Combinatorial Market Scoring Rules

Robin Hanson
Department of Economics
George Mason University

Old Tech Meets New

- To gain info, elicit probs $\mathbf{p} = \{p_i\}_i$, $E_{\mathbf{p}}[x \mid A]$
  (Let verify state i later, N/Q = people/questions)
- Old tech (~1950+): *Proper Scoring Rules*
  N/Q $\ll$ 1: works well, N/Q $\gg$ 1: hard to combine
- New tech (~1990+): *Info/Predict Markets*
  N/Q $\gg$ 1: works well, N/Q $\ll$ 1: thin markets
- The best of both: *Market Scoring Rules*
  – modular, lab tests, compute issues, …

Old Tech: Proper Scoring Rules

- When report **r**, state is i, reward is $s_i(\mathbf{r})$
  $\mathbf{p} = \text{argmax}_{\mathbf{r}} \Sigma_i p_i s_i(\mathbf{r})$, $\Sigma_i p_i s_i(\mathbf{p}) \geq 0$
- Quadratic (G. Brier 1950) $s_i \propto 2r_i - \Sigma_k r_k^2$
- Logarithmic (I. Good '52) $s_i \propto \log(r_i)$
  - Unique: reward = likelihood (R. Winkler 1969)
- Offers info effort incentive (R. Clemen '02)
  - Stronger for simul. pick: pivot mech. (T. Page '88)
- In principle, can elicit complex joint distributions
- Long used in forecasting, test scoring, lab expers

Old Tech Issues

Problems

- Incentives
- Number shy
- Cognitive bias
- Non risk-neutral
- State-dependent utility
- Combo explosion
- Disagreements

Solutions

- Proper scoring rules
- Prob wheel, word menu
- Corrections
- Lottery payoffs
- Insurance game
- Dependence network
- Dictator per Q, ??

Opinion Pool "Impossibile"

- Task: pool prob. T(A) from opinions $p^n(A)$
- Any 2 of IPP, MP, EB $\Rightarrow$ dictator ($T = p^d$) !

  IPP = if A,B indep. in all p^n, are indep. in T

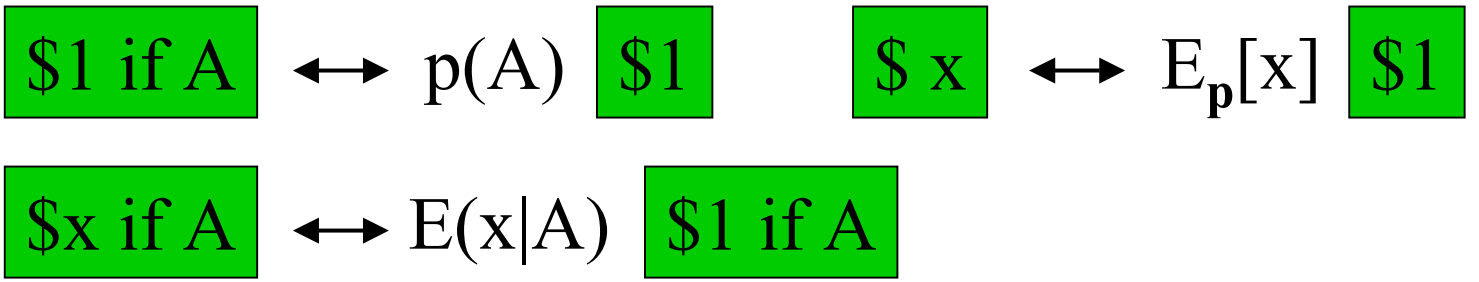
  EB = commutes: pool, update on info

  MP = commutes: pool, coarsen states ($\sigma-$field)

  (MP $\Rightarrow T = \Sigma_{n=0} w_n p^n$, with w_n indep. of A)
- Really want pool via belief origin theory
  - General solution: let best traders figure it out?

New Tech: Info/Predict Markets

- Most markets aggregate info as side effect
- Markets unbeaten in institution field comparisons
  - OJ weather (Roll '84), horse races (Figlewski '79), Oscar winners (Pennock et al '01), gas demand (Spencer '04), NASA accident (Maloney & Mulherin '03), HP printer sales 6/8 (Chen & Plott '98), I.E.M. beat election polls 451/596 (Berg et al '01)



