DISTRIBUTIONS OF TREATMENT EFFECTS IN SWITCHING REGIMES MODELS: PARTIAL IDENTIFICATION, CONFIDENCE SETS, AND AN APPLICATION

By

Jisong Wu

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Economics

August, 2009

Nashville, Tennessee

Approved:

Professor Yanqin Fan

Professor Tong Li

Professor Ronald Masulis

Professor Peter Rousseau

Professor Bryan Shepherd

To my wife Xia, my son Kylin, my mother-in-law Zailan

and

To my parents Yanfan and Chunmei

ACKNOWLEDGMENTS

First, and most deeply, I would thank my advisor, Professor Yanqin Fan. Her profound knowledge and excellent teaching incited my interest in econometrics when I just entered the PhD program. I'm very lucky to be her student and learn a great deal from her. Her guidance, encouragement and support help me get through many difficulties in my research and beyond. The dissertation is based on the joint work with her and it could not have been possible without her dedication of countless hours on it.

I am also very thankful to my committee members, Professor Tong Li, Professor Ronald Masulis, Professor Peter Rousseau, and Professor Bryan Shepherd for their valuable comments and suggestions that have improved the quality of the dissertation substantially. In special, I would thank Professor Bryan Shepherd for his generous financial support in my last year of PhD study.

Other faculty members, staff and classmates in the Department of Economics also gave me tremendous help over the years of my PhD study. I would like to thank them all.

My final, and most heartfelt, acknowledgment must go to my wife Xia He. Without her love, support and taking care of our new born baby, I could not have been able to focus on the dissertation.

TABLE OF CONTENTS

Page

ACKNOWLEDGMENTS iii

LIST OF TABLES vi

LIST OF FIGURES vii

Chapter

I INTRODUCTION 1

Introduction 1

II SIMPLE ESTIMATORS OF AVERAGE TREATMENT EFFECTS IN SWITCHING REGIMES MODELS WITH NORMALMEAN-VARIANCEMIXTURE COPULAS 9

Introduction 9

NMVMC-SRMs 13

NMVM Distribution and its Copula 14

NMVM-SRMs and NMVMC-SRMs 16

Two-Step Estimation of NMVMC-SRMs 18

Two-Step Estimation of NMVM-SRMs 21

Two-Step Estimation of NMVMC-SRMs 24

Asymptotic Properties of the Two-Step Estimator in NMVMC-SRMs. 26

Simple Estimators of Four Treatment Parameters 28

Expressions for ATE, TT, LATE, and MTE 28

Estimators and Their Asymptotic Properties 30

Monte Carlo Simulation 33

Conclusion 37

III PARTIAL IDENTIFICATION OF THE DISTRIBUTION OF TREATMENT EFFECTS IN SWITCHING REGIMES MODELS AND ITS CONFIDENCE SETS . 38

Introduction 38

Normal Mean Variance Mixture Switching Regimes Models and Parameter Identification 44

Parameter Identification/Partial Identification 44

Distribution Bounds in NMVM-SRMs 46

Sharp Bounds on the Joint Distribution of Potential Outcomes 47

Sharp Bounds on the Distribution of Treatment Effects 50

Distribution Bounds in Semiparametric SRMs 54

Sharp Bounds on the Distribution of a Difference of Two Random Variables 55

Semiparametric SRMs 59

Some Applications of the Distribution Bounds 63

A Comparison of the two sets of Bounds 64

Bounds on F_{10}^{Y} in Semiparametric SRMs with Bivariate NMVM Distributions 65
Bounds on F_{Δ} in Semiparametric SRMs with Bivariate NMVM Distributions 66
Bounds on $F_{\Delta}(\cdot|D=1)$ and the Propensity Score 71
Confidence Sets for $F_{\Delta}(\delta)$ in Semiparametric SRMs 75
Conclusion 78

IV TREATMENT EFFECT STUDY FOR FIRMS WITH ACCELERATED EQUITY OFFERINGS 80
Introduction 80
Data and Issuer Characteristics 83
Models and Main Results 84
Conclusion 87

APPENDIX 89

BIBLIOGRAPHY 118

LIST OF TABLES

Table Page

1 Bias of Estimators of ATE and TT 101

2 Bias of Parameters in SRM (multiplied by 100) 101

3 The number of offerings and total proceeds 102

4 Firms and offer characteristics (mean) 102

5 Determinants of the choice of flotation (Accelerated=1) 103

6 Estimates of the ROA equation for the accelerated underwriting 103

7 Estimates of the ROA equation for the non-accelerated underwriting 104

8 Estimates of the CAR equation for the accelerated underwriting 104

9 Estimates of the CAR equation for the non-accelerated underwriting 105

10 Average Treatment Effects 105

LIST OF FIGURES

Figure Page

1 Sample correction terms for skewed t model and Gaussian model, and density functions of skew t distribution and Gaussian distribution with different values of the skewness parameter in the selection error 106

2 Sharp bounds on the distribution of treatment effects, $\sigma_1 = \sigma_0 = 1$. Solid curves are bounds assuming bivariate normality for (U_{ji}, ϵ_i), $j = 1, 0$, and dashed curves are bounds under the trivariate normality assumption for $(U_{1i}, U_{0i}, \epsilon_i)$. 107

3 Treatment effect at a given δ — $\sigma_1 = 1$, $\rho_{1\varepsilon} = 0.5$, and $\rho_{0\varepsilon} = 0.5$. Solid curves are bounds assuming bivariate normality for (U_{ji}, ϵ_i), $j = 1, 0$, and dashed curves are bounds under the trivariate normality assumption for $(U_{1i}, U_{0i}, \epsilon_i)$. 108

4 Sharp bounds on the distribution of treatment effects, $\sigma_1 = \sigma_0 = 1$. Solid curves are bounds assuming (U_{ji}, ϵ_i), $j = 0, 1$, follows bivariate student's t distribution with 4 degrees of freedom and dashed curves are bounds assuming $(U_{1i}, U_{0i}, \epsilon_i)$ follows trivariate student's t distribution with 4 degrees of freedom. 109

5 Sharp bounds on the distribution of treatment effects for the treated — $ATE = 0$, $\sigma_1 = \sigma_0 = 1$, and the Propensity Score = 0.1. In (a) and (c), $\rho_{1\varepsilon} = 0.5$ and $\rho_{0\varepsilon} = -0.5$, while in (b) and (d), $\rho_{1\varepsilon} = \rho_{0\varepsilon} = 0.5$. 110

6 Treatment effect for the treated for subpopulations and the probability that a person in a subpopulation benefits from the treatment, where $ATE = 0$, $\rho_{1\varepsilon} = 0.5$, $\rho_{0\varepsilon} = -0.5$, and $\sigma_1 = \sigma_0 = 1$. 111

7 Treatment effect for the treated for subpopulations and the probability that a person in a subpopulation benefits from the treatment, where $ATE = 0$, $\rho_{1\varepsilon} = -0.5$, $\rho_{0\varepsilon} = 0.5$, and $\sigma_1 = \sigma_0 = 1$. 112

8 Treatment effect for the treated for subpopulations and the probability that a person in a subpopulation benefits from the treatment, where $ATE = -0.5$, $\rho_{1\varepsilon} = 0.5$, $\rho_{0\varepsilon} = -0.5$, and $\sigma_1 = \sigma_0 = 1$. 113

9 Treatment effect for the treated for subpopulations and the probability that a person in a subpopulation benefits from the treatment, where $ATE = 0.5$, $\rho_{1\varepsilon} = -0.5$, $\rho_{0\varepsilon} = 0.5$, and $\sigma_1 = \sigma_0 = 1$. 114

10 Treatment effect for the treated for subpopulations and the probability that a person in a subpopulation benefits from the treatment, where $ATE = 0$, $\rho_{1\varepsilon} = 0.95$, $\rho_{0\varepsilon} = -0.5$, and $\sigma_1 = \sigma_0 = 1$. 115

11 Treatment effect for the treated for subpopulations and the probability that a person in a subpopulation benefits from the treatment, where $ATE = 0$, $\rho_{1\varepsilon} = 0.95$, $\rho_{0\varepsilon} = 0.5$, and $\sigma_1 = \sigma_0 = 1$. 116

12 Expected gain from choosing accelerated underwriting in terms of return on assets one year after the equity issuance 116

13 Expected gain from choosing accelerated underwriting in terms of cumulative abnormal return one year after the equity issuance 117

CHAPTER I

INTRODUCTION

Introduction

The term ‘treatment effect’ refers to the causal effect of a variable (treatment) on an outcome variable of interest. It originates in the medical literature concerned with the causal effects of a treatment, such as an experimental drug or a new surgical procedure. Concepts and tools in the treatment effect study can be used to analyze economic data. We can view any economic factor or policy on an individual or organization as a treatment, and many interesting or important empirical questions in economics can be viewed as questions about some types of treatment effects. Examples include the effects of government programmes and policies, such as those that subsidize training for disadvantaged workers, and the effects of individual choices like college attendance.

The notion of a causal effect can be made more precise using a conceptual framework that postulates a set of potential outcomes that could be observed in alternative states of the world. Originally introduced by statisticians in the 1920s as a way to discuss treatment effects in randomized experiments, the potential outcomes framework has become the conceptual workhouse for non-experimental (observational) as well as experimental studies in many fields (see Holland, 1986, for a survey and Rubin, 1974, 1977, for influential early contributions). In this dissertation, we focus on a binary treatment and there are two potential outcomes: with and without the treatment. The difference of two potential outcomes is considered as treatment effect, which is only caused by the change of treatment status. A standard setting in the treatment effect study requires two groups of individuals, a treatment group with individuals taking the treatment and a control group with individuals not taking the treatment, to learn about marginal distributions of two potential outcomes and average treatment effects. Selection of the two groups becomes an essential issue in the treatment effect study. We distinguish two types of selections, selection on observables and selection on unobservables. Selection on observables means the selection of individuals into either the treatment group or the control group is randomized or randomized after controlling for observable factors that affect both the selection and the outcome. Selection on unobservables means that there are factors affecting the outcome in the selection that we can not observe or control for. These uncontrolled factors in the selection are also called private information. Efforts are made by statisticians and econometricians to develop models to deal with data involving selection on observables or/and selection on unobservables. The models that can deal with both selection issues are also called sample-selection models or self-selection models and they are distinct from models that are only able to deal with selection on observables. Most data in economics are observational data involving self selection or individual choice. The private information in self selection might also have impact on the outcome. Treatment effect study without accounting for the impact of private information will lead to misleading conclusions. Thus our main focus in this dissertation is on self-selection models. Specifically, we use switching regimes models (SRMs) to address the self-selection, in which there are two outcome equations corresponding to two potential states and a selection equation to decide which state can be observed.

One of the earlier econometric models to address self-selection is the Roy model (1951). The income maximizing Roy model was developed to explain occupational choice and its consequences for the distribution of earnings when individuals differ in their endowments of occupation-specific skills. Switching regimes models (SRMs) extend the Roy model of self-selection by allowing a more general decision rule for selecting into different states. By allowing a more general decision/selection rule, SRMs enjoy a much wider scope of applications than the Roy model. Recently, SRMs have been used to evaluate average effects of a policy intervention using choice data. Heckman, Tobias, and Vytlacil (2003) derived expressions for four average treatment effect parameters for a Gaussian copula SRM and a Student's t copula SRM with normal outcome errors and non-normal selection errors[1]. The four average treatment effect parameters are the average treatment effect (ATE), the treatment effect for the treated (TT), the local average treatment effect (LATE, Imbens and Angrist, 1994), and the marginal treatment effect (MTE, Bjorklund and Moffitt, 1987; Heckman, 1997; Heckman and Vytlacil, 1999, 2000a, 2000b).

One of the most commonly used SRMs in empirical work is the Gaussian SRM in which the error vector (consisting of two outcome errors and one selection error) follows a trivariate normal distribution. This is partly due to the simplicity of Heckman's two-step estimation procedure for Gaussian SRM introduced in Heckman (1976). The Gaussian SRM has been extended to allow for non-normal marginal distributions in the errors in Lee (1982, 1983), Heckman, Tobias, and Vytlacil (2003), and Li, Poirier, and Tobias (2004). The models in Heckman, Tobias, and Vytlacil (2003) essentially assume that the trivariate error vector follows a distribution with either the Gaussian copula or a trivariate Student's t copula. When the outcome errors are normal or Student's t and the selection error has an arbitrary distribution, Lee (1982, 1983), Heckman, Tobias, and Vytlacil (2003) show that the model parameters can be consistently estimated by a two-step estimation procedure extending Heckman's two-step procedure for Gaussian SRM. By extending Heckman's two-step procedure to Student's t outcome error(s), fat tailed outcome error(s) can be accounted

[1] They didn't use the concept of copulas, but their models can be interpreted this way.

for in the two-step procedure of Lee (1982, 1983), Heckman, Tobias, and Vytlacil (2003). Indeed simulation results in Heckman, Tobias, and Vytlacil (2003) show that their two-step estimation procedure works well for fat-tailed distributions when used to estimate ATE, TT, and LATE and that for ATE, TT, even Heckman's two-step procedure based on Gaussian SRM yields minor biases. Although not reported in their paper, Heckman, Tobias, and Vytlacil (2003) mention in their Footnote 7 that 'When generating data from highly asymmetric distributions, such as a $\chi^2(3)$, we do see larger biases' (in the estimates of ATE and TT). Although the SRM with Student's t outcome errors in Heckman, Tobias, and Vytlacil (2003) allow for fat tails, it does not allow for skewness in the outcome errors and existing two-step estimation procedures do not account for skewness in the outcome errors either. The purpose of Chapter II is to bridge this gap in the existing literature.

Although average treatment effects are identifiable in SRMs, they are also limited in their ability to address a wide range of interesting economic/policy questions because of the non-identifiability[2] of the joint distribution of potential outcomes in SRMs. Even in the 'textbook' Gaussian SRM, the correlation coefficient between the potential outcomes or equivalently the joint distribution of the potential outcomes is not identifiable. In a study of a sectoral labor market using the Gaussian SRM, Vijverberg (1993) showed that a number of interesting economic questions including the share of 'productive' workers employed in a sector can not be answered without knowledge of the joint distribution of the two potential outcomes. When used to study treatment effect defined as the difference between the two potential outcomes, important distributional aspects of the treatment effect other than its mean are not identified in SRMs. This partly explains why the current literature has mainly focussed on various measures of average treatment effects.

[2]In this dissertation, we use identification and point identification interchangeably. We say a parameter is not identified if it is not point identified.

Recently two approaches have been proposed to deal with the non-identifiability problem of the joint distribution of potential outcomes in the 'textbook' Gaussian SRM and some of its extensions. By employing the positive semidefiniteness of the covariance matrix of the outcome errors and the selection error, Vijverberg (1993) showed that in the 'textbook' Gaussian SRM, although unidentified, useful bounds can be placed on the correlation coefficient between the potential outcomes, that is, it is partially identified. Koop and Poirier (1997), Poirier (1998), and Poirier and Tobias (2003) demonstrated via Bayesian approach that these bounds often provide informative information on the unidentified correlation coefficient. Since the joint distribution of the potential outcomes in the 'textbook' Gaussian SRM depends on the unidentified correlation coefficient only (besides the identified marginal parameters), it is possible to place bounds on the joint distribution of the potential outcomes and on the distribution of the difference between the potential outcomes. In the second approach, restrictions are imposed on the dependence structure between the potential outcomes such that their joint distribution and the distribution of treatment effects are identified, see, e.g., Heckman, Smith, and Clements (1997), Biddle, Boden, and Reville (2003), Carneiro, Hansen, and Heckman (2003), Aakvik, Heckman, and Vytlacil (2005), Cunha and Heckman (2007), among others. Among other things, they demonstrated that knowledge of the joint distribution of potential outcomes and the distribution of treatment effects allows a much richer analysis of policy effects than average treatment effects. Questions that can be addressed include the proportion of people participating in the program who benefit from it in terms of having positive treatment effects; the proportion of the total population that benefits from the program; and which groups in an initial position benefit or lose from the program.

We take the first approach in Chapter III and extend the partial identification results in Vijverberg (1993), Koop and Poirier (1997), Poirier (1998), Poirier and Tobias (2003), and Li, Poirier, and Tobias (2004) to a general class of SRMs in which the joint distribution of the outcome errors and the selection error is assumed to follow a trivariate NMVM distribution referred to as NMVM-SRMs. The 'textbook' Gaussian and Student's t SRMs are members of NMVM-SRMs. For NMVM-SRMs, we provide sharp bounds or partial identification results on the correlation coefficient of the potential outcomes, their joint distribution, and the distribution of treatment effects. The distribution bounds established in NMVM-SRMs rely on two special properties of NMVM-SRMs: (i) the only unidentified parameter in a NMVM-SRM is the correlation coefficient between the two potential outcomes and (ii) the joint distribution of the two potential outcomes in a NMVM-SRM is also NMVM. The fact that the joint distribution of the potential outcomes in a SRM is not identifiable raises two issues: (i) is it possible to test the NMVM specification of the joint distribution of the potential outcomes? and (ii) are the results for NMVM-SRMs robust to the implied joint distribution of the potential outcomes? To address these issues, we establish sharp bounds on the joint distribution of potential outcomes and the distribution of treatment effects in the general class of semiparametric SRMs in Heckman (1990) in which the joint distribution of the trio of errors is completely unspecified. Our results rely on and supplement the point identification results in Heckman (1990).

Methodologically, the approach we use to bound the joint distribution of the potential outcomes and the distribution of treatment effects in semiparametric SRMs differs from that in NMVM-SRMs. Without specifying the joint distribution of the outcome errors and the selection error, the approach used to bound the distribution of the potential outcomes and the distribution of treatment effects in NMVM-SRMs breaks down. The new tool that we employ in Chapter III to establish bounds on the joint distribution of potential outcomes is the Fréchet-Hoeffding inequality on copulas. A straightforward application of this inequality allows us to bound the joint distribution of potential outcomes using the bivariate distributions of each outcome error and the selection error, where the latter distributions are known to be identified under general conditions, see Heckman (1990). To bound the distribution of treatment effects, we make use of existing results on sharp bounds on the distributions of functions of two random variables including the four simple arithmetic operations, see Williamson and Downs (1990). For a sum of two random variables, Makarov (1981), Rüschendorf (1982), and Frank, Nelsen, and Schweizer (1987) establish sharp bounds on its distribution, see also Nelsen (1999). These results have been used in Fan and Park (2006, 2008) to bound the distribution of treatment effects and the quantile function of treatment effects in the context of ideal social experiments where selection is random. Other applications of the Fréchet-Hoeffding inequality include Heckman, Smith, and Clements (1997) in which they bound the variance of treatment effects under the assumption of random selection; Manski (1997b) in which he established bounds on the mixture of two potential outcomes when the distribution of each outcome is known; Lee (2002) in which he presented bounds on the correlation coefficient between the potential outcomes in SRMS, and Fan (2005) in which she provided a systematic study on the estimation and inference on the sharp bounds on the correlation bounds.

The partial identification results established in Chapter III can be used to develop inference procedures for the joint distribution of potential outcomes and the distribution of treatment effects. There is a recent, but rapidly growing literature on inference for partially identified parameters, including Imbens and Manski (2004), Bugni (2007), Canay (2007), Chernozhukov, Hong, and Tamer (2007), Fan and Park (2007), Romano and Shaikh (2006), Stoye (2007), Andrews and Guggenberger (2007), and Andrews and Soares (2007), among others. We refer the reader to Fan and Park (2007) for more references. A complete treatment of this important issue is beyond the scope of this dissertation. However, we demonstrate the feasibility of inference by constructing an asymptotically uniformly valid and non-conservative confidence set (CS) for the distribution of treatment effects in a semiparametric SRM.

In Chapter IV, we focus on an application of the models developed in Chapter II and Chapter III in corporate finance. One very interesting phenomenon in corporate finance is the global rise of accelerated equity offers (Bortolotti, Megginson and Smart, 2008) in recent years. Autore, Hutton and Kovacs (2009) find that shelf registered firms with accelerated underwriting underperform those with non-accelerated underwriting. They hypothesize that the choice of flotation methods can be served as a signal of issuer quality and lower quality firms intend to use accelerated underwriting. However, their hypothesis can not explain the increasing popularity of accelerated underwriting among seasoned equity offerings (SEOs) in recent years. In addition, we notice that shelf registered firms are required by security and exchange commission (SEC) to be large and financially sound (Eckbo, Masulis, Norli, 2007). For those shelf registered firms, they can either choose accelerated underwriting or traditional non-accelerated underwriting. There is private information involved in their decision to choose a flotation method of their SEOs. In Chapter IV, we use our models to study the impact of a firm's accelerated underwriting on its performance one year after the equity issuance.

CHAPTER II

SIMPLE ESTIMATORS OF AVERAGE TREATMENT EFFECTS IN SWITCHING REGIMES MODELS WITH NORMALMEAN-VARIANCEMIXTURE COPULAS

Introduction

Consider the following switching regimes model[1] (SRM):

$$
\begin{aligned}
Y_{1i} &= X_i'\beta_1 + U_{1i}, \\
Y_{0i} &= X_i'\beta_0 + U_{0i}, \qquad \text{(II.1)} \\
D_i &= I_{\{W_i'\gamma+\epsilon_i>0\}}, \; i = 1, \ldots, n,
\end{aligned}
$$

where D_i is a binary variable taking the value 1 if individual i participates in the program and taking the value zero if he chooses not to participate in the program, Y_{1i} is the outcome of individual i we observe if he participates in the program, and Y_{0i} is his outcome if he chooses not to participate in the program. The errors $\{U_{1i}, U_{0i}, \epsilon_i\}$ are assumed to be independent of the covariates $\{X_i, W_i\}$. We also assume the existence of an exclusion restriction, i.e., there exists at least one element of W_i which is not contained in X_i.

Switching regimes models (SRMs) extend the Roy model of self-selection by allowing a more general decision rule for selecting into different states. The income maximizing Roy model of self-selection was developed to explain occupational choice and its consequences for the distribution of earnings when individuals differ in their endowments of occupation-specific skills, see Heckman and Honore (1990). By allowing a more general decision/selection rule, SRMs enjoy a much wider scope of applications than the Roy model.

[1] Switching regimes models are also referred to as generalized sample selection models or generalized Roy models.

Recently, SRMs have been used to evaluate average effects of a policy intervention using choice data. Heckman, Tobias, and Vytlacil (2003) derived expressions for four average treatment effect parameters for a Gaussian copula SRM and a Student's t copula SRM with normal outcome errors and non-normal selection errors. The four average treatment effect parameters are the average treatment effect (ATE), the treatment effect for the treated (TT), the local average treatment effect (LATE, Imbens and Angrist, 1994), and the marginal treatment effect (MTE, Bjorklund and Moffitt, 1987; Heckman, 1997; Heckman and Vytlacil, 1999, 2000a, 2000b).

One of the most commonly used SRMs in empirical work is the Gaussian SRM in which $\{U_{1i}, U_{0i}, \epsilon_i\}$ follows a trivariate normal distribution. This is partly due to the simplicity of Heckman's two-step estimation procedure for Gaussian SRM introduced in Heckman (1976). The Gaussian SRM has been extended to allow for non-normal marginal distributions in the errors in Lee (1982, 1983), Heckman, Tobias, and Vytlacil (2003), and Li, Poirier, and Tobias (2004). The models in Heckman, Tobias, and Vytlacil (2003) essentially assume that the trivariate error vector $\{U_{1i}, U_{0i}, \epsilon_i\}$ follows a distribution with either the Gaussian copula or a trivariate Student's t copula. When the outcome errors are normal or Student's t and the selection error has an arbitrary distribution, Lee (1982, 1983), Heckman, Tobias, and Vytlacil (2003) show that the model parameters can be consistently estimated by a two-step estimation procedure extending Heckman's two-step procedure for Gaussian SRM. By extending Heckman's two-step procedure to Student's t outcome error(s), fat tailed outcome error(s) can be accounted for in the two-step procedure of Lee (1982, 1983), Heckman, Tobias, and Vytlacil (2003). Indeed simulation results in Heckman, Tobias, and Vytlacil (2003) show that their two-step estimation procedure works well for fat-tailed distributions when used to estimate ATE, TT, and LATE and that for ATE, TT, even Heckman's two-step procedure based on Gaussian SRM yields minor biases. Although not reported in their paper, Heckman, Tobias, and Vytlacil (2003) mention in their Footnote 7 that 'When generating data from highly asymmetric distributions, such as a $\chi^2(3)$, we do see larger biases' (in the estimates of ATE and TT). Although the SRM with Student's t outcome errors in Heckman, Tobias, and Vytlacil (2003) allow for fat tailes, it does not allow for skewness in the outcome errors and existing two-step estimation procedures do not account for skewness in the outcome errors either. The purpose of Chapter II is to bridge this gap in the existing literature.

First we propose a new class of SRMs, referred to as the Normal Mean-Variance Mixture Copula SRMs (NMVMC-SRMs) in which the bivariate distributions of $\{U_{1i}, \epsilon_i\}$ and $\{U_{0i}, \epsilon_i\}$ are constructed via the copula approach with NMVM copulas and arbitrary marginal distributions. The Gaussian copula and Student's t copula SRMs in Heckman, Tobias, and Vytlacil (2003) are members of this class.[2] In addition, the class of NMVM copulas includes asymmetric copulas as well allowing for asymmetric dependence between each outcome error and the selection error. An important subclass of NMVMC-SRMs is the class of NMVM-SRMs given by (II.1) with each pair of errors (U_{ji}, ϵ_i) following a bivariate NMVM distribution. The class of NMVM distributions includes Gaussian and Student's t distributions as special cases. In addition, it includes multi-modal and skewed distributions, see McNeil, Frey and Embrechts (2005). Thus, the class of NMVM-SRMs includes Gaussian and Student's t SRMs in Chib (2005) as special cases.

The second contribution of Chapter II is to develop a simple two-step estimation procedure for the class of NMVM-SRMs extending Heckman's two-step procedure for Gaussian SRM. Using the two-step procedure, we construct estimators of ATE, TT,

[2] In fact, Heckman, Tobias, and Vytlacil (2003) impose a trivariate Gaussian or Student's t copula structure on the error vector $(U_{1i}, U_{0i}, \epsilon_i)$ as opposed to bivariate Gaussian or Student's t copula structure on (U_{1i}, ϵ_i) and (U_{0i}, ϵ_i) separately.

LATE, and MTE in NMVM-SRMs and establish their asymptotic properties. In contrast to Heckman's two-step procedure, there are two correction terms in the second step for NMVM-SRMs; one for the dependence between each outcome error and the selection error as in the second step for Gaussian SRM and the other for skewness in each outcome error distribution. As a result, applying Heckman's two-step procedure to SRMs with skewed outcome distributions in general leads to inconsistent estimators of parameters in the potential outcomes equations and of the four average treatment effect parameters. This provides a theoretical explanation for the large biases in the estimates of ATE and TT mentioned in Footnote 7 in Heckman, Tobias, and Vytlacil (2003). Our simulation results using asymmetric distributions also reveal significant biases in the estimates of parameters in SRMs and/or of ATE and TT if skewness in the distribution of the error terms is not accounted for in the estimation procedure. More importantly, we find that our two-step procedure correcting for both skewness and dependence performs very well. We also extend our two-step estimation procedure for NMVM-SRMs to the subclass of NMVMC-SRMs in which the outcome errors follow univariate NMVM distributions and the selection error follows an arbitrary distribution. In general, the second step for the subclass of NMVMC-SRMs may involve a nonlinear regression, but for certain members of NMVM copulas such as Gaussian copula or Student's t copula with a known degree of freedom in Lee (1982, 1983), Heckman, Tobias, and Vytlacil (2003), the second step regression is linear.

The rest of Chapter II is organized as follows. In Section 2, we introduce the class of NMVMC-SRMs and some special cases. In Section 3, we propose a simple two-step estimation procedure for parameters in the potential outcomes equations for NMVM-SRMs and NMVMC-SRMs when the outcome errors follow NMVM distributions. In Section 4, we use our two-step estimation procedure to construct simple estimators of ATE, TT, LATE, and MTE, extending the estimators of Heckman, Tobias, and Vytlacil (2003) to a much wider class of SRMs. We present results from a small Monte Carlo simulation study in Section 5. Section 6 concludes. Techncial proofs are relegated to the Appendix.

NMVMC-SRMs

In this section, we introduce a general class of SRMs via the copula approach. Let $\epsilon_i \sim F(\epsilon; \alpha)$ for some $\alpha \in \mathcal{A} \subset \mathcal{R}^p$, where $\{F(\epsilon; \alpha) : \alpha \in \mathcal{A}\}$ is any family of parametric distribution functions with zero mean and variance 1. Let $U_{ji} \sim F_j(u; \alpha_j)$ for some $\alpha_j \in \mathcal{A}_j \subset \mathcal{R}^{p_j}$, where $\{F_j(u; \alpha_j) : \alpha_j \in \mathcal{A}_j\}$ is any family of parametric distribution functions with zero mean and variance σ_j^2, $j = 1, 0$. By Sklar's theorem (1959), for any copula function $C_j(u, v)$: $0 \leq u, v \leq 1$, the function $C_j(F(\epsilon; \alpha), F_j(u; \alpha_j))$ is a bivariate distribution function with marginal distributions $F(\epsilon; \alpha)$ and $F_j(u; \alpha_j)$ and copula function $C_j(u, v)$. This motivates us to assume

$$(\epsilon_i, U_{ji}) \sim C_j(F(\epsilon; \alpha), F_j(u; \alpha_j); \theta_j), \ j = 1, 0, \tag{II.2}$$

where $\{C_j(u, v; \theta_j) : \theta_j \in \times_j\}$ is a parametric family of copula functions. We refer the reader to Joe (1997) and Nelsen (1999) for excellent discussions of copulas and their properties. Instead of modeling the trivariate distribution of $U_{1i}, U_{0i}, \epsilon_i$ via the copula approach as in Heckman, Tobias, and Vytlacil (2003), we model the bivariate distributions of U_{1i}, ϵ_i and of U_{0i}, ϵ_i, as either Y_{1i} or Y_{0i} is observed for any given individual i but never both, so the joint distribution of the trio of errors $\{U_{1i}, U_{0i}, \epsilon_i\}$ is not identified. In addition to the Gaussian and Student's t copulas used in Lee (1982, 1983) and Heckman, Tobias, and Vytlacil (2003), Smith (2003, 2005) and Prieger (2002) respectively used Archimedean and Farlie-Gumbel-Morgenstern copulas.

The class of NMVMC-SRMs we propose in Chapter II is given by (II.1) and (II.2) with $\{C_j(u,v;\theta_j) : \theta_j \in \times_j\}$ specified as the class of NMVM copulas. This section contains two subsections. In the first subsection, we introduce the general class of NMVM distributions and their copulas. In the second subsection, we introduce the class of NMVMC-SRMs and the class of NMVM-SRMs.

NMVM Distribution and its Copula

We first present a definition of a d-dimensional normal mean variance mixture (NMVM) distribution. It can be found in McNeil, Frey, and Embrechts (2005).

Definition 1 *We say $V \sim NMVM_d(\xi, \mu, \Sigma, \zeta)$ with $\Sigma = AA'$ if V has the same distribution as $\mu + S\zeta + \sqrt{S}AZ$, where μ, ζ, A are respectively $d \times 1$, $d \times 1$, and $d \times k$ constant matrices, $Z \sim N_k(0, I_k)$, and $S \geq 0$ is independent of Z and follows a distribution with a parametric density function or probability function $f_S(\cdot;\xi)$.*

We use $NMVM_d(\cdot;\xi,\mu,\Sigma,\zeta)$ and $nmvm_d(\cdot;\xi,\mu,\Sigma,\zeta)$ to denote respectively the distribution function and density function of the NMVM distribution introduced in Definition 2.1. Based on Definition 2.1, we can put the parameters in $NMVM_d(\xi,\mu,\Sigma,\zeta)$ into two groups: those in the distribution of the mixing variable S denoted as ξ and the remaining parameters (μ, Σ, ζ). When $\zeta = 0$, the NMVM distribution or the distribution of V belongs to the class of symmetric NMVM distributions. A non-zero value of ζ introduces asymmetry into the distribution of V. It follows from the above definition that $V|S = s \sim N_d(\mu + s\zeta, s\Sigma)$. Thus, μ and Σ play the roles of the location vector and the dispersion matrix respectively.

The class of NMVM distributions includes two important subclasses. When the mixing variable S is a discrete random variable taking a finite number of values, we obtain the class of finite normal mixture distributions. When S follows the generalized inverse

Gaussian (GIG) distribution, denoted as $N^{-1}(\lambda, \chi, \psi)$, with density function:

$$f_S(s; \xi) = \frac{(\psi/\chi)^{\lambda/2}}{2K_\lambda\left(\sqrt{\chi\psi}\right)} s^{\lambda-1} \exp\left(-\frac{\chi s^{-1} + \psi s}{2}\right), \ \xi = (\lambda, \chi, \psi)', \tag{II.3}$$

we obtain the class of generalized hyperbolic (GH) distributions denoted as $GH_d(\lambda, \chi, \psi, \mu, \Sigma, \zeta)$. Among the parameters in ξ in a GH distribution, ψ is a scale parameter and (λ, χ) allow for flexible tails in the distribution of V, see Jogensen (1982) and Mencia and Sentana (2005) for more discussion. Let V denote a d-dimensional random vector following a GH distribution with parameters $\lambda, \chi, \psi, \mu, \Sigma, \zeta$, where μ and ζ are $d \times 1$ vectors, Σ is a $d \times d$ matrix, and parameters λ, χ, ψ satisfy: $\chi > 0$, $\psi \geq 0$ if $\lambda < 0$; $\chi > 0, \psi > 0$, if $\lambda = 0$; and $\chi \geq 0, \psi > 0$, if $\lambda > 0$. To simplify exposition, we use $GH_d(\lambda, \chi, \psi, \mu, \Sigma, \zeta)$ to denote this distribution and $GH_d(\cdot; \lambda, \chi, \psi, \mu, \Sigma, \zeta)$, $gh_d(\cdot; \lambda, \chi, \psi, \mu, \Sigma, \zeta)$ to denote respectively its distribution function and density function. If Σ is positive definite, then the probability density function of V has the following closed-form expression (see Mencia and Sentana (2005)):

$$gh_d(x; \lambda, \chi, \psi, \mu, \Sigma, \zeta) = c\frac{K_{\lambda-(d/2)}\left(\sqrt{\left(\chi + (x-\mu)' \Sigma^{-1}(x-\mu)\right)\left(\psi + \zeta'\Sigma^{-1}\zeta\right)}\right) e^{(x-\mu)'\Sigma^{-1}\zeta}}{\left[\sqrt{\left(\chi + (x-\mu)' \Sigma^{-1}(x-\mu)\right)\left(\psi + \zeta'\Sigma^{-1}\zeta\right)}\right]^{(d/2)-\lambda}}, \tag{II.4}$$

where $K_v(\cdot)$ is the modified Bessel function of the third kind and

$$c = \frac{\left(\sqrt{\chi\psi}\right)^{-\lambda} \psi^\lambda \left(\psi + \zeta'\Sigma^{-1}\zeta\right)^{(d/2)-\lambda}}{(2\pi)^{d/2} |\Sigma^{-1}|^{1/2} K_\lambda\left(\sqrt{\chi\psi}\right)}.$$

The class of GH distributions is very general. It includes skewed t distributions obtained when $\lambda = -\frac{1}{2}v$, $\chi = v$ and $\psi = 0$ (S follows a inverse gamma distribution with parameter $(v/2, v/2)$). When $\zeta = 0$, a skewed t distribution becomes a Student's t distribution with degree of freedom v.

For any $NMVM_d(\xi, \mu, \Sigma, \zeta)$, let $\mu = (\mu_1, ..., \mu_d)'$, $\zeta = (\zeta_1, ..., \zeta_d)'$, and

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \cdots & \rho_{1d}\sigma_1\sigma_d \\ \vdots & \ddots & \vdots \\ \rho_{d1}\sigma_1\sigma_d & \cdots & \sigma_d^2 \end{pmatrix}.$$

It is known that the marginal distributions of $NMVM_d(\xi, \mu, \Sigma, \zeta)$ are $NMVM_1\left(\xi, \mu_k, \sigma_k^2, \zeta_k\right)$, $k = 1, ..., d$. By the Sklar's theorem (1959), there is a unique copula function corresponding to $NMVM_d(\xi, \mu, \Sigma, \zeta)$. This associated copula is called normal mean variance mixture copula (NMVMC) denoted as $C^{NMVM}(\cdot; \xi, \mu, \Sigma, \zeta)$:

$$C^{NMVM}(u; \xi, \mu, \Sigma, \zeta) = NMVM_d(NMVM_1^{-1}(u_1; \eta_1), ..., NMVM_1^{-1}(u_d; \eta_d); \xi, \mu, \Sigma, \zeta), \tag{II.5}$$

where $u = (u_1, ..., u_d)$ and $\eta_k = \left[\xi, \mu_k, \sigma_k^2, \zeta_k\right]$, $k = 1, ..., d$.

Like the class of NMVM distributions, the class of NMVM copulas is also very general including Gaussian, Student's t, and skewed Student's t copulas as special cases.

