

A Full-Scale Test of The Language Farming Dispersal Hypothesis

Harald Hammarström
harald2@cs.chalmers.se

January 17, 2009

Language Families

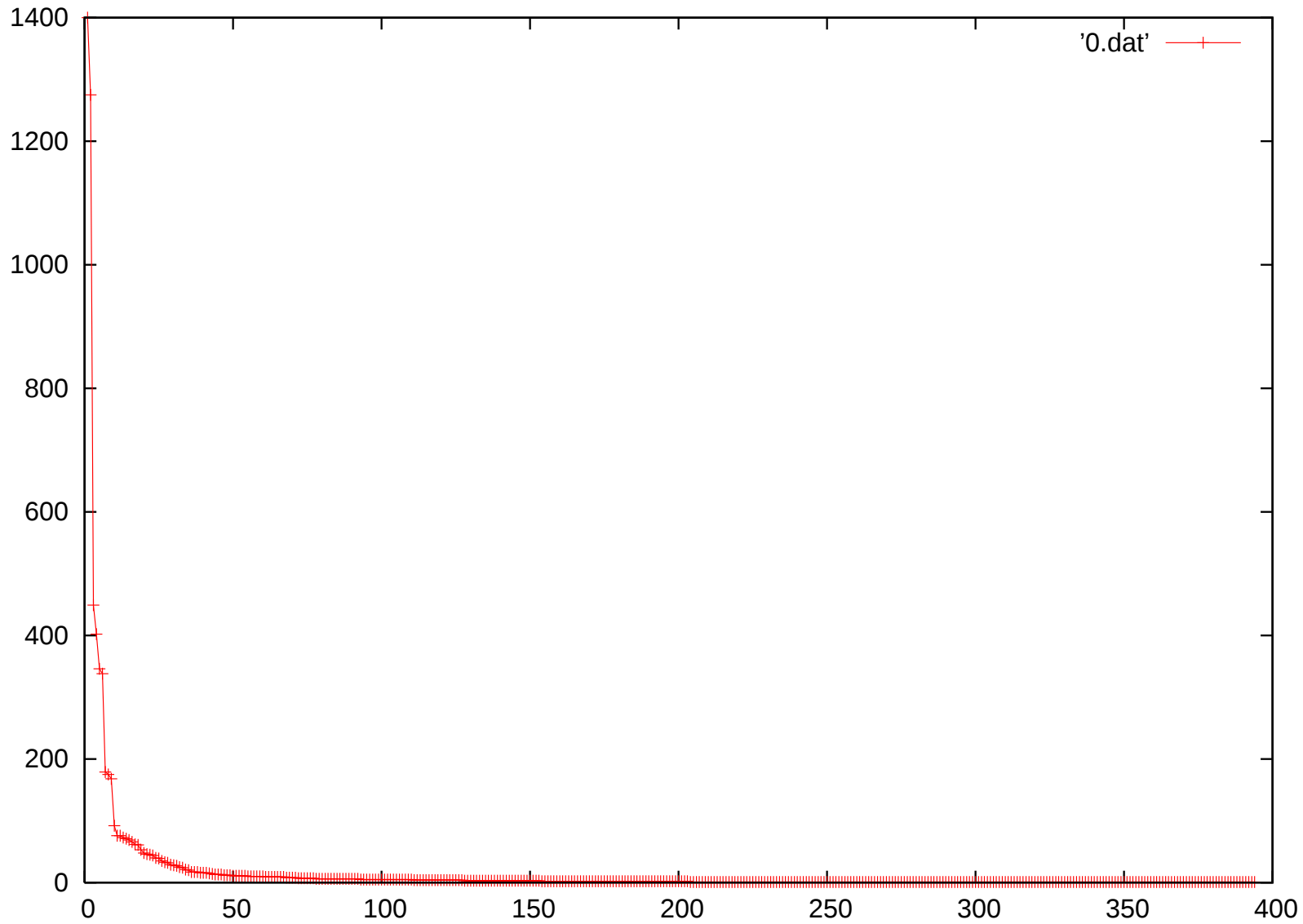
- There are some 7 000 languages in the world
- Language family defined as
 - a **set of languages** (possibly a one-member set)
 - with at least one **sufficiently attested** member language
 - that has been **demonstrated in publication**
 - to **stem from a common ancestor**
 - by **orthodox comparative methodology**
 - for which there are **no** convincing published attempts to demonstrate a **wider affiliation**
- Application of this definition yields some 400 families for the 7 000 languages (shown on handout!)

