COPULA APPROACHES

NAVEEN ELURU

Motivation

- Dependency across choice dimension
  - Household location and household vehicle mileage
    - Individuals willing to reside in neo-urbanist neighborhoods are more likely to bicycle and as a result drive less. On the other hand, individuals residing in conventional neighborhoods are more likely to drive more

Decision Maker: **Household**
Choices:
- Household Location
- Daily VMT travelled

Motivation

- Current approach to the problem
- Lets say there are three choices that we are considering r_q, s_q and t_q
- The latent propensity for these equations is given by:

$$r_q^* = \beta' x_q + \varepsilon_q$$

$$s_q^* = \alpha' z_q + \eta_q$$

$$t_q^* = \gamma' w_q + \xi_q$$

- The dependency across choice dimensions is obtained by correlating the error terms $\varepsilon_q, \eta_q, \xi_q$ using
  - Multivariate normal assumption on the error terms
  - Simulation based approaches

Motivation

- Problems with current approaches
  - Inflexible
    - Typically impose a bivariate (multivariate) normal structure to model dependencies
  - No closed form solutions
    - Computationally challenging because they require simulation
    - Potentially inaccurate when we have high dimensionality
  - Downright infeasible in some cases
    - The likelihood functions are of such high dimensionality we cannot mathematically compute them!

COPULA METHODOLOGY

Copula Approach

- The copula approach has recently revived interest in a whole set of alternative couplings that can allow non-linear and asymmetric dependencies.
  - Copula (Sklar, 1959): tie, bond, connect
- A copula is essentially a multivariate functional form for the joint distribution of random variables derived purely from pre-specified parametric marginal distributions of each random variable.
- Does not pre-impose a particular multivariate error structure
  - Allows different specifications for the univariate marginal distributions and the dependence structure
  - Empirically tests different multivariate dependence functions
  - Chooses the one that best fits the data

Copula Approach: Formulation

- The precise definition of a copula is that it is a multivariate distribution function defined over the unit cube linking uniformly distributed marginals

- Let C be a Q-dimensional copula of uniformly distributed random variables $U_1, U_2, U_3, \ldots, U_Q$ with support contained in $[0,1]^Q$

- $C_\theta (u_1, u_2, \ldots, u_q) = \Pr(U_1 < u_1, U_2 < u_2, \ldots, U_Q < u_Q)$ where θ is a parameter vector of the copula commonly referred to as the dependence parameter vector

Copula Approach: Formulation

- Consider Q random variables $V_1, V_2, V_3, \ldots, V_Q$
  - Each with standard univariate continuous marginal distribution functions $F_q(v_q) = \Pr(V_q < v_q)$, $q = 1, 2, 3, \ldots, Q$
  - By the integral transform result

$$F_q(v_q) = \Pr(V_q < v_q) = \Pr(F_q^{-1}(U_q) < v_q) = \Pr(U_q < F_q(v_q))$$

  - Finally, by Sklar's (1973) theorem

$$F(v_1, v_2, \ldots, v_Q) = \Pr(V_1 < v_1, V_2 < v_2, \ldots, V_Q < v_Q) = \Pr(U_1 < F_1(v_1), U_2 < F_2(v_2), \ldots, U_Q < F_Q(v_Q))$$

$$= C_\theta\,(u_1 = F_1(v_1), u_2 = F_2(v_2), \ldots, u_Q = F_Q(v_Q))$$

Different Types of Copulas

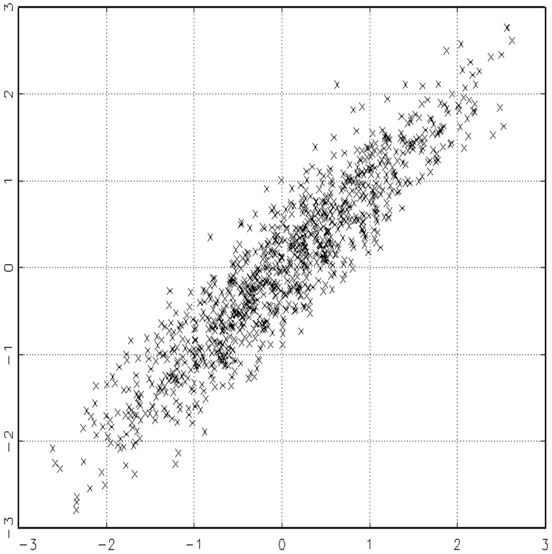
- A rich set of bivariate copula types have been generated using the inversion and other methods
  - Gaussian copula
  - Farlie-Gumbel-Morgenstern (FGM) copula
  - Archimedean class of copulas
    - Clayton
    - Gumbel
    - Frank
    - Joe