New Tech Issues

Problems

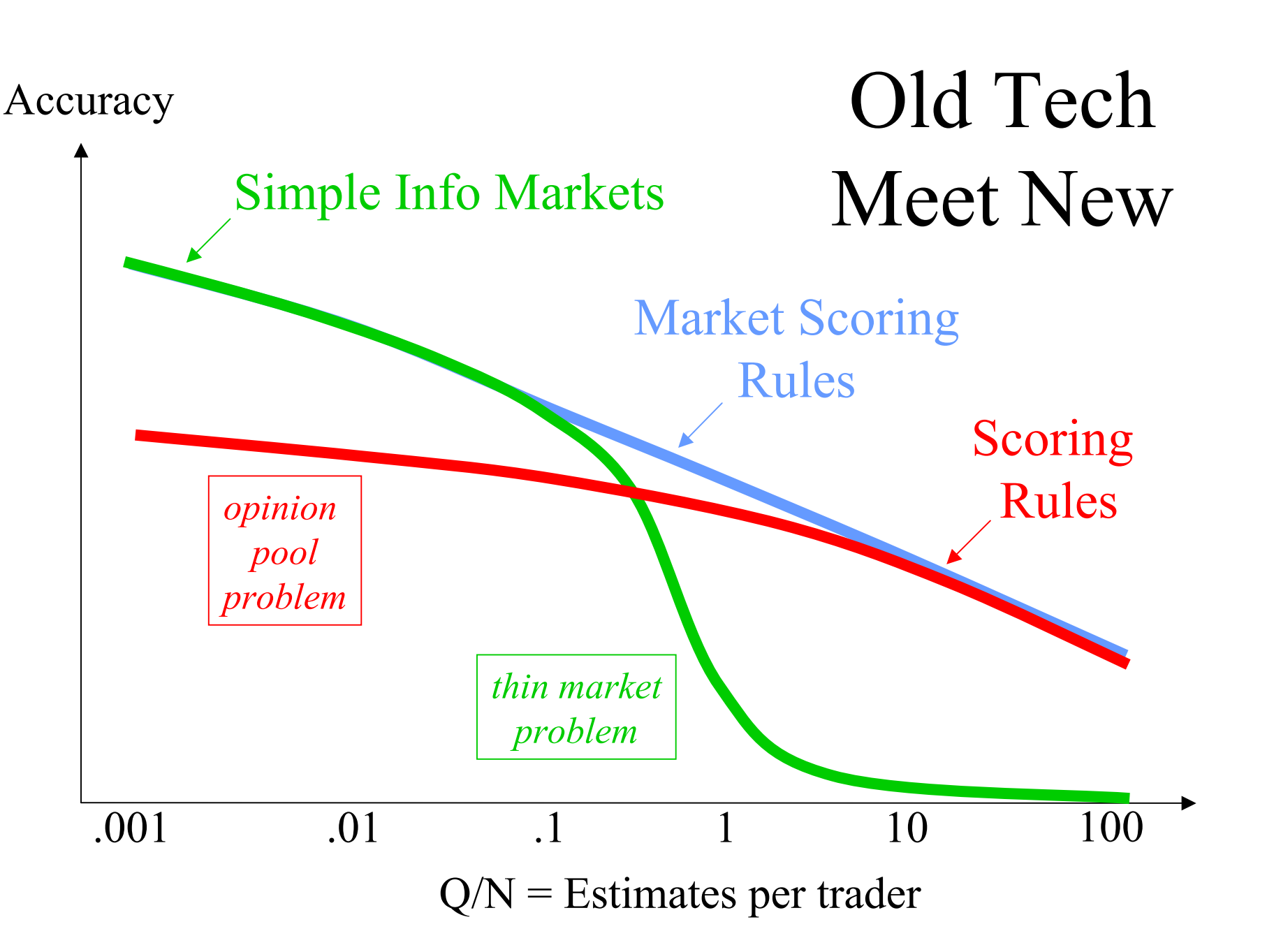
- Incentives
- Shy, complex utility
- Combo explosion
- Who expert on what
- Cognitive bias, pooling
- Thin markets (N/Q <~1)
- What is independent

