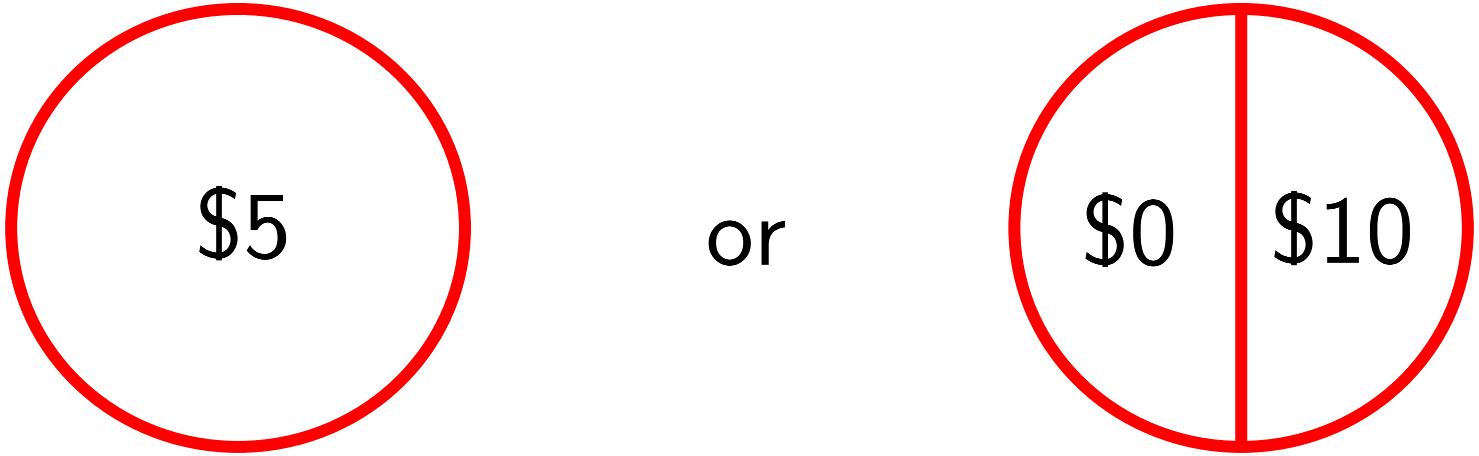
# Estimation of Risk Aversion Parameters: Binary Lottery Choice
# 估計風險偏好: 二選一風險決策

Joseph Tao-yi Wang (王道一)
EE-BGT, Experimetrics Module 3a

# Thaler and Johnson (1990); Keasey and Moon (1996)

- You received an endowment, and now have a choice:

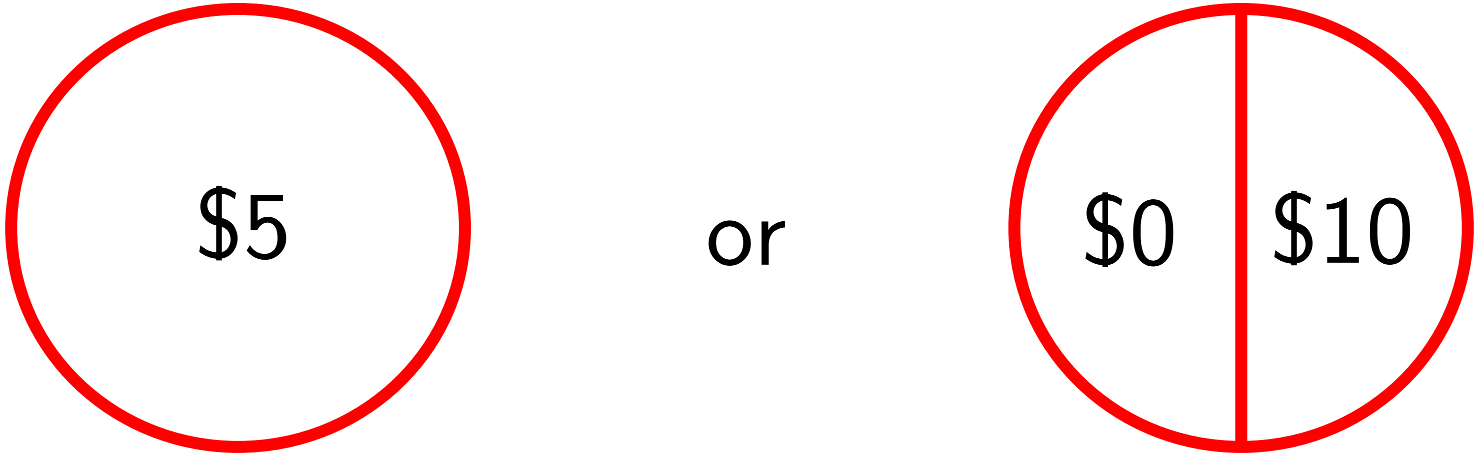



Safe Choice (S): $y=1$ Risky Choice (R): $y=0$

- Which would you choose if your endowment is $10?
- Which would you choose if your endowment is $1,000?

# Apply Binary Data Models to Risky Choice Experiments

- Ask subjects (with different endowment) to choose:




Safe Choice (S): $y=1$ Risky Choice (R): $y=0$

- To test House Money Effect:
  - Choices more risk-seeking when initial endowment is high
  - Or: Take more risk when reinvesting prior profit

Thaler and Johnson (1990)
Keasey and Moon (1996)

# Apply Binary Data Models to Risky C

- 1,050 subjects with wealth $w_i$
- Binary Outcome: Choose
  - Safe ($y = 1$), or Risky ($y = 0$)
- Simulated experiment data
  - `House_money_sim.dta`
  - STATA: `table w, contents(n y mean y)`
- 92% choose safe at $w_i = \$0$
- 50% choose safe at $w_i = \$10$

| w | N(y) | mean(y) |
|---|---|---|
| 0 | 50 | .92 |
| .5 | 50 | .88 |
| 1 | 50 | .88 |
| 1.5 | 50 | .84 |
| 2 | 50 | .84 |
| 2.5 | 50 | .9 |
| 3 | 50 | .84 |
| 3.5 | 50 | .72 |
| 4 | 50 | .78 |
| 4.5 | 50 | .7 |
| 5 | 50 | .7 |
| 5.5 | 50 | .74 |
| 6 | 50 | .72 |
| 6.5 | 50 | .72 |
| 7 | 50 | .5 |
| 7.5 | 50 | .64 |
| 8 | 50 | .5 |
| 8.5 | 50 | .48 |
| 9 | 50 | .56 |
| 9.5 | 50 | .5 |
| 10 | 50 | .5 |

# Probit Model for Choosing Safe Under Wealth Level $w_i$

- Model this as Probit: $\Pr(y_i = 1|w_i) = \Phi(\beta_0 + \beta_1 w_i)$

  where $\Phi(z) = \Pr(Z < z) = \int_{-\infty}^{z} \phi(z)dz$ is standard Normal cdf

