

Inferring cognitive heterogeneity from aggregate choices

Article (Supplemental Material)

Dardanoni, Valentino, Manzini, Paola, Mariotti, Marco and Tyson, Christopher J (2020) Inferring cognitive heterogeneity from aggregate choices. *Econometrica*, 88 (3). pp. 1269-1296. ISSN 0012-9682

This version is available from Sussex Research Online: <http://sro.sussex.ac.uk/id/eprint/89905/>

This document is made available in accordance with publisher policies and may differ from the published version or from the version of record. If you wish to cite this item you are advised to consult the publisher's version. Please see the URL above for details on accessing the published version.

Copyright and reuse:

Sussex Research Online is a digital repository of the research output of the University.

Copyright and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable, the material made available in SRO has been checked for eligibility before being made available.

Copies of full text items generally can be reproduced, displayed or performed and given to third parties in any format or medium for personal research or study, educational, or not-for-profit purposes without prior permission or charge, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

SUPPLEMENT TO “INFERRING COGNITIVE HETEROGENEITY FROM AGGREGATE CHOICES”

VALENTINO DARDANONI

Dipartimento SEAS, Università degli Studi di Palermo

PAOLA MANZINI

Department of Economics, University of Sussex, and IZA

MARCO MARIOTTI

School of Economics and Finance, Queen Mary University of London

CHRISTOPHER J. TYSON

School of Economics and Finance, Queen Mary University of London

February 7, 2020

APPENDIX B: A PRIMER ON TENSOR DECOMPOSITIONS

A *tensor* is a multidimensional array. The *order* of a tensor is the number of dimensions it possesses; e.g., a vector is a tensor of order 1, and a matrix is a tensor of order 2. The *dimensions* of a tensor are the numbers of components along each side of the array. A tensor \mathbf{X} of order I with dimensions $n_1 \times \cdots \times n_I$ has typical element $x_{k_1 \dots k_I}$, where $1 \leq k_i \leq n_i$ for each $i = 1, \dots, I$.¹

For example, consider the tensor of order 3 and dimensions $4 \times 3 \times 2$ given by

$$\mathbf{X} = \left[\left[\begin{array}{c|ccc|ccc} k_3 = 1 & k_2 = 1 & k_2 = 2 & k_2 = 3 & k_3 = 2 & k_2 = 1 & k_2 = 2 & k_2 = 3 \\ k_1 = 1 & x_{111} & x_{121} & x_{131} & k_1 = 1 & x_{112} & x_{122} & x_{132} \\ k_1 = 2 & x_{211} & x_{221} & x_{231} & k_1 = 2 & x_{212} & x_{222} & x_{232} \\ k_1 = 3 & x_{311} & x_{321} & x_{331} & k_1 = 3 & x_{312} & x_{322} & x_{332} \\ k_1 = 4 & x_{411} & x_{421} & x_{431} & k_1 = 4 & x_{412} & x_{422} & x_{432} \end{array} \right] \right]. \quad (1)$$

This tensor can be viewed variously as a list of two 4×3 matrices (at left and at right), as a list of six 4-vectors (in the columns), or as a list of 24 individual entries. Note that in Equation 1 the values of the indices k_i are shown in the margins, for clarity, but these are usually suppressed in line with vector and matrix notation.

A *decomposition* of a tensor of a given order is a representation as the outer product of lower-order tensors.² For instance, any tensor of the form

$$\mathbf{X} = \left[\left[\begin{array}{ccc|ccc} y_{11}w_1 & y_{12}w_1 & y_{13}w_1 & y_{11}w_2 & y_{12}w_2 & y_{13}w_2 \\ y_{21}w_1 & y_{22}w_1 & y_{23}w_1 & y_{21}w_2 & y_{22}w_2 & y_{23}w_2 \\ y_{31}w_1 & y_{32}w_1 & y_{33}w_1 & y_{31}w_2 & y_{32}w_2 & y_{33}w_2 \\ y_{41}w_1 & y_{42}w_1 & y_{43}w_1 & y_{41}w_2 & y_{42}w_2 & y_{43}w_2 \end{array} \right] \right]$$

Valentino Dardanoni: valentino.dardanoni@unipa.it

Paola Manzini: p.manzini@sussex.ac.uk

Marco Mariotti: m.mariotti@qmul.ac.uk

Christopher J. Tyson: c.j.tyson@qmul.ac.uk

¹Tensors are widely used in science and engineering; e.g., in the fields of chemometrics, electrodynamics, general relativity, and signal processing. For a textbook treatment of the subject, see [Danielson \(2003\)](#).

