Sensitivity Analysis in Triangular Systems of Equations with Binary Endogenous Variables

Sungwon Lee
(joint with Sukjin Han)

The University of Texas at Austin

Job Seminar at Portland State University

January 17, 2018
Introduction

- Bivariate probit (BVP) models
  - Frequently used in empirical analysis with two binary endogenous variables
  - Strong distributional assumption: Bivariate normality
  - Model misspecification

- This paper considers
  - Parametric / semiparametric estimation and inference for generalized BVP models
  - Sensitivity analysis
  - Failure of identification
A triangular model with binary endogenous variables

\[ Y_i = 1\{D_i \delta_0 + X'_i \beta_0 - \epsilon_i \geq 0\}, \]
\[ D_i = 1\{Z'_i \gamma_0 + X'_i \alpha_0 - \nu_i \geq 0\} \]

\((\epsilon, \nu)' \sim F_{\epsilon \nu}(\epsilon, \nu)\) and \(F_{\epsilon \nu}\) may not be fully known

Special case: BVP model

\(F_{\epsilon \nu} \sim \text{Bivariate Normal}\)
Example 1: The effect of Catholic school on high school graduation

- **Endogenous variables:**
  
  \[
  Y = \begin{cases} 
  1 & \text{if completed high school} \\
  0 & \text{o.w} 
  \end{cases}
  \]
  
  \[
  D = \begin{cases} 
  1 & \text{if attended Catholic school} \\
  0 & \text{o.w} 
  \end{cases}
  \]

- **Instruments** \( Z \)
  
  - (Parents’) religious affiliation (Evans and Schwab (1995), Altonji et al. (2005))
  - Number of Catholic schools in an area (Neal (1997))
  - Distance to the nearest Catholic school (Altonji et al. (2005))
Example 2: The effect of insurance on mortality

- **Endogenous variables:**
  \[
  Y = \begin{cases} 
  1 & \text{if died after 1 year} \\
  0 & \text{o.w} 
  \end{cases}
  \]
  \[
  D = \begin{cases} 
  1 & \text{if had an insurance} \\
  0 & \text{o.w} 
  \end{cases}
  \]

- **Instruments** \( Z \)
  - Policies that expand eligibility for medicaid (Goldman et al. (2001))
Example 3: The relationship between education and fertility

- **Endogenous variables:**

  \[
  Y = \begin{cases} 
  1 & \text{if a woman had at least one child} \\
  0 & \text{o.w} 
  \end{cases}
  \]

  \[
  D = \begin{cases} 
  1 & \text{if a woman had at least 8 years of education} \\
  0 & \text{o.w} 
  \end{cases}
  \]

- **Instruments** \( Z \)

  - Whether a woman was born during the first 6 months of the year (Marra and Radice (2011))
Generalization of BVP Models

- BVP models are frequently used in many applied studies
  - All of empirical examples mentioned above use BVP models
- Model misspecification in BVP models
  - Bivariate normality assumption
  - Inconsistent estimators, misleading policy implications
- Generalized BVP models: Relax the bivariate normality and generalize the BVP models
Copula functions: $C(\cdot, \cdot; \rho)$

- A copula, together with marginal distributions, fully characterizes a joint distribution:

$$F_{\epsilon \nu}(\epsilon, \nu) = C(F_{\epsilon}(\epsilon), F_{\nu}(\nu); \rho)$$

- The dependence structure between $\epsilon$ and $\nu$ is characterized by the copula function $C(\cdot, \cdot; \rho)$
- The degree of dependence: a scalar parameter $\rho$

Marginal distributions: $F_{\epsilon}$ and $F_{\nu}$

- Parametric $F_{\epsilon \nu}$: known marginals
- Semiparametric $F_{\epsilon \nu}$: unknown marginals
Generalized BVP

- Han and Vytlačil (2017, HV17): Identification
  - Provide identification of model primitives in a class of parametric/semiparametric models
  - Exclusion restriction and first-order stochastic dominance (FOSD) ordering for copulas

- This paper: Estimation and inference/ Sensitivity Analysis
  - Parametric/Semiparametric estimation framework
  - Model specification
Results and Contributions of This Paper

- Semiparametric sieve maximum likelihood estimation (MLE) and inference of a class of generalized bivariate probit models
  - Develop the asymptotic theory for the sieve ML estimator
  - Sensitivity analysis
- Failure of identification
  - Some believe that instruments may not be required
  - Exclusion restriction is important
- Provide a guideline of estimation procedure to help empirical researchers
Sensitivity Analysis

- Sensitivity to model specifications
  - Parametric and semiparametric MLE
  - Misspecification of marginals and/or copula

- Main findings
  - Marginal misspecification
    - Lower mean squared errors (MSEs) from semiparametric models
    - Parametric estimates of the average treatment effects (ATEs) are misleading
  - Copula and marginals misspecification
    - Even more distortions of estimates
Outline