NMVM-SRMs and NMVMC-SRMs

We first introduce the class of NMVM-SRMs and then extend it to the class of NMVMC-SRMs. For $j = 1, 0$, let $V_{ji} \equiv (U_{ji}, \epsilon_i) \sim NMVM_2\left(\xi, \mu_j, \Sigma_j, \zeta_j\right)$, where

$$\mu_j = \begin{pmatrix} \mu_{Uj} \\ \mu_\epsilon \end{pmatrix}, \ \zeta_j = \begin{pmatrix} \zeta_{Uj} \\ \zeta_\epsilon \end{pmatrix}, \ \Sigma_j = \begin{pmatrix} \sigma_j^2 & \rho_j\sigma_j\sigma_\epsilon \\ \rho_j\sigma_j\sigma_\epsilon & \sigma_\epsilon^2 \end{pmatrix}.$$

It follows from Definition 2.1 that V_{ji} has the same distribution as $\left[\mu_j + S\zeta_j + \sqrt{S}A_jZ\right]$, where $A_jA_j' = \Sigma_j$. Since $\left[\mu_j + (aS)\left(\zeta_j/a\right) + \sqrt{aS}\left(A_j/\sqrt{a}\right)Z\right]$ has the same distribution as $\left[\mu_j + S\zeta_j + \sqrt{S}A_j Z\right]$ for any $a > 0$, the parameters $\xi, \mu_j, \Sigma_j, \zeta_j$ are not separately identifiable without normalization. Following the convention in the literature on SRMs, we normalize the variance of the selection error ϵ_i to be 1, so the parameters $\xi, \mu_j, \Sigma_j, \zeta_j$, $j = 1, 0$, are restricted such that $E(V_{ji}) = 0$ and $Var(\epsilon_i) = 1$. To ensure $E(V_{ji}) = 0$, we let

$$\mu_j = -\zeta_j E(S). \tag{II.6}$$

Note that in general Σ_j is not the variance-covariance matrix of the error vector V_{ji}, as

$$Var(V_{ji}) = E(S)\Sigma_j + Var(S)\zeta_j\zeta_j'. \tag{II.7}$$

It follows that $Var(\epsilon_i) = E(S)\sigma_\epsilon^2 + \zeta_\epsilon^2 Var(S)$. Restricting $Var(\epsilon_i) = 1$ leads to

$$\sigma_\epsilon^2 = \left[1 - \zeta_\epsilon^2 Var(S)\right]/E(S). \tag{II.8}$$

The class of NMVM-SRMs is characterized by (II.1) with $(U_{ji}, \epsilon_i) \sim NMVM_2\left(\xi, \mu_j, \Sigma_j, \zeta_j\right)$, where $\xi, \mu_j, \Sigma_j, \zeta_j$ satisfy (II.6) and (II.8). This class of models is very general. It includes the class of SRMs with finite normal mixture error distributions and SRMs with GH error distributions. We denote the latter as GH-SRMs. The class of GH-SRMs includes the commonly used Gaussian and Student's t SRMs as special cases. Furthermore, it allows the errors $(U_{ji}, \epsilon_i)'$ to have skewed and kurtotic distributions. For GH-SRMs, we have

$$E(S) = \left(\frac{\chi}{\psi}\right)^{1/2} \frac{K_{\lambda+1}\left(\sqrt{\chi\psi}\right)}{K_\lambda\left(\sqrt{\chi\psi}\right)}, \tag{II.9}$$

and

$$Var(S) = \left(\frac{\chi}{\psi}\right)\left(\frac{K_{\lambda+2}\left(\sqrt{\chi\psi}\right)K_\lambda\left(\sqrt{\chi\psi}\right) - K_{\lambda+1}^2\left(\sqrt{\chi\psi}\right)}{K_\lambda^2\left(\sqrt{\chi\psi}\right)}\right). \tag{II.10}$$

The fact that the marginal distributions of a multivariate NMVM distribution are univariate NMVM distributions implies that all three errors $U_{1i}, U_{0i}, \epsilon_i$ must have univariate NMVM distributions. To allow for arbitrary marginal distributions and more flexibility in the dependence structures between U_{ji} and ϵ_i, we now introduce the class of NMVMC-SRMs. Specifically, for $j = 1, 0$, let

$$(U_{ji}, \epsilon_i) \sim C^{NMVM}(F_j(u; \alpha_j), F(\epsilon; \alpha); \xi_j, \mu_j, \Sigma_j, \zeta_j). \tag{II.11}$$

Then the class of NMVMC-SRMs is given by (II.1) and (II.11). A special class of NMVMC-SRMs is obtained when the outcomes errors follow univariate NMVM distributions, but the selection error follows an arbitrary distribution. This includes the models in Lee (1982, 1983) and Heckman, Tobias, and Vytlacil (2003) with Gaussian and Student's t outcome errors. By allowing the outcome errors to follow general NMVM distributions, we can accommodate both asymmetric and fat tailed outcome errors. An alternative construction of this class of models follows Lee (1982, 1983). Suppose $U_{ji} \sim NMVM_1\left(\xi_j, \mu_{j1}, \sigma_{j1}^2, \zeta_{j1}\right)$ for $j = 1, 0$, where $\mu_{j1} = -\zeta_{j1} E\left(S_j\right)$, in which S_j is the mixing variable with a known distribution characterized by parameter ξ_j. For $j = 1, 0$, let $\Gamma_j = \left[\xi_j, 0, 1, 0\right]$. We assume:

$$\left(U_{ji}, NMVM_1^{-1}\left(F(\epsilon_i; \alpha); \Gamma_j\right)\right) \sim NMVM_2\left(\xi_j, \mu_j, \Sigma_j, \zeta_j\right),\ j = 0, 1,$$

where $\mu_j = \left(\mu_{j1}, 0\right)'$, $\zeta_j = \left(\zeta_{j1}, 0\right)'$, and

$$\Sigma_j = \begin{pmatrix} \sigma_{j1}^2 & \rho_j \sigma_{j1} \\ \rho_j \sigma_{j1} & 1 \end{pmatrix}.$$

Compared with the Gaussian and Student's t copula SRMs in Heckman, Tobias, and Vytlacil (2003), we allow the outcome errors to have different parametric distributions such as Gaussian for U_{1i} and skewed Student's t for U_{0i}.

Two-Step Estimation of NMVMC-SRMs

In general, SRMs with the distribution specification in (II.2) including NMVMC-SRMs can be efficiently estimated by MLE. Let $\theta = (\beta_1', \beta_0', \gamma', \alpha', \alpha_1', \alpha_0', \theta_1', \theta_0')'$. The contribution of the i-th observation to the likelihood function is either

$$\begin{aligned} p(Y_{1i}, D_i = 1) &= f_1(Y_{1i} - X_i'\beta_1, \alpha_1) \int_{-W_i'\gamma,\alpha}^{\infty} f_{\epsilon|1}(\epsilon|Y_{1i} - X_i'\beta_1) d\epsilon \\ &= f_1(Y_{1i} - X_i'\beta_1, \alpha_1) \int_{F(-W_i'\gamma,\alpha)}^{1} c_1(F_1(Y_{1i} - X_i'\beta_1, \alpha_1), u, \theta_1) du, \end{aligned}$$

or

$$\begin{aligned} p(Y_{0i}, D_i = 0) &= f_0(Y_{0i} - X_i'\beta_0, \alpha_0) \int_{-\infty}^{-W_i'\gamma} f_{\epsilon|0}(\epsilon|Y_{0i} - X_i'\beta_0) d\epsilon \\ &= f_0(Y_{0i} - X_i'\beta_0, \alpha_0) \int_{0}^{F(-W_i'\gamma,\alpha)} c_0(F_0(Y_{0i} - X_i'\beta_0, \alpha_0), u, \theta_0) du. \end{aligned}$$

Hence the log-likelihood function is given by

$$\begin{aligned} L(\theta) = \sum_{i=1}^{n} & [D_i \ln f_1(Y_{1i} - X_i\prime\beta_1, \alpha_1) + (1 - D_i) \ln f_0(Y_{0i} - X_i\prime\beta_0, \alpha_0)] \\ + \sum_{i=1}^{n} & [D_i \ln \int_{F(-W_i'\gamma,\alpha)}^{1} c_1(F_1(Y_{1i} - X_i'\beta_1, \alpha_1), u, \theta_1) du \\ +(1 - D_i) & \ln \int_{0}^{F(-W_i'\gamma,\alpha)} c_0(F_0(Y_{0i} - X_i'\beta_0, \alpha_0), u, \theta_0) du]. \end{aligned} \tag{II.12}$$

Noting that in general the log-likelihood function in (II.12) is a complicated non-linear function of the unknown parameter θ, the MLE may be computationally difficult. Heckman (1976) first proposed a two-step estimator for the Gaussian SRM in which parameters in the selection equation are estimated by MLE in the first step and parameters in the potential outcomes equations are estimated by OLS with inverse mill's ratio added as an addition regressor in each regression in the second step. Heckman, Tobias, and Vytlacil (2003) extended Heckman's two-step estimator to SRMs characterized by Gaussian copula with normal outcomes errors or Student's t copula with Student's t outcomes errors when the degree of freedom is known (the selection error in both cases can have an arbitrary marginal distribution). They show that parameters in the potential outcomes equations can be consistently estimated by OLS with an extension of the inverse mill's ratio for Gaussian SRM added as an addition regressor in each regression in the second step.

In this section, we first construct a two-step estimation procedure for the class of NMVM-SRMs and then extend it to the sub-class of NMVMC-SRMs in which the outcomes errors follow NMVM distributions when the distributions of the mixing variables are known and the selection error follows an arbitrary distribution. In sharp contrast to existing two-step procedures, we find that for skewed outcomes errors, an additional correction term is needed in each regression in the second step besides extensions of the inverse mill's ratio. Ignoring it will in general lead to inconsistent estimators.

The first step in the two-step procedures for both NMVM-SRMs and NMVMC-SRMs involves the estimation of (γ, α) in the selection model. Note that

$$\begin{aligned} P\left(D_i = 1|W_i\right) &= P\left(\epsilon_i > -W_i'\gamma|W_i\right) = 1 - F(-W_i'\gamma;\alpha), \\ P\left(D_i = 0|W_i\right) &= P\left(\epsilon_i \leq -W_i'\gamma|W_i\right) = F(-W_i'\gamma;\alpha). \end{aligned}$$

The log-likelihood function for (γ, α) can be written as

$$L\left(\alpha,\gamma\right) = \sum_{i=1}^{n} D_i \ln\left(1 - F\left(-W_i'\gamma;\alpha\right)\right) + \left(1-D_i\right)\ln\left(F\left(-W_i'\gamma;\alpha\right)\right). \tag{II.13}$$

Let $(\widehat{\gamma}, \widehat{\alpha})$ denote the MLE of (γ, α) from maximizing $L\left(\alpha,\gamma\right)$. Let $\theta/\left[\gamma,\alpha\right]$ denote parameters in θ excluding γ and α. The second step involves estimation of $\theta/\left[\gamma,\alpha\right]$.

Two-Step Estimation of NMVM-SRMs

In a NMVM-SRM, $(U_{ji}, \epsilon_i) \sim NMVM_2\left[\xi, \mu_j, \Sigma_j, \zeta_j\right], j = 0, 1,$ where

$$\mu_j = \begin{bmatrix} -E(S)\zeta_{Uj} \\ -E(S)\zeta_\epsilon \end{bmatrix}, \tag{II.14}$$

$$\Sigma_j = \begin{bmatrix} \sigma_j^2 & \rho_j\sigma_j\sigma_\epsilon \\ \rho_j\sigma_j\sigma_\epsilon & \sigma_\epsilon^2 \end{bmatrix}, \tag{II.15}$$

$$\zeta_j = \begin{bmatrix} \zeta_{Uj} \\ \zeta_\epsilon \end{bmatrix}. \tag{II.16}$$

Thus, $F(\epsilon;\alpha) = NMVM_1\left(\epsilon;\xi,\mu_\epsilon,\sigma_\epsilon^2,\zeta_\epsilon\right)$ in which $\alpha = \left(\xi,\mu_\epsilon,\sigma_\epsilon^2,\zeta_\epsilon\right)$. To estimate $\theta/[\gamma,\alpha]$, we note that

$$\begin{aligned} E(Y_{1i}|D_i=1,X_i,W_i) &= X_i'\beta_1 + E\left(U_{1i}|\epsilon_i > -W_i'\gamma, W_i\right) = X_i'\beta_1 + \int\limits_{-W_i'\gamma}^{\infty} E(U_{1i}|\epsilon_i=\epsilon)\,dF(\epsilon;\alpha), \\ E(Y_{0i}|D_i=1,X_i,W_i) &= X_i'\beta_0 + E\left(U_{0i}|\epsilon_i \le -W_i'\gamma, W_i\right) = X_i'\beta_0 + \int\limits_{-\infty}^{-W_i'\gamma} E(U_{0i}|\epsilon_i=\epsilon)\,dF(\epsilon;\alpha). \end{aligned}$$

We will employ the following theorem to find simple expressions for $\int_{-W_i'\gamma}^{\infty} E(U_{1i}|\epsilon_i=\epsilon)\,dF(\epsilon;\alpha)$ and $\int_{-\infty}^{-W_i'\gamma} E(U_{0i}|\epsilon_i=\epsilon)\,dF(\epsilon;\alpha)$ in terms of $\theta/[\gamma,\alpha]$.

Theorem 1 *Suppose* $(U,\epsilon) \sim NMVM_2[\xi,\mu,\Sigma,\zeta]$, *where*

$$\mu = \begin{bmatrix} -E(S)\zeta_U \\ -E(S)\zeta_\epsilon \end{bmatrix},\ \Sigma = \begin{bmatrix} \sigma_U^2 & \rho\sigma_U\sigma_\epsilon \\ \rho\sigma_U\sigma_\epsilon & \sigma_\epsilon^2 \end{bmatrix},\ \zeta = \begin{bmatrix} \zeta_U \\ \zeta_\epsilon \end{bmatrix},$$

in which S is a non-negative random variable with distribution function $F_S(s) \equiv F_S(s;\xi)$. Let $f_\epsilon(\cdot)$ and $F_\epsilon(\cdot)$ denote respectively the density and distribution functions of ϵ. Then we have:

$$\begin{aligned} E(U|\epsilon=-x) &= \zeta_U H(x) + \rho\sigma_U G(x), && \text{(II.17)} \\ E(U|\epsilon\ge -x) &= \zeta_U\lambda_{1a}(x) + \rho\sigma_U\lambda_{1b}(x), && \text{(II.18)} \\ E(U|\epsilon< -x) &= \zeta_U\lambda_{0a}(x) + \rho\sigma_U\lambda_{0b}(x), && \text{(II.19)} \end{aligned}$$

where the function $\lambda_{ja}(x)$, $\lambda_{jb}(x)$, $j=1,0$, $H(x)$, *and* $G(x)$ *are defined as*

$$\begin{aligned}
\lambda_{1a}(x) &= \frac{\lambda_a(x)}{1-F_\epsilon(-x)},\ \ \lambda_{1b}(x) = \frac{\lambda_b(x)}{1-F_\epsilon(-x)}, \\
\lambda_{0a}(x) &= -\frac{\lambda_a(x)}{F_\epsilon(-x)},\ \ \lambda_{0b}(x) = -\frac{\lambda_b(x)}{F_\epsilon(-x)}, \\
\lambda_a(x) &= E_S\left((S-E(S))\,\Phi\left(\frac{x+\zeta_\epsilon(S-E(S))}{\sqrt{S}\sigma_\epsilon}\right)\right), \\
\lambda_b(x) &= E_S\left(\sqrt{S}\phi\left(\frac{x+\zeta_\epsilon(S-E(S))}{\sqrt{S}\sigma_\epsilon}\right)\right), \\
H(x) &= \int (s-E(S))\,dF_{S|\epsilon=x}(s), \\
G(x) &= \frac{1}{\sigma_\epsilon}\int (x-(s-E(S))\,\xi_\epsilon)\,dF_{S|\epsilon=x}(s),
\end{aligned}$$

in which

$$F_{S|\epsilon=x}(s) = \int_0^s \frac{1}{f_\epsilon(x)\sqrt{2\pi t\sigma_\epsilon^2}} \exp\left\{-\frac{(x-((t-E(S))\,\zeta_\epsilon))^2}{2t\sigma_\epsilon^2}\right\} dF_S(t).$$

From Theorem 3.1, we obtain:

$$\begin{aligned}
\int_{-W_i'\gamma}^{\infty} E(U_{1i}|\epsilon_i=\epsilon)\,dF(\epsilon;\alpha) &= \zeta_{U1}\lambda_{1a}(W_i'\gamma) + \rho_1\sigma_1\lambda_{1b}(W_i'\gamma), \\
\int_{-\infty}^{-W_i'\gamma} E(U_{0i}|\epsilon_i=\epsilon)\,dF(\epsilon;\alpha) &= \zeta_{U0}\lambda_{0a}(W_i'\gamma) + \rho_0\sigma_0\lambda_{0b}(W_i'\gamma),
\end{aligned}$$

where $\lambda_{ja}(W_i'\gamma)$, $\lambda_{jb}(W_i'\gamma)$, $j=1,0$, are functions of α and γ only. For the treatment group and the control group, we have the following estimating equations:

$$E(Y_{1i}|X_i,W_i,D_i=1) = X_i'\beta_1 + a_1\lambda_{1a}(W_i'\gamma) + b_1\lambda_{1b}(W_i'\gamma), \tag{II.20}$$

$$E(Y_{0i}|X_i,W_i,D_i=0) = X_i'\beta_0 + a_0\lambda_{0a}(W_i'\gamma) + b_0\lambda_{0b}(W_i'\gamma), \tag{II.21}$$

where $a_j=\zeta_{Uj}$, $b_j=\rho_j\sigma_j$, $j=1,0$. We note that in each estimating equation, there are two correction terms: $\lambda_{ja}(W_i'\gamma)$ and $\lambda_{jb}(W_i'\gamma)$, $j=1,0$. The correction term $\lambda_{jb}(W_i'\gamma)$ is the extension of the inverse mill's ration in Heckman's two-step procedure for Gaussian SRM and the correction term $\lambda_{ja}(W_i'\gamma)$ is new and it is present only when $\zeta_{Uj}\neq 0$, i.e., when the outcome error U_{ji} has an asymmetric distribution. In general, ignoring the presence of skewness in outcome errors would lead to inconsistent estimators of parameters in the potential outcome equations. With both correction terms added as additional regressors in (II.20) and (II.21), simple ordinary least squares (OLS) method can be applied to estimating β_1, a_1, b_1, β_0, a_0, b_0 in the second step.

Extension of Heckman's two-step procedure to NMVM-SRMs:

Step 1. Estimate (α, γ) by $(\widehat{\alpha}, \widehat{\gamma})$. Then estimate $\lambda_{ja}\left(W_i'\gamma\right), \lambda_{jb}\left(W_i'\gamma\right), j = 1, 0$ by replacing (α, γ) with $(\widehat{\alpha}, \widehat{\gamma})$ in their expressions and let $\widehat{\lambda}_{ja}\left(W_i'\widehat{\gamma}\right), \widehat{\lambda}_{jb}\left(W_i'\widehat{\gamma}\right), j = 1, 0$ denote the resulting estimators;

Step 2. Estimate $\left(\beta_j, a_j, b_j\right)$ by OLS regression of Y_{ji} on $X_i, \widehat{\lambda}_{ja}\left(W_i'\widehat{\gamma}\right), \widehat{\lambda}_{jb}\left(W_i'\widehat{\gamma}\right)$, $j = 1, 0$.

Example 3.1 Suppose $(U_{ji}, \epsilon_i) \sim GH_2\left(\lambda, \chi, \psi, \mu_j, \Sigma_j, \zeta_j\right)$, $j = 0, 1$, with μ_j, Σ_j, ζ_j defined in (II.14), (II.15), and (II.16). In the Appendix, we show that the correction terms are:

$$\begin{aligned}
\lambda_{1a}\left(w'\gamma\right) &= \frac{\left(GH_1\left(-w'\gamma; \eta_0\right) - GH_1\left(-w'\gamma; \eta_1\right)\right)}{1 - GH_1\left(-w'\gamma; \eta_0\right)} E\left(S\right), \\
\lambda_{1b}\left(w'\gamma\right) &= \frac{gh_1\left(-w'\gamma; \eta_1\right)}{1 - GH_1\left(-w'\gamma; \eta_0\right)} \left(E\left(S\right)\sigma_\epsilon\right), \\
\lambda_{0a}\left(w'\gamma\right) &= -\frac{\left(GH_1\left(-w'\gamma; \eta_0\right) - GH_1\left(-w'\gamma; \eta_1\right)\right)}{GH_1\left(-w'\gamma; \eta_0\right)} E\left(S\right), \\
\lambda_{0b}\left(w'\gamma\right) &= -\frac{gh_1\left(-w'\gamma; \eta_1\right)}{GH_1\left(-w'\gamma; \eta_0\right)} \left(E\left(S\right)\sigma_\epsilon\right),
\end{aligned}$$

where

$$\eta_0 = \left[\lambda, \chi, \psi, -E\left(S\right)\zeta_\epsilon, \sigma_\epsilon, \zeta_\epsilon\right], \ \eta_1 = \left[\lambda + 1, \chi, \psi, -E\left(S\right)\zeta_\epsilon, \sigma_\epsilon, \zeta_\epsilon\right].$$

(i) If $\lambda = -v/2$, $\chi = v$, $\psi \to 0$, and $\zeta_j = 0$, then generalized hyperbolic distributions become Student's t distributions. In this case, the correction terms λ_{1a} and λ_{0a} for skewness disappear in the second step since $\zeta_j = 0$. The correction terms for dependence between the selection error and the outcomes errors are:

$$\begin{aligned}
\lambda_{1b}\left(w'\gamma\right) &= \left(\frac{v+\left(w'\gamma\right)^2}{v-1}\right)\frac{t_{[v]}\left(w'\gamma\right)}{T_{[v]}\left(w'\gamma\right)}, \\
\lambda_{0b}\left(w'\gamma\right) &= -\left(\frac{v+\left(w'\gamma\right)^2}{v-1}\right)\frac{t_{[v]}\left(w'\gamma\right)}{1-T_{[v]}\left(w'\gamma\right)}.
\end{aligned}$$

(ii) Let $v \to +\infty$ in (i). The Student's t distributions become normal distributions and λ_{1b} and λ_{0b} reduce to the well known inverse mills ratio in Heckman's two-step procedure:

$$\lambda_{1b}\left(w'\gamma\right) = \frac{\phi\left(w'\gamma\right)}{\Phi\left(w'\gamma\right)},\ \lambda_{0b}\left(w'\gamma\right) = -\frac{\phi\left(w'\gamma\right)}{1-\Phi\left(w'\gamma\right)}.$$

Example 3.2 Suppose S follows a discrete distribution, then the corresponding NMVM distribution becomes a finite normal mixture distribution. Let S have the following distribution:

$$P\left(S = s_i\right) = p_i,\ i = 1, ...m,\ \sum_{i=1}^{m} p_i = 1.$$

Then we have

$$\begin{aligned}
\lambda_{1a}\left(w'\gamma\right) &= \frac{1}{1-F_\epsilon\left(-w'\gamma\right)}\sum_{i=1}^{m} p_i\left(s_i - E\left(S\right)\right)\Phi\left(\frac{w'\gamma+\zeta_\epsilon\left(s_i-E\left(S\right)\right)}{\sqrt{s_i}\sigma_\epsilon}\right), \\
\lambda_{1b}\left(w'\gamma\right) &= \frac{1}{1-F_\epsilon\left(-w'\gamma\right)}\sum_{i=1}^{m} p_i\sqrt{s_i}\phi\left(\frac{w'\gamma+\zeta_\epsilon\left(s_i-E\left(S\right)\right)}{\sqrt{s_i}\sigma_\epsilon}\right), \\
\lambda_{0a}\left(w'\gamma\right) &= -\frac{1}{F_\epsilon\left(-w'\gamma\right)}\sum_{i=1}^{m} p_i\left(s_i - E\left(S\right)\right)\Phi\left(\frac{w'\gamma+\zeta_\epsilon\left(s_i-E\left(S\right)\right)}{\sqrt{s_i}\sigma_\epsilon}\right), \\
\lambda_{0b}\left(w'\gamma\right) &= -\frac{1}{F_\epsilon\left(-w'\gamma\right)}\sum_{i=1}^{m} p_i\sqrt{s_i}\phi\left(\frac{w'\gamma+\zeta_\epsilon\left(s_i-E\left(S\right)\right)}{\sqrt{s_i}\sigma_\epsilon}\right), \\
F_\epsilon\left(-w'\gamma\right) &= \sum_{i=1}^{m} p_i\Phi\left(\frac{-w'\gamma-\zeta_\epsilon\left(s_i-E\left(S\right)\right)}{\sqrt{s_i}\sigma_\epsilon}\right).
\end{aligned}$$

Two-Step Estimation of NMVMC-SRMs

We now extend the two-step estimation procedure for NMVM-SRMs to a subclass of NMVMC-SRMs in which $U_{ji} \sim NMVM_1\left(\xi_j, \mu_{j1}, \sigma_{j1}^2, \zeta_{j1}\right)$ for $j = 1, 0$, where $\mu_{j1} = -\zeta_{j1} E\left(S_j\right)$ and S_j is the mixing variable with a known distribution characterized by parameter ξ_j. For $j = 1, 0$, let $\Gamma_j = \left[\xi_j, 0, 1, 0\right]$ and

$$\left(U_{ji}, NMVM_1^{-1}\left(F(\epsilon_i; \alpha); \Gamma_j\right)\right) \sim NMVM_2\left(\xi_j, \mu_j, \Sigma_j, \zeta_j\right),\ j = 0, 1,$$

where $\mu_j = \left(\mu_{j1}, 0\right)'$, $\zeta_j = \left(\zeta_{j1}, 0\right)'$,

$$\Sigma_j = \begin{pmatrix} \sigma_{j1}^2 & \rho_j \sigma_{j1} \\ \rho_j \sigma_{j1} & 1 \end{pmatrix}.$$

For notational convenience, let

$$Q_{\Gamma_j}^{-1}(F(-W_i'\gamma; \alpha)) = NMVM_1^{-1}(F(-W_i'\gamma; \alpha); \Gamma_j).$$

Then we have

$$\begin{aligned} E\left(U_{1i} | D_i = 1, W_i\right) &= \zeta_{11} \lambda_{1a}\left(Q_{\Gamma_j}^{-1}(F(-W_i'\gamma; \alpha))\right) + \rho_1 \sigma_{11} \lambda_{1b}\left(Q_{\Gamma_j}^{-1}(F(-W_i'\gamma; \alpha))\right), \\ E\left(U_{0i} | D_i = 0, W_i\right) &= \zeta_{01} \lambda_{0a}\left(Q_{\Gamma_j}^{-1}(F(-W_i'\gamma; \alpha))\right) + \rho_0 \sigma_{01} \lambda_{0b}\left(Q_{\Gamma_j}^{-1}(F(-W_i'\gamma; \alpha))\right), \end{aligned}$$

where

$$
\begin{aligned}
\lambda_{1a}(x) &= \frac{\lambda_a^1(x)}{1-F_\epsilon(-x)},\ \lambda_{1b}(x) = \frac{\lambda_b^1(x)}{1-F_\epsilon(-x)}, \\
\lambda_{0a}(x) &= -\frac{\lambda_a^0(x)}{F_\epsilon(-x)},\ \lambda_{0b}(x) = -\frac{\lambda_b^0(x)}{F_\epsilon(-x)}, \\
\lambda_a^j(x) &= E_{S_j}\left((S_j - E(S_j))\,\Phi\left(\frac{x}{\sqrt{S_j}}\right)\right),\ j=1,0, \\
\lambda_b^j(x) &= E_{S_j}\left(\sqrt{S_j}\phi\left(\frac{x}{\sqrt{S_j}}\right)\right),\ j=1,0.
\end{aligned}
$$

Thus, the two estimating equations are:

$$
\begin{aligned}
E(Y_{1i}|X_i,W_i,D_i=1) &= X_i'\beta_1 + \zeta_{11}\lambda_{1a}\left(Q_{\Gamma_1}^{-1}(F(-W_i'\gamma;\alpha))\right) + \rho_1\sigma_{11}\lambda_{1b}\left(Q_{\Gamma_1}^{-1}(F(-W_i'\gamma;\alpha))\right), \\
E(Y_{0i}|X_i,W_i,D_i=0) &= X_i'\beta_0 + \zeta_{01}\lambda_{0a}\left(Q_{\Gamma_0}^{-1}(F(-W_i'\gamma;\alpha))\right) + \rho_0\sigma_{01}\lambda_{0b}\left(Q_{\Gamma_0}^{-1}(F(-W_i'\gamma;\alpha))\right).
\end{aligned}
$$

Extension of Heckman's two-step procedure to NMVMC-SRMs:

Step 1. Estimate (α,γ) by $(\widehat{\alpha},\widehat{\gamma})$;

Step 2. **(i)** If ξ_j is known, then estimate the above equations by OLS regression of Y_{ji} on $X_i, \lambda_{ja}\left(Q_{\Gamma_j}^{-1}(F(-W_i'\widehat{\gamma};\widehat{\alpha}))\right), \lambda_{jb}\left(Q_{\Gamma_j}^{-1}(F(-W_i'\widehat{\gamma};\widehat{\alpha}))\right), j=1,0$; **(ii)** If ξ_j is unknown, then estimate the above equations by NLS regressions.

For Gaussian and Student's t outcome errors, the above procedure with Step 2.(i) reduces to that in Lee (1982, 1983), and Heckman, Tobias, and Vytlacil (2003). For skewed outcome errors, the extra correction terms $\lambda_{ja}\left(Q_{\Gamma_1}^{-1}(F(-W_i'\gamma;\alpha))\right)$ are essential to the consistent estimation of parameters in the second step.

Asymptotic Properties of the Two-Step Estimator in NMVMC-SRMs

Suppose we have a random sample $\{Y_i,X_i,W_i,D_i\}_{i=1}^n$, where $Y_i = D_iY_{1i} + (1-D_i)\,Y_{0i}$. For notational compactness, we let $Z_i = (Y_i,X_i,W_i,D_i)$. The asymptotic properties of two-step estimators in NMVMC-SRMs can be established by applying general properties of GMM estimators for the stacked moment conditions with the identity matrix as the weighting matrix, see Newey and Mcfadden (1994), Heckman, Tobias and Vytiacil (2003). To illustrate, consider the two-step estimators in NMVM-SRMs. Define the moment functions as follows:

$$\begin{aligned}
g_1\left(Z_i;\tilde{\beta}_1,\gamma,\alpha\right) &= D_i\begin{bmatrix} X_i \\ \lambda_{1a}\left(W_i'\gamma\right) \\ \lambda_{1b}\left(W_i'\gamma\right)\end{bmatrix}\left[Y_i - X_i'\beta_1 - a_1\lambda_{1a}\left(W_i'\gamma\right) - b_1\lambda_{1b}\left(W_i'\gamma\right)\right], \\
g_0\left(Z_i;\tilde{\beta}_0,\gamma,\alpha\right) &= \left(1-D_i\right)\begin{bmatrix} X_i \\ \lambda_{0a}\left(W_i'\gamma\right) \\ \lambda_{0b}\left(W_i'\gamma\right)\end{bmatrix}\left[Y_i - X_i'\beta_0 - a_0\lambda_{0a}\left(W_i'\gamma\right) - b_0\lambda_{0b}\left(W_i'\gamma\right)\right], \\
m\left(W_i;\gamma,\alpha\right) &= \frac{\partial L\left(\gamma,\alpha\right)}{\partial\left(\gamma',\alpha'\right)},
\end{aligned}$$

where $\tilde{\beta}_1 = \left(\beta_1', a_1, b_1\right)'$, $\tilde{\beta}_0 = \left(\beta_0', a_0, b_0\right)'$, and $L\left(\gamma,\alpha\right)$ is the log-likelihood function in (II.13). Further, let

$$g\left(Z_i;\theta\right) = \begin{pmatrix} g_1\left(Z_i;\tilde{\beta}_1,\gamma,\alpha\right) \\ g_0\left(Z_i;\tilde{\beta}_0,\gamma,\alpha\right) \\ m\left(W_i;\gamma,\alpha\right)\end{pmatrix}.$$

Then the true value of the parameter θ satisfies the moment condition: $E\left[g\left(Z_i;\theta\right)\right] = 0$. It follows from the arguments in Newey and McFadden (1994, Theorem 6.1) that under some regularity conditions, the two-step estimator $\widehat{\theta}$ satisfies:

$$\sqrt{n}\left(\hat{\theta}-\theta\right) \to N\left(0,V\right),$$

where $V = G^{-1}\Omega G^{-1\prime}$ in which $\Omega = E\left(g\left(Z_i;\theta\right)g\left(Z_i;\theta\right)'\right)$ and

$$G = E\left(\frac{\partial g\left(Z_i;\theta\right)}{\partial\theta}\right),$$

Noting that $E\left(g_j\left(Z_i;\tilde{\beta}_j,\gamma,\alpha\right)|X_i,W_i,D_i\right) = 0$, $j = 1,0$, and $D_i\left(1-D_i\right) = 0$, we have

$$\Omega = E\left(\left[\begin{array}{ccc} g_1\left(Z_i\right)g_1\left(Z_i\right)' & 0 & 0 \\ 0 & g_0\left(Z_i\right)g_0\left(Z_i\right)' & 0 \\ 0 & 0 & m\left(W_i\right)m\left(W_i\right)' \end{array}\right]\right),$$

where $g_j\left(Z_i\right)$ denotes $g_j\left(Z_i;\tilde{\beta}_1,\gamma,\alpha\right)$, $j = 0,1$, and $m\left(W_i\right)$ denotes $m\left(W_i;\gamma,\alpha\right)$.

Simple Estimators of Four Treatment Parameters

In this section, we provide expressions and estimators of ATE, TT, LATE, and MTE in the class of NMVMC-SRMs. These extend the corresponding results for Gaussian and Student's t copula SRMs in Heckman, Tobias, and Vytlacil (2003). We'll use the same notation as Heckman, Tobias, and Vytlacil (2003).

Expressions for ATE, TT, LATE, and MTE

Let $\Delta_i = Y_{1i} - Y_{0i}$ denote individual i's treatment effect. Then the conditional ATE on $X_i = x$ can be written as:

$$ATE\left(x\right) = E\left(\Delta_i|X_i = x\right) = x'\left(\beta_1 - \beta_0\right) \qquad \text{(II.22)}$$

and the unconditional ATE is given by

$$ATE \equiv E\left(ATE\left(X_i\right)\right) = \left(E\left(X_i\right)\right)'\left(\beta_1 - \beta_0\right).$$

Let $D(W_i) = D_i$ denote individual i's observed decision, i.e., $D(W_i) = I_{\{W_i'\gamma+\epsilon_i>0\}}$. Then TT is the expected treatment effect for those actually received the treatment:.

$$\begin{aligned} TT(x, w, D(w) = 1) &= E(\Delta_i | X_i = x, W_i = w, D(w) = 1) \\ &= x'(\beta_1 - \beta_0) + E\left(U_{1i} - U_{0i} | \epsilon_i \geq -w'\gamma\right). \end{aligned} \tag{II.23}$$

The unconditional TT can be calculated by integrating the above equation over the joint distribution of X and W for those who actually received the treatment:

$$TT = E(\Delta_i | D_i = 1) = E(TT(X_i, W_i, D_i = 1) | D_i = 1).$$

LATE is defined as the expected treatment effect for those induced to receive treatment through a change in the instrument from $W_k = w_k$ to $W_k = \tilde{w}_k$, where W_k is W's kth coordinate assumed to affect the treatment decision, but not to affect the potential outcomes Y_{1i} and Y_{0i}. Let w and $\widetilde{w}$ be identical except for their kth coordinates. LATE can be defined as

$$\begin{aligned} & LATE(D(w) = 0, D(\tilde{w}) = 1, X_i = x) \\ = \; & E(\Delta_i | D(w) = 0, D(\tilde{w}) = 1, X_i = x) \\ = \; & x'(\beta_1 - \beta_0) + E\left(U_{1i} - U_{0i} | -\tilde{w}'\gamma \leq \epsilon_i \leq -w'\gamma\right) \end{aligned} \tag{II.24}$$

and the unconditional LATE can be calculated as follows:

$$E(\Delta_i | D(w) = 0, D(\widetilde{w}) = 1, X_i = x) = E(LATE(D(w) = 0, D(\tilde{w}) = 1, X_i)).$$

The unconditional LATE corresponds to the treatment effect for individuals who would decide not to receive the treatment if their vector W were set to w, but would decide to have the treatment if W were set to $\tilde{w}$.

Finally, MTE is defined as the treatment effect for individuals with a given value of ϵ,

$$
\begin{aligned}
MTE(x,\epsilon) &= E(\Delta_i | X_i = x, \epsilon_i = \epsilon) \\
&= x'(\beta_1 - \beta_0) + E(U_{1i} - U_{0i} | \epsilon_i = \epsilon) \qquad \text{(II.25)}
\end{aligned}
$$

and the unconditional MTE can be calculated as

$$
MTE(\epsilon) = E(MTE(X_i,\epsilon) | \epsilon_i = \epsilon).
$$

It follows from (II.23), (II.24), and (II.25) that we need to provide expressions for $E(U_{1i} - U_{0i} | \epsilon_i = \epsilon)$. Given the general distribution specification for (U_{ji}, ϵ_i) in (II.2), we get:

$$
\begin{aligned}
E(U_{1i} - U_{0i} | \epsilon_i = \epsilon) &= E(U_{1i} | \epsilon_i = \epsilon) - E(U_{0i} | \epsilon_i = \epsilon) \\
&= \int_{-\infty}^{\infty} c_1(F(\epsilon;\alpha), F_1(u;\alpha_1);\theta_1) f_1(u;\alpha_1)\, u du \\
&\quad - \int_{-\infty}^{\infty} c_0(F(\epsilon;\alpha), F_1(u;\alpha_0);\theta_0) f_0(u;\alpha_0)\, u du,
\end{aligned}
$$

where $c_j(u,v;\theta_j)$ is the copula density function of $C_j(u,v;\theta_j)$ and $f_j(u;\alpha_j)$ is the density function of $F_j(u;\alpha_j)$, $j = 1, 0$. In addition, we have:

$$
\begin{aligned}
E\left(U_{1i} - U_{0i} | \epsilon_i \geq -w'\gamma\right) &= \int_{-w'\gamma}^{\infty} E(U_{1i} - U_{0i} | \epsilon_i = \epsilon)\, dF(\epsilon;\alpha), \\
E\left(U_{1i} - U_{0i} | -\tilde{w}'\gamma \leq \epsilon_i \leq -w'\gamma\right) &= \int_{-\tilde{w}'\gamma}^{-w'\gamma} E(U_{1i} - U_{0i} | \epsilon_i = \epsilon)\, dF(\epsilon;\alpha).
\end{aligned}
$$

As a result, the four treatment parameters of interest (ATE, TT, LATE, MTE) can be estimated once the parameters in the SRM are estimated.

Estimators and Their Asymptotic Properties

For NMVMC-SRMs, expressions for $E\left(U_{1i}-U_{0i}|\epsilon_i=\epsilon\right)$ are provided in Section 3. Based on these expressions, we can construct simple plug-in estimators of ATE, TT, LATE, and MTE using the two-step estimation procedure for NMVMC-SRMs proposed in the previous section. To illustrate, let $\hat{\alpha}$, $\hat{\gamma}$, $\hat{\beta}_1$, $\hat{a}_1$, $\hat{b}_1, \hat{\beta}_0$, $\hat{a}_0$, $\hat{b}_0$ denote the two-step estimators of parameters α, γ and β_1, a_1, b_1, β_0, a_0, b_0 in a NMVM-SRM respectively. Then the four treatment parameters of interest can be estimated by

$$\begin{aligned}
\widehat{ATE}\left(x\right) &= x'\left(\hat{\beta}_1-\hat{\beta}_0\right), \\
\widehat{TT}\left(x,w,D\left(w\right)=1\right) &= x'\left(\hat{\beta}_1-\hat{\beta}_0\right)+\left(\hat{a}_1-\hat{a}_0\right)\lambda_{1a}\left(w'\hat{\gamma}\right)+\left(\hat{b}_1-\hat{b}_0\right)\lambda_{1b}\left(w'\hat{\gamma}\right), \\
\widehat{LATE}\left(x,D\left(\tilde{w}\right)=1,D\left(w\right)=0\right) &= x'\left(\hat{\beta}_1-\hat{\beta}_0\right)+\left(\hat{a}_1-\hat{a}_0\right)\left(\frac{\lambda_a\left(\tilde{w}'\hat{\gamma}\right)-\lambda_a\left(w'\hat{\gamma}\right)}{F_\epsilon\left(-w'\hat{\gamma}\right)-F_\epsilon\left(-\tilde{w}'\hat{\gamma}\right)}\right) \\
&\quad +\left(\hat{b}_1-\hat{b}_0\right)\left(\frac{\lambda_b\left(\tilde{w}'\hat{\gamma}\right)-\lambda_b\left(w'\hat{\gamma}\right)}{F_\epsilon\left(-w'\hat{\gamma}\right)-F_\epsilon\left(-\tilde{w}'\hat{\gamma}\right)}\right), \\
\widehat{MTE}\left(x,-w'\gamma\right) &= x'\left(\hat{\beta}_1-\hat{\beta}_0\right)+\left(\hat{a}_1-\hat{a}_0\right)\hat{H}\left(w'\hat{\gamma}\right)+\left(\hat{b}_1-\hat{b}_0\right)\hat{G}\left(w'\hat{\gamma}\right),
\end{aligned}$$

where $\hat{H}\left(\cdot\right)$ and $\hat{G}\left(\cdot\right)$ are estimators of $H\left(\cdot\right)$ and $G\left(\cdot\right)$ in (II.17) from parameters estimated in the first step. The expressions for estimators of TT, LATE, and MTE reveal two effects of skewness in the outcome errors. First, ignoring skewness in outcome errors would in general lead to inconsistent estimators of α, γ and β_1, a_1, b_1, β_0, a_0, b_0; Second, there is an extra term in estimators of TT, LATE, and MTE accounting for skewness in outcome errors. For ATE, there is only the indirect effect through estimation of β_1, β_0.

