Extended Twin Family Designs: The Motivation for Using Them

Matthew Keller

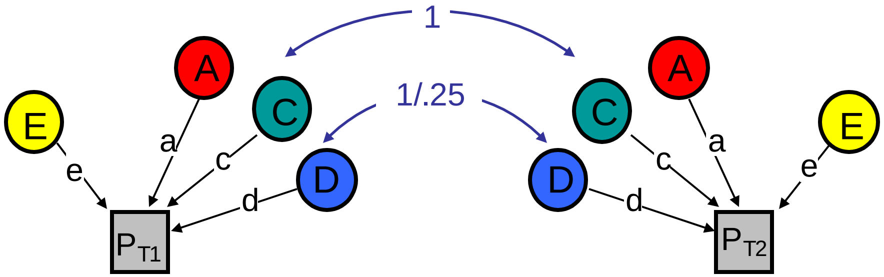
International Statistical Genetics Workshop 2022

The point of this lecture

- Extending the Classical Twin Design (CTD) to include additional relative types not only allows estimation of additional quantities, but it also probably produces less biased estimates
- Unlike the effects of violations of assumptions in the CTD, understanding the influence of biased assumptions in ETFDs requires simulation.

Classical Twin Design (CTD)

Assumption	biased up	biased down
Either V_D or V_C is zero	$\hat{V}_A$	$\hat{V}_C$ & $\hat{V}_D$
No assortative mating	$\hat{V}_C$	$\hat{V}_D$ & $\hat{V}_A$
No A-C covariance	$\hat{V}_C$	$\hat{V}_A$ & $\hat{V}_D$



Assortative mating consequence on V_A

- AM: phenotypic correlation between mating partners
- Many examples (e.g., height ~.25; IQ ~ .4; Social attitudes ~ .6)
- If AM leads to genetic similarity in partners (as it does if it's primary phenotypic AM), there are genetic consequences:
  - Height V_A increases in the population because 'tall' ('short') alleles are more concentrated in individuals than expected.
  - E.g., if you're a 'tall' allele sitting in an egg and are waiting around to see what other height alleles you'll get paired with from that sperm swimming to you, they are more likely than chance to be other 'tall' alleles (both at the same locus and at others; & this just considers the effects on V_A in 1st gen)

$$V_{A.equil} = \frac{V_{A0}}{1 - rh^2_{equil}}$$

AM consequence on relative covariance

- AM increases genetic covariances and correlations between relatives (e.g., sibs, parents, cousins, etc). E.g.:

$$r_{A,DZ} = \frac{1}{2}(1 + rh^2_{equil})$$

- Consider again being a ‘tall’ allele in a zygote. This time you are watching your co-twin’s zygote get formed. Regardless of whether you exist (are IBD) in your co-twin’s zygote, you can expect more tall alleles in your co-twin’s zygote.
- Thus, you can also expect to share more ‘tall’ alleles with your sibling(s) than expected under random mating.
- AM biases $\hat{V}_C$ upwards and $\hat{V}_A$ (or $\hat{V}_D$) downwards
  - $r_{MZ.A1,A2}$ is already 1 so AM cannot increase it
  - AM will increase $r_{DZ.A1,A2} > .50$

Passive G-E Covariance (aka “genetic nurture”)

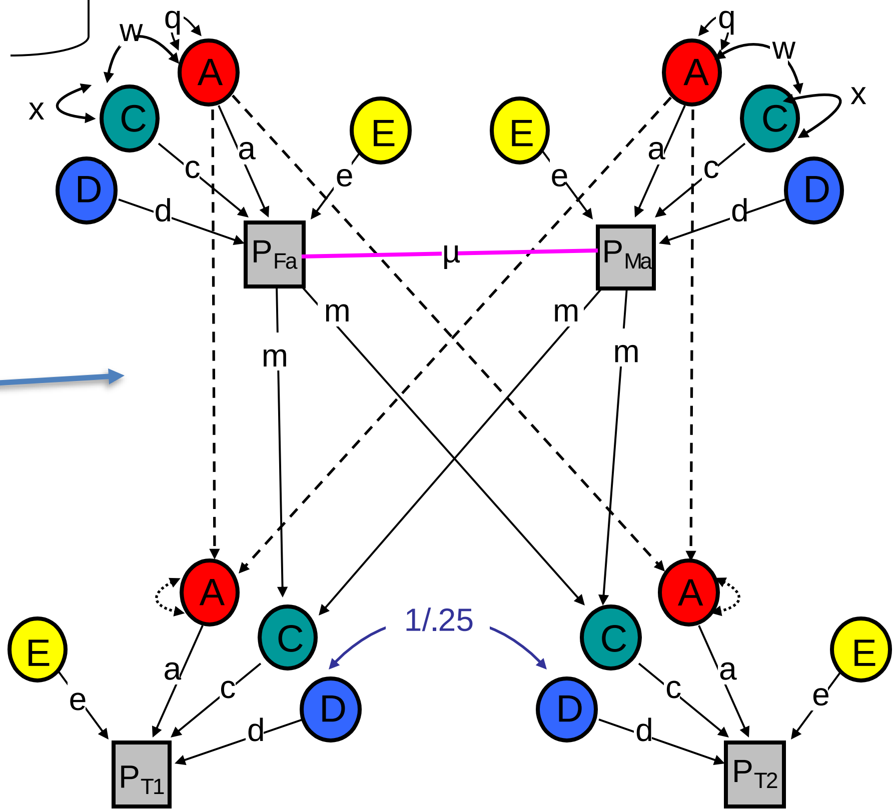
- Covariance between the average genetic effect in a family and the “familial environment” that occurs when parental trait directly influences the offspring trait, leading to covariance between genes affective a trait and the offspring rearing environment
- E.g., educated parents pass on both genes and environments that predispose to higher education
- Because it is a covariance between the *average* genetic effect and the familial environment, it increases the covariance of all siblings by the same amount regardless of their relatedness (e.g., MZ vs. DZ). It therefore mimics, and leads to an overestimate, of V_C in the CTD.

Adding parents gets us around all these assumptions

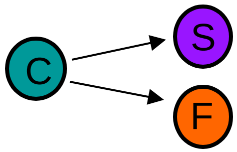
Assumption	biased up	biased down
Either V_D or V_C is zero		
No assortative mating		
No A-C covariance		

We don't have to make these

Nuclear Twin Family Design



We can model V_C as either V_S or V_F

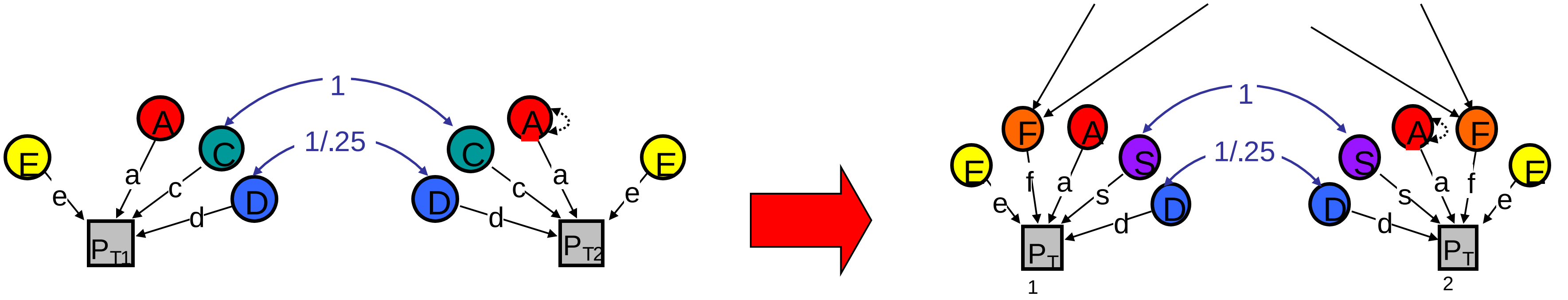


With parents, we can break V_C up into:

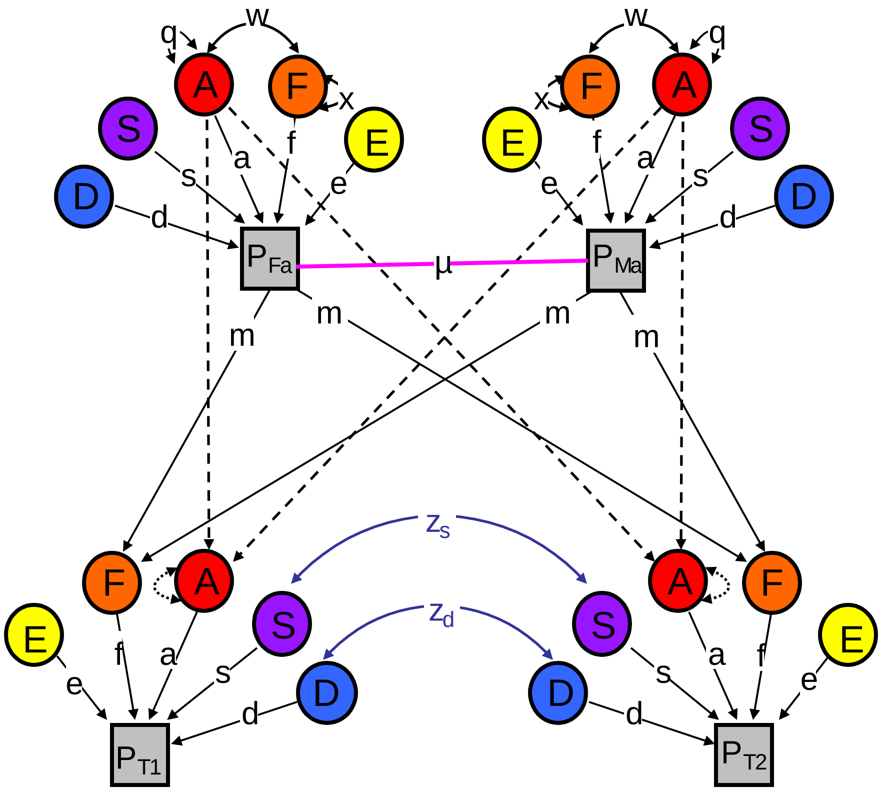
S = env. factors shared only between **sibs**

F = **familial** env factors passed from parents to offspring

But we can only estimate one of these (or more technically, three of V_A, V_S, V_F, & V_D)



Nuclear Twin Family Design (NTFD)



Note: m estimated and f fixed to 1

On average across 38 traits CTD vs. ETFD results*

- $\hat{V}_A$ 65% higher in CTD
- $\hat{V}_D$ 43% lower in CTD

$\hat{V}_G$ 18% higher in CTD

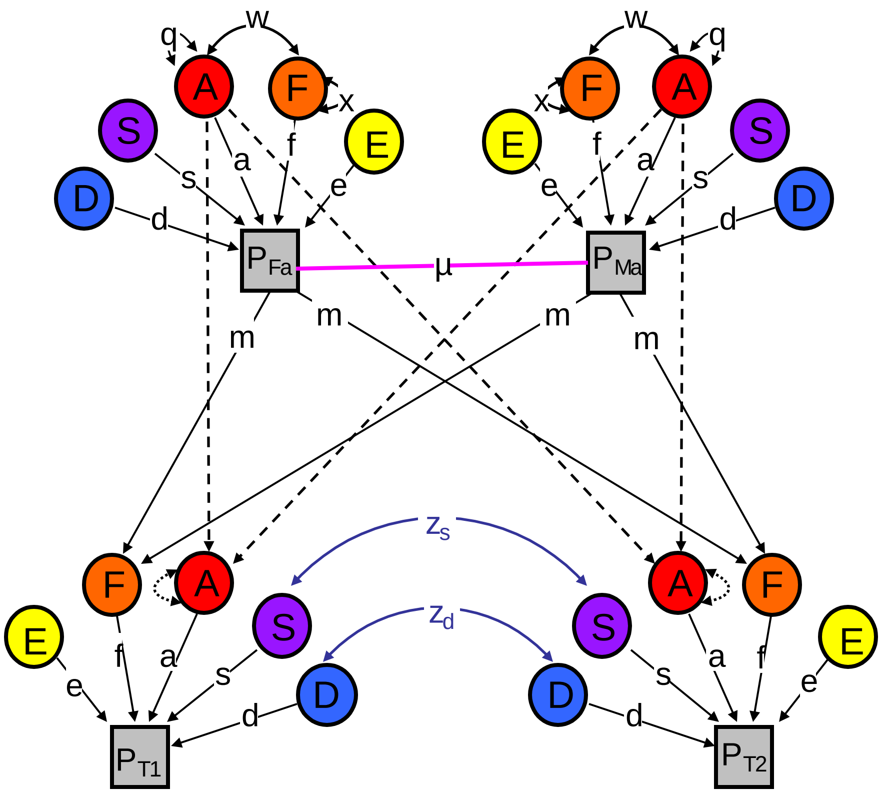
- $\hat{V}_C$ 45% lower in CTD when r(spouse)~0
- $\hat{V}_C$ 100% higher in CTD when r(spouse)>0

- ETFD results are not perfect, but theory and simulation suggest they are, on average, much more accurate than CTD results.
  - Accuracy across all sims: CTD=.14; NTF=.07; ETFD=.045

---

* Coventry & Keller, 2005

Nuclear Twin Family Design (NTFD)



Note: m estimated and f fixed to 1

- Assumptions:
  - Only can estimate 3 of 4: V_A, V_D, V_S, V_F (bias is variable)
  - Assortative mating due to primary phenotypic assortment (bias is variable)

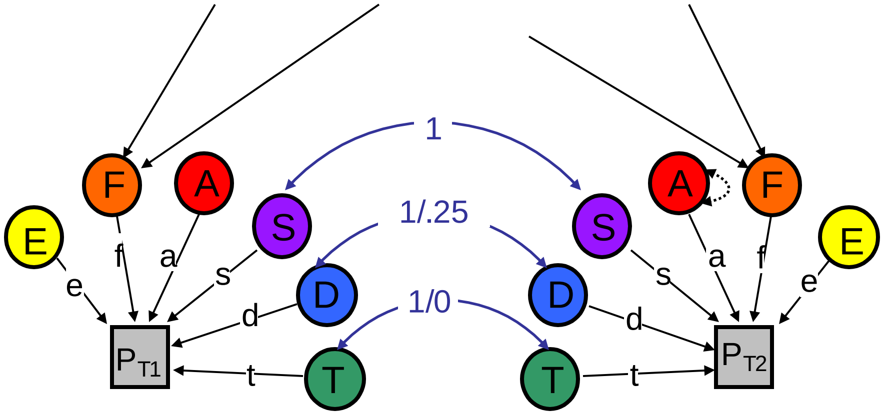
Stealth

- Include twins and their sibs, parents, spouses, and offspring…
  - Gives 17 unique covariances (MZ, DZ, Sib, P-O, Spousal, MZ avunc, DZ avunc, MZ cous, DZ cous, GP-GO, and 7 in-laws)
  - 88 covariances with sex effects