1. Introduction
2. Model and Identification
3. Estimation
4. Asymptotic Theory
5. Simulations
6. Conclusions
7. Appendix
Model

- Triangular threshold crossing model

\[ Y_i = 1\{D_i \delta_0 + X_i' \beta_0 - \epsilon_i \geq 0\}, \]
\[ D_i = 1\{Z_i' \gamma_0 + X_i' \alpha_0 - \nu_i \geq 0\}, \]

where \(\epsilon_i\) and \(\nu_i\) are unobservables

- A copula, together with marginal distributions, characterizes the joint distribution function of \((\epsilon, \nu)\)

  - Dependence structure between \(\epsilon\) and \(\nu\) is captured by a copula function:

\[ C(\cdot, \cdot; \rho) : [0, 1] \times [0, 1] \rightarrow [0, 1] \]

- Marginal distributions of \(\epsilon\) and \(\nu\): Either parametric or nonparametric
Identification conditions

- Exclusion restriction: Instrumental variables
  - If not, non-identification or at most weak identification

- Stochastic dominance ordering
  - The copula needs to satisfy this stochastic ordering w.r.t. $\rho$
Exclusion restriction

- The instrumental variables $Z$ should not directly affect $Y$ and $\gamma_0 \neq 0$, where $\gamma_0$ is the coefficient on $Z$
  - There is an excluded variable that directly affects $D$ but not $Y$
- HV17 showed that this exclusion restriction is a sufficient condition for identification
Let $C(u_1|u_2)$ be a conditional copula.

First-order stochastic dominance (FOSD) of $C(u_1|u_2)$ with respect to $u_2$:
- e.g. For $u_2 > \tilde{u}_2$, $C(u_1|u_2)$ first-order stochastically dominates $C(u_1|\tilde{u}_2)$

We can rank copulas with respect to the degree of FOSD, denoted by $\prec_S$. 

Identification: Stochastic Dominance Ordering
Identification: Stochastic Dominance Ordering

- Stochastic dominance ordering of copulas with respect to $\rho$:

$$C(u_1|u_2; \rho_1) \prec_S C(u_1|u_2; \rho_2) \text{ for any } \rho_1 < \rho_2.$$  

- The copula function is ordered in $\rho$ in the sense of FOSD:
  - The larger $\rho$ is, the greater degree of the stochastic dominance is.
  - The bivariate normal distribution also satisfies this property.
Recall the model:

\[ Y_i = \mathbf{1}\{ D_i \delta_0 + X_i' \beta_0 - \epsilon_i \geq 0 \}, \]
\[ D_i = \mathbf{1}\{ Z_i' \gamma_0 + X_i' \alpha_0 - \nu_i \geq 0 \}, \]

where \( F_{\epsilon \nu_0}(e, v) \sim C(F_{\epsilon 0}(e), F_{\nu 0}(v); \rho_0) \)

**Proposition (Theorem 6.1 in HV17)**

Suppose that the identification conditions are satisfied. Then \( \psi_0 \equiv (\alpha_0', \beta_0', \delta_0, \gamma_0, \rho_0)' \) is identified. In addition, the marginal distributions \( F_{\epsilon 0}(\cdot) \) and \( F_{\nu 0}(\cdot) \) are also identified over the support of \( X \).
Failure of Identification: No Exclusion Restriction

- This paper shows that identification may fail without exclusion restrictions.
- An exclusion restriction is necessary and sufficient for identification when $X$ is binary.

**Theorem (Failure of Identification)**

In the model with $X = (1, X_1)$ where $X_1 \in \{0, 1\}$, suppose that the assumptions in Theorem 5.1 in HV17 hold, except that $\gamma_0 = 0$. If $\rho_0 \neq 0$, then there exist two distinct sets of $(\delta_1, \rho, \mu, \sigma)$ that generate the same observed data.
Failure of Identification: No Exclusion Restriction

- Mourifié and Méango (2014)
  - Failure of identification in the same situation
  - Not all available information is used
  - Simulation study
- Even a larger variation in $X$ causes a problem in identification
  - In a BVP model with continuous covariates $X$, the parameters are at best weakly identified
### Maximum Likelihood Estimation

- **Parameters**
  - $\psi_0 \equiv (\alpha'_0, \beta'_0, \delta_0, \gamma'_0, \rho_0)' \in \Psi$, $f_{\epsilon 0}(\cdot) \in {\mathcal{F}}_{\epsilon}$, and $f_{\nu 0}(\cdot) \in {\mathcal{F}}_{\nu}$
  - $\theta_0 \equiv (\psi_0, f_{\epsilon 0}(\cdot), f_{\nu 0}(\cdot))' \in \Theta = \Psi \times {\mathcal{F}}_{\epsilon} \times {\mathcal{F}}_{\nu}$