  And its pdf is $\phi(z) = \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-z^2}{2}\right)$

- Likelihood:

$$L = \prod_{i=1}^{n} [\Phi(\beta_0 + \beta_1 w_i)]^{y_i} [1 - \Phi(\beta_0 + \beta_1 w_i)]^{1-y_i}$$

# Probit Model for Choosing Safe Under Wealth Level $w_i$

- Log-Likelihood: (Easier for numerical maximization!)

$$\log L = \sum_{i=1}^{n} y_i \ln\left[\Phi(\beta_0 + \beta_1 w_i)\right] + (1 - y_i) \ln\left[1 - \Phi(\beta_0 + \beta_1 w_i)\right]$$

- Since $\Phi(-z) = 1 - \Phi(z)$,
  - Rewrite log-Likelihood with Safe ($yy_i$ = 1) & Risky ($yy_i$ = -1)

$$\log L = \sum_{i=1}^{n} \ln\left[\Phi(yy_i \times (\beta_0 + \beta_1 w_i))\right]$$

  - `probit y w` in STATA to perform MLE to find $\beta_0, \beta_1$

# Probit Model for Choosing Safe Under Wealth Level $w_i$

- STATA `probit y w`
  Results:

```
Iteration 0: log likelihood = -634.4833
Iteration 1: log likelihood = -584.91375
Iteration 2: log likelihood = -584.5851
Iteration 3: log likelihood = -584.58503
Iteration 4: log likelihood = -584.58503
Probit regression                          Number of obs   =      1050
                                           LR chi2(1)      =     99.80
                                           Prob > chi2     =    0.0000
Log likelihood = -584.58503                Pseudo R2       =    0.0786
------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           w |  -.1409882   .0145377    -9.70   0.000    -.1694816   -.1124948
       _cons |   1.301654   .0911155    14.29   0.000     1.123071    1.480237
------------------------------------------------------------------------------
```

1,050 Subjects of 1 Round each

Strong House Money Effect ($z =$ -9.70)

# Wald Test for the House Money Effect (Probit)

- Wald Test for $\beta_1 = 0$ : $W = \dfrac{(\hat{\beta}_1 - 0)^2}{\mathrm{Var}(\hat{\beta}_1)} \sim \chi^2(1)$

  - STATA `test w=0`

    Results:

```
( 1)  [y]w = 0

           chi2(  1) =    94.05
         Prob > chi2 =    0.0000
```

Strong House Money Effect:
$W = (\text{-}9.70)^2 = 94.05$
$>> 3.84 = \chi^2_{1,0.05}$

# Prediction: Pr(Safe) for Each Wealth Level

- Graph estimated probability:

$$\Phi(\beta_0 + \beta_1 w) = \Phi(1.302 - 0.141w)$$

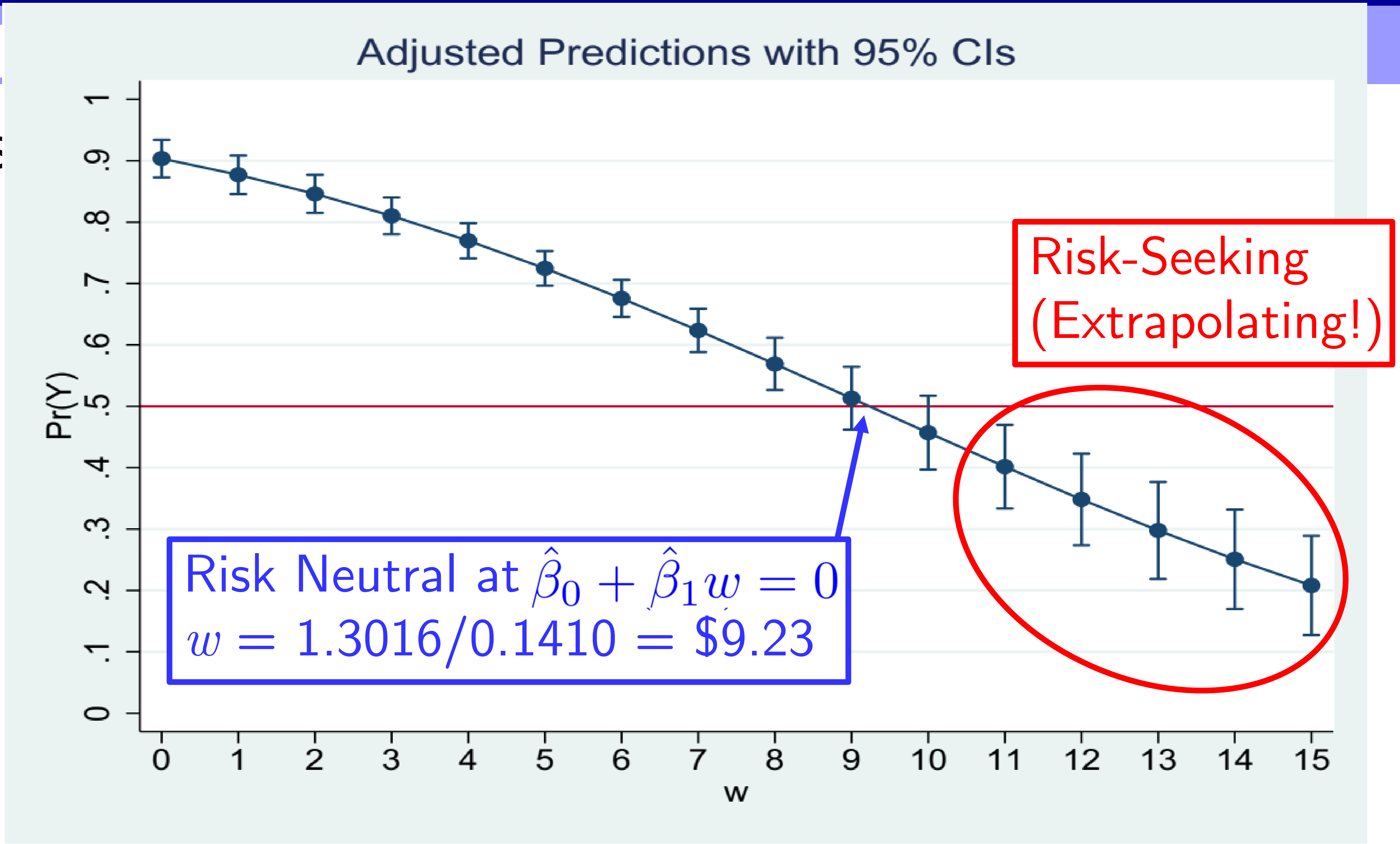
- STATA Command:

```
margins, at(w=(0(1)15)
marginsplot, ylabel(0(0.1)1) yline(0.5)
```

  - STATA Results:

# Prediction: I

- Graph es




- STATA
  - STATA
    Results:

# Conditional Marginal Effect at No Wealth ($w = \$0$)

- Predict change in Pr(Safe) due to change in $w$ ($w = \$0$)

```
margins, dydx(w) at(w=0)
```

