Copulas for Tunable Markov Processes

JMM January 2010

Kobi Abayomi [1] Lee Hawkins [2]

[1] Asst. Professor
Industrial Engineering - Statistics Group
Georgia Institute of Technology

[2] Wall Street Journal and CNBC

October 2009

Setup

The Parametric Copula - a marginally fixed distributional representation - has a role in the Chapman-Kolmogorov equations, the engines of Markovian Processes. We look to investigate, via tunable depictions of Markov Chains using Copulas, heuristics for career paths for African-American males. Specifically, we look to compare salary outcomes for black male M.B.A.'s vs. black male professional Athletes.

- Copula approach
- Copula as Markov Chain
- Application

Setup

The Parametric Copula - a marginally fixed distributional representation - has a role in the Chapman-Kolmogorov equations, the engines of Markovian Processes. We look to investigate, via tunable depictions of Markov Chains using Copulas, heuristics for career paths for African-American males. Specifically, we look to compare salary outcomes for black male M.B.A.'s vs. black male professional Athletes.

- Copula approach
- Copula as Markov Chain
- Application

Setup

The Parametric Copula - a marginally fixed distributional representation - has a role in the Chapman-Kolmogorov equations, the engines of Markovian Processes. We look to investigate, via tunable depictions of Markov Chains using Copulas, heuristics for career paths for African-American males. Specifically, we look to compare salary outcomes for black male M.B.A.'s vs. black male professional Athletes.

- Copula approach
- Copula as Markov Chain
- Application

Name the author

I could play Basketball, with a consuming passion...On the basketball court I could find a community of sorts, with an inner life all its own..., on turf where blackness couldn't be a disadvantage.

Name the author

I could play Basketball, with a consuming passion...On the basketball court I could find a community of sorts, with an inner life all its own..., on turf where blackness couldn't be a disadvantage. ***Obama, Barack (1995) Dreams From My Father***

Outline

A copula is a distribution function...

Take $X_1 \sim F_{X_1}$, $X_2 \sim F_{X_2}$

...on the marginal distributions of random variables

Set $U = F_{X_1}$ and $V = F_{X_2}$.

- The pair (U, V) are the '*grades*' of (X_1, X_2)
- i.e. the mapping of (X_1, X_2) in F_{X_1}, F_{X_2} space.

...on the marginal distributions of random variables

Set $U = F_{X_1}$ and $V = F_{X_2}$.

- The pair (U, V) are the '*grades*' of (X_1, X_2)
- i.e. the mapping of (X_1, X_2) in F_{X_1}, F_{X_2} space.

Copula specification

A copula is a function that takes the 'grades' as arguments and returns a joint distribution function, with marginals F_{X_1}, F_{X_2}.

