Sponsored Search Seminar Group Project: Approach and Initial Results

Edi Bice, Kuzman Ganchev, Alex Kulesza, Qian Liu, Jinsong Tan, Qiuye Zhao

Initial Ideas

- Analyze keyword markets related by use of modifiers
  – ‘lexus’ vs. ‘used lexus’
  – ‘electrician’ vs. ‘chicago electrician’
- Develop better models for real-world bidding behavior
  – Not “uniformly at random from [a,b]”

Combined Approach

- Develop a parameterized model for bids on a single query
  - E.g., parameters are first bid and rate of falloff
- Collect bids for a wide array of keyword/modifier pairs
  - Fit the model to each bid viewer result
- Use model parameters as analysis quantities
  - E.g., the ‘free’ modifier lowers the first bid and decreases rate of falloff

Analysis

- Noise is an issue
  - Consider keywords and modifiers in groups
- Generate matrices showing effect of each modifier group on each base keyword group
  - Preliminary results today
- So far, groups created by hand
  - Can we automate this? More later…

Outline

1. Keywords and modifiers (Edi)
2. Initial results (Qian and Qiuye)
3. Ongoing work: Modeling bids (Kuzman)
4. Ongoing work: Modeling values (Jinsong)
5. Ongoing work: Automatic clustering (Alex)

Keyword Base Groups and Modifiers

- relevant, popular, diverse, and interesting
  - what some people search for
  - affected differently by modifiers
  - differ in several aspects (spatial, temporal, expense)
- eight groups with ~600 keywords total
  - one or two groups per person
- six modifier groups ~50 modifiers
  - modeling phases of consumer interaction
  - not necessarily applicable to all base keywords
- 32K base-modifier pairs
  - sparsity
  - data collection (Tales from the (s)Crypt)

Base Keywords

- **Cars** (Alex)
  - toyota camry, chevy, ford suv, porsche 911
- **Drugs, medical** (Edi)
  - zoloft, cialis, psoriasis, sciatica, liposuction
- **Electronics, software** (Jinsong)
  - xbox, mp3, pda, oracle, world of warcraft
- **Travel** (Kuzman)
  - airfare, cruise, safari, sailing, vacation
- **Local and non-local services** (Qian)
  - electrician, locksmith, ** insurance, ** loan
- **Subscription services** (Qiuye)
  - cable tv, gym membership, magazine subscription

Keyword Modifiers

- INFO:
  - info, information,
  - specs, specifications,
  - reviews, ratings,
  - prices,
  - coupon, rebate,
  - guide, news
- QUALITY:
  - best
  - luxury
  - favorite
  - inclusive, exclusive
  - preferred
  - used, new
- LOCATION:
  - 20 U.S. States
  - 20 U.S. Cities
- PRICE:
  - cheap, free
  - bargain, discount, deal
  - special, sale
  - budget, affordable
  - expensive
- ACTION:
  - buy, sell, purchase
  - lease, rent, hire
- POST:
  - support
  - parts
  - repair
  - mechanic
  - manufacturer
  - warranty

Base groups vs. modifier groups

	drugs	electronics	local service	medical	non-local service	software	subscription	travel
quality	0.2425	0.522697	0.454516	0.763077	1.38907	0.262778	0.3745283	0.718875
price	-1	0.554332	0.509529	0.558636	1.644848	0.186486	0.4626552	0.777972
post	-1	0.385122	0.713784	-1	0.730417	0.233529	0.3529545	0.11
location	-1	0.38551	1.58505	1.651934	2.829279	0.264792	0.5842331	0.729871
info	0.285461	0.259721	0.487416	1.008542	1.341488	0.223793	0.2439355	0.303883
action	-1	0.665185	0.498667	0.3125	1.409231	0.297857	0.4692941	0.323721
null	0.878548	1.026176	1.264545	1.742041	4.6692	0.662973	0.7644633	0.815219

