A Random Utility Approach to Strategic Models

Curtis S. Signorino
University of Rochester

July 1, 2003

Introduction

- Theories where one or more actors — be they individuals, firms, parties, or states — make choices over discrete sets of actions (or options) leading to a discrete set of outcomes.

- Most common method of statistically analyzing discrete categorical data has been to use one variant or another of logit or probit. Most of these statistical discrete choice models can be derived from assumptions of utility maximization, where the actor's utilities over outcomes include a random element with a known distribution (see, for example, McFadden 1974a, 1974b, 1976; Hausman & Wise 1978; Alvarez & Nagler 1998).

- The randomness in the utility is often attributed to one of two sources:
  - **regressor error**
  - **“bounded rationality”** due to misperception or implementation error

- Experimental game theory has recently given rise to statistical equilibrium solution concepts for extensive form games (see, for example, McKelvey & Palfrey 1996, 1998; Zauner 1996; Chen, Friedman, Thisse 1996). These are also based on random utility assumptions, except that the sources of error do not overlap exactly with the statistical literature:
  - incomplete information due to outcome **payoff disturbances**
  - bounded rationality

- Signorino & Yilmaz (2003) formally characterize “strategic” specification error — structural or functional form misspecification.

- Little attention to the source of uncertainty in these models, whether it is **theoretically** or **econometrically** motivated, and how that uncertainty interacts with the structural assumptions to produce different probability models.

On “Theoretical” and “Statistical” Models

- For our inferences to be valid, we need a statistical model that accurately represents the theory being tested. How can we derive a *statistical model* from our *theoretical model*?

- Not mutually exclusive.
  - Game-theoretic model of war with perfect and complete information would be a theoretical but not statistical model.
  - Logit models of war where the regressors were randomly selected or selected via stepwise regression only to improve fit would be examples of statistical but not necessarily theoretical models
  - A game-theoretic model of war that employed a statistical equilibrium solution concept — such as McKelvey & Palfrey’s (1998) logit quantal response equilibrium — would fall in the category of both theoretical and statistical

Recipe for RUM

One derives a statistical model from a theoretical model using the same general steps:

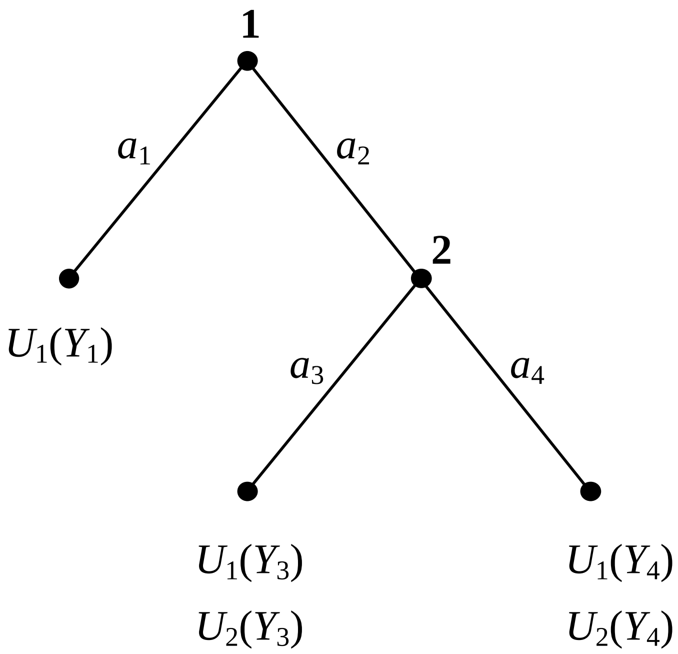
1. specify the theoretical choice model
2. add a random component (i.e., source of uncertainty) if none exists
3. derive the probability model associated with one's dependent variable
4. construct a likelihood equation based on that probability model

Choice Model

- The choice model identifies not only the set of players, their sequence of moves, their options, and their incentives, but also the rule players use to choose among options.

- A player's "true" utility for an action or outcome will often consist of an observable component, related to the explanatory variables, and an unobservable component, related to the type of uncertainty. Player m's true utility for outcome Y_k will be written as $U_m^*(Y_k)$. The true utility for an action a_j takes the same form: $U_m^*(a_j)$.

- Players maximize their expected utility.



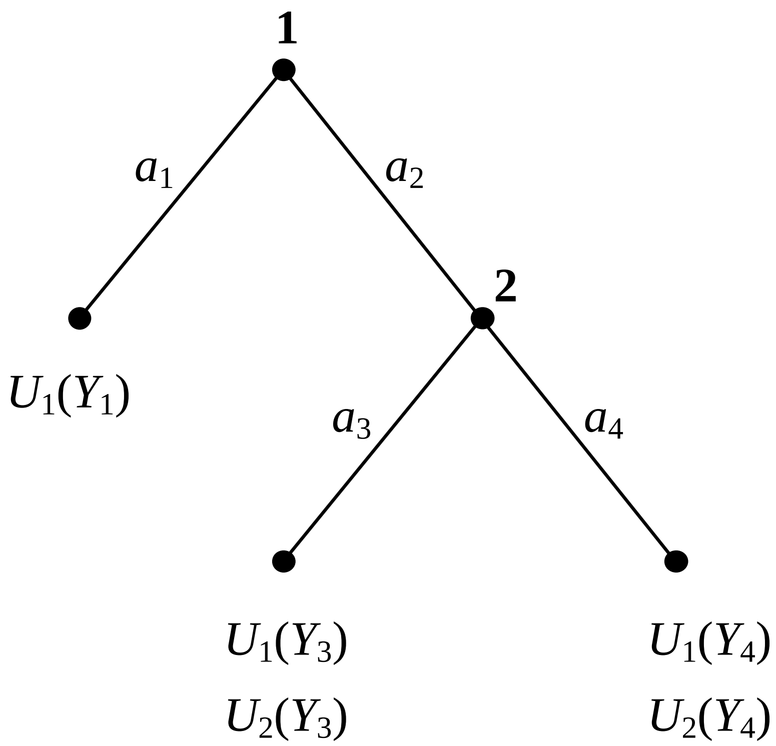
- Formally specifying the choice structure and choice rule precisely relates the dependent variable — whether it is an action or outcome in the model — to the explanatory variables, which enter through the utilities.