Asymptotic properties of estimators of ATE, TT, LATE, and MTE can be established by the Delta method. We provide results for NMVM-SRMs using the two-step estimation procedure. Let ι_i be a selection matrix which means $\iota_i'\theta$ will equal to ith component of θ and $\beta_1', a_1, b_1, \beta_0', a_0, b_0, \gamma', \alpha'$ are eight components in θ. Define $V_{ij}=\iota_i'V\iota_j$.

We thus have

$$\sqrt{n}\left(\widehat{ATE}\left(x\right)-ATE\left(x\right)\right)\rightarrow N\left(0,x'\left(V_{11}-2V_{14}+V_{44}\right)x\right).$$

For TT, we get:

$$\sqrt{n}\left(\widehat{TT}\left(x,w,D\left(w\right)=1\right)-TT\left(x,w,D\left(w\right)=1\right)\right)\rightarrow N\left(0,S_1VS_1'\right),$$

where

$$S_1=\frac{\partial TT\left(x,w,D\left(w\right)=1\right)}{\partial\left(\tilde{\beta}_1',\tilde{\beta}_0',\gamma',\alpha'\right)}.$$

Similarly, for LATE, we get:

$$\begin{aligned} & \sqrt{n}\left(\widehat{LATE}\left(x,D\left(\tilde{w}\right)=1,D\left(w\right)=0\right)-LATE\left(x,D\left(\tilde{w}\right)=1,D\left(w\right)=0\right)\right) \\ \rightarrow\ & N\left(0,S_2VS_2'\right), \end{aligned}$$

where

$$S_2=\frac{\partial LATE\left(x,D\left(\tilde{w}\right)=1,D\left(w\right)=0\right)}{\partial\left(\tilde{\beta}_1',\tilde{\beta}_0',\gamma',\alpha'\right)}.$$

Finally, the estimator for MTE conditional on $X_i=x$, $\epsilon_i=\epsilon$ is

$$\widehat{MTE}\left(x,\epsilon\right)=x'\left(\hat{\beta}_1-\hat{\beta}_0\right)+\left(\hat{a}_1-\hat{a}_0\right)\hat{H}\left(-\epsilon\right)+\left(\hat{b}_1-\hat{b}_0\right)\hat{G}\left(-\epsilon\right),$$

satisfying

$$\sqrt{n}\left(\widehat{MTE}\left(x,\epsilon\right)-MTE\left(x,\epsilon\right)\right)\rightarrow N\left(0,S_3VS_3'\right),$$

where

$$S_3=\frac{\partial MTE\left(x,\epsilon\right)}{\partial\left(\tilde{\beta}_1',\tilde{\beta}_0',\gamma',\alpha'\right)}.$$

Monte Carlo Simulation

In this section, we provide simulation results to supplement those reported in Heckman, Tobias, and Vytlacil (2003). Their main finding is that Heckman's two-step procedure for Gaussian SRM and its extensions to Gaussian and Student's t copula SRMs work well in terms of estimating ATE and TT for symmetric error distributions, but do not work well for estimating LATE. In their Footnote 7, they stated that when the distributions are asymmetric, neither Heckman's two-step procedure nor its extensions to Gaussian and Student's t copula SRMs work well.

The purpose of our simulation study is two-fold. First, we provide additional evidence on the significant biases of estimates of ATE and TT when skewness in the distributions is not accounted for; Second, we confirm our theoretical finding that when skewness is accounted for in the two-step estimation procedure, the estimates of ATE and TT are much improved. Throughout the experiment, we used the same basic model as Heckman, Tobias, and Vytlacil (2003) below:

$$Y_{1i} = \beta_1 + U_{1i}, \tag{II.26}$$

$$Y_{0i} = \beta_0 + U_{0i}, \tag{II.27}$$

$$D_i = I_{\{\gamma_0+\gamma_1 W_i+\epsilon>0\}},\ i = 1, ..., n, \tag{II.28}$$

where $\beta_1 = 2$, $\beta_0 = 1$, $\gamma_0 = 0$, $\gamma_1 = 1$, and $W_i \sim N(0,1)$. Different from Heckman, Tobias, and Vytlacil (2003), we assume $(U_{1i}, U_{0i}, \epsilon_i)$ follows a skewed t distribution[3] with

[3] We also used skewed finite mixture distributions and obtained the same qualitative results as reported in this section.

degree of freedom $v = 5$ and skewness parameter $\zeta = (\zeta_{U1}, \zeta_{U0}, \zeta_\epsilon)$. We set $Var(U_{1i}) = Var(U_{0i}) = Var(\epsilon_i) = 1$, $\rho_1 = 0.9$, and $\rho_0 = 0.1$. The sample size is $n = 1500$ and the number of replications is 1000. The skewness parameter for each error term takes the values: -0.3721, 0, or 0.3721, corresponding to left skewed, symmetric or right skewed distributions. For notational convenience, we use code -1, 0, 1 to denote the left skewed, symmetric and right skewed distributions respectively, i.e., a code $(1, -1, 0)$ for ζ corresponds to $\zeta = [0.3721, -0.3721, 0]$. The non-zero values for the skewness parameter were chosen such that $\sigma_k = |\zeta_k|$, $k = U1, U0, \epsilon$, and from (II.8), we have $\sigma_\epsilon = |\zeta_\epsilon| = 0.3721$.

To gauge the separate effects of skewness in the selection error and each outcome error, we generated data from four classes of SRMs categorized according to the value of their skewness parameter. In category 1, only the selection error is skewed ($\zeta_\epsilon \neq 0$), while in category 2, only one of the outcome errors is skewed ($\zeta_{U1} \neq 0$ or $\zeta_{U0} \neq 0$). Category 3 includes cases where both outcome errors are skewed but the selection error is not, and category 4 includes cases that all three errors are skewed. For each generated data, we first use the two-step method to estimate parameters in the SRM under correctly specified model (skewed t model) and misspecified model (Gaussian model) respectively. Then we use the resulting parameter estimates to estimate ATE and TT, where ATE and TT are computed as unconditional values by integrating the conditional estimates for ATE and TT. In Tables 1 and 2, we report respectively the percentage bias of estimates of ATE, TT and bias of estimates of the parameters in the model. For comparison purposes, in Table 1, we also report estimates of ATE by OLS without any correction in the second step. Several interesting findings emerge from Tables 1 and 2. First, simple OLS without any correction in the second step leads to severely biased estimates of ATE. Second, our two-step procedure based on the correctly specified model, skewed t model, performs well in all four categories, for all parameters including model parameters and ATE, TT. Third, the bias of the two-step estimates using Gaussian specification varies with the value of the skewness parameter. We provide a detailed discussion on this below.

In category 1, we find that ignoring skewness in the selection error leads to significant biases in the estimates of parameters in the selection equation, but does not incur serious bias in the estimates of ATE, TT, and parameters in the potential outcomes equations. To see why the bias in the first step estimation of the selection equation does not cause significant bias in the second step estimation of parameters in the potential outcomes equations and of ATE, TT, in Figure 1, we plot the correction terms for Student's t and Gaussian distributions respectively as functions of $w'\gamma$ and the corresponding density functions when ζ_ϵ takes values -0.3721, 0, 0.3721. As we see from the third panel in Figure 1, the density function of skewed t distribution is in general more concentrated around the mean than that of the Gaussian case. Thus using a misspecified Gaussian model to estimate a skewed t model will cause the estimate of the scaling coefficient γ_1 inflated and the estimate of the location coefficient γ_0 having a bias with the same sign as ζ_ϵ. The combined effects make the estimate $w'\hat{\gamma}$ biased upward ($w'\hat{\gamma} \gg 0$) or downward ($w'\hat{\gamma} \ll 0$). However, the correction term of the Gaussian model is biased downward (the correction term for the treated) or upward (the correction term for the control) with respect to the corresponding correction term λ_{bj} ($j = 1, 0$) for skewed t. The correction term λ_{aj} does not have a role in category 1 since $\zeta_{U1} = \zeta_{U0} = 0$. These opposite effects cancel each other out in the second step estimation, leading to negligible biases in the estimates of parameters in the potential outcomes equations and ATE, TT. We also note that when $\zeta = 0$, using the misspecified Gaussian model to estimate β_1, β_0, and ATE, TT in a fat-tailed Student's t SRM does not yield significant biases. This finding is consistent with Heckman, Tobias, and Vytlacil

(2003). However, we observe that the bias in the estimate of the scaling coefficient γ_1 in the selection error is relatively large.

In category 2, we find that ignoring skewness in the outcome errors yields significant biases in the estimates of ATE and TT. The magnitudes of the biases of ATE are similar among all four cases in this category, but the direction of the bias depends on the sign of the skewness parameter in the outcome error. In contrast, the effect of ignoring skewness on the estimate of TT depends on whether there is skewness in the treated or in the untreated group. If skewness only appears in the outcome error for the treated group, the bias is minimal; otherwise the bias is significant. To understand the different effects of ignoring skewness in the outcome error on ATE and TT, we recall that when there is skewness in the outcome errors, there are additional terms, λ_{aj}, $(j = 1, 0)$, in the second step regression. Further, λ_{a1} has the same sign as λ_{b1}, while λ_{a0} has the opposite sign as $\lambda_{b0.}$ Note that individuals who self-select into the treated state have higher values of $w'\gamma$, and individuals who self-select into the untreated state have lower values of $w'\gamma$. When we use a Gaussian model to estimate skewed data, the omitted λ_{aj} will be projected onto the outcome coefficient and the correction term λ_{bj}. For the treated group, λ_{a1} is more concave than λ_{b1}, thus causes downward bias in the estimate of β_1 and upward bias in the coefficient of λ_{b1} if $\zeta_{U1} > 0$. This will cause bias in the estimation of ATE and the impact on the estimation of TT will be minimized. However, for the untreated group, λ_{a0} has the opposite sign as λ_{b0}, thus it causes downward bias in the estimate of β_0 and of the coefficient of λ_{b1}if $\zeta_{U0} > 0$. So the bias remains in the estimates of both ATE and TT if the outcome error in the untreated state is skewed.

In category 3, both outcome errors are skewed. If the skewness parameters have the same sign, then the bias in the estimate of ATE due to model misspecification is minimal, otherwise the bias is significant. The reason for this is that the sign of the skewness parameter determines the direction of bias in the estimate of the coefficient in the corresponding potential outcome equation. When skewness parameters in both outcome equations have the same sign, the estimates of β_1 and β_0 are biased in the direction and thus the bias in the estimate of ATE will be minimized. As discussed for category 2, only the skewness in the untreated group will significantly bias the estimate of TT, so the bias of TT remains significant in all cases in category 3.

In category 4, as we discussed for the cases in category 1 and category 3, ignoring skewness in the selection error will not cause significant biases in the estimates of ATE and TT. Thus the impact of model misspecification on the estimates of ATE and TT remains similar to the cases in category 3.

Conclusion

Chapter II makes two contributions to the literature on specification and estimation of SRMs. First, we propose a general class of SRMs allowing for asymmetric and fat-tailed error distributions and general dependence structure between each outcome error and the selection error. Second, we extend Heckman's two-step estimation procedure for Gaussian SRM to a much wider class of SRMs allowing for asymmetric and fat-tailed error distributions. Our two-step estimation procedure for asymmetric distributions incorporates an additional correction term besides the well-known inverse mill's ration in Heckman's two-step procedure. This additional correction term accounts for skewness in the outcome error distribution. Igoring it would in general lead to inconsistent estimators of model parameters and of average treatment effect parameters. Our simulation results confirm this and also show that our two-step procedure accounting for skewness works well.

CHAPTER III

PARTIAL IDENTIFICATION OF THE DISTRIBUTION OF TREATMENT EFFECTS IN SWITCHING REGIMES MODELS AND ITS CONFIDENCE SETS

Introduction

The class of switching regimes models (SRMs) or generalized sample selection models extends the Roy model of self-selection by allowing a more general decision rule for selecting into different states. The income maximizing Roy model of self-selection was developed to explain occupational choice and its consequences for the distribution of earnings when individuals differ in their endowments of occupation-specific skills. Heckman and Honore (1990) demonstrated that the identification of the joint distribution of potential outcomes is essential to the empirical content of the Roy model.

By allowing a more general decision/selection rule, SRMs enjoy a much wider scope of applications than the Roy model, but in any particular application, they are also limited in their ability to address a wide range of interesting economic/policy questions because of the non-identifiability[1] of the joint distribution of potential outcomes in SRMs. Even in the 'textbook' Gaussian SRM, the correlation coefficient between the potential outcomes or equivalently the joint distribution of the potential outcomes is not identifiable. In a study of a sectoral labor market using the Gaussian SRM, Vijverberg (1993) showed that a number of interesting economic questions including the share of 'productive' workers employed in a sector can not be answered without knowledge of the joint distribution of the two potential outcomes. When used to study treatment effect defined as the difference between the two

[1] In this paper, we use identification and point identification interchangeably. We say a parameter is not identified if it is not point identified.

potential outcomes, important distributional aspects of the treatment effect other than its mean are not identified in SRMs. This partly explains why the current literature has mainly focussed on various measures of average treatment effect including the average treatment effect (ATE), the treatment effect for the treated (TT), the local average treatment effect (LATE), and the marginal treatment effect (MTE). Heckman, Tobias, and Vytlacil (2003) derived expressions for these four average treatment effect parameters for a Gaussian copula SRM and a Student's t copula SRM with normal outcome errors and non-normal selection errors[2]. Heckman and Vytlacil (2005), among other things, showed that in a latent variable framework, ATE, TT, and LATE can be expressed in terms of MTE.

Recently two approaches have been proposed to deal with the non-identifiability problem of the joint distribution of potential outcomes in the 'textbook' Gaussian SRM and some of its extensions. By employing the positive semidefiniteness of the covariance matrix of the outcome errors and the selection error, Vijverberg (1993) showed that in the 'textbook' Gaussian SRM, although unidentified, useful bounds can be placed on the correlation coefficient between the potential outcomes, that is, it is partially identified. Koop and Poirier (1997), Poirier (1998), and Poirier and Tobias (2003) demonstrated via Bayesian approach that these bounds often provide informative information on the unidentified correlation coefficient. Since the joint distribution of the potential outcomes in the 'textbook' Gaussian SRM depends on the unidentified correlation coefficient only (besides the identified marginal parameters), it is possible to place bounds on the joint distribution of the potential outcomes and on the distribution of the difference between the potential outcomes. In the second approach, restrictions are imposed on the dependence structure between the potential outcomes such that their joint distribution and the distribution of treatment effects

[2]They didn't use the concept of copulas, but their models can be interpreted this way.

are identified, see, e.g., Heckman, Smith, and Clements (1997), Biddle, Boden, and Reville (2003), Carneiro, Hansen, and Heckman (2003), Aakvik, Heckman, and Vytlacil (2005), Cunha and Heckman (2007), among others. Among other things, they demonstrated that knowledge of the joint distribution of potential outcomes and the distribution of treatment effects allows a much richer analysis of policy effects than average treatment effects. Questions that can be addressed include the proportion of people participating in the program who benefit from it in terms of having positive treatment effects; the proportion of the total population that benefits from the program; and which groups in an initial position benefit or lose from the program.

Chapter III takes the first approach and makes several contributions to the current literature. First, we extend the partial identification results in Vijverberg (1993), Koop and Poirier (1997), Poirier (1998), Poirier and Tobias (2003), and Li, Poirier, and Tobias (2004) to a general class of SRMs in which the joint distribution of the outcome errors and the selection error is assumed to follow a trivariate Normal Mean-Variance Mixture (NMVM) distribution referred to as NMVM-SRMs. The 'textbook' Gaussian and Student's t SRMs are members of NMVM-SRMs. In addition, NMVM-SRMs also allow the errors to follow asymmetric distributions. For NMVM-SRMs, we provide sharp bounds or partial identification results on the correlation coefficient of the potential outcomes, their joint distribution, and the distribution of treatment effects. The distribution bounds established in NMVM-SRMs rely on two special properties of NMVM-SRMs: (i) the only unidentified parameter in a NMVM-SRM is the correlation coefficient between the two potential outcomes and (ii) the joint distribution of the two potential outcomes in a NMVM-SRM is also NMVM. The fact that the joint distribution of the potential outcomes in a SRM is not identifiable raises two issues: (i) is it possible to test the NMVM specification of the joint distribution of the potential outcomes? and (ii) are the results for NMVM-SRMs robust to the implied joint distribution of the potential outcomes? To address these issues, we establish sharp bounds on the joint distribution of potential outcomes and the distribution of treatment effects in the general class of semiparametric SRMs in Heckman (1990) in which the joint distribution of the trio of errors is completely unspecified. Our results rely on and supplement the point identification results in Heckman (1990). This is the second contribution of Chapter III.

Two interesting conclusions emerge from our results. First, in NMVM-SRMs with symmetric outcome errors, we find that the sharp bounds on the joint distribution of potential outcomes are robust to the implied joint distribution of the potential outcomes in the sense that these bounds remain valid for any distribution of the trio of errors as long as the implied bivariate distributions for each outcome error and the selection error are NMVM with symmetric outcome errors. In contrast, the sharp bounds on the treatment effect distribution in NMVM-SRMs are not robust to the implied joint distribution of the potential outcomes. We provide a detailed numerical comparison between sharp bounds on the treatment effect distribution relying on the trivariate Gaussian and Student's t distributions with those that do not specify the joint distribution of potential outcomes. Our numerical results show that bounds relying on the trivariate Gaussian or Student's t assumption can be misleading. Second, we find that in general information on individual's selection decision can help improve sharp bounds on the joint distribution of the potential outcomes and the distribution of treatment effects. When the conditional distribution of one of the potential outcome errors given the selection error is degenerate at a finite value, our sharp bounds point identify the joint distribution of potential outcomes and the distribution of treatment effects.

The partial identification results established in Chapter III can be used to develop inference procedures for the joint distribution of potential outcomes and the distribution of treatment effects. There is a recent, but rapidly growing literature on inference for partially identified parameters, including Imbens and Manski (2004), Bugni (2007), Canay (2007), Chernozhukov, Hong, and Tamer (2007), Fan and Park (2007), Romano and Shaikh (2006), Stoye (2007), Andrews and Guggenberger (2007), and Andrews and Soares (2007), among others. We refer the reader to Fan and Park (2007) for more references. A complete treatment of this important issue is beyond the scope of Chapter III. However, we demonstrate this feasibility by constructing an asymptotically uniformly valid and non-conservative confidence set (CS) for the distribution of treatment effects in a semiparametric SRM.

Methodologically, the approach we use to bound the joint distribution of the potential outcomes and the distribution of treatment effects in semiparametric SRMs differs from that in NMVM-SRMs. Without specifying the joint distribution of the outcome errors and the selection error, the approach used to bound the distribution of the potential outcomes and the distribution of treatment effects in NMVM-SRMs breaks down. The new tool that we employ in Chapter III to establish bounds on the joint distribution of potential outcomes is the Fréchet-Hoeffding inequality on copulas. A straightforward application of this inequality allows us to bound the joint distribution of potential outcomes using the bivariate distributions of each outcome error and the selection error, where the latter distributions are known to be identified under general conditions, see Heckman (1990). To bound the distribution of treatment effects, we make use of existing results on sharp bounds on the distributions of functions of two random variables including the four simple arithmetic operations, see Williamson and Downs (1990). For a sum of two random variables, Makarov (1981), Rüschendorf (1982), and Frank, Nelsen, and Schweizer (1987) establish sharp bounds on its distribution, see also Nelsen (1999). These results have been used in

Fan and Park (2006, 2008) to bound the distribution of treatment effects and the quantile function of treatment effects in the context of ideal social experiments where selection is random. Other applications of the Fréchet-Hoeffding inequality include Heckman, Smith, and Clements (1997) in which they bound the variance of treatment effects under the assumption of random selection; Manski (1997b) in which he established bounds on the mixture of two potential outcomes when the distribution of each outcome is known; Lee (2002) in which he presented bounds on the correlation coefficient between the potential outcomes in SRMS, and Fan (2005) in which she provided a systematic study on the estimation and inference on the sharp bounds on the correlation bounds.

The rest of Chapter III is organized as follows. In Section 2, we introduce the class of NMVM-SRMs and discuss the identification/partial identification of the parameters in NMVM-SRMs. In particular, we extend existing work on correlation bounds in SRMs with trivariate normal or Student's t errors to our NMVM-SRMs. In Section 3, we establish sharp bounds on the joint distribution of potential outcomes and on the distribution of treatment effects for the whole population and the subpopulation receiving treatment in NMVM-SRMs. In Section 4, we establish sharp bounds on the joint distribution of potential outcomes and on the distribution of treatment effects for the whole population and the subpopulation receiving treatment in semiparametric SRMs in Heckman (1990). In Section 5 we provide a systematic comparison of the two sets of bounds when the two identified bivariate marginal distributions are NMVM. Section 6 presents an asymptotically uniformly valid and non-conservative CS for the distribution of treatment effects in a special class of semiparametric SRMs. The last section concludes. Technical proofs are relegated in the Appendix.

Normal Mean Variance Mixture Switching Regimes Models and Parameter Identification

Parametric SRMs supplement model (II.1) by distributional specifications for the vector of errors $(U_{1i}, U_{0i}, \epsilon_i)'$. Commonly used distributions include the trivariate normal and Student's t distributions. We introduce a general flexible class of parametric SRMs characterized by (II.1) with the trivariate error vector $(U_{1i}, U_{0i}, \epsilon_i)'$ following a NMVM distribution. We discuss identification or partial identification of the parameters in NMVM-SRMs. For notational compactness, we omit the subscript i in the rest of Section 2 and Section 3.

Parameter Identification/Partial Identification

It follows from an extention of Proposition 3.13 in McNeil, Frey, and Embrechts (2005) that

$$\begin{aligned}
(U_1, \epsilon)' &\sim NMVM_2\left(\xi, \mu_{1\epsilon}, \Sigma_{1\epsilon}, \zeta_{1\epsilon}\right), \\
(U_0, \epsilon)' &\sim NMVM_2\left(\xi, \mu_{0\epsilon}, \Sigma_{0\epsilon}, \zeta_{0\epsilon}\right), \\
(U_1, U_0)' &\sim NMVM_2\left(\xi, \mu_{10}, \Sigma_{10}, \zeta_{10}\right),
\end{aligned}$$

where for $i, j = 0, 1, \epsilon$,

$$\mu_{ij} = \begin{pmatrix} \mu_i \\ \mu_j \end{pmatrix}, \; \zeta_{ij} = \begin{pmatrix} \zeta_i \\ \zeta_j \end{pmatrix}, \; \Sigma_{ij} = \begin{pmatrix} \sigma_i^2 & \sigma_i \sigma_j \rho_{ij} \\ \sigma_i \sigma_j \rho_{ij} & \sigma_j^2 \end{pmatrix}.$$

Since either Y_1 or Y_0 is observed for any given individual but never both, the joint distribution of U_1 and U_0 is in general not point identified in NMVM-SRMs. As a result,

ρ_{10} is in general not point identified, as it only appears in the joint distribution of U_1, U_0, whereas the remaining parameters including $\rho_{1\epsilon}$ and $\rho_{0\epsilon}$ are point identified. However, since $\Sigma = Var\left(s^{-1/2}V|S=s\right)$ is positive semi-definite, we obtain: $\rho_{10} \in [\rho_L, \rho_U]$, where

$$\rho_L = \rho_{1\epsilon}\rho_{0\epsilon} - \sqrt{(1-\rho_{1\epsilon}^2)(1-\rho_{0\epsilon}^2)},\ \rho_U = \rho_{1\epsilon}\rho_{0\epsilon} + \sqrt{(1-\rho_{1\epsilon}^2)(1-\rho_{0\epsilon}^2)}.$$

This result was first established in Vijverberg (1993) for Gaussian SRMs by using the fact that the variance covariance matrix of the error vector $(U_1, U_0, \epsilon)'$ is positive semi-definite, see also Koop and Poirier (1997), Poirier (1998), and Poirier and Tobias (2003).

In Gaussian SRMs, Σ is the variance-covariance matrix of $(U_1, U_0, \epsilon)'$ and ρ_{10} is the correlation coefficient between U_1 and U_0. In general NMVM-SRMs, ρ_{10} is not the correlation coefficient between U_1 and U_0. Let $\overline{\rho}_{10}$ denote the correlation coefficient between U_1 and U_0. It is related to ρ_{10} through the following expression:

$$\overline{\rho}_{10} = \frac{E\left(S\right)\rho_{10}\sigma_1\sigma_0 + \zeta_1\zeta_0 Var\left(S\right)}{\sqrt{\left(E\left(S\right)\sigma_1^2 + \zeta_1^2 Var\left(S\right)\right)\left(E\left(S\right)\sigma_0^2 + \zeta_0^2 Var\left(S\right)\right)}}. \tag{III.1}$$

The bounds on ρ_{10} and (III.1) yield bounds on $\overline{\rho}_{10}$: $\overline{\rho}_L \leq \overline{\rho}_{10} \leq \overline{\rho}_U$, where

$$\begin{aligned} \bar{\rho}_L &= \frac{E\left(S\right)\rho_L\sigma_1\sigma_0 + \zeta_1\zeta_0 Var\left(S\right)}{\sqrt{\left(E\left(S\right)\sigma_1^2 + \zeta_1^2 Var\left(S\right)\right)\left(E\left(S\right)\sigma_0^2 + \zeta_0^2 Var\left(S\right)\right)}}, \\ \bar{\rho}_U &= \frac{E\left(S\right)\rho_U\sigma_1\sigma_0 + \zeta_1\zeta_0 Var\left(S\right)}{\sqrt{\left(E\left(S\right)\sigma_1^2 + \zeta_1^2 Var\left(S\right)\right)\left(E\left(S\right)\sigma_0^2 + \zeta_0^2 Var\left(S\right)\right)}}. \end{aligned}$$

Since $\rho_{1\epsilon}, \rho_{0\epsilon}$ are point identified, the bounds ρ_L, ρ_U are point identified and thus $\bar{\rho}_L$, $\bar{\rho}_U$ are point identified. If $[\overline{\rho}_L, \overline{\rho}_U] \neq [-1, 1]$, then we say $\overline{\rho}_{10}$ is partially identified with identified interval (set) given by $[\overline{\rho}_L, \overline{\rho}_U]$. We point out that the bounds $\overline{\rho}_L$, $\overline{\rho}_U$ are sharp, as ρ_L and ρ_U are sharp for ρ_{10}. The following theorem shows that the identified interval for $\overline{\rho}_{10}$ may identify its sign and may even point identify $\overline{\rho}_{10}$.

Theorem 2 *Let* $\overline{\rho}_{1\epsilon} = Corr\left(U_1, \epsilon\right)$ *and* $\overline{\rho}_{0\epsilon} = Corr\left(U_0, \epsilon\right)$*. (i) Suppose* $Var\left(S\right) > 0$*. Then* $[\overline{\rho}_L, \overline{\rho}_U] = [-1, 1]$ *if and only if* $\overline{\rho}_{1\epsilon} = \overline{\rho}_{0\epsilon} = 0$ *and* $\zeta_1 = \zeta_0 = 0$*. Suppose* $Var\left(S\right) = 0$*. Then*

$[\overline{\rho}_L, \overline{\rho}_U] = [-1, 1]$ *if and only if* $\overline{\rho}_{1\epsilon} = \overline{\rho}_{0\epsilon} = 0$; *(ii) If* $\overline{\rho}_{1\epsilon}^2 + \overline{\rho}_{0\epsilon}^2 > 1$ *and* $\overline{\rho}_{1\epsilon}$, $\overline{\rho}_{0\epsilon}$ *have the same sign, then* $\overline{\rho}_L > 0$; *(iii) If* $\overline{\rho}_{1\epsilon}^2 + \overline{\rho}_{0\epsilon}^2 > 1$ *and* $\overline{\rho}_{1\epsilon}$, $\overline{\rho}_{0\epsilon}$ *have the opposite sign, then* $\overline{\rho}_U < 0$; *(iv) If* $\overline{\rho}_{1\epsilon}^2 = 1$ *or* $\overline{\rho}_{0\epsilon}^2 = 1$, *then* $\bar{\rho}_L = \bar{\rho}_U$ *implying that* $\overline{\rho}_{10}$ *is point identified and*

$$\overline{\rho}_{10} = \frac{E(S)\,\rho_{1\epsilon}\rho_{0\epsilon}\sigma_1\sigma_0 + \zeta_1\zeta_0 Var(S)}{\sqrt{\left(E(S)\,\sigma_1^2 + \zeta_1^2 Var(S)\right)\left(E(S)\,\sigma_0^2 + \zeta_0^2 Var(S)\right)}}.$$

Theorem 2.1 (i) implies that when $(U_1, U_0)'$ follows a symmetric NMVM distribution, the bounds $\overline{\rho}_L, \overline{\rho}_U$ are informative or $\overline{\rho}_{10}$ is partially identified as long as at least one of the potential outcomes is correlated with the selection error. In addition, Theorem 2.1 (ii) and (iii) imply that it is possible to identify the sign of $\overline{\rho}_{10}$. The conditions in (ii) and (iii) require that the selection error must be correlated with both outcome errors. If the selection error is perfectly correlated with at least one of the outcomes errors, then Theorem 2.1 (iv) implies that $\overline{\rho}_{10}$ is point identified. The inequality: $\overline{\rho}_L \leq \overline{\rho}_{10} \leq \overline{\rho}_U$ characterizes the class of NMVM-SRMs consistent with the sample information; any NMVM-SRM with $\overline{\rho}_{10}$ violating it is inconsistent with the sample information.

Distribution Bounds in NMVM-SRMs

Let $\Delta = Y_1 - Y_0$ denote the individual treatment effect. Heckman, Tobias, and Vytlacil (2003) derived expressions for four treatment parameters of interest for a Gaussian copula model and a Student's t copula model with normal outcome errors and non-normal selection errors. They are respectively ATE, TT, LATE, and MTE. Let ATE denote the average treatment effect conditional on X, W: $ATE \equiv E(\Delta|X, W) = X'(\beta_1 - \beta_0)$. Note that

$$\Delta = ATE + (U_1 - U_0).$$

The individual treatment effect Δ may differ across individuals with the same observable covariates because of the unobserved heterogeneity $(U_1 - U_0)$. This motivates the study of the distribution of treatment effect Δ conditional on the observed covariates. Throughout the rest of Chapter III, we focus on establishing sharp bounds on distributions conditional on the observed covariates without explicitly mentioning it. We emphasize that sharp bounds on the corresponding unconditional distributions are given by the respective expectations of sharp bounds on the conditional distributions with respect to the observable covariates.

The distribution of Δ depends on ρ_{10} or $\overline{\rho}_{10}$ and hence is not point identified in general. In this section, we establish partial identification results for the joint distribution of potential outcomes and the distribution of Δ for a randomly chosen individual from the whole population and from the population participating in the treatment.

Sharp Bounds on the Joint Distribution of Potential Outcomes

Let F_{10}^{Y} denote the joint distribution of potential outcomes Y_1, Y_0 conditional on $X = x$, $W = w$. Let $\alpha_{10} = (\xi, \mu_{10}, \Sigma_{10}, \zeta_{10})$ and α_{10}^{-} denote all the parameters in α_{10} except ρ_{10}.

In a NMVM-SRM, $F_{10}^{Y}(y_1, y_0) = NMVM_2\left(y_1 - x'\beta_1, y_0 - x'\beta_0; \alpha_{10}\right)$,where

$$NMVM_2\left(u_1, u_0; \alpha_{10}\right) = \int_0^\infty f_S\left(s\right)\Phi_{\rho_{10}}\left(\frac{u_1 - \left(\zeta_1 s - \zeta_1 E\left(S\right)\right)}{\sigma_1\sqrt{s}}, \frac{u_0 - \left(\zeta_0 s - \zeta_0 E\left(S\right)\right)}{\sigma_0\sqrt{s}}\right) ds,$$

in which $\Phi_{\rho}(\cdot, \cdot)$ is the distribution function of a bivariate normal variable with zero means, unit variances, and correlation coefficient ρ. We now show that the bounds on ρ_{10} place bounds on the joint distribution $F_{10}^{Y}(y_1, y_0)$. Let C^{Gau} denote the Gaussian copula given by

$$C^{Gau}(u, v, \rho) = \Phi_{\rho}\left(\Phi^{-1}\left(u\right), \Phi^{-1}\left(v\right)\right),\ (u, v) \in [0, 1]^2.$$

Then we can write

$$
\begin{aligned}
& NMVM_2\left(u_1, u_0; \alpha_{10}\right) \\
= & \int_0^{\infty} f_S(s)\, C^{Gau}\left(\Phi\left(\frac{u_1-\left(\zeta_1 s-\zeta_1 E(S)\right)}{\sigma_1 \sqrt{s}}\right), \Phi\left(\frac{u_0-\left(\zeta_0 s-\zeta_0 E(S)\right)}{\sigma_0 \sqrt{s}}\right), \rho_{10}\right) ds.
\end{aligned}
$$

Since the Gaussian copula is increasing in concordance in ρ_{10} (see Joe (1997)), we obtain the following sharp bounds on the joint distribution of Y_1, Y_0:

$$
\begin{aligned}
& NMVM_2\left(y_1-x'\beta_1, y_0-x'\beta_0; \left[\alpha_{10}^{-}, \rho_L\right]\right) \\
\leq\ & F_{10}^{Y}(y_1, y_0) \\
\leq\ & NMVM_2\left(y_1-x'\beta_1, y_0-x'\beta_0; \left[\alpha_{10}^{-}, \rho_U\right]\right). \qquad \text{(III.2)}
\end{aligned}
$$

For any fixed x and y_1, y_0, the bounds above are informative for $F_{10}^{Y}(y_1, y_0)$ as long as

$$
NMVM_2\left(y_1-x'\beta_1, y_0-x'\beta_0; \left[\alpha_{10}^{-}, \rho_L\right]\right) \neq 0 \text{ or } NMVM_2\left(y_1-x'\beta_1, y_0-x'\beta_0; \left[\alpha_{10}^{-}, \rho_U\right]\right) \neq 1.
$$

Moreover, if either $\overline{\rho}_{1\epsilon}^2=1$ or $\overline{\rho}_{0\epsilon}^2=1$, Theorem 2.1 (iv) implies that $\bar{\rho}_L=\bar{\rho}_U$ or equivalently $\rho_L=\rho_U$ and thus (III.2) point identifies $F_{10}^{Y}(y_1, y_0)$.

Theorem 3 *Let $C_L(s,t)=\max(s+t-1,0)$ denote the Fréchet lower bound copula and $C_U(s,t)=\min(s,t)$ denote the Fréchet upper bound copula. Suppose $\zeta_1=\zeta_0=0$ and $\bar{\rho}_{1\epsilon}=\bar{\rho}_{0\epsilon}=0$. Then*

$$
\begin{aligned}
NMVM_2\left(y_1-x'\beta_1, y_0-x'\beta_0; \left[\alpha_{10}^{-}, \rho_L\right]\right) &= C_L\left(NMVM_1\left(y_1-x'\beta_1; \theta_1\right), NMVM_1\left(y_0-x'\beta_0; \theta_0\right)\right), \\
NMVM_2\left(y_1-x'\beta_1, y_0-x'\beta_0; \left[\alpha_{10}^{-}, \rho_U\right]\right) &= C_U\left(NMVM_1\left(y_1-x'\beta_1; \theta_1\right), NMVM_1\left(y_0-x'\beta_0; \theta_0\right)\right),
\end{aligned}
$$

where $\theta_1=\left(\xi, \mu_1, \sigma_1^2, \zeta_1\right)$ and $\theta_0=\left(\xi, \mu_0, \sigma_0^2, \zeta_0\right)$.

It is interesting to observe from (III.2) and Theorem 3.1 that in NMVM-SRMs, two sources of information contribute to the partial identification of $F_{10}^{Y}(y_1, y_0)$: (i) the partial identification of $\overline{\rho}_{10}$ or ρ_{10}; (ii) the point identification of the marginal distributions of the potential outcomes. When selection is random, i.e., the selection error is independent of the outcome errors, we have

$$
\begin{aligned}
& C_L\left(NMVM_1\left(y_1-x'\beta_1;\theta_1\right), NMVM_1\left(y_0-x'\beta_0;\theta_0\right)\right) \\
\leq\ & F_{10}^Y(y_1,y_0) \\
\leq\ & C_U\left(NMVM_1\left(y_1-x'\beta_1;\theta_1\right), NMVM_1\left(y_0-x'\beta_0;\theta_0\right)\right). \qquad \text{(III.3)}
\end{aligned}
$$

This is a straightforward applicationof the Fréchet-Hoeffding inequality:

$$
C_L(s,t)\leq C(s,t)\leq C_U(s,t),\ \text{for all } (s,t)\in[0,1]^2, \qquad \text{(III.4)}
$$

where $C(\cdot,\cdot)$ is any copula function, as $Y_j \sim NMVM_1\left(\cdot - x'\beta_j;\theta_j\right)$, $j=1,0$. Theorem 3.1 shows that when the outcome errors follow symmetric NMVM distributions and are uncorrelated with the selection error, the bounds in (III.2) are the same as those in (III.3). In general, the bounds in (III.2) are sharper than those in (III.3). Thus, taking into account self-selection tightens the bounds on the joint distribution of potential outcomes.