$$C(U, V) = F_{X_1, X_2}$$

Copula generation

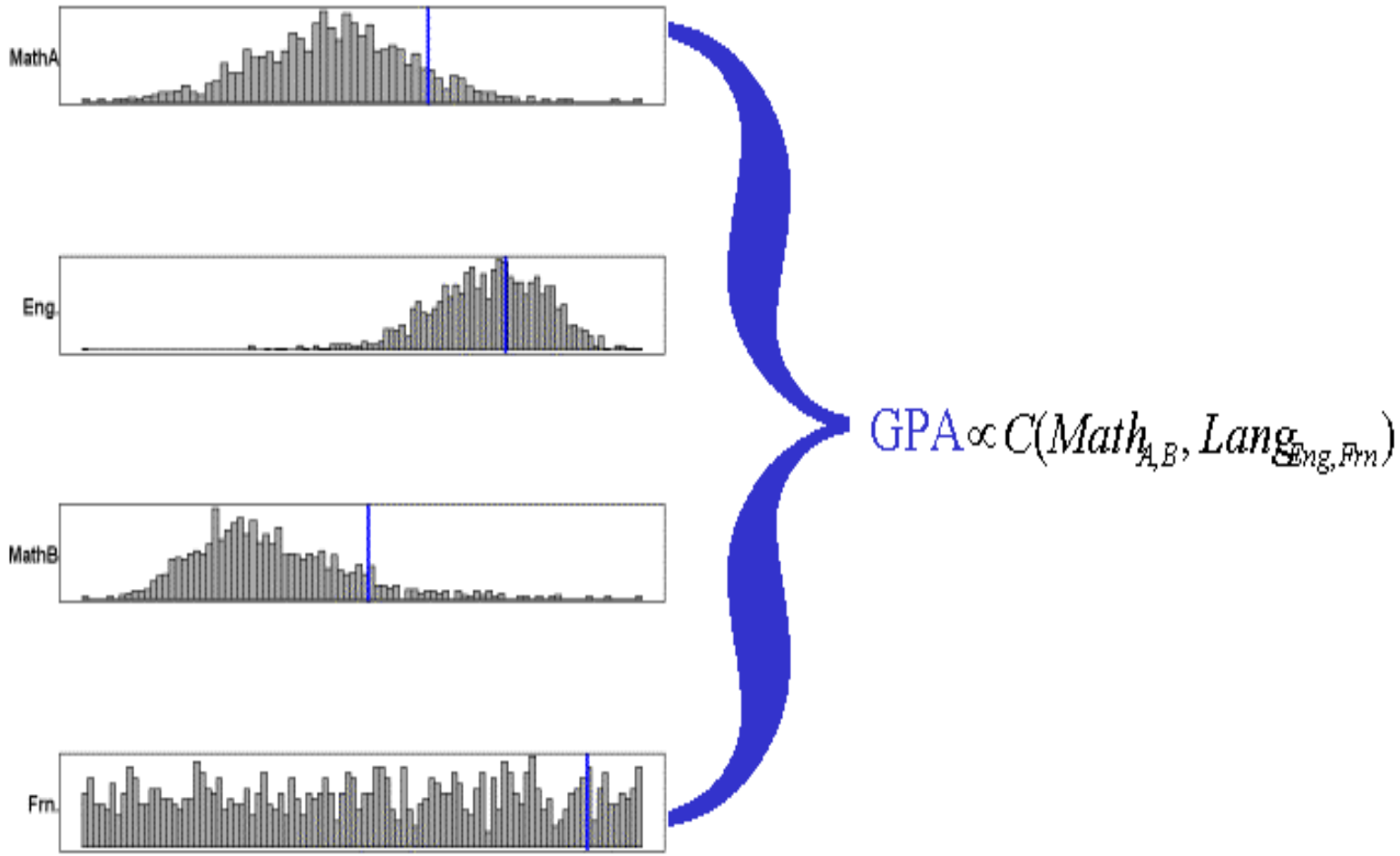
Any multivariate distribution function can yield a copula function.

$$F_{X_1,X_2}(F_{X_1}^{-1}(U), F_{X_2}^{-1}(V)) = C'(U, V)$$

Heuristically speaking

The correspondence which assigns the value of the joint distribution function to each ordered pair of values (F_{X_1}, F_{X_2}) for each X_1, X_2 is a distribution function called a Copula.

GPA



Simple Example: Bivariate Distribution

$$H_\theta(x_1, x_2) = 1 - e^{-x_1} - e^{-x_2} + e^{-(x_1+x_2+\theta x_1 x_2)}$$

for $x_1, x_2 \in \mathbb{R}^+$. $H = 0$ otherwise.

