On the other hand, $\theta = 0$ indicates independence for the Gaussian copula. Therefore, we report Kendall's τ as a measure of dependence.[16] Kendall's τ ranges from -1 to 1, and when it takes a value of zero,

[16]The most familiar measure of dependence is the correlation coefficient. The correlation coeffi-

random variables are independent.

For Gumbel and Gaussian copulas (and also for other Archimedean copulas), Kendall's τ can be computed with the dependence parameter θ: $\tau = \frac{\theta-1}{\theta}$ for the Gumbel copula, and $\tau = 2\pi^{-1}\sin^{-1}(\theta)$ for the Gaussian copula. In each case, the estimated measure is close to 0, which indicates weak dependence. As shown earlier, the weak dependence (correlation) between ε_{it}^{L} and ε_{it}^{K} is also found under the assumption of bivariate normality. Table 3.3 reports Kendall's τ in addition to the correlation coefficient, and it is also close to zero.

When comparing the Gumbel and Gaussian copula models, we see the results are essentially the same. This is reasonable given the finding that Kendall's τ are small. Although the contours of a pdf from those copulas have very different shapes when Kendall's τ is large, the contours become more and more circular as Kendall's τ approaches zero (Smith, 2003). At the values close to zero like the estimates here, the contours have essentially the same shape. Therefore, there is no difference in the estimation results between these copulas.

In comparison with the bivariate normal model, the estimations by the copula method return similar results of the estimates of $\widetilde{\gamma}_L$. On the other hand, the coefficients $\widetilde{\gamma}_K$ are estimated differently. The coefficients on Y_{t-1}/L_{t-1} and its square become almost zero and insignificant. The different distribution assumptions also affect the estimation of the structural parameters. For labor adjustment, the fixed components are now smaller while the quadratic components become larger. For capital adjustment, all components are now larger. The estimates of $\widetilde{c}_K^-$ are now more than ten times as large as the estimate from the assumption of bivariate normality.

However, the direct comparisons of the parameter estimates across the models can be misleading since the parameters are only estimated proportional to the scale

cient is limited in that it only measures linear dependence and it is not invariant under nonlinear transformations of random variables (Trivedi and Zimmer, 2007b). Kendall's τ does not have such limitations.

Table 3.7: Reparameterized Estimates of Adjustment Cost Parameters

	Labor			Capital		
	B.N.[a]	Gumbel[b]	Gaussian[b]	B.N.[a]	Gumbel[b]	Gaussian[b]
$\tilde{a}^{+}$	0.40667	0.17352	0.16070	0.35259	0.28610	0.29059
$\tilde{a}^{-}$	0.63085	0.32664	0.32178	1.87040	1.44001	1.43616
$\tilde{c}^{+}$	2.29560	2.47459	2.49244	4.99242	2.55225	2.56030
$\tilde{c}^{-}$	7.35974	6.00632	5.98868	5.96579	13.31807	13.31610
$\tilde{b}^{+}+\tilde{b}^{-}$	0.71766	0.46874	0.46645	2.28557	0.76824	0.76825

[a] B.N. stands for bivariate normal.

[b] The estimated parameters in Tables 3.5 and 3.6 are reparameterized by dividing by $\sqrt{3}$ for labor and $\sqrt{21}$ for capital.

parameters σ_L and σ_K. In the case of normal distribution, σ_J can be interpreted as a standard deviation of ε_{it}^{J}. However, if ε_{it}^{J} has a t distribution with degrees of freedom ν_J, the standard deviation is not σ_J but $\sigma_J\sqrt{\frac{\nu_J}{\nu_J-2}}$. Therefore, $\widetilde{\varepsilon}_{it}^{J}$ has a standard deviation of $\sqrt{\frac{\nu_J}{\nu_J-2}}$. To make the estimates from the bivariate normal model and the copula models comparable, I reparameterize the copula models so that $\widetilde{\varepsilon}_{it}^{J}$ has a standard deviation (and a variance) equal to 1. Since this does not affect probabilities and thus a likelihood function, this operation is equivalent to dividing the already estimated parameters by $\sqrt{\frac{\nu_J}{\nu_J-2}}$.

Table 3.7 shows that the reparameterized structural parameters from the copula models. It also shows the estimated parameters from the bivariate normal model, which are the same as Table 3.3. For the parameters of the labor adjustment cost function, the estimates from the copula models are smaller than those from the bivariate normal model except a positive quadratic component. For the capital adjustment, the estimates of $\tilde{c}^{-}$ from the copula models are more than twice as large as the one from the bivariate normal model.

Despite the differences in the magnitudes of the estimated parameters, I still find nonconvexity in the adjustment costs as in the case of bivariate normality. The adjustment costs also are found to be asymmetric. The differences between positive and negative components of the adjustment costs are significant at the 1% significance level, based on the bootstrap confidence intervals.

These comparisons across the different distribution models imply that the assumption of marginal distributions plays a more important role than the assumption of joint distribution in the estimation of this study. The comparison of the log likelihood values indicates that the model with t distributions as margins (Table 3.6) fits better than the model with normal margins (Table 3.3). Next, in order to see how well the models fit the data, I calculate the predictions of the optimal employment and investment rates H_{it}/L_{it-1} and I_{it}/K_{it-1} and compare them with the sample means.

3.5.2 Model Predictions

Table 3.8 shows the sample and prediction of unconditional and conditional means of H_{it}/L_{it-1} and I_{it}/K_{it-1}. The first column shows that the corresponding means from the samples. The second, third and fourth columns show the model predictions from the different distribution assumptions. See Appendix 3.D for how I calculate the predictions.

First of all, the model predicts both the unconditional mean of H_{it}/L_{it-1} and its means conditional on labor adjustment decisions considerably accurately under each distribution assumption. The predicted unconditional means are quite close to the sample mean. In terms of the relative size of the prediction to the sample, the model prediction under the bivariate normality is slightly superior to the the model predictions with the copulas. The relative size is approximately 0.98 from the bivariate normal model while the relative sizes are around 0.94 and 0.96 from the Gumbel and Gaussian models, respectively.

On the other hand, for the conditional means, the model predictions by the copulas are closer to the sample counterparts than the bivariate normal model. The relative sizes of the prediction from the copula models to the sample means are around 99% for both of the conditional means from each copula model. Still, the bivariate normal

Table 3.8: The Sample and Predicted Means of H_{it}/L_{it-1} and I_{it}/K_{it-1}

	Sample		Prediction [a]	
		B.N.	Gumbel	Gaussian
The means of H_{it}/L_{it-1}				
Unconditional	0.0088	0.0086	0.0083	0.0084
Conditional on				
$H_{it} > 0$	0.2029	0.1961	0.2002	0.2005
$H_{it} < 0$	-0.1489	-0.1463	-0.1469	-0.1470
$H_{it} > 0$ & $I_{it} > 0$	0.2117	0.2073	0.2421	0.2122
$H_{it} > 0$ & $I_{it} = 0$	0.1990	0.1926	0.1831	0.1924
$H_{it} > 0$ & $I_{it} < 0$	0.1984	0.1746	0.1668	0.1769
$H_{it} < 0$ & $I_{it} > 0$	-0.1365	-0.1423	-0.1443	-0.1404
$H_{it} < 0$ & $I_{it} = 0$	-0.1532	-0.1472	-0.1474	-0.1485
$H_{it} < 0$ & $I_{it} < 0$	-0.1562	-0.1539	-0.1432	-0.1583
The means of I_{it}/K_{it-1}				
Unconditional	0.0429	0.0257	0.0365	0.0365
Conditional on				
$I_{it} > 0$	0.1899	0.0879	0.1949	0.1947
$I_{it} < 0$	-0.1081	-0.1390	-0.1369	-0.1369
$I_{it} > 0$ & $H_{it} > 0$	0.2193	0.1587	0.2131	0.1988
$I_{it} > 0$ & $H_{it} = 0$	0.1610	0.1546	0.1780	0.1983
$I_{it} > 0$ & $H_{it} < 0$	0.1770	0.1508	0.1841	0.1805
$I_{it} < 0$ & $H_{it} > 0$	-0.1025	-0.1370	-0.1185	-0.1087
$I_{it} < 0$ & $H_{it} = 0$	-0.1284	-0.1384	-0.1373	-0.1343
$I_{it} < 0$ & $H_{it} < 0$	-0.1006	-0.1400	-0.1369	-0.1329

[a] See Appendix 3.D for a description of how the predicted values are calculated.

model predictions seem close enough: the relative size is 0.97 for the mean conditional on $H_{it} > 0$ and 0.98 for $H_{it} < 0$.

The predictions of the means of H_{it}/L_{it-1} conditional on the joint adjustment decisions are also moderately accurate. The accuracy of the predictions vary: there is not any particular model that predicts all of the conditional means more accurately than the other two models. It should be mentioned that the conditional means from the copula models are numerically computed while those from the bivariate normal model are computed analytically (see Appendix 3.D).

In contrast to the predictions of employment rates H_{it}/L_{it-1}, the models, especially the bivariate normal model, end up with relatively poor predictions of the means of investment rates I_{it}/K_{it-1}. The predicted unconditional mean from the bivariate normal model is almost half as large as the sample counterpart. The prediction improves with the copula method. The relative sizes of the prediction to the sample are now about 0.85 for both Gumbel and Gaussian models.

The improvement apparently stems from the improvement of the predicted mean conditional on $I_{it} > 0$. The predictions of I_{it}/K_{it-1} conditional on $I_{it} < 0$ by all of the models are almost the same, and these are overestimating the sample counterpart in absolute value by about 30 percent. The bivariate normal model underestimates the mean conditional on $I_{it} > 0$ considerably. The predictions from the Gumbel and Gaussian copula models are reasonably accurate even though overestimated to some extent. The improvement in the prediction of this conditional mean can be explained by the different assumptions of the distribution of $\widetilde{\varepsilon}_{it}^{K}$. A t distribution has thicker tails compared to a normal distribution. Therefore, the truncated mean of $\widetilde{\varepsilon}_{it}^{K}$ is larger under a t distribution than a normal distribution. The difference in the estimates of $\widetilde{\gamma}_K$ (the coefficients on Y_{t-1}/L_{t-1} may also be important. Given the smaller estimates from the copula models, the threshold value $\widetilde{A}_{Kit}^{+}$ tends to be larger in the copula models than the bivariate normal model. It could make the truncated mean in a right tail even larger. On the other hand, the larger threshold $\widetilde{A}_{Kit}^{-}$ can cancel out the effect of a thicker negative tail. That may be why the predicted mean conditional on $I_{it} < 0$ are similar between the bivariate normal model and the copula models.

Thicker tails of a t distribution also indicate that an extreme value is more likely to realize in a t distribution than a normal distribution. There are several pieces of empirical evidence that investment can be really large even though small investment also occurs (for example, Doms and Dunne 1998 and Sakellaris 2004). In this circumstance, a t distribution is likely to fit better than a normal distribution.

The coexistence of small and large investment episodes may be due to heterogeneity of capital goods. Investment expenditures such as maintenance/repairs can be small but more likely to occur. On the other hand, expenditures such as new construction of buildings are less likely to occur, but these are noticeably large when they occur. Letterie and Pfann (2007) construct an empirical model for the two different investment regimes, which switch endogenously. Their model is similar to the model in this study, but the model in this study ignores this heterogeneity to avoid additional complexity.

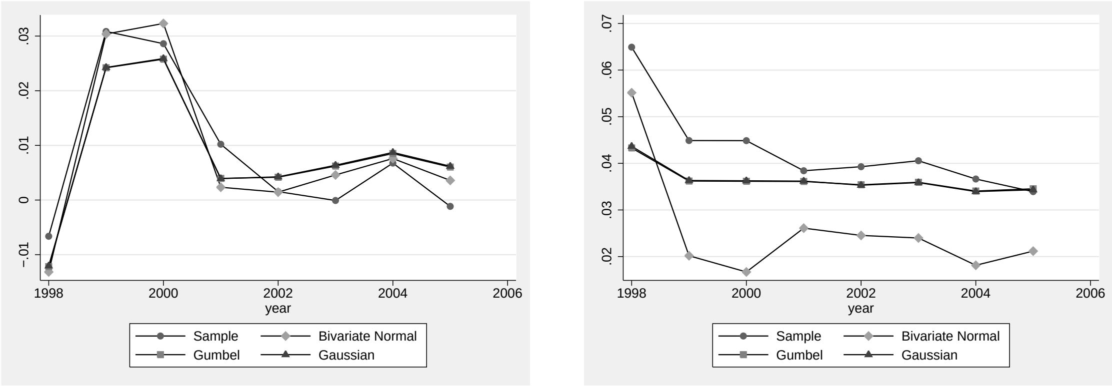
For the conditional means of I_{it}/K_{it-1} on joint factor adjustment decisions, the model predictions fails to yield the revealed pattern in the sample. In the bivariate normal model, the predicted mean conditional on $I_{it} > 0$ and $H_{it} > 0$ is larger than the conditional mean on $I_{it} > 0$ and $H_{it} = 0$, which in turn is larger than the conditional mean on $I_{it} > 0$ and $H_{it} < 0$ given the positive correlation between ε_{it}^L and ε_{it}^K.[17] A similar pattern can be seen in the capital contraction with three labor adjustments. However, the sample means do not follow this pattern. Even though the predictions from the copula models reveal this non-monotonic pattern, the numerical computation of the means are sensitive to the seed of random-number, and the pattern varies with different seeds.[18]

In this model, the interrelation between the adjustment decisions of labor and capital is only due to the dependence between the two random variable ε_{it}^L and ε_{it}^K. However, there is the possibility that the adjustment cost functions are interrelated.

[17]If the correlation is negative, the order of sizes reverses.

[18]The number of draws is the same as the number of observations, that is, 60,440. An increase in the number of draws did not solve this problem.

Figure 3.1: The Unconditional Means of H_{it}/L_{it-1} and I_{it}/K_{it-1}



(a) H_{it}/L_{it-1} (b) I_{it}/K_{it-1}

The sample means of I_{it}/K_{it-1} indicate that investments tend to be large when labor is adjusted simultaneously. By adjusting at the same time, a plant may be able to reduce the adjustment costs relative to the case where a plant adjust factors separately because a plant does not need to interrupt the production process over and over. Asphjell et al. (2010) assume that there is an interrelated fixed component in the cost function. Their empirical finding with Norwegian plant data is that the sign of the cost of interrelation is negative. It implies that a plant has an incentive to adjust both labor and capital simultaneously in order to reduce the adjustment costs.[19] However, their model assumes a symmetric adjustment cost function unlike the model in this study. With asymmetry, modeling the interrelated adjustment costs is complicated and this study leave it as a task for future research.

Overall, the predictions of the means of H_{it}/L_{it-1} do not change much by the assumption of marginal distributions. On the other hand, the means of I_{it}/K_{it-1} do change considerably with the distribution assumptions. Figure 3.1 shows the sample means and predicted means of H_{it}/L_{it-1} and I_{it}/K_{it-1} by year. The unconditional means of H_{it}/L_{it-1} are plotted by year on the left. This figure shows that there is

[19]Polder and Verick (2004) assume that there is an interrelated linear components and Eslava et al. (2010) also make a similar assumption. However, these studies do not estimate structural parameters.

no considerable difference between the sample and the predictions and no difference across the models. The predictions from Gumbel and Gaussian copulas completely overlap. The predictions of I_{it}/K_{it-1} are better with the copulas, which are again coincident, than the bivariate normal model as the right figure shows. This result reinforces the speculation that a marginal distribution matters in modeling, especially for capital adjustment. The results are in favor of t-distributions.