Similar conclusions hold for the joint distribution of the potential outcomes for participants. To simplify the notation for sharp bounds on $F_{10}^Y(y_1,y_0|D=1)$, we let $F_{10|D=1}^*\left(u_1,u_0;\rho_{10|\epsilon}\right)$ denote the conditional distribution function of (U_1,U_0) given $D=1$ when $\rho_{1\epsilon}^2\neq 1$ and $\rho_{0\epsilon}^2\neq 1$, where

$$
\rho_{10|\epsilon}=\frac{\rho_{10-}\rho_{1\epsilon}\rho_{0\epsilon}}{\sqrt{\left(1-\rho_{1\epsilon}^2\right)\left(1-\rho_{0\epsilon}^2\right)}}.
$$

It is easy to show that

$$
\begin{aligned}
& F_{10|D=1}^*\left(u_1,u_0;\rho_{10|\epsilon}\right) \\
=\ & \frac{\int_{-(w,x_c)'\gamma}^{+\infty}\int \frac{f_S(s)}{\sigma_\epsilon\sqrt{s}}\phi\left(\frac{\epsilon-\mu_\epsilon(s)}{\sigma_\epsilon\sqrt{s}}\right)\Phi_{\rho_{10|\epsilon}}\left(\frac{u_1-\bar{\mu}_1(s)}{\sqrt{(1-\rho_{1\epsilon}^2)\sigma_1^2 s}},\frac{u_0-\bar{\mu}_0(s)}{\sqrt{(1-\rho_{0\epsilon}^2)\sigma_0^2 s}}\right)dsdF_\epsilon\left(\epsilon\right)}{1-F_\epsilon(-(w,x_c)'\gamma)},
\end{aligned}
$$

where

$$\mu_j(s) = \zeta_j(s - E(S)),\ j = 0, 1, \epsilon,$$

$$\bar{\mu}_j(s) = \mu_j(s) + \rho_{j\epsilon}\frac{\sigma_j}{\sigma_\epsilon}\epsilon,\ j = 0, 1.$$

When ρ_{10} reaches ρ_L (ρ_U), $\rho_{10|\epsilon}$ will reach -1 (1), so the sharp bounds on $F_{10}^Y(y_1, y_0|D = 1)$ are given by $F_{10}^L(y_1, y_0|D = 1)$ and $F_{10}^U(y_1, y_0|D = 1)$ respectively, where

$$\begin{aligned} F_{10}^Y(y_1, y_0|D &= 1) = F_{10|D=1}^*\left(y_1 - x'\beta_1, y_0 - x'\beta_0; \rho_{10|\epsilon}\right), \\ F_{10}^L(y_1, y_0|D &= 1) = F_{10|D=1}^*\left(y_1 - x'\beta_1, y_0 - x'\beta_0; -1\right), \\ F_{10}^U(y_1, y_0|D &= 1) = F_{10|D=1}^*\left(y_1 - x'\beta_1, y_0 - x'\beta_0; 1\right). \end{aligned}$$

When $\rho_{1\epsilon}^2 = 1$ or $\rho_{0\epsilon}^2 = 1$, we have $\rho_{10} = \rho_{1\epsilon}\rho_{0\epsilon}$ and thus $F_{10}^Y(y_1, y_0|D = 1) = F_{10}^L(y_1, y_0|D = 1) = F_{10}^U(y_1, y_0|D = 1)$. For example, when $\rho_{1\epsilon}^2 = 1$ and $\rho_{0\epsilon}^2 \neq 1$, we have

$$\begin{aligned} &F_{10}^Y(y_1, y_0|D = 1) \\ = &\frac{\int_{-(w,x_c)'\gamma}^{+\infty}\int \frac{f_S(s)}{\sigma_\epsilon\sqrt{s}}\phi\left(\frac{\epsilon-\mu_\epsilon(s)}{\sigma_\epsilon\sqrt{s}}\right) C_L\left(I\left(y_1 - x'\beta_1 - \bar{\mu}_1(s) \geq 0\right), \Phi\left(\frac{y_0-x'\beta_0-\bar{\mu}_0(s)}{\sqrt{(1-\rho_{0\epsilon}^2)\sigma_0^2 s}}\right)\right) ds dF_\epsilon(\epsilon)}{1 - F_\epsilon(-(w, x_c)'\gamma)}, \end{aligned}$$

where $I(\bullet)$ is a indicator function. If $\rho_{1\epsilon}^2 = 1$ and $\rho_{0\epsilon}^2 = 1$, then

$$\begin{aligned} &F_{10}^Y(y_1, y_0|D = 1) \\ = &\frac{\int_{-(w,x_c)'\gamma}^{+\infty}\int \frac{f_S(s)}{\sigma_\epsilon\sqrt{s}}\phi\left(\frac{\epsilon-\mu_\epsilon(s)}{\sigma_\epsilon\sqrt{s}}\right) C_L\left(I\left(y_1 - x'\beta_1 - \bar{\mu}_1(s) \geq 0\right), I\left(y_0 - x'\beta_0 - \bar{\mu}_0(s) \geq 0\right)\right) ds dF_\epsilon(\epsilon)}{1 - F_\epsilon(-(w, x_c)'\gamma)}. \end{aligned}$$

Sharp Bounds on the Distribution of Treatment Effects

In NMVM-SRMs, the individual treatment effect when $X = x$ is given by

$$\Delta = x' (\beta_1 - \beta_0) + U_1 - U_0.$$

Define $\gamma_1 = \sigma_1 \rho_{1\epsilon} - \sigma_0 \rho_{0\epsilon}$ and $\gamma_2^2 = \sigma_1^2 + \sigma_0^2 - 2\sigma_1\sigma_0\rho_{10}$. Then γ_2^2 satisfies $\sigma_L^2 \leq \gamma_2^2 \leq \sigma_U^2$, where

$$\sigma_U^2 = \sigma_1^2 + \sigma_0^2 - 2\rho_L\sigma_1\sigma_0, \ \sigma_L^2 = \sigma_1^2 + \sigma_0^2 - 2\rho_U\sigma_1\sigma_0.$$

Since $(U_1, U_0, \epsilon)' \sim NMVM_3(\xi, \mu, \Sigma, \zeta)$ and $(\Delta - x'(\beta_1 - \beta_0), U_0, \epsilon)' = B(U_1, U_0, \epsilon)'$, where

$$B = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

An extention of Proposition 3.13 in McNeil, Frey, and Embrechts (2005) implies that

$$\left(\Delta - x'(\beta_1 - \beta_0), U_0, \epsilon\right)' \sim NMVM_3\left(\xi, B\mu, B\Sigma B', B\zeta\right),$$

where

$$\begin{aligned} B\mu &= \begin{pmatrix} \mu_1 - \mu_0 \\ \mu_0 \\ \mu_\epsilon \end{pmatrix}, \ B\zeta = \begin{pmatrix} \zeta_1 - \zeta_0 \\ \zeta_0 \\ \zeta_\epsilon \end{pmatrix}, \text{ and} \\ B\Sigma B' &= \begin{pmatrix} \gamma_2^2 & \sigma_1\sigma_0\rho_{10} - \sigma_0^2 & \gamma_1\sigma_\epsilon \\ \sigma_1\sigma_0\rho_{10} - \sigma_0^2 & \sigma_0^2 & \sigma_0\sigma_\epsilon\rho_{0\epsilon} \\ \gamma_1\sigma_\epsilon & \sigma_0\sigma_\epsilon\rho_{0\epsilon} & \sigma_\epsilon^2 \end{pmatrix}. \end{aligned}$$

Applying an extention of Proposition 3.13 in McNeil, Frey, and Embrechts (2005) again, we get

$$\begin{aligned}
\left(\Delta - x'\left(\beta_1 - \beta_0\right), \epsilon\right)' &\sim NMVM_2\left(\xi, (B\mu)_{1\epsilon}, \left(B\Sigma B'\right)_{1\epsilon}, (B\zeta)_{1\epsilon}\right), \\
\Delta - x'\left(\beta_1 - \beta_0\right) &\sim NMVM_1\left(\xi, \mu_1 - \mu_0, \gamma_2^2, \zeta_1 - \zeta_0\right), \\
\epsilon &\sim NMVM_1\left(\xi, \mu_\epsilon, \sigma_\epsilon^2, \zeta_\epsilon\right),
\end{aligned} \tag{III.5}$$

where

$$(B\mu)_{1\epsilon} = \begin{pmatrix} \mu_1 - \mu_0 \\ \mu_\epsilon \end{pmatrix}, \ \left(B\Sigma B'\right)_{1\epsilon} = \begin{pmatrix} \gamma_2^2 & \gamma_1\sigma_\epsilon \\ \gamma_1\sigma_\epsilon & \sigma_\epsilon^2 \end{pmatrix}, \ (B\zeta)_{1\epsilon} = \begin{pmatrix} \zeta_1 - \zeta_0 \\ \zeta_\epsilon \end{pmatrix}.$$

Let $F_\Delta(\delta)$ and $F_\Delta(\delta|D=1)$ denote respectively the distribution of Δ conditional on $X = x$, $W = w$ and the distribution of Δ conditional on $X = x$, $W = w$, and $D = 1$. The Theorem below provides sharp bounds on $F_\Delta(\delta)$ and $F_\Delta(\delta|D=1)$.

Theorem 4 *(i) It holds that* $F_\Delta^L(\delta) \le F_\Delta(\delta) \le F_\Delta^U(\delta)$, *where*

$$\begin{aligned}
F_\Delta^U(\delta) &= \max_{\sigma_L^2 \le \gamma_2^2 \le \sigma_U^2} NMVM_1\left(\delta - x'\left(\beta_1 - \beta_0\right); \xi, \mu_1 - \mu_0, \gamma_2^2, \zeta_1 - \zeta_0\right), \\
F_\Delta^L(\delta) &= \min_{\sigma_L^2 \le \gamma_2^2 \le \sigma_U^2} NMVM_1\left(\delta - x'\left(\beta_1 - \beta_0\right); \xi, \mu_1 - \mu_0, \gamma_2^2, \zeta_1 - \zeta_0\right).
\end{aligned}$$

If $\zeta_1 = \zeta_0$, *then*

$$\begin{aligned}
F_\Delta^L(\delta) &= \begin{cases} NMVM_1\left(\delta - x'\left(\beta_1 - \beta_0\right); \xi, \mu_1 - \mu_0, \sigma_U^2, 0\right) & \textit{if } \delta \ge ATE \\ NMVM_1\left(\delta - x'\left(\beta_1 - \beta_0\right); \xi, \mu_1 - \mu_0, \sigma_L^2, 0\right) & \textit{if } \delta < ATE \end{cases}; \\
F_\Delta^U(\delta) &= \begin{cases} NMVM_1\left(\delta - x'\left(\beta_1 - \beta_0\right); \xi, \mu_1 - \mu_0, \sigma_L^2, 0\right) & \textit{if } \delta \ge ATE \\ NMVM_1\left(\delta - x'\left(\beta_1 - \beta_0\right); \xi, \mu_1 - \mu_0, \sigma_U^2, 0\right) & \textit{if } \delta < ATE \end{cases}.
\end{aligned}$$

(ii) It holds that $F_\Delta^L(\delta|D=1) \le F_\Delta(\delta|D=1) \le F_\Delta^U(\delta|D=1)$, *where*

$$\begin{aligned}
&F_\Delta^U(\delta|D=1) \\
&= \max_{\sigma_L^2 \le \gamma_2^2 \le \sigma_U^2} \int_{-\infty}^{\delta} \int_{-w'\gamma}^{\infty} \frac{nmvm_2\left(u - x'\left(\beta_1 - \beta_0\right), \epsilon; \xi, (B\mu)_{1\epsilon}, (B\Sigma B')_{1\epsilon}, (B\zeta)_{1\epsilon}\right)}{1 - NMVM_1\left(-w'\gamma; \xi, \mu_\epsilon, \sigma_\epsilon^2, \zeta_\epsilon\right)} du d\epsilon, \\
&F_\Delta^L(\delta|D=1) \\
&= \min_{\sigma_L^2 \le \gamma_2^2 \le \sigma_U^2} \int_{-\infty}^{\delta} \int_{-w'\gamma}^{\infty} \frac{nmvm_2\left(u - x'\left(\beta_1 - \beta_0\right), \epsilon; \xi, (B\mu)_{1\epsilon}, (B\Sigma B')_{1\epsilon}, (B\zeta)_{1\epsilon}\right)}{1 - NMVM_1\left(-w'\gamma; \xi, \mu_\epsilon, \sigma_\epsilon^2, \zeta_\epsilon\right)} du d\epsilon.
\end{aligned}$$

For any fixed x and fixed δ, the bounds on $F_\Delta(\delta)$ are informative as long as $F_\Delta^L(\delta) \neq 0$ or $F_\Delta^U(\delta) \neq 1$. When either $\overline{\rho}_{1\epsilon}^2 = 1$ or $\overline{\rho}_{0\epsilon}^2 = 1$, we have $\rho_L = \rho_U$ and thus $\sigma_L^2 = \sigma_U^2$. Theorem 3.1 (i) implies that in this case, $F_\Delta(\cdot)$ is point identified. In general, $F_\Delta(\delta)$ is partially identified. However, consider the case that $\zeta_1 = \zeta_0$. In this case, the distribution of Δ is symmetric. Theorem 3.2 (i) implies that when $\delta = ATE$, $F_\Delta^L(\delta) = F_\Delta^U(\delta) = 0.5$. Hence $F_\Delta(ATE) = 0.5$, implying that the value of the distribution of Δ at the ATE is identified and that the median of the distribution of the outcome gain is the same as ATE.

We note that the same two sources of information contributing to the partial identification of the joint distribution also contribute to the partial identification of the distribution of Δ. Suppose $\zeta_1 = \zeta_0 = 0$ and $\bar{\rho}_{1\epsilon} = \bar{\rho}_{0\epsilon} = 0$. Theorem 2.1 (i) implies that $[\overline{\rho}_L, \overline{\rho}_U] = [-1, 1]$. In this case, $\sigma_L^2 = (\sigma_1 - \sigma_0)^2$ and $\sigma_U^2 = (\sigma_1 + \sigma_0)^2$. In general, $(\sigma_1 - \sigma_0)^2 \leq \sigma_L^2 \leq \sigma_U^2 \leq (\sigma_1 + \sigma_0)^2$. Theorem 3.2 (i) implies that taking into account self-selection tightens the bounds on $F_\Delta(\delta)$. Moreover, the following simple algebra demonstrates that the stronger the self-selection is, the tighter the bounds. For any δ, the width of the distribution bounds depends on σ_U and σ_L. Noting that

$$\sigma_U^2 - \sigma_L^2 = 4\sigma_1\sigma_0\sqrt{(1 - \rho_{1\epsilon}^2)(1 - \rho_{0\epsilon}^2)},$$

we conclude that the width of the distribution bounds on $F_\Delta(\delta)$ becomes narrower as the correlation between the selection error and the outcome errors become stronger.

Example 2.1. (Cont.) For SRMs with trivariate Gaussian errors, Theorem 3.2

(i) implies that $F^L_\Delta(\delta) \leq F_\Delta(\delta) \leq F^U_\Delta(\delta)$, where

$$\begin{aligned} F^L_\Delta(\delta) &= \begin{cases} \Phi\left(\frac{\delta - ATE}{\sigma_U}\right) & \text{if } \delta \geq ATE \\ \Phi\left(\frac{\delta - ATE}{\sigma_L}\right) & \text{if } \delta < ATE \end{cases}; \\ F^U_\Delta(\delta) &= \begin{cases} \Phi\left(\frac{\delta - ATE}{\sigma_L}\right) & \text{if } \delta \geq ATE \\ \Phi\left(\frac{\delta - ATE}{\sigma_U}\right) & \text{if } \delta < ATE \end{cases}. \end{aligned} \tag{III.6}$$

Example 2.2. (Cont.) Let $T_{[v]}(\cdot)$ denote the distribution function of the Student's t distribution with v degrees of freedom. For SRMs with trivariate Student's t errors, Theorem 3.2 (i) implies that $F^L_\Delta(\delta) \leq F_\Delta(\delta) \leq F^U_\Delta(\delta)$, where

$$\begin{aligned} F^L_\Delta(\delta) &= \begin{cases} T_{[v]}\left(\frac{\delta - ATE}{\sigma_U}\sqrt{\frac{v}{v-2}}\right) & \text{if } \delta \geq ATE \\ T_{[v]}\left(\frac{\delta - ATE}{\sigma_L}\sqrt{\frac{v}{v-2}}\right) & \text{if } \delta < ATE \end{cases}; \\ F^U_\Delta(\delta) &= \begin{cases} T_{[v]}\left(\frac{\delta - ATE}{\sigma_L}\sqrt{\frac{v}{v-2}}\right) & \text{if } \delta \geq ATE \\ T_{[v]}\left(\frac{\delta - ATE}{\sigma_U}\sqrt{\frac{v}{v-2}}\right) & \text{if } \delta < ATE \end{cases}. \end{aligned}$$

Distribution Bounds in Semiparametric SRMs

The bounds for NMVM-SRMs established in Section 3 depend crucially on the parametric distribution assumption, especially the implied joint NMVM distribution of the potential outcomes. In this section, we dispense with the parametric distribution assumption on $(U_{1i}, U_{0i}, \epsilon_i)'$. In particular, we consider the semiparametric SRM in Heckman

(1990):

$$
\begin{aligned}
Y_{1i} &= g_1(X_{1i}, X_{ci}) + U_{1i}, \\
Y_{0i} &= g_0(X_{0i}, X_{ci}) + U_{0i}, \\
D_i &= I_{\{(W_i, X_{ci})'\gamma + \epsilon_i > 0\}}, \; i = 1, \ldots, n,
\end{aligned} \tag{III.7}
$$

where both $g_1(x_1, x_c)$, $g_0(x_0, x_c)$ and the distribution of $(U_{1i}, U_{0i}, \epsilon_i)'$ are completely unknown. Heckman (1990) provided conditions under which the distributions of $(U_{1i}, \epsilon_i)'$ and $(U_{0i}, \epsilon_i)'$, $g_1(x_1, x_c)$, $g_0(x_0, x_c)$, and γ are point identified from the sample information alone. However, the joint distribution of $(U_{1i}, U_{0i})'$ is not (point) identified.

In this section, we provide sharp bounds on the joint distribution of U_{1i}, U_{0i} or Y_{1i}, Y_{0i} and the distribution of Δ_i conditional on the observed covariates in (III.7). We assume independence of the errors $U_{1i}, U_{0i}, \epsilon_i$ and the regressors $X_{1i}, X_{0i}, X_{ci}, W_i$. Again we emphasize that sharp bounds on the corresponding unconditional distributions are given by the respective expectations of sharp bounds on the conditional distributions with respect to the observable covariates. We note that the covariance approach used in Section 3 is not applicable here, as the distribution of $(U_{1i}, U_{0i}, \epsilon_i)'$ is completely unknown. Instead we make use of the Fréchet-Hoeffding inequality in (III.4) and existing results on bounding the distribution of a difference of two random variables each having a given distribution function. Again, we omit the subscript i in the rest of Section 4 and Section 5.

Sharp Bounds on the Distribution of a Difference of Two Random Variables

Sharp bounds on distributions of functions of random variables Y_1 and Y_0 including the four simple arithmetic operations are presented in Williamson and Downs (1990). For a sum of two random variables, Makarov (1981), Rüschendorf (1982), and Frank, Nelsen, and Schweizer (1987) establish sharp bounds on its distribution, see also Nelsen (1999). Frank, Nelsen, and Schweizer (1987) demonstrate that their proof based on copulas can be extended to more general functions than the sum. In this subsection, we will present the relevant results for the difference between two random variables. Specifically, let $\Delta = Y_1 - Y_0$ and $F_\Delta(\cdot)$ denote the distribution function of Δ. The following lemma presents sharp bounds on $F_\Delta(\cdot)$ when only F_1 and F_0 are known.

Lemma 1 *Let* $F_{\min}(\delta) = \sup_{y_1} \max(F_1(y_1) - F_0(y_1 - \delta), 0)$ *and* $F_{\max}(\delta) = 1 + \inf_{y_1} \min(F_1(y_1) - F_0(y_1 - \delta), 0)$*. Then* $F_{\min}(\delta) \leq F_\Delta(\delta) \leq F_{\max}(\delta)$*.*

Viewed as an inequality among all possible distribution functions, the sharp bounds $F_{\min}(\delta)$ and $F_{\max}(\delta)$ cannot be improved, because it is easy to show that if either F_1 or F_0 is the degenerate distribution at a finite value, then for all δ, we have $F_{\min}(\delta) = F_\Delta(\delta) = F_{\max}(\delta)$. In fact, given any pair of distribution functions F_1 and F_0, the inequality: $F_{\min}(\delta) \leq F_\Delta(\delta) \leq F_{\max}(\delta)$ cannot be improved, that is, the bounds $F_{\min}(\delta)$ and $F_{\max}(\delta)$ for $F_\Delta(\delta)$ are point-wise best-possible, see Frank, Nelsen, and Schweizer (1987) for a proof of this for a sum of random variables and Williamson and Downs (1990) for a general operation on two random variables. Unlike the sharp bounds on the correlation coefficient between Y_1, Y_0 or the joint distribution of Y_1, Y_0 which are reached at the Fréchet-Hoeffding lower and upper bounds for the distribution of Y_1, Y_0 when Y_1 and Y_0 are perfectly negatively dependent or perfectly positive dependent (see Fan (2005)), the sharp bounds on the distribution of Δ are not reached at the Fréchet-Hoeffding lower and upper bounds for the distribution of Y_1, Y_0. Frank, Nelsen, and Schweizer (1987) provided explicit expressions for copulas that reach the bounds on the distribution of Δ.

Explicit expressions for bounds on the distribution of a sum of two random variables are available for the case where the distributions of both random variables belong to the same family which includes the uniform, the normal, the Cauchy, and the exponential families, see Alsina (1981), Frank, Nelsen, and Schweizer (1987), and Denuit, Genest, and Marceau (1999). Below we provide expressions for $F_{\min}(\delta)$ and $F_{\max}(\delta)$ when both Y_1 and Y_0 are normal or Student's t.

Example 4.1. Let $Y_1 \sim N\left(\mu_1, \sigma_1^2\right)$ and $Y_0 \sim N\left(\mu_0, \sigma_0^2\right)$. Fan and Park (2006) provide the following expressions for the bounds $F_{\min}(\delta)$ and $F_{\max}(\delta)$:

(i) If $\sigma_1 = \sigma_0 = \sigma$, then

$$F_{\min}(\delta) = \begin{cases} 0 & \text{if } \delta < \mu_1 - \mu_0, \\ 2\Phi\left(\frac{\delta-(\mu_1-\mu_0)}{2\sigma}\right) - 1 & \text{if } \delta \geq \mu_1 - \mu_0, \end{cases} \tag{III.8}$$

$$F_{\max}(\delta) = \begin{cases} 2\Phi\left(\frac{\delta-(\mu_1-\mu_0)}{2\sigma}\right) & \text{if } \delta < \mu_1 - \mu_0, \\ 1 & \text{if } \delta \geq \mu_1 - \mu_0, \end{cases} \tag{III.9}$$

(ii) If $\sigma_1 \neq \sigma_0$, then

$$\begin{aligned} F_{\min}(\delta) &= \Phi\left(\frac{\sigma_1 s - \sigma_0 t}{\sigma_1^2 - \sigma_0^2}\right) + \Phi\left(\frac{\sigma_1 t - \sigma_0 s}{\sigma_1^2 - \sigma_0^2}\right) - 1, \\ F_{\max}(\delta) &= \Phi\left(\frac{\sigma_1 s + \sigma_0 t}{\sigma_1^2 - \sigma_0^2}\right) - \Phi\left(\frac{\sigma_1 t + \sigma_0 s}{\sigma_1^2 - \sigma_0^2}\right) + 1, \end{aligned}$$

where $s = \delta - (\mu_1 - \mu_0)$ and $t = \left(s^2 + 2\left(\sigma_1^2 - \sigma_0^2\right)\ln\left(\frac{\sigma_1}{\sigma_0}\right)\right)^{\frac{1}{2}}$.

Example 4.2. For $j = 0, 1$, we assume $\frac{Y_j-\mu_j}{\sigma_j}\sqrt{\frac{v_j}{v_j-2}} \sim t_{[v_j]}$, where $v_j > 2$, so that $E\left(Y_j\right) = \mu_j$, $Var(Y_j) = \sigma_j^2$ and $F_j(\delta) = T_{[v_j]}\left(\left(\frac{\delta-\mu_j}{\sigma_j}\right)\sqrt{\frac{v_j}{v_j-2}}\right)$.

By Lemma 4.1, $F_{\min}(\delta) = \max(F_1(x_1^*) - F_0(x_1^* - \delta), 0)$ and $F_{\max}(\delta) = 1 + \min(F_1(x_2^*) - F_0(x_2^* - \delta), 0)$, where x_1^* and x_2^* are the maximizer and minimizer of the function $[F_1(x) - F_0(x - \delta)]$ respectively, i.e., x_1^*, x_2^* satisfy the equation:

$$\frac{1}{\sigma_1}\sqrt{\frac{v_1}{v_1-2}}t_{[v_1]}\left(\left(\frac{x-\mu_1}{\sigma_1}\right)\sqrt{\frac{v_1}{v_1-2}}\right)=\frac{1}{\sigma_0}\sqrt{\frac{v_0}{v_0-2}}t_{[v_0]}\left(\left(\frac{x-\mu_0-\delta}{\sigma_0}\right)\sqrt{\frac{v_0}{v_0-2}}\right).$$

In general, one must solve the above equation and hence evaluate $F_{\min}(\delta)$ and $F_{\max}(\delta)$ numerically. When $v_1 = v_0 \equiv v$ (say), we are able to get closed-form expressions for $F_{\min}(\delta)$ and $F_{\max}(\delta)$ as follows:

(i) If $\sigma_1 = \sigma_0 = \sigma$, then

$$F_{\min}(\delta) = \begin{cases} 0 & \text{if } \delta < \mu_1 - \mu_0, \\ 2T_{[v]}\left(\left(\frac{\delta-(\mu_1-\mu_0)}{2\sigma}\right)\sqrt{\frac{v}{v-2}}\right) - 1 & \text{if } \delta \geq \mu_1 - \mu_0, \end{cases} \tag{III.10}$$

$$F_{\max}(\delta) = \begin{cases} 2T_{[v]}\left(\left(\frac{\delta-(\mu_1-\mu_0)}{2\sigma}\right)\sqrt{\frac{v}{v-2}}\right) & \text{if } \delta \geq \mu_1 - \mu_0, \\ 1 & \text{if } \delta \geq \mu_1 - \mu_0. \end{cases} \tag{III.11}$$

(ii) If $\sigma_1 \neq \sigma_0$, then

$$\begin{aligned} & F_{\min}(\delta) \\ = \; & T_{[v]}\left(\left(\frac{\sigma_1^{2\kappa-1}s - \sigma_0^{\kappa}\sigma_1^{\kappa-1}t}{\sigma_1^{2\kappa}-\sigma_0^{2\kappa}}\right)\sqrt{\frac{v}{v-2}}\right) + T_{[v]}\left(\left(\frac{\sigma_1^{\kappa}\sigma_0^{\kappa-1}t - \sigma_0^{(2\kappa-1)}s}{\sigma_1^{2\kappa}-\sigma_0^{2\kappa}}\right)\sqrt{\frac{v}{v-2}}\right) - 1, \\ & F_{\max}(\delta) \\ = \; & T_{[v]}\left(\left(\frac{\sigma_1^{2\kappa-1}s + \sigma_0^{\kappa}\sigma_1^{\kappa-1}t}{\sigma_1^{2\kappa}-\sigma_0^{2\kappa}}\right)\sqrt{\frac{v}{v-2}}\right) - T_{[v]}\left(\left(\frac{\sigma_1^{\kappa}\sigma_0^{\kappa-1}t + \sigma_0^{(2\kappa-1)}s}{\sigma_1^{2\kappa}-\sigma_0^{2\kappa}}\right)\sqrt{\frac{v}{v-2}}\right) + 1, \end{aligned}$$

where $s = \delta - (\mu_1 - \mu_0)$, $\kappa = \frac{v}{v+1}$, and

$$t = \left(s^2 + \left(\sigma_1^{2\kappa} - \sigma_0^{2\kappa}\right)\left(\sigma_1^{2(1-\kappa)} - \sigma_0^{2(1-\kappa)}\right)(v-2)\right)^{\frac{1}{2}}.$$

It is easy to see that in both cases, the expressions for $F_{\min}(\delta)$ and $F_{\max}(\delta)$ reduce to those in Example 4.1 as $v \to +\infty$. For instance, consider the case where $\sigma_1 \neq \sigma_0$. As $v \to +\infty$, we have $\kappa \to 1$, $\sqrt{\frac{v}{v-2}} \to 1$, and $\left(\sigma_1^{2(1-\kappa)} - \sigma_0^{2(1-\kappa)}\right)(v-2) \to 2\log\left(\frac{\sigma_1}{\sigma_0}\right)$.

Semiparametric SRMs

Let $F_{1\epsilon}(u_1, \epsilon)$ and $F_{0\epsilon}(u_0, \epsilon)$ denote respectively the distribution functions of $(U_1, \epsilon)'$ and $(U_0, \epsilon)'$ in model (III.7). Since $F_{1\epsilon}(u_1, \epsilon)$ and $F_{0\epsilon}(u_0, \epsilon)$ are identified from the sample information, the joint distribution of U_1, U_0, ϵ belongs to the Fréchet class of trivariate distributions for which the (1,3) and (2,3) bivariate margins are given or fixed, denoted as $\mathcal{F}(F_{1\epsilon}, F_{0\epsilon})$. Joe (1997) showed that for any $F_{10\epsilon} \in \mathcal{F}(F_{1\epsilon}, F_{0\epsilon})$, it must satisfy

$$\int_{-\infty}^{\epsilon} C_L \left[F_{1|\epsilon}(u_1), F_{0|\epsilon}(u_0)\right] dF_\epsilon(\epsilon) \leq F_{10\epsilon}(u_1, u_0, \epsilon) \leq \int_{-\infty}^{\epsilon} C_U \left[F_{1|\epsilon}(u_1), F_{0|\epsilon}(u_0)\right] dF_\epsilon(\epsilon), \tag{III.12}$$

where $F_{j|\epsilon}(u_j)$ denote the conditional distribution of U_j given ϵ, $j = 1, 0$ and $F_\epsilon(\epsilon)$ the marginal distribution function of ϵ. Inequality (III.12) follows from the Fréchet-Hoeffding inequality and the expression: $F_{10\epsilon}(u_1, u_0, \epsilon) = \int_{-\infty}^{\epsilon} F_{10|\epsilon}(u_1, u_0) dF_\epsilon(\epsilon)$, where $F_{10|\epsilon}(u_1, u_0)$ is the conditional joint distribution of U_1, U_0 given ϵ.

Theorem 5 *In a semiparametric SRM, the following inequalities hold.*

(i) ATE: The joint distribution of potential outcomes satisfies

$$F_{10}^L(y_1, y_0) \leq F_{10}^Y(y_1, y_0) \leq F_{10}^U(y_1, y_0), \tag{III.13}$$

where

$$F_{10}^L(y_1, y_0) = \int_{-\infty}^{\infty} C_L \left[F_{1|\epsilon}(y_1 - g_1(x_1, x_c)), F_{0|\epsilon}(y_0 - g_0(x_0, x_c))\right] dF_\epsilon(\epsilon), \tag{III.14}$$

$$F_{10}^U(y_1, y_0) = \int_{-\infty}^{\infty} C_U \left[F_{1|\epsilon}(y_1 - g_1(x_1, x_c)), F_{0|\epsilon}(y_0 - g_0(x_0, x_c))\right] dF_\epsilon(\epsilon). \tag{III.15}$$

(ii) TT: The joint distribution of potential outcomes for the treated satisfies

$$F_{10}^L(y_1, y_0|D = 1) \leq F_{10}^Y(y_1, y_0|D = 1) \leq F_{10}^U(y_1, y_0|D = 1),$$

where

$$F_{10}^{L}(y_1, y_0|D=1) = \frac{\int_{-(w,x_c)'\gamma}^{\infty} C_L\left(F_{1|\epsilon}(y_1 - g_1(x_1,x_c)), F_{0|\epsilon}(y_0 - g_0(x_0,x_c))\right) dF_\epsilon(\epsilon)}{1 - F_\epsilon(-(w,x_c)'\gamma)},$$
$$F_{10}^{U}(y_1, y_0|D=1) = \frac{\int_{-(w,x_c)'\gamma}^{\infty} C_U\left(F_{1|\epsilon}(y_1 - g_1(x_1,x_c)), F_{0|\epsilon}(y_0 - g_0(x_0,x_c))\right) dF_\epsilon(\epsilon)}{1 - F_\epsilon(-(w,x_c)'\gamma)}.$$

The result in (i) is presented in Lee (2002). It is an immediate consequence of (III.12) when $\epsilon = \infty$. To prove (ii), we note that

$$\begin{aligned} & F_{10}^{Y}(y_1, y_0|D=1) = P(U_{1i} \le y_1 - g_1(x_1,x_c), U_{0i} \le y_0 - g_0(x_0,x_c)|\epsilon_i > -(w,x_c)'\gamma) \\ = & \frac{P(U_{1i} \le y_1 - g_1(x_1,x_c), U_{0i} \le y_0 - g_0(x_0,x_c), \epsilon_i > -(w,x_c)'\gamma)}{P(\epsilon_i > -(w,x_c)'\gamma)} \\ = & \frac{\int_{-(w,x_c)'\gamma}^{\infty} F_{10|\epsilon}(y_1 - g_1(x_1,x_c), y_0 - g_0(x_0,x_c)) dF_\epsilon(\epsilon)}{1 - F_\epsilon(-(w,x_c)'\gamma)}. \qquad \text{(III.16)} \end{aligned}$$

Now since $F_{10|\epsilon}(y_1 - g_1(x_1,x_c), g_0(x_0,x_c))$ satisfies the Fréchet-Hoeffding inequality, we obtain the inequality in (ii).

The lower or upper bounds in Theorem 4.2 are reached when the two potential outcomes are conditionally (on ϵ) perfectly negatively or positively dependent on each other. One example is $\epsilon = U_1 - U_0$ in which U_1, U_0 are perfectly positively dependent conditional on ϵ and the upper bound is reached. These bounds take into account the self-selection process and are tighter than the bounds obtained under random selection. For instance, if selection is random, i.e., both U_1 and U_0 are independent of ϵ, then the bounds in Theorem 4.2 (i) become

$$F_{10}^{LI}(y_1, y_0) = C_L\left[F_1(y_1 - g_1(x_1,x_c)), F_0(y_0 - g_0(x_0,x_c))\right], \qquad \text{(III.17)}$$

$$F_{10}^{UI}(y_1, y_0) = C_U\left[F_1(y_1 - g_1(x_1,x_c)), F_0(y_0 - g_0(x_0,x_c))\right]. \qquad \text{(III.18)}$$

In general, $F_{10}^{LI}(y_1, y_0) \le F_{10}^{L}(y_1, y_0)$ and $F_{10}^{UI}(y_1, y_0) \ge F_{10}^{U}(y_1, y_0)$ implying that the dependence between the outcome errors and the selection error improves on the bounds on

$F_{10}^{Y}(y_1, y_0)$. When the distribution of either U_1 or U_0 conditional on ϵ is degenerate at a finite value, the lower and upper bounds in Theorem 4.2 (i) coincide and thus point identify $F_{10}(y_1, y_0)$ for any y_1, y_0. For example, if U_0 conditional on ϵ is degenerate at a finite value, then

$$F_{0|\epsilon}(u_0) = \begin{cases} 1 & \text{if } u_0 \geq E\left(U_0|\epsilon\right) \\ 0 & \text{if } u_0 < E\left(U_0|\epsilon\right) \end{cases},$$

and

$$\begin{aligned} & F_{10}(y_1, y_0) \\ = & \int_{-\infty}^{\infty} C_U \left[F_{1|\epsilon}(y_1 - g_1(x_1, x_c)), F_{0|\epsilon}(y_0 - g_0(x_0, x_c))\right] dF_\epsilon(\epsilon) \\ = & \int_{-\infty}^{\infty} C_L \left[F_{1|\epsilon}(y_1 - g_1(x_1, x_c)), F_{0|\epsilon}(y_0 - g_0(x_0, x_c))\right] dF_\epsilon(\epsilon) \\ = & \int_{-\infty}^{\infty} C_L \left(F_{1|\epsilon}(y_1 - g_1(x_1, x_c)), I(y_0 - g_0(x_0, x_c) > E\left(U_0|\epsilon\right))\right) dF_\epsilon(\epsilon). \end{aligned}$$

We now consider sharp bounds on the distribution of $\Delta = Y_1 - Y_0$. Note that

$$ATE \equiv E(\Delta|X = x) = g_1(x_1, x_c) - g_0(x_0, x_c)$$

and $F_\Delta(\delta) = E\left[P(U_1 - U_0 \leq \{\delta - ATE\}|\epsilon)\right]$. Applying Lemma 4.1 to $P(U_1 - U_0 \leq \{\delta - ATE\}|\epsilon)$, we obtain the sharp bounds on the distribution of the treatment effect in Theorem 4.3 (i) below. Other bounds presented in Theorem 4.3 can be obtained in the same way.

Theorem 6 *In a semiparametric SRM, the following inequalities hold.*

(i) ATE: $F_\Delta^L(\delta) \leq F_\Delta(\delta) \leq F_\Delta^U(\delta)$, *where*

$$\begin{aligned} F_\Delta^L(\delta) &= \int_{-\infty}^{+\infty} \left[\sup_u \max\left\{F_{1|\epsilon}(u) - F_{0|\epsilon}(u - \{\delta - ATE\}), 0\right\}\right] dF_\epsilon(\epsilon), \\ F_\Delta^U(\delta) &= \int_{-\infty}^{+\infty} \left[\inf_u \min\left\{1 - F_{0|\epsilon}(u - \{\delta - ATE\}) + F_{1|\epsilon}(u), 1\right\}\right] dF_\epsilon(\epsilon). \end{aligned}$$

(ii) TT: The distribution of Δ *for the treated satisfies*

$$F_\Delta^L(\delta|D = 1) \leq F_\Delta(\delta|D = 1) \leq F_\Delta^U(\delta|D = 1),$$

where

$$F^L_\Delta(\delta|D=1) = \frac{\int_{-(w,x_c)'\gamma}^{\infty}\left[\sup_u \max\left\{F_{1|\epsilon}(u) - F_{0|\epsilon}(u - \{\delta - ATE\}), 0\right\}\right] dF_\epsilon(\epsilon)}{1 - F_\epsilon(-(w,x_c)'\gamma)},$$

$$F^U_\Delta(\delta|D=1) = \frac{\int_{-(w,x_c)'\gamma}^{\infty}\left[\inf_u \min\left\{1 - F_{0|\epsilon}(u - \{\delta - ATE\}) + F_{1|\epsilon}(u), 1\right\}\right] dF_\epsilon(\epsilon)}{1 - F_\epsilon(-(w,x_c)'\gamma)}.$$

In contrast to sharp bounds on the joint distribution of potential outcomes, the sharp bounds on the distribution of treatment effects are not reached at conditional perfect positive or negative dependence. Again, self-selection and the identified marginals of $F^Y_{10}(y_1, y_0)$ contribute to the partial identification of the distribution of Δ. When the distribution of either U_1 or U_0 conditional on ϵ is degenerate at a finite value, the lower and upper bounds in Theorem 4.3 (i) coincide and thus point identify $F_\Delta(\delta)$ for any δ. For example, if U_0 conditional on ϵ is degenerate at a finite value, then

$$\begin{aligned} F_\Delta(\delta) &= \int_{-\infty}^{+\infty}\left[\sup_u \max\left\{F_{1|\epsilon}(u) - F_{0|\epsilon}(u - \{\delta - ATE\}), 0\right\}\right] dF_\epsilon(\epsilon) \\ &= \int_{-\infty}^{+\infty}\left[\inf_u \min\left\{1 - F_{0|\epsilon}(u - \{\delta - ATE\}) + F_{1|\epsilon}(u), 1\right\}\right] dF_\epsilon(\epsilon) \\ &= \int_{-\infty}^{+\infty}\left[F_{1|\epsilon}(\delta - ATE + E\left(U_0|\epsilon\right))\right] dF_\epsilon(\epsilon). \end{aligned}$$

Remark 4.1. When $\epsilon = U_1 - U_0$, the potential outcome errors are perfectly positively dependent conditional on ϵ. Let $F^R_{\Delta|\epsilon}$ and F^R_Δ denote respectively the conditional distribution of Δ on ϵ and the unconditional distribution of Δ in this case. Fan and Park (2006) shows that $F^R_{\Delta|\epsilon}$ second order stochastically dominates any outcome gain distribution conditional on ϵ, $F_{\Delta|\epsilon}$. Taking expectation with respect to ϵ, we conclude that in a semiparametric SRM, F^R_Δ second order stochastically dominates any F_Δ consistent with the sample information.