Different Types of Copulas

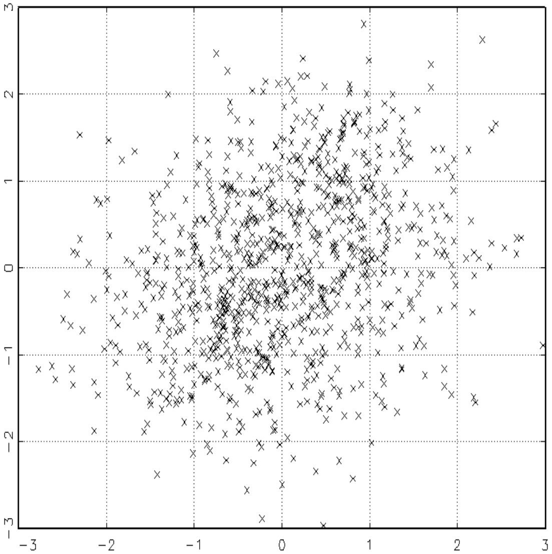
- To better understand the generated dependence structures between the random variables (Y_1,Y_2) we define a scalar measure (The traditional dependence concept of correlation coefficient (ρ)) is limited *
  - Kendall's τ
  - Spearmans ρ_S

*For more details see Bhat and Eluru (2009)