Base groups vs. modifier groups

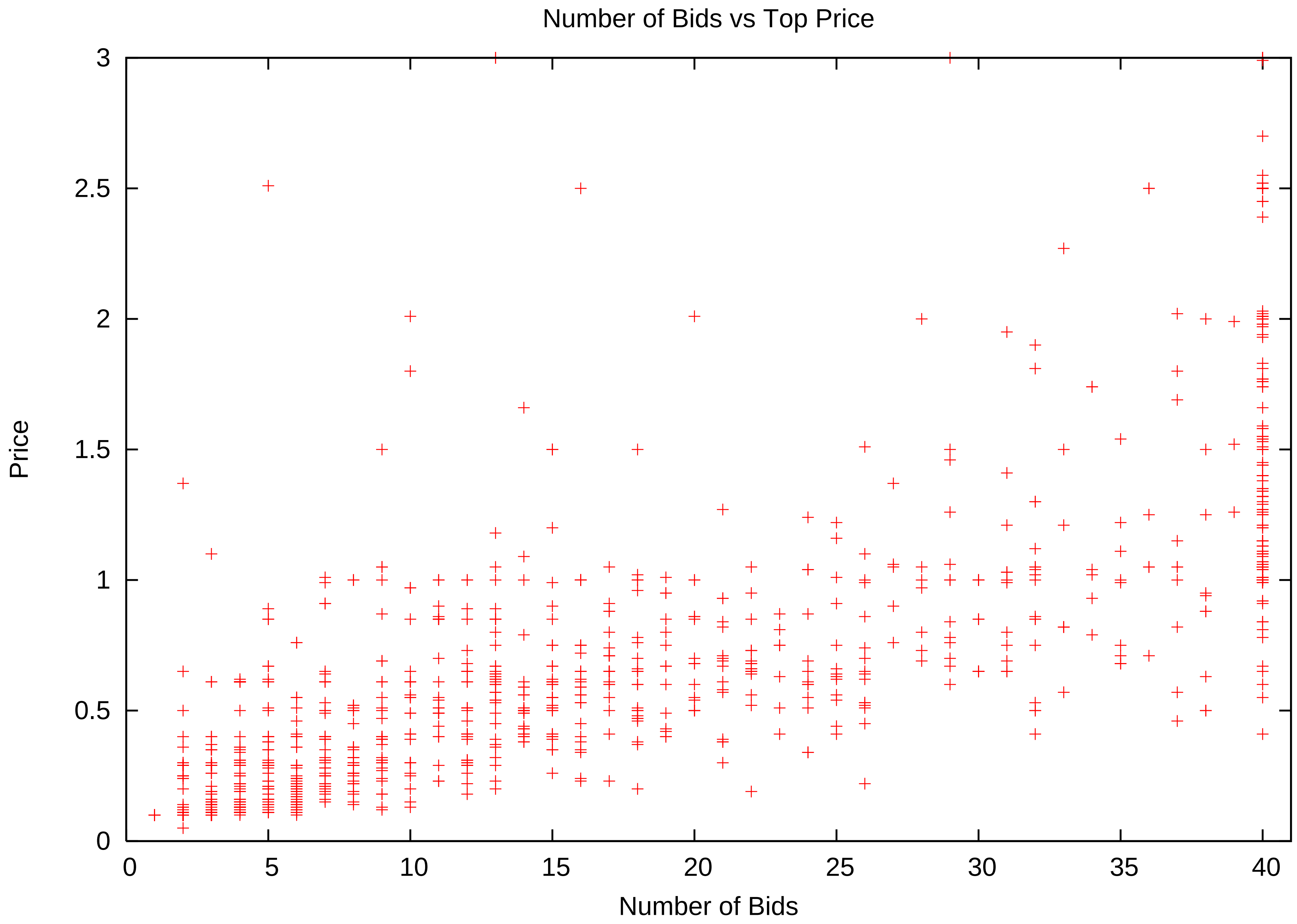
quality	0.2425	0.522697	0.454516	0.763077	1.38907	0.262778	0.3745283	0.718875
price	-1	0.554332	0.509529	0.558636	1.644848	0.186486	0.4626552	0.777972
post	-1	0.385122	0.713784	-1	0.730417	0.233529	0.3529545	0.11
location	-1	0.38551	1.58505	1.651934	2.829279	0.264792	0.5842331	0.729871
info	0.285461	0.259721	0.487416	1.008542	1.341488	0.223793	0.2439355	0.303883
action	-1	0.665185	0.498667	0.3125	1.409231	0.297857	0.4692941	0.323721
null	0.878548	1.026176	1.264545	1.742041	4.6692	0.662973	0.7644633	0.815219
	drugs	electronics	local service	medical	non-local service	software	subscription	travel