Remark 4.2. Unlike the average treatment parameters such as ATE and TT, the quantile of Δ is in general not identified. By inverting the distribution bounds in Theorem 4.3, we obtain sharp bounds on the quantile of the treatment effect[3] for the whole population and the subpopulation receiving treatment.

Some Applications of the Distribution Bounds

By using the distribution bounds established in the previous subsection, we can provide informative bounds on many interesting effects other than the average treatment effect. Some illustrative examples are discussed below, see Heckman, Smith, and Clements (1997) and Vijverberg (1993) for more examples.

1. The proportion of people participating in the program who benefit from it,

$$P(Y_1 > Y_0|D = 1) = P(\Delta > 0|D = 1) = 1 - F_\Delta(0|D = 1).$$

2. The proportion of the total population that benefits from the program,

$$P(Y_1 > Y_0|D = 1)P(D = 1) = \{1 - F_\Delta(0|D = 1)\}\, P(D = 1).$$

3. The share of 'productive' workers employed in sector 1,

$$P\left(D = 1|Y_1 > Y_0\right) = \frac{\{1 - F_\Delta(0|D = 1)\}\, P(D = 1)}{1 - F_\Delta(0)}.$$

4. The distribution of the potential outcome Y_1 of an individual with an above average

[3] Recently, $\left[F_1^{-1}(q) - F_0^{-1}(q)\right]$ has been used to study treatment effect heterogeneity and is referred to as the quantile treatment effect (QTE), see e.g., Heckman, Smith, and Clements (1997), Abadie, Angrist, and Imbens (2002), Chen, Hong, and Tarozzi (2004), Chernozhukov and Hansen (2005), Firpo (2007), Imbens and Newey (2005), among others, for more discussion and references on the estimation of QTE. Manski (1997a) referred to QTE as ΔD-parameters and the quantile of the treatment effect distribution as $D\Delta$-parameters. Assuming monotone treatment response, Manski (1997a) provided sharp bounds on the quantile of the treatment effect distribution.

Y_0,

$$P(Y_1 \leq y_1 | U_0 > 0) = \frac{F_1\left(y_1 - g_1(x_1, x_c)\right) - F_{10}\left(y_1 - g_1(x_1, x_c), 0\right)}{1 - F_0(0)}.$$

5. The variance of the treatment effect for participants (Lee, 2002),

$$\sigma^2_{L,D=1} \leq Var(\Delta | D = 1) \leq \sigma^2_{U,D=1},$$

where

$$\begin{aligned}
\sigma^2_{U,D=1} &= Var\left(Y_1 | D = 1\right) + Var\left(Y_0 | D = 1\right) \\
&\quad -2 \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \left(F^L_{10}(y_1, y_0 | D = 1) - F_1\left(y_1 | D = 1\right) F_0\left(y_0 | D = 1\right)\right) dy_1 dy_0, \\
\sigma^2_{L,D=1} &= Var\left(Y_1 | D = 1\right) + Var\left(Y_0 | D = 1\right) \\
&\quad -2 \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \left(F^U_{10}(y_1, y_0 | D = 1) - F_1\left(y_1 | D = 1\right) F_0\left(y_0 | D = 1\right)\right) dy_1 dy_0.
\end{aligned}$$

Similar techniques used in the previous subsection may help establish bounds on other parameters of interest. For example, the distribution of the potential outcome Y_1 of an individual with an above average Y_0 who selects into the program is given by

$$P(Y_1 \leq y_1 | D = 1, U_0 > 0) = \frac{P\left(Y_1 \leq y_1, \epsilon \geq -(w, x_c)'\gamma\right) - \int_{-(w,x_c)'\gamma}^{\infty} F_{10|\epsilon}(y_1 - g_1(x_1, x_c), 0) dF_\epsilon(\epsilon)}{P\left(\epsilon \geq -(w, x_c)'\gamma, U_0 > 0\right)},$$

where the probability in the denominator and the first probability in the numerator are identified from the sample information and the second term in the numerator can be bounded by applying the Fréchet-Hoeffding inequality to $F_{10|\epsilon}(y_1 - g_1(x_1, x_c), 0)$.

A Comparison of the two sets of Bounds

The distribution bounds developed in Section 4 depend on the bivariate distributions of (U_1, ϵ) and (U_0, ϵ) which can be parametric or nonparametric. In this section, we first study these bounds when (U_j, ϵ) , $j = 1, 0$, follow bivariate NMVM distributions and then compare them with those established in Section 3 for NMVM-SRMs. The difference between these two sets of bounds is that the former bounds are valid for any joint distribution of the errors U_1, U_0, ϵ, provided that the bivariate marginal distributions corresponding to $\{U_1, \epsilon\}$ and $\{U_0, \epsilon\}$ are bivariate NMVM distributions, while the bounds in Section 3 depend crucially on the joint NMVM distribution for the trio of errors $\{U_1, U_0, \epsilon\}$.

Bounds on F_{10}^Y in Semiparametric SRMs with Bivariate NMVM Distributions

For $j = 0, 1$, assume $(U_j, \epsilon)' \sim NMVM_2\left(\xi, \mu_{j\epsilon}, \Sigma_{j\epsilon}, \zeta_{j\epsilon}\right)$, where

$$\mu_{j\epsilon} = \begin{pmatrix} \mu_j \\ \mu_\epsilon \end{pmatrix},\ \zeta_{j\epsilon} = \begin{pmatrix} \zeta_j \\ \zeta_\epsilon \end{pmatrix},\ \Sigma_{j\epsilon} = \begin{pmatrix} \sigma_j^2 & \sigma_j \sigma_\epsilon \rho_{j\epsilon} \\ \sigma_j \sigma_\epsilon \rho_{j\epsilon} & \sigma_\epsilon^2 \end{pmatrix}.$$

We show in the Appendix that the following theorem holds.

Theorem 7 *In a semiparametric SRM with bivariate NMVM distributions for $\{U_j, \epsilon\}$ for $j = 0, 1$ with $\zeta_1 = \zeta_0 = 0$, we have:*

$$\begin{aligned} F_{10}^L(y_1, y_0) &= \int_{-\infty}^{\infty} \max\{F_{1|\epsilon}(y_1 - g_1(x_1, x_c)) + F_{0|\epsilon}(y_0 - g_0(x_0, x_c)) - 1, 0\} dF_\epsilon(\epsilon) \\ &= NMVM_2\left(y_1 - g_1(x_1, x_c), y_0 - g_0(x_0, x_c); \left[\alpha_{10}^-, \rho_L\right]\right), \end{aligned} \tag{III.19}$$

$$\begin{aligned} F_{10}^U(y_1, y_0) &= \int_{-\infty}^{\infty} \min\{F_{1|\epsilon}(y_1 - g_1(x_1, x_c)), F_{0|\epsilon}(y_0 - g_0(x_0, x_c))\} dF_\epsilon(\epsilon) \\ &= NMVM_2\left(y_1 - g_1(x_1, x_c), y_0 - g_0(x_0, x_c); \left[\alpha_{10}^-, \rho_U\right]\right). \end{aligned} \tag{III.20}$$

We observe immediately that these bounds are the same as the bounds on the joint distribution of potential outcomes in NMVM-SRMs presented in Section 3. This is interesting, because it implies that the NMVM assumption on the joint distribution of the potential outcomes in NMVM-SRMs with symmetric outcome errors does not improve on the bounds of this joint distribution. Heuristically, this is because the conditional copula for

$\{U_1, U_0\}$ given ϵ implied by the trivariate NMVM distribution assumption in NMVM-SRMs is still a NMVM copula. Since the partial correlation between U_1 and U_0 ranges from -1 to 1 when $\zeta_1 = \zeta_0 = 0$, the conditional copula for $\{U_1, U_0\}$ given ϵ interpolates between the lower bound copula and the upper bound copula, resulting in the same bounds as if the conditional copula for $\{U_1, U_0\}$ is unrestricted at all.

Bounds on F_Δ in Semiparametric SRMs with Bivariate NMVM Distributions

Theorem 4.3 (i) provides bounds on F_Δ for any pair of bivariate distributions for $\{U_j, \epsilon\}$, $j = 0, 1$ including NMVM distributions. In this subsection, we use Examples 4.1 and 4.2 to simplify the expressions for bivariate normal marginals and bivariate Student's t marginals.

Suppose $\{U_j, \epsilon\}$ follows a bivariate normal distribution:

$$\begin{pmatrix} U_j \\ \epsilon \end{pmatrix} \sim N \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_j^2 & \sigma_j \rho_{j\epsilon} \\ \sigma_j \rho_{j\epsilon} & 1 \end{pmatrix} \right].$$

The distribution of U_j given ϵ follows a univariate normal distribution with mean $\sigma_j \rho_{j\epsilon} \epsilon$ and variance $\sigma_j^2(1 - \rho_{j\epsilon}^2)$, $j = 1, 0$. Example 4.1 provides bounds on the distribution of Δ given ϵ, i.e., expressions for

$$\sup_u \max \left\{ F_{1|\epsilon}(u) - F_{0|\epsilon}(u - \{\delta - ATE\}), 0 \right\}$$

and

$$\inf_u \min \left\{ 1 - F_{0|\epsilon}(u - \{\delta - ATE\}) + F_{1|\epsilon}(u), 1 \right\}$$

in Theorem 4.3 (i). Taking their expectations with respect to ϵ leads to the following bounds on $F_\Delta(\delta)$.

Theorem 8 *In a semiparametric SRM with bivariate normal distributions for $\{U_j, \epsilon\}$ for $j = 0, 1$, we have:*

(i) Suppose $\rho_{j\epsilon}^2 \neq 1$ for both $j = 0, 1$. If $\sigma_1\sqrt{1-\rho_{1\epsilon}^2} = \sigma_0\sqrt{1-\rho_{0\epsilon}^2}$, then

$$F_\Delta^L(\delta) = 2\int_A \Phi\left(\frac{\{\delta - ATE\} - (\sigma_1\rho_{1\epsilon} - \sigma_0\rho_{0\epsilon})\,\epsilon}{2\sigma_1\sqrt{1-\rho_{1\epsilon}^2}}\right)\phi(\epsilon)d\epsilon - P(A),$$

$$F_\Delta^U(\delta) = 2\int_{A^C} \Phi\left(\frac{\{\delta - ATE\} - (\sigma_1\rho_{1\epsilon} - \sigma_0\rho_{0\epsilon})\,\epsilon}{2\sigma_1\sqrt{1-\rho_{1\epsilon}^2}}\right)\phi(\epsilon)d\epsilon + P(A),$$

where $A = \{\epsilon : \{\delta - ATE\} \geq (\sigma_1\rho_{1\epsilon} - \sigma_0\rho_{0\epsilon})\,\epsilon\}$ and A^C is the complement of A. When $(\sigma_1\rho_{1\epsilon} - \sigma_0\rho_{0\epsilon}) = 0$, A is the whole real line if $\delta \geq ATE$, else A is an empty set;

(ii) Suppose $\rho_{j\epsilon}^2 \neq 1$ for both $j = 0, 1$. If $\sigma_1\sqrt{1-\rho_{1\epsilon}^2} \neq \sigma_0\sqrt{1-\rho_{0\epsilon}^2}$, then

$$\begin{aligned} F_\Delta^L(\delta) &= \int_{-\infty}^{+\infty} \Phi\left(\frac{\sigma_1\sqrt{(1-\rho_{1\epsilon}^2)}s - \sigma_0\sqrt{(1-\rho_{0\epsilon}^2)}t}{\sigma_1^2(1-\rho_{1\epsilon}^2) - \sigma_0^2(1-\rho_{0\epsilon}^2)}\right)\phi(\epsilon)d\epsilon \\ &\quad + \int_{-\infty}^{+\infty} \Phi\left(\frac{\sigma_1\sqrt{(1-\rho_{1\epsilon}^2)}t - \sigma_0\sqrt{(1-\rho_{0\epsilon}^2)}s}{\sigma_1^2(1-\rho_{1\epsilon}^2) - \sigma_0^2(1-\rho_{0\epsilon}^2)}\right)\phi(\epsilon)d\epsilon - 1, \end{aligned}$$

$$\begin{aligned} F_\Delta^U(\delta) &= \int_{-\infty}^{+\infty} \Phi\left(\frac{\sigma_1\sqrt{(1-\rho_{1\epsilon}^2)}s + \sigma_0\sqrt{(1-\rho_{0\epsilon}^2)}t}{\sigma_1^2(1-\rho_{1\epsilon}^2) - \sigma_0^2(1-\rho_{0\epsilon}^2)}\right)\phi(\epsilon)d\epsilon \\ &\quad - \int_{-\infty}^{+\infty} \Phi\left(\frac{\sigma_1\sqrt{(1-\rho_{1\epsilon}^2)}t + \sigma_0\sqrt{(1-\rho_{0\epsilon}^2)}s}{\sigma_1^2(1-\rho_{1\epsilon}^2) - \sigma_0^2(1-\rho_{0\epsilon}^2)}\right)\phi(\epsilon)d\epsilon + 1, \end{aligned}$$

where $s = \{\delta - ATE\} - (\sigma_1\rho_{1\epsilon} - \sigma_0\rho_{0\epsilon})\,\epsilon$ and

$$t = \left(s^2 + 2\left[\sigma_1^2(1-\rho_{1\epsilon}^2) - \sigma_0^2(1-\rho_{0\epsilon}^2)\right]\ln\left(\frac{\sigma_1\sqrt{1-\rho_{1\epsilon}^2}}{\sigma_0\sqrt{1-\rho_{0\epsilon}^2}}\right)\right)^{\frac{1}{2}}.$$

(iii) Suppose $\rho_{j\epsilon}^2 = 1$ for at least one $j = 0, 1$. Then $F_\Delta(\delta)$ is point identified. For example,

if $\rho_{1\epsilon}^2 = 1$, then

$$F_\Delta(\delta) = F_\Delta^L(\delta) = F_\Delta^U(\delta) = \Phi\left(\frac{\delta - ATE}{\sqrt{\sigma_1^2 + \sigma_0^2 - 2\rho_{0\epsilon}\sigma_1\sigma_0}}\right), \text{ if } \rho_{1\epsilon} = 1, \quad \text{(III.21)}$$

$$F_\Delta(\delta) = F_\Delta^L(\delta) = F_\Delta^U(\delta) = \Phi\left(\frac{\delta - ATE}{\sqrt{\sigma_1^2 + \sigma_0^2 + 2\rho_{0\epsilon}\sigma_1\sigma_0}}\right), \text{ if } \rho_{1\epsilon} = -1 \quad \text{(III.22)}$$

In contrast to the sharp bounds on the joint distribution of the potential outcomes in Theorem 5.1, the bounds given above on the distribution of the outcome gain differ from the corresponding bounds in Gaussian SRMs and are in general wider, because they are valid for any trivariate distribution with bivariate normal marginals for (U_1, ϵ) and (U_0, ϵ), not necessarily the trivariate Normal distribution in Gaussian SRMs. On the one hand, imposing the trivariate normality assumption narrows the width of the bounds, but on the other hand, it may lead to misleading conclusions if the implied normality assumption for the joint distribution of potential outcomes is violated. To see the seriousness of this problem, remember in Gaussian SRMs, the value of the treatment effect distribution at its mean is always identified: $F_\Delta(ATE) = 0.5$. However, if the joint distribution of the potential outcomes is unknown, then $F_\Delta(ATE)$ is in general not identified and the bounds on $F_\Delta(ATE)$ depend on the parameters of the identified bivariate distributions.

In Figure 2, we plotted the two sets of bounds on $F_\Delta(\cdot)$ in Gaussian SRMs and semiparametric SRMs with bivariate normal marginals. We fixed $ATE = 0$, $\sigma_1^2 = 1$ and $\sigma_0^2 = 1$. For $\rho_{1\epsilon} = 0.5$, we chose a range of values for $\rho_{0\epsilon}$. We also plotted the bounds when $\rho_{1\epsilon} = \rho_{0\epsilon} = 0$. Solid curves are bounds in Theorem 5.2 assuming bivariate normality (BN) for (U_j, ϵ) only, while dashed curves are bounds in (6) assuming trivariate normality (TN) for (U_j, U_0, ϵ). Several general conclusions emerge from Figure 2. First, for any given set of parameter values, the bounds under bivariate normal marginals are always wider than the bounds under the trivariate normal assumption; Second, for given δ, the bounds in general become narrower as the dependence between U_0 and ϵ as measured by the magnitude of $\rho_{0\epsilon}$ increases except when $\delta = 0$ in Gaussian SRMs in which case the lower and upper bounds coincide and become 0.5. In the extreme cases where either $\rho_{1\epsilon}^2 = 1$ or $\rho_{0\epsilon}^2 = 1$, the two sets of bounds coincide and both identify the distribution of Δ. Third, the bounds corresponding to $(\rho_{1\epsilon}, \rho_{0\epsilon}) = (0, 0)$ are wider than the bounds when $(\rho_{1\epsilon}, \rho_{0\epsilon}) \neq (0, 0)$, because the former does not account for the information through self-selection.

To see how these bounds change with the variance parameters. In Figure 3, we plotted the bounds on $F_\Delta(\delta)$ against σ_0 at $\delta = 0, 1, 4$ when $\sigma_1^2 = 1$, $\rho_{1\epsilon} = 0.5$ and $\rho_{0\epsilon} = 0.5$. One interesting fact we observe is that the distribution bounds under both trivariate normality and bivariate normality become wider to some point and then narrower as σ_0 goes to ∞.

Suppose $\{U_j, \epsilon\}$ follows a bivariate Student's t distribution:

$$\left\{\sqrt{\frac{v}{v-2}}\frac{U_j}{\sigma_j}, \sqrt{\frac{v}{v-2}}\epsilon\right\} \sim t_{[v]}\left(\bullet, \bullet, \rho_{j\epsilon}\right), \ j = 1, 0.$$

To derive bounds on the distribution of Δ in this case, we make use of the fact that $U_j|\epsilon$ follows the univariate Student's t distribution with degrees of freedom $v+1$, mean $\sigma_j \rho_{j\epsilon}\epsilon$, and variance $\sigma_j^2(1-\rho_{j\epsilon}^2)\left(\frac{(v-2)+\epsilon^2}{v-1}\right)$, $j = 1, 0$. Example 4.2 provides bounds on the distribution of Δ given ϵ, i.e., expressions for

$$\sup_u \max\left\{F_{1|\epsilon}(u) - F_{0|\epsilon}(u - \{\delta - ATE\}), 0\right\}$$

and

$$\inf_u \min\left\{1 - F_{0|\epsilon}(u - \{\delta - ATE\}) + F_{1|\epsilon}(u), 1\right\}$$

in Theorem 4.3 (i). Taking their expectations with respect to ϵ leads to the bounds on $F_\Delta(\delta)$.

Theorem 9 *In a semiparametric SRM with bivariate Student's t distributions for $\{U_j, \epsilon\}$ for $j = 0, 1$, we have:*

(i) Suppose $\rho_{j\epsilon}^2 \neq 1$ for both $j = 0, 1$ and $\sigma_1\sqrt{1-\rho_{1\epsilon}^2} = \sigma_0\sqrt{1-\rho_{0\epsilon}^2} \equiv \sigma$. Let $\overline{\sigma} =$

$\sigma\sqrt{\left(\frac{(v-2)+\epsilon^2}{v-1}\right)}$. *Then*

$$F_\Delta^L(\delta) = 2\int_A T_{[v+1]}\left(\left(\frac{\delta - ATE - (\sigma_1\rho_{1\epsilon} - \sigma_0\rho_{0\epsilon})\,\epsilon}{2\overline{\sigma}}\right)\sqrt{\frac{v+1}{v-1}}\right)t_{[v]}(\epsilon)d\epsilon - P(A),$$

$$F_\Delta^U(\delta) = 2\int_{A^C} T_{[v+1]}\left(\left(\frac{\delta - ATE - (\sigma_1\rho_{1\epsilon} - \sigma_0\rho_{0\epsilon})\,\epsilon}{2\overline{\sigma}}\right)\sqrt{\frac{v+1}{v-1}}\right)t_{[v]}(\epsilon)d\epsilon + P(A),$$

where $A = \{\epsilon : \{\delta - ATE\} \geq (\sigma_1\rho_{1\epsilon} - \sigma_0\rho_{0\epsilon})\,\epsilon\}$ *and* A^C *is the complement of* A.

(ii) Suppose $\rho_{j\epsilon}^2 \neq 1$ *for both* $j = 0, 1$ *and* $\sigma_1\sqrt{1-\rho_{1\epsilon}^2} \neq \sigma_0\sqrt{1-\rho_{0\epsilon}^2}$. *Let* $\overline{\sigma}_1 = \sigma_1\sqrt{\left(1-\rho_{1\epsilon}^2\right)\left(\frac{(v-2)+\epsilon^2}{v-1}\right)}$ *and* $\overline{\sigma}_0 = \sigma_0\sqrt{\left(1-\rho_{0\epsilon}^2\right)\left(\frac{(v-2)+\epsilon^2}{v-1}\right)}$. *Then*

$$\begin{aligned} F_\Delta^L(\delta) &= \int_{-\infty}^{+\infty} T_{[v+1]}\left(\left(\frac{\overline{\sigma}_1^{2\kappa-1}s - \overline{\sigma}_0^{\kappa-1}\overline{\sigma}_1^{\kappa-1}t}{\overline{\sigma}_1^{2\kappa} - \overline{\sigma}_0^{2\kappa}}\right)\sqrt{\frac{v+1}{v-1}}\right)t_{[v]}(\epsilon)d\epsilon \\ &\quad + \int_{-\infty}^{+\infty} T_{[v+1]}\left(\left(\frac{\overline{\sigma}_1^{\kappa}\overline{\sigma}_0^{\kappa-1}t - \overline{\sigma}_0^{2\kappa-1}s}{\overline{\sigma}_1^{2\kappa} - \overline{\sigma}_0^{2\kappa}}\right)\sqrt{\frac{v+1}{v-1}}\right)t_{[v]}(\epsilon)d\epsilon - 1, \end{aligned}$$

$$\begin{aligned} F_\Delta^U(\delta) &= \int_{-\infty}^{+\infty} T_{[v+1]}\left(\left(\frac{\overline{\sigma}_1^{2\kappa-1}s + \overline{\sigma}_0^{\kappa}\overline{\sigma}_1^{\kappa-1}t}{\overline{\sigma}_1^{2\kappa} - \overline{\sigma}_0^{2\kappa}}\right)\sqrt{\frac{v+1}{v-1}}\right)t_{[v]}(\epsilon)d\epsilon \\ &\quad - \int_{-\infty}^{+\infty} T_{[v+1]}\left(\left(\frac{\overline{\sigma}_1^{\kappa}\overline{\sigma}_0^{\kappa-1}t + \overline{\sigma}_0^{2\kappa-1}s}{\overline{\sigma}_1^{2\kappa} - \overline{\sigma}_0^{2\kappa}}\right)\sqrt{\frac{v+1}{v-1}}\right)t_{[v]}(\epsilon)d\epsilon + 1, \end{aligned}$$

where

$$s = \{(\delta - ATE) - (\sigma_1\rho_{1\epsilon} - \sigma_0\rho_{0\epsilon})\,\epsilon\},\ \kappa = \frac{v+1}{v+2},$$

and

$$t = \left(s^2 + \left(\overline{\sigma}_1^{2\kappa} - \overline{\sigma}_0^{2\kappa}\right)\left(\left(\overline{\sigma}_1^{2(1-\kappa)} - \overline{\sigma}_0^{2(1-\kappa)}\right)(v-1)\right)\right)^{\frac{1}{2}}.$$

(iii) Suppose $\rho_{j\epsilon}^2 = 1$ *for at least one* $j = 0, 1$. *Then* $F_\Delta(\delta)$ *is point identified. For example, if* $\rho_{1\epsilon}^2 = 1$, *then*

$$\begin{aligned} F_\Delta(\delta) &= F_\Delta^L(\delta) = F_\Delta^U(\delta) = T_{[v]}\left(\frac{\delta - ATE}{\sqrt{\sigma_1^2 + \sigma_0^2 - 2\rho_{0\epsilon}\sigma_1\sigma_0}}\right), \ \textit{if } \rho_{1\epsilon} = 1, \\ F_\Delta(\delta) &= F_\Delta^L(\delta) = F_\Delta^U(\delta) = T_{[v]}\left(\frac{\delta - ATE}{\sqrt{\sigma_1^2 + \sigma_0^2 + 2\rho_{0\epsilon}\sigma_1\sigma_0}}\right), \ \textit{if } \rho_{1\epsilon} = -1. \end{aligned}$$

We evaluated these bounds for the same set of parameters used in the normal case for $v = 4$, see Figure 4. The same general qualitative conclusions hold as in the normal case. Comparing Figures 1 and 3, we observe that the degree of freedom parameter has little effect on the bounds at the ATE, but it has large effect on the bounds away from ATE. This is due to the fact that Student's t distribution has fatter tails than the normal distribution.

Bounds on $F_\Delta(\cdot|D=1)$ and the Propensity Score

In a SRM, the propensity score is given by

$$P(D=1|W,X_c) = P(\epsilon > -(W,X_c)'\gamma) = 1 - F_\epsilon(-(W,X_c)'\gamma).$$

Hence the smaller the value of $(W,X_c)'\gamma$, the less likely the individual with the value of $(W,X_c)'\gamma$ will participate in the program or the smaller the propensity score. Since there is a one-to-one relation between the propensity score and $(W,X_c)'\gamma$, we can group individuals in the population via their propensity score. For a given value of the propensity score, Theorem 4.3 (ii) provides sharp bounds on the distribution of Δ for participants with the given propensity score in semiparametric SRMs. Figure 5 depicts the distribution bounds for Δ for participants with $(W,X_c)'\gamma = -1.28$ or propensity score 0.1. Figures 5 (a) and 5 (b) are based on the normal assumption, while Figures 5 (c) and 5 (d) are based on the Student's t assumption with degree of freedom 4. We observe that the distribution bounds in Student's t case are generally wider than those in normal case. Moreover, plots with different values of the propensity score and/or the degree of freedom in the Student's t case reveal that the degree of skewness of each bound increases as the propensity score decreases and the bounds get tighter as the degree of freedom increases.

One important and potentially useful application of the distribution bounds established in Theorem 4.3 (ii) is to predict or bound the probability that an individual with a given propensity score will benefit from participating in the program in terms of Δ. Note that

$$F^L_\Delta(0|D=1) \leq P(\Delta \leq 0|D=1) \leq F^U_\Delta(0|D=1).$$

Hence $1-F^L_\Delta(0|D=1)$ is the maximum probability that an individual with a given propensity score will benefit from participating in the program and $1-F^U_\Delta(0|D=1)$ is the minimum probability that an individual with a given propensity score will benefit from participating in the program. To see how these probabilities change with respect to the propensity score, we plotted them against the propensity score in SRMs with bivariate normal distributions in Figures 6(b)-6(b). The expressions[4] for $F^L_\Delta(\delta|D=1)$ and $F^U_\Delta(\delta|D=1)$ are derived by using Theorem 4.3 (ii) and a similar argument to Theorem 5.2. Using these expressions, one can show[5] that the bounds $F^L_\Delta(\delta|D=1)$ and $F^U_\Delta(\delta|D=1)$ approach either 0 or 1 as the propensity score approaches zero. As a result, the bounds are informative for individuals with low propensity score and once they participate, with high probability, they either get hurt or benefit from the treatment.

In a SRM with bivariate normal distributions, TT is given by

$$TT = ATE + (\rho_{1\epsilon}\sigma_1 - \rho_{0\epsilon}\sigma_0)\,\lambda\left((W, X_c)'\gamma\right),$$

where $\lambda(\cdot)$ is the inverse mills ratio. For a given value of $(W, X_c)'\gamma$ or a given value of the propensity score, TT measures the average treatment effect for the subpopulation of participants with the given propensity score. It is composed of two terms: the first term is the average treatment effect for the population with covariates X_1, X_0, X_c, W and the

[4] They are tedious and hence not provided here, but they are available upon request.
[5] The proofs are elementary, but tedious. They are available upon request.

second term is the effect due to selection on unobservables. Figures 6((a)-11(a) plotted TT and the second term in TT due to unobservables against the propensity score. Also plotted in each graph are the bounds on the median of the distribution $F_\Delta(\cdot|D=1)$.

In Figures 6 and 7, ATE is zero. In Figure 6(, $(\rho_{1\epsilon}, \rho_{0\epsilon}) = (0.5, -0.5)$ and TT is non-negative for all values of the propensity score. However, when the propensity score is greater than 0.54, there is a positive probability that an individual with the given propensity score will get hurt by participating in the program. This probability increases as the value of the propensity score increases. And for all values of the propensity score, there is always a positive probability that an individual with the given propensity score will benefit from participating in the program and this probability decreases as the value of the propensity score increases. Consequently, people with low propensity score would benefit from the program with high probability once they participate. In Figure 7, $(\rho_{1\epsilon}, \rho_{0\epsilon}) = (-0.5, 0.5)$ and TT is non-positive for all values of the propensity score. However, when the propensity score is less than 0.54, there is a positive probability that an individual with the given propensity score will benefit from participating in the program and this probability increases as the value of the propensity score increases. In addition, for all values of the propensity score, there is always a positive probability that an individual with the given propensity score will get hurt from participating in the program and this probability decreases as the value of the propensity score increases. The seemingly reversal roles of the two probabilities in Figures 6 and 7 are due to the reversal of the correlation values. Consider Figure 6 with $(\rho_{1\epsilon}, \rho_{0\epsilon}) = (0.5, -0.5)$. Heuristically, for small values of the propensity score, individuals participating in the program tend to have large selection errors ϵ. Given the positive correlation between Y_1 and ϵ, Y_1 would tend to be large for those participants. By the same token, the negative correlation between Y_0 and ϵ imply small Y_0. As a result, Δ tend to be large for participants with small propensity score. Figures 6 and 7 demonstrate clearly that average treatment effect parameters such as ATE and TT do not provide a complete picture of the effects of treatment when there is selection on unobserved variables, and the distribution bounds we established in Chapter III provide useful information that are missed by ATE and TT.

Figures 8 and 9 further support the conclusions we drew from Figures 6 and 7. They are similar to Figures 6 and 7 except that $ATE = -0.5$ in Figure 8 and $ATE = 0.5$ in Figure 9. In both figures, TT is positive for some values of the propensity score and negative for other values of the propensity score. The patterns of $\left[1 - F_{\Delta}^{L}(0|D=1)\right]$ and $\left[1 - F_{\Delta}^{U}(0|D=1)\right]$ as functions of the propensity score remain the same as in Figures 6 and 7. It is interesting to observe from Figures 8 and 9 that even when the ATE for the whole population is negative (-0.5) or positive (0.5), some subpopulations (those with the propensity score less than 0.73) will in general benefit or get hurt from the program if they join the program. The proportion of people in each subgroup who will benefit or get hurt from being in the program will also change with the level of ATE.

In Figures 10 and 11, we increased $\rho_{1\epsilon}$ to 0.95. Comparing these figures with Figures 6 -9, we see clearly that the distribution bounds get tighter as $\rho_{1\epsilon}$ ($\rho_{0\epsilon}$) gets larger. When the magnitudes of $\rho_{1\epsilon}, \rho_{0\epsilon}$ are the same, the bounds are more informative when $\rho_{1\epsilon}$ and $\rho_{0\epsilon}$ have different signs than when they have the same sign.

Summarizing Figures 6 -9, we conclude that the unobserved selection error has a large effect on those with low propensity score. That is, those who are less likely to participate in the program will most likely be affected by the program once they participate in the program. Whether they gain or lose from participating in the program once they participate depends on the sign of $(\rho_{1\epsilon}\sigma_1 - \rho_{0\epsilon}\sigma_0)$.

Confidence Sets for $F_\Delta(\delta)$ in Semiparametric SRMs

Given the sharp bounds established in Chapter III, statistical inference on the joint distribution of potential outcomes and the distribution of treatment effects falls in the currently active research area: inference for partially identified parameters. However, due to the complicated functional dependence of these bounds, especially the bounds on the distribution of treatment effects, on $g_1(x_1, x_c)$, $g_0(x_0, x_c)$, $F_{1\epsilon}$, and $F_{0\epsilon}$, existing inference procedures for partially identified parameters are not easily extendable to the general class of semiparametric SRMs in (III.7). We leave this important issue to future work. In this section, we demonstrate its feasibility by constructing an asymptotically uniformly valid and non-conservative confidence set for $F_\Delta(\delta)$ in a special class of semiparametric SRMs in which $g_1(x_1, x_c)$, $g_0(x_0, x_c)$, $F_{1\epsilon}$, and $F_{0\epsilon}$ are parametric. This is a semiparametric SRM, as the joint distribution of U_1, U_0 is completely unspecified.

Theorem 4.2 (i) implies that $F_\Delta^L(\delta) \leq F_\Delta(\delta) \leq F_\Delta^U(\delta)$, where

$$\begin{aligned} F_\Delta^L(\delta) &= \int_{-\infty}^{+\infty} \left\{ \sup_u \left[F_{1|\epsilon}(u) - F_{0|\epsilon}(u - \{\delta - [g_1(x_1, x_c) - g_0(x_0, x_c)]\}) \right] \right\}_+ dF_\epsilon(\epsilon), \\ F_\Delta^U(\delta) &= 1 + \int_{-\infty}^{+\infty} \left\{ \inf_u \left[F_{1|\epsilon}(u) - F_{0|\epsilon}(u - \{\delta - [g_1(x_1, x_c) - g_0(x_0, x_c)]\}) \right] \right\}_- dF_\epsilon(\epsilon), \end{aligned}$$

in which $(x)_+ = \max(x, 0)$ and $(x)_- = \min(x, 0)$. For notational compactness, we let $\theta^0 = F_\Delta(\delta)$, $\theta_L = F_\Delta^L(\delta)$, and $\theta_U = F_\Delta^U(\delta)$.

Let

$$\tau_\epsilon(u) = F_{1|\epsilon}(u) - F_{0|\epsilon}(u - \{\delta - [g_1(x_1, x_c) - g_0(x_0, x_c)]\})$$

and θ_τ denote the collection of unknown parameters in $\tau_\epsilon(u) : \tau_\epsilon(u) = \tau_\epsilon(u; \theta_\tau)$. Further, let $F_\epsilon(\epsilon) = F_\epsilon(\epsilon; \theta_\epsilon)$. Let $\vartheta_o \in \Theta$ be the vector of all the parameters in the model, i.e.,

$\vartheta_o = \left(\theta_\tau', \theta_\epsilon'\right)'$. Then

$$\begin{aligned}\theta_L &= F_\Delta^L(\delta;\vartheta) = \int_{-\infty}^{+\infty} \left\{\sup_u \left[\tau_\epsilon\left(u;\theta_\tau\right)\right]\right\}_+ dF_\epsilon(\epsilon;\theta_\epsilon),\\ \theta_U &= F_\Delta^U(\delta;\vartheta) = 1 + \int_{-\infty}^{+\infty} \left\{\inf_u \left[\tau_\epsilon\left(u;\theta_\tau\right)\right]\right\}_- dF_\epsilon(\epsilon;\theta_\epsilon).\end{aligned}$$

Let $\widehat{\theta}_\tau$ and $\widehat{\theta}_\epsilon$ denote consistent estimators of θ_τ and θ_ϵ respectively and $\widehat{\vartheta} = \left(\widehat{\theta}_\tau', \widehat{\theta}_\epsilon'\right)'$. Examples of $\widehat{\vartheta}$ include the maximum likelihood estimator and the two-step estimator. The plug-in estimators of θ_L and θ_U are given by

$$\begin{aligned}\widehat{\theta}_L &= \widehat{F}_\Delta^L(\delta) = \int_{-\infty}^{+\infty} \left\{\sup_u \left[\tau_\epsilon\left(u;\widehat{\theta}_\tau\right)\right]\right\}_+ dF_\epsilon(\epsilon;\widehat{\theta}_\epsilon),\\ \widehat{\theta}_U &= \widehat{F}_\Delta^U(\delta) = 1 + \int_{-\infty}^{+\infty} \left\{\inf_u \left[\tau_\epsilon\left(u;\widehat{\theta}_\tau\right)\right]\right\}_- dF_\epsilon(\epsilon;\widehat{\theta}_\epsilon).\end{aligned}$$

For $\varepsilon > 0$, let $B_\Theta(\varepsilon) = \{\vartheta \in \Theta : \|\vartheta - \vartheta_o\| < \varepsilon\}$. Let $\mathcal{W}_i = (Y_i, X_{1i}', X_{0i}', X_{ci}', W_i', D_i)'$ with $Y_i = D_i Y_{1i} + (1-D_i) Y_{0i}$ and $\mathcal{P}$ denote the collection of all the potential distributions of the random sample $\{\mathcal{W}_i\}_{i=1}^n$ such that Assumption 1 below holds.