$$y = \begin{cases} Y_1 & \text{if } U_2^*(Y_4) > U_2^*(Y_3) \text{ and } U_1^*(Y_4) < U_1^*(Y_1), \text{ or} \\ & \text{if } U_2^*(Y_3) > U_2^*(Y_4) \text{ and } U_1^*(Y_3) < U_1^*(Y_1) \\ Y_3 & \text{if } U_2^*(Y_3) > U_2^*(Y_4) \text{ and } U_1^*(Y_3) > U_1^*(Y_1) \\ Y_4 & \text{if } U_2^*(Y_4) > U_2^*(Y_3) \text{ and } U_1^*(Y_4) > U_1^*(Y_1) \end{cases} \tag{1}$$

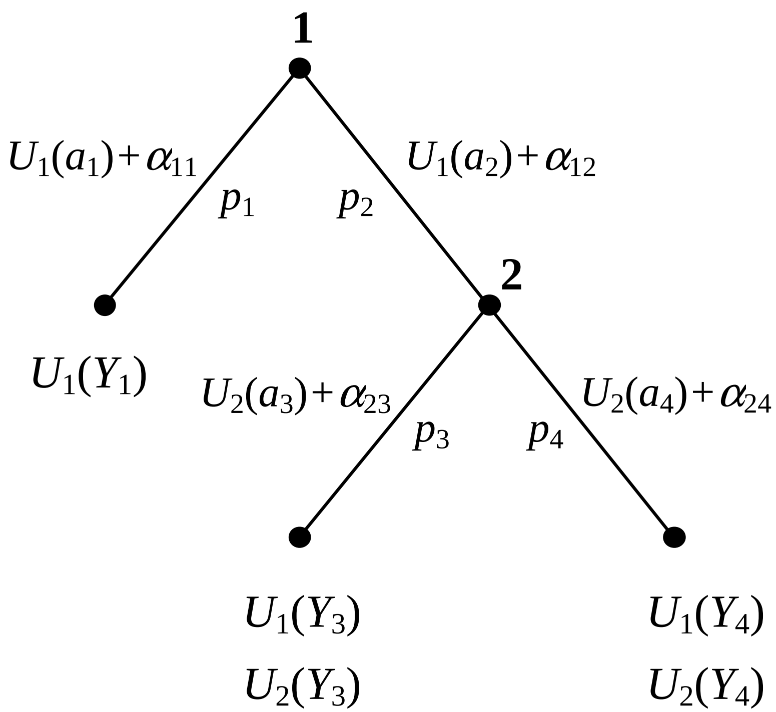
Sources of Uncertainty

- The random component enters as an assumption about
  1. the information available to the analyst concerning the players' utilities
  2. the information available to each player concerning the other's utilities
  3. both
- Major difference between our undertaking and that of traditional game theorists: game theorists are omniscient with respect to their models — they know (and specify) each of the players' utilities when they solve for the equilibria. The poor empirical analyst is not so fortunate. We may assume an underlying game-theoretic model. However, we must allow for the fact that we cannot know the players' true utilities.

No Uncertainty: SPE

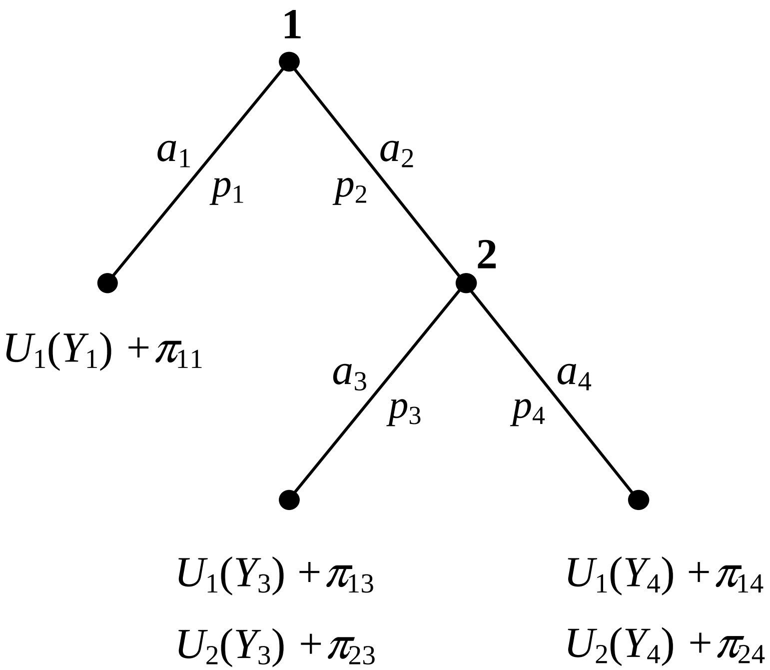


Agent Error



- Players sometimes misperceive each other's utilities or err in implementing their actions.
- α_{ij} is private to i
- Other players *and* analyst do not observe it