3.6 Changes in Labor Market Regulations

As the last exercise in this study, I investigate the effect of the change in labor market regulations in 2003 on the structural parameters and its effect on employment adjustments. In 2003, with the passage of the Manpower Law (No.13/2003), Indonesia tightened the labor market regulations and make it more difficult for employers to fire workers. The law covers a range of labor issues and consists of a number of clauses. Among them, the clauses that have been controversial are related to severance payment, the use of temporary contracts (fixed-term contracts), and the setting of minimum wages (Manning and Roesad, 2007).

The law raised the severance payment rates, and the severance rates in Indonesia are high by international standards. Moreover, the law also made it complicated to determine the rates: the rates are not only based on years of service but also on the reasons for firing. The complicated way to determine the rates makes it more difficult for employers to anticipate total employment costs (World Bank, 2010). Since severance payment rates are based on wages, the rates can be affected by an increase in minimum wages. Legal restrictions on the use of temporary contracts reduce flexibility in Indonesian labor markets by limiting employers' employment behavior in response to changes in economic conditions. One of the main business concerns and main political issues in Indonesia is tightened labor market regulations. See Manning and Roesad (2007) for more details of the law and related discussions.

It is natural to raise a question of whether and how this law affects the parameters of the adjustment costs. It is reasonable to expect that tighter regulations increase the structural parameters of the adjustment cost function of labor. The effects on the costs of capital adjustment are not clear.

In order to examine how the law affects the adjustment cost parameters, I first split the sample between the years earlier than 2003 and the years of 2003 and later, and then, compare the estimated parameters of these two subsamples. The estimation strategy, however, does not allow identification of the structural parameters apart from the scale parameters σ_L and σ_K. If the subsamples are estimated separately, it is not possible to identify whether the difference in the estimated parameters is due to changes in the structural parameters or due to changes in the scale parameters. To avoid this problem, I estimate the structural parameters from the split sample simultaneously by assuming that the scale parameters do not change.

More specifically, first I create a dummy for the years up to 2002 and a dummy for the years from 2003 onward. Then, I interact the dummies with all the variables in the estimation (both in the first-step ML estimation and the second-step OLS), and I estimate the model with the interacted variables simultaneously. There is only one likelihood function to maximize and there are four OLS equations (expansion and contraction for each of labor and capital) as before, but the number of variables doubles (less the number of year dummies). With the assumption that the scale parameters σ_L and σ_K are stable across the split samples, it can be said that changes in the parameter estimates are due to changes in the structural parameters.

Table 3.9 shows the results from the bivariate normal and Gumbel-t models; Gaussian-t results are almost the same as Gumbel-t and these are suppressed.[20] Since

[20] A likelihood ratio test rejects the hypothesis that the parameters to estimate in the first-step ML estimation are equal across the two periods (before and after 2003) at any level of significance for each model.

Table 3.9: The Structural Parameters Before and After the Year of 2003

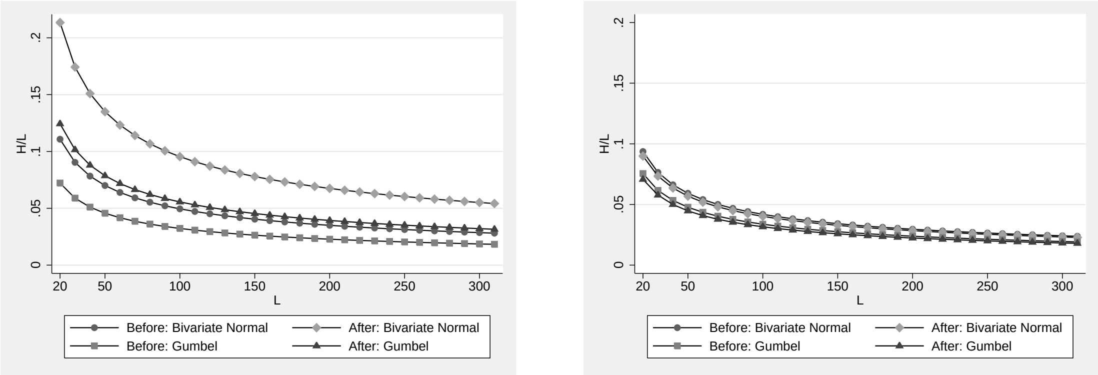
	Bivariate Normal				Gumbel			
	before	after	difference		before	after	difference	
Labor								
$\widetilde{a}_L^+$	0.3021	0.8687	0.5667	***	0.1297	0.3702	0.2405	***
$\widetilde{a}_L^-$	0.6638	0.5669	-0.0969		0.3442	0.2989	-0.0453	
$\widetilde{c}_L^+$	2.4622	1.9071	-0.5551	***	2.4881	2.3937	-0.0945	
$\widetilde{c}_L^-$	7.5674	7.0103	-0.5771	***	6.0430	5.9736	-0.0694	
$\widetilde{b}_L^+ + \widetilde{b}_L^-$	0.5387	1.0633	0.5246	***	0.3488	0.7100	0.3611	***
Capital								
$\widetilde{a}_K^+$	0.5016	0.1954	-0.3062	*	0.2933	0.2726	-0.0206	
$\widetilde{a}_K^-$	1.4566	2.5569	1.1003	**	1.0572	2.0651	1.0080	**
$\widetilde{c}_K^+$	4.2071	7.2822	3.0751	*	2.3635	2.9377	0.5742	***
$\widetilde{c}_K^-$	5.3702	6.8767	1.5064	**	12.4771	14.5712	2.0941	
$\widetilde{b}_K^+ + \widetilde{b}_K^-$	2.3382	2.2112	-0.1270	*	0.8533	0.6516	-0.2017	***

Notes: The figures in parentheses are the differences in the estimates. *,**, and *** indicate that the difference is significant at 10%, 5% and 1% significance levels, repsepctively. The inferences are based on the bootstrap confidence intervals. The estimates from the Gumbel model are rescaled so that these estimates are comparable with those from the bivariate normal model.

the results are almost the same between the Gumbel and Gaussian models, I suppress the results from the Gaussian model here. The differences in the estimated parameters are in round brackets, and the statistical inferences are based on whether the bootstrap confidence intervals of the differences contain zero or not.

$\widehat{\widetilde{c}}_L^+$ and $\widehat{\widetilde{c}}_L^-$ change statistically significantly at the 1% level in the bivariate normal model while the differences in those parameters are not significant in the Gumbel model. Nevertheless, the directions of changes in the parameters are the same across the models. Furthermore, the results from both models show that the change in $\widehat{\widetilde{a}}_L^+$ is significant in magnitude and statistically. The estimate for the later period is almost three times as large as the one for the earlier period, and the difference is statistically significant at the 1% level in both models. The importance in the fixed component of employment expansion increases relative to the convex component. Figure 3.2 shows this point.

Figure 3.2: Relative Size of Fixed and Convex Costs in Labor Adjustment



(a) $H_{it} > 0$

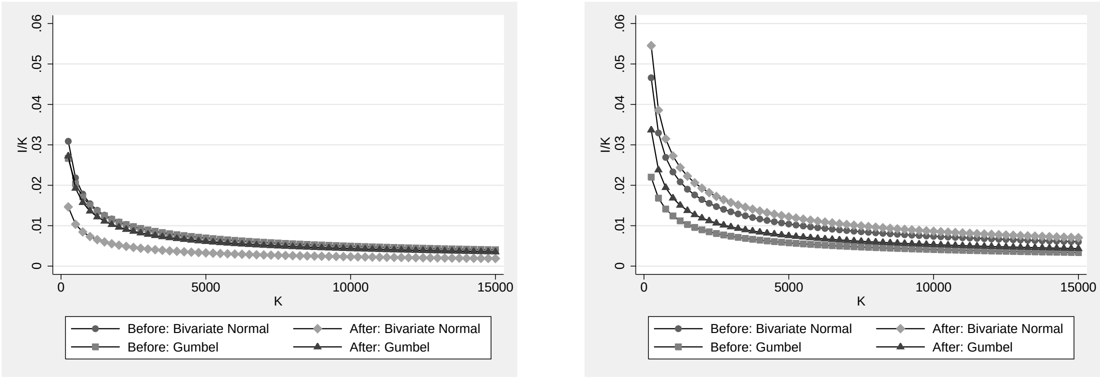
(b) $H_{it} < 0$

Following Nilsen et al. (2007), I plot the combination of values of employment rate H_{it}/L_{it-1} and of employment level L_{it-1} that equalize the sizes of the fixed costs and the convex costs in Figure 3.2. If a point is above the curve, the convex costs are greater than the fixed costs. In both the bivariate normal model and the Gumbel model, the increase in $\widetilde{\widetilde{a}}_L^+$ leads the curve to shift upward in Figure 3.2 (a), indicating the increase in the relative importance of the fixed costs. It is not clear why a_L^+ increases when the labor market regulations change. The possible explanation is that in response to increasing difficulty of firing a worker, a plant sets stricter hiring criteria, which involves more extensive search activities no matter how many workers it plans to hire. It may result in the increase in a_L^+.

On the other hand, the relative importance of the fixed and convex costs does not change in contracting employment levels, as shown in Figure 3.2 (b). Against the expectation, the estimates of $\widetilde{a}_L^-$ and $\widetilde{c}_L^-$ increase even though only the change of either of the parameters is significant.

It is unfortunate that the estimation strategy cannot identify the linear components of the adjustment costs separately. The changes in labor market regulations, especially the increase in severance payment rates, are the most likely to be associated with an increase in b_L^+. In all the models, $\widetilde{\widetilde{b}}_L^+ + \widetilde{\widetilde{b}}_L^-$ for the later period doubles from

Figure 3.3: Relative Size of Fixed and Convex Costs in Capital Adjustment



(a) $I_{it} > 0$ (b) $I_{it} < 0$

the earlier period, and the difference is significant at the 1% level. This result may be the indication of the increase in b_L^+. It is a future task to examine how the law in 2003 affects the linear components by the estimation strategy that is able to identify the linear components separately.

There is no prior expectation about how the adjustment costs of capital change. Changes in all the estimates in the bivariate normal model are statistically significant at some level, and the changes in the estimates in the copula models are also significant except the convex component in contraction $\widetilde{c}_K^-$. The change in $\widetilde{a}_K^-$ is sizable in all the models. The parameter estimate nearly doubles. Even though $\widetilde{c}_K^-$ also increases (insignificantly in the copula models) at the same time, the relative importance of the fixed component in contraction of capital slightly increases (Figure 3.3 (b)). In the bivariate normal model, the fixed component parameter in expansion of capital $\widetilde{a}_K^+$ becomes less than one half. Along with a large increase in the convex component parameter $\widetilde{c}_K^+$, it results in a decline in the relative importance of the fixed costs to the convex costs (Figure 3.3 (a)). On the other hand, there is no apparent change in the relative importance in the copula models. The effects of the labor market policies on capital adjustments need to be analyzed in a more careful manner in future. One possible explanation for changes in capital adjustment parameters is that there may

Table 3.10: Prediction Comparisons in 2003

	Bivariate Normal				Gumbel			
	(1)	(2)	(3)=(1)-(2)		(1)	(2)	(3)=(1)-(2)	
$E(H_{it})$	0.9964	1.6146	-0.6182		1.4245	0.7328	0.6917	
$E(H_{it}\|H_{it} > 0)$	51.80	48.91	2.89	**	53.62	48.76	4.86	***
$Pr(H_{it} > 0)$	0.2449	0.2845	-0.0396	***	0.2414	0.2749	-0.0335	***
$Pr(H_{it} = 0)$	0.4444	0.3041	0.1403	***	0.4443	0.3007	0.1436	***
$E(H_{it}\|H_{it} < 0)$	-26.34	-26.15	-0.19		-25.62	-26.05	0.43	
$Pr(H_{it} < 0)$	0.3107	0.4114	-0.1007	***	0.3143	0.4244	-0.1101	***

Column (1): Predictions with the "after" parameters.

Column (2): Predictions with the "before" parameters.

Notes: **, and *** indicate that the difference is significant at 5% and 1% significance levels, repsepctively, based on the bootstrap confidence intervals. Unconditional expectations and conditional expectations and probabilities are separately calculated and do not match exactly.

be interacted adjustment costs of labor and capital, which are omitted in this study. The policy change may have affected capital adjustment through this channel.

In order to see how the policy change affects employment adjustments, I calculate the predictions with the parameters from the earlier period (the "before" parameters) and those with the parameters from the later period (the "after" parameters). The former can be considered as counterfactual values, which would have realized if the policy had not changed. The comparisons of these predictions show the effect of the policy change. Furthermore, the effect on overall changes in employment is decomposed into four components: expected employment changes conditional on adjustments and probabilities of adjustments. Note, however, that in the model of this study, it is not possible to generate counterfactual flows of all the variables. The predictions for only the year of 2003, which use the same variables from the year of 2002, are comparable. Therefore, the result should be viewed only as temporary effects of the policy change.

Table 3.10 shows the results. To see the effects more clearly, the predictions here are of changes in employees, H_{it}, rather than of employment growth rates, H_{it}/L_{it-1}.[21] Column (1) shows the predictions with the "after" parameters. Column (2) shows the predictions with the "before" parameters, and these are counterfactual predictions.

In both bivariate normal and Gumbel models, the sizes of the effect are small. The difference in changes in employment per plant is only less than one employee on average. Furthermore, the signs of the overall effect are opposite in both models and the effects are not statistically significant in both models.

Interestingly but not surprisingly, the probability of decreasing employment level drops considerably and statistically significantly because of the policy change. The probability with the policy change is 10 percentage points lower than the probability without the policy change. It is reasonable since employment protection regulations were tightened and it became more difficult to fire employees. This result is consistent with the nature of the policy change.

Furthemore, the probability of expanding the level of employment also drops. Since it is more difficult to fire employees in future, a plant is less likely to increase its employment level at a current period. As a consequence, the probability of not adjusting is much higher than it would have been if the policy had not changed. The difference is about 14 percentage points.

However, as mentioned above, this shows only temporary effects of the policy effects in 2003. The policy change needs to be evaluated from more longer time perspectives.

3.7 Conclusion

This study investigates the structure of the adjustment costs of labor and capital simultaneously using the plant-level panel dataset from the Indonesian manufacturing

[21]To obtain the predictions, I simply multiply the predictions of H_{it}/L_{it-1}, which are described in Appendix 3.D in details, by L_{it-1}, which is a predetermined variable.

sector. The results show the nonconvexity and asymmetry of the adjustment cost function in Indonesia similar to the evidence from developed countries such as the U.S., Norway, and so on.

Moreover, I relax the distribution assumption made by the previous studies. In the previous studies, it was standard to assume a normal distribution to implement ML estimation. This study makes the alternative assumption of t distributions. In order to construct the joint distribution, I use the copula method. The estimation results show that the assumption of distribution is important, and t distributions fit the data better than normal distributions. In future research, we can relax the assumption further using more general distributions.

Lastly, I examine the impact of the law that tightened labor market regulations. The results provide empirical evidence that the parameters change in response to the law and that the probability of decreasing the level of employment in 2003 is much lower than the case without the law even though the overall effect on employment adjustments is small. However, it is necessary to examine the effects of the law more carefully using estimation methods that identify the structural parameters fully. Furthermore, this study only examines the effects on the parameters of the adjustment costs and temporary effects on employment adjustments. It is important to investigate the effects on total employment, aggregated investment activities, total factor productivity, and so on. This study is, therefore, a starting point for future research.

