Applications of Symmetries to Econometrics: Case Studies

Vladik Kreinovich

University of Texas at El Paso
El Paso, Texas 79968, USA
vladik@utep.edu

largely based on a joint work with
Thongchai Dumrongpokaphan
Department of Mathematics, Faculty of Science
Chiang Mai University, Thailand, tcd43@hotmail.com

1. Introduction

- In physics, the results of many processes can be accurately predicted:
  - by appropriate mathematical models
  - such as partial differential equations.
- In contrast, in economics, there are no *general* models capable of accurate prediction.
- Because of the importance of economics, many *specific* models have been developed.
- However, practitioners are often reluctant to use these (mostly empirical) models.
- World economy is dynamic, constantly changing.
- So, practitioners are wary to extrapolate current empirical dependencies to the future.

2. Introduction (cont-d)

- To make these models more reliable, it is desirable to come up with their theoretical justification.
- In physics, such a justification often comes from:
  - considering appropriate symmetries,
  - and showing that the empirical model is the only one invariant w.r.t. such symmetries.
- This is how most fundamental physical equations can be derived: Maxwell's, Einstein's, Schroedinger's.
- In this talk, on several examples, we show that:
  - in many cases,
  - a similar symmetry-based approach can also help to justify empirical models in econometrics.

Part I
Econometric Models of Probabilistic Choice: Beyond McFadden's Formulas

3. Traditional (Deterministic Choice) Approach to Decision Making

- In the traditional approach to decision making, we assume that for every two alternatives a and b:
  - either the decision maker always prefers a,
  - or the decision maker always prefers b,
  - or, to the decision maker, a and b are equivalent.
- Then, decision maker's preferences can be described by *utilities* defined as follows.
- We select two alternatives which are not present in the original choices:
  - a very bad alternative a_0, and
  - a very good alternative a_1.
- Then, each actual alternative a is better than a_0 and worse than a_1: $a_0 < a < a_1$.

4. Traditional Decision Making (cont-d)

- To gauge the quality of the alternative a to the decision maker, we can consider lotteries $L(p)$ in which:
  - we get a_1 with probability p and
  - we get a_0 with the remaining probability $1 - p$.
- For every p, we either have $L(p) < a$ or $a < L(p)$ or $L(p) \sim a$.
- When $p = 1$, $L(1) = a_1$, thus $a < L(1)$.
- When $p = 0$, $L(0) = a_0$, thus $L(0) < a$.
- Clearly, the larger the probability p of the very good outcome, the better the lottery; thus, if $p < p'$, then:
  - $a < L(p)$ implies $a < L(p')$, and
  - $L(p') < a$ implies $L(p) < a$.

5. Traditional Decision Making (cont-d)

- Therefore, we can conclude that

$$\sup\{p : L(p) < a\} = \inf\{p : a < L(p)\}.$$

- $u(a) \stackrel{\text{def}}{=} \sup\{p : L(p) < a\} = \inf\{p : a < L(p)\}$ has the following properties:
  - if $p < u$, then $L(p) < a$; and
  - if $p > a$, then $a < L(p)$.
- In particular, for every small $\varepsilon > 0$, we have

$$L(u(a) - \varepsilon) < a < L(u(a) + \varepsilon).$$

- So, a is "equivalent" to the lottery $L(p)$ in which a_1 is selected with the probability $p = u(a)$: $a \equiv L(u(a))$.
- This probability $u(a)$ is what is known as *utility*.
- Once we know all the utility values, we select the alternative with the largest utility.

6. Traditional Decision Making (final)

- Indeed, as we have mentioned, $p < p'$ implies that $L(p) < L(p')$, so when $u(a) < u(b)$, we have

$$a \equiv L(u(a)) < L(u(b)) \equiv b \text{ and thus } a < b.$$

- The above definition of utility depends on the choice of two alternatives a_0 and a_1.
- If we select different a'_0 and a'_1, then, as one can show, we get $u'(a) = k \cdot u(a) + \ell$ for some $k > 0$ and ℓ.
- Thus, *utility is defined modulo linear transformation.*

7. Actual Choices Are Often Probabilistic

- People sometimes make different choices when repeatedly presented with the same alternatives a and b.
- This is especially true when the compared alternatives a and b are close in value.
- In such situations, we would like to predict the probability $P(a, A)$ of a from a set A.
- We can still have a deterministic distinction: $b > a$ if the person selects a more frequently than b:

$$P(a, b) > 0.5.$$

- Based on $>$, we can determine the utilities $u(a)$.
- It is reasonable to assume that $P(a, A)$ depends only on the utilities $u(a), \ldots, u(b)$.

8. McFadden's Formulas for Probabilistic Selection

- The 2001 Nobelist D. McFadden proposed the following formula for the desired probability $P(a, A)$:

$$P(a, A) = \frac{\exp(\beta \cdot u(a))}{\sum\limits_{b \in A} \exp(\beta \cdot u(b))}.$$

- In many practical situations, this formula indeed describes people's choices really well.
- In some case, alternative formulas provide a better explanation of the empirical choices.
- In this talk, we use natural symmetries to come up with an appropriate generalization of McFadden's formulas.