Solutions

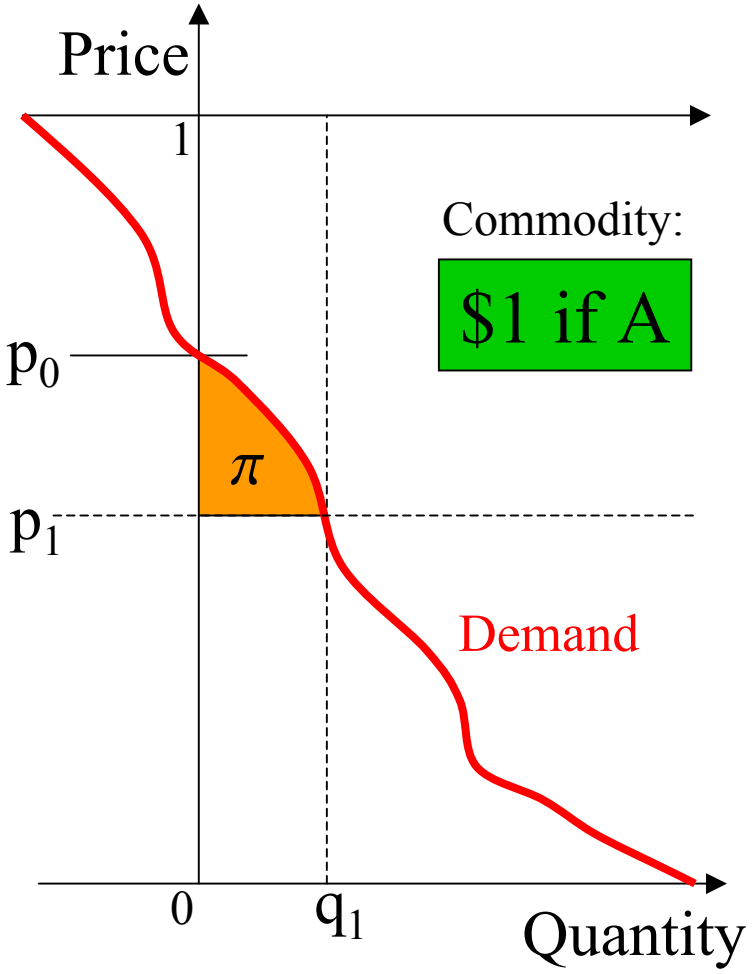
- Bet with each other
- Same solutions
- Same solutions
- Self-select
- Specialist traders
- Market scoring rules
- ??

Thin Market, No-Trade Problems

- Trade requires *coordination*
  - *In Time*: waiting offers suffer adverse selection
  - *In Assets:* far more possible assets than trades
    - combo match call markets help some, but matching is NP-hard for all or nothing orders
  - If expect traders rare, don't bother to offer
    - Most possible markets do not exist (also illegal)
- No-trade among rational, info-motivated
  - Need fools, risk-hedgers, or outside subsidy


Old Tech Meet New
Accuracy
Simple Info Markets
Market Scoring Rules
Scoring Rules
opinion pool problem
thin market problem
.001
.01
.1
1
10
100
Q/N = Estimates per trader

Scoring Rule = Demand Curve



- (L. Savage 1971)
- Rule is demand $q^D(p)$
  - p_0 solves $q=0$
- User sells q_1, price discr.
  - Let p_1 be user belief
- Note: $\pi = 0$ if $p_1 = p_0$
  - p_0 is like rule belief
- Note: can *reuse* rule as $q^D(p) - q_1$ (i.e., $p_0 \rightarrow p_1$)
  - So this is a market maker!

Market Scoring Rule (MSR)

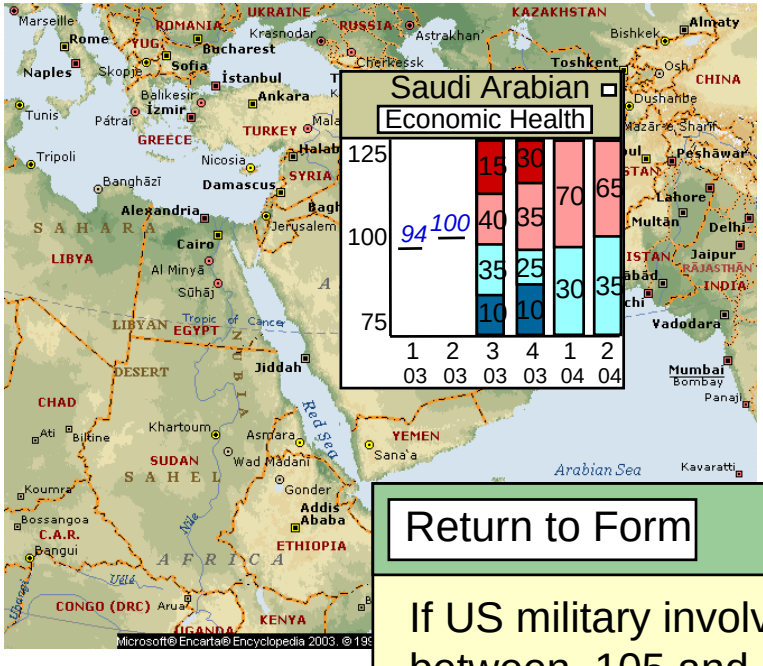
- MSRs act as scoring rules *and* info market makers
  - are sequentially reused combinatorial scoring rules
- Score rule: user t faces $ rule: $\Delta s_i = s_i(\mathbf{p}^t) - s_i(\mathbf{p}^{t-1})$