3.A Data

This section illustrates the data cleaning and sample construction procedures.

Capital stock K_{it} is measured by the perpetual inventory method:

$$K_{it} = (1 - \delta_K)K_{it-1} + I_{it},$$

where I_{it} is gross investment and δ_K is the depreciation rate. Gross investment (disinvestment) I_{it} is measured by purchase net of sales during period t of fixed capital: buildings, machinery, vehicles, and other.[22] Land is excluded from fixed capital. Purchases or sales of buildings are deflated by the wholesale price index of construction materials for "residential and non residential buildings." Purchases or sales of machinery is deflated by a simple average of the wholesale price indexes of "machinery except electrical equipment" and "electrical machinery, apparatus, and supplies." Purchases or sales of vehicle and other fixed capital are deflated by the wholesale price index of "transportation materials" and the general price index, respectively. These price indexes come from the Monthly Statistical Bulletin of Economic Indicators (Indikator Ekonomi). The base year of the price indexes is 1993. The monetary unit is 1,000,000 Indonesia Rupiah. In the base year of 1993, the exchange rate against US dollar was approximately 2,100 Indonesian Rupiah per US dollar.

The book value of fixed capital in 1997, which is also deflated with the appropriate price index, is used as the initial value of capital stock. There is no official value for a depreciation rate. Schundeln (2007) estimates the depreciation rate to be between 8% and 10% in Indonesia, using the same dataset as in this paper, though covering a different time period. Here, I set $\delta_K = 0.09$. When capital stock turns to be negative, we drop those observations.

To reduce noise in the data, we remove outliers: the observations with Y_t/L_t,

[22] "Purchase" here includes "purchase," "additions," "construction," and "major repair."

w_t, or Y_t/K_t above the 99 percentile or below the 1 percentile. We also discard the observations if the employment rate H_t/L_{t-1} or the investment rate I_t/K_{t-1} exceeds the 99 percentile. When I_t/K_{t-1} is less than -100%, the observation is dropped since this cannot happen in reality. This sort of extreme values does not happen for the employment rate H_t/L_{t-1}.

For estimation, I only use the plants with 8 consecutive usable observations. It results in a balanced panel dataset containing total 60,440 observations (8 time periods of 7,555 plants). In terms of employment level H_t, the sample covers around $36 \sim 39\%$ of the raw data by year. When estimating with an unbalanced panel of all usable observations, the results are similar to those from the balanced panel. Therefore, I keep using the balanced panel so that it reduces the computational time needed for bootstrapping.

The raw data only covers plants with more than 20 employees, which leads to low coverage rates of total employment in the manufacturing sector and the entire nation. In Indonesia, agricultural sector accounts for around 40% of total employment, and manufacturing sector accounts for only 11 to 13%. In manufacturing sector, the size of plants tends to be small so that the raw data only covers 35 to 42% of manufacturing employment by year. The coverage rates of the sample used for the estimations are about 2% of total employment and around 14% of manufacturing employment. The data on total and manufacturing employment in Indonesia are taken from the website of ILO: laborsta.ilo.org.

3.B The Probabilities of Adjustments under Bivariate Normality

In this appendix, I show the probabilities of the adjustment decisions, which are suppressed in the text, under the assumption of a bivariate normality. The specific forms of the probabilities are necessary to implement the first-step ML estimation.

For example, a plant contracts both factors at time t when $\varepsilon_{it}^L < A_{Lit}^-$ and $\varepsilon_{it}^K < A_{Kit}^-$. With the assumption of a bivariate normal distribution of ε_{it}^L and ε_{it}^K, the probability of this state is;

$$
\begin{aligned}
Pr(\Omega_{it}^{-,-}) &= Pr(\tilde{\varepsilon}_{it}^L < A_{Lit}^-, \tilde{\varepsilon}_{it}^K < A_{Kit}^-) \\
&= \Phi_2\left(\widetilde{A}_{Lit}^-, \widetilde{A}_{Kit}^-, \rho\right),
\end{aligned}
$$

where $\Phi_2(\cdot)$ is a cdf of bivariate normal distribution. Similarly, the probabilities of other eight decision states can be expressed as follows:

$$
\begin{aligned}
Pr(\Omega_{it}^{-,0}) &= \Phi_2\left(\widetilde{A}_{Lit}^-, \widetilde{A}_{Kit}^+, \rho\right) - \Phi_2\left(\widetilde{A}_{Lit}^-, \widetilde{A}_{Kit}^-, \rho\right), \\
Pr(\Omega_{it}^{-,+}) &= \Phi_2\left(\widetilde{A}_{Lit}^-, -\widetilde{A}_{Kit}^+, -\rho\right), \\
Pr(\Omega_{it}^{0,-}) &= \Phi_2\left(\widetilde{A}_{Lit}^+, \widetilde{A}_{Kit}^-, \rho\right) - \Phi_2\left(\widetilde{A}_{Lit}^-, \widetilde{A}_{Kit}^-, \rho\right), \\
Pr(\Omega_{it}^{0,0}) &= \Phi_2\left(\widetilde{A}_{Lit}^+, \widetilde{A}_{Kit}^+, \rho\right) - \Phi_2\left(\widetilde{A}_{Lit}^+, \widetilde{A}_{Kit}^-, \rho\right) - \Phi_2\left(\widetilde{A}_{Lit}^-, \widetilde{A}_{Kit}^+, \rho\right) \\
&\quad + \Phi_2\left(\widetilde{A}_{Lit}^-, \widetilde{A}_{Kit}^-, \rho\right), \\
Pr(\Omega_{it}^{0,+}) &= \Phi_2\left(\widetilde{A}_{Lit}^+, -\widetilde{A}_{Kit}^+, -\rho\right) - \Phi_2\left(\widetilde{A}_{Lit}^-, -\widetilde{A}_{Kit}^+, -\rho\right), \\
Pr(\Omega_{it}^{+,-}) &= \Phi_2\left(-\widetilde{A}_{Lit}^+, \widetilde{A}_{Kit}^-, -\rho\right), \\
Pr(\Omega_{it}^{+,0}) &= \Phi_2\left(-\widetilde{A}_{Lit}^+, \widetilde{A}_{Kit}^+, -\rho\right) - \Phi_2\left(-\widetilde{A}_{Lit}^+, \widetilde{A}_{Kit}^-, -\rho\right), \\
Pr(\Omega_{it}^{+,+}) &= \Phi_2\left(-\widetilde{A}_{Lit}^+, -\widetilde{A}_{Kit}^+, \rho\right),
\end{aligned}
$$

Specifying the probabilities as above, I estimate the model by ML estimation. To implement the ML estimation, I utilize Stata's ml command with Newton-Raphson maximization algorithm . I provide analytical first derivatives of the log likelihood function, but the second order derivatives are computed numerically.

3.C Estimation with the Copula Method

This section describes the first-step ML estimation of the joint adjustment decisions under distribution assumptions other than bivariate normality. To estimate the joint adjustment decisions, it is necessary to specify the joint probabilities in the log likelihood function (3.5). In order to specify the joint probabilities, we use the copula method. The copula method is introducted in Chapter 1. Then, this section discusses specifically to the model in this chapter.

In the model, there are two continuous random variables ε_{it}^L and ε_{it}^K. Then, for example, the probability of the joint decision of contractions of both factors by a plant i at time t is $Pr(\Omega^{-,-}) = Pr(\varepsilon_{it}^L < A_{Lit}^-, \varepsilon_{it}^K < A_{Kit}^-)$. If a bivariate normal distribution of ε_{it}^L and ε_{it}^K is assumed, this probability and the probabilities of other joint decisions can be expressed as in Appendix 3.B.

At this time, we alternatively assume that each of the random variables ε_{it}^L and ε_{it}^K marginally follows a t distribution. Therefore, the marginal probability of employment contraction is $Pr(\tilde{\varepsilon}_{it}^L < \widetilde{A}_{Lit}^-) = F_{\nu_L}(\widetilde{A}_{Lit}^-)$, where $F_{\nu_L}(\cdot)$ is the cdf of a (standard) t distribution with degrees of freedom ν_L. Notice that the random variable ε_{it}^L is now standardized by dividing it by the scaler parameter σ_L. Similarly, the marginal probability of disinvestment is $Pr(\tilde{\varepsilon}_{it}^K < \widetilde{A}_{Kit}^-) = F_{\nu_K}(\widetilde{A}_{Kit}^-)$. Given these marginal distributions, a copula function generate the joint probabilities of the state where a plant i contracts both factors at time t:

$$
\begin{aligned}
Pr(\Omega_{it}^{-,-}) &= Pr(\tilde{\varepsilon}_{it}^L < \widetilde{A}_{Lit}^-, \tilde{\varepsilon}_{it}^K < \widetilde{A}_{Kit}^-) \\
&= C(F_{\nu_L}(\widetilde{A}_{Lit}^-), F_{\nu_K}(\widetilde{A}_{Kit}^-); \theta),
\end{aligned}
$$

where the specific form of $C(\cdot)$ depends on the choice of a copula function.

Likewise, the probabilities of other eight adjustment decision states are expressed:

$$
\begin{aligned}
Pr(\Omega_{it}^{-,0}) &= C(F_{\nu_L}(\widetilde{A}_{Lit}^{-}), F_{\nu_K}(\widetilde{A}_{Kit}^{+});\theta) - C(F_{\nu_L}(\widetilde{A}_{Lit}^{-}), F_{\nu_K}(\widetilde{A}_{Kit}^{-});\theta), \\
Pr(\Omega_{it}^{-,+}) &= F_{\nu_L}(\widetilde{A}_{Lit}^{-}) - C(F_{\nu_L}(\widetilde{A}_{Lit}^{-}), F_{\nu_K}(\widetilde{A}_{Kit}^{+});\theta), \\
Pr(\Omega_{it}^{0,-}) &= C(F_{\nu_L}(\widetilde{A}_{Lit}^{+}), F_{\nu_K}(\widetilde{A}_{Kit}^{-});\theta) - C(F_{\nu_L}(\widetilde{A}_{Lit}^{-}), F_{\nu_K}(\widetilde{A}_{Kit}^{-});\theta), \\
Pr(\Omega_{it}^{0,0}) &= C(F_{\nu_L}(\widetilde{A}_{Lit}^{+}), F_{\nu_K}(\widetilde{A}_{Kit}^{+});\theta) - C(F_{\nu_L}(\widetilde{A}_{Lit}^{-}), F_{\nu_K}(\widetilde{A}_{Kit}^{+});\theta) \\
&\quad - C(F_{\nu_L}(\widetilde{A}_{Lit}^{+}), F_{\nu_K}(\widetilde{A}_{Kit}^{-})) + C(F_{\nu_L}(\widetilde{A}_{Lit}^{-}), F_{\nu_K}(\widetilde{A}_{Kit}^{-});\theta), \\
Pr(\Omega_{it}^{0,+}) &= F_{\nu_L}(\widetilde{A}_{Lit}^{+}) - F_{\nu_L}(\widetilde{A}_{Lit}^{-}) - C(F_{\nu_L}(\widetilde{A}_{Lit}^{+}), F_{\nu_K}(\widetilde{A}_{Kit}^{+});\theta) \\
&\qquad + C(F_{\nu_L}(\widetilde{A}_{Lit}^{-}), F_{\nu_K}(\widetilde{A}_{Kit}^{+});\theta), \\
Pr(\Omega_{it}^{+,-}) &= F_{\nu_K}(\widetilde{A}_{Kit}^{-}) - C(F_{\nu_L}(\widetilde{A}_{Lit}^{+}), F_{\nu_K}(\widetilde{A}_{Kit}^{-});\theta), \\
Pr(\Omega_{it}^{+,0}) &= F_{\nu_K}(\widetilde{A}_{Kit}^{+}) - F_{\nu_K}(\widetilde{A}_{Kit}^{-}) - C(F_{\nu_L}(\widetilde{A}_{Lit}^{+}), F_{\nu_K}(\widetilde{A}_{Kit}^{+});\theta) \\
&\qquad + C(F_{\nu_L}(\widetilde{A}_{Lit}^{+}), F_{\nu_K}(\widetilde{A}_{Kit}^{-});\theta), \\
Pr(\Omega_{it}^{+,+}) &= 1 - F_{\nu_L}(\widetilde{A}_{Lit}^{+}) - F_{\nu_K}(\widetilde{A}_{Kit}^{+}) + C(F_{\nu_L}(\widetilde{A}_{Lit}^{+}), F_{\nu_K}(\widetilde{A}_{Kit}^{+});\theta).
\end{aligned}
$$

Each copula function considered in this study has a specific functional form $C(\cdot)$. Given this selected function, the log likelihood function (3.5) is specified. It is now feasible to implement the first-step ML estimation. As in the case of a bivariate normal distribution, each log likelihood function is maximized with Stata's ml module. Only first derivatives are provided analytically.

3.D The Predictions from the Model

This appendix describes how the predicted means of H_{it}/L_{it-1} unconditional and conditional on the adjustment decisions, which are reported in Table 3.8. The predicted values for I_{it}/K_{it-1} are obtained in the analogous way. The means of H_{it}/L_{it-1} conditional on expansion $H_{it} > 0$ and contraction $H_{it} < 0$ are obtained from the second-step OLS estimations (3.6). For each observation regardless of its outcome, we calculate the fitted value of $\left(\widehat{\widetilde{\gamma}_0^L - \widetilde{b}_L^+} + \widehat{\widetilde{\gamma}}_L{}'Z_{it} + \widehat{\lambda}_{Lit}^+\right)/\widehat{\widetilde{c}}_L^+$, and then, we take the unweighted average over the samples. This is the predicted mean of H_{it}/L_{it-1} conditional on $H_{it} > 0$. Likewise, the predicted mean of H_{it}/L_{it-1} conditional on $H_{it} < 0$ is the average of $\left(\widehat{\widetilde{\gamma}_0^L + \widetilde{b}_L^-} + \widehat{\widetilde{\gamma}}_L{}'Z_{it} + \widehat{\lambda}_{Lit}^-\right)/\widehat{\widetilde{c}}_L^-$.

Analytically, the unconditional mean of H_{it}/L_{it-1} can be expressed as

$$E[H_{it}/L_{it-1}] = Pr(H_{it} > 0)E[H_{it}/L_{it-1}|H_{it} > 0]+Pr(H_{it-1} < 0)E[H_{it}/L_{it-1}|H_{it} < 0].$$

For each observation, we calculate the fitted value and take the average. The conditional means are calculated as described above. The specific forms of $Pr(H_{it} > 0)$ and $Pr(H_{it} < 0)$ depend on the assumption of the marginal distribution of ε_{it}^L. Under the assumption of normality, $Pr(H_{it} > 0) = \Phi(-\widetilde{A}_{Lit}^+)$ and $Pr(H_{it} < 0) = \Phi(\widetilde{A}_{Lit}^-)$. Under the assumption that $\widetilde{\varepsilon}_{it}^L$ has a t distribution, $Pr(H_{it} > 0) = 1 - F_{\nu_L}(\widetilde{A}_{Lit}^+)$ and $Pr(H_{it} < 0) = F_{\nu_L}(\widetilde{A}_{Lit}^-)$, where $F_{\nu_L}(\cdot)$ is the cdf of a (standard) t distribution with degrees of freedom ν_L.