9. Analysis of the Problem

- We may have many different alternatives a, b, ...
- In some cases, we prefer a, in other cases, we prefer b.
- It is reasonable to require that:
  - once we have decided on selecting either a or b, then
  - the relative frequency of selecting a should be the same as when we simply select between a and b:

$$\frac{P(a,A)}{P(b,A)} = \frac{P(a,b)}{P(b,a)} = \frac{P(a,b)}{1-P(a,b)}.$$

- Let us add a new alternative a_n to our list, then:

$$\frac{P(a,A)}{P(a_n,A)} = \frac{P(a,a_n)}{1-P(a,a_n)}, \text{ so } P(a,A) = P(a_n,A) \cdot f(a),$$

$$\text{where } c \stackrel{\text{def}}{=} P(a_n,A) \text{ and } f(a) \stackrel{\text{def}}{=} \frac{P(a,a_n)}{1-P(a,a_n)}.$$

10. Analysis of the Problem (cont-d)

- c can be found from the condition that one of $b \in A$ will be selected: $\sum\limits_{b\in A} P(b, A) = 1$, so $P(a, A) = \dfrac{f(a)}{\sum\limits_{b\in A} f(b)}$.
- We assumed that the probabilities depend only on the utilities $u(a)$.
- We thus conclude that $f(a)$ must depend only on the utilities: $f(a) = F(u(a))$ for some $F(u)$, and

$$P(a, A) = \frac{F(u(a))}{\sum\limits_{b\in A} F(u(b))}.$$

- Thus, all we need is to find an appropriate function $F(u)$.

11. Properties of $F(u)$

- The better the alternative a, i.e., the larger $u(a)$, the higher should be the probability $P(a, A)$.
- Thus, $F(u)$ is an increasing function of the utility u.
- If we multiply all the values of $F(u)$ by a constant, we will get the exact same probabilities.
- Utilities are defined modulo a general linear transformation.
- In particular, it is possible to add a constant to all the utility values $u(a) \to u'(a) = u(a) + c$.
- Since this shift does not change the preferences, it is reasonable to require that for $u'(a)$, we get the same probabilities.
- Using new utility values $u'(a) = u(a) + c$ means that we replace $F(u(a))$ with $F(u'(a)) = F(u(a) + c)$.

12. Deriving McFadden's Formula

- Using new utility values $u'(a) = u(a) + c$ means that we replace $F(u(a))$ with $F(u'(a)) = F(u(a) + c)$.
- This is equivalent to using the *original* utility values but with a *new function* $F'(u) \stackrel{\text{def}}{=} F(u + c)$.
- The functions $F(u)$ and $F'(u)$ describe the same probabilities if and only if $F'(u) = C \cdot F(u)$ for some C.
- So, $F(u + c) = C(c) \cdot F(u)$ for some $C(c)$.
- It is known that every monotonic solution to this function equation has the form $F(u) = C_0 \cdot \exp(\beta \cdot u)$.
- This is exactly McFadden's formula.

13. Discussion

- The proof of the function-equation result is somewhat complicated.
- However, under a natural assumption that $F(u)$ is differentiable, this result can be proven rather easily.
- $C(u)$ is a ratio of two differentiable functions $F(u+c)$ and $F(u)$, and is, thus, differentiable.
- Since $F(u)$ and $C(c)$ are differentiable, we can differentiate both sides of the equality by c and take $c=0$.
- As a result, we get $\dfrac{dF}{du} = \beta \cdot F$, where $\beta \stackrel{\text{def}}{=} \dfrac{dC}{dc}_{|c=0}$.
- By moving all the terms with F to one side and all others to the other side, we get: $\dfrac{dF}{F} = \beta \cdot du$.
- Integrating both sides, we get $\ln(F) = \beta \cdot u + C_1$, so $F(u) = \exp(\ln(F)) = C_0 \cdot \exp(\beta \cdot u)$, w/$C_0 \stackrel{\text{def}}{=} e^{C_1}$.

14. Main Idea of This Case Study

- Multiplying all the utility values by a constant is also a legitimate transformation for utilities.
- However, this *does* change McFadden's probabilities.
- So, we cannot require that the probability formula not change for *all* possible linear transformations of utility:
  - once we require shift-invariance,
  - we get McFadden's formula
  - which is not scale-invariant.
- So, we should require invariance with respect to *some* family of re-scalings.