Assumption 1. (i) For some $\varepsilon > 0$, there exists a function $\Gamma_P^j(\delta)[\vartheta - \vartheta_o]$ of $(\vartheta - \vartheta_o)$, $\vartheta \in B_\Theta(\varepsilon)$, such that for $j = L, U$,

$$\left|F_\Delta^j(\delta;\vartheta) - F_\Delta^j(\delta;\vartheta_o) - \Gamma_P^j(\delta)[\vartheta - \vartheta_o]\right| \le M_\delta \|\vartheta - \vartheta_o\|^2 \tag{III.23}$$

with a constant M_δ that does not depend on P, and for each $\varepsilon > 0$,

$$\limsup_{n\ge 1} \sup_{P\in\mathcal{P}} P\left\{\left|\sqrt{n}\Gamma_P^j(\delta)\left[\widehat{\vartheta} - \vartheta_o\right] - \frac{1}{\sqrt{n}}\sum_{i=1}^n \psi_\delta^j(\mathcal{W}_i)\right| > \varepsilon\right\} = 0, \tag{III.24}$$

where $\psi_\delta^j(\mathcal{W}_i)$ satisfies that there exists $\eta > 0$ such that $E\left[\psi_\delta^j(\mathcal{W}_i)\right] = 0$ and $\sup_{P\in\mathcal{P}} \left\|\psi_\delta^j(\mathcal{W}_i)\right\|_{P,2+\eta} < \infty$. (ii) $\left\|\widehat{\vartheta} - \vartheta_o\right\| = o_P\left(n^{-1/4}\right)$ uniformly in $P \in \mathcal{P}$.

The first condition in Assumption 1 (i) is concerned with differentiability of $F_\Delta^j(\delta;\vartheta)$, $j = L, U$, with respect to $\vartheta \in B_\Theta(\varepsilon)$. This holds under mild differentiability conditions on $\tau_\epsilon(u;\theta_\tau)$ with respect to θ_τ and $F_\epsilon(\epsilon;\theta_\epsilon)$ with respect to θ_ϵ. The second condition in Assumption 1 (i) imposes an asymptotic linear representation on $\Gamma_P^j(\delta)\left[\widehat{\vartheta}-\vartheta_o\right]$. This can be established using the asymptotic linear representation of $\left[\widehat{\vartheta}-\vartheta_o\right]$. Assumption 1 (ii) can be established by following the procedure of Theorem 3.2.5 of van der Vaart and Wellner (1996).

Under Assumption 1, one can show that unformly in $P \in \mathcal{P}$, we have

$$\sqrt{n}\begin{pmatrix}\widehat{\theta}_L-\theta_L\\ \widehat{\theta}_U-\theta_U\end{pmatrix}\Longrightarrow N(0,\Omega), \tag{III.25}$$

where Ω is the variance-covariance matrix of $\left(\psi_\delta^L(\mathcal{W}_i),\psi_\delta^U(\mathcal{W}_i)\right)'$. As a result, CSs developed in Andrews and Guggenberger (2007), Andrews and Soares (2007), Fan and Park (2007), and Stoye (2007), among others, are applicable to θ^0. To illustrate, we provide a brief summary of the CS in Fan and Park (2007). Note that $\theta_L \leq \theta^0 \leq \theta_U$ is equivalent to $\theta^0 = \arg\min_{\theta\in[0,1]}\left\{(\theta_L-\theta)_+^2+(\theta_U-\theta)_-^2\right\}$. We define the sample criterion function as

$$T_n(\theta^0)=n\left(\widehat{\theta}_L-\theta^0\right)_+^2+n\left(\widehat{\theta}_U-\theta^0\right)_-^2. \tag{III.26}$$

Then a $(1-\alpha)$ level CS for θ^0 can be constructed as

$$CS_n=\{\theta\in[0,1]:T_n(\theta)\leq c_{1-\alpha}(\theta)\} \tag{III.27}$$

for an appropriately chosen critical value $c_{1-\alpha}(\theta)$.

Let $(Z_L,Z_U)'\sim N(0,\Omega)$. It follows from (III.25) that

$$T_n(\theta)\Longrightarrow\left(Z_L-h^L(\theta)\right)_+^2+\left(Z_U+h^U(\theta)\right)_-^2$$

where

$$h^L(\theta)=-\lim_{n\to\infty}\sqrt{n}\,[\theta_L-\theta]\text{ and }h^U(\theta)=\lim_{n\to\infty}\sqrt{n}\,[\theta_U-\theta].$$

Let $\widehat{\Delta}^S = \widehat{\Delta} I \left\{ \widehat{\Delta} > b_n \right\}$, where $\widehat{\Delta} \equiv \widehat{\theta}_U - \widehat{\theta}_L$ and b_n is a pre-specified sequence of positive numbers satisfying $b_n \to 0$ and $\sqrt{n} b_n \to \infty$. Then an asymptotically uniformly valid and non-conservative CS for θ_0 is given by CS_n with

$$c_{1-\alpha}(\theta) = \max \left\{ cv_{1-\alpha}\left(0, \sqrt{n}\widehat{\Delta}^S, \widehat{\Omega}\right), cv_{1-\alpha}\left(\sqrt{n}\widehat{\Delta}^S, 0, \widehat{\Omega}\right) \right\},$$

where $cv_{1-\alpha}\left(h^L, h^U, \Omega\right)$ is the $1-\alpha$ quantile of the random variable $\left(Z_L - h^L\right)_+^2 + \left(Z_U + h^U\right)_-^2$ and $\widehat{\Omega}$ is a uniformly consistent estimator of Ω.

Theorem 10 *Suppose Assumption 1 holds and $0 < \alpha < 1/2$. Then CS_n satisfies*

$$\lim_{n \to \infty} \inf_{\theta \in [0,1]} \inf_{P \in \mathcal{P}: \theta_0(P) = \theta} \Pr\left(\theta^0 \in CS_n\right) = 1 - \alpha.$$

Remark 6.1. Note that $c_{1-\alpha}(\theta) = c_{1-\alpha}$ does not depend on θ. It follows from Fan and Park (2007) that

$$CS_n = \begin{cases} \left[\hat{\theta}_L - \sqrt{c_{1-\alpha}} \frac{1}{\sqrt{n}},\ \hat{\theta}_U + \sqrt{c_{1-\alpha}} \frac{1}{\sqrt{n}}\right] & \text{if } \sqrt{n}\widehat{\Delta} \geq -\sqrt{c_{1-\alpha}} \\ \left[\widehat{A}, \widehat{B}\right] & \text{if } -\sqrt{2c_{1-\alpha}} \leq \sqrt{n}\widehat{\Delta} < -\sqrt{c_{1-\alpha}} \\ \varnothing & \text{if } \sqrt{n}\widehat{\Delta} < -\sqrt{2c_{1-\alpha}} \end{cases}$$

where

$$\widehat{A} \equiv \frac{\hat{\theta}_L + \hat{\theta}_U}{2} - \frac{1}{2\sqrt{n}} \sqrt{c_{1-\alpha} - \frac{n\widehat{\Delta}^2}{2}},\ \widehat{B} \equiv \frac{\hat{\theta}_L + \hat{\theta}_U}{2} + \frac{1}{2\sqrt{n}} \sqrt{c_{1-\alpha} - \frac{n\widehat{\Delta}^2}{2}}.$$

Conclusion

In Chapter III we have established sharp bounds on the joint distribution of potential outcomes and the distribution of treatment effects in NMVM-SRMs and in semiparametric SRMs of Heckman (1990). The means of the distributions of treatment effects that we considered correspond to the average treatment effect and the treatment effect for the treated respectively. The results we obtain reveal the important role played by self selection, i.e., it helps tighten the bounds on these distributions. The approach that we used to establish sharp bounds in semiparametric SRMs of Heckman (1990) is general in that it is applicable to any semiparametric SRMs in which the only unidentified bivariate marginal distribution of the joint distribution of $(U_1, U_0, \epsilon)'$ is that of $(U_1, U_0)'$. For example, sharp bounds on the distribution of potential outcomes and the distribution of treatment effects in the semiparametric SRMs of Carneiro and Lee (2008) can be established by using our approach.

As a first step, Chapter III has focused on partial identification. Estimation of the distribution bounds developed in Chapter III is straightforward in view of the identification results in Heckman (1990) and existing work on estimation of parametric/semiparametric sample selection models. Heckman (1990) provides a review of various nonparametric/semiparametric methods for estimating $g_1(x_1, x_c)$ and $g_0(x_0, x_c)$ without specifying the bivariate margins for (U_1, ϵ) and (U_0, ϵ), see also Ai (1997), Andrews and Schafgans (1998), Schafgans and Zinde-Walsh (2002), Das, Newey, and Vella (2003), Chen (2006), and Chen and Zhou (2006). Gallant and Nychka (1987) provide estimators of the unknown bivariate marginal distributions $F_{1\epsilon}$ and $F_{0\epsilon}$. It remains to establish a complete set of inference tools for the joint distribution of potential outcomes and the distribution of treatment effects in general semiparametric SRMs.

CHAPTER IV

TREATMENT EFFECT STUDY FOR FIRMS WITH ACCELERATED EQUITY OFFERINGS

Introduction

Equity offerings are important in the life of a firm. Firms can use equity offerings to raise capital for their capital expenditures, new investment projects, mergers and acquisitions, and other various motives. For firms decided to have equity offerings, they also face a variety of flotation methods, which include firm commitment underwriting, shelf offer, rights offer, standby rights offer, private placement, etc. We refer the reader to Eckbo, Masulis, Norli (2007) for their excellent survey and discussion of alternative flotation methods for equity issuance. Firm commitment underwriting is the primary choice of publicly traded U.S. firms and it has lengthy and complicated processes for both the registration with the security and exchange commission (SEC) and the bookbuilding to solicit tentative offers from investors. We refer to it as the "traditional bookbuilding" method. Shelf offer refers to an issue that has been pre-registered with the SEC under Rule 415, which was introduced in 1983. Only financially strong firms are allowed to use shelf offers and shelf registration increases the flexibility and speed of their issues over a two year period.

For those shelf-registered firms to issue seasoned equity offerings (SEOs), they are already entitled to speeding up their underwriting process of SEOs. However, it is still a choice for them to further speed up the underwriting process through accelerated underwriting. Here we adopt the same definition for an accelerated underwriting by Bortolotti, Megginson and Smart (2008). Two of the three forms of the accelerated underwriting include block trades (BTs) and bought deals (BDs) to sell large blocks of shares, at an auction-determined market price, directly to an investment bank by the issuing firm or selling shareholder, with little need or capacity for information production with respect to pricing or demand. The winning bank is then responsible for reselling the shares to institutional investors. The third and most popular type of accelerated underwriting, accelerated bookbuilt offering (ABO), is executed much more rapidly than conventional bookbuilds, but is similar to traditional underwritings in that banks are responsible for the quality of the order book, price stabilization, and transparency of the allocation.

SEOs executed through accelerated underwriting have increased global market share recently, raising over $850 billion since 1998, and now account for over half (two-thirds) of the value of US SEOs (Bortolotti, Megginson and Smart, 2008). Autore, Hutton and Kovacs (2009) finds shelf registered firms with accelerated underwriting underperform those with non-accelerated underwriting. They hypothesize that the choice of flotation methods can be served as a signal of issuer quality and lower quality firms intend to use accelerated underwriting. However, their hypothesis can not explain the increasing popularity of accelerated underwriting among SEOs in recent years. In addition, we note that shelf registered firms are required by the SEC to be large and financially sound. We are interested in studying the impact of accelerated underwriting on firm's performance in order to better understand the global rise of accelerated underwriting.

To correctly evaluate the impact of accelerated underwriting, we need to have knowledge of a firm's counterfactual outcome, the outcome of an alternative choice of flotation methods. The difference between a firm's actual outcome and its counterfactual outcome is due to the choice of flotation methods. This is a direct application of treatment effect studies to corporate finance. A comparison between accelerated firms and matched non-accelerated firms may not be valid due to the fact that corporate finance decisions are usually deliberate decisions by firms or their managers to self-select into their preferred choices. Matching can not account for the impact of private information in firm's financial decision. Alternatives that could be appropriate to address the issue are self-selection models. Self-selection models are not new to corporate finance. Li and Prabhala (2005) survey econometrics models of self selection and their applications in corporate finance. In Chapter II, we extend existing self-selection models (SRMs) to allow more flexibility in modeling the relationship between private information and the outcomes, and we are interested in applying them to study the impact of accelerated underwriting on firm's performance in Chapter IV. For firms facing uncertain outcomes in their financial decisions, the ability of managers to make decisions associated with a higher probability of wealth increasing for their shareholders is an important question in corporate governance. The partial identification results in Chapter III can also help to look into this issue and it will be on my future research agenda.

We use two types of performance measures in our study: cumulative abnormal stock returns (CAR) one year after the equity issuance and industry adjusted return on assets (ROA) one year after the equity issuance. Stock return is a forward looking measure of performance, while return on assets is a backward looking measure of performance. Performance measured by ROA is also considered as operating performance. Our initial findings show that private information in firms' decisions to choose accelerated underwriting has a significant positive impact on issuing firm's ROA one year after its equity issuance. We also find a significant negative average treatment effect on firm's ROA for the control group (with non-accelerated underwriting).

Chapter IV is organized as follows. Section 2 describes the data and issuer characteristics. In section 3 we present the models and main results. We conclude in section 4.

Data and Issuer Characteristics

Our data is collected from Securities Data Company's (SDC) Global New Issues database and consists of shelf-registered equity offers of common shares during the period 1997-2006. All sample issuers are U.S. non-financial and non-real estate firms that are listed on the NYSE, AMEX, or NASDAQ, and that have the required stock return data available from the Center for Research in Securities Prices (CRSP) as well as financial data available from COMPUSTAT.

Table 3 presents the annual frequency of and total proceeds reaped from accelerated and non-accelerated shelf offers. The sample contains 559 bookbuilt and 420 accelerated offerings, and all of them are shelf registered. Although the total number of traditional offers outnumbers the total number of shelf offers, the annual frequency of traditional offers has declined over time. The annual frequency of shelf offerings, however, has increased and even surpassed that of traditional SEOs.

Table 4 provides descriptive statistics for firm and offer characteristics. Variable definitions are presented in the Appendix. Compared to bookbuilt issuers, accelerated issuers are significantly larger based on book assets and market capitalization. Bookbuilt and accelerated issuers raise similar proceeds, but bookbuilt issuers offer more shares relative to the number of shares outstanding. There are few differences between bookbuilt and accelerated offerings in terms of the market-to-book ratios and return on assets. By contrast, the typical traditional SEO issuer is considerably different from either type of shelf issuer across most dimensions. These characteristics indicate firms in our sample are more homogeneous with respect to firm and offer characteristics than firms using traditional flotation methods.

Models and Main Results

Consider each firm in our sample faces a binary choice, accelerated underwriting or non-accelerated underwriting of shelf registered firms, and each choice associated with an outcome. Use D as an indicator for the choices, $D = 1$ indicates the choice of accelerated underwriting, while $D = 0$ indicates the choice of non-accelerated underwriting. We assume firms are independent and we exclude firms who have other equity issuances a year before and after the equity issuance. The sample size is reduced to 249 accelerated underwriting and 272 non-accelerated underwriting. We assume each firm has private information, ϵ, in its decision to choose a flotation method. We can not observe ϵ, but we can observe their choice D. $D = 1$ if $W'\gamma + \epsilon > 0$ and $D = 0$ otherwise, where W represents observed firm and offer characteristics. We have the following switching regimes model as in Chapter II and Chapter III. :

$$\begin{aligned} Y_{1i} &= X_i'\beta_1 + U_{1i}, \\ Y_{0i} &= X_i'\beta_0 + U_{0i}, \\ D_i &= I_{\{W_i'\gamma + \epsilon_i > 0\}}, \; i = 1, \ldots, n, \end{aligned} \tag{IV.1}$$

to model firm i's choice and outcome. Y_{1i} is the outcome of firm i with accelerated underwriting, and Y_{0i} is its outcome with non-accelerated underwriting.

The outcome variables in this chapter are variables to measure firm's performance after its equity issuance. We include two types of performance measure in our study: cumulative abnormal stock returns (CAR) and industry adjusted return on assets (ROA). The observed covariates are key variables of firm and offer characteristics, which include relative offer size, firm's log value of market capitalization, a dummy variable to indicate whether a firm is listed in NASDAQ, average turnover of daily shares between 1 year and 30 days prior to the issuance scaled by the number of shares outstanding, runup of stock price 90 trading days prior to the issuance, return on assets in the fiscal year of the issuance, market to book ratio, percentage of secondary shares in the offer and dummy variables to indicate if a firm has a simultaneous offer for convertible debts, or if it has convertible debt issued before or after its equity issuance, or if it is in the S&P 500. Cumulative abnormal stock return is calculated using Fama-French three-factor model and return on assets is calculated using earnings before interest, taxes, depreciation and amortization (EBITDA) divided by average total assets. We use industry (classified by a 2-digit standard industrial classification code) median ROA to adjust our ROA for each firm in the sample. The relative offer size and residual volatility might only affect the choice of flotation methods but will probably not have much effect on firm's performance. In our analysis, we use relative offer size and residual volatility as instrument variables in our model.

The errors $\{U_{1i}, U_{0i}, \epsilon_i\}$ are assumed to be independent of the covariates $\{X_i, W_i\}$ and we impose parametric distribution assumptions on the bivariate distribution between U_{ji} and ϵ, $j = 0, 1$. To address the impact of skewness, we use skewed t model with degree of freedom $v = 5$ and compare the results with conventional Heckman's sample selection model. In Table 5, we present the results for the choice of flotation methods in the first step. Results are qualitatively similar between two models and they are also qualitatively similar to the results by Autore, Hutton and Kovacs (2009). The choice of an accelerated offer is inversely related to the relative offer size, residual volatility, return on assets and whether there is convertible debt issued prior to the issuance, and positively related to average turnover and the percentage of secondary shares in the offer. The log-likelihood of skewed t model and Heckman's model are close to each other, while the skewed t model has slightly larger log-likelihood value and it shows that the skewness parameter in the select is also significant at 90 percent level.

In Table 6 and Table 7, we present the results of the outcome equations of ROA. The sign and magnitude of coefficients in the equations are similar between skewed t model and Heckman's model. In skewed t model, correction term 1 is corresponding to inverse mill's ratio in Heckman's model and the coefficient of correction term 2 is the skewness parameter in the outcome. The results show that log(market capitalization), being a Nasdaq firm, average turnover, return on assets, a firm's convertible debt issued after its equity issuance within a year have a positive impact on firm's operating performance in the treated group (with accelerated underwriting) , while a firm's convertible debt issued prior to its equity issuance, whether it is a simultaneous offer have a negative impact on firm's operating performance in the treated group. In addition, the skewed t model captures a significant (at 90 percent level) negative skewness parameter in the outcome equation for the treated group. It also shows that being a S&P 500 firm has a negative impact, while the private information in firm's flotation choice and the percentage of secondary in the offer have a positive impact on a firm's operating performance. In the control group, return on assets and a firm's convertible debt issued prior to its equity issuance have a positive impact and a firm's convertible debt issued after its equity issuance has a negative impact on its operating performance.

In Table 8 and Table 9, we present the results of the outcome equations of cumulative abnormal return one year after a firm's equity issuance. We do not find any significance (at 90 percent level) for the skewness parameter in both outcome equations. The sign and magnitude of most coefficients (except the intercept and the percentage of secondary shares in the offer) in the equations are similar between skewed t model and Heckman's model. The results show that return on assets, a firm's convertible debt issued after its equity issuance within a year have a positive impact and stock runup 90 days prior to the issuance has a negative impact on firm's cumulative abnormal return in the treated group. In the control group, return on assets and a firm's convertible debt issued prior to its equity issuance have a positive impact on its stock performance. Two models give similar results in terms of significance at 90 percent level.

In Table 10, we also present the results for various average treatment effects. The average effect on the whole population, the effect on the treated and the effect on the control are insignificant for cumulative abnormal return 1 year, 6 months, 3 months and 3 days after the equity issuance. However, we document a significant (at 90 percent level) negative treatment effect on operating performance (ROA) for firms in the control group. From the estimated outcome equations, we can also calculate the expected gain from choosing accelerated underwriting for each firm in our sample based on its observed characteristics. We plot these expected gains for ROA and CAR one year after the equity issuance in Figure 12 and Figure 13 respectively for both the treated group and the control group with respect to their propensity score (probability to choose an accelerated underwriting). The skewed t model captures more difference in the expected gains between the treated group and the control group than Heckman's model does.

Conclusion

We document the findings for the impact of accelerated underwriting on a firm's performance one year after its equity issuance: industry adjusted return on assets and cumulative abnormal return. It shows that the skewness is a significant component in the outcome equations for return on assets. With the ability to account for skewness in the outcome equations, our skewed t model is able to capture the impact of private information involved in a firm's flotation choice, the impact of being a S&P 500 firm and the impact of the percentage of secondary shares on its operating performance (ROA). When the skewness parameter is insignificant, the skewed t model and conventional Heckman's model yield similar results.

.

Appendix: Technical Proofs

Chapter II

Proof of Theorem 3.1: Let $V = (U, \epsilon) \sim NMVM_2 [\xi, \mu, \Sigma, \zeta]$, then by definition, V has the same distribution as $\mu + S\zeta + \sqrt{S}AZ$, where $AA' = \Sigma$. Thus, $V|S = s \sim N_d (\mu + s\zeta, s\Sigma)$. We have $\mu = -\zeta E(S)$. Let $E(U|\epsilon = x) = m(x)$, then

$$\begin{aligned} m(x) &= E(U|\epsilon = x) \\ &= E(E(U|\epsilon = x, S)) \\ &= E\left(\zeta_U (S - E(S)) + \rho\frac{\sigma_U}{\sigma_\epsilon}(x - \zeta_\epsilon (S - E(S)))\,|\epsilon = x\right) \\ &= \int \zeta_U (S - E(S)) + \rho\frac{\sigma_U}{\sigma_\epsilon}(x - \zeta_\epsilon (S - E(S)))\,dF_{S|\epsilon=x}(s), \end{aligned}$$

where

$$F_{S|\epsilon=x}(s) = \left(2\pi s\sigma_\epsilon^2\right)^{-1/2}\int_0^s \exp\left\{-(x - \zeta_\epsilon (s - E(S)))^2 / \left(2s\sigma_\epsilon^2\right)\right\}/f_\epsilon(x)\,dF_S(s).$$

Thus, we get

$$\begin{aligned} m(x) &= \zeta_U \int (s - E(S))\,dF_{S|\epsilon=x}(s) + \rho\sigma_U \int \frac{1}{\sigma_\epsilon}(x - (s - E(S))\,\xi_\epsilon)\,dF_{S|\epsilon=x}(s) \\ &= \zeta_U H(x) + \rho\sigma_U G(x), \end{aligned}$$

where

$$H(x) = \int (s - E(S))\,dF_{S|\epsilon=x}(s),\ G(x) = \int \frac{1}{\sigma_\epsilon}(x - (s - E(S))\,\xi_\epsilon)\,dF_{S|\epsilon=x}(s).$$

The expressions for $E(U|\epsilon > -x)$ and $E(U|\epsilon < -x)$ can be written as

$$
\begin{aligned}
E\left(U|\epsilon>-x\right) &= \int_{-x}^{\infty} m\left(u\right)\frac{f_{\epsilon}\left(u\right)}{1-F_{\epsilon}\left(-x\right)}du \\
&= \zeta_U\int_{-x}^{\infty} H\left(u\right)\frac{f_{\epsilon}\left(u\right)}{1-F_{\epsilon}\left(-x\right)}du+\rho\sigma_U\int_{-x}^{\infty} G\left(u\right)\frac{f_{\epsilon}\left(u\right)}{1-F_{\epsilon}\left(-x\right)}du, \\
E\left(U|\epsilon<-x\right) &= \int_{-\infty}^{-x} m\left(u\right)\frac{f_{\epsilon}\left(u\right)}{F_{\epsilon}\left(-x\right)}du \\
&= \zeta_U\int_{-\infty}^{-x} H\left(u\right)\frac{f_{\epsilon}\left(u\right)}{F_{\epsilon}\left(-x\right)}du+\rho\sigma_U\int_{-\infty}^{-x} G\left(u\right)\frac{f_{\epsilon}\left(u\right)}{F_{\epsilon}\left(-x\right)}du.
\end{aligned}
$$

Let $\lambda_{1a}\left(x\right)=\int_{-x}^{\infty}H\left(u\right)\frac{f_{\epsilon}(u)}{1-F_{\epsilon}(-x)}du$. Then we have

$$
\begin{aligned}
\lambda_{1a}\left(x\right) &= \frac{1}{1-F_{\epsilon}\left(-x\right)}\left\{\int_{-x}^{\infty}\int_{0}^{\infty}\left(s-E\left(S\right)\right)f_{\epsilon}\left(u\right)dF_{S|\epsilon=u}\left(s\right)du\right\} \\
&= \frac{1}{1-F_{\epsilon}\left(-x\right)}\int_{0}^{\infty}\left(s-E\left(S\right)\right)\Phi\left(\frac{x+\xi_{\epsilon}\left(s-E\left(S\right)\right)}{\sqrt{s}\sigma_{\epsilon}}\right)dF_S(s) \\
&= \frac{1}{1-F_{\epsilon}\left(-x\right)}E\left(\left(S-E\left(S\right)\right)\Phi\left(\frac{x+\xi_{\epsilon}\left(S-E\left(S\right)\right)}{\sqrt{S}\sigma_{\epsilon}}\right)\right).
\end{aligned}
$$

Similarly, let $\lambda_{1b}\left(x\right)=\int_{-x}^{\infty}G\left(u\right)\frac{f_{\epsilon}(u)}{1-F_{\epsilon}(-x)}du$, $\lambda_{0a}\left(x\right)=\int_{-\infty}^{-x}H\left(u\right)\frac{f_{\epsilon}(u)}{F_{\epsilon}(-x)}du$, and $\lambda_{0b}\left(x\right)=\int_{-\infty}^{-x}G\left(u\right)\frac{f_{\epsilon}(u)}{1-F_{\epsilon}(-x)}du$. We obtain:

$$
\begin{aligned}
\lambda_{1b}\left(x\right) &= \frac{1}{1-F_{\epsilon}\left(-x\right)}\left\{\int_{-x}^{\infty}\int_{0}^{\infty}\frac{1}{\sigma_{\epsilon}}\left(u-\xi_{\epsilon}\left(s-E\left(S\right)\right)\right)f_{\epsilon}\left(u\right)dF_{S|\epsilon=u}\left(s\right)du\right\} \\
&= \frac{1}{1-F_{\epsilon}\left(-x\right)}\left\{\int_{0}^{\infty}\sqrt{s}\phi\left(\frac{x+\xi_{\epsilon}\left(s-E\left(S\right)\right)}{\sqrt{s}\sigma_{\epsilon}}\right)dF_S\left(s\right)\right\} \\
&= \frac{1}{1-F_{\epsilon}\left(-x\right)}E\left(\sqrt{S}\phi\left(\frac{x+\xi_{\epsilon}\left(S-E\left(S\right)\right)}{\sqrt{S}\sigma_{\epsilon}}\right)\right),
\end{aligned}
$$

$$\begin{aligned}
\lambda_{0a}(x) &= \frac{1}{F_\epsilon(-x)}\left\{\int_{-\infty}^{-x}\int_{0}^{\infty}(s-E(S))\,f_\epsilon(u)\,dF_{S|\epsilon=u}(s)\,du\right\} \\
&= \frac{1}{F_\epsilon(-x)}\int_{0}^{\infty}(s-E(S))\left(1-\Phi\left(\frac{x+\xi_\epsilon(s-E(S))}{\sqrt{s}\sigma_\epsilon}\right)\right)dF_S(s) \\
&= -\frac{1}{F_\epsilon(-x)}\int_{0}^{\infty}(s-E(S))\,\Phi\left(\frac{x+\xi_\epsilon(s-E(S))}{\sqrt{s}\sigma_\epsilon}\right)dF_S(s) \\
&= -\frac{1}{F_\epsilon(-x)}E\left((S-E(S))\,\Phi\left(\frac{x+\xi_\epsilon(S-E(S))}{\sqrt{S}\sigma_\epsilon}\right)\right),
\end{aligned}$$

and

$$\begin{aligned}
\lambda_{0b}(x) &= \frac{1}{F_\epsilon(-x)}\left\{\int_{-\infty}^{-x}\int_{0}^{\infty}\frac{1}{\sigma_\epsilon}(u-\xi_\epsilon(s-E(S)))\,f_\epsilon(u)\,dF_{S|\epsilon=u}(s)\,du\right\} \\
&= -\frac{1}{F_\epsilon(-x)}\left\{\int_{0}^{\infty}\sqrt{s}\phi\left(\frac{x+\xi_\epsilon(s-E(S))}{\sqrt{s}\sigma_\epsilon}\right)dF_S(s)\right\} \\
&= -\frac{1}{F_\epsilon(-x)}E\left(\sqrt{S}\phi\left(\frac{x+\xi_\epsilon(S-E(S))}{\sqrt{S}\sigma_\epsilon}\right)\right).
\end{aligned}$$

For notational convenience, let

$$\lambda_a(x) = E\left((S-E(S))\,\Phi\left(\frac{x+\xi_\epsilon(S-E(S))}{\sqrt{S}\sigma_\epsilon}\right)\right),\ \ \lambda_b(x) = E\left(\sqrt{S}\phi\left(\frac{x+\xi_\epsilon(S-E(S))}{\sqrt{S}\sigma_\epsilon}\right)\right).$$

Then we have:

$$\begin{aligned}
E(U|\epsilon=-x) &= \zeta_U H(x) + \rho\sigma_U G(x), \\
E(U|\epsilon>-x) &= \zeta_U \lambda_{1a}(x) + \rho\sigma_U \lambda_{1b}(x), \\
E(U|\epsilon<-x) &= \zeta_U \lambda_{0a}(x) + \rho\sigma_U \lambda_{0b}(x).
\end{aligned}$$

Technical Derivations in Example 3.1: Let $f_S(s)$ denote the density function of a generalized inverse Gaussian distribution with parameters (λ, χ, ψ). Then we have:

$$f_S(s) = \begin{cases} \frac{\chi^{-\lambda}}{\Gamma(-\lambda)2^{-\lambda}} s^{\lambda-1} \exp\left\{-\frac{\chi}{2s}\right\} & \text{if } \lambda < 0, \chi > 0, \psi = 0; \\ \frac{\psi^{\lambda}}{\Gamma(\lambda)2^{\lambda}} s^{\lambda-1} \exp\left\{-\frac{\psi s}{2}\right\} & \text{if } \lambda > 0, \chi = 0, \psi > 0; \\ \frac{\chi^{-\lambda}\left(\sqrt{\chi\psi}\right)^{\lambda}}{2K_{\lambda}\left(\sqrt{\chi\psi}\right)} s^{\lambda-1} \exp\left\{-\frac{\psi s}{2} - \frac{\chi}{2s}\right\} & \text{if } \chi > 0, \psi > 0. \end{cases}$$

and

$$E(S) = \begin{cases} \frac{\chi\Gamma(-\lambda-1)}{2\Gamma(-\lambda)} & \text{if } \lambda + 1 < 0, \chi > 0, \psi = 0; \\ \frac{2\Gamma(\lambda+1)}{\psi\Gamma(\lambda)} & \text{if } \lambda > 0, \chi = 0, \psi > 0; \\ \frac{\chi^{1/2}K_{\lambda+1}\left(\sqrt{\chi\psi}\right)}{(\psi)^{1/2}K_{\lambda}\left(\sqrt{\chi\psi}\right)} & \text{if } \chi > 0, \psi > 0. \end{cases}$$

So, $\frac{sf_S(s)}{E(S)}$ is a density function and it is also a generalized inverse Gaussian density function with parameters $(\lambda + 1, \chi, \psi)$.

Let

$$\eta_0 = [\lambda, \chi, \psi, -E(S)\zeta_\epsilon, \sigma_\epsilon, \zeta_\epsilon], \ \eta_1 = [\lambda + 1, \chi, \psi, -E(S)\zeta_\epsilon, \sigma_\epsilon, \zeta_\epsilon].$$

Then we obtain the sample correction terms for generalized hyperbolic distributions given below:

$$\begin{aligned}
\lambda_{1a}\left(w'\gamma\right) &= \frac{1}{1-F_{\epsilon}\left(-w'\gamma\right)}\int_{0}^{\infty}\left(s-E\left(S\right)\right)\Phi\left(\frac{w'\gamma+\xi_{\epsilon}\left(s-E\left(S\right)\right)}{\sqrt{s}\sigma_{\epsilon}}\right)f_{S}(s)ds \\
&= \frac{\left(GH_{1}\left(-w'\gamma;\eta_{0}\right)-GH_{1}\left(-w'\gamma;\eta_{1}\right)\right)}{1-GH_{1}\left(-w'\gamma;\eta_{0}\right)}E\left(S\right), \\
\lambda_{1b}\left(w'\gamma\right) &= \frac{1}{1-F_{\epsilon}\left(-w'\gamma\right)}\left\{\int_{0}^{\infty}\sqrt{s}\phi\left(\frac{w'\gamma+\xi_{\epsilon}\left(s-E\left(S\right)\right)}{\sqrt{s}\sigma_{\epsilon}}\right)f_{S}\left(s\right)ds\right\} \\
&= \frac{gh_{1}\left(-w'\gamma;\eta_{1}\right)}{1-GH_{1}\left(-w'\gamma;\eta_{0}\right)}\left(E\left(S\right)\sigma_{\epsilon}\right), \\
\lambda_{0a}\left(w'\gamma\right) &= -\frac{1}{F_{\epsilon}\left(-w'\gamma\right)}\int_{0}^{\infty}\left(s-E\left(S\right)\right)\Phi\left(\frac{w'\gamma+\xi_{\epsilon}\left(s-E\left(S\right)\right)}{\sqrt{s}\sigma_{\epsilon}}\right)f_{S}(s)ds \\
&= \frac{\left(GH_{1}\left(-w'\gamma;\eta_{1}\right)-GH_{1}\left(-w'\gamma;\eta_{0}\right)\right)}{GH_{1}\left(-w'\gamma;\eta_{0}\right)}E\left(S\right), \\
\lambda_{0b}\left(w'\gamma\right) &= -\frac{1}{F_{\epsilon}\left(-w'\gamma\right)}\left\{\int_{0}^{\infty}\sqrt{s}\phi\left(\frac{w'\gamma+\xi_{\epsilon}\left(s-E\left(S\right)\right)}{\sqrt{s}\sigma_{\epsilon}}\right)f_{S}\left(s\right)ds\right\} \\
&= -\frac{gh_{1}\left(-w'\gamma;\eta_{1}\right)}{GH_{1}\left(-w'\gamma;\eta_{0}\right)}\left(E\left(S\right)\sigma_{\epsilon}\right).
\end{aligned}$$

Chapter III

Proof of Theorem 2.1: (i) The 'if part' is obvious. Now we prove the 'only if' part.

First, we consider the case that $Var\left(S\right)>0$. Let $a_j=\sqrt{E\left(S\right)\sigma_j^2}$, $b_j=sign\left(\zeta_j\right)\sqrt{\zeta_j^2 Var\left(S\right)}$, $j=1,0,\epsilon$. Then for any ρ, we have

$$\begin{aligned}
&\left(a_1^2+b_1^2\right)\left(a_0^2+b_0^2\right)-\left(a_1a_0\rho+b_1b_0\right)^2 \\
=\ &\left(1-\rho^2\right)\left(a_1^2a_0^2+a_1^2b_0^2\right)+\left(\rho a_1b_0-a_0b_1\right)^2. \qquad \text{(A.1)}
\end{aligned}$$

Note that

$$\begin{aligned}\bar{\rho}_L &= \frac{E(S)\rho_L\sigma_1\sigma_0+\zeta_1\zeta_0 Var(S)}{\sqrt{\left(E(S)\sigma_1^2+\zeta_1^2 Var(S)\right)\left(E(S)\sigma_0^2+\zeta_0^2 Var(S)\right)}}\\ &= \frac{a_1a_0\rho_L+b_1b_0}{\sqrt{\left(a_1^2+b_1^2\right)\left(a_0^2+b_0^2\right)}}.\end{aligned}$$

If $\bar{\rho}_L=-1$, then the left hand and the right hand side expressions in (A.1) with ρ replaced by ρ_L equal to 0, which implies that $\rho_L^2=1$ and $\rho_L a_1b_0-a_0b_1=0$. Similarly, since

$$\begin{aligned}\bar{\rho}_U &= \frac{E(S)\rho_U\sigma_1\sigma_0+\zeta_1\zeta_0 Var(S)}{\sqrt{\left(E(S)\sigma_1^2+\zeta_1^2 Var(S)\right)\left(E(S)\sigma_0^2+\zeta_0^2 Var(S)\right)}}\\ &= \frac{a_1a_0\rho_U+b_1b_0}{\sqrt{\left(a_1^2+b_1^2\right)\left(a_0^2+b_0^2\right)}},\end{aligned}$$

$\bar{\rho}_U=1$ implies that $\rho_U^2=1$ and $\rho_U a_1b_0-a_0b_1=0$. Since $a_j>0$ and $Var(S)>0$, we must have $(\rho_L-\rho_U)b_0=0$ which implies that either $\rho_L=\rho_U$ or $b_0=0$. Since $\rho_L=\rho_U$ implies $\bar{\rho}_L=\bar{\rho}_U$ violating the condition, we must have $b_0=0$ implying $\zeta_0=0$. Now since $\rho_L\neq\rho_U$, it follows from $\rho_L^2=\rho_U^2=1$ that $\rho_L=-1$ and $\rho_U=1$. This in turn implies that $\rho_L=\bar{\rho}_L=-1$ and $\rho_U=\bar{\rho}_U=1$. $\zeta_1=0$ follows from $a_1b_0+a_0b_1=0$ and $a_1b_0-a_0b_1=0$. From the expressions for ρ_L and ρ_U, it follows immediately that $\rho_{1\epsilon}=\rho_{0\epsilon}=0$. Since

$$\bar{\rho}_{j\epsilon}=\frac{E(S)\rho_{j\epsilon}\sigma_1\sigma_0+\zeta_1\zeta_0 Var(S)}{\sqrt{\left(E(S)\sigma_1^2+\zeta_1^2 Var(S)\right)\left(E(S)\sigma_0^2+\zeta_0^2 Var(S)\right)}},\ j=1,0,$$

with $\zeta_1=\zeta_0=0$, we get $\bar{\rho}_{j\epsilon}=\rho_{j\epsilon}=0$, $j=1,0$.

It remains to consider the case that $Var(S)=0$. In this case, $\bar{\rho}_L=\rho_L$ and $\bar{\rho}_U=\rho_U$. It follows that $\rho_{1\epsilon}=\rho_{0\epsilon}=0$. But when $Var(S)=0$, $\bar{\rho}_{j\epsilon}=\rho_{j\epsilon}=0$.