The predicted means conditional on the decision of adjusting both factors can be calculated in a similar way to calculating the prediction of the conditional mean on the decision of adjusting a single factor. The difference is the conditional (truncated) mean of the random variable ε_{it}^L. Previously, truncation is only with respect to the threshold of labor adjustment. At this time, however, truncation is with respect to both its own threshold and the threshold for capital adjustment. Define $\mu_{Lit}^{+,+} = E[\widetilde{\varepsilon}_{it}^L|\widetilde{\varepsilon}_{it}^L > \widetilde{A}_{Lit}^+, \widetilde{\varepsilon}_{it}^K > \widetilde{A}_{Kit}^+]$, $\mu_{Lit}^{+,0} = E[\widetilde{\varepsilon}_{it}^L|\widetilde{\varepsilon}_{it}^L > \widetilde{A}_{Lit}^+, \widetilde{A}_{Kit}^- < \widetilde{\varepsilon}_{it}^K < \widetilde{A}_{Kit}^+]$ and $\mu_{Lit}^{+,-} = E[\widetilde{\varepsilon}_{it}^L|\widetilde{\varepsilon}_{it}^L > \widetilde{A}_{Lit}^+, \widetilde{\varepsilon}_{it}^K < \widetilde{A}_{Kit}^-]$. To obtain the predicted means of H_{it}/L_{it-1} conditional on $H_{it} > 0$ and $I_{it} > 0$, on $H_{it} > 0$, and on $I_{it} = 0$, and $H_{it} > 0$ and $I_{it} < 0$, replace $\widehat{\lambda}_{Lit}^+$ in calculating the predicted mean conditional on $H_{it} > 0$ with $\widehat{\mu}_{Lit}^{+,+}$, $\widehat{\mu}_{Lit}^{+,0}$, and $\widehat{\mu}_{Lit}^{+,-}$, respectively. The predicted means of H_{it}/L_{it-1} conditional on employment contraction and capital adjustment decisions can be calculated in similar ways.

Under the assumption of the bivariate normal distribution between ε_{it}^{L} and ε_{it}^{K}, the analytical expressions for $\mu_{Lit}^{+,+}$, $\mu_{Lit}^{+,0}$ and $\mu_{Lit}^{-,-}$ exist. In particular, following the derivation by Rosenbaum (1961), the doubly truncated mean $\mu_{Lit}^{+,+}$ is expressed as:

$$\mu_{Lit}^{+,+} = \frac{\phi(\widetilde{A}_{Lit}^{+})\Phi\left(\frac{-\widetilde{A}_{Kit}^{+}+\rho\widetilde{A}_{Lit}^{+}}{\sqrt{1-\rho^2}}\right) + \rho\phi\left(\widetilde{A}_{Kit}^{+}\right)\Phi\left(\frac{-\widetilde{A}_{Lit}^{+}+\rho\widetilde{A}_{Kit}^{+}}{\sqrt{1-\rho^2}}\right)}{\Phi_2\left(-\widetilde{A}_{Lit}^{+}, -\widetilde{A}_{Kit}^{+}, \rho\right)}.$$

Notice that if there is no correlation between ε_{it}^{L} and ε_{it}^{K}, that is, $\rho = 0$, this expression can be reduced to the truncation mean only with respect to its own threshold value $\widetilde{A}_{Lit}^{+}$, which is exactly λ_{Lit}^{+}. Similarly,

$$\mu_{Lit}^{+,0} = \frac{\phi(\widetilde{A}_{Lit}^{+})\left[\Phi\left(\frac{\widetilde{A}_{Kit}^{+}-\rho\widetilde{A}_{Lit}^{+}}{\sqrt{1-\rho^2}}\right) - \Phi\left(\frac{\widetilde{A}_{Kit}^{-}-\rho\widetilde{A}_{Lit}^{+}}{\sqrt{1-\rho^2}}\right)\right]}{\Phi_2\left(-\widetilde{A}_{Lit}^{+}, \widetilde{A}_{Kit}^{+}, -\rho\right) - \Phi_2\left(-\widetilde{A}_{Lit}^{+}, \widetilde{A}_{Kit}^{-}, -\rho\right)} - \frac{\rho\left[\phi(\widetilde{A}_{Kit}^{+})\Phi\left(\frac{-\widetilde{A}_{Lit}^{+}+\rho\widetilde{A}_{Kit}^{+}}{\sqrt{1-\rho^2}}\right) - \phi(\widetilde{A}_{Kit}^{-})\Phi\left(\frac{-\widetilde{A}_{Lit}^{+}+\rho\widetilde{A}_{Kit}^{-}}{\sqrt{1-\rho^2}}\right)\right]}{\Phi_2\left(-\widetilde{A}_{Lit}^{+}, \widetilde{A}_{Kit}^{+}, -\rho\right) - \Phi_2\left(-\widetilde{A}_{Lit}^{+}, \widetilde{A}_{Kit}^{-}, -\rho\right)},$$

and

$$\mu_{Lit}^{+,-} = \frac{\phi(\widetilde{A}_{Lit}^{+})\Phi\left(\frac{\widetilde{A}_{Kit}^{-}-\rho\widetilde{A}_{Lit}^{+}}{\sqrt{1-\rho^2}}\right) - \rho\phi\left(\widetilde{A}_{Kit}^{-}\right)\Phi\left(\frac{-\widetilde{A}_{Lit}^{+}+\rho\widetilde{A}_{Kit}^{-}}{\sqrt{1-\rho^2}}\right)}{\Phi_2\left(-\widetilde{A}_{Lit}^{+}, \widetilde{A}_{Kit}^{-}, -\rho\right)}.$$

The truncated means of $\widetilde{\varepsilon}_{it}^{L}$ conditional on employment contraction and three capital adjustment decisions can be derived with appropriate changes of signs from the above expressions.

When we assume that ε_{it}^{L} and ε_{it}^{K} are marginally t-distributed and the joint distribution is created by the copula method, there is no analytical expression for the μ's. Therefore, we calculate the μ's numerically. To do so, first we randomly generate $\widetilde{\varepsilon}^{L}$ and $\widetilde{\varepsilon}^{K}$ with the appropriate margins and copula, and then, calculate the means conditional on the appropriate thresholds. Note that the threshold values differ across observations. For drawing pseudo-random variables from copulas, see the appendix in Trivedi and Zimmer (2007b). It is fair to mention that the predictions by this numerical method take different values when we change the seed of random-number generation.

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

with Wim Vijverberg

4.1 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.

In particular, we model the marginal distributions with the Generalized Tukey lambda (GTL) distribution. First proposed by Pregibon (1980) and Freimer et al.

[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.

(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 six applications on a wide variety of topics: wages of married women in two different countries, wages of children in a lower-income country, health expenditures, speeding tickets, and international disputes. In all six, the joint normality assumption is rejected, and eleven of the twelve 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 six 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 4.3 lays out the copula approach with special reference to the sample selection model. This is followed in Section 4.4 by an introduction of the GTL distribution and a description of its attractive properties. In Section 4.5, 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 4.6, we examine the relevance of the GTL-copula estimator in the six real-world applications that were mentioned above, in comparison with the familiar estimator of the joint normal sample selection model. Section 4.7 concludes the paper.

4.2 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

$$s_i = 1({z_i}'\gamma + \nu_i > 0) \tag{4.1}$$

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

$$y_i = {x_i}'\beta + \sigma\varepsilon_i, \tag{4.2}$$

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 (4.2) with the subsample for which $s_i = 1$ yields inconsistent estimates of β. This is the well-known selectivity bias problem (Heckman, 1979).

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

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 (4.1)-(4.2) 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

$$L = \prod_{i=1}^{N} \left[\int_{-\infty}^{-z_i{}'\gamma} f_\nu(\nu) d\nu \right]^{s_i=0} \left[\int_{-z_i{}'\gamma}^{\infty} f_{\nu\varepsilon}(\nu, \varepsilon_i) d\nu \right]^{s_i=1}, \tag{4.3}$$

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. 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).

4.3 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 (2007a). 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 $C(F_1, F_2; \theta) = C(F_2, F_1; \theta)$.

Using a copula, the likelihood function of the sample selection model is

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

where $F_\nu(\cdot)$ and $F_\varepsilon(\cdot)$ are the cdf of ν and ε, respectively, and $f_\varepsilon(\cdot)$ is the pdf of ε.

Many different copula functions are available, each with its own characteristics.

[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 $C(F_{1b}, F_{2b}) - C(F_{1b}, F_{2a}) - C(F_{1a}, F_{2b}) + C(F_{1a}, F_{2a}) \geq 0$.

Table 4.1: Dependence parameter θ and Kendall's τ

Copula	$C(F_1, F_2)$	Range of θ	θ_{ind}	Kendall's $\tau(\theta)$	Range of τ
Product	$F_1 F_2$	—	—	0	0
Gaussian	$\Phi_2(\Phi^{-1}(F_1), \Phi^{-1}(F_2); \theta)$	$[-1, 1]$	0	$2\sin^{-1}(\theta)/\pi$	$[-1, 1]$
FGM	$F_1 F_2(1 + \theta(1 - F_1)(1 - F_2))$	$[-1, 1]$	0	$2\theta/9$	$[-2/9, 2/9]$
Archimedean Family					
Frank	$-\theta^{-1} \ln \left(1 + \frac{(e^{-\theta F_1} - 1)(e^{-\theta F_2} - 1)}{(e^{-\theta} - 1)}\right)$	$(-\infty, \infty)$	0	$1 - \frac{4[1 - D_1(\theta)]}{\theta}$	$[-1, 1]$
Clayton	$\left(F_1^{-\theta} + F_2^{-\theta} - 1\right)^{-1/\theta}$	$[0, \infty)$	0	$\theta/(\theta + 2)$	$[0, 1]$
Gumbel	$\exp\left(-\left((-\ln F_1)^\theta + (-\ln F_2)^\theta\right)^{1/\theta}\right)$	$[1, \infty)$	1	$(\theta - 1)/\theta$	$[0, 1]$
Joe	$1 - \left[(\tilde{F}_1)^\theta + (\tilde{F}_2)^\theta - (\tilde{F}_1 \tilde{F}_2)^\theta\right]^{1/\theta}$	$[1, \infty)$	1	$\cdot$	$[0, 1]$

θ_{ind} is the value of θ if independent. For Gaussian, $\Phi_2(\cdot)$ is a cdf of a bivariate normal distribution with a coefficient of correlation θ, and $\Phi(\cdot)$ is a cdf of a standard normal distribution. 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. It is evaluated numerically.

Table 4.1 lists six of them that we will use later in this paper. In the context of sample selection models, the Gaussian copula appears as part of a two-step estimator in Lee (1982) and in the context of a FIML estimator in Lee (1983). Prieger (2002) inserts the FGM copula into a selection model of hospitalization duration. Smith (2003) illustrates sample selection models using Archimedean copulas. 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.

One of attractive features of copulas is their various dependence structures. 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.[4] In this aspect, the Frank copula is similar to the Gaussian and FGM copulas, but its dependence is weaker in

[4]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.

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.[5] Alternatively, Vuong's (1989) test can be used to weigh one copula against another.[6] To allow for potential copula misspecification, Trivedi and Zimmer (2007a) recommend that the standard errors be estimated in robust sandwich form under the theory of quasi-maximum likelihood (White, 1982).

Even though the dependence parameter θ governs the degree of dependence, it is not comparable across different copulas. A common measure of dependence is Kendall's τ,[7] In principle, τ ranges from -1 to 1. The lower and upper bounds correspond to perfect negative and positive dependence, respectively. A copula for

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

[6]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.

[7]Kendall's τ may be computed for general copula functions as

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

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

which τ attains both bounds is called comprehensive. When $\tau = 0$, the two random variables are independent.[8] The copula corresponding to independence is the Product copula (also referred to as the Independence copula). 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 4.1.

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 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).[9]

Not all copulas are comprehensive as Table 4.1 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 (4.2) 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: $-1 \leq \tau \leq 0$.[10] We will refer to these sign-switched copula

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

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

[10] Equivalently, we can specify the selection equation (4.1) 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

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 rather than a priori given.

4.4 Specifying the Marginal Distributions: The Case for GTL

4.4.1 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

between the selection equation and the other outcome equation intact.

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.[11] 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 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.[12]

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

[11] 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.

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

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.

4.4.2 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).[13] The random variable ϵ from the GTL distribution is given by a quantile function $Q(u)$,

$$\epsilon = Q(u) = \mu + \sigma \left(\frac{u^{\alpha-\delta} - 1}{\alpha - \delta} - \frac{(1-u)^{\alpha+\delta} - 1}{\alpha + \delta} \right), \tag{4.4}$$

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

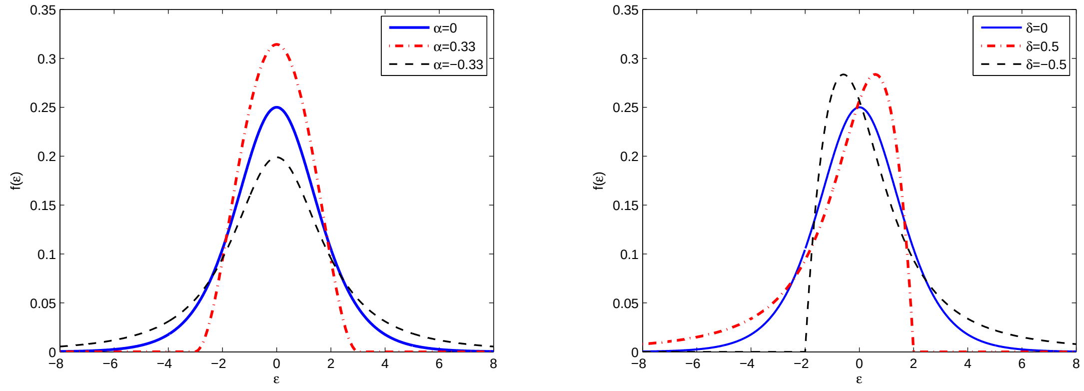
$$F(\epsilon) = u = Q^{-1}(\epsilon).$$

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.[14] The pdf $f(\epsilon)$ is given

[13] 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.

[14] 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

Figure 4.1: GTL distributions with different α and δ



(a) δ is fixed at 0

(b) α is fixed at 0

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 4.1a). 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 4.1b).[15] For $\delta = 0$. the distribution is symmetric. The shape of the density does not even have to be bell-shaped as Figure 4.1 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.

the exponential distribution.

[15]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).

Although no analytical expression of the pdf exists, the moments are analytically calculated. For convenience in notation, define $\lambda_1 = \alpha - \delta$ and $\lambda_2 = \alpha + \delta$. The first and second moments are given by

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

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

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.[16] 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 14), 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[17] with $(\alpha, \delta) = (0.1422, -0.2290)$

[16]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).

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

(Freimer et al., 1988, Vijverberg and Vijverberg, 2012).[18]

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 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

[18] 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).

$\max(\lambda_1, \lambda_2) < 1/2$. This property extends to the data-fitting estimator of Su (2007).

4.4.3 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.[19] 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 (4.4) 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

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

discrete choice models. But even during estimation, standardization is beneficial: it tends to speed up the optimization of the 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 4.6, we apply the copula-GTL sample selection model to actual data. All the estimations in the following sections are implemented in STATA.[20]

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

4.5 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.

4.5.1 Varying the Dependence Structure

In our first research design, we draw ν_i and ε_i from each of the six copulas[21] that were described in Section 4.3 with standard normal marginals,[22] and we examine the consequences of estimating (i) a common normal-Gaussian selection model, (ii)

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

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

a selection model with GTL marginals and the correct copula function, or (iii) an optimally selected general GTL-copula selection model.[23] 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 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 4.2 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.[24] 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 4.A.

[23] 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.

[24] 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 4.2 uses standard normal marginals, β_0^* in fact equals 1. For the estimator under the joint normality, $\hat{\beta}_0^* = \hat{\beta}_0$.