15. Main Idea (cont-d)

- If a formula does not change when we apply each transformation, it will also not change:
  - if we apply them one after another,
  - i.e., if we consider a composition of transformations.
- Each shift can be represented as a superposition of many small (infinitesimal) shifts $u \to u + B \cdot dt$.
- Similarly, each scaling can be represented as a superposition of many small scalings $u \to (1 + A \cdot dt) \cdot u$.
- Thus, it is sufficient to consider invariance with respect to an infinitesimal transformation

$$u \to u' = (1 + A \cdot dt) \cdot u + B \cdot dt.$$

16. Main Idea (cont-d)

- Invariance means that the values $F(u')$ lead to the same probabilities as the original values $F(u)$, so:

$$F(u + (A \cdot u + B) \cdot dt) = F(u) + C \cdot F(u) \cdot dt.$$

- Here, by definition of the derivative, $F(u + q \cdot dt) = F(u) + \frac{dF}{du} \cdot q \cdot dt$, so $(A \cdot u + B) \cdot \frac{dF}{du} = C \cdot F(u)$.
- We can separate the variables by moving all the terms with F to one side and all the terms with u to another:

$$\frac{dF}{F} = C \cdot \frac{du}{A \cdot u + b}.$$

- $A = 0$ leads to McFadden's formulas; when $A \neq 0$, then for $x \stackrel{\text{def}}{=} u + \frac{B}{A}$, $\frac{dF}{F} = c \cdot \frac{dx}{x}$, w/$c \stackrel{\text{def}}{=} \frac{C}{A}$.
- Integration leads to $\ln(F) = c \cdot \ln(x) + C_0$, thus $F = C_1 \cdot x^c$ for $C_1 \stackrel{\text{def}}{=} \exp(C_0)$, and $F(u) = C_1 \cdot (u + k)^c$.

17. Conclusions and Discussion

- In addition to the original McFadden's formula, we now have another option $P(a, A) = \dfrac{(u(a)+k)^c}{\sum\limits_{b\in A}(u(b)+k)^c}$.
- This is in good accordance with empirical data.
- This formula is a generalization of McFadden's.
- Indeed, $\exp(\beta\cdot u) = \lim\limits_{n\to\infty}\left(1+\dfrac{\beta\cdot u}{n}\right)^n$, so for large n, $\exp(\beta\cdot u)$ is indistinguishable from
$$\left(1+\frac{\beta\cdot u}{n}\right)^n = \left(\frac{\beta}{n}\right)^n\cdot(u+k)^c \text{ for } c=n, k=\frac{n}{\beta}.$$
- So, instead of a 1-parametric McFadden's formula, we now have a 2-parametric formula.
- We can use this additional parameter to get an even more accurate description of the probabilistic choice.

Part II

How to Explain Ubiquity of Constant Elasticity of Substitution (CES) Production and Utility Functions Without Explicitly Postulating CES

18. Outline

- The dependence of production on various factors is often described by CES functions.
- These functions are usually explained by postulating two requirements:
  - that the formulas should not change if we change a measuring unit, and
  - a less convincing CES requirement.
- In this paper, we show that the CES requirement can be replaced by a more convincing requirement:
  - that the combined effect of all the factors
  - should not depend on the order in which we combine these factors.

19. CES Production Functions and CES Utility Functions Are Ubiquitous

- Most observed data about production y is well described by the *CES production function*

$$y = \left(\sum_{i=1}^{n} a_i \cdot x_i^r\right)^{1/r}.$$

- Here x_i are the numerical measures of the factors that influence production, such as:
  - amount of capital,
  - amount of labor, etc.
- A similar formula describes how the person's utility y depends on different factors x_i such as:
  - amounts of different types of consumer goods,
  - utilities of other people, etc.

20. How This Ubiquity Is Explained Now

- The current explanation for the empirical success of CES function is based on two requirements.
- The first requirement is that the corresponding function $y = f(x_1, \ldots, x_n)$ is *homogeneous*:

$$f(\lambda \cdot x_1, \ldots, \lambda \cdot x_n) = \lambda \cdot f(x_1, \ldots, x_n).$$

- Meaning: we can describe different factors by using different monetary units.
- The results should not change if we replace the original unit by a one which is λ times smaller.
- After this replacement, the numerical value of each factor changes from x_i to $\lambda \cdot x_i$ and y is replace by $\lambda \cdot y$.
- So, we get exactly the above requirement.

21. Second Requirement

- The second requirement is that $f(x_1, \ldots, x_n)$ should provide *constant elasticity of substitution* (CES).
- The requirement is easier to explain for the case of two factors $n = 2$.
- In this case, this requirement deals with "substitution" situations in which:
  - we change x_1 and then
  - change the original value x_2 to the new value $x_2(x_1)$
  - so that the overall production or utility remain the same.
- The corresponding substitution rate can then be calculated as $s \stackrel{\text{def}}{=} \dfrac{dx_2}{dx_1}$.

22. Second Requirement (cont-d)

- The substitution function $x_2(x_1)$ is explicitly defined by the equation $f(x_1, x_2(x_1)) = \text{const}$, then

$$s = -\frac{f_{,1}(x_1, x_2)}{f_{,2}(x_1, x_2)}, \text{ where } f_{,i}(x_1, x_2) \stackrel{\text{def}}{=} \frac{\partial f}{\partial x_i}(x_1, x_2).$$

- The requirement is that:
  - for each percent of the change in ratio $\dfrac{x_2}{x_1}$,
  - we get the same constant number of percents change in s: $\dfrac{ds}{d\left(\dfrac{x_2}{x_1}\right)} = \text{const}$.
- *Problem:* the CES condition is too mathematical to be convincing for economists.
- *We provide:* more convincing arguments.