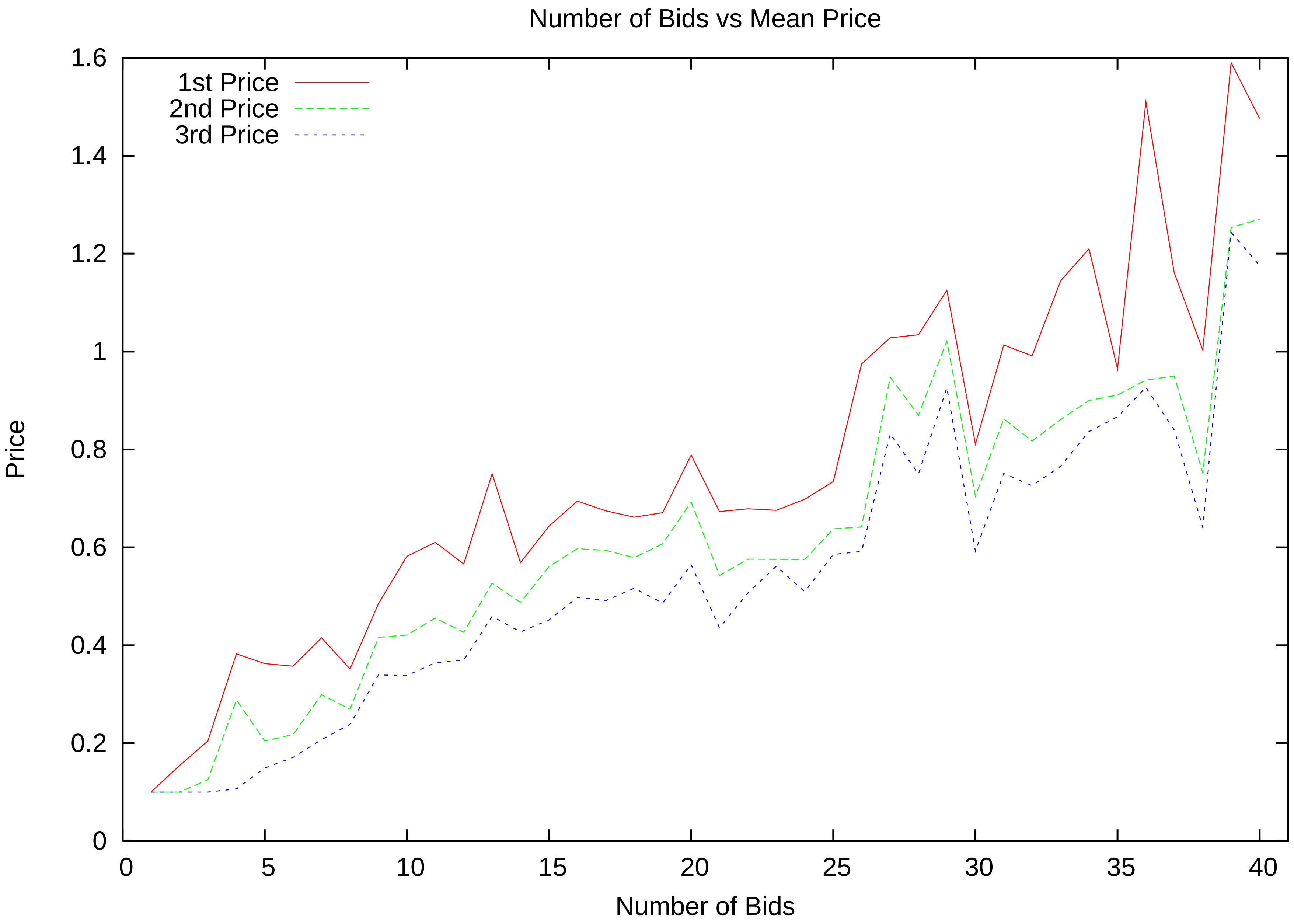
Base groups vs. price modifier sub-groups

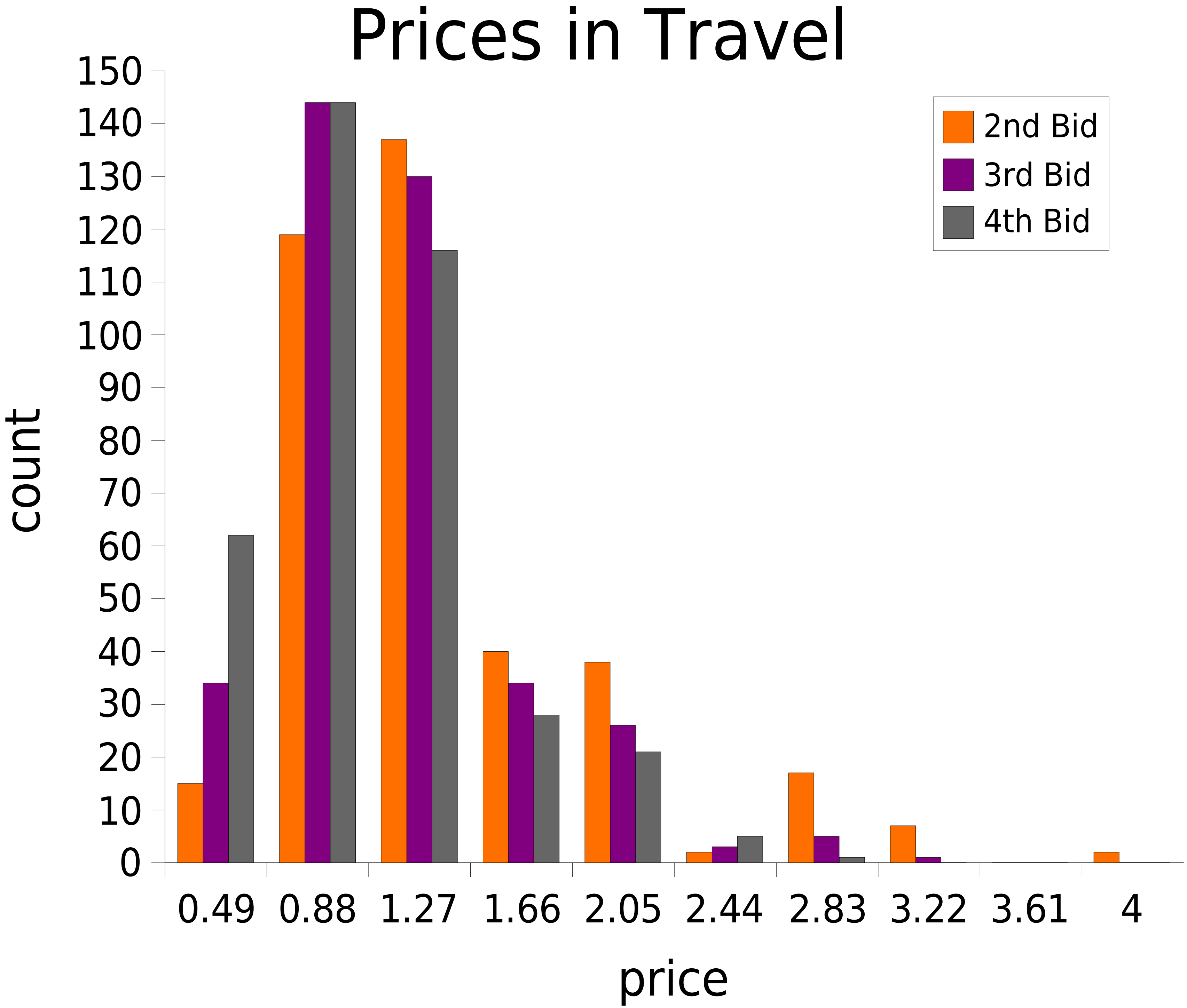
	drugs	electronics	local service	medical	non-local service	software	subscription	travel
expensive	-1	0.155556	0.115	-1	0.82	-1	0.12	0.124
sale	-1	0.63037	0.395238	0.2	0.795625	0.2525	0.3460606	0.538889
affordable	-1	0.414615	0.682	0.87375	2.181333	-1	0.3394737	0.481333
bargain	-1	0.314737	0.214286	-1	0.893846	0.1	0.4636364	0.896842
special	-1	0.213	0.346364	0.1	1.1	0.15	0.3727778	0.824286
budget	-1	0.236667	0.908571	-1	1	-1	0.173	0.618889
deal	-1	0.6528	0.321538	0.185	1.572308	0.34	0.5138235	1.119375
discount	-1	0.833214	0.550303	0.68	2.51	0.132	0.6111864	0.964681
free	-1	0.297895	0.4605	0.221538	1.115909	0.14	0.3020408	0.235
cheap	-1	0.806774	0.589722	0.781818	3.125263	0.166667	0.6505769	0.723011
null	0.878548	1.026176	1.264545	1.742041	4.6692	0.662973	0.7644633	0.815219