Table 4.2: 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.

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 4.2): assuming independence is more harmful than choosing a wrong dependence structure.

Second, as the third line of each panel of Table 4.2 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 4.3 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 4.A)—but other copulas are occasionally erroneously preferred.

Table 4.3: 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.

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 4.2, 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 4.3 also indicates that even when the true copula is not selected, a copula similar to the true one tends to be selected. For example, both Gumbel and Joe copulas exhibit strong upper tail but weaker dependence in a lower tail. 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 rarely selected. 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 resemblant 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.[25]

Sixth, the shape parameters of the GTL distribution for the outcome equation (α_ε and δ_ε) are estimated precisely with small biases.[26] 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.

[25]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.

[26]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.

4.5.2 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.

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 4.4 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

Table 4.4: 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)
DGP: $\alpha = 0.33$ 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)
DGP: $\alpha = 0.33$ 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)
DGP: $\alpha = -0.33$ 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)
DGP: $\alpha = -0.33$ 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)
DGP: $\alpha = -0.33$ 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 4.2. For each DGP, the copula is Gaussian with $\tau = 0.333$ and marginals that both are GTL with the specified shape parameters.

$\hat{\beta}_0^*$ 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.

4.5.3 A Bivariate χ^2 Distribution

We have so far investigated various dependence structure and various marginal distributions and have shown the flexibility of our proposed approach. Yet, the simulation designs are still limited in that the data are generated from the copulas we use in the estimation and from the GTL distribution. We strengthen our claim of the flexibility of our approach by considering the DGP that is not exactly nested in our estimation. For this purpose, we generate the error terms from the bivariate χ^2 distribution with 10 degrees of freedom and 0.25 of correlation.[27] As before, the disturbances are standardized such that their population mean is 0 and their population variance is 1.

Table 4.5 reports the simulation results with the bivariate χ^2 distributed errors. The normal-Gaussian estimator performs very poorly. Its bias and standard deviation are quite large. Even the GTL-Product estimator, which assumes independence, outperforms the normal-Gaussian estimator. Our proposed estimator still performs well. As the second panel of Table 4.5 shows, the Gaussian copula is selected most frequently, followed by the Joe and Gumbel copulas. In line with the previous findings, the "Pretest" estimator has a smaller bias but less precision than using the GTL-Gaussian estimator all the time. This simulation result adds evidence of the flexibility of our proposed estimator.

[27] We generate the bivariate χ^2 distributed error terms, following Zuehlke and Zeman (1991). Suppose $\epsilon_j = (\epsilon_{1j}, \epsilon_{2j})$ be an independent vector for $j = 1, \cdots, r$. ϵ_{1j} and ϵ_{2j} are drawn from a bivariate standard normal distribution with a coefficient of correlation ρ. Then, $X = \sum_{j=1}^{r} \epsilon_j$ is a draw from a bivariate χ^2 distribution with r degrees of freedom and correlation ρ^2.

Table 4.5: Simulation results when DGP is a bivariate χ^2 distribution

Biases and Standard deviations[a]	β_0^*	β_1	α_ε	δ_ε	τ	α_ν	δ_ν
Normal-Gaussian	-0.3227	0.1793			0.2260		
	(0.3889)	(0.2347)			(0.2597)		
GTL-Product	0.1940	-0.0932	0.0136	0.0041		0.0510	-0.0388
	(0.0392)	(0.0350)	(0.0315)	(0.0166)		(0.1606)	(0.0604)
GTL-Gaussian	-0.0337	0.0224	0.0343	-0.0099	0.0389	0.0574	-0.0442
	(0.1127)	(0.0679)	(0.0336)	(0.0181)	(0.0892)	(0.2425)	(0.1688)
Pretest	-0.0030	0.0138	0.0315	-0.0079	0.0176	0.0568	-0.0471
	(0.1477)	(0.0979)	(0.0389)	(0.0251)	(0.0944)	(0.2454)	(0.1694)

Frequencies of selecting copulas[b]

Gaussian	FGM	Frank	Clayton	Gumbel	Joe
0.362	0.064	0.086	0.080	0.170	0.244

[a] The biases are evaluated by assuming that the true parameters are $\alpha = 0.1614$ and $\delta = -0.1834$, representing the GTL approximation of the χ^2 distribution with 10 degrees of freedom, and $\tau = 0.1495$, which is computed numerically by generating 500,000 draws from the bivariate χ^2 distribution.

[b] Copula selection is based on the largest likelihood value.

4.5.4 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 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

Table 4.6: 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 $\hat{\beta}_1$.

about a half of observations have observable outcomes.

Tables 4.6 and 4.7 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 4.6 repeats the distributional assumptions of Table 4.2; thus, the column for Specification 1 repeats the results for $\hat{\beta}_1$ in first and third lines of each panel in Table 4.2. Table 4.7 is similarly linked with Table 4.4.

In the first panel of Table 4.6 where the DGP uses a joint normal distribution to generate disturbances, the normal-Gaussian estimator yields slightly smaller biases

Table 4.7: 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)
DGP: $\alpha = -0.33$ 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 $\hat{\beta}_1$.

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 4.6, 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 4.7 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 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 4.8: GTL-copula model selection: log-likelihood values and Vuong tests

Estimation method	Applications Wages of married women Portugal	Wages of married women USA	Wages of school-aged children	Health expenditures	Speeding tickets	International disputes
Normal-Gaussian [a]	-2487.94	-36938.67	-4927.80	-10170.11	-45173.91	-9209.59
Copula [b]						
Product	-2317.81	-36958.80	-4473.11	-10135.11	-10402.06	-9095.19
Gaussian	-2307.33	-36681.59	-4468.23	*-10132.26*	-10397.16	-9092.53
FGM	-2312.97	-36828.43	-4471.31	-10135.05	-10305.82	-9075.48
Frank	-2307.70	-36700.49	-4470.39	-10135.11	-10382.39	*-9071.47*
Clayton	-2311.63	*-36682.17*	-4473.11	**-10132.24**	-10399.98	**-9063.70**
Gumbel	*-2300.68*	**-36676.24**	-4473.11	-10134.55	-10367.10	-9095.19
Joe	**-2295.18**	-36721.86	-4473.11	-10134.71	-10360.02	-9095.19
nClayton	-2317.81	-36958.80	-4471.52	-10135.11	-5630.22	-9095.19
nGumbel	-2317.81	-36958.80	*-4461.85*	-10135.11	*-4320.31*	-9095.19
nJoe	-2317.81	-36958.80	**-4461.63**	-10136.83	**-4098.00**	-9095.19
Vuong tests [d]						
Second best	2.50	0.72	0.44	0.01	13.30	3.10
GTL-Gaussian	3.26	0.86	1.57	0.01	24.12	8.60
Normal-Gaussian	10.29	7.97	11.73	4.30	76.04	4.08
Number of observations						
Selection	2,339	36,803	15,526	5,574	68,357	149,004
Outcome	1,400	23,496	1,657	4,281	31.674	972
Share (%)	59.9	63.8	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).

4.6 Applications

Greater flexibility is desirable only if it has practical relevance. Thus, we turn to six 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), those wages in the US, 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 4.8 gives an overview of these six 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.

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, 5 and 6). Other times, the Gaussian copula is adequate but the marginals are nonnormal (columns 2, 3 and 4). Note also that a higher log-likelihood value does not always translate into a larger Vuong test value: in column 6, 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 4.B. For better readability, we have adjusted the acronyms of the variables from what the authors used in their papers.

4.6.1 Wages of Married Women, Portugal

The first application is a study of wages among married women in Portugal (Martins, 2001).[28] 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.

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

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 4.8). 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 4.9 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.[29] The estimated GTL distribution of ν has thicker tails than the standard normal, as is shown by the dashed curves

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

Table 4.9: Estimation results: Log-wages of married women

Variables	Normal-Gaussian Coeff.	(S.E.) [a]	GTL-Joe 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 $\times$ CHILDY18	0.035	(0.019)	0.008	(0.011)
PEXP2 $\times$ 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.

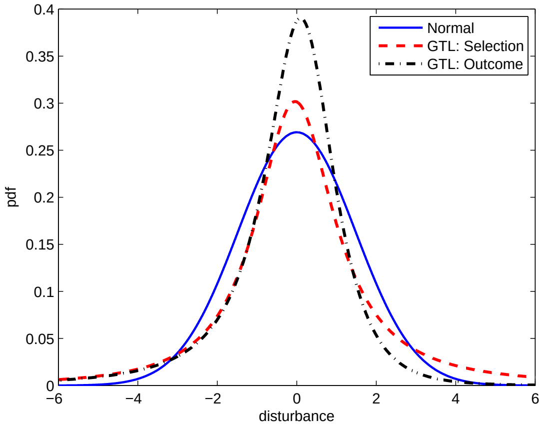
[b] The selection equation is standardized by the median and interquartile range.

in Figure 4.2a and 4.2b. δ_ν is close to 0 and statistically not 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:[30] our GTL estimates are not inconsistent with ν being distributed logistically (Genius and Strazzera, 2008) or even normally (Martins, 2001).

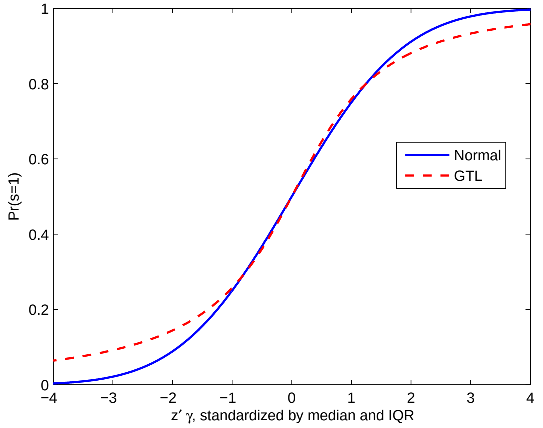
The estimated coefficients of the selection equation are mostly similar between the two estimators except the husband's wage: the GTL-Joe estimate is statistically

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

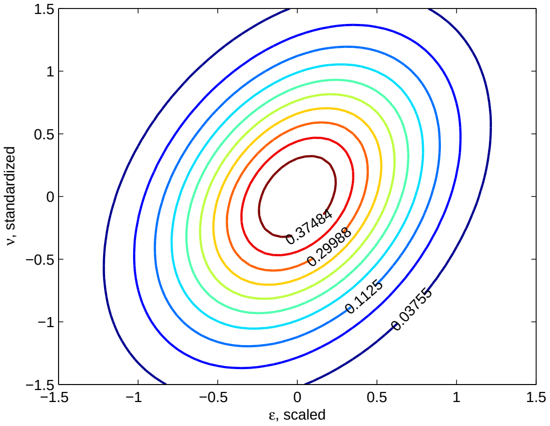
Figure 4.2: Differences between normal-Gaussian and GTL-Joe: Wages of married women



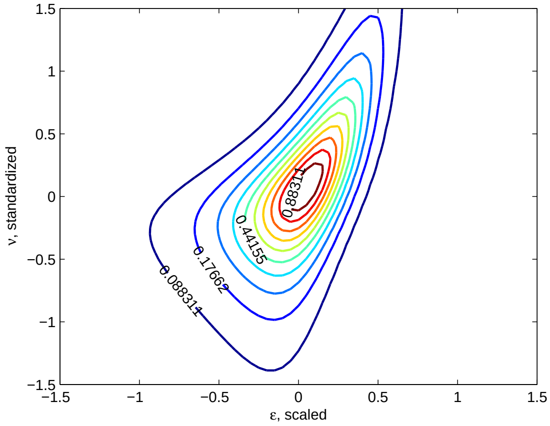
(a) PDF comparison



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



(c) Contour plot: normal-Gaussian



(d) Contour plot: GTL-Joe

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 freedom parameter of their assumed t distribution to be 2.95 and thus find thick tails as well.[31] 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.[32] In principle, this skewness could matter for the estimated wage equation (e.g., see footnote 34 below). To illustrate, Figure 4.3 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 4.4 plots expected

[31]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.

[32]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.

Figure 4.3: Density of scaled disturbances of the wage equation

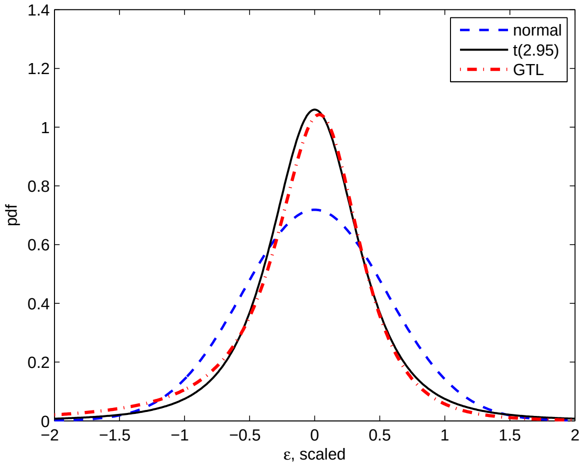
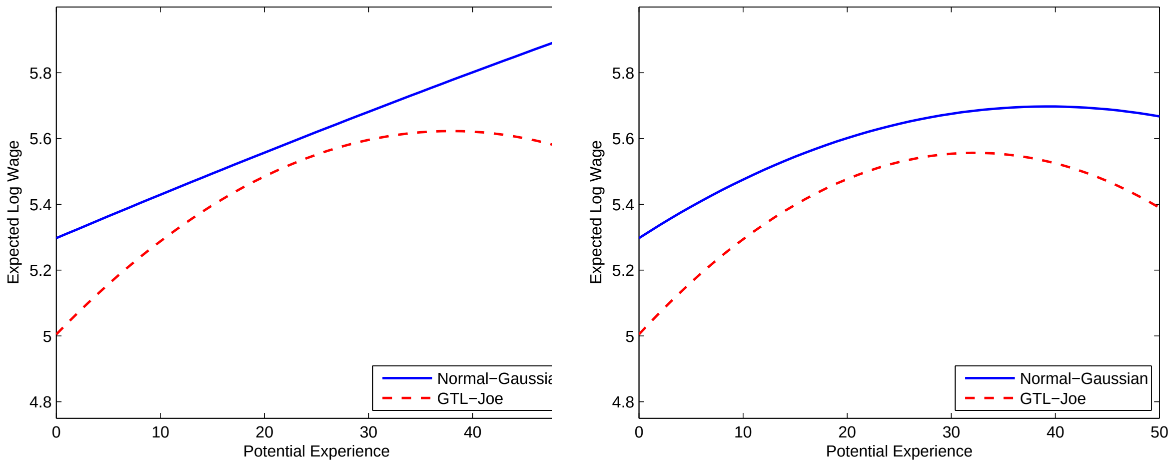


Figure 4.4: Log-wage profiles, married women in Portugal



(a) Log-wage profile for CHILD18 = 0 (b) Log-wage profile for CHILD18 = 2 log-wages as a function of potential experience, evaluated at the average years of schooling, 7.24.[33] Figure 4.4a shows that for a woman without a child 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 4.4b), but only slightly so for the GTL-Joe estimates.[34]

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 4.2c and 4.2d 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 4.4, the expected (unconditional) GTL-Joe wage profile lies substantially below the normal-Gaussian profile.

4.6.2 Wages of Married Women, USA

The second application is also a study of wages among married women as the first application. For this time, we derive the data from the US: Merged Outgoing Rotation Groups (MORG) of the Current Population Survey (CPS) in the year of 2000. After cleaning the data, we have 36,803 usable observations, out of which 23,496 married

[33] 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 (4.2).