- **Fitted probabilities**
  - Let $p_{y_d,x,z,i}(\theta) \equiv \Pr(Y_i = y, D_i = d | X_i = x, Z_i = z)$; then
    
    \[
    p_{11,x,z,i}(\theta) = C(F_{\epsilon}(x' \beta + \delta), F_{\nu}(x' \alpha + z' \gamma); \rho)
    \]
    
    \[
    p_{10,x,z,i}(\theta) = F_{\epsilon}(x' \beta) - C(F_{\epsilon}(x' \beta), F_{\nu}(x' \alpha + z' \gamma); \rho)
    \]
    
    \[
    p_{01,x,z,i}(\theta) = F_{\nu}(x' \alpha + z' \gamma) - p_{11,x,z,i}(\theta)
    \]
    
    \[
    p_{00,x,z,i}(\theta) = 1 - (p_{11,x,z,i}(\theta) + p_{10,x,z,i}(\theta) + p_{01,x,z,i}(\theta))
    \]
Maximum Likelihood Estimation

- Log-likelihood function

\[
\hat{Q}_n(\theta; W) = \frac{1}{n} \sum_{i=1}^{n} \sum_{y,d \in \{0,1\}} 1_{yd,i} \log(p_{yd,xz,i}(\theta)) \\
\equiv \frac{1}{n} \sum_{i=1}^{n} l(W_i; \theta)
\]

where \(1_{yd,i} \equiv 1\{Y_i = y, D_i = d\}\) and \(W \equiv (Y, D, X', Z')'\)

- ML estimator \(\tilde{\theta}_n\)

\[
\tilde{\theta}_n \equiv \arg \max_{\theta \in \Theta} \hat{Q}_n(\theta; W)
\]
Let $Q_0(\theta) \equiv \mathbb{E}[l(W_i; \theta)]$

To use MLE, we need to show that $\theta_0$ is the unique maximizer of $Q_0(\theta)$

**Lemma**

*Under the identification conditions of Theorem 6.1 in HV17, $\theta_0$ is the unique maximizer of $Q_0(\theta)$*

The main difficulty is that the optimization problem is defined over an infinite-dimensional space $\Theta = \Psi \times \mathcal{F}_\epsilon \times \mathcal{F}_\nu$
Sieve Space

- Let $\mathcal{F}$ be a space of functions with some degree of smoothness.
- Consider a sequence of spaces of functions $(\mathcal{F}_k)_k$ such that (i) $\mathcal{F}_k \subseteq \mathcal{F}_{k+1} \subseteq \cdots$ for all $k \in \mathbb{N}$ and (ii) $\bigcup_{k=1}^{\infty} \mathcal{F}_k = \mathcal{F}$.

Example

Suppose that $\mathcal{F}$ is the space of continuous functions on $[0, 1]$ and $f \in \mathcal{F}$. Then we can find a sequence of polynomial functions $(p_k(x))_{k=1}^{\infty}$ such that

$$p_k(x) = a_{k0} + a_{k1}x + a_{k2}x^2 + \cdots + a_{kk}x^k$$

for some $a_k \equiv (a_{k0}, \ldots, a_{kk}) \in \mathbb{R}^{k+1}$ and $\sup_{x \in [0,1]} |f(x) - p_k(x)| \to 0$ as $k \to \infty$. Then, $\mathcal{F}_k = \{a_{k0} + a_{k1}x + a_{k2}x^2 + \cdots + a_{kk}x^k : a_k \in \mathbb{R}^{k+1}\}$ is a sieve space for $\mathcal{F}$. 
Sieve MLE

- Main idea: Replace the original parameter space with appropriate sieve spaces
  - Sieve space for $\Theta = \Psi \times \mathcal{F}_\epsilon \times \mathcal{F}_\nu$:
    \[ \Theta_n \equiv \Psi \times \mathcal{F}_{n\epsilon} \times \mathcal{F}_{n\nu} \]
    with appropriately chosen $(\mathcal{F}_{n\epsilon})_n$ and $(\mathcal{F}_{n\nu})_n$
  - The marginal density functions are approximated over these sieve spaces
  - Then, the sieve ML estimator is defined as
    \[ \hat{\theta}_n \equiv \arg \max_{\theta \in \Theta_n} \hat{Q}_n(\theta; W) \]
Asymptotic Theory

- Model:

\[ Y_i = 1 \{ D_i \delta_0 + X_i' \beta_0 - \epsilon_i \geq 0 \}, \]
\[ D_i = 1 \{ Z_i' \gamma_0 + X_i' \alpha_0 - \nu_i \geq 0 \}, \]

where \( F_{\epsilon \nu_0}(\epsilon, \nu) \sim C(F_{\epsilon 0}(\epsilon), F_{\nu 0}(\nu); \rho_0) \) and \( F_{\epsilon 0} \) and \( F_{\nu 0} \) are marginal distributions.