  "Anyone can use scoring rule if pay off last user"
- Automated market maker: price from net sales **s**
  - Tiny sale $\$ \epsilon_i$ if i fee: $\cong p_i(\mathbf{s})\, \epsilon_i$ $(s_i \rightarrow s_i + \epsilon_i)$
  - Big sale $\$ \mathbf{s}(1) - \mathbf{s}(0)$ fee: $\int_0^1 \Sigma_i\, p_i(\mathbf{s}(t))\, s_i'(t)\, dt$
  - Log score rule gives: $p_i(\mathbf{s}) = \exp(\lambda s_i) / \Sigma_k \exp(\lambda s_k)$

MSR Usage Concept

- User browses current probs, expectations
  - Can set assumptions, browse other values given that
  - Can see market value of portfolio given assumptions
- Upon finding an "odd" current value E(x|A)
  - Can see value's price/trade history
  - Can see how far long/short are from past trades
- User proposes new value to replace old
  - Told exact bet required to implement change
  - Can accept, and/or make book orders & trading agents

PAM Example



Return to Focus | Trade | Update | ?

	Select	New Price	Payoffs: If SAum3 03 105-125 & IQcs4 03 <85	IQcs4 03 >85	Ave. pay <85 65%
Buy	Max Up	95.13%	+$34.74	-$85.18	-$19.72
	10% Up	68.72%	+$2.74	-$3.28	-$1.07
	You Pick	65 %	+1.43	-2.04	+0.34
	No Trade	62.47%	$0.00	$0.00	$0.00
Sell	10% Dn	56.79%	-$2.61	+$2.74	-$1.12
	Exit Issue	48.54%	-$15.34	+$26.02	-$6.31
	Max Dn	22.98%	-$120.74	+$96.61	-$22.22

Return to Form | Execute a Trade | ?

If US military involvement in Saudi Arabia in 3rd Quarter 2003 is not between 105 and 125, this trade is null and void. Otherwise, if Iraq civil stability in 4th Quarter 2003 is below 85, then I will receive $1.43, but if it is not below 85, I will pay $2.04.

Abort trade if price has changed ☐ | **Execute**

Who is Using What

- Most forecasting: cheap talk, no formal eval.
- Most formally scored forecasts: scoring rules
- Most info markets: simple double auctions
  - HP switched: info markets → scoring rules!
- Using 1-Dim. MSR (= sales-based market maker):
  - Hollywood Stock Exchange, Microsoft, Yahoo?
- Using Combinatorial Market Scoring Rules:
  - Common Knowledge Markets, YooNew, US Air Force Research, Commerce Net, … (my public sample code)

Laboratory Tests

- Joint work with John Ledyard (Caltech), Takashi Ishida (Net Exchange)
- Caltech students, get ~$30/session
- 6 periods/session, 12-15 minutes each
- Trained in 3var session, return for 8var
- Metric: *Kulback-Leibler* $\Sigma_i \, q_i \log(p_i / q_i)$ distance from mechanism estimate to Bayesian beliefs given all group info

Environments: Goals, Training

- ***Want in Environment:***
  - Many vars, few related, guess which
  - Use both theory and data
  - Few people, specialize in variable sets
  - Can compute rational group estimates
  - Explainable, fast, neutral
- ***Training Environment:***
  - 3 binary variables X,Y,Z, $2^3 = 8$ combos
  - P(X=0) = .3, P(X=Y) = .2, P(Z=1)= .5
  - 3 people, see 10 cases of: AB, BC, AC
  - Random map XYZ to ABC

(Actually: X Z Y)

Case	A	B	C
1	1	-	1
2	1	-	0
3	1	-	0
4	1	-	0
5	1	-	0
6	1	-	1
7	1	-	1
8	1	-	0
9	1	-	0
10	0	-	0
Sum:	9	-	3