[34] 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.

women reports wages. We use a set of covariates slightly different from the previous application. We exclude the variables regarding children in order to avoid a potential endogeneity problem, and we add more demographic and geographic variables since the United States is demographically and geographically more diversified than Portugal.

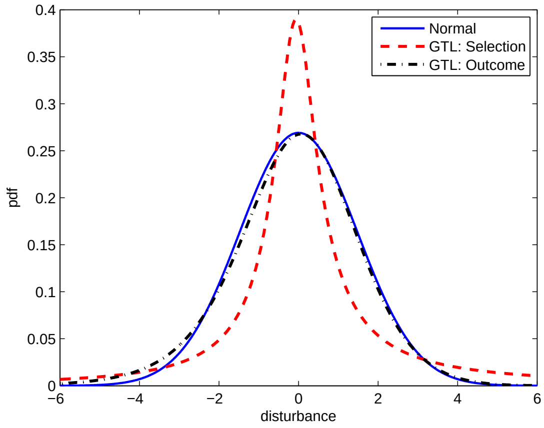
Table 4.10 summarizes the results of the estimation. Although the Gumbel copula attains the largest log likelihood value, Vuong tests cannot statistically discriminate it against the Clayton copula, which attains the second largest log likelihood value, and even the Gaussian copula. Yet, the Vuong test against normal-Gaussian is statistically in favor of the GTL-Gumbel estimator (Table 4.8). For comparison, we report the results of the Gaussian copula as well.

The disturbance of the selection equation has very thick tails. The estimated α_ν implied that the tails are too thick to define even the first moment. The negative $\widehat{\delta}_\nu$ implies that the distribution is right-skewed though it is not statistically significant from zero. These estimates are jointly statistically different from the parameters approximating the standard normal, $(\alpha, \delta) = (0.1436, 0)$, based on the Wald test.[35] The difference in the underlying distributions is reflected in the predicted probabilities of working. Figure 4.6 plots the predictions from the GTL-Gumbel and the normal-Gaussian model for each observation. The considerable differences appear in the tails.

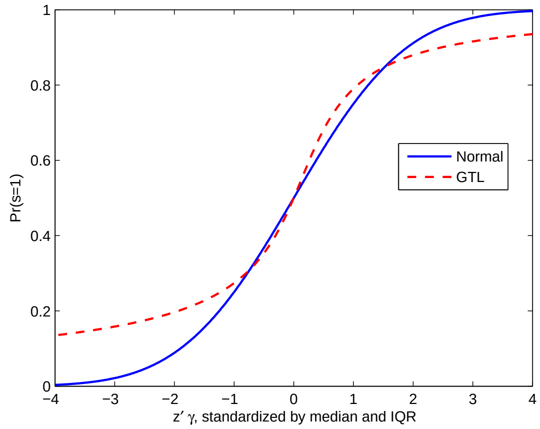
The disturbance of the wage equation is also nonnormal. The estimated shape parameters are statistically different from $(\alpha, \delta) = (0.1436, 0)$. The implied distribution is thicker than the normal distribution and left-skewed. Even though the estimated coefficients of the GTL-Gumbel and the GTL-Gaussian models are virtually the same and the contour plots of the joint pdf look similar (Figure 4.5d and 4.5e), the estimated shape parameters are comparatively different. While the implied skewness and

[35] The test statistics are 27.56 and 31.98, under the GTL-Gumbel and the GTL-Gaussian, respectively.

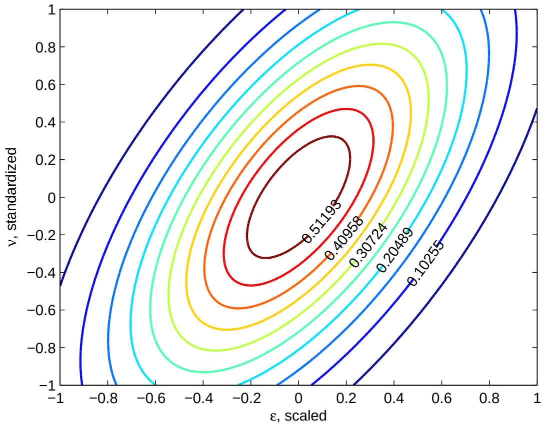
Figure 4.5: Differences between normal-Gaussian and GTL-copula: Wages of Married Women-CPS



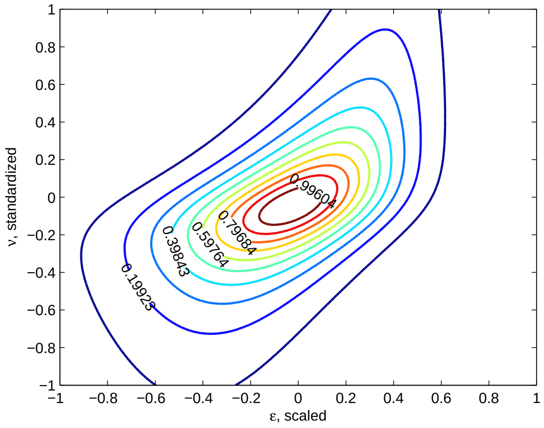
(a) PDF comparison



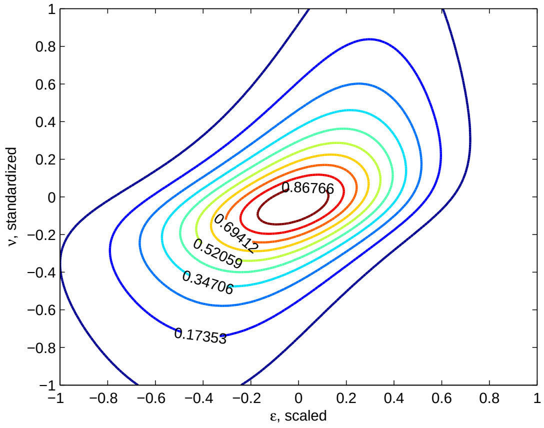
(b) $Pr(s_i = 1)$: GTL versus probit



(c) Contour plot: Normal-Gaussian



(d) Contour plot: GTL-Gubmel



(e) Contour plot: GTL-Gaussian

Table 4.10: Estimation results: Log Wage of Married Women using CPS-MORG data

	Normal-Gaussian		GTL-Gumbel		GTL-Gaussian	
Variables	Coeff.	(S.E.) [a]	Coeff.	(S.E.) [a]	Coeff.	(S.E.) [a]
The Selection Equation [b]						
YRSCH	0.076	(0.003)	0.064	(0.018)	0.068	(0.018)
AGE	0.065	(0.083)	0.085	(0.106)	0.088	(0.107)
$AGE^2 \times 10^{-2}$	-0.366	(0.318)	-0.414	(0.453)	-0.431	(0.456)
$AGE^3 \times 10^{-4}$	0.944	(0.527)	0.960	(0.861)	1.003	(0.861)
$AGE^4 \times 10^{-6}$	-0.849	(0.317)	-0.810	(0.604)	-0.845	(0.601)
lnHUSBW	-0.218	(0.011)	-0.198	(0.064)	-0.202	(0.062)
AGEdif	-0.020	(0.016)	-0.003	(0.012)	-0.003	(0.013)
BLACK	0.071	(0.023)	0.037	(0.021)	0.037	(0.021)
HISPANIC	-0.022	(0.022)	-0.026	(0.018)	-0.025	(0.018)
ASIAN	0.041	(0.036)	0.035	(0.032)	0.035	(0.033)
INDIAN	-0.152	(0.057)	-0.126	(0.053)	-0.129	(0.053)
constant	0.195	(0.773)	-0.135	(0.783)	-0.161	(0.807)
α_ν			-1.340	(0.431)	-1.328	(0.407)
δ_ν			-0.311	(0.214)	-0.262	(0.209)
The Wage Equation						
YRSCH	0.129	(0.002)	0.137	(0.002)	0.137	(0.002)
EXP	0.027	(0.001)	0.031	(0.001)	0.030	(0.001)
$EXP^2 \times 10^{-2}$	-0.057	(0.003)	-0.065	(0.003)	-0.063	(0.003)
BLACK	0.000	(0.013)	0.004	(0.014)	0.002	(0.014)
HISPANIC	-0.044	(0.014)	-0.045	(0.014)	-0.043	(0.014)
ASIAN	0.050	(0.023)	0.055	(0.024)	0.054	(0.024)
INDIAN	-0.102	(0.038)	-0.116	(0.037)	-0.114	(0.036)
constant	0.392	(0.033)	0.235	(0.038)	0.250	(0.037)
α_ε			0.043	(0.010)	-0.039	(0.015)
δ_ε			0.053	(0.010)	0.130	(0.016)
σ	0.495	(0.005)	0.322	(0.007)	0.315	(0.005)
ρ or θ	0.638	(0.019)	2.428	(0.120)	0.793	(0.021)
τ	0.441		0.588		0.583	
$\ln L$	-36938.67		-36676.24		-36681.59	
AIC	73953.34		73436.48		73447.18	

Both selection and wage equations also contains dummy variables for geographic divisions. In the estimation, the data are weighted by the inverse of sampling probability.

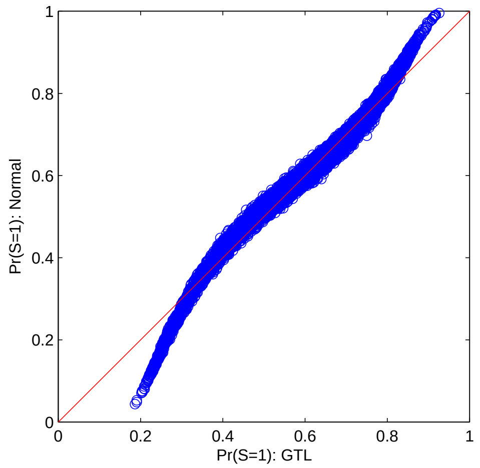
[a] Robust Standard Error.

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

the kurtosis are -0.3768 and 4.0181, respectively, in the GTL-Gumbel model, these are -1.5760 and 13.7385, respectively, in the GTL-Gaussian model.

The estimated coefficients of the normal-Gaussian model and the GTL-copula models are slightly different. For example, the normal-Gaussian model estimates the

Figure 4.6: Predicted probability $Pr(s_i = 1)$: GTL-Gumbel vs normal-Gaussian



coefficient on YRSCH to be 0.129. Whereas, the GTL-Gumbel model estimates it to be 0.137. Although this difference of 0.008 may seem small, if the difference is multiplied by the sample mean of YRSCH, which is 13.86, the sample average of the predicted wage offers will differ by approximately 0.11 log points, which is not trivially small.

Figure 4.7 shows the predicted wage distribution both unconditional and conditional on working.[36] The predicted distribution of wage offer from the GTL-Gumbel model is slightly left-skewed, compared that from the normal-Gaussian model. Figure 4.7b shows that both models predict the data well.

4.6.3 Wages of School-Aged Children, Mexico

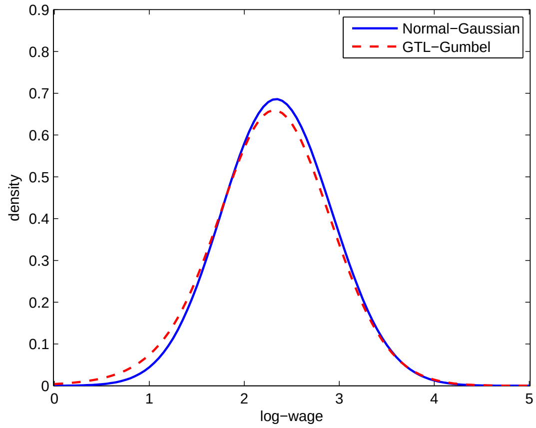
The sample selection model is widely used. Its estimates not only inform on the role of explanatory variables, but also permit prediction of the outcome variable among

[36]The predicted wage unconditional density of $\ln w$ is computed as

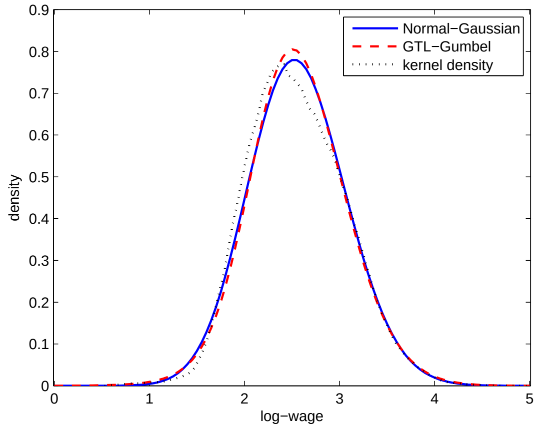
$$\int f_\varepsilon(\ln w|X)dX = N^{-1}\sum_i^N f_\varepsilon(\ln w|x_i).$$

It is evaluated at every 0.05 points of $\ln w$ from 0 to 5. The conditional distribution is similarly obtained by replacing the density with the conditional density and averaging over the samples with $s_i = 1$.

Figure 4.7: Differences between normal-Gaussian and GTL-Gumbel: Predicted Wage Distribution



(a) unconditional on working (log-wage offers)



(b) conditional on working (observed log-wages)

those for whom outcomes are not observed. Attanasio et al. (2012) use such predicted outcomes in 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 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.[37]

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 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

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

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.[38]

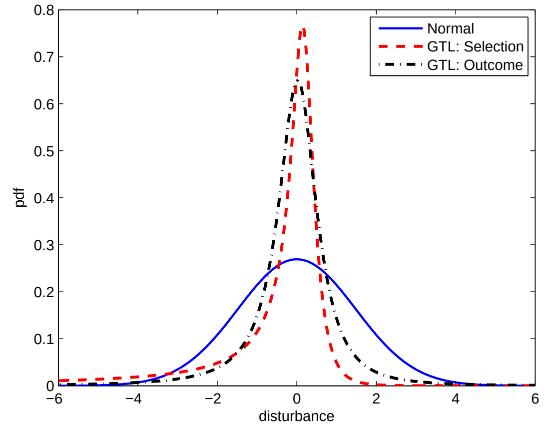
Estimates are reported in Table 4.11. The best copula is nJoe, based on the log likelihood values (Table 4.8). However, a Vuong test of GTL-nJoe against GTL-nGumbel, which attains the second largest log likelihood value, is inconclusive. Even in comparison with GTL-Gaussian, the Vuong test is not significantly 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 4.8a demonstrates the large difference of these GTL densities with the normal density (standardized by the interquartile range). The cdf that yields selection probabilities (Figure 4.8b) is therefore quite different as well. According to Table 4.11, 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 4.8 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.

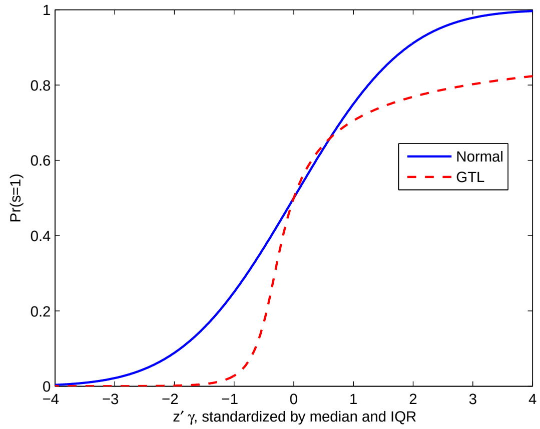
Interestingly, the GTL-nJoe and GTL-Gaussian contour plots are visually very similar (Figures 4.8d and 4.8e). This is due to weak dependence that the estimated τ indicates. With weak dependence, the difference in the Joe and Gaussian copula functions is not visible. The coefficients in the outcome equation are only different

[38] 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 log wage 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.