Simple Example: Bivariate Distribution

$$H_{X_1}^{(-1)}(u) = -\ln(1-u); H_{x_2}^{(-1)}(v) = -\ln(1-v)$$

Simple Example: Bivariate Distribution

$$C_\theta(u, v) = H(-ln(1-u), -ln(1-v)) =$$

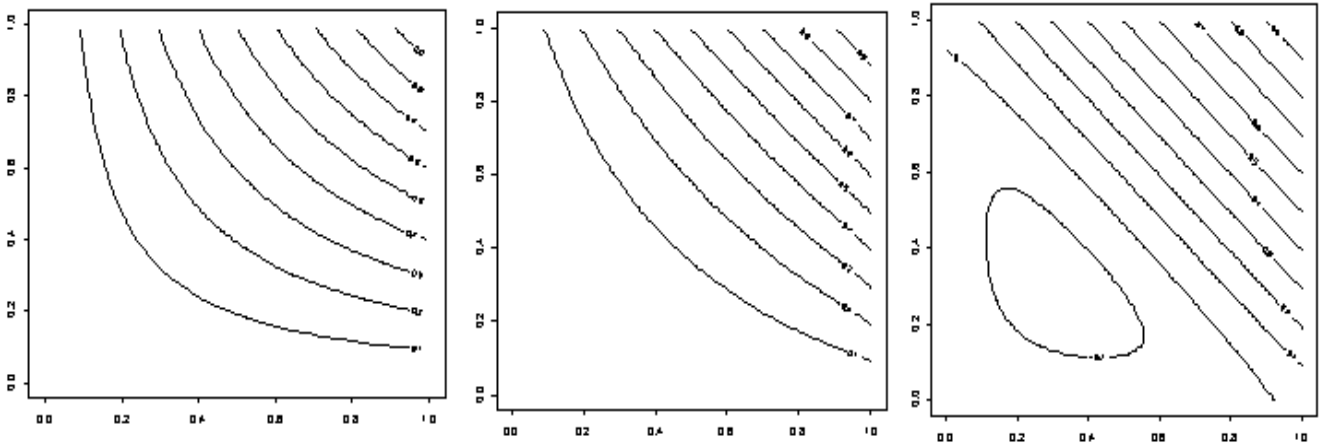
$$= (u + v - 1) + (1 - u)(1 - v) * e^{-\theta \ln(1-u) \ln(1-v)}$$

Notice if $\theta = 0$

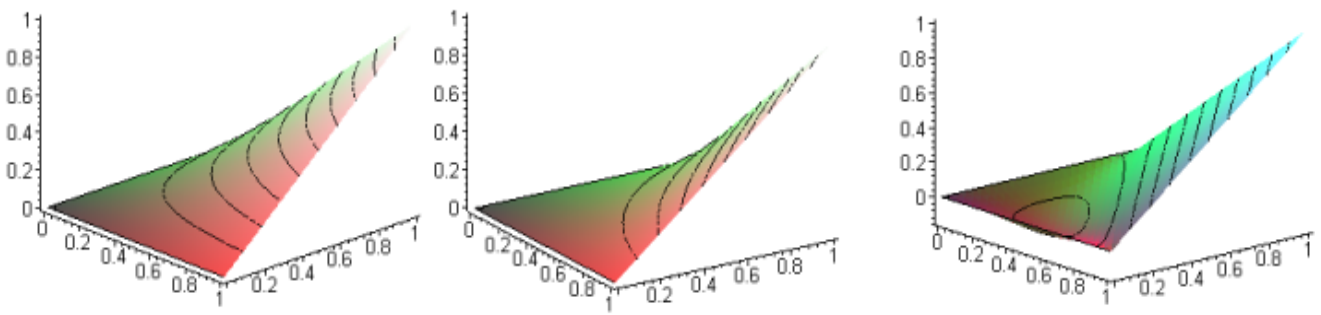
$$C_\theta(u, v) = uv$$

...the independence copula.

