Information Theory

in Horseracing, the Share Markets and in Life

Overview

1. What Is Information? How do we Measure it?

2. Using Information Theory for Financial Gain

3. Entropy Modeling: Using Information Theory to Learn

4. Financial Gain Revisited: Universal Portfolios

5. Conclusions

1. What is Information?

The Description of Structure

Information is conveyed between a **Transmitter** and a **Receiver**, sharing a common **Subjectivity**, as made specific by a **Language** or 'Code'.

What is Information Used For?

Assumption: Information may be stored, but is ultimately used for Taking Action and/or Communicating.

How does one Quantify Information?

Approach: Money. We shall eventually arrive at the approach of Shannon, via Entropy, and then come back to Money.

Editorial Remark

Economic Rationalism: ’include it if it fits on a balance sheet. If it’s too hard to figure out how to put it there, then don’t’

It shall be argued that if one understands the structure of a Market well enough to compress Market data better than the Market could, then one can make money investing in that Market. The **Rate of Growth** which can be achieved is the **same** as the excess amount by which the data can be **compressed** relative to the Market's capability.

The **units** of compression may represented either by bits, or equivalently in terms of Shannon Information, or Entropy, as used in Thermodynamics, or equivalently, by Complexity as defined in Coding Theory, Number Theory, Engineering, and Computer Science.

On Investment

Money M_0 invested at fixed rate r will grow over time t like

$$M_t = M_0 e^{rt}$$

r is the slope, or 'Rate of Growth' of $\log(M_t)$:

$$\log(M_t) = \log(M_0) + rt$$

Idea (can be justified mathematically): optimise average Rate of Growth of $\log(M_t)$ even if returns are random:

$$E\{\log(M_{t+1})) - \log(M_t)\} = \max!$$

$\Longrightarrow$ Many Optimality Properties, including minimising Expected Waiting time to any Fixed Financial Goal

App: Log-Optimality in Investment ’Kelly Criterion’

Example Four-horse race.

no.	Horse	Market Div.	Your Div.
1.	Armstrong	4	2
2.	Bix	4	4
3.	Cheetham	4	8
4.	Dizzy	4	8

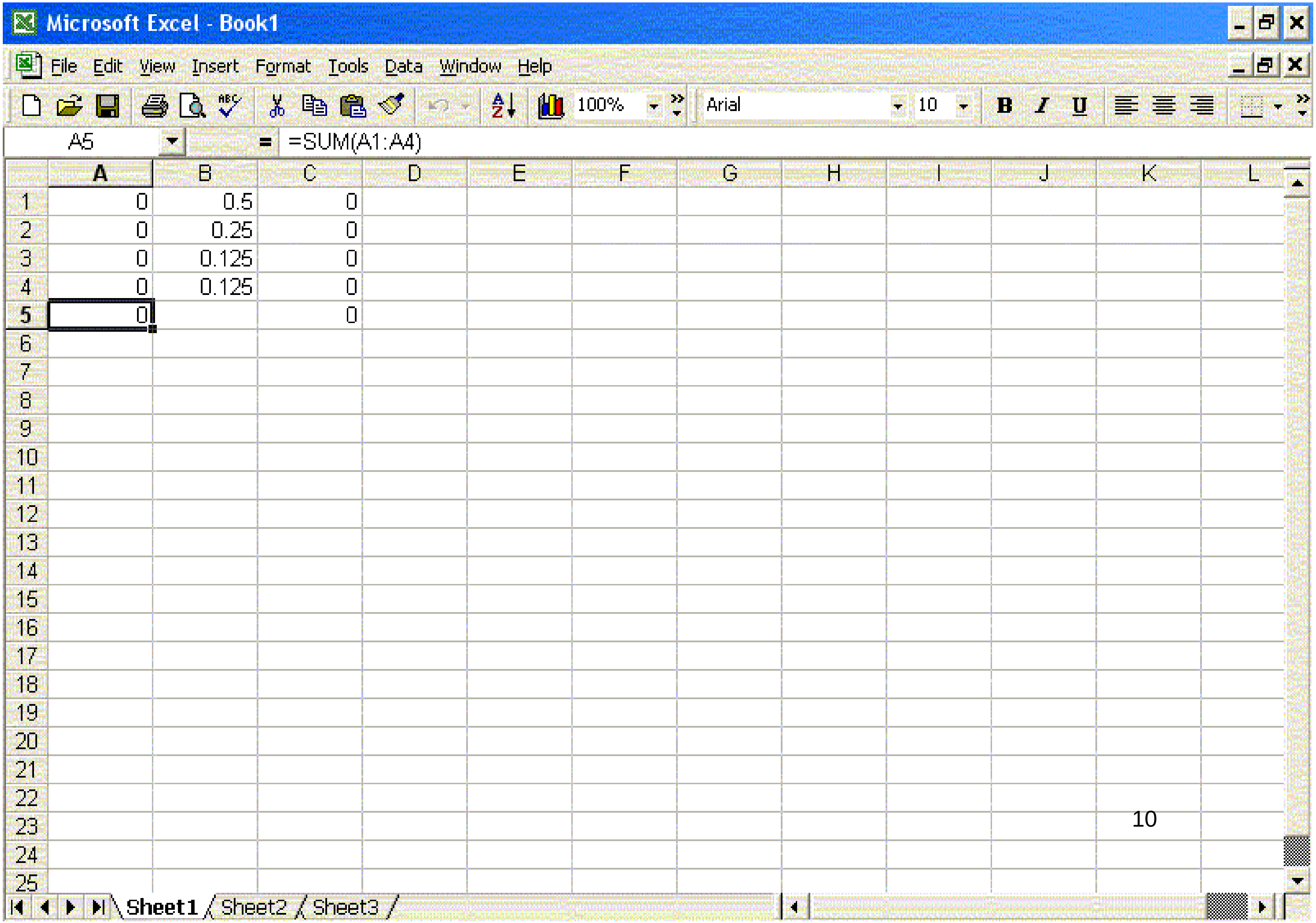
Assuming you are right about chances, how do you use your information optimally, and what can you achieve by way of logarithmic growth?

Bet prop. of wealth c_1 on Armstrong, c_2, c_3, and c_4 on the others, where c's are all positive or zero, and total is not more than 1 (100% of your wealth -no borrowing).