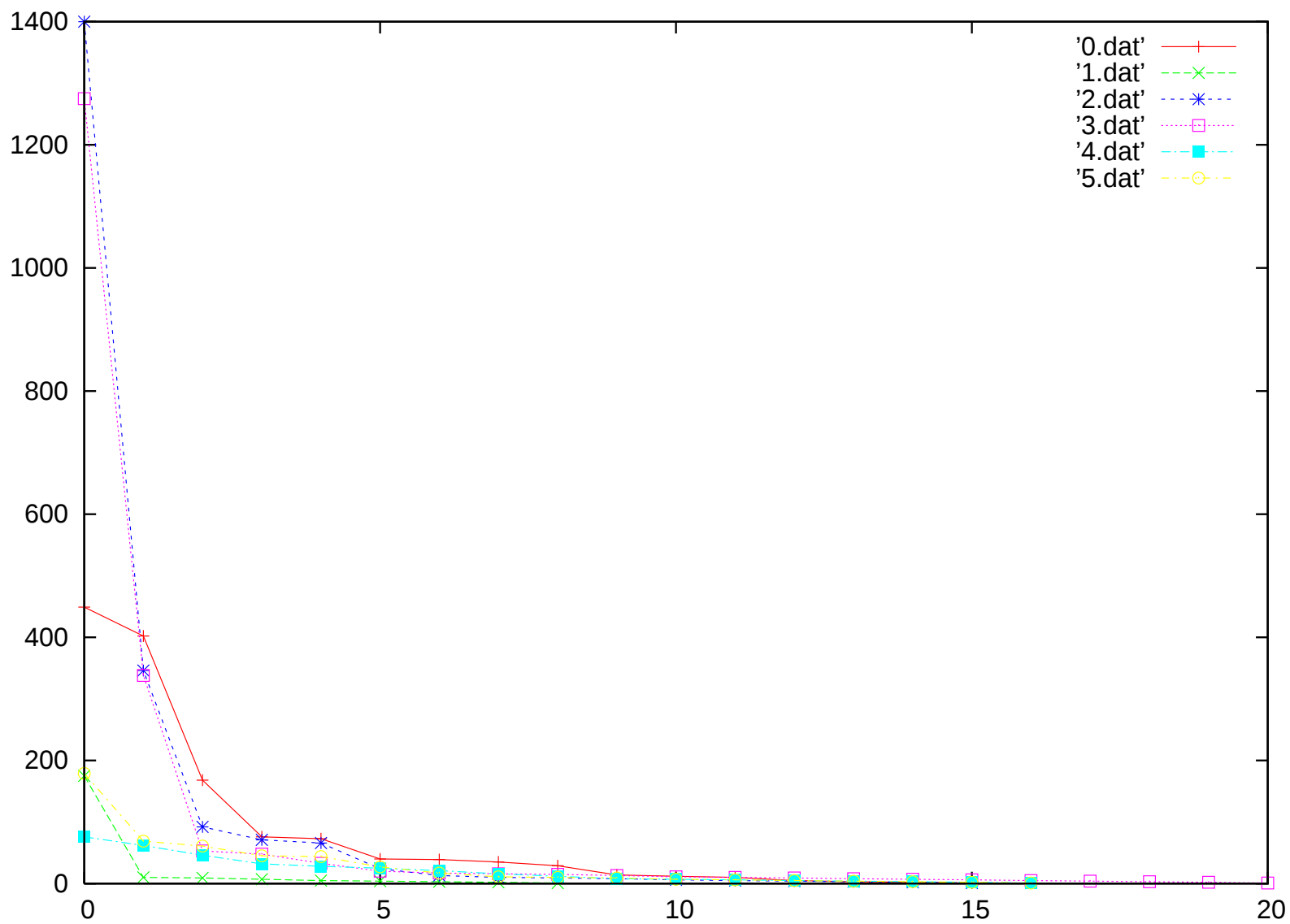
Language Family Sizes

Size of a family = the number of languages belonging to it

- The ca 400 families are of very unequal size
- A few are very big and very many are tiny
- Their sizes are not normally distributed
- In fact, the rank-size distribution follows a power-law (aka Zipfian, log-normal etc.)