Bivariate Copula $\theta = 0, 1, 5$



$\theta = 0$ $\theta = 1$ $\theta = 5$



Outline

Markov Process

Remember a Markov Process, on some state space, is an indexed set of random variables

$$X_1, ..., X_T$$

where

$$f_{X_1,...,X_T}(X_1, ..., X_T) = f_{X_1}(X_1) \cdot f_{X_2|X_1}(X_2|X_1) \cdots f_{X_T|X_{T-1}}(X_T|X_{T-1})$$

The $f_{X_i|X_j}(X_i|X_j)$ are *transition probabilities* from *index j* to index *i*

Markov Process

A Chapman-Kolmogorov Equation is

$$f_{X_{t_1},\ldots,X_{t_n}} = \int_{-\infty}^{\infty} f_{X_{t_n}|X_{t_{n-1}}}(X_{t_n}|X_{t_{n-1}}) \cdots f_{X_{t_2}|X_{t_1}}(X_{t_2}|X_{t_1}) dX_{t_2} \cdots dX_{t_{n-1}}$$

$*$ Notation

Following Darsow, Nguyen, Olsen (1991)
Define

$$(A * B)(x_1, x_2) = \int_0^1 \frac{\partial A(x_1, t)}{\partial x_2} \cdot \frac{\partial B(t, x_2)}{\partial x_1} dt$$

for A, B copulas and x_1, x_2 in I

CK Equivalence

'But' since, for $X_1, X_2 \sim F_{X_1}, F_{X_2}, C$

$$\mathbb{P}(X_1 < x_1 | X_2 = x_2) = \frac{\partial C(F_{X_1}, F_{X_2})}{\partial X_2}$$

and

$$\mathbb{P}(X_2 < x_2 | X_1 = x_1) = \frac{\partial C(F_{X_1}, F_{X_2})}{\partial X_1}$$

CK Equivalence

Then,
For any three random variables X_1, X_2, X_3,
where $(X_1 \perp X_3)|X_2$

$$C_{X_1X_3} = C_{X_1X_2} * C_{X_2X_3}$$

CK Equivalence

Calling $C_{t_i t_j}$ the copula of the random variables X_{t_i}, X_{t_j}
Then, for $t_i < t_j < t_k$

$$C_{t_i t_k} = C_{t_i t_j} * C_{t_j t_k} \tag{1}$$

is an equivalent representation of the CK equations.

CK Equivalence

And

$$\mathbb{P}(X_t \in A | X_s = x) = \frac{\partial C_{st}(F_s(x), F_t(a))}{\partial X_s}$$

is the copula version of the CK transition probability

CK Equivalence - Summary

Specifying a Markov Process in the 'conventional' approach

- Set initial distributions F_0
- Specify family of transition probabilities $f_{X_i|X_j}(X_i|X_j)$

As an estimation problem, the goal is to estimate the transition probabilities from data.

CK Equivalence - Summary

Specifying a Markov Process in the 'conventional' approach

- Set initial distributions F_0
- Specify family of transition probabilities $f_{X_i|X_j}(X_i|X_j)$

As an estimation problem, the goal is to estimate the transition probabilities from data.

CK Equivalence - Summary

Via the Copulae

- Specify the marginal distributions for each state $F_{X_1}, ..., F_{X_m}$
- Specify family of 2-copulas satisfying (1)

As an estimation problem the goal is to estimate the *copulae*, i.e. the transition dependence between states, from data.