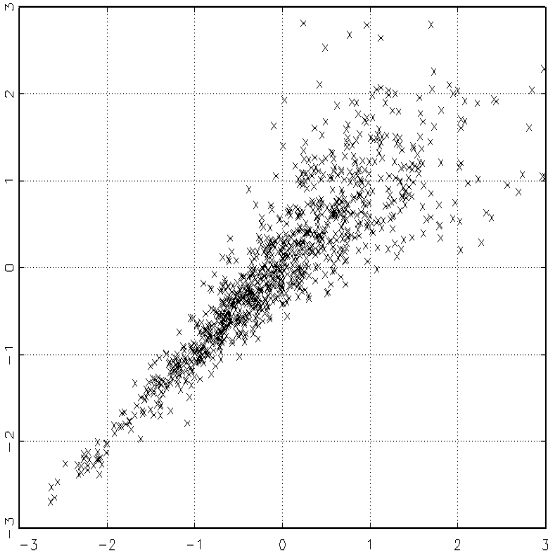
Different Types of Copulas



Gaussian Copula



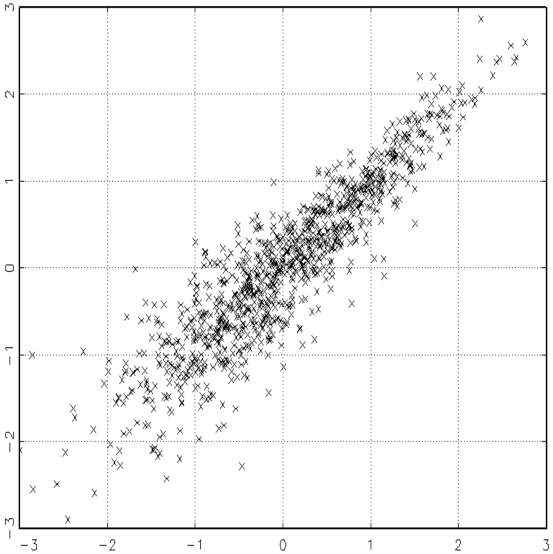
FGM Copula



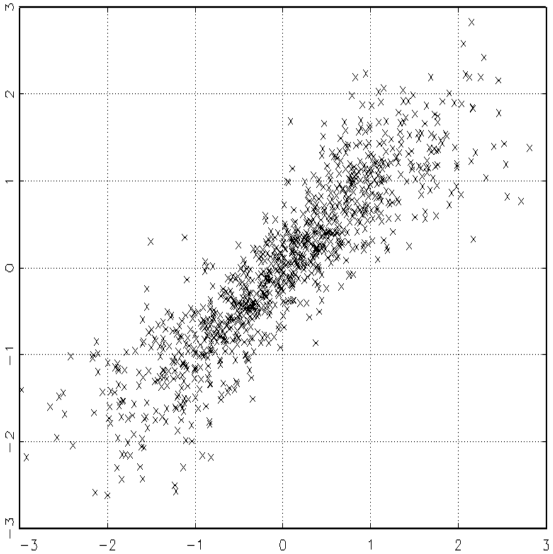
Clayton Copula

Kendall's $\tau = 0.75$

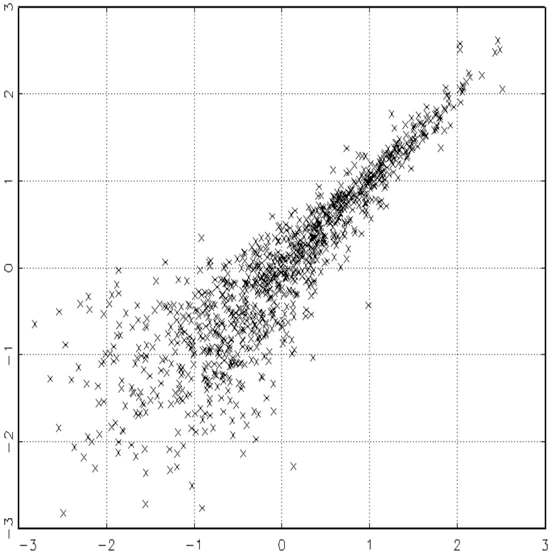
Different Types of Copulas



Gumbel Copula



Frank Copula



Joe Copula

Kendall's $\tau = 0.75$

Case Study of a Copula Application

- Modeling Household Location and household vehicle mileage
- Why?
  - Self-selection versus influence of built environment
  - Residential location is an important pre-cursor to activity-travel patterns
  - With increasing emphasis on green house gases (GHG) accurately modeling vehicle miles is very important

Case Study of a Copula Application

- Location categorized as:

Neo-urbanist	Conventional
➢High population density	➢Low population density
➢High bicycle lane and roadway street density	➢Low bicycle lane and roadway street density
➢Good land-use mix	➢Primarily single use residential land use
➢Efficient transit	➢Auto-dependent urban design
➢Non-motorized mode accessibility/facilities	

- Household vehicle mileage
  - Logarithm of vehicle miles travelled
- 2000 San Francisco Bay Area Household Travel Survey (BATS)

Case Study of a Copula Application

- Endogenous switching system

$$r_q^* = \beta' x_q + \varepsilon_q, \quad r_q = 1 \text{ if } r_q^* > 0, \quad r_q = 0 \text{ if } r_q^* \le 0,$$ Binary Choice

$$m_{q0}^* = \alpha' z_q + \eta_q, \quad m_{q0} = 1[r_q = 0] m_{q0}^*$$ Neo-urbanist

$$m_{q1}^* = \gamma' w_q + \xi_q, \quad m_{q1} = 1[r_q = 1] m_{q1}^*$$ Conventional

- The correlated pairs are (ε_q, η_q) (ε_q, ξ_q)
- The log-likelihood function is given by

$$L = \prod_{q=1}^{Q} \left[\left\{ \Pr[m_{q0} \mid r_q^* \le 0] \times \Pr[r_q^* \le 0] \right\}^{1-r_q} \times \left\{ \Pr[m_{q1} \mid r_q^* > 0] \times \Pr[r_q^* > 0] \right\}^{r_q} \right]$$

Case Study of a Copula Application

- Computing the previous expression

$$L = \prod_{q=1}^{Q} \left[\frac{1}{\sigma_\eta} \cdot f_\eta \left(\frac{m_{q0} - \alpha' z_q}{\sigma_\eta} \right) \cdot \frac{\partial}{\partial u_{q2}^0} C_{\theta_0} \left(u_{q1}^0, u_{q2}^0 \right) \right]^{1-r_q} \times$$

$$\left[\frac{1}{\sigma_\xi} \cdot f_\xi \left(\frac{m_{q1} - \gamma' w_q}{\sigma_\xi} \right) \left\{ 1 - \frac{\partial}{\partial u_{q2}^1} C_{\theta_1} \left(u_{q1}^1, u_{q2}^1 \right) \right\} \right]^{r_q},$$

$$\text{where} \quad u_{q1}^0 = F_\varepsilon(-\beta' x_q), \qquad u_{q2}^0 = F_\eta \left(\frac{m_{q0} - \alpha' z_q}{\sigma_\eta} \right),$$

$$u_{q1}^1 = u_{q1}^0, \qquad u_{q2}^1 = F_\xi \left(\frac{m_{q1} - \gamma' w_q}{\sigma_\xi} \right)$$

- Any copula possible
  - In fact $C_{\theta o}$,C_{θ_1} need not be same!