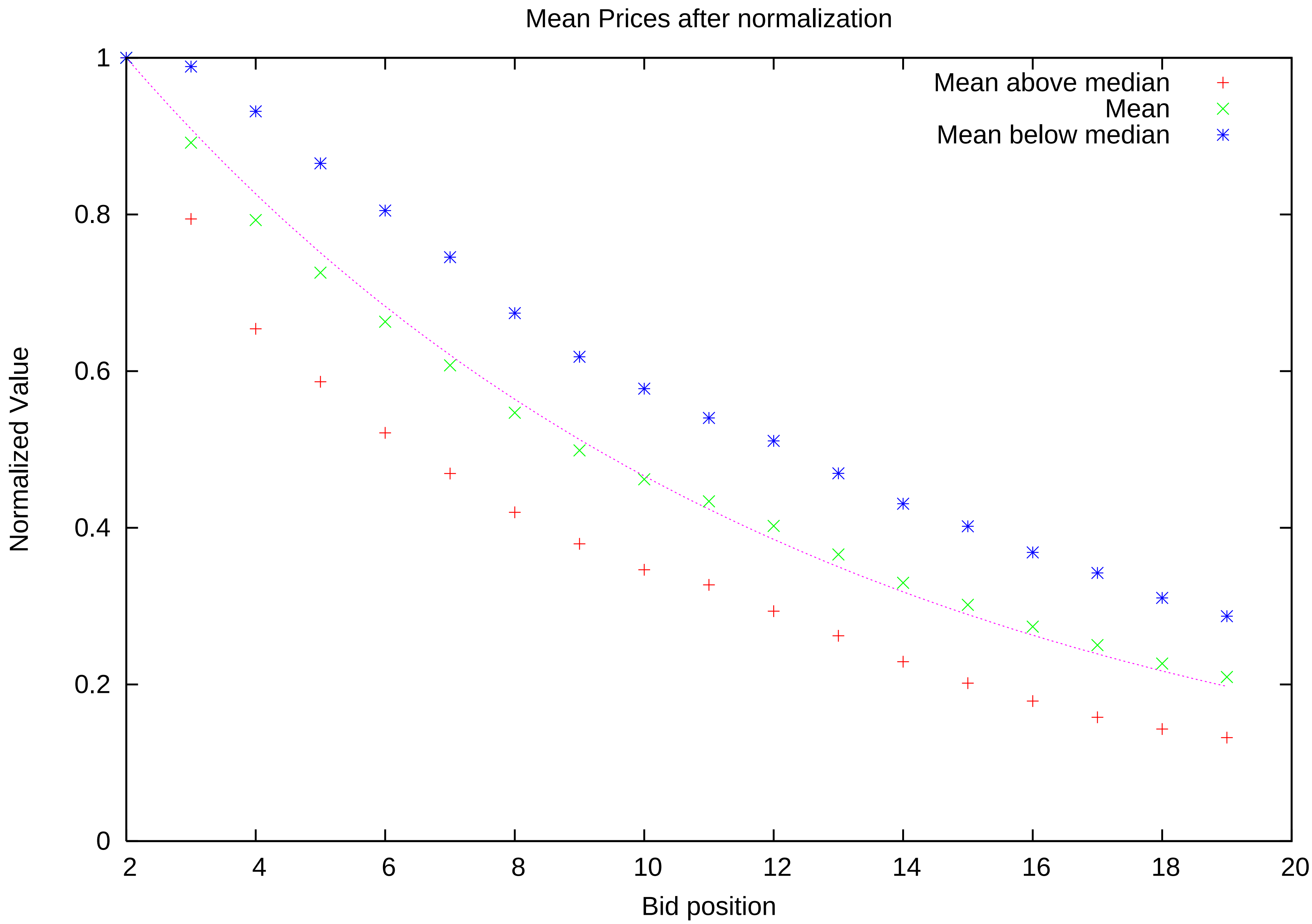
Base groups vs. location modifier sub-groups