Rank-Size Plot





Why Some Big and Some Small?

Two explanations so far proposed:

a) *Farming/Language Dispersal Hypothesis*

Some families are big because their speakers acquired farming, which allowed unprecedented expansion

We follow up this line today.

b) A power-law distribution are the expectation of stochastic branching processes

Not discussed today.

Farming/Language Dispersal Hypothesis

The farming/language dispersal hypothesis makes the ... proposal that the present-day distributions of many of the world's ... language families can be traced back to the early developments and dispersals of farming ... (Bellwood & Renfrew 2002:i)

- There are many case studies of individual families which support the FLDH
- There are many counterexamples
 - Individual widespread families with no association to farming
 - Presence of farming without expansion

Questions Discussed Today

On a worldwide scale, i.e., with all families taken into account

- Does the farming have any explanatory power in predicting which families are large (and which are not)?
- Does the geospatial distribution of the observed farming language families show an east-west spread (rather than a north-south) as predicted if the cause of their spread is farming, cf. Diamond 1997?

Database of Farming Families

Every family is judged **AGR**icultural (AGR) or **H**unter-**G**atherer (HG)

- A language is a Hunter-Gatherer (HG) language iff
 - its speakers subsist more than **50%** on
 - **hunted/gathered** food (= reproduction of species not controlled)
 - as of **ethnographic evidence** at
 - **first eyewitness documentation time**
- A family is HG iff all of its member languages are HG (otherwise AGR)

AGR		HG	
Atlantic-Congo	1400	Pama-Nyungan	175
Austronesian	1275	Sepik	48
Indo-European	449	Eyak-Athapaskan-Tlingit	45
Sino-Tibetan	402	Algic	44
Afro-Asiatic	346	Lower Sepik-Ramu	33
Trans New Guinea	338	Carib	32
Otomanguean	179	Panoan	28
Austroasiatic	168	Salishan	27
East Sudanic	92	Tucanoan	25
Tai-Kadai	76	Lakes Plain	20
Tupí	76	Tor-Orya	13
Dravidian	73	Cenderawasih Bay	11
Mande	71	Eskimo-Aleut	11
Mayan	69	Bosavi	10
Central Sudanic	66	Great Andamanese	10
Arawak	62	Miwok-Costanoan	10
Uto-Aztecan	61	Western Daly	10
...	...	10	

Farming-Size Correlation

	AGR	HG	ALL
# families	165	229	394
Σ -size	6012	1027	7039
Mean size	36.44	4.48	17.87
Median size	2	1	2

Is the correlation **AGR** vs. mean size statistically significant?

Test: Sample 1000 subsets S_i of size 165, and check how many have a sum size ≥ 6012

AGR-families and Size

- The correlation between AGR and (average) size is highly significant ($p < 0.001$)
- What about the Small **AGR** families?
 - If small ≤ 10 then there are some 164 small **AGR** families
 - A majority (ca 100) of these are found surrounded by other **AGR** families in East Papua (i.e., islands off New Guinea), Sahel, Mexico, Andes, Eurasia
 - The rest are found in **HG** surroundings in the Amazon and New Guinea

So FLDH passes first round!

FLDH and Geographic Distribution

- Agriculture spreads east-west easier than north-south
- If agriculture is indeed the cause of large families then the large families should show east-west expansion rather than north-south
- Measure the geospatial distribution of a family:
 - Database of center coordinates for all languages
 - East-west (EW) expansion is the difference between the eastern and western endpoint languages of the family
 - North-South (NS) expansion is the difference between the northern and southern endpoint languages of the family
 - Define **HOR** horizontality as the ratio between east-west expansion and north-south expansion $HOR = \frac{EW}{NS}$
- NOTE: Isolates are excluded [198 points remaining]