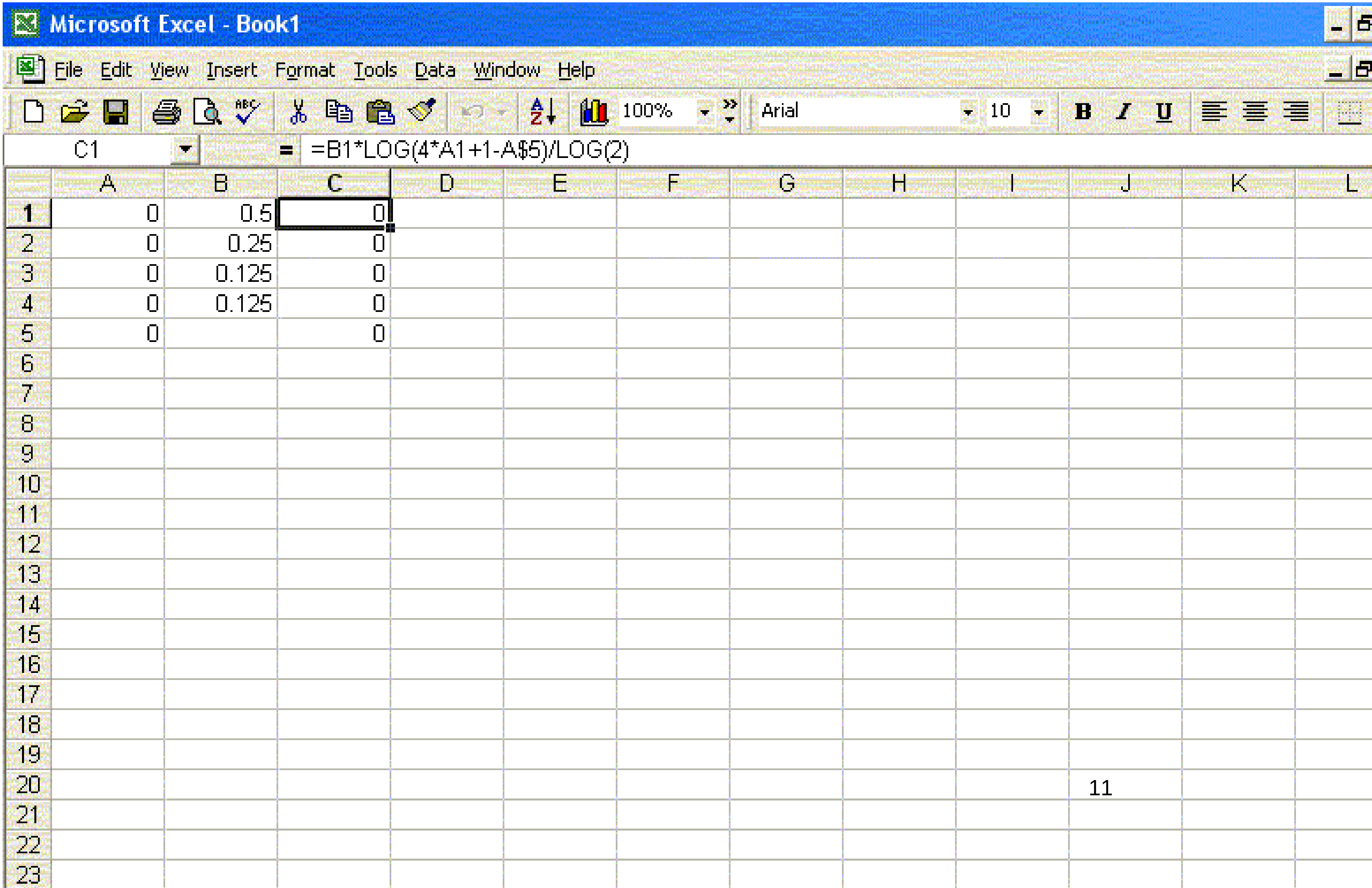
$$\frac{1}{2}\log_2(4c_1+c_0)+\frac{1}{4}\log_2(4c_2+c_0)+\frac{1}{8}\log_2(4c_3+1-c_0)+\frac{1}{8}\log_2(4c_4+c_0),$$

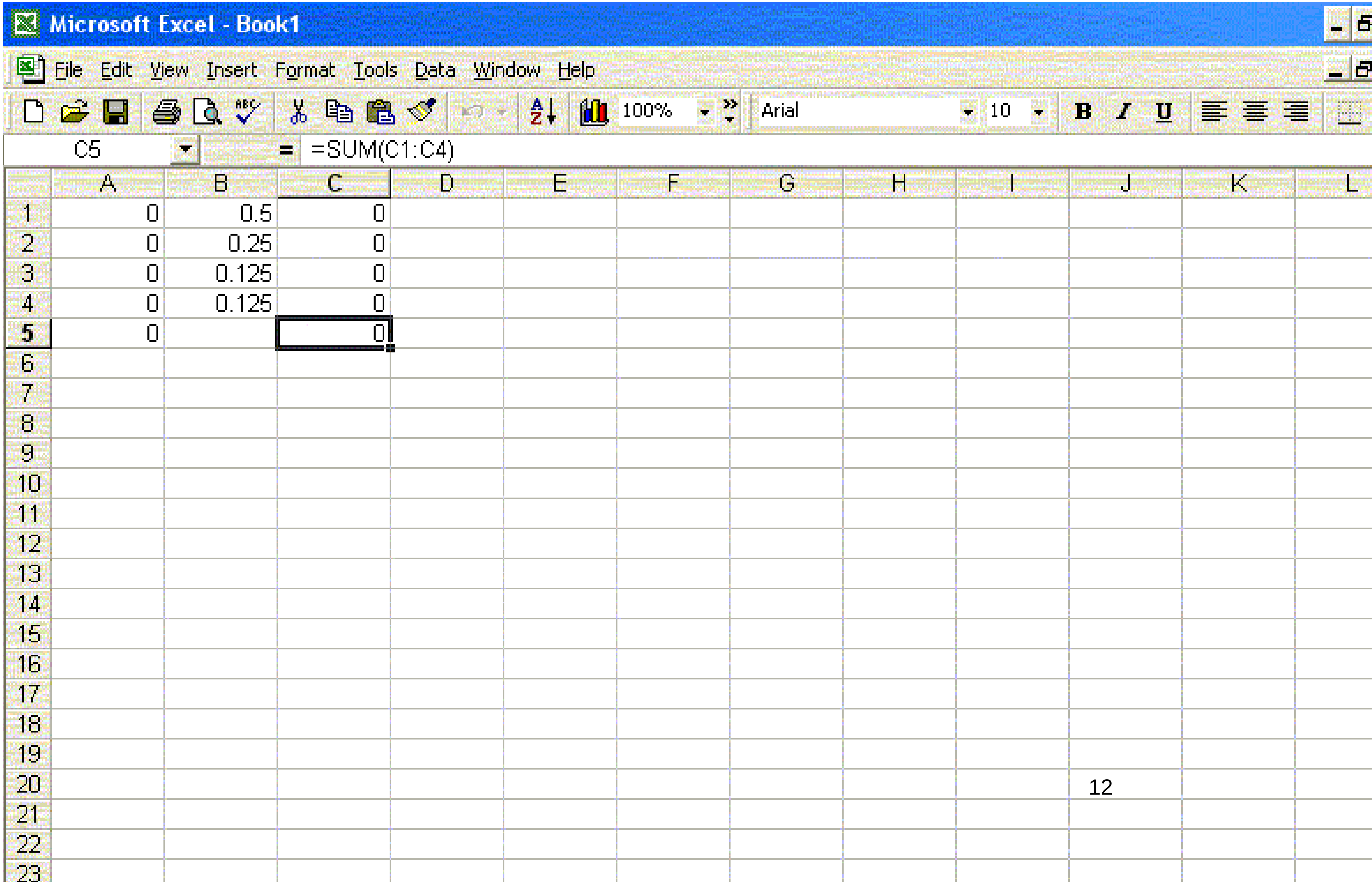
where $c_0 = 1-\sum c_i$ is amount retained in cash.