Same	A	B	C
A	--	--	4
B	--	--	--
C	--	--	--

Experiment Environment

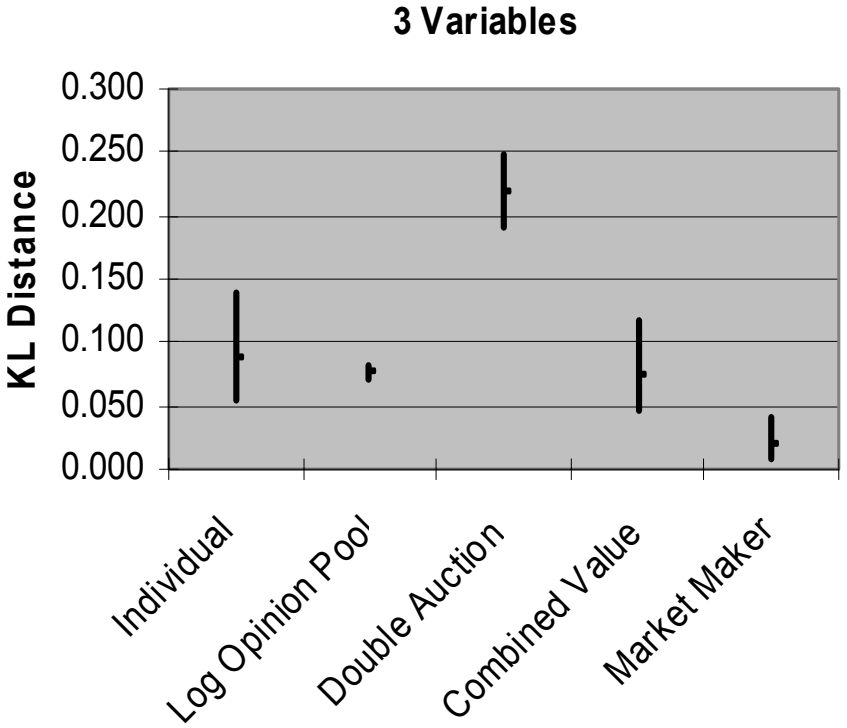
- 8 binary vars: STUVWXYZ
- $2^8 = 256$ combinations
- 20% = P(S=0) = P(S=T) = P(T=U) = P(U=V) = … = P(X=Y) = P(Y=Z)
- 6 people, each see 10 cases: ABCD, EFGH, ABEF, CDGH, ACEG, BDFH
- random map STUVWXYZ to ABCDEFGH

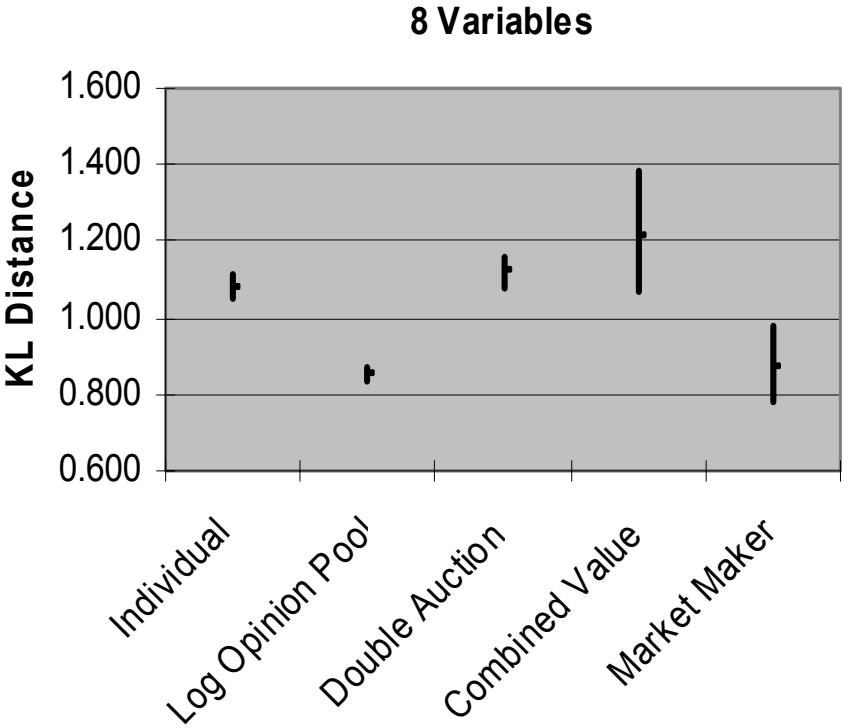
(*Really:* *W* *V* *X* *S* *U* *Z* *Y* *T*)

Case	A	B	C	D	E	F	G	H
1	0	1	0	1	-	-	-	-
2	1	0	0	1	-	-	-	-
3	0	0	1	1	-	-	-	-
4	1	0	1	1	-	-	-	-
5	0	1	1	1	-	-	-	-
6	1	0	0	1	-	-	-	-
7	0	1	1	1	-	-	-	-
8	1	0	0	1	-	-	-	-
9	1	0	0	1	-	-	-	-
10	1	0	0	1	-	-	-	-
Sum	6	3	4	10	-	-	-	-