Additional obs. covs with *Stealth* allow estimation of V_A, V_S, V_D, V_F, V_T

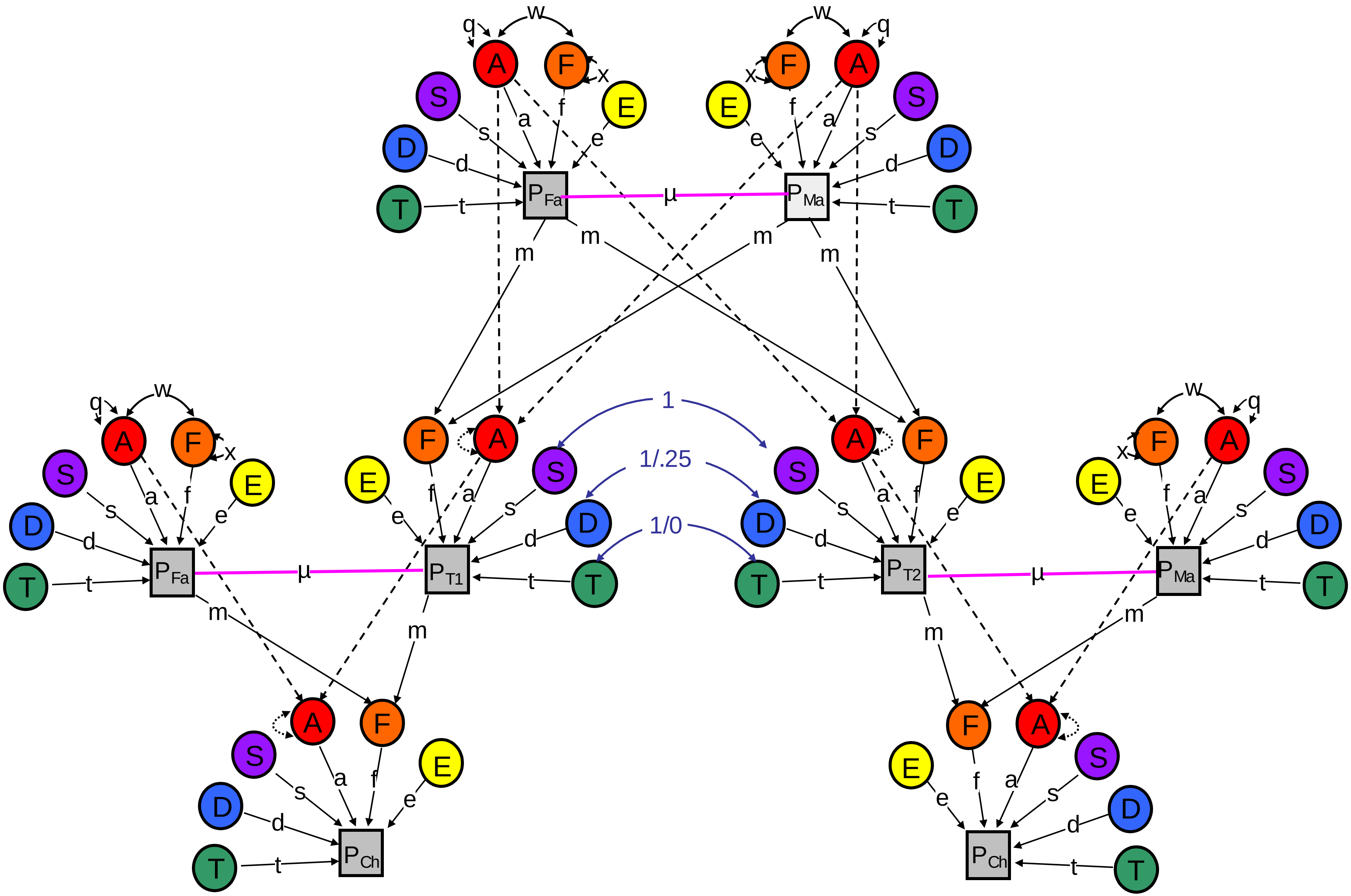
A S F D T can be estimated simultaneously

T = env. factors shared only between **twins**



(Remember: we're not just estimating more effects. More importantly, we're reducing the bias in estimated effects – although perhaps at the expense of more variance in estimates)

Stealth



Stealth

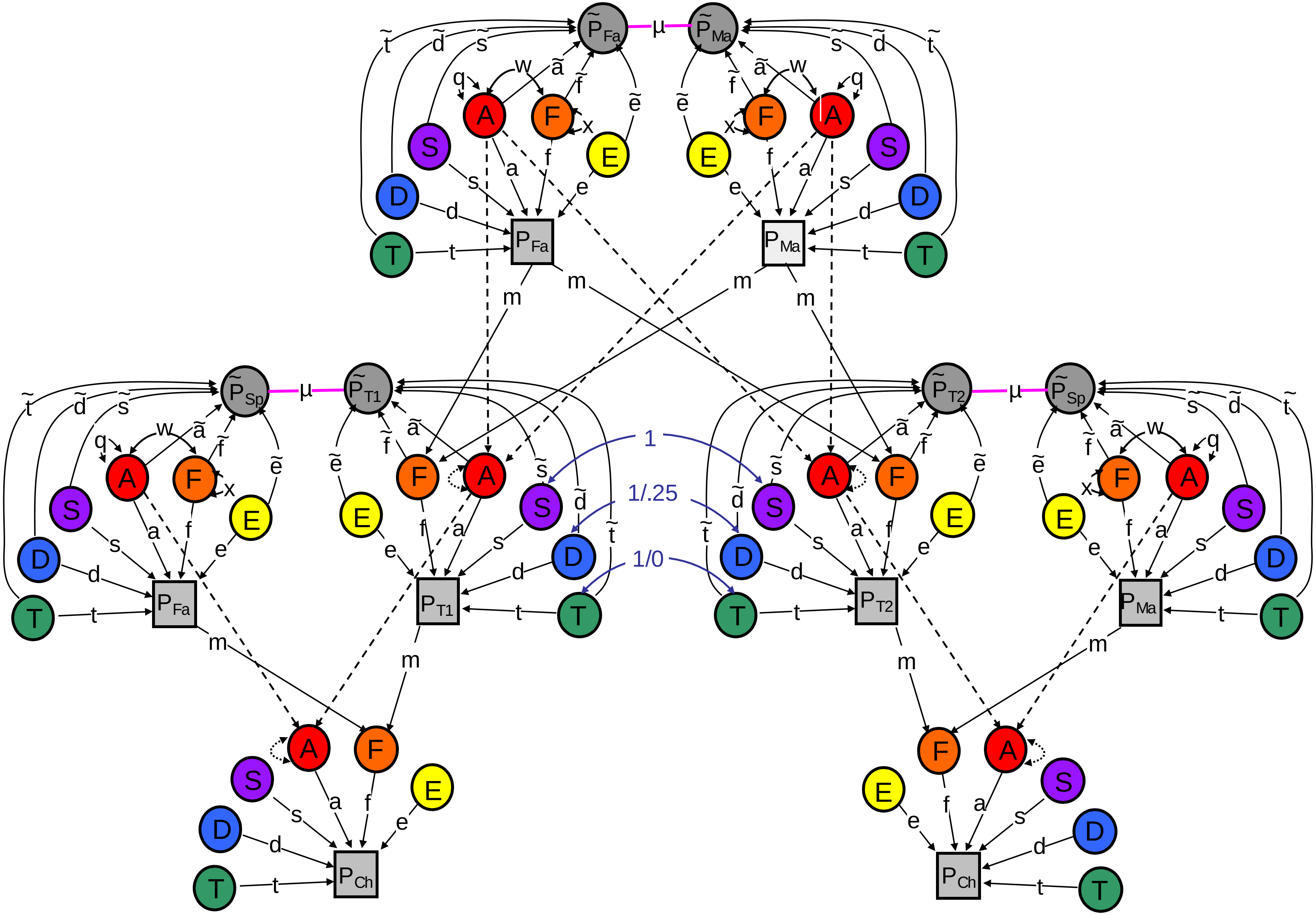
Assumption	biased up	biased down
Primary assortative mating	$\hat{V}_A, \hat{V}_D, \hat{V}_F$	$\hat{V}_A, \hat{V}_D, \hat{V}_F$
No epistasis	$\hat{V}_A, \hat{V}_D$	$\hat{V}_S$
No AxAge	$\hat{V}_D, \hat{V}_S$	$\hat{V}_A$

