Invertible Additive Random Utility models

André de Palma (ENS Paris-Saclay, CREST, France)
Karim Kilani (CNAM, LISRA, France)

Workshop on Discrete choice models

Lausanne, April 2019

Please do not quote and do not circulate

Summary

DCM models associate to a vector of (a) deterministic components and of (b) random components of the utility a set of choice probabilities.

We show that one choice probability allows to uniquely recover :

(1) all the choice probabilities,

(2) the differences between any random component and a predefined one

(3) all random components consistent with one choice probability.

When the choice probabilities satisfy the IIA property, we can construct random components positively and negatively correlated.

We characterize all expected maximum utilities consistent with IIA.

Finally, we find all Generalized Extreme Value models satisfying IIA.

Additive Random Utility Models

Standard ARUM models :$\left\{v_i\right\}+\left\{\varepsilon_i\right\} \rightarrow P_{i,C}\left(\boldsymbol{v}\right);\boldsymbol{v} \equiv v_1,...,v_n$

$P_{i,C}\left(\boldsymbol{v}\right)=\Pr\left\{v_i+\varepsilon_i \geq v_k+\varepsilon_k, k \in C\right\}, \quad C=\left\{1,....n\right\}.$

$P_{i,A}\left(\boldsymbol{v}\right)$ for $A \subset C,\ i \in A$ is defined similarly.

$n-1$ integrations of the random components $\left\{\varepsilon_k\right\}$ leads $\mathrm{P}_{i,C}\left(\boldsymbol{v}\right)=$

$$\int_{-\infty}^{\infty}\int_{-\infty}^{v_i+x_i-v_1}...\left[\int_{-\infty}^{v_i+x_i-v_i}\quad\right]...\int_{-\infty}^{v_i+x_i-v_n} f_{\boldsymbol{\varepsilon}}\left(x_1...x_i...x_n\right)dx_n....\left[dx_i\right]....dx_1dx_i$$

See McFadden, 1977, MBA-L, 1985, A-D-T, 1992.

Different ARUM can be generated

$\{\varepsilon_i\}$ i.i.d. Gumbel $\rightarrow$ Logit

$\{\varepsilon_i\}$ GEV $\rightarrow$ GEV

$\{\varepsilon_i\}$ Normal distrib. $\rightarrow$ Probit

$P_{i,C}(\mathbf{v}) \rightarrow CS(\mathbf{v})$ by integration (Small) = indirect utility function $[$price (or $\mathbf{v}$)$]$

$\rightarrow$ Direct utility function [Quantity] by sustitution of the demand.

$*$ Shannon entropy direct utility $\rightarrow$ Logit (A-D-T, 1988)

$*$ Generalised entropy $\rightarrow$ All ARUMs and more (with complementarity)

(F-M-D-S 2018 & F-D-M, 2019) Trick: Start with the inverse model:

OLS estimations for market shares for MNL, any Nested, etc...

Standard inversion: Recover the deterministic part, **v**

Berry, 1994: Logit $P_{i,C}\left(\boldsymbol{v}\right)+\text{IIA}$, where $P_{i,C}\left(\boldsymbol{v}\right)$ are the observed market share

$$\frac{P_{i,C}\left(\boldsymbol{v}\right)}{P_{n,C}\left(\boldsymbol{v}\right)}=\exp\left(v_i-v_n\right), i, n\in C, \text{ or}$$

$$v_i-v_n=\log\left(\frac{P_{1,C}(\boldsymbol{v})}{P_{n,C}(\boldsymbol{v})}\right),$$

i.e. all derministic components $\boldsymbol{v}$ (w.r.t. to component n)

can be recovered numerically (using a variant of Brower's Theorem).

This recovery result is known as the "Berry Inversion"; it holds for any ARUM.

[Early results: Existence of a Random Utility Model, given $P_{1,A},\ldots,P_{n,A}$?]

- DEFINITION A system of choice probabilities, $\{P_{i,A}\}$, is stochastically rationalizable if it can be generated by a RUM (Random Utility Model): Block & Marschak, 1960, McFadden & Richter, 1990.

- Block and Marschak, derived conditions ("B&M polynomials") on $\{P_{i,A}\}$ that guarantee the system $\{P_{i,A}\}$ is stochastically rationalizable.

Assumption A1

Let: δ =: $(\varepsilon_1\text{-}\varepsilon_n,\ldots, \varepsilon_1\text{-}\varepsilon_{n-1})$.