23. Main Idea Behind a New Explanation

- In our explanation, we will use the fact that in most practical situations, we combine several factors.
- We can combine these factors in different order. For example:
  - we can first combine the effects of capital and labor into a single characteristic,
  - and then combine it with other factors.
- Alternatively:
  - we can first combine capital with other factors,
  - and only then combine the resulting combined factor with labor, etc.
- The result should not depend on the order in which we perform these combinations.
- We show that this idea implies the CES functions.

24. Derivation of the CES Functions from the Above Idea

- Let us denote a function that combines factors i and j into a single quantity x_{ij} by $f_{i,j}(x_i, x_j)$.
- Similarly, let's denote a function that combines x_{ij} and $x_{k\ell}$ into a single quantity $x_{ijk\ell}$ by $f_{ij,k\ell}(x_{ij}, x_{k\ell})$.
- In these terms, the requirement that the resulting values do not depend on the order means that

$$f_{12,34}(f_{1,2}(x_1, x_2), f_{3,4}(x_3, x_4)) = f_{13,24}(f_{1,3}(x_1, x_3), f_{2,4}(x_2, x_4)).$$

- In both production and utility situations, for each i and j, $f_{i,j}(x_i, x_j)$ is increasing in x_i and x_j.
- It is also reasonable to require that:
  - the function $f_{i,j}(x_i, x_j)$ is continuous, and
  - when one of the factors tends to infinity, the result also tends to infinity.

25. Derivation (cont-d)

- Under these assumptions, $f(a,b)$ is *invertible*:
  - for every $a \in A$ and for every $c \in C$, there exists a unique value $b \in B$ for which $c = f(a,b)$;
  - for every $b \in B$ and for every $c \in C$, there exists a unique value $a \in A$ for which $f(a,b) = c$.
- It is known that:
  - for every set of invertible operations that satisfy the generalized associativity requirement,
  - there exists an Abelian group G and 1-1 mappings $r_i : X_i \to G$, $r_{ij} : X_{ij} \to G$ and $r_X : X \to G$
  - for which, for all $x_i \in X_i$ and $x_{ij} \in X_{ij}$, we have

$$f_{ij}(x_i, x_j) = r_{ij}^{-1}(g(r_i(x_i), r_j(x_j))) \text{ and}$$

$$f_{ij,kl}(x_{ij}, x_{k\ell}) = r_X^{-1}(g(r_{ij}(x_{ij}), r_{k\ell}(x_{k\ell}))).$$

26. Groups and Abelian Groups: Reminder

- A set G with an associative operation $g(a, b)$ and a unit element e ($g(a, e) = g(e, a) = a$) is called a *group*
  - if every element is invertible, i.e.,
  - if for every a, there exists an a' for which
$$g(a, a') = e.$$
- A group in which the operation $g(a, b)$ is commutative is known as *Abelian*.

27. Discussion

- All continuous 1-D Abelian groups with order-preserving operations are isomorphic to $(\mathbb{R}, +)$.
- Here, $(\mathbb{R}, +)$ is the additive group of real numbers, with
$$g(a, b) = a + b.$$
- Thus, we can conclude that all combining operations have the form
$$f_{ij}(x_i, x_j) = r_{ij}^{-1}(r_i(x_i) + r_j(x_j)).$$
- Equivalently, $f_{ij}(x_i, x_j) = y$ means that
$$r_{ij}(y) = r_i(x_i) + r_j(x_j).$$

28. Let Us Use Homogeneity

- We will now prove that homogeneity leads exactly to the desired CES combinations.
- Homogeneity means that if $r_{ij}(y) = r_i(x_i)+r_j(x_j)$, then for every λ: $r_{ij}(\lambda \cdot y) = r_i(\lambda \cdot x_i) + r_j(\lambda \cdot x_j)$.
- For each $x'_i = x_i + \Delta x_i$, let us find $x'_j = x_j + \Delta x_j$ for which $r_i(x_i) + r_j(x_j)$ remains the same (thus, the combined value y remains the same):

$$r_i(x'_1)+r_j(x'_j) = r_i(x_i+\Delta x_i)+r_j(x_j+\Delta x_j) = r_i(x_i)+r_j(x_j).$$

- For small Δx_i, $\Delta x_j = k \cdot \Delta x_i + o(\Delta x_i)$ for some k.
- Here, $r_i(x_i+\Delta x_i) = r_i(x_i)+r'_i(x_i)\cdot\Delta x_i+o(\Delta x_i)$, where, as usual, f' denotes the derivative.
- Similarly, $r_j(x_j + \Delta x_j) = r_j(x_j + k \cdot \Delta x_i + o(\Delta x_i)) =$

$$r_j(x_j) + k \cdot r'_j(x_j) \cdot \Delta x_i + o(\Delta x_i).$$

29. Let Us Use Homogeneity (cont-d)

- Thus, the above condition takes the form

$$r_i(x_i)+r_j(x_j)+(r_i'(x_i)+k\cdot r_j'(x_j))\cdot\Delta x_i+o(\Delta x_i) = r_i(x_i)+r_j(x_j).$$

- Thus, $r_i'(x_i)+k\cdot r_j'(x_j)=0$, and $k=-\dfrac{r_i'(x_i)}{r_j'(x_j)}$.