Stealth

Assumption	biased up	biased down
Primary assortative mating	$\hat{V}_A, \hat{V}_D, \hat{V}_F$	$\hat{V}_A, \hat{V}_D, \hat{V}_F$
No epistasis	$\hat{V}_A, \hat{V}_D$	$\hat{V}_S$
No AxAge	$\hat{V}_D, \hat{V}_S$	$\hat{V}_A$

- Phenotypic homogamy (primary AM): matching based on phenotypic similarity
- Social homogamy: matching based on environmental similarity (e.g., religion)
- Convergence: mates become more similar to each other (e.g., becoming more conservative when dating a conservative)

Cascade



$\tilde{P}_{Fa}$
μ
$\tilde{P}_{Ma}$
$\tilde{t}$
$\tilde{d}$
$\tilde{s}$
$\tilde{a}$
$\tilde{f}$
$\tilde{e}$
w
q
x
A
F
S
E
D
T
a
f
s
e
d
t
P_{Fa}
P_{Ma}
m
$\tilde{P}_{Sp}$
$\tilde{P}_{T1}$
$\tilde{P}_{T2}$
P_{T1}
P_{T2}
1
1/.25
1/0
P_{Ch}

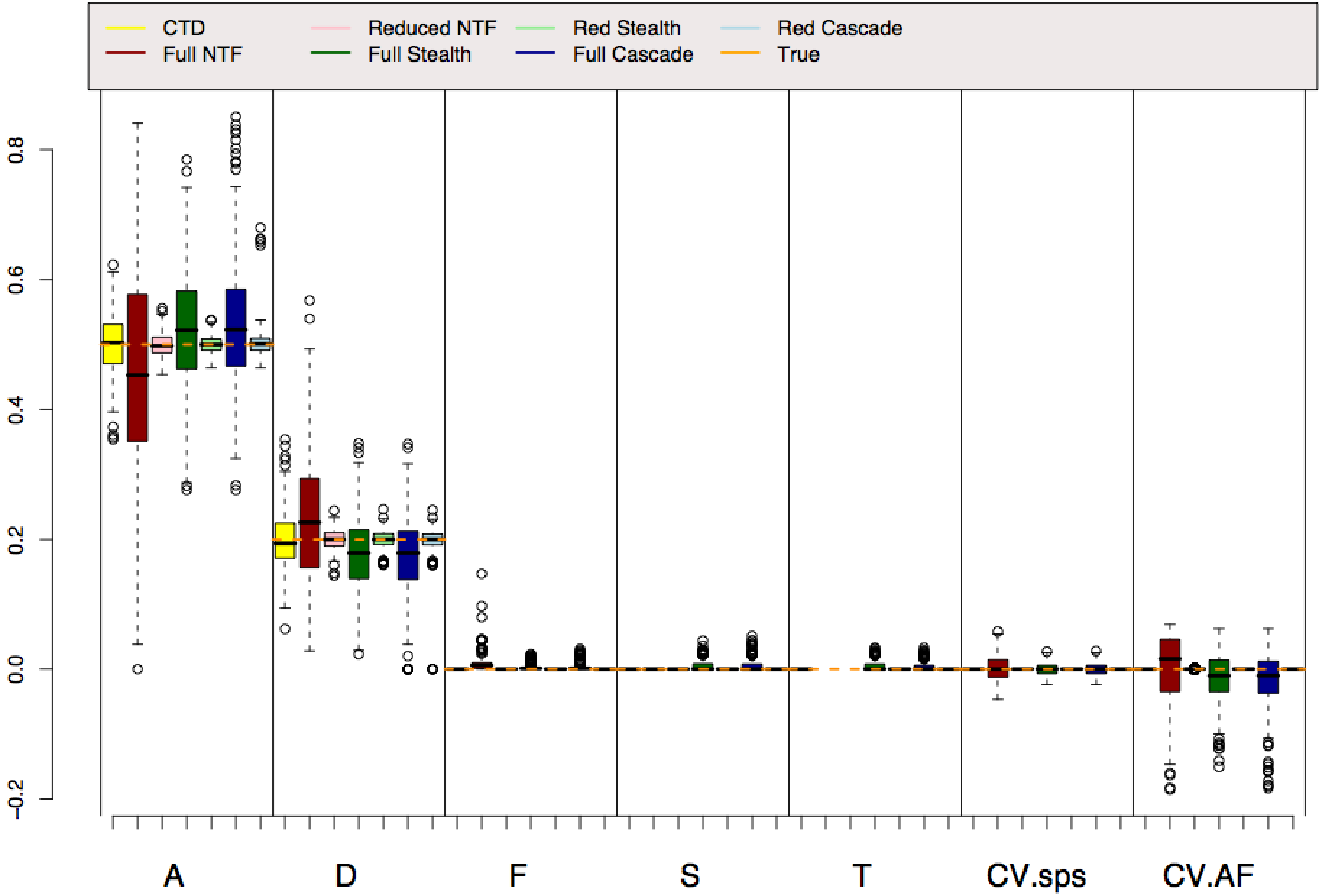
Simulation program: GeneEvolve

Get GeneEvolve:

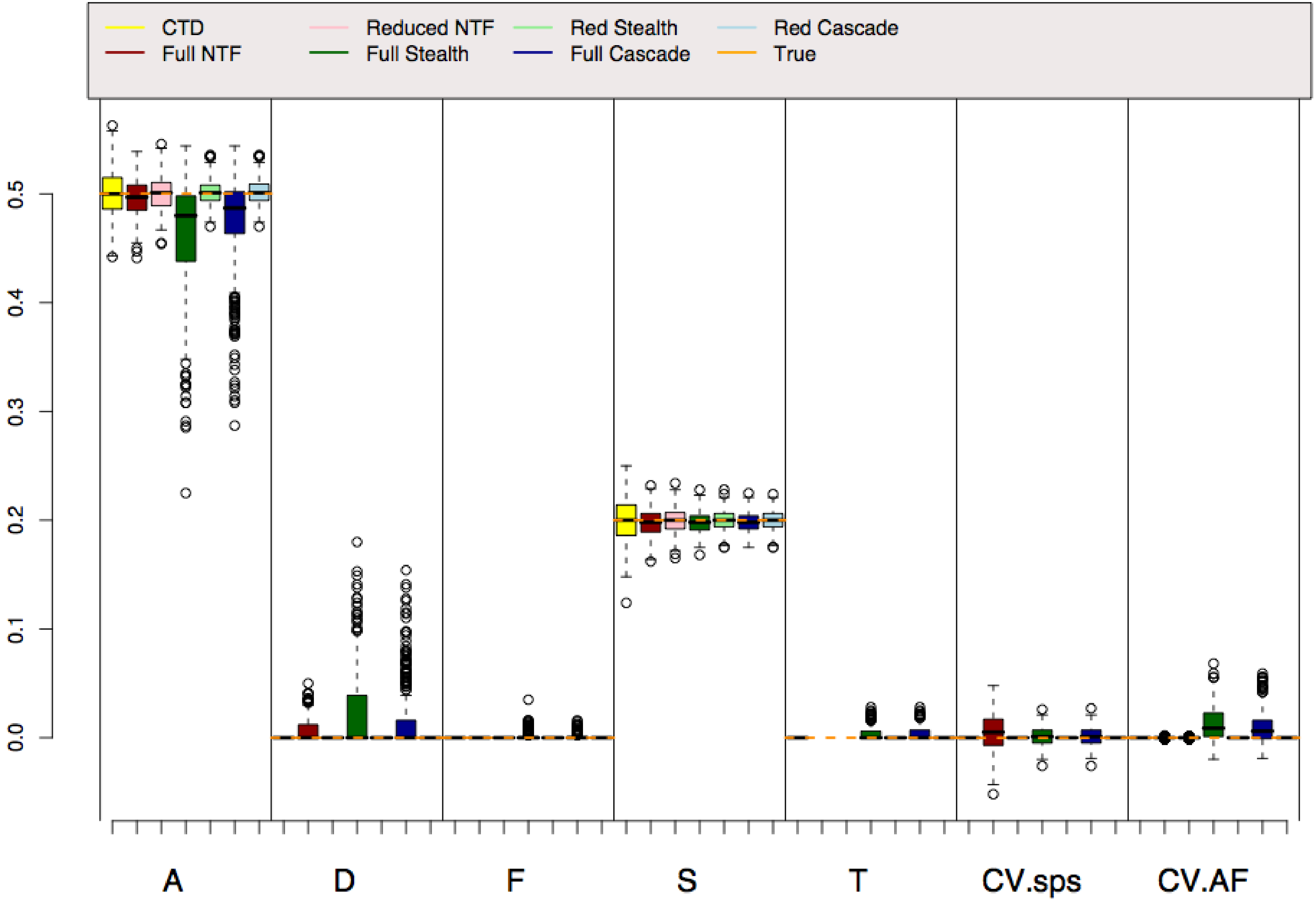
- GeneEvolve65.zip

Note: *GeneEvolve* is still a 'beta-version'. Breakdowns are likely! You can help by

