Greek Letters

αA	`\alpha A`	νN	`\nu N`
βB	`\beta B`	$\xi \Xi$	`\xi \Xi`
$\gamma \Gamma$	`\gamma \Gamma`	$o O$	`o O`
$\delta \Delta$	`\delta \Delta`	$\pi \Pi$	`\pi \Pi`
$\epsilon \varepsilon E$	`\epsilon \varepsilon E`	$\rho \varrho P$	`\rho \varrho P`
ζZ	`\zeta Z`	$\sigma \Sigma$	`\sigma \Sigma`
ηH	`\eta H`	τT	`\tau T`
$\theta \vartheta \Theta$	`\theta \vartheta \Theta`	$\upsilon \Upsilon$	`\upsilon \Upsilon`
ιI	`\iota I`	$\phi \varphi \Phi$	`\phi \varphi \Phi`
κK	`\kappa K`	χX	`\chi X`
$\lambda \Lambda$	`\lambda \Lambda`	$\psi \Psi$	`\psi \Psi`
μM	`\mu M`	$\omega \Omega$	`\omega \Omega`

Greek Letters

- Notations have their common meaning
- π for profit, θ for parameter, ϵ for idiosyncratic shocks
- Greek letters are usually quantities unobserved to economists

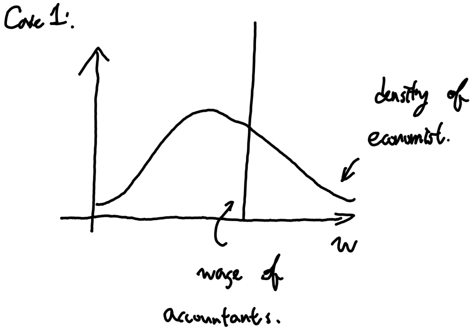
Review of Some Statistic Properties

- Truncated normal formula:
  - Let $X \sim N(\mu, \sigma^2)$
  - $E[X|X > a] = \mu + \sigma \frac{\phi(a)}{1-\Phi(a)}$
- Joint normality:
  - Suppose X and Y are joint normal
  - $E[Y|X = x] = \frac{\sigma_{XY}}{\sigma_X^2} x$
  - Note that $\frac{\sigma_{XY}}{\sigma_X^2}$ is the regression coefficient of Y on X

Motivating Question

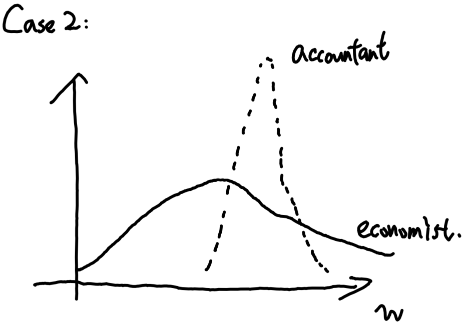
- What would be your earnings if I become an accountant instead of an economist?
- Maybe Look at the difference in avg. earning?
  - $\bar{y}_A$ - $\bar{y}_E$
- Is this correct?
- Under what circumstance would it be correct?
- Think about this as if you're making this choice

Case I



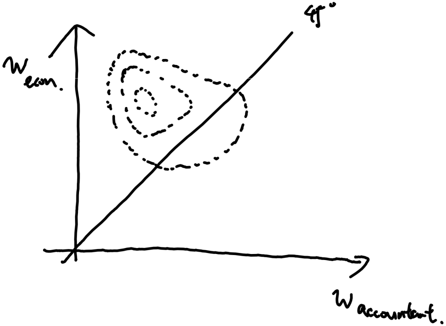
- Who would become an economist? Can that be shown on the graph?

Case II



- Who would become an economist? Can that be shown on the graph?

Case II



- Who would become an economist? Can that be shown on the graph?
- It's the joint distribution that matters!

Back to the Motivation Question

- What would be your earning if I become an accountant instead of economist?
- Think about Case I again
- Does comparing the difference in mean make sense?
- No!
    - It depends on who you are
    - On average, it's incorrect either
- Roy model is a formal way to think about self-selection and how it relates to data
- You might be aware, but (almost) everything is Roy model in economics
- Definitely includes those econometric techniques, such as IV, ..., etc.

Example: Who migrates?

- Simple model:

$$\begin{cases} w_{i0} &= \mu_0 + \epsilon_{i0}, \text{ stay} \\ w_{i1} &= \mu_1 + \epsilon_{i1}, \text{ migrate} \end{cases}$$

- w_{i0} and w_{i1}: **potential** earnings of staying and migrating
- μ_0 and μ_1: mean of **potential** earnings of staying and migrating
- ϵ_{i0} and ϵ_{i1}: idiosyncratic **potential** earnings of staying and migrating
  - e.g., someone who specifically doesn't have the skill for the origin country
- Assumptions:

$$\begin{pmatrix} \epsilon_{i0} \\ \epsilon_{i1} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_0^2 & \sigma_{01} \\ \sigma_{01} & \sigma_1^2 \end{pmatrix} \right)$$

- Migrant cost: C
- Correlation coefficient: $\rho \equiv \frac{\sigma_{01}}{\sigma_0 \sigma_1}$