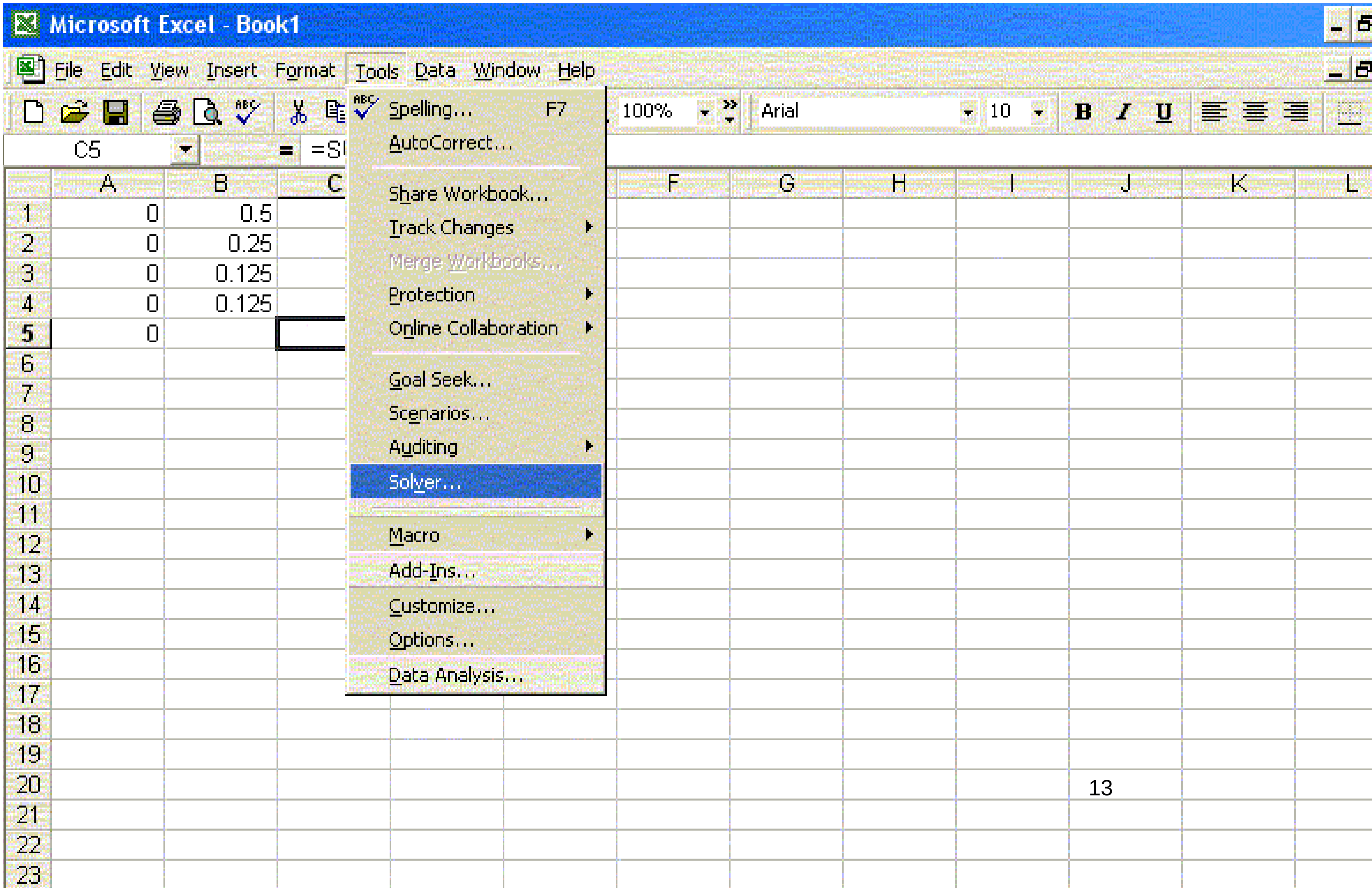
Log-Growth may be Maximised over $c's$. (maths or Excel Solver)