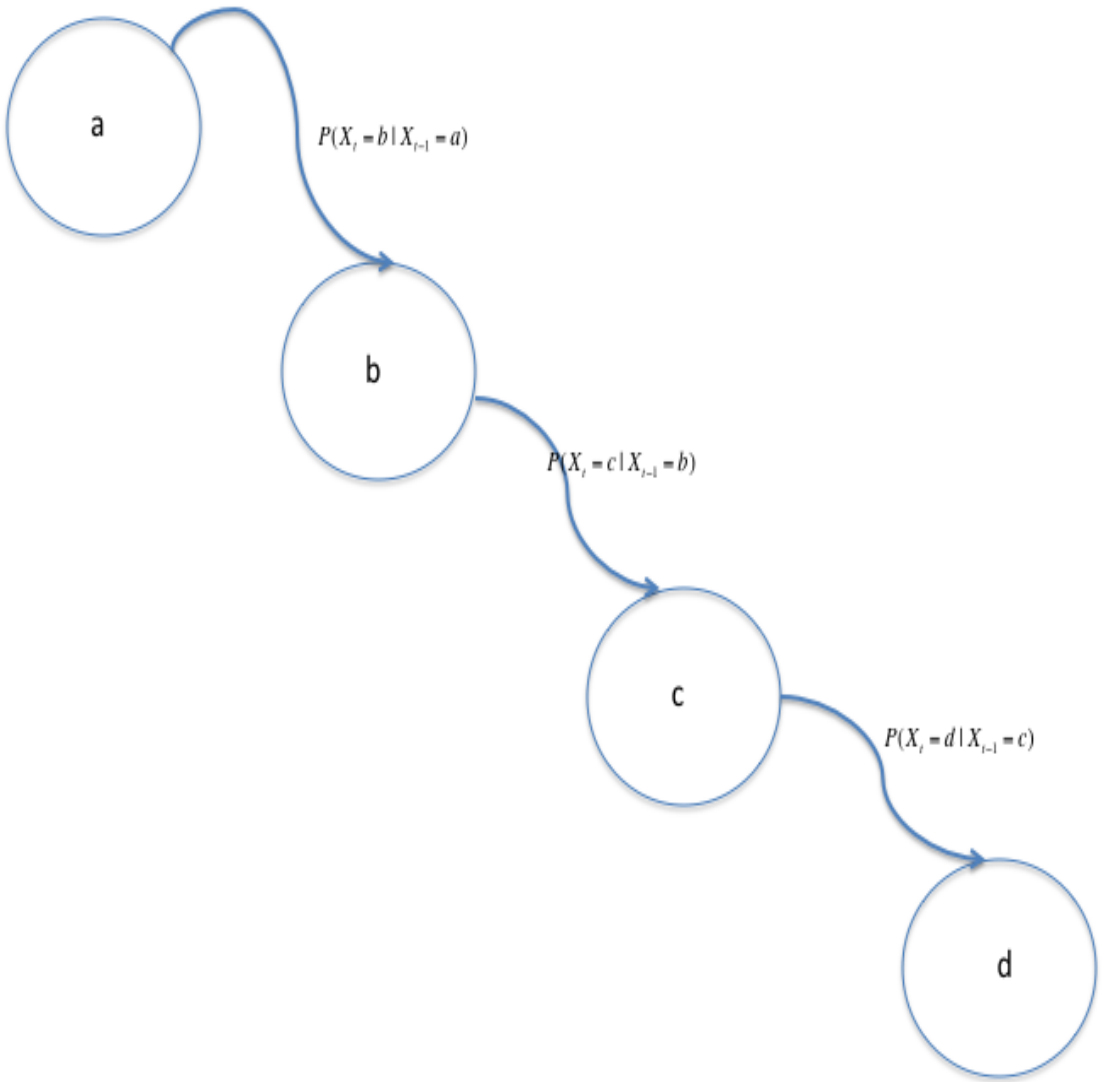
CK Equivalence - Summary

Via the Copulae

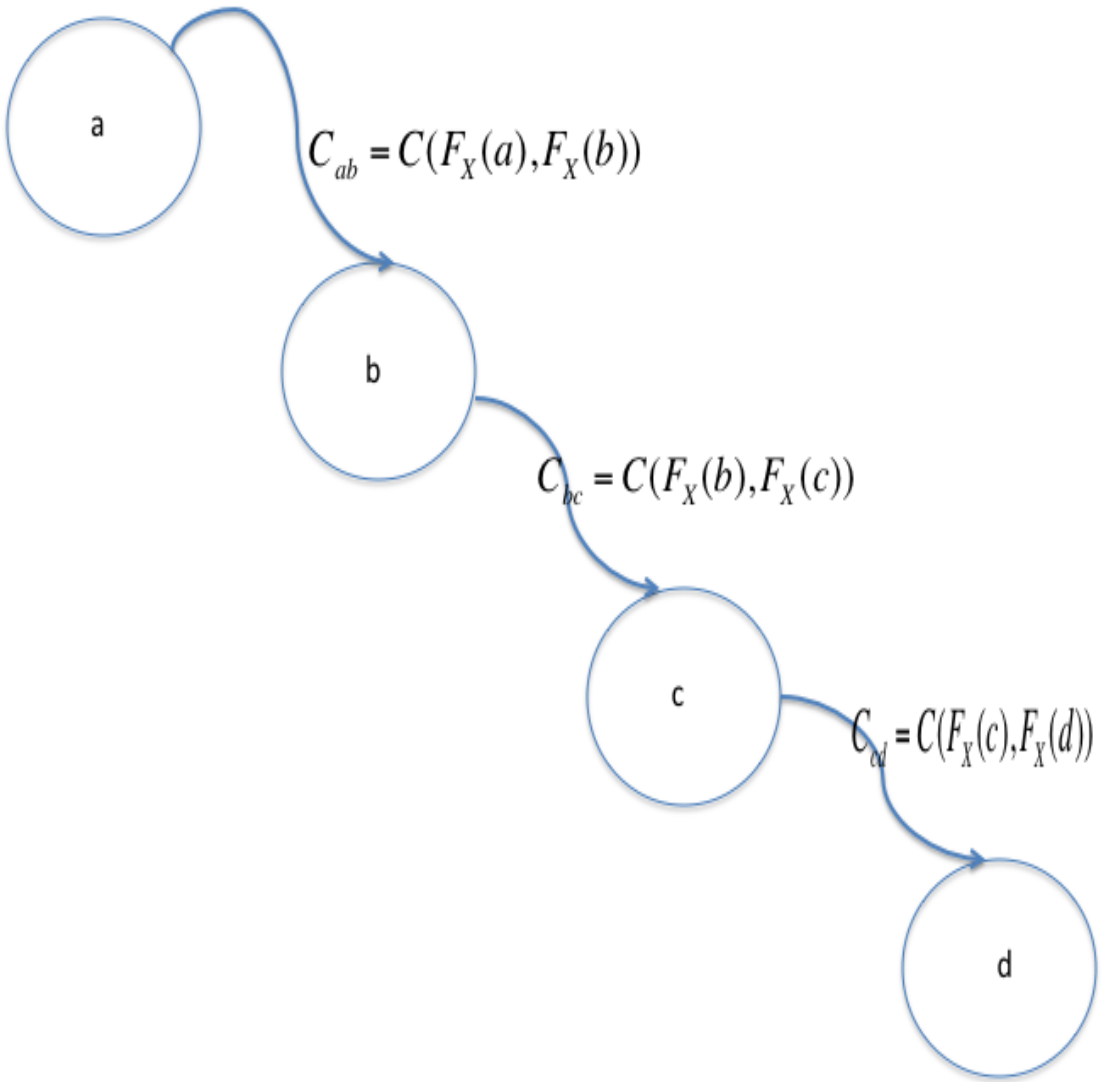
- Specify the marginal distributions for each state $F_{X_1}, ..., F_{X_m}$
- Specify family of 2-copulas satisfying (1)

As an estimation problem the goal is to estimate the *copulae*, i.e. the transition dependence between states, from data.