Private Information about Payoffs

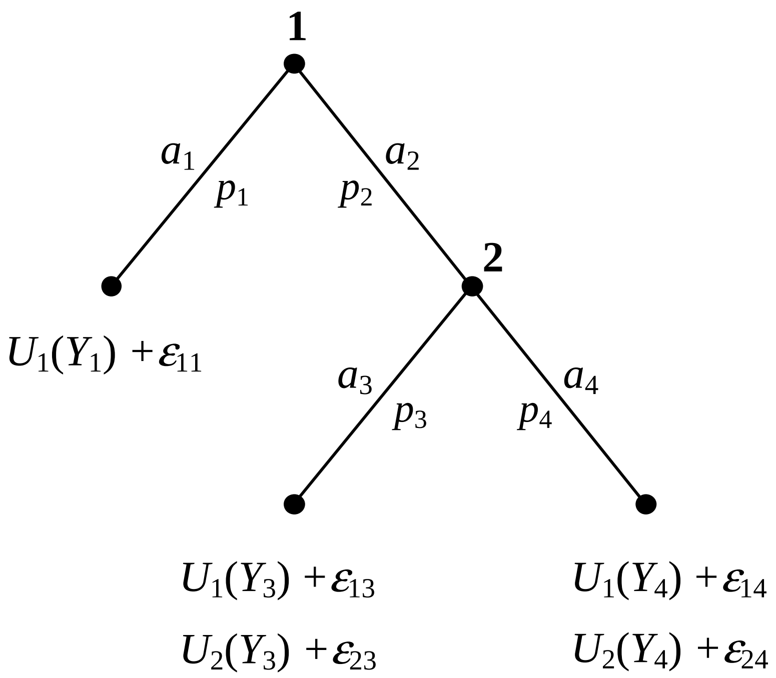


- Players have private information concerning their own outcome payoffs.

- π_{ij} is private to i

- Other players *and* analyst do not observe it

Regressor Error



- Players have no uncertainty concerning each other's payoffs
- Analyst does not perfectly specify utilities
- ϵ_{ij} similar to traditional regression disturbance

Likelihood Equation.

- Given a choice model, a source of uncertainty, and an appropriate distribution for the random components, we can derive probabilities for the actions and outcomes in the model.

- Let $\delta_k = 1$ if outcome Y_k occurred in the current observation, and zero otherwise. Then the likelihood function is

$$L = \prod^{N} p_{Y_1}^{\delta_1} p_{Y_3}^{\delta_3} p_{Y_4}^{\delta_4} \tag{2}$$

- For a dependent variable denoting whether a_1 or a_2 occurred, the likelihood function would be

$$L = \prod^{N} p_1^{\delta_{a_1}} p_2^{\delta_{a_2}}, \tag{3}$$

  where $\delta_{a_j} = 1$ if action a_j was chosen, and zero otherwise.

A General Class of Statistical Discrete Choice Models

- Choice Tree
- Information Sets
- Actions and Outcomes
- Utilities and "Error" Terms
- Agent Utility Maximization
- Statistical Discrete Choice Model

  Assume now that we have N observations of a dependent variable y that takes on the values $y_i \in Y$. Define $j \in Y$ and let the dummies $y_{i,j}$ take on the value

$$y_{i,j} = \begin{cases} 1 & \text{if } j = y_i \\ 0 & \text{otherwise} \end{cases} .$$

In other words, $y_{i,j}$ is a dummy that indicates whether outcome j occurred or not in observation i. Assuming that the utilities are specified in terms of explanatory variables (e.g., $U_{mki}(Y_k) = X_{mki}\beta_{mk}$), then we can denote the probability of Y_k occurring in observation i as $p_{Y_k i}$ and estimate the parameters β_{mk} by maximizing the multinomial likelihood function

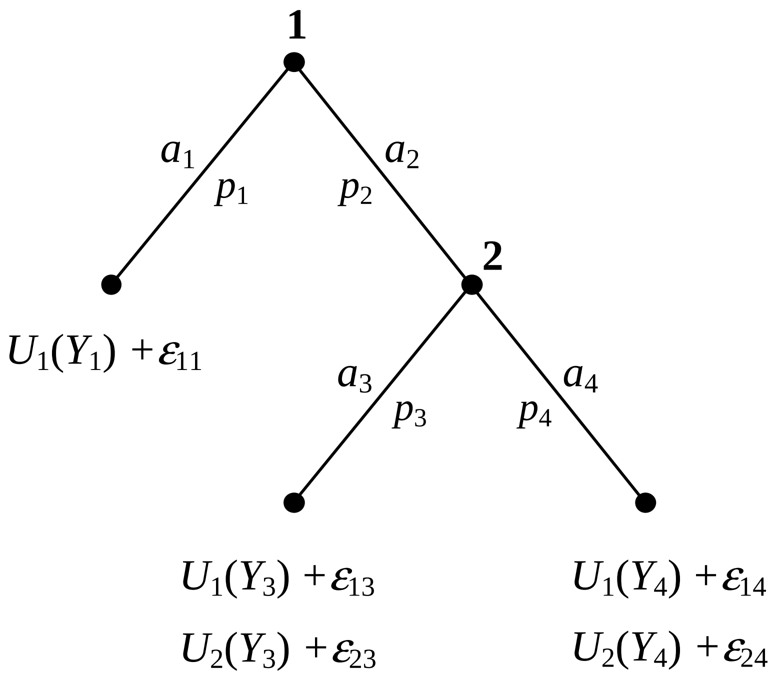
$$L = \prod_{i=1}^{N} \prod_{j \in Y} p_{ji}^{y_{i,j}} \tag{4}$$