²Recall that the outer product of a k_1 -vector \mathbf{u} and a k_2 -vector \mathbf{v} is the matrix with dimensions $k_1 \times k_2$ given by $\mathbf{u} \otimes \mathbf{v} = \mathbf{u}\mathbf{v}^\top$ (where \mathbf{v}^\top is the transpose of \mathbf{v}). Each further outer product operation then adds another dimension to the resulting array.

can be decomposed into the outer product of a 4×3 matrix and a 2-vector as

$$\mathbf{X} = \underbrace{\begin{bmatrix} y_{11} & y_{12} & y_{13} \\ y_{21} & y_{22} & y_{23} \\ y_{31} & y_{32} & y_{33} \\ y_{41} & y_{42} & y_{43} \end{bmatrix}}_{\mathbf{Y}} \otimes \underbrace{\begin{bmatrix} w_1 \\ w_2 \end{bmatrix}}_{\mathbf{w}},$$

since here each $x_{k_1 k_2 k_3} = y_{k_1 k_2} w_{k_3}$. Moreover, if

$$\mathbf{Y} = \begin{bmatrix} u_1 v_1 & u_1 v_2 & u_1 v_3 \\ u_2 v_1 & u_2 v_2 & u_2 v_3 \\ u_3 v_1 & u_3 v_2 & u_3 v_3 \\ u_4 v_1 & u_4 v_2 & u_4 v_3 \end{bmatrix} = \underbrace{\begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{bmatrix}}_{\mathbf{u}} \otimes \underbrace{\begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}}_{\mathbf{v}},$$

then we have the further decomposition $\mathbf{X} = \mathbf{u} \otimes \mathbf{v} \otimes \mathbf{w}$. A tensor is said to be of *rank 1* if it can be decomposed into an outer product of vectors in this way. More generally, the rank of any tensor \mathbf{X} is the minimum number of rank-1 tensors that sum to \mathbf{X} .

Intuitively, decomposition of a tensor separates out independent influences on its entries. As an application of the concept (for which we make no claim of originality), imagine a dinner party at which numerous guests are all talking at the same time. The verbal barrage can be better understood if it is separated into a number of distinct, albeit simultaneous conversations. Each is characterized by particular words or topics, and they proceed for the most part independently of each other, even while interacting in the auditory space.

In the paper we study a joint choice share tensor \mathbf{S} of order I (the number of choice ‘‘occasions’’) and dimensions $n \times \dots \times n$, with typical entry

$$s_{k_1 \dots k_I} = \bar{p}(k_1 \dots k_I) = \sum_{\gamma=1}^n \bar{\pi}(\gamma) p_{\gamma}(k_1 \dots k_I).$$

Here $k_1 \dots k_I$ are the choices on the I occasions, γ is the cognitive type of the agent, and $p_{\gamma}(k_1 \dots k_I)$ is the type-conditional probability of the joint choice share observation. In the consideration capacity model, the latter probability can be expressed as

$$p_{\gamma}(k_1 \dots k_I) = \prod_{i=1}^I \underbrace{\sum_{r=1}^{n-\gamma+1} \frac{\binom{n-r}{\gamma-1}}{\binom{n}{\gamma}} \sum_{h: \varphi_h(k_i)=r} \tau_{ih}}_{\mathbf{1}_{k_i}^{\top} [\mathbf{B}_i \mathbf{C}] \mathbf{1}_{\gamma}}. \quad (2)$$

Defining the matrices $\mathbf{Z}_1 = [\mathbf{B}_1 \mathbf{C}] \mathbf{D}(\bar{\pi})$ and $\mathbf{Z}_i = \mathbf{B}_i \mathbf{C}$ for each $i = 2, \dots, I$, we have that the joint choice shares make up the tensor $\mathbf{S} = \sum_{\gamma=1}^n \otimes_{i=1}^I \mathbf{Z}_i \mathbf{1}_{\gamma}$.³ In other words, \mathbf{S} is the sum of n rank-1 tensors corresponding to the distinguishable cognitive types.

If there are at least three choice occasions, then we can use the fundamental result of [Kruskal \(1977\)](#), as adapted by [Rhodes \(2010\)](#), to show that the decomposition of \mathbf{S} into \mathbf{Z}_i matrices is effectively unique.

Lemma 1 (Kruskal; Rhodes). *Given any triad $\langle \mathbf{Z}_1, \mathbf{Z}_2, \mathbf{Z}_3 \rangle$ of invertible $n \times n$ matrices, the tensor $\mathbf{T} = \sum_{\gamma=1}^n [\mathbf{Z}_1 \mathbf{1}_{\gamma} \otimes \mathbf{Z}_2 \mathbf{1}_{\gamma} \otimes \mathbf{Z}_3 \mathbf{1}_{\gamma}]$ uniquely determines each \mathbf{Z}_i up to column rescaling and permutation.*

We then demonstrate that uniqueness of the decomposition guarantees generic identification of both the $\bar{\pi} = \langle \bar{\pi}(\gamma) \rangle_{\gamma=1}^n$ vector and the \mathbf{B}_i matrices; since \mathbf{C} is constant, known, and invertible. Hence the joint choice shares in \mathbf{S} yield full knowledge of the cognitive distribution (over types $\gamma < n$) as well as partial knowledge of the taste distributions τ_i .