  - STATA Results:

```
Conditional marginal effects                    Number of obs     =       1050
Model VCE    : OIM

Expression   : Pr(y), predict()
dy/dx w.r.t. : w
at           : w               =           0

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           w |   -.024109   .0013299   -18.13   0.000    -.0267155   -.0215026
------------------------------------------------------------------------------
```

Pr(Safe) drops by 2.41% when $w$ rises from \$0 to \$1

# Conditional Marginal Effect at Higher Wealth ($w = \$10$)

- Predict change in Pr(Safe) due to change in $w$ ($w = \$10$)

```
margins, dydx(w) at(w=10)
```

- STATA Results:

```
Conditional marginal effects                    Number of obs     =      1050
Model VCE   : OIM

Expression  : Pr(y), predict()
dy/dx w.r.t. : w
at          : w                    =          10

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           w |  -.0559177   .0053804   -10.39   0.000    -.0664631   -.0453724
------------------------------------------------------------------------------
```

Pr(Safe) drops by 5.59% as $w$ rises from $10 to $11 (Steeper slope in Figure)

# Average Marginal Effect

- Average predict change in Pr(Safe) due to change in $w$

`margins, dydx(w) ~~at(w=10)~~`

- STATA Results:

```
Average marginal effects                          Number of obs     =      1050
Model VCE     : OIM

Expression   : Pr(y), predict()
dy/dx w.r.t. : w

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           w |  -.0444259    .0039929   -11.13   0.000    -.0522518      -.0366
------------------------------------------------------------------------------
```

Pr(Safe) drops by 4.44% on average as $w$ rises by 1 across all observations

# Likelihood Ratio Test for House Money Effect (Probit)

- Wald Test for $\beta_1 = 0$ : $W = \frac{\left(\hat{\beta}_1 - 0\right)^2}{\text{Var}\left(\hat{\beta}_1\right)} \sim \chi^2(1)$

- Likelihood Ratio (LR) Test between:
  - Unrestricted: $\log L_U = \sum_{i=1}^{n} \ln\left[\Phi(yy_i \times (\beta_0 + \beta_1 w_i))\right]$
  - Restricted: $\log L_R = \sum_{i=1}^{n} \ln\left[\Phi(yy_i \times (\beta_0))\right]$

- LR for $\beta_1 = 0$ : $LR = 2\left(\log L_U - \log L_R\right) \sim \chi^2(1)$

# Unrestricted Probit Model

$$\log L_U = \sum_{i=1}^{n} \ln \left[ \Phi(yy_i \times (\beta_0 + \beta_1 w_i)) \right]$$

- STATA `probit y w`

Results:

1,050 Subjects of 1 Round each

```
Iteration 0: log likelihood = -634.4833
Iteration 1: log likelihood = -584.91375
Iteration 2: log likelihood = -584.5851
Iteration 3: log likelihood = -584.58503
Iteration 4: log likelihood = -584.58503
Probit regression                                 Number of obs   =       1050
                                                  LR chi2(1)      =      99.80
                                                  Prob > chi2     =     0.0000
Log likelihood = -584.58503                       Pseudo R2       =     0.0786
------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           w |  -.1409882   .0145377    -9.70   0.000    -.1694816   -.1124948
       _cons |   1.301654   .0911155    14.29   0.000     1.123071    1.480237
------------------------------------------------------------------------------
```

$$\log L_U = -584.59$$

# Restricted Probit Model

- Restricted: $\log L_R = \sum_{i=1}^{n} \ln\left[\Phi(yy_i \times (\beta_0))\right]$
- STATA Results: `probit y`

1,050 Subjects of 1 Round each

```
Iteration 0: log likelihood = -634.4833
Iteration 1: log likelihood = -634.4833
Probit regression                              Number of obs   =      1050
                                               LR chi2(0)      =      0.00
                                               Prob > chi2     =         .
Log likelihood = -634.4833                     Pseudo R2       =    0.0000
------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |   .5464424   .0408516    13.38   0.000     .4663746    .6265101
------------------------------------------------------------------------------
```

$\log L_R = -634.48$

# Likelihood Ratio Test for House Money Effect (Probit)

- STATA `probit y w`

  Results:

```
Iteration 0:  log likelihood = -634.4833
Iteration 1:  log likelihood = -584.91375
Iteration 2:  log likelihood = -584.5851
Iteration 3:  log likelihood = -584.58503
Iteration 4:  log likelihood = -584.58503
Probit regression                          Number of obs   =       1050
                                           LR chi2(1)      =      99.80
                                           Prob > chi2     =     0.0000
Log likelihood = -584.58503                Pseudo R2       =     0.0786
--------------------------------------------------------------------------
           |  Co                                          . Interval]
-----------+--------------------------------------------------------------
         w |  -.1409                                         -.1124948
     _cons |   1.301                                          1.480237
--------------------------------------------------------------------------
```

1,050 Subjects of 1 Round each

$\log L_U = -584.59$

$\log L_R = -634.48$

$$LR = 2\left(\log L_U - \log L_R\right) = 2\left(-584.59 + 634.48\right) = 99.8 \gg \chi^2_{1,0.05} = 3.84$$

# Likelihood Ratio Test for House Money Effect (Probit)

- STATA Command:

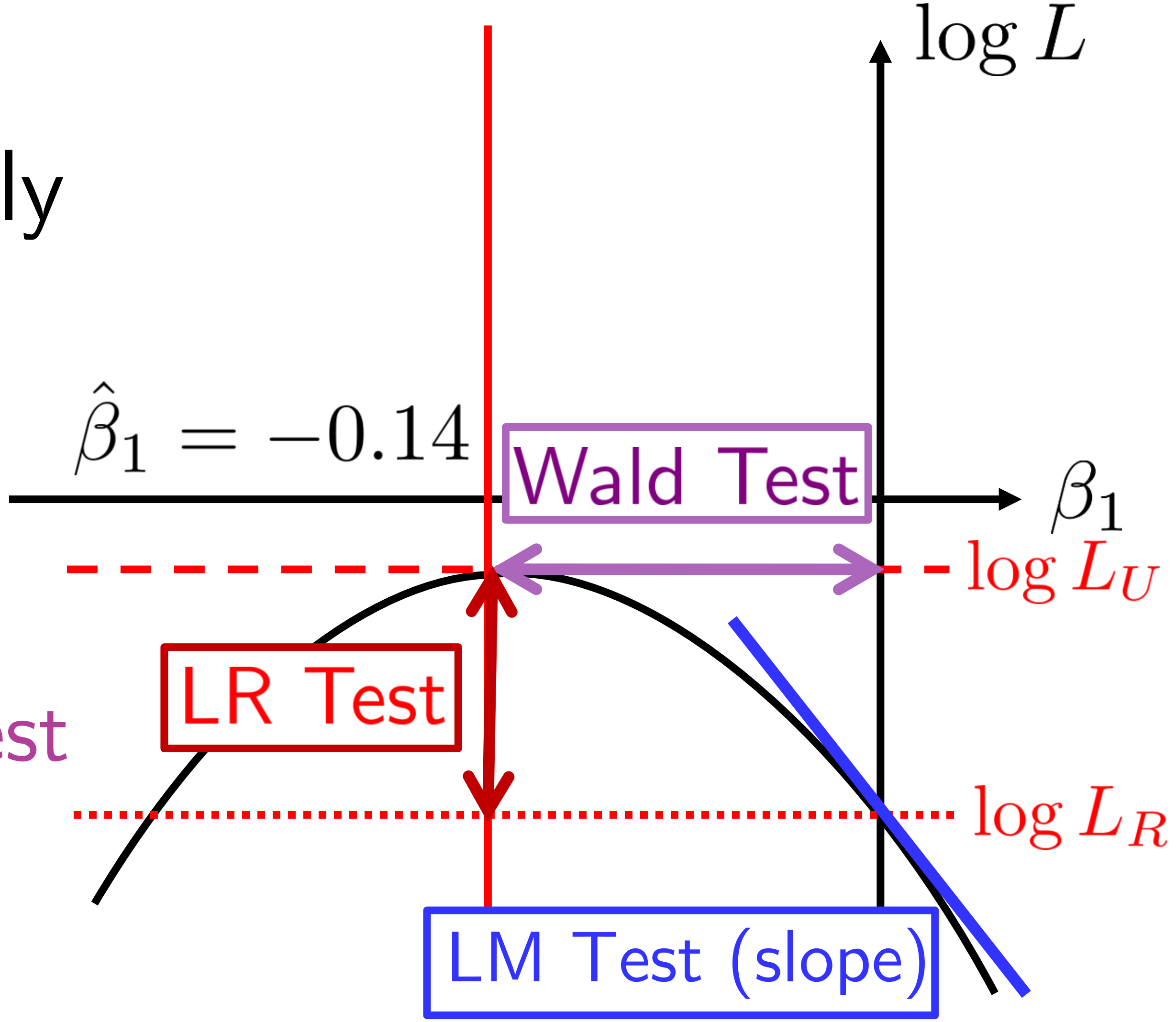
```
probit y w
est store with_w

probit y
est store without_w

lrtest with_w without_w
```

- STATA Results:

```
Likelihood-ratio test                                 LR chi2(1)  =     99.80
(Assumption: without_w nested in with_w)              Prob > chi2 =    0.0000
```

# Relation between Likelihood Ratio, Wald (and LM) Test

- All Three are Asymptotically Equivalent!
  - Likelihood Ratio (LR) Test
  - Wald Test
  - Lagrange Multiplier (LM) Test



# Maximum Likelihood Estimation (MLE) of Probit

- Estimate probit with MLE
- Use binary lottery choice of: `house_money_sim.dta`
  - STATA: `myprobit.do`
- Use `ml` to maximize

$$\log L = \sum_{i=1}^{n} \ln \left[\Phi(yy_i \times (\beta_0 + \beta_1 w_i))\right]$$

```
* LOG-LIKELIHOOD EVALUATION PROGRAM "myprobit" STARTS HERE:

program define myprobit

* SPECIFY NAME OF QUANTITY WHOSE SUM WE WISH TO MAXIMISE (logl)
* AND ALSO PARAMETER NAMES (EMBODIED IN xb)
* PROVIDE LIST OF TEMPORARY VARIABLES (p ONLY)

args logl xb
tempvar p

* GENERATE PROBABILITY OF CHOICE MADE BY EACH SUBJECT (p):

quietly gen double `p'=normal(yy*`xb')

* TAKE NATURAL LOG OF p AND STORE THIS AS logl

quietly replace `logl'=ln(`p')

* END "myprobit" PROGRAM:

end
```

# Maximum Likelihood Estimation (MLE) of Probit

- To ignore errors if the following command is not applicable, can add at the beginning

```
* LOG-LIKELIHOOD EVALUATION PROGRAM "myp
capture program drop myprobit
program define myprobit

* SPECIFY NAME OF QUANTITY WHOSE SUM WE
* AND ALSO PARAMETER NAMES (EMBODIED IN
* PROVIDE LIST OF TEMPORARY VARIABLES (p

args logl xb
tempvar p
```

- Maximize `logl`
- Over `xb` $= \beta_0 + \beta_1 w_i$
  - local variables like other defined by `tempvar`
  - Such as ‘`p`’

# Maximum Likelihood Estimation (MLE) of Probit

- Unlike global variables like yy
  - Does not need single quotation marks like local variables ‘p’ or ‘logl’

```
* GENERATE PROBABILITY OF CHOICE MADE BY
quietly gen double ‘p’=normal(yy*‘xb’)
* TAKE NATURAL LOG OF p AND STORE THIS A
quietly replace ‘logl’=ln(‘p’)
* END "myprobit" PROGRAM:
end
```

# Maximum Likelihood Estimation (MLE) of Probit

- The `ml` Routine uses the `lf` likelihood evaluator
  - Run on each row of the data set, unlike the `d-family` evaluator (which runs on each block of rows)

```
* READ DATA
use "house money_sim", clear
* GENERATE (INTEGER) yy FROM y:
gen int yy=2*y-1
* SPECIFY LIKELIHOOD EVALUATOR (lf), EVALUATION PROGRAM (myprobit),
* AND EXPLANATORY VARIABLE LIST.
* RUN MAXIMUM LIKELIHOOD PROCEDURE
ml model lf myprobit ( = w)
ml maximize
```

# Maximum Likelihood Estimation (MLE) of Probit

- STATA Results:

```
ml model lf myprobit ( = w)
ml maximize
```

```
initial:       log likelihood = -727.80454
alternative:   log likelihood = -635.1321
rescale:       log likelihood = -635.1321
Iteration 0:   log likelihood = -635.1321
Iteration 1:   log likelihood = -584.84039
Iteration 2:   log likelihood = -584.58503
Iteration 3:   log likelihood = -584.58503
                                                  Number of obs   =       1050
                                                  Wald chi2(1)    =      94.05
Log likelihood = -584.58503                       Prob > chi2     =     0.0000
------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           w |  -.1409882    .0145377    -9.70   0.000    -.1694816   -.1124948
       _cons |   1.301654    .0911155    14.29   0.000     1.123071    1.480237
------------------------------------------------------------------------------
```

# Same as Probit Model

- STATA `probit y w`

  Results:

```
Iteration 0:  log likelihood = -634.4833
Iteration 1:  log likelihood = -584.91375
Iteration 2:  log likelihood = -584.5851
Iteration 3:  log likelihood = -584.58503
Iteration 4:  log likelihood = -584.58503
Probit regression                                 Number of obs   =       1050
                                                  LR chi2(1)      =      99.80
                                                  Prob > chi2     =     0.0000
Log likelihood = -584.58503                       Pseudo R2       =     0.0786
------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           w |  -.1409882   .0145377    -9.70   0.000    -.1694816   -.1124948
       _cons |   1.301654   .0911155    14.29   0.000     1.123071    1.480237
------------------------------------------------------------------------------
```