Illustration



Illustration



Features of copula approach for MP/MC

Estimation problems

- Where marginal distributions are available for each state
- Where initial distribution is non-informative
- Where parametric models for *transition dependence* are desirable

Features of copula approach for MP/MC

Estimation problems

- Where marginal distributions are available for each state
- Where initial distribution is non-informative
- Where parametric models for *transition dependence* are desirable

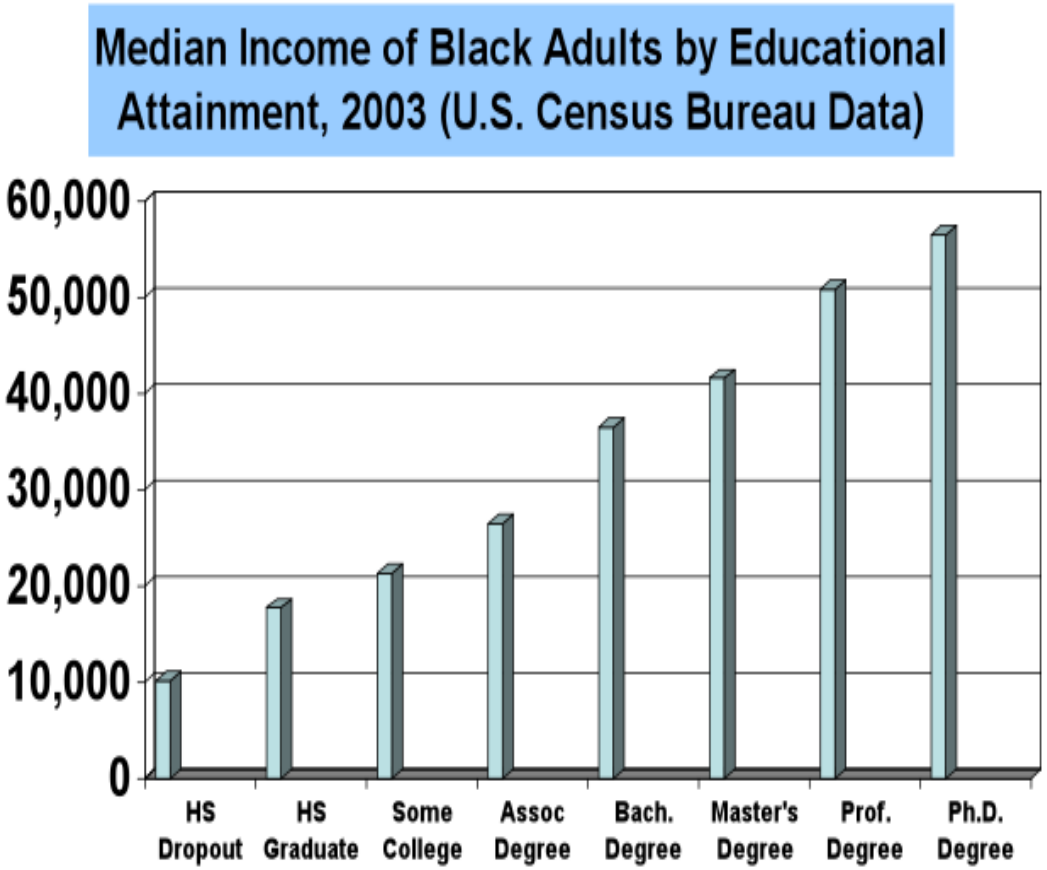
Features of copula approach for MP/MC

Estimation problems

- Where marginal distributions are available for each state
- Where initial distribution is non-informative
- Where parametric models for *transition dependence* are desirable

Outline

Comparison of salary attainment for (African American Male) MBA's vs. Professional Athletes.



Setup

- Compare outcomes for professions/individuals at 'extremes' of salary distribution
- Marginal data at each state (educational/employment waystation)
- Transitional Dependence — the copula parameter estimates — is a proxy for throughput

Setup

- Compare outcomes for professions/individuals at 'extremes' of salary distribution
- Marginal data at each state (educational/employment waystation)
- Transitional Dependence — the copula parameter estimates — is a proxy for throughput

Setup

- Compare outcomes for professions/individuals at 'extremes' of salary distribution
- Marginal data at each state (educational/employment waystation)
- Transitional Dependence — the copula parameter estimates — is a proxy for throughput

Initial Comments

Data from

- National Center on Educational Statistics [NCES]
- National Education Longitudinal Study [NELS] of 1988
- NCAA 2007 Probability of Competition Report
- Graduate Student researchers at Political Science, U Wisc.