- Parameter to be estimated; \( \theta_0 = (\psi_0', f_{\epsilon 0}, f_{\nu 0})' \)
Asymptotic Theory

- We consider the following specification of the marginals:

\[ F_\epsilon(x) = H_\epsilon(G(x)), \]
\[ F_\nu(x) = H_\nu(G(x)), \]

where \( G : \mathbb{R} \to [0, 1] \) is a strictly increasing and differentiable function whose the derivative is bounded away from zero on \( \mathbb{R} \).

- Let \( h_{\epsilon 0}(x) \equiv \frac{dH_{\epsilon 0}(x)}{dx}, h_{\nu 0}(x) \equiv \frac{dH_{\nu 0}(x)}{dx}, \) and \( g(x) \equiv \frac{dG(x)}{dx} \).

- Parameter of interest is redefined: \( \theta_0 = (\psi'_0, h_{\epsilon 0}, h_{\nu 0})' \)

  - \( f_{\epsilon 0}(x) = h_{\epsilon 0}(G(x))g(x) \) and \( f_{\nu 0}(x) = h_{\nu 0}(G(x))g(x) \)
Consistency

Regularity Conditions

- Smoothness of $f_\epsilon$ and $f_\nu$, compactness of $\Theta$ under the pseudo-metric $d_c$ induced by the consistency norm
  - We use the sup-norm to define $d_c$:
    \[ d_c(\theta, \tilde{\theta}) \equiv ||\psi - \tilde{\psi}||_E + \sup_{t \in [0,1]} |h_\epsilon(t) - \tilde{h}_\epsilon(t)| + \sup_{t \in [0,1]} |h_\nu(t) - \tilde{h}_\nu(t)|, \]

  where $|| \cdot ||_E$ is the Euclidean norm

- $\{ W_i \}_{i=1}^n$ are i.i.d with the finite second moment

- Sieve space is chosen to approximate $\theta_0 \in \Theta$ well over $\Theta_n$ w.r.t. $d_c$

- Envelope function with finite moment conditions

- Complexity of $\Theta_n$ w.r.t. $d_c$
Consistency

Remarks on the conditions

- Choice of norm is important to ensure compactness of the parameter space
- Choice of sieve space depends on
  - boundedness of the support
  - class of functions (e.g. Hölder space)
- Complexity of the sieve space
  - Controlled by the number of approximating functions, denoted by $k_n$
Consistency

Theorem (Consistency)

Suppose that the identification conditions are satisfied. Under the conditions above, the sieve ML estimator is consistent w.r.t. \( d_c(\cdot, \cdot) \), i.e.,
\[
d_c(\hat{\theta}_n, \theta_0) \xrightarrow{p} 0.
\]
Convergence Rate

- Convergence rate plays an important role in deriving the asymptotic normality.
- We establish the convergence rate with respect to $L^2$-norm:

$$||\theta - \theta_0||_2 \equiv ||\psi - \psi_0||_E + ||h_\epsilon - h_{\epsilon_0}||_2 + ||h_\nu - h_{\nu_0}||_2,$$

where

$$||h - \tilde{h}||_2^2 \equiv \int_0^1 (h(t) - \tilde{h}(t))^2 dt$$
Theorem (Convergence Rate)

Let $p$ be the degree of smoothness of unknown densities. Under some regularity conditions, we have

$$||\hat{\theta}_n - \theta_0||_2 = O_p(\max\{\sqrt{\frac{k_n}{n}}, k_n^{-p}\}).$$

Furthermore, if we choose $k_n \propto n^{\frac{1}{2p+1}}$, then we have

$$||\hat{\theta}_n - \theta_0||_2 = O_p(n^{-\frac{p}{2p+1}}).$$

- Variance-bias trade-off: When $k_n$ increases,
  - the rate of the variance term ($\sqrt{\frac{k_n}{n}}$) increases
  - the rate of the bias term ($k_n^{-p}$) decreases
Asymptotic Normality of a Smooth Functional $T$

- **Functional** $T : \Theta \rightarrow \mathbb{R}$
  - Many quantities of interest can be considered as a functional of the parameter $\theta$
- $\sqrt{n}$-asymptotic normality of a class of functionals
  - Restrict our attention to the class of smooth functionals
- **Examples**
  - Elements of $\psi_0$: $T(\theta_0) = \delta_0$ or $T(\theta_0) = \rho_0$
  - Average treatment effect conditional on $X = x$:
    $T(\theta_0) = ATE(x)$, where

$$ATE(x) \equiv \mathbb{E}[Y|X = x, D = 1] - \mathbb{E}[Y|X = x, D = 0] = F_{\epsilon_0}(\delta_0 + x'\beta_0) - F_{\epsilon_0}(x'\beta_0)$$

$$= H_{\epsilon_0}(G(\delta_0 + x'\beta_0)) - H_{\epsilon_0}(G(x'\beta_0))$$
Smoothness of $T$

- Consider

$$\frac{\partial T(\theta_0)}{\partial \theta'}[v] = \lim_{t \to 0} \frac{T(\theta_0 + tv) - T(\theta_0)}{t}$$