with respect to the β_{mk}. Again,

Players M	Info Sets I	Termnl Nodes Q^t	Y	SPECIAL CASES $\Omega \sim$ T1EV	$\Omega \sim$ N(0,Σ)
				NONSTRATEGIC	
1	1	2	2	Binomial Logit	Binomial Probit
1	1	≥ 2	$\#Q^t$	Multinomial Logit Nested Logit†	Multinomial Probit
1	> 1	$> \#I$	> 1	Sequential Logit	Sequential Probit
1 or 2	2	3	3		Heckman Selection
				STRATEGIC	
> 1	$\geq \#M$	$> \#I$	> 1	Logit	Probit

†$\Omega \sim$ GEV

Examples: Strategic Probit Models



Regressor Error. What are the conditions for Y_k to be realized? Here, that question is equivalent to asking when Y_k will be the subgame perfect equilibrium (SPE).

$$y = \begin{cases} Y_1 & \text{if } U_2^*(Y_3) > U_2^*(Y_4) \text{ and } U_1^*(Y_1) > U_1^*(Y_3) \text{ or} \\ & \text{if } U_2^*(Y_4) > U_2^*(Y_3) \text{ and } U_1^*(Y_1) > U_1^*(Y_4) \\ Y_3 & \text{if } U_2^*(Y_3) > U_2^*(Y_4) \text{ and } U_1^*(Y_3) > U_1^*(Y_1) \\ Y_4 & \text{if } U_2^*(Y_4) > U_2^*(Y_3) \text{ and } U_1^*(Y_4) > U_1^*(Y_1) \end{cases} \tag{5}$$

$$y = \begin{cases} Y_1 & \text{if } U_2^*(Y_3) > U_2^*(Y_4) \text{ and } U_1^*(Y_1) > U_1^*(Y_3) \text{ or} \\ & \text{if } U_2^*(Y_4) > U_2^*(Y_3) \text{ and } U_1^*(Y_1) > U_1^*(Y_4) \\ Y_3 & \text{if } U_2^*(Y_3) > U_2^*(Y_4) \text{ and } U_1^*(Y_3) > U_1^*(Y_1) \\ Y_4 & \text{if } U_2^*(Y_4) > U_2^*(Y_3) \text{ and } U_1^*(Y_4) > U_1^*(Y_1) \end{cases} \tag{6}$$

Since we do not perfectly observe the true utilities, we can only make probabilistic statements concerning the outcomes. So, for example, the probability of Y_1 occurring is

$$\begin{aligned} p_{Y_1} &= \Pr[U_2^*(Y_3) > U_2^*(Y_4), U_1^*(Y_1) > U_1^*(Y_3)] + \\ &\quad \Pr[U_2^*(Y_4) > U_2^*(Y_3), U_1^*(Y_1) > U_1^*(Y_4)] \\ &= \Pr[U_2(Y_3) + \epsilon_{23} > U_2(Y_4) + \epsilon_{24}, U_1(Y_1) + \epsilon_{11} > U_1(Y_3) + \epsilon_{13}] + \\ &\quad \Pr[U_2(Y_4) + \epsilon_{24} > U_2(Y_3) + \epsilon_{23}, U_1(Y_1) + \epsilon_{11} > U_1(Y_4) + \epsilon_{14}] \end{aligned}$$

Denote the variance of ϵ_{ij} as $\sigma^2_{\epsilon_{ij}}$ and the covariance of ϵ_{ij} with ϵ_{ik} as $\sigma_{\epsilon_{ijk}}$. If we let $\eta_{ijk} = \epsilon_{ij} - \epsilon_{ik}$, then we can rewrite Equation 7 as

$$
\begin{aligned}
p_{Y_1} &= \Pr[\eta_{243} < U_2(Y_3) - U_2(Y_4), \eta_{131} < U_1(Y_1) - U_1(Y_3)] + \\
&\quad \Pr[\eta_{234} < U_2(Y_4) - U_2(Y_3), \eta_{141} < U_1(Y_1) - U_1(Y_4)] \\
&= \int_{-\infty}^{U_2(Y_3)-U_2(Y_4)} \int_{-\infty}^{U_1(Y_1)-U_1(Y_3)} \phi(\eta_{243}, \eta_{131}) d\eta_{131} d\eta_{243} + \\
&\quad \int_{-\infty}^{U_2(Y_4)-U_2(Y_3)} \int_{-\infty}^{U_1(Y_1)-U_1(Y_4)} \phi(\eta_{234}, \eta_{141}) d\eta_{141} d\eta_{233}. \quad (7)
\end{aligned}
$$