Example: Saharan

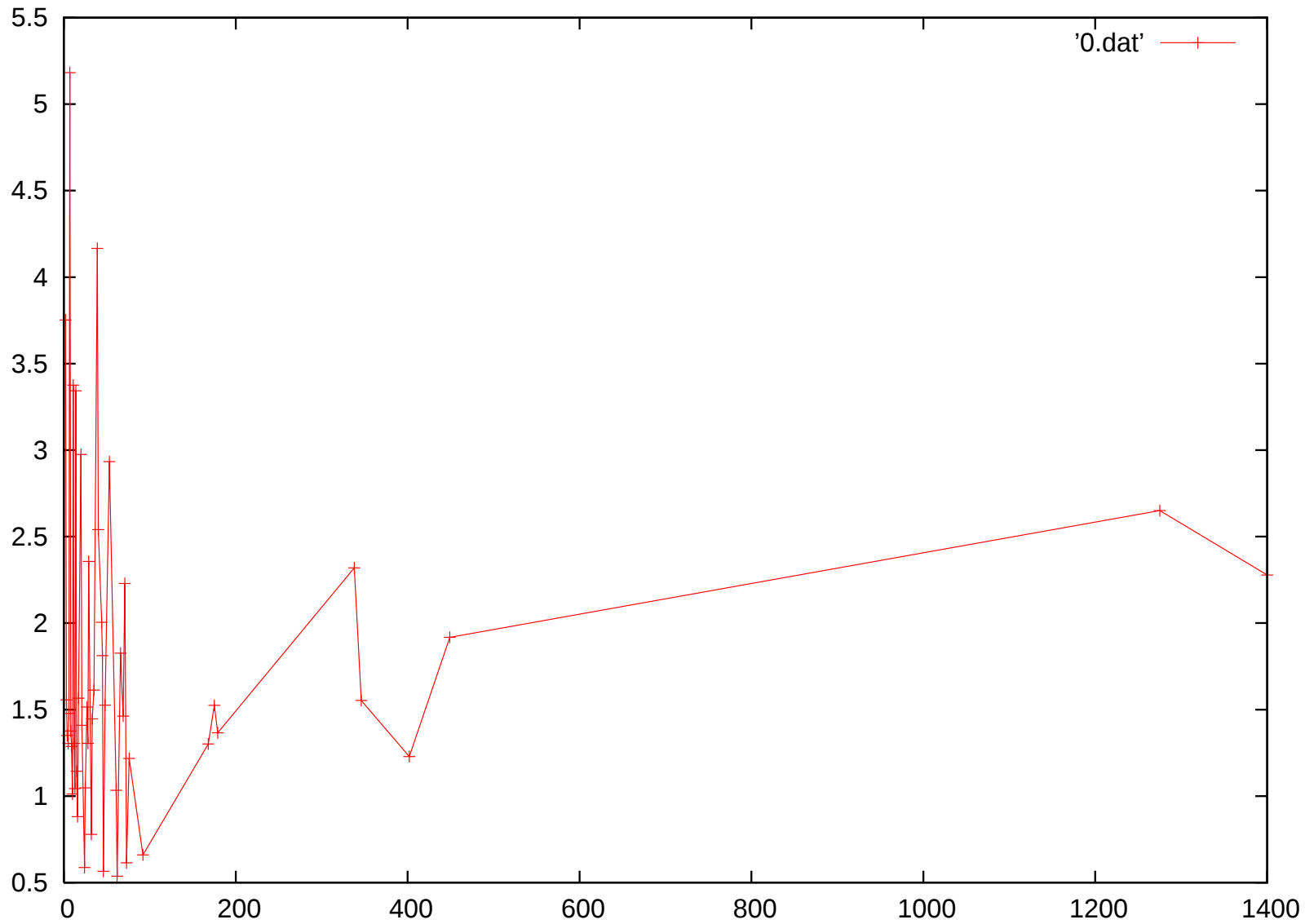
W Endpoint	Kanuri, Manga	kby	Niger	10.85 E
E Endpoint	Berti	byt	Sudan	32.72 E
S Endpoint	Kanuri, Central	knc	Nigerian	11.01 N
N Endpoint	Berti	byt	Sudan	20.61 N