- Pathwise derivative in the direction $v$

- Let

$$\left\| \frac{\partial T(\theta_0)}{\partial \theta'} \right\| \equiv \sup_{v \in V, \|v\| > 0} \frac{\left| \frac{\partial T(\theta_0)}{\partial \theta'}[v] \right|}{\|v\|}$$

be the operator norm of the pathwise derivative of $T$
**Smoothness of** $T$

- Smoothness is characterized by first-order approximation of $T$ and the operator norm of the pathwise derivative:

$$|T(\theta_0 + v) - T(\theta_0) - \frac{\partial T(\theta_0)}{\partial \theta'}[v]| = O(||v||^w)$$

for some $w > 0$, and

$$||\frac{\partial T(\theta_0)}{\partial \theta'}|| < \infty$$

- Generalization of the delta-method
Asymptotic Normality of a Smooth Functional $T$

Theorem (Asymptotic Normality)

Under several conditions, we have

$$\sqrt{n}(T(\hat{\theta}_n) - T(\theta_0)) \xrightarrow{d} N(0, \left\| \frac{\partial T(\theta_0)}{\partial \theta'} \right\|^2),$$

where $\left\| \frac{\partial T(\theta_0)}{\partial \theta'} \right\| \equiv \sup_{v \in \mathcal{V}, ||v|| > 0} \frac{|\frac{\partial T(\theta_0)}{\partial \theta'}[v]|}{||v||}$.

Corollary

Let $S_{\psi_0}$ be the efficient score function and $\mathcal{I}_*(\psi_0) = E[S_{\psi_0} S'_{\psi_0}]$. Under several conditions, we have

$$\sqrt{n}(\hat{\psi}_n - \psi_0) \xrightarrow{d} N(0, \mathcal{I}_*(\psi_0)^{-1}),$$

provided $\mathcal{I}_*(\psi_0)$ is non-singular.
Simulation Design

- Data Generating Process (DGP)

\[
Y_i = 1\{D_i\delta + X_i\beta_1 \geq \epsilon_i\}
\]

\[
D_i = 1\{Z_i\gamma + X_i\alpha_1 \geq \nu_i\},
\]

where \((\epsilon, \nu) \sim C(F_\epsilon(\epsilon), F_\nu(\nu); \rho)\)

- Covariates:

\[
\begin{pmatrix} X \\ Z \end{pmatrix} \sim N \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & -0.1 \\ -0.1 & 1 \end{pmatrix} \right) \]

- Parameters of interest

  - Finite-dimensional parameter \((\gamma, \delta, \rho)\)
  - ATE at \(x = \mu_x\); \(F_\epsilon(\mu_x\beta_1 + \delta) - F_\epsilon(\mu_x\beta_1)\), where \(\mu_x = E[X]\)
Simulation Design

- Marginal distribution

Table: Marginal distributions

<table>
<thead>
<tr>
<th>DGP</th>
<th>Parametric model</th>
<th>Semiparametric model</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\epsilon, \nu \sim N(0, 1)$</td>
<td>$\epsilon \sim N(\mu_\epsilon, \sigma_\epsilon^2)$</td>
<td>Sieve approximation</td>
</tr>
<tr>
<td>$\epsilon, \nu \sim 0.6N(-1, \sigma^2) + 0.4N(1.5, \sigma^2)$</td>
<td>$\nu \sim N(\mu_\nu, \sigma_\nu^2)$</td>
<td></td>
</tr>
</tbody>
</table>

- Dependence measure: Spearman's $\rho$
- Sieve space: Polynomial sieve with $k_n \propto n^{1/7}$

$^1\sigma^2$ is set to have $\text{Var}(\epsilon) = \text{Var}(\nu) = 1.$
Results: Correctly Specified Model

**Table: Result of simulation** ($n = 500$)

<table>
<thead>
<tr>
<th>Gaussian</th>
<th>Parametric model</th>
<th>Semiparametric model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\gamma$</td>
<td>$\delta$</td>
</tr>
<tr>
<td>True value</td>
<td>0.8000</td>
<td>1.1000</td>
</tr>
<tr>
<td>S.D.</td>
<td>0.0934</td>
<td>0.3954</td>
</tr>
<tr>
<td>Abs. Bias</td>
<td>0.0074</td>
<td>0.0469</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.0936</td>
<td>0.3982</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Frank</th>
<th>Parametric model</th>
<th>Semiparametric model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\gamma$</td>
<td>$\delta$</td>
</tr>
<tr>
<td>True value</td>
<td>0.8000</td>
<td>1.1000</td>
</tr>
<tr>
<td>S.D.</td>
<td>0.0936</td>
<td>0.3379</td>
</tr>
<tr>
<td>Abs. Bias</td>
<td>0.0027</td>
<td>0.0450</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.0936</td>
<td>0.3409</td>
</tr>
</tbody>
</table>
### Results: Correctly Specified Model