³Recall that in our notation $\mathbf{1}_{\ell}$ is the unit vector for component ℓ and $\mathbf{D}(\mathbf{v})$ the diagonal matrix with entries drawn from the vector \mathbf{v} . Thus $[\mathbf{B}_i \mathbf{C}] \mathbf{1}_{\gamma}$ is the γ th column of the matrix $\mathbf{B}_i \mathbf{C}$, and $\mathbf{1}_{k_i}^{\top} [\mathbf{B}_i \mathbf{C}] \mathbf{1}_{\gamma}$ is the k_i th entry in this column. Similarly, $[\mathbf{B}_1 \mathbf{C}] \mathbf{D}(\bar{\pi})$ is the matrix $\mathbf{B}_1 \mathbf{C}$ with each column weighted by the corresponding scalar $\bar{\pi}(\gamma)$.

It is apparent that nothing about our approach to identification depends in any essential way on the details of the consideration capacity model. The capacity γ can be replaced by an arbitrary cognitive type $\theta \in \{\theta_1, \dots, \theta_n\}$, distributed according to $\boldsymbol{\pi} = \langle \pi(\theta_j) \rangle_{j=1}^n$; with \mathbf{C} likewise replaced by an $n \times n$ matrix \mathbf{A} (still constant and known) whose typical entry a_{kj} is the probability that the k th best alternative is chosen by an agent of cognitive type θ_j . Equation 2 then becomes

$$p_{\theta_j}(k_1 \cdots k_I) = \prod_{i=1}^I \underbrace{\sum_{r=1}^n a_{rj}}_{\mathbf{1}_{k_i}^\top [\mathbf{B}_i \mathbf{A}] \mathbf{1}_j} \sum_{h: \varphi_h(k_i)=r} \tau_{ih}, \quad (3)$$

and as before we can define $\mathbf{Z}_1 = [\mathbf{B}_1 \mathbf{A}] \mathbf{D}(\boldsymbol{\pi})$ and $\mathbf{Z}_i = \mathbf{B}_i \mathbf{A}$ for each $i = 2, \dots, I$. As long as the matrix \mathbf{A} is invertible, the decomposition $\mathbf{S} = \sum_{j=1}^n \otimes_{i=1}^I \mathbf{Z}_i \mathbf{1}_j$ will be effectively unique, and both $\boldsymbol{\pi}$ and the \mathbf{B}_i matrices will be generically identified.⁴

To illustrate the broad scope of our methodology, consider a primitive model of satisficing in which with probability $\pi(\theta_j)$ the agent randomizes uniformly among the j best alternatives available. For $n = 3$ we then have an invertible (and indeed upper triangular) matrix

$$\mathbf{A} = \begin{bmatrix} 1 & 1/2 & 1/3 \\ 0 & 1/2 & 1/3 \\ 0 & 0 & 1/3 \end{bmatrix},$$

and can compute each product

$$\mathbf{B}_i \mathbf{A} = \begin{bmatrix} \tau_{i1} + \tau_{i2} & [\tau_{i1} + \tau_{i2} + \tau_{i3} + \tau_{i4}]/2 & 1/3 \\ \tau_{i3} + \tau_{i5} & [\tau_{i1} + \tau_{i3} + \tau_{i5} + \tau_{i6}]/2 & 1/3 \\ \tau_{i4} + \tau_{i6} & [\tau_{i2} + \tau_{i4} + \tau_{i5} + \tau_{i6}]/2 & 1/3 \end{bmatrix}.$$

With $I = 3$, the resulting joint choice share tensor decomposition

$$\mathbf{S} = \pi(1) \otimes_{i=1}^3 \begin{bmatrix} \tau_{i1} + \tau_{i2} \\ \tau_{i3} + \tau_{i5} \\ \tau_{i4} + \tau_{i6} \end{bmatrix} + \pi(2) \otimes_{i=1}^3 \begin{bmatrix} [\tau_{i1} + \tau_{i2} + \tau_{i3} + \tau_{i4}]/2 \\ [\tau_{i1} + \tau_{i3} + \tau_{i5} + \tau_{i6}]/2 \\ [\tau_{i2} + \tau_{i4} + \tau_{i5} + \tau_{i6}]/2 \end{bmatrix} + \pi(3) \otimes_{i=1}^3 \begin{bmatrix} 1/3 \\ 1/3 \\ 1/3 \end{bmatrix}$$

is effectively unique, and generic identification holds both for the distribution $\boldsymbol{\pi}$ of satisficing types and for each matrix \mathbf{B}_i of rank-position probabilities.

REFERENCES

- DANIELSON, D. A. (2003): *Vectors and Tensors in Engineering and Physics*, Cambridge, MA: Westview Press.
- KRUSKAL, J. B. (1977): “Three-way arrays: Rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics,” *Linear Algebra and its Applications*, 18, 95–138.
- RHODES, J. A. (2010): “A concise proof of Kruskal’s theorem on tensor decomposition,” *Linear Algebra and its Applications*, 432, 1818–1824.

⁴For simplicity we have continued to assume that there are n distinguishable cognitive types, and therefore the matrix \mathbf{A} is square. In the absence of this assumption, a version of Kruskal’s result more general than Lemma 1 would be needed to show uniqueness of the decomposition.