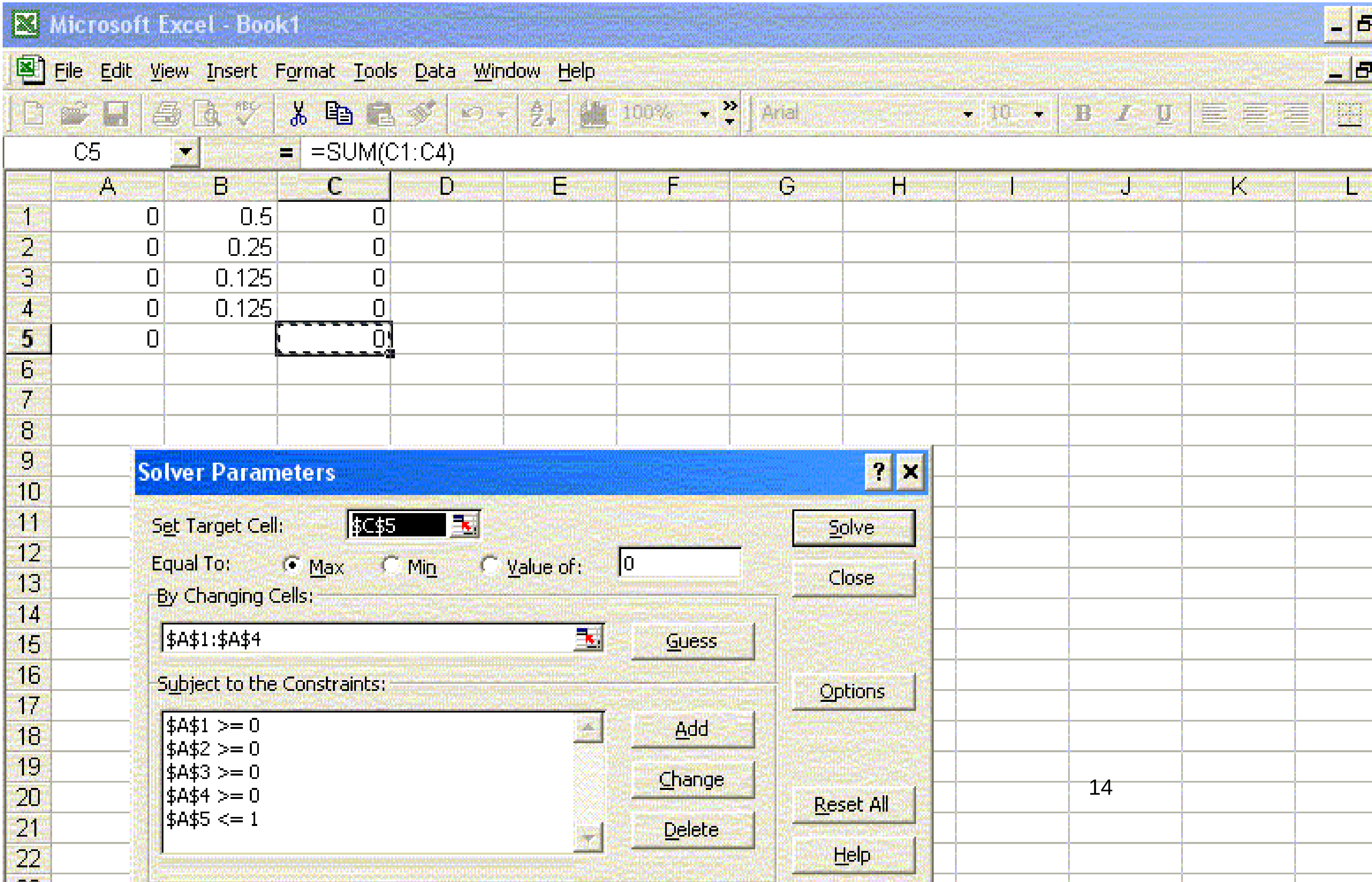


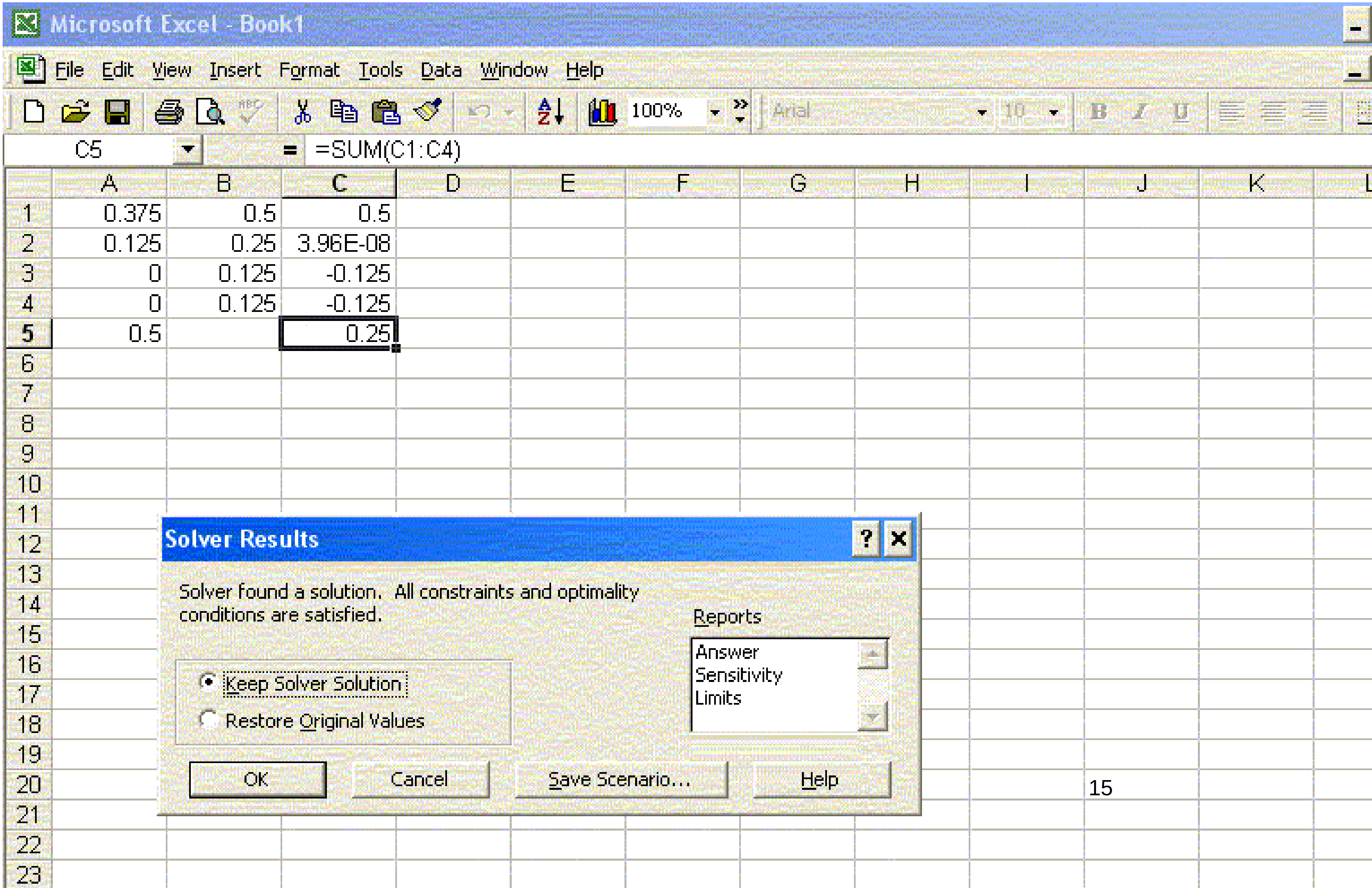
	A	B	C
1	0	0.5	0
2	0	0.25	0
3	0	0.125	0
4	0	0.125	0
5	0		0


Microsoft Excel - Book1
File Edit View Insert Format Tools Data Window Help
100%
Arial
10
C1
=B1*LOG(4*A1+1-A$5)/LOG(2)
A B C D E F G H I J K L
1 0 0.5 0
2 0 0.25 0
3 0 0.125 0
4 0 0.125 0
5 0 0


Microsoft Excel - Book1
File Edit View Insert Format Tools Data Window Help
100%
Arial
10
C5
=SUM(C1:C4)
A B C D E F G H I J K L
1 0 0.5 0
2 0 0.25 0
3 0 0.125 0
4 0 0.125 0
5 0 0


Microsoft Excel - Book1
File Edit View Insert Format Tools Data Window Help
100%
Arial
10
C5
=SI
A B C F G H I J K L
0 0.5
0 0.25
0 0.125
0 0.125
0
Spelling... F7
AutoCorrect...
Share Workbook...
Track Changes
Merge Workbooks...
Protection
Online Collaboration
Goal Seek...
Scenarios...
Auditing
Solver...
Macro
Add-Ins...
Customize...
Options...
Data Analysis...