$$EW = 32.72 - 10.85 = 21.88$$

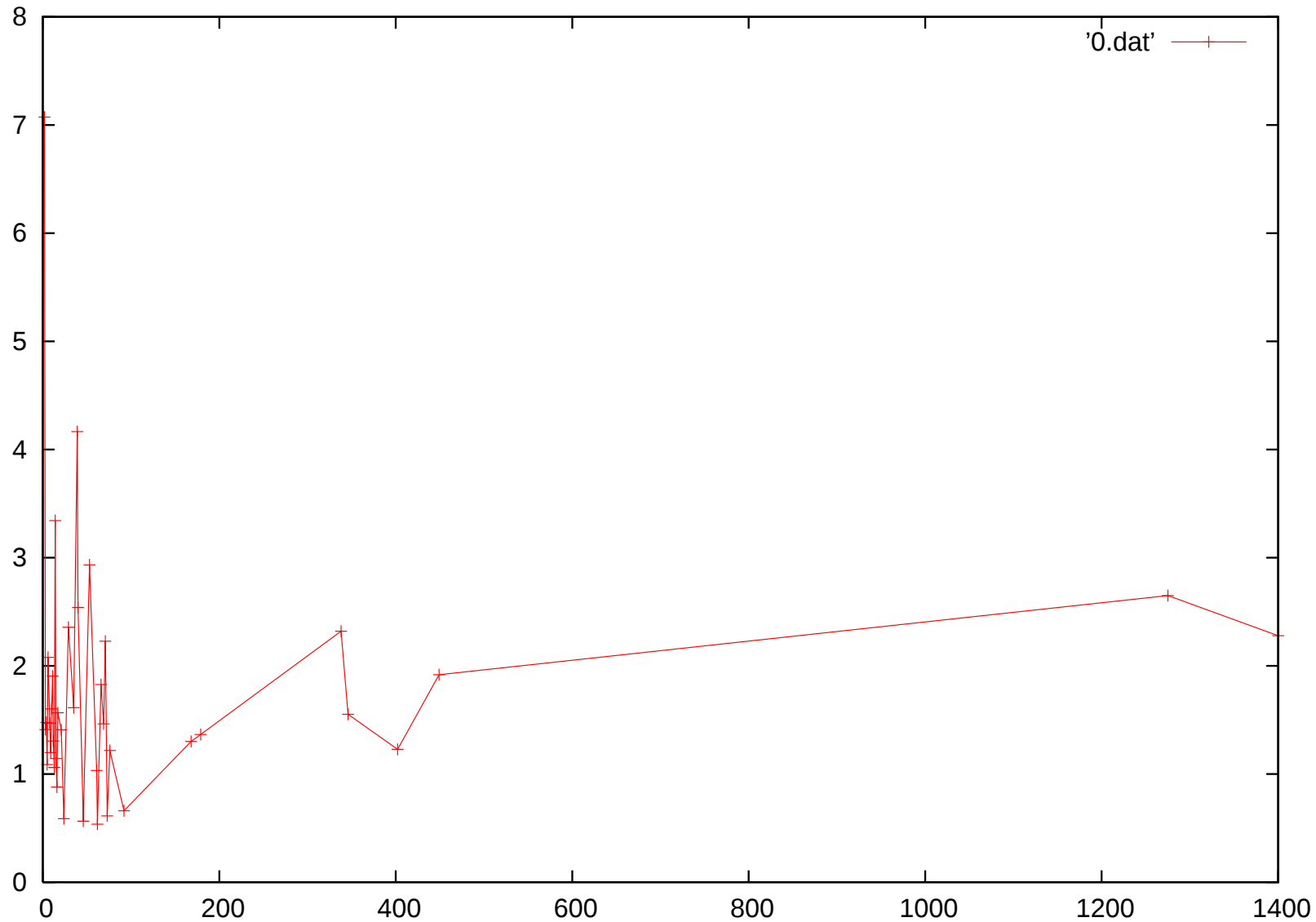
$$NS = 20.61 - 11.01 = 10.60$$

$$HOR = \frac{21.88}{10.60} = 2.06$$

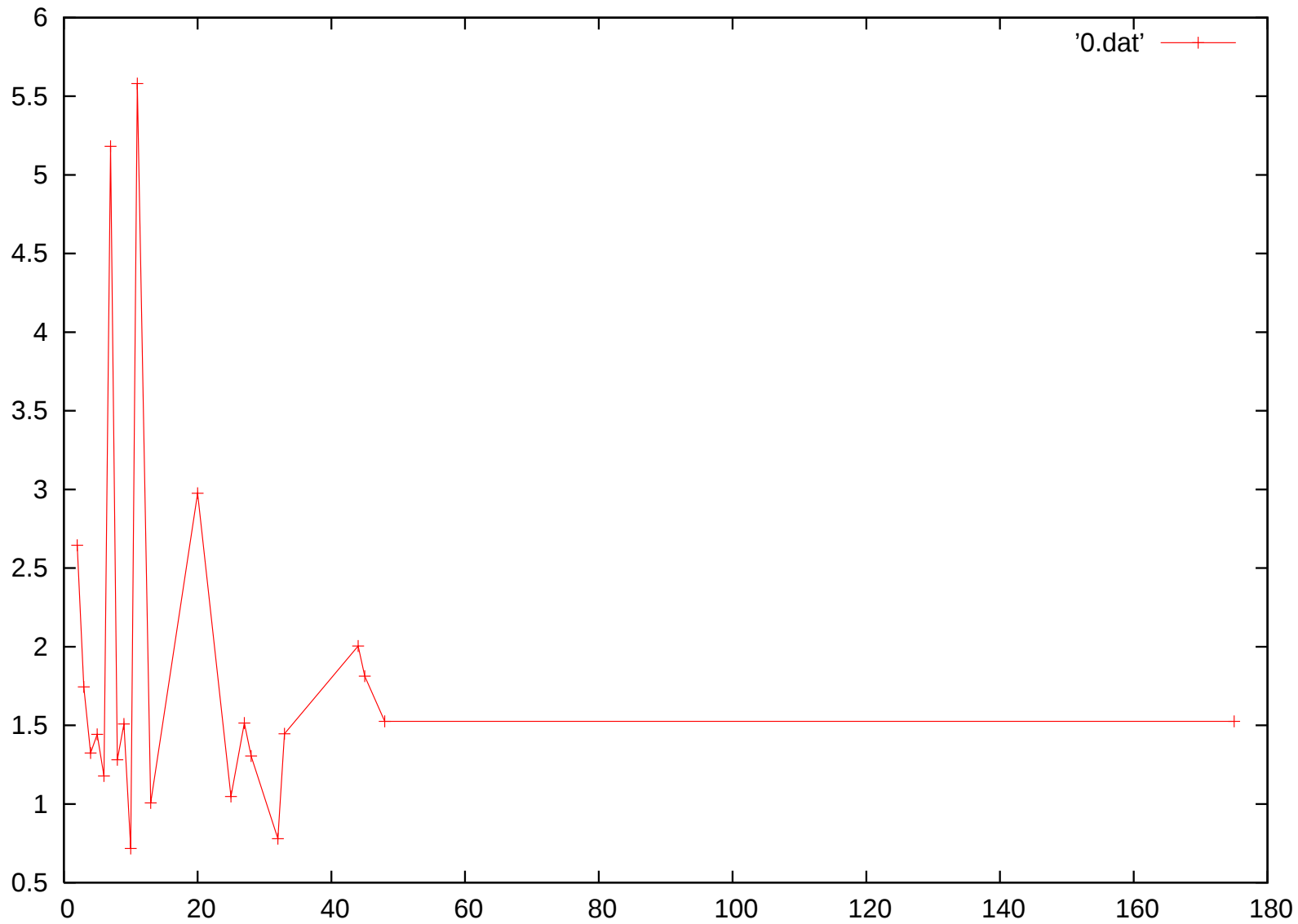
HOR-Size Correlation



HOR-Size for AGR only



HOR-Size for HG only



HORizontality and AGR/HG-families

It appears that there is not significantly more horizontality in AGR families than in HG (not even for the largest families)

	AGR	HG	ALL
# families	90	108	198
Mean HOR	2.24	2.11	2.17
Median HOR	1.30	1.21	1.25

Conclusions

On a shallow but world-wide test:

- Most families are small, whether agricultural or hunter-gatherer
- Agricultural families are significantly larger than hunter-gatherer families (on average)
- Small agricultural families more often than not have (only) agricultural neighbours
- If agriculture was the cause of the larger agricultural families, one would expect them to show more horizontalness than the corresponding hunter-gatherer families

This is not the case