A1: The vector δ admits a strictly positive and continuous PDF, $f_\delta(.)$, with respect to the Lebesgue measure of $\mathbf{R}^{n-1}$.

Note: No finite expectations are required. for ε_i.

Include for ε_i Gauchy, Gumbel Laplace, Logistic, Normal, ….

but ε_i can be defined on a semi-interval as for:

Chi, Exponential, Fréchet, Gamma or Lognormal.

Inversion (1) : One choice probability is enough?

Given $P_{n,C}(\mathbf{v})$ with ARUM and A1:

can we infer all other choice probabilities

$$P_{n,C}(\mathbf{v}) \rightarrow P_{1,C}(\mathbf{v}),\ldots,P_{n-1,C}(\mathbf{v})?$$

Inversion (2) : Find the differences of random components?

Given $P_{n,C}\left(\boldsymbol{v}\right)$ with ARUM and A1:

can we infer the differences of random components,

which generate this choice probability?

$$P_{n,C}\left(\boldsymbol{v}\right) \to \left\{\delta_1,\ldots,\delta_{n-1}\right\}?$$

Inversion (3) : Find the error terms?

Given $P_{n,C}(\boldsymbol{v})$, with ARUM and A1:

can we infer all random components,

which generates this choice probability?

$$P_{n,C}(\boldsymbol{v}) \rightarrow \{\varepsilon_i\}?$$

[Yellott and Strauss]

If $\{\varepsilon_i\}$ are i.i.d. with ARUM and IIIA(Yellott, 1977)

shows $\{\varepsilon_i\}$ are necessarily Gumbelly distributed.

But could $\{\varepsilon_i\}$ be correlated ?

Strauss (1979) has an example, with positive correlations.

But coul $\boldsymbol{\varepsilon} \equiv \{\varepsilon_1, \ldots, \varepsilon_n\}$ also be negative correlated?

Is the i.i.d. hypothesis needed?

[Hint! Only differences matter!]

$$P_{1,C}(\boldsymbol{v}) = \Pr\left\{v_1 + \varepsilon_1 \geq v_k + \varepsilon_k, k \in C\right\}; \boldsymbol{v} = \left\{v_1, \ldots, v_n\right\}$$

$$P_{1,C}(\boldsymbol{v}) = \Pr\left\{v_1 + \delta_1 > v_2 + \delta_2, \ldots, v_1 + \delta_1 > v_n\right\}$$

Recall: $\delta_i \equiv \varepsilon_i - \varepsilon_n, i = 1, \ldots, n-1.$

So it suffices to find:

$\boldsymbol{\delta} \equiv \left\{\delta_1, \ldots, \delta_{n-1}\right\}$; but we need to backup $\boldsymbol{\varepsilon} = \varepsilon_1, \ldots, \varepsilon_n$ consistent with $\boldsymbol{\delta}$ (presumably $\boldsymbol{\varepsilon}$ is not unique).

Handling (1) Recover the $P_{1,C},\ldots.P_{n-1,C}$ from $P_{n,C}$?

Recall: δ are uniquely determined by one choice probability, e.g. $P_{n,A}$.

Since δ generate $P_{1,C},\ldots.P_{n-1,C}$,then $P_{n,C}$ is enough to recover all others

LEMMA

Under A1, all RUM choice probabilities

can be derived from $P_{n,C}$:

$$P_{i,C}(\boldsymbol{v}) = -\int_{v_n}^{\infty} \frac{\partial P_{n,C}(v_1,\cdots,v_{n-1},x)}{\partial v_i} dx,\ \boldsymbol{v} \in R^n, i = 1,\ldots,n-1.$$

$$P_{i,C}(\boldsymbol{v}) > 0.$$

Handling (2) Recover the δ from an unique $P_{n,C}$

LEMMA

Under A1, and ARUM and given $P_{n,C}(\boldsymbol{v})$

$F_\delta(z_1,\dots,z_{n-1}) = P_{n,C}(-z_1,-z_2,\dots - z_{n-1},0); \boldsymbol{z} \in \mathbb{R}^{n-1},$

$z_1 = x_1 - x_n,\dots,z_{n-1} = x_{n-1} - x_n,\ C = \{1,\dots.n\}.$

Note : We only use the functional form of the choice probabilities $P_{n,C}(\boldsymbol{v})$.

Standard conditions on $P_{n,C}(\boldsymbol{v})$ so that $F_\delta(z_1,\dots,z_{n-1})$ is a CDF.

f(δ) can be recovered by deriving the above $F_\delta(z_1,\dots,z_{n-1})$ n times.