Both terms in Equation 7 require integration over standardized bivariate Normal densities. If we assume that the disturbances are independent of each other, then the equation further simplifies to

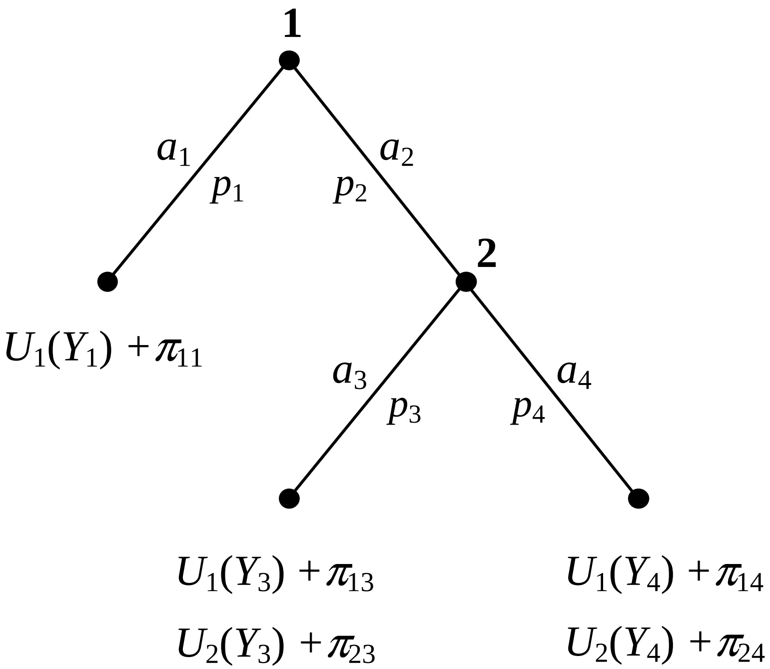
$$
\begin{aligned}
p_{Y_1} \;=\; & \Phi\left[\frac{U_2(Y_3)-U_2(Y_4)}{\sqrt{\sigma^2_{\epsilon_{23}}+\sigma^2_{\epsilon_{24}}}}\right]\Phi\left[\frac{U_1(Y_1)-U_1(Y_3)}{\sqrt{\sigma^2_{\epsilon_{11}}+\sigma^2_{\epsilon_{13}}}}\right]+ \\
& \Phi\left[\frac{U_2(Y_4)-U_2(Y_3)}{\sqrt{\sigma^2_{\epsilon_{24}}+\sigma^2_{\epsilon_{23}}}}\right]\Phi\left[\frac{U_1(Y_1)-U_1(Y_4)}{\sqrt{\sigma^2_{\epsilon_{11}}+\sigma^2_{\epsilon_{14}}}}\right].
\end{aligned}
\tag{8}
$$

Carrying out the same steps to calculate p_{Y_3} and p_{Y_4} gives

$$
p_{Y_3} \;=\; \Phi\left[\frac{U_2(Y_3)-U_2(Y_4)}{\sqrt{\sigma^2_{\epsilon_{23}}+\sigma^2_{\epsilon_{24}}}}\right]\Phi\left[\frac{U_1(Y_3)-U_1(Y_1)}{\sqrt{\sigma^2_{\epsilon_{13}}+\sigma^2_{\epsilon_{11}}}}\right]
\tag{9}
$$

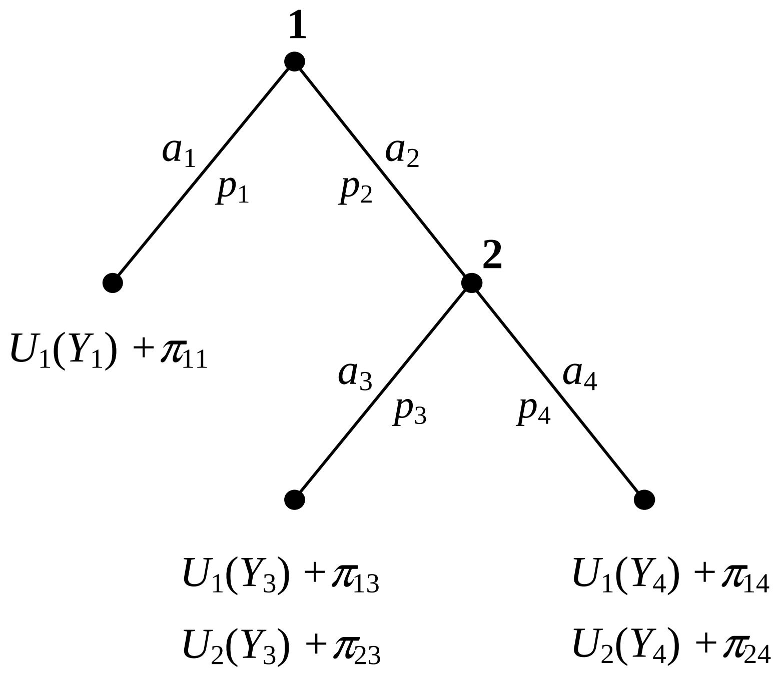
Private Information about Outcome Payoffs

To derive the choice probabilities, work "up the tree."