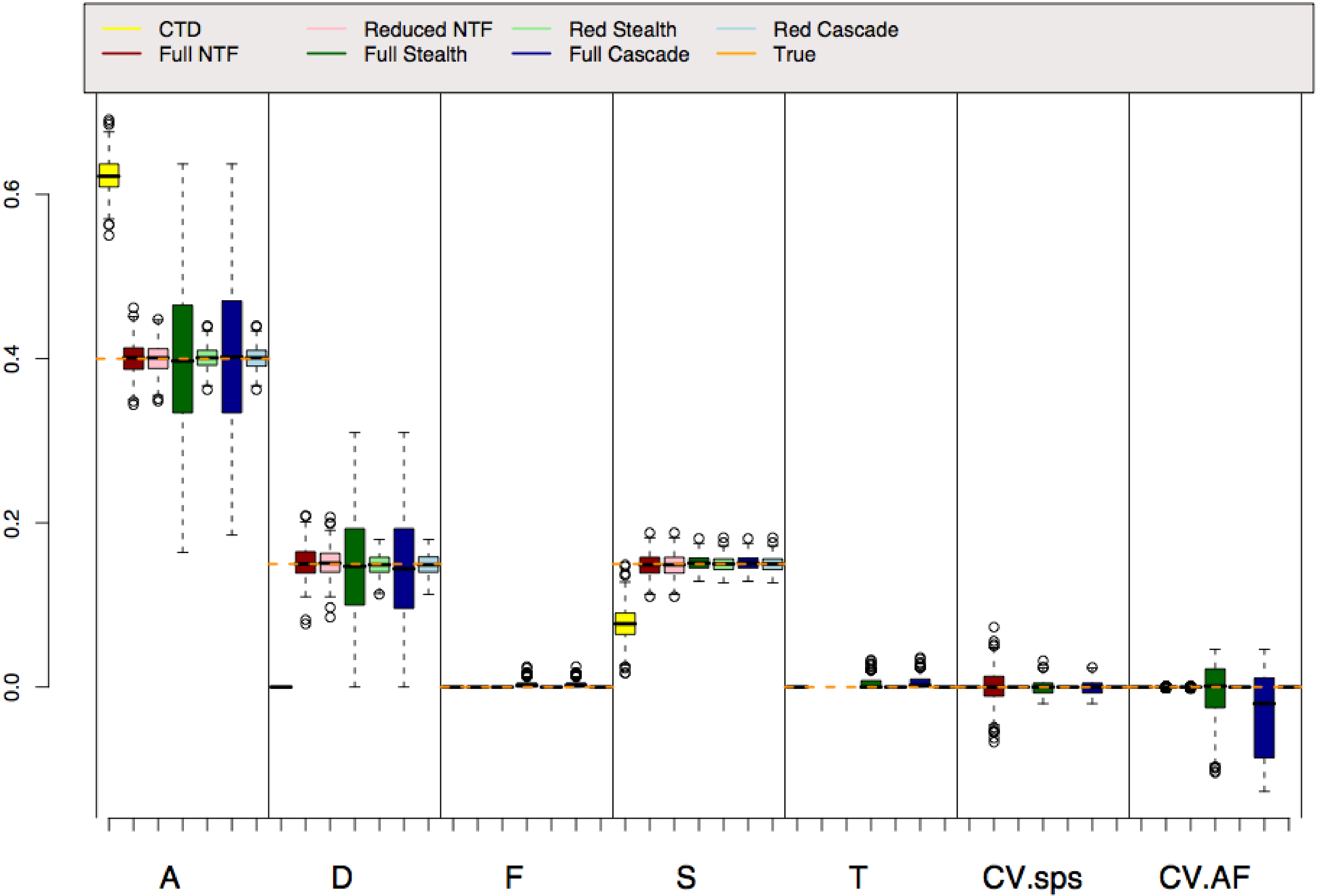
Reality: V_A =.5, V_D=.2



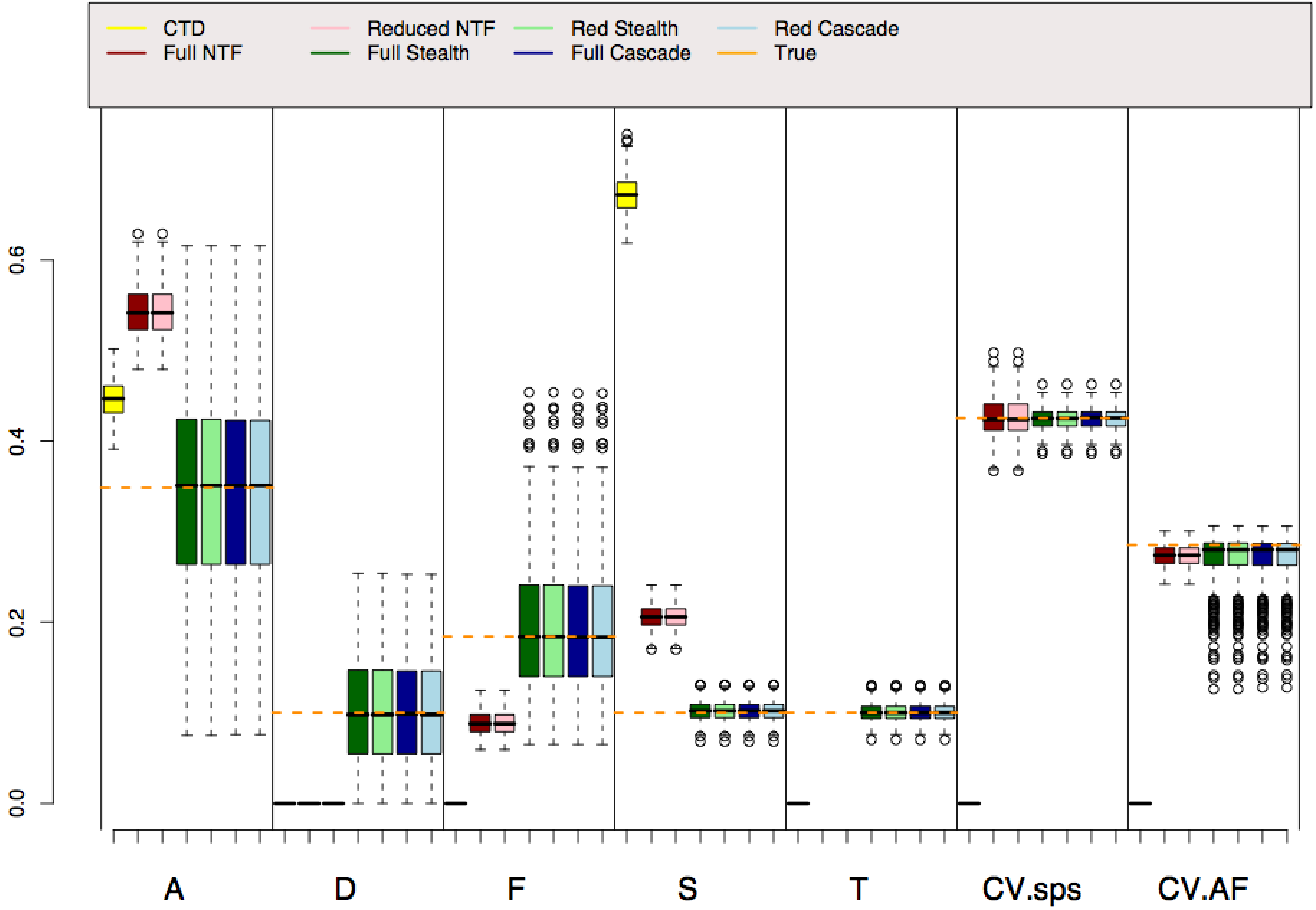
Reality: V_A =.5, V_S =.2



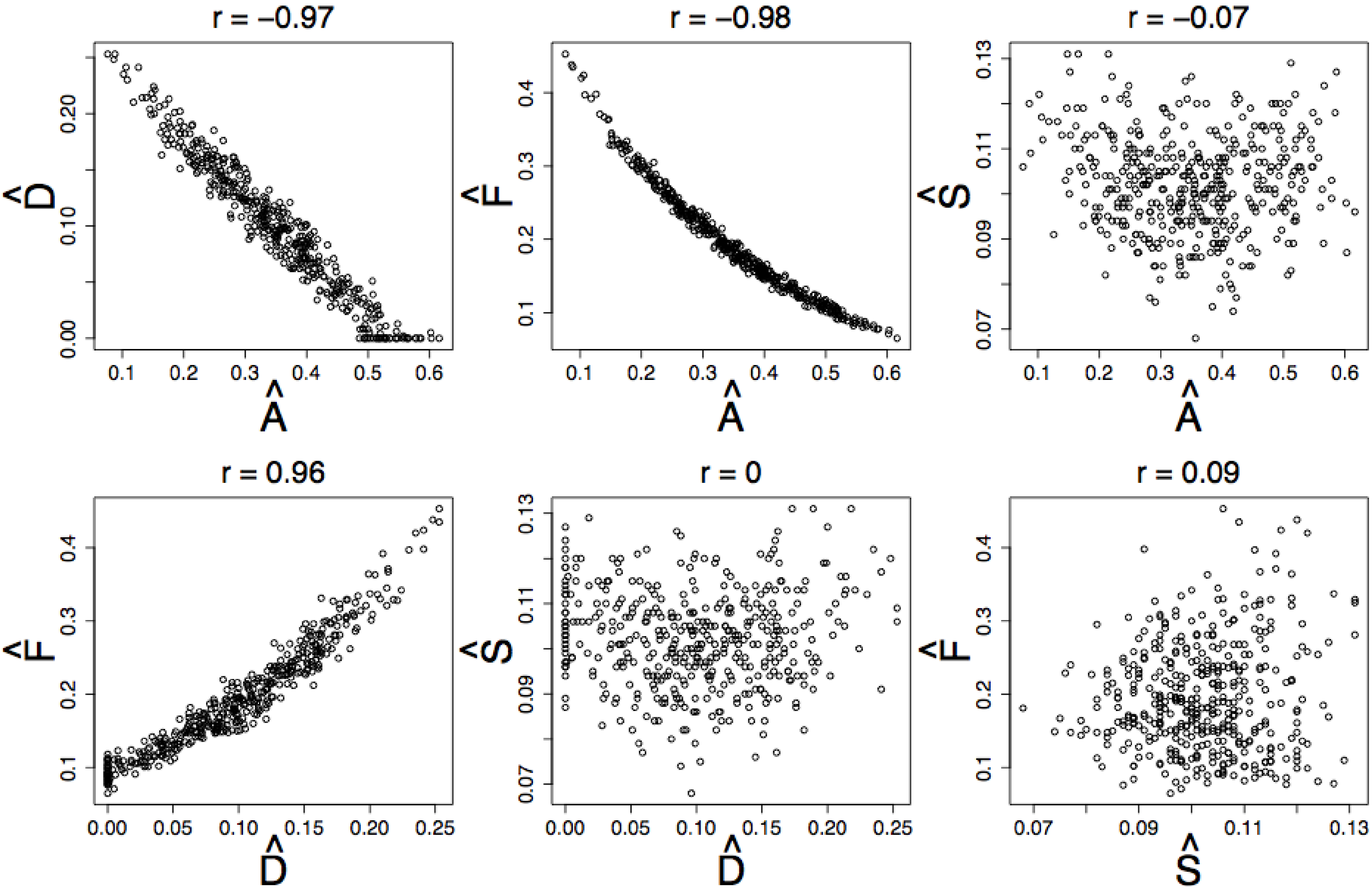
Reality: V_A=.4, V_D=.15, V_S=.15