	drugs	electronics	local service	medical	non-local service	software	subscription	travel
pacific	-1	0.365	1.481158	1.494248	3.099717	0.179	0.553864	0.875042
mountain	-1	0.231905	1.449853	1.376667	2.850753	0.575	0.520769	0.633766
central	-1	0.344068	1.631514	1.684847	2.793631	0.217391	0.562696	0.68305
eastern	-1	0.590606	1.718561	1.961528	2.767216	0.188571	0.710263	0.77661
null	0.87855	1.02618	1.26455	1.74204	4.6692	0.66297	0.76446	0.81522

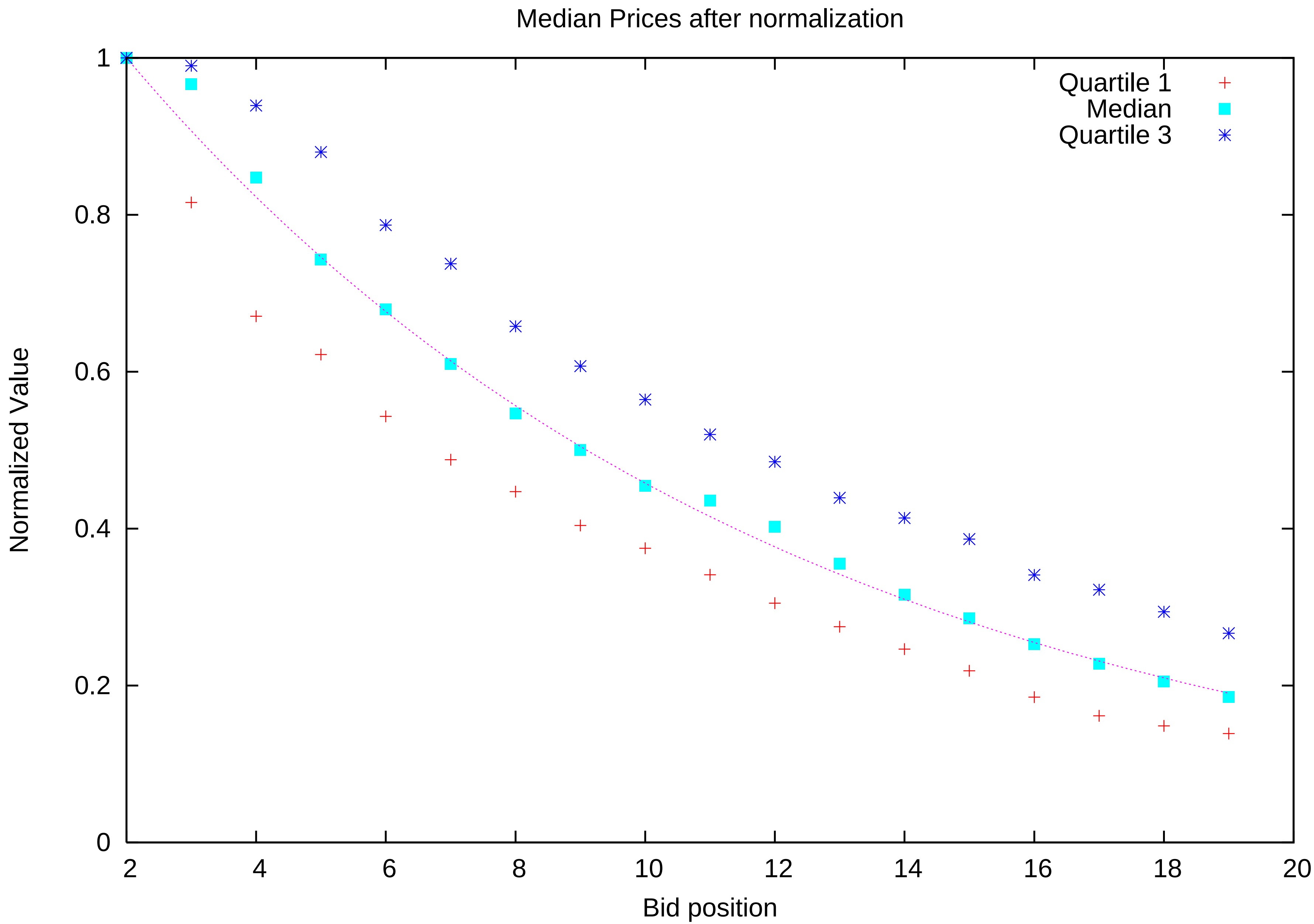

Number of Bids vs Top Price
Price
Number of Bids
0
0.5
1
1.5
2
2.5
3
0
5
10
15
20
25
30
35
40