1,050 Subjects
of 1 Round each

Strong House Money Effect ($z$ = -9.70)

# Homogeneous Structural Modelling

- Reduced form models simply attempt to explain the data
- Structural Models: Assume all individuals have the same utility function:

$$U(x) = \frac{x^{1-r}}{1-r}, \; r \neq 1$$
$$= \ln(x), \; r = 1$$

  - Constant Relative Risk Aversion (CRRA): $r = \mathrm{RRA}$
  - Higher $r =$ more risk averse
  - Negative $r =$ risk-seeking

# Homogeneous Structural Modelling

- In Binary Lottery Choice, Expected Utility for choosing Safe and Risky are:

$$EU(S) = \frac{(w+5)^{1-r}}{1-r}$$

$$EU(R) = 0.5\frac{(w)^{1-r}}{1-r} + 0.5\frac{(w+10)^{1-r}}{1-r}$$

  - Choose Safe if $EU(S) - EU(R) + \epsilon > 0$
- Fechner Error Term: When computing EU difference, individuals make computational error $\epsilon \sim N\left(0, \sigma^2\right)$

# Homogeneous Structural Modelling

- Probability of Safe choice being made is:

$$\begin{aligned}
\Pr(S) &= \Pr[EU(S) - EU(R) + \epsilon > 0] \\
&= \Pr[\epsilon > EU(R) - EU(S)] \\
&= \Pr\left[\frac{\epsilon}{\sigma} > \frac{EU(R) - EU(S)}{\sigma}\right] \\
&= 1 - \Phi\left[\frac{EU(R) - EU(S)}{\sigma}\right] = \Phi\left[\frac{EU(S) - EU(R)}{\sigma}\right]
\end{aligned}$$

# Homogeneous Structural Modelling

- Filling in EU (and using the yy trick), we obtain the log-likelihood function:

$$\log L = \sum_{i=1}^{n} \ln \Phi \left[ yy_i \times \frac{\frac{(w_i+5)^{1-r}}{1-r} - \left(0.5\frac{(w_i)^{1-r}}{1-r} + 0.5\frac{(w_i+10)^{1-r}}{1-r}\right)}{\sigma} \right]$$

- Choose $r$ and $\sigma$ to maximize $\log L$
  - Need to program this in STATA using the `ml` command

# Homogeneous Structural Modelling

- Choose $r$ and $\sigma$ to maximize $\log L$
  - STATA command: `args logl r sig`

```
program drop structural
program structural
args logl r sig
tempvar eus eur diff p
quietly gen double `eus'=(w+5)^(1-`r')/(1-`r')
quietly gen double `eur'=0.5*w^(1-`r')/(1-`r')+0.5*(w+10)^(1-`r')/(1-`r')
quietly gen double `diff'=(`eus'-`eur')/`sig'
quietly gen double `p'=normal(yy*`diff')
quietly replace `logl'=ln(`p')
end
```

# Homogeneous Structural Modelling

- STATA Results

```
initial: log likelihood = -<inf> (could not be evaluated)
feasible: log likelihood = -601.45646
rescale: log likelihood = -601.45646
rescale eq: log likelihood = -600.78259
Iteration 0: log likelihood = -600.78259
Iteration 1: log likelihood = -595.2424
Iteration 2: log likelihood = -595.22797
Iteration 3: log likelihood = -595.22739
Iteration 4: log likelihood = -595.22739
Number of obs = 1050
Wald chi2(0) = .
Log likelihood = -595.22739 Prob > chi2 = .
------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
r            |
       _cons |     .21765    .0976928     2.23   0.026     .0261757    .4091244
-------------+----------------------------------------------------------------
sig          |
       _cons |   .3585733    .1046733     3.43   0.001     .1534174    .5637292
------------------------------------------------------------------------------
```

$$r = 0.2177$$

$$\sigma = 0.3586$$

# Homogeneous Structural Modelling

- The estimated (homogeneous) utility function is:

$$U(x) = \frac{x^{1-0.2177}}{1-0.2177} = \frac{x^{0.7823}}{0.7823}$$

  - With Fechner computation error $\epsilon \sim N(0, 0.3586^2)$

- Note that this estimation assumes every individual has the same risk preference
  - We can relax this assumption...

# Heterogeneous Agent Model

- Now assume each subject has his/her own $r$ for the CRRA utility function:

$$U(x) = \frac{x^{1-r}}{1-r}$$

  - Assume $r$ has population distribution $r \sim N(\mu, \sigma^2)$
- Subject respond to Multiple Price List (MPL) of Holt and Laury (2002)
  - Choose Safe or Risky lottery for each question

# Heterogeneous Agent Model

- Indifferent between S and R at threshold risk attitude $r^*$

| Problem | Safe(S) | Risky(R) | $r^*$ |
|---|---|---|---|
| 1 | (0.1, \$2.00; 0.9, \$1.60) | (0.1, \$3.85; 0.9, \$0.10) | −1.72 |
| 2 | (0.2, \$2.00; 0.8, \$1.60) | (0.2, \$3.85; 0.8, \$0.10) | −0.95 |
| 3 | (0.3, \$2.00; 0.7, \$1.60) | (0.3, \$3.85; 0.7, \$0.10) | −0.49 |
| 4 | (0.4, \$2.00; 0.6, \$1.60) | (0.4, \$3.85; 0.6, \$0.10) | −0.15 |
| 5 | (0.5, \$2.00; 0.5, \$1.60) | (0.5, \$3.85; 0.5, \$0.10) | 0.15 |
| 6 | (0.6, \$2.00; 0.4, \$1.60) | (0.6, \$3.85; 0.4, \$0.10) | 0.41 |
| 7 | (0.7, \$2.00; 0.3, \$1.60) | (0.7, \$3.85; 0.3, \$0.10) | 0.68 |
| 8 | (0.8, \$2.00; 0.2, \$1.60) | (0.8, \$3.85; 0.2, \$0.10) | 0.97 |
| 9 | (0.9, \$2.00; 0.1, \$1.60) | (0.9, \$3.85; 0.1, \$0.10) | 1.37 |
| 10 | (1.0, \$2.00; 0.0, \$1.60) | (1.0, \$3.85; 0.0, \$0.10) | $\infty$ |

# Heterogeneous Agent Model

- Threshold risk attitude $r^*$ can be calculated with Excel in:
  - `risk aversion calculations.xlsx`
- Each subject is only asked 1 of the 10 problems
  - (Pseudo) data for 100 subjects: `holtlaury_sim.dta`
- Subject $i$ asked choice problem with threshold $r_i^*$
- Safe Choice Dummy:
  - $y_i = 1$ if chose S
  - $y_i = 0$ if chose R