$$\begin{aligned} p_4 &= \Pr[U_2^*(a_4) > U_2^*(a_3)] \\ &= \Pr[U_2(Y_4) + \pi_{24} > U_2(Y_3) + \pi_{23}] \\ &= \Phi\left[\frac{U_2(Y_4) - U_2(Y_3)}{\sqrt{\sigma^2_{\pi_{24}} + \sigma^2_{\pi_{24}}}}\right] \end{aligned}$$

$$\begin{aligned} U_1^*(a_2) &= p_3 U_1^*(a_3) + p_4 U_1^*(a_4) \\ &= p_3[U_1(Y_3) + \pi_{13}] + p_4[U_1(Y_4) + \pi_{14}] > U_1(Y_1) + \pi_{11} \end{aligned}$$

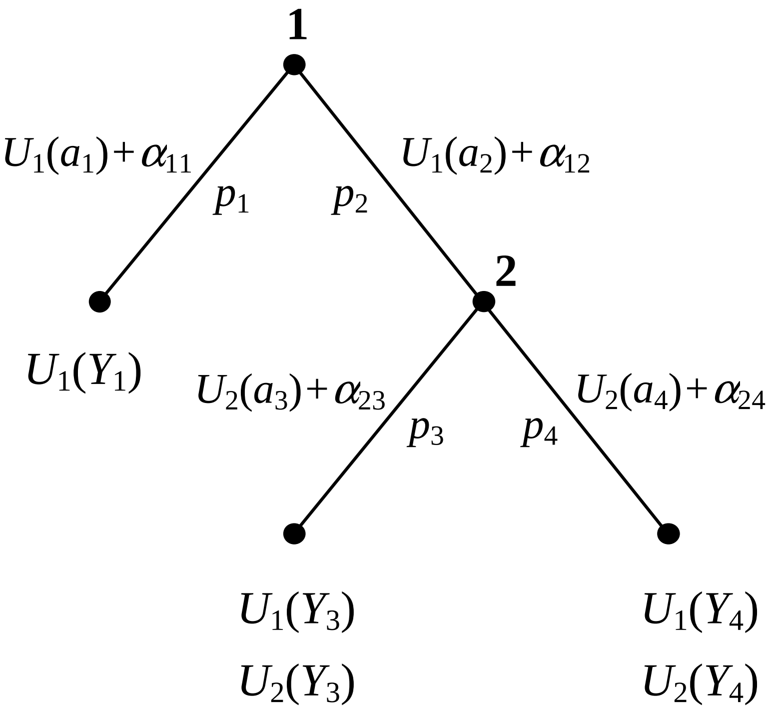


The equilibrium outcome probabilities are the product of the action probabilities along their paths.

$$\begin{aligned} p_{Y_1} &= p_1 \\ p_{Y_3} &= p_2 p_3 \\ p_{Y_4} &= p_2 p_4 \end{aligned}$$

$$\begin{aligned} p_2 &= \Pr[U_1^*(a_2) > U_1^*(a_1)] \\ &= \Pr[p_3 U_1^*(Y_3) + p_4 U_1^*(Y_4) > U_1^*(Y_1)] \\ &= \Pr[p_3(U_1(Y_3)+\pi_{13}) + p_4(U_1(Y_4)+\pi_{14}) > U_1(Y_1)+\pi_{11}] \\ &= \Phi\left[\frac{p_3 U_1(Y_3) + p_4 U_1(Y_4) - U_1(Y_1)}{\sqrt{p_3^2\sigma_{\pi_{13}}^2 + p_4^2\sigma_{\pi_{14}}^2 + \sigma_{\pi_{11}}^2}}\right] \end{aligned}$$

Agent Error



$$\begin{aligned} p_4 &= \Pr[U_2(Y_4) + \alpha_{24} > U_2(Y_3) + \alpha_{23}] \\ &= \Phi\left[\frac{U_2(Y_4) - U_2(Y_3)}{\sqrt{\sigma^2_{\alpha_{23}} + \sigma^2_{\alpha_{24}}}}\right] \end{aligned}$$

$$\begin{aligned} U_1^*(a_2) &= U_1(a_2) + \alpha_{12} \\ &= [p_3 U_1(a_3) + p_4 U_1(a_4)] + \alpha_{12} \end{aligned}$$

$$
\begin{aligned}
p_2 &= \Pr[U_1^*(a_2) > U_1^*(a_1)] \\
&= \Pr[p_3 U_1(a_3) + p_4 U_1(a_4) + \alpha_2 > U_1(a_1) + \alpha_1] \\
&= \Phi\left[\frac{p_3 U_1(Y_3) + p_4 U_1(Y_4) - U_1(Y_1)}{\sqrt{\sigma_{\alpha_1}^2 + \sigma_{\alpha_2}^2}}\right]
\end{aligned}
$$

The equilibrium outcome probabilities are, again, the product of the action probabilities along their paths.

$$
\begin{aligned}
p_{Y_1} &= p_1 \\
p_{Y_3} &= p_2 p_3 \\
p_{Y_4} &= p_2 p_4
\end{aligned}
$$

The Type of Uncertainty Matters... To a Degree

- theoretical sources of uncertainty sometimes generate different statistical models

- One implication is that, in some circumstances, we should be able to formulate tests to differentiate between the types of uncertainty.

- It also implies, however, that misspecifying the source of uncertainty may lead to incorrect inferences.