- For re-scaled values, we similarly get $k=-\dfrac{r_i'(\lambda\cdot x_i)}{r_j'(\lambda\cdot x_j)}$,

  so $\dfrac{r_i'(\lambda\cdot x_i)}{r_i'(x_i)}=\dfrac{r_j'(\lambda\cdot x_j)}{r_j'(x_j)}$.

- The right-hand side does not depend on x_i, the left-hand side does not depend on x_j, so:

$$\frac{r_i'(\lambda\cdot x_i)}{r_i'(x_i)} = c(\lambda) \text{ for some } c(\lambda).$$

30. Let Us Use Homogeneity (cont-d)

- Thus, the derivative $R_i(x_i) \stackrel{\text{def}}{=} r_i'(x_i)$ satisfies the functional equation:

$$R_i(\lambda \cdot x_i) = R_i(x_i) \cdot c(\lambda) \text{ for all } \lambda \text{ and } x_i.$$

- It is known that every continuous solution to this equation has the form $r_i'(x_i) = R_i(x_i) = A_i \cdot x_i^{\alpha_i}$.
- For differentiable functions, this can be proven if we differentiate both sides by c and take $c = 1$.
- Then, we get $x_i \cdot \dfrac{dR_i}{dc_i} = c \cdot R_i$.
- Separating variables, we get $\dfrac{dR_i}{R_i} = c \cdot \dfrac{dx_i}{x_i}$.
- Integration leads to $\ln(R_i) = c \cdot \ln(x_i) + C_1$ and thus, to the desired formula $r_i'(x_i) = A_i \cdot x_i^{\alpha_i}$.

31. Let Us Use Homogeneity (cont-d)

- Integrating the above expression for $r_i'(x_i)$, we get $r_i(x_i) = a_i \cdot x_i^{\beta_i} + C_i$ and similarly, $r_j(x_j) = a_j \cdot x_j^{\beta_j} + C_j$.
- One can easily check that homogeneity implies that $\beta_i = \beta_j$ and $C_i + C_j = 0$, so

$$r_i(x_i) + r_j(x_j) = a_i \cdot x_i^r + a_j \cdot x_j^r.$$

- By considering a similar substitution between x_i and y, we conclude that $r_{ij}(y) = \text{const} \cdot y^r$.
- So, we indeed get the desired formula

$$r_{ij}(x_i, x_j) = (a_i \cdot x_i^r + a_j \cdot x_j^r)^{1/r}.$$

- By using similar formulas to combine x_{ij} with x_k, etc., we get the desired CES combination function.

32. Possible Application to Copulas

- A 1-D probability distribution of a random variable X can be described by its cdf $F_X(x) \stackrel{\text{def}}{=} \text{Prob}(X \le x)$.
- A 2-D distribution of a random vector (X, Y) can be similarly described by its 2-D cdf

$$F_{XY}(x, y) = \text{Prob}(X \le x \,\&\, Y \le y).$$

- It turns out that we can always describe $F(x, y)$ as

$$F_{XY}(x, y) = C_{XY}(F_X(x), F_Y(y)) \text{ for some } C_{XY}(a, b).$$

- This function C_{XY} is called a *copula*.
- For a joint distribution of several random variables X, Y, ..., Z, we can similarly write

$$F_{XY\ldots Z}(x, y, \ldots, z) \stackrel{\text{def}}{=} \text{Prob}(X \le x \,\&\, Y \le y \,\&\, \ldots \,\&\, Z \le z) =$$
$$C_{XY\ldots Z}(F_X(x), F_Y(y), \ldots, F_Z(z)) \text{ for some } C_{XY\ldots Z}.$$

33. Copulas (cont-d)

- When we have many ($n \gg 1$) random variables, then we need to describe a function of n variables.
- Even if we use two values for each variable, we get 2^n combinations.
- For large n this is astronomically large.
- Thus, a reasonable idea is to approximate the multi-D distribution.
- A reasonable way to approximate is to use 2-D copulas.

34. Copulas (cont-d)

- For example, to describe a joint distribution of three variables X, Y, and Z:
  - we first describe the joint distribution of X and Y as $F_{XY}(x, y) = C_{XY}(F_X(x), F_Y(y))$,
  - and then use a copula $C_{XY,Z}$ to combine it with $F_Z(z)$: $F_{XYZ}(x, y, z) \approx C_{XY,Z}(F_{XY}(x, y), F_Z(z)) =$

$$C_{XY,Z}(C_{XY}(F_X(x), F_Y(y), F_Z(z)).$$

- Such an approximation, when copulas are applied to one another like a vine, are known as *vine copulas*.
- It is reasonable to require that:
  - the result of the vine copula approximation
  - should not depend on the order in which we combine the variables.