Initial Comments

Data from

- National Center on Educational Statistics [NCES]
- National Education Longitudinal Study [NELS] of 1988
- NCAA 2007 Probability of Competition Report
- Graduate Student researchers at Political Science, U Wisc.

Initial Comments

Data from

- National Center on Educational Statistics [NCES]
- National Education Longitudinal Study [NELS] of 1988
- NCAA 2007 Probability of Competition Report
- Graduate Student researchers at Political Science, U Wisc.

Initial Comments

Data from

- National Center on Educational Statistics [NCES]
- National Education Longitudinal Study [NELS] of 1988
- NCAA 2007 Probability of Competition Report
- Graduate Student researchers at Political Science, U Wisc.

Some Anecdotal Numbers

- Total salaries for Black Male Professional (U.S.) athletes: $3.7 Billion (2007-2008)
- Total compensation for Black Male Fortune 500 CEOs (there were three! in 2007-2008): $96 Million

Some Anecdotal Numbers

- Total salaries for Black Male Professional (U.S.) athletes: $3.7 Billion (2007-2008)
- Total compensation for Black Male Fortune 500 CEOs (there were three! in 2007-2008): $96 Million

Very Initial Results

Using GH Copula ($0 \leq \theta \leq 1$) on NELS data; probabilities in percent:

	Professional Athlete	Salaried MBA
8th Grade	-	.2
HS Senior (B, F, Ba)	.03, .08, .45	-
College Senior (B, F, Ba)	1.2, 1.8, .9.4	-

$\hat{\theta}$ for 8th grade to MBA much less than $\hat{\theta}$ for other transitions.

Very Initial Comments

- Frequency of African American Males earning Ph.D. approx .1 pct
- Given participation in NCAA athletics, 10-90 times higher prob. of playing professional sports than an 8th grader does to earn Master's degree
- Prob of High School Baseball player, any race, playing professional ball is 2.25 greater than Af. Am 8th grader earning masters, 4.5 times Af. Am male going on to earn a professional degree (including Ph.D.)

Very Initial Comments

- Frequency of African American Males earning Ph.D. approx .1 pct
- Given participation in NCAA athletics, 10-90 times higher prob. of playing professional sports than an 8th grader does to earn Master's degree
- Prob of High School Baseball player, any race, playing professional ball is 2.25 greater than Af. Am 8th grader earning masters, 4.5 times Af. Am male going on to earn a professional degree (including Ph.D.)

Very Initial Comments

- Frequency of African American Males earning Ph.D. approx .1 pct
- Given participation in NCAA athletics, 10-90 times higher prob. of playing professional sports than an 8th grader does to earn Master's degree
- Prob of High School Baseball player, any race, playing professional ball is 2.25 greater than Af. Am 8th grader earning masters, 4.5 times Af. Am male going on to earn a professional degree (including Ph.D.)

Final comments

- Need Better/More/Recent Data - Af. Am MBA salary data shamefully hard to acquire
- Very initial findings suggest greater, persistent barriers to MBA (salary) path — $\hat{\theta}$ low
- Heuristic for *not* staying in school?

Final comments

- Need Better/More/Recent Data - Af. Am MBA salary data shamefully hard to acquire
- Very initial findings suggest greater, persistent barriers to MBA (salary) path — $\hat{\theta}$ low
- Heuristic for *not* staying in school?

Final comments

- Need Better/More/Recent Data - Af. Am MBA salary data shamefully hard to acquire
- Very initial findings suggest greater, persistent barriers to MBA (salary) path — $\hat{\theta}$ low
- Heuristic for *not* staying in school?