Handling (3) Recover all the ε from a unique $P_{n,C}$

THEOREM

Under A1, and ARUM and given $P_{i,n}(\boldsymbol{v})$, the $\boldsymbol{\varepsilon}$ has a CDF, satisfy:

$$\frac{\partial F_{\boldsymbol{\varepsilon}}(x)}{\partial x_n} = P_{n,C}\left(-x_1,\ldots,-x_n\right) g_{\boldsymbol{z}}(x_n), x \in \mathbb{R}^n,$$

where $g_{\boldsymbol{z}}(.)$ is any PDF parametrized by $\boldsymbol{z}=\left(x_1 - x_n, x_2 - x_n, \ldots..x_{n-1} - x_n\right)$.

Note that the error compoments are not unique!

Note : In ADP-1992, $g_{\boldsymbol{z}}(.) \rightarrow g(.)$, so they do not generated all $\boldsymbol{\varepsilon}$.

[It amounts to compute $\boldsymbol{\delta}$ (associated to $P_{n,C}$) and freely choose ε_{n}.]

(Difference of) of random terms in RUM with IIA

Assume more structure for $\{P_{i,A}\}$?

Work plan: assume that $\{P_{i,A}\}$ satisfy the IIA properties.

- We show that δ is uniquely determined.
- We investigate all classes of ε consistent with all ARUM probabilities $\{P_{i,A}\}$ satisfying IIA.

Handling (1) Given $P_{n,C}(v)$

LEMMA

One Logit choice probabilities allows to recover all other choice probabilities.

i.e. If $P_{n,C}(\boldsymbol{v}) = \dfrac{e^{v_1}}{\sum_{k=1}^{n} e^{v_k}}$ is given,

then all other choice probabilities can be recovered.

Hint: $\dfrac{\partial P_{i,C}(v)}{\partial v_n} = \dfrac{\partial P_{n,C}(v)}{\partial v_i}.$

Handling (2) Recover all ε under the IIA

THEOREM:

Under A1, an ARUM satisfies IIA for any $\boldsymbol{v} \in R^n$ iif $\boldsymbol{\varepsilon}$ has a CDF such that its derivative satisfies:

$$\frac{\partial F_{\boldsymbol{\varepsilon}}(x_1,..x_n)}{\partial x_n} = \frac{e^{-x_n}}{\sum_{k=1}^{n} e^{-x_k}} g_{\mathbf{z}}(x_n), x \in \mathbb{R}^n,$$

where $\mathrm{g}_{\mathbf{z}}(.)$ is any PDF parametrized by $\mathbf{z}=\left(x_1 - x_n, x_2 - x_n, ..., x_{n-1} - x_n\right)$.

Note : we do not assume that $\boldsymbol{\varepsilon}$ are defined on R (ε_i could be defined on semi-intervals).
We just assume IIA.

handling (3) Recover the δ under IIA

PROPOSITION

Under A1, an ARUM system satisfies IIA for any $\mathbf{v} \in \mathrm{R}^n$ iif $\boldsymbol{\delta}$ has a CDF given by:

$$F_\delta(\boldsymbol{z}) = \left[1 + \sum_{k=1}^{n-1} e^{-(z_k - \alpha_k)/\sigma}\right]^{-1}, \ \boldsymbol{z} \in \mathbb{R}^{n-1},$$

where $\alpha_1, \ldots, \alpha_{n-1}$ are location parameters and σ is a positive scale parameter.

Note: this is the multivariate logistic.

Class of models to generate the ε

Consider the following class of **ε** and the function Φ(x).

$$F_{\boldsymbol{\varepsilon}}(\boldsymbol{x}) = \Phi\left(-\ln\sum_{k=1}^{n} e^{-x_k}\right), \ \boldsymbol{x} \in \mathbb{R}^n,$$

$$\text{with } \Phi''(x) + \Phi'(x) \geq 0, \ x \in (a, \infty).$$

Note: These conditions guarantee that $F_{\epsilon}(\boldsymbol{x})$ is a CDF.