35. Copulas (cont-d)

- In particular, for X, Y, Z, and T, we should get the same result in the following two situations:
  - if we first combine X with Y, Z and T, and then combine the two results; or
  - if we first combine X with Z, Y with T, and then combine the two results.
- Thus, we require that for all possible real numbers x, y, z, and t, we get

$$C_{XY,ZT}(C_{XY}(F_X(x),F_Y(y)),C_{ZT}(F_Z(z),F_T(t)))=$$

$$C_{XZ,YT}(C_{XZ}(F_X(x),F_Z(z)),C_{YT}(F_Y(y),F_T(t))).$$

- If we denote $a=F_X(x)$, $b=F_Y(y)$, $c=F_Z(z)$, and $d=F_T(t)$, we conclude that for every a, b, c, and d:

$$C_{XY,ZT}(C_{XY}(a,b),C_{ZT}(c,d))=C_{XZ,YT}(C_{XZ}(a,c),C_{YT}(b,d)).$$

36. Copulas: Conclusion

- We have:

$$C_{XY,ZT}(C_{XY}(a,b), C_{ZT}(c,d)) = C_{XZ,YT}(C_{XZ}(a,c), C_{YT}(b,d)).$$

- This is exactly the above generalized associativity requirement; thus:
  - if we extend copulas to invertible operations,
  - then we can conclude that copulas can be *re-scaled to associative operations* $\circ$:

$$C(a,b) = f(g(a) \circ h(b)) \text{ for some } f, g, h.$$

Part III

Why Cannot We Have a Strongly Consistent Family of Skew Normal (and Higher Order) Distributions

37. Formulation of the Problem

- Often, the only information that we have about the probability distribution is its first few moments.
- Many statistical techniques requires us to select a single distribution.
- It is therefore desirable to select,
  - out of all possible distributions with these moments,
  - a single "most representative" one.
- When we know the first two moments, a natural idea is to select a normal distribution.
- This selection is *strongly consistent* in the sense that:
  - if a random variable is a sum of several ones,
  - and we select normal distribution for all of them,
  - then the sum is also normally distributed.

38. Need for Strong Consistency

- Often, the random variable of interest has several components.
- For example, an overall income consists of salaries, pensions, unemployment benefits, interest, etc.
- Each of these categories, in its turn, can be subdivided into more subcategories.
- If for each of these categories, we only know the first moments, then we can apply the selection:
  - either to the overall sum,
  - or separately to each term.
- It seems reasonable to require that the resulting distribution for the overall sum should be the same.

39. What We Do in This Case Study

- When we know three moments, there is also a widely used selection – a skew-normal distribution.
- However, this selection is not strongly consistent in the above sense.
- In this talk, we show that this absence of strong consistency:
  - is not a fault of a specific selection but a general feature of the problem;
  - namely, for third and higher order moments, no strongly consistent selection is possible.

40. Skew Normal Distributions

- In addition to the first two moments μ and M_2, we may also know the third moment M_3.
- This can be described by the mean μ, the variance $V = \sigma^2$, and the third central moment $m_3 \stackrel{\text{def}}{=} E[(X-\mu)^3]$.
- There is a widely used selection, called *skew normal*:

$$\rho(x) = \frac{1}{2\omega} \cdot \phi\left(\frac{x-\eta}{\omega}\right) \cdot \Phi\left(\alpha \cdot \frac{x-\eta}{\omega}\right), \text{ where}$$

$$\rho(x) = \frac{1}{2\omega} \cdot \exp\left(-\frac{x^2}{2}\right), \text{ and } \Phi(x) = \int_{-\infty}^{x} \phi(t)\,dt.$$

- Here, $\mu = \eta + \omega \cdot \delta \cdot \sqrt{\frac{2}{\pi}}$, where $\delta \stackrel{\text{def}}{=} \frac{\alpha}{\sqrt{1+\alpha^2}}$,

$$\sigma^2 = \omega^2 \cdot \left(1 - \frac{2\delta^2}{\pi}\right), \text{ and } m_3 = \frac{4-\pi}{2} \cdot \sigma^3 \cdot \frac{(\delta \cdot \sqrt{2/\pi})^3}{(1-2\delta^2/\pi)^{3/2}}.$$

41. Analysis of the Problem

- We want to assign, to each triple (μ, V, m_3), a probability distribution $\rho(x, \mu, V, m_3)$.
- Let us list the natural properties of this assignment.
- Moments are rarely known exactly, we usually know them with some accuracy.
- It is reasonable to require that if the moments change slightly, then $\rho(x, \mu, V, m_3)$ should not change much.
- In other words, it is reasonable to require that the function $\rho(x, \mu, V, m_3)$ is continuous.
- *Comment:* in our proof, we will only use that $\rho(x)$ is measurable.
- *Strong consistency:* if X_1 and X_2 are independent, $X_1 \sim \rho(x, \mu_1, V_1, m_{31})$, and $X_2 \sim \rho(x, \mu_2, V_2, m_{32})$, then
$$X_1 + X_2 \sim \rho(x, \mu_1 + \mu_2, V_1 + V_2, m_{31} + m_{32}).$$