(ii) and (iii) Since the positive semi-definiteness of Σ implies the positive semi-definiteness of $E(S)\Sigma+Var(S)[\zeta_1,\zeta_0,\zeta_\epsilon]'[\zeta_1,\zeta_0,\zeta_\epsilon]$, which is the variance-covariance matrix of $(U_1,U_0,\epsilon)'$. The bounds $[\bar{\rho}_L,\bar{\rho}_U]$ implied by the positive semi-definiteness of Σ are in general tighter than the bounds implied by that of the variance-covariance matrix of

$(U_1, U_0, \epsilon)'$. So we have:

$$\bar{\rho}_L \geq \bar{\rho}_{1\epsilon}\bar{\rho}_{0\epsilon} - \sqrt{\left(1-\bar{\rho}_{1\epsilon}^2\right)\left(1-\bar{\rho}_{0\epsilon}^2\right)} \text{ and } \bar{\rho}_U \leq \bar{\rho}_{1\epsilon}\bar{\rho}_{0\epsilon} - \sqrt{\left(1-\bar{\rho}_{1\epsilon}^2\right)\left(1-\bar{\rho}_{0\epsilon}^2\right)}.$$

When $\bar{\rho}_{1\epsilon}^2 + \bar{\rho}_{0\epsilon}^2 > 1$ and $\bar{\rho}_{1\epsilon}$, $\bar{\rho}_{0\epsilon}$ have the same sign, we have

$$\begin{aligned} \bar{\rho}_L &\geq \bar{\rho}_{1\epsilon}\bar{\rho}_{0\epsilon} - \sqrt{\left(1-\bar{\rho}_{1\epsilon}^2\right)\left(1-\bar{\rho}_{0\epsilon}^2\right)} \\ &= \bar{\rho}_{1\epsilon}\bar{\rho}_{0\epsilon} - \sqrt{1-\left(\bar{\rho}_{1\epsilon}^2+\bar{\rho}_{0\epsilon}^2\right)+\bar{\rho}_{1\epsilon}^2\bar{\rho}_{0\epsilon}^2} \\ &> \bar{\rho}_{1\epsilon}\bar{\rho}_{0\epsilon} - \sqrt{\bar{\rho}_{1\epsilon}^2\bar{\rho}_{0\epsilon}^2} \\ &= 0. \end{aligned}$$

When $\bar{\rho}_{1\epsilon}^2 + \bar{\rho}_{0\epsilon}^2 > 1$ and $\bar{\rho}_{1\epsilon}$, $\bar{\rho}_{0\epsilon}$ have the opposite sign, we have

$$\begin{aligned} \bar{\rho}_U &\leq \bar{\rho}_{1\epsilon}\bar{\rho}_{0\epsilon} + \sqrt{\left(1-\bar{\rho}_{1\epsilon}^2\right)\left(1-\bar{\rho}_{0\epsilon}^2\right)} \\ &= \bar{\rho}_{1\epsilon}\bar{\rho}_{0\epsilon} + \sqrt{1-\left(\bar{\rho}_{1\epsilon}^2+\bar{\rho}_{0\epsilon}^2\right)+\bar{\rho}_{1\epsilon}^2\bar{\rho}_{0\epsilon}^2} \\ &< \bar{\rho}_{1\epsilon}\bar{\rho}_{0\epsilon} + \sqrt{\bar{\rho}_{1\epsilon}^2\bar{\rho}_{0\epsilon}^2} \\ &= 0. \end{aligned}$$

(iv) Following the notation in **(i)**, we have

$$\begin{aligned} \bar{\rho}_{j\epsilon} &= \frac{E(S)\,\rho_{j\epsilon}\sigma_1\sigma_\epsilon + \zeta_j\zeta_\epsilon Var(S)}{\sqrt{\left(E(S)\,\sigma_j^2+\zeta_j^2 Var(S)\right)\left(E(S)\,\sigma_\epsilon^2+\zeta_\epsilon^2 Var(S)\right)}} \\ &= \frac{a_j a_\epsilon \rho_{j\epsilon} + b_j b_\epsilon}{\sqrt{\left(a_j^2+b_j^2\right)\left(a_\epsilon^2+b_\epsilon^2\right)}},\ j=0,1. \end{aligned}$$

Similar to **(i)**, we conclude that $\bar{\rho}_{j\epsilon}^2 = 1$ implies $\rho_{j\epsilon}^2 = 1$. If $\rho_{1\epsilon}^2 = 1$ or $\rho_{0\epsilon}^2 = 1$, we have $\rho_L = \rho_U$ which leads to $\bar{\rho}_L = \bar{\rho}_U$ and the result that $\bar{\rho}_{10}$ is point identified.

Proof of Theorem 3.1: When $\zeta_1 = \zeta_0 = 0$ and $\bar{\rho}_{1\epsilon} = \bar{\rho}_{0\epsilon} = 0$, Theorem 2.1

implies that $\rho_L = -1$ and $\rho_U = 1$. So we have:

$$\begin{aligned}
& NMVM_2\left(y_1 - x'\beta_1, y_0 - x'\beta_0; \left[\alpha_{10}^{-}, \rho_L\right]\right) \\
= & \int_0^\infty f_S(s)\, C^{Gau}\left(\Phi\left(\frac{y_1 - x'\beta_1}{\sigma_1\sqrt{s}}\right), \Phi\left(\frac{y_0 - x'\beta_0}{\sigma_0\sqrt{s}}\right), -1\right) ds \\
= & \int_0^\infty f_S(s) \max\left\{\Phi\left(\frac{y_1 - x'\beta_1}{\sigma_1\sqrt{s}}\right) + \Phi\left(\frac{y_0 - x'\beta_0}{\sigma_0\sqrt{s}}\right) - 1, 0\right\} ds,
\end{aligned}$$

where the second equality follows from the fact that $C^{Gau}(\cdot,\cdot,-1) = C_L(\cdot,\cdot)$, see e.g., Joe (1997). When $\frac{y_1 - x'\beta_1}{\sigma_1} \geq (\leq) - \frac{y_0 - x'\beta_0}{\sigma_0}$, we have $\frac{y_1 - x'\beta_1}{\sigma_1\sqrt{s}} \geq (\leq) - \frac{y_0 - x'\beta_0}{\sigma_1\sqrt{s}}$ and

$$\Phi\left(\frac{y_1 - x'\beta_1}{\sigma_1\sqrt{s}}\right) + \Phi\left(\frac{y_0 - x'\beta_0}{\sigma_0\sqrt{s}}\right) - 1 = \Phi\left(\frac{y_1 - x'\beta_1}{\sigma_1\sqrt{s}}\right) - \Phi\left(-\frac{y_0 - x'\beta_0}{\sigma_0\sqrt{s}}\right) \geq (\leq)\, 0.$$

Thus, the sign of $\left[\Phi\left(\frac{y_1 - x'\beta_1}{\sigma_1\sqrt{s}}\right) + \Phi\left(\frac{y_0 - x'\beta_0}{\sigma_0\sqrt{s}}\right) - 1\right]$ does not depend on s. We obtain:

$$\begin{aligned}
& NMVM_2\left(y_1 - x'\beta_1, y_0 - x'\beta_0; \left[\alpha_{10}^{-}, \rho_L\right]\right) \\
= & \int_0^\infty f_S(s) \max\left\{\Phi\left(\frac{y_1 - x'\beta_1}{\sigma_1\sqrt{s}}\right) + \Phi\left(\frac{y_0 - x'\beta_0}{\sigma_0\sqrt{s}}\right) - 1, 0\right\} ds \\
= & \max\left\{\int_0^\infty f_S(s)\left\{\Phi\left(\frac{y_1 - x'\beta_1}{\sigma_1\sqrt{s}}\right) + \Phi\left(\frac{y_0 - x'\beta_0}{\sigma_0\sqrt{s}}\right) - 1\right\} ds, 0\right\} \\
= & \max\left\{NMVM_1\left(y_1 - x'\beta_1; \theta_1\right) + NMVM_1\left(y_0 - x'\beta_0; \theta_0\right) - 1, 0\right\} \\
= & C_L\left(NMVM_1\left(y_1 - x'\beta_1; \theta_1\right), NMVM_1\left(y_0 - x'\beta_0; \theta_0\right)\right),
\end{aligned}$$

where $\theta_1 = \left(\xi, \mu_1, \sigma_1^2, \zeta_1\right)$ and $\theta_0 = \left(\xi, \mu_0, \sigma_0^2, \zeta_0\right)$. Similarly, we can show that

$$NMVM_2\left(y_1 - x'\beta_1, y_0 - x'\beta_0; \left[\alpha_{10}^{-}, \rho_U\right]\right) = C_U\left(NMVM_1\left(y_1 - x'\beta_1; \theta_1\right), NMVM_1\left(y_0 - x'\beta_0; \theta_0\right)\right).$$

Proof of Theorem 3.2: It follows from (III.5) that

$$F_\Delta(\delta) = NMVM_1\left(\delta - x'\left(\beta_1 - \beta_0\right); \xi, \mu_1 - \mu_0, \gamma_2^2, \zeta_1 - \zeta_0\right), \tag{A.2}$$

$$
\begin{aligned}
& F_{\Delta}(\delta|D=1) \\
= & \int_{-\infty}^{\delta} \int_{-w'\gamma}^{\infty} \frac{nmvm_2\left(u - x'(\beta_1 - \beta_0), \epsilon; \xi, (B\mu)_{1\epsilon}, (B\Sigma B')_{1\epsilon}, (B\zeta)_{1\epsilon}\right)}{1 - NMVM_1\left(-w'\gamma; \xi, \mu_\epsilon, \sigma_\epsilon^2, \zeta_\epsilon\right)} du d\epsilon.
\end{aligned}
$$

Since the only unidentified parameter in $F_{\Delta}(\delta)$ and $F_{\Delta}(\delta|D=1)$ is γ_2^2, we get

$$
\begin{aligned}
F_{\Delta}^{L}(\delta) &= \min_{\sigma_L^2 \leq \gamma_2^2 \leq \sigma_U^2} NMVM_1\left(\delta - x'(\beta_1 - \beta_0); \xi, \mu_1 - \mu_0, \gamma_2^2, \zeta_1 - \zeta_0\right) \leq F_{\Delta}(\delta) \\
&\leq \max_{\sigma_L^2 \leq \gamma_2^2 \leq \sigma_U^2} NMVM_1\left(\delta - x'(\beta_1 - \beta_0); \xi, \mu_1 - \mu_0, \gamma_2^2, \zeta_1 - \zeta_0\right) = F_{\Delta}^{U}(\delta)
\end{aligned}
$$

and

$$
\begin{aligned}
& F_{\Delta}^{L}(\delta|D=1) \\
= & \min_{\sigma_L^2 \leq \gamma_2^2 \leq \sigma_U^2} \int_{-\infty}^{\delta} \int_{-w'\gamma}^{\infty} \frac{nmvm_2\left(u - x'(\beta_1 - \beta_0), \epsilon; \xi, (B\mu)_{1\epsilon}, (B\Sigma B')_{1\epsilon}, (B\zeta)_{1\epsilon}\right)}{1 - NMVM_1\left(-w'\gamma; \xi, \mu_\epsilon, \sigma_\epsilon^2, \zeta_\epsilon\right)} du d\epsilon \\
\leq & F_{\Delta}(\delta|D=1) \\
\leq & \max_{\sigma_L^2 \leq \gamma_2^2 \leq \sigma_U^2} \int_{-\infty}^{\delta} \int_{-w'\gamma}^{\infty} \frac{nmvm_2\left(u - x'(\beta_1 - \beta_0), \epsilon; \xi, (B\mu)_{1\epsilon}, (B\Sigma B')_{1\epsilon}, (B\zeta)_{1\epsilon}\right)}{1 - NMVM_1\left(-w'\gamma; \xi, \mu_\epsilon, \sigma_\epsilon^2, \zeta_\epsilon\right)} du d\epsilon \\
= & F_{\Delta}^{U}(\delta|D=1).
\end{aligned}
$$

If $\zeta_1 = \zeta_0$, then

$$
F_{\Delta}(\delta) = NMVM_1\left(\delta - x'(\beta_1 - \beta_0); \xi, \mu_1 - \mu_0, \gamma_2^2, 0\right) = \int_{0}^{\infty} f_S(s)\, \Phi\left(\frac{\delta - x'(\beta_1 - \beta_0)}{\gamma_2 \sqrt{s}}\right) ds.
$$

Thus, $F_{\Delta}(\delta)$ is an increasing function of γ_2 when $\delta < ATE\ (= x'(\beta_1 - \beta_0))$ and a decreasing function of γ_2 when $\delta \geq ATE$. We get the second result in (i).

Proof of Theorem 5.1: Since $(U_j, \epsilon) \sim NMVM_2\left(\xi, \mu_{j\epsilon}, \Sigma_{j\epsilon}, \zeta_{j\epsilon}\right)$, it follows that the conditional distribution function of U_j given ϵ is given by

$$
F_{j|\epsilon}(u_j|\epsilon) = \frac{\int \frac{f_S(s)}{\sigma_\epsilon \sqrt{s}} \phi\left(\frac{\epsilon - \mu_\epsilon(s)}{\sigma_\epsilon \sqrt{s}}\right) \Phi\left(\frac{u_j - \bar{\mu}_j(s)}{\sqrt{\left(1 - \rho_{j\epsilon}^2\right)\sigma_j^2 s}}\right) ds}{f_\epsilon(\epsilon)}, \ j = 0, 1,
$$

where

$$
\begin{aligned}
\mu_j(s) &= \zeta_j(s - E(S)),\ j = 0, 1, \epsilon, \\
\bar{\mu}_j(s) &= \mu_j(s) + \rho_{j\epsilon}\frac{\sigma_j}{\sigma_\epsilon}\epsilon,\ j = 0, 1.
\end{aligned}
$$

Let

$$
m_1(s) = \frac{y_1 - g_1(x_1, x_c) - \bar{\mu}_1(s)}{\sqrt{\left(1 - \rho_{1\epsilon}^2\right)\sigma_1^2 s}} \text{ and } m_0(s) = \frac{y_0 - g_0(x_0, x_c) - \bar{\mu}_0(s)}{\sqrt{\left(1 - \rho_{0\epsilon}^2\right)\sigma_0^2 s}}.
$$

If $\sigma_0\sqrt{\left(1-\rho_{0\epsilon}^2\right)}\zeta_1 = -\sigma_1\sqrt{\left(1-\rho_{1\epsilon}^2\right)}\zeta_0$, then

$$
\begin{aligned}
& m_1(s) + m_0(s) \\
= \ & \frac{\sigma_0\sqrt{\left(1-\rho_{0\epsilon}^2\right)}\left(y_1 - g_1(x_1, x_c) - \rho_{1\epsilon}\frac{\sigma_1}{\sigma_\epsilon}\epsilon\right) + \sigma_1\sqrt{\left(1-\rho_{1\epsilon}^2\right)}\left(y_0 - g_0(x_0, x_c) - \rho_{0\epsilon}\frac{\sigma_0}{\sigma_\epsilon}\epsilon\right)}{\sqrt{\left(1-\rho_{1\epsilon}^2\right)\left(1-\rho_{0\epsilon}^2\right)\sigma_1^2\sigma_0^2 s}},
\end{aligned}
$$

which will not change sign with respect to s, so $\Phi(m_1(s)) + \Phi(m_0(s)) - 1$ $(= \Phi(m_1(s)) + \Phi(-m_0(s)))$ will not change sign with respect to s. Thus, we obtain

$$
\begin{aligned}
& C_L\left(F_{1|\epsilon}(y_1 - g_1(x_1, x_c)), F_{0|\epsilon}(y_0 - g_0(x_0, x_c))\right) \\
= \ & \max\left\{\frac{\int \frac{f_S(s)}{\sigma_\epsilon\sqrt{s}}\phi\left(\frac{\epsilon - \mu_\epsilon(s)}{\sigma_\epsilon\sqrt{s}}\right)\{\Phi(m_1(s)) + \Phi(m_0(s)) - 1\}\, ds}{f_\epsilon(\epsilon)}, 0\right\} \\
= \ & \frac{\int \frac{f_S(s)}{\sigma_\epsilon\sqrt{s}}\phi\left(\frac{\epsilon - \mu_\epsilon(s)}{\sigma_\epsilon\sqrt{s}}\right)\max\{\Phi(m_1(s)) + \Phi(m_0(s)) - 1, 0\}\, ds}{f_\epsilon(\epsilon)} \\
= \ & \frac{\int \frac{f_S(s)}{\sigma_\epsilon\sqrt{s}}\phi\left(\frac{\epsilon - \mu_\epsilon(s)}{\sigma_\epsilon\sqrt{s}}\right)\Phi_{-1}(m_1(s), m_0(s))\, ds}{f_\epsilon(\epsilon)}.
\end{aligned}
\tag{A.3}
$$

Similarly, if $\sigma_0\sqrt{\left(1-\rho_{0\epsilon}^2\right)}\zeta_1 = \sigma_1\sqrt{\left(1-\rho_{1\epsilon}^2\right)}\zeta_0$, we have

$$
\begin{aligned}
& C_U\left(F_{1|\epsilon}(y_1 - g_1(x_1, x_c)), F_{0|\epsilon}(y_0 - g_0(x_0, x_c))\right) \\
= \ & \frac{\int \frac{f_S(s)}{(\sigma_\epsilon^2 s)^{1/2}}\phi\left(\frac{\epsilon - \mu_\epsilon(s)}{\sigma_\epsilon\sqrt{s}}\right)\Phi_{-1}(m_1(s), m_0(s))\, ds}{f_\epsilon(\epsilon)}.
\end{aligned}
\tag{A.4}
$$

Now suppose $Z = (Z_1, Z_0, Z_\epsilon)' \sim NMVM_3(\xi, \mu, \Sigma, \zeta)$. Then the density function of Z_ϵ is the same as that of ϵ and the conditional density function of Z_1, Z_0 given $Z_\epsilon = \epsilon$ is:

$$\begin{aligned} f_{Z_1,Z_0|Z_\epsilon=\epsilon}\left(z_1, z_0; \rho_{10|\epsilon}\right) &= \frac{\int \frac{f_S(s)}{(2\pi s)^{3/2}|\Sigma|^{1/2}} \exp\left\{-(z-\mu(s))' \Sigma^{-1}(z-\mu(s))/(2s)\right\} ds}{f_\epsilon(\epsilon)} \\ &= \frac{\int \frac{f_S(s)}{(s)^{3/2}\sigma_\epsilon|\Sigma_2|^{1/2}} \phi\left(\frac{\epsilon-\mu_\epsilon(s)}{\sigma_\epsilon\sqrt{s}}\right) \phi_{\rho_{10|\epsilon}}\left(\frac{z_1-\bar{\mu}_1(s)}{\sqrt{(1-\rho_{1\epsilon}^2)\sigma_1^2 s}}, \frac{z_0-\bar{\mu}_0(s)}{\sqrt{(1-\rho_{0\epsilon}^2)\sigma_0^2 s}}\right) ds}{f_\epsilon(\epsilon)}, \end{aligned}$$

where

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho_{10}\sigma_1\sigma_0 & \rho_{1\epsilon}\sigma_1\sigma_\epsilon \\ \rho_{10}\sigma_1\sigma_0 & \sigma_0^2 & \rho_{0\epsilon}\sigma_0\sigma_\epsilon \\ \rho_{1\epsilon}\sigma_1\sigma_\epsilon & \rho_{0\epsilon}\sigma_0\sigma_\epsilon & \sigma_\epsilon^2 \end{pmatrix}, \ \Sigma_2 = \begin{pmatrix} \left(1-\rho_{1\epsilon}^2\right)\sigma_1^2 & \left(\rho_{10} - \rho_{1\epsilon}\rho_{0\epsilon}\right)\sigma_1\sigma_0 \\ \left(\rho_{10} - \rho_{1\epsilon}\rho_{0\epsilon}\right)\sigma_1\sigma_0 & \left(1-\rho_{0\epsilon}^2\right)\sigma_0^2 \end{pmatrix},$$

$$\rho_{10|\epsilon} = \frac{\rho_{10} - \rho_{1\epsilon}\rho_{0\epsilon}}{\sqrt{\left(1-\rho_{1\epsilon}^2\right)\left(1-\rho_{0\epsilon}^2\right)}}, \ \mu(s) = \left(\mu_1(s), \mu_0(s), \mu_\epsilon(s)\right)'.$$

Thus the conditional distribution function of Z_1, Z_0 given $Z_\epsilon = \epsilon$ is

$$F_{Z_1,Z_0|Z_\epsilon=\epsilon}\left(z_1, z_0; \rho_{10|\epsilon}\right) = \frac{\int \frac{f_S(s)}{\sigma_\epsilon\sqrt{s}} \phi\left(\frac{\epsilon-\mu_\epsilon(s)}{\sigma_\epsilon\sqrt{s}}\right) \Phi_{\rho_{10|\epsilon}}\left(\frac{z_1-\bar{\mu}_1(s)}{\sqrt{(1-\rho_{1\epsilon}^2)\sigma_1^2 s}}, \frac{z_0-\bar{\mu}_0(s)}{\sqrt{(1-\rho_{0\epsilon}^2)\sigma_0^2 s}}\right) ds}{f_\epsilon(\epsilon)}. \quad (A.5)$$

Comparing $(A.3)$ and $(A.4)$ with $(A.5)$, we obtain: if $\sigma_0\sqrt{\left(1-\rho_{0\epsilon}^2\right)}\zeta_1 = -\sigma_1\sqrt{\left(1-\rho_{1\epsilon}^2\right)}\zeta_0$, then

$$\begin{aligned} F_{10}^L(y_1, y_0) &= \int C_L\left(F_{1|\epsilon}\left(y_1 - g_1(x_1, x_c)\right), F_{0|\epsilon}\left(y_0 - g_0(x_0, x_c)\right)\right) dF_\epsilon(\epsilon) \\ &= \int F_{Z_1,Z_0|Z_\epsilon=\epsilon}\left(y_1 - g_1(x_1, x_c), y_0 - g_0(x_0, x_c); -1\right) dF_\epsilon(\epsilon) \\ &= F_{Z_1,Z_0}\left(y_1 - g_1(x_1, x_c), y_0 - g_0(x_0, x_c); \rho_L\right) \\ &= NMVM_2\left(y_1 - g_1(x_1, x_c), y_0 - g_0(x_0, x_c); \left[\alpha_{10}^-, \rho_L\right]\right), \end{aligned}$$

where the last equality follows, as $(Z_1, Z_0)' \sim NMVM_2\left(\left[\alpha_{10}^-, \rho_{10}\right]\right)$; If $\sigma_0\sqrt{\left(1-\rho_{0\epsilon}^2\right)}\zeta_1 =$

$\sigma_1\sqrt{\left(1-\rho_{1\epsilon}^2\right)}\zeta_0$, then

$$\begin{aligned}
F_{10}^{U}\left(y_1,y_0\right) &= \int C_U\left(F_{1|\epsilon}\left(y_1-g_1\left(x_1,x_c\right)\right),F_{0|\epsilon}\left(y_0-g_0\left(x_0,x_c\right)\right)\right)dF_\epsilon\left(\epsilon\right) \\
&= \int F_{Z_1,Z_0|Z_\epsilon=\epsilon}\left(y_1-g_1\left(x_1,x_c\right),y_0-g_0\left(x_0,x_c\right);1\right)dF_\epsilon\left(\epsilon\right) \\
&= F_{Z_1,Z_0}\left(y_1-g_1\left(x_1,x_c\right),y_0-g_0\left(x_0,x_c\right);\rho_U\right) \\
&= NMVM_2\left(y_1-g_1\left(x_1,x_c\right),y_0-g_0\left(x_0,x_c\right);\left[\alpha_{10}^{-},\rho_U\right]\right).
\end{aligned}$$

So when $\zeta_1=\zeta_0=0$, we have

$$\begin{aligned}
F_{10}^{L}\left(y_1,y_0\right) &= NMVM_2\left(y_1-g_1\left(x_1,x_c\right),y_0-g_0\left(x_0,x_c\right);\left[\alpha_{10}^{-},\rho_L\right]\right), \\
F_{10}^{U}\left(y_1,y_0\right) &= NMVM_2\left(y_1-g_1\left(x_1,x_c\right),y_0-g_0\left(x_0,x_c\right);\left[\alpha_{10}^{-},\rho_U\right]\right).
\end{aligned}$$

Chapter IV

Variable Definition

Market capitalization: the number of shares outstanding times the price at the end of the most recent month prior to the issue.

Market-to-book: market capitalization plus total assets minus book value of equity minus deferred taxes at the end of the most recent fiscal year prior to the issue, divided by total book assets.

Relative offer size: number of shares issued relative to the number of shares outstanding prior to the issue.

Runup: buy-and-hold return in the 90 trading days prior to the issue net of the value-weighted market index.

Average turnover: average daily share volume scaled by the number of shares outstanding in the [-390, -30] window relative to the offer date.

Return on assets: income before extraordinary items scaled by book assets.

Table 1. Bias of Estimators of ATE and TT

Data	% Bias of ATE			% Bias of TT		True TT
$\zeta = (\zeta_{U1}, \zeta_{U0}, \zeta_\epsilon)$						
Category 1	Skew t	Gaussian	No Correction	Skew t	Gaussian	
$(0,0,0)$	0.22807	-1.454	50.228	0.53729	-0.97482	1.4038
$(0,0,1)$	2.2377	0.062898	27.907	3.5022	0.069725	1.2289
$(0,0,-1)$	0.67477	-0.54789	26.223	-0.07958	-0.79053	1.207
Category 2						
$(1,0,0)$	0.73928	-10.521	27.422	0.36456	0.33516	1.1687
$(-1,0,0)$	0.19997	10.875	27.196	-0.02432	0.076538	1.1672
$(0,1,0)$	-0.9086	9.8268	47.241	-0.6891	15.879	1.4292
$(0,-1,0)$	-1.6763	-12.037	47.766	-2.0352	-15.674	1.4313
Category 3						
$(1,1,0)$	0.54094	-0.74301	24.294	0.90689	18.421	1.1944
$(-1,-1,0)$	0.75034	-1.3118	24.275	1.422	-19.356	1.1937
$(1,-1,0)$	0.64949	-22.218	25.024	1.9992	-17.473	1.1975
$(-1,1,0)$	-1.1989	22.219	24.167	-1.4086	20.008	1.1924
Category 4						
$(1,1,1)$	-0.1968	-1.2218	47.422	0.20834	20.488	1.1099
$(1,-1,1)$	1.2981	-22.424	15.928	1.2922	-15.873	1.4778

Table 2. Bias of Parameters in SRM (multiplied by 100)

Data	β_1		β_0		γ_0		γ_1	
$(\zeta = (\zeta_{U1}, \zeta_{U0}, \zeta_\epsilon))$								
Category 1	Skew t	Gaussian	Skew t	Gaussian	Skew t	Gaussian	Skew t	Gaussian
$(0,0,0)$	0.011951	-0.63355	-0.21612	0.82047	0.029441	-0.41827	-2.8651	11.258
$(0,0,1)$	0.21275	0.013267	-2.025	-0.0496	-0.44802	-10.945	0.96954	59.771
$(0,0,-1)$	0.83001	0.053632	0.15524	0.60153	0.34164	10.356	-0.02209	58.377
Category 2								
$(1,0,0)$	0.46302	-10.763	-0.27626	-0.2418	-0.25278	-0.31216	-3.3182	10.847
$(-1,0,0)$	0.18102	10.799	-0.01895	-0.0758	0.6213	0.39025	-2.8652	11.254
$(0,1,0)$	-0.46955	-1.5318	0.43905	-11.359	-0.24259	-0.43806	-4.136	10.944
$(0,-1,0)$	0.014604	-0.60671	1.6909	11.43	-0.20528	-0.39002	-3.7125	10.927
Category 3								
$(1,1,0)$	-0.20199	-11.948	-0.74293	-11.205	0.19988	0.090188	-3.6397	10.993
$(-1,-1,0)$	-0.1156	10.179	-0.86594	11.491	-0.4089	-0.47814	-2.8559	11.572
$(1,-1,0)$	-0.30633	-11.519	-0.95582	10.699	0.4359	0.091921	-3.6534	10.588
$(-1,1,0)$	-0.38346	10.309	0.81539	-11.91	0.01114	-0.02306	-3.4427	10.994
Category 4								
$(1,1,1)$	-0.39261	-12.073	-0.19582	-10.851	-0.44793	-10.679	2.5966	60.718
$(1,-1,1)$	0.3948	-11.382	-0.90328	11.042	-0.07596	-10.045	0.6578	59.671

Table 3. The number of offerings and total proceeds

	Number of offers			Proceeds ($mil)		
	Shelf		Traditional	Shelf		Traditional
Year	Non-Accelerated	Accelerated		Non-Accelerated	Accelerated	
1997	25	4	225	3492.5	597.7	18156.6
1998	35	4	138	4426.9	453.2	13880.8
1999	35	8	159	8001.3	4047.7	28656
2000	42	15	175	15007.8	2688.9	34733.3
2001	57	29	124	10535.8	5381.6	21158.5
2002	56	34	121	11546.4	5579.8	15942.2
2003	56	77	133	8162.8	14156.3	14916.9
2004	54	156	119	7948.9	31681	13920
2005	46	115	90	6325.8	19307.4	15415.1
2006	76	79	90	12418.3	14900.5	17160.6
Total	482	521	1,374	87,867	98,794	193,940

Table 4. Firms and offer characteristics (mean)

	Shelf		Traditional
	Non-Accelerated	Accelerated	
Proceeds	182.3	189.6	141.1
Relative Offer Size	0.149	0.099	0.133
Market capitalization ($mil)	2305.2	3656.4	1452.1
Total Assets ($mil)	3661.5	5899.7	1575.2
Market-to-Book	1.398	1.305	2.398
Residual Volatility	0.029	0.026	0.038
Turnover	0.007	0.008	0.008
Return on Assets	0.039	0.052	0.081
SP500	0.129	0.142	0.044
Percentage of Secondary Shares	0.119	0.239	0.307
Simultaneous Offer	0.039	0.046	0.005
Convertible Debt Issued Prior	0.124	0.587	0.399
Convertible Debt Issued After	0.122	0.201	0.238

Table 5. Determinants of the choice of flotation (Accelerated=1)

	Skewed t		Heckman	
	estimates	t-stat	estimates	t-stat
Intercept	0.520	9.600	0.710	5.070
Relative offer size	-0.831	-3.145	-1.348	-4.270
Ln(Market capitalization)	0.016	0.916	0.084	1.780
Nasdaq dummy	0.011	0.286	0.030	0.262
Residual volatility	-5.033	-3.282	-13.612	-3.723
Average turnover	5.422	3.366	18.778	2.944
Runup	0.046	1.419	0.179	1.401
Return on assets	-0.266	-3.030	-0.873	-3.314
Market to book	-0.015	-1.182	-0.058	-1.544
SP500	-0.027	-0.565	-0.087	-0.549
Secondary Percentage	0.175	4.648	0.644	5.162
Simultaneous offer	-0.069	-0.987	-0.105	-0.472
Convertible debt issued prior	-0.082	-2.519	-0.275	-2.525
Skewness parameter	2.305	2.683		
Log-Likelihood	614.620		620.964	
Number of obs	521		521	

Table 6. Estimates of the ROA equation for the accelerated underwriting

	Skewed t		Normal	
	estimates	t-stat	estimates	t-stat
Intercept	-0.166	-3.271	-0.076	-2.338
Ln(Market capitalization)	0.030	3.176	0.022	2.392
Nasdaq dummy	0.046	2.517	0.044	2.339
Turnover	2.436	2.095	2.157	1.825
Runup	-0.021	-0.866	-0.019	-0.755
Return on assets	0.619	13.190	0.632	12.822
Market to Book	-0.006	-1.077	-0.005	-0.873
SP 500	-0.052	-1.897	-0.043	-1.573
Secondary Percentage	0.059	2.136	0.031	1.153
Simultaneous offer	-0.083	-2.323	-0.065	-1.853
Convertible debt issued prior	-0.046	-2.715	-0.036	-2.034
Convertible debt issued after	0.031	1.965	0.030	1.938
Correction term 1	0.133	2.394	0.051	1.119
Correction term 2	-0.125	-1.829		
Number of obs	249		249	

Table 7. Estimates of the ROA equation for the non-accelerated underwriting

	Skewed t		Normal	
	estimates	t-stat	estimates	t-stat
Intercept	-0.035	-0.360	0.00	-0.07
Ln(Market capitalization)	0.012	1.478	0.01	1.38
Nasdaq dummy	-0.023	-1.466	-0.02	-1.40
Turnover	-0.247	-0.313	-0.20	-0.24
Runup	0.023	0.889	0.03	0.97
Return on assets	0.479	11.943	0.48	11.03
Market to Book	0.002	0.395	0.00	0.33
SP 500	-0.008	-0.351	-0.01	-0.33
Secondary Percentage	-0.006	-0.203	-0.01	-0.24
Simultaneous offer	-0.022	-0.621	-0.02	-0.59
Convertible debt issued prior	0.029	1.649	0.03	1.66
Convertible debt issued after	-0.028	-1.891	-0.03	-1.86
Correction term 1	0.022	0.636	0.03	0.67
Correction term 2	-0.089	-0.316		
Number of obs	272		272	

Table 8. Estimates of the CAR equation for the accelerated underwriting

	Skewed t		Normal	
	estimates	t-stat	estimates	t-stat
Intercept	0.071	0.431	-0.020	-0.193
Ln(Market capitalization)	-0.035	-1.152	-0.029	-0.974
Nasdaq dummy	-0.037	-0.612	-0.033	-0.543
Turnover	-6.181	-1.629	-5.841	-1.530
Runup	-0.269	-3.367	-0.268	-3.365
Return on assets	0.539	3.524	0.528	3.317
Market to Book	0.013	0.647	0.011	0.568
SP 500	0.120	1.350	0.114	1.286
Secondary Percentage	-0.003	-0.031	0.023	0.269
Simultaneous offer	-0.152	-1.306	-0.164	-1.434
Convertible debt issued prior	-0.027	-0.480	-0.036	-0.637
Convertible debt issued after	0.120	2.373	0.122	2.392
Correction term 1	-0.124	-0.682	-0.011	-0.077
Correction term 2	0.159	0.715		
Number of obs	249		249	

Table 9. Estimates of the CAR equation for the non-accelerated underwriting

	Skewed t		Normal	
	estimates	t-stat	estimates	t-stat
Intercept	0.093	0.261	-0.26	-1.51
Ln(Market capitalization)	-0.018	-0.605	-0.01	-0.46
Nasdaq dummy	-0.015	-0.267	-0.03	-0.48
Turnover	-3.021	-1.060	-2.63	-0.85
Runup	0.043	0.454	0.03	0.31
Return on assets	0.479	3.315	0.46	2.91
Market to Book	-0.006	-0.345	-0.01	-0.38
SP 500	0.088	1.029	0.08	0.95
Secondary Percentage	-0.009	-0.087	0.03	0.25
Simultaneous offer	0.046	0.350	0.04	0.34
Convertible debt issued prior	0.145	2.324	0.13	2.04
Convertible debt issued after	0.080	1.523	0.08	1.43
Correction term 1	-0.172	-1.354	-0.09	-0.53
Correction term 2	1.108	1.090		
Number of obs	272		272	

Table 10. Average Treatment Effects

	Skewed t		Heckman	
	estimates	t-statistics	estimates	t-statistics
	Industry adjusted ROA a year after issuance			
Average Treatment Effect	-0.129	-1.162	-0.068	-1.294
Treatment Effect for the Treated	0.021	0.111	-0.047	-0.621
Treatment Effect for the non-Treated	-0.301	-2.635	-0.090	-1.236
	Cumulative abnormal return a year after issuance			
Average Treatment Effect	-0.181	-0.463	0.089	0.490
Treatment Effect for the Treated	-0.560	-0.818	0.155	0.574
Treatment Effect for the non-Treated	0.227	0.608	0.020	0.084
	Cumulative abnormal return 6 months after issuance			
Average Treatment Effect	-0.151	-0.536	0.128	1.005
Treatment Effect for the Treated	-0.235	-0.472	0.237	1.209
Treatment Effect for the non-Treated	-0.069	-0.280	0.014	0.086
	Cumulative abnormal return 3 months after issuance			
Average Treatment Effect	0.026	0.143	0.046	0.547
Treatment Effect for the Treated	0.188	0.581	0.205	1.618
Treatment Effect for the non-Treated	-0.154	-0.916	-0.121	-1.144
	Cumulative abnormal return 3 days after issuance			
Average Treatment Effect	0.048	0.709	-0.016	-0.517
Treatment Effect for the Treated	0.085	0.717	-0.032	-0.683
Treatment Effect for the non-Treated	0.011	0.175	0.001	0.019

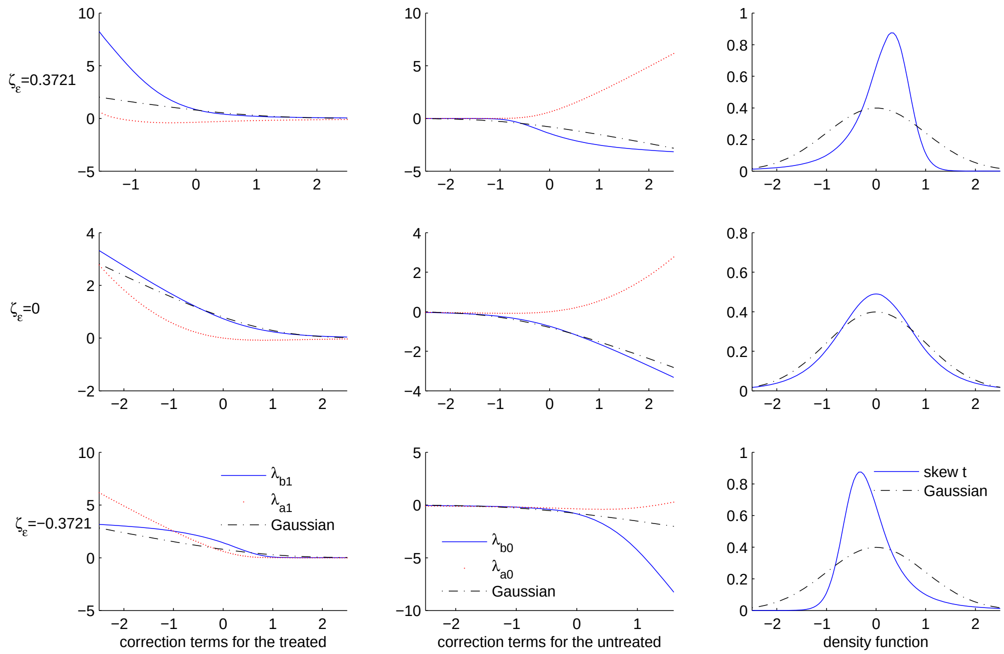

Figure 1. Sample correction terms for skewed t model and Gaussian model, and density functions of skew t distribution and Gaussian distribution with different values of the skewness parameter in the selection error

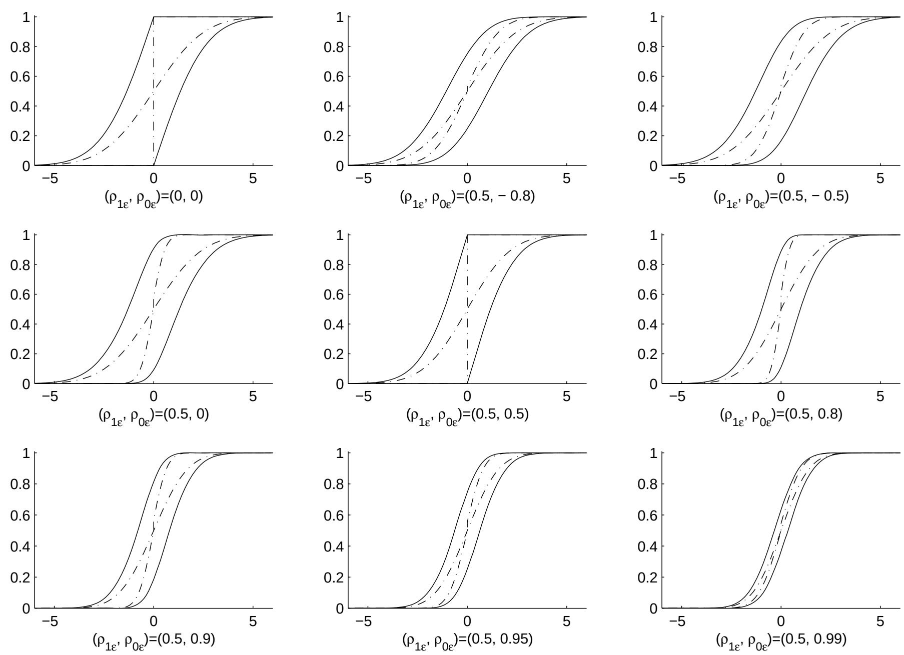

Figure 2. Sharp bounds on the distribution of treatment effects, $\sigma_1 = \sigma_0 = 1$. Solid curves are bounds assuming bivariate normality for (U_{ji}, ϵ_i), $j = 1, 0$, and dashed curves are bounds under the trivariate normality assumption for $(U_{1i}, U_{0i}, \epsilon_i)$.