How do the probability models differ?

- How much will the agent error and payoff disturbance probabilities differ, given the same observed utilities $U_m(Y_k)$?

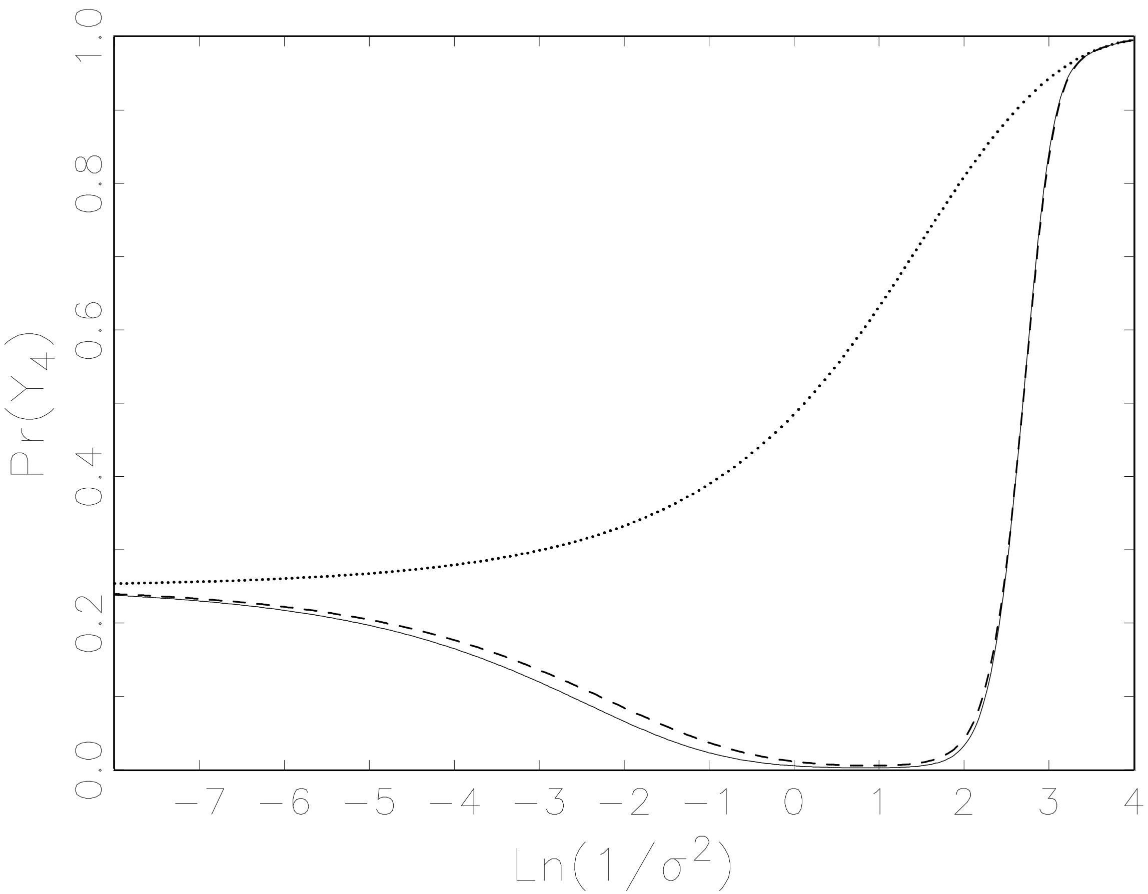
- Under certain conditions, they will not differ at all. All three strategic probit models have the same limiting conditions.

- For simplicity, let us continue to assume that the variance is constant across disturbances. As the variance σ^2 of the error terms approaches zero, the choice probabilities approach the complete information SPE.

- As σ^2 approaches infinity, the observed utilities offer no information to distinguish which actions are more likely, and the action probabilities at information sets become uniformly distributed.

- Where the models differ is in the choice probabilities between these limiting conditions. The differences in probabilities may at times be large, and will depend on the complexity of the model, on the payoffs, and on the "signal to noise" ratio of the regressors and disturbances.

$U_1(Y_1) = 0$
$U_1(Y_3) = -10$
$U_1(Y_4) = 1$
$U_2(Y_3) = 0$
$U_2(Y_4) = .5$



Bias and Inconsistency

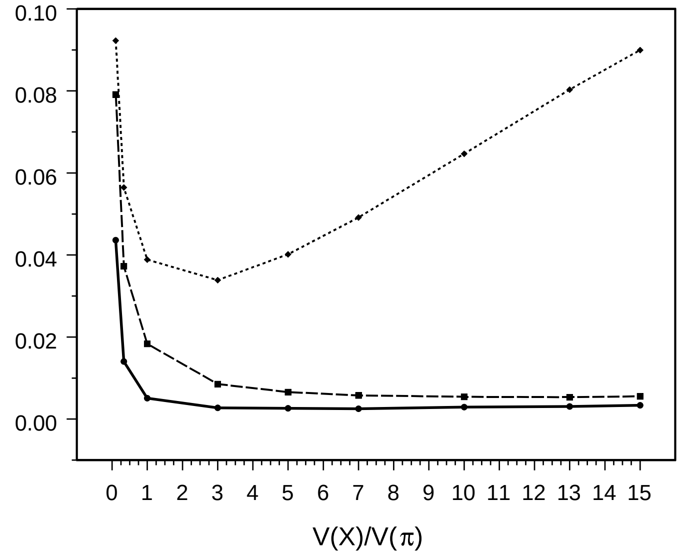
Monte Carlo simulations

- The utilities are specified as
  $U_1(Y_1) = 0, \quad U_1(Y_3) = \beta_{13}X_{13}, \quad U_1(Y_4) = \beta_{14}X_{14},$
  $U_2(Y_3) = 0, \quad U_2(Y_4) = \beta_{24}X_{24}.$