# Heterogeneous Agent Model

- Given $r_i \sim N(\mu, \sigma^2)$, we have:

$$\Pr(y_i = 1) = \Pr(r_i > r_i^*) = \Pr\left(z > \frac{r_i^* - \mu}{\sigma}\right)$$

$$= \Pr\left(z < \frac{\mu - r_i^*}{\sigma}\right) = \Phi\left[\frac{\mu}{\sigma} + \left(-\frac{1}{\sigma}\right) r_i^*\right] = \Phi\left[\alpha + \beta r_i^*\right]$$

  - Can estimate a probit model: `probit y rstar`
  - Then, apply delta method:

```
nlcom (mu: -_b[_cons]/_b[rstar]) (sig: -1/_b[rstar])
```

# Delta Method

- (`nlcom` in STATA) used to obtain standard errors of $\hat{\mu}, \hat{\sigma}$
- More generally, consider reduced form estimates $\hat{\alpha}, \hat{\beta}$
  - Variance matrix is: $\hat{V}\begin{pmatrix} \hat{\beta} \\ \hat{\alpha} \end{pmatrix} = \begin{pmatrix} \mathrm{Var}(\hat{\beta}) & \mathrm{Cov}(\hat{\beta}, \hat{\alpha}) \\ \mathrm{Cov}(\hat{\alpha}, \hat{\beta}) & \mathrm{Var}(\hat{\alpha}) \end{pmatrix}$
  - After estimating `probit`, can see it in STATA: `mat V=e(V)`
  - STATA Results:

```
symmetric V[2,2]
                   y:          y:
               rstar       _cons
y:rstar    .12119211
y:_cons   -.04842685   .05126459
```

```
mat list V
```

# The Delta Method

- Can uncover structure parameters $\mu, \sigma$ from reduced form estimates of $\alpha, \beta$ through
$$\alpha = \frac{\mu}{\sigma}, \beta = -\frac{1}{\sigma} \Rightarrow \mu = -\frac{\alpha}{\beta}, \sigma = -\frac{1}{\beta}$$
- Estimate matrix $D = \begin{pmatrix} \frac{\partial\mu}{\partial\beta} & \frac{\partial\mu}{\partial\alpha} \\ \frac{\partial\sigma}{\partial\beta} & \frac{\partial\sigma}{\partial\alpha} \end{pmatrix} = \begin{pmatrix} \frac{\alpha}{\beta^2} & -\frac{1}{\beta} \\ \frac{1}{\beta^2} & 0 \end{pmatrix}$
- Use square root of diagonal in $\hat{V}\begin{pmatrix} \hat{\mu} \\ \hat{\sigma} \end{pmatrix} = \hat{D}\left[\hat{V}\begin{pmatrix} \hat{\beta} \\ \hat{\alpha} \end{pmatrix}\right]\hat{D}'$

# Heterogeneous Agent Model

- STATA Results `probit y rstar`

```
Iteration 0: log likelihood = -68.994376
Iteration 1: log likelihood = -32.754689
Iteration 2: log likelihood = -31.899974
Iteration 3: log likelihood = -31.896643
Iteration 4: log likelihood = -31.896643
Probit regression Number of obs = 100
LR chi2(1) = 74.20
Prob > chi2 = 0.0000
Log likelihood = -31.896643 Pseudo R2 = 0.5377

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.       z    P>|z|      [95% Conf. Interval]
-------------+----------------------------------------------------------------
       rstar |  -1.826082   .3481266    -5.25    0.000     -2.508398   -1.143767
       _cons |   .7306556   .2264169     3.23    0.001      .2868867    1.174424
------------------------------------------------------------------------------
Note: 10 failures and 0 successes completely determined.
```

# Heterogeneous Agent Model

STATA Results

```
nlcom (mu: -_b[_cons]/_b[rstar]) (sig: -1/_b[rstar])
```

```
mu: -_b[_cons]/_b[rstar]
sig: -1/_b[rstar]
```

| y | Coef. | Std. Err. | z | P>\|z\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| mu | .400122 | .0978294 | 4.09 | 0.000 | .2083799 | .5918641 |
| sig | .5476205 | .104399 | 5.25 | 0.000 | .3430021 | .7522389 |

$\hat{\mu} = 0.4001$

$\hat{\sigma} = 0.5476$

- Hence, every subject has RRA coefficient drawn from:

$$r \sim N(0.4001, 0.5476^2)$$

  - And calculate EU to make decision without error

# Interval Data

- Revisit Holt and Laury (2002)
- Still assume subjects have CRRA utility function:

$$U(x) = \frac{x^{1-r}}{1-r}$$

- But subject are asked each problem in order, revealing where in the list they switch
  - More precise information available regarding subject risk preference

# Interval Data

- EU-maximizing subject has RRA $r$ between 0.15 and 0.41

| Problem | Safe(S) | Risky(R) | $r^*$ |
|---|---|---|---|
| 1 | (0.1, \$2.00; 0.9, \$1.60) | (0.1, \$3.85; 0.9, \$0.10) | $-1.72$ |
| 2 | (0.2, \$2.00; 0.8, \$1.60) | (0.2, \$3.85; 0.8, \$0.10) | $-0.95$ |
| 3 | (0.3, \$2.00; 0.7, \$1.60) | (0.3, \$3.85; 0.7, \$0.10) | $-0.49$ |
| 4 | (0.4, \$2.00; 0.6, \$1.60) | (0.4, \$3.85; 0.6, \$0.10) | $-0.15$ |
| 5 | (0.5, \$2.00; 0.5, \$1.60) | (0.5, \$3.85; 0.5, \$0.10) | 0.15 |
| 6 | (0.6, \$2.00; 0.4, \$1.60) | (0.6, \$3.85; 0.4, \$0.10) | 0.41 |
| 7 | (0.7, \$2.00; 0.3, \$1.60) | (0.7, \$3.85; 0.3, \$0.10) | 0.68 |
| 8 | (0.8, \$2.00; 0.2, \$1.60) | (0.8, \$3.85; 0.2, \$0.10) | 0.97 |
| 9 | (0.9, \$2.00; 0.1, \$1.60) | (0.9, \$3.85; 0.1, \$0.10) | 1.37 |
| 10 | (1.0, \$2.00; ~~0.0, \$1.60~~) | (1.0, \$3.85; 0.0, \$0.10) | $\infty$ |

# Interval Data: Interval Regression Model

- Estimate population $r$ from subject-specific intervals
  - (Pseudo) data for 100 subjects: `interval_data_sim.dta`
- For $r_i \sim N(\mu, \sigma^2)$, subject $i$ with $l_i < r_i < h_i$ has

$$L_i = \Pr(l_i < r < h_i) = \Pr(r_i < h_i) - \Pr(r < l_i)$$

$$= \Phi\left(\frac{u_i - \mu}{\sigma}\right) - \Phi\left(\frac{l_i - \mu}{\sigma}\right)$$