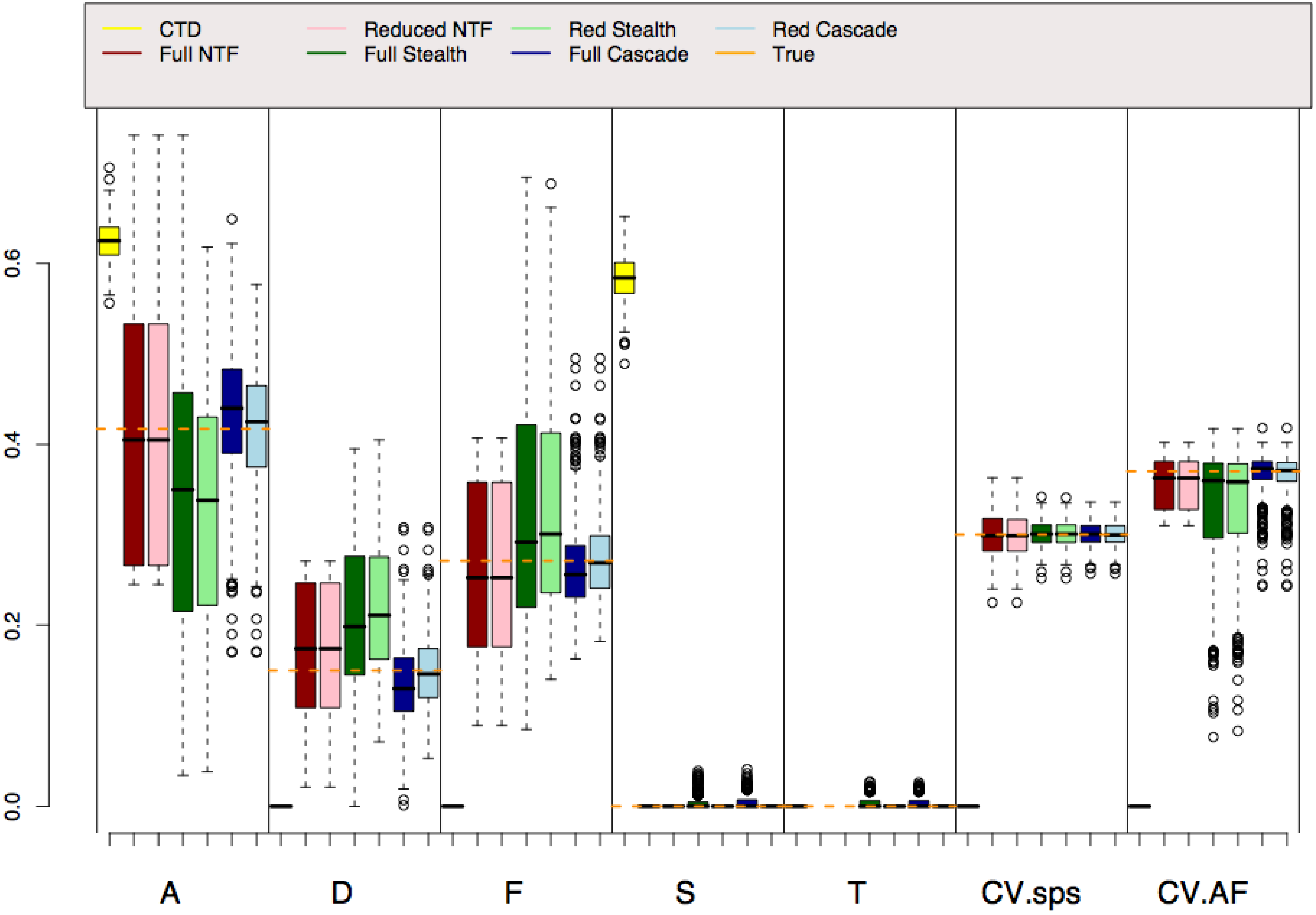
Reality: V_A=.35, V_D=.15, V_F=.2, V_S=.15, V_T=.15, AM=.3



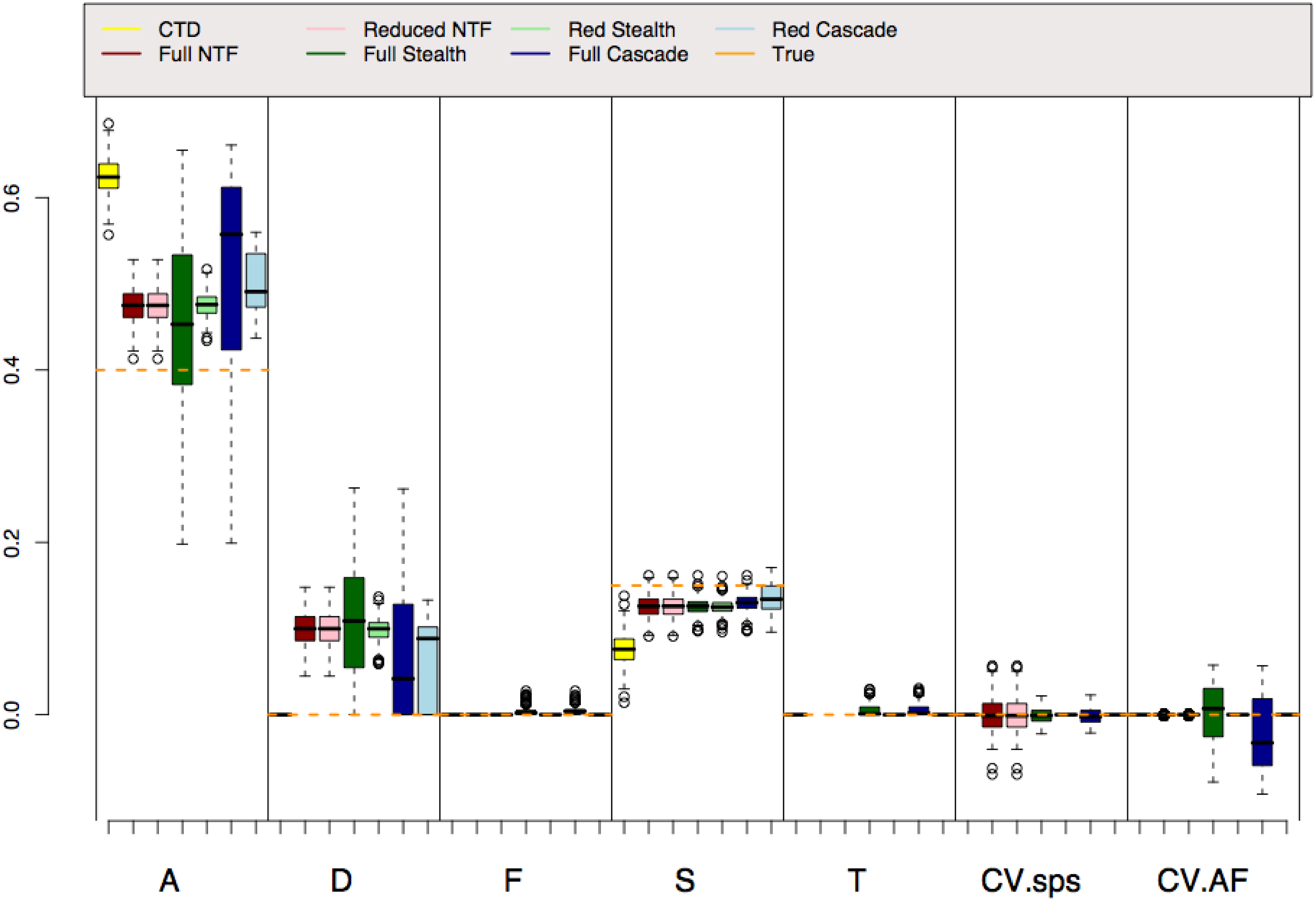
V_A, V_D, V_F estimates are highly correlated in Stealth & Cascade