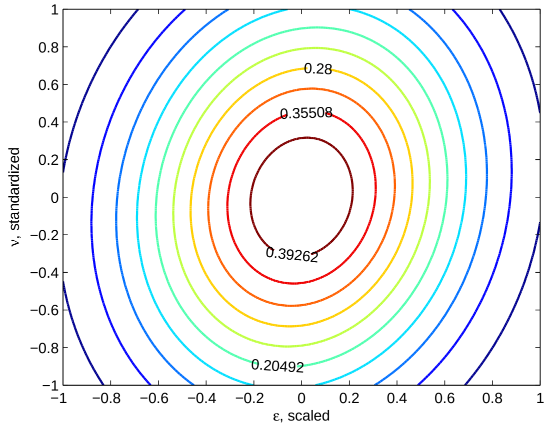
Figure 4.8: Differences between normal-Gaussian and GTL-copula: Wages of school-aged children



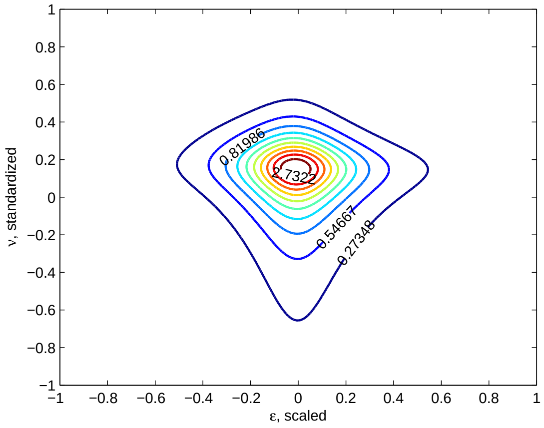
(a) PDF comparison



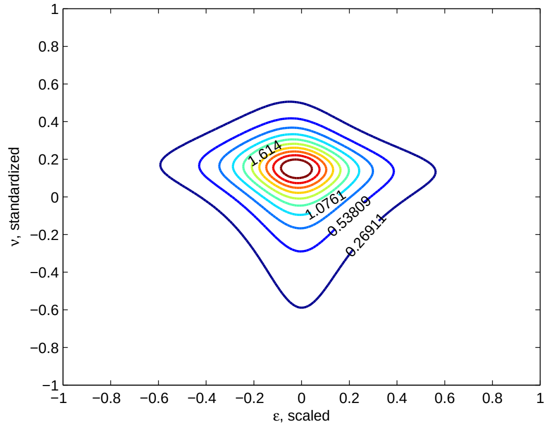
(b) $Pr(s_i = 1)$: GTL versus probit



(c) Contour plot: Normal-Gaussian



(d) Contour plot: GTL-nJoe



(e) Contour plot: GTL-Gaussian

Table 4.11: 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)
constant	-5.647	(0.947)	-4.108	(3.824)	-3.684	(3.540)
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	

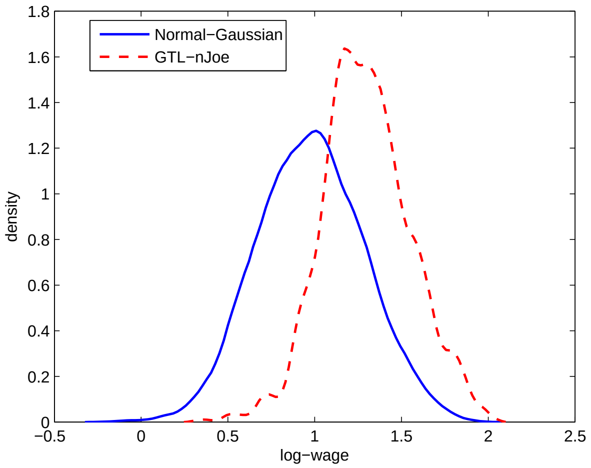
[a] Standard errors are clustered at a community level.

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

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

Figure 4.9: Kernel density plots of predicted log-wages



community'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.

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 4.9, 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.

4.6.4 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.[39] 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 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.[40] 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 4.5.4.

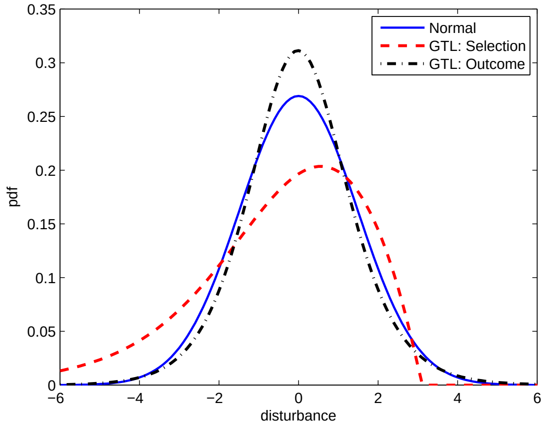
Table 4.12 reports the estimation results; demographic variables are suppressed to save space. In this application, the optimal copula is Clayton, but the maximized likelihood value of the Clayton copula is only slightly larger than that of the Gaussian copula, and Vuong's test does not discriminate 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.

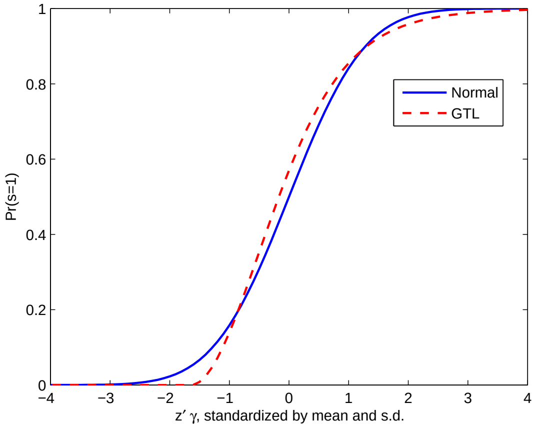
[39] The data are available online at `http://cameron.econ.ucdavis.edu/mmabook/mmaprograms.html`.

[40] The variables are defined in Table 4.B.4; see also Cameron and Trivedi (2005, Table 20.4).

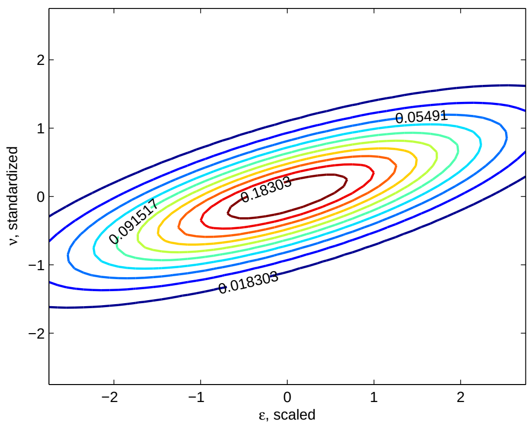
Figure 4.10: Differences between normal-Gaussian and GTL-copula: Health expenditures



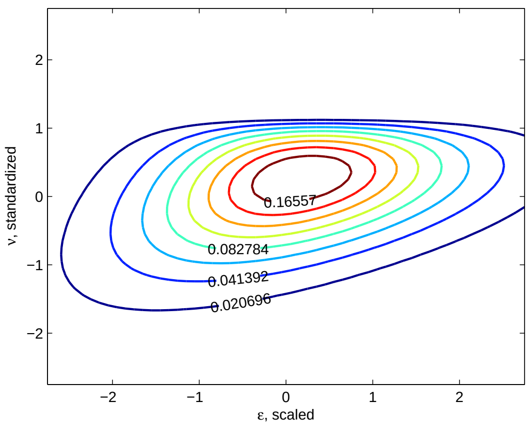
(a) PDF comparison



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



(c) Contour plot: Normal-Gaussian



(d) Contour plot: GTL-Clayton

Table 4.12: 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 4.B.4.
[a] White-robust standard errors.
[b] The selection equation is standardized by mean and standard deviation.

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.

4.6.5 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.[41]

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 + \$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 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

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

Table 4.13: 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.

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.

Table 4.14: Estimation results: Speeding tickets (outcome equation)

	Normal-Gaussian		GTL-nJoe		GTL-Gaussian	
Variables	Coef.	(S.E.) [a]	Coef.	(S.E.) [a]	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.

Tables 4.13 and 4.14 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. 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 4.8).

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 4.11c and 4.11d shows the contour plots implied by these estimators. As an intermediate form, the GTL-Gaussian estimator yields a small and insignificant degree of dependence.

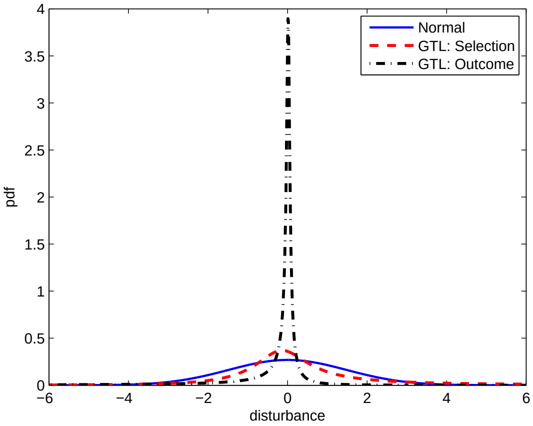
The values of $\hat{\alpha_\nu}$ and $\hat{\delta_\nu}$ indicate that the disturbances driving the selection equation are considerably skewed and heavy-tailed. Even the first moment of this GTL distribution cannot be defined. Accordingly, the selection equation is standardized by median and interquartile 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 4.15 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 4.12): 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.

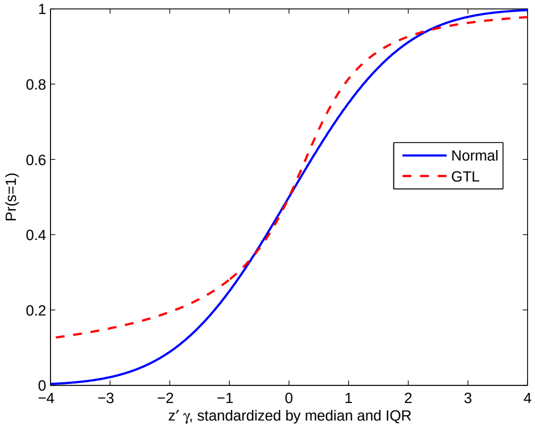
The difference in the outcome equation is even more striking. The disturbances are sharply peaked and have thick tails (Figure 4.11a). Even though the coefficients on some of variables are still statistically significant, no variable other than lnMPHOVER is economically significant any longer. The results 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 im-

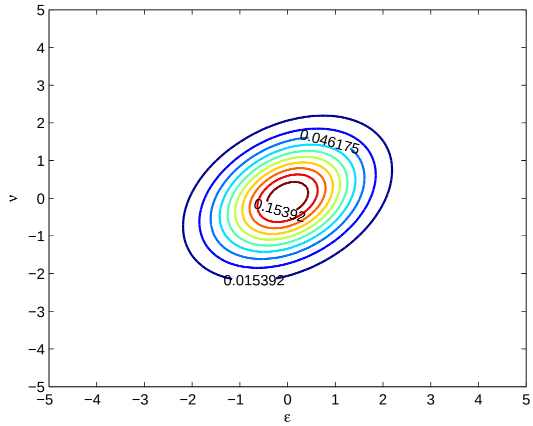
Figure 4.11: Differences between normal-Gaussian and GTL-copula: Speeding tickets



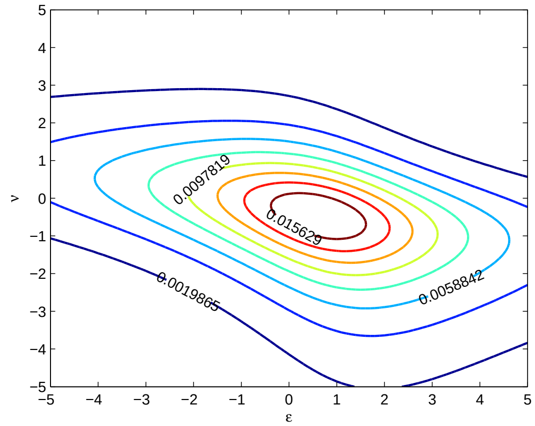
(a) PDF comparison



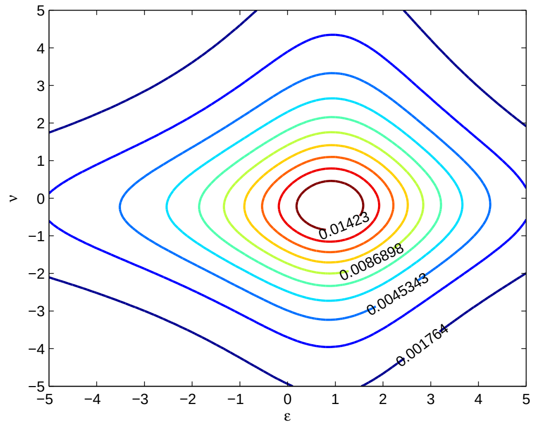
(b) $Pr(s_i = 1)$: GTL versus probit



(c) Contour Plot: Normal-Gaussian



(d) Contour Plot: GTL-nJoe



(e) Contour Plot: GTL-Gaussian

Table 4.15: Marginal effects on the probability of issuing a speeding ticket

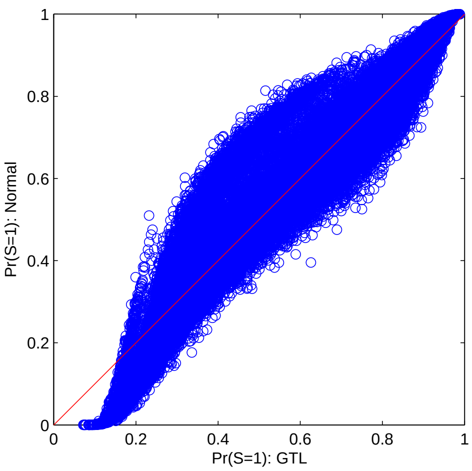
Variables	Normal-Gaussian Estimate	(S.E.) [a]	GTL-nJoe Estimate	(S.E.) [a]	GTL-Gaussian Estimate	(S.E.) [a]
lnMPHOVER	0.489	(0.025)	0.508	(0.096)	0.510	(0.186)
OUTTOWN [b]	0.078	(0.010)	0.053	(0.009)	0.065	(0.013)
OR [b]	0.147	(0.046)	0.066	(0.029)	0.123	(0.057)
FEMALE [b]	-0.061	(0.005)	-0.042	(0.006)	-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 4.12: Predicted probability $Pr(s_i = 1)$: GTL-nJoe vs normal-Gaussian



plications. 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.[42]

[42]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.