- Hence, $\log L = \sum_{i=1}^{n} \ln\left[\Phi\left(\frac{u_i - \mu}{\sigma}\right) - \Phi\left(\frac{l_i - \mu}{\sigma}\right)\right]$

# Interval Data: Interval Regression Model

- Interval Regression: Estimate likelihood-maximizing $\hat{\mu}, \hat{\sigma}$
- STATA command: `intreg rlower rupper`
- Result:

```
Fitting constant-only model:
Iteration 0: log likelihood = -199.07231
Iteration 1: log likelihood = -198.96851
Iteration 2: log likelihood = -198.96849
Fitting full model:
Iteration 0: log likelihood = -198.96849
Iteration 1: log likelihood = -198.96849
Interval regression Number of obs = 100
LR chi2(0) = 0.00
Log likelihood = -198.96849 Prob > chi2 = .
------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |    .613146   .0597808    10.26   0.000     .4959777    .7303143
-------------+----------------------------------------------------------------
    /lnsigma | -.5323404   .0764651    -6.96   0.000    -.6822092   -.3824716
```

# Interval Dat

- STATA
  Results

$$r \sim N\left(0.6131, 0.5872^2\right)$$

```
Fitting constant-only model:
Iteration 0: log likelihood = -199.07231
Iteration 1: log likelihood = -198.96851
Iteration 2: log likelihood = -198.96849
Fitting full model:
Iteration 0: log likelihood = -198.96849
Iteration 1: log likelihood = -198.96849
Interval regression Number of obs = 100
LR chi2(0) = 0.00
Log likelihood = -198.96849 Prob > chi2 = .
------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |    .613146   .0597808    10.26   0.000     .4959777    .7303143
-------------+----------------------------------------------------------------
    /lnsigma |  -.5323404   .0764651    -6.96   0.000    -.6822092   -.3824716
-------------+----------------------------------------------------------------
       sigma |    .587229   .0449025                       .505499    .6821733
------------------------------------------------------------------------------
Observation summary: 0 left-censored observations
0 uncensored observations
6 right-censored observations
94 interval observations
```

# Interval Data: Interval Regression Model

- If risk attitude depends on age and gender:

$$r_i = \beta_0 + \beta_1 age_i + \beta_2 male_i + \epsilon_i, \quad \epsilon_i \sim N(0, \sigma^2)$$

  - Explanatory variables $\vec{x}_i = (1 \quad age_i \quad male_i)'$
  - Have coefficients $\vec{\beta} = (\beta_0 \quad \beta_1 \quad \beta_2)'$

$$r_i = \vec{x}_i{}'\vec{\beta} + \epsilon_i, \quad \epsilon_i \sim N(0, \sigma^2) \Rightarrow r_i \sim N\left(\vec{x}_i{}'\vec{\beta}, \sigma^2\right)$$

- Hence,

$$\log L = \sum_{i=1}^{n} \ln \left[ \Phi\left( \frac{u_i - \vec{x}_i{}'\vec{\beta}}{\sigma} \right) - \Phi\left( \frac{l_i - \vec{x}_i{}'\vec{\beta}}{\sigma} \right) \right]$$

# Interval Data: Interval Regression Model

- STATA `intreg rlower rupper age male`
  Result:

```
Fitting constant-only model:
Iteration 0: log likelihood = -199.07231
Iteration 1: log likelihood = -198.96851
Iteration 2: log likelihood = -198.96849
Fitting full model:
Iteration 0: log likelihood = -197.24143
Iteration 1: log likelihood = -197.17109
Iteration 2: log likelihood = -197.17108
Interval regression Number of obs = 100
LR chi2(2) = 3.59
Log likelihood = -197.17108 Prob > chi2 = 0.1657
------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |     .02213    .0196956     1.12   0.261    -.0164727    .0607327
        male |  -.2165679    .1341118    -1.61   0.106    -.4794222    .0462864
       _cons |   .1592841    .4565128     0.35   0.727    -.7354646    1.054033
-------------+----------------------------------------------------------------
```

# Interval Dat

- STATA Result:

```
Fitting constant-only model:
Iteration 0: log likelihood = -199.07231
Iteration 1: log likelihood = -198.96851
Iteration 2: log likelihood = -198.96849
Fitting full model:
Iteration 0: log likelihood = -197.24143
Iteration 1: log likelihood = -197.17109
Iteration 2: log likelihood = -197.17108
Interval regression Number of obs = 100
LR chi2(2) = 3.59
Log likelihood = -197.17108 Prob > chi2 = 0.1657
------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |     .02213   .0196956     1.12   0.261    -.0164727    .0607327
        male |  -.2165679   .1341118    -1.61   0.106    -.4794222    .0462864
       _cons |   .1592841   .4565128     0.35   0.727    -.7354646    1.054033
-------------+----------------------------------------------------------------
    /lnsigma |  -.5507208   .0764747    -7.20   0.000    -.7006085   -.4008332
-------------+----------------------------------------------------------------
       sigma |   .5765341   .0440903                      .4962832    .6697618
------------------------------------------------------------------------------
6 right-censored observations
94 interval observations
```

But none significant

$$\hat{r}_i = 0.159 + 0.022age_i - 0.217male_i)$$

$$\hat{\sigma} = 0.577$$

# Continuous (Exact) Data

- Ask subject Certainty Equivalent (CE) for a lottery
  - Amount for sure indifferent with lottery
  - Exact information of subject risk preference
- With the CRRA utility function: $U(x) = \dfrac{x^{1-r}}{1-r}$
- If $(0.3,\ \$3.85;\ 0.7,\ \$0.10)$ has CE $= \$0.75$, can find $r$ so:

$$0.3\frac{(3.85)^{1-r}}{1-r} + 0.7\frac{(0.1)^{1-r}}{1-r} = \frac{(0.75)^{1-r}}{1-r}$$

  - $r = 0.41$! (See `risk aversion calculations.xlsx`)