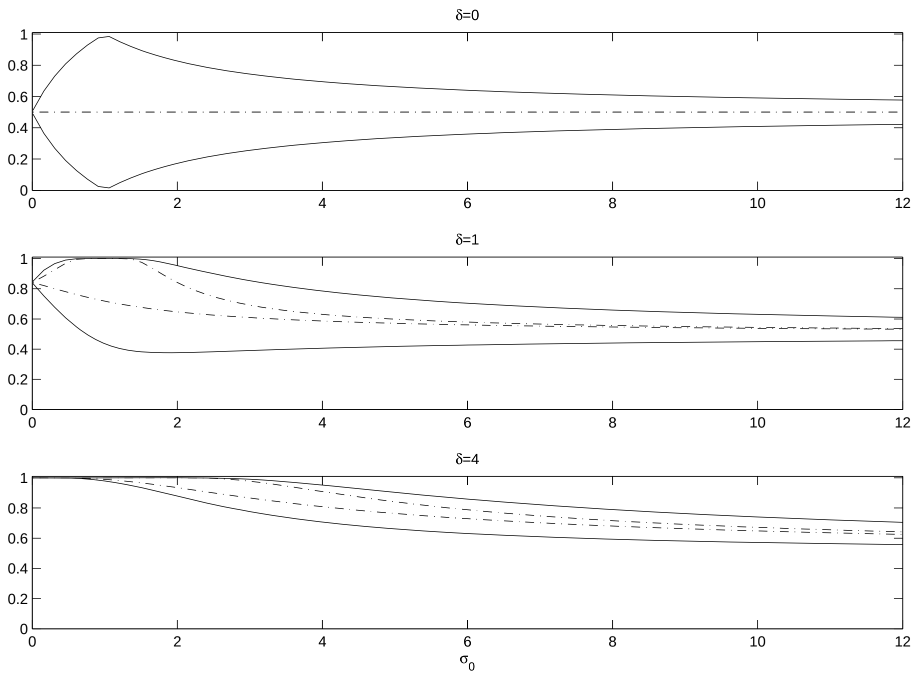

Figure 3. Treatment effect at a given δ — $\sigma_1 = 1$, $\rho_{1\varepsilon} = 0.5$, and $\rho_{0\varepsilon} = 0.5$. Solid curves are bounds assuming bivariate normality for (U_{ji}, ϵ_i), $j = 1, 0$, and dashed curves are bounds under the trivariate normality assumption for $(U_{1i}, U_{0i}, \epsilon_i)$.

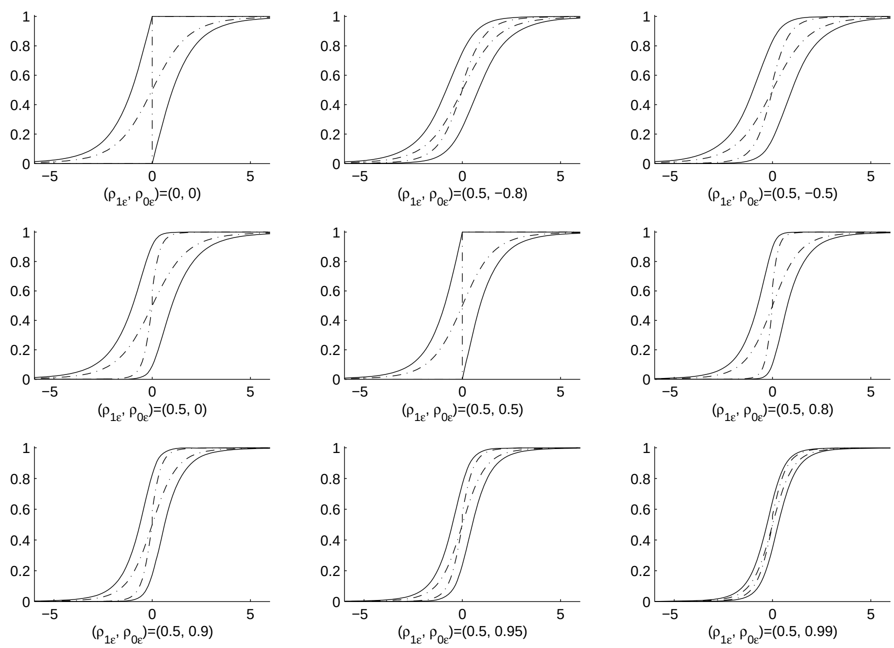

Figure 4. Sharp bounds on the distribution of treatment effects, $\sigma_1 = \sigma_0 = 1$. Solid curves are bounds assuming (U_{ji}, ϵ_i), $j = 0, 1$, follows bivariate student's t distribution with 4 degrees of freedom and dashed curves are bounds assuming $(U_{1i}, U_{0i}, \epsilon_i)$ follows trivariate student's t distribution with 4 degrees of freedom.

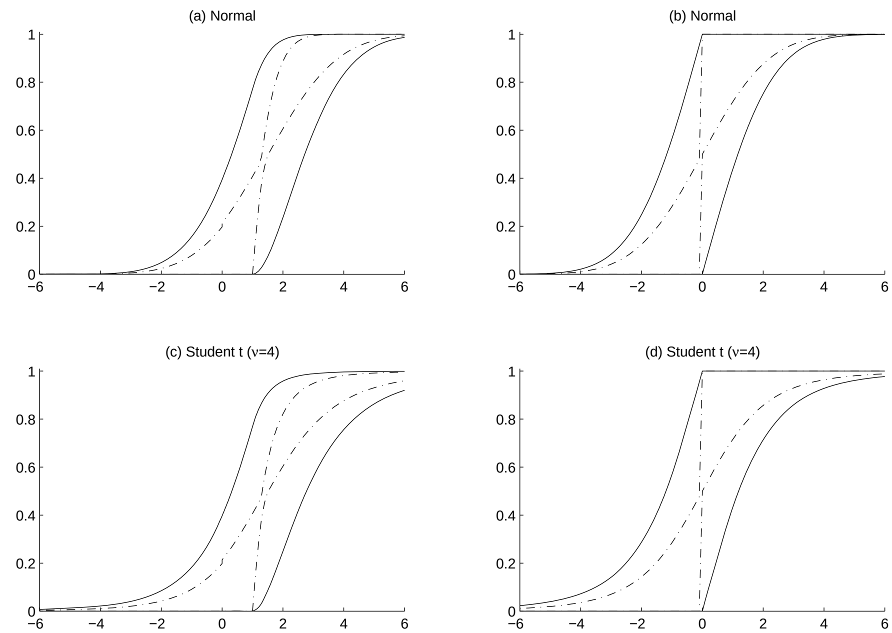

Figure 5. Sharp bounds on the distribution of treatment effects for the treated — $ATE = 0$, $\sigma_1 = \sigma_0 = 1$, and the Propensity Score $= 0.1$. In (a) and (c), $\rho_{1\varepsilon} = 0.5$ and $\rho_{0\varepsilon} = -0.5$, while in (b) and (d), $\rho_{1\varepsilon} = \rho_{0\varepsilon} = 0.5$.

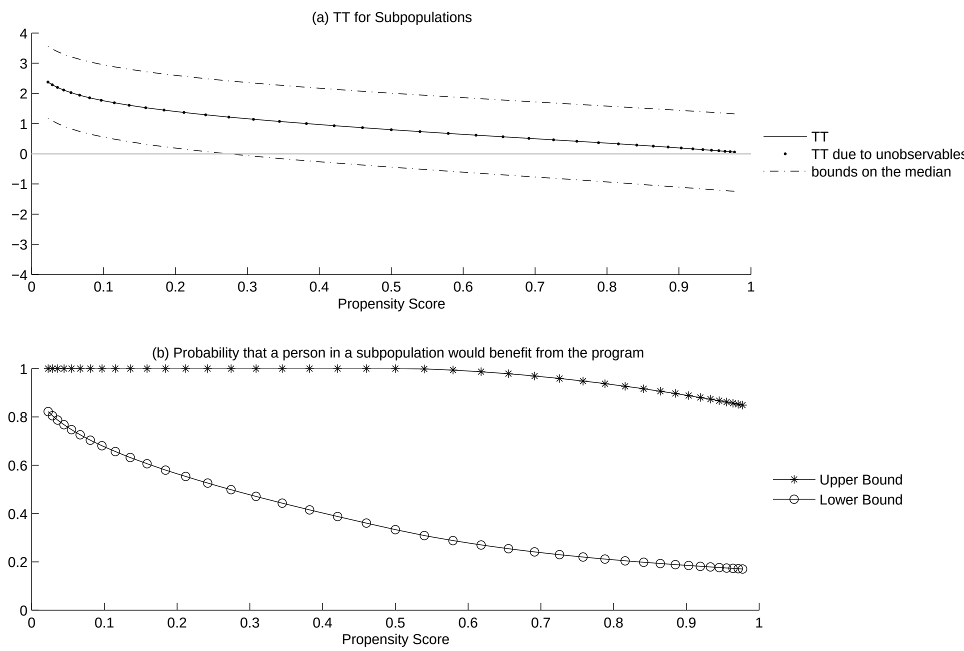

Figure 6. Treatment effect for the treated for subpopulations and the probability that a person in a subpopulation benefits from the treatment, where $ATE = 0$, $\rho_{1\varepsilon} = 0.5$, $\rho_{0\varepsilon} = -0.5$, and $\sigma_1 = \sigma_0 = 1$.

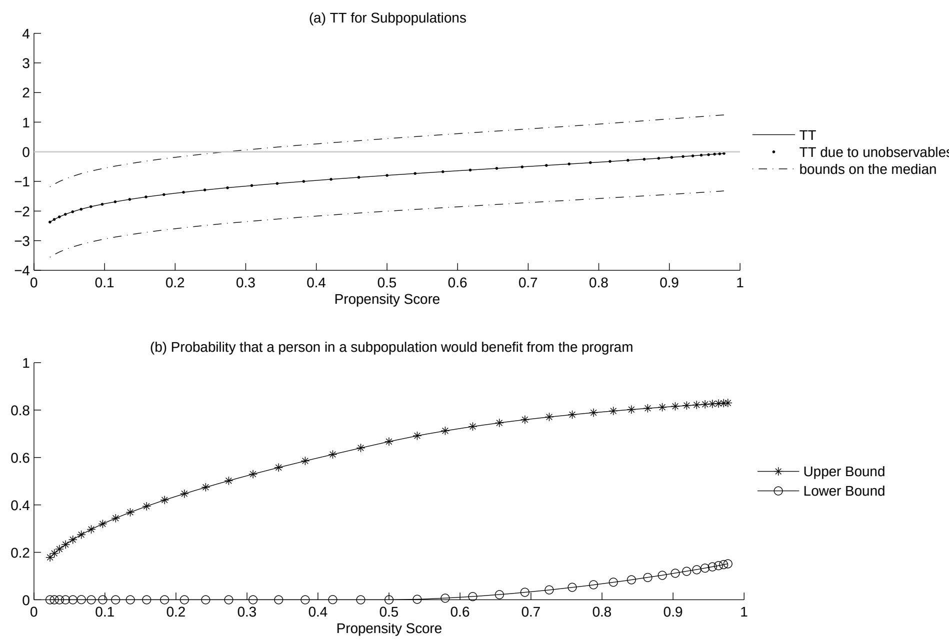

Figure 7. Treatment effect for the treated for subpopulations and the probability that a person in a subpopulation benefits from the treatment, where $ATE = 0$, $\rho_{1\varepsilon} = -0.5$, $\rho_{0\varepsilon} = 0.5$, and $\sigma_1 = \sigma_0 = 1$.

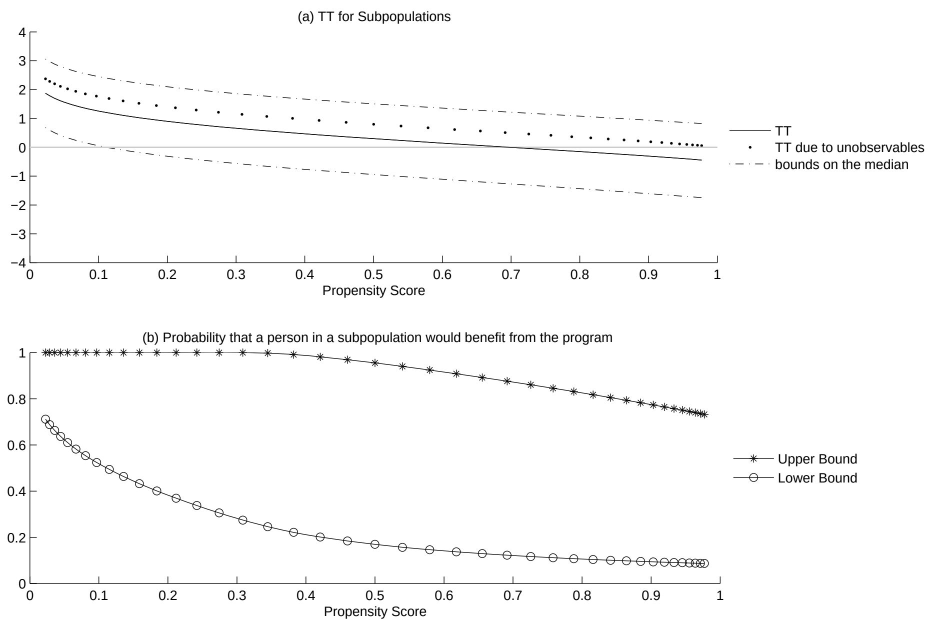

Figure 8. Treatment effect for the treated for subpopulations and the probability that a person in a subpopulation benefits from the treatment, where $ATE = -0.5$, $\rho_{1\varepsilon} = 0.5$, $\rho_{0\varepsilon} = -0.5$, and $\sigma_1 = \sigma_0 = 1$.

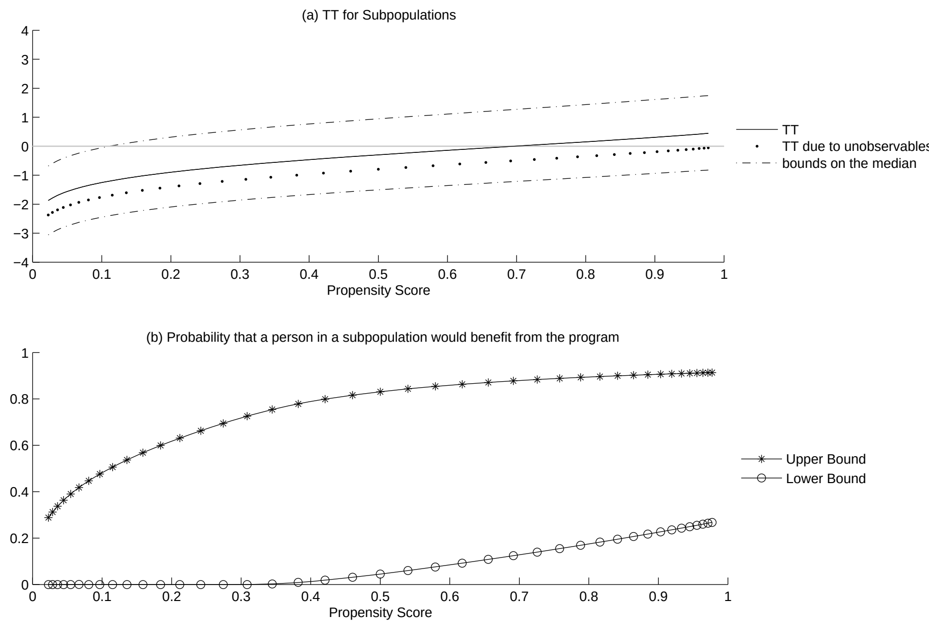

Figure 9. Treatment effect for the treated for subpopulations and the probability that a person in a subpopulation benefits from the treatment, where $ATE = 0.5$, $\rho_{1\varepsilon} = -0.5$, $\rho_{0\varepsilon} = 0.5$, and $\sigma_1 = \sigma_0 = 1$.

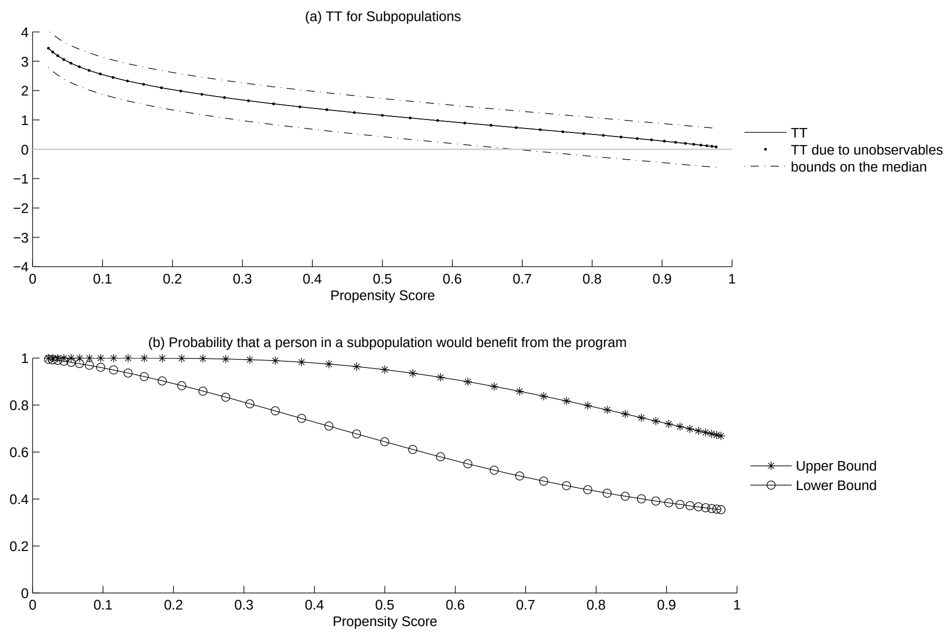

Figure 10. Treatment effect for the treated for subpopulations and the probability that a person in a subpopulation benefits from the treatment, where $ATE = 0$, $\rho_{1\varepsilon} = 0.95$, $\rho_{0\varepsilon} = -0.5$, and $\sigma_1 = \sigma_0 = 1$.

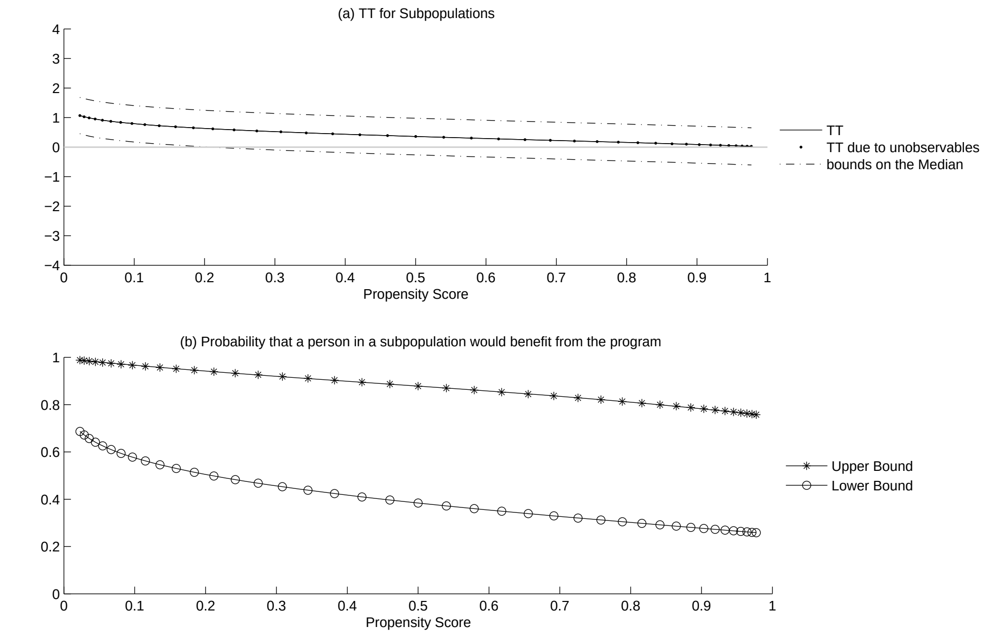

Figure 11. Treatment effect for the treated for subpopulations and the probability that a person in a subpopulation benefits from the treatment, where $ATE = 0$, $\rho_{1\varepsilon} = 0.95$, $\rho_{0\varepsilon} = 0.5$, and $\sigma_1 = \sigma_0 = 1$.

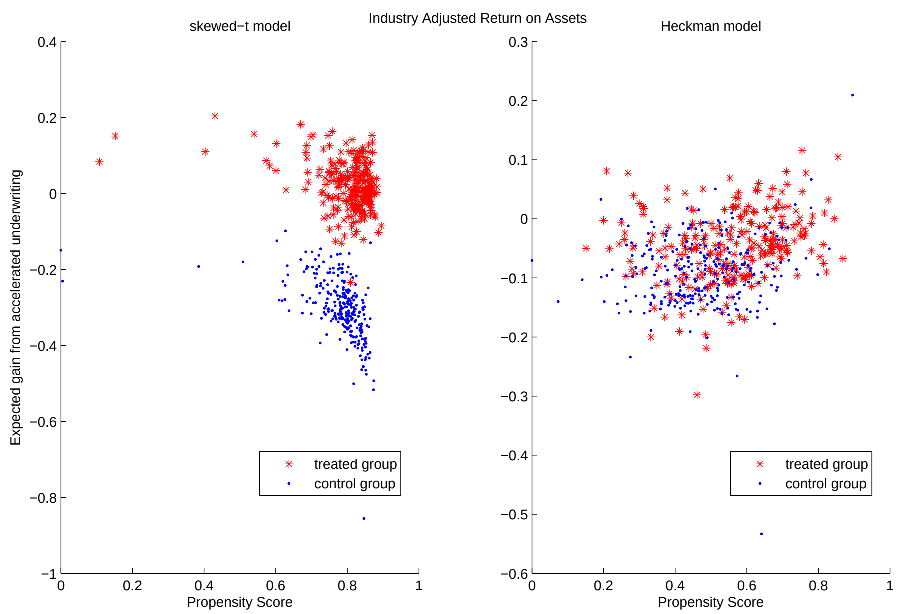

Figure 12. Expected gain from choosing accelerated underwriting in terms of return on assets one year after the equity issuance

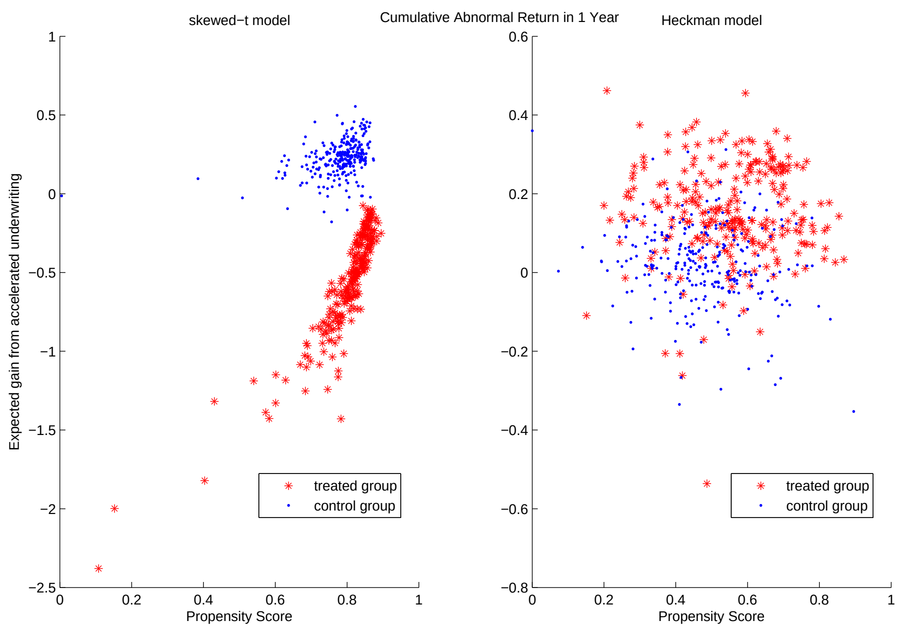

Figure 13. Expected gain from choosing accelerated underwriting in terms of cumulative abnormal return one year after the equity issuance

BIBLIOGRAPHY

[1] Aakvik, A. , J. Heckman, and E. Vytlacil (2005), "Treatment Effects for Discrete Outcomes When Responses to Treatment Vary Among Observationally Identical Persons: An Application to Norwegian Vocational Rehabilitation Programs," *Journal of Econometrics* 125, 15-51.

[2] Abadie, A., J. Angrist, and G. Imbens (2002), "Instrumental Variables Estimation of Quantile Treatment Effects," *Econometrica* 70, 91-117.

[3] Ai, C. (1997), "A Semiparametric Maximum Likelihood Estimator," *Econometrica* 65, 933-963.

[4] Alsina, C. (1981), "Some Functional Equations in the Space of Uniform Distribution Functions," *Equationes Mathematicae* 22, 153-164.

[5] Andrews, D.W.K. and P. Guggenberger (2007), "Validity of Subsampling and 'Plug-in Asymptotic' Inference for Parameters Defined by Moment Inequalities," Unpublished Manuscript, Cowles Foundation, Yale University.

[6] Andrews, D. W. K. and M. M. A. Schafgans (1998), "Semiparametric Estimation of the Intercept of a Sample Selection Model," *Review of Economic Studies* 65, 497-517.

[7] Andrews, D.W.K. and G. Soares (2007), "Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection," Manuscript, Yale University.

[8] Autore, D., Kumar, R., Shome, D.K. (2008), "The revival of shelf-registered corporate equity offerings," Journal of Corporate Finance 14, 32-50.

[9] Autore, D., Hutton, I., Kovacs, T. (2009), "Certification, Firm Quality, and Seasoned Equity Offerings," SSRN working paper.

[10] Bethal. J., Krigman, L. (2009), "Managing the costs of issuing common equity: The role of registration choice," Quarterly Journal of Finance and accounting 47, 35-57.

[11] Biddle, J., L. Boden and R. Reville (2003), "A Method for Estimating the Full Distribution of a Treatment Effect, With Application to the Impact of Workfare Injury on Subsequent Earnings." Mimeo.

[12] Bjorklund, A. and R. Moffitt (1987), "The Estimation of Wage Gains and Welfare Gains in Self-Selection Models," *Review of Economics and Statistics* 69, 42-49.

[13] Bortolotti, B., Megginson, W.L., Smart, S.B.(2008), "The rise of accelerated seasoned equity underwritings," Journal of Applied Corporate Finance 20, 35-57.

[14] Bugni, F. A. (2007), "Bootstrap Inference in Partially Identified Models," mimeo, Northwestern University.

[15] Canay, I.A. (2007), "EL Inference for Partially Identified Models: Large Deviations Optimality and Bootstrap Validity," Manuscript, University of Wisconsin.

[16] Carneiro, P. , K. T. Hansen, and J. Heckman (2003), "Estimating Distributions of Treatment Effects With an Application to the Returns to Schooling and Measurement of the Effects of Uncertainty on College Choice," *International Economic Review* 44(2), 361-422.

[17] Carneiro, P. and S. Lee (2008), "Estimating Potential Outcome Distributions in Local Instrumental Variables Models with an Application to Changes in College Enrollment and Wage Inequality," Working paper, University College London.

[18] Chen, S. (2006), "Nonparametric Identification and Estimation of Truncated Regression Models," Working Paper, The Hong Kong University of Science and Technology.

[19] Chen, X., H. Hong, and A. Tarozzi (2004), "Semiparametric Efficiency in GMM Models of Nonclassical Measurement Errors, Missing Data and Treatment Effects," Working paper.

[20] Chen, S. and Y. Zhou (2006), "Semiparametric and Nonparametric Estimation of Sample Selection Models Under Symmetry," Working Paper, The Hong Kong University of Science and Technology.

[21] Chernozhukov, V. and C. Hansen (2005), "An IV Model of Quantile Treatment Effects," *Econometrica* 73, 245-261.

[22] Chernozhukov, V., H. Hong, and E. Tamer (2007), "Parameter Set Inference in a Class of Econometric Models," *Econometrica* 75, 1243-1284.

[23] Chib, S. (2005), "Analysis of Treatment Response Data Without the Joint Distribution of Potential Outcomes," Manuscript, Washington University in St. Louis.

[24] Cunha, F. and J. Heckman (2007), "A New Framework for the Analysis of Inequality," IZA DP No. 2565.

[25] Das, M., W. Newey, and F. Vella (2003), "Nonparametric Estimation of Sample Selection Models," *Review of Economic Studies* 70, 33-58.

[26] Denuit, M. , C. Genest, and E. Marceau (1999), "Stochastic Bounds on Sums of Dependent Risks," *Insurance: Mathematics and Economics* 25, 85-104.

[27] Eckbo, E., Masulis, R., Norli, O. (2007), "Security Offerings," Handbook of Corporate F: Empirical Corporate Finance, Vol. 1.

[28] Fan, Y. (2005), "Sharp Correlation Bounds and Their Applications," Mimeo.

[29] Fan, Y. and S. Park (2006), "Sharp Bounds on the Distribution of the Treatment Effect and Their Statistical Inference," Manuscript, Vanderbilt University.

[30] Fan, Y. and S. Park (2007), "Confidence Sets for Some Partially Identified Parameters," Manuscript, Vanderbilt University.

[31] Fan, Y. and S. Park (2008), "Confidence Sets for the Quantile of Treatment Effects," Manuscript, Vanderbilt University.

[32] Firpo, S. (2007), "Efficient Semiparametric Estimation of Quantile Treatment Effects," *Econometrica* 75, 259-276.

[33] Frank, M. J. , R. B. Nelsen, and B. Schweizer (1987), "Best-Possible Bounds on the Distribution of a Sum—a Problem of Kolmogorov," *Probability Theory and Related Fields* 74, 199-211.

[34] Gallant, R. and D. Nychka (1987), "Semi-Nonparametric Maximum Likelihood Estimation," *Econometrica* 55, 363-390.

[35] Heckman, J. J. (1976), "The Common Structure of Statistical Models of Truncation, Sample Selection, and Limited Dependent Variables and a Simple Estimator for Such Models," *Annals of Economic and Social Measurement* 5, 475–492.

[36] —(1990), "Varieties of Selection Bias," *American Economic Review, Papers and Proceedings* 80, 313-318.

[37] — (1997), "Instrumental Variables: A Study of Implicit Behavioral Assumptions Used in Making Program Evaluations," *Journal of Human Resources* 32, 441-462.

[38] Heckman, J. J. and Bo E. Honore (1990), "The Empirical Content of the Roy Model," *Econometrica* 58, 1121-1149.

[39] Heckman, J. J. , J. Smith, and N. Clements (1997), "Making The Most Out Of Programme Evaluations and Social Experiments: Accounting For Heterogeneity in Programme Impacts," *Review of Economic Studies* 64, 487-535.

[40] Heckman, J., J. L. Tobias, and E. Vytlacil (2003), "Simple Estimators for Treatment Parameters in a Latent Variable Framework," *Review of Economics and Statistics* 85, 748-755.

[41] Heckman, J. and E. Vytlacil (1999), "Local Instrumental Variables and Latent Variable Models for Identifying and Bounding Treatment Effects," *Proceedings of the National Academy of Sciences* 96, 4730–4734.

[42] — (2000a), "The Relationship between Treatment Parameters within a Latent Variable Framework," *Economics Letters* 66, 33–39.

[43] — (2000b), "Local InstrumentalVariables," in C. Hsiao, K. Morimune, and J. Powell (Eds.), *Nonlinear Statistical Inference: Essays in Honor of Takeshi Amemiya* (Cambridge, U.K.: Cambridge University Press).

[44] — (2005), "Structural Equations, Treatment Effects, and Econometric Policy Evaluation," *Econometrica* 73, 669-738.

[45] Holland, P., (1986), "Statistics and Causal Inference," *Journal of the American Statistical Association*, 81, 945-960.

[46] Imbens, G. and J. Angrist (1994), "Identification and Estimation of Local Average Treatment Effects," *Econometrica* 62, 467-476.

[47] Imbens, G. W. and C. F. Manski (2004), "Confidence Intervals For Partially Identified Parameters." *Econometrica* 72, 1845–1857.

[48] Imbens, G. W. and W. Newey (2005), "Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity," Working Paper.

[49] Joe, H. (1997). *Multivariate Models and dependence Concepts*. Chapman & Hall/CRC, London.

[50] Jogensen, B. (1982). *Statistical Properties of the Generalized Inverse Gaussian Distribution*, Lecture Notes in Statistics 9. New York–Berlin: Springer-Verlag.

[51] Koop, G. and D. J. Poirier (1997), "Learning About the Across-Regime Correlation in Switching Regression Models," *Journal of Econometrics* 78, 217-227.

[52] Lee, L. F. (1982), "Some Approaches to the Correction of Selectivity Bias," *Review of Economic Studies* XLIX, 355-372.

[53] — (1983), "Generalized Econometric Models With Selectivity," *Econometrica* 51, 507-512.

[54] — (2002), "Correlation Bounds for Sample Selection Models with Mixed Continuous, Discrete and Count Data Variables," Manuscript, The Ohio State University.

[55] Li, K., Prabhala, N. R. (2005), "Self-Selection Models in Corporate Finance," Handbook of Corporate Finance, Empirical Corporate Finance, Volume 1.

[56] Li, M., D. J. Poirier, and J. L. Tobias (2004), "Do Dropouts Suffer From Dropping Out? Estimation and Prediction of Outcome Gains in Generalized Selection Models," *Journal of Applied Econometrics* 19, 203-225.

[57] Loughran, T., Ritter, J.R. (1997), "The operating performance of firms conducting seasoned equity offerings," Journal of Finance 52, 1823-1850.

[58] Makarov, G. D. (1981), "Estimates for the Distribution Function of a Sum of two Random Variables When the Marginal Distributions are Fixed," *Theory of Probability and its Applications* 26, 803-806.

[59] Manski, C. F. (1997a), "Monotone Treatment Effect," *Econometrica* 65, 1311-1334.

[60] Manski, C. F. (1997b), "The Mixing Problem in Programme Evaluation," *Review of Economic Studies* 64, 537-553.

[61] Masulis, R.W., Korwar, A.N. (1986), "Seasoned equity offerings: An empirical investigation," Journal of Financial Economics 15, 1-118.

[62] McNeil, Frey, and Embrechts (2005), *Quantitative Risk Management: Concepts, Techniques, and Tools*, Princeton University Press.

[63] Mencia, J. and Sentana, E. (2005), "Estimation and Testing of Dynamic Models with Generalized Hyperbolic Innovations," CEPR Discussion Papers, No. 5177.

[64] Nelsen, R. B. (1999). *An Introduction to Copulas.* Springer, New York.

[65] Newey, W., and D. McFadden (1994), "Large Sample Estimation and Hypothesis Testing," in R. Engle and D. McFadden (Eds.), Handbook of Econometrics: Volume IV (Amsterdam: Elsevier).

[66] Poirier, D. J. (1998), "Revising Beliefs in Non-Identified Models," *Econometric Theory* 14, 483-509.

[67] Poirier, D. J. and J. L. Tobias (2003), "On the Predictive Distributions of Outcome Gains in the Presence of an Unidentified Parameter," *Journal of Business & Economic Statistics* 21, 258-268.

[68] Prieger, J. E. (2002), "A Flexible Parametric Selection Model for Non-Normal Data with Application to Health Care Usage," *Journal of Applied Econometrics* 17, 367-392.

[69] Romano, J. and A. M. Shaikh (2006), "Inference for Identifiable Parameters in Partially Identified Econometric Models," Working Paper.

[70] Rubin, D. (1974), "Estimating causal effects of treatments in randomized and non-randomized studies", *Journal of Educational Psychology* 66, 688–701.

[71] — (1977), "Assignment to a treatment group on the basis of a covariate", *Journal of Educational Statistics* 2, 1–26.

[72] Rüschendorf, L. (1982), "Random Variables With Maximum Sums," *Advances in Applied Probability* 14, 623-632.

[73] Schafgans, M. M. A. and V. Zinde-Walsh (2002), "On Intercept Estimation in the Sample Selection Model," *Econometric Theory* 18, 40-50.

[74] Sklar A. (1959), "Fonctions de réartition à n dimensions et leures marges.," *Publications de l'Institut de Statistique de L'Université de Paris* 8, 229-231.

[75] Stoye, J. (2007), "More on Confidence Intervals for Partially Identified Parameters," Working paper, NYU.

[76] Smith, M. D. (2003), "Modelling Sample Selection Using Archimedean Copulas," *Econometrics Journal* 6, 99-123.

[77] — (2005), “Using Copulas to Model Switching Regimes with an Application to Child Labour,” *The Economic Record* 81, S47-S57.

[78] van der Vaart , A. W. and Jon A. Wellner (1996). *Weak Convergence and Empirical Processes*. Springer.

[79] Vijverberg, W. P. M. (1993), “Measuring the Unidentified Parameter of the Extended Roy Model of Selectivity,” *Journal of Econometrics* 57, 69-89.

[80] Williamson, R. C. and T. Downs (1990), “Probabilistic Arithmetic I: Numerical Methods for Calculating Convolutions and Dependency Bounds,” *International Journal of Approximate Reasoning* 4, 89-158.