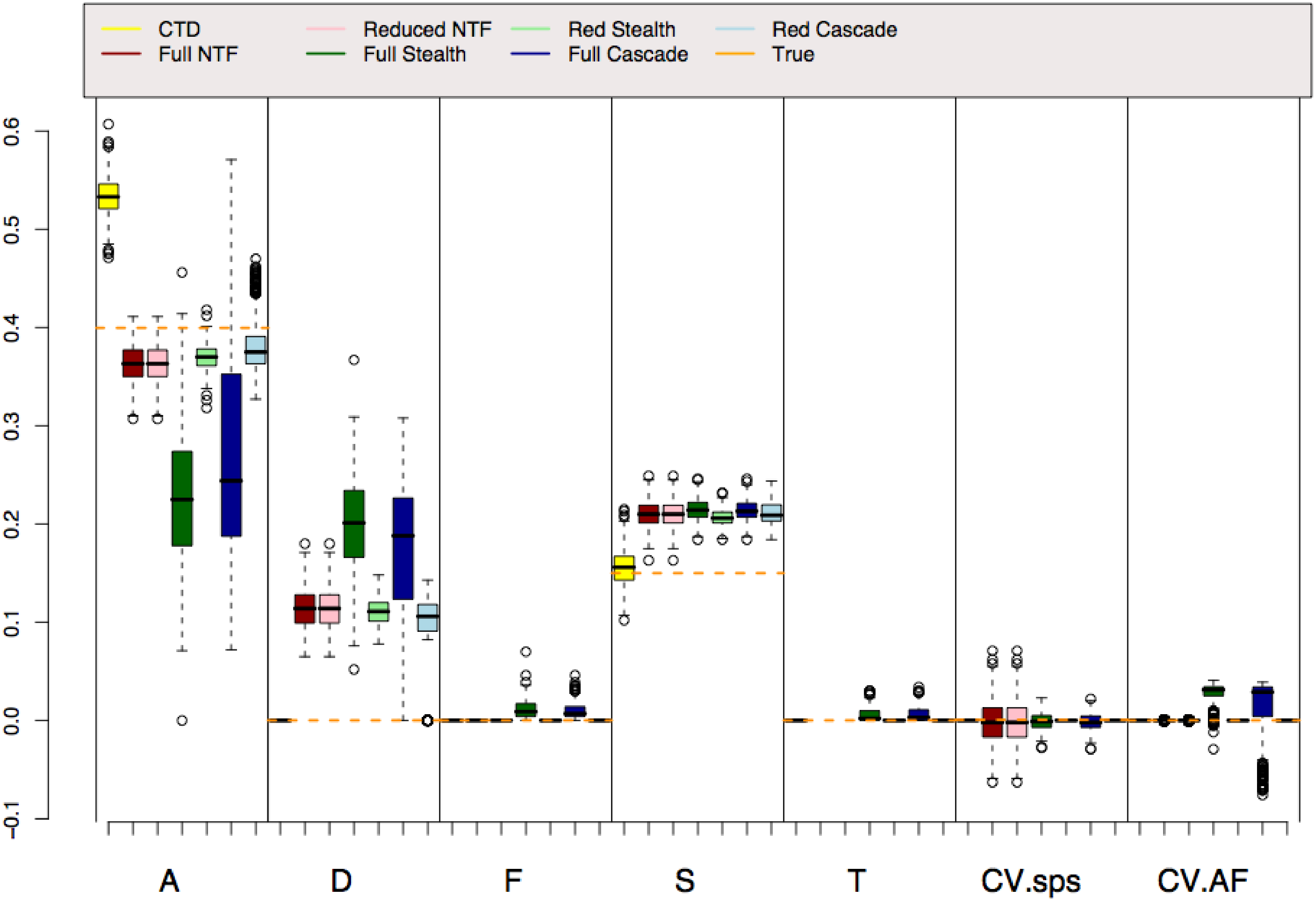
Reality: V_A=.45, V_D=.15, V_F=.25, AM=.3 (Soc Hom)



Reality: V_A=.4, $V_{A\times A}$=.15, V_S=.15



Reality: V_A=.4, $V_{A\times Age}$=.15, V_S=.15



Conclusions

- All models require assumptions. More assumptions = more biased estimates on average
- Simulations provide assessments of NTFD, *Stealth*, and *Cascade*. These complicated models work as designed and are less biased and provide more nuanced understanding than CTDs
- But ETFDs have drawbacks:
    - Complicated; easy to make mistakes
    - Require large datasets (e.g., n > 20k individuals) that are rarely collected
    - Require many assumptions (but fewer than CTD)

Further reading on this lecture

- Eaves LJ, Last KA, Young PA, Martin NG (1978) Model-fitting approaches to the analysis of human behaviour. *Heredity* 41:249-320
- Fulker DW (1982) Extensions of the classical twin method. Human Genetics. Part A: The Unfolding Genome (Progress in Clinical and Biological Research Vol 103A). p. 395-406
- Fulker DW (1988) Genetic and cultural transmission in human behavior. Proceedings of the Second International conference on Quantitative Genetics
- Eaves LJ, Heath AC, Martin NG, Neale MC, Meyer JM, Silberg JL, Corey LA, Truett K, Walter E (1999) Comparing the biological and cultural inheritance of stature and conservatism in the kinships of monozygotic and dizygotic twins. In: Cloninger CR (Ed) Proceedings of 1994 APPA Conference. p. 269-308
- Keller MC & Coventry WL (2005). Quantifying and addressing parameter indeterminacy in the classical twin design. *Twin Research and Human Genetics*, 8, 201-213
- Keller MC, Medland SE, Duncan LE, Hatemi PK, Neale MC, Maes HHM, Eaves LJ. Modeling extended twin family data I: Description of the Cascade Model. *Twin Research and Human Genetics*, 29, 8-18.
- Keller MC, Medland SE, & Duncan LE (2010). Are extended twin family designs worth the trouble? A comparison of the bias, precision, and accuracy of parameters estimated in four twin family models. *Behavior Genetics*.