Microsoft Excel - Book1
File Edit View Insert Format Tools Data Window Help
C5
=SUM(C1:C4)
A B C D E F G H I J K L
1 0 0.5 0
2 0 0.25 0
3 0 0.125 0
4 0 0.125 0
5 0 0
Solver Parameters
Set Target Cell: $C$5
Equal To: Max Min Value of: 0
By Changing Cells:
$A$1:$A$4
Guess
Subject to the Constraints:
$A$1 >= 0
$A$2 >= 0
$A$3 >= 0
$A$4 >= 0
$A$5 <= 1
Add
Change
Delete
Solve
Close
Options
Reset All
Help


Microsoft Excel - Book1
File Edit View Insert Format Tools Data Window Help
100%
Arial
C5
=SUM(C1:C4)
A B C D E F G H I J K
1 0.375 0.5 0.5
2 0.125 0.25 3.96E-08
3 0 0.125 -0.125
4 0 0.125 -0.125
5 0.5 0.25
Solver Results
Solver found a solution. All constraints and optimality conditions are satisfied.
Reports
Answer
Sensitivity
Limits
Keep Solver Solution
Restore Original Values
OK
Cancel
Save Scenario...
Help

Optimal 'portfolio': $c_1 = \frac{3}{8}$, $c_2 = \frac{1}{8}$, and $c_3 = c_4 = 0$, and the expected logarithmic Growth Rate is

$$\frac{1}{2}\log_2(4(\frac{3}{8}) + \frac{1}{2}) + \frac{1}{4}\log_2(4(\frac{1}{8}) + \frac{1}{2}) + \frac{2}{8}\log_2(\frac{1}{2})$$

$$= \frac{1}{2}\log_2(2) + \frac{1}{4}\log_2(1) + \frac{1}{4}\log_2(\frac{1}{2}) = \frac{1}{4}$$

Achievable Logarithmic Growth Rate (base 2): $\frac{1}{4}$
For m such opportunities

$$Wealth \sim 2^{\frac{1}{4}m}$$

indicating doubling of wealth every 4 races.

Aside: if T is a stopping time req'd time to reach A,

$$A = E(2^{\sum_1^T LGR_i}) = 2^{E(LGR)E(T)} \quad \text{(Wald)}$$

so

$$E(T) = \frac{\log_2(A)}{E(LGR)} \geq \frac{\log_2(A)}{1/4}$$

Example - cont'd - Data Compression of Outcome

no.	Horse	Naïve Coding
1.	Armstrong	00
2.	Bix	01
3.	Cheetham	10
4.	Dizzy	11

2-Bit code - Best the 'Market' can do if Divs. are 4,4,4, and 4

Your Market - Required Dividends

no.	Horse	Div.
1.	Armstrong	2.00
2.	Bix	4.00
3.	Cheetham	8.00
4.	Dizzy	8.00

With Prefix Coding & Description Length

no.	Horse	Div.	Code	DL (bits)
1.	Armstrong	2.00	1	1
2.	Bix	4.00	01	2
3.	Cheetham	8.00	001	3
4.	Dizzy	8.00	000	3

$$E(DL) = \frac{1}{2}(1) + \frac{1}{4}(2) + \frac{1}{4}(3) = 1\frac{3}{4}$$

$\frac{1}{4}$-bit below 2!   Equal to LGR (no coincidence)

With 'Likelihood Score' - Negative Log-Probability

no.	Horse	Div.	Code	(DL) bits	$\log_2(Div)$
1.	Armstrong	2.00	1	1	1
2.	Bix	4.00	01	2	2
3.	Cheetham	8.00	001	3	3
4.	Dizzy	8.00	000	3	3

Define Entropy: $\sum -p\log(p)$, here,

$$Entropy = \frac{1}{2}(1) + \frac{1}{4}(2) + \frac{1}{4}(3) = 1\frac{3}{4}$$

Note: DL of optimal code equals Log-Likelihood

$E(DL)$ equals Entropy (units: Information)

’Management’ of Entropy enables Financial Gain here

Units: $\sum -p\log(p)$, or ’Shannon Information’

Qualitative Interpretation: The better one understands something, the more succinctly one can describe it. In competitive Markets, succinctness can be related directly to profitability.

Mathematical Theorem [Edelman, 2000]: This holds in general, for continuous distributions,and in continuous time, not just simple discrete gambling games [Breiman 1965].

Data Compression – Comment: ’Zipping’ files on your computer

What happens?

Large file reduced in size, content entirely reproducible.

How Does it Work?

Program knows structure(e.g. English text with lots of spaces) and exploits it to *encode* as efficiently as possible, decoding ’key’ included.

Compression and Knowledge

The more completely structure is understood, the more efficiently a file may be compressed.
[Note that digital pictures do not compress much using standard 'ZIP' programs, but text does.]

A simple definition of **Knowledge** is the ability to compress data.

The process of discovering how to compress data is **Knowledge Creation**

In a given context, this is made manifest by the ability to design an efficient Language of Concepts, as communicated via words or symbols, amounting to a **Code**.

Other Conceptual Examples

People are awarded Nobel Prizes for discoveries which drastically reduce the complexity of describing or characterising natural phenomena. Awards do not go generally to the most **complicated** theories, but rather usually the **simplest** which explain a broad range of phenomena.

Example: Harry Markowitz's *Modern Portfolio Theory*

Harry Markowitz's *Modern Portfolio Theory*

He had found that by characterising investors' desire for Gain as **Expected Return** and distaste for Risk as **Variance of Return**, a wide variety of preference profiles and pricing phenomena in the Financial Markets could be explained. He took the formerly nebulous notions of Risk and Return and greatly simplified the characterisation of Markets. The Mathematics is fairly simple, but he was recognised (albeit somewhat belatedly).

3. Entropy Modeling

Mathematical Definition of Entropy

$$\mathcal{E} = -\sum p_i \log(p_i)$$

or

$$\mathcal{E} = -\int f(x) \log(f(x)) dx$$

(differential Entropy)

Remark: The second law of Thermodynamics says that Entropy (Complexity) of Physical systems increases over time. By analogy Economic, Social, Political systems, etc have been argued to follow suit. Therefore, for systems in/approaching steady state, assuming the Entropy of unconstrained components of a system is a large as possible is not an unreasonable modeling assumption.