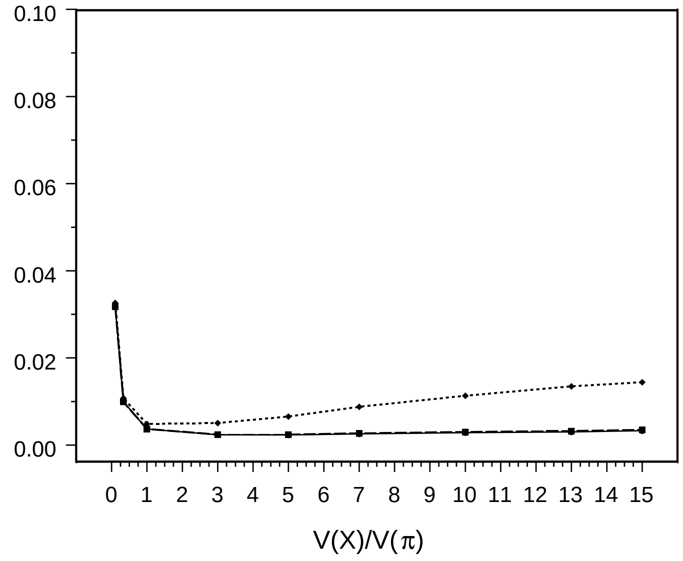
- The uncertainty is due to private information concerning payoffs, where the disturbances are Normally distributed with mean zero and variance $V(\pi)$

- The regressors are each uniformaly distributed with mean zero and variance $V(X)$.

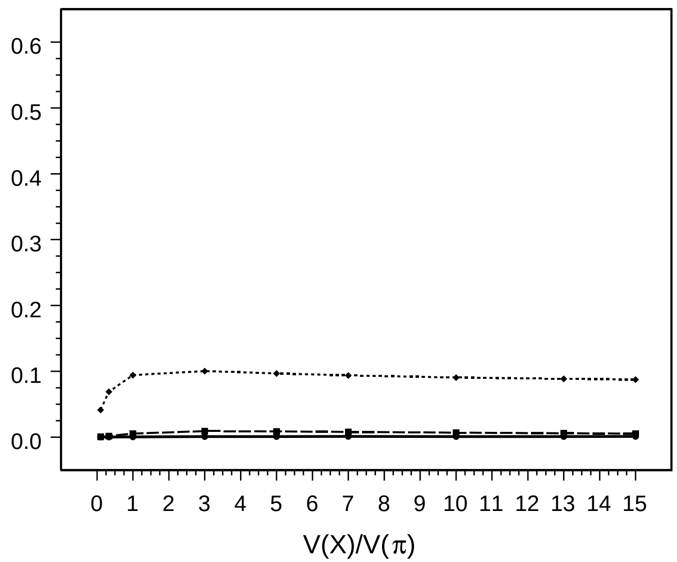
- signal-to-noise ratios $V(X)/V(\pi) \in \{\frac{1}{10}, \frac{1}{3}, 1, 3, 5, 7, 10, 13, 15\}$



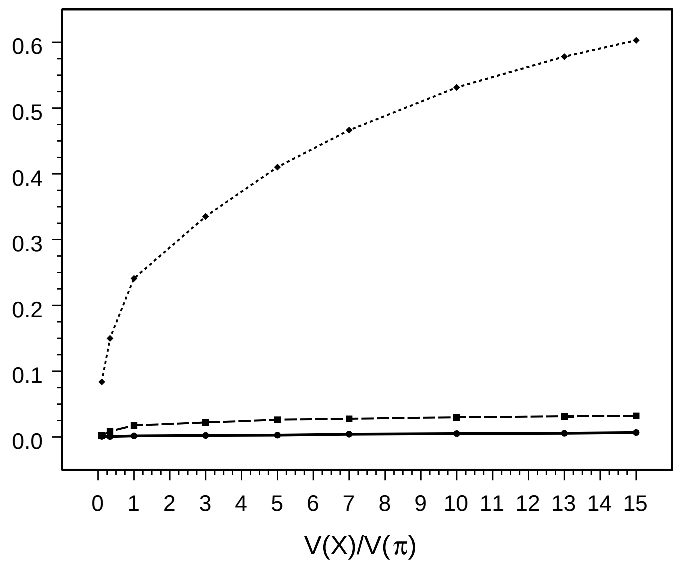
(a) Estimated MSE of $\widehat{\beta}_{14}$



(b) Estimated MSE of $\widehat{\beta}_{24}$



(c) Mean Distance



(d) Max Distance

Concluding Remarks

- examine how three plausible sources of uncertainty interact with structural assumptions to produce different statistical models and possibly affect our inferences

- demonstrate how to derive such statistical models from theoretical first principles

- generalize a broad class of statistical discrete choice models