4.6.6 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%).[43]

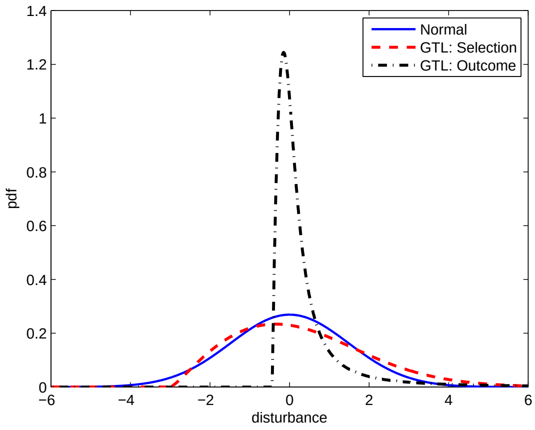
Dispute severity is measured by an index that combines the level of hostility and the number of fatalities. States express their interest similarity by being involved in similar alliances. Military balance is expressed as the ratio of the military capacity of the strongest member of a pair of states 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.[44]

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

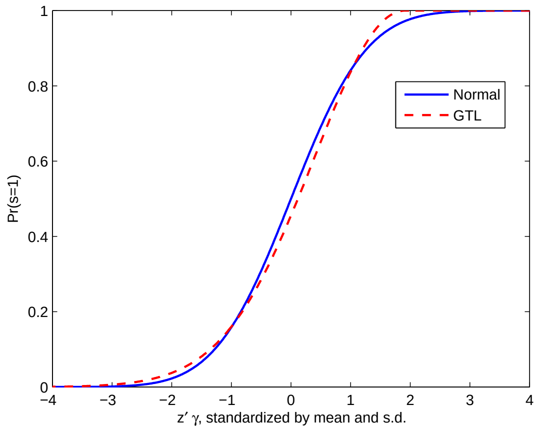
[43]The data are available online at `http://jcr.sagepub.com/content/47/6.toc`. For more detail on the variables, see also Oneal and Russett (1999).

[44]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.

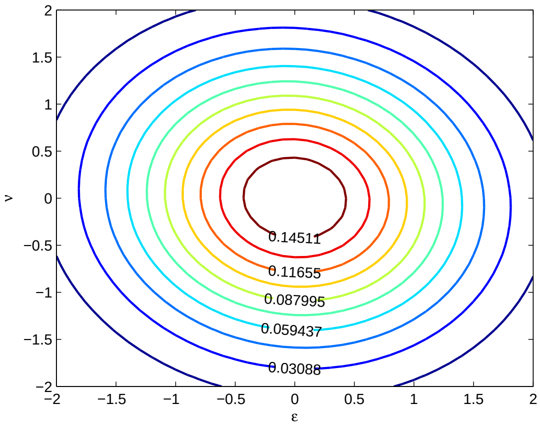
Figure 4.13: Differences between normal-Gaussian and GTL-Clayton: Severity of interstate disputes



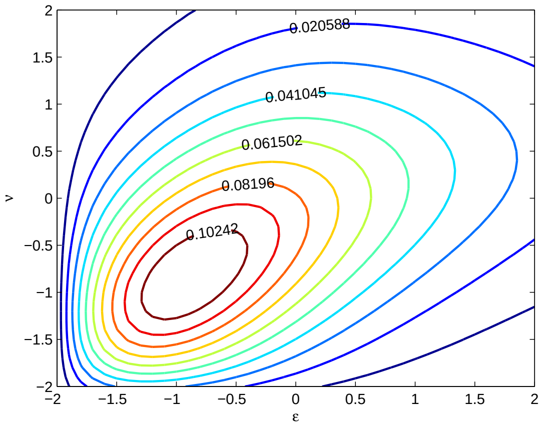
(a) PDF comparison



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



(c) Contour Plot: Normal-Gaussian



(d) Contour Plot: GTL-Clayton

Table 4.16: 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.

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 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 4.16 and Figure 4.13. The marginal distribution of the selection disturbances is nonnormal mostly because it lacks a left tail (Figure 4.13a): 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 4.17.

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

Table 4.17: 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.

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.

4.7 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. The GTL-copula estimator indicates slightly higher returns to education and experience than the normal-Gaussian estimator in the U.S.

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 4.8), 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.

4.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 4.2 and 4.3 with a reduced ($\tau = 0.2$) and a strengthened ($\tau = 0.5$) level of dependence.

Table 4.18: 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 4.2.

Table 4.19: 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 4.2.

Table 4.20: 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 4.3.

Table 4.21: 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 4.3.

4.B Variable Definitions and Summary Statistics

Table 4.B.1: Wages of married women, Portugal [a]

		Whole Sample		Selected Sample	
Number of Obs.		2,339		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 4.B.2: Wages of married women, USA-CPS [a]

		Whole Sample		Selected Sample	
Number of Obs.		36,803		23,496	
Variable	Definition	Mean	Std. dev.	Mean	Std. dev.
Selection	1 if at work	0.638	0.480		
Outcome	log of hourly wage (in dollars)			2.548	0.508
YRSCH	years of schooling	13.860	2.119	14.070	2.105
AGE	Age in years	41.100	10.332	40.572	9.749
EXP	AGE - YRSCH - 6	21.239	10.617	20.502	10.006
lnHUSBW	log of husband's wage	6.587	0.638	6.573	0.597
AGEdif	Age difference between husband and wife	0.171	0.339	0.170	0.333
BLACK	1 if black	0.063	0.242	0.067	0.250
HISPANIC	1 if hispanic	0.066	0.249	0.062	0.241
ASIAN	1 if Asian	0.027	0.161	0.028	0.165
INDIAN	1 if Indian	0.011	0.103	0.010	0.098

[a] Source: Derived from the MORG of CPS in 2000; The raw data are downloadable from the NBER website `www.nber.org`.

Table 4.B.3: Wages of school-aged children, Mexico [a]

		Whole Sample		Selected Sample	
Number of Obs.		15,526		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 [c]	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 4.B.4: Health expenditure, USA [a]

		Whole Sample		Selected Sample	
Number of Obs.		5,574		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 4.B.5: 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 4.B.6: Severity of interstate disputes [a]

		Whole Sample		Selected Sample	
Number of Obs.		149,004		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.

Bibliography

Abel, A. B., Eberly, J. C., 1994. A unified model of investment under uncertainty. The American Economic Review 84 (5), 1369–1384.

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, C., Garcia-Serrano, C., Hernanz, V., 2005. Firm-provided traning and temporary contracts. Spanish Economic Review 7 (1), 67–88.

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.

Amuedo-Dorantes, C., Serrano-Padial, R., 2007. Wage growth implications of fixed-term employment: An analysis by contract duration and job mobility. Labour Economics 14 (5), 829–847.

Arulampalam, W., Booth, A. L., Bryan, M. L., 2004. Training in europe. Journal of the European Economic Association 2 (2-3), 346–360.

Asphjell, M. K., Letterie, W. A., Nilsen, O. A., Pfann, G. A., 2010. Sequentiality versus simultaneity: Interrelated factor demand. Discussion Paper 5359, IZA.

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.

Bentolila, S., Dolado, J. J., 1994. Labour flexibility and wages: Lessons from Spain. Economic Policy 9 (18), 53–99.

Bhat, C. R., Sener, I. N., 2009. A copula-based closed-form binary logit choice model for accommodating spatial correlation across observational units. Journal of Geographical Systems 11, 243–272.

Blanchard, O., Landier, A., 2002. The perverse effects of partial labour market reform: Fixed-term contracts in France. The Economic Journal 112 (480), F214–F244.

Bloom, N., 2009. The impact of uncertainty shocks. Econometrica 77 (3), 623–685.

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.

Boockmann, B., Hagen, T., 2008. Fixed-term contracts as sorting mechanisms: Evidence from job durations in West Germany. Labour Economics 15 (5), 984–1005.

Booth, A. L., Francesconi, M., Frank, J., 2002. Temporary jobs: Stepping stones or dead ends? The Economic Journal 112 (480), F189–F213.

Brown, S., Sessions, J. G., 2003. Earnings, education, and fixed-term contracts. Scottish Journal of Political Economy 50 (4), 492–506.

Brown, S., Sessions, J. G., 2005. Employee attitudes, earnings and fixed-term contracts: International evidence. Review of World Economics 141 (2), 296–317.

Caballero, R. J., Engel, E. M. R. A., Haltiwanger, J., 1997. Aggregate employment dynamics: Building from microeconomic evidence. The American Economic Review 87 (1), 115–137.

Caballero, R. J., Engel, E. M. R. A., Haltiwanger, J. C., 1995. Plant-level adjustment and aggregate investment dynamics. Brookings Papers on Economic Activity 1995 (2), 1–54.

Cahuc, P., Postel-Vinay, F., 2002. Temporary jobs, employment protection and labor market performance. Labour Economics 9 (1), 63–91.

Cameron, C., Li, T., 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.

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

Chen, X., Fan, Y., 2005. Pseudo-likelihood ratio tests for semiparametric multivariate copula model selection. Canadian Journal of Statistics 33 (3), 389–414.

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

Connolly, H., Gottschalk, P., 2008. Wage cuts as investment in future wage growth. Labour 22 (1), 1–22.

Cooper, R., Willis, J. L., 2009. The cost of labor adjustment: Inferences from the gap. Review of Economic Dynamics 12 (4), 632–647.

Cooper, R. W., Haltiwanger, J. C., 2006. On the nature of capital adjustment costs. Review of Economic Studies 73 (3), 611–633.

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.

Davia, M. A., Hernanz, V., 2004. Temporary employment and segmentation in the Spanish labour market: An empirical analysis through the study of wage differentials. Spanish Economic Review 6 (4), 291–318.

de Graaf-Zijl, M., 2012. Compensation of on-call and fixed-term employment: The role of uncertainty. The Manchester School 80 (1), 6–27.

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.

Deb, P., Trivedi, P. K., Zimmer, D. M., 2009. Dynamic cost-offsets of prescription drug expenditures: Panel data analysis using a copula-based hurdle model. Working Paper 15191, NBER.

Dixit, A., January 01 1997. Investment and employment dynamics in the short run and the long run. Oxford Economic Papers 49 (1), 1–20.

Doms, M., Dunne, T., 1998. Capital adjustment patterns in manufacturing plants,. Review of Economic Dynamics 1 (2), 409–429.

Eberly, J. C., Mieghem, J. A. V., 1997. Multi-factor dynamic investment under uncertainty,. Journal of Economic Theory 75 (2), 345–387.

Elia, L., 2010. Temporary/permanent workers' wage gap: A brand-new form of wage inequality. Labour 24 (2), 178–200.

Eslava, M., Haltiwanger, J., Kugler, A., Kugler, M., 2010. Factor adjustments after deregulation: Panel evidence from Colombian plants. Review of Economics and Statistics 92 (2), 378–391.

Fazzari, S. M., Hubbard, R. G., Petersen, B. C., 1988. Financing constraints and corporate investment. Brookings Papers on Economic Activity (1), 141–206.

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.

Genest, C., Quessy, J., Rémilard, B., 2006. Goodness-of-Fit procedures for copula models based on the probability integral transfomation. Scadinavian Journal of Statistics 33 (2), 337–366.

Genest, C., Rivest, L.-P., 1993. Statistical inference procedures for bivariate Archimedean copulas. Journal of the American Statistical Association 88 (423), 1034–1043.

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.

Greene, W. H., 2008. Econometric analysis, 6th Edition. Pearson/Prentice Hall, Upper Saddle River, N.J.

Hagen, T., 2002. Do temporary workers receive risk premiums? assessing the wage effects of fixed-term contracts in West Germany by a matching estimator compared with parametric approaches. Labour 16 (4), 667–705.

Hamermesh, D. S., 1989. Labor demand and the structure of adjustment costs. The American Economic Review 79 (4), pp. 674–689.

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.

Hayashi, F., 1982. Tobin's marginal q and average q: A neoclassical interpretation. Econometrica 50 (1), 213–224.

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

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

Huard, D., Evin, G., Favre, A.-C., 2006. Bayesian copula selection. Computational Statistics & Data Analysis 51 (2), 809–822.

Imbens, G. W., Wooldridge, J. M., 2009. Recent developments in the econometrics of program evaluation. Journal of Economic Literature 47 (1), 5–86.

Jaramillo, F., Schiantarelli, F., Sembenelli, A., 1993. Are adjustment costs for labor asymmetric? An econometric test on panel data for Italy. The Review of Economics and Statistics 75 (4), 640–648.

Jimeno, J. F., Toharia, L., 1993. The effects of fixed-term employment on wages: Theory and evidence from Spain. Investigaciones Ecnomicas 17 (3), 475–494.

Kahn, L. M., 2007. The impact of employment protection mandates on demographic temporary employment patterns: International microeconomic evidence. The Economic Journal 117 (521), F333–F356.

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

Kim, H.-J., 2008. Moments of truncated Student-t distribution. Journal of the Korean Statistical Society 37 (1), 81–87.

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.

Lazear, E. P., 1990. Job security provisions and employment. The Quarterly Journal of Economics 105 (3), 699–726.

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.

Letterie, W. A., Pfann, G. A., 2007. Structural identification of high and low investment regimes. Journal of Monetary Economics 54 (3), 797–819.

Leung, S. F., Yu, S., 1996. On the choice between sample selection and two-part models. Journal of Econometrics 72 (12), 197–229.

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.

Manning, C., Roesad, K., 2007. The Manpower Law of 2003 and its implementing regulations: Genesis, key articles and potential impact. Bulletin of Indonesian Economic Studies 43 (1), 59–86.

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.

Mertens, A., Gash, V., McGinnity, F., 2007. The cost of flexibility at the margin. comparing the wage penalty for fixed-term contracts in Germany and Spain using quantile regression. Labour 21 (4), 637–666.

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., 1984. A method of moments interpretation of sequential estimators. Economics Letters 14 (2), 201–206.

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.

Nilsen, O. A., Salvanes, K. G., Schiantarelli, F., 2007. Employment changes, the structure of adjustment costs, and plant size. European Economic Review 51 (3), 577–598.

Nilsen, O. A., Schiantarelli, F., 2003. Zeros and lumps in investment: Empirical evidence on irreversibilities and nonconvexities. The Review of Economics and Statistics 85 (4), 1021–1037.

Oi, W. Y., 1962. Labor as a quasi-fixed factor. The Journal of Political Economy 70 (6), 538–555.

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.

Pfann, G. A., Palm, F. C., 1993. Asymmetric adjustment costs in non-linear labour demand models for the Netherlands and U.K. manufacturing sectors. The Review of Economic Studies 60 (2), 397–412.

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

Polder, M., Verick, S., 2004. Dynamics of labour and capital adjustment - a comparison of Germany and the Netherlands. Discusson Paper 1212, IZA.

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.

Rosenbaum, P. R., Rubin, D. B., 1983. The central role of the propensity score in observational studies for causal effects. Biometrika 70 (1), 41–55.

Rosenbaum, S., 1961. Moments of a truncated bivariate normal distribution. Journal of the Royal Statistical Society.Series B (Methodological) 23 (2), 405–408.

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

Sakellaris, P., 2004. Patterns of plant adjustment. Journal of Monetary Economics 51 (2), 425–450.

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

Schundeln, M., 2007. Appreciating depreciation: Physical capital depreciation in a developing country, Unpublished Manuscript.

Shapiro, M. D., 1986. The dynamic demand for capital and labor. The Quarterly Journal of Economics 101 (3), 513–542.

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.

Tobin, J., 1969. A general equilibrium approach to monetary theory. Journal of Money, Credit and Banking 1 (1), 15–29.

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

Trivedi, P. K., Zimmer, D. M., 2007b. 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., 1993. Measuring the unidentified parameter of the roy model of selectivity. Journal of Econometrics 57 (1), 69–89.

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.

White, H., 2000. A reality check for data snooping. Econometrica 68 (5), 1097–1126.

Wilde, J., 2000. Identification of multiple equation probit models with endogenous dummy regressors. Economics Letters 69 (3), 309–312.

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

World Bank, 2005. Doing business in 2005. The World Bank, Washington D.C.

World Bank, 2010. Indonesia jobs report: Towards better jobs and security for all. Tech. rep.

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.