42. Scale Invariance (Reminder)

- Numerical values of different quantities depend on the choice of a measuring unit.
- E.g.: income can be described in Baht or in dollars.
- If we change the unit to λ times smaller one, then:
  - the actual incomes will not change,
  - but the numerical values will change $x \to x' = \lambda \cdot x$.
- If we perform the selection in the original units, then we get $\rho(x, \mu, V, m_3)$.
- If we simply re-scale x to $x' = \lambda \cdot x$, then for x', we get a new distribution $\rho'(x') = \dfrac{1}{\lambda} \cdot \rho\left(\dfrac{x'}{\lambda}, \mu, V, m_3\right)$.

43. Scale Invariance (cont-d)

- If we re-scale $\rho(x, \mu, V, m_3)$, we get

$$\rho'(x') = \frac{1}{\lambda} \cdot \rho\left(\frac{x'}{\lambda}, \mu, V, m_3\right).$$

- We should get the exact same distribution if we make a selection *after* the re-scaling, i.e., for

$$\mu' = \lambda \cdot \mu, \quad V' = \lambda^2 \cdot V, \quad m_3' \to \lambda^3 \cdot m_3.$$

- In the new units, we get $\rho(x', \lambda \cdot \mu, \lambda^2 \cdot V, \lambda^3 \cdot m_3)$.
- A natural requirement is that the resulting selection should be the same:

$$\frac{1}{\lambda} \cdot \rho\left(\frac{x'}{\lambda}, \mu, V, m_3\right) = \rho(x', \lambda \cdot \mu, \lambda^2 \cdot V, \lambda^3 \cdot m_3).$$

44. Definitions

- We say that a tuple (μ, V, m_3) is *possible* if there exists a distr. with mean μ, variance V, and moment m_3.
- By a *3-selection*, we mean a measurable mapping $\rho(x, \mu, V, m_3)$ defined for all possible tuples.
- We say that a 3-selection is *strongly consistent* if $X_i \sim \rho(x, \mu_i, V_i, m_{3i})$ for independent X_i implies

$$X_1 + X_2 \sim \rho(x, \mu_1 + \mu_2, V_1 + V_2, m_{31} + m_{32}).$$

- We say that a 3-selection is *scale-invariant* if for every possible tuple (μ, V, m_3), for every $\lambda > 0$ and x':

$$\frac{1}{\lambda} \cdot \rho\left(\frac{x'}{\lambda}, \mu, V, m_3\right) = \rho(x', \lambda \cdot \mu, \lambda^2 \cdot V, \lambda^3 \cdot m_3).$$

45. Main Result

- **Proposition.** *No 3-selection is strongly consistent and scale-invariant.*
- A similar result can be formulated for the case when we also know higher order moments.
- In this case, instead of the original moments, we can consider *cumulants* κ_n.
- Cumulants are terms at $\dfrac{\mathrm{i}^n \cdot t^n}{n!}$ in the Taylor expansion of $\ln(E[\exp(\mathrm{i} \cdot t \cdot X)])$.
- For $n = 1$, $n = 2$, and $n = 3$, we get exactly the mean, the variance, and the central third moment.
- Cumulants are additive: if $X = X_1 + X_2$ and X_1 and X_2 are independent, then $\kappa_n(X) = \kappa_n(X_1) + \kappa_n(X_2)$.

46. Discussion

- Since we cannot make a strongly consistent selection, what should we do?
- min and max are also natural operations in many applications; for example, in econometrics:
  - if there are several ways to invest money with the same level of risk,
  - then an investor selects the one that leads to the largest interest rate.
- From this viewpoint, it is reasonable to consider minima and maxima of normal variables.
- In some cases, these minima and maxima are distributed according to the skew normal distribution.
- This may be an additional argument in favor of using these distributions.

47. Proof

- For sums of independent random variables $X = X_1 + X_2$, it is convenient to use *characteristic functions*

  $\chi_X(\omega) \stackrel{\text{def}}{=} E[\exp(\mathrm{i}\cdot\omega\cdot X)]$ for which $\chi_X(\omega) = \chi_{X_1}(\omega)\cdot\chi_{X_2}(\omega)$.