Case Study of a Copula Application

- Empirical application
  - Variables considered
    - Household demographics and employment characteristics
    - Neighborhood characteristics including
      - Population density
      - Employment density
      - Accessibility measures
      - Population by ethnicity in the neighborhood
      - Presence/number of schools and physically active centers
      - Density of bicycle lanes and street blocks

Estimation results

- Total number of models estimated
  - 6 models with the same copula dependency structure
  - 24 models with different combinations of the six copula dependency structures
  - A model that assumed independence for comparison

Estimation results

- Five best copula dependency structure combinations (based on the BIC)
  - Frank-Frank (-6842.2)
  - Frank-Joe (-6844.2)
  - FGM-Joe (-6851.0)
  - Independent-Joe (-6863.7) and
  - FGM-Gumbel (-6866.2)
- The value at convergence for Gaussian-Gaussian copula is -6877.9
- This is simply an artifact of the normal dependency structure
  - indicative of the kind of incorrect results that can be obtained by placing restrictive distributional assumptions

Estimation results – Binary Component

Variables	Independence-Independence Copula		Frank-Frank Copula	
	Parameter	t-stat	Parameter	t-stat
Propensity to choose conventional neighborhood				
Constant	0.201	4.15	0.275	5.72
Age of householder < 35 years	-0.131	-2.35	-0.143	-2.75
Number of children (of age < 16 years) in the household	0.164	4.62	0.161	4.59
Household lives in a single family dwelling unit	0.382	6.79	0.337	6.28
Own household	0.597	10.37	0.497	8.81

Estimation results – Log(VMT) Neo

Variables	Independence-Independence Copula		Frank-Frank Copula	
	Parameter	t-stat	Parameter	t-stat
Log of vehicle miles of travel in a neo-urbanist neighborhood				
Constant	-0.017	-0.16	-0.638	-5.48
Household vehicle ownership				
Household Vehicles = 1	2.617	21.50	2.744	24.26
Household Vehicles ≥ 2	3.525	25.44	3.518	27.40
Number of full-time students in the household	0.183	2.13	0.112	1.41
Copula dependency parameter (θ)	--	--	-2.472	-6.98
Scale parameter of the continuous component	1.301	40.62	1.348	34.31

Estimation results – Log(VMT) Conv

Variables	Independence-Independence Copula		Frank-Frank Copula	
	Parameter	t-stat	Parameter	t-stat
Log of vehicle miles of travel in a conventional neighborhood				
Constant	0.379	2.28	0.163	1.08
Household vehicle ownership				
Household Vehicles = 1	3.172	21.77	3.257	25.43
Household Vehicles = 2	3.705	25.32	3.854	29.92
Household Vehicles ≥ 3	3.931	25.92	4.102	30.41
Number of employed individuals in the household	0.229	7.24	0.208	6.66
Number of full-time students in the household	0.104	5.06	0.131	6.27
Density of bicycle lanes	-0.023	-3.08	-0.024	-3.24
Accessibility to shopping (Hansen measure)	-0.024	-7.34	-0.027	-8.19
Copula dependency parameter (θ)	--	--	3.604	7.22
Scale parameter of the continuous component	0.891	75.78	0.920	63.59

Treatment Effects

- You would note that for each household we have either the mileage in neo-urbanist zone or conventional zone
- However, using the switching model, we would like to assess the impact of the neighborhood on VMT
- In the social science terminology, we would like to evaluate the expected gains (*i.e.*, VMT increase) from the receipt of treatment (*i.e.*, residing in a conventional neighborhood)
- Typical measures
  - Average Treatment Effect (ATE)
  - Effect of Treatment on the Treated (TT)

Treatment Effects

- ATE - expected VMT increase for a random household if it were to reside in a conventional neighborhood as opposed to a neo-urbanist neighborhood.
- TT- average impact of treatment on the treated
- TNT - average impact of treatment on the non-treated
- TTNT - average impact of treatment on the (currently) treated and (currently) non-treated

Treatment Effects

Copulas	Independence-Independence Copula (I-I)	FGM-Joe Copula (FGM-J)	Frank-Joe Copula (F-J)	Frank-Frank Copula (F-F)
ATE	0.49 (1.75)	10.75 (1.03)	19.99 (4.42)	21.37 (5.21)
TT	3.04 (1.49)	31.04 (3.30)	42.45 (7.46)	41.76 (8.16)
TNT	-8.38 (1.38)	-31.55 (10.06)	-33.66 (10.82)	-30.74 (9.55)
TTNT	0.49 (1.75)	17.07 (0.88)	25.46 (3.03)	25.59 (4.75)

Summary

- A copula based approach to model residential neighborhood choice and daily household vehicle miles of travel (VMT)

- A variety of copula-based models are estimated
  - Results indicate that using a bivariate normal dependency structure suggests the absence of residential self-selection effects.
  - However, other copula structures reveal a high and statistically significant level of residential self-selection
  - Frank-Frank copula yields the best results

- The Treatment measures clearly provide the influence of self-selection
  - 83% due to built environment and 17% due to self-selection
  - If we could redesign the urban region as a neo-urban neighborhood it will reduce overall mileage by 43%

Summary

- Copula approach offers flexibility in modeling joint choices
- For other examples involving copula based methodology check my website