#### Table: Result of simulation \((n = 1,000)\)

<table>
<thead>
<tr>
<th></th>
<th>Gaussian Parametric model</th>
<th></th>
<th>Gaussian Semiparametric model</th>
<th></th>
<th>Frank Parametric model</th>
<th></th>
<th>Frank Semiparametric model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(\gamma)</td>
<td>(\delta)</td>
<td>(\rho_{sp})</td>
<td>(ATE)</td>
<td>(\gamma)</td>
<td>(\delta)</td>
<td>(\rho_{sp})</td>
</tr>
<tr>
<td>True value</td>
<td>0.8000</td>
<td>1.1000</td>
<td>0.5000</td>
<td>0.3643</td>
<td>0.8000</td>
<td>1.1000</td>
<td>0.5000</td>
</tr>
<tr>
<td>S.D.</td>
<td>0.0654</td>
<td>0.2737</td>
<td>0.1081</td>
<td>0.0656</td>
<td>0.0655</td>
<td>0.2939</td>
<td>0.1092</td>
</tr>
<tr>
<td>Abs. Bias</td>
<td>0.0025</td>
<td>0.0165</td>
<td>0.0004</td>
<td>0.0011</td>
<td>0.0026</td>
<td>0.0205</td>
<td>0.0031</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.0655</td>
<td>0.2742</td>
<td>0.1081</td>
<td>0.0656</td>
<td>0.0655</td>
<td>0.2946</td>
<td>0.1092</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Frank Parametric model</td>
<td></td>
<td>Frank Semiparametric model</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(\gamma)</td>
<td>(\delta)</td>
<td>(\rho_{sp})</td>
<td>(ATE)</td>
<td>(\gamma)</td>
<td>(\delta)</td>
<td>(\rho_{sp})</td>
</tr>
<tr>
<td>True value</td>
<td>0.8000</td>
<td>1.1000</td>
<td>0.5000</td>
<td>0.3643</td>
<td>0.8000</td>
<td>1.1000</td>
<td>0.5000</td>
</tr>
<tr>
<td>S.D.</td>
<td>0.0658</td>
<td>0.2605</td>
<td>0.1023</td>
<td>0.0620</td>
<td>0.0652</td>
<td>0.2663</td>
<td>0.1066</td>
</tr>
<tr>
<td>Abs. Bias</td>
<td>0.0017</td>
<td>0.0188</td>
<td>0.0010</td>
<td>0.0009</td>
<td>0.0007</td>
<td>0.0164</td>
<td>0.0042</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.0658</td>
<td>0.2612</td>
<td>0.1023</td>
<td>0.0620</td>
<td>0.0652</td>
<td>0.2668</td>
<td>0.1067</td>
</tr>
</tbody>
</table>
## Results: Misspecification of Marginals

**Table: Result of simulation ($n = 500$)**

<table>
<thead>
<tr>
<th>Gaussian</th>
<th>Parametric model</th>
<th></th>
<th></th>
<th>Semiparametric model</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\gamma$</td>
<td>$\delta$</td>
<td>$\rho_{sp}$</td>
<td>$ATE$</td>
<td>$\gamma$</td>
</tr>
<tr>
<td>True value</td>
<td>0.8000</td>
<td>1.1000</td>
<td>0.5000</td>
<td>0.1066</td>
<td>0.8000</td>
</tr>
<tr>
<td>S.D.</td>
<td>0.1281</td>
<td>0.6285</td>
<td>0.1651</td>
<td>0.1129</td>
<td>0.1113</td>
</tr>
<tr>
<td>Abs. Bias</td>
<td>0.0006</td>
<td>0.0075</td>
<td>0.0504</td>
<td>0.1377</td>
<td>0.0562</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.1281</td>
<td>0.6285</td>
<td>0.1726</td>
<td>0.1780</td>
<td>0.1247</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Frank</th>
<th>Parametric model</th>
<th></th>
<th></th>
<th>Semiparametric model</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\gamma$</td>
<td>$\delta$</td>
<td>$\rho_{sp}$</td>
<td>$ATE$</td>
<td>$\gamma$</td>
</tr>
<tr>
<td>True value</td>
<td>0.8000</td>
<td>1.1000</td>
<td>0.5000</td>
<td>0.1066</td>
<td>0.8000</td>
</tr>
<tr>
<td>S.D.</td>
<td>0.1272</td>
<td>0.5093</td>
<td>0.1221</td>
<td>0.0883</td>
<td>0.1141</td>
</tr>
<tr>
<td>Abs. Bias</td>
<td>0.0056</td>
<td>0.2088</td>
<td>0.1024</td>
<td>0.1827</td>
<td>0.0377</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.1273</td>
<td>0.5504</td>
<td>0.1594</td>
<td>0.2030</td>
<td>0.1202</td>
</tr>
</tbody>
</table>
### Table: Result of simulation ($n = 1,000$)