[Examples (1/2)]

If $\Phi\left(x\right)=\exp\left(-e^{-x/\theta}\right),\ x\in R,\ \theta\geq 1:$

$$F_{\boldsymbol{\varepsilon}}\left(\boldsymbol{x}\right)=\exp\left[-\left(\sum_{k=1}^{n}e^{-x_k}\right)^{1/\theta}\right],\ x\in R^n;\ \text{Correlation } \rho\text{=1-}\theta^{-2}\geq 0.$$

If $\Phi\left(x\right)=\left(1+e^{-x/\theta}\right),x\in R,\theta\geq 1:$

$$F_{\boldsymbol{\varepsilon}}\left(\boldsymbol{x}\right)=\left[1+\left(\sum_{k=1}^{n}e^{-x_k}\right)^{1/\theta}\right]^{-1},\ x\in R^n;\ \text{Correlation } \rho\text{=}1-\theta^{-2}/2>0.$$

Examples (2/2)

If $\Phi(x) = 1 - e^{(-x/\theta)},\ x \in [0, \infty),\ \theta \geq 1.$

$F_{\boldsymbol{\varepsilon}}(\boldsymbol{x}) = 1 - Y^{1/\theta},\ Y = \sum_{k=1}^{n} e^{-x_k} < 1$ and $F_{\boldsymbol{\varepsilon}}(\boldsymbol{x}) = 0$ otherwise:

$\rho = 1 - \left(\pi^2 / 6\right) / \theta^2$ so correlation is negative if $\theta > \pi / \sqrt{6}$.

Cauchy works as well but Normal $\Phi(x)$ does not works!

[Yellott (1977) revisited]

THEOREM

Let A1 hold and assume that $\boldsymbol{\varepsilon}$ has independent components.

An ARUM satisfies IIA for any $\boldsymbol{v} \in R^n$ iif the components are

i.d. Gumbel: $F_{\epsilon_k}(x) = \exp\left[-e^{-(x-\alpha_k)/\sigma}\right], x \in R, k \in C,$

Note:

We do not need R as the support for ε_k;

we do not need identical distributions, for $\varepsilon_1,\ldots,\varepsilon_n$

as required by Yellott.

IIA expected maximum utility

Definitions and notations

A2 : The vector of $\boldsymbol{\varepsilon}$ has finite expectations.

The expected maximum utility is given by:

$$\Omega(\boldsymbol{v}) \equiv E\left[\max_{k=1,\ldots,n}\left(v_k + \varepsilon_k\right)\right], \boldsymbol{v} \in \mathbb{R}^n.$$

Expected utility under IIA

LEMMA (Williams-Daly-Zachary in McFadden, 1981)

Let A1 and A2 hold. For the ARUM, with the associated $\Omega(v)$, the choice probabilities are given by:

$$\frac{\partial \Omega(\boldsymbol{v})}{\partial v_i} = P_{i,C}\left(\boldsymbol{v}\right),\ \boldsymbol{v} \in \mathbb{R}^n,\ i = 1,\dots,n.$$

Expected utility under IIA

Assume ε_k has a mean normalized to zero, k=1,..,n.

THEOREM

Let A1-A2 hold. For the ARUM, the choice probabilities satisfy IIA for any $\boldsymbol{v}$ iif the expected maximum utility can be written as:

$$\Omega(\boldsymbol{v}) = \ln\left(\sum_{k=1}^{n} e^{v_k}\right), \ \boldsymbol{v} \in R^n.$$

GEV families satisfying IIA

GEV with IIA

THEOREM

A GEV model satisfies IIA iif it is a multivariate Gumbel distribution with CDF given by:

$$F_\epsilon(\boldsymbol{x}) = \exp\left[-\left(\sum_{k=1}^{n} e^{-x_k}\right)^{1/\theta}\right],\ \theta \geq 1, \boldsymbol{x} \in \mathbb{R}^n.$$

Note: $F_\epsilon(\boldsymbol{x})$ generates the Logit (as seen above).

Conclusions – more to go....

These results (potentially) be can be applied to:

- ❖ the Logit kernel model, and any probability Kernel model,
- ❖ the CES demand functions,
- ❖ more general $F_\varepsilon(x)$.

I;e. the "logit" Logsum formula : $F_{\boldsymbol{\varepsilon}}(\boldsymbol{x}) = \Phi\left(-\ln \sum_{k=1}^{n} e^{-x_k}\right), \ \boldsymbol{x} \in \mathbb{R}^n,$

with $\Phi''(x) + \Phi'(x) \geq 0, x \in (a, \infty)$

can be generalized as : $F_{\boldsymbol{\varepsilon}}(\boldsymbol{x}) = \Phi\left(-\Omega(-\boldsymbol{v})\right),$

where $\Omega(\boldsymbol{v})$ is the expected maximum utility.