Number of Bids vs Mean Price
1st Price
2nd Price
3rd Price
Price
Number of Bids
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
0
5
10
15
20
25
30
35
40

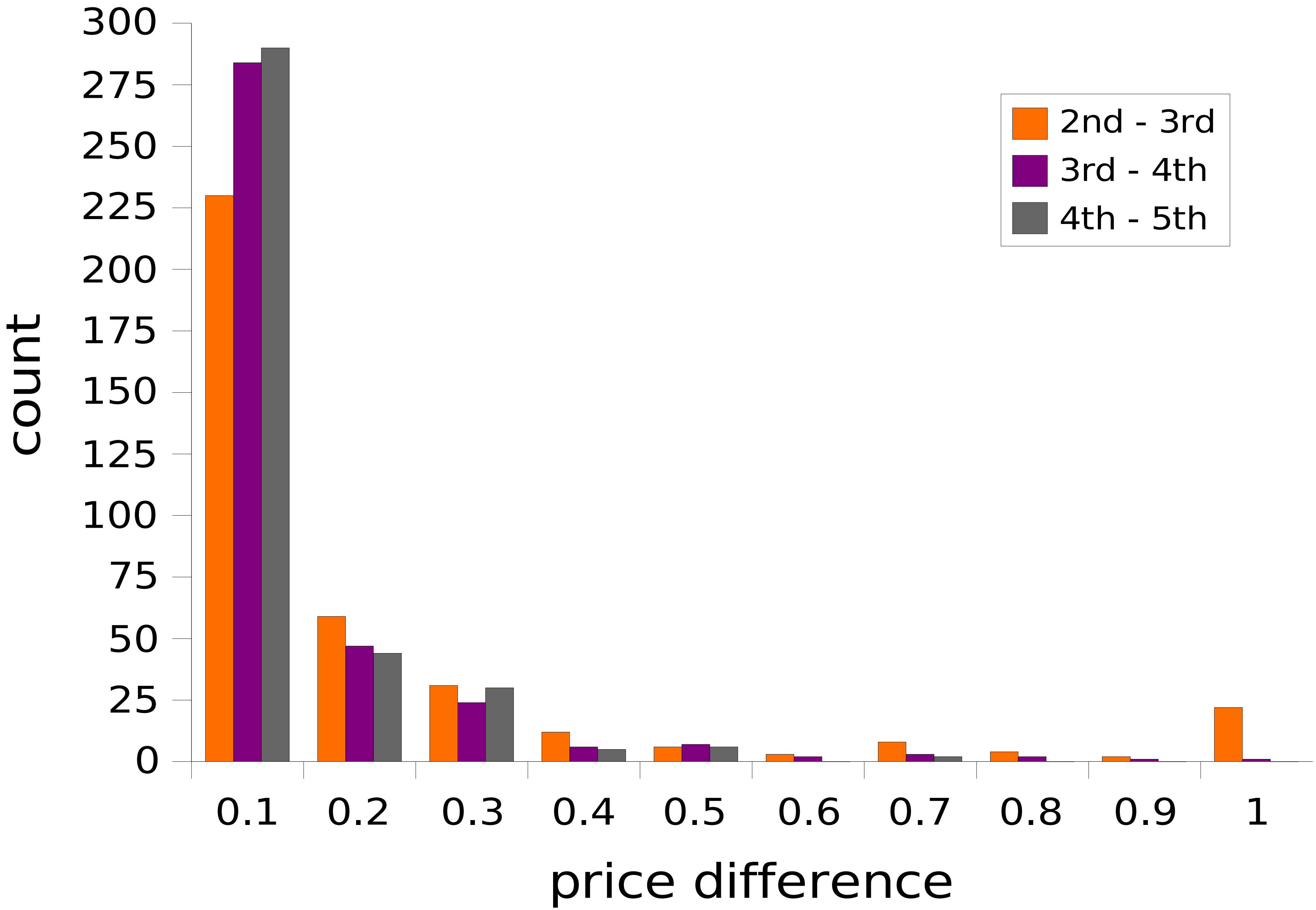

Prices in Travel
count
150
140
130
120
110
100
90
80
70
60
50
40
30
20
10
0
0.49
0.88
1.27
1.66
2.05
2.44
2.83
3.22
3.61
4
price
2nd Bid
3rd Bid
4th Bid