Same	A	B	C	D	E	F	G	H
A	--	1	2	6	--	--	--	--
B	--	--	7	3	--	--	--	--
C	--	--	--	4	--	--	--	--
D	--	--	--	--	--	--	--	--
…								

Accuracy 95% Confidence Levels

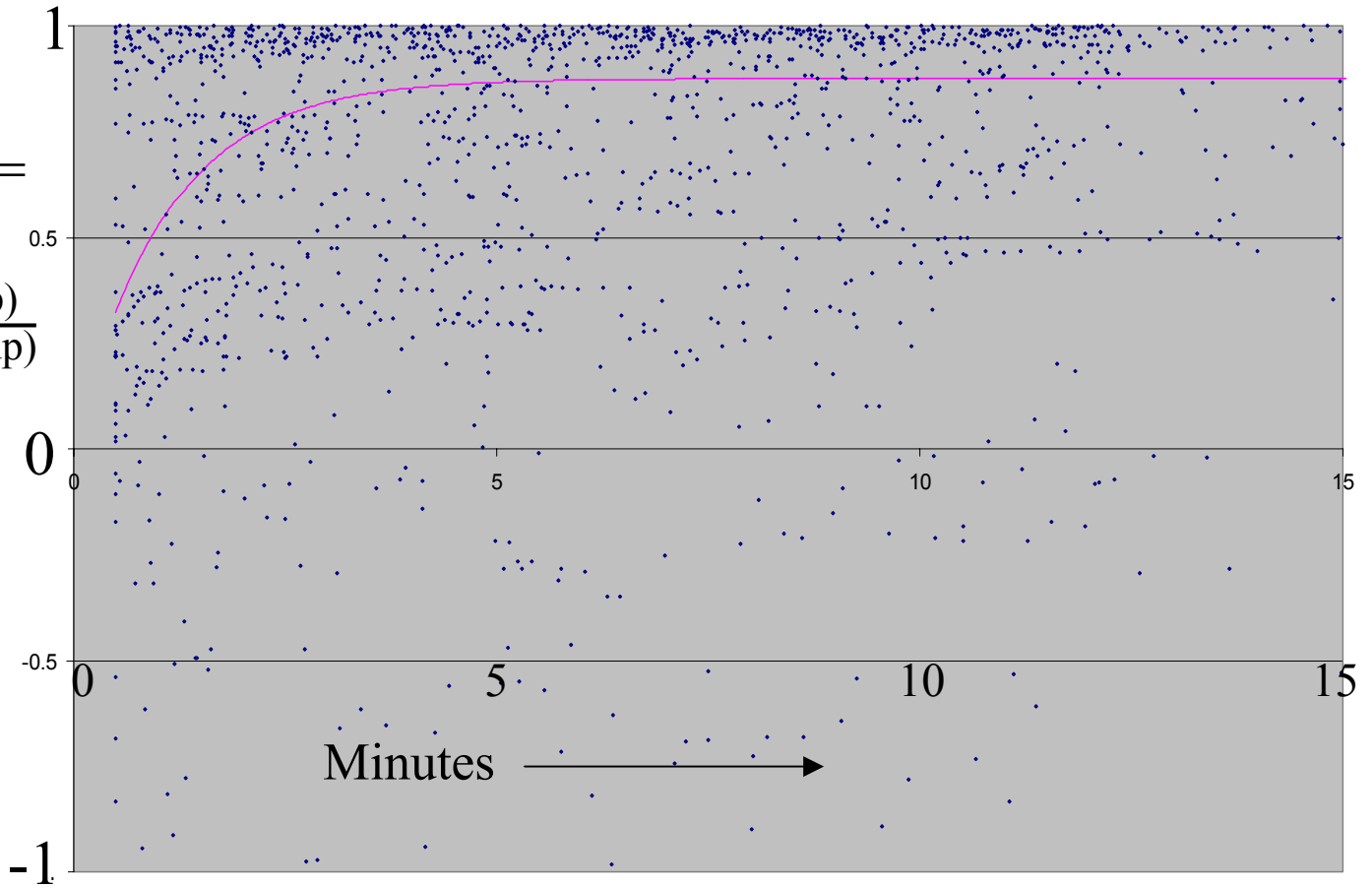




MSR Info vs. Time – 3 Variables

% Info Agg. =

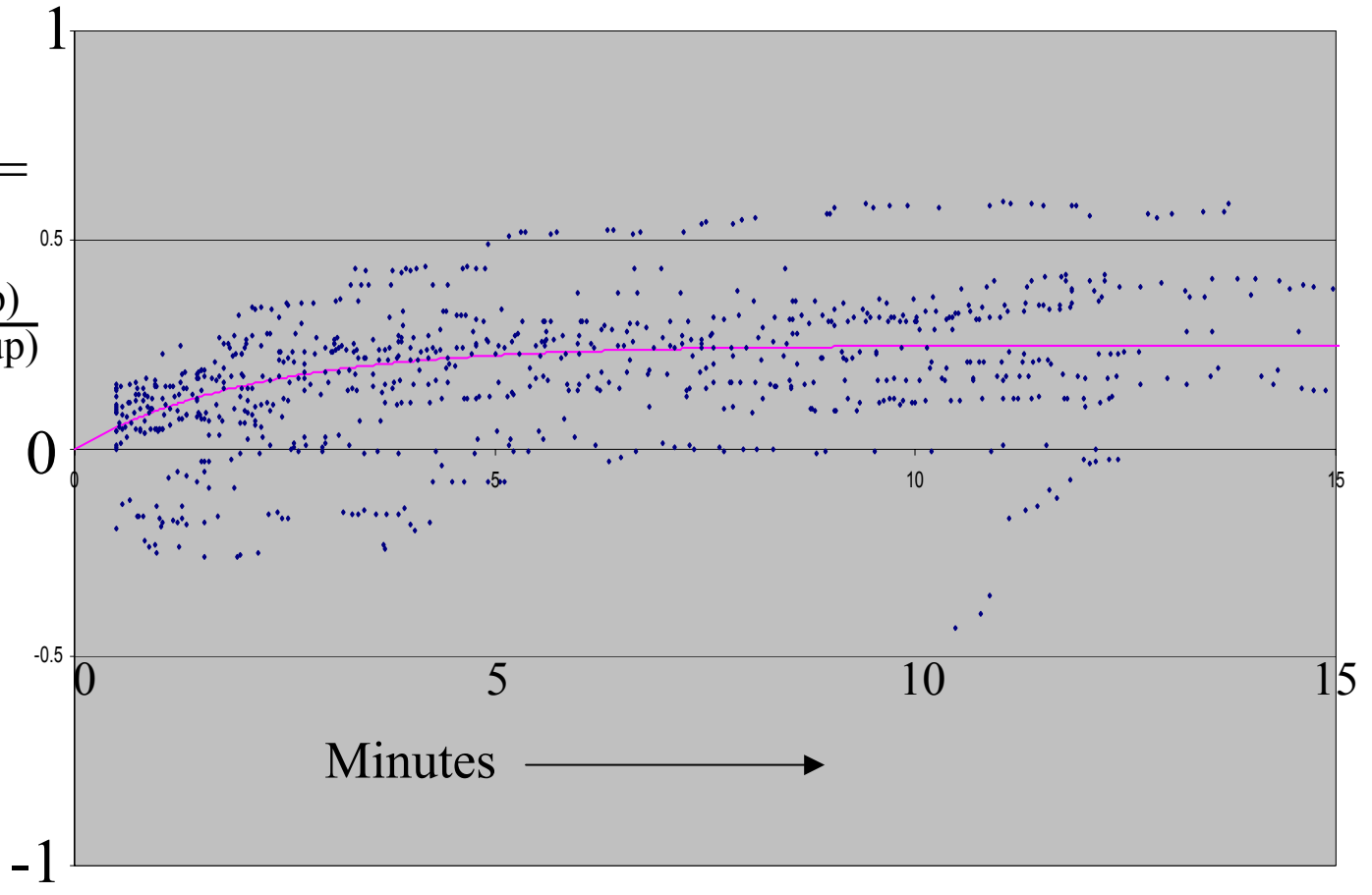
$$1-\frac{KL(prices,group)}{KL(uniform,group)}$$