Maximum Entropy and Constraints

Obviously, equiprobable is 'hardest' to forecast, hence Maximum Entropy.

Constrained Optimisation Problem: Maximise

$$-\sum p_i \log_e(p_i) \quad \text{subject to} \quad \sum p_i = 1$$

Lagrange Multiplier Characterisation

$$\frac{\partial}{\partial p_k}\{-\sum p_i \log(p_i) - \lambda(\sum p_i - 1)\} = -1 - \log(p_k) - \lambda$$

[Note:

$$\frac{\partial}{\partial \lambda}\{-\sum p_i \log(p_i) - \lambda(\sum p_i - 1)\} = \sum p_i - 1$$

showing how the optimisation problem enforces $\sum p_i = 1$.]

Solution (in 'character'): $p_k = e^{-1-\lambda}$, showing p_k are constant over k. Applying constraint, $p_k \equiv \frac{1}{n}$, where n is the number of possibilities.

Example Waiting for a bus. Average waiting time is 'known' to be 15 mins. You wait 20 mins. What is your further expected waiting time?

First, we need to fi nd a 'reasonable' candidate for the distribution of waiting time.

Suppose the waiting time t is measured in minutes. Lagrangian is

$$-\sum p_i \log(p_i) - \lambda_0(\sum p_i - 1) - \lambda_1(\sum t_i p_i - 20)$$

Differential w.r.t. p_k:

$$-1 - \log(p_k) - \lambda_0 - \lambda_1 t_k$$

Set to zero and get

$$p_k = e^{-1-\lambda_0-\lambda_1 t_k},$$

recognisable as the general form of the Negative Exponential distribution:

$$f(t) = e^{-t/\mu}/\mu, \;\; t > 0 \qquad\qquad P(T > t) = 1 - F(t) = e^{-t/\mu}$$

The Conditional Distribution of Wait, given we have already waited 20 min.

$$P(T > t | T > 20) = \frac{F(t)}{F(20)} = e^{-t/15+20/15} = e^{-(t-20)/15}, \;\; t > 20$$

i.e., your additional wait is Negative Exponential with average 15 min!

['memorylessness' of the Negative Exponential distribution]

Suppose, that, instead of the mean, the variance is constrained.

Lagrangian

$$-\sum p_i \log(p_i) - \lambda_0(\sum p_i - 1) - \lambda_1(\sum x_i^2 p_i - \sigma^2)$$

Differential w.r.t. p_k:

$$-1 - \log(p_k) - \lambda_0 - \lambda_1 x_k^2$$

Set to zero and get

$$p_k = e^{-1-\lambda_0-\lambda_1 x_k^2},$$

recognisable as the general form of the Gaussian

$$f(x) = e^{-\frac{1}{2}\frac{x^2}{\sigma^2}}/\sqrt{2\pi}\sigma$$

Uniform, Negative Exponential, and Gaussian are the most common Maximum Entropy distributions.

Horseracing. Given 1-2 Place probabilities $p_1, p_2, \ldots, p_n$, what are 'reasonable' Quinella (1-2 combination) probabilities?

Lagrangian: (suppose $q_{ij} = 0$ for $i \leq j$)

$$-\sum_{i,j} q_{ij} \log(q_{ij}) - \sum_i \lambda_i((\sum_j q_{ij}) - p_i)$$

Differential (w.r.t. $q_{i_0 j_0}$)

$$-1 - \log(q_{i_0 j_0}) - \lambda_{i_0} - \lambda_{j_0}$$

Of the form

$$q_{ij} = C \exp(-\lambda_i - \lambda_j) = Ce^{-\lambda_i} e^{-\lambda_j}$$

or equivalently

$$q_{ij} = \frac{2\tilde{p}_i \tilde{p}_j}{1 - \sum \tilde{p}^2}, \quad (\text{where} \quad \sum \tilde{p}_i \equiv 1)$$

One may solve interatively for $\tilde{p}_1, \ldots, \tilde{p}_n$ in terms of $p_1, \ldots, p_n$.

Note also that

$$\sum_j q_{ij} = \frac{2\tilde{p}_i(1-\tilde{p}_i)}{1-\sum \tilde{p}^2}$$

with $\tilde{p}_i$ being, arguably, monotonic functions of the Win probabilities, suggesting a reasonable 1-2 Place model.

Aside: One possibility

$$\tilde{p}_i = \frac{(p_i^{(Win)})^\alpha}{\sum (p_j^{(Win)})^\alpha}$$

($\alpha \sim .94$ from data).

Finance

Extracting Price Distribution information from an Options Market

Let $f(\cdot)$ denote the (unknown) risk-neutral distribution of (logarithmic) return

Constraints:

$$\int f(x)dx = 1$$

$$\int e^x f(x)dx = e^{rt} \quad (\mathrm{Risk-Neutrality})$$

$$\int [P_0 e^x - K_i]^+ f(x)dx = c_i, \;\; i = 1, m \quad (\mathrm{Mkt.OptionsPrices})$$

Idea (Edelman and Buchen [1994], Buchen and Kelly 1996 [JFQA]): Solve for $f(\cdot)$ by Maximising Entropy subject to above.

Previous (Rubinstein): optimise ’lack-of-smoothness’ penalty term $\int f''(x)^2$.

Comment 1: In Options context, Maximum Entropy seems less ’arbitrary’ and appears to give reasonable solutions for a wide variety of problems. However, there are problems of instability (particularly for small numbers of Options Price constraints) and ultimately, a degree of Subjectivity can be argued.

Comment 2: Maximising Entropy can also be seen as Minimising the distance of a distribution from the Uniform.

Cross-Entropy:

Cross-Entropy, or *Information Distance*, or *Kullback-Liebler* distance from two distributions P and Q is given by