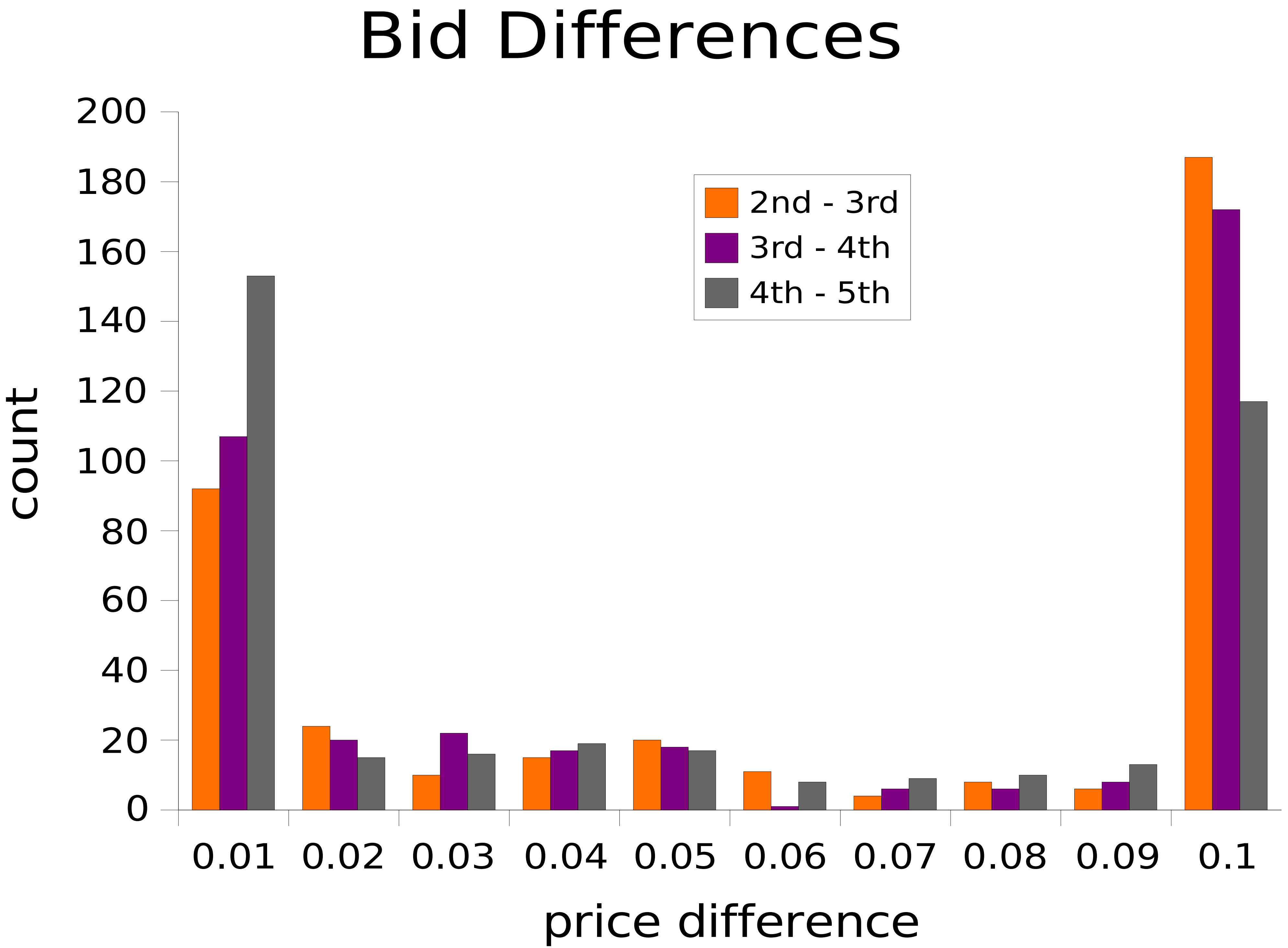


Mean Prices after normalization
Mean above median
Mean
Mean below median
Normalized Value
Bid position
0
0.2
0.4
0.6
0.8
1
2
4
6
8
10
12
14
16
18
20

