RATIONAL INATTENTION

Alexandr Shapoval

Financial University under the Government of the Russian Federation
The Institute of Earth Physics of Paris

Nizhnij, 2012

Our world is imperfect

- Dynamic stochastic general equilibrium models (DSGE)
- Sticky price, Calvo, 1983
- Sticky information, Mankiw, Reis, 2002
- Rational inattention, Sims, 2003
- Information constrained state-dependent pricing Woodford, 2009.
- Handbook of monetary economics, 2010

How much information in “YES”?

To be or not to be?

Does she love me?

Equally likely “Yes”/“No”

Choice	Yes	No
Prob	1/2	1/2

Pre-supposed “Yes”

Choice	Yes	No
Prob	0.99	0.01

Now answer “Yes” gives too small information. What about ”No”?

How much information in “YES”?

Does she love me?

Equally likely “Yes”/“No”

Choice	Yes	No
Prob	1/2	1/2

Pre-supposed “Yes”

Choice	Yes	No
Prob	0.99	0.01

Now answer “Yes” gives too small information. What about ”No”?

How much information in “YES”?

Does she love me?

Equally likely “Yes”/“No”

Choice	Yes	No
Prob	1/2	1/2

Pre-supposed “Yes”

Choice	Yes	No
Prob	0.99	0.01

Now answer “Yes” gives too small information. What about ”No”?

Information in “Yes”

“Yes” is very unlikely

Choice	Yes	I would think	...
Prob	$1/n$	$1/n$	$1/n$

Intuitive answer

Information from “Yes” depends on the probability of the alternatives.

Formal approach

Idea

Information about SOMETHING is the differences of chances for SOMETHING after and before the experiment

Formally,

$\mathsf{P}(H_1) \Longrightarrow$ experiment (I ask her, does she love me) $\Longrightarrow \mathsf{P}(H_1|x)$

Bayes rule (the altermative appears)

$$\mathsf{P}\{H_1|x\} = \frac{\mathsf{P}\{x|H_1\}\,\mathsf{P}\{H_1\}}{\mathsf{P}\{x|H_1\}\,\mathsf{P}\{H_1\} + \mathsf{P}\{x|H_2\}\,\mathsf{P}\{H_2\}}$$

Under hypothesis H_i the random variable has distribution $f_i(x)$

$$\log\frac{\mathsf{P}\{H_1|x\}}{\mathsf{P}\{H_2|x\}} = \log\frac{f_1(x)}{f_2(x)} + \log\frac{\mathsf{P}\{H_1\}}{\mathsf{P}\{H_1\}}$$

Formal approach

Idea

Information about SOMETHING is the differences of chances for SOMETHING after and before the experiment

Formally,

$\mathsf{P}(H_1) \Longrightarrow$ experiment (I ask her, does she love me) $\Longrightarrow \mathsf{P}(H_1|x)$

Bayes rule (the altermative appears)

$$\mathsf{P}\{H_1|x\} = \frac{\mathsf{P}\{x|H_1\}\,\mathsf{P}\{H_1\}}{\mathsf{P}\{x|H_1\}\,\mathsf{P}\{H_1\} + \mathsf{P}\{x|H_2\}\,\mathsf{P}\{H_2\}}$$

Under hypothesis H_i the random variable has distribution $f_i(x)$

$$\log\frac{\mathsf{P}\{H_1|x\}}{\mathsf{P}\{H_2|x\}} = \log\frac{f_1(x)}{f_2(x)} + \log\frac{\mathsf{P}\{H_1\}}{\mathsf{P}\{H_1\}}$$

Formal approach

Idea

Information about SOMETHING is the differences of chances for SOMETHING after and before the experiment

Formally,

$\mathsf{P}(H_1) \Longrightarrow$ experiment (I ask her, does she love me) $\Longrightarrow \mathsf{P}(H_1|x)$

Bayes rule (the altermative appears)

$$\mathsf{P}\{H_1|x\} = \frac{\mathsf{P}\{x|H_1\}\,\mathsf{P}\{H_1\}}{\mathsf{P}\{x|H_1\}\,\mathsf{P}\{H_1\} + \mathsf{P}\{x|H_2\}\,\mathsf{P}\{H_2\}}$$

Under hypothesis H_i the random variable has distribution $f_i(x)$

$$\log\frac{\mathsf{P}\{H_1|x\}}{\mathsf{P}\{H_2|x\}} = \log\frac{f_1(x)}{f_2(x)} + \log\frac{\mathsf{P}\{H_1\}}{\mathsf{P}\{H_1\}}$$

Definition

Information for H_1-selection versus H_2 at the point x

$$\log \frac{f_1(x)}{f_2(x)} = \underbrace{\log \frac{\mathsf{P}\{H_1|x\}}{\mathsf{P}\{H_2|x\}}}_{\text{chances for } H_1 \text{ after observation}} - \underbrace{\log \frac{\mathsf{P}\{H_1\}}{\mathsf{P}\{H_1\}}}_{\text{chances for } H_1 \text{ before observation}}$$

Can it be negative?

Definition

Information for H_1-selection versus H_2 at the point x

$$\log \frac{f_1(x)}{f_2(x)} = \underbrace{\log \frac{\mathsf{P}\{H_1|x\}}{\mathsf{P}\{H_2|x\}}}_{\text{chances for } H_1 \text{ after observation}} - \underbrace{\log \frac{\mathsf{P}\{H_1\}}{\mathsf{P}\{H_1\}}}_{\text{chances for } H_1 \text{ before observation}}$$

Can it be negative?

Average info for H_1 vs. H_2

$$I(1:2) = \sum_x \left(\log \frac{f_1(x)}{f_2(x)} \right)$$

Definition

Information for H_1-selection versus H_2 at the point x

$$\log \frac{f_1(x)}{f_2(x)} = \underbrace{\log \frac{\mathsf{P}\{H_1|x\}}{\mathsf{P}\{H_2|x\}}}_{\text{chances for } H_1 \text{ after observation}} - \underbrace{\log \frac{\mathsf{P}\{H_1\}}{\mathsf{P}\{H_1\}}}_{\text{chances for } H_1 \text{ before observation}}$$

Can it be negative?

Average info for H_1 vs. H_2

$$I(1:2) = \sum_x \left(\log \frac{f_1(x)}{f_2(x)} \right) f_1(x)$$

Definition

Information for H_1-selection versus H_2 at the point x

$$\log \frac{f_1(x)}{f_2(x)} = \underbrace{\log \frac{\mathsf{P}\{H_1|x\}}{\mathsf{P}\{H_2|x\}}}_{\text{chances for } H_1 \text{ after observation}} - \underbrace{\log \frac{\mathsf{P}\{H_1\}}{\mathsf{P}\{H_1\}}}_{\text{chances for } H_1 \text{ before observation}}$$

Can it be negative?

Average info for H_1 vs. H_2

$$I(1:2) = \sum_x \left(\log \frac{f_1(x)}{f_2(x)} \right) f_1(x) = \int f_1(x) \log \frac{f_1(x)}{f_2(x)} dx$$

Definition

Information for H_1-selection versus H_2 at the point x

$$\log \frac{f_1(x)}{f_2(x)} = \underbrace{\log \frac{\mathsf{P}\{H_1|x\}}{\mathsf{P}\{H_2|x\}}}_{\text{chances for } H_1 \text{ after observation}} - \underbrace{\log \frac{\mathsf{P}\{H_1\}}{\mathsf{P}\{H_1\}}}_{\text{chances for } H_1 \text{ before observation}}$$

Can it be negative?

Average info for H_1 vs. H_2

$$I(1:2) = \sum_x \left(\log \frac{f_1(x)}{f_2(x)} \right) f_1(x) = \int f_1(x) \log \frac{f_1(x)}{f_2(x)} dx$$

Deviation

$$J(1:2) = I(1:2) + I(2:1)$$

Absolute average information about H_1

Example

$$\mathsf{P}\{H_2\} = 1, \quad H_1 \subset H_2$$

$$I(1:2) = \log\big(\mathsf{P}\{H_1|x\}\big) - \log\big(\mathsf{P}\{H_1\}\big)$$

Let $\mathsf{P}\{H_1|x\} = 1$. Then $I(1:2) = -\log \mathsf{P}\{H_1\}$. The information is large, if unconditional probability of H_1 is small.

Entropy

n hypotheses: H_1, H_2, ..., H_n
for each hypothesis: $(H_i, \bar{H}_i)$, $I(1:2) = -\log \mathsf{P}\{H_i\}$
Average information from an observation:

$$\mathcal{H} = -\sum \mathsf{P}\{H_i\} \log \mathsf{P}\{H_i\}$$

Absolute average information about H_1

Example

$$\mathsf{P}\{H_2\} = 1, \quad H_1 \subset H_2$$

$$I(1:2) = \log\big(\mathsf{P}\{H_1|x\}\big) - \log\big(\mathsf{P}\{H_1\}\big)$$

Let $\mathsf{P}\{H_1|x\} = 1$. Then $I(1:2) = -\log \mathsf{P}\{H_1\}$. The information is large, if unconditional probability of H_1 is small.

Entropy

n hypotheses: H_1, H_2, ..., H_n
for each hypothesis: $(H_i, \bar{H}_i)$, $I(1:2) = -\log \mathsf{P}\{H_i\}$
Average information from an observation:

$$\mathcal{H} = -\sum \mathsf{P}\{H_i\} \log \mathsf{P}\{H_i\}$$

Computation of the entropy

Equally likely Yes/No

Yes	No
1/2	1/2

$$\mathcal{H} = -\frac{1}{2}\log\frac{1}{2} - \frac{1}{2} = \log\frac{1}{2}\log 2$$

Units: $\mathcal{H} = \log_2 2 = 1$ bit or $\quad \mathcal{H} = \ln 2$ bits $= 1$ nat.

Examples

Yes	No
0.99	0.01

$$\mathcal{H} = -0.99\log 0.99 - 0.01\log 0.01 = 0.08 \text{ bits}$$

$I(1:2, \text{Yes}) = -\mathrm{P}\{\text{Yes}\}\log(\mathrm{P}\{\text{Yes}\}) = 0.0099.$

Computation of the entropy

Equally likely Yes/No

Yes	No
1/2	1/2

$$\mathcal{H} = -\frac{1}{2}\log\frac{1}{2} - \frac{1}{2} = \log\frac{1}{2}\log 2$$

Units: $\mathcal{H} = \log_2 2 = 1$ bit or $\mathcal{H} = \ln 2$ bits $= 1$ nat.

Examples

Yes	No
0.99	0.01

$$\mathcal{H} = -0.99\log 0.99 - 0.01\log 0.01 = 0.08 \text{ bits}$$

$I(1:2,\text{Yes}) = -\mathsf{P}\{\text{Yes}\}\log(\mathsf{P}\{\text{Yes}\}) = 0.0099.$

Entropy vs Uncertainty

Equally likely two values

x_1	x_2
1/2	1/2

$\mathcal{H} = 1$ Shown before

Equally likely four values

x_1	x_2	x_3	x_4
1/4	1/4	1/4	1/4

$\mathcal{H} =$ Who knows?

- $\mathcal{H}(\text{const}) = 0$
- Increase of choices $\Longrightarrow$ Increase of entropy
- Entropy increases within uncertainty

Entropy vs Uncertainty

Equally likely two values

x_1	x_2
$1/2$	$1/2$

$\mathcal{H} = 1$ Shown before

Equally likely four values

x_1	x_2	x_3	x_4
$1/4$	$1/4$	$1/4$	$1/4$

$\mathcal{H} = 1$

- $\mathcal{H}(\text{const}) = 0$
- Increase of choices $\Longrightarrow$ Increase of entropy
- Entropy increases within uncertainty

Entropy vs Uncertainty

Equally likely two values

x_1	x_2
$1/2$	$1/2$

$\mathcal{H} = 1$ Shown before

Equally likely four values

x_1	x_2	x_3	x_4
$1/4$	$1/4$	$1/4$	$1/4$

$$\mathcal{H} = -4 \times \frac{1}{4} \log \frac{1}{4} = 2$$

- $\mathcal{H}(\text{const}) = 0$
- Increase of choices $\Longrightarrow$ Increase of entropy
- Entropy increases within uncertainty

Entropy vs Uncertainty

Equally likely two values

x_1	x_2
1/2	1/2

$\mathcal{H} = 1$ Shown before

Equally likely four values

x_1	x_2	x_3	x_4
1/4	1/4	1/4	1/4

$\mathcal{H} =$

- $\mathcal{H}(\text{const}) = 0$
- Increase of choices $\Longrightarrow$ Increase of entropy
- Entropy increases within uncertainty

Mutual information

Information about X in Y

$$X \xrightarrow{\text{noise}, N} Y = X + N.$$

$$H_2 \text{ is always true}, \quad \mathsf{P}\{H_2\} = 1$$

H_1: the joint distribution (X, Y) is given by the probability density $f(x, y)$; $X = x$.
Information when observe $Y = y$:

$$I(1:2, Y = y) = \log \mathsf{P}\{H_1 | Y = y\} - \log \mathsf{P}\{H_1\}$$

Compare with (discussed before)

$$I(1:2) = \log \big(\mathsf{P}\{H_1 | x\} \big) - \log \big(\mathsf{P}\{H_1\} \big)$$

Mutual information

Average information in Y about X

$$\langle I(1:2, Y=y)\rangle_{x,y} = \langle \log \mathsf{P}\{X=x|Y=y\}\rangle_{x,y} - \langle \log \mathsf{P}\{X=x\}\rangle_x$$

$$\langle \log \mathsf{P}\{X=x, Y=y\}\rangle_{x,y} - \langle \log \mathsf{P}\{Y=y\}\rangle_y - \langle \log \mathsf{P}\{X=x\}\rangle_x$$

$$= \int\int f(x,y)\log f(x,y)dxdy - \int h(y)\log h(y)dy - \int g(x)\log g(x)dx =$$

$$-\mathcal{H}(X,Y) + \mathcal{H}(X) + \mathcal{H}(y)$$

Capacity of the chanel

A noisy telegraph line, the dot or dash input reproduces itself in the output with probability of p:

$$Y = X + N, \quad \mathsf{P}\{Y = X\} = p, \quad \mathsf{P}\{X = 0\} = \alpha$$

Mutual information $I_\alpha(X, Y)$ depends on α
Capacity is $\max_\alpha I_\alpha(X, Y)$
Simple algebra gives evidence that I_α attains its maximum at $\alpha = 1/2$, the dashs and dots are equally likely. In this case

$$\mathsf{P}\{X = 0, Y = 0\} = \mathsf{P}\{X = 1, Y = 1\} = p/2,$$
$$\mathsf{P}\{X = 0, Y = 1\} = \mathsf{P}\{X = 1, Y = 0\} = (1 - p)/2 = (1 - p)/2$$

$$C = I = -\mathcal{H}(X, Y) + \mathcal{H}(X) + \mathcal{H}(Y) = p \log p + (1 - p) \log(1 - p) - 4 \frac{1}{2} \log \frac{1}{2}$$

Capacity of the chanel

A noisy telegraph line, the dot or dash input reproduces itself in the output with probability of p:

$$Y = X + N, \quad \mathsf{P}\{Y = X\} = p, \quad \mathsf{P}\{X = 0\} = \alpha$$

Mutual information $I_\alpha(X, Y)$ depends on α
Capacity is $\max_\alpha I_\alpha(X, Y)$
Simple algebra gives evidence that I_α attains its maximum at $\alpha = 1/2$, the dashs and dots are equally likely. In this case

$$\mathsf{P}\{X = 0, Y = 0\} = \mathsf{P}\{X = 1, Y = 1\} = p/2,$$
$$\mathsf{P}\{X = 0, Y = 1\} = \mathsf{P}\{X = 1, Y = 0\} = (1 - p)/2 = (1 - p)/2$$

$$C = I = -\mathcal{H}(X, Y) + \mathcal{H}(X) + \mathcal{H}(Y) = p \log p + (1 - p) \log(1 - p) + 2$$

Capacity of the chanel

A noisy telegraph line, the dot or dash input reproduces itself in the output with probability of p:

$$Y = X + N, \quad \mathsf{P}\{Y = X\} = p, \quad \mathsf{P}\{X = 0\} = \alpha$$

Mutual information $I_\alpha(X, Y)$ depends on α
Capacity is $\max_\alpha I_\alpha(X, Y)$
Simple algebra gives evidence that I_α attains its maximum at $\alpha = 1/2$, the dashs and dots are equally likely. In this case

$$\mathsf{P}\{X = 0, Y = 0\} = \mathsf{P}\{X = 1, Y = 1\} = p/2,$$
$$\mathsf{P}\{X = 0, Y = 1\} = \mathsf{P}\{X = 1, Y = 0\} = (1 - p)/2 = (1 - p)/2$$

If N is normal noise that capacity is attained for normal X

A noisy telegraph line, the dot or dash input reproduces itself in the output with probability of p:

$$Y = X + N, \quad \mathsf{P}\{Y = X\} = p, \quad \mathsf{P}\{X = 0\} = \alpha$$

Mutual information $I_\alpha(X, Y)$ depends on α
Capacity is $\max_\alpha I_\alpha(X, Y)$
Simple algebra gives evidence that I_α attains its maximum at $\alpha = 1/2$, the dashs and dots are equally likely. In this case

$$\mathsf{P}\{X = 0, Y = 0\} = \mathsf{P}\{X = 1, Y = 1\} = p/2,$$
$$\mathsf{P}\{X = 0, Y = 1\} = \mathsf{P}\{X = 1, Y = 0\} = (1 - p)/2 = (1 - p)/2$$

If N is normal noise that capacity is attained for normal X
Can we get full information about X, if N is continious (say, normal) random variable?

Coding

Input $\overset{\text{coding}}{\Longrightarrow}$ Equally likely 0s and 1s $\overset{\text{chanel}}{\Longrightarrow}$ Output
The graph is scanned into a 100×100 grid of pixels
The most of symbols are zeros, their fraction is 0.98.
If we send symbols one by one, we have to send 1 bit information for each symbol, totally 10 000 bits.
Coding: 0 represents the sequence 000, $1001 \Longrightarrow 001$
Coding: 1010 represents the sequence 010, $1011 \Longrightarrow 011$
Coding: 1100 represents the sequence 100, $1101 \Longrightarrow 101$
Coding: 1110 represents the sequence 110, $1111 \Longrightarrow 111$
Then approximately $0.98^3 = 0.94$ of three-pixel blocks are represented by a single 0, 0.06 of them are by four-element sequence. The average number of bits to transfer is

$$(0.94 \times 1 + 0.06 \times 4) \times 10\,000/3 = 3934 < 10\,000$$

Entropy of the normal distribution

$$\varphi(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}.\right)$$

$$\mathcal{H}(x) = -\int_{-\infty}^{+\infty} \left(\log\left(\frac{1}{\sqrt{2\pi}\sigma}\right) - \frac{(x-\mu)^2}{2\sigma^2}\right)\varphi(x)dx =$$

$$\log\left(\sqrt{2\pi}\right) + \log\sigma + \frac{1}{\sigma^2}\underbrace{\int_{-\infty}^{+\infty}\frac{(x-\mu)^2}{2}\varphi(x)dx}_{\text{variance}} = \frac{1}{2}\log\left(2\pi\right) + \log\sigma + 1/2.$$

2-dimensional vector

$$\mathcal{H}(X,Y) = \log(2\pi) + \log\sigma_x + \log\sigma_y + \frac{1}{2}\log(1-\rho^2) + 1$$

Entropy of the normal distribution

$$\varphi(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}.\right)$$

$$\mathcal{H}(x) = -\int_{-\infty}^{+\infty} \left(\log\left(\frac{1}{\sqrt{2\pi}\sigma}\right) - \frac{(x-\mu)^2}{2\sigma^2}\right)\varphi(x)dx =$$

$$\log\left(\sqrt{2\pi}\right) + \log\sigma + \frac{1}{\sigma^2}\underbrace{\int_{-\infty}^{+\infty}\frac{(x-\mu)^2}{2}\varphi(x)dx}_{\text{variance}} = \frac{1}{2}\log\left(2\pi\right) + \log\sigma + 1/2.$$

2-dimensional vector

$$\mathcal{H}(X,Y) = \log(2\pi) + \log\sigma_x + \log\sigma_y + \frac{1}{2}\log(1-\rho^2) + 1$$

Information, observing normal RV

$$I(X,Y) = -\frac{1}{2}\big(\log(1-\rho^2)\big)$$

Model

$$Y = X + N, \quad \mathbf{Cov}(X,N) = 0, (X,Y) \sim \text{ normal}, \sigma_x, \sigma_y, \rho$$

X is input, Y is output, N is noise; $\rho = \sigma_x/\sigma_y$, $\mathbf{Var}\, N = \nu^2$

$$I(X,Y) = -\frac{1}{2}\log\left(1 - \frac{\sigma_x^2}{\sigma_y^2}\right).$$

The amount of information (in bits) obtained from observation X

$$I(X,Y) = \frac{1}{2}\log\left(1 + \frac{\sigma_x^2}{\sigma_y^2 - \sigma_x^2}\right) = \frac{1}{2}\log\left(1 + \frac{\sigma_x^2}{\nu^2}\right).$$

Decrease of uncertainty

Observe $X_i + N_i$, $\mathbf{E} N_i = 0$, m times and choose the mean of the observations as estimation of X,

$$\mathbf{Var}\left(\frac{X_1 + N_1 + X_2 + N_2 + \ldots + X_m + N_m}{m}\right) =$$

$$\frac{1}{m^2}\mathbf{Var}(X_1 + N_1 + \ldots + X_m + N_m) = \frac{\sigma_x^2 + \sigma_N^2}{m}$$

At cost κ for each bit of information

$$\kappa m \frac{1}{2} \log\left(1 + \frac{\sigma_x^2}{\sigma_N^2}\right)$$

Economic behaviour

A decision maker is uncertain about some economic variable x. Observing $y = x + N$ as many times as she wishes she obtaines information I about x at a cost and chooses $\tilde{x}$ based on this information:

$$U(\tilde{x}, x) \to \max$$
$$I \leqslant I^* \quad \text{information constraint}$$

Optimization problem

Autoregression process

$$y_t = a y_{t-1} + \varepsilon_t, \quad (y_t, y_{t-1}) \sim N(\sigma_t^2, \sigma_{t-1}^2, \rho), \mathbf{Var}(\varepsilon_t) = \nu^2$$

$$\mathbf{Cov}(y_t, y_{t-1}) = a\sigma_{t-1}^2, \quad \rho = \frac{a\sigma_{t-1}}{\sigma_t}$$

$$I(y_t, y_{t-1}) = -\frac{1}{2}\log\left(1-\rho^2\right) = \frac{1}{2}\log\frac{a^2\sigma_{t-1}^2 + \nu^2}{\nu^2}$$

Minimization of losses, given an observation of y_t

$$\sum \beta^t \left(\mathbf{E}(y_t - x_t)^2 + \frac{\kappa}{2}\log\frac{a^2\sigma_{t-1}^2 + \nu^2}{\nu^2}\right) \xrightarrow{x_t} \min$$

$$\sum \beta^t \left(\sigma_t^2 + \frac{\kappa}{2}\log\frac{a^2\sigma_{t-1}^2 + \nu^2}{\nu^2}\right) \xrightarrow{x_t} \min$$

Minimization of losses, given m_t observations of y_t

$$\sum \beta^t \left(\sigma_t^2 \qquad\qquad\qquad + \kappa m_t \log \frac{a^2 \sigma_{t-1}^2 + \nu^2}{\nu^2} \right) \xrightarrow{m_t} \min$$

Given yellow constants

$$m_t = \sqrt{C/\kappa}$$

Minimization of losses, given m_t observations of y_t

$$\sum \beta^t \left(\frac{a^2\sigma_{t-1}^2 + \nu^2}{m_t} + \kappa m_t \log \frac{a^2\sigma_{t-1}^2 + \nu^2}{\nu^2} \right) \xrightarrow{m_t} \min$$

Given yellow constants

$$m_t = \sqrt{C/\kappa}$$

Minimization of losses, given m_t observations of y_t

$$\sum \beta^t \left(\frac{a^2\sigma_{t-1}^2 + \nu^2}{m_t} + \kappa m_t \log \frac{a^2\sigma_{t-1}^2 + \nu^2}{\nu^2} \right) \xrightarrow{m_t} \min$$

School algebra

$$\frac{A}{z} + Bz \xrightarrow{z} \min$$

Solution:

$$\frac{A}{z^*} = Bz^*, \quad z^* = \sqrt{A/B}$$

Given yellow constants

$$m_t = \sqrt{C/\kappa}$$

Minimization of losses, given m_t observations of y_t

$$\sum \beta^t \left(\frac{a^2\sigma_{t-1}^2 + \nu^2}{m_t} + \kappa m_t \log \frac{a^2\sigma_{t-1}^2 + \nu^2}{\nu^2} \right) \xrightarrow{m_t} \min$$

Given yellow constants

$$m_t = \sqrt{C/\kappa}$$

Qualitative behaviour of the solution

$$m_t = \sqrt{C/\kappa}$$

- $\kappa \to 0$, then $m_t \to \infty$, information is cheap and therefore in use
- $\kappa \to \infty$, then $m_t \to 0$; if integer $m_t < 1$, the information is not processed.

Let σ_{t-1} and ν be such that $m_t^* < 1$ and $\sigma_t^2 = a^2\sigma_{t-1}^2 + \nu^2 > \sigma_{t-1}^2$. Then the information is not processed, $x_t = ax_{t-1}$ and the variance increases. The first term $(x_t - y_t)^2$ in the objective is increasing in t, optimal (unconditional) solution m_t^* is increasing in σ_t. The choice x_t deviates from y_t more and more with t so that (since m_t^* increases) the optimal stratege at some moment t' is to collect information.

Dixit-Stiglitz consumption I

$$U = \left(\int_0^1 q^{1-\nu}(i)di \right)^{1/(1-\nu)} \longrightarrow \max$$

$$\int_0^1 p(i)q(i) = I$$

Solution

$$\mathcal{L} = \left(\int_0^1 q^{1-\nu}(i)di \right)^{1/(1-\nu)} - \lambda \int_0^1 p(i)q(i) = I$$

FOC:

$$q(i) = \tilde{\lambda}^{-1/\nu} p^{-1/\nu}(i), \quad \tilde{\lambda} = \lambda U^{-\nu/(1-\nu)}$$

Dixit-Stiglitz consumption II

Price index

Substitution of FOC into budget constraint:

$$\tilde{\lambda}^{-1/\nu} P^{-(1-\nu)/\nu} = I, \quad P = \left(\int_0^1 p^{-(1-\nu)/\nu}(i) di \right)^{-\nu/(1-\nu)}$$

The Lagrange multiplier

$$\tilde{\lambda} = I^{-\nu} P^{-(1-\nu)}$$

Optimal consumption

$$q(i) = I P^{(1-\nu)/\nu} p^{-1/\nu}(i)$$

Firms

Choose x_t for unknown q_t to get profit

$$\pi_t = \begin{cases} p_t x_t - c q_t, & \text{if } x_t > q_t; \\ p_t x_t - c x_t, & \text{if } x_t < q_t \end{cases}$$

Expected profit for optimal $q_t = p_{t-1}^{-1/\nu}$ at period t

$$\mathbf{E}\pi_t = (p_t x_t - c q_t)\,\mathsf{P}\{x_t > q_t\} + (p_t x_t - c x_t)\,\mathsf{P}\{x_t < q_t\} =$$

$$(p_t x_t - c p_{t-1}^{1/\nu}) \left(1 - \Phi\left(\frac{\log x_t - p_{t-1}^{-1/\nu}}{\sigma_t \sqrt{m_t}}\right)\right) + (p_t - c) x_t \Phi\left(\frac{\log x_t - p_{t-1}^{-1/\nu}}{\sigma_t \sqrt{m_t}}\right)$$

$$- \text{info\_cost}$$

Φ is the normal distribution function

Maximization of profit

$$\sum_{t=1}^{\infty}\left(\left(p_t x_t - c p_{t-1}^{1/\nu}\right)\left(1-\Phi\left(\frac{\log x_t - p_{t-1}^{-1/\nu}}{\sigma_t\sqrt{m_t}}\right)\right)+\right.$$

$$\left.(p_t - c)x_t\Phi\left(\frac{\log x_t - p_{t-1}^{-1/\nu}}{\sigma_t\sqrt{m_t}}\right)\right) - \kappa m_t \frac{1}{2}\log\frac{\sigma_t^2}{\nu^2} \xrightarrow{x_t,p_t,m_t} \max$$

subject to

$$\begin{cases} \sigma_t^2 = (\sigma_{t-1}^2 + \nu^2)/m_t, & \text{if } m \geqslant 1; \\ \sigma_t^2 = \sigma_{t-1}^2 + \nu^2, & \text{if } m = 0 \end{cases}$$

Solution

1. $m_t = 0$: price is not adjusted on $[t_1, t_2]$
2. $m_t > 1$: price is adjusted on $[t_2, t_3]$, and so on

Woodford, 2009

Model

- Cost of price adjustment
- Information cost
- Objective: given the distribution of uncertain price, set (probabilistic) strategy of price adjustment

Results

- Stationary distribution of prices (under optimal price review)
- The cost of the information in terms of the firms revenue per time unit.
- The average rate of price review

THANK YOU!

$$x_t = a x_{t-1}\theta + (1-\theta) y_t + \xi_t$$

random variable ξ_t does not depend on x_t, x_{t-1}, y_t.

$$\mathbf{Var}(a x_{t-1}\theta + (1-\theta) y_t + \xi_t) \xrightarrow{\theta} \min$$

$$\sigma_t^2\theta^2 + \sigma_{t-1}^2(1-\theta)^2 + 2\rho\theta(1-\theta)\sigma_t\sigma_{t-1} \xrightarrow{\theta} \min$$

$$\theta^* = \frac{\sigma_t^2 - \rho\sigma_{t-1}\sigma_t}{\sigma_{t-1}^2 + \sigma_t^2 - 2\rho\sigma_{t-1}\sigma_t}$$