<table>
<thead>
<tr>
<th></th>
<th>Gaussian</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Parametric model</td>
<td></td>
<td>Semiparametric model</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\gamma$</td>
<td>$\delta$</td>
<td>$\rho_{sp}$</td>
<td>$ATE$</td>
<td>$\gamma$</td>
<td>$\delta$</td>
<td>$\rho_{sp}$</td>
<td>$ATE$</td>
</tr>
<tr>
<td>True value</td>
<td></td>
<td>0.8000</td>
<td>1.1000</td>
<td>0.5000</td>
<td>0.1066</td>
<td>0.8000</td>
<td>1.1000</td>
<td>0.5000</td>
<td>0.1066</td>
</tr>
<tr>
<td>S.D.</td>
<td></td>
<td>0.0911</td>
<td>0.4256</td>
<td>0.1156</td>
<td>0.0807</td>
<td>0.0778</td>
<td>0.2576</td>
<td>0.0721</td>
<td>0.0463</td>
</tr>
<tr>
<td>Abs. Bias</td>
<td></td>
<td>0.0059</td>
<td>0.0451</td>
<td>0.0504</td>
<td>0.1381</td>
<td>0.0641</td>
<td>0.2030</td>
<td>0.0222</td>
<td>0.0195</td>
</tr>
<tr>
<td>RMSE</td>
<td></td>
<td>0.0913</td>
<td>0.4279</td>
<td>0.1261</td>
<td>0.1599</td>
<td>0.1008</td>
<td>0.3279</td>
<td>0.0755</td>
<td>0.0502</td>
</tr>
</tbody>
</table>

|               | Frank |               |               |               |               |               |               |               |               |
|---------------|-------|---------------|---------------|---------------|---------------|---------------|---------------|               |
|               |       |               |               |               |               |               |               |               |
|               |       | $\gamma$ | $\delta$ | $\rho_{sp}$ | $ATE$ | $\gamma$ | $\delta$ | $\rho_{sp}$ | $ATE$ |               |               |               |               |               |               |
| True value    |       | 0.8000 | 1.1000 | 0.5000 | 0.1066 | 0.8000 | 1.1000 | 0.5000 | 0.1066 |               |               |               |               |               |               |
| S.D.          |       | 0.0899 | 0.3876 | 0.0966 | 0.0684 | 0.0837 | 0.2577 | 0.0690 | 0.0500 |               |               |               |               |               |               |
| Abs. Bias     |       | 0.0044 | 0.2066 | 0.1060 | 0.1853 | 0.0525 | 0.1802 | 0.0223 | 0.0225 |               |               |               |               |               |               |
| RMSE          |       | 0.0901 | 0.4392 | 0.1434 | 0.1975 | 0.0988 | 0.3145 | 0.0725 | 0.0549 |               |               |               |               |               |               |
Results: Misspecification of Marginals and Copula

**Table: Result of simulation ($n = 500$, DGP: Gumbel copula)**

<table>
<thead>
<tr>
<th>Gaussian</th>
<th>Parametric model</th>
<th>Semiparametric model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\gamma$</td>
<td>$\delta$</td>
</tr>
<tr>
<td>True value</td>
<td>0.8000</td>
<td>1.1000</td>
</tr>
<tr>
<td>S.D.</td>
<td>0.1304</td>
<td>0.6489</td>
</tr>
<tr>
<td>Abs. Bias</td>
<td>0.0022</td>
<td>0.0488</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.1304</td>
<td><strong>0.6508</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Frank</th>
<th>Parametric model</th>
<th>Semiparametric model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\gamma$</td>
<td>$\delta$</td>
</tr>
<tr>
<td>True value</td>
<td>0.8000</td>
<td>1.1000</td>
</tr>
<tr>
<td>S.D.</td>
<td>0.1290</td>
<td>0.5211</td>
</tr>
<tr>
<td>Abs. Bias</td>
<td>0.0140</td>
<td>0.3128</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.1297</td>
<td><strong>0.6078</strong></td>
</tr>
</tbody>
</table>
Results: Misspecification of Marginals and Copula

Table: Result of simulation ($n = 1,000$, DGP: Gumbel copula)

<table>
<thead>
<tr>
<th>Gaussian</th>
<th>Parametric model</th>
<th>Semiparametric model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\gamma$</td>
<td>$\delta$</td>
</tr>
<tr>
<td>True value</td>
<td>0.8000</td>
<td>1.1000</td>
</tr>
<tr>
<td>S.D.</td>
<td>0.0896</td>
<td>0.4412</td>
</tr>
<tr>
<td>Abs. Bias</td>
<td>0.0095</td>
<td>0.0059</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.0901</td>
<td>0.4412</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Frank</th>
<th>Parametric model</th>
<th>Semiparametric model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\gamma$</td>
<td>$\delta$</td>
</tr>
<tr>
<td>True value</td>
<td>0.8000</td>
<td>1.1000</td>
</tr>
<tr>
<td>S.D.</td>
<td>0.0901</td>
<td>0.3917</td>
</tr>
<tr>
<td>Abs. Bias</td>
<td>0.0123</td>
<td>0.3374</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.0910</td>
<td>0.5169</td>
</tr>
</tbody>
</table>
Conclusions