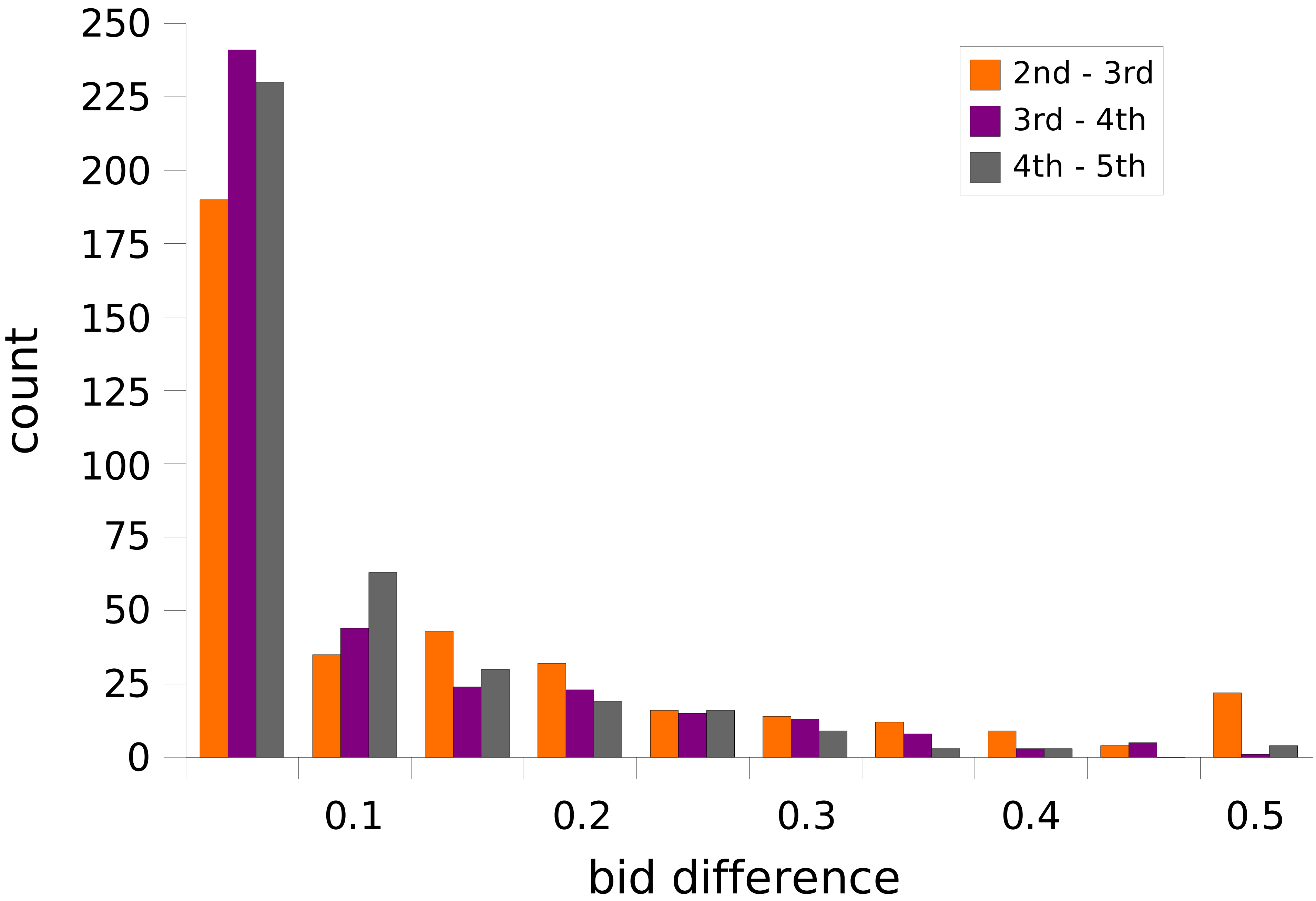

Median Prices after normalization
Quartile 1
Median
Quartile 3
Normalized Value
Bid position
0
0.2
0.4
0.6
0.8
1
2
4
6
8
10
12
14
16
18
20

Bid differences




Bid Differences
count
200
180
160
140
120
100
80
60
40
20
0
2nd - 3rd
3rd - 4th
4th - 5th
0.01
0.02
0.03
0.04
0.05
0.06
0.07
0.08
0.09
0.1
price difference

Normalized price differences



Modeling Bidder Values

- The questions:
  - What is the distribution of values (of all potential bidders) for a random keyword?
  - What is the distribution of values for a keyword from a specific category?
  - How does a modifier affects the distribution of value?
- In the current literature, such distributions are often assumed to be uniformly distributed over some interval [a,b]
  - An oversimplification
- Our experiments set to answer these questions

Modeling bidder values

- The problem:
  - bidder values are *never* directly observable
- Estimate bidder b_i*'s* value with the maximum bid ever observed during some period of time
- Assumptions:
  - 1. the Max bid is highly correlated with her value (and positively).
  - 2. the bid value of any bidder does not vary too much over the period of time we observe

Experiment Setup

- 1. sample a set of keywords;
- 2. observe the bids over, say, a few weeks for each keyword X;
- 3. record the max bid, *max*(b_i,X) for each bidder b_i
- 4. normalize these data according to some criteria
  - e.g. by dividing by the highest max bid for X among all bidders
    - $max(b_i,x) \rightarrow max(b_i,x)/max_j\{ max(b_j,x) \}$
  - Or by further take into consideration n_X, the avg num of bidders for X
    - $max(b_i,x) \rightarrow \mathbf{[\ (n_X+1)/n_X\ ]} * [\ max(b_i,x)/max_j\{ max(b_j,x) \}\ ]$
  - now each (keyword, bidder value) pair maps to a point in [0,1]
- 5. plot all such data points in [0,1] will give us a rough idea of the "prior" distribution of bidder values for a random keyword.
- 6. come up with some statistical model that fits the data
  - Hopefully also come up with a theory explains it

Automatic Clustering

- Can we choose keyword/modifier groups automatically?
- Idea: use data to guide clustering
  - Modifiers that have similar effects should be grouped together
  - Base keywords that are effected similarly should be grouped together
- Might rediscover original groups, or find interesting new ones (or garbage)

Algorithmic Ideas

- Suppose l fixed base keyword groups
- Compute vector of length l per modifier
  - i^{th} dimension is average effect of modifier on keyword group i
- Can run k-means (or something else…)
  - Should produce clusters with desired property
- Now suppose k fixed modifer groups
  - Can do the same thing for base keywords

Algorithmic Ideas

- Idea: alternate k/l-means steps for base keywords and modifiers
  - Recompute vectors at each step

Randomly initialize prototypes

Assign modifiers to clusters

Re-center keyword prototypes

Assign keywords to clusters

Re-center modifier prototypes