MSR Info vs. Time – 8 Variables

% Info Agg. =

$$1-\frac{KL(prices,group)}{KL(uniform,group)}$$



Experiment Conclusions

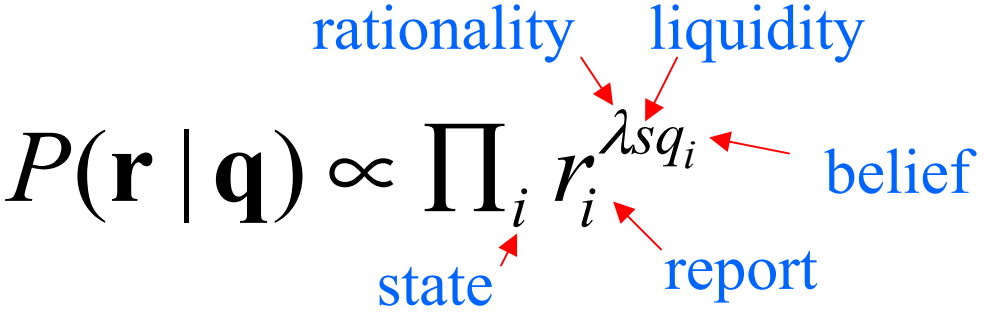
- Experiments on complex info problems
  - Bayesian estimates far too high a standard
- 7 indep. prices from 3 people in < 4 minutes
  - Simple DA < Indiv. Score Rule ~ Opinion Pool ~ Combined Value < Market Scoring Rule
- 255 indep. prices from 6 people in < 4 min.
  - Combined Value ~ Simple DA ~ Indiv. Score Rule < Opinion Pool ~ Market Scoring Rule

Log Rule is Modular

- Consider bet: \$1 if A&B ↔ p(A|B) \$1 if B
- Changes p(A|B); only log rule keeps p(B)
- Also keeps p(C|A&B), p(C|¬A&B), p(C|¬B), I(A,B,X), I(B,A,X), I(X,A,B), I(X,B,A)
  - A is var, one of whose values is A, etc.
  - I(A,B,X) iff p(A|B&C) = p(A|B) for all values
- Log rule uniquely keeps changes modular

Quantal Response Modularity

- Noisy choice: prob(act) $\propto$ exp(λ*payoff)
- When apply to a log MSR, get user reports (= new prices) independent of the last price:

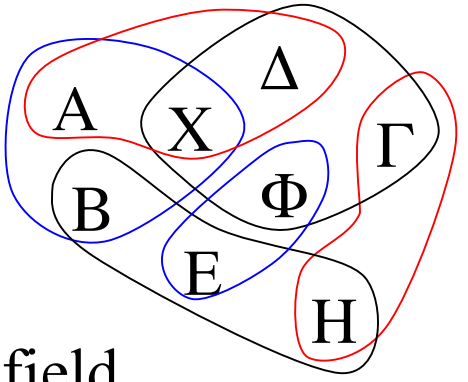


- Simplifies inferences about beliefs from acts
- Ignores that harder to make complex changes

No Cost For Combos!

- Total cost: $C = s_{i\,true}(\mathbf{p}^{final}) - s_{i\,true}(\mathbf{p}^{initial})$
- Expected cost: $E_{\boldsymbol{\pi}}[C] \leq \Sigma_i\, \pi_i\, (s_i(\mathbf{1}_i) - s_i(\boldsymbol{\pi}))$
- For log MSR: $= S(\boldsymbol{\pi}) = -\Sigma_i\, \pi_i \log(\pi_i)$
  - Let state i = combination of base var values v_i
  - $S(\boldsymbol{\pi}_{all}) \leq \Sigma_{var}\, S(\boldsymbol{\pi}_{var})$, for $\boldsymbol{\pi}_{var} = \{\pi_{var\ value\ v}\}_v$
  - So compared to cost of log rule for each var, all var/value combos cost no more!

Computation Issues



- Simple: store 2^N probs, asset values
  - Can integrate book orders
- Feasible: overlapping var patches
  - With simple MSR per patch, is Markov field
  - Allow trade only if all vars in same patch
- How pick/change patch structure?
- How make patches agree?
  - Arbitrage neighbors is robust, but non-modular
  - Bayes rule modular, but NP-hard w/o patch tree
    - Can approximate update avoid arbitrage?

A garden path …

- *Market Maker:* if subsidize or allow single/impatient user, "system" must trade
- *Dutch book:* if let tiny "fair" bets on enough E(x|A), prices = prob. distr., e.g., p_i for $1 if i
- *Sales-based:* if Δ price rule arbitrage free, known, $\mathbf{p} = f(\mathbf{s})$, for sales $\mathbf{s} = (s_i)_i$, $s_i = \#$ $1 if i
- *Exp:* if p(A|B) bet keeps p(B) , $p_i \propto \exp(s_i)$

Summary

- How elicit informed estimates?
  - Scoring rules if N/Q $\ll 1$, info/predict markets if $\gg 1$
- Market scoring rules do both:
  - Key insight … reused scoring rules *are* market makers
  - Browse billions of estimates, change ones want via bets
  - Lab tests confirm ability; are growing # groups using
- Log rule has many advantages
  - If bet on p(A|B), keeps p(B), I(A,B,X), I(B,A,X) …
  - Noisy choices easier to interpret
  - Costs no more for all var/value combos
  - Computation simplified, but still issues to explore