- Propose a semiparametric sieve ML estimation and inference method
  - Provide the asymptotic theory of the sieve ML estimator

- Sensitivity analysis
  - Performance of the sieve ML estimator
  - Marginal misspecification yields distortions of estimates

- Failure of identification
  - Exclusion restriction is important

- Empirical application (in progress)
  - The effect of attending Catholic school on high school completion
### Table: Examples of Copulas

<table>
<thead>
<tr>
<th>Copula family</th>
<th>Functional form</th>
<th>( \rho )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gaussian family(^2)</td>
<td>( C(u_1, u_2; \rho) = \Phi_2(\Phi^{-1}(u_1), \Phi^{-1}(u_2); \rho) )</td>
<td>( \rho \in [-1, 1] )</td>
</tr>
<tr>
<td>Frank family</td>
<td>( C(u_1, u_2; \rho) = -\frac{1}{\rho} \ln{1 + \frac{(e^{-\rho u_1} - 1)(e^{-\rho u_2} - 1)}{e^{-\rho} - 1}} )</td>
<td>( \rho \in \mathbb{R} - {0} )</td>
</tr>
<tr>
<td>Clayton family</td>
<td>( C(u_1, u_2; \rho) = (u_1^{-\rho} + u_2^{-\rho} - 1)^{-1/\rho} )</td>
<td>( \rho \in [0, \infty) )</td>
</tr>
</tbody>
</table>

\(^2\)\( \Phi_2(\cdot, \cdot; \rho) \) is the bivariate standard normal distribution function and \( \Phi(\cdot) \) is the standard normal distribution function.
Suppose that $(F_\epsilon(\epsilon), F_\nu(\nu))' \sim C(\cdot, \cdot; \rho)$. Then,

$$C_{1|2}(u_1|u_2; \rho) = \Pr(F_\epsilon(\epsilon) \leq u_1 | F_\nu(\nu) = u_2; \rho)$$

$$= \Pr(\epsilon \leq F_\epsilon^{-1}(u_1) | \nu = F_\nu^{-1}(u_2); \rho)$$

$$= F_\epsilon|\nu(F_\epsilon^{-1}(u_1) | F_\nu^{-1}(u_2); \rho)$$

$$\equiv F_\epsilon|\nu(e|\nu; \rho),$$

where $e = F_\epsilon^{-1}(u_1)$ and $\nu = F_\nu^{-1}(u_2)$
Example

Suppose that

$$\begin{pmatrix} \epsilon \\ \nu \end{pmatrix} \sim N \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right).$$

Then,

$$F_{\epsilon \nu}(e, v; \rho) = C(\Phi(e), \Phi(v); \rho),$$

where $C(\cdot, \cdot; \rho)$ is a Gaussian copula. In this case, $C_{1|2}(u_1|u_2; \rho)$ is the conditional distribution function of the normal random variable with mean $\rho \Phi^{-1}(u_2)$ and variance $(1 - \rho^2)$. 
Stochastic Dominance Ordering

Figure: Stochastic Dominance Ordering

Stochastic Dominance with $\rho = 0.2$

Stochastic Dominance with $\rho = 0.5$
Sieve Extremum Estimation

Sieve Extremum Estimation

- Idea: Replace the original parameter space with appropriate sieve spaces
- The complexity of sieve spaces grows as $n$ increases

Example

Let $f_0(x) \equiv \mathbb{E}[Y|X = x]$ and $f_0 \in \mathcal{F}$.

A nonparametric estimator of $f_0$:

$$\tilde{f}(x) = \arg \max_{f \in \mathcal{F}} -\frac{1}{n} \sum_i (Y_i - f(X_i))^2$$

A sieve estimator of $f_0$:

$$\hat{f}(x) = \arg \max_{f \in \mathcal{F}_k} -\frac{1}{n} \sum_i (Y_i - f(X_i))^2$$
Example

For a random sample \( \{X_i \in \mathbb{R} : i = 1, 2, \ldots, n\} \) with \( EX_i = \mu \) and \( EX_i^2 = \sigma^2 < \infty \), define \( \bar{X} \equiv \frac{1}{n} \sum_i^n X_i \). Letting \( T(\mu) = \frac{1}{\mu} \), we can show that

\[
\sqrt{n}(T(\bar{X}) - T(\mu)) = \sqrt{n} \left( \frac{1}{\bar{X}} - \frac{1}{\mu} \right) = T'(\hat{\mu}) \sqrt{n}(\bar{X} - \mu) \xrightarrow{d} N(0, \sigma^2 \frac{1}{\mu^4}),
\]

where \( \hat{\mu} \) lies between \( \bar{X}_n \) and \( \mu \) and \( T'(\mu) = -\frac{1}{\mu^2} \).