$$D(P||Q) = E_P(\log(\frac{P}{Q}))$$

$$= \sum p_i \log(\frac{p_i}{q_i})$$

for discrete, or

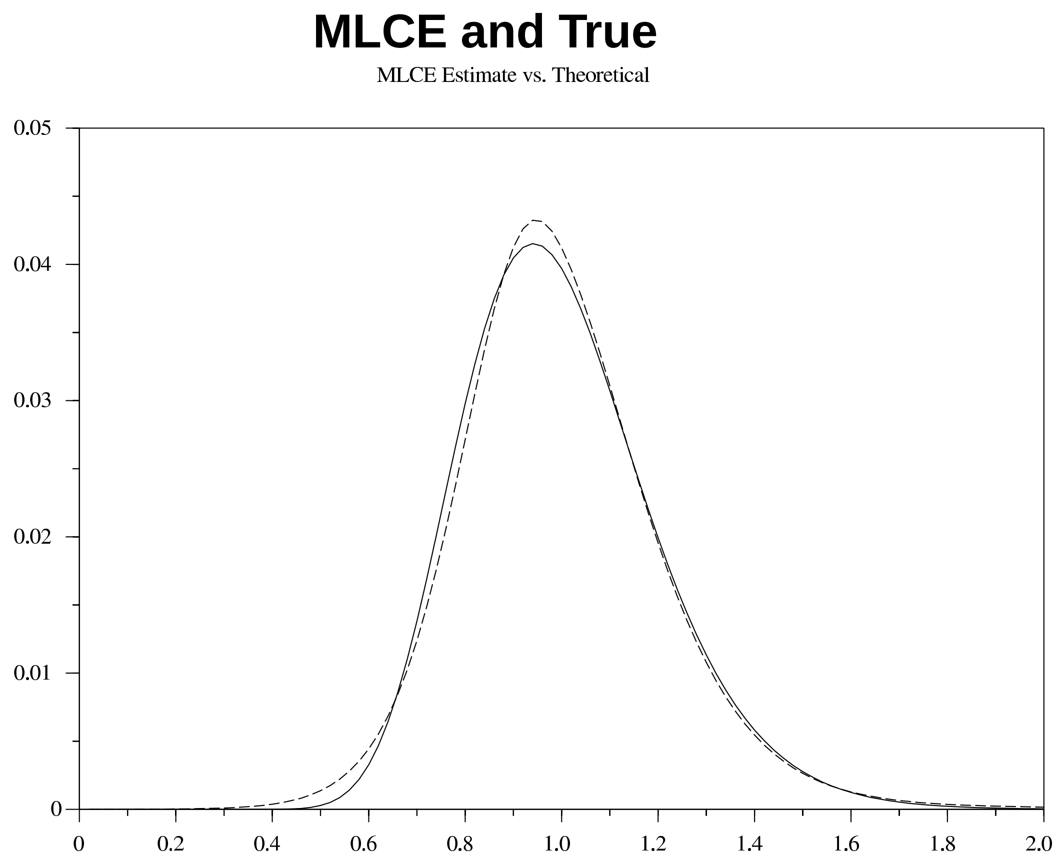
$$\int \log(\frac{p(x)}{q(x)}) p(x) dx$$

continuous. Helpful to think of this as the "compression advantage" of coding (or betting!) w.r.t. P instead of Q when P is true.

Remark: Maximising Entropy may be regarded as Minimising Cross-Entropy to the Uniform.

Alternative Idea (*Risk Magazine*, July '04): Why not Maximise Entropy *Locally*? Analogous to Perturbation Theory. If price distribution is in Equilibrium, then perturbations should have minimal effect on Entropy.
Called the **Local Cross-Entropy** approach.

Example, $r = 0$, $P_0 = 1$, B-S prices for exercises .95,1,1.05 and vol. $\sigma = .2$



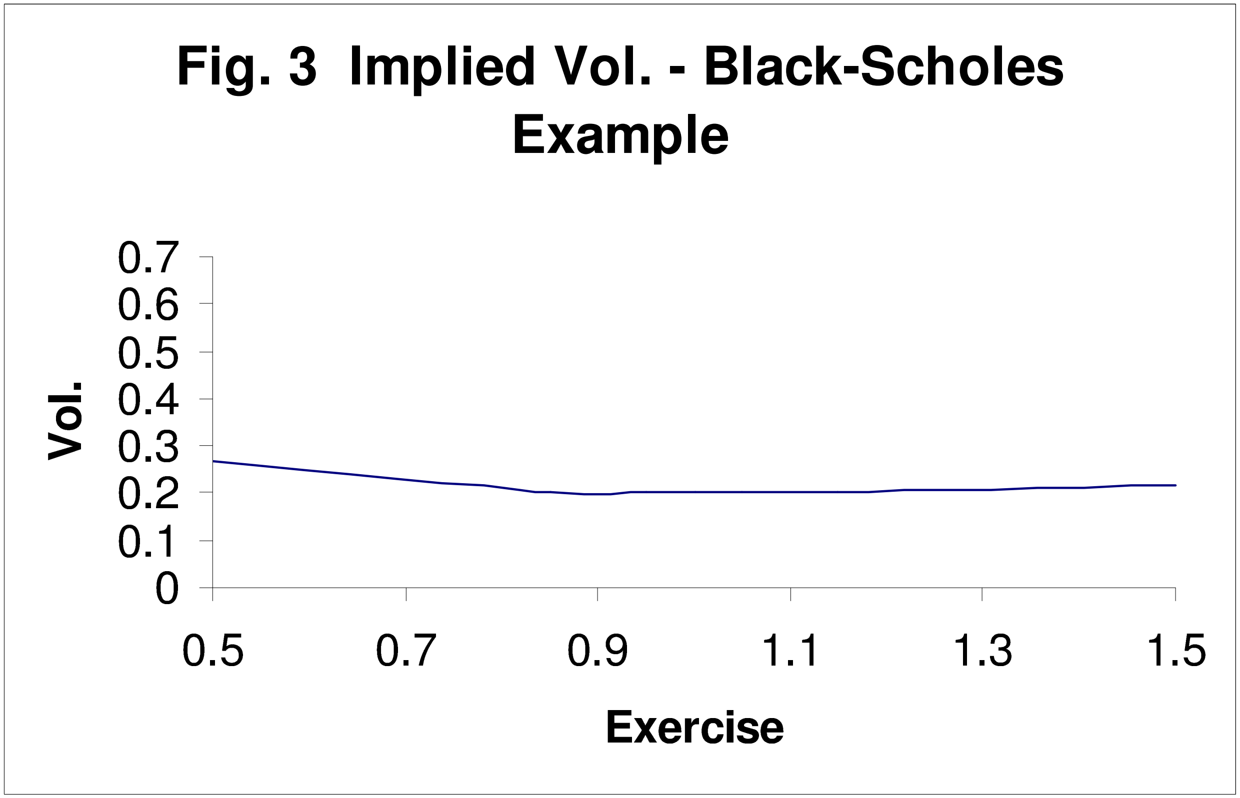


Fig. 3 Implied Vol. - Black-Scholes
Example
Vol.
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0.5
0.7
0.9
1.1
1.3
1.5
Exercise


Fig. 4 Implied Vol. - Skew Example
Vol.
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0.5
0.7
0.9
1.1
1.3
1.5
Exercise

Conclusion - Part 4: Entropy modeling is a useful approach for ’underconstrained’ problems.

4. Information Theory and Cover's Universal Portfolios The opening line (hook?), Cover's '96 UP paper:

"We exhibit an algorithm for portfolio selection that asymptotically outperforms the best stock in the market"

Specifics of Cover's main UP result

Background: Define a 'Constant Rebalanced Portfolio' as one which preserves the proportion of investment in the various assets (i.e., 'shedding' for stocks that rise, 'topping up' for those that fall).

Important Special Case: 100% investment in a single asset.

Gist of Main Mathematical Result: It is possible, through Dynamic Allocation, to outperform any individual stock asymptotically. This may be achieved by approaching the performance of the best Constant Rebalanced Portfolio asymptotically.

Aside: At present, UP results contain certain practical limitations, specifi cally, with the results not being established for the case of no borrowing or short-selling, and the requirement that for each stock, $E(R)/V(R)$ must be greater than a certain unversal value, not satis-fi ed by many stocks in World Markets.

In a forthcoming monograph (Edelman, as a corrollary to Edelman AOR(1999)), these (unnecessary) limitations are lifted.

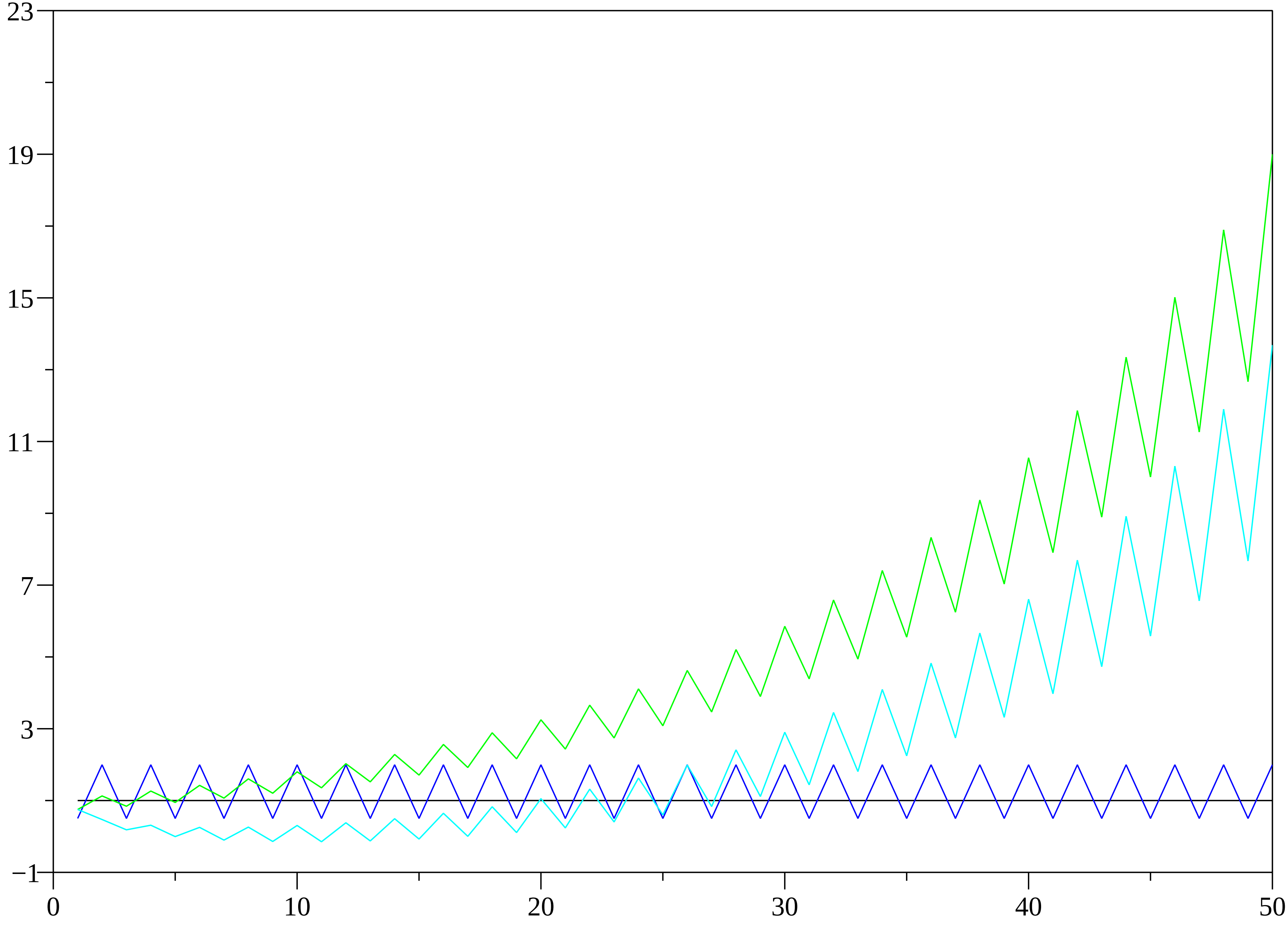
More General Caveat: Thus far, Cover's results as stated are asymptotic only.

Relation to Information Theory: Implicitly based on log-Optimality (as in horse example), intimately related to Information Theory.

Example 2 stocks, one constant, one alternatively halving and doubling in value ’sawtooth’.

First, note that 50-50 ’Constant Rebalanced’ portfolio (CRP) outperforms each individual stock, and that it is possible to ’track’ the best CRP via a simple rule (such as applying what would have been the optimal weights from time zero up until just now, and applying these weights next).

See graphical illustration of performance


23
19
15
11
7
3
−1
0
10
20
30
40
50

The ’Hindsight Bias’ or ’Datasnooping’ Issue

If we develop models for ’beating the Market’ in hindsight, and keep trying until successful, then aren’t we likely to come up with spurious results?

’Bootstrap Snooper’ (H White et al, JoF 1999)

After finding your model for the real data series, simulate ’copies’ of the series, assuming ’random walk’-type prices. Apply the same modeling methdology in each case as was applied on the real data series, taking the best of many attempts to fit a trading model in each case, and see how the performance of this ’spurious best’ compares to what the results had been in your analysis of the ’real’ data series. Then adjust your ’necessarily optimistic’ view accordingly.

Hindsight Bias from the point of view of Universal Portfolios

Imagine a finite set of rules which might have been considered as candidates to be used to ’outperform’ the market.

Imagine, as a thought experiment, application of a Universal Portfolio Algorithm to this group, a dynamic allocation strategy which then must (by Cover’s results) asymptotically outperform the best individual rule in the group.

Therefore: If any one (i.e., the best) of these candidates applied in hindsight appears to outperform the Market in the long-run, the Efficient Markets Hypothesis (based on the appropriate Information Set) would appear to be violated.

Main Conclusions:

Knowledge of structure may be represented by the notion of Data Compression, in a manner which can be quantified precisely in Information-Theoretic units or terms of Financial Gain.

Lack of Knowledge of structure may be incorporated into modeling via Maximum Entropy methods, in order to quantify that which cannot be predicted.

Information-Theoretic approaches to Investment Analysis yield strategies with optimal growth properties

Information-Theoretic approaches borrow from and can enrich many disciplines, with results from each area yielding potential insight into many others.