Migration Decision

- Migrate if and only if: $w_{i1} > w_{i0} + C$
- Define: $\nu \equiv \epsilon_{i1} - \epsilon_{i0}$
- Probability of migration (or share of people who migrate):

$$
\begin{aligned}
Pr(w_{i1} > w_{i0} + C) &= Pr(\mu_1 + \epsilon_{i1} > \mu_0 + \epsilon_{i0} + C) \\
&= Pr(\epsilon_{i1} - \epsilon_{i0} > \mu_0 - \mu_1 + C) \\
&= Pr\left(\frac{\epsilon_{i1} - \epsilon_{i0}}{\sigma_\nu} > \frac{\mu_0 - \mu_1 + C}{\sigma_\nu}\right) \\
&\equiv Pr\left(\frac{\epsilon_{i1} - \epsilon_{i0}}{\sigma_\nu} > z\right) \\
&= 1 - \Phi(z)
\end{aligned}
$$

In the Data

- What do we observe in the data?
- We observe
  - earnings for those who stay if they stay: $E[w_{i0}|\text{Stay}]$
  - earnings for those who migrate if they migrate: $E[w_{i1}|\text{Migrate}]$
- We don't observe
  - earnings for those who stay if they migrate: $E[w_{i1}|\text{Stay}]$
  - earnings for those who migrate if they stay: $E[w_{i0}|\text{Migrate}]$
- Can we infer what would one's earnings change if she migrates or stays, $w_{i1} - w_{i0}$?
- Or at least the average amount, $E[w_{i1} - w_{i0}]$?

Model Implications

- Turns out the model provides a rich framework for thinking about these
- Let's start with the quantity $E[w_{i0}|\text{Migrate}]$

$$E[w_{i0}|\text{Migrate}] = \mu_0 + E[\epsilon_{i0}|\frac{\nu}{\sigma_\nu} > z]$$

- Using the regression coefficient formula:

$$E[\epsilon_{i0}|\nu] = \frac{\sigma_{0\nu}}{\sigma_\nu^2}\nu \qquad (1)$$

-

$$E[\frac{\epsilon_{i0}}{\sigma_0} \mid \frac{\nu}{\sigma_\nu}] = \underbrace{\frac{1}{\sigma_0}}_{\because\ \text{divide by } \sigma_0} \overbrace{\frac{\sigma_{0\nu}}{\sigma_\nu^2}\frac{\nu}{\sigma_\nu}}^{\text{original eq1}} \underbrace{\frac{1}{\sigma_\nu^{-2}}}_{\nu \text{ becomes } \frac{\nu}{\sigma_\nu}} \overbrace{\frac{1}{\sigma_\nu}}^{\sigma_{0\nu} \text{ changed}} = \frac{\sigma_{0\nu}}{\sigma_0\sigma_\nu}\frac{\nu}{\sigma_\nu} = \rho_{0\nu}\frac{\nu}{\sigma_\nu}$$

Model Implications

$$
\begin{aligned}
E[\frac{\epsilon_{i0}}{\sigma_0} \mid \frac{\nu}{\sigma_\nu}] &= \rho_{0\nu}\frac{\nu}{\sigma_\nu} \\
E[w_{i0} \mid \text{Migrate}] &= \mu_0 + \sigma_0 E[\frac{\epsilon_{i0}}{\sigma_0} \mid \frac{\nu}{\sigma_\nu} > z] \\
&= \mu_0 + \rho_{0\nu}\sigma_0 E(\frac{\nu}{\sigma_\nu} \mid \frac{\nu}{\sigma_\nu} > z) \\
&= \mu_0 + \rho_{0\nu}\sigma_0 \frac{\phi(z)}{1 - \Phi(z)}
\end{aligned}
$$

- Similarly, $E[w_{i1} \mid \text{Migrate}] = \mu_1 + E[\epsilon_{i1} \mid \frac{\nu}{\sigma_\nu} > z] = \mu_1 + \rho_{1\nu}\sigma_1 \frac{\phi(z)}{\Phi(-z)}$

Understanding Migration

- Putting together:

$$E[w_{i0}|\text{Migrate}] = \mu_0 + \frac{\sigma_0\sigma_1}{\sigma_\nu}(\rho - \frac{\sigma_0}{\sigma_1})(\frac{\phi(z)}{1-\Phi(z)})$$
$$E[w_{i1}|\text{Migrate}] = \mu_1 + \frac{\sigma_0\sigma_1}{\sigma_\nu}(\frac{\sigma_1}{\sigma_0} - \rho)(\frac{\phi(z)}{1-\Phi(z)})$$

- For convenience, write the latter part
  - $Q_0 \equiv E[\epsilon_{i0}|\text{Migrate}]$
  - $Q_1 \equiv E[\epsilon_{i1}|\text{Migrate}]$

Who Migrates?

- Case 1: Positive Selection
  - Conditions: $Q_0 > 0$, $Q_1 > 0$
  - Meaning: Bright people move for opportunities
- Case 2: Negative Selection
  - Conditions: $Q_0 < 0$, $Q_1 < 0$
  - Meaning: Worst people move for "insurance"
- Case 3: Refugee, Sorting
  - Conditions: $Q_0 < 0$, $Q_1 > 0$
  - Meaning: People sort where most needed
- Case 4: Open Question
  - Is $Q_0 > 0$, $Q_1 < 0$ possible?

Revisit the Motivation Issue

- Comparing the earnings of accountants and economists is like:

$$E[w_{i1}|\text{Migrate}] - E[w_{i0}|\text{Stay}]$$

- Different from the parameter of interest:

$$E[w_{i1} - w_{i0}]$$

- Or questions such as: What are the gains from migration?

$$E[w_{i1}|\text{Migrate}] - E[w_{i0}|\text{Migrate}]$$

- Note that this is a counterfactual question
- Most interesting questions in economics are counterfactual
- With the above expressions, possible to answer these