- For characteristic functions $\chi(\omega, \mu, V, m_3)$, strong consistency takes the form:

$$\chi(\omega, \mu_1 + \mu_2, V_1 + V_2, m_{31} + m_{32}) =$$

$$\chi(\omega, \mu_1, V_1, m_{31}) \cdot \chi(\omega_2, \mu_2, V, m_{32}).$$

- This requirement becomes even simpler if we take logarithm of both sides: for $\ell \stackrel{\text{def}}{=} \ln(\chi)$:

$$\ell(\omega, \mu_1 + \mu_2, V_1 + V_2, m_{31} + m_{32}) =$$

$$\ell(\omega, \mu_1, V_1, m_{31}) + \ell(\omega_2, \mu_2, V, m_{32}).$$

48. Proof (cont-d)

- It is known that the only measurable functions with this additivity property are linear functions, so

  $\ell(\omega,\mu,V,m_3)=\mu\cdot\ell_1(\omega)+V\cdot\ell_2(\omega)+m_3\cdot\ell_3(\omega)$ for some $\ell_i(\omega)$.

- Let us now use the scale invariance requirement.
- When we replace x with $x'=\lambda\cdot x$, then

$$\chi_{X'}(\omega)=\chi_X(\lambda\cdot\omega).$$

- Thus re-scaled $\chi(\lambda\cdot\omega,\mu,V,m_3)$ should be equal to what we get from re-scaled moments: $\chi(\omega,\lambda\cdot\mu,\lambda^2\cdot V,\lambda^3\cdot m_3)$:

$$\chi(\lambda\cdot\omega,\mu,V,m_3)=\chi(\omega,\lambda\cdot\mu,\lambda^2\cdot V,\lambda^3\cdot m_3).$$

- Their logarithms should also be equal:

$$\ell(\lambda\cdot\omega,\mu,V,m_3)=\ell(\omega,\lambda\cdot\mu,\lambda^2\cdot V,\lambda^3\cdot m_3).$$

49. Proof (cont-d)

- Substituting the above linear expression for the function $\ell(\omega, \mu, V, m_3)$ into this equality, we conclude that

$$\mu \cdot \ell_1(\lambda \cdot \omega) + V \cdot \ell_2(\omega \cdot \omega) + m_3 \cdot \ell_3(\lambda \cdot \omega) =$$

$$\lambda \cdot \mu \cdot \ell_1(\omega) + \lambda^2 \cdot V \cdot \ell_2(\omega) + \lambda^3 \cdot m_3 \cdot \ell_3(\omega).$$

- This equality must hold for all possible triples (μ, V, m_3).
- Thus, the coefficient at μ, V, and m_3 on both sides must coincide.

50. Proof (final)

- By equating coefficients at μ, we conclude that

$$\ell_1(\lambda \cdot \omega) = \lambda \cdot \ell_1(\omega).$$

- In particular, for $\omega = 1$, we conclude that $\ell_1(\lambda) = \lambda \cdot \ell_1(1)$, i.e., that $\ell_1(\omega) = c_1 \cdot \omega$ for some constant c_1.
- By equating coefficients at V and m_3, we similarly get $\ell_2(\omega) = c_2 \cdot \omega^2$ and $\ell_3(\omega) = c_3 \cdot \omega^3$.
- Thus, $\ell(\omega, \mu, V, m_3) = c_1 \cdot \mu \cdot \omega + c_2 \cdot V \cdot \omega^2 + c_3 \cdot m_3 \cdot \omega^3$, and

$$\chi(\omega,\ u, V.m_3) = \exp(c_1 \cdot \mu \cdot \omega + c_2 \cdot V \cdot \omega^2 + c_3 \cdot m_3 \cdot \omega^3).$$

- However, the Fourier transform of the above expression is, in general, *not* an everywhere non-negative function.
- Thus, it cannot serve as a probability density function.

51. Comment

- If we only consider two moments, then the above proof leads to the characteristic function

$$\chi(\omega, \mu, V) = \exp(c_1 \cdot \mu \cdot \omega + c_2 \cdot V \cdot \omega^2).$$

- This characteristic function describes the Gaussian distribution.
- Thus, we have, in effect proven the following auxiliary result.

52. Auxiliary Definitions

- By a *2-selection*, we mean a measurable mapping $\rho(x, \mu, V)$ defined for all possible tuples.
- We say that a 2-selection is *strongly consistent* if $X_i \sim \rho(x, \mu_i, V_i)$ for independent X_i implies

$$X_1 + X_2 \sim \rho(x, \mu_1 + \mu_2, V_1 + V_2).$$

- We say that a 3-selection is *scale-invariant* if for every possible tuple (μ, V), for every $\lambda > 0$ and x':

$$\frac{1}{\lambda} \cdot \rho\left(\frac{x'}{\lambda}, \mu, V\right) = \rho(x', \lambda \cdot \mu, \lambda^2 \cdot V).$$

- **Proposition.** *Every strongly consistent and scale-invariant 2-selection assigns:*
  - *to each possible tuple* (μ, V),
  - *Gaussian distribution with mean* μ *and variance* V.

53. Acknowledgments

This work was supported in part:

- by the Chiang Mai University,
- by the National Science Foundation grants:
    - HRD-0734825 and HRD-1242122 (Cyber-ShARE Center of Excellence) and
    - DUE-0926721, and
- by an award from Prudential Foundation.