Econometrica Supplementary Material

SUPPLEMENT TO "INFERRING COGNITIVE HETEROGENEITY FROM AGGREGATE CHOICES"

(*Econometrica*, Vol. 88, No. 3, May 2020, 1269–1296)

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

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; for example, 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 $\boldsymbol{X}$ of order I with dimensions $n_1 \times \cdots \times n_I$ has typical element $x_{k_1 \cdots k_I}$, where $1 \leq k_i \leq n_i$ for each $i = 1, \ldots, I$.[1]

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

$$\boldsymbol{X} = \left[\!\!\left[\begin{array}{c|ccc||c|ccc} k_3 = 1 & k_2 = 1 & k_2 = 2 & k_2 = 3 & k_3 = 2 & k_2 = 1 & k_2 = 2 & k_2 = 3 \\ \hline k_1 = 1 & x_{111} & x_{121} & x_{131} & k_1 = 1 & x_{112} & x_{122} & x_{132} \\ k_1 = 2 & x_{211} & x_{221} & x_{231} & k_1 = 2 & x_{212} & x_{222} & x_{232} \\ k_1 = 3 & x_{311} & x_{321} & x_{331} & k_1 = 3 & x_{312} & x_{322} & x_{332} \\ k_1 = 4 & x_{411} & x_{421} & x_{431} & k_1 = 4 & x_{412} & x_{422} & x_{432} \end{array} \right]\!\!\right]. \tag{S1}$$

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 (S1) 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.

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

[1] Tensors 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).

 https://doi.org/10.3982/ECTA16382

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

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

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

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

since here each $x_{k_1k_2k_3} = y_{k_1k_2}w_{k_3}$. Moreover, if

$$\boldsymbol{Y} = \begin{bmatrix} u_1v_1 & u_1v_2 & u_1v_3 \\ u_2v_1 & u_2v_2 & u_2v_3 \\ u_3v_1 & u_3v_2 & u_3v_3 \\ u_4v_1 & u_4v_2 & u_4v_3 \end{bmatrix} = \underbrace{\begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{bmatrix}}_{\boldsymbol{u}} \otimes \underbrace{\begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}}_{\boldsymbol{v}},$$

then we have the further decomposition $\boldsymbol{X} = \boldsymbol{u} \otimes \boldsymbol{v} \otimes \boldsymbol{w}$. 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 $\boldsymbol{X}$ is the minimum number of rank-1 tensors that sum to $\boldsymbol{X}$.

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 $\boldsymbol{S}$ of order I (the number of choice "occasions") and dimensions $n \times \cdots \times n$, with typical entry

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

Here, $k_1 \cdots k_I$ are the choices on the I occasions, γ is the cognitive type of the agent, and $p_\gamma(k_1 \cdots k_I)$ is the type-conditional probability of the joint choice share observation. In

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

the consideration capacity model, the latter probability can be expressed as

$$p_\gamma(k_1\cdots k_I)=\prod_{i=1}^{I}\underbrace{\sum_{r=1}^{n-\gamma+1}\frac{\binom{n-r}{\gamma-1}}{\binom{n}{\gamma}}\sum_{h:\varphi_h(k_i)=r}\tau_{ih}}_{\mathbf{1}_{k_i}^\top[\boldsymbol{B}_i\boldsymbol{C}]\mathbf{1}_\gamma}. \tag{S2}$$

Defining the matrices $\boldsymbol{Z}_1=[\boldsymbol{B}_1\boldsymbol{C}]\boldsymbol{D}(\overline{\boldsymbol{\pi}})$ and $\boldsymbol{Z}_i=\boldsymbol{B}_i\boldsymbol{C}$ for each $i=2,\dots,I$, we have that the joint choice shares make up the tensor $\boldsymbol{S}=\sum_{\gamma=1}^{n}\bigotimes_{i=1}^{I}\boldsymbol{Z}_i\mathbf{1}_\gamma$.[3] In other words, $\boldsymbol{S}$ 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 $\boldsymbol{S}$ into $\boldsymbol{Z}_i$ 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 $\boldsymbol{T}=\sum_{\gamma=1}^{n}[\mathbf{Z}_1\mathbf{1}_\gamma\otimes\mathbf{Z}_2\mathbf{1}_\gamma\otimes\mathbf{Z}_3\mathbf{1}_\gamma]$ 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 $\overline{\boldsymbol{\pi}}=\langle\overline{\pi}(\gamma)\rangle_{\gamma=1}^{n}$ vector and the $\boldsymbol{B}_i$ matrices; since $\boldsymbol{C}$ is constant, known, and invertible. Hence the joint choice shares in $\boldsymbol{S}$ yield full knowledge of the cognitive distribution (over types $\gamma<n$) as well as partial knowledge of the taste distributions $\boldsymbol{\tau}_i$.

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 $\boldsymbol{C}$ likewise replaced by a $n\times n$ matrix $\boldsymbol{A}$ (still constant and known) whose typical entry a_{kj} is the probability that the kth best alternative is chosen by an agent of cognitive type θ_j. Equation (S2) then becomes

$$p_{\theta_j}(k_1\cdots k_I)=\prod_{i=1}^{I}\underbrace{\sum_{r=1}^{n}a_{rj}\sum_{h:\varphi_h(k_i)=r}\tau_{ih}}_{\mathbf{1}_{k_i}^\top[\boldsymbol{B}_i\boldsymbol{A}]\mathbf{1}_j}, \tag{S3}$$

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

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

[4] For simplicity, we have continued to assume that there are n distinguishable cognitive types and, therefore, the matrix $\boldsymbol{A}$ 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.

alternatives available. For $n = 3$, we then have an invertible (and indeed upper triangular) matrix

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

and can compute each product

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

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

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

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

REFERENCES

DANIELSON, D. A. (2003): *Vectors and Tensors in Engineering and Physics*. Cambridge, MA: Westview Press. [1]

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. [3]

RHODES, J. A. (2010): "A Concise Proof of Kruskal's Theorem on Tensor Decomposition," *Linear Algebra and its Applications*, 432, 1818–1824. [3]

Co-editor Itzhak Gilboa handled this manuscript.

Manuscript received 25 May, 2018; final version accepted 31 January, 2020; available online 5 February, 2020.