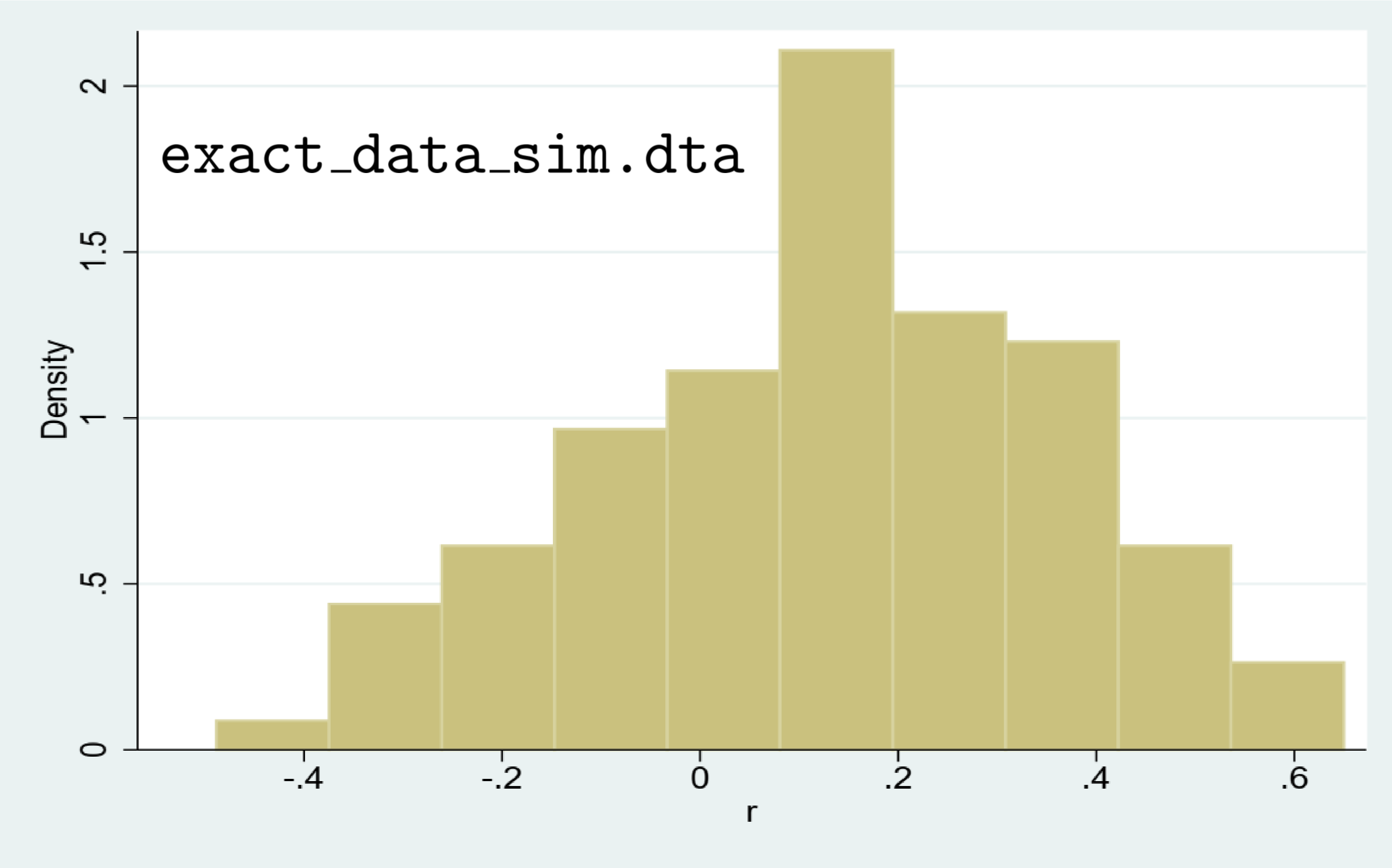
# Continuous (Exact) Data

- To elicit subject CE with Incentive Compatibility (IC)
  - Provide incentives for truthful report
- Use Becker-DeGroot-Marschak (BDM) Mechanism
  - Becker et al. (1964)

  1. Report CE. Then, computer draw a random price
  2. If random price is higher than CE, earn random price
  3. If random price is lower than CE, play the lottery
- Why is it IC to report truthfully?

# Continuous (Exact) Data

- Simulated Data for N=100:



# Continuous (Exact) Data

- Given population distribution $r \sim N\left(\mu, \sigma^2\right)$
  - Want to estimate $\hat{\mu}, \hat{\sigma}$
- Density for observation $r_i$ is

$$f(r_i; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(r_i-\mu)^2}{2\sigma^2}\right) = \frac{1}{\sigma}\phi\left(\frac{r_i-\mu}{\sigma}\right)$$

- Hence, the sample log-likelihood function is:

$$\log L = \sum_{i=1}^{n} \ln\left[\frac{1}{\sigma}\phi\left(\frac{r_i-\mu}{\sigma}\right)\right]$$

# Continuous (Exact) Data

- Choose $r$ and $\sigma$ to maximize $\log L$
  - STATA command: `args lnl xb sig`

```
program define exact
args lnf xb sig
tempvar y p
quietly gen double ‘y’=$ML_y1
quietly gen double ‘p’=(1/‘sig’)*normalden((‘y’-‘xb’)/‘sig’)
quietly replace ‘lnf’=ln(‘p’)
end
ml model lf exact (r= ) ()
ml maximize
```

# Continuous (Exact) Data

- STATA Results

```
initial: log likelihood = -<inf> (could not be evaluated)
feasible: log likelihood = -60.251905
rescale: log likelihood = -7.5739988
rescale eq: log likelihood = 3.1167494
Iteration 0: log likelihood = 3.1167494
Iteration 1: log likelihood = 3.2682025
Iteration 2: log likelihood = 3.6372157
Iteration 3: log likelihood = 3.637384
Iteration 4: log likelihood = 3.637384
Number of obs = 100
Wald chi2(0) = .
Log likelihood = 3.637384 Prob > chi2 = .
------------------------------------------------------------------------------
           r |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
eq1          |
       _cons |   .1340463   .0233327     5.74   0.000     .0883149    .1797776
-------------+----------------------------------------------------------------
eq2          |
       _cons |   .2333275   .0164987    14.14   0.000     .2009905    .2656644
------------------------------------------------------------------------------
```

$$\hat{\mu} = 0.1340$$

$$\hat{\sigma} = 0.2333$$

# Sample Mean and Sample Variance are MLE!

- This is exactly the sample mean and variance of $r$
  - STATA command: `summ r`

| Variable | Obs | Mean | Std. Dev. | Min | Max |
|---|---|---|---|---|---|
| r | 100 | .1340463 | .2345029 | -.4884877 | .6499107 |

  - Except sample variance is divided by $(n - 1)$ instead of $n$
- Recall from your Econometrics class
  - Sample mean and sample variance maximizes likelihood!
  - But ML applies to other continuous data (even censored)

# CE Closer to Risk Neutrality?

- $\hat{\mu}^{\text{MLE}} = 0.1340$ much closer to 0 than previous ones:
  - Homogeneous Agent Model: $\hat{r}^{\text{Homo}} = 0.2177$
  - Heterogeneous Agent Model: $\hat{\mu}^{\text{Hetero}} = 0.4001$
  - Interval Data: $\hat{\mu}^{\text{Interval}} = 0.6131$
- Subjects tend toward risk neutrality when asked CE
  - As if they compute EV and report something near
- Explains: Prefer safer lottery in binary choice (P-bet)
  - But place higher valuation on riskier lottery ($-bet)

Preference Reversals!!

# Acknowledgment

- This presentation is based on
  - Section 3.1, 3.3-10 of the lecture notes of Experimetrics,
- prepared for a mini-course taught by Peter G. Moffatt (UEA) at National Taiwan University in Spring 2019
  - We thank 沈汶鋒, 陳詠駿、張學仁, 蘇靖勛、胡南圳 and